



Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Estatística

# **Uma classe de distribuições log-simétricas discretas**

por

**Leonardo de Sousa Paiva**

Brasília, 27 de fevereiro de 2020

# **Uma classe de distribuições log-simétricas discretas**

**por**

**Leonardo de Sousa Paiva**

Dissertação apresentada ao Departamento de Estatística da Universidade de Brasília, como requisito parcial para obtenção do título de Mestre em Estatística.

Orientador: Prof. Dr. Helton Saulo Bezerra dos Santos

Coorientador: Prof. Dr. Roberto Vila Gabriel

Brasília, 27 de fevereiro de 2020

Dissertação apresentada ao Programa de Pós-Graduação em Estatística do Departamento de Estatística da Universidade de Brasília como parte dos requisitos para a obtenção do grau de Mestre em Estatística.

Banca:

Prof. Dr. Helton Saulo Bezerra dos Santos  
Orientador, Departamento de Estatística/UnB

Prof. Dr. Roberto Vila Gabriel  
Coorientador, Departamento de Estatística/UnB

Prof. Dr. Eduardo Yoshio Nakano  
Departamento de Estatística/UnB

Prof. Dr. Rodrigo Nobre Fernandez  
Departamento de Economia/UFPel

---

---

*Aprender é a única coisa de que a mente nunca se cansa, nunca tem medo e nunca se arrepende.*

(Leonardo da Vinci)

---

---

*A Deus.*

*A minha família, em especial à minha avó, Eunice,  
e às memórias do meu avô José Primo e do meu irmão Ricardo Paiva, que viveram seus  
tempos com dignidade.*

*Aos bons amigos.*

---

---



# Agradecimentos

Agradeço primeiramente a Deus por ser meu guia e por todas as oportunidades.

Aos professores do PPGEST/UnB, em especial aos professores Helton Saulo (orientador) e Roberto Vila (coorientador) pelo apoio e paciência.

Aos professores Eduardo Nakano (UnB) e Rodrigo Fernandez (UFPel) por aceitarem participar da avaliação deste trabalho e, assim, deixarem suas contribuições.

Aos servidores da UnB, em especial aos da secretaria do curso de Estatística pela prestimosidade.

À Universidade Federal do Piauí (UFPI) pela concessão do afastamento integral para que eu pudesse cursar o mestrado.

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

---

---

# Resumo

O uso de distribuições de probabilidade contínuas tem sido difundido em problemas com natureza puramente discreta. Em geral, essas distribuições não são apropriadas nesse cenário. Neste trabalho, é introduzida uma classe de distribuições de probabilidade discretas baseada na família de distribuições log-simétricas contínuas, incluindo como casos especiais as seguintes distribuições discretas: log-normal, log- $t$ -Student, log-exponencial-potência, log-normal-contaminada, Birnbaum-Saunders e Birnbaum-Saunders- $t$  estendidas, entre outras. Algumas propriedades são discutidas, assim como a estimação pelo método da máxima verossimilhança. No contexto de modelos regressão, as variáveis explicativas são incluídas no parâmetro de escala  $\lambda$ , interpretado diretamente como a mediana no caso contínuo. Estudos de simulação de Monte Carlo são realizados para avaliar o desempenho dos estimadores dos modelos propostos com e sem covariáveis, e na ausência e na presença de censura (10% e 30%). Três conjuntos de dados reais, incluindo um com observações censuradas, são usados para ilustrar a metodologia proposta. Em geral, as distribuições log-simétricas discretas propostas mostraram-se capazes de ajustar bem os conjuntos de dados considerados.

**Palavras-chave:** Distribuições discretas, análise de sobrevivência, método da máxima verossimilhança, simulações de Monte Carlo, software R.

---

---

# Abstract

The use of continuous probability distributions has been widespread in problems with a purely discrete nature. In general, these distributions are not appropriate in this scenario. In this paper, a class of discrete probability distributions based on continuous log-symmetric distributions is introduced, including as special cases the following discrete distributions: log-normal, log-Student- $t$ , log-power-exponential, log-contaminated-normal, extended Birnbaum-Saunders, and extended Birnbaum-Saunders- $t$ , among others. Some properties are discussed, as well as the estimation by the maximum likelihood method. In the context of regression models, covariates are included in the scale parameter  $\lambda$ , interpreted directly as the median in the continuous case. Monte Carlo simulation studies are carried out to evaluate the performance of the estimators of the proposed models both with and without covariates, and in both the absence and presence of censoring (10% and 30%). Three real-world data sets, including one with censored observations, are used to illustrate the proposed methodology. In general, the proposed discrete log-symmetric distributions provided good adjustments to the data sets considered.

**Keywords:** Discrete distributions, survival analysis, maximum likelihood method, Monte Carlo simulations, R software.

---

---

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Conceitos e definições básicas</b>	<b>5</b>
2.1	Análise de Sobrevivência . . . . .	5
2.1.1	Função de sobrevivência . . . . .	6
2.1.2	Função de risco . . . . .	6
2.1.3	Vida média residual . . . . .	7
2.1.4	Censura . . . . .	8
2.1.5	Estimador de Kaplan-Meier . . . . .	9
2.2	Distribuições simétricas contínuas . . . . .	10
2.2.1	Distribuição normal . . . . .	10
2.2.2	Distribuição $t$ de Student . . . . .	11
2.2.3	Distribuição exponencial potência . . . . .	12
2.3	Distribuições log-simétricas contínuas . . . . .	14
2.3.1	Distribuição log-normal . . . . .	15
2.3.2	Distribuição log- $t$ -Student . . . . .	16
2.3.3	Distribuição log-exponencial-potência . . . . .	17
2.4	Critérios para seleção o modelo . . . . .	19
2.4.1	Critério de Informação Akaike (AIC) . . . . .	19
2.4.2	Critério de Informação Bayesiano (BIC) . . . . .	19

---

2.5	Testes da qualidade do ajuste . . . . .	20
2.5.1	Teste qui-quadrado . . . . .	20
2.5.2	Teste de Cramér-von Mises . . . . .	21
<b>3</b>	<b>Distribuições log-simétricas discretas</b>	<b>23</b>
3.1	Distribuições log-simétricas discretas . . . . .	23
3.2	Propriedades Matemáticas . . . . .	27
3.2.1	Momentos e variância . . . . .	28
3.2.2	O $p$ -Quantil . . . . .	31
3.3	Estimadores de Máxima Verossimilhança . . . . .	34
3.3.1	Dados não censurados . . . . .	34
3.3.2	Dados censurados . . . . .	36
3.4	Simulações . . . . .	37
3.5	Aplicações . . . . .	46
3.5.1	Dados não censurados . . . . .	46
3.5.2	Dados censurados . . . . .	49
<b>4</b>	<b>Modelos de regressão log-simétricos discretos</b>	<b>53</b>
4.1	Modelos de regressão log-simétricos discretos . . . . .	53
4.2	Análise de resíduos . . . . .	54
4.3	Simulações . . . . .	56
4.4	Aplicações . . . . .	68
4.4.1	Dados não censurados . . . . .	69
4.4.2	Análise do ajuste dos modelos . . . . .	73
4.4.3	Dados censurados . . . . .	77
4.4.4	Análise do ajuste dos modelos . . . . .	81
<b>5</b>	<b>Considerações Finais</b>	<b>89</b>

---



---

<b>Bibliografia</b>	<b>91</b>
<b>A Dados utilizados na aplicação</b>	<b>95</b>

---

---

---

# Lista de Tabelas

2.1	Distribuições log-simétricas com suas respectivas funções geradoras de densidade.	18
3.1	Valores de média, viés e EQM para dados simulados da distribuição log-normal discreta considerando os estimadores de máxima verossimilhança e $\lambda = 2$ .	40
3.2	Valores de média, viés e EQM para dados simulados da distribuição log- <i>t</i> -Student discreta considerando os estimadores de máxima verossimilhança, $\xi = 4$ e $\lambda = 2$ .	40
3.3	Valores de média, viés e EQM para dados simulados da distribuição log-normal-contaminada discreta considerando os estimadores de máxima verossimilhança, $\lambda = 2$ e $\xi = (0.5, 0.5)^\top$ .	41
3.4	Valores de média, viés e EQM para dados simulados da distribuição log-exponencial-potência discreta considerando os estimadores de máxima verossimilhança, $\lambda = 2$ e $\xi = -0.5$ .	41
3.5	Valores de média, viés e EQM para dados simulados da distribuição Birnbaum-Saunders estendida discreta considerando os estimadores de máxima verossimilhança, $\lambda = 2$ e $\xi = 0.5$ .	42
3.6	Valores de média, viés e EQM para dados simulados da distribuição Birnbaum-Saunders- <i>t</i> discreta considerando os estimadores de máxima verossimilhança, $\lambda = 2$ and $\xi = (0.5, 4)^\top$ .	42
3.7	Dados de falha de um computador.	47

---

3.8	Estimativas de máxima verossimilhança e medidas de seleção de modelo para ajuste dos dados de falhas de computador. . . . .	48
3.9	Estimativas de máxima verossimilhança e medidas de seleção de modelo para ajuste dos dados de evasão no curso de Estatística. . . . .	51
3.10	Estimativas da função de sobrevivência via Kaplan-Meier e via distribuições log-simétricas discretas. . . . .	51
4.1	Valores de média, viés e EQM para dados simulados da distribuição log-normal discreta considerando os estimadores de máxima verossimilhança do modelo de regressão ( $\beta_0 = 1$ , e $\beta_1 = 2.5$ ). . . . .	57
4.2	Valores de média, viés e EQM para dados simulados da distribuição log- <i>t</i> -Student discreta considerando os estimadores de máxima verossimilhança do modelo de regressão ( $\beta_0 = 1$ , $\beta_1 = 2.5$ e $\xi = 4$ ). . . . .	58
4.3	Valores de média, viés e EQM para dados simulados da distribuição log-normal-contaminada discreta considerando os estimadores de máxima verossimilhança do modelo de regressão ( $\beta_0 = 1$ , $\beta_1 = 2.5$ e $\xi = (0.5, 0.5)^T$ ). . . . .	59
4.4	Valores de média, viés e EQM para dados simulados da distribuição log-exponencial-potência discreta considerando os estimadores de máxima verossimilhança do modelo de regressão ( $\beta_0 = 1$ , $\beta_1 = 2.5$ e $\xi = -0.5$ ). . . . .	60
4.5	Valores de média, viés e EQM para dados simulados da distribuição Birnbaum-Saunders estendida discreta considerando os estimadores de máxima verossimilhança do modelo de regressão ( $\beta_0 = 1$ , $\beta_1 = 2.5$ e $\xi = 0.5$ ). . . . .	61
4.6	Valores de média, viés e EQM para dados simulados da distribuição Birnbaum-Saunders- <i>t</i> estendida discreta considerando os estimadores de máxima verossimilhança do modelo de regressão ( $\beta_0 = 1$ , $\beta_1 = 2.5$ e $\xi = (0.5, 4)^T$ ). . . . .	62
4.7	Descrição das variáveis explicativas. . . . .	70
4.8	Estatísticas descritivas para as variáveis explicativas do número de projetos. . .	70

---

---

4.9	Modelos de regressão log-simétricos discretos para o número de projetos. . . . .	72
4.10	Critérios AIC e BIC para seleção dos modelos. . . . .	73
4.11	Erro absoluto médio para os valores estimados do número de projetos. . . . .	73
4.12	Perfil dos alunos do curso de Estatística (UFPI). . . . .	77
4.13	Testes de comparação das curvas de sobrevivência entre as categorias das variáveis explicativas. . . . .	78
4.14	Estimativas dos parâmetros dos modelos de regressão log-simétricos discretos para o tempo até a evasão. . . . .	80
4.15	Critérios AIC e BIC para seleção dos modelos. . . . .	81
4.16	Valores estimados de $\lambda$ para o número de semestres até a evasão sob os modelos log-simétricos discretos. . . . .	81
4.17	Frequências das combinações das categorias das covariáveis. . . . .	86
A.1	Dados não censurados referentes ao número de projetos em contratos de concessões dos estados brasileiros. Onde <i>lpop</i> : logaritmo do tamanho da população (IBGE), <i>respripib</i> : resultado primário - % do PIB, <i>servdivtotpib</i> : serviço da dívida pública - % do PIB e <i>desprec</i> : despesas com pessoal/receita corrente (Secretaria do Tesouro Nacional). . . . .	95
A.2	Dados censurados referentes ao número de semestres até a evasão no curso de Estatística da Universidade Federal do Piauí, do primeiro semestre de 2010 ao primeiro semestre de 2019. Censura (0: censura, 1: falha), Idade (0: > 20 anos, 1: ≤ 20 anos), Sexo (0: feminino, 1: masculino), Origem (0: Teresina, 1: outro município), Ingresso (0: enem, 1: vestibular). . . . .	108

---

---

---

# Lista de Figuras

2.1	Densidades de uma distribuição normal. . . . .	11
2.2	Densidades das distribuições $t$ de Student (linha tracejada), com $\xi = 4$ e $\xi = 12$ graus de liberdade, e normal padrão (linha sólida). . . . .	12
2.3	Densidades das distribuições exponencial potência (linha tracejada), com $\xi = -0.5$ e $\xi = 1$ , e normal padrão (linha sólida). . . . .	13
2.4	Densidades de uma distribuição log-normal. . . . .	16
2.5	Densidades das distribuições log- $t$ -Student (linha tracejada), com $\xi = 4$ e $\xi = 12$ , e log-normal (linha sólida). . . . .	17
2.6	Densidades das distribuições log-exponencial-potência (linha tracejada), com $\xi = -0.5$ e $\xi = 1$ , e log-normal (linha sólida). . . . .	18
3.1	Comportamento das distribuição log-normal discreta e função de risco para $\lambda = 20$ e $\phi$ . . . . .	25
3.2	Comportamento da distribuição log- $t$ -Student discreta e função de risco para $\lambda = 20$ , $\xi = 4$ e $\phi$ . . . . .	25
3.3	Comportamento da distribuição log-normal-contaminada discreta e função de risco para $\lambda = 20$ , $\xi_1 = 0.5$ , $\xi_2 = 0.5$ e $\phi$ . . . . .	26
3.4	Comportamento da distribuição log-exponencial-potência discreta e função de risco para $\lambda = 20$ , $\xi = 0.5$ e $\phi$ . . . . .	26

---

3.5	Comportamento da distribuição Birnbaum-Saunders estendida discreta e função de risco para $\lambda=20$ , $\xi=6$ e $\phi$ . . . . .	27
3.6	Comportamento da distribuição Birnbaum-Saunders estendida discreta e função de risco para $\lambda=20$ , $\xi_1=7$ , $\xi_2=9$ e $\phi$ . . . . .	27
3.7	Razão entre variância e esperança dos modelos log-simétricos discretos. . . . .	31
3.8	Comportamento do erro quadrático médio para as estimativas $\hat{\phi}$ segundo a censura, valores verdadeiros do parâmetro $\phi$ e tamanho da amostra. . . . .	43
3.9	Comportamento do erro quadrático médio para as estimativas $\hat{\lambda}$ segundo a censura, valores verdadeiros do parâmetro $\phi$ e tamanho da amostra. . . . .	44
3.10	Comportamento do viés para as estimativas $\hat{\phi}$ segundo a censura, valores verdadeiros do parâmetro $\phi$ e tamanho da amostra. . . . .	45
3.11	Comportamento do viés para as estimativas $\hat{\lambda}$ segundo a censura, valores verdadeiros do parâmetro $\phi$ e tamanho da amostra. . . . .	46
3.12	Distribuição e boxplot dos dados de falha de um computador. . . . .	47
3.13	Distribuição empírica ( $\bullet$ ) <i>versus</i> distribuições log-simétricas discretas ajustadas para os dados de falha de computador ( $\circ$ ). . . . .	49
3.14	Curva de sobrevivência (com intervalo de confiança) estimada pelo método Kaplan-Meier (a) e boxplot para os tempos de falha e censura (b). . . . .	50
3.15	Curvas de sobrevivência estimadas por Kaplan-Meier (—) e pelos modelos log-simétricos discretos (- - -). . . . .	52
4.1	Comportamento do erro quadrático médio para as estimativas $\hat{\phi}$ segundo a censura, valores verdadeiros do parâmetro $\phi$ e tamanho da amostra. . . . .	63
4.2	Comportamento do erro quadrático médio para as estimativas do intercepto $\hat{\beta}_0$ segundo a censura, valores verdadeiros do parâmetro $\phi$ e tamanho da amostra. . . . .	64
4.3	Comportamento do erro quadrático médio para as estimativas $\hat{\beta}_1$ segundo a censura, valores verdadeiros do parâmetro $\phi$ e tamanho da amostra. . . . .	65

---



---

4.4	Comportamento do viés para as estimativas $\hat{\phi}$ segundo a censura, valores verdadeiros do parâmetro $\phi$ e tamanho da amostra. . . . .	66
4.5	Comportamento do viés para as estimativas $\hat{\beta}_0$ segundo a censura, valores verdadeiros do parâmetro $\phi$ e tamanho da amostra. . . . .	67
4.6	Comportamento do viés para as estimativas $\hat{\beta}_1$ segundo a censura, valores verdadeiros do parâmetro $\phi$ e tamanho da amostra. . . . .	68
4.7	Distribuição do número de projetos (a) e boxplot (b). . . . .	69
4.8	Variáveis explicativas versus variáveis respostas. . . . .	71
4.9	Gráfico quantil-quantil (a) e histograma (b) para os resíduos do modelo log-normal discreto. . . . .	74
4.10	Gráfico quantil-quantil (a) e histograma (b) para os resíduos do modelo log-Student discreto. . . . .	74
4.11	Gráfico quantil-quantil (a) e histograma (b) para os resíduos do modelo log-normal-contaminado discreto. . . . .	75
4.12	Gráfico quantil-quantil (a) e histograma (b) para os resíduos do modelo log-exponencial-potência discreto. . . . .	75
4.13	Gráfico quantil-quantil (a) e histograma (b) para os resíduos do modelo Birnbaum-Saunders estendido discreto. . . . .	76
4.14	Gráfico quantil-quantil (a) e histograma (b) para os resíduos do modelo Birnbaum-Saunders- $t$ estendido discreto. . . . .	76
4.15	Curvas de sobrevivência de Kaplan-Meier para as covariáveis (— 0, --- 1). . . . .	78
4.16	Sobrevivências dos resíduos $e^{r_{q,i}}$ do modelo log-normal discreto estimadas pelo método de Kaplan-Meier e pelo modelo log-normal padrão (a) e suas respectivas curvas estimadas (b). . . . .	82
4.17	Sobrevivências dos resíduos $e^{r_{q,i}}$ do modelo log- $t$ -Student discreto estimadas pelo método de Kaplan-Meier e pelo modelo log-normal padrão (a) e suas respectivas curvas estimadas (b). . . . .	83

---

---

4.18	Sobrevivências dos resíduos $e^{r_{q,i}}$ do modelo log-normal-contaminado discreto estimadas pelo método de Kaplan-Meier e pelo modelo log-normal padrão (a) e suas respectivas curvas estimadas (b). . . . .	83
4.19	Sobrevivências dos resíduos $e^{r_{q,i}}$ do modelo log-exponencial-potência discreto estimadas pelo método de Kaplan-Meier e pelo modelo log-normal padrão (a) e suas respectivas curvas estimadas (b). . . . .	84
4.20	Sobrevivências dos resíduos $e^{r_{q,i}}$ do modelo BS discreto estimadas pelo método de Kaplan-Meier e pelo modelo log-normal padrão (a) e suas respectivas curvas estimadas (b). . . . .	84
4.21	Sobrevivências dos resíduos $e^{r_{q,i}}$ do modelo BS estendido discreto estimadas pelo método de Kaplan-Meier e pelo modelo log-normal padrão (a) e suas respectivas curvas estimadas (b). . . . .	85
4.22	Sobrevivências dos resíduos $e^{r_{q,i}}$ do modelo BS- $t$ discreto estimadas pelo método de Kaplan-Meier e pelo modelo log-normal padrão (a) e suas respectivas curvas estimadas (b). . . . .	85
4.23	Sobrevivências dos resíduos $e^{r_{q,i}}$ do modelo BS- $t$ estendido discreto estimadas pelo método de Kaplan-Meier e pelo modelo log-normal padrão (a) e suas respectivas curvas estimadas (b). . . . .	86
4.24	Ajuste dos modelos por categoria (0, 1) das covaráveis idade, origem e forma de ingresso (— Kaplan-Meier - - - modelo log-simétrico discreto). . . . .	87

---

# Abreviações e Siglas

FDP	Função densidade de probabilidade
FDA	Função de distribuição acumulada
FMP	Função massa de probabilidade
LS <sub>d</sub>	Log-simétrica discreta
EQM	Erro quadrático médio
CVM	Cramér-von Mises
KM	Kaplan-Meier
Ln	Log-normal
Lt	Log- <i>t</i> -Student
Lep	Log-exponencial-potência
Lnc	Log-normal-contaminada
BS	Birnbaum-Saunders
BSe	Birnbaum-Saunders estendida
BSt	Birnbaum-Saunders- <i>t</i>
BSte	Birnbaum-Saunders- <i>t</i> estendida
$H_0$	Hipótese nula
$H_1$	Hipótese alternativa
IC	Intervalo de confiança
EP	Erro padrão

---

---

# Capítulo 1

## Introdução

As distribuições discretas são caracterizadas por assumirem valores em conjunto enumerável, mas não necessariamente finito. Formalmente, diz-se que  $X$  tem uma distribuição discreta ou que  $X$  é uma variável aleatória discreta se  $X$  pode assumir um número  $k$  finito de diferentes valores  $x_1, \dots, x_k$  ou, no máximo, uma sequência infinita de diferentes valores  $x_1, x_2, \dots$  (DEGROOT e SCHERVISH, 2012).

As distribuições de probabilidade mais conhecidas que são utilizadas para modelar dados discretos são Bernoulli, binomial, hipergeométrica, Poisson, geométrica, binomial negativa, entre outras. Seguindo Bakouch et al. (2014), essas distribuições discretas tradicionais têm aplicabilidade limitada. À vista disso, foram desenvolvidas outras distribuições baseadas em distribuições contínuas utilizando alguma metodologia de discretização. Dentre essas distribuições pode-se destacar a distribuição normal discreta (Lisman e Van Zuylen (1972); Kemp (1997); Liang (1999); e Szablowski (2001)); a distribuição exponencial discreta (Sato et al., 1999); a distribuição gama discreta (Sato et al., 1999) e várias outras distribuições (ver CHAKRABORTY, 2015).

Por outro lado, quando uma variável aleatória  $X$  pode receber qualquer valor dentro de um determinado intervalo, diz-se que essa variável é contínua. Na definição de Meyer (2015),  $X$  é uma variável aleatória contínua se existir uma função  $f$ , denominada função densidade de

probabilidade (FDP), tal que  $f(x) \geq 0$  para todo  $x$ ; a integral em todo o espaço dessa função é 1, isto é,  $\int_{-\infty}^{\infty} f(x)dx = 1$ ; e para  $a$  e  $b$  pertencentes aos reais ( $a < b$ ),  $P(a \leq X \leq b) = \int_a^b f(x)dx$ .

A distribuição normal é uma das distribuições mais importantes dentre as distribuições contínuas por ser capaz de modelar muitos fenômenos aleatórios. Pode ser usada para aproximar probabilidades de outras distribuições, inclusive de distribuições discretas, por exemplo, a distribuição binomial. Outras distribuições contínuas de destaque são a exponencial, beta, beta generalizada, gama, gama-inversa, exponencial potência,  $t$  de Student, etc. As distribuições normal,  $t$  de Student e exponencial potência pertencem à classe de distribuições simétricas.

A classe de distribuições simétricas se caracteriza pela simetria em torno do parâmetro de posição no gráfico da função densidade de probabilidade. Essa classe contém distribuições com caudas mais leves e mais pesadas em comparação com a distribuição normal. Para dados simetricamente distribuídos e com observações atípicas a distribuição  $t$  de Student é mais indicada do que a normal, porque apresenta caudas mais pesadas. Aplicações de distribuições simétricas em diversas áreas são encontradas em Chmielewski (1981). Pode-se destacar outra classe de distribuições presente na literatura cujo suporte varia no intervalo  $(0, \infty)$  e deriva das distribuições simétricas, a classe de distribuições log-simétricas.

A classe de distribuições log-simétricas contínuas é caracterizada por distribuições assimétricas e estritamente positivas, sendo obtida de distribuições simétricas por meio de uma transformação exponencial. Distribuições pertencentes a essa classe são comumente utilizadas para modelar dados de sobrevivência (ou confiabilidade). As principais distribuições que pertencem a essa classe são a log-normal, log- $t$ -Student, log-hiperbólica, log-slash, log-normal-contaminada, log-exponencial-potência, Birnbaum-Saunders estendida, Birnbaum-Saunders- $t$  estendida, entre outras. Uma distribuição log-simétrica apresenta duas propriedades interessantes: é fechada sob mudança de escala e também sob reciprocidade (PUIG, 2008). Isso quer dizer que, se uma variável aleatória  $Y$  segue uma distribuição pertencente à classe de distribuições log-simétricas, então, para qualquer constante  $c > 0$ ,  $cY$  e  $1/Y$  seguirão a mesma distribuição

de  $Y$  (VANEGAS e PAULA, 2017).

Em geral, o uso de distribuições de probabilidade contínuas, e aqui são inseridas as distribuições log-simétricas contínuas, tem sido difundido em problemas com natureza discreta. Além disso, em muitas situações reais a obtenção de dados de uma distribuição contínua não é apropriada. Seguindo Chakraborty (2015), quase sempre os dados observados são de natureza discreta e, mesmo em casos em que as medições forem feitas em uma escala contínua, o registro das observações pode ser feito de maneira a tornar o modelo discreto mais apropriado. Em análise de confiabilidade, por exemplo, a utilização da distribuição Birnbaum-Saunders estendida para modelar o número de ciclos que um equipamento leva para falhar pode não ser adequado, porque o número de ciclos resulta de uma contagem. Além disso, o suporte das distribuições log-simétricas contínuas não contém o zero. Com a finalidade de tornar o modelo mais adequado para a modelagem de dados com essa natureza, foram propostos alguns métodos de discretização de distribuições contínuas.

Muitas metodologias para a obtenção de uma distribuição discreta a partir de uma distribuição contínua foram propostas na literatura. Stein e Dattero (1984) propuseram a discretização usando a função de risco. Nakagawa e Osaki (1975) propuseram um método de discretização que se baseia na função de sobrevivência. Roy e Dasgupta (2001) se basearam na função de distribuição acumulada para obtenção de distribuições discretas. Esses métodos de discretização, e outros mais, foram recentemente abordados por Chakraborty (2015).

A distribuição Weibull foi uma das primeiras no rol de distribuições contínuas a passar por um processo de discretização. Nakagawa e Osaki (1975) propuseram a distribuição Weibull discreta para modelagem de tempos de falha. Posteriormente, alguns autores estudaram as propriedades dessa distribuição, tais como Roy (2002), Murthy, Xie e Jiang (2004), Almalki e Nadarajah (2014), Brunello e Nakano (2015) e, mais recentemente, Vila, Nakano e Saulo (2019).

Nesse contexto, apesar da extensa aplicabilidade dos modelos log-simétricos contínuos, o uso deles em dados de natureza discreta pode não ser adequado. Partindo desse pressuposto, o

objetivo deste trabalho é apresentar uma classe de distribuições log-simétricas discretas, abordando algumas de suas propriedades, bem como sua aplicação a dados reais. As distribuições log-simétricas discretas consideradas neste trabalho são: log-normal, log- $t$ -Student, log-normal-contaminada, log-exponencial-potência, Birnbaum-Saunders- $t$  estendida e Birnbaum-Saunders estendida.

A relevância dessa classe discreta de distribuições se dá para evitar que as distribuições log-simétricas contínuas sejam utilizadas em dados discretos ou discretizados. A discretização de distribuições é interessante, porque preserva uma mais características importantes presentes nas distribuições contínuas. Além disso, em um processo de contagem, comumente o zero está presente no conjunto de dados e o uso de distribuições da classe log-simétrica contínua pode não ser adequado por excluir o zero do seu suporte. Portanto, este trabalho contribui para aplicações em dados discretos preservando pelo menos um atributo da classe contínua e inserindo o zero no suporte.

Este trabalho está segmentado em cinco capítulos, iniciados por esta introdução. No Capítulo 2 são introduzidos alguns conceitos em análise de sobrevivência, além da apresentação das distribuições simétricas contínuas estudadas por Cysneiros, Paula e Galea (2005) e log-simétricas contínuas abordadas recentemente por Vanegas e Paula (2016). O Capítulo 3 é referente às distribuições log-simétricas discretas, onde são apresentados alguns conceitos, propriedades e relações com as distribuições log-simétricas contínuas. Nesse capítulo também é feito um estudo de simulação para avaliar os estimadores de verossimilhança dos modelos e aplicações com dados reais. No Capítulo 4, são estudados os modelos de regressão baseados nas distribuições discretas propostas, e são avaliados seus respectivos estimadores de máxima verossimilhança para os parâmetros, além de aplicações a dados reais. No Capítulo 5 estão dispostas as considerações finais, além de sugestões para pesquisas futuras.



# Capítulo 2

## Conceitos e definições básicas

Este capítulo trata de algumas definições em análise de sobrevivência, traz a definição de distribuição log-simétrica contínua e algumas de suas propriedades. Além disso, são apresentados os critérios adotados para escolha dos modelos.

### 2.1 Análise de Sobrevivência

A análise de sobrevivência é definida por um conjunto de técnicas que tem como intuito analisar o tempo até a ocorrência de determinado evento, definido como falha. Essa falha pode ser, por exemplo, o tempo decorrido entre o diagnóstico de uma doença e a morte do paciente, o número de semanas que um equipamento leva até apresentar um defeito, etc. Conforme Colosimo e Giolo (2006), os elementos compostos pelo tempo inicial, escala de medida e evento de interesse integram o tempo de falha e devem ser nitidamente definidos. Para o tempo de início do estudo deve ser utilizado o mesmo critério para todas as observações e o acompanhamento dessas observações pode ser iniciado em tempos distintos. A escala de medida frequentemente utilizada é o tempo cronológico, mas há outras alternativas como, por exemplo, em testes de engenharia em que as escalas de medida são definidas como o número de ciclos até falha de um equipamento, o número de quilômetros que um automóvel percorreu ou outra medida se-

melhante. O evento de interesse (falha) é bem compreensível em certas situações, por exemplo, a morte de um paciente após o diagnóstico de um câncer. Porém, em alguns casos, a falha deve ser expressamente definida, por exemplo, se o evento de interesse é o tempo até a falha de um computador, é necessário explicitar o que será considerado como falha, se é a queima de algum equipamento, um defeito em algum software, etc.

### 2.1.1 Função de sobrevivência

Uma função de sobrevivência,  $S(x)$ , é definida como a probabilidade de não ocorrer falha até o tempo  $x$ . Essa função é monótona não crescente e, quando a variável  $X$  é contínua,  $S(0) = 1$  e  $\lim_{x \rightarrow \infty} S(x) = 0$ , ou seja, é certo que não irá ocorrer falha no tempo  $X = 0$  e que a probabilidade de sobreviver por um período muito longo de tempo é 0. A equação que descreve a função de sobrevivência é dada por:

$$S(x) = P(X > x) = 1 - F(x), \quad x \geq 0. \quad (2.1)$$

Quando  $X$  é uma variável aleatória discreta, assumindo os valores  $x_i, i = 1, 2, 3, \dots$ , com função massa de probabilidade  $p(x_i) = P(X = x_i)$ , onde  $x_1 < x_2 \dots$ , a função de sobrevivência é dada por (KLEIN e MOESCHBERGER, 1997):

$$S(x) = P(X > x) = \sum_{x_i > x} p(x_i). \quad (2.2)$$

### 2.1.2 Função de risco

A função de risco, ou função taxa de falha condicional, representa o risco instantâneo de uma observação experimentar o evento de interesse num intervalo entre  $x$  e  $x + \epsilon$  ( $\epsilon \rightarrow 0$ ),  $x > 0$ , considerando que essa observação sobreviveu até esse instante. Se  $X$  é uma variável aleatória contínua, a função de risco é expressa por:

$$h(x) = \lim_{\epsilon \rightarrow 0} \frac{P(x \leq X < x + \epsilon | X \geq x)}{\epsilon} = \frac{f(x)}{S(x)}, \quad (2.3)$$

onde  $f(x)$  e  $S(x)$  representam a função densidade de probabilidade e a função de sobrevivência de  $X$ , respectivamente.

A função de risco, quando  $X$  é uma variável aleatória discreta, é dada por (KLEIN e MOESCHBERGER, 1997):

$$h(x_i) = P(X = x_i | X \geq x_i) = \frac{P(X = x_i)}{S(x_{i-1})}, \quad i = 1, 2, \dots, \quad (2.4)$$

em que  $S(x_0) = 1$ . Como  $P(X = x_i) = S(x_{i-1}) - S(x_i)$ , a equação 2.4 pode ser reescrita como:

$$h(x_i) = \frac{P(X = x_i)}{P(X = x_i) + S(x_i)}, \quad i = 1, 2, \dots \quad (2.5)$$

### 2.1.3 Vida média residual

Outra função de grande interesse na análise de sobrevivência é a vida média residual, que representa o tempo restante que uma observação pode viver, em média, sabendo que essa observação sobreviveu até o tempo  $x$ . Matematicamente, a vida média residual é expressa pela fórmula:

$$L_x = E(X - x | X > x), \quad x > 0, \quad (2.6)$$

onde percebe-se que, quando  $x$  assume o valor 0, a vida residual média é a esperança de  $X$ , ou seja,  $L_0 = E(X)$ .

#### 2.1.4 Censura

Em análise de sobrevivência são frequentes os casos em que o tempo é observado parcialmente, isto é, o evento de interesse não ocorre durante o estudo. Essa intermissão no monitoramento do tempo até a falha é denominada censura. De acordo com Colosimo e Giolo (2006), as observações censuradas fornecem informações úteis para o estudo e não devem ser omitidas, pois isso pode gerar conclusões viciadas no cálculo das estatísticas. Portanto, as técnicas de análise de sobrevivência são recomendadas por incorporarem as censuras às análises estatísticas. Os tipos de censura mais usuais são: censura à direita, censura à esquerda, censura intervalar e censura não informativa. Censura à direita, que será considerada neste trabalho, ocorre quando é possível apenas saber o tempo de vida que determinada observação excede, ou seja, a falha não ocorre até o último instante de acompanhamento da observação. A censura à direita é dividida em outros três tipos: censura tipo I, censura tipo II e censura aleatória.

- Censura tipo I: Ocorre quando o estudo tem um período predeterminado para o seu término ( $x_0$ ). Assim, um par de variáveis  $(Y, \delta)$  é associado a cada observação do estudo, em que  $\delta$  é 0 se o tempo de vida  $X$  for superior a  $x_0$  e 1 caso a falha ocorra antes de  $x_0$ . Desse modo,  $Y = \min(X, x_0)$  e a quantidade de falhas observadas é aleatória.
- Censura tipo II: Nesse tipo de censura é fixada uma quantidade  $m$  de falhas em um estudo com  $n$  observações ( $m < n$ ), onde as  $n - m$  observações são censuradas. O tempo inicial para todas as observações é o mesmo e o tempo de duração do estudo é aleatório.
- Censura aleatória: Nesse caso, os tempos de censura são variáveis aleatórias independentes entre si e independentes dos tempos de sobrevivência. Esse tipo de censura acontece nos casos em que uma observação é removida por um motivo impertinente ao estudo, sem ter ocorrido a falha. Por exemplo, quando a causa da falha é diferente da considerada no estudo.

### 2.1.5 Estimador de Kaplan-Meier

O estimador de Kaplan-Meier (KAPLAN e MEIER, 1958), ou estimador produto-limite, é um estimador não paramétrico utilizado para estimar a função de sobrevivência quando há observações censuradas na amostra. Para a construção desse estimador, o número de intervalos de tempo considerado é equivalente ao número de falhas distintas.

Considerando os seguintes pontos:

- $x_1 < x_2 < \dots < x_k$ , os tempos distintos e ordenados de falha,
- $d_j$  o número de falhas em  $x_j$ ,  $j = 1, 2, \dots, k$ , e
- $n_j$  o número de indivíduos sob risco em  $x_j$ , ou seja, as observações que não falharam e não foram censuradas até o instante imediatamente anterior a  $x_j$ ,

o estimador de Kaplan-Meier é dado por:

$$\hat{S}(x) = \prod_{j:x_j < x} \left( \frac{n_j - d_j}{n_j} \right) = \prod_{j:x_j < x} \left( 1 - \frac{d_j}{n_j} \right). \quad (2.7)$$

De acordo com Colosimo e Giolo (2006), as principais propriedades do estimador de Kaplan-Meier são:

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

Os dados de sobrevivência podem ser modelados por qualquer distribuição probabilística que apresente suporte positivo. Portanto, as distribuições pertencentes à classe log-simétrica, cujo suporte pertence ao intervalo  $(0, +\infty)$ , podem ser utilizadas para estudos com dados dessa natureza. As distribuições log-simétricas derivam das distribuições simétricas.

## 2.2 Distribuições simétricas contínuas

Seja  $Y$  uma variável aleatória contínua com parâmetro de localização  $-\infty < \mu < \infty$  e parâmetro de escala  $\phi > 0$ .  $Y$  pertence à classe de distribuições simétricas (CYSNEIROS; PAULA; GALEA, 2005) se sua FDP é dada por:

$$f_Y(y) = \frac{1}{\sqrt{\phi}} g \left[ \left( \frac{y - \mu}{\sqrt{\phi}} \right)^2 \right], \quad y \in \mathbb{R}, \quad (2.8)$$

em que  $g(\cdot)$  é a função geradora de densidade, sendo as condições  $g(u) > 0$ , com  $u > 0$  e  $\int_0^\infty u^{-1/2} g(u) du = 1$  necessárias e suficientes para que  $f_Y(y)$  seja uma FDP. A função  $g(\cdot)$  pode envolver um ou mais parâmetros extras que, neste trabalho, serão definidos por  $\xi$ , ou um vetor de parâmetros  $\boldsymbol{\xi}$ . Uma variável aleatória simétrica  $Y$  é representada por  $Y \sim \mathcal{S}(\mu, \phi, g(\cdot))$ .

### 2.2.1 Distribuição normal

Uma das distribuições simétricas mais conhecidas e importantes é a distribuição normal. A principal razão para a vasta utilização dessa distribuição é decorrente do Teorema Central do Limite que demonstra que, satisfazendo determinadas condições de regularidade, a soma de quaisquer  $n$  variáveis aleatórias independentes terá distribuição aproximadamente normal para  $n$  suficientemente grande.

A FDP da distribuição normal  $Y \sim N(\mu, \phi)$  é dada por:

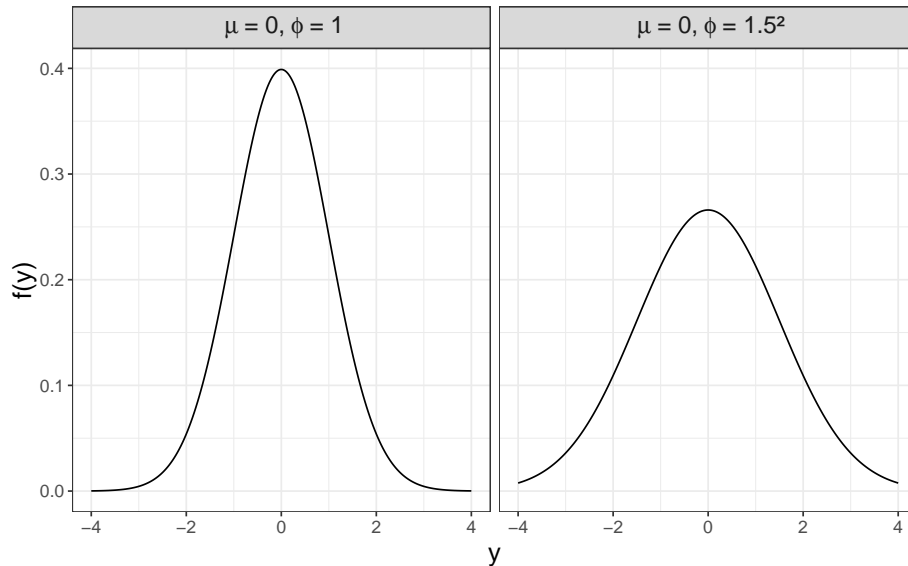
$$f_Y(y) = \frac{1}{\sqrt{2\pi\phi}} \exp \left[ -\frac{1}{2} \left( \frac{y - \mu}{\sqrt{\phi}} \right)^2 \right], \quad y \in \mathbb{R}, \quad (2.9)$$

e sua função geradora de densidade  $g(\cdot)$  é representada como:

$$g(u) \propto \exp \left[ -\frac{1}{2} u \right], \quad u > 0. \quad (2.10)$$

A distribuição normal é dita normal padrão quando  $\mu = 0$  e  $\phi = 1$ .

A Figura 2.1 apresenta as curvas de densidade da distribuição normal com mesma média ( $\mu=0$ ) e variâncias  $\phi=1$  e  $\phi = 1.5^2$ .



**Figura 2.1:** Densidades de uma distribuição normal.

### 2.2.2 Distribuição $t$ de Student

A distribuição  $t$  de Student tem maior variabilidade em relação a distribuição normal, possuindo caudas mais pesadas. Portanto, a distribuição  $t$  é mais indicada para modelar dados com observações extremas.

Uma variável aleatória  $Y$  que segue uma distribuição  $t$  de Student é denotada por  $Y \sim t(\mu, \phi, \xi)$ , em que  $\xi$  são os graus de liberdade. Quando  $\xi \rightarrow \infty$  essa distribuição se aproxima de uma distribuição normal com média  $-\infty < \mu < +\infty$  e variância  $\phi > 0$ . A FDP de  $Y$  é da seguinte forma:

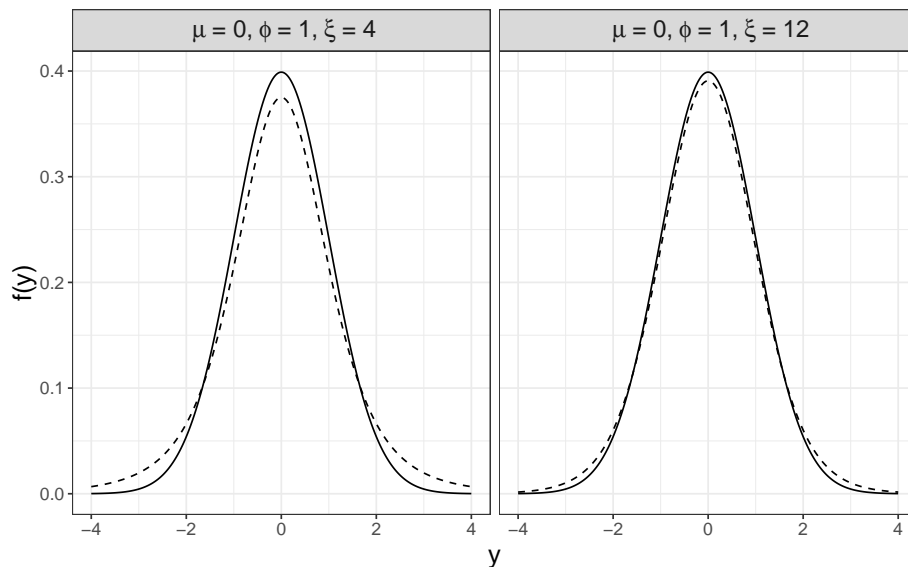
$$f_Y(y) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left[ 1 + \frac{\left(\frac{y-\mu}{\sqrt{\phi}}\right)^2}{\xi} \right]^{-\frac{\xi+1}{2}}, \quad \xi > 0 \text{ e } y \in \mathbb{R}, \quad (2.11)$$

em que  $\Gamma(\cdot)$  representa a função gama, calculada como  $\Gamma(k) = \int_0^\infty u^{k-1} e^{-u} du$ .

A função geradora de densidade da distribuição  $t$  de Student pode ser escrita como:

$$g(u) \propto \left[ 1 + \frac{u}{\xi} \right]^{-\frac{\xi+1}{2}}, \quad u, \xi > 0. \quad (2.12)$$

A Figura 2.2 representa as densidades das distribuições  $t$  de Student e normal (linha sólida).



