



UNIVERSIDADE
ESTADUAL DE LONDRINA

MARA CAROLINE TORRES DOS SANTOS

UMA NOVA EXTENSÃO DA DISTRIBUIÇÃO BURR XII

Londrina

2022

MARA CAROLINE TORRES DOS SANTOS

UMA NOVA EXTENSÃO DA DISTRIBUIÇÃO BURR XII

Dissertação apresentada ao Departamento de Matemática da Universidade Estadual de Londrina, como requisito parcial para a obtenção do Título de MESTRE em Matemática Aplicada e Computacional.

Orientador: Prof. Dr. Rodrigo Rossetto Pescim

Londrina
2022

**Catálogo elaborado pela Divisão de Processos Técnicos da Biblioteca Central da
Universidade Estadual de Londrina**

Dados Internacionais de Catalogação -na-Publicação (CIP)

S237n Santos, Mara Caroline Torres dos.
Uma nova extensão da distribuição Burr XII /
Mara Caroline Torres dos Santos. – Londrina, 2022.
67 f. : il.

Orientador: Rodrigo Rossetto Pescim.
Dissertação (Mestrado em Matemática Aplicada e Computacional) Universidade
Estadual de Londrina, Centro de Ciências Exatas, Programa de Pós-Graduação em
Matemática Aplicada e Computacional, 2022.

Inclui Bibliografia.

1. Análise de sobrevivência - Teses. 2. Distribuição Burr XII - Teses. 3. Distri-
buição odd-log-logística Burr XII - Teses. 4. Máxima verossimilhança -Teses. 5. Re-
gressão - Teses. I. Pescim, Rodrigo Rossetto. II. Universidade Estadual de Londrina.
Centro de Ciências Exatas. Programa de Pós-Graduação em Matemática Aplicada e
Computacional. III. Título.

CDU 51

MARA CAROLINE TORRES DOS SANTOS

UMA NOVA EXTENSÃO DA DISTRIBUIÇÃO BURR XII

Dissertação apresentada ao Departamento de Matemática da Universidade Estadual de Londrina, como requisito parcial para a obtenção do Título de MESTRE em Matemática Aplicada e Computacional.

BANCA EXAMINADORA

Prof. Dr. Rodrigo Rossetto Pescim
Universidade Estadual de Londrina

Profa. Dra. Mariana Ragassi Urbano
Universidade Estadual de Londrina

Prof. Dr. Thiago Gentil Ramires
Universidade Tecnológica Federal do Paraná

Londrina, 27 de JANEIRO de 2022.

*À minha mãe Joana de Deus Torres e à minha
irmã Carla Jaqueline Torres Alves*

AGRADECIMENTOS

Primeiramente a Deus por permitir que tudo isso acontecesse.

Agradeço imensamente a minha família, pelo incentivo e apoio incondicional. São eles quem me fortalecem e fizeram com que tudo isso se tornasse realidade.

À Universidade Estadual de Londrina pela oportunidade de fazer o curso de pós-graduação.

Agradeço a todos os professores do PGMAC, em especial ao meu orientador prof. Rodrigo, por sua disponibilidade e pela paciência na orientação deste trabalho.

Aos professores do PGMAC que, direta ou indiretamente, contribuíram com a realização deste trabalho.

Aos amigos e companheiros de Mestrado que me motivaram a melhorar.

SANTOS, Mara Caroline Torres dos. **Uma nova extensão da distribuição Burr XII**. 2022. 67. Dissertação (Mestrado em Matemática Aplicada e Computacional) – Universidade Estadual de Londrina, Londrina, 2022.

RESUMO

O objetivo deste trabalho é apresentar a nova distribuição odd-log-logística Burr XII (OLLBXII) que pode ser utilizada para estudar dados de análise de sobrevivência. A distribuição OLLBXII contém vários casos especiais como as distribuições Burr XII, log-logística, Weibull, dentre outras. Observa-se que a nova função densidade pode ser expressa como uma combinação linear da distribuição Burr XII, assim algumas propriedades matemáticas como momentos ordinários e incompletos, além da função geradora de momentos podem ser calculadas para a nova distribuição a partir da distribuição base. Utilizou-se o método de máxima verossimilhança para estimar os parâmetros do modelo para dados censurados. Além disso, realizou-se sucessivas simulações de Monte Carlo, com 3000 repetições, para diferentes valores dos parâmetros e tamanhos de amostras, com o intuito de estudar a performance da nova distribuição. Analisou-se também três conjunto de dados com e sem censura, comparando a nova distribuição com modelos encaixados e não encaixados, anteriormente descritos na literatura. Iniciou-se uma análise de regressão, com as informações necessárias para utilizar o novo modelo com suporte computacional. Observa-se que todos os estudos realizados tiveram o intuito de verificar a aplicabilidade e flexibilidade do modelo. A distribuição OLLBXII obteve um bom desempenho, sendo possível concluir que a nova distribuição pode ser considerada uma alternativa interessante para a aplicação de dados de sobrevivência.

Palavras-chave: Análise de sobrevivência. Distribuição Burr XII. Distribuição odd-log-logística Burr XII. Máxima verossimilhança. Regressão.

SANTOS, Mara Caroline Torres dos. **A new extension of the Burr XII distribution.** 2022. 67. Dissertação (Mestrado em Matemática Aplicada e Computacional) – Universidade Estadual de Londrina, Londrina, 2022.

ABSTRACT

The aim of this work is to present the new odd-log-logistic Burr XII distribution (OLLBXII) that can be used to study survival data analysis. The OLLBXII distribution contains several special cases such as the Burr XII, log-logistic, Weibull distributions, among others. It is observed that the new density function can be expressed as a linear combination of the Burr XII distribution, then some mathematical properties such as ordinary and incomplete moments, in addition to the moment generating function can be calculated for the new distribution from the base distribution. The maximum likelihood method is used to estimate the model parameters for censored data. Also, successive Monte Carlo simulations are carried out, with 3000 repetitions, for different parameter values and sample sizes to study the performance of the new distribution. Three datasets with and without censorship are also analyzed, comparing the new distribution with nested and non-nested models, previously described in the literature. A regression analysis is started, with the necessary information to use the new model with computational support. It is observed that all the studies carried out are intended to verify the applicability and flexibility of the model. The OLLBXII distribution obtained a good performance, and it is possible to conclude that the new distribution can be considered an interesting alternative for application in survival data.

Keywords: Burr XII distribution. Maximum likelihood. Odd-log-logistic Burr XII distribution. Regression. Survival analysis.

SUMÁRIO

1	INTRODUÇÃO	13
2	REVISÃO DE LITERATURA	15
2.1	ANÁLISE DE SOBREVIVÊNCIA	15
2.1.1	Censura	16
2.1.2	Conceitos básicos	16
2.1.3	Estimador de Kaplan-Meier	18
2.1.4	Função de Verossimilhança	19
2.2	COMPARAÇÃO DE MODELOS	20
2.3	FAMÍLIA ODD LOG-LOGÍSTICA	21
2.3.1	Definições	21
2.3.2	Estudos Relacionados	22
2.4	SIMULAÇÃO DE MONTE CARLO	23
2.5	DISTRIBUIÇÃO BURR TIPO XII (BXII)	23
2.5.1	Definições	23
2.5.2	Propriedades matemáticas	25
2.5.3	Função de verossimilhança	26
2.5.4	Estudos Relacionados	26
2.6	REGRESSÃO DE LOCAÇÃO	27
2.6.1	Definições	27
2.6.2	Estudos Relacionados	28
2.6.3	Modelo Log-Burr XII (LBXII)	28
2.6.4	Estudos Relacionados ao modelo LBXII	29
2.6.5	Análise de resíduos	30
2.6.6	Envelope simulado	31
3	A DISTRIBUIÇÃO ODD LOG-LOGÍSTICA BURR TIPO XII (OLLBXII)	32
4	EXPANSÕES PARA AS FUNÇÕES DENSIDADE E ACUMULADA	35
4.1	PROPRIEDADES MATEMÁTICAS	37
4.1.1	Momentos Ordinários	38
4.1.2	Momentos Incompletos	38
4.1.3	Função geradora de momentos	39

5	ESTIMAÇÃO E INFERÊNCIA	41
5.1	CONJUNTO DE DADOS NÃO CENSURADOS	41
5.2	CONJUNTO DE DADOS CENSURADOS	41
6	ESTUDOS DE SIMULAÇÕES	44
7	MODELO DE REGRESSÃO LOLLBXII	47
8	APLICAÇÕES	49
8.1	CONJUNTO DE DADOS 1: ATUARIA	51
8.2	CONJUNTO DE DADOS 2: MELANOMA	54
8.3	CONJUNTO DE DADOS 3: LINFOMA	55
8.4	CONJUNTO DE DADOS 4: TEOR ALCOÓLICO	58
9	CONSIDERAÇÕES FINAIS	61
	REFERÊNCIAS	62

LISTA DE FIGURAS

2.1	Representação dos tipos de censura à direita, em que \bullet representa falha e \circ a censura.	17
2.2	Gráfico de Kaplan-Meier para o tempo de recorrência à infecção do conjunto de dados de câncer no rim, dividido em masculino e feminino.	19
2.3	Função densidade de probabilidade para a distribuição BXII.	24
2.4	Função taxa de falha para a distribuição BXII.	25
2.5	Função densidade de probabilidade da distribuição LBXII.	30
3.1	Função densidade de probabilidade para a distribuição OLLBXII.	33
3.2	Função densidade de probabilidade para a distribuição OLLBXII.	33
3.3	Função taxa de falha para a distribuição OLLBXII.	34
4.1	Medidas de assimetria e curtose para a distribuição OLLBXII, com parâmetros $s = 3.5$, $k = 2.5$, $c = 1.5$ e α uma sequência entre 1.1 e 10.1.	39
6.1	Vieses dos EMVs para os parâmetros s, k, c, α versus diferentes tamanhos da amostra.	46
7.1	Função densidade de probabilidade para o modelo de regressão LOLLBXII.	48
8.1	Estimação da densidade para a variável tempo até a morte de mulheres aposentadas.	53
8.2	Função de sobrevivência estimada e empírica das distribuições BXII e OLLBXII para a variável tempo do tratamento pós-operatório até a morte de pacientes com melanoma.	56
8.3	Função de sobrevivência estimada e empírica para os modelos BXII e OLLBXII para a variável tempo até a morte de pacientes com linfoma que foram transplantados.	57
8.4	Resíduos para a regressão LOLLBXII dos dados teor alcoólico.	60

LISTA DE TABELAS

6.1	Resultado da simulação de Monte Carlo: Tamanho da amostra, média estimada dos parâmetros, EP e REQMs dos parâmetros $s = 4, k = 12, c = 4,2$ e $\alpha = 5, 5$	44
6.2	Resultado da simulação de Monte Carlo: Tamanho da amostra, média estimada dos parâmetros, EP e REQMs dos parâmetros $s = 6, k = 1, c = 1$ e $\alpha = 0, 5$	44
6.3	Resultado da simulação de Monte Carlo: Tamanho da amostra, média estimada dos parâmetros, EP e REQMs dos parâmetros $s = 3, k = 10, c = 0,2$ e $\alpha = 15$	45
6.4	Resultado da simulação de Monte Carlo: Tamanho da amostra, média estimada dos parâmetros, EP e REQMs dos parâmetros $s = 4, k = 12, c = 4,2$ e $\alpha = 5, 5$	45
8.1	Principais medidas descritivas da variável tempo até a morte de mulheres aposentadas.	51
8.2	EMVs, erro padrão em parênteses e as estatísticas AIC, CAIC e BIC para a variável tempo até a morte de mulheres aposentadas.	52
8.3	Teste RV para a variável tempo até a morte de mulheres aposentadas.	52
8.4	Teste qualidade do ajuste para a variável tempo até a morte de mulheres aposentadas.	54
8.5	EMVs, erro padrão (entre parênteses) e as estatísticas AIC, CAIC e BIC para a variável tempo do tratamento pós-operatório até a morte de pessoas com melanoma.	55
8.6	Teste RV para a variável tempo do tratamento pós-operatório até a morte de pessoas com melanoma.	55
8.8	Teste RV para a variável tempo até a morte de pacientes com linfoma que foram transplantados.	56
8.7	EMVs, erro padrão (entre parênteses) e as estatísticas AIC, BIC e CAIC para a variável tempo até a morte de pacientes com linfoma que foram transplantados.	57
8.9	EMVs, erro padrão (entre parênteses), p-valor [.] e as estatísticas AIC, CAIC e BIC para os dados de teor alcoólico.	58
8.10	Intervalo de confiança de 95% para os parâmetros da regressão LOLLBXII da variável explicativa teor alcoólico.	58
8.11	EMVs, erro padrão (entre parênteses), p-valor [.] e as estatísticas AIC, CAIC e BIC para os dados de teor alcoólico, sem a covariável x_2	59

1 INTRODUÇÃO

As distribuições de probabilidades univariadas são fundamentais para a ciência estatística, sendo uma ferramenta indispensável no desenvolvimento da inferência e dos métodos estatísticos. Essas distribuições têm sido amplamente utilizadas nas últimas décadas para ajustar conjuntos de dados em vários campos de pesquisa. No entanto, em muitas áreas aplicadas como ciências médicas e biológicas, confiabilidade, engenharias, sociologia, epidemiologia, criminologia, finanças e seguros, existe uma clara necessidade de formas estendidas dessas distribuições univariadas. Neste sentido, foram realizados grandes avanços para a generalização de algumas distribuições bem conhecidas e suas aplicações em diversos problemas. Irving W. Burr em 1942 introduziu um sistema de doze distribuições baseado em generalizações da equação diferencial de Pearson [9]. A distribuição Burr do tipo XII é uma das soluções dessa equação diferencial que foi discutida em detalhes no artigo em [9, 38].

A distribuição Burr XII (BXII) é muito utilizada para modelagem de dados de tempo de vida e fenômenos com taxas de falha monótonas e unimodais. Esse modelo probabilístico possui inúmeras aplicações nas ciências da saúde, agronomia, ciências econômicas e atuariais. Ao modelar taxas de risco monótonas, as distribuições Weibull e gama podem ser uma escolha inicial por causa de suas formas de densidade assimétricas. No entanto, elas não fornecem um ajuste razoável para taxas de falha não monótonas, como as formas de banheira e unimodal, que são comuns em estudos de confiabilidade e nas ciências biológicas. Para mais detalhes, ver [50] e [51]. Além disso, a função de distribuição acumulada (fda) da distribuição BXII tem forma fechada e, por esta razão, os cálculos da função de verossimilhança são matematicamente tratáveis para dados censurados. Devido à flexibilidade da distribuição BXII em acomodar os formatos unimodal e monótono da função de risco, é considerado um modelo importante em uma variedade de problemas em análise de sobrevivência e estudos de confiabilidade.

No contexto das distribuições generalizadas, a distribuição BXII tem sido amplamente estudada nas últimas décadas por muitos autores como em [50, 51] que introduziram as distribuições Beta Burr XII (BBXII) e Kumaraswamy Burr XII (KwBXII) com aplicações para dados censurados usando uma abordagem Bayesiana. Em [3] propuseram a distribuição exponenciada Burr XII (EBXII) para estudos de confiabilidade. Em [21] foi estudada a distribuição Burr XII com sete parâmetros, chamada Burr XII estendida (EB), utilizando o gerador Marshal-Olkin estendido. Mais recentemente, [4] e [37] trabalharam também com o gerador estendido Marshal-Olkin baseado na distribuição BXII, em que [4] propuseram a distribuição Marshall-Olkin estendida Burr tipo XII (MOEBXII), já [37] procuraram explicitar algumas expressões para encontrar propriedades matemáticas como a função inversa dos momentos. Em [27] definiu-se o modelo McDonald Burr XII (McBXII) a fim de fornecer mais aplicabilidade para as funções densidade e taxa de falha. Em [19] foi proposto uma extensão para a Burr XII

chamada de série de potência Burr XII (BXIISP). A distribuição Poisson exponenciada Burr XII (Exp-BXIIP) foi proposta em [65] com cinco parâmetros. Em [28] foram estudadas as extensões das distribuições Zografos-Balakrishnan Burr XII, Ristié-Balakrishnan Burr XII e a Weibull Burr XII. Em [30] foi estudada uma extensão com quatro parâmetros, a distribuição Marshal-Olkin Burr XII estendida (GMOBXII). Em [25] foi analisada a extensão Burr XII Burr XII (BXII-BXII). A distribuição Burr XII modificada foi proposta por [33].

Apesar das distribuições de probabilidade BBXII, KwBXII, EB e McBXII possuírem entre cinco e sete parâmetros, os quais permitem densidades mais flexíveis do que da distribuição BXII, suas funções densidade são relativamente complexas e muitas vezes matematicamente intratáveis. Além disso, distribuições com muitos parâmetros (cinco ou mais) podem ter problemas de identificabilidade e complicações no processo de estimação de parâmetros usando o método de máxima verossimilhança [55]. Portanto, existe uma necessidade evidente de novas extensões da distribuição BXII considerando famílias de distribuições de probabilidade com menos parâmetros que são flexíveis para acomodar uma ampla variedade de formas de funções de sobrevivência. Para uma distribuição base arbitrária com fda $G(x)$, em [18] foi proposta a função de densidade de probabilidade (fdp) e fda da distribuição *odd log-logistic-G* (OLL-G) com um parâmetro de forma adicional $\alpha > 0$. Exceto por algumas escolhas especiais das funções $g(x)$ e $G(x)$, a função densidade $f(x)$ será difícil de lidar com alguma generalidade.

Observa-se que o fato da sua fda ter um formato fechado é particularmente útil para análise de dados de sobrevivência com censura. Nota-se que não existem funções especiais em contraste com outros geradores, como a família beta generalizada (Beta-G) [22], que inclui dois parâmetros extras e também envolve a função beta incompleta, a família McDonald generalizada (Mc-G) [5], que inclui três parâmetros extras e também envolve a função beta incompleta e a família Kummer beta generalizada (KB-G) [54], que inclui três parâmetros extras e envolve a função hipergeométrica confluyente.

Nesse contexto, o objetivo deste trabalho, foi definir uma nova distribuição de probabilidade, pertencente à família de distribuições OLL-G, denominada *odd log-logística Burr XII* (OLLBXII). Observa-se que a nova distribuição é muito mais flexível do que a distribuição BXII, além de ser um modelo com menos parâmetros, o que a torna mais simples do que os modelos probabilísticos discutidos anteriormente e ao mesmo tempo permite uma melhor interpretação além de ampla aplicabilidade.

2 REVISÃO DE LITERATURA

Este capítulo permeia a revisão dos principais conceitos estudados como, uma introdução à análise de sobrevivência, algumas estatísticas utilizadas para comparar modelos, além da família odd log-logística generalizada, a simulação de Monte Carlo, exposição das características da distribuição Burr XII, modelos de regressão de locação, e por fim, são ressaltadas as características do modelo de regressão log-Burr XII.

2.1 ANÁLISE DE SOBREVIVÊNCIA

A análise de sobrevivência, teve um crescimento muito amplo nas últimas décadas, impulsionada pelo desenvolvimento de computadores mais potentes que permitiram o processamento de grande volume de dados [13]. A variável resposta para a análise de dados de sobrevivência é o tempo até o evento de interesse. Desta forma, as distribuições de probabilidade com funções densidade assimétrica são muito utilizadas nesse tipo de estudo.

A variável tempo, também chamada de tempo até a falha ou tempo de sobrevivência, indica qual foi o tempo que um indivíduo sobreviveu durante o período estudado. Pode se referir a horas, dias, semanas, meses, trimestres, anos, assim como a idade de um indivíduo quando um evento ocorreu. Assim, alguns exemplos de tempo de sobrevivência incluem nascimento até a morte, tempo ao entrar em uma pesquisa clínica até a morte, progressão, cura ou recidiva da doença, tempo de falha de componentes mecânicos e eletrônicos, tempo até a aposentadoria, tempo até a migração de um país ao outro [13, 41], etc.

Devido ao crescimento da utilização de análise de sobrevivência, é possível observar diversas áreas de aplicabilidade como em [13], em que o foco principal do livro é a área médica, no qual são apresentadas várias aplicações a diversos tipos de dados, como os de HIV, aleitamento materno, hepatite, leucemia dentre outros.

Em [35] o foco principal são dados atuarias, com a descrição de diversas distribuições, como Pareto, lognormal e gama. Já [59] evidenciou o modelo de Cox e suas extensões. Observa-se que [68] utilizou o método de máxima verossimilhança para estimar a distribuição Burr tipo XII, que será estudada posteriormente neste trabalho.

Quando se trabalha com dados de tempo de vida, em que a observação é dada pelo acompanhamento de um grupo de pessoas durante um longo período de tempo, pode ocorrer que certo indivíduo não continue o acompanhamento, não sendo possível assim, determinar se o evento de interesse ocorreu. Assim, observa-se que a característica principal de dados de sobrevivência é a observação parcial da resposta, nesse caso se diz que os dados são censurados.

2.1.1 Censura

A censura ocorre quando existe alguma informação sobre o tempo de sobrevivência, mas não se sabe quando o evento ocorreu ou se ocorreu. Nesse sentido, pode-se dizer que censura é a observação parcial da resposta. Os dados censurados podem ser classificados em censura à esquerda, que ocorre quando o tempo de sobrevivência é menor ou igual ao início do estudo, ou seja, quando se sabe que o evento ocorreu antes de um tempo especificado; censura intervalar em que o tempo de sobrevivência está dentro de um intervalo conhecido; censura à direita acontece quando o evento não ocorre antes do estudo ser finalizado, isto é, ao acompanhar um indivíduo por certo período e este sair do estudo por qualquer razão que não seja o evento de interesse, o tempo registrado é considerado censura à direita [13].

Existem três tipos de censura à direita: tipo I, tipo II e aleatória [13, 41, 36].

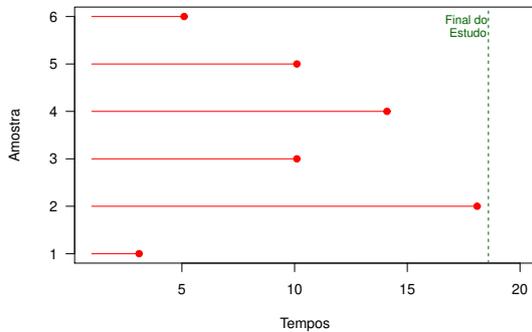
- Tipo I: O tempo de censura é pré-definido, como por exemplo em um experimento com animais em que inicia-se em um tempo específico, e todos os animais são acompanhados até o tempo já especificado para finalizar o estudo, tornando-se censura os animais que não atingiram o evento de interesse.
- Tipo II: A análise é realizada até um número especificado de eventos. Um exemplo pode ser no setor industrial em que o tempo até a falha de certo dispositivo é estudado. Assim quando atingirem um valor n especificado de eventos, o experimento é interrompido.
- Censura aleatória: Como o próprio nome diz é uma censura que ocorre por qualquer razão. Um exemplo é o paciente desistir e não querer mais ser acompanhado pela pesquisa. Se a desistência não está atrelada ao fato do paciente estar próximo a morte, então essa análise pode ser continuada, mas se uma quantidade considerável de pacientes desistem quando estão perto da morte, então esse tempo de censura pode conter um viés muito grande.