**Figura 2.2:** Densidades das distribuições  $t$  de Student (linha tracejada), com  $\xi = 4$  e  $\xi = 12$  graus de liberdade, e normal padrão (linha sólida).

### 2.2.3 Distribuição exponencial potência

Outra distribuição com vasta aplicação é a distribuição exponencial potência, conhecida também como normal generalizada. Essa distribuição é uma generalização da distribuição normal e foi proposta inicialmente por Subbotin (1923), sendo reintroduzida por Nadarajah (2005) como uma generalização das distribuições normal e Laplace (exponencial dupla). A FDP dessa distribuição é da forma:

$$f_Y(y) = \frac{1}{\sqrt{\phi} \Gamma\left(1 + \frac{1}{2\xi}\right) 2^{1+\frac{1}{2\xi}}} \exp\left[-\frac{1}{2} \left|\frac{y - \mu}{\sqrt{\phi}}\right|^{2\xi}\right], \quad \xi > 0 \text{ e } y \in \mathbb{R}. \quad (2.13)$$

A FDP da exponencial potência descrita na equação (2.13) é apresentada com outra para-



metrização em Box e Tiao (1973), p. 157, equação (3.2.3), sendo escrita da seguinte forma:

$$f_Y(y) = \frac{k}{\sqrt{\phi}} \exp \left[ -\frac{1}{2} \left| \frac{y - \mu}{\sqrt{\phi}} \right|^{\frac{2}{1+\xi}} \right], \quad (2.14)$$

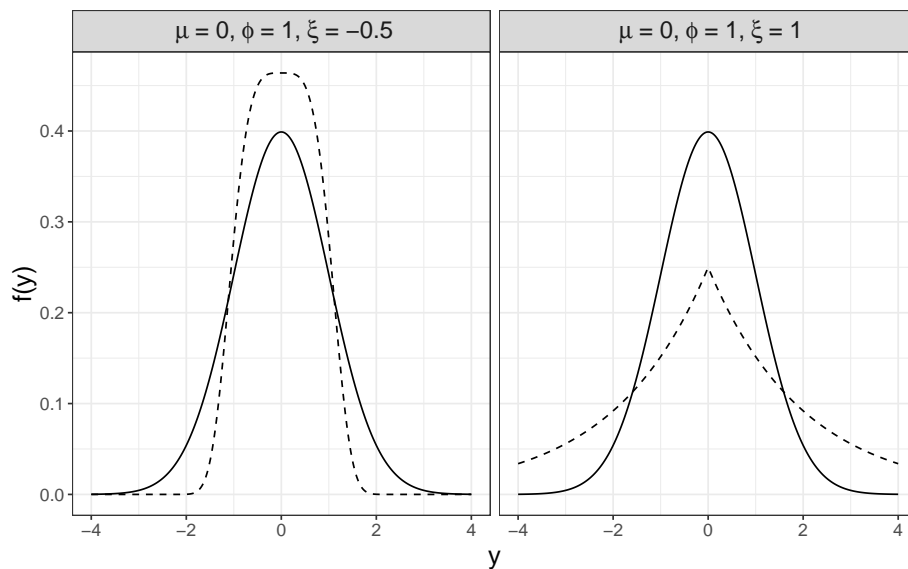
em que  $k^{-1} = \Gamma \left( 1 + \frac{1+\xi}{2} \right) 2^{1+(\frac{1+\xi}{2})}$ ,  $-1 < \xi \leq 1$ .

Quando  $\xi = 0$  e  $\xi = 1$  a densidade (2.14) se reduz às distribuições normal e exponencial dupla, respectivamente.

A função geradora de densidade da distribuição exponencial potência é:

$$g(u) \propto \exp \left[ -\frac{1}{2} u^{\left( \frac{1}{1+\xi} \right)} \right], \quad -1 < \xi \leq 1. \quad (2.15)$$

A Figura 2.3 ilustra as densidades da distribuição exponencial potência para os valores  $\xi = -1/2$  e  $\xi = 1$  e da normal padrão (equivalente a exponencial potência quando  $\xi = 0$  na equação (2.14)).



**Figura 2.3:** Densidades das distribuições exponencial potência (linha tracejada), com  $\xi=-0.5$  e  $\xi=1$ , e normal padrão (linha sólida).

Mais detalhes dessas e de outras distribuições simétricas podem ser consultados em Cysneiros, Paula e Galea (2005).

### 2.3 Distribuições log-simétricas contínuas

Distribuições log-simétricas contínuas são de particular interesse para descrever dados estritamente positivos e assimétricos, com a possibilidade de observações discrepantes. Na literatura existem algumas discussões e aplicações que consideram os modelos log-simétricos contínuos; ver, por exemplo, Jones (2008), Vanegas e Paula (2016a,b), Saulo e Leao (2017), Medeiros e Ferrari (2017) e Ventura et al. (2019). Uma variável aleatória contínua  $Y$  segue uma distribuição log-simétrica se sua função de densidade de probabilidade (FDP) for dada por:

$$f_Y(y|\boldsymbol{\theta}) \propto \frac{1}{\sqrt{\phi} y} g(a_{\boldsymbol{\theta}}^2(y)), \quad a_{\boldsymbol{\theta}}(y) = \log\left(\frac{y}{\lambda}\right)^{1/\sqrt{\phi}}, \quad y > 0; \quad (2.16)$$

onde  $\boldsymbol{\theta} = (\lambda, \phi)$ ,  $\lambda > 0$ ,  $\phi > 0$ , e a constante de proporcionalidade é exatamente a função de partição (ou inversa)  $\int_{-\infty}^{\infty} g(w^2) dw$ .

A função  $g$  é uma geradora de densidade, de modo que  $g(u) > 0$  para  $u > 0$ . Essa função está associada a um parâmetro extra  $\xi$  (ou vetor de parâmetros  $\boldsymbol{\xi}$ ).

Os parâmetros  $\lambda$  e  $\phi$  são os parâmetros de escala e forma e representam, respectivamente, a mediana e a assimetria (ou dispersão relativa) da distribuição de  $Y$ . Neste trabalho é usada a notação  $Y \sim \text{LS}(\boldsymbol{\theta}, g)$ . Nota-se que, se  $g(u)$  em (2.16) é  $\exp(-u/2)$ ;  $[1 + (u/\xi)]^{-(\xi+1)/2}$ ,  $\xi > 0$ ;  $\exp(-u^{1/(1+\xi)}/2)$ ,  $-1 < \xi \leq 1$ ;  $\sqrt{\xi_2} \exp(-\xi_2 u/2) + [(1 - \xi_1)/\xi_1] \exp(-u/2)$ ,  $0 < \xi_1, \xi_2 < 1$ ;  $\cosh(u^{1/2}) \exp(-(2/\xi^2) \sinh^2(u^{1/2}))$ ,  $\xi > 0$ ; ou  $\cosh(u^{1/2}) [\xi_2 \xi_1^2 + 4 \sinh^2(u^{1/2})]^{-(\xi_2+1)/2}$ ,  $\xi_1, \xi_2 > 0$ ; tem-se, respectivamente, as distribuições log-normal, log- $t$ -Student, log-exponencial-potência, log-normal-contaminada, Birnbaum-Saunders estendida ou Birnbaum-Saunders- $t$  estendida; ver Vanegas e Paula (2016) e Ventura et al. (2019). Se  $Y \sim \text{LS}(\boldsymbol{\theta}, g)$ , a função de distribuição acumulada associada (FDA) é dada por  $F_Y(y|\boldsymbol{\theta}) = (G \circ a_{\boldsymbol{\theta}})(y)$ , onde a função

$G : \mathbb{R} \rightarrow [0, 1]$  é definida como:

$$G(r) = (Z_g)^{-1} \int_{-\infty}^r g(z^2) dz, \quad r \in \mathbb{R}, \quad (2.17)$$

onde  $Z_g = \int_{-\infty}^{\infty} g(w^2) dw$ . Nota-se que  $G(0) = 1/2$ ,  $G(+\infty) = 1$ ,  $G(-\infty) = 0$ , de modo que  $G(\cdot)$  é uma função contínua, estrita e monotonicamente crescente. Portanto,  $G(\cdot)$  tem uma função inversa indicada por  $G^{-1}(\cdot)$ . Além disso, nota-se que  $G(\cdot)$  é uma FDA e  $G^{-1}(1 - p) = -G^{-1}(p)$  para  $p \in (0, 1)$ .

Algumas propriedades dessa classe de distribuições foram exploradas no trabalho de Vane-gas e Paula (2016). Considerando  $X \sim \mathcal{LS}(\lambda, \phi, g(\cdot))$ , esses autores verificaram que:

- (a) A função de distribuição acumulada (FDA) de  $X$  pode ser escrita como  $F_X(x) = F_Z(\tilde{x})$ , em que  $F_Z(\cdot)$  é a FDA de  $Z = (Y - \mu)/\sqrt{\phi} \sim \mathcal{S}(0, 1, g(\cdot))$ .
- (b)  $X^* = (X/\lambda)^{\frac{1}{\sqrt{\phi}}} \sim \mathcal{LS}(1, 1, g(\cdot))$ , isto é,  $X^*$  segue uma distribuição log-simétrica padrão.
- (c)  $cX \sim \mathcal{LS}(c\lambda, \phi, g(\cdot))$  para todo  $c > 0$ .
- (d)  $X^c \sim \mathcal{LS}(\lambda^c, c^2\phi, g(\cdot))$  para todo  $c \neq 0$ .
- (e)  $(X/\lambda)$  e  $(\lambda/X)$  são variáveis aleatórias identicamente distribuídas.
- (f) A função quantil de  $X$  é dada por  $Q(p) = \lambda \exp(\sqrt{\phi} Z_{\xi}^{(p)})$ , onde  $Z_{\xi}^{(p)}$  é o quantil 100(p)% de  $Z = (Y - \mu)/\sqrt{\phi} \sim \mathcal{S}(0, 1, g(\cdot))$ .

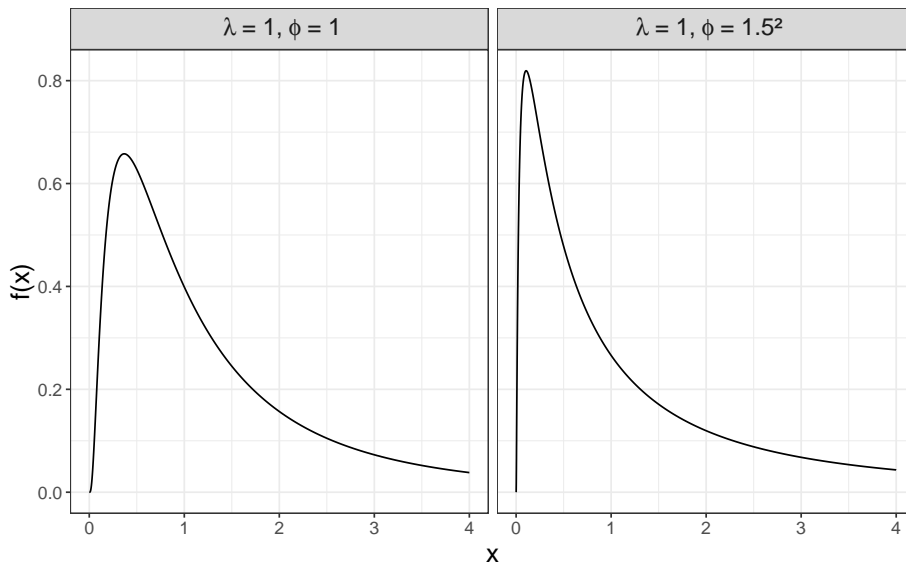
### 2.3.1 Distribuição log-normal

Se  $X$  tem distribuição log-normal, então uma variável aleatória  $Y = \log(X)$  é normalmente distribuída. Essa distribuição também é referida na literatura como distribuição de Galton, em referência ao estatístico inglês Francis Galton.

Uma variável aleatória  $X$ , denotada por  $X \sim \text{LN}(\lambda, \phi)$ , com  $\lambda > 0$  e  $\phi > 0$ , é log-normalmente distribuída se sua FDP é definida como:

$$f_X(x) = \frac{1}{x\sqrt{2\pi\phi}} \exp \left\{ -\frac{1}{2} \left[ \frac{\log(x) - \log(\lambda)}{\sqrt{\phi}} \right]^2 \right\}, \quad x > 0. \quad (2.18)$$

A Figura 2.4 apresenta as curvas de densidade da distribuição log-normal com  $\lambda = 1$  e  $\phi = (1, 1.5^2)$ .



**Figura 2.4:** Densidades de uma distribuição log-normal.

### 2.3.2 Distribuição log-*t*-Student

A distribuição log-*t*-Student pode ser considerada uma extensão da distribuição log-normal. A FDP da distribuição log-*t*-Student é escrita da seguinte forma:

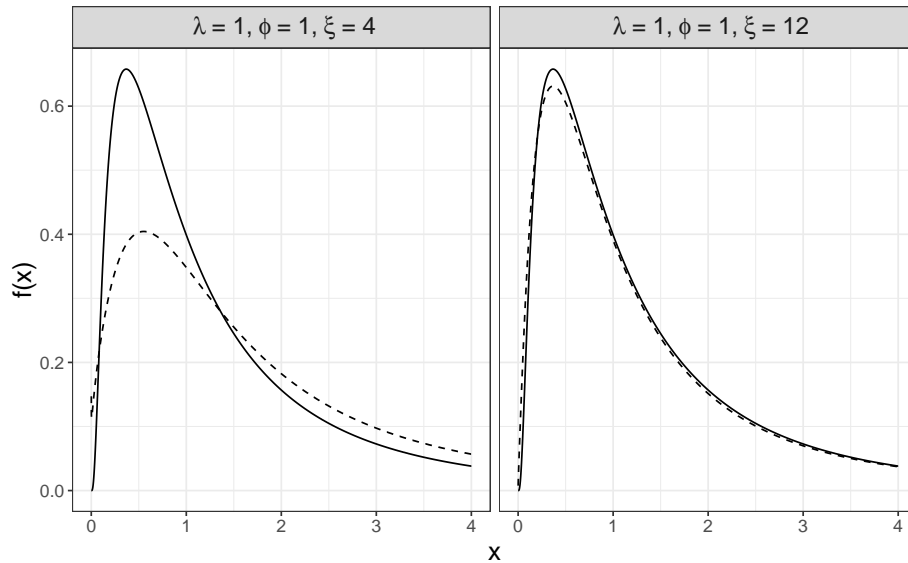
$$f_X(x) = \frac{\Gamma\left(\frac{\xi+1}{2}\right)}{x\sqrt{\xi\pi}\Gamma\left(\frac{\xi}{2}\right)} \left[ 1 + \frac{1}{\xi} \left( \frac{\log(x) - \log(\lambda)}{\sqrt{\phi}} \right)^2 \right]^{-\frac{\xi+1}{2}}, \quad x, \xi > 0. \quad (2.19)$$

Na forma mais simples, quando  $\lambda = 1$  e  $\phi = 1$ , sua densidade é dada por:

$$f_X(x) = \left[ \sqrt{\xi} B\left(\frac{1}{2}, \frac{\xi}{2}\right) \left( 1 + \frac{(\log(x))^2}{\xi} \right)^{(\xi+1)/2} x \right]^{-1}, \quad x, \xi > 0, \quad (2.20)$$

em que  $B(a, b)$  é a função beta dada por  $\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ . Quando  $\xi \rightarrow \infty$ , a distribuição log- $t$ -Student se aproxima de uma distribuição log-normal.

A Figura 2.5 apresenta as curvas de densidade da distribuição log- $t$ -Student ( $\xi_1=6$ ) e da distribuição log-normal.



**Figura 2.5:** Densidades das distribuições log- $t$ -Student (linha tracejada), com  $\xi = 4$  e  $\xi = 12$ , e log-normal (linha sólida).

### 2.3.3 Distribuição log-exponencial-potência

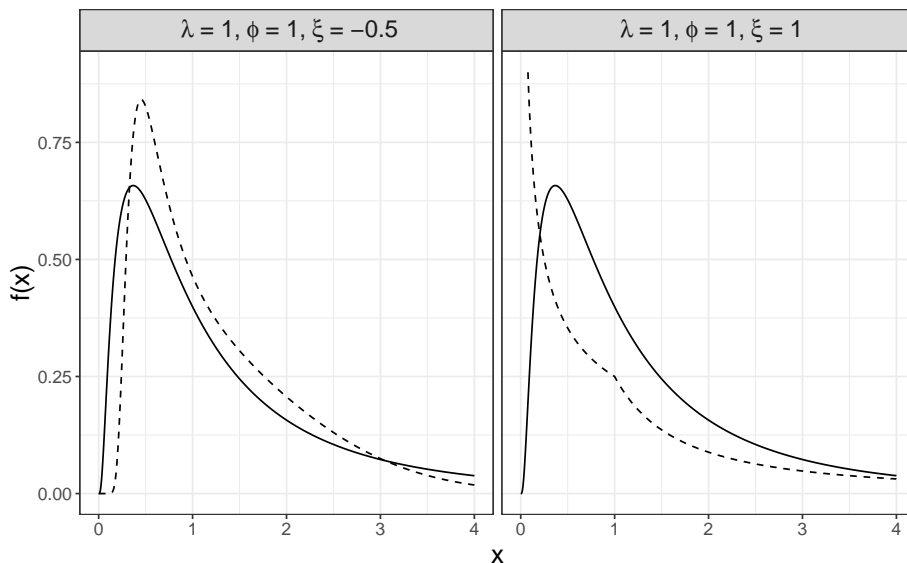
A distribuição log-exponencial-potência tem a FDP (ver LINDSEY, 2001, p. 201), considerando  $\mu=\log(\lambda)$ , escrita como:

$$f_X(x) = \frac{1}{x\sqrt{\phi}\Gamma\left(1 + \frac{1}{2\xi}\right) 2^{1+\frac{1}{2\xi}}} \exp\left[-\frac{1}{2} \left| \frac{\log(x) - \log(\lambda)}{\sqrt{\phi}} \right|^{2\xi}\right], \quad \xi > 0. \quad (2.21)$$

Considerando  $k$  e a reparametrização da equação (2.14), a FDP da log-exponencial-potência é expressa da seguinte forma:

$$f_X(x) = \frac{k}{x\sqrt{\phi}} \exp\left[-\frac{1}{2} \left| \frac{\log(x) - \log(\lambda)}{\sqrt{\phi}} \right|^{\frac{2}{1+\xi}}\right], \quad -1 < \xi \leq 1. \quad (2.22)$$

Quando  $\xi=1$  e  $\xi=0$  a FDP dada na equação (2.22) é equivalente às distribuições log-Laplace e log-normal, respectivamente. A Figura 2.6 apresenta as curvas de densidade da distribuição log-exponencial-potência ( $\xi=-0.5, \xi = 1$ ) e da distribuição log-normal.



**Figura 2.6:** Densidades das distribuições log-exponencial-potência (linha tracejada), com  $\xi=-0.5$  e  $\xi = 1$ , e log-normal (linha sólida).

A Tabela 2.1 resume as funções geradoras de densidade e seus respectivos parâmetros extras para as distribuições log-normal-contaminada, Birnbaum-Saunders estendida e Birnbaum-Saunders- $t$  estendida.

**Tabela 2.1:** Distribuições log-simétricas com suas respectivas funções geradoras de densidade.

Distribuição	$g(u)$	Parâmetro extra ( $\xi$ )
Log-normal-contaminada( $\lambda, \phi, \xi_1, \xi_2$ )	$\propto \sqrt{\xi_2} \exp \left[ -\frac{1}{2} \xi_2 u \right] + \frac{(1-\xi_1)}{\xi_1} \exp \left[ -\frac{1}{2} u \right]$	$0 < \xi_1 < 1, 0 < \xi_2 < 1$
Birnbaum-Saunders estendida ( $\lambda, \phi, \xi$ )	$\propto \cosh(u^{\frac{1}{2}}) \exp \left[ -\frac{2}{\xi^2} \sinh^2(u^{\frac{1}{2}}) \right]$	$\xi > 0$
Birnbaum-Saunders- $t$ estendida ( $\mu, \phi, \xi_1, \xi_2$ )	$\propto \cosh(u^{\frac{1}{2}}) [\xi_2 \xi_1^2 + 4 \sinh^2(u^{\frac{1}{2}})]^{-\frac{\xi_2+1}{2}}$	$\xi_1 > 0, \xi_2 > 0$

## 2.4 Critérios para seleção o modelo

Ao utilizar o mesmo conjunto de dados para ajustá-lo a diferentes modelos é preciso definir segundo algum critério qual desses modelos é o melhor. Na literatura existem diversas metodologias que auxiliam nas escolha desses modelos, dentre as quais pode-se destacar o Critério de Informação Akaike (AIC) e o Critério de Informação Bayesiano (BIC).

### 2.4.1 Critério de Informação Akaike (AIC)

Esse critério foi desenvolvido por Akaike (1974) e admite que existe um modelo verdadeiro e desconhecido que descreve os dados. Baseado na informação de Kullback-Leiber (K-L), esse critério seleciona dentre os modelos avaliados aquele que minimiza a informação de K-L. A estimativa de AIC para um determinado modelo é dada por:

$$AIC = -2\ell(\hat{\theta}) + 2k,$$

em que  $\ell(\hat{\theta})$  é o valor da função de log-verossimilhança do modelo baseado nas estimativas de máxima verossimilhança ( $\hat{\theta}$ ) e  $k$  é o número de parâmetros que compõem o modelo. O modelo com menor medida de AIC será selecionado.

### 2.4.2 Critério de Informação Bayesiano (BIC)

O Critério de Informação Bayesiano foi proposto Schwartz (1978) e seleciona os modelos considerando a probabilidade a posteriori. Esse critério pressupõe que exista um modelo verdadeiro que melhor descreve os dados, entre os modelos considerados. Desse modo, o critério é definido como a estatística que maximiza a probabilidade de detectar o modelo verdadeiro dentre os modelos propostos. O valor desse critério para um modelo qualquer é expresso por:

$$BIC = -2\ell(\hat{\theta}) + 2k\ln(n),$$

em que  $\ell(\hat{\theta})$  é o valor da função de log-verossimilhança do modelo baseado nas estimativas de máxima verossimilhança ( $\hat{\theta}$ ),  $k$  é o número de parâmetros que compõem o modelo e  $n$  é o tamanho da amostra. Assim como no critério AIC, o modelo selecionado é o que apresenta menor valor de BIC.

## 2.5 Testes da qualidade do ajuste

### 2.5.1 Teste qui-quadrado

O teste mais usual para verificar a adequabilidade dos dados a uma determinada distribuição é o teste qui-quadrado de Pearson (PEARSON, 1900). Considerando uma amostra aleatória de tamanho  $n$ , as hipóteses nula e alternativa desse teste são escritas da seguinte forma:

$$\begin{cases} H_0 : \text{Os dados amostrais seguem a distribuição hipotética.} \\ H_1 : \text{Os dados amostrais não seguem a distribuição hipotética.} \end{cases}$$

A estatística utilizada para testar a qualidade do ajuste é dada por:

$$Q = \sum_{i=1}^k \frac{(f_{x_i} - nP_{x_i})^2}{nP_{x_i}}, \quad (2.23)$$

em que  $k$  representa o número de classes,  $f_{x_i}$  e  $P_{x_i}$  representam a frequência observada e a probabilidade relacionada à distribuição hipotética na classe  $i$ , respectivamente. A estatística  $Q$  segue, sob  $H_0$ , uma distribuição qui-quadrado com  $k - p - 1$  graus de liberdade, sendo  $p$  o número de parâmetros estimados no modelo. Assim, considerando um nível de significância  $\alpha$  a região de rejeição é definida comparando a estatística  $Q$  com uma qui-quadrado com  $k - p - 1$  graus de liberdade ( $\chi_{\alpha, k-p-1}^2$ ).



### 2.5.2 Teste de Cramér-von Mises

Originalmente o teste de Cramér-von Mises foi desenvolvido por Harald Cramér e Richard von Mises (CRAMER, 1928; von MISES, 1928) apenas para dados contínuos e não contemplava dados de natureza discreta. Porém, Choulakian et al. (1994) generalizaram os testes de Cramér-von Mises para distribuições discretas. Analiticamente, é difícil obter o valor crítico desse teste. Entretanto, pode ser obtido no software R por meio da função `cmv.test` do pacote `dgof` (ARNOLD e EMERSON, 2012). As hipóteses desse teste são similares às hipóteses do teste qui-quadrado de aderência, isto é, hipótese nula é que os dados da amostra seguem a distribuição hipotética. Portanto, nos testes qui-quadrado e Cramér-von Mises, a não rejeição da hipótese nula indica uma boa aderência do modelo aos dados.



# Capítulo 3

## Distribuições log-simétricas discretas

Neste capítulo é estudada uma classe de distribuições log-simétricas discretas obtidas a partir das distribuições log-simétricas contínuas, estudadas por Vanegas e Paula (2016), por meio de um processo de discretização. Algumas propriedades dessa classe são obtidas.

### 3.1 Distribuições log-simétricas discretas

Na definição de Kemp (2004), se  $X$  é uma variável aleatória contínua e não negativa com função de sobrevivência  $S_X(x) = p(X \geq x)$ , então a variável aleatória  $Y = \lfloor X \rfloor$  tem função massa de probabilidade (FMP) dada por:

$$\begin{aligned} p(Y = k) &= p(k \leq X < k + 1) & (3.1) \\ &= F_X(k + 1) - F_X(k) \\ &= S_X(k) - S_X(k + 1), \quad k = 0, 1, 2, \dots, \end{aligned}$$

em que  $\lfloor X \rfloor$  representa o maior inteiro menor ou igual a  $X$ . Uma variável aleatória discreta  $X$ , tomando valores no conjunto  $\{0, 1, \dots\}$ , segue uma distribuição log-simétrica discreta com vetor de parâmetros  $\theta = (\lambda, \phi)$ , onde  $\lambda > 0, \phi > 0$ , denotado por  $X \sim \text{LS}_d(\theta, g)$ , se sua função

de massa de probabilidade (FMP) for dada por:

$$p(x|\boldsymbol{\theta}) = G(a_{\boldsymbol{\theta}}(x+1)) - G(a_{\boldsymbol{\theta}}(x)), \quad x = 0, 1, \dots, \quad (3.2)$$

onde  $a_{\boldsymbol{\theta}}(\cdot)$  e  $G(\cdot)$  são como em (2.16) e (2.17), respectivamente. Note que  $G(a_{\boldsymbol{\theta}}(0)) = G(-\infty) = 0$ . Dada a função geradora de densidade  $g$ , definida no Item (2.16), os parâmetros  $\lambda$  e  $\phi$  determinam completamente a FMP (3.2) em  $x = 0$ . Como as funções  $G(\cdot)$  e  $a_{\boldsymbol{\theta}}(\cdot)$  são estritamente crescentes, e

$$\lim_{n \rightarrow \infty} \sum_{x=0}^n p(x|\boldsymbol{\theta}) = \lim_{n \rightarrow \infty} G(a_{\boldsymbol{\theta}}(n+1)) = G(+\infty) = 1,$$

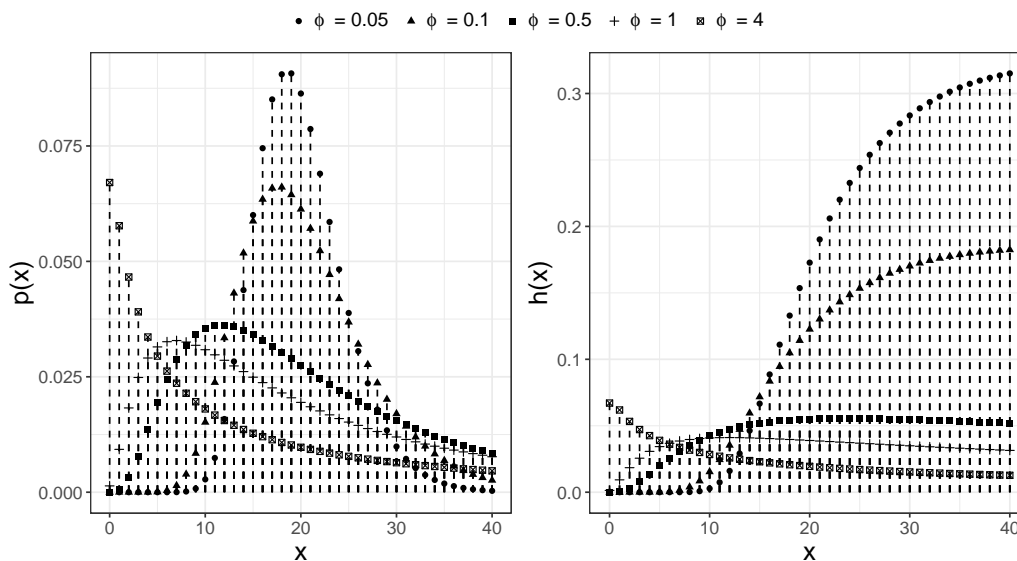
fica claro que  $p(x|\boldsymbol{\theta})$  é uma FMP.

A função de distribuição acumulada (FDA), função de sobrevivência ou confiabilidade ( $S$ ) e a taxa de risco ( $h$ ) da distribuição  $LS_d$  são dados, respectivamente, por:

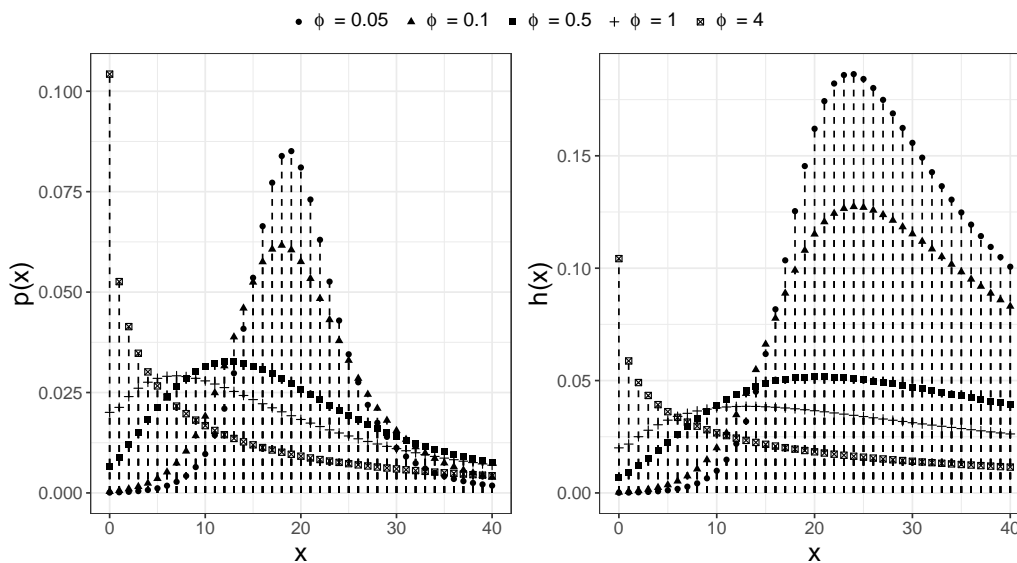
$$F(x|\boldsymbol{\theta}) = 1 - S(x|\boldsymbol{\theta}) = G(a_{\boldsymbol{\theta}}(\lfloor x \rfloor + 1)), \quad x \geq 0,$$

$$h(x|\boldsymbol{\theta}) = \frac{p(x|\boldsymbol{\theta})}{p(x|\boldsymbol{\theta}) + S(x|\boldsymbol{\theta})} = \frac{G(a_{\boldsymbol{\theta}}(x+1)) - G(a_{\boldsymbol{\theta}}(x))}{1 - G(a_{\boldsymbol{\theta}}(x))}, \quad x = 0, 1, \dots$$

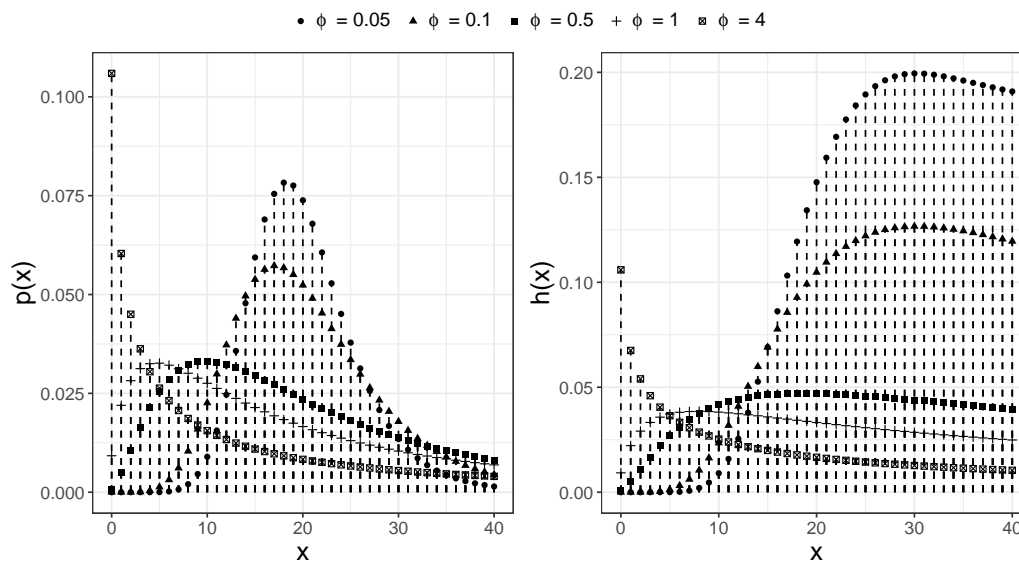
O comportamento das distribuições log-simétricas discretas e das suas respectivas funções de risco pode ser observado nas Figuras (3.1-3.6). Observando o comportamento da função de probabilidade nas Figuras 3.1, 3.2, 3.3 e 3.4, nota-se que, à medida que o parâmetro  $\phi$  diminui, a distribuição vai tornando simétrica em torno do parâmetro  $\lambda$ . As distribuições Birnbaum-Saunders e Birnbaum-Saunders- $t$  estendidas, como se observa nas Figuras 3.5 e 3.6, respectivamente, apresentam casos unimodais e bimodais em suas densidades. Quanto à função de risco, todos os modelos apresentam função de risco crescente e decrescente.



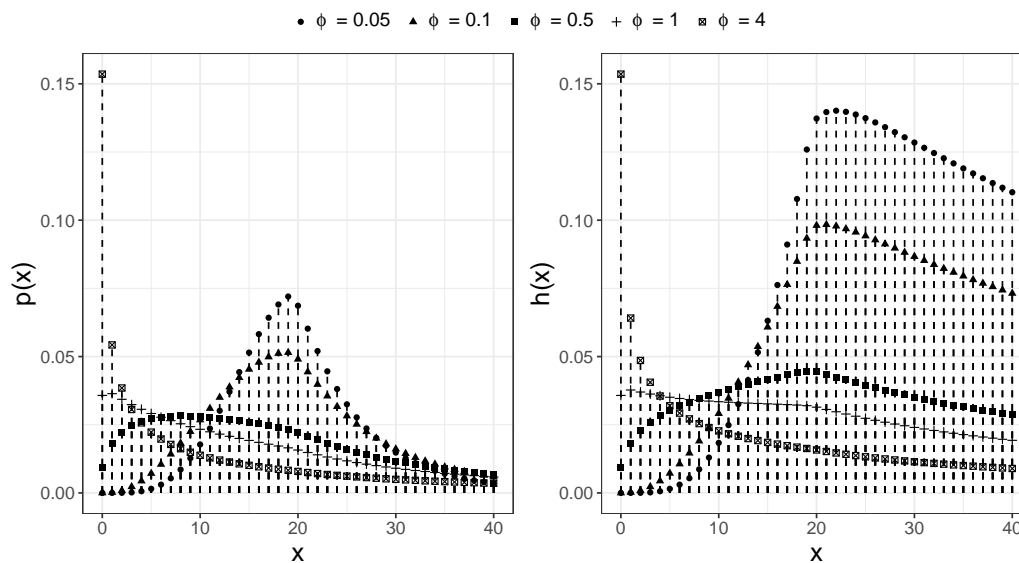
**Figura 3.1:** Comportamento das distribuição log-normal discreta e função de risco para  $\lambda=20$  e  $\phi$ .



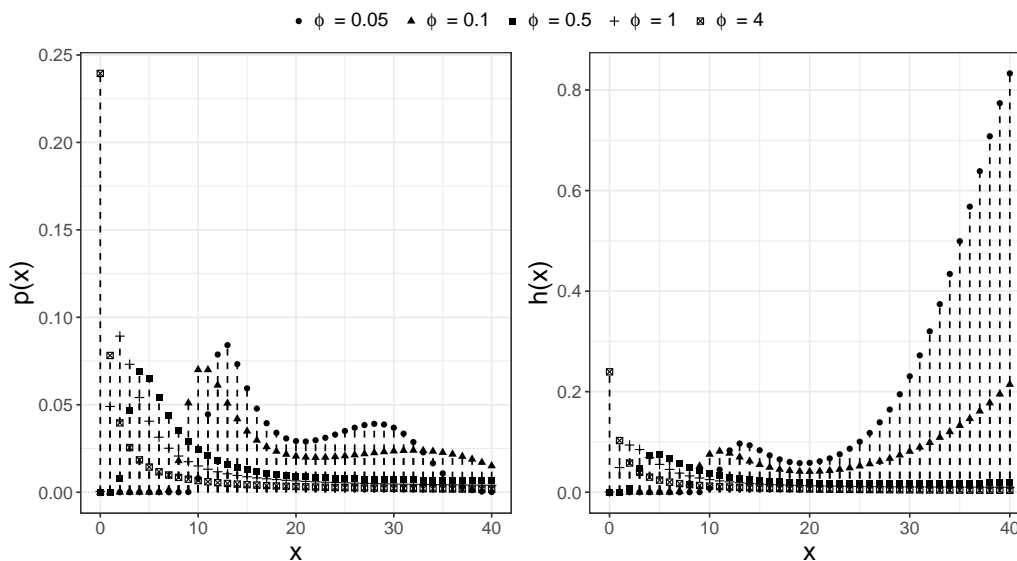
**Figura 3.2:** Comportamento da distribuição log- $t$ -Student discreta e função de risco para  $\lambda=20$ ,  $\xi = 4$  e  $\phi$ .



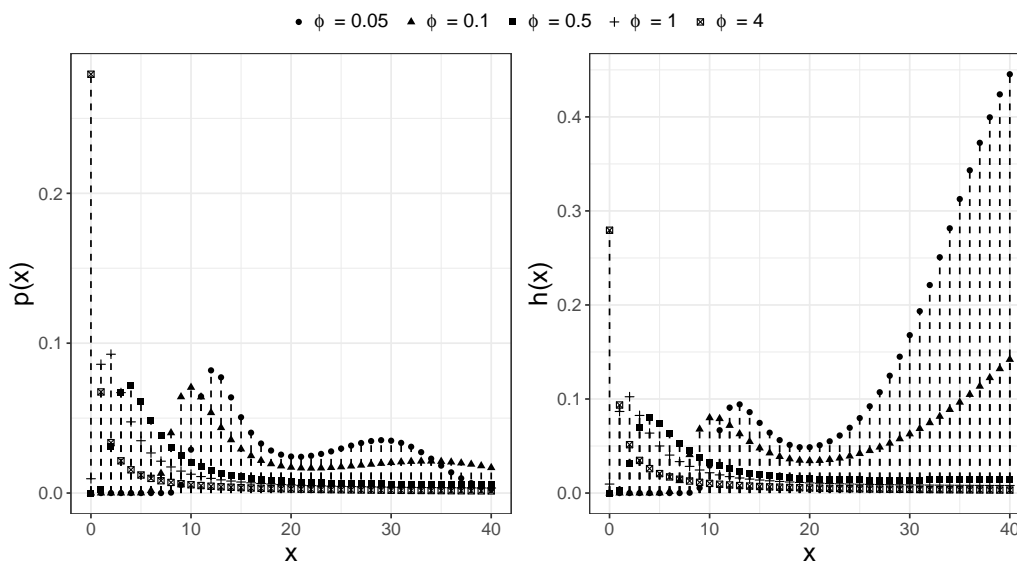
**Figura 3.3:** Comportamento da distribuição log-normal-contaminada discreta e função de risco para  $\lambda=20$ ,  $\xi_1=0.5$ ,  $\xi_2=0.5$  e  $\phi$ .



**Figura 3.4:** Comportamento da distribuição log-exponencial-potência discreta e função de risco para  $\lambda=20$ ,  $\xi=0.5$  e  $\phi$ .



**Figura 3.5:** Comportamento da distribuição Birnbaum-Saunders estendida discreta e função de risco para  $\lambda=20$ ,  $\xi=6$  e  $\phi$ .



**Figura 3.6:** Comportamento da distribuição Birnbaum-Saunders estendida discreta e função de risco para  $\lambda=20$ ,  $\xi_1=7$ ,  $\xi_2=9$  e  $\phi$ .

### 3.2 Propriedades Matemáticas

Esta seção, se não mencionada explicitamente, consiste em propriedades matemáticas válidas para qualquer variável aleatória discreta  $X$  com suporte  $\{0, 1, \dots\}$ .

Seja  $(b_n)$  uma sequência de números reais. Por razões técnicas, no próximo resultado, é usada a convenção  $\prod_{y=0}^{-1} b_y = 1$ . Esse resultado fornece uma caracterização da FMP e da função de sobrevivência (confiabilidade) de uma distribuição discreta em termos da taxa de risco.

**Proposição 1.** Se  $X$  é uma variável aleatória discreta, então, para cada  $x = 0, 1, 2, \dots$ ,

- $p(x|\boldsymbol{\theta}) = \frac{h(x|\boldsymbol{\theta})}{1-h(x|\boldsymbol{\theta})} \prod_{y=0}^{x-1} [1 - h(y|\boldsymbol{\theta})]$ ,
- $S(x|\boldsymbol{\theta}) = \prod_{y=0}^{x-1} [1 - h(y|\boldsymbol{\theta})]$ ,

onde  $h(\cdot|\boldsymbol{\theta})$  é a taxa de risco.

*Demonstração.* Usando a identidade  $p(x|\boldsymbol{\theta}) = S(x|\boldsymbol{\theta}) - S(x+1|\boldsymbol{\theta}) = [p(x|\boldsymbol{\theta}) + S(x|\boldsymbol{\theta})]h(x|\boldsymbol{\theta})$ ,  $x = 0, 1, \dots$ , tem-se

$$1 = h(x|\boldsymbol{\theta}) + \frac{S(x|\boldsymbol{\theta})h(x|\boldsymbol{\theta})}{p(x|\boldsymbol{\theta})}, \quad x = 0, 1, 2, \dots$$

Desde que  $p(x|\boldsymbol{\theta})/h(x|\boldsymbol{\theta}) = p(x|\boldsymbol{\theta}) + S(x|\boldsymbol{\theta}) = S(x-1|\boldsymbol{\theta})$ , segue que

$$\frac{S(x|\boldsymbol{\theta})}{S(x-1|\boldsymbol{\theta})} = 1 - h(x|\boldsymbol{\theta}), \quad x = 0, 1, 2, \dots$$

Trocando  $x$  por  $y$  na identidade acima e multiplicando de  $y = 0$  para  $y = x - 1$ , obtém-se

$$S(x|\boldsymbol{\theta}) = \prod_{y=0}^{x-1} \frac{S(y|\boldsymbol{\theta})}{S(y-1|\boldsymbol{\theta})} = \prod_{y=0}^{x-1} [1 - h(y|\boldsymbol{\theta})], \quad x = 0, 1, 2, \dots,$$

verificando a identidade para  $S(x|\boldsymbol{\theta})$ . Por outro lado, combinando a identidade acima com a definição de taxa de risco, a identidade para  $p(x|\boldsymbol{\theta})$  segue.  $\square$

### 3.2.1 Momentos e variância

**Proposição 2.** Se  $X$  é uma variável aleatória discreta tal que  $\mathbb{E}X^r < \infty$ ,  $r \in \mathbb{N}$ , então

- $\mathbb{E}X^r = \sum_{x=0}^{\infty} [(x+1)^r - x^r] S(x|\boldsymbol{\theta})$ ,



- $\mathbb{E}X^r = \sum_{x=0}^{\infty} \sum_{k=0}^r \sum_{i=0}^{r-k} \binom{r-k}{i} x^{k+i} S(x|\boldsymbol{\theta}),$
- $\text{Var}(X) = 2 \sum_{x=0}^{\infty} xS(x|\boldsymbol{\theta}) + \sum_{x=0}^{\infty} S(x|\boldsymbol{\theta}) [1 - \sum_{x=0}^{\infty} S(x|\boldsymbol{\theta})],$

onde  $S(\cdot|\boldsymbol{\theta})$  é a função de sobrevivência (confiabilidade).