A Figura 2.1 exibe os três tipos de censura à direita apresentados anteriormente. Observa-se que (a) representa uma pesquisa em que o evento de interesse ocorreu em toda a amostra estudada, em (b) é indicado que o tempo de estudo acabou e algumas observações não experimentaram o evento de interesse, em (c) assim que o número de eventos estabelecidos ocorreu o estudo foi finalizado, por fim em (d) a censura é a aleatória, ou seja, por algum motivo algumas observações saíram do estudo e outras não apresentaram a falha até o estudo acabar. Como esse tipo de censura é muito comum em análise de dados de tempo de vida, este estudo tem por escopo censuras à direita não informativas.

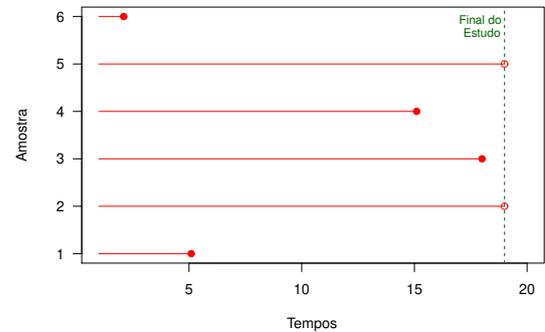
2.1.2 Conceitos básicos

Seja X uma variável aleatória contínua (não-negativa) que representa o tempo de sobrevivência, isto é, o tempo até o evento de interesse ocorrer. Além disso, quando se deseja

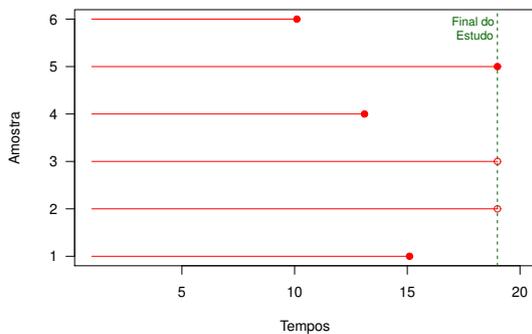
Figura 2.1: Representação dos tipos de censura à direita, em que ● representa falha e ○ a censura.



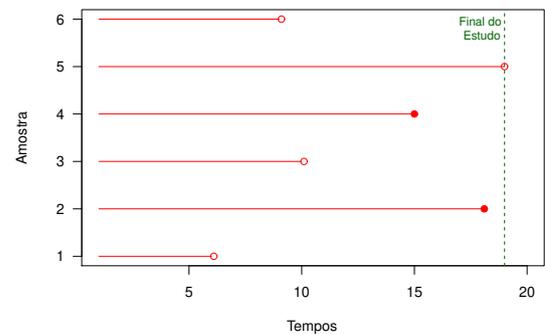
(a) Dados Completos



(b) Dados com censura tipo I



(c) Dados com censura tipo II



(d) Dados com censura aleatória

Fonte: Autora.

representar a censura, C pode ser considerada uma variável aleatória que informa o tempo até a censura. O indicador de censura é dado por,

$$\delta_i = \begin{cases} 1, & \text{se } X_i \leq C_i \\ 0, & \text{se } X_i > C_i, \end{cases}$$

e as observações são então definidas como $x_i = \min(X_i, C_i)$, para $i = 1, \dots, n$. Ao estudar dados de análise de sobrevivência, uma das principais funções é a função de sobrevivência $S(x)$ que representa a probabilidade de sobreviver, ou não falhar, até o tempo x especificado, isto é, a probabilidade da variável aleatória X ser maior que o tempo especificado x . A função $S(x)$ é representada por

$$S(x) = P(X \geq x),$$

em que $0 < x < \infty$, podendo ser reescrita em termos da fda, $F(x) = 1 - S(x)$, é interpretada como sendo a probabilidade de uma observação não sobreviver ao tempo x [13, 61].

A função de risco ou taxa instantânea de falha, fornece o potencial instantâneo por unidade de tempo para o evento ocorrer, dado que o evento ocorre [13, 18], e é dada por,

$$h(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X \leq x + \Delta x | X \geq x)}{\Delta x}.$$

Note que, é possível estabelecer uma relação entre as funções de sobrevivência e taxa de falha [13, 59], isto é

$$S(x) = \exp \left[- \int_0^x h(u) du \right] = \exp [-H(x)],$$

em que $H(x)$ representa a função taxa de falha acumulada, que é a área sob a função de risco no tempo x . A fdp também pode ser escrita em termos da função de sobrevivência, isto é

$$f(x) = \frac{d}{dx} F(x) = -\frac{d}{dx} S(x).$$

Além disso, a função de risco pode ser reescrita a partir da função densidade [61],

$$h(x) = \frac{f(x)}{S(x)}.$$

O tempo médio de sobrevivência é definido como o valor esperado, sendo representado por,

$$\mu = E(X) = \int_0^{+\infty} x f(x) dx.$$

2.1.3 Estimador de Kaplan-Meier

O estimador não paramétrico de Kaplan-Meier (KM) é uma ótima opção, principalmente quando existe a presença de censura, para estimar a função de sobrevivência $S(x)$. Também denominado de estimador limite-produto, o estimador de KM, foi desenvolvido por Kaplan e Meier em 1958 [34], em que $\hat{S}(x)$ é uma função escada, com degraus que aparecem quando ocorre a falha ou evento de interesse.

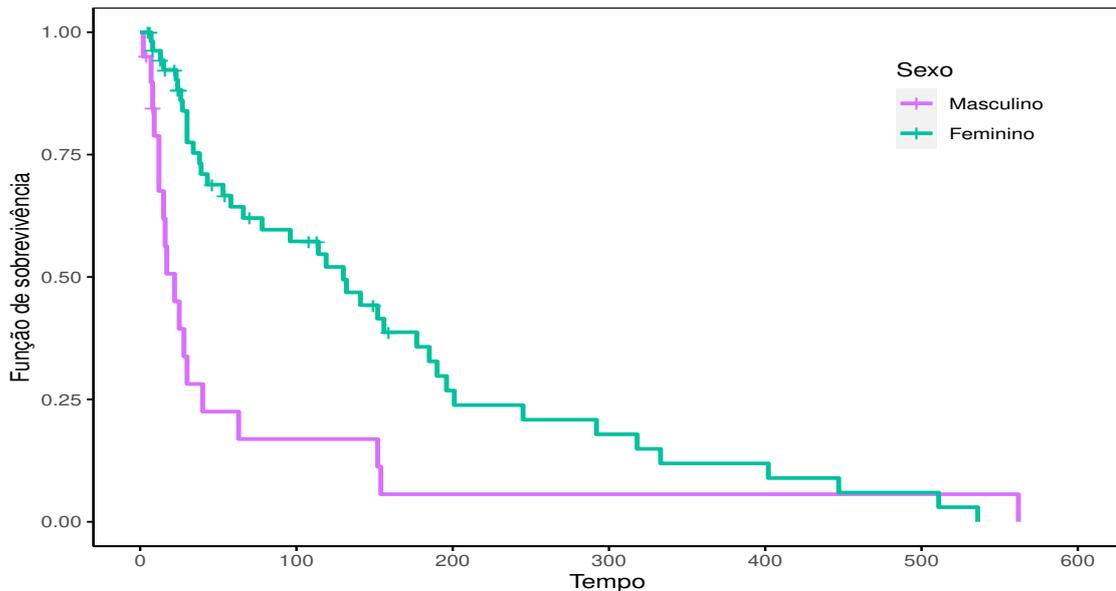
Assim, o gráfico de KM é utilizado para expressar o tempo de sobrevivência, devido a facilidade de incluir censura e é muito utilizado em estudos clínicos e também em estudos de confiabilidade [13]. O estimador de KM é definido por

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

em que, x_1, \dots, x_k são os k tempos ordenados onde o evento de interesse ocorreu, d_j são as falhas em x_j e n_j é o número de indivíduos em risco em x_j , para $j = 1, \dots, k$. A Figura 2.2 apresenta um exemplo da aplicação do gráfico de KM para o conjunto de dados *kidney* obtido

na biblioteca *survival* no *software* R, em que os dados referem-se aos tempos de recorrência à infecção no ponto de inserção do cateter em pacientes renais usando equipamento de diálise portátil.

Figura 2.2: Gráfico de Kaplan-Meier para o tempo de recorrência à infecção do conjunto de dados de câncer no rim, dividido em masculino e feminino.



Fonte: Autora.

Desta forma, são propriedades do estimador de KM [13]: não viciado para amostras grandes; ser fracamente consistente; convergir assintoticamente a um processo Gaussiano e ser um estimador de máxima verossimilhança para $S(x)$.

2.1.4 Função de Verossimilhança

A função de verossimilhança é muito importante em estatística, sendo utilizada no processo de estimação de parâmetros. A função de verossimilhança para dados não censurados é dada por,

$$L(\theta; x_1, \dots, x_n) = \prod_{i \in F} f(x_i, \theta),$$

em que, x_1, \dots, x_n são as realizações de um amostra aleatória, F representa o conjunto de observações das falhas [13].

Se algumas observações são censuradas, é necessário ajustar a expressão, em que, quando é observada a falha usa-se a fdp, e para censura à direita é utilizado a função de sobrevivência, pois tem-se conhecimento da observação mas apenas até um tempo particular [13]. Assim,

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i, \theta)^{\delta_i} S(x_i, \theta)^{1-\delta_i} = \prod_{i=1}^n h(x_i, \theta)^{\delta_i} S(x_i, \theta) = \prod_{i \in F} f(x_i, \theta) \prod_{i \in C} S(x_i, \theta),$$

em que, F representa o conjunto de observações que apresentaram falha e C o conjunto de observações que apresentaram censura.

2.2 COMPARAÇÃO DE MODELOS

Ao analisar diferentes distribuições, é importante utilizar algum método para compara-las, em que para as estatísticas utilizadas, o menor resultado é um indicativo de maior qualidade e simplicidade. Sendo assim, a seguir tem-se as estatísticas que serão utilizadas para realizar a comparação de distribuições encaixadas ou não.

- Critério de informação de Akaike (**AIC**), é uma forma de estimar a quantidade de informações que foram perdidas de determinado conjunto de dados quando utiliza-se certo modelo, assim, seja l o valor máximo do logaritmo da função de verossimilhança e d o número de parâmetros, a estatística é representada da seguinte forma

$$\text{AIC} = 2l + 2d.$$

Para mais informações, ver em [2].

- Critério de informação consistente de Akaike (**CAIC**), leva em consideração o AIC e o tamanho da amostra n , e é dado por

$$\text{CAIC} = \text{AIC} + \left[\frac{(2(d+2)(d+3))}{(n-d-3)} \right].$$

- Critério de informação de Akaike corrigido (**AICc**), é utilizado quando o tamanho amostral é muito pequeno, realizando-se uma modificação no AIC e é dada por

$$\text{AICc} = \text{AIC} + 2 \frac{[d(d+1)]}{(n-d-1)}.$$

- Critério de informação Bayesiano (**BIC**) leva em consideração l , d e o logaritmo de n , é representado por

$$\text{BIC} = l + d \log(n).$$

- Estatística de Cramér-Von Mises (W^*), leva em consideração a variável resposta orde-

nada z , o tamanho da amostra n , a sequencia de 1 até n dada por i , foi proposto

$$W^2 = \sum_{i=1}^n \left(z - \frac{i-1}{2 * n} \right)^2 + \left(\frac{1}{12n} \right)$$

inicialmente, depois foram feitas algumas alterações e é dados por

$$W^* = W^2 \left[1 + \left(\frac{0.5}{n} \right) \right].$$

Para mais informações, ver em [10, 11].

- Estatística de Anderson Darling (A^*), leva em consideração a variável resposta ordenada z , o tamanho da amostra n , a seqüência de 1 até o tamanho da amostra i , e

$$A^2 = -n - \left(\frac{1}{n} \right) \sum_{i=1}^n [(2i-1)(\log z) + (2n+1-2i)(\log(1-z))]$$

e é dada por

$$A^* = A^2 \left[1 + \left(\frac{0.75}{n} \right) + \left(\frac{2.25}{n^2} \right) \right].$$

Para mais informações, ver em [6, 10].

2.3 FAMÍLIA ODD LOG-LOGÍSTICA

Desenvolver uma nova distribuição baseada em uma já conhecida é uma prática comum em estatística. Dessa forma, a família odd log-logística pretende agregar flexibilidade as distribuições conhecidas e ao mesmo tempo permitir fácil interpretação [18, 57].

2.3.1 Definições

O método da família de distribuições odd log-logística (OLL-G) gera uma nova distribuição a partir de uma distribuição base $G(x)$, em que acrescenta um novo parâmetro $\alpha > 0$, ao modelo existente. Observa-se que o parâmetro α representa, o quociente do logaritmo da distribuição gerada com a distribuição base. Assim, as funções fdp e fda dessa nova distribuição respectivamente $f(x)$ e $F(x)$ são construídas a partir do novo parâmetro inserido. Logo, a fdp é dada em (2.1) e a fda em (2.2) [60, 18, 16],

$$f(x) = \frac{\alpha g(x) \{G(x) [1 - G(x)]\}^{\alpha-1}}{\{G(x)^\alpha + [1 - G(x)]^\alpha\}^2} \quad (2.1)$$

e

$$F(x) = \int_0^{\frac{G(x)}{G(x)}} \frac{\alpha t^{\alpha-1}}{(1+t^\alpha)^2} dt = \frac{G(x)^\alpha}{G(x)^\alpha + \bar{G}(x)^\alpha}, \quad (2.2)$$

em que α e $\bar{G}(x)$ são dados por

$$\alpha = \frac{\log \left[\frac{F(x)}{\bar{F}(x)} \right]}{\log \left[\frac{G(x)}{\bar{G}(x)} \right]} \quad \text{e} \quad \bar{G}(x) = 1 - G(x).$$

Além disso, a função de risco da nova distribuição é dada por

$$h(x) = \frac{\alpha g(x) G(x)^{\alpha-1}}{\bar{G}(x) [G(x)^\alpha - \bar{G}(x)^\alpha]}. \quad (2.3)$$

Ao obter a nova distribuição da família odd log-logística, é possível simular dados, por meio da função quantílica, isto é,

$$x = Q_G \left(\frac{u^{1/\alpha}}{(1-u)^{\frac{1}{\alpha}} + u^{\frac{1}{\alpha}}} \right),$$

sendo Q_G a função quantílica da distribuição base $G(x)$, U uma variável aleatória que segue distribuição uniforme no intervalo $(0,1)$. Além disso, a função de log-verossimilhança para esta generalização para $\gamma = (\alpha, \boldsymbol{\xi}^T)^T$, em que $\boldsymbol{\xi}$ é um vetor de parâmetros, é dada por,

$$\begin{aligned} \ell(\gamma) = & r \log \alpha + \sum_{i \in F} \log[g(x_i; \boldsymbol{\xi})] + (\alpha - 1) \sum_{i \in F} \{\log[G(x_i; \boldsymbol{\xi})] + \log[1 - G(x_i; \boldsymbol{\xi})]\} \\ & - 2 \sum_{i \in F} \{\log\{G(x_i; \boldsymbol{\xi})^\alpha + [1 - G(x_i; \boldsymbol{\xi})]^\alpha\}\} + \sum_{i \in C} \log \left\{ \frac{\bar{G}(x_i; \boldsymbol{\xi})^\alpha}{G(x_i; \boldsymbol{\xi})^\alpha + \bar{G}(x_i; \boldsymbol{\xi})^\alpha} \right\} \end{aligned}$$

em que r representa o número de falhas e $g(x)$ a fdp da distribuição base G . Assim, o vetor escore $U(\gamma)$ é encontrado derivando-se $\ell(\gamma)$ em relação aos parâmetros.

2.3.2 Estudos Relacionados

Foi proposto por [15] uma generalização da distribuição Weibull, na qual observou a família odd Weibull, mostrando a relação da fda da distribuição Weibull com sua inversa, além da aplicação e do cálculo da verossimilhança. Já [26] fizeram uma extensiva revisão da generalização da família log-logística, além de uma aplicação. A distribuição odd log-logística Weibull modificada foi proposta por [60] que além de aplicações, obtiveram propriedades matemáticas da distribuição, determinaram a função de máxima verossimilhança e um estudo de simulação. Observa-se que [16] propuseram a distribuição odd log-logística half-normal generalizada. Além disso, [18] fizeram uso da família odd-log logística e estudaram as extensões das distribuições Weibull, Fréchet, half-normal generalizada, log-logística e lognormal. Na mesma linha, [57] propuseram a distribuição odd log-logística gamma generalizada (OLLGG) obtendo-se propriedades como a função geradora de momentos, momentos ordinários e incompletos dentre outras.

2.4 SIMULAÇÃO DE MONTE CARLO

Conhecida devido a mesas de cassino chamadas de mesas de Monte Carlo, é muito utilizada em estatística para simular dados, como por exemplo, quando o objetivo é compreender melhor certa distribuição estudada. Outras áreas como economia, física, engenharia, área de finanças em geral, dentre outras também utilizam a simulação de Monte Carlo.

Para simular dados utilizando a simulação de Monte Carlo a partir de uma distribuição definida, é necessário seguir os passos:

- Estabelecer o tamanho da amostra;
- Gerar valores pseudo-aleatórios para a distribuição uniforme, $X \sim U(0, 1)$;
- Determinar a inversa da fda da distribuição que está sendo estudada;
- Estabelecer os parâmetros da distribuição;
- Encontrar os valores pseudo-aleatórios para a inversa da distribuição;
- Estimar os parâmetros pelo método de máxima verossimilhança;
- Replicar o processo da simulação de Monte Carlo k-vezes;
- Calcular a média e as estatísticas adequadas.

Após [1, 18, 28, 33, 51, 60] proporem suas respectivas distribuições, realizaram estudos de simulação para verificar se o modelo é adequado, ou seja, se conforme o tamanho da amostra aumenta, mais próximo os parâmetros estimados ficam dos parâmetros verdadeiros.

2.5 DISTRIBUIÇÃO BURR TIPO XII (BXII)

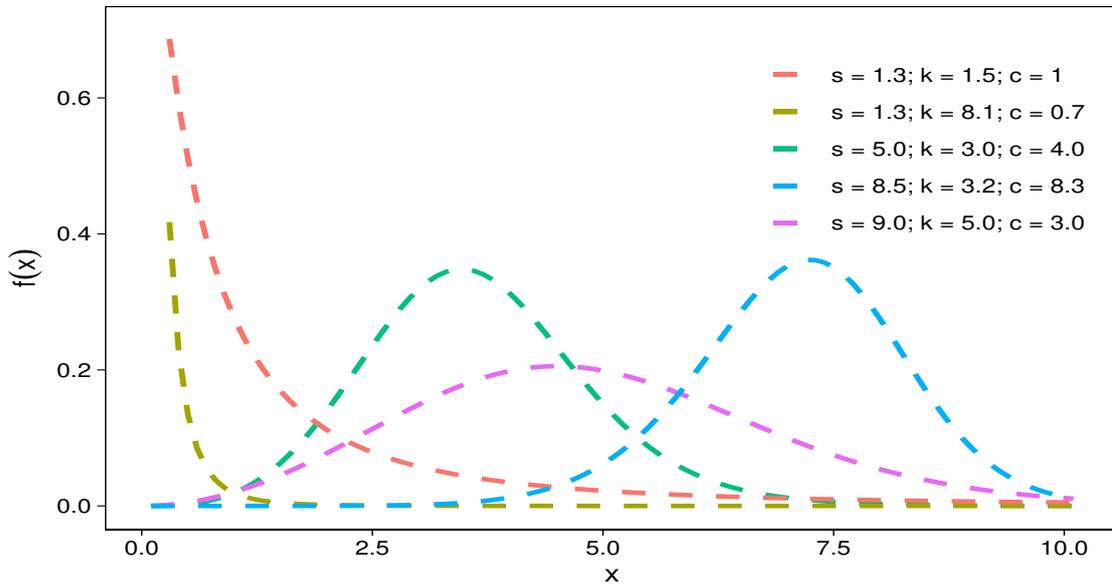
Irving W. Burr em 1942 propôs um sistema com doze equações diferenciais [9]. A 12ª distribuição de Burr, abreviada para BXII, é considerada uma distribuição flexível ao ajustar dados, principalmente devido as funções de risco, sendo comumente utilizada para estudar dados de sobrevivência [49].

2.5.1 Definições

A fdp da distribuição BXII, com parâmetros de forma $c > 0$ e $k > 0$ e parâmetro de escala $s > 0$ é denotada por

$$g(x) = c k s^{-c} \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k-1} x^{c-1}, \quad (2.4)$$

Figura 2.3: Função densidade de probabilidade para a distribuição BXII.



Fonte: Autora.

em que $0 \leq x < \infty$ [63, 50, 49]. A Figura 2.3 mostra as principais formas que a fdp pode assumir.

Uma das características da distribuição BXII é que a fda pode ser escrita de forma fechada, o que facilita na obtenção da função de verossimilhança para dados censurados (descritos na Seção 2.1.4), e pode ser expressa respectivamente por

$$G(x) = 1 - \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k} \quad (2.5)$$

As funções de sobrevivência e risco correspondentes a variável aleatória X que segue distribuição BXII são, respectivamente, representadas por

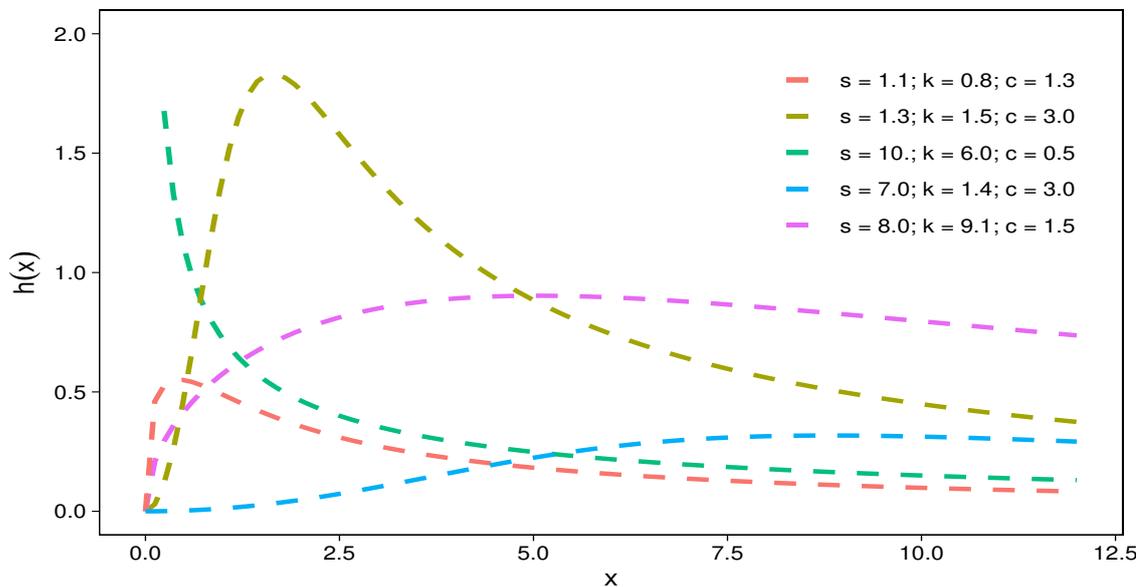
$$S_G(x) = \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k} \quad (2.6)$$

e

$$h_G(x) = \frac{c k \left(\frac{x}{s} \right)^{c-1}}{s \left[1 + \left(\frac{x}{s} \right)^c \right]}. \quad (2.7)$$

Observa-se que a função de risco (2.7) assume as forma decrescente, unimodal e constante, sendo assim, uma alternativa para as distribuições log-logística (LL) e log-normal (LN), pois é uma opção para modelar a função taxa de falha monótona [63], e a unimodalidade que ocorre com frequência na área de confiabilidade e em estudos biológicos. Além disso, a distribuição LL é um casos especial da BXII [49, 51]. A Figura 2.4 mostra diferentes formas da função de risco para a distribuição BXII.