*Demonstração.* Para provar o primeiro item, usando a série telescópica  $\sum_{x=0}^{i-1} [(x+1)^r - x^r] = i^r$ , segue que

$$\begin{aligned} \mathbb{E}X^r &= \sum_{i=0}^{\infty} \sum_{x=0}^{\infty} \mathbb{1}_{\{x < i\}} [(x+1)^r - x^r] p(i|\boldsymbol{\theta}) \\ &= \sum_{x=0}^{\infty} [(x+1)^r - x^r] \sum_{i=0}^{\infty} \mathbb{1}_{\{i > x\}} p(i|\boldsymbol{\theta}), \end{aligned}$$

onde na segunda igualdade foram trocadas as ordens das somas porque

$$\begin{aligned} \sum_{x=0}^{\infty} \mathbb{1}_{\{x < i\}} [(x+1)^r - x^r] p(i|\boldsymbol{\theta}) &= \sum_{x=0}^{\infty} \mathbb{1}_{\{x < i\}} \cdot [(x+1)^r - x^r] p(i|\boldsymbol{\theta}) \\ &= i^r p(i|\boldsymbol{\theta}) =: M_i < \infty, \quad i = 0, 1, \dots, \end{aligned}$$

e  $\sum_{i=0}^{\infty} M_i = \mathbb{E}X^r < \infty$  (por hipótese). Isso prova o primeiro item.

O segundo item segue combinando a expressão para  $\mathbb{E}X^r$  dado no primeiro item com a identidade polinomial  $a^n - b^n = (a - b) \sum_{k=0}^{n-1} a^{n-k} b^k$  e com a expansão binomial.

A prova do terceiro item segue imediatamente do segundo item. Assim, a prova está completa □

Seja  $X$  uma variável aleatória discreta com suporte  $\{0, 1, \dots\}$ . Seguindo Gupta e Gupta (1983), a função de vida média residual associada é dada por:

$$L_x := \mathbb{E}[X - x | X \geq x], \quad x = 0, 1, \dots$$

**Proposição 3.** Se  $X$  é uma variável aleatória discreta tal que  $\sum_{y=x}^{\infty} S(y|\boldsymbol{\theta}) < \infty$  para cada  $x$ ,

então

$$L_x = \frac{\sum_{y=x}^{\infty} S(y|\boldsymbol{\theta})}{S(x-1|\boldsymbol{\theta})}, \quad x = 0, 1, \dots, \quad (3.3)$$

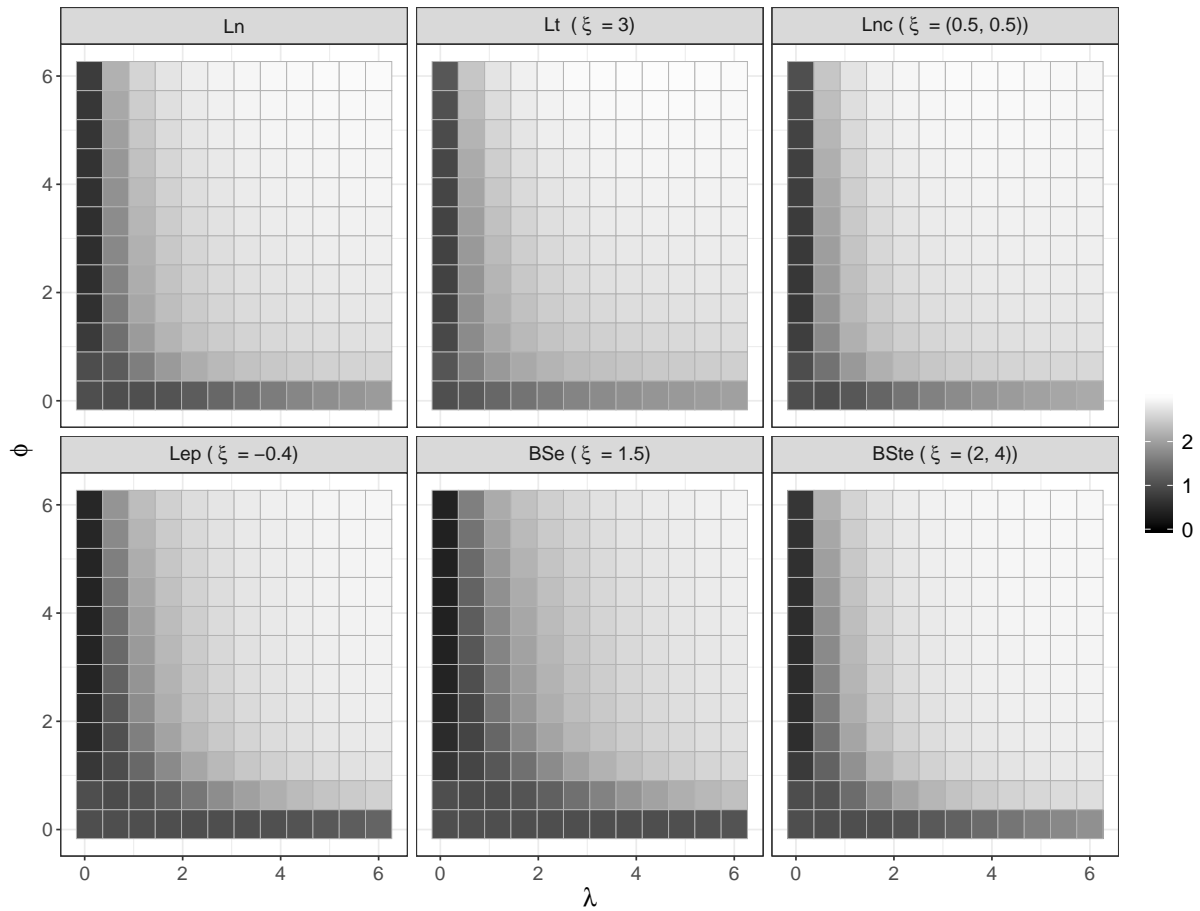
onde  $S(\cdot|\boldsymbol{\theta})$  é a função de sobrevivência (confiabilidade).

*Demonstração.* A prova segue combinando as identidades

$$L_x = \frac{\mathbb{E}[\mathbb{1}_{\{X \geq x\}}(X-x)]}{S(x-1|\boldsymbol{\theta})}, \quad \mathbb{E}[\mathbb{1}_{\{X \geq x\}}(X-x)] = \sum_{y=x}^{\infty} S(y|\boldsymbol{\theta}), \quad x = 0, 1, \dots,$$

porque  $X - x$  é uma variável aleatória discreta não-negativa. □

A Figura 3.7 a seguir mostra a razão entre a variância e a esperança dos modelos log-simétricos discretos para diversos valores de  $\lambda$  e  $\phi$ . Nota-se que esses modelos podem envolver dados com superdispersão e subdispersão.



**Figura 3.7:** Razão entre variância e esperança dos modelos log-simétricos discretos.

### 3.2.2 O $p$ -Quantil

**Proposição 4.** Seja  $X = \lfloor Y \rfloor$  uma variável aleatória discreta obtida de uma variável aleatória contínua positiva  $Y$  com  $p$ -quantil  $Q(p) = F_Y^{-1}(p|\boldsymbol{\theta})$ ,  $p \in (0, 1)$ , onde  $F_Y(\cdot|\boldsymbol{\theta})$  é a FDA de  $Y$  que é estritamente crescente.

- (a) Se  $0 < Q(p) < 1$ , então 0 é um  $p$ -quantil de  $X$ .
- (b) Se  $Q(p) > 1$  e  $Q(p) \notin \mathbb{N}$ , então  $\lfloor Q(p) \rfloor - 1$  é um  $p$ -quantil de  $X$ .
- (c) Se  $Q(p) = 1$ , então todo  $y \in [0, 1]$  é um  $p$ -quantil de  $X$ .
- (d) Se  $Q(p) > 1$  e  $Q(p) \in \mathbb{N}$ , então todo  $y \in [Q(p) - 1, Q(p)]$  é um  $p$ -quantil de  $X$ .

*Demonstração.* Desde que  $X = \lfloor Y \rfloor$ , tem-se que  $\mathbb{P}(X = x|\boldsymbol{\theta}) = F_Y(x + 1|\boldsymbol{\theta}) - F_Y(x|\boldsymbol{\theta})$ ,  $x = 0, 1, \dots$ , com  $F_Y(0|\boldsymbol{\theta}) = 0$ . Então  $F_X(x|\boldsymbol{\theta}) = F_Y(\lfloor x \rfloor + 1|\boldsymbol{\theta})$ .

(a) Se  $0 < Q(p) < 1$  então  $F_X(0|\boldsymbol{\theta}) = F_Y(1|\boldsymbol{\theta}) > p$ . Isto é, 0 é um  $p$ -quantil de  $X$ .

(b) Desde que  $Q(p) > 1$  e  $Q(p) \notin \mathbb{N}$ , tem-se que  $Q(p) - 2 < \lfloor Q(p) \rfloor - 1 < Q(p) - 1$ . Isso implica que  $F_Y(\lfloor Q(p) \rfloor|\boldsymbol{\theta}) < p < F_Y(\lfloor Q(p) \rfloor + 1|\boldsymbol{\theta})$  ou equivalentemente  $F_X(\lfloor Q(p) \rfloor - 1|\boldsymbol{\theta}) < p < F_X(\lfloor Q(p) \rfloor|\boldsymbol{\theta})$ . Então o resultado segue.

(c) A identidade  $Q(p) = 1$  implica que  $F_X(0|\boldsymbol{\theta}) = F_Y(1|\boldsymbol{\theta}) = p$ . Consequentemente, todo  $y \in [0, 1]$  representa um  $p$ -quantil de  $X$ .

(d) Se  $Q(p) > 1$  e  $Q(p) \in \mathbb{N}$ ,  $F_X(Q(p) - 1|\boldsymbol{\theta}) = F_Y(Q(p)|\boldsymbol{\theta}) = p$ . Então todo  $y \in [Q(p) - 1, Q(p)]$  representa um  $p$ -quantil de  $X$ .  $\square$

Os dois resultados a seguir são aplicados exclusivamente a variáveis aleatórias com distribuição log-simétrica discreta.

**Proposição 5.** Seja  $X$  uma variável aleatória seguindo uma distribuição log-simétrica discreta  $\text{LS}_d(\boldsymbol{\theta}, g)$ .

(a) Se  $\lambda < 1$ , então 0 é uma mediana de  $X$ .

(b) Se  $\lambda > 1$  e  $\lambda \notin \mathbb{N}$ , então  $\lfloor \lambda \rfloor - 1$  é uma mediana de  $X$ .

(c) Se  $\lambda = 1$ , então todo  $y \in [0, 1]$  é uma mediana de  $X$ .

(d) Se  $\lambda > 1$  e  $\lambda \in \mathbb{N}$ , então todo  $y \in [\lambda - 1, \lambda]$  é uma mediana de  $X$ .

*Demonstração.* Seja  $X \sim \text{LS}_d(\boldsymbol{\theta}, g)$ . Desde que  $G(\cdot)$  e  $a_\theta(\cdot)$  sejam funções estritamente crescentes, a função  $(G \circ a_\theta)(\cdot)$  é uma FDA estritamente crescente de alguma variável aleatória contínua  $Y$  com distribuição log-simétrica  $\text{LS}(\boldsymbol{\theta}, g)$  definida em (2.16). Além disso, observa-se que a mediana  $Q(1/2)$  para  $Y$  pode ser escrita como:

$$Q(1/2) = (G \circ a_\theta)^{-1}(1/2) = a_\theta^{-1}(G^{-1}(1/2)) = \lambda \exp(\sqrt{\phi} G^{-1}(1/2)) = \lambda,$$

porque  $G^{-1}(1/2) = 0$ . Então, pela Proposição 4, a prova segue.  $\square$

Seja  $X \sim \text{LS}_d(\boldsymbol{\theta}, g)$  e  $Q_d(p)$ ,  $p \in (0, 1)$ , seja o  $p$ -quantil de  $X$ . As seguintes quantidades são definidas:

Dispersão:  $\zeta = Q_d(0.75) - Q_d(0.25)$ ,  $0 < \zeta < \infty$ ;

Dispersão relativa:  $\varpi = \frac{\zeta}{\zeta + 2Q_d(0.25)}$ ,  $0 < \varpi < 1$ ;

Assimetria:  $\varkappa(p) = \frac{Q_d(p) + Q_d(1-p) - 2Q_d(1/2)}{Q_d(1-p) + Q_d(p)}$ ,  $0 < \varkappa(p) < 1$ ,  $0 < p < 1/2$ ;

Curtose:  $\varsigma = \frac{Q_d(7/8) - Q_d(5/8) + Q_d(3/8) - Q_d(1/8)}{Q_d(6/8) - Q_d(2/8)}$ ,  $0 \leq \varsigma < \infty$ .

A dispersão relativa, assimetria e curtose apareceram em Zwillinger e Kokoska (2000), Hinkley (1975) e Moors (1988), respectivamente.

**Proposição 6.** Se  $X \sim \text{LS}_d(\boldsymbol{\theta}, g)$ ,  $Q(p) > 1$  e  $Q(p) \in \mathbb{N}$ ,  $p \in (0, 1)$ , onde  $Q(p)$  é o  $p$ -quantil de uma distribuição log-simétrica contínua, então

- $\zeta = 2\lambda \sinh(\sqrt{\phi} G^{-1}(0.75))$ ;
- $\varpi = [\text{cotanh}(\sqrt{\phi} G^{-1}(0.75)) - \text{cosech}(\sqrt{\phi} G^{-1}(0.75))]^{-1}$ ;
- $\varkappa(p) = \text{cosech}(\sqrt{\phi} G^{-1}(p)) - \text{cotanh}(\sqrt{\phi} G^{-1}(p))$ ;
- $\varsigma = \frac{\sinh(\sqrt{\phi} G^{-1}(7/8)) - \sinh(\sqrt{\phi} G^{-1}(5/8))}{\sinh(\sqrt{\phi} G^{-1}(6/8))}$ ;

onde  $G(\cdot)$  foi definido em (2.17).

*Demonstração.* Para  $Q(p) > 1$  e  $Q(p) \in \mathbb{N}$ , pela Proposição 4, todo  $y \in [Q(p) - 1, Q(p)]$  é o  $p$ -quantil de  $X$ , onde  $Q(p) = \lambda \exp(\sqrt{\phi} G^{-1}(p))$ . Considerando  $Q_d(p) \in \{Q(p) - 1, Q(p)\}$ , usando a identidade  $G^{-1}(1-p) = -G^{-1}(p)$  e a Proposição 5, um simples cálculo algébrico completa a prova. □

### 3.3 Estimadores de Máxima Verossimilhança

#### 3.3.1 Dados não censurados

Seja  $(X_1, \dots, X_n)$  uma amostra aleatória do tamanho  $n$  de uma variável aleatória  $X$  com FDA dado por (3.2) e  $\mathbf{x} = (x_1, \dots, x_n)$  suas observações (dados). Então, a função de log-verossimilhança para o vetor de parâmetros  $\boldsymbol{\theta} = (\lambda, \phi)$  é dado por

$$\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^n \log p(x_i|\boldsymbol{\theta}) = \sum_{i=1}^n \log\{G(a_{\boldsymbol{\theta}}(x_i + 1)) - G(a_{\boldsymbol{\theta}}(x_i))\}. \quad (3.4)$$

As raízes do sistema formado pelas derivadas parciais da função de log-verossimilhança  $\ell(\boldsymbol{\theta})$  em relação a  $\lambda$  e  $\phi$  são as estimativas desses parâmetros, respectivamente. Portanto, deve-se resolver o seguinte sistema de equações:

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta} = (Z_g)^{-1} \sum_{i=1}^n \sum_{j=0}^1 (-1)^{j+1} \frac{g(a_{\boldsymbol{\theta}}^2(x_i + j))}{p(x_i|\boldsymbol{\theta})} \frac{\partial a_{\boldsymbol{\theta}}(x_i + j)}{\partial \theta} = 0, \quad \theta \in \{\lambda, \phi\},$$

onde  $Z_g = \int_{-\infty}^{\infty} g(w^2)dw$  e,

$$\frac{\partial a_{\boldsymbol{\theta}}(x)}{\partial \lambda} = -(\lambda\phi^{1/2})^{-1}, \quad \frac{\partial a_{\boldsymbol{\theta}}(x)}{\partial \phi} = \log\left(\frac{x}{\lambda}\right)^{1/(2\phi^{3/2})}, \quad x > 0. \quad (3.5)$$

Observa-se que eles devem ser resolvidos por um procedimento iterativo para otimização não linear, como o método quase-Newton de Broyden-Fletcher-Goldfarb-Shanno (BFGS) (ver MITTELHAMMER et al., 2000, p. 199).

A inferência para  $\boldsymbol{\theta}$  do modelo  $LS_d$  pode ser baseada na distribuição assintótica do estimador de máxima verossimilhança  $\hat{\boldsymbol{\theta}}$ . Sob condições clássicas de regularidade, esse estimador tem distribuição normal bivariada com média  $\boldsymbol{\theta}$  e matriz de covariância  $\Sigma_{\hat{\boldsymbol{\theta}}}$ , ou seja,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{D} N_2(\mathbf{0}, \Sigma_{\hat{\boldsymbol{\theta}}} = \mathcal{J}(\boldsymbol{\theta})^{-1}),$$

como  $n \rightarrow \infty$ , onde  $\xrightarrow{\mathcal{D}}$  significa “convergência em distribuição”,  $\mathcal{I}(\boldsymbol{\theta})$  é a matriz de informações Fisher esperada, e  $\mathcal{J}(\boldsymbol{\theta}) = \lim_{n \rightarrow \infty} (1/n)[\mathcal{I}(\boldsymbol{\theta})]$ . Observe que  $[\widehat{\mathcal{I}}(\boldsymbol{\theta})]^{-1}$  é um estimador consistente da matriz de covariância assintótica de  $\widehat{\boldsymbol{\theta}}$ . Observa-se também que é possível usar a matriz Hessiana para obter a versão observada da matriz de informação de Fisher esperada.

A matriz Hessiana de  $\ell(\boldsymbol{\theta})$  é dada por  $(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta \partial \theta'})_{\theta, \theta' \in \{\lambda, \phi\}}$ , onde seus elementos são dados por

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta \partial \theta'} = (Z_g)^{-1} \sum_{i=1}^n \sum_{j=0}^1 (-1)^{j+1} \frac{g(a^2(x_i + j))}{p(x_i | \boldsymbol{\theta})} \left[ \frac{\partial^2 a_{\boldsymbol{\theta}}(x_i + j)}{\partial \theta \partial \theta'} + \Theta_j(x_i) - \Upsilon_j(x_i) \right],$$

para cada  $\theta, \theta' \in \{\lambda, \phi\}$ . Aqui,

$$\Theta_j(x) = 2a_{\boldsymbol{\theta}}(x + j) g'(a^2(x + j)) \frac{\partial a_{\boldsymbol{\theta}}(x+j)}{\partial \theta} \frac{\partial a_{\boldsymbol{\theta}}(x+j)}{\partial \theta'}, \quad (3.6)$$

$$\Upsilon_j(x) = (Z_g)^{-1} \frac{\partial a_{\boldsymbol{\theta}}(x+j)}{\partial \theta} \sum_{k=0}^1 (-1)^{k+1} \frac{g(a^2(x+k))}{p(x | \boldsymbol{\theta})} \frac{\partial a_{\boldsymbol{\theta}}(x+k)}{\partial \theta'}, \quad (3.7)$$

sempre que a função geradora de densidade  $g$  é diferenciável. Aqui, as derivadas parciais de segunda ordem de  $a_{\boldsymbol{\theta}}(\cdot)$ , com relação aos parâmetros, são dadas por

$$\frac{\partial^2 a_{\boldsymbol{\theta}}(x)}{\partial \lambda^2} = (\lambda^2 \phi^{1/2})^{-1}, \quad \frac{\partial^2 a_{\boldsymbol{\theta}}(x)}{\partial \phi^2} = \log\left(\frac{y}{\lambda}\right)^{-3/(4\phi^{5/2})}, \quad \frac{\partial a_{\boldsymbol{\theta}}(x)}{\partial \lambda \partial \phi} = \frac{\partial a_{\boldsymbol{\theta}}(x)}{\partial \phi \partial \lambda} = (2\lambda \phi^{3/2})^{-1}, \quad x > 0. \quad (3.8)$$

A matriz de informação de Fisher  $(\mathbb{E}[-\frac{\partial^2 \log p(X|\boldsymbol{\theta})}{\partial \theta \partial \theta'}])_{\theta, \theta' \in \{\lambda, \phi\}}$ , sob certas condições de regularidade, onde  $X \sim \text{LS}_d(\boldsymbol{\theta}, g)$ , possui elementos da seguinte forma

$$\mathbb{E}[-\frac{\partial^2 \log p(X|\boldsymbol{\theta})}{\partial \theta \partial \theta'}] = (Z_g)^{-1} \sum_{j=0}^1 (-1)^j \sum_{x=0}^{\infty} g(a^2(x + j)) \left[ \frac{\partial^2 a_{\boldsymbol{\theta}}(x+j)}{\partial \theta \partial \theta'} + \Theta_j(x) - \Upsilon_j(x) \right],$$

para cada  $\theta, \theta' \in \{\lambda, \phi\}$ , onde  $\Theta_j(\cdot)$  e  $\Upsilon_j(\cdot)$  são como em (3.6) e (3.7), respectivamente, e sempre que a série acima convergir absolutamente.

O parâmetro extra  $\xi$  (ou vetor de parâmetro  $\boldsymbol{\xi}$ ) associado a  $g$  é selecionado usando a função log-verossimilhança perfilada. Por exemplo, no caso da distribuição discreta log- $t$ -Student, são

necessárias duas etapas:

- i) Seja  $\xi_k = k$  e para cada  $k = 1, \dots, 100$  é calculada a  $k$ -ésima estimativa de máxima verossimilhança de  $\boldsymbol{\theta}_k = (\lambda_k, \phi_k)^\top$ ,  $\hat{\boldsymbol{\theta}}_k = (\hat{\lambda}_k, \hat{\phi}_k)^\top$ . É calculado também o valor da  $k$ -ésima estimativa de log-verossimilhança  $\ell_k(\hat{\boldsymbol{\theta}}_k)$ ;
- ii) A estimativa final de  $\xi$ ,  $\hat{\xi} = \xi_k$ , é aquela que maximiza a função de log-verossimilhança, isto é,  $\hat{\xi} \in \{\operatorname{argmax}_{\xi_k} \ell_k(\hat{\boldsymbol{\theta}}_k)\}$ , e a estimativa de  $\boldsymbol{\theta}$  é  $\hat{\boldsymbol{\theta}} = (\hat{\lambda}_k, \hat{\phi}_k)^\top$ .

### 3.3.2 Dados censurados

Seja  $Y_i \sim \text{LS}(\boldsymbol{\theta}, g)$  o tempo de falha do  $i$ -ésimo indivíduo e  $\delta_i$  o indicador de censura ou não censura para o  $i$ -ésimo indivíduo.

Define-se  $d_k =$  “número de falhas no tempo  $t_k$ ”,  $q_k =$  “número de censuras no tempo  $t_k$ ” e  $N_k = \sum_{i=k}^{\infty} (d_i + q_i)$ . Note que  $N_k - d_k$  representa o número sobrevivido imediatamente antes do tempo  $t_k + 1$ . Ou seja, em cada tempo especificado  $t_k$ , tem-se  $d_k$  falhas e  $N_k - d_k$  sobrevivências.

Como os dados são discretos, observando  $\{(Y_i, \delta_i)\}$  equivale a observar  $\{(d_k, q_k)\}$ , a função de probabilidade para a censura aleatória é dada por

$$L^S(\boldsymbol{\theta}) = \prod_{i=1}^n [f_Y(y_i|\boldsymbol{\theta})]^{\delta_i} [1 - F_Y(y_i|\boldsymbol{\theta})]^{1-\delta_i} = \prod_{k=1}^{\infty} [p(x_k|\boldsymbol{\theta})]^{d_k} [p(x_k|\boldsymbol{\theta}) + S(x_k|\boldsymbol{\theta})]^{q_k}.$$

Esse tipo de censura tem como caso especial os tipos I e II. A correspondente função de log-verossimilhança é

$$\begin{aligned} \ell^S(\boldsymbol{\theta}) &= \sum_{k=1}^{\infty} [d_k \log p(x_k|\boldsymbol{\theta}) + q_k \log \{p(x_k|\boldsymbol{\theta}) + S(x_k|\boldsymbol{\theta})\}] \\ &= \sum_{k=1}^{\infty} [d_k \log \{G(a_{\boldsymbol{\theta}}(x_k + 1)) - G(a_{\boldsymbol{\theta}}(x_k))\} + q_k \log \{1 - G(a_{\boldsymbol{\theta}}(x_k))\}]. \end{aligned} \quad (3.9)$$



**Observação 1.** Pela Proposição 1, a log-verossimilhança (3.9) pode ser reescrito em termos da taxa de risco como

$$\ell^S(\boldsymbol{\theta}) = \sum_{k=1}^{\infty} (d_k + q_k) \left[ \frac{d_k}{d_k + q_k} \log h(x_k | \boldsymbol{\theta}) - \log(1 - h(x_k | \boldsymbol{\theta})) + \sum_{y=0}^{x_k-1} \log(1 - h(y | \boldsymbol{\theta})) \right],$$

sempre que a série acima convergir absolutamente.

Para cada  $\theta \in \{\lambda, \phi\}$ , derivado de (3.9), um cálculo direto mostra que

$$\begin{aligned} \frac{\partial \ell^S(\boldsymbol{\theta})}{\partial \theta} = (Z_g)^{-1} \sum_{k=1}^{\infty} \left[ d_k \sum_{j=0}^1 (-1)^{j+1} \frac{g(a_{\boldsymbol{\theta}}^2(x_k + j))}{p(x_k | \boldsymbol{\theta})} \frac{\partial a_{\boldsymbol{\theta}}(x_k + j)}{\partial \theta} \right. \\ \left. - q_k \frac{g(a_{\boldsymbol{\theta}}^2(x_k))}{p(x_k | \boldsymbol{\theta}) + S(x_k | \boldsymbol{\theta})} \frac{\partial a_{\boldsymbol{\theta}}(x_k)}{\partial \theta} \right], \end{aligned}$$

onde  $Z_g = \int_{-\infty}^{\infty} g(w^2) dw$ . As derivadas parciais mistas de  $\ell^S(\boldsymbol{\theta})$  são dadas por

$$\begin{aligned} \frac{\partial^2 \ell^S(\boldsymbol{\theta})}{\partial \theta \partial \theta'} = (Z_g)^{-1} \sum_{k=1}^{\infty} \left\{ d_k \sum_{j=0}^1 (-1)^{j+1} \frac{g(a_{\boldsymbol{\theta}}^2(x_k + j))}{p(x_k | \boldsymbol{\theta})} \left[ \frac{\partial^2 a_{\boldsymbol{\theta}}(x_k + j)}{\partial \theta \partial \theta'} + \Theta_j(x_k) - \Upsilon_j(x_k) \right] \right. \\ \left. - q_k \frac{g(a_{\boldsymbol{\theta}}^2(x_k))}{p(x_k | \boldsymbol{\theta}) + S(x_k | \boldsymbol{\theta})} \left[ \frac{\partial^2 a_{\boldsymbol{\theta}}(x_k)}{\partial \theta \partial \theta'} + \Theta_0(x_k) + \hat{\Upsilon}(x_k) \right] \right\}, \end{aligned}$$

para cada  $\theta, \theta' \in \{\lambda, \phi\}$ , onde

$$\hat{\Upsilon}(x) = (Z_g)^{-1} \frac{g(a^2(x))}{p(x|\boldsymbol{\theta})+S(x|\boldsymbol{\theta})} \frac{\partial a_{\boldsymbol{\theta}}(x)}{\partial \theta} \frac{\partial a_{\boldsymbol{\theta}}(x)}{\partial \theta'}, \quad x > 0,$$

e  $\Theta_j(\cdot), \Upsilon_j(\cdot)$  são com em (3.6) e (3.7), respectivamente. As primeira e segunda derivadas da função  $a_{\boldsymbol{\theta}}(\cdot)$  são dadas em (3.5) e (3.8), respectivamente.

### 3.4 Simulações

Um estudo de simulação de Monte Carlo foi realizado para avaliar o desempenho dos estimadores de máxima verossimilhança para os modelos  $LS_d$ , particularmente os casos da log-

normal, log- $t$ -Student, log-normal-contaminada, log-exponencial-potência, Birnbaum-Saunders estendida e Birnbaum-Saunders- $t$  estendida. As estimativas de máxima verossimilhança foram obtidas por meio da função `optim` aplicando o método de otimização BFGS. Todas as avaliações numéricas foram feitas no software R; ver R Core Team (2016).

O cenário de simulação considera: tamanho da amostra  $n \in \{40, 120, 400\}$ , valores verdadeiros dos parâmetros  $\phi \in \{1, 4, 8\}$  e  $\lambda \in \{2\}$ , e 1000 réplicas de Monte Carlo para cada tamanho de amostra. Os valores reais dos parâmetros extras são apresentados na legenda de cada tabela.

Para a obtenção dos valores da média, do viés e o do erro quadrático médio (EQM), os seguintes passos foram seguidos:

- (1) Foram geradas 1000 amostras de tamanho  $n$  de cada distribuição considerada, com os percentuais de censura de 0%, 10% e 30%. Para gerar as amostras censuradas, considerou-se que a distribuição do tempo do evento segue uma  $LS_d$  e a distribuição do tempo de censura segue uma  $cLS_d$ , em que  $c$  é uma constante positiva. Assim, foram geradas  $n$  observações para os tempos do evento e  $n$  observações para os tempos de censura. Para cada uma das  $n$  observações o mínimo desses tempos é observado ( $Y$ ) e então são gerados  $n$  pares de observações  $(Y, \delta)$ , em que  $\delta = 1$  se  $Y$  é igual ao tempo do evento e  $\delta = 0$  se  $Y$  é igual ao tempo de censura. Os percentuais de censura desejados foram obtidos modificando o valor da constante  $c$  da distribuição dos tempos de censura.
- (2) Foram calculadas as estimativas de máxima verossimilhança para as 1000 amostras geradas no passo (1), para cada distribuição.
- (3) Foram calculadas, para cada tamanho de amostra ( $n$ ), as seguintes medidas: média, viés e EQM. Essas medidas são obtidas conforme as seguintes equações:

$$\text{média}(\hat{\theta}) = \frac{1}{1000} \sum_{i=1}^{1000} \hat{\theta}_i,$$

$$\text{viés} = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\theta}_i - \theta)$$

e

$$\text{EQM} = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\theta}_i - \theta)^2,$$

com  $\theta$  representando os parâmetros  $\lambda$  e  $\phi$ .

Esses passos foram repetidos para cada tamanho de amostra  $n \in \{40, 120, 400\}$  combinado com cada  $\phi \in \{1, 4, 8\}$ .

Os resultados da estimativa de máxima verossimilhança estão apresentados nas Tabelas 3.1–3.6. As seguintes estatísticas da amostra para as estimativas de máxima verossimilhança são relatadas: média empírica, viés e EQM. Observando os resultados nas Tabelas 3.1–3.6, é possível concluir que, à medida que o tamanho da amostra aumenta, o viés e o EQM de todos os estimadores diminuem, indicando que eles são assintoticamente não-viesados, como esperado. Além disso, pelas Figuras 3.8-3.11, nota-se que o EQM aumenta quando o percentual de censura na amostra aumenta. Em geral, todos esses resultados mostram um bom desempenho dos modelos propostos.

**Tabela 3.1:** Valores de média, viés e EQM para dados simulados da distribuição log-normal discreta considerando os estimadores de máxima verossimilhança e  $\lambda = 2$ .

n	Cen.		$\phi = 1$			$\phi = 4$			$\phi = 8$		
			Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
40	0%	$\hat{\phi}$	1.0119	0.0119	0.0851	4.0752	0.0752	1.7247	8.1388	0.1388	7.5728
		$\hat{\lambda}$	2.0262	0.0262	0.1166	2.1126	0.1126	0.5629	2.2586	0.2586	1.4043
120	0%	$\hat{\phi}$	1.0059	0.0059	0.0260	4.0413	0.0413	0.5237	8.0768	0.0768	2.1683
		$\hat{\lambda}$	2.0086	0.0086	0.0394	2.0317	0.0317	0.1759	2.0750	0.0750	0.3897
400	0%	$\hat{\phi}$	1.0041	0.0041	0.0077	4.0184	0.0184	0.1522	8.0404	0.0404	0.6641
		$\hat{\lambda}$	2.0021	0.0021	0.0110	2.0086	0.0086	0.0464	2.0198	0.0198	0.1020
40	10%	$\hat{\phi}$	1.0189	0.0189	0.0998	4.1507	0.1507	2.2755	8.3002	0.3002	10.0324
		$\hat{\lambda}$	2.0301	0.0301	0.1198	2.1154	0.1154	0.5850	2.2528	0.2528	1.4323
120	10%	$\hat{\phi}$	1.0114	0.0114	0.0290	4.0519	0.0519	0.5978	8.1162	0.1162	2.5822
		$\hat{\lambda}$	2.0099	0.0099	0.0396	2.0416	0.0416	0.1795	2.0842	0.0842	0.4037
400	10%	$\hat{\phi}$	1.0015	0.0015	0.0090	4.0262	0.0262	0.1817	8.0675	0.0675	0.8152
		$\hat{\lambda}$	2.0012	0.0012	0.0109	2.0014	0.0014	0.0466	2.0073	0.0073	0.0984
40	30%	$\hat{\phi}$	1.0562	0.0562	0.1839	4.4113	0.4113	5.7506	8.9965	0.9965	28.9825
		$\hat{\lambda}$	2.0435	0.0435	0.1407	2.1396	0.1396	0.6858	2.2925	0.2925	1.7228
120	30%	$\hat{\phi}$	1.0218	0.0218	0.0458	4.1116	0.1116	1.1065	8.2539	0.2539	5.0762
		$\hat{\lambda}$	2.0124	0.0124	0.0438	2.0450	0.0450	0.1915	2.0868	0.0868	0.4270
400	30%	$\hat{\phi}$	1.0093	0.0093	0.0133	4.0840	0.0840	0.3158	8.1961	0.1961	1.5102
		$\hat{\lambda}$	2.0046	0.0046	0.0119	2.0076	0.0076	0.0501	2.0131	0.0131	0.1043

**Tabela 3.2:** Valores de média, viés e EQM para dados simulados da distribuição log-*t*-Student discreta considerando os estimadores de máxima verossimilhança,  $\xi = 4$  e  $\lambda = 2$ .

n	Cen.		$\phi = 1$			$\phi = 4$			$\phi = 8$		
			Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
40	0%	$\hat{\phi}$	1.0250	0.0250	0.1313	4.1432	0.1432	2.6039	8.3035	0.3035	14.7992
		$\hat{\lambda}$	2.0351	0.0351	0.1589	2.1416	0.1416	0.7463	2.3098	0.3098	1.9332
120	0%	$\hat{\phi}$	1.0093	0.0093	0.0408	4.0316	0.0316	0.8793	7.9743	-0.0257	4.0502
		$\hat{\lambda}$	2.0179	0.0179	0.0497	2.0657	0.0657	0.2195	2.1694	0.1694	0.7087
400	0%	$\hat{\phi}$	0.9974	-0.0026	0.0123	3.9778	-0.0222	0.4006	7.9731	-0.0269	2.8371
		$\hat{\lambda}$	1.9986	-0.0014	0.0137	2.0148	0.0148	0.0684	2.0657	0.0657	0.2455
40	10%	$\hat{\phi}$	1.0332	0.0332	0.1499	4.2117	0.2117	3.2204	8.5131	0.5131	14.0700
		$\hat{\lambda}$	2.0318	0.0318	0.1692	2.1338	0.1338	0.8105	2.2965	0.2965	2.1397
120	10%	$\hat{\phi}$	1.0149	0.0149	0.0466	4.0564	0.0564	0.9162	8.1134	0.1134	4.0340
		$\hat{\lambda}$	2.0068	0.0068	0.0510	2.0434	0.0434	0.2086	2.0916	0.0916	0.4621
400	10%	$\hat{\phi}$	1.0026	0.0026	0.0126	4.0285	0.0285	0.2729	8.0335	0.0335	1.1661
		$\hat{\lambda}$	1.9987	-0.0013	0.0146	2.0000	0.0000	0.0598	2.0137	0.0137	0.1208
40	30%	$\hat{\phi}$	1.0772	0.0772	0.2681	4.5456	0.5456	8.5155	9.3122	1.3122	41.0324
		$\hat{\lambda}$	2.0430	0.0430	0.2007	2.1605	0.1605	1.0059	2.3570	0.3570	3.1707
120	30%	$\hat{\phi}$	1.0148	0.0148	0.0666	4.0801	0.0801	1.4871	8.1426	0.1426	6.8379
		$\hat{\lambda}$	2.0032	0.0032	0.0534	2.0429	0.0429	0.2210	2.0866	0.0866	0.4900
400	30%	$\hat{\phi}$	1.0129	0.0129	0.0168	4.0890	0.0890	0.4097	8.1679	0.1679	1.8873
		$\hat{\lambda}$	2.0021	0.0021	0.0160	2.0040	0.0040	0.0645	2.0179	0.0179	0.1297

**Tabela 3.3:** Valores de média, viés e EQM para dados simulados da distribuição log-normal-contaminada discreta considerando os estimadores de máxima verossimilhança,  $\lambda = 2$  e  $\xi = (0.5, 0.5)^\top$ .

$n$	Cen.		$\phi = 1$			$\phi = 4$			$\phi = 8$		
			Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
40	0%	$\hat{\phi}$	1.0057	0.0057	0.0947	4.0459	0.0459	1.8954	8.0944	0.0944	7.9832
		$\hat{\lambda}$	2.0437	0.0437	0.1706	2.1750	0.1750	0.8474	2.3687	0.3687	2.1616
120		$\hat{\phi}$	1.0024	0.0024	0.0300	4.0207	0.0207	0.6042	8.0165	0.0165	2.5295
		$\hat{\lambda}$	2.0161	0.0161	0.0545	2.0589	0.0589	0.2503	2.1297	0.1297	0.5661
400		$\hat{\phi}$	1.0020	0.0020	0.0085	4.0061	0.0061	0.1779	7.9912	-0.0088	0.8225
		$\hat{\lambda}$	2.0024	0.0024	0.0165	2.0161	0.0161	0.0745	2.0410	0.0410	0.1751
40	10%	$\hat{\phi}$	1.0336	0.0336	0.1484	4.1971	0.1971	3.1525	8.4153	0.4153	12.8221
		$\hat{\lambda}$	2.0562	0.0562	0.1819	2.1952	0.1952	0.9257	2.3944	0.3944	2.3858
120		$\hat{\phi}$	1.0154	0.0154	0.0396	4.0676	0.0676	0.7806	8.1552	0.1552	3.3793
		$\hat{\lambda}$	2.0182	0.0182	0.0588	2.0688	0.0688	0.2649	2.1338	0.1338	0.5975
400		$\hat{\phi}$	1.0067	0.0067	0.0103	4.0278	0.0278	0.2113	8.0581	0.0581	0.9135
		$\hat{\lambda}$	1.9988	-0.0012	0.0163	2.0077	0.0077	0.0735	2.0222	0.0222	0.1586
40	30%	$\hat{\phi}$	1.0644	0.0644	0.2448	4.4774	0.4774	7.4371	9.1280	1.1280	33.9336
		$\hat{\lambda}$	2.0635	0.0635	0.2088	2.2106	0.2106	1.0845	2.4422	0.4422	3.1094
120		$\hat{\phi}$	1.0310	0.0310	0.0591	4.1681	0.1681	1.3522	8.4147	0.4147	6.3718
		$\hat{\lambda}$	2.0237	0.0237	0.0635	2.0767	0.0767	0.2895	2.1436	0.1436	0.6536
400		$\hat{\phi}$	1.0126	0.0126	0.0161	4.0609	0.0609	0.3648	8.1301	0.1301	1.6252
		$\hat{\lambda}$	2.0005	0.0005	0.0176	2.0102	0.0102	0.0771	2.0237	0.0237	0.1628

**Tabela 3.4:** Valores de média, viés e EQM para dados simulados da distribuição log-exponencial-potência discreta considerando os estimadores de máxima verossimilhança,  $\lambda = 2$  e  $\xi = -0.5$ .