Figura 2.4: Função taxa de falha para a distribuição BXII.



Fonte: Autora.

2.5.2 Propriedades matemáticas

Ao analisar a primeira derivada da função de risco igual a zero $h'_G(x) = 0$, para $x > 0$, tem-se:

- $c \leq 1$: Como $h'_G(x) < 0$ então $h_G(x)$ é decrescente.
- $0 < c < 2$: A função taxa de falha pode ser considerada aproximadamente constante, dependendo do parâmetro s .
- $c > 2$: A função taxa de falha atinge um máximo e decresce, em que o parâmetro s determina o alcance da função de risco.

Para $c > 1$, a função de risco é unimodal. Observa-se que $h'_G(x^*) = 0$ tem-se que $c - 1 - \left(\frac{x^*}{s}\right)^c = 0$. Logo, o ponto crítico é dado por $x^* = s(c-1)^{\frac{1}{c}}$. Além disso, $h_G(x) \rightarrow 0$ quando $x \rightarrow 0$ ou $x \rightarrow \infty$ [63]. A moda é igual a $x = s\left[\frac{(c-1)}{(ck+1)}\right]^{\frac{1}{c}}$, e tem formato em L se $c = 1$ [49].

Por fim, se $ck > n$ tem-se que n é o n -ésimo momento próximo de zero, assim a esperança de X é dada por,

$$E(X^n) = \mu'_n = s^n k B(k - n c^{-1}, 1 + n c^{-1}),$$

em que $B(., .)$ é a função beta [49].

2.5.3 Função de verossimilhança

A função de verossimilhança para dados censurados à direita para $x_i = \min(X_i, C_i)$ é dada por

$$L(c, k, s) = (kc)^r \prod_{i \in F} \left[\left(\frac{x_i}{s} \right)^c x_i^{-1} \left(1 + \left(\frac{x_i}{s} \right)^c \right)^{-(k+1)} \right] \prod_{i \in C} \left[\left(1 + \left(\frac{x_i}{s} \right)^c \right)^{-k} \right],$$

desta forma, a função log-verossimilhança é representada por

$$\begin{aligned} \ell(c, k, s) = & r \log k + r \log c - (k+1) \sum_{i \in F} \log \left[1 + \left(\frac{x_i}{s} \right)^c \right] + \sum_{i \in F} \log \left(\frac{x_i^{c-1}}{s^c} \right) \\ & - k \sum_{i \in C} \log \left[1 + \left(\frac{x_i}{s} \right)^c \right]. \end{aligned}$$

O vetor escore $\mathbf{U} = \left(\frac{\partial \ell}{\partial c}, \frac{\partial \ell}{\partial k}, \frac{\partial \ell}{\partial s} \right)^T$ é denotado por

$$\begin{aligned} \frac{\partial \ell(c, k, s)}{\partial c} &= \frac{r}{c} - (k+1) \sum_{i \in F} \frac{\left(\frac{x_i}{s} \right)^c \log \left(\frac{x_i}{s} \right)}{\left(1 + \left(\frac{x_i}{s} \right)^c \right)} + \sum_{i \in F} \log \left(\frac{x_i}{s} \right) \\ &\quad - k \sum_{i \in C} \frac{\left(\frac{x_i}{s} \right)^c \log \left(\frac{x_i}{s} \right)}{\left[1 + \left(\frac{x_i}{s} \right)^c \right]}, \\ \frac{\partial \ell(c, k, s)}{\partial k} &= \frac{r}{k} - \sum_{i \in F} \log \left[1 + \left(\frac{x_i}{s} \right)^c \right] - \sum_{i \in C} \log \left[1 + \left(\frac{x_i}{s} \right)^c \right], \\ \frac{\partial \ell(c, k, s)}{\partial s} &= c(k+1) \sum_{i \in F} \frac{x_i^c s^{-(c+1)}}{\left[1 + \left(\frac{x_i}{s} \right)^c \right]} - \frac{nc}{s} + c \sum_{i \in C} \frac{x_i^c s^{-(c+1)}}{\left[1 + \left(\frac{x_i}{s} \right)^c \right]}. \end{aligned}$$

2.5.4 Estudos Relacionados

Em [69] foi estudada a BXII com dois parâmetros, com o objetivo de construir os intervalos e as regiões de confiança para a distribuição considerando censura tipo II progressivamente aumentando o número de observações censuradas, [68] estudaram a estimação dos parâmetros da distribuição BXII de dois parâmetros por máxima verossimilhança para ajustar conjunto de dados. Também para a BXII com dois parâmetros [38] apresentaram algumas propriedades estatísticas como quantis, momentos, função geradora de momentos, momentos condicionais, curva Bonferroni e Lorenz, estatística de ordem além da caracterização baseada nos momentos condicionais.

Observa-se em [70] que partindo-se da distribuição BXII de dois parâmetros, pode-se determinar a distribuição BXII de três parâmetros. Dessa maneira o modelo torna-se mais flexível que a distribuição log-normal e permanece com as vantagens computacionais da distribuição log-logística.

Em [49] antes de iniciar um estudo sobre extensões da distribuição BXII, foi feita uma revisão das características da distribuição BXII, características da função de risco, além de apresentar propriedades da distribuição. Além disso, estudaram a estimação da máxima verossimilhança da distribuição BXII, apresentando um algoritmo usado para a implementação e diversos exemplos numéricos que permeiam as características deste modelo, como os limitantes para as distribuições Weibull e Pareto foi estudada por [62].

Em relação as generalizações da distribuição BXII, foram realizados diferentes estudos como [1] que estudaram a extensão de quatro parâmetros chamada Marshall-Olkin Burr XII. A extensão para a BXII utilizando série de potência chamada família BXIIPS foi estudada por [19]. Em relação a distribuição BXII com sete parâmetros, [21] realizaram uma análise teórica extensiva da nova distribuição. Observa-se que [25] analisaram a extensão Burr XII Burr XII chamada de BXII-BXII, já [28] estudaram as extensões Zografos-Balakrishnan Burr XII, Ristić-Balakrishnan Burr XII e a Weibull Burr XII. As distribuições BBXII, log-BBXII e KwB-XII foram propostas por [49]. Ainda, [65] prouseram a distribuição Poisson exponenciada da BXII. Em relação a uma extensão de quatro parâmetros, [30] propuseram a distribuição generalizada Marshall-Olkin Burr-XII estendida chamada de GMOBXII. Já [33] propuseram a distribuição Burr XII modificada.

2.6 REGRESSÃO DE LOCAÇÃO

Diversas pesquisas científicas tem o tempo de sobrevivência sendo afetado por variáveis explicativas. Um exemplo em epidemiologia, é o tempo de um indivíduo se infectar até evoluir para um caso grave de COVID-19 (*coronavirus disease 2019*), em que essa resposta pode depender de fatores como idade, peso, comorbidades e assim por diante. Assim, incluir covariáveis ao modelo pode ser uma ferramenta importante para explicar de maneira mais adequada a variável resposta. Desta forma, estudar modelos que levam em consideração a influência de covariáveis tem relevância na análise de sobrevivência podendo este ser desenvolvido a partir do estudo de regressão. Dentre eles o modelo de regressão de locação [39] é amplamente estudado, frequentemente em pesquisas clínicas, mas também na engenharia, hidrologia e análise de sobrevivência [55, 64].

2.6.1 Definições

Para determinar um modelo de regressão paramétrico de locação a partir de uma variável aleatória Y no intervalo de $(-\infty, \infty)$, utiliza-se a seguinte equação

$$f(y) = \frac{1}{\sigma} g\left(\frac{y - \mu}{\sigma}\right), \quad -\infty < y < \infty, \quad (2.8)$$

em que $-\infty < \mu < \infty$ e $\sigma > 0$ são os respectivos parâmetros de locação e escala e $g(\cdot)$ representa a fdp de uma distribuição base. Para isso é necessário considerar uma transformação da variável aleatória $Y = \log(X)$, no qual X é a variável aleatória de uma distribuição base, sendo necessário utilizar o método do Jacobiano [31, 40], para que $Y \sim \log(X)$. Assim, a equação (2.8) representa a fdp da variável aleatória Y . Além disso, a fda da distribuição é dada por

$$F(z) = \int_{-\infty}^z g(w)dw = 1 - S_G(z), \quad (2.9)$$

no qual, $z = (y - \mu)/\sigma$, $S_G(z)$ é a função de sobrevivência da distribuição base aplicada em z , ou ainda $F(y) = G[y - \mu/\sigma]$ [39]. Observa-se que para $\mu = 0$ e $\sigma = 1$, encontra-se a forma padrão da distribuição. Logo, o modelo de regressão linear para a variável aleatória Y é dado por

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \sigma z_i = \mu(\mathbf{x}) + \sigma z_i, \quad i = 1, \dots, n,$$

em que $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$ representa a matriz de covariáveis, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ representa o vetor de parâmetros desconhecidos que devem ser estimados, σ é o parâmetro de escala e z_i representa os valores observados da variável aleatória que assume distribuição log-G padrão.

2.6.2 Estudos Relacionados

O modelo de regressão de locação descrito em [39] foi aplicado em diversas distribuições base, como os modelos log-gama generalizado dentre outros. O modelo log-beta half-normal generalizado (LBGHN) foi estudado por [56], em que além da aplicação em conjunto de dados, obteve-se algumas propriedades matemáticas. Já [48] também encontraram propriedades como a função geradora de momentos, além da estimação dos parâmetros via a máxima verossimilhança para o modelo log- β -Birnbau-Saunders ($L\beta$ BS).

A distribuição gama exponenciada generalizada (EGG) e também o modelo de regressão log-gama exponenciada generalizada (LEGG) foram propostos por [32]. Ainda, [64] propuseram dois modelos de regressão log-gama-Weibull e log-gama-logística, e [31] estudaram propriedades dos modelos log-gama-Weibull e log-gama-logística.

A análise de resíduos, simulações e aplicações foram estudadas por [55] para a distribuição log-odd log-logística half-normal generalizada (LOLLGHN). Por fim [57] estudaram o modelo log-odd log-logística gamma generalizada (LOLLGG).

2.6.3 Modelo Log-Burr XII (LBXII)

Sabendo-se que a distribuição BXII tem como caso especial a distribuição log-logística (LL) e é uma distribuição muito utilizada em pesquisas clínicas, aplicar o modelo de regressão de locação pode ser uma alternativa viável para o estudo da influência de covariáveis no tempo de falha. Assim, o modelo de regressão log-Burr XII (LBXII), refere-se a regressão

do tipo locação [63] sendo apresentado nesta subsecção.