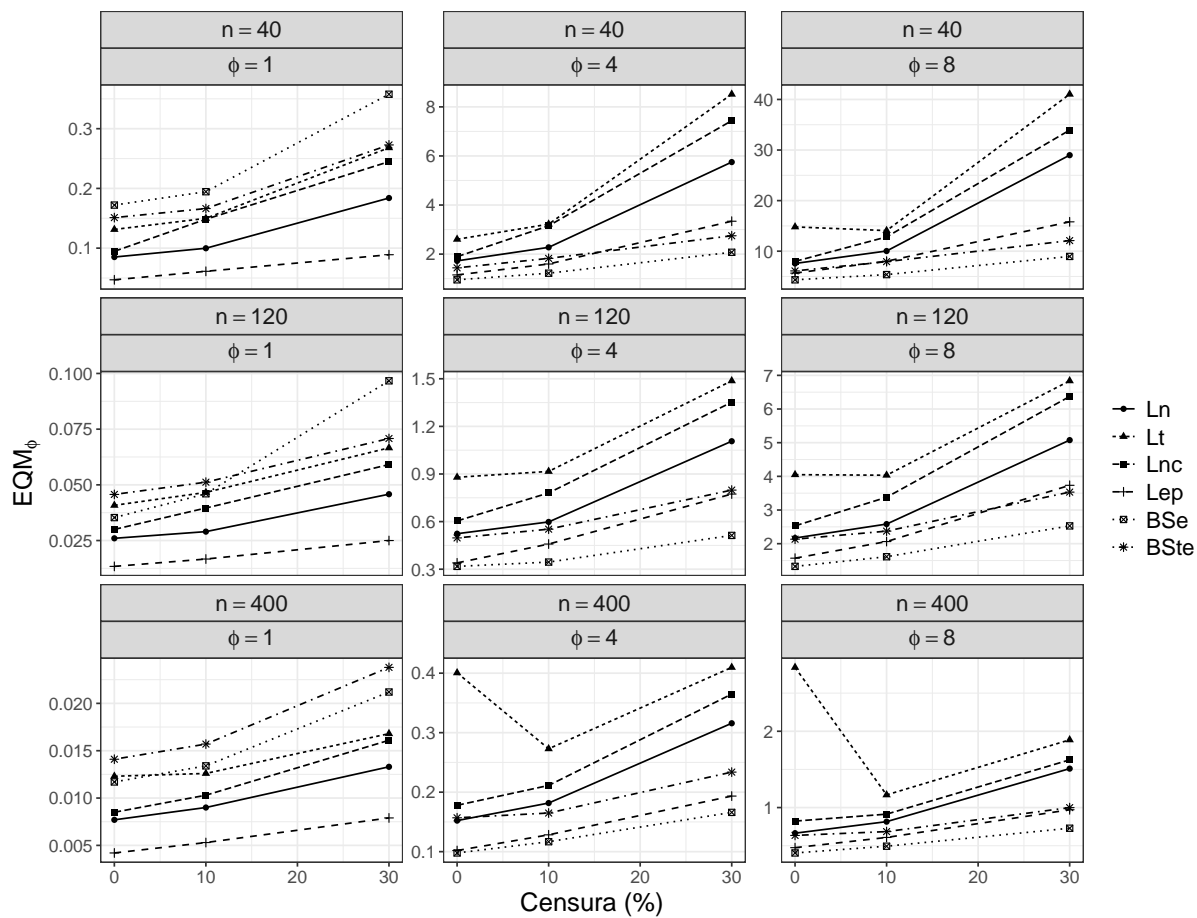
$n$	Cen.		$\phi = 1$			$\phi = 4$			$\phi = 8$		
			Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
40	0%	$\hat{\phi}$	0.9783	-0.0217	0.0470	3.9497	-0.0503	1.1520	7.9186	-0.0814	5.6648
		$\hat{\lambda}$	2.0072	0.0072	0.0481	2.0493	0.0493	0.2856	2.1332	0.1332	0.6806
120		$\hat{\phi}$	0.9970	-0.0030	0.0134	4.0014	0.0014	0.3399	7.9966	-0.0034	1.5695
		$\hat{\lambda}$	2.0032	0.0032	0.0153	2.0185	0.0185	0.0863	2.0482	0.0482	0.1996
400		$\hat{\phi}$	0.9969	-0.0031	0.0042	3.9906	-0.0094	0.1019	7.9864	-0.0136	0.4742
		$\hat{\lambda}$	1.9998	-0.0002	0.0047	2.0036	0.0036	0.0282	2.0098	0.0098	0.0640
40	10%	$\hat{\phi}$	0.9941	-0.0059	0.0612	4.0123	0.0123	1.5967	8.0169	0.0169	8.0172
		$\hat{\lambda}$	2.0046	0.0046	0.0510	2.0472	0.0472	0.2900	2.1398	0.1398	0.6989
120		$\hat{\phi}$	0.9999	-0.0001	0.0167	4.0151	0.0151	0.4585	8.0260	0.0260	2.0633
		$\hat{\lambda}$	2.0010	0.0010	0.0165	2.0099	0.0099	0.0866	2.0349	0.0349	0.1968
400		$\hat{\phi}$	1.0020	0.0020	0.0053	4.0178	0.0178	0.1283	8.0441	0.0441	0.6083
		$\hat{\lambda}$	2.0010	0.0010	0.0050	2.0044	0.0044	0.0298	2.0084	0.0084	0.0661
40	30%	$\hat{\phi}$	1.0093	0.0093	0.0890	4.1602	0.1602	3.3402	8.2982	0.2982	15.8037
		$\hat{\lambda}$	2.0111	0.0111	0.0646	2.0512	0.0512	0.3185	2.1408	0.1408	0.7509
120		$\hat{\phi}$	1.0040	0.0040	0.0250	4.0485	0.0485	0.7736	8.1109	0.1109	3.7337
		$\hat{\lambda}$	2.0019	0.0019	0.0191	2.0104	0.0104	0.0927	2.0350	0.0350	0.2090
400		$\hat{\phi}$	1.0026	0.0026	0.0079	4.0215	0.0215	0.1934	8.0665	0.0665	0.9711
		$\hat{\lambda}$	2.0010	0.0010	0.0056	2.0045	0.0045	0.0318	2.0093	0.0093	0.0694

**Tabela 3.5:** Valores de média, viés e EQM para dados simulados da distribuição Birnbaum-Saunders estendida discreta considerando os estimadores de máxima verossimilhança,  $\lambda = 2$  e  $\xi = 0.5$ .

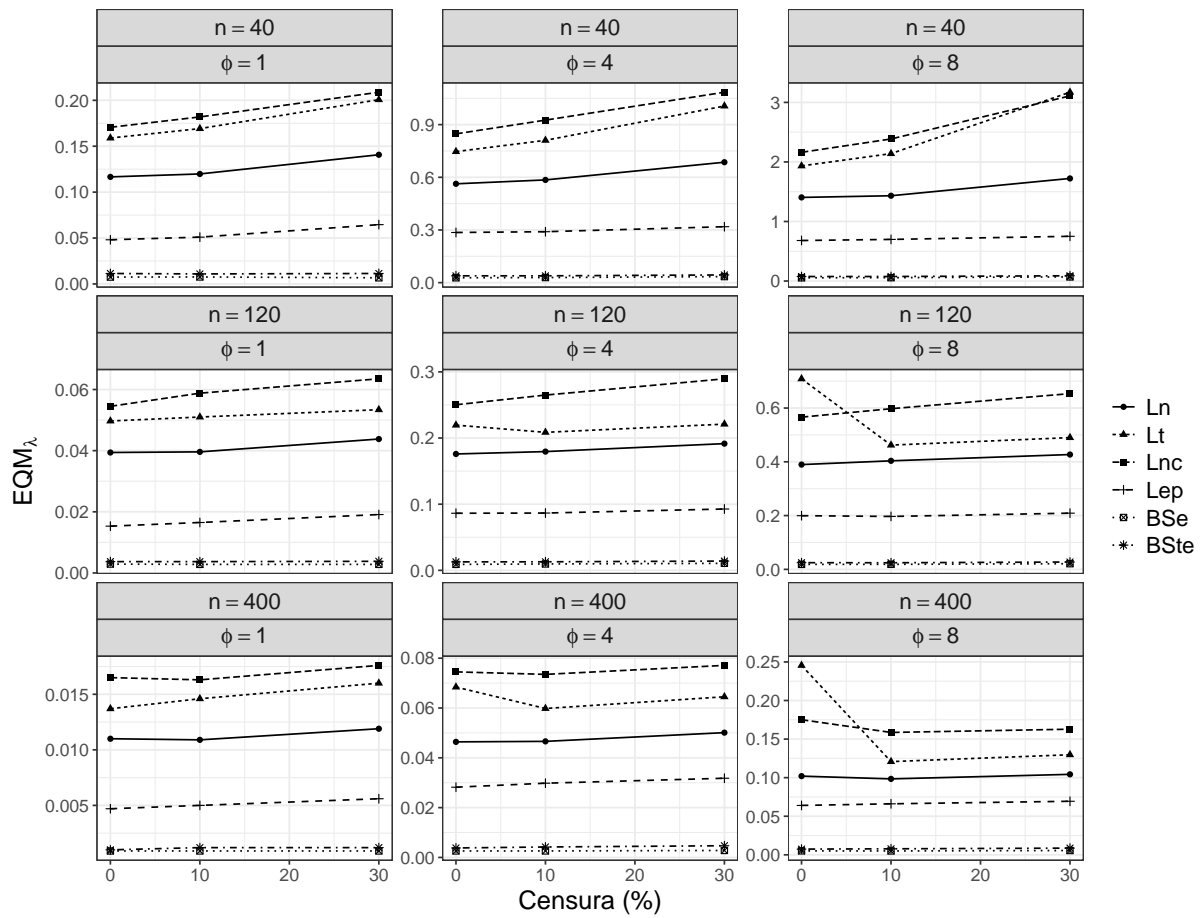
$n$	Cen.		$\phi = 1$			$\phi = 4$			$\phi = 8$			
			Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM	
40	0%	$\hat{\phi}$	0.8741	-0.1259	0.1721	3.9719	-0.0281	0.9531	7.9871	-0.0129	4.3427	
		$\hat{\lambda}$	2.0126	0.0126	0.0075	2.0117	0.0117	0.0275	2.0152	0.0152	0.0542	
120		$\hat{\phi}$	0.9747	-0.0253	0.0353	4.0090	0.0090	0.3178	8.0289	0.0289	1.3258	
		$\hat{\lambda}$	2.0049	0.0049	0.0029	2.0032	0.0032	0.0090	2.0037	0.0037	0.0185	
400		$\hat{\phi}$	0.9913	-0.0087	0.0117	4.0042	0.0042	0.0978	8.0266	0.0266	0.4048	
		$\hat{\lambda}$	2.0020	0.0020	0.0009	2.0011	0.0011	0.0026	2.0004	0.0004	0.0051	
40		10%	$\hat{\phi}$	0.8825	-0.1175	0.1945	4.0107	0.0107	1.2217	8.1117	0.1117	5.3717
			$\hat{\lambda}$	2.0128	0.0128	0.0076	2.0120	0.0120	0.0289	2.0172	0.0172	0.0565
120			$\hat{\phi}$	0.9699	-0.0301	0.0460	4.0048	0.0048	0.3449	8.0872	0.0872	1.6128
			$\hat{\lambda}$	2.0047	0.0047	0.0028	2.0043	0.0043	0.0096	2.0031	0.0031	0.0187
400	$\hat{\phi}$		0.9907	-0.0093	0.0134	4.0000	0.0000	0.1167	8.0092	0.0092	0.4932	
	$\hat{\lambda}$		2.0005	0.0005	0.0009	1.9997	-0.0003	0.0026	2.0005	0.0005	0.0050	
40	30%		$\hat{\phi}$	0.7674	-0.2326	0.3577	4.0841	0.0841	2.0748	8.3517	0.3517	8.9625
			$\hat{\lambda}$	2.0152	0.0152	0.0067	2.0178	0.0178	0.0343	2.0280	0.0280	0.0692
120			$\hat{\phi}$	0.9263	-0.0737	0.0967	4.0323	0.0323	0.5129	8.1723	0.1723	2.5286
			$\hat{\lambda}$	2.0052	0.0052	0.0028	2.0057	0.0057	0.0106	2.0055	0.0055	0.0210
400		$\hat{\phi}$	0.9852	-0.0148	0.0212	4.0188	0.0188	0.1660	8.0638	0.0638	0.7290	
		$\hat{\lambda}$	2.0009	0.0009	0.0009	2.0011	0.0011	0.0028	2.0028	0.0028	0.0056	

**Tabela 3.6:** Valores de média, viés e EQM para dados simulados da distribuição Birnbaum-Saunders- $t$  discreta considerando os estimadores de máxima verossimilhança,  $\lambda = 2$  and  $\xi = (0.5, 4)^T$ .

$n$	Cen.		$\phi = 1$			$\phi = 4$			$\phi = 8$			
			Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM	
40	0%	$\hat{\phi}$	0.9871	-0.0129	0.1511	3.9863	-0.0137	1.4404	8.0258	0.0258	6.0994	
		$\hat{\lambda}$	2.0028	0.0028	0.0114	2.0100	0.0100	0.0387	2.0167	0.0167	0.0756	
120		$\hat{\phi}$	0.9998	-0.0002	0.0457	4.0319	0.0319	0.4972	8.0689	0.0689	2.1333	
		$\hat{\lambda}$	2.0026	0.0026	0.0037	2.0061	0.0061	0.0128	2.0086	0.0086	0.0248	
400		$\hat{\phi}$	0.9984	-0.0016	0.0141	3.9996	-0.0004	0.1569	8.0282	0.0282	0.6348	
		$\hat{\lambda}$	2.0018	0.0018	0.0010	2.0031	0.0031	0.0038	2.0035	0.0035	0.0074	
40		10%	$\hat{\phi}$	1.0002	0.0002	0.1663	4.1093	0.1093	1.8263	8.2395	0.2395	7.9333
			$\hat{\lambda}$	2.0009	0.0009	0.0109	2.0029	0.0029	0.0387	2.0095	0.0095	0.0760
120			$\hat{\phi}$	1.0055	0.0055	0.0512	4.0492	0.0492	0.5522	8.1287	0.1287	2.3725
			$\hat{\lambda}$	2.0027	0.0027	0.0037	2.0044	0.0044	0.0129	2.0080	0.0080	0.0245
400	$\hat{\phi}$		1.0003	0.0003	0.0157	3.9944	-0.0056	0.1649	8.0085	0.0085	0.6848	
	$\hat{\lambda}$		2.0011	0.0011	0.0012	2.0020	0.0020	0.0042	2.0028	0.0028	0.0078	
40	30%		$\hat{\phi}$	0.9829	-0.0171	0.2728	4.1695	0.1695	2.7474	8.4158	0.4158	12.0941
			$\hat{\lambda}$	2.0027	0.0027	0.0114	2.0055	0.0055	0.0438	2.0151	0.0151	0.0882
120			$\hat{\phi}$	1.0115	0.0115	0.0708	4.1024	0.1024	0.7985	8.2566	0.2566	3.5320
			$\hat{\lambda}$	2.0029	0.0029	0.0038	2.0065	0.0065	0.0140	2.0119	0.0119	0.0273
400		$\hat{\phi}$	1.0009	0.0009	0.0238	4.0094	0.0094	0.2337	8.0494	0.0494	0.9990	
		$\hat{\lambda}$	2.0014	0.0014	0.0012	2.0037	0.0037	0.0047	2.0049	0.0049	0.0087	

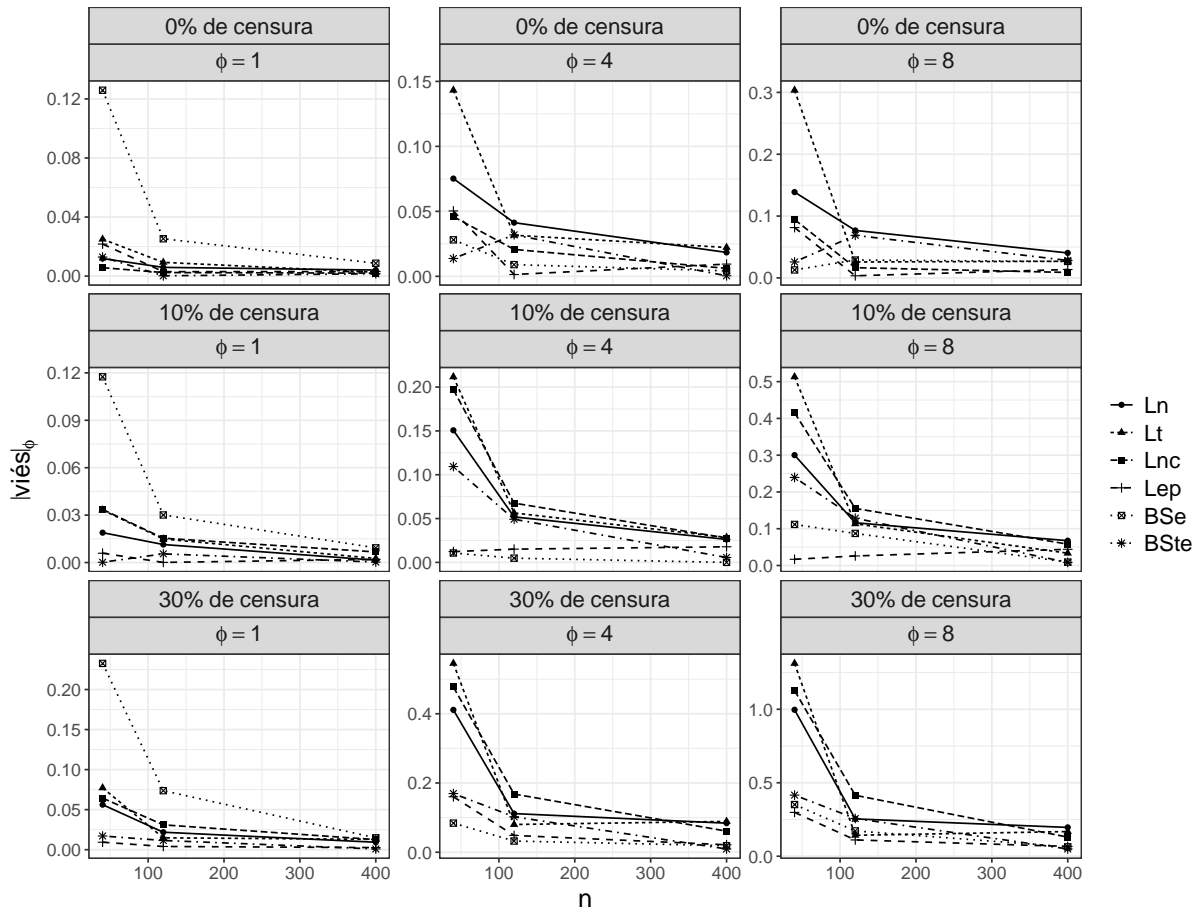


**Figura 3.8:** Comportamento do erro quadrático médio para as estimativas  $\hat{\phi}$  segundo a censura, valores verdadeiros do parâmetro  $\phi$  e tamanho da amostra.

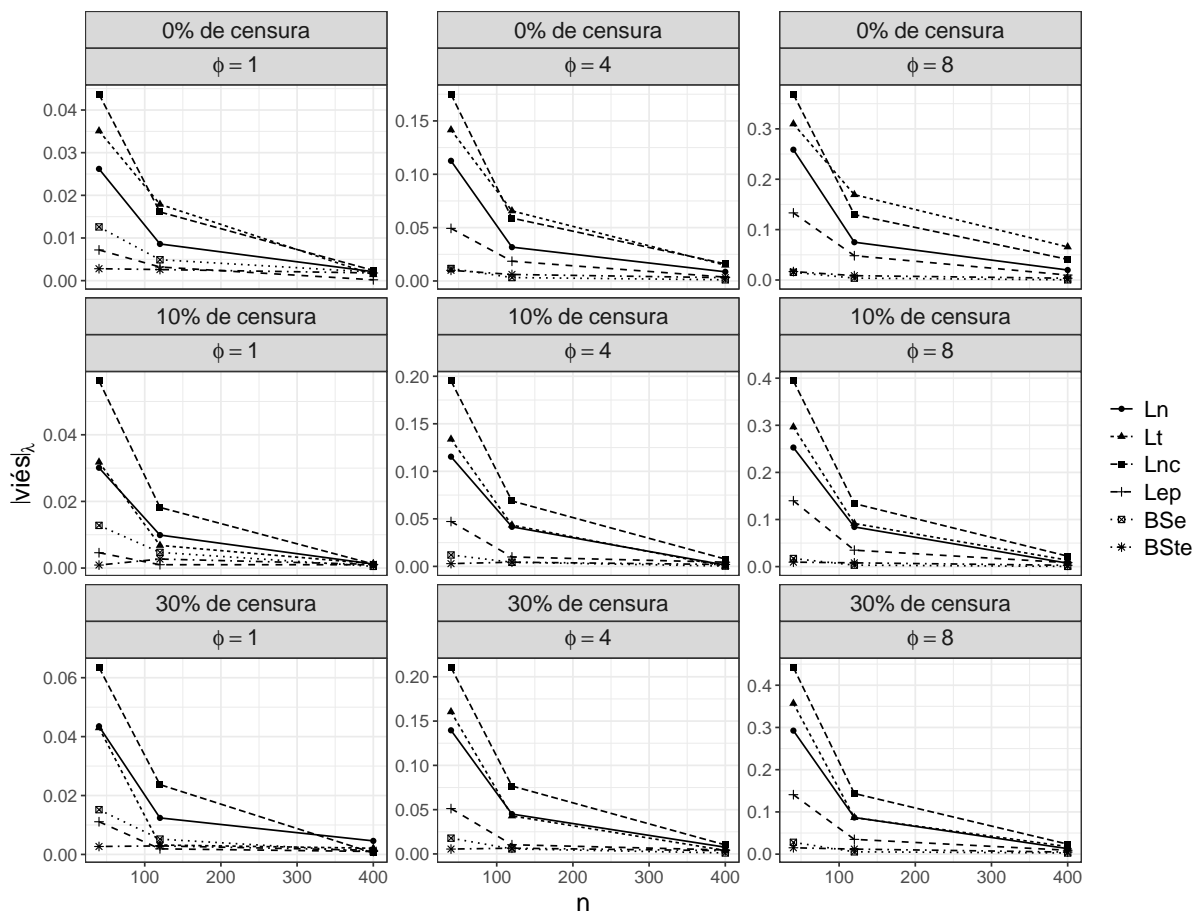


**Figura 3.9:** Comportamento do erro quadrático médio para as estimativas  $\hat{\lambda}$  segundo a censura, valores verdadeiros do parâmetro  $\phi$  e tamanho da amostra.





**Figura 3.10:** Comportamento do viés para as estimativas  $\hat{\phi}$  segundo a censura, valores verdadeiros do parâmetro  $\phi$  e tamanho da amostra.



**Figura 3.11:** Comportamento do viés para as estimativas  $\hat{\lambda}$  segundo a censura, valores verdadeiros do parâmetro  $\phi$  e tamanho da amostra.

### 3.5 Aplicações

Os modelos  $LS_d$  estudados na seção de simulação agora são usados para analisar conjuntos de dados do mundo real, com e sem observações censuradas.

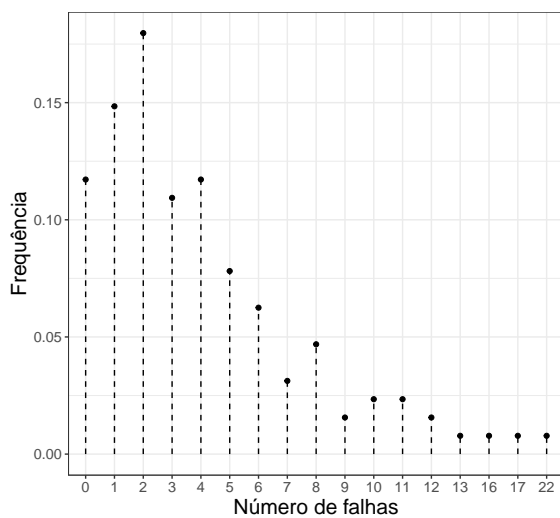
#### 3.5.1 Dados não censurados

O conjunto de dados considerado corresponde ao número de vezes que um computador DEC-20 falhou em cada uma das 128 semanas consecutivas de operação. Tal computador operou na Open University durante os anos 80; veja Tabela 3.7 e Trenkler (1995). As estatísticas

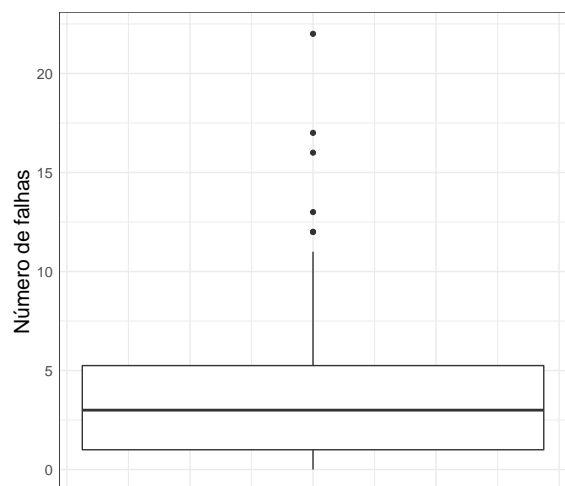
descriptivas para esse conjunto de dados são as seguintes: 128 (tamanho da amostra), 0 (mínimo), 22 (máximo), 3 (mediana), 4.016 (média), 3.808 (desvio padrão), 94.839 (coeficiente de variação), 1.732 (coeficiente de assimetria) e 3.995 (coeficiente de curtose). A partir desses resultados, observa-se que esses dados têm um comportamento assimétrico e curtose elevada. A Figura 3.12 refere-se ao plot dos dados de falha de computador (a) e ao box-plot desses dados (b), de onde se confirma uma assimetria à direita.

**Tabela 3.7:** Dados de falha de um computador.

x (número de falhas)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	16	17	22
frequência	15	19	23	14	15	10	8	4	6	2	3	3	2	1	1	1	1



(a) Distribuição dos dados.



(b) Boxplot para o número de falhas.

**Figura 3.12:** Distribuição e boxplot dos dados de falha de um computador.

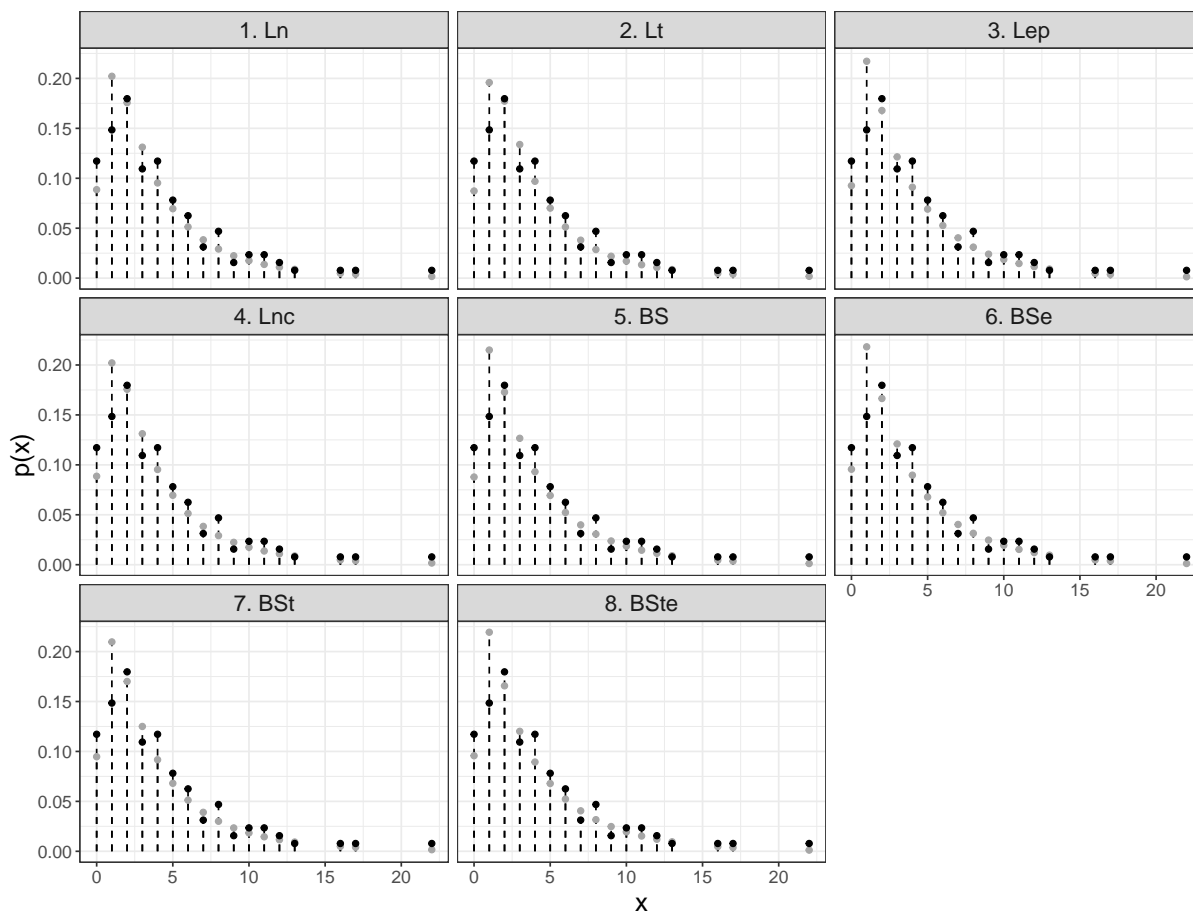
A Tabela 3.8 mostra as estimativas de máxima verossimilhança, calculadas pelo método BFGS, e os  $p$ -valores das estatísticas  $\chi^2$  e CVM. Também são considerados os critérios AIC e BIC. Aqui, todos os modelos  $LS_d$  fornecem ajustes adequados com base nos  $p$ -valores de  $\chi^2$  e CVM. Levando em consideração os valores de AIC e BIC, os melhores ajustes são dos modelos Birnbaum-Saunders estendido e log-exponencial-potência.

**Tabela 3.8:** Estimativas de máxima verossimilhança e medidas de seleção de modelo para ajuste dos dados de falhas de computador.

Modelo	Estimativas			$p$ -valor		AIC	BIC
	$\hat{\lambda}$ (EP)	$\hat{\phi}$ (EP)	$\hat{\xi}$ ( $\xi$ )	$\chi^2$	CMV		
Ln	3.2280 (0.2526)	0.7541 (0.1048)	—	0.7841	0.6959	643.5141	652.0702
Lt	3.2653 (0.2574)	0.7065 (0.1026)	20.0	0.8306	0.7762	644.2248	652.7809
Lnc	3.2283 (0.2526)	0.6858 (0.0953)	(0.9, 0.9)	0.7996	0.6967	645.5205	656.9287
Lep	3.1624 (0.2770)	1.0176 (0.0555)	-0.2	0.7096	0.5151	642.8785	651.4346
BS	3.1704 (0.0589)	*	0.9	0.6007	0.5283	640.6061	646.3102
BSe	3.1436 (0.2090)	2.9392 (0.0438)	1.1	0.6517	0.4700	642.4045	650.9605
BSt	3.1803 (0.3008)	**	(0.9, 20.0)	0.7966	0.5799	642.9026	651.4587
BSte	3.1406 (0.2193)	2.0820 (0.0818)	(1.3, 20.0)	0.6492	0.4364	644.5534	655.9615

EP: erro padrão; \* corresponde à BS estendida com  $\phi = 4$ ; \*\* corresponde à BS- $t$  estendida com  $\phi = 4$ .

Na Figura 3.13 estão configurados a distribuição empírica dos dados e o ajuste das distribuições log-simétricas discretas. Em geral, os pontos estimados ficaram próximos dos pontos observados em todos os modelos considerados no estudo.



**Figura 3.13:** Distribuição empírica (.) *versus* distribuições log-simétricas discretas ajustadas para os dados de falha de computador (\*).

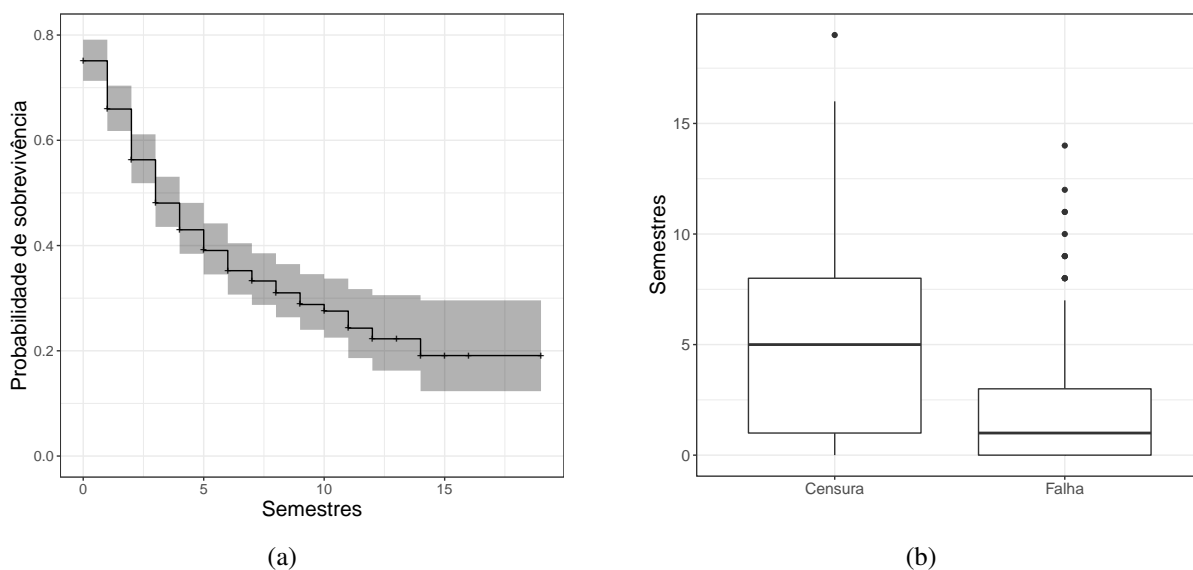
### 3.5.2 Dados censurados

Para a segunda aplicação serão considerados os dados sobre evasão de alunos que ingressaram no curso de Estatística da Universidade Federal do Piauí (UFPI), campus de Teresina, do primeiro semestre de 2010 ao primeiro semestre de 2019. Esse banco é composto por 470 observações, tendo aproximadamente 40% de censura.

Neste trabalho, o tempo falha foi definido como o tempo (em semestres) até o aluno evadir (cancelamento ou trancamento). Se esse evento não foi observado (aluno ativo ao final do acompanhamento ou tendo concluído o curso), então é caracterizada a presença de censura.

Na Figura 3.14 estão as curvas de sobrevivência de Kaplan-Meier (a) e os boxplots para

os tempos de falha e censura (b). Pela Figura 3.14 (a), observa-se que o tempo mediano está próximo de 2.5 semestres. Além disso, com o aumento do tempo a curva de sobrevivência tende a se estabilizar em torno de 20%. Quanto aos tempos de falha e censura, nota-se, pela Figura 3.14 (b), que a mediana dos tempos de censura é de 5 semestres e a mediana dos tempos de falha é próxima de 1 semestre.



**Figura 3.14:** Curva de sobrevivência (com intervalo de confiança) estimada pelo método Kaplan-Meier (a) e boxplot para os tempos de falha e censura (b).

As estimativas de máxima verossimilhança calculadas pelo método BFGS e os valores de AIC e BIC para os dados de tempo de até a evasão são mostrados na Tabela 3.9. Levando em consideração os valores de AIC e BIC, os modelos que melhor se ajustaram aos dados foram os modelos log- $t$ -Student e log-exponencial-potência.

**Tabela 3.9:** Estimativas de máxima verossimilhança e medidas de seleção de modelo para ajuste dos dados de evasão no curso de Estatística.

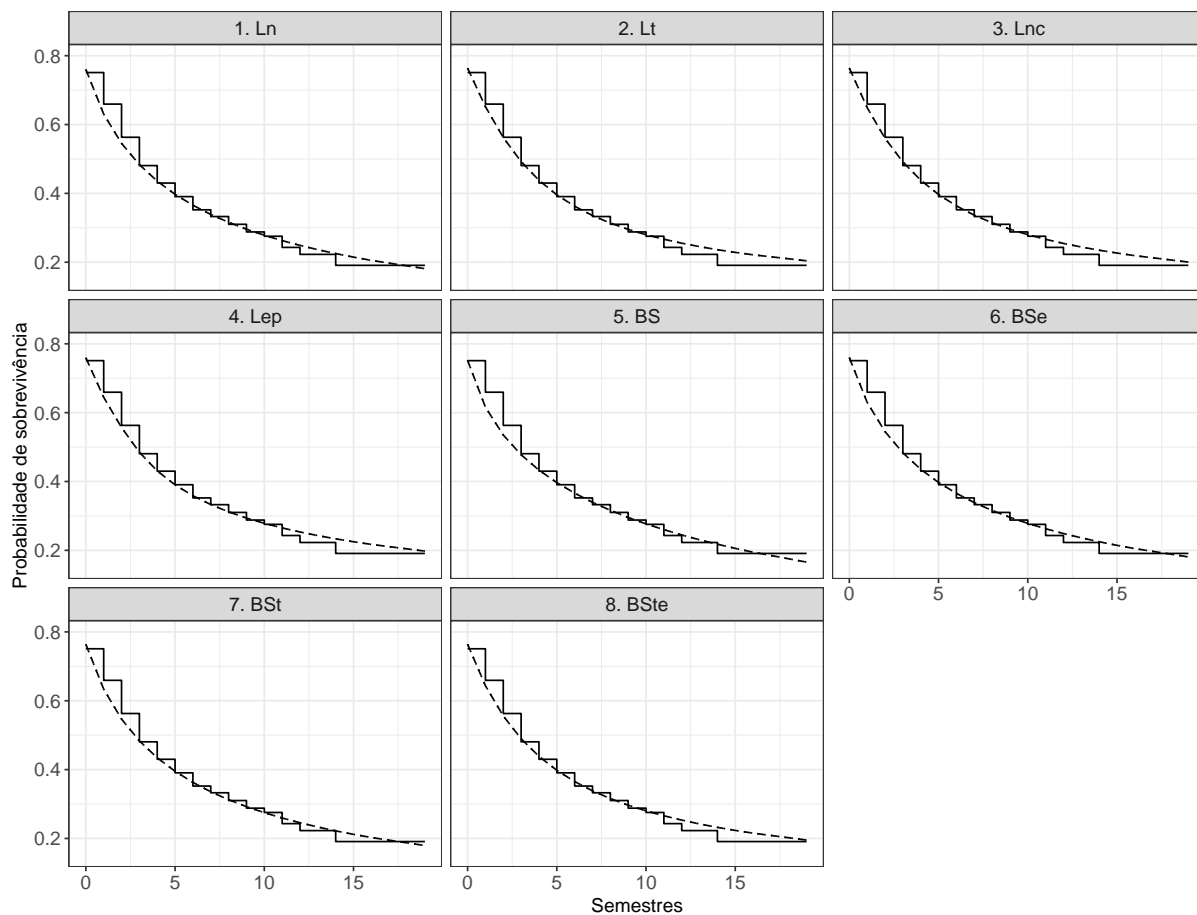
Distribuição discreta	Estimativas			AIC	BIC
	$\hat{\lambda}$ (EP)	$\hat{\phi}$ (EP)	$\hat{\xi}$ ( $\xi$ )		
Ln	3.7039 (0.3577)	3.4309 (0.4355)	–	1591.294	1603.752
Lt	3.8761 (0.3593)	1.7887 (0.4228)	1.2	1585.895	1598.353
Lnc	3.8593 (0.3577)	0.6724 (0.3962)	(0.74, 0.10)	1588.580	1605.190
Lep	3.7723 (0.3347)	1.1552 (0.1691)	0.8	1586.022	1598.480
BS	3.5753 (0.4812)	*	2.0	1593.202	1601.508
BSe	3.7001 (0.3227)	153.0782 (0.2981)	0.3	1591.389	1603.847
BSt	3.7104 (0.4977)	**	(1.6, 2.0)	1590.891	1603.350
BSte	3.8253 (0.3249)	103.7957 (0.2906)	(0.3, 2.0)	1589.999	1606.610

EP: erro padrão; \* corresponde à BS estendida com  $\phi = 4$ ; \*\* corresponde à BS- $t$  estendida com  $\phi = 4$ .

Conforme Tabela 3.10 e Figura 3.15, verifica-se que as estimativas das sobrevivências dos modelos log-simétricos discretos ficaram bem próximas das estimativas de Kaplan-Meier, o que indica um bom ajuste desses modelos aos dados.

**Tabela 3.10:** Estimativas da função de sobrevivência via Kaplan-Meier e via distribuições log-simétricas discretas.

x	KM	Modelo								IC 95% (KM)
		Ln	Lt	Lnc	Lep	BS	BSe	BSt	BSte	
0	0.7511	0.7602	0.7638	0.7641	0.7595	0.7521	0.7600	0.7640	0.7643	0.7130–0.7912
1	0.6594	0.6303	0.6520	0.6491	0.6446	0.6158	0.6299	0.6337	0.6438	0.6177–0.7038
2	0.5630	0.5453	0.5623	0.5600	0.5569	0.5350	0.5450	0.5468	0.5559	0.5185–0.6114
3	0.4807	0.4834	0.4923	0.4914	0.4849	0.4776	0.4833	0.4834	0.4897	0.4354–0.5307
4	0.4300	0.4357	0.4380	0.4383	0.4307	0.4331	0.4356	0.4344	0.4385	0.3843–0.4811
5	0.3906	0.3973	0.3958	0.3968	0.3902	0.3968	0.3972	0.3952	0.3980	0.3451–0.4421
6	0.3522	0.3656	0.3625	0.3636	0.3586	0.3661	0.3655	0.3628	0.3653	0.3068–0.4044
7	0.3327	0.3388	0.3356	0.3367	0.3330	0.3396	0.3388	0.3357	0.3383	0.2872–0.3853
8	0.3101	0.3159	0.3137	0.3145	0.3118	0.3163	0.3158	0.3125	0.3158	0.2637–0.3647
9	0.2880	0.2959	0.2954	0.2959	0.2939	0.2956	0.2959	0.2924	0.2966	0.2399–0.3457
10	0.2754	0.2784	0.2799	0.2801	0.2785	0.2769	0.2783	0.2749	0.2801	0.2250–0.3372
11	0.2430	0.2628	0.2667	0.2665	0.2651	0.2601	0.2627	0.2594	0.2658	0.1862–0.3173
12	0.2228	0.2489	0.2553	0.2546	0.2533	0.2447	0.2488	0.2456	0.2533	0.1623–0.3057
13	0.2228	0.2364	0.2453	0.2442	0.2428	0.2306	0.2363	0.2332	0.2421	0.1623–0.3057
14	0.1910	0.2251	0.2364	0.2349	0.2334	0.2177	0.2249	0.2220	0.2322	0.1233–0.2958
15	0.1910	0.2148	0.2286	0.2266	0.2249	0.2057	0.2145	0.2119	0.2233	0.1233–0.2958
16	0.1910	0.2053	0.2215	0.2191	0.2172	0.1946	0.2050	0.2027	0.2153	0.1233–0.2958
19	0.1910	0.1813	0.2041	0.2004	0.1977	0.1657	0.1809	0.1794	0.1953	0.1233–0.2958



**Figura 3.15:** Curvas de sobrevivência estimadas por Kaplan-Meier (—) e pelos modelos log-simétricos discretos (- - -).



# Capítulo 4

## Modelos de regressão log-simétricos discretos

### 4.1 Modelos de regressão log-simétricos discretos

Na maioria dos estudos estatísticos e, sobretudo, em estudos de análise de sobrevivência é interessante estudar a influência que algumas covariáveis exercem sobre a variável resposta. Para Lawless (2011), a inclusão de covariáveis nos modelos é importante porque pode revelar a heterogeneidade de uma população.

Seja  $X_1, \dots, X_n$  variáveis aleatórias independentes, em que  $X_i \sim \text{LS}_d(\lambda_i, \phi, g(\cdot))$ , para  $i = 1, \dots, n$ , e  $\mathbf{x} = (x_1, \dots, x_n)^\top$  as correspondentes observações. O modelo de regressão log-simétrico discreto é definido pela seguinte relação

$$h(\lambda_i) = \eta_i = \mathbf{w}_i^\top \boldsymbol{\beta}, \quad i = 1, \dots, n, \quad (4.1)$$

em que  $\mathbf{w}_i^\top = (1, w_{i1}, \dots, w_{ik})$  representa os valores das covariáveis,  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)^\top$  é um vetor de parâmetros desconhecidos, e  $h : \mathbb{R} \rightarrow \mathbb{R}^+$  é uma função de ligação invertível, estritamente monótona, positiva e pelo menos duas vezes diferenciável. Neste trabalho,  $h(\lambda) =$

$\log(\lambda)$ .

Os parâmetros do modelo de regressão log-simétrico discreto são estimados usando o método da máxima verossimilhança. As funções de verossimilhança e log-verossimilhança do modelo, considerando dados censurados à direita e uma amostra de tamanho  $n$ , são definidas, respectivamente, como:

$$L(\boldsymbol{\theta}) \propto \prod_{i=1}^n [p(x_i|w)]^{\delta_i} [S(x_i|w)]^{1-\delta_i} \quad (4.2)$$

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \{\delta_i \log[p(x_i|w)] + (1 - \delta_i) \log[S(x_i|w)]\} + C \quad (4.3)$$

sendo  $C$  uma constante não dependente de  $\boldsymbol{\theta}$ .

em que  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi)$  e  $\delta_i$  é o indicador de censura. As estimativas de máxima verossimilhança de  $\phi$  e  $\boldsymbol{\beta}$  são obtidas por meio da maximização de  $\ell(\boldsymbol{\theta})$ .