Definições do modelo LBXII

Seja X uma variável aleatória que segue distribuição BXII com parâmetros c , k e s . Fazendo-se a transformação $Y = \log(X)$, além das reparametrizações $c = 1/\sigma$ e $s = \exp(\mu)$, a fdp da variável aleatória $Y \sim \text{LBXII}$, é dada por,

$$f(y; k, \sigma, \mu) = \frac{k}{\sigma} \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right) \right]^{-(k+1)} \exp\left(\frac{y - \mu}{\sigma}\right) \quad -\infty < y < \infty, \quad (2.10)$$

em que $k > 0$, $\sigma > 0$ e $-\infty < \mu < \infty$. Além disso, a fda e a função geradora de momentos (fgm) são, respectivamente, dadas por

$$F(y; k, \sigma, \mu) = 1 - \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right) \right]^{-k} \quad (2.11)$$

e

$$M_Y(t) = ks^t B [k - tc^{-1}, 1 + tc^{-1}], \quad \text{se} \quad kc > t.$$

Observa-se que o modelo pode ser escrito como um modelo log-linear, Subsecção 2.6.1, ou seja $Y = \mu(x) + \sigma Z$, em que a variável aleatória Z tem densidade dada por

$$f(z) = k[1 + \exp(z)]^{-(k+1)} \exp(z),$$

sendo, $k > 0$, $z = (y - \mu(x))/\sigma$ e $-\infty < z < \infty$. Para as n observações da amostra aleatória $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$, em que, y_i são as observações de uma variável aleatória Y , e x_i representam as correspondentes covariáveis dos dados, a função log-verossimilhança para dados censurados assume a seguinte forma,

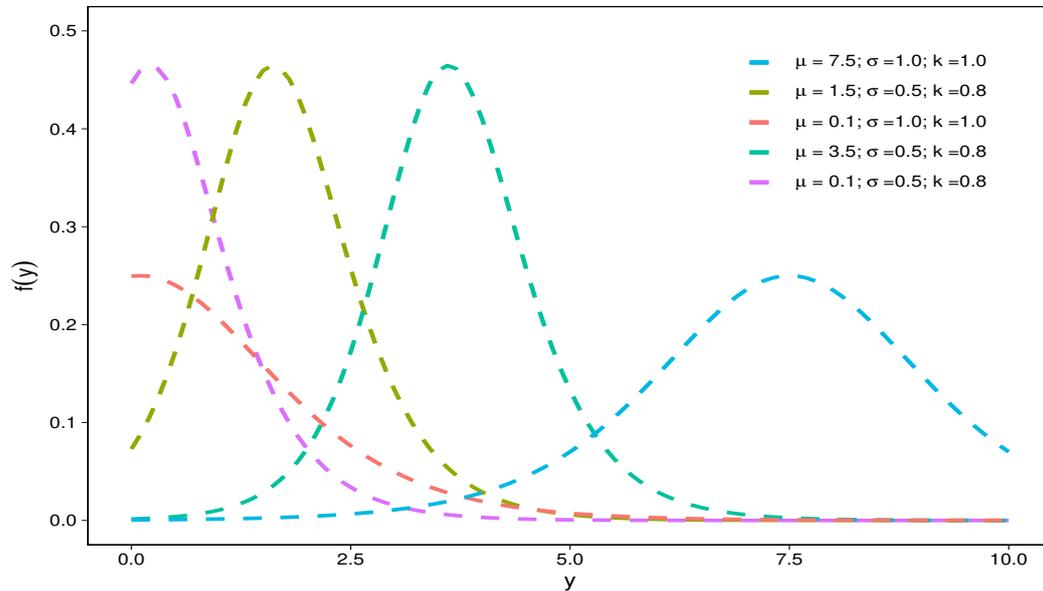
$$\ell(\beta^T s, k, \sigma) = r \log(k) - r \log(\sigma) + \sum_{i \in F} z_i \sum_{i \in F} \log[1 + \exp(z_i)] - k \sum_{i \in C} \log[1 + \exp(z_i)],$$

em que r representa o número de falhas. A Figura 2.5 apresentam as principais formas que a função densidade do modelo de regressão das distribuições LBXII.

2.6.4 Estudos Relacionados ao modelo LBXII

O estudo desenvolvido em [49] faz uma revisão dos principais pontos do modelo de regressão LBXII, pois o mesmo é um caso particular do modelo proposto pela autora. A regressão LBXII foi amplamente estudada por [63], em que para estimar os parâmetros do modelo de regressão de locação para a distribuição BXII efetuou-se uma análise clássica, Bayesiana e jackknife. Além disso, propuseram a simulação de dados com o intuito de comparar

Figura 2.5: Função densidade de probabilidade da distribuição LBXII.



Fonte: Autora.

as performances dos modelos log-logístico (LL) e LBXII, além de fazer análise de resíduos e aplicar a um banco de dados.

2.6.5 Análise de resíduos

A análise de resíduos é realizada como uma maneira de verificar se o ajuste do modelo é adequado aos dados. Sabendo disso, existem várias formas de encontrar os resíduos de um modelo, para identificar possíveis outliers, se o ajuste realmente é adequado ou se é necessário realizar modificações no modelo. Os principais resíduos estudados em análise de sobrevivência são os resíduos de Cox-Snell [20], os resíduos padronizados [13], os resíduos martingale [24], os resíduos deviance modificado [12, 67], caso-deleção [14], etc. Nesta dissertação o foco será nos resíduos martingale e deviance modificado.

Os resíduos martingale [24] são uma modificação dos resíduos Cox-Snell, muito utilizados quando o conjunto de dados vem de processos de contagem, para estimar o número de falhas em excesso que não foi previsto pelo modelo e para verificar se alguma covariável necessita de transformação ou da inclusão de um termo quadrático etc, é dado por

$$r_{M_i} = \delta_i + \log[S(y_i, \hat{\theta})],$$

sendo δ_i o indicador de censura e $S(y_i, \hat{\theta})$ a função de sobrevivência estimada.

Verifica-se que quando os resíduos muito assimétricos é possível realizar um refinamento na obtenção dos resíduos, chamado de resíduo deviance modificado [12, 67] que é

dada pela seguinte expressão

$$r_{D_i} = \text{sinal}(r_{\hat{M}_i}) \{-2[r_{\hat{M}_i} + \delta_i \log(\delta_i - r_{\hat{M}_i})]\}^{(1/2)}.$$

Ao comparar o resíduo deviance modificado com o resíduo martigale, observa-se que o primeiro tende a ser simétrico em torno de zero, devido ao seu maior refinamento, fazendo-se a detecção de pontos atípicos de forma mais assertiva.

Se o ajuste do modelo é adequado, pode-se construir o gráfico dos resíduos versus os valores ajustados ou seu índice, que é uma forma simples e eficaz de detectar a adequabilidade do modelo de regressão permitindo verificar a existência de outliers.

2.6.6 Envelope simulado

O envelope simulado é um gráfico de probabilidade com duas bandas e é utilizado para verificar se o modelo é adequado aos dados. O envelope simulado foi proposto por [7], e para obtê-lo deve-se seguir os seguintes passos:

1. Encontrar o valor esperado, e ordenado

$$E(Z_{[i]}) \approx \phi^{-1} \left(\frac{i - 3/8}{n + 1/4} \right),$$

que é o valor esperado da i -ésima estatística de ordem e ϕ é a função de distribuição acumulada da normal padrão;

2. Gerar n observações $y_{i,1}^*$;
3. Ajustar os dados obtidos no item anterior, sem mudar as variáveis preditoras;
4. Calcular os resíduos para o item anterior e coloca-los de forma crescente;
5. Repetir os três itens anteriores m vezes ($m = 19$, de acordo com [7]);
6. Inserir os valores obtidos no item 1, ou seja, colocar no gráfico $E(Z_{[i]}, t_{[i]})$;
7. Inserir no gráfico na banda superior os valores máximos obtidos dos resíduos;
8. Inserir no gráfico na banda inferior os valores mínimos obtidos dos resíduos;

Para mais informações, consultar [7, 23].

O trabalho segue com a introdução da distribuição proposta neste estudo, que tem como distribuição base a BXII, chamada de distribuição odd log-logística Burr XII (OLLB-XII), tendo por interesse aumentar a flexibilidade da função taxa de falha, o que auxiliaria na aplicação do modelo em diferentes bancos de dados.

3 A DISTRIBUIÇÃO ODD LOG-LOGÍSTICA BURR TIPO XII (OLLBXII)

Acrescentar um parâmetro para distribuições existentes pode ser muito interessante, pois as propriedades desta nova distribuição podem ser estudadas por meio dos resultados da distribuição base, além de ser possível utilizar a mesma para verificar se o novo modelo desenvolvido se adequa melhor ao ser aplicado a um banco de dados.

Com o que foi visto nas Seções 2.3 e 2.5, é possível utilizar a distribuição BXII como base para obter uma nova distribuição ao acrescentar um novo parâmetro α . Assim, a nova distribuição apresentada neste capítulo será chamada Odd Log-Logística Burr XII ou $X \sim \text{OLLBXII}(c, k, s, \alpha)$ ou distribuição de OLLBXII. A fda da nova distribuição é dada por (3.1) e é calculada substituindo-se (2.5) em (2.2)

$$F(x) = \frac{\left[1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right]^\alpha}{\left[1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right]^\alpha + \left[1 + \left(\frac{x}{s}\right)^c\right]^{-\alpha k}}, \quad (3.1)$$

em que $x > 0$ representa o tempo até o evento de interesse, c e $k > 0$ são os parâmetros de forma, $s > 0$ é o parâmetro de escala e $\alpha > 0$ é o parâmetro extra de forma. Observa-se que ao considerar $\alpha = 1$ obtém-se a distribuição BXII. Derivando-se a fda da OLLBXII em relação a x , ou utilizando-se a fórmula (2.1) e substituindo-se por (2.4) e (2.5), obtém-se a fdp da distribuição OLLBXII que é dada por

$$\begin{aligned} f(x) &= \frac{\alpha c k s^{-c} \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k-1} x^{c-1} \left\{\left(1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right) \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right\}^{\alpha-1}}{\left[\left[1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right]^\alpha + \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k\alpha}\right]^2} \quad (3.2) \\ &= c \alpha k x^{-1} \left(\frac{x}{s}\right)^c \frac{\left[1 + \left(\frac{x}{s}\right)^c\right]^{k-1} \left\{\left[1 + \left(\frac{x}{s}\right)^c\right]^k - 1\right\}^{\alpha-1}}{\left\{\left[\left[1 + \left(\frac{x}{s}\right)^c\right]^k - 1\right]^\alpha + 1\right\}^2}. \end{aligned}$$

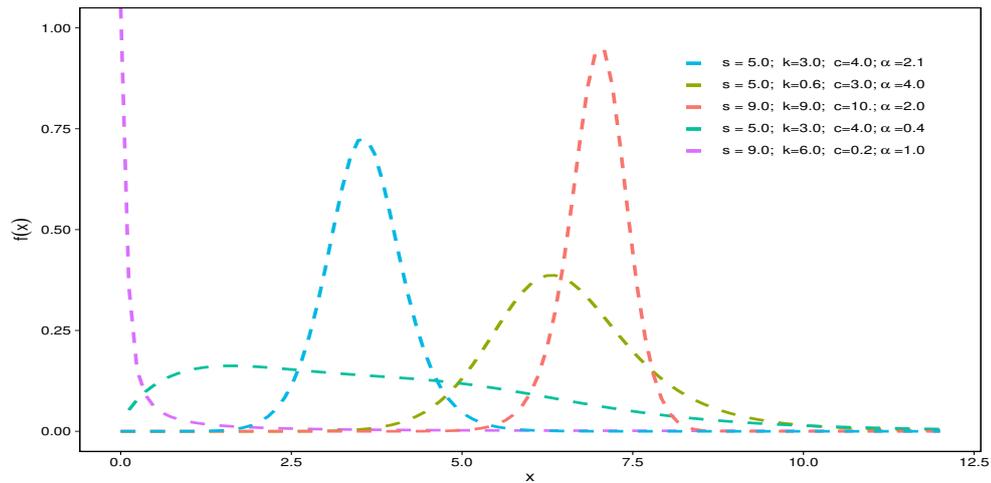
As Figuras 3.1 e 3.2 exibem as principais formas da fdp da nova distribuição, destacando-se as formas decrescente unimodal e bimodal, essa última forma que não ocorria com a distribuição BXII. A função de sobrevivência corresponde a (3.1) e (3.2) é dada por

$$S(x) = \frac{\left[1 + \left(\frac{x}{s}\right)^c\right]^{-\alpha k}}{\left[1 - \left(1 + \left(\frac{x}{s}\right)^c\right)^{-k}\right]^\alpha + \left[1 + \left(\frac{x}{s}\right)^c\right]^{-\alpha k}}. \quad (3.3)$$

A função de taxa de falha dada em (3.4), pode assumir diferentes formas, conforme apresentado pela Figura 3.3, que exhibe as principais formas dessa função,

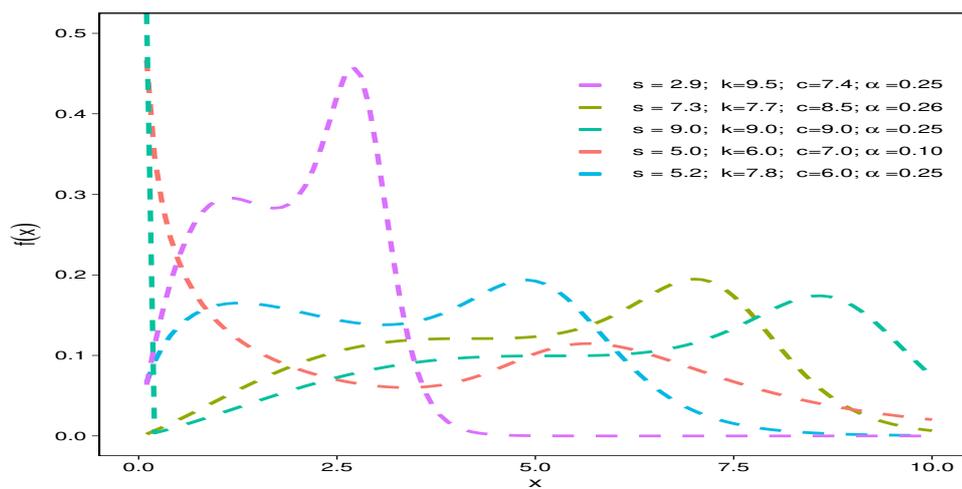
$$h(x) = \frac{\alpha c k s^{-c} \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k-1} x^{c-1} \left\{1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right\}^{\alpha-1}}{\left[1 + \left(\frac{x}{s}\right)^c\right]^{-k} \left\{\left[1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right]^\alpha + \left[1 + \left(\frac{x}{s}\right)^c\right]^{-\alpha k}\right\}}. \quad (3.4)$$

Figura 3.1: Função densidade de probabilidade para a distribuição OLLBXII.



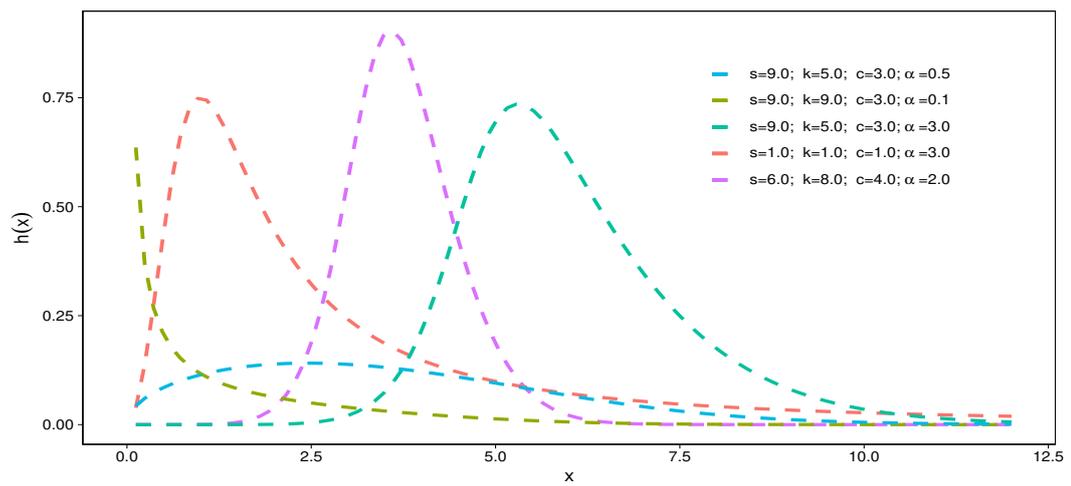
Fonte: Autora.

Figura 3.2: Função densidade de probabilidade para a distribuição OLLBXII.



Fonte: Autora.

Figura 3.3: Função taxa de falha para a distribuição OLLBXII.



Fonte: Autora.

4 EXPANSÕES PARA AS FUNÇÕES DENSIDADE E ACUMULADA

Quando se propõe uma nova distribuição é importante estudar as principais propriedades matemáticas relacionadas com a média, variância, assimetria e curtose. Observa-se que [18, 45, 49, 50, 51, 60], dentre muitos outros autores, propuseram novos modelos aos quais calculou-se as principais propriedades. Com base nestas informações, este capítulo contém os momentos ordinários e incompletos e a função geradora de momentos referentes a distribuição OLLBXII. Assim, primeiramente, realizam-se algumas expansões da fdp e fda necessárias para esta análise.

A família de distribuições G-exponenciada (EG) também chamada distribuição Lehmann tipo I para uma base qualquer $G(x)$, segue distribuição $EG(\psi)$, desta forma a fda e a fdp pode ser escrita, para $\psi > 0$, respectivamente como

$$H_{\psi}(x) = G(x)^{\psi} \quad \text{e} \quad h_{\psi}(x) = \psi g(x) G(x)^{\psi-1}.$$

Para a família EG a fda de G elevada α pode ser escrita como série de potência da seguinte forma,

$$G(x)^{\alpha} = \sum_{i=0}^{\infty} a_i G(x)^i,$$

em que

$$a_i = a_i(\alpha) = \sum_{j=i}^{\infty} (-1)^{i+j} \binom{\alpha}{j} \binom{j}{i},$$

além disso, pode-se reescrever $[1 - G(x)]^{\alpha}$ por meio da expansão binomial da seguinte forma

$$[1 - G(x)]^{\alpha} = \sum_{i=0}^{\infty} (-1)^i \binom{\alpha}{i} G(x)^i.$$

Assim, a fda de $F(x)$ pode ser escrita em função das expansões de $G(x)$,

$$F(x) = \frac{\sum_{i=0}^{\infty} a_i G(x)^i}{\sum_{i=0}^{\infty} b_i G(x)^i},$$

no qual $b_i = a_i + (-1)^i \binom{\alpha}{i}$, que pode ser reescrito como,

$$F(x) = \sum_{i=0}^{\infty} d_i G(x)^i,$$

sendo,

$$d_i = b_0^{-1} \left(a_i - b_0^{-1} \sum_{r=1}^i b_r c_{i-r} \right).$$

Por fim a fdp é dada por,

$$f(x) = \sum_{i=0}^{\infty} d_{i+1} h_{i+1}(x),$$

em que $h_{i+1}(x) = (i+1)g(x)G(x)^i$. A distribuição G-exponenciada é utilizada para expandir novas famílias de distribuições odd log-logística em [18, 60], além de diversas outras distribuições [29, 42, 43, 44, 47, 53].

Desta forma, para encontrar uma expansão para (3.1) se faz uso de uma expansão em séries de potência na fda da distribuição BXII isto é,

$$\left\{ 1 - \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k} \right\}^\alpha = \sum_{i=0}^{\infty} \sum_{j=i}^{\infty} (-1)^{i+j} \binom{\alpha}{j} \binom{j}{i} \left\{ 1 - \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k} \right\}^i. \quad (4.1)$$

Para a expansão binomial generalizada em que $\alpha > 0$,

$$\left\{ 1 - \left\{ 1 - \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k} \right\} \right\}^\alpha = \sum_{i=0}^{\infty} (-1)^i \binom{\alpha}{i} \left\{ 1 - \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k} \right\}^i. \quad (4.2)$$

Assim, substituindo-se (4.1) e (4.2) em (3.1), tem-se

$$F(x) = \frac{\sum_{i=0}^{\infty} a_i \left\{ 1 - \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k} \right\}^i}{\sum_{i=0}^{\infty} b_i \left\{ 1 - \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k} \right\}^i},$$

em que

$$a_i = a_i(\alpha) = \sum_{j=i}^{\infty} (-1)^{i+j} \binom{\alpha}{j} \binom{j}{i} \quad \text{e} \quad b_i = (-1)^i \binom{\alpha}{i} + a_i.$$

A proporção das duas séries de potências pode ser expressa como

$$F(x) = \sum_{i=0}^{\infty} d_i \left\{ 1 - \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k} \right\}^i, \quad (4.3)$$

em que os coeficientes d_i 's (para $i \geq 0$) determinados pela equação

$$d_i = b_0^{-1} \left(a_i - b_0^{-1} \sum_{r=1}^i b_r \cdot d_{i-r} \right).$$

Utilizando-se a equação (4.3), a fdp de X pode ser escrita como

$$f(x) = \sum_{i=0}^{\infty} d_{i+1} h_{i+1}(x), \quad (4.4)$$

em que,

$$h_{i+1}(x) = (i+1) c k s^{-c} x^{c-1} \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k-1} \left\{ 1 - \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k} \right\}^i,$$

é a função densidade da distribuição Burr XII exponenciada (EBXII) com parâmetro de potência $(i+1)$.

Além disso, utilizando-se novamente a expansão binomial em $\left\{ 1 - \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k} \right\}^i$, a fdp (4.4) pode ser reescrita como segue

$$\begin{aligned} f(x) &= \sum_{i=0}^{\infty} d_{i+1} (i+1) c k x^{-1} \left(\frac{x}{s} \right)^c \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k-1} \sum_{p=0}^i (-1)^p \binom{i}{p} \left[1 + \left(\frac{x}{s} \right)^c \right]^{-kp} \\ &= \sum_{i=0}^{\infty} \sum_{p=0}^i d_{i+1} (i+1) (-1)^p c k x^{-1} \left(\frac{x}{s} \right)^c \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k-1} \left[1 + \left(\frac{x}{s} \right)^c \right]^{-kp} \\ &= \sum_{p=0}^{\infty} \sum_{i=p}^{\infty} d_{i+1} (i+1) (-1)^p \frac{(1+p)}{(1+p)} c k x^{-1} \left(\frac{x}{s} \right)^c \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k(1+p)-1} \\ &= \sum_{p=0}^{\infty} \omega_p g(x; s, k(p+1), c), \end{aligned} \quad (4.5)$$

no qual, $\omega_p = \sum_{i=p}^{\infty} d_{i+1} (i+1) \binom{i}{p} \frac{(-1)^p}{p+1}$ e $g(x; s, k(p+1), c)$ representa a fdp da distribuição BXII. A equação (4.5) revela que a função densidade de X é uma combinação linear das densidades da distribuição BXII. Portanto, algumas propriedades estruturais da distribuição OLLB-XII pode ser derivada de propriedades bem estabelecidas da distribuição EBXII.

4.1 PROPRIEDADES MATEMÁTICAS

Os momentos são usados para calcular tendência, dispersão, assimetria e curtose, entre outras características da distribuição estudada. Por esta razão, nesta seção, os momentos ordinários, incompletos e a função geradora de momentos são apresentados.