## 4.2 Análise de resíduos

A análise dos resíduos é útil para avaliar a adequação do modelo e detectar a presença de pontos extremos. Em virtude da diversidade dos modelos log-simétricos discretos, os resíduos utilizados para os modelos lineares não são adequados para esses modelos, pois, frequentemente, os resíduos podem não apresentar distribuição normal, embora o modelo esteja bem ajustado aos dados. Essa peculiaridade é inerente aos modelos com a variável resposta discreta, especialmente quando os dados assumem valores pequenos. Portanto, neste trabalho, será utilizado o resíduo quantílico aleatorizado proposto por Dunn e Smyth (1996). Esse tipo de resíduo é obtido da seguinte maneira:

$$r_{q,i} = \Phi^{-1}(F(x_i; \hat{\boldsymbol{\theta}}_i)), \quad i = 1, 2, \dots, n, \quad (4.4)$$

em que  $\Phi(\cdot)$  é a FDA da normal padrão;  $x_i$  é a  $i$ -ésima observação da variável resposta;  $F(x_i; \hat{\boldsymbol{\theta}}_i)$

é a FDA do modelo e  $\hat{\boldsymbol{\theta}}_i$  é o vetor de parâmetros estimados. Para modelos de distribuições discretas,  $r_{q,i}$  é expresso como:

$$r_{q,i} = \Phi^{-1}(u_i), \quad (4.5)$$

em que  $u_i$  é uma variável aleatória uniformemente distribuída no intervalo  $(a_i, b_i]$ ,  $a_i$  e  $b_i$  definidos da seguinte forma:

$$a_i = F(x_i - 1; \hat{\boldsymbol{\theta}}_i), \quad b_i = F(x_i; \hat{\boldsymbol{\theta}}_i). \quad (4.6)$$

Para os modelos log-simétricos discretos, as funções  $a_i$  e  $b_i$  são escritas de acordo com a função de sobrevivência, usando a relação  $F(x_i - 1; \hat{\boldsymbol{\theta}}_i) = 1 - S(x_i - 1; \hat{\boldsymbol{\theta}}_i)$ . Portanto,  $a_i$  e  $b_i$  são expressos como:

$$a_i = G(b(\lfloor x_i - 1 \rfloor + 1))\mathbb{1}_{\mathbb{N}^*}, \quad b_i = G(b(\lfloor x_i \rfloor + 1))\mathbb{1}_{\mathbb{N}}, \quad (4.7)$$

onde

$$b(x) = \frac{\log(x) - \log(\exp(\mathbf{w}^\top \boldsymbol{\beta}))}{\sqrt{\phi}} = \frac{\log(x) - \mathbf{w}^\top \boldsymbol{\beta}}{\sqrt{\phi}}. \quad (4.8)$$

O resíduo quantílico aleatorizado,  $r_{q,i}$ , é distribuído segundo uma normal padrão caso o modelo esteja bem ajustado aos dados. Portanto, fazendo o *QQplot* dos resíduos em função de uma distribuição normal padrão, é possível constatar a normalidade desses resíduos, além de identificar discrepâncias entre as estimativas e os valores observados.

### 4.3 Simulações

Para a obtenção dos dados simulados da regressão, considerou-se apenas uma covariável para cada indivíduo da amostra. Os tamanhos amostrais foram fixados em  $n \in \{40, 120, 400\}$  e a covariável foi obtida a partir de uma distribuição Bernoulli com probabilidade de sucesso  $p = 0.5$ . Os percentuais de censura foram obtidos usando o mesmo procedimento adotado para os modelos sem covariáveis (ver Seção 3.4). Os valores reais considerados para os parâmetros foram:  $\beta_0 = 1$ ,  $\beta_1 = 2.5$  e  $\phi \in \{1, 4, 8\}$ . Nesse cenário, os modelos foram avaliados por meio de simulações de Monte Carlo (1000 réplicas). Assim como na Seção 3.4, as estimativas dos parâmetros foram obtidas por máxima verossimilhança por meio da função `optim` (método BFGS).

Os resultados das simulações estão denotados nas Tabelas 4.1–4.6 e ilustrados nas Figuras 4.1, 4.2 e 4.3. Em geral, os resultados obtidos na simulação foram como esperados: viés diminui conforme aumenta o tamanho da amostra e o erro quadrático médio aumenta conforme aumenta o percentual de censura.

**Tabela 4.1:** Valores de média, viés e EQM para dados simulados da distribuição log-normal discreta considerando os estimadores de máxima verossimilhança do modelo de regressão ( $\beta_0 = 1$ , e  $\beta_1 = 2.5$ ).

$n$	Cen.		$\phi = 1$			$\phi = 4$			$\phi = 8$		
			Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
40	0%	$\hat{\phi}$	0.9662	-0.0338	0.0617	3.8973	-0.1027	1.0548	7.8430	-0.1570	4.7602
		$\hat{\beta}_0$	1.0007	0.0007	0.0410	1.0118	0.0118	0.3446	1.0096	0.0096	0.7066
		$\hat{\beta}_1$	2.5091	0.0091	0.1221	2.4875	-0.0125	0.5118	2.4837	-0.0163	1.0467
120	0%	$\hat{\phi}$	0.9935	-0.0065	0.0194	3.9813	-0.0187	0.3414	7.9973	-0.0027	1.4937
		$\hat{\beta}_0$	0.9987	-0.0013	0.0140	0.9979	-0.0021	0.1108	0.9927	-0.0073	0.2268
		$\hat{\beta}_1$	2.5024	0.0024	0.0396	2.5017	0.0017	0.1615	2.5032	0.0032	0.3322
400	0%	$\hat{\phi}$	0.9996	-0.0004	0.0062	4.0075	0.0075	0.1022	8.0172	0.0172	0.4435
		$\hat{\beta}_0$	0.9994	-0.0006	0.0039	0.9963	-0.0037	0.0342	0.9956	-0.0044	0.0698
		$\hat{\beta}_1$	2.5012	0.0012	0.0117	2.5031	0.0031	0.0502	2.5029	0.0029	0.1018
40	10%	$\hat{\phi}$	0.9755	-0.0245	0.0640	3.9292	-0.0708	1.2240	7.9272	-0.0728	5.7356
		$\hat{\beta}_0$	1.0092	0.0092	0.0859	1.0225	0.0225	0.3551	1.0223	0.0223	0.7286
		$\hat{\beta}_1$	2.4905	-0.0095	0.1233	2.4739	-0.0261	0.5094	2.4685	-0.0315	1.0565
120	10%	$\hat{\phi}$	0.9935	-0.0065	0.0198	3.9985	-0.0015	0.3589	8.0143	0.0143	1.6294
		$\hat{\beta}_0$	1.0059	0.0059	0.0293	1.0085	0.0085	0.1241	1.0158	0.0158	0.2501
		$\hat{\beta}_1$	2.4927	-0.0073	0.0428	2.4867	-0.0133	0.1807	2.4753	-0.0247	0.3648
400	10%	$\hat{\phi}$	0.9972	-0.0028	0.0067	4.0001	0.0001	0.1244	8.0061	0.0061	0.5485
		$\hat{\beta}_0$	0.9966	-0.0034	0.0077	0.9931	-0.0069	0.0333	0.9893	-0.0107	0.0686
		$\hat{\beta}_1$	2.5036	0.0036	0.0115	2.5063	0.0063	0.0483	2.5097	0.0097	0.0984
40	30%	$\hat{\phi}$	0.9852	-0.0148	0.0943	4.0006	0.0006	1.9467	8.1342	0.1342	9.7994
		$\hat{\beta}_0$	1.0131	0.0131	0.0967	1.0304	0.0304	0.3929	1.0300	0.0300	0.8084
		$\hat{\beta}_1$	2.4916	-0.0084	0.1398	2.4771	-0.0229	0.5765	2.4754	-0.0246	1.2095
120	30%	$\hat{\phi}$	0.9940	-0.0060	0.0267	4.0106	0.0106	0.5171	8.0639	0.0639	2.5059
		$\hat{\beta}_0$	1.0048	0.0048	0.0323	1.0061	0.0061	0.1377	1.0138	0.0138	0.2751
		$\hat{\beta}_1$	2.4948	-0.0052	0.0484	2.4922	-0.0078	0.2064	2.4824	-0.0176	0.4155
400	30%	$\hat{\phi}$	1.0008	0.0008	0.0091	4.0190	0.0190	0.1717	8.0557	0.0557	0.7819
		$\hat{\beta}_0$	0.9977	-0.0023	0.0083	0.9923	-0.0077	0.0356	0.9914	-0.0086	0.0740
		$\hat{\beta}_1$	2.5040	0.0040	0.0128	2.5106	0.0106	0.0540	2.5125	0.0125	0.1100

**Tabela 4.2:** Valores de média, viés e EQM para dados simulados da distribuição log- $t$ -Student discreta considerando os estimadores de máxima verossimilhança do modelo de regressão ( $\beta_0 = 1$ ,  $\beta_1 = 2.5$  e  $\xi = 4$ ).

$n$	Cen.		$\phi = 1$			$\phi = 4$			$\phi = 8$		
			Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
40		$\hat{\phi}$	0.9735	-0.0265	0.0904	3.9013	-0.0987	1.5487	7.8770	-0.1230	7.0533
		$\hat{\beta}_0$	0.9955	-0.0045	0.1090	0.9882	-0.0118	0.4418	1.0150	0.0150	1.0758
		$\hat{\beta}_1$	2.5030	0.0030	0.1768	2.5082	0.0082	0.7032	2.4517	-0.0483	1.4709
120	0%	$\hat{\phi}$	0.9916	-0.0084	0.0316	3.9568	-0.0432	0.5491	7.8748	-0.1252	2.2273
		$\hat{\beta}_0$	0.9995	-0.0005	0.0378	1.0021	0.0021	0.1491	1.0079	0.0079	0.2971
		$\hat{\beta}_1$	2.5047	0.0047	0.0576	2.5055	0.0055	0.2286	2.4988	-0.0012	0.4579
400		$\hat{\phi}$	0.9942	-0.0058	0.0096	3.9706	-0.0294	0.1572	7.9123	-0.0877	0.6245
		$\hat{\beta}_0$	0.9946	-0.0054	0.0108	0.9873	-0.0127	0.0419	0.9833	0.0167	0.0854
		$\hat{\beta}_1$	2.5046	0.0046	0.0163	2.5109	0.0109	0.0641	2.5095	0.0095	0.1381
40		$\hat{\phi}$	0.9800	-0.0200	0.1026	3.9559	-0.0441	1.7840	7.9800	-0.0200	8.1794
		$\hat{\beta}_0$	0.9955	-0.0045	0.1100	0.9812	-0.0188	0.4738	0.9572	-0.0428	1.0059
		$\hat{\beta}_1$	2.4985	-0.0015	0.1797	2.5082	0.0082	0.7560	2.5274	0.0274	1.5607
120	10%	$\hat{\phi}$	0.9933	-0.0067	0.0323	3.9789	-0.0211	0.5758	7.9678	-0.0322	2.5767
		$\hat{\beta}_0$	0.9951	-0.0049	0.0394	0.9921	-0.0079	0.1583	0.9847	-0.0153	0.3199
		$\hat{\beta}_1$	2.5040	0.0040	0.0586	2.5064	0.0064	0.2356	2.5132	0.0132	0.4765
400		$\hat{\phi}$	0.9968	-0.0032	0.0097	3.9909	-0.0091	0.1726	7.9819	-0.0181	0.7554
		$\hat{\beta}_0$	0.9960	-0.0040	0.0114	0.9904	-0.0096	0.0456	0.9863	-0.0137	0.0923
		$\hat{\beta}_1$	2.5025	0.0025	0.0168	2.5068	0.0068	0.0679	2.5098	0.0098	0.1377
40		$\hat{\phi}$	0.9948	-0.0052	0.1455	4.0459	0.0459	2.7599	8.2371	0.2371	13.2247
		$\hat{\beta}_0$	0.9999	-0.0001	0.1233	0.9863	-0.0137	0.5261	0.9648	-0.0352	1.1184
		$\hat{\beta}_1$	2.4954	-0.0046	0.2006	2.5086	0.0086	0.8500	2.5291	0.0291	1.7631
120	30%	$\hat{\phi}$	0.9867	-0.0133	0.0406	3.9604	-0.0396	0.7475	7.9368	-0.0632	3.5671
		$\hat{\beta}_0$	0.9934	-0.0066	0.0413	0.9923	-0.0077	0.1629	0.9855	-0.0145	0.3309
		$\hat{\beta}_1$	2.5029	0.0029	0.0615	2.5004	0.0004	0.2439	2.5044	0.0044	0.4988
400		$\hat{\phi}$	1.0015	0.0015	0.0119	4.0136	0.0136	0.2166	8.0324	0.0324	0.9932
		$\hat{\beta}_0$	0.9964	-0.0036	0.0118	0.9909	-0.0091	0.0480	0.9859	-0.0141	0.0958
		$\hat{\beta}_1$	2.5034	0.0034	0.0182	2.5095	0.0095	0.0752	2.5150	0.0150	0.1493

**Tabela 4.3:** Valores de média, viés e EQM para dados simulados da distribuição log-normal-contaminada discreta considerando os estimadores de máxima verossimilhança do modelo de regressão ( $\beta_0 = 1$ ,  $\beta_1 = 2.5$  e  $\xi = (0.5, 0.5)^\top$ ).

$n$	Cen.		$\phi = 1$			$\phi = 4$			$\phi = 8$		
			Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
40	0%	$\hat{\phi}$	0.9673	-0.0327	0.0651	3.8956	-0.1044	1.1743	7.8287	-0.1713	5.3106
		$\hat{\beta}_0$	1.0143	0.0143	0.1277	1.0254	0.0254	0.5352	1.0325	0.0325	1.0775
		$\hat{\beta}_1$	2.4801	-0.0199	0.1916	2.4608	-0.0392	0.8083	2.4430	-0.0570	1.6381
120	0%	$\hat{\phi}$	0.9931	-0.0069	0.0199	3.9755	-0.0245	0.3771	7.9604	-0.0396	1.6600
		$\hat{\beta}_0$	0.9986	-0.0014	0.0391	1.0008	0.0008	0.1623	1.0000	0.0000	0.3267
		$\hat{\beta}_1$	2.5031	0.0031	0.0568	2.5014	0.0014	0.2337	2.5023	0.0023	0.4724
400	0%	$\hat{\phi}$	0.9976	-0.0024	0.0064	3.9996	-0.0004	0.1178	7.9985	-0.0015	0.5074
		$\hat{\beta}_0$	1.0043	0.0043	0.0111	1.0058	0.0058	0.0460	1.0101	0.0101	0.0952
		$\hat{\beta}_1$	2.4930	-0.0070	0.0170	2.4880	-0.0120	0.0686	2.4805	-0.0195	0.1402
40	10%	$\hat{\phi}$	0.9795	-0.0205	0.0818	3.9481	-0.0519	1.5214	7.9792	-0.0208	7.3393
		$\hat{\beta}_0$	1.0192	0.0192	0.1180	1.0334	0.0334	0.5046	1.0373	0.0373	1.0258
		$\hat{\beta}_1$	2.4832	-0.0168	0.1835	2.4689	-0.0311	0.7764	2.4625	-0.0375	1.5665
120	10%	$\hat{\phi}$	0.9971	-0.0029	0.0237	3.9995	-0.0005	0.4473	7.9990	-0.0010	2.0590
		$\hat{\beta}_0$	1.0032	0.0032	0.0422	1.0049	0.0049	0.1800	1.0131	0.0131	0.3636
		$\hat{\beta}_1$	2.5002	0.0002	0.0585	2.5005	0.0005	0.2467	2.4938	-0.0062	0.4971
400	10%	$\hat{\phi}$	0.9990	-0.0010	0.0069	4.0050	0.0050	0.1285	8.0279	0.0279	0.5742
		$\hat{\beta}_0$	1.0026	0.0026	0.0112	1.0032	0.0032	0.0466	1.0036	0.0036	0.0990
		$\hat{\beta}_1$	2.4934	-0.0066	0.0170	2.4881	-0.0119	0.0698	2.4823	-0.0177	0.1446
40	30%	$\hat{\phi}$	0.9813	-0.0187	0.1083	3.9969	-0.0031	2.2990	8.1638	0.1638	12.6151
		$\hat{\beta}_0$	1.0188	0.0188	0.1264	1.0355	0.0355	0.5419	1.0392	0.0392	1.0991
		$\hat{\beta}_1$	2.4834	-0.0166	0.2012	2.4680	-0.0320	0.8623	2.4655	-0.0345	1.7373
120	30%	$\hat{\phi}$	1.0009	0.0009	0.0302	4.0265	0.0265	0.5935	8.0789	0.0789	2.9376
		$\hat{\beta}_0$	1.0037	0.0037	0.0459	1.0066	0.0066	0.1960	1.0171	0.0171	0.3907
		$\hat{\beta}_1$	2.5014	0.0014	0.0647	2.5028	0.0028	0.2740	2.4964	-0.0036	0.5492
400	30%	$\hat{\phi}$	0.9997	-0.0003	0.0097	4.0140	0.0140	0.1854	8.0604	0.0604	0.8748
		$\hat{\beta}_0$	1.0029	0.0029	0.0125	1.0046	0.0046	0.0513	1.0055	0.0055	0.1088
		$\hat{\beta}_1$	2.4929	-0.0071	0.0193	2.4870	-0.0130	0.0793	2.4817	-0.0183	0.1637

**Tabela 4.4:** Valores de média, viés e EQM para dados simulados da distribuição log-exponencial-potência discreta considerando os estimadores de máxima verossimilhança do modelo de regressão ( $\beta_0 = 1$ ,  $\beta_1 = 2.5$  e  $\xi = -0.5$ ).

n	Cen.	$\phi = 1$			$\phi = 4$			$\phi = 8$			
		Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM	
40	0%	$\hat{\phi}$	0.9335	-0.0665	0.0339	3.7510	-0.2490	0.5575	7.5486	-0.4514	2.4972
		$\hat{\beta}_0$	1.0024	0.0024	0.0164	0.9931	-0.0069	0.1600	0.9830	-0.0170	0.3403
		$\hat{\beta}_1$	2.5060	0.0060	0.0451	2.5090	0.0090	0.2168	2.5130	0.0130	0.4523
120	0%	$\hat{\phi}$	0.9834	-0.0166	0.0096	3.9512	-0.0488	0.1616	7.9173	-0.0827	0.7284
		$\hat{\beta}_0$	1.0045	0.0045	0.0051	0.9790	-0.0210	0.0499	0.9680	-0.0320	0.1063
		$\hat{\beta}_1$	2.4933	-0.0067	0.0146	2.5301	0.0301	0.0693	2.5424	0.0424	0.1440
400	0%	$\hat{\phi}$	0.9919	-0.0081	0.0030	3.9723	-0.0277	0.0488	7.9456	-0.0544	0.2145
		$\hat{\beta}_0$	0.9994	-0.0006	0.0015	0.9998	-0.0002	0.0156	1.0006	0.0006	0.0339
		$\hat{\beta}_1$	2.5024	0.0024	0.0043	2.4988	-0.0012	0.0220	2.4971	-0.0029	0.0461
40	10%	$\hat{\phi}$	0.9431	-0.0569	0.0369	3.7771	-0.2229	0.6802	7.6080	-0.3920	3.0023
		$\hat{\beta}_0$	1.0046	0.0046	0.0342	0.9958	-0.0042	0.1719	0.9912	-0.0088	0.3691
		$\hat{\beta}_1$	2.4973	-0.0027	0.0482	2.5060	0.0060	0.2297	2.5058	0.0058	0.4846
120	10%	$\hat{\phi}$	0.9831	-0.0169	0.0112	3.9458	-0.0542	0.2057	7.8949	-0.1051	0.9115
		$\hat{\beta}_0$	1.0019	0.0019	0.0110	0.9921	-0.0079	0.0536	0.9904	-0.0096	0.1154
		$\hat{\beta}_1$	2.4984	-0.0016	0.0162	2.5084	0.0084	0.0767	2.5084	0.0084	0.1619
400	10%	$\hat{\phi}$	0.9962	-0.0038	0.0033	3.9885	-0.0115	0.0571	7.9801	-0.0199	0.2540
		$\hat{\beta}_0$	1.0002	0.0002	0.0032	0.9976	-0.0024	0.0166	0.9977	-0.0023	0.0364
		$\hat{\beta}_1$	2.5009	0.0009	0.0049	2.5046	0.0046	0.0235	2.5047	0.0047	0.0502
40	30%	$\hat{\phi}$	0.9432	-0.0568	0.0515	3.7876	-0.2124	1.0016	7.6435	-0.3565	4.6024
		$\hat{\beta}_0$	1.0114	0.0114	0.0392	1.0076	0.0076	0.1928	1.0068	0.0068	0.4179
		$\hat{\beta}_1$	2.4882	-0.0118	0.0552	2.4903	-0.0097	0.2645	2.4846	-0.0154	0.5663
120	30%	$\hat{\phi}$	0.9823	-0.0177	0.0159	3.9491	-0.0509	0.2982	7.9088	-0.0912	1.3537
		$\hat{\beta}_0$	1.0034	0.0034	0.0132	0.9942	-0.0058	0.0667	0.9927	-0.0073	0.1432
		$\hat{\beta}_1$	2.4956	-0.0044	0.0195	2.5055	0.0055	0.0976	2.5052	0.0052	0.2078
400	30%	$\hat{\phi}$	0.9963	-0.0037	0.0047	3.9894	-0.0106	0.0830	7.9838	-0.0162	0.3745
		$\hat{\beta}_0$	1.0011	0.0011	0.0038	0.9987	-0.0013	0.0194	0.9991	-0.0009	0.0422
		$\hat{\beta}_1$	2.4998	-0.0002	0.0056	2.5033	0.0033	0.0278	2.5029	0.0029	0.0596

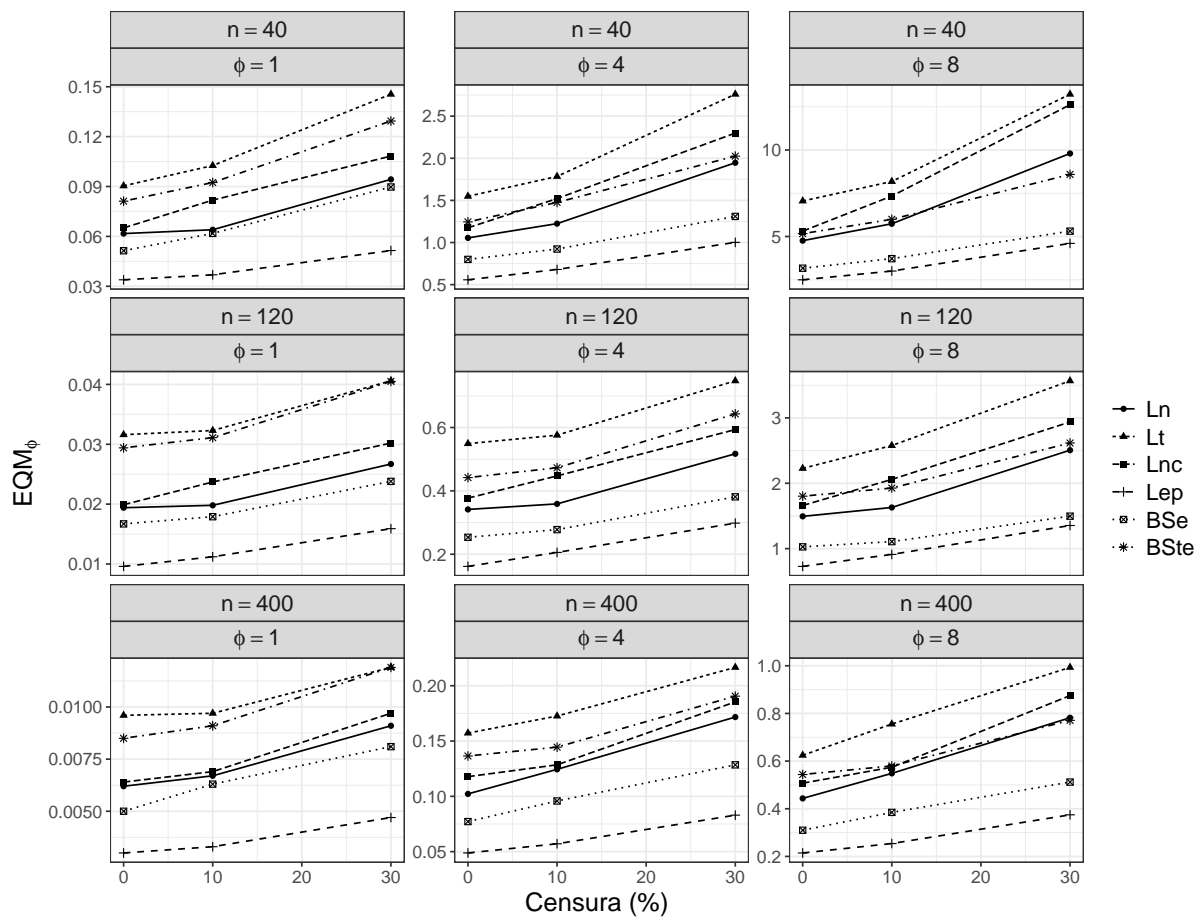


**Tabela 4.5:** Valores de média, viés e EQM para dados simulados da distribuição Birnbaum-Saunders estendida discreta considerando os estimadores de máxima verossimilhança do modelo de regressão ( $\beta_0 = 1$ ,  $\beta_1 = 2.5$  e  $\xi = 0.5$ ).

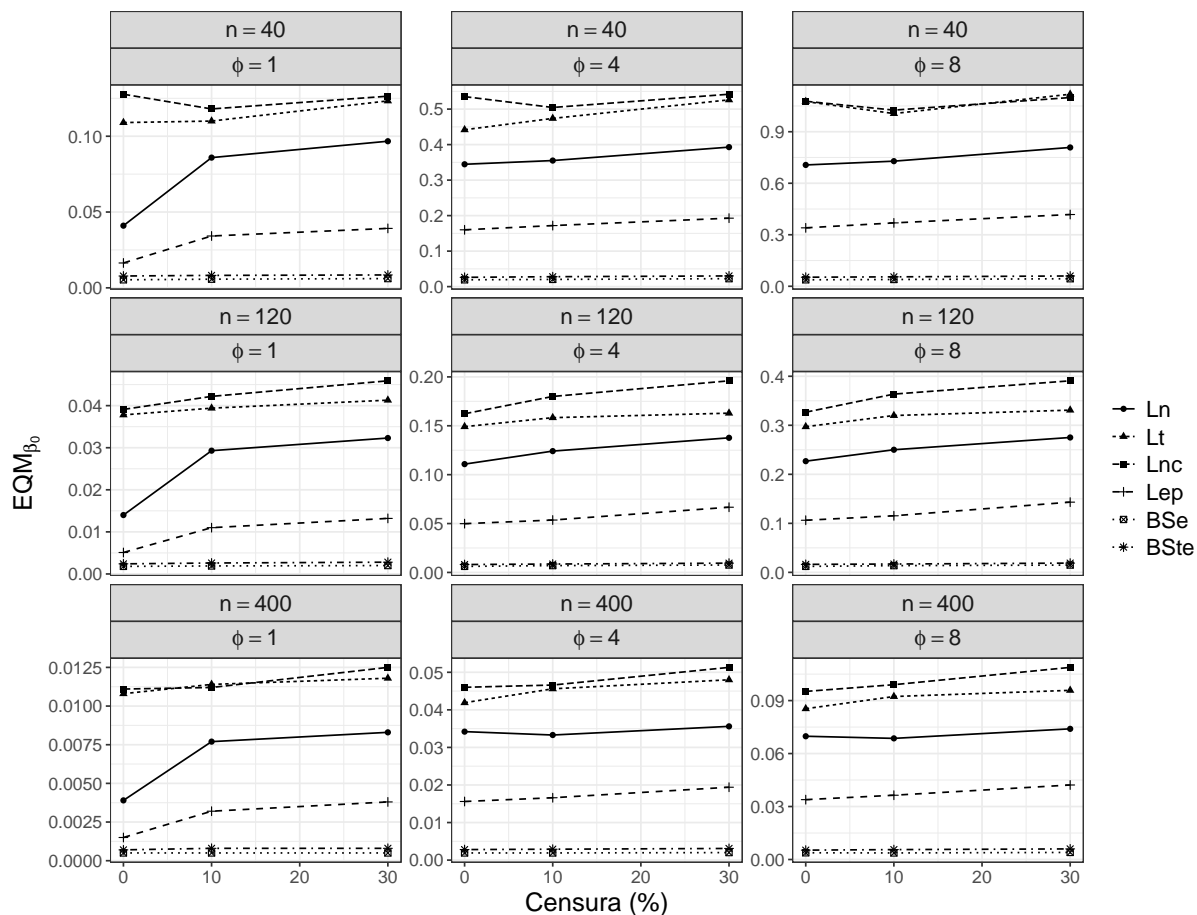
$n$	Cen.	$\phi = 1$			$\phi = 4$			$\phi = 8$			
		Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM	
40	0%	$\hat{\phi}$	0.9622	-0.0378	0.0514	3.8305	-0.1695	0.7996	7.6936	-0.3064	3.1667
		$\hat{\beta}_0$	1.0001	0.0001	0.0053	1.0045	0.0045	0.0192	1.0045	0.0045	0.0377
		$\hat{\beta}_1$	2.5002	0.0002	0.0079	2.4961	-0.0039	0.0289	2.4964	-0.0036	0.0573
120	0%	$\hat{\phi}$	0.9903	-0.0097	0.0167	3.9535	-0.0465	0.2534	7.9214	-0.0786	1.0272
		$\hat{\beta}_0$	0.9997	-0.0003	0.0018	1.0002	0.0002	0.0062	1.0006	0.0006	0.0122
		$\hat{\beta}_1$	2.5006	0.0006	0.0026	2.5002	0.0002	0.0091	2.4999	-0.0001	0.0180
400	0%	$\hat{\phi}$	0.9974	-0.0026	0.005	3.9882	-0.0118	0.0772	7.9807	-0.0193	0.3098
		$\hat{\beta}_0$	0.9998	-0.0002	0.0005	1.0002	0.0002	0.0019	0.9999	-0.0001	0.0038
		$\hat{\beta}_1$	2.5003	0.0003	0.0007	2.5000	0.0000	0.0028	2.5003	0.0003	0.0056
40	10%	$\hat{\phi}$	0.9670	-0.0330	0.0618	3.8460	-0.1540	0.9225	7.7258	-0.2742	3.7201
		$\hat{\beta}_0$	1.0017	0.0017	0.0057	1.0069	0.0069	0.0203	1.0091	0.0091	0.0395
		$\hat{\beta}_1$	2.4985	-0.0015	0.0080	2.4933	-0.0067	0.0294	2.4912	-0.0088	0.0570
120	10%	$\hat{\phi}$	0.9894	-0.0106	0.0179	3.9524	-0.0476	0.2777	7.9074	-0.0926	1.1073
		$\hat{\beta}_0$	1.0023	0.0023	0.0019	1.0035	0.0035	0.0070	1.0054	0.0054	0.0137
		$\hat{\beta}_1$	2.4975	-0.0025	0.0027	2.4962	-0.0038	0.0102	2.4939	-0.0061	0.0201
400	10%	$\hat{\phi}$	0.9944	-0.0056	0.0063	3.9800	-0.0200	0.0958	7.9663	-0.0337	0.3848
		$\hat{\beta}_0$	0.9995	-0.0005	0.0005	0.9995	-0.0005	0.0019	0.9989	-0.0011	0.0037
		$\hat{\beta}_1$	2.5006	0.0006	0.0007	2.5007	0.0007	0.0028	2.5013	0.0013	0.0055
40	30%	$\hat{\phi}$	0.9740	-0.0260	0.0897	3.8667	-0.1333	1.3096	7.7840	-0.2160	5.3092
		$\hat{\beta}_0$	1.0025	0.0025	0.0061	1.0101	0.0101	0.0224	1.0136	0.0136	0.0439
		$\hat{\beta}_1$	2.4985	-0.0015	0.0087	2.4911	-0.0089	0.0330	2.4887	-0.0113	0.0646
120	30%	$\hat{\phi}$	0.9878	-0.0122	0.0238	3.9480	-0.0520	0.3812	7.9023	-0.0977	1.4965
		$\hat{\beta}_0$	1.0022	0.0022	0.0020	1.0031	0.0031	0.0076	1.0046	0.0046	0.0149
		$\hat{\beta}_1$	2.4975	-0.0025	0.0030	2.4965	-0.0035	0.0114	2.4950	-0.0050	0.0225
400	30%	$\hat{\phi}$	0.9973	-0.0027	0.0081	3.9886	-0.0114	0.1285	7.9953	-0.0047	0.5119
		$\hat{\beta}_0$	0.9996	-0.0004	0.0005	0.9994	-0.0006	0.0020	0.9996	-0.0004	0.0040
		$\hat{\beta}_1$	2.5009	0.0009	0.0008	2.5013	0.0013	0.0030	2.5017	0.0017	0.0062

**Tabela 4.6:** Valores de média, viés e EQM para dados simulados da distribuição Birnbaum-Saunders- $t$  estendida discreta considerando os estimadores de máxima verossimilhança do modelo de regressão ( $\beta_0 = 1$ ,  $\beta_1 = 2.5$  e  $\xi = (0.5, 4)^\top$ ).

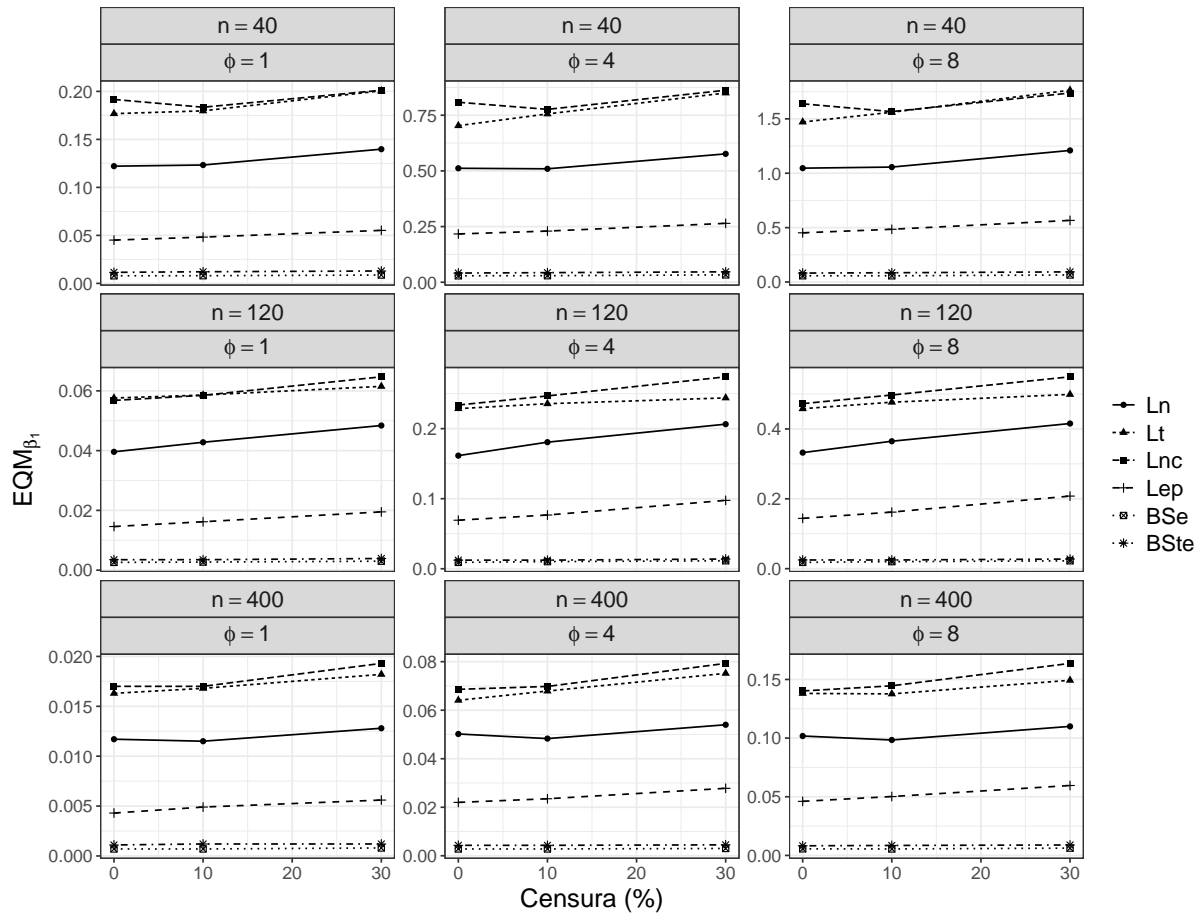
$n$	Cen.		$\phi = 1$			$\phi = 4$			$\phi = 8$		
			Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
40		$\hat{\phi}$	0.9716	-0.0284	0.0812	3.8890	-0.1110	1.2447	7.7889	-0.2111	5.1651
		$\hat{\beta}_0$	1.0013	0.0013	0.0078	1.0045	0.0045	0.0262	1.0058	0.0058	0.0528
		$\hat{\beta}_1$	2.4969	-0.0031	0.0115	2.4918	-0.0082	0.0413	2.4890	-0.0110	0.0822
120	0%	$\hat{\phi}$	1.0015	0.0015	0.0294	4.0035	0.0035	0.4419	8.0026	0.0026	1.8024
		$\hat{\beta}_0$	1.0009	0.0009	0.0024	1.0027	0.0027	0.0081	1.0057	0.0057	0.0163
		$\hat{\beta}_1$	2.4985	-0.0015	0.0035	2.4960	-0.0040	0.0123	2.4925	-0.0075	0.0249
400		$\hat{\phi}$	1.0006	0.0006	0.0085	3.9995	-0.0005	0.1365	7.9946	-0.0054	0.5438
		$\hat{\beta}_0$	0.9991	-0.0009	0.0007	0.9979	-0.0021	0.0028	0.9976	-0.0024	0.0053
		$\hat{\beta}_1$	2.5015	0.0015	0.0011	2.5033	0.0033	0.0043	2.5041	0.0041	0.0082
40		$\hat{\phi}$	0.9878	-0.0122	0.0924	3.9628	-0.0372	1.4762	7.9412	-0.0588	5.9928
		$\hat{\beta}_0$	0.9966	-0.0034	0.0082	0.9968	-0.0032	0.0279	0.9942	-0.0058	0.0549
		$\hat{\beta}_1$	2.5017	0.0017	0.0121	2.4997	-0.0003	0.0431	2.5006	0.0006	0.0851
120	10%	$\hat{\phi}$	1.0032	0.0032	0.0311	4.0185	0.0185	0.4730	8.0396	0.0396	1.9264
		$\hat{\beta}_0$	1.0021	0.0021	0.0026	1.0037	0.0037	0.0086	1.0058	0.0058	0.0170
		$\hat{\beta}_1$	2.4968	-0.0032	0.0035	2.4941	-0.0059	0.0123	2.4912	-0.0088	0.0248
400		$\hat{\phi}$	0.9974	-0.0026	0.0091	3.9940	-0.006	0.1445	7.9759	-0.0241	0.5798
		$\hat{\beta}_0$	0.9998	-0.0002	0.0008	1.0002	0.0002	0.0029	1.0014	0.0014	0.0056
		$\hat{\beta}_1$	2.5000	0.0000	0.0012	2.4995	-0.0005	0.0043	2.4982	-0.0018	0.0085
40		$\hat{\phi}$	0.9855	-0.0145	0.1294	3.9710	-0.0290	2.0250	7.9753	-0.0247	8.5906
		$\hat{\beta}_0$	0.9962	-0.0038	0.0085	0.9984	-0.0016	0.0301	0.9955	-0.0045	0.0600
		$\hat{\beta}_1$	2.5024	0.0024	0.0129	2.4987	-0.0013	0.0467	2.5001	0.0001	0.0931
120	30%	$\hat{\phi}$	1.0086	0.0086	0.0405	4.0467	0.0467	0.6430	8.0944	0.0944	2.6162
		$\hat{\beta}_0$	1.0031	0.0031	0.0028	1.0051	0.0051	0.0095	1.0079	0.0079	0.0189
		$\hat{\beta}_1$	2.4956	-0.0044	0.0039	2.4922	-0.0078	0.0139	2.4884	-0.0116	0.0275
400		$\hat{\phi}$	0.9985	-0.0015	0.0119	3.9980	-0.0020	0.1906	7.9885	-0.0115	0.7710
		$\hat{\beta}_0$	1.0002	0.0002	0.0008	1.0010	0.0010	0.0031	1.0033	0.0033	0.0060
		$\hat{\beta}_1$	2.4997	-0.0003	0.0012	2.4985	-0.0015	0.0045	2.4957	-0.0043	0.0089



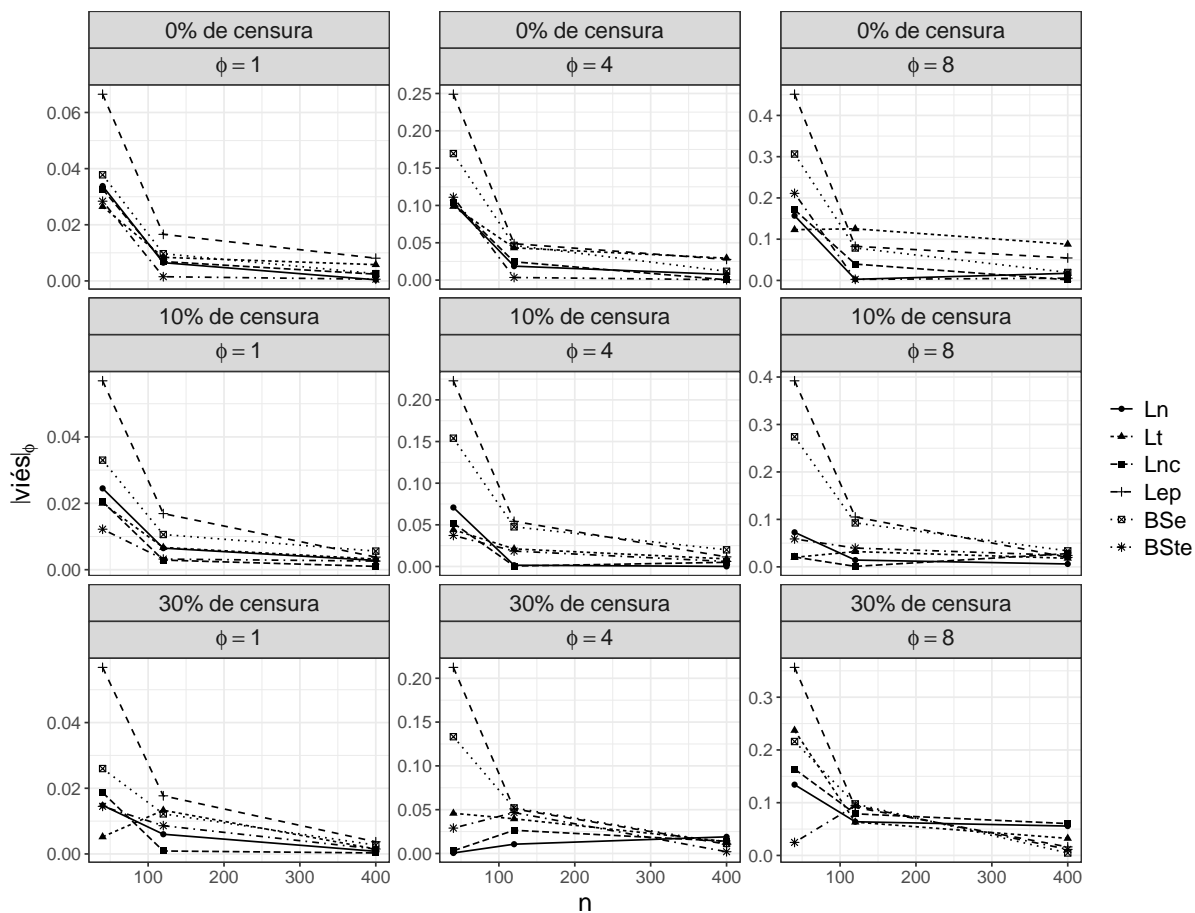
**Figura 4.1:** Comportamento do erro quadrático médio para as estimativas  $\hat{\phi}$  segundo a censura, valores verdadeiros do parâmetro  $\phi$  e tamanho da amostra.



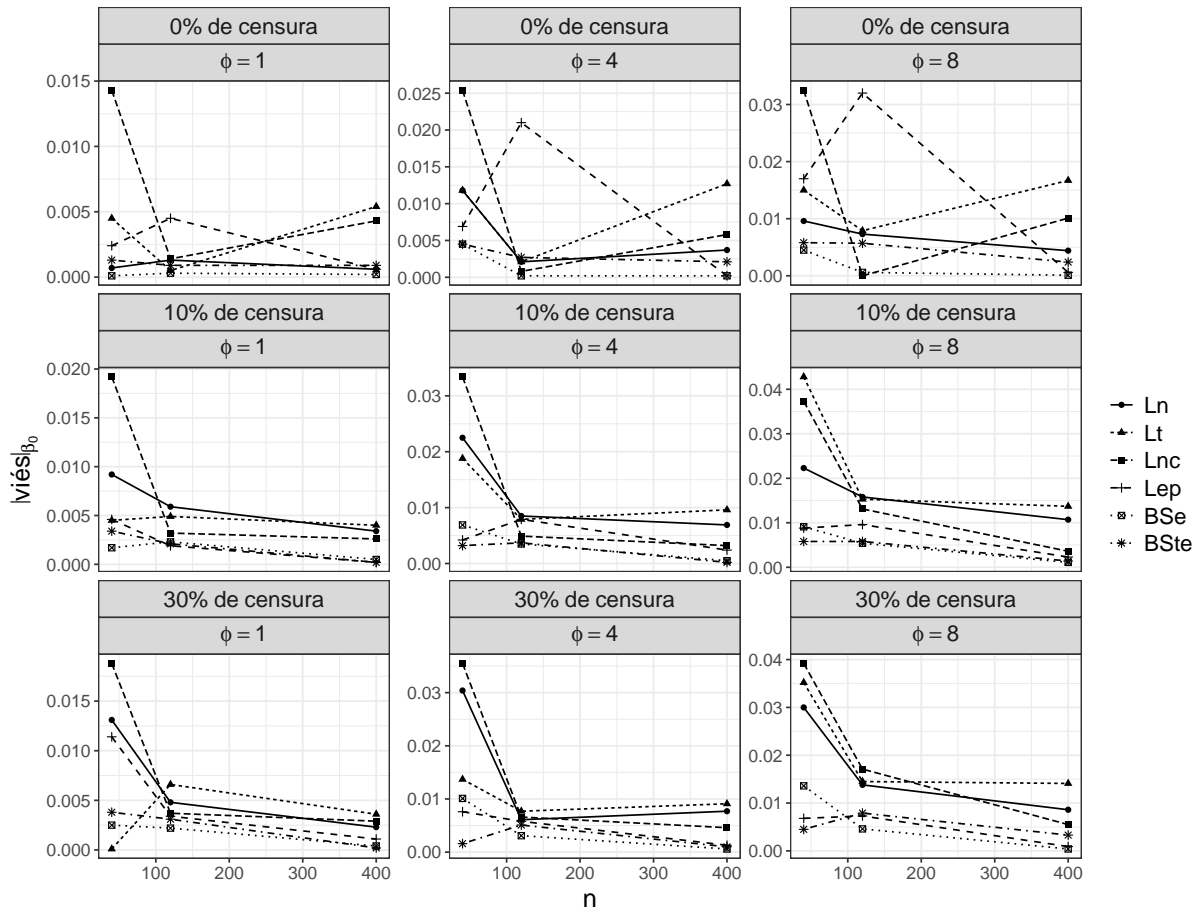
**Figura 4.2:** Comportamento do erro quadrático médio para as estimativas do intercepto  $\hat{\beta}_0$  segundo a censura, valores verdadeiros do parâmetro  $\phi$  e tamanho da amostra.



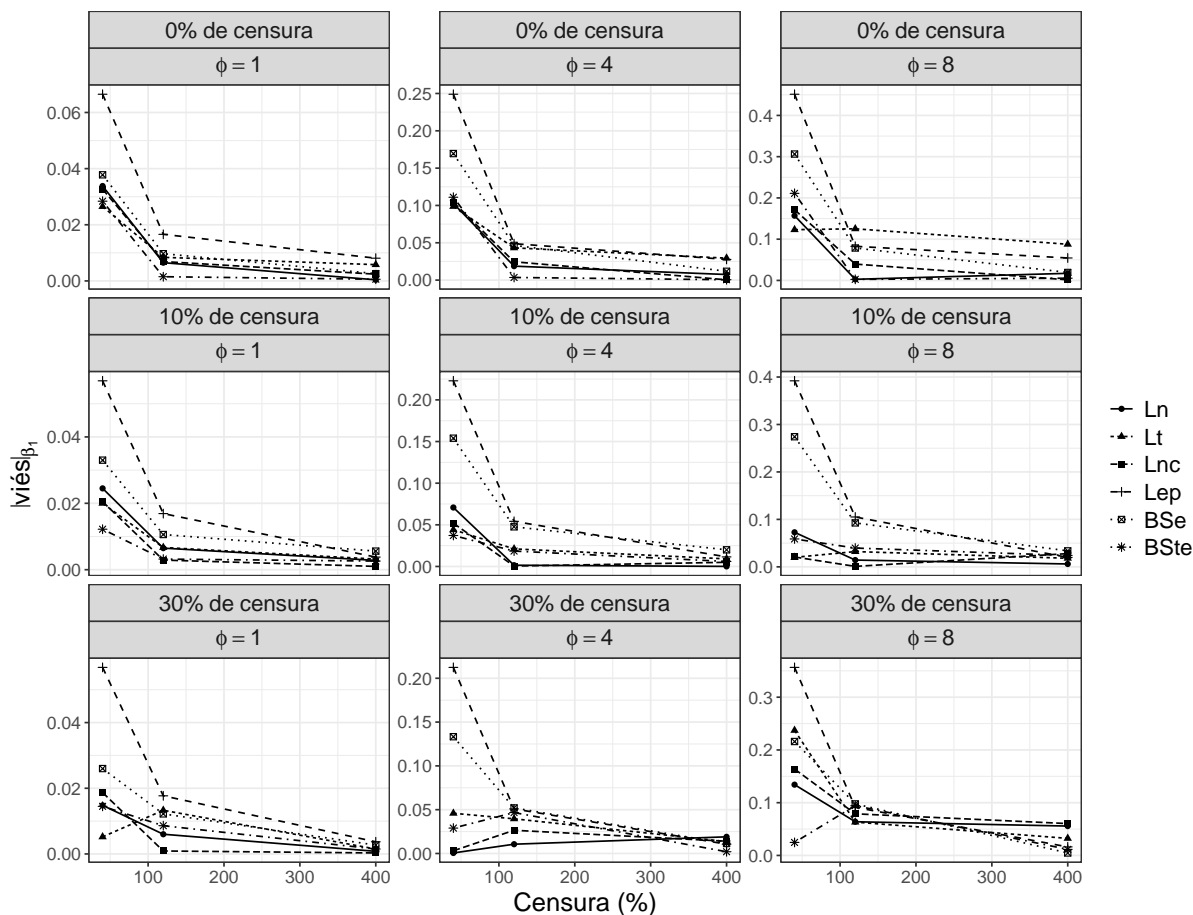
**Figura 4.3:** Comportamento do erro quadrático médio para as estimativas  $\hat{\beta}_1$  segundo a censura, valores verdadeiros do parâmetro  $\phi$  e tamanho da amostra.



**Figura 4.4:** Comportamento do viés para as estimativas  $\hat{\phi}$  segundo a censura, valores verdadeiros do parâmetro  $\phi$  e tamanho da amostra.



**Figura 4.5:** Comportamento do viés para as estimativas  $\hat{\beta}_0$  segundo a censura, valores verdadeiros do parâmetro  $\phi$  e tamanho da amostra.



**Figura 4.6:** Comportamento do viés para as estimativas  $\hat{\beta}_1$  segundo a censura, valores verdadeiros do parâmetro  $\phi$  e tamanho da amostra.

#### 4.4 Aplicações

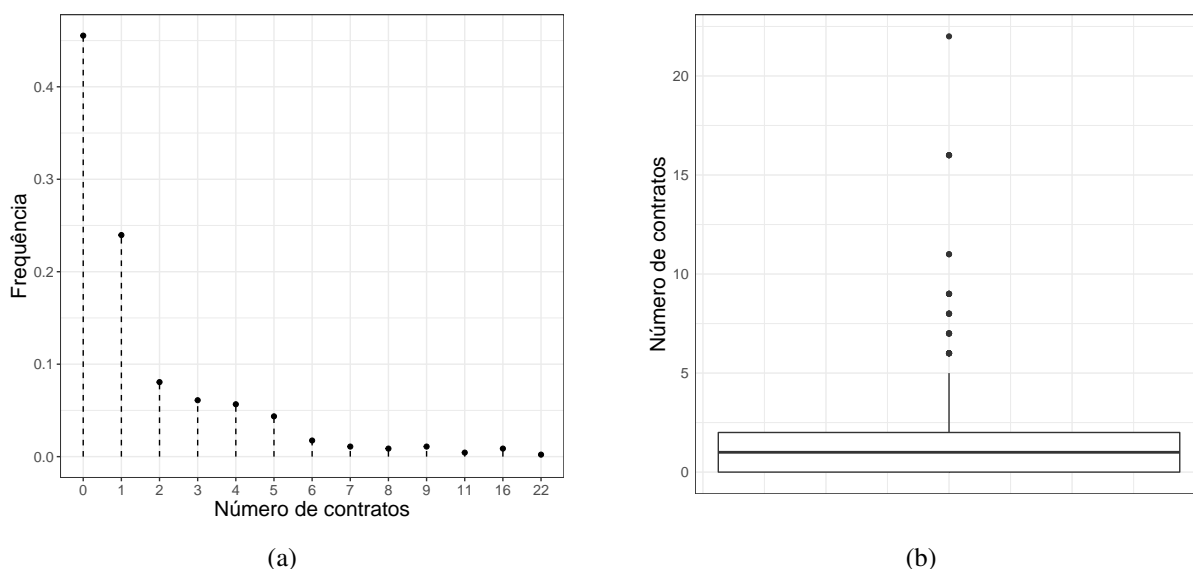
Nesta seção serão estudadas aplicações dos modelos de regressão log-simétricos discretos a dados reais não censurados e censurados. Os dados sem censura são referentes ao número de projetos de infraestrutura em contratos de concessões (concessões e parcerias público privadas) realizados em todos os estados brasileiros entre 2000 e 2017 (Banco Mundial). Esses dados foram estudados por Fernandez (2019) em aplicações do modelo de Poisson inflado de zeros (ZIP) na estrutura de dados de painel. Neste trabalho, não há controle para grupos de indivíduos em cada ano, caracterizando o conjunto de dados como uma seção transversal agrupada



(*pooled cross-section*). Os dados censurados são referentes à evasão no curso de Estatística da Universidade Federal do Piauí, já descritos na Seção 3.5.2.

#### 4.4.1 Dados não censurados

A variável resposta correspondente ao número de projetos em contratos de concessões apresenta as seguintes medidas descritivas: 459 (tamanho da amostra), 0 (mínimo), 22 (máximo), 1 (mediana), 1.6122 (média), 2.6337 (desvio padrão), 3.2039 (coeficiente de assimetria) e 14.8155 (coeficiente de curtose). Essas medidas sugerem assimetria à direita (média > mediana), confirmada pela Figura 4.7, e uma elevada curtose.



**Figura 4.7:** Distribuição do número de projetos (a) e boxplot (b).

Na Tabela 4.7 estão descritas as variáveis explicativas que podem estar relacionadas com o número de projetos em contratos e concessões. O resultado primário é a diferença entre as receitas e as despesas do estado e será considerado, neste trabalho, como um percentual do PIB (Produto Interno Bruto). O serviço da dívida pública (% PIB) refere-se à totalidade dos pagamentos que os estados realizaram para fins de custeio com juros e amortizações de dívidas contraídas por meio de empréstimos. As demais covariáveis são a razão entre as despesas com

pessoal e a receita corrente do estado e o logaritmo do tamanho da população do estado.

**Tabela 4.7:** Descrição das variáveis explicativas.

Tipo	Definição	Legenda	Fonte
Restrição orçamentária	Resultado Primário (% do PIB)	respripib	Secretaria do Tesouro Nacional
	Serviço da Dívida Pública (% do PIB)	servdivtotpib	
	Despesas com Pessoal/Receita Corrente	desprec	
Variável de controle	Número de habitantes	pop	IBGE

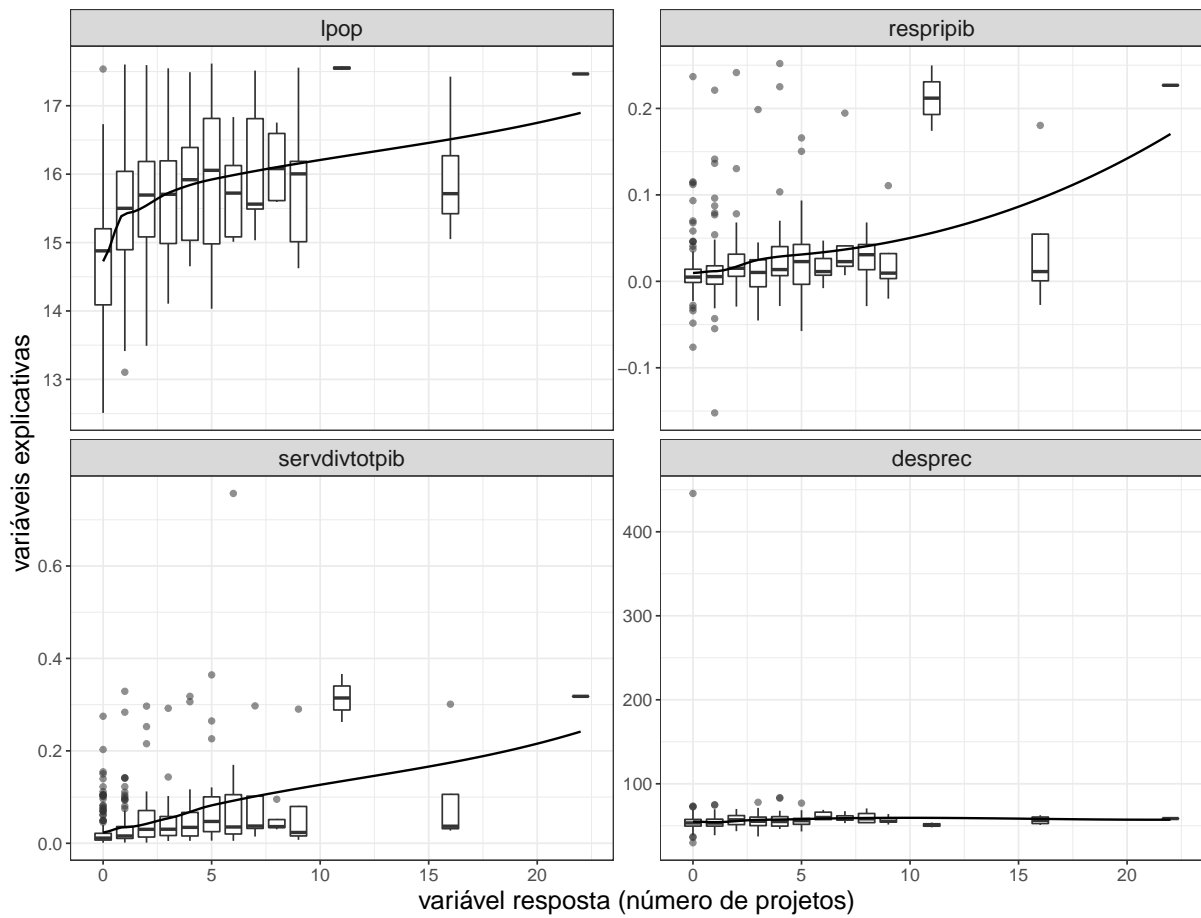
Fonte: Fernandez (2019)

O resumo descritivo com medidas de tendência central e dispersão das covariáveis está denotado na Tabela 4.8.

**Tabela 4.8:** Estatísticas descritivas para as variáveis explicativas do número de projetos.

Variáveis	Média	Desvio Padrão	Mediana	Mínimo	Máximo
lpop	15.2323	1.0764	15.1028	12.5099	17.6166
respripib	0.0168	0.0427	0.0077	-0.1522	0.2519
servdivtotpib	0.0411	0.0695	0.0165	0.0011	0.7570
desprec	55.4763	19.7065	54.2600	29.7900	445.6200

A Figura 4.8 apresenta o cruzamento entre as variáveis explicativas e a variável resposta. Nota-se que parece haver correlação positiva entre o número de projetos e as covariáveis tamanho da população (em logaritmo), resultado primário e serviço da dívida pública. Assim, um aumento nessas variáveis pode representar um aumento no número de projetos. A variável *desprec* apresenta um valor discrepante e não há relação aparente com o número de projetos, ou seja, não é observada nenhuma tendência.



**Figura 4.8:** Variáveis explicativas versus variáveis respostas.

A Tabela 4.9 denota os ajustes dos modelos log-simétricos discretos aos dados referentes ao número de projetos. Foram considerados os modelos discretos log-normal, log- $t$ -Student, log-normal-contaminado, log-exponencial-potência, BS estendido e BS- $t$  estendido.

**Tabela 4.9:** Modelos de regressão log-simétricos discretos para o número de projetos.

Modelo	Parâmetros	Estimativas	EP	$z$	$e^{\beta}$	IC (95%)	p-valor
Ln	$\beta_0$	-7.7517	1.0414	-7.4435	0.0004	0.0001–0.0033	0.0000
	$\beta_1$ (lpop)	0.5198	0.0693	7.5007	1.6817	1.4681–1.9264	0.0000
	$\beta_2$ (respripiB)	-0.7783	1.5249	-0.5104	0.4592	0.0231–9.1203	0.6098
	$\beta_3$ (servdivtotpiB)	0.9810	1.0474	0.9366	2.6671	0.3424–20.7782	0.3490
	$\beta_4$ (desprec)	-0.0011	0.0029	-0.3793	0.9989	0.9932–1.0046	0.7045
	$\phi$	0.9495	0.0930				
Lt	$\beta_0$	-7.6975	1.0313	-7.4639	0.0005	0.0001–0.0034	0.0000
	$\beta_1$ (lpop)	0.5164	0.0686	7.5277	1.6760	1.4651–1.9172	0.0000
	$\beta_2$ (respripiB)	-0.7796	1.5124	-0.5155	0.4586	0.0237–8.8880	0.6062
	$\beta_3$ (servdivtotpiB)	1.0487	1.0211	1.0270	2.8539	0.3857–21.1165	0.3044
	$\beta_4$ (desprec)	-0.0012	0.0029	-0.4138	0.9988	0.9931–1.0045	0.6790
	$\phi$	0.8762	0.0894				
	$\xi$	20.0					
Lnc	$\beta_0$	-7.7510	1.0413	-7.4436	0.0004	0.0001–0.0033	0.0000
	$\beta_1$ (lpop)	0.5195	0.0693	7.4964	1.6812	1.4677–1.9258	0.0000
	$\beta_2$ (respripiB)	-0.7784	1.5248	-0.5105	0.4591	0.0231–9.1176	0.6097
	$\beta_3$ (servdivtotpiB)	0.9818	1.0471	0.9376	2.6693	0.3428–20.7826	0.3485
	$\beta_4$ (desprec)	-0.0010	0.0029	-0.3448	0.9990	0.9933–1.0047	0.7302
	$\phi$	0.8634	0.0847				
	$\xi$	(0.90, 0.90)					
Lep	$\beta_0$	-7.8366	1.1376	-6.8887	0.0004	0.0000–0.0037	0.0000
	$\beta_1$ (lpop)	0.5254	0.0751	6.9960	1.6911	1.4597–1.9593	0.0000
	$\beta_2$ (respripiB)	-0.7335	1.4380	-0.5101	0.4802	0.0287–8.0444	0.6100
	$\beta_3$ (servdivtotpiB)	0.8531	0.9603	0.8884	2.3469	0.3573–15.4142	0.3743
	$\beta_4$ (desprec)	-0.0010	0.0027	-0.3704	0.9990	0.9937–1.0043	0.7111
	$\phi$	1.1188	0.0454				
	$\xi$	-0.10					
BSe	$\beta_0$	-8.3549	1.2344	-6.7684	0.0002	0.0000–0.0026	0.0000
	$\beta_1$ (lpop)	0.5607	0.0798	7.0263	1.7519	1.4982–2.0485	0.0000
	$\beta_2$ (respripiB)	-0.7984	1.5353	-0.5200	0.4500	0.0222–9.1229	0.6031
	$\beta_3$ (servdivtotpiB)	0.3192	0.9949	0.3208	1.3760	0.1958–9.6717	0.7484
	$\beta_4$ (desprec)	-0.0009	0.0046	-0.1957	0.9991	0.9901–1.0081	0.8448
	$\phi$	1.8279	0.1259				
	$\xi$	1.9					
BSte	$\beta_0$	-7.8390	1.1129	-7.0438	0.0004	0.0000–0.0035	0.0000
	$\beta_1$ (lpop)	0.5255	0.0744	7.0632	1.6913	1.4618–1.9568	0.0000
	$\beta_2$ (respripiB)	-0.7618	1.5348	-0.4964	0.4668	0.0231–9.4537	0.6196
	$\beta_3$ (servdivtotpiB)	0.8652	0.9621	0.8993	2.3755	0.3604–15.6570	0.3685
	$\beta_4$ (desprec)	-0.0010	0.0031	-0.3226	0.9990	0.9929–1.0051	0.7470
	$\phi$	5.0153	0.0995				
	$\xi$	(0.9, 20.0)					

EP: Erro-Padrão

Dentre os modelos estudados, se forem considerados os critérios AIC e BIC para seleção, os modelos log-exponencial-potência e log-normal seriam os escolhidos, pois apresentam as menores medidas (Tabela 4.10).

**Tabela 4.10:** Critérios AIC e BIC para seleção dos modelos.

Medida	Ln	Lt	Lnc	Lep	BSe	BSte
AIC	1470.765	1472.711	1472.779	1470.195	1485.569	1472.045
BIC	1499.669	1501.615	1505.811	1499.098	1514.472	1505.077

#### 4.4.2 Análise do ajuste dos modelos

Para mensurar o afastamento das estimativas dos modelos em relação aos valores observados do número de contratos levou-se em consideração o Erro Absoluto Médio (EAM) dado pela seguinte equação:

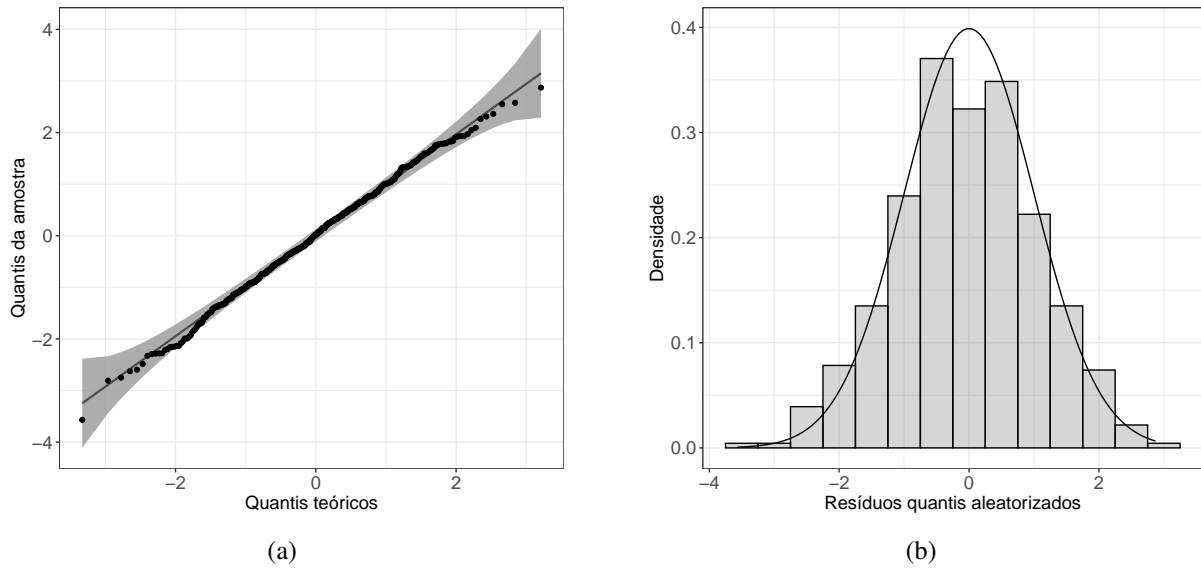
$$\text{EAM} = \frac{\sum_{i=1}^n |\text{valor observado}_i - \text{valor estimado}_i|}{\text{número de observações}(n)}. \quad (4.9)$$

Desse modo, o ideal seria um valor bem pequeno para o EAM, indicando que os valores estimados se distanciam pouco dos valores observados, em média. Observando a Tabela 4.11, nota-se que os valores estimados para o número de contratos diferiram dos valores observados em 1.4 contrato, em média. O menor EAM foi observado no modelo log-*t*-Student (1.4452) e o maior no modelo Birnbaum-Saunders estendido (1.4550).

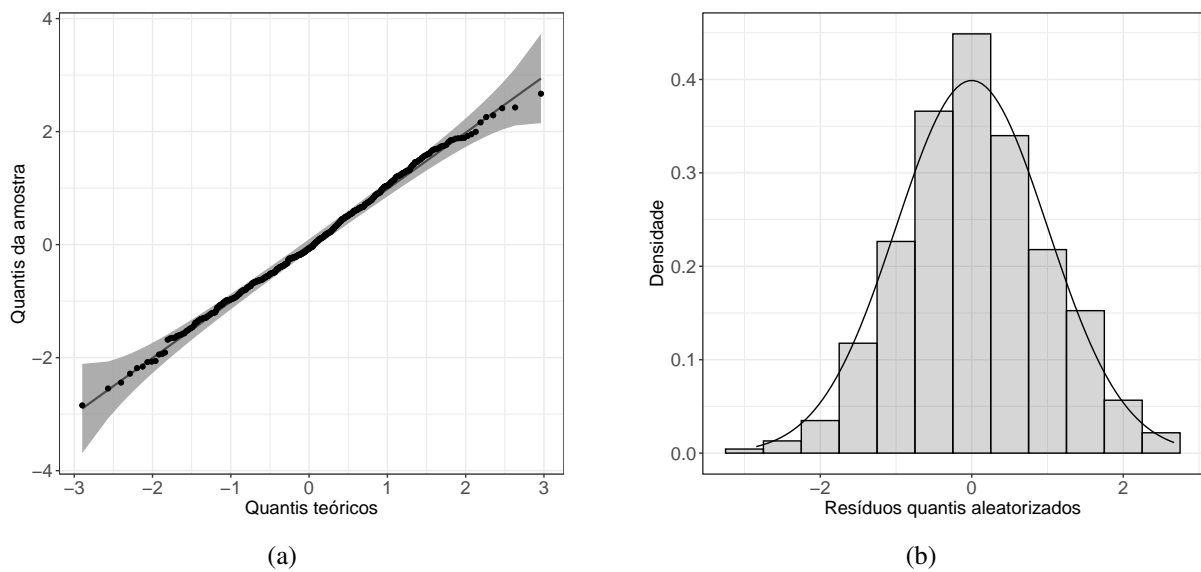
**Tabela 4.11:** Erro absoluto médio para os valores estimados do número de projetos.

Erro	Modelo					
	Ln	Lt	Lnc	Lep	BSe	BSte
Erro absoluto médio (EAM)	1.4461	1.4452	1.4464	1.4478	1.4550	1.4476

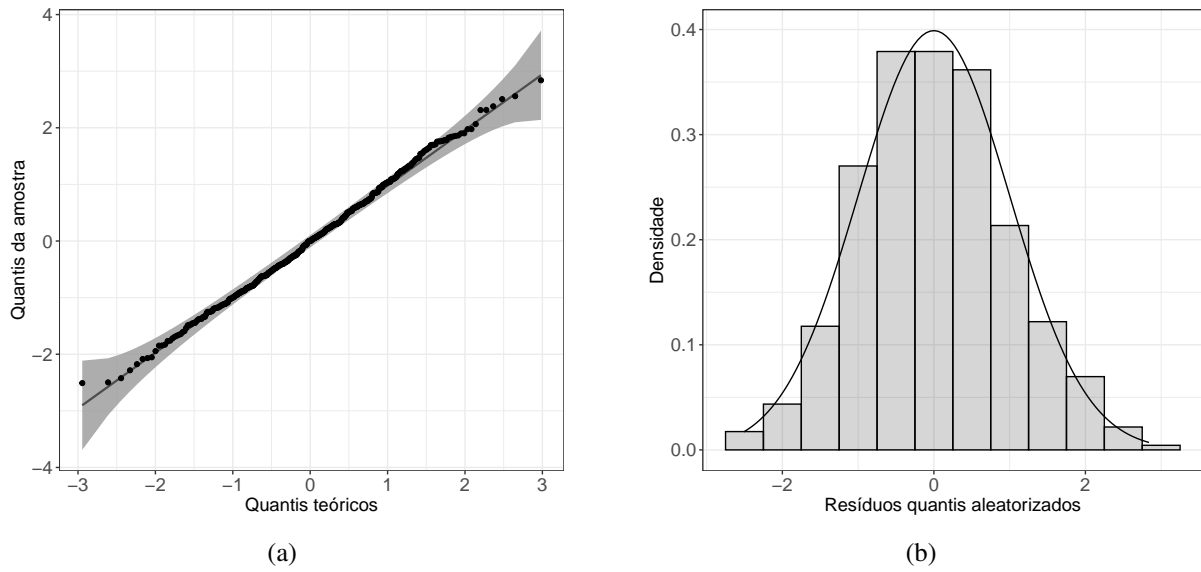
As Figuras 4.9–4.14 representam os resíduos quantílicos aleatorizados para os modelos estudados. Nota-se que, em geral, os modelos parecem estar bem ajustados, pois os resíduos apresentam boa aderência à distribuição normal padrão.



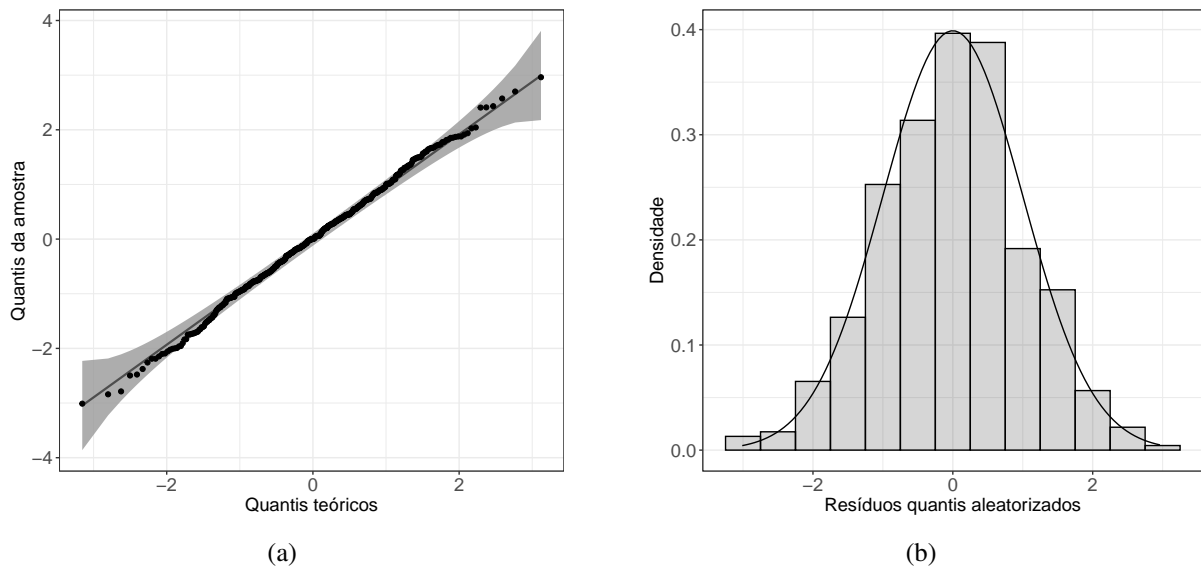
**Figura 4.9:** Gráfico quantil-quantil (a) e histograma (b) para os resíduos do modelo log-normal discreto.



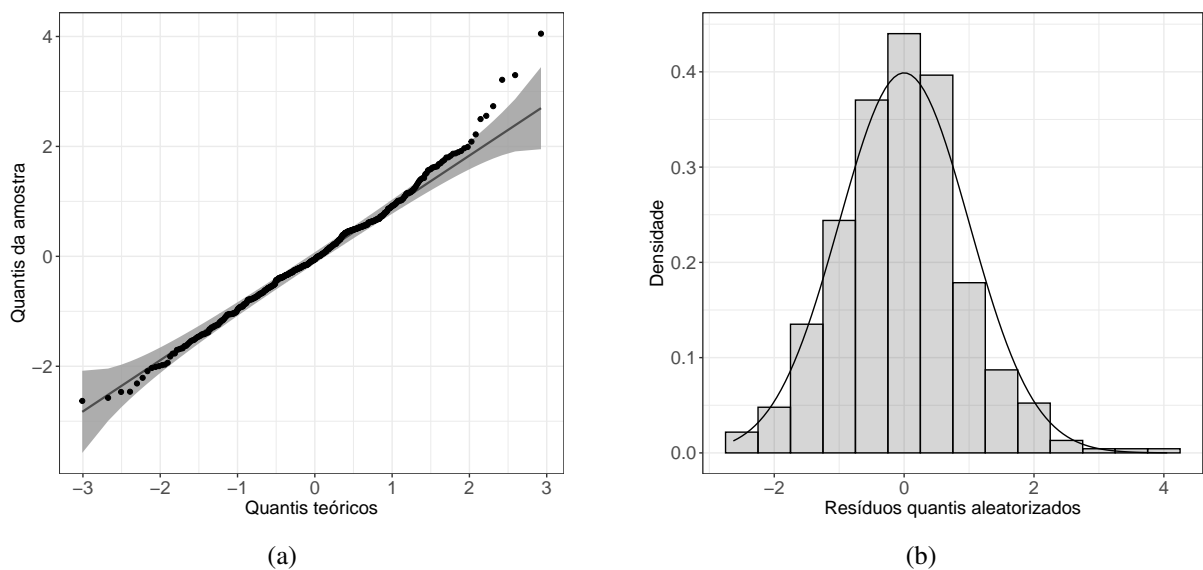
**Figura 4.10:** Gráfico quantil-quantil (a) e histograma (b) para os resíduos do modelo log- $t$ -Student discreto.



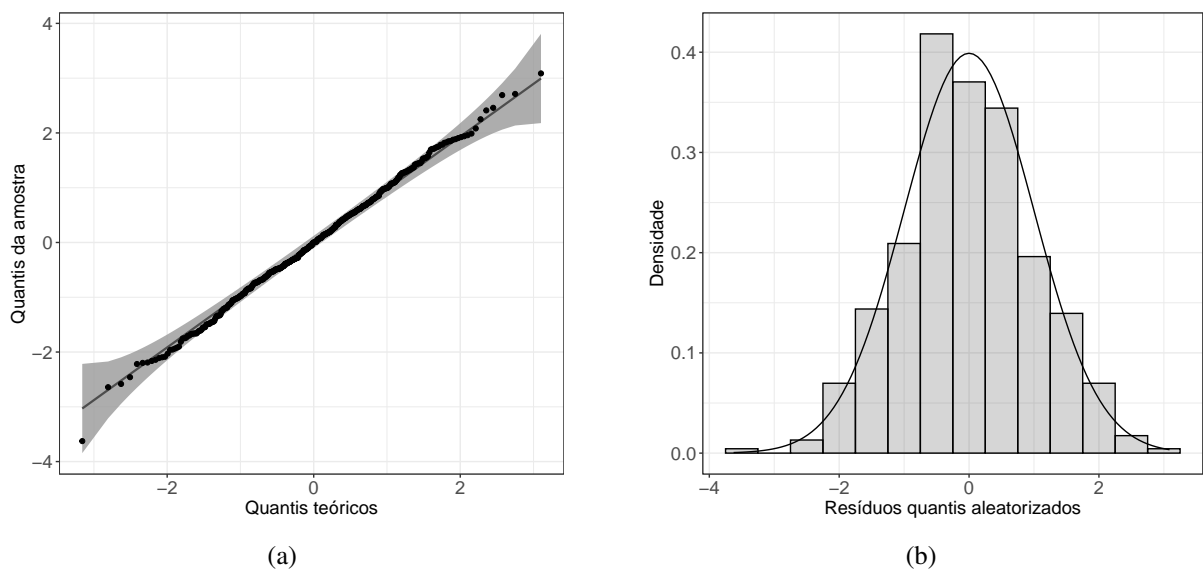
**Figura 4.11:** Gráfico quantil-quantil (a) e histograma (b) para os resíduos do modelo log-normal-contaminado discreto.



**Figura 4.12:** Gráfico quantil-quantil (a) e histograma (b) para os resíduos do modelo log-exponencial-potência discreto.



**Figura 4.13:** Gráfico quantil-quantil (a) e histograma (b) para os resíduos do modelo Birnbaum-Saunders estendido discreto.



**Figura 4.14:** Gráfico quantil-quantil (a) e histograma (b) para os resíduos do modelo Birnbaum-Saunders- $t$  estendido discreto.



### 4.4.3 Dados censurados

Nesta seção é estudado como se comporta a variável referente ao número de semestres até o aluno evadir na presença de algumas covariáveis. Essas covariáveis estão resumidas na Tabela 4.12. Nota-se que 72.77% dos indivíduos são do sexo masculino, 62.13% ingressaram no curso com idade inferior a 20 anos, 85.96% são oriundos da capital Teresina (sede do curso) e 85.11% ingressaram por meio do Enem.

**Tabela 4.12:** Perfil dos alunos do curso de Estatística (UFPI).

Variável	n	%
<b>Sexo</b>		
Feminino (0)	128	27.23
Masculino (1)	342	72.77
<b>Idade</b>		
Maior que 20 (0)	292	62.13
Menor ou igual a 20 (1)	178	37.87
<b>Local de origem</b>		
Teresina (0)	404	85.96
Outro município (1)	66	14.04
<b>Forma de ingresso</b>		
Enem (0)	400	85.11
Vestibular (1)	70	14.89

A Tabela 4.13 apresenta os valores dos testes de logrank e Wilcoxon. Esses testes são utilizados para verificar se as covariáveis exercem alguma influência sobre o tempo de sobrevivência. Desse modo, as hipóteses desses testes são:

$$\begin{cases} H_0 : \text{As curvas de sobrevivência são iguais,} \\ H_1 : \text{As curvas de sobrevivência não são iguais.} \end{cases}$$

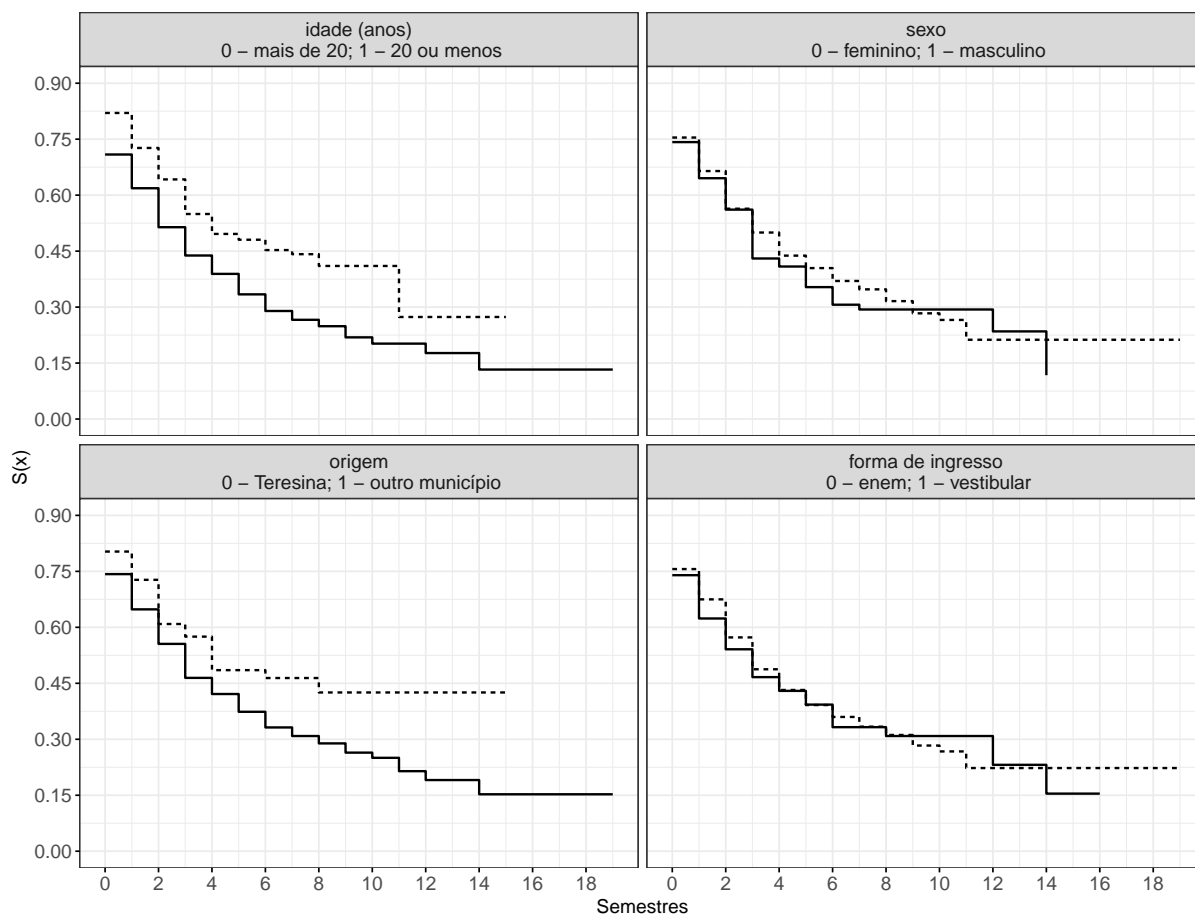
Neste trabalho, os testes de logrank e Wilcoxon foram utilizados como critério de inclusão da variável no modelo. Assim, a variável que apresentou um valor- $p$  de até 0.25 em pelo menos um dos testes citados foi incluída no modelo, seguindo as recomendações de Bendel e Afifi (1977) para regressão linear e de Mickey e Greenland (1989) para regressão logística. Portanto,

a única variável excluída da regressão foi o sexo.

**Tabela 4.13:** Testes de comparação das curvas de sobrevivência entre as categorias das variáveis explicativas.

Variável	Teste de logrank		Teste de Wilcoxon	
	Estatística	p-valor	Estatística	p-valor
Idade	11.57	0.0007	3.15	0.0016
Sexo	0.41	0.5224	0.64	0.5218
Origem	4.42	0.0355	-1.62	0.1046
Forma de ingresso	0.91	0.3391	-1.82	0.0686

A Figura 4.15 ilustra as formas das curvas de sobrevivência de Kaplan-Meier de cada categoria das covariáveis. Nota-se uma diferença mais acentuada entre as categorias das variáveis idade e origem, como confirmada pelos testes de logrank e Wilcoxon.



**Figura 4.15:** Curvas de sobrevivência de Kaplan-Meier para as covariáveis (— 0, - - - 1).

A Tabela 4.14 denota os ajustes dos modelos log-simétricos discretos aos dados referentes ao número de semestres até o aluno evadir. Foram considerados os modelos discretos log-normal, log- $t$ -Student, log-normal-contaminado, log-exponencial-potência, BS, BS- $t$ , BS estendido e BS- $t$  estendido.

**Tabela 4.14:** Estimativas dos parâmetros dos modelos de regressão log-simétricos discretos para o tempo até a evasão.

Modelo	Parâmetros	Estimativas	EP	$z$	$e^\beta$	IC (95%)	p-valor
Ln	$\beta_0$	0.9184	0.1292	7.1084	2.5053	1.9448–3.2273	0.0000
	$\beta_1$ (idade $\leq 20$ anos)	0.6766	0.1965	3.4433	1.9672	1.3384–2.8914	0.0006
	$\beta_2$ (origem: outra cidade)	0.5519	0.2737	2.0164	1.7365	1.0156–2.9694	0.0438
	$\beta_3$ (ingresso: vestibular)	0.3600	0.2559	1.4068	1.4333	0.8680–2.3669	0.1595
	$\phi$	3.2586	0.4132				
Lt	$\beta_0$	0.9251	0.1294	7.1491	2.5221	1.9571–3.2502	0.0000
	$\beta_1$ (idade $\leq 20$ anos)	0.6730	0.1969	3.4180	1.9601	1.3325–2.8833	0.0006
	$\beta_2$ (origem: outra cidade)	0.5525	0.2752	2.0076	1.7376	1.0132–2.9799	0.0447
	$\beta_3$ (ingresso: vestibular)	0.3478	0.2553	1.3623	1.4159	0.8585–2.3354	0.1731
	$\phi$	3.1364	0.4017				
Lnc	$\xi$	20.0					
	$\beta_0$	0.9185	0.1292	7.1091	2.5055	1.9450–3.2276	0.0000
	$\beta_1$ (idade $\leq 20$ anos)	0.6766	0.1965	3.4433	1.9672	1.3384–2.8914	0.0006
	$\beta_2$ (origem: outra cidade)	0.5519	0.2737	2.0164	1.7365	1.0156–2.9694	0.0438
	$\beta_3$ (ingresso: vestibular)	0.3599	0.2559	1.4064	1.4332	0.8679–2.3666	0.1596
Lep	$\phi$	3.2248	0.3759				
	$\xi$	(0.1, 0.9)					
	$\beta_0$	0.9044	0.1303	6.9409	2.4704	1.9136–3.1893	0.0000
	$\beta_1$ (idade $\leq 20$ anos)	0.6876	0.1921	3.5794	1.9889	1.3649–2.8983	0.0003
	$\beta_2$ (origem: outra cidade)	0.5533	0.2586	2.1396	1.7390	1.0475–2.8868	0.0324
BS	$\beta_3$ (ingresso: vestibular)	0.3817	0.2330	1.6382	1.4648	0.9278–2.3126	0.1014
	$\phi$	3.6315	0.2385				
	$\xi$	-0.1					
	$\beta_0$	0.8725	0.1625	5.3692	2.3929	1.7402–3.2904	0.0000
	$\beta_1$ (idade $\leq 20$ anos)	0.6667	0.2345	2.8431	1.9478	1.2301–3.0843	0.0045
BSe	$\beta_2$ (origem: outra cidade)	0.5059	0.3388	1.4932	1.6585	0.8537–3.2218	0.1354
	$\beta_3$ (ingresso: vestibular)	0.4424	0.2720	1.6265	1.5564	0.9133–2.6525	0.1038
	$\xi$	1.9					
	$\beta_0$	0.8699	0.0710	12.2521	2.3867	2.0766–2.7430	0.0000
	$\beta_1$ (idade $\leq 20$ anos)	0.6859	0.1172	5.8524	1.9856	1.5780–2.4983	0.0000
BSt	$\beta_2$ (origem: outra cidade)	0.5191	0.1272	4.0810	1.6805	1.3097–2.1563	0.0000
	$\beta_3$ (ingresso: vestibular)	0.4488	0.1314	3.4155	1.5664	1.2108–2.0266	0.0006
	$\phi$	4.1739	0.0170				
	$\xi$	1.9					
	$\beta_0$	0.9013	0.1758	5.1268	2.4628	1.7450–3.4760	0.0000
BSte	$\beta_1$ (idade $\leq 20$ anos)	0.6900	0.2468	2.7958	1.9937	1.2291–3.2340	0.0052
	$\beta_2$ (origem: outra cidade)	0.5591	0.3367	1.6605	1.7491	0.9041–3.3839	0.0968
	$\beta_3$ (ingresso: vestibular)	0.3850	0.3396	1.1337	1.4696	0.7553–2.8594	0.2569
	$\xi$	(1.7, 3.0)					
	$\beta_0$	0.9001	0.1226	7.3418	2.4598	1.9344–3.1280	0.0000
BSt	$\beta_1$ (idade $\leq 20$ anos)	0.6922	0.1869	3.7036	1.9981	1.3852–2.8821	0.0002
	$\beta_2$ (origem: outra cidade)	0.5636	0.2455	2.2957	1.7570	1.0859–2.8428	0.0217
	$\beta_3$ (ingresso: vestibular)	0.3817	0.2463	1.5497	1.4648	0.9039–2.3737	0.1212
	$\phi$	2.6712	0.2858				
	$\xi$	(2.0, 2.0)					

EP: Erro-Padrão

Pelos valores de AIC e BIC (Tabela 4.15), nota-se que o modelo selecionado seria o log-normal-contaminado, pois apresenta as menores medidas de AIC e BIC, sendo 1577.963 e

1598.727, respectivamente.

**Tabela 4.15:** Critérios AIC e BIC para seleção dos modelos.

Medida	Ln	Lt	Lnc	Lep	BS	BSe	BSt	BSte
AIC	1579.150	1579.177	1577.963	1579.116	1581.150	1579.827	1579.094	1581.053
BIC	1604.066	1604.094	1598.727	1604.033	1610.219	1604.743	1604.010	1610.122

Com base nas estimativas da Tabela 4.14, o valor estimado de  $\lambda$  associado ao número de semestres sob os modelos log-simétricos discretos é obtido por:

$$\hat{\lambda}(\text{idade}, \text{origem}, \text{ingresso}) = \exp(\hat{\beta}_0 + \hat{\beta}_1 \text{idade} + \hat{\beta}_2 \text{origem} + \hat{\beta}_3 \text{ingresso}), \quad (4.10)$$

em que os valores assumidos pelas covariáveis, como já visto, são: idade (0:  $> 20$  anos, 1:  $\leq 20$  anos), origem (0: Teresina, 1: outro município) e ingresso (0: Enem, 1: vestibular). Portanto, tem-se, para cada modelo, 8 tempos estimados associados a  $\lambda$ , conforme Tabela 4.16. Nota-se que o tempo até a evasão associado a  $\lambda$  se reduz quando o aluno ingressante tem mais de 20 anos, é oriundo da capital Teresina e ingressa por meio do Enem.

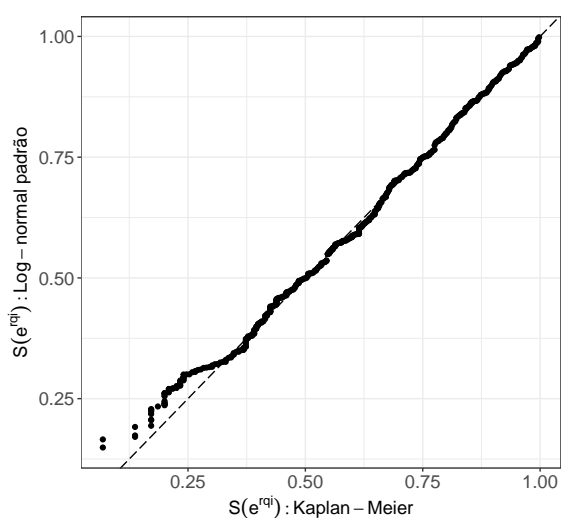
**Tabela 4.16:** Valores estimados de  $\lambda$  para o número de semestres até a evasão sob os modelos log-simétricos discretos.

Modelo	$\hat{\lambda}(0,0,0)$	$\hat{\lambda}(0,0,1)$	$\hat{\lambda}(0,1,0)$	$\hat{\lambda}(0,1,1)$	$\hat{\lambda}(1,0,0)$	$\hat{\lambda}(1,0,1)$	$\hat{\lambda}(1,1,0)$	$\hat{\lambda}(1,1,1)$
Ln	2.51	3.59	4.35	6.24	4.93	7.06	8.56	12.27
Lt	2.52	3.57	4.38	6.21	4.94	7.00	8.59	12.16
Lnc	2.51	3.59	4.35	6.24	4.93	7.06	8.56	12.27
Lep	2.47	3.62	4.30	6.29	4.91	7.20	8.54	12.52
BS	2.39	3.72	3.97	6.18	4.66	7.25	7.73	12.03
BSe	2.39	3.74	4.01	6.28	4.74	7.42	7.96	12.47
BSt	2.46	3.62	4.31	6.33	4.91	7.22	8.59	12.62
BSte	2.46	3.60	4.32	6.33	4.92	7.20	8.64	12.65

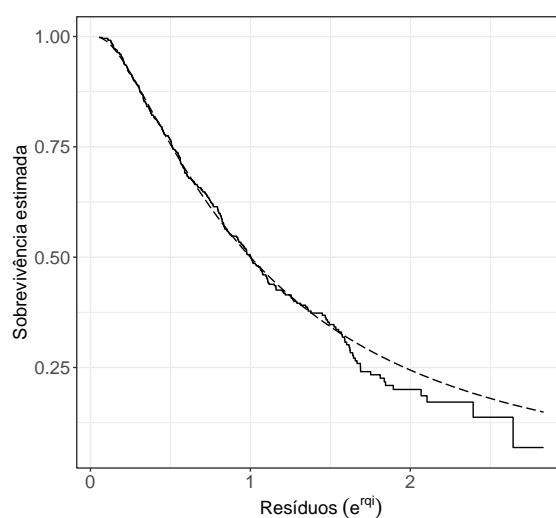
#### 4.4.4 Análise do ajuste dos modelos

Se os modelos log-simétricos discretos para o tempo até a evasão estiverem bem ajustados a esses dados, a distribuição exponencial dos resíduos quantílicos aleatorizados ( $r_{q,i}$ ) deve es-

tar próxima à distribuição log-normal padrão. De acordo com Colosimo e Giolo (2006), essa transformação é importante por ter distribuição conhecida (log-normal padrão) e por deixar os resíduos estritamente positivos, resolvendo problemas de estimação da função de sobrevivência por Kaplan-Meier. As Figuras 4.16–4.23 apresentam as sobrevivências dos resíduos dos respectivos modelos estimadas por Kaplan-Meier e pelo modelo log-normal padrão. Nota-se que os modelos log-simétricos discretos parecem bem ajustados aos dados do número de semestres até a evasão.

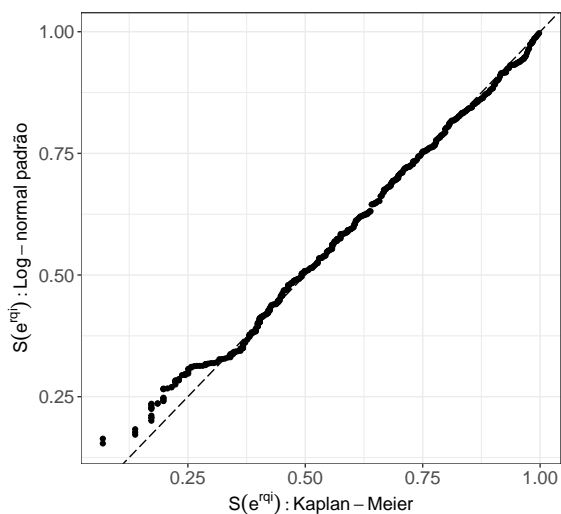


(a) Kaplan-Meier vs log-normal padrão.

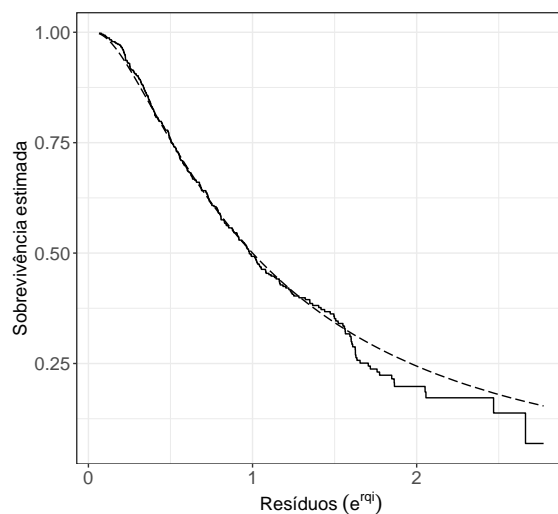


(b) Kaplan-Meier (—) e log-normal padrão (---).

**Figura 4.16:** Sobrevivências dos resíduos  $e^{r_{q,i}}$  do modelo log-normal discreto estimadas pelo método de Kaplan-Meier e pelo modelo log-normal padrão (a) e suas respectivas curvas estimadas (b).

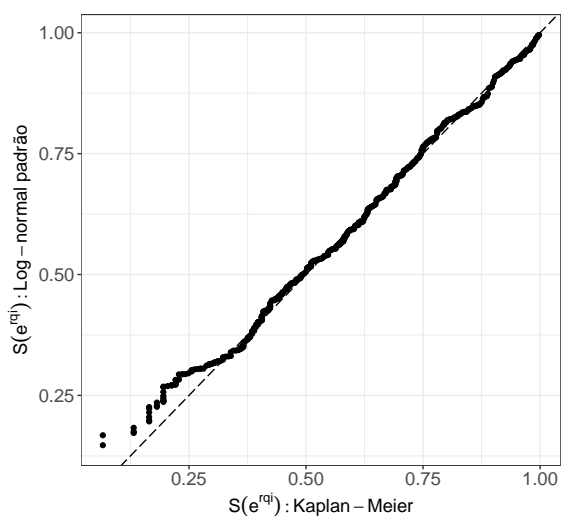


(a) Kaplan-Meier vs log-normal padrão.

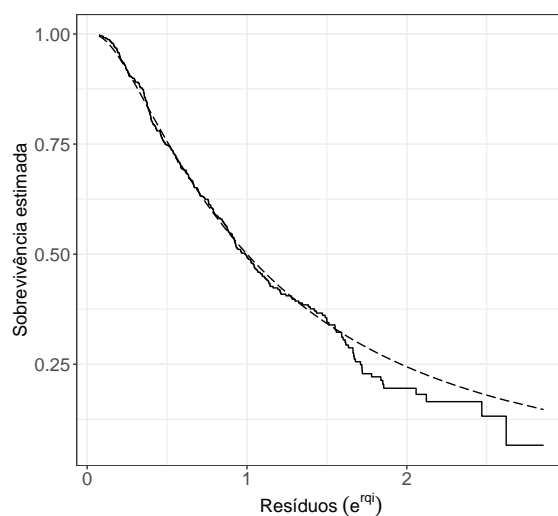


(b) Kaplan-Meier (—) e log-normal padrão (---).

**Figura 4.17:** Sobrevivências dos resíduos  $e^{r_{q,i}}$  do modelo log- $t$ -Student discreto estimadas pelo método de Kaplan-Meier e pelo modelo log-normal padrão (a) e suas respectivas curvas estimadas (b).

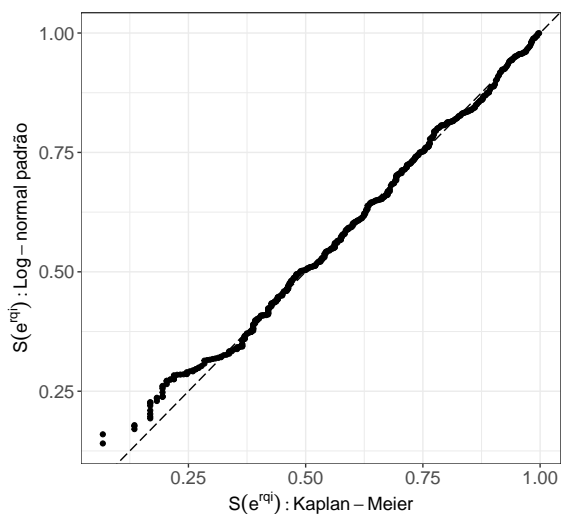


(a) Kaplan-Meier vs log-normal padrão.

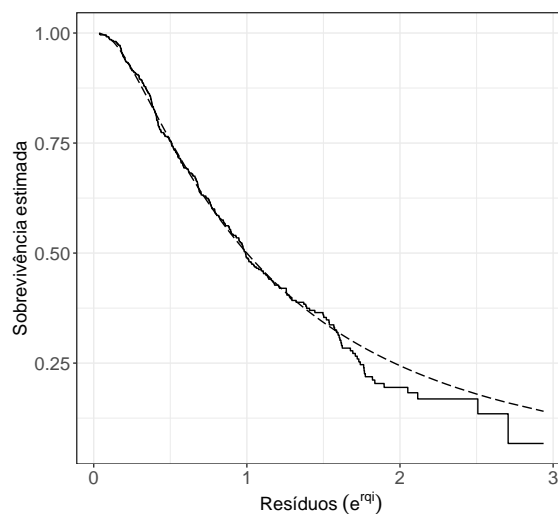


(b) Kaplan-Meier (—) e log-normal padrão (---).

**Figura 4.18:** Sobrevivências dos resíduos  $e^{r_{q,i}}$  do modelo log-normal-contaminado discreto estimadas pelo método de Kaplan-Meier e pelo modelo log-normal padrão (a) e suas respectivas curvas estimadas (b).

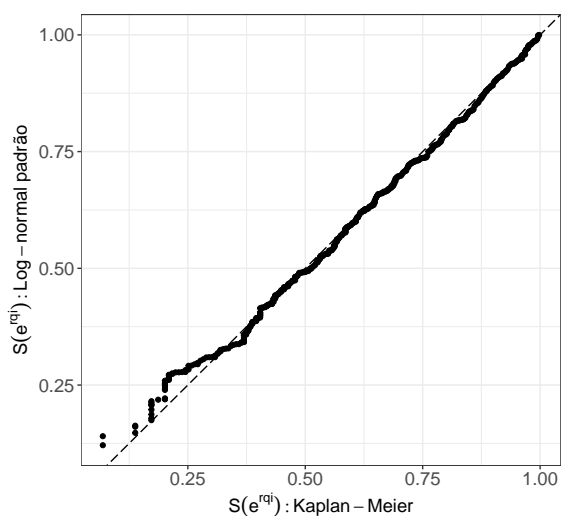


(a) Kaplan-Meier vs log-normal padrão.

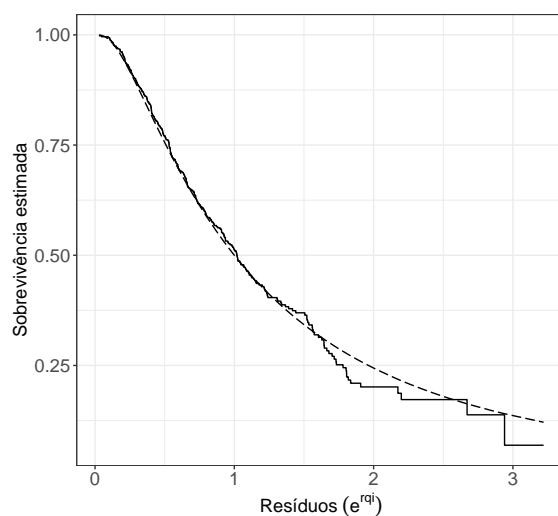


(b) Kaplan-Meier (—) e log-normal padrão (---).

**Figura 4.19:** Sobrevivências dos resíduos  $e^{r_{q,i}}$  do modelo log-exponencial-potência discreto estimadas pelo método de Kaplan-Meier e pelo modelo log-normal padrão (a) e suas respectivas curvas estimadas (b).



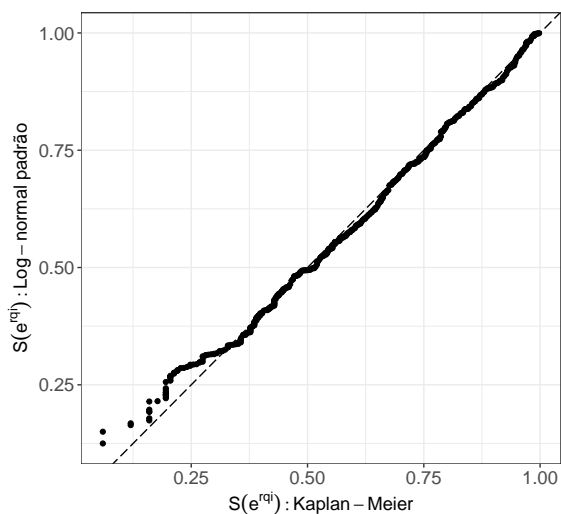
(a) Kaplan-Meier vs log-normal padrão.



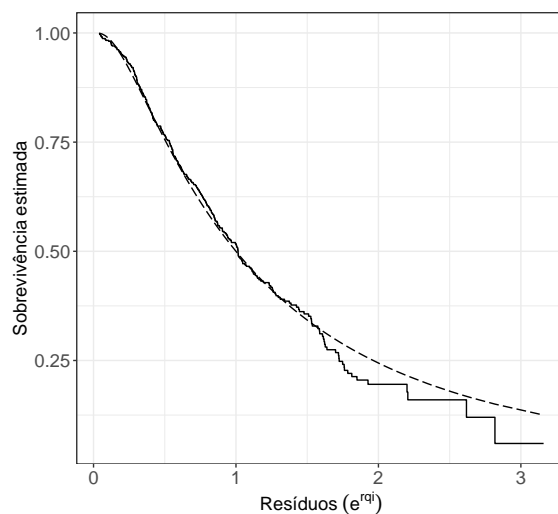
(b) Kaplan-Meier (—) e log-normal padrão (---).

**Figura 4.20:** Sobrevivências dos resíduos  $e^{r_{q,i}}$  do modelo BS discreto estimadas pelo método de Kaplan-Meier e pelo modelo log-normal padrão (a) e suas respectivas curvas estimadas (b).



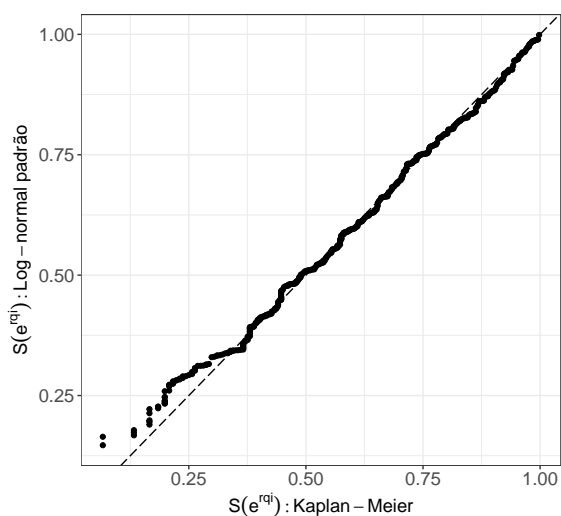


(a) Kaplan-Meier vs log-normal padrão.

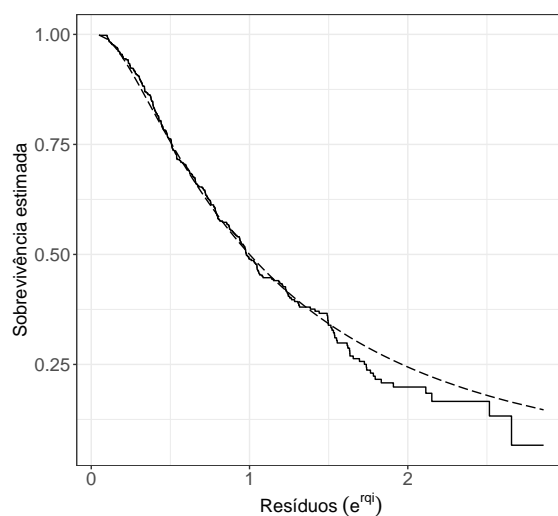


(b) Kaplan-Meier (—) e log-normal padrão (---).

**Figura 4.21:** Sobrevivências dos resíduos  $e^{r_{q,i}}$  do modelo BS estendido discreto estimadas pelo método de Kaplan-Meier e pelo modelo log-normal padrão (a) e suas respectivas curvas estimadas (b).

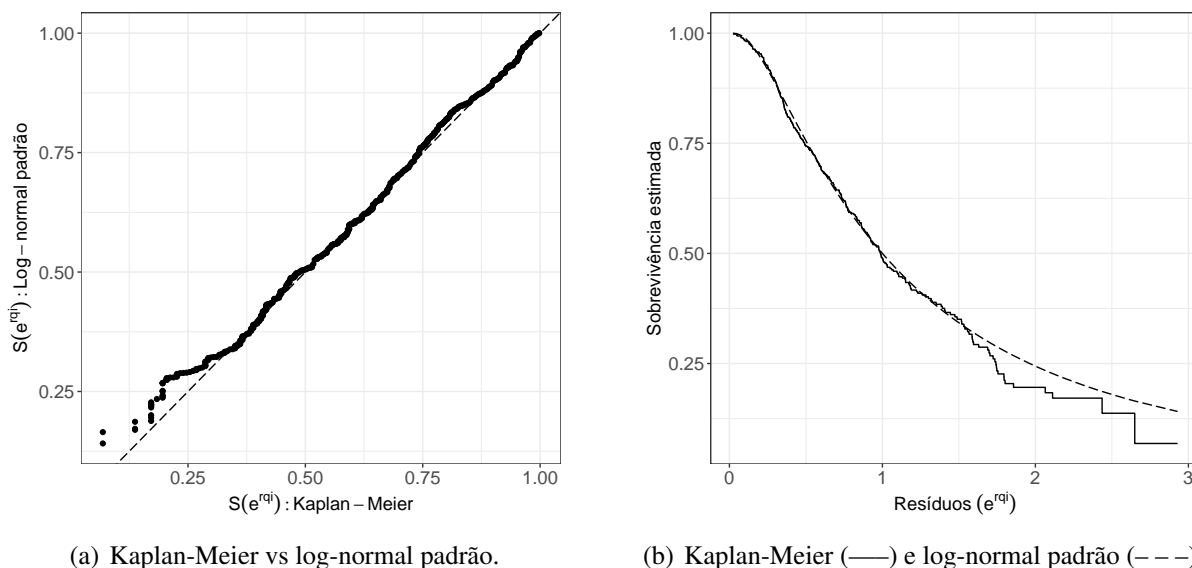


(a) Kaplan-Meier vs log-normal padrão.



(b) Kaplan-Meier (—) e log-normal padrão (---).

**Figura 4.22:** Sobrevivências dos resíduos  $e^{r_{q,i}}$  do modelo BS- $t$  discreto estimadas pelo método de Kaplan-Meier e pelo modelo log-normal padrão (a) e suas respectivas curvas estimadas (b).

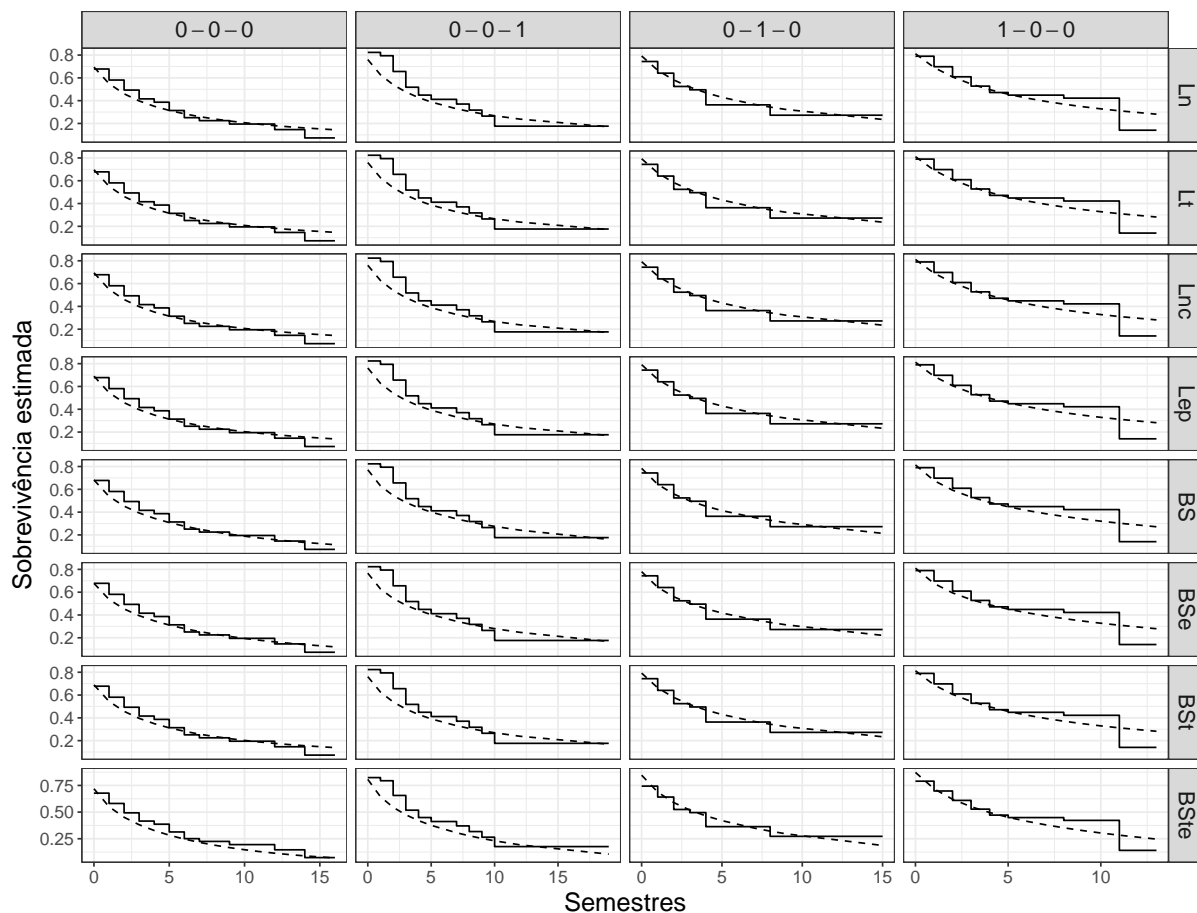


**Figura 4.23:** Sobrevivências dos resíduos  $e^{r_{q,i}}$  do modelo BS- $t$  estendido discreto estimadas pelo método de Kaplan-Meier e pelo modelo log-normal padrão (a) e suas respectivas curvas estimadas (b).

Outra forma de analisar o ajuste dos modelos é plotando as curvas de Kaplan-Meier e as curvas de sobrevivência estimadas pelos modelos log-simétricos discretos para os indivíduos de cada combinação possível entre as categorias das variáveis regressoras. Assim, como os modelos têm 3 covariáveis e cada uma tem 2 categorias de resposta (0 ou 1), tem-se 8 combinações possíveis. No entanto, considerou-se para o ajustamento apenas as combinações com frequência superior a 30 observações na Tabela 4.17. Pela Figura 4.24, nota-se que os modelos estão se ajustando bem aos dados.

**Tabela 4.17:** Frequências das combinações das categorias das covariáveis.

idade—origem—ingresso	frequência
0-0-0	211
1-0-0	138
0-1-0	39
0-0-1	34
1-0-1	21
1-1-0	12
0-1-1	8
1-1-1	7



**Figura 4.24:** Ajuste dos modelos por categoria (0, 1) das covaráveis idade, origem e forma de ingresso (— Kaplan-Meier - - - modelo log-simétrico discreto).



# Capítulo 5

## Considerações Finais

Neste trabalho uma nova classe de distribuições para os casos em que os dados são discretos, assimétricos e não-negativos é proposta. A abordagem proposta é uma versão discreta da família de distribuições log-simétricas contínuas e pode ser usada para evitar o uso de distribuições contínuas em dados discretos ou discretizados. Algumas propriedades matemáticas das distribuições log-simétricas discretas foram elencadas, tais como momentos e variância e quantis. Foram ilustradas as funções massa de probabilidade e de risco para algumas distribuições, notando casos unimodais e bimodais e funções de risco crescentes e decrescentes. Considerou-se as estimativas sobre os parâmetros dos modelos com base no método da máxima verossimilhança abordando modelos com e sem covariáveis. Nesses dois casos, considerou-se ainda dados não censurados e dados com censura nos percentuais de 10% e 30% da amostra. As covariáveis foram incluídas no parâmetro  $\lambda_i$  que é interpretado diretamente como a mediana no caso contínuo. Estudos de simulação de Monte Carlo foram realizados para avaliar o comportamento dos estimadores de máxima verossimilhança, indicando bons desempenhos para os estimadores. Os modelos propostos foram aplicados a conjuntos de dados do mundo real, ilustrando casos em que esses dados são não censurados e casos com dados censurados. Os resultados mostraram que a família discreta proposta provou ser um modelo útil para os dados considerados.

Como proposta para trabalhos futuros, pode-se destacar os seguintes pontos: extensão para o caso multivariado; implementação de técnicas de diagnósticos de influência local/global; aplicação em modelos para séries temporais de valores inteiros; aplicação para modelos de dados em painel e implementação de modelos com fração de cura.

# Bibliografia

- [1] AGRESTI, A. *An Introduction to Categorical Data Analysis*. Wiley, Florida, 2007.
- [2] AKAIKE, H. *A new look at the statistical model identification*. Automatic Control, IEEE Transactions on, Ieee, v. 19, n. 6, p. 716-723, 1974.
- [3] ALMALKI, S. J.; NADARAJAH, S. *Modifications of the weibull distribution: a review*. Reliability Engineering & System Safety, Elsevier, v. 124, p. 32-55, 2014.
- [4] ARNOLD, T. B.; EMERSON, J. W. *dgof: Discrete Goodness-of-Fit Tests*, 2012. URL <<http://CRAN.R-project.org/package=dgof>>. R package version 1.2.
- [5] BAKOUCH, H. S.; JAZI, M. A.; NADARAJAH, S. *A new discrete distribution*. Statistics, Taylor & Francis, v. 48, n. 1, p. 200-240, 2014.
- [6] BENDEL, R. B.; AFIFI, A. A. *Comparison of stopping rules in forward "stepwise" regression*. Journal of the American Statistical Association. 1977.
- [7] BOX, G. E. P.; TIAO G. C. *Bayesian Inference in Statistical Analysis*. Wiley, New York, 1973.
- [8] BRUNELLO, G.; NAKANO, E. *Inferência bayesiana no modelo Weibull discreto em dados com presença de censura*. TEMA (São Carlos), SciELO Brasil, v. 16, n. 2, p. 97-110, 2015.
- [9] CHAKRABORTY, S. *Generating discrete analogues of continuous probability distributions-a survey of methods and constructions*. JOURNAL of Statistical Distributions and Applications, Springer, v. 2, n. 1, p. 1-30, 2015.
- [10] CHMIELEWSKI, M. A. *Elliptically Symmetric Distributions: A Review and Bibliography*. International Statistical Review/Revue Internationale de Statistique Vol. 49, No. 1 (Apr., 1981), pp. 67-74.
- [11] CHOULAKIAN, V.; LOCKHART, R. A.; STEPHENS, M. A. *Cramér-von Mises statistics for discrete distributions*. The Canadian Journal of Statistics, 22(1):125-137, 1994.

- [12] COLOSIMO, E. A.; GIOLO, S. R. *Análise de sobrevivência aplicada*. São Paulo: Edgard Blucher Ltda, 2006.
- [13] COX, D. R.; HINKLEY D. V. *Theoretical Statistics*. Londres: Chapman & Hall, 1974.
- [14] CRAMÉR, H. *On the composition of elementary errors*. Skand. Akt., 11:141-180, 1928.
- [15] CYSNEIROS, F. J. A.; PAULA, G. A.; GALEA, M. *Modelos simétricos aplicados*. ABE, XI Escola de Modelos de Regressão, São Paulo, 2005. 23, 25, 37, 38
- [16] DEGROOT, M. H.; SCHERVISH, M. J. *Probability and Statistics*. Boston: Addison-Wesley, 2012.
- [17] DUNN, P. K.; SMYTH, G. K. *Randomized quantile residuals*. Journal of Computational and Graphical Statistics, Taylor & Francis, v. 5, n. 3, p. 236-244, 1996.
- [18] FERNANDEZ, R. N. *Restrições fiscais e infraestrutura: uma análise para os estados brasileiros*. XXII Encontro de Economia da Região Sul - ANPECSUL, Maringá, 2019.
- [19] GUPTA, P. L.; GUPTA, R. C. *On the moments of residual life in reliability and some characterization results*. Communications in Statistics - Theory and Methods, 12(4):449-461, 1983.
- [20] HINKLEY, D. V. *On power transformations to symmetry*. Biometrika, 62:101-111, 1975.
- [21] HUBERT, M.; VANDERVEEKEN, S. *Outlier detection for skewed data*. Journal of Chemometrics, 2:235-246, 2008.
- [22] JOHNSON, N. L.; KOTZ, S.; KEMP, A. *Univariate Discrete Distributions*. Wiley, New York, US, 1993.
- [23] JONES, M. C. *On reciprocal symmetry*. Journal of Statistical Planning and Inference, 138:3039-3043, 2008.
- [24] KAPLAN, E. L.; MEIER, P. *Nonparametric estimation from incomplete observations*. Journal of the American Statistical Association 53, 457-481. 1958.
- [25] KEMP, A. W. *Characterizations of a discrete normal distribution*. Journal of Statistical Planning and Inference, v. 63, n. 2, p. 223-229, 1997.
- [26] KEMP, A. W. *Classes of discrete lifetime distributions*. Taylor & Francis, 2004.
- [27] KLEIN, J. P.; MOESCHBERGER, M. L. *Survival analysis: techniques for censored and truncated data*. New York: Springer, 1997.
- [28] LAWLESS, J. F. *Statistical Models and Methods for Lifetime Data*. New York: John Wiley & Sons, 2011.



- [29] LIANG, T. C. *Monotone empirical Bayes tests for a discrete normal distribution*. Stat. Probabil. Lett. 44, 241-249, 1999.
- [30] LINDSEY, J. K. *The Statistical Analysis of Stochastic Processes in Time*. New York: Cambridge University Press, 2001.
- [31] LISMAN, J.; ZUYLEN, M. V. *Note on the generation of most probable frequency distributions*. Statistica Neerlandica, Wiley Online Library, v. 26, n. 1, p. 19-23, 1972.
- [32] MEDEIROS, F. M. C.; FERRARI, S. L. P. *Small-sample testing inference in symmetric and logsymmetric linear regression models*. Statistica Neerlandica, 71:200-224, 2017.
- [33] MEYER, P. L. *Probabilidade: aplicações a estatística*. Rio de Janeiro: LTC, 2015.
- [34] MICKEY, J.; GREENLAND, S. *A study of the impact of confounder-selection criteria on effect estimation*. American Journal of Epidemiology. 1989.
- [35] MISES von, R. E. *Wahrscheinlichkeit, Statistik und Wahrheit*. Julius Springer, Vienna, Austria, 1928.
- [36] MITTELHAMMER, R. C.; JUDGE, G. G.; MILLER, D. J. *Econometric Foundations*. Cambridge University Press, New York, US, 2000.
- [37] MOORS, J. J. A. *A quantile alternative for kurtosis*. The Statistician, 37:25-32, 1988.
- [38] MURTHY, D. P.; XIE, M.; JIANG, R. *Weibull models*. [S.l.]: John Wiley & Sons. v. 505, 2004.
- [39] NADARAJAH, S. *A generalized normal distribution*. Journal of Applied Statistics 32.7: 685-694, 2005.
- [40] NAKAGAWA, T.; OSAKI, S. *The discrete Weibull distribution*. IEEE Transactions on Reliability, v. 5, p. 300-301, 1975.
- [41] PEARSON, K. *On the criterion that a given system of deviations from the probable in the case of a correlated system of a variables in such that in can be reasonably supposed to have arisen from random sampling*. Philos. Mag., v.50, p.157-172, 1900.
- [42] R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.
- [43] ROY, D.; DASGUPTA, T. *A discretizing approach for evaluating reliability of complex systems under stress-strength model*. IEEE Transactions on Reliability, IEEE, v. 50, 2001.
- [44] ROY, D. *Discretization of continuous distributions with an application to stress-strength reliability*. Bulletin of the Calcutta Statistical Association, Calcutta: The Association,[1947]-, v. 52, n. 205-208, p. 295-314, 2002.

- [45] SATO, H.; IKOTA, M.; ARITOSHI, S.; MASUDA, H. *A new defect distribution meteorology with a consistent discrete exponential formula and its applications*. IEEE. Trans. Semicond. Manufactur. 12(4), 409-418, 1999.
- [46] SAULO, H.; LEÃO, J. *On log-symmetric duration models applied to high frequency financial data*. Economics Bulletin, 37:1089-1097, 2017.
- [47] SCHWARZ, G. *Estimating the dimension of a model*. The annals of statistics, Institute of Mathematical Statistics, v. 6, n. 2, p. 461-464, 1978.
- [48] STEIN, W. E.; DATTERO, R. *A new discrete weibull distribution*. IEEE Transactions on Reliability, Institute of Electrical and Electronics Engineers, v. 33, n. 2, p. 196-197, 1984.
- [49] SUBBOTIN, M. T. *On the Law of Frequency of Error*. Mat. Sb. 31.2: 206-301, 1923.
- [50] SZABLowski, P. J. *Discrete normal distribution and its relationship with Jacobi Theta functions*. Stat. Probabil. Lett. 52, 289-299, 2001.
- [51] TRENKLER, D. *A handbook of small data sets*: Hand, d.j., daly, f., lunn, a.d., mcconway, k.j. & ostrowski, e. (1994): Chapman & hall, london. Computational Statistics & Data Analysis, 19(1):101-101, 1995.
- [52] PUIG, P. *A note on harmonic law: A two-parameter family of distributions for ratios*. Statistics and Probability Letters, 78, 320-326, 2008.
- [53] VANEGAS, L. H., PAULA, G. A. *Log-symmetric regression models under the presence of non-informative left-or right-censored observations*. TEST, 26(2), 405-428, 2017.
- [54] VANEGAS, L. H.; PAULA, G. A. *Log-symmetric distributions: statistical properties and parameter estimation*. Brazilian Journal of Probability and Statistics, Brazilian Statistical Association, v. 30, n. 2, p. 196-220, 2016.
- [55] VANEGAS, L. H.; PAULA, G. A. *An extension of log-symmetric regression models: R codes and applications*. Journal of Statistical Simulation and Computation, 86:1709-1735, 2016a.
- [56] VANEGAS, L. H.; PAULA, G. A. *ssym: Fitting Semi-Parametric log-Symmetric Regression Models*. R package version 1.5.7, 2016b.
- [57] VENTURA, M.; SAULO, H.; LEIVA, V.; MONSUETO, S. E. *Log-symmetric regression models: information criteria and application to movie business and industry data*. Applied Stochastic Models in Business and Industry, <https://doi.org/10.1002/asmb.2433>, 2019.
- [58] VILA, R.; NAKANO, E. Y.; SAULO, H. *Theoretical results on the discrete weibull distribution of nakagawa and osaki*. Statistics, 53(2):339-363, 2019.
- [59] ZWILLINGER, D.; KOKOSKA, S. *Standard Probability and Statistical Tables and Formula*. Chapman & Hall, Boca Raton, 2000.

# Apêndice A

## Dados utilizados na aplicação

Este apêndice lista os conjuntos de dados utilizados para ilustrar os modelos de regressão log-simétricos discretos.

**Tabela A.1:** Dados não censurados referentes ao número de projetos em contratos de concessões dos estados brasileiros. Onde *lpop*: logaritmo do tamanho da população (IBGE), *respripi*: resultado primário - % do PIB, *servdivtotpi*: serviço da dívida pública - % do PIB e *desprec*: despesas com pessoal/receita corrente (Secretaria do Tesouro Nacional).

Observação	Contratos	lpop	respripi	servdivtotpi	desprec
1	0	13.23073	0.0060	0.0056	58.47
2	0	12.90917	0.0059	0.0069	56.71
3	0	12.94324	0.0041	0.0073	55.14
4	0	12.97849	0.0052	0.0070	57.81
5	0	13.38626	0.0038	0.0043	57.87
6	0	13.40905	0.0053	0.0044	53.33
7	0	13.42985	-0.0052	0.0046	55.29
8	0	13.45023	0.0040	0.0043	54.90
9	0	13.47161	0.0064	0.0045	51.59
10	0	13.49255	-0.0090	0.0046	56.70
11	0	13.50566	-0.0123	0.0053	56.16
12	0	13.53049	0.0016	0.0061	51.90
13	0	13.54894	-0.0077	0.0078	55.94
14	0	13.56705	-0.0096	0.0072	53.94

Continua na próxima página

**Tabela A.1 – continuação da página anterior**

Observação	Contratos	lpop	respripib	servdivtotpib	desprec
15	0	13.58358	-0.0094	0.0070	53.00
16	0	13.59984	-0.0017	0.0095	57.97
17	2	13.61301	0.0079	0.0099	54.56
18	0	14.85195	0.0093	0.0232	62.63
19	0	14.89630	0.0114	0.0253	57.87
20	1	14.90944	0.0041	0.0073	55.14
21	0	14.92176	0.0112	0.0070	57.81
22	0	14.93360	0.0123	0.0146	57.66
23	0	14.94433	0.0160	0.0177	53.41
24	0	14.95494	0.0156	0.0168	55.09
25	0	14.96450	0.0235	0.0157	55.90
26	0	14.97365	0.0138	0.0166	57.79
27	0	14.98209	0.0042	0.0203	56.99
28	0	14.95350	0.0132	0.0152	50.29
29	0	14.99754	0.0091	0.0156	52.14
30	1	15.00457	0.0073	0.0163	53.20
31	1	15.01125	0.0063	0.0147	53.21
32	0	15.01728	-0.0059	0.0145	55.43
33	0	15.02298	0.0122	0.0175	57.12
34	0	15.02714	0.0210	0.0092	51.30
35	0	13.07284	0.0047	0.0013	50.30
36	0	13.05408	0.0033	0.0021	39.06
37	0	13.09190	0.0100	0.0070	36.12
38	0	13.12636	-0.0033	0.0022	53.61
39	0	13.27594	-0.0009	0.0018	52.74
40	0	13.30635	0.0049	0.0012	47.54
41	0	13.33747	-0.0005	0.0011	50.30
42	0	13.36609	0.0004	0.0012	52.19
43	0	13.39239	-0.0018	0.0011	49.45
44	0	13.41950	-0.0056	0.0013	57.59
45	1	13.41433	-0.0033	0.0016	57.03
46	0	13.46737	0.0005	0.0031	29.79
47	2	13.49116	-0.0127	0.0013	61.93
48	0	13.51305	-0.0010	0.0073	53.82
49	1	13.53447	0.0003	0.0032	53.18
50	0	13.55544	-0.0051	0.0047	58.91

Continua na próxima página

**Tabela A.1 – continuação da página anterior**

Observação	Contratos	lpop	respripib	servdivtotpib	desprec
51	0	13.56999	0.0170	0.0048	50.92
52	0	14.84979	0.0182	0.0139	44.63
53	0	14.62243	-0.0148	0.0159	45.74
54	4	14.65362	0.0111	0.0148	47.79
55	0	14.68345	0.0101	0.0138	50.09
56	0	14.97522	0.0096	0.0128	47.71
57	2	14.99816	0.0079	0.0113	47.16
58	3	15.02058	0.0060	0.0102	48.45
59	0	15.04222	0.0126	0.0089	44.95
60	0	15.06283	0.0053	0.0084	45.05
61	2	15.08274	-0.0292	0.0086	51.82
62	0	15.06369	-0.0085	0.0084	46.93
63	1	15.12087	0.0022	0.0100	47.40
64	0	15.13913	0.0025	0.0111	46.79
65	0	15.15655	-0.0181	0.0120	49.28
66	0	15.17366	-0.0227	0.0125	50.90
67	0	15.18973	-0.0050	0.0152	53.69
68	0	15.20222	0.0027	0.0158	50.82
69	1	16.38559	0.0311	0.0623	47.11
70	0	16.43283	0.0463	0.0769	45.62
71	4	16.44315	0.0479	0.0700	46.27
72	0	16.45300	0.0287	0.0811	50.75
73	1	16.46232	0.0457	0.0713	50.57
74	0	16.47121	0.0461	0.0685	49.07
75	2	16.47982	0.0508	0.0709	51.42
76	0	16.48787	0.0677	0.0692	51.98
77	1	16.49544	0.0426	0.0691	52.94
78	1	16.50261	0.0193	0.0579	58.57
79	1	16.45577	0.0169	0.0380	53.16
80	3	16.51587	0.0129	0.0339	53.99
81	3	16.52203	0.0141	0.0345	55.24
82	4	16.52775	0.0078	0.0608	55.17
83	5	16.53312	0.0218	0.0290	55.64
84	5	16.53812	-0.0224	0.0341	58.48
85	8	16.54183	-0.0287	0.0302	53.29
86	0	15.81948	-0.0056	0.0361	50.42

Continua na próxima página

**Tabela A.1 – continuação da página anterior**

Observação	Contratos	lpop	respripib	servdivtotpib	desprec
87	0	15.86295	-0.0046	0.0389	50.33
88	3	15.87704	0.0015	0.0449	47.95
89	1	15.89044	0.0057	0.0411	51.30
90	1	15.90292	0.0193	0.0347	49.65
91	0	15.91488	0.0278	0.0322	47.69
92	0	15.92633	0.0092	0.0494	46.37
93	1	15.93705	0.0349	0.0241	47.23
94	2	15.94707	0.0314	0.0215	44.21
95	4	15.95663	-0.0093	0.0244	48.64
96	1	15.94996	-0.0312	0.0131	49.40
97	4	15.97423	-0.0054	0.0136	46.71
98	3	15.98239	-0.0083	0.0198	49.16
99	5	15.99004	-0.0030	0.0165	49.52
100	3	15.99728	-0.0336	0.0190	49.68
101	9	16.00426	-0.0201	0.0234	55.42
102	4	16.00869	0.0130	0.0280	50.34
103	0	14.53001	0.0055	0.0162	66.62
104	2	14.58377	0.0155	0.0148	65.63
105	1	14.60306	0.0089	0.0138	65.82
106	1	14.62332	0.0052	0.0110	42.03
107	0	14.64406	0.0067	0.0097	39.30
108	0	14.66566	0.0044	0.0083	43.75
109	0	14.68806	-0.0025	0.0080	49.67
110	0	14.71037	0.0209	0.0076	47.72
111	0	14.73260	0.0048	0.0076	51.90
112	0	14.75513	-0.0181	0.0080	48.82
113	0	14.75948	0.0035	0.0071	48.99
114	0	14.80058	0.0012	0.0066	53.06
115	0	14.82310	-0.0092	0.0066	52.28
116	0	14.84549	-0.0201	0.0069	51.51
117	0	14.86738	-0.0483	0.0079	56.98
118	0	14.88881	0.0042	0.0123	52.13
119	1	14.90650	-0.0103	0.0089	56.78
120	0	14.94510	-0.0025	0.0174	56.25
121	0	14.99632	0.0158	0.0315	51.02
122	1	15.01185	0.0142	0.0199	55.60

Continua na próxima página

**Tabela A.1 – continuação da página anterior**

Observação	Contratos	lpop	respripib	servdivtotpib	desprec
123	0	15.02745	0.0262	0.0258	47.92
124	0	15.04252	0.0280	0.0177	43.08
125	0	15.05707	0.0237	0.0165	40.41
126	1	15.07170	0.0108	0.0157	40.35
127	0	15.08583	0.0126	0.0132	41.48
128	3	15.09949	0.0260	0.0125	37.32
129	1	15.11297	-0.0137	0.0121	43.90
130	1	15.07254	-0.0205	0.0094	46.53
131	2	15.13860	0.0007	0.0091	44.31
132	1	15.15130	0.0043	0.0141	43.51
133	0	15.16332	-0.0209	0.0108	49.75
134	0	15.17495	-0.0309	0.0112	51.21
135	0	15.18644	0.0042	0.0123	52.13
136	1	15.19521	0.0005	0.0128	52.74
137	1	15.42424	0.0186	0.0271	65.26
138	3	15.46839	0.0108	0.0633	65.68
139	7	15.49036	0.0174	0.0322	67.44
140	6	15.51113	0.0259	0.0317	59.14
141	0	15.53129	0.0017	0.0324	61.05
142	0	15.55052	0.0302	0.0335	58.09
143	2	15.56922	0.0317	0.0368	65.09
144	2	15.58707	0.0307	0.0339	65.58
145	5	15.60410	0.0377	0.0345	57.60
146	8	15.62051	0.0341	0.0355	62.64
147	4	15.60790	-0.0205	0.0094	46.53
148	0	15.65145	0.0457	0.0718	57.38
149	5	15.66632	0.0270	0.0713	59.95
150	0	15.68052	0.0112	0.0494	59.78
151	3	15.69406	-0.0059	0.0576	60.58
152	4	15.70727	0.0109	0.0568	63.84
153	3	15.71700	0.0289	0.0418	60.90
154	16	15.54592	0.0127	0.0330	50.99
155	0	15.59012	0.0049	0.0216	45.11
156	1	15.60560	0.0123	0.0241	46.30
157	1	15.62035	-0.0050	0.0237	55.36
158	1	15.63424	0.0203	0.0229	48.80

Continua na próxima página

**Tabela A.1 – continuação da página anterior**

Observação	Contratos	lpop	respripib	servdivtotpib	desprec
159	0	15.64761	0.0292	0.0230	43.72
160	0	15.66049	0.0245	0.0237	43.19
161	0	15.67259	0.0260	0.0225	42.38
162	0	15.68392	0.0226	0.0247	39.84
163	2	15.69482	0.0058	0.0257	45.52
164	0	15.69875	0.0025	0.0236	46.42
165	1	15.71478	0.0175	0.0242	42.48
166	2	15.72402	0.0127	0.0257	43.58
167	1	15.73273	-0.0069	0.0805	45.52
168	1	15.74093	-0.0137	0.0106	46.28
169	0	15.74848	-0.0032	0.0162	50.31
170	0	15.75483	0.0140	0.0200	445.62
171	3	14.73271	0.0226	0.0230	53.46
172	1	14.78021	0.0256	0.0271	49.64
173	5	14.79946	0.0195	0.0268	49.93
174	5	14.81798	0.0323	0.0288	48.25
175	1	14.83616	0.0281	0.0294	45.53
176	4	14.85367	0.0286	0.0324	48.46
177	3	14.87122	0.0246	0.0267	50.40
178	1	14.88744	0.0237	0.0247	49.47
179	1	14.90340	0.0241	0.0234	45.96
180	3	14.91878	0.0098	0.0258	50.34
181	2	14.92576	0.0151	0.0232	53.36
182	4	14.94820	0.0135	0.0230	54.56
183	4	14.96260	0.0137	0.0612	55.62
184	3	14.97585	-0.0133	0.0134	55.38
185	1	14.98892	0.0042	0.0156	54.12
186	2	15.00152	0.0097	0.0227	59.72
187	6	15.01111	0.0104	0.0209	60.66
188	0	14.54541	0.0010	0.0611	63.88
189	1	14.59116	0.0140	0.0127	52.58
190	5	14.60805	0.0121	0.0203	52.40
191	9	14.62466	0.0095	0.0159	54.38
192	0	14.64056	0.0141	0.0159	53.00
193	3	14.65622	0.0110	0.0169	55.34
194	1	14.67205	0.0169	0.0172	57.83

Continua na próxima página



**Tabela A.1 – continuação da página anterior**

Observação	Contratos	lpop	respripib	servdivtotpib	desprec
195	0	14.68680	0.0304	0.0195	57.00
196	0	14.70175	0.0291	0.0164	48.63
197	2	14.71567	0.0144	0.0197	57.95
198	1	14.71120	0.0056	0.0158	61.95
199	3	14.74333	0.0055	0.0159	58.31
200	0	14.75669	0.0112	0.0168	59.60
201	4	14.76910	0.0047	0.0189	60.69
202	1	14.78173	0.0021	0.0224	58.39
203	0	14.79346	0.0265	0.0258	73.49
204	0	14.80222	0.0016	0.0117	65.49
205	1	16.69843	0.0042	0.1406	75.03
206	0	16.73096	-0.0092	0.1223	72.93
207	3	17.17147	0.0450	0.1022	71.33
208	8	16.75377	0.0681	0.0953	70.68
209	2	16.76447	0.0651	0.0904	64.11
210	1	16.77468	0.0791	0.0933	58.78
211	4	16.78469	0.0701	0.0993	58.59
212	5	16.79405	0.0720	0.0954	59.64
213	5	16.80301	0.0571	0.1004	56.52
214	7	16.81154	0.0410	0.1019	57.99
215	4	16.79090	0.0396	0.0896	60.34
216	2	16.82751	0.0458	0.0905	63.01
217	6	16.83506	0.0471	0.1697	65.74
218	2	16.84216	-0.0145	0.2154	65.14
219	5	16.84888	-0.0364	0.1209	68.59
220	3	16.85526	-0.0453	0.1434	78.00
221	5	16.85992	-0.0575	0.0600	77.00
222	0	15.63837	0.0107	0.0186	54.17
223	0	15.29027	0.0105	0.0159	53.48
224	2	15.31290	0.0099	0.0135	51.78
225	1	15.33502	0.0079	0.0136	54.26
226	1	15.75296	0.0101	0.0139	52.31
227	0	15.77095	0.0142	0.0131	48.17
228	1	15.78876	0.0000	0.0122	51.31
229	1	15.80530	0.0153	0.0116	52.28
230	1	15.82157	0.0057	0.0123	52.19

Continua na próxima página

**Tabela A.1 – continuação da página anterior**

Observação	Contratos	lpop	respripib	servdivtotpib	desprec
231	0	15.83693	-0.0013	0.0128	53.55
232	1	15.84116	-0.0115	0.0121	53.37
233	1	15.86642	0.0169	0.0124	53.29
234	4	15.88047	0.0167	0.0123	55.03
235	1	15.89370	0.0009	0.0127	56.69
236	1	15.90651	-0.0013	0.0113	56.03
237	0	15.91879	0.0011	0.0134	57.00
238	1	15.92847	0.0024	0.0119	57.41
239	16	15.05079	0.0099	0.0275	53.30
240	0	15.07255	-0.0054	0.0219	51.06
241	0	15.08274	-0.0025	0.0186	54.50
242	6	15.09283	0.0058	0.0178	67.70
243	0	15.10281	0.0038	0.0164	65.47
244	3	15.11243	0.0122	0.0164	58.13
245	0	15.12222	0.0091	0.0175	58.71
246	0	15.13138	0.0099	0.0137	59.06
247	0	15.14020	0.0140	0.0112	56.55
248	1	15.38288	0.0063	0.0110	63.49
249	1	15.14166	-0.0036	0.0089	66.30
250	0	15.16566	0.0109	0.0081	57.58
251	0	15.17366	-0.0036	0.0078	64.13
252	2	15.18135	-0.0024	0.0079	62.19
253	0	15.18872	-0.0099	0.0070	63.29
254	2	15.19579	-0.0057	0.0098	64.44
255	0	15.20166	0.0053	0.0111	61.18
256	0	16.07294	0.0582	0.2029	58.51
257	2	16.09963	0.0231	0.0650	61.52
258	6	16.11107	0.0279	0.7570	57.34
259	5	16.12219	0.0379	0.0709	62.63
260	0	16.13288	0.0407	0.0614	62.90
261	3	16.14328	0.0423	0.0589	60.31
262	1	16.15356	0.0234	0.0522	64.19
263	0	16.16335	0.0323	0.0489	63.98
264	0	16.17277	0.0373	0.0454	60.38
265	3	16.18182	0.0301	0.0458	59.97
266	2	16.16159	0.0302	0.0383	60.81

Continua na próxima página

**Tabela A.1 – continuação da página anterior**

Observação	Contratos	lpop	respripib	servdivtotpib	desprec
267	0	16.19912	0.0346	0.0320	62.25
268	2	16.20748	0.0099	0.0300	63.24
269	1	16.21540	0.0035	0.0304	66.22
270	1	16.22291	-0.0002	0.0308	65.36
271	1	16.23026	0.0268	0.0343	63.45
272	1	16.23523	0.0180	0.0185	60.65
273	16	15.88388	-0.0273	0.0409	62.59
274	1	15.92500	-0.0273	0.0363	60.43
275	6	15.93513	-0.0079	0.0389	57.34
276	2	15.94600	0.0199	0.0354	57.14
277	0	15.95652	0.0218	0.0317	56.41
278	0	15.96669	0.0308	0.0319	53.97
279	2	15.97676	0.0253	0.0307	54.17
280	0	15.98628	0.0258	0.0256	54.53
281	1	15.99548	0.0178	0.0245	53.43
282	1	16.00437	-0.0035	0.0246	55.65
283	1	15.98986	0.0017	0.0163	52.72
284	0	16.02136	-0.0084	0.0164	51.03
285	1	16.02948	-0.0219	0.0192	55.18
286	4	16.03721	-0.0285	0.0233	50.96
287	1	16.04466	-0.0431	0.0244	53.04
288	3	16.05185	0.0042	0.0304	52.92
289	0	16.05732	0.0186	0.0281	50.95
290	0	14.85974	0.0159	0.0394	55.80
291	0	14.88744	0.0143	0.0181	54.66
292	1	14.89901	0.0143	0.0194	54.61
293	0	14.90944	0.0077	0.0151	62.58
294	1	14.91911	0.0083	0.0141	58.87
295	0	14.92770	0.0128	0.0144	50.56
296	0	13.88036	0.0113	0.0142	50.07
297	0	14.94336	0.1120	0.0142	50.08
298	1	14.95013	0.0100	0.0152	48.87
299	1	14.95590	-0.0053	0.0159	52.54
300	1	14.95282	-0.0003	0.0145	52.33
301	0	14.96640	0.0077	0.0167	51.90
302	0	14.97082	0.0077	0.0228	50.76

Continua na próxima página

**Tabela A.1 – continuação da página anterior**

Observação	Contratos	lpop	respripib	servdivtotpib	desprec
303	1	14.97459	-0.0109	0.0069	57.79
304	5	14.97772	-0.0032	0.0074	57.95
305	0	14.98085	-0.0006	0.0083	58.03
306	5	14.98246	-0.0140	0.0083	54.00
307	0	16.48051	0.0933	0.0960	50.11
308	0	16.52316	0.0699	0.1050	51.88
309	0	16.53232	-0.0277	0.1395	62.91
310	1	16.54120	0.0962	0.1416	56.49
311	0	16.54975	0.1140	0.1053	47.95
312	4	16.55790	0.1035	0.1167	49.44
313	2	16.56598	0.0781	0.1122	50.59
314	0	16.57349	0.1151	0.1115	49.85
315	1	16.58076	0.1414	0.1042	48.23
316	1	16.58766	0.0873	0.1105	49.16
317	1	16.58747	0.0539	0.0946	49.63
318	1	16.60071	0.0769	0.0959	48.23
319	0	16.60692	0.0158	0.1050	53.18
320	1	16.61273	-0.0233	0.1226	48.28
321	1	16.61827	-0.0548	0.1418	52.71
322	0	16.62359	-0.0762	0.1546	62.84
323	1	16.62708	-0.1522	0.1000	74.73
324	0	14.83491	0.0060	0.0093	51.74
325	0	14.87884	0.0046	0.0091	54.47
326	1	14.89460	0.0039	0.0106	59.94
327	4	14.90944	0.0105	0.0100	60.97
328	1	14.92407	0.0072	0.0091	57.89
329	0	14.93816	0.0062	0.0084	54.82
330	1	14.95174	0.0013	0.0081	56.28
331	0	14.96482	0.0015	0.0077	60.06
332	0	14.97741	0.0067	0.0080	55.16
333	3	14.98954	-0.0074	0.0086	58.91
334	1	14.96862	-0.0077	0.0075	58.78
335	9	15.01246	0.0032	0.0075	59.36
336	4	15.02357	0.0006	0.0069	61.10
337	7	15.03398	0.0071	0.0146	61.88
338	1	15.04427	0.0017	0.0046	61.00

Continua na próxima página

**Tabela A.1 – continuação da página anterior**

Observação	Contratos	lpop	respripib	servdivtotpib	desprec
339	6	15.05388	0.0075	0.0053	57.48
340	4	15.06110	0.0139	0.0053	60.01
341	0	16.13611	-0.0341	0.1012	72.45
342	2	16.15655	-0.0117	0.0744	70.05
343	0	16.16488	0.0209	0.0824	69.09
344	6	16.17277	0.0122	0.0837	68.96
345	1	16.18013	0.0090	0.0749	67.72
346	9	16.18715	0.0321	0.0797	63.99
347	0	16.19385	0.0328	0.0744	65.47
348	4	16.20004	0.0404	0.0687	67.62
349	2	16.20583	0.0681	0.1058	60.37
350	1	16.21131	0.0483	0.0662	64.30
351	2	16.18519	0.0274	0.0778	59.57
352	1	16.22110	0.0376	0.0563	62.38
353	1	16.22569	0.0213	0.0560	65.41
354	4	16.22982	0.0062	0.0605	67.89
355	3	16.23365	-0.0218	0.0686	70.31
356	3	16.23721	-0.0098	0.0786	70.64
357	1	16.23912	-0.0141	0.0361	69.73
358	2	14.13599	0.0065	0.0120	58.88
359	0	13.75257	0.0082	0.0113	53.14
360	0	13.78814	0.0025	0.0102	55.80
361	0	13.82050	0.0090	0.0107	51.53
362	0	14.23488	0.0078	0.0096	50.22
363	0	14.25247	0.0096	0.0098	48.56
364	1	14.26913	0.0054	0.0087	52.81
365	0	14.28426	0.0076	0.0079	48.93
366	0	14.29917	0.0077	0.0073	42.60
367	1	14.31386	-0.0107	0.0071	48.70
368	1	14.26174	-0.0010	0.0065	49.95
369	2	14.34083	0.0035	0.0073	49.29
370	1	14.35317	0.0003	0.0057	54.96
371	1	14.36537	-0.0079	0.0056	56.92
372	0	14.37684	-0.0006	0.0040	55.14
373	1	14.38818	-0.0050	0.0022	56.08
374	1	14.39620	0.0054	0.0022	53.24

Continua na próxima página

**Tabela A.1 – continuação da página anterior**

Observação	Contratos	lpop	respripib	servdivtotpib	desprec
375	0	12.68897	0.0025	0.0016	41.48
376	0	12.50987	-0.0035	0.0017	42.72
377	0	12.54966	0.0040	0.0014	41.66
378	0	12.58451	-0.0031	0.0021	50.74
379	0	12.88664	0.0044	0.0022	41.54
380	0	12.91902	0.0024	0.0025	40.58
381	0	12.94801	0.0018	0.0022	38.39
382	0	12.97386	0.0014	0.0020	41.34
383	0	12.99907	0.0018	0.0021	40.89
384	0	13.02144	-0.0059	0.0020	44.84
385	0	13.01807	0.0000	0.0020	47.49
386	0	13.06473	0.0020	0.0026	43.80
387	0	13.08570	-0.0055	0.0081	53.81
388	1	13.10420	-0.0068	0.0100	52.95
389	0	13.12236	0.0013	0.0032	53.67
390	0	13.14020	-0.0019	0.0036	59.76
391	0	13.15042	0.0045	0.0049	55.71
392	0	15.49253	0.0300	0.1502	64.44
393	3	15.53111	0.0358	0.0304	57.88
394	0	15.54646	0.0211	0.1023	58.81
395	7	15.56192	0.0228	0.0373	56.87
396	1	15.57733	0.0284	0.0365	57.37
397	8	15.59266	0.0275	0.0367	53.93
398	0	15.60810	0.0311	0.0379	57.05
399	4	15.62331	0.0415	0.0365	54.50
400	1	15.63828	0.0376	0.0459	52.69
401	4	15.65320	0.0342	0.0396	52.09
402	2	15.64784	0.0274	0.0333	54.73
403	2	15.68222	0.0326	0.0351	54.22
404	5	15.69681	0.0239	0.1007	57.92
405	1	15.71088	0.0062	0.0489	54.16
406	2	15.72461	-0.0156	0.0302	56.29
407	0	15.73815	-0.0049	0.0392	57.47
408	0	15.74856	-0.0073	0.0283	55.57
409	16	17.42560	0.1804	0.3010	59.71
410	22	17.46508	0.2268	0.3180	58.62

Continua na próxima página

**Tabela A.1 – continuação da página anterior**

Observação	Contratos	lpop	respripib	servdivtotpib	desprec
411	4	17.47797	0.2251	0.3063	83.53
412	4	17.49044	0.2519	0.3181	82.89
413	1	17.50249	0.2211	0.2837	55.66
414	7	17.51414	0.1946	0.2973	53.23
415	11	17.52564	0.2497	0.3663	53.82
416	0	17.53654	0.2368	0.2749	50.71
417	3	17.54702	0.1987	0.2921	49.90
418	9	17.55717	0.1107	0.2903	51.50
419	5	17.53546	0.1659	0.2647	48.05
420	11	17.57641	0.1741	0.2625	48.13
421	2	17.58567	0.2415	0.2525	48.69
422	2	17.59444	0.1304	0.2968	49.38
423	1	17.60288	0.1365	0.3289	50.41
424	5	17.61100	0.1504	0.3644	53.29
425	5	17.61660	0.0935	0.2261	54.27
426	0	14.39309	0.0082	0.0080	64.44
427	0	14.43877	0.0100	0.0097	52.04
428	1	14.45578	0.0110	0.0106	51.97
429	1	14.47199	0.0049	0.0091	55.69
430	0	14.48743	0.0016	0.0085	55.62
431	0	14.50264	0.0125	0.0091	52.04
432	0	14.51712	0.0000	0.0084	57.81
433	0	14.53091	0.0160	0.0075	59.19
434	0	14.54451	0.0134	0.0069	56.23
435	0	14.55745	-0.0040	0.0072	60.78
436	0	14.54210	-0.0083	0.0064	60.08
437	0	14.58191	-0.0034	0.0065	59.78
438	1	14.59346	-0.0033	0.0015	60.32
439	0	14.60488	0.0004	0.0172	58.46
440	0	14.61527	0.0005	0.0078	58.15
441	0	14.62600	0.0017	0.0088	55.19
442	0	14.63343	-0.0003	0.0081	56.77
443	1	13.96040	0.0113	0.0053	38.75
444	0	14.01272	0.0125	0.0057	39.98
445	5	14.03304	-0.0042	0.0057	43.28
446	1	14.05216	-0.0037	0.0063	40.71

Continua na próxima página

**Tabela A.1 – continuação da página anterior**

Observação	Contratos	lpop	respripib	servdivtotpib	desprec
447	0	14.07093	0.0026	0.0078	42.83
448	0	14.08935	0.0105	0.0059	37.04
449	3	14.10669	0.0057	0.0050	46.50
450	1	14.12300	0.0042	0.0047	46.23
451	1	14.13904	-0.0057	0.0052	43.09
452	0	14.15412	-0.0074	0.0051	53.47
453	0	14.14009	-0.0090	0.0038	56.82
454	0	14.18293	-0.0027	0.0043	57.29
455	0	14.19668	-0.0026	0.0049	61.08
456	1	14.20958	-0.0060	0.0070	59.98
457	0	14.22164	-0.0137	0.0073	58.54
458	0	14.23356	0.0018	0.0103	57.07
459	0	14.24267	0.0023	0.0114	52.63

**Tabela A.2:** Dados censurados referentes ao número de semestres até a evasão no curso de Estatística da Universidade Federal do Piauí, do primeiro semestre de 2010 ao primeiro semestre de 2019. Censura (0: censura, 1: falha), Idade (0: > 20 anos, 1: ≤ 20 anos), Sexo (0: feminino, 1: masculino), Origem (0: Teresina, 1: outro município), Ingresso (0: enem, 1: vestibular).

Indivíduo	Tempo (semestres)	Censura	Idade	Sexo	Origem	Ingresso
1	0	0	1	1	0	0
2	0	1	0	1	0	0
3	0	1	0	1	0	0
4	13	0	1	0	0	0
5	2	1	1	0	0	0
6	0	1	0	0	0	0
7	1	0	0	0	0	0
8	4	1	1	1	0	0
9	0	1	0	1	0	0
10	5	1	0	1	0	0
11	1	0	0	1	0	0
12	0	1	0	1	0	1
13	10	1	0	1	0	1

Continua na próxima página



**Tabela A.2 – continuação da página anterior**

Indivíduo	Tempo (semestres)	Censura	Idade	Sexo	Origem	Ingresso
14	1	0	0	1	0	1
15	0	1	0	0	0	0
16	15	0	0	1	1	0
17	3	0	0	1	0	0
18	5	1	0	1	0	0
19	6	0	1	1	1	0
20	1	1	0	1	0	0
21	3	0	1	1	0	0
22	0	0	1	0	0	0
23	2	1	1	0	1	1
24	3	0	1	1	0	0
25	0	1	0	0	0	0
26	4	1	0	1	1	0
27	2	1	0	1	0	0
28	1	0	1	1	0	0
29	6	1	0	1	0	0
30	0	1	0	1	0	0
31	1	0	0	1	0	0
32	2	1	0	1	0	0
33	0	1	0	0	1	0
34	7	1	0	1	0	0
35	1	0	0	0	1	1
36	4	1	1	1	0	0
37	8	0	1	1	1	0
38	4	1	1	0	0	1
39	3	1	0	1	0	1
40	7	0	0	1	0	1
41	0	1	0	1	0	0
42	2	1	0	1	0	1
43	5	0	1	1	0	0
44	9	0	0	1	0	0
45	0	1	1	1	0	0
46	5	0	0	1	1	0
47	0	1	0	1	0	0
48	3	1	1	0	0	0
49	0	1	1	1	0	0

Continua na próxima página

**Tabela A.2 – continuação da página anterior**

Indivíduo	Tempo (semestres)	Censura	Idade	Sexo	Origem	Ingresso
50	3	1	1	0	0	1
51	0	1	1	1	1	0
52	7	0	0	0	0	0
53	1	0	1	0	0	0
54	19	0	0	1	0	1
55	3	1	1	0	0	0
56	0	1	1	1	0	0
57	0	1	1	0	0	0
58	1	0	1	0	0	0
59	1	1	0	0	0	0
60	1	1	0	1	0	0
61	1	1	1	1	0	0
62	0	1	0	1	1	0
63	0	1	0	1	0	0
64	8	1	0	1	1	0
65	1	0	0	1	0	0
66	3	1	1	1	0	1
67	7	1	1	0	0	1
68	9	1	0	1	0	0
69	1	0	1	1	0	0
70	6	0	0	1	0	0
71	0	1	0	1	1	0
72	2	1	1	1	0	0
73	0	1	1	1	0	0
74	5	0	0	1	0	1
75	2	1	1	0	0	1
76	3	0	1	1	0	0
77	0	1	0	1	0	0
78	3	0	0	1	0	0
79	9	1	0	1	0	1
80	1	1	0	0	0	1
81	12	0	0	1	1	1
82	3	1	1	1	0	0
83	3	1	0	1	0	0
84	5	1	0	0	0	0
85	0	1	0	1	0	0

Continua na próxima página

**Tabela A.2 – continuação da página anterior**

Indivíduo	Tempo (semestres)	Censura	Idade	Sexo	Origem	Ingresso
86	1	0	0	0	0	0
87	9	0	1	1	0	0
88	3	0	1	1	0	0
89	2	1	1	1	0	0
90	3	1	0	1	0	0
91	15	0	1	1	1	1
92	3	0	0	1	0	0
93	7	0	1	0	0	0
94	9	0	0	1	0	1
95	0	1	0	1	0	1
96	1	0	1	0	0	0
97	3	0	0	1	0	0
98	6	0	1	1	0	0
99	3	1	1	0	0	0
100	4	1	0	1	0	1
101	1	0	0	0	1	0
102	3	0	1	1	0	0
103	0	1	1	1	0	0
104	0	1	0	1	0	0
105	3	1	0	0	0	0
106	1	1	1	0	0	1
107	2	1	0	1	0	0
108	2	1	0	0	0	1
109	4	1	0	1	0	0
110	7	0	0	1	1	0
111	3	0	0	1	0	0
112	10	0	1	1	0	1
113	1	0	0	1	0	0
114	0	1	1	1	0	0
115	2	0	0	1	0	0
116	7	0	0	1	0	0
117	7	0	1	0	0	0
118	1	0	0	1	0	1
119	0	1	1	0	0	0
120	0	1	0	1	0	0
121	1	1	0	0	1	1

Continua na próxima página

**Tabela A.2 – continuação da página anterior**

Indivíduo	Tempo (semestres)	Censura	Idade	Sexo	Origem	Ingresso
122	1	0	0	1	0	0
123	0	1	0	1	0	0
124	6	1	0	1	0	0
125	2	1	0	1	1	1
126	5	0	1	1	0	0
127	7	0	0	1	0	0
128	2	1	0	1	1	0
129	0	1	0	1	0	0
130	7	1	0	1	0	0
131	2	1	1	1	0	0
132	9	0	0	0	0	0
133	3	0	0	1	0	0
134	7	0	1	1	0	1
135	0	1	0	1	0	0
136	8	0	1	1	0	0
137	1	1	1	1	0	0
138	6	1	0	0	0	0
139	0	0	0	1	0	0
140	0	1	0	0	0	0
141	0	1	0	1	0	0
142	7	0	0	1	1	0
143	9	0	0	1	0	0
144	3	0	1	0	0	0
145	0	1	0	1	0	0
146	0	1	1	1	0	0
147	7	0	1	0	0	0
148	0	1	0	1	1	0
149	0	1	0	1	0	1
150	7	0	0	0	0	0
151	0	1	0	1	0	0
152	1	0	1	1	0	0
153	2	1	0	1	0	0
154	1	1	0	0	0	0
155	5	1	0	0	0	0
156	6	0	1	1	0	0
157	0	1	1	0	1	1

Continua na próxima página

**Tabela A.2 – continuação da página anterior**

Indivíduo	Tempo (semestres)	Censura	Idade	Sexo	Origem	Ingresso
158	10	0	0	1	0	0
159	4	1	0	1	0	0
160	2	1	0	1	0	0
161	0	1	1	0	0	0
162	2	1	1	0	0	0
163	2	1	0	1	0	0
164	14	0	1	0	0	1
165	1	1	1	1	0	0
166	3	0	0	1	0	0
167	0	1	0	0	0	0
168	5	0	0	1	0	0
169	0	1	0	1	0	0
170	1	1	1	1	0	0
171	9	0	1	0	0	0
172	4	0	0	1	0	1
173	3	0	1	0	0	0
174	3	1	0	1	0	0
175	7	0	1	1	0	0
176	0	1	0	0	0	0
177	6	0	1	1	0	0
178	7	0	0	0	0	0
179	3	0	0	1	0	0
180	9	0	1	1	0	0
181	9	0	1	1	0	0
182	3	0	1	1	0	0
183	3	0	1	0	0	0
184	1	0	0	1	0	0
185	5	0	0	1	0	0
186	1	1	1	0	0	0
187	3	1	1	0	0	1
188	1	1	1	1	0	0
189	6	1	1	0	1	1
190	0	1	1	0	0	0
191	0	1	0	1	0	0
192	7	0	1	1	0	0
193	10	0	1	1	0	0

Continua na próxima página

**Tabela A.2 – continuação da página anterior**

Indivíduo	Tempo (semestres)	Censura	Idade	Sexo	Origem	Ingresso
194	5	1	0	1	0	0
195	0	1	0	1	0	0
196	1	1	1	1	0	0
197	0	1	0	1	0	0
198	1	0	0	1	0	0
199	13	0	0	0	0	0
200	0	1	0	1	0	0
201	8	1	1	1	0	1
202	11	0	0	0	0	0
203	1	1	1	1	0	1
204	1	1	0	1	1	0
205	5	1	0	1	0	0
206	1	1	1	1	0	1
207	0	1	1	1	0	0
208	2	0	1	1	0	0
209	0	1	0	0	1	1
210	5	0	1	0	1	0
211	12	0	0	1	0	1
212	0	1	0	1	0	0
213	7	0	1	1	0	0
214	1	1	0	0	1	0
215	8	0	0	1	0	0
216	3	1	1	1	0	0
217	0	1	1	0	0	0
218	1	0	0	1	1	1
219	9	0	1	0	0	0
220	0	1	0	0	0	0
221	2	1	0	1	0	0
222	8	0	1	1	1	0
223	3	1	0	0	0	0
224	5	0	0	1	1	0
225	6	1	0	0	0	0
226	3	1	0	0	0	1
227	1	0	1	1	0	0
228	3	1	1	0	0	0
229	1	1	0	1	0	0

Continua na próxima página

**Tabela A.2 – continuação da página anterior**

Indivíduo	Tempo (semestres)	Censura	Idade	Sexo	Origem	Ingresso
230	2	1	0	1	0	0
231	4	1	0	1	0	0
232	0	1	0	1	0	0
233	5	0	0	1	1	0
234	0	1	0	1	0	0
235	8	0	1	0	1	0
236	1	1	0	1	0	0
237	6	1	1	1	0	1
238	0	1	1	1	0	0
239	5	0	1	1	0	0
240	0	1	0	1	0	0
241	14	0	1	1	0	1
242	9	0	1	1	1	0
243	1	1	0	0	0	0
244	2	0	1	1	0	0
245	0	1	1	1	0	1
246	0	1	0	1	0	0
247	2	1	0	1	0	0
248	0	1	0	1	0	0
249	3	0	0	1	1	0
250	7	0	1	1	0	0
251	0	1	0	0	0	0
252	3	1	0	1	1	0
253	0	1	1	0	0	0
254	1	0	1	0	0	0
255	9	0	0	1	0	0
256	7	1	0	1	0	1
257	4	1	0	1	1	0
258	0	1	0	1	1	0
259	0	1	1	1	0	0
260	0	1	1	1	0	0
261	1	1	1	1	0	1
262	9	0	0	0	1	0
263	5	0	1	1	0	0
264	6	0	0	1	0	0
265	0	1	0	0	0	0

Continua na próxima página

**Tabela A.2 – continuação da página anterior**

Indivíduo	Tempo (semestres)	Censura	Idade	Sexo	Origem	Ingresso
266	0	1	0	1	0	0
267	3	1	0	0	0	1
268	10	0	1	1	0	0
269	1	0	1	1	0	0
270	0	1	1	0	0	0
271	7	0	1	1	0	0
272	0	1	0	1	0	0
273	7	0	0	1	0	0
274	9	0	1	0	0	1
275	1	0	0	1	0	0
276	2	1	1	0	0	0
277	8	0	1	0	0	0
278	2	1	0	1	0	0
279	2	1	1	1	0	0
280	16	0	0	1	0	0
281	1	1	0	0	0	0
282	0	1	1	1	0	0
283	7	0	1	1	1	1
284	0	1	0	1	1	0
285	6	0	0	1	0	0
286	4	0	1	0	0	0
287	3	1	0	0	0	0
288	0	1	0	1	0	0
289	5	0	1	1	0	0
290	3	1	0	1	0	0
291	11	1	1	1	0	0
292	3	1	1	1	0	1
293	3	0	0	1	0	0
294	3	1	1	0	1	0
295	8	0	1	1	0	0
296	1	1	1	1	0	0
297	1	1	1	1	0	0
298	4	1	1	1	0	0
299	8	0	0	1	0	0
300	1	0	0	1	0	0
301	6	0	1	0	1	0

Continua na próxima página



**Tabela A.2 – continuação da página anterior**

Indivíduo	Tempo (semestres)	Censura	Idade	Sexo	Origem	Ingresso
302	2	1	1	1	0	0
303	0	1	0	1	0	0
304	1	1	0	0	0	0
305	6	1	0	1	0	0
306	8	1	1	1	0	0
307	1	0	0	1	1	0
308	7	0	0	1	0	1
309	0	1	0	1	0	0
310	0	1	0	0	0	0
311	8	0	0	1	0	0
312	8	0	1	1	0	0
313	1	1	0	1	1	0
314	0	0	1	1	0	0
315	6	0	1	0	0	0
316	0	1	0	0	0	0
317	5	1	0	0	0	0
318	0	1	0	0	0	0
319	2	1	0	1	0	0
320	2	1	0	1	0	0
321	0	0	1	1	0	0
322	8	0	0	1	1	0
323	9	0	0	1	0	1
324	0	1	0	0	0	0
325	11	0	0	0	0	0
326	5	1	1	1	0	0
327	1	0	1	0	0	0
328	2	1	0	1	0	0
329	5	0	0	1	0	0
330	3	1	0	1	0	1
331	2	1	0	1	1	0
332	1	1	0	1	0	0
333	0	1	0	1	0	0
334	1	1	1	1	0	0
335	1	1	1	1	0	0
336	3	0	0	1	0	0
337	4	1	0	1	0	1

Continua na próxima página

**Tabela A.2 – continuação da página anterior**

Indivíduo	Tempo (semestres)	Censura	Idade	Sexo	Origem	Ingresso
338	7	0	1	0	0	0
339	1	0	0	1	0	0
340	7	0	0	0	1	0
341	6	0	1	1	1	0
342	8	0	1	1	0	0
343	7	0	1	1	1	0
344	5	0	1	1	0	0
345	0	1	0	1	0	0
346	8	0	1	1	0	0
347	1	1	0	1	0	0
348	6	1	1	1	0	1
349	0	1	1	0	0	0
350	5	1	0	1	0	0
351	7	1	0	1	0	0
352	0	1	0	1	0	0
353	3	1	0	1	0	0
354	11	0	0	1	0	0
355	0	1	1	1	0	0
356	2	1	1	1	0	0
357	3	0	0	1	1	0
358	1	1	0	1	0	0
359	1	1	0	0	0	0
360	3	1	0	1	0	0
361	2	1	0	1	0	0
362	1	0	0	1	0	0
363	0	1	1	1	0	0
364	9	1	0	1	0	0
365	0	0	0	0	0	0
366	1	1	0	1	0	0
367	0	1	1	1	0	0
368	1	1	0	1	0	0
369	0	1	0	1	0	0
370	0	1	0	1	0	0
371	0	1	0	1	0	0
372	5	1	0	1	0	0
373	0	1	0	1	0	0

Continua na próxima página

**Tabela A.2 – continuação da página anterior**

Indivíduo	Tempo (semestres)	Censura	Idade	Sexo	Origem	Ingresso
374	1	0	1	1	0	0
375	0	1	0	1	0	0
376	0	1	0	0	0	0
377	3	1	1	1	0	0
378	1	0	0	1	0	1
379	2	1	0	1	0	1
380	4	1	1	1	0	0
381	1	1	0	1	0	0
382	5	0	1	1	0	0
383	3	1	0	1	0	0
384	2	1	1	1	0	1
385	1	1	0	1	0	0
386	4	1	0	1	1	0
387	3	0	0	0	0	0
388	1	0	0	1	0	0
389	0	1	0	1	0	0
390	5	1	0	1	0	1
391	0	1	0	1	0	0
392	0	1	0	1	1	0
393	0	1	0	1	0	0
394	3	0	1	1	0	0
395	0	0	1	1	0	0
396	3	0	1	1	0	0
397	7	0	0	1	1	0
398	3	1	1	0	0	0
399	1	0	0	0	0	1
400	6	1	0	0	0	0
401	0	1	0	1	0	0
402	4	1	1	1	1	1
403	0	1	1	0	0	0
404	1	0	0	1	0	0
405	1	0	0	1	1	0
406	2	1	0	0	1	1
407	9	0	0	1	1	1
408	0	1	0	1	0	0
409	1	0	0	1	0	0

Continua na próxima página

**Tabela A.2 – continuação da página anterior**

Indivíduo	Tempo (semestres)	Censura	Idade	Sexo	Origem	Ingresso
410	0	1	0	0	0	1
411	0	1	1	0	0	0
412	6	1	0	1	0	0
413	0	1	1	1	0	0
414	0	1	0	0	0	1
415	3	0	0	0	0	0
416	11	1	1	1	0	0
417	2	1	0	1	1	0
418	1	1	1	1	0	0
419	0	1	0	1	1	0
420	7	0	0	1	0	0
421	0	0	1	0	0	0
422	0	1	0	1	0	0
423	1	1	0	0	0	0
424	4	1	1	1	0	0
425	5	1	1	0	0	0
426	2	1	0	1	1	0
427	3	1	0	0	0	0
428	0	1	0	1	0	0
429	5	0	1	1	0	0
430	2	1	1	1	0	0
431	2	1	0	0	0	1
432	1	1	0	1	0	0
433	0	1	0	1	0	0
434	0	1	0	1	0	0
435	4	1	0	1	0	0
436	0	1	0	0	1	0
437	2	1	0	0	0	0
438	3	1	0	1	0	0
439	0	1	1	1	0	0
440	10	0	1	1	1	1
441	0	1	0	1	0	0
442	3	0	0	0	0	0
443	1	0	0	1	0	0
444	0	1	0	0	0	1
445	0	1	0	1	0	0

Continua na próxima página

**Tabela A.2 – continuação da página anterior**

---

Indivíduo	Tempo (semestres)	Censura	Idade	Sexo	Origem	Ingresso
446	0	1	0	0	0	0
447	1	0	0	1	0	0
448	0	1	1	0	0	0
449	4	1	0	0	1	0
450	0	1	0	1	1	0
451	6	0	1	1	1	0
452	8	1	0	1	0	1
453	5	0	1	0	0	0
454	3	1	0	1	0	0
455	1	0	0	1	0	0
456	0	0	0	1	0	0
457	1	1	0	1	1	0
458	1	0	0	0	0	0
459	3	1	1	1	0	0
460	0	1	1	1	0	0
461	14	1	0	0	0	0
462	9	0	0	0	0	0
463	0	1	0	1	0	0
464	12	1	0	0	0	0
465	1	0	0	1	0	0
466	6	1	0	1	0	0
467	5	1	0	0	0	0
468	7	0	0	1	0	0
469	1	1	0	1	0	0
470	1	0	1	1	0	0

---