4.1.1 Momentos Ordinários

O n -ésimo momento ordinário da variável aleatória X pode ser calculada como $\mu'_n = E(X^n) = \int_0^\infty x^n f(x) dx$. Utilizando-se a expansão (4.5), é possível calcular os momentos que são apresentados da seguinte forma,

$$\mu'_n = \sum_{p=0}^{\infty} \omega_p k(p+1) \int_0^\infty x^{n-1} \left(\frac{x}{s}\right)^c \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k(p+1)-1} dx, \quad (4.6)$$

em que $ck > n$. Resolvendo-se a integral em (4.6), o n -ésimo momento da distribuição OLLB-XII é escrito por

$$\mu'_n = \sum_{p=0}^{\infty} \omega_p s^n k(p+1) B(k(p+1) - nc^{-1}, nc^{-1} + 1). \quad (4.7)$$

Pode-se notar que os cumulantes (κ_n) de X podem ser calculados a partir de (4.7) usando-se a seguinte expressão

$$\kappa_n = \mu'_n - \sum_{r=1}^{n-1} \binom{n-1}{r-1} \kappa_r \mu'_{n-r},$$

em que, $\kappa_1 = \mu'_1$, $\kappa_2 = \mu'_2 - (\mu'_1)^2$, $\kappa_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2(\mu'_1)^3$, $\kappa_4 = \mu'_4 - 4\mu'_3\mu'_1 - 3(\mu'_2)^2 + 12\mu'_2(\mu'_1)^2 - 6(\mu'_1)^4$, etc.

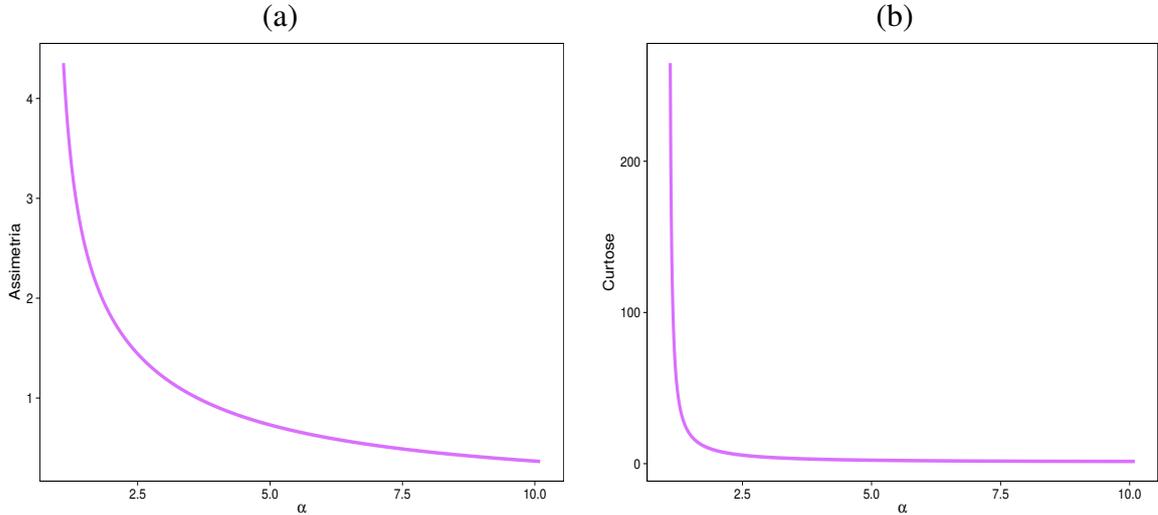
A assimetria $\gamma_1 = \kappa_3/\kappa_2^{3/2}$ e curtose $\gamma_2 = \kappa_4/\kappa_2^2$ são medidas que podem ser calculadas a partir do terceiro e quarto cumulantes padronizados usando essas relações. Gráficos das medidas de assimetria e curtose para algumas escolhas dos parâmetros c , k e s em função do parâmetro α são exibidos na Figura 4.1 (a) e (b), respectivamente.

4.1.2 Momentos Incompletos

O n -ésimo momento incompleto de X pode ser calculado como $T_n(t) = \int_0^t x^n f(x) dx$, novamente usando a expansão (4.5), o momento incompleto pode ser escrito como,

$$\begin{aligned} T_n(t) &= \sum_{p=0}^{\infty} \omega_p c k(p+1) \int_0^t x^{n-1} \left(\frac{x}{s}\right)^c \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k(p+1)-1} dx \\ &= \sum_{p=0}^{\infty} \omega_p g(x; s, k(p+1), c), \end{aligned}$$

Figura 4.1: Medidas de assimetria e curtose para a distribuição OLLBXII, com parâmetros $s = 3.5$, $k = 2.5$, $c = 1.5$ e α uma sequência entre 1.1 e 10.1.



Fonte: Autora.

observa-se que, o momento incompleto da distribuição BXII [28, 65] é dado por $T_{n:BXII}(t) = s^n k B_{s^c/s^c + t^c}(k - n c^{-1}, n c^{-1} + 1)$, desta forma os momentos incompletos é dado por

$$T_n(t) = \sum_{p=0}^{\infty} \omega_p s^n k (p + 1) B_{s^c/s^c + t^c}(k (p + 1) - n c^{-1}, n c^{-1} + 1). \quad (4.8)$$

4.1.3 Função geradora de momentos

A função geradora de momentos (fgm) pode ser calculada por

$$M(t) = E \left[\sum_{i=0}^{\infty} \frac{(tX)^n}{n!} \right] = \sum_{i=0}^{\infty} \frac{(t)^n}{n!} E[X^n],$$

e verificando-se que a expansão obtida da fdp na seção anterior $E(X^n)$ pode ser escrita como $E(X^n) = \sum_{i=0}^{\infty} \omega_i E(X_{s,k(p+1),c}^n)$. Portanto, fgm para a OLLBXII pode ser escrita como

$$M(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \sum_{p=0}^{\infty} \omega_p E(X_{s,k(p+1),c}^n) = \sum_{p=0}^{\infty} \omega_p M_{p+1}(t),$$

em que $M_{p+1}(t)$ é fgm da distribuição BXII($s, k(p + 1), c$). Assim, considerando-se a representação da fgm dada por [51, 50] para BXII e a integral definida em [58], para m e $k(p + 1)$ inteiros positivos, a fgm da distribuição OLLBXII é dada por (4.9)

$$M(t) = m \sum_{n=0}^{\infty} \omega_n I' \left(-st, \frac{m}{k(p+1)} - 1, \frac{m}{k(p+1)}, -k(p+1) - 1 \right), \quad t < 0, \quad (4.9)$$

sendo $c = m/k(p+1)$ e assumindo m inteiros positivos. A função especial I pode ser resolvida usando a integral definida em [58] como,

$$\begin{aligned} I &= I \left(-st, \frac{m}{k(p+1)} - 1, \frac{m}{k(p+1)}, -k(p+1) - 1 \right) = \\ &= \int_0^{\infty} \exp(stx) x^{\frac{m}{k(p+1)} - 1} \left[1 + x^{\frac{m}{k(p+1)}} \right]^{-k(p+1) - 1} dx \\ &= \frac{[k(p+1)]^{k(p+1)+1} m^{\frac{m}{k(p+1)} - \frac{1}{2}}}{(2\pi)^{(m-1)/2} \Gamma(k(p+1) + 1) (-st)^{m/k(p+1)}} \\ &\times G_{k(p+1)+m \quad k(p+1)}^{k(p+1) \quad k(p+1)+m} \left(\begin{matrix} \Delta(m, 1 - \frac{m}{k(p+1)}), \Delta(k(p+1), -k(p+1)) \\ \Delta(k(p+1), 0) \end{matrix} \middle| \frac{m^m}{(-st)^m} \right), \end{aligned}$$

em que $G_{k+m \quad k}^{k \quad k+m} (:|\cdot)$ é a função G-Meijer para $\Delta(k, a) = a/k, (a+1)/k, \dots, (a+k)/k$, [50, 58].

5 ESTIMAÇÃO E INFERÊNCIA

5.1 CONJUNTO DE DADOS NÃO CENSURADOS

Observa-se que para os dados não censurados, considera-se apenas o conjunto F dos tempos de falha. Assim,

$$\begin{aligned} \ell(c, k, s, \alpha) = & r \log \alpha \sum_{i \in F} \log \left[\frac{ckx_i^{c-1}}{s^c z_i^{k+1}} \right] + \sum_{i \in F} \log \left[\frac{z_i^k - 1}{z_i^{2k}} \right] \\ & - 2 \sum_{i \in F} \log \left[\frac{(z_i^k - 1)^\alpha + 1}{z_i^{k\alpha}} \right], \end{aligned}$$

sendo, $z_i = z_i(x_i) = 1 + \left(\frac{x_i}{s}\right)^c$.

5.2 CONJUNTO DE DADOS CENSURADOS

Várias abordagens para estimação de parâmetros são propostas na literatura, sendo que o método de máxima verossimilhança é uma alternativa frequentemente utilizada, pois o estimador de máxima verossimilhança (EMV) tem importantes propriedades que permitem calcular o intervalo de confiança, além de realizar diferentes testes estatístico. A teoria de grandes amostras para essas estimativas fornece aproximações simples que funcionam bem em amostras finitas. Se X segue a distribuição OLLBXII com vetor de parâmetros $\gamma = (c, k, s, \alpha)^T$, o logaritmo da função de verossimilhança considerando γ uma única observação x de X é dada por (5.1)

$$\begin{aligned} \ell(c, k, s, \alpha) = & r [\log(\alpha) + \log(c) + \log(k) - c \log(s)] \\ & + (c - 1) \sum_{i \in F} \log(x_i) + \left(\alpha + \frac{1}{k}\right) \sum_{i \in F} \log(1 - z_i) + (\alpha - 1) \sum_{i \in F} \log(z_i) \\ & - 2 \sum_{i \in F} \log [z_i^\alpha + (1 - z_i)^\alpha] + \alpha \sum_{i \in C} \log(1 - z_i) \\ & + \sum_{i \in C} \log [z_i^\alpha + (1 - z_i)^\alpha], \end{aligned} \quad (5.1)$$

sendo, $z_i = z(x_i) = 1 - [1 + (x_i/s)^c]^{-k}$, r é o número de falhas, C e F são grupos de observações censuradas e não censuradas respectivamente.

Os componentes do vetor score $\mathbf{U} = \left(\frac{\partial \ell}{\partial c}, \frac{\partial \ell}{\partial k}, \frac{\partial \ell}{\partial s}, \frac{\partial \ell}{\partial \alpha}\right)^T$ são dados por,

$$\begin{aligned} \frac{\partial \ell}{\partial c} &= \frac{r}{s} - r \log(c) + \sum_{i \in F} \log(x_i) + \frac{\left(\frac{x_i}{s}\right)^c \log\left(\frac{x_i}{s}\right)^c}{(1-z_i)^{(-1/k)}} \left[\frac{k(\alpha+1)(1-z_i)}{z_i} - (\alpha k + 1) \right] \\ &\quad - 2k\alpha \sum_{i \in F} \frac{\left(\frac{x_i}{s}\right)^c \log\left(\frac{x_i}{s}\right)^c}{(1-z_i)^{(-1/k)}} \left[\left[\frac{[(1-z_i)^{-1} - 1]^{(\alpha-1)} (1-z_i)^{-1}}{[(1-z_i)^{-1} - 1]^\alpha + 1} \right] - 1 \right] \\ &\quad + k\alpha \sum_{i \in C} \frac{\left(\frac{x_i}{s}\right)^c \log\left(\frac{x_i}{s}\right)^c}{(1-z_i)^{(-1/k)}} \left[\left[\frac{[(1-z_i)^{-1} - 1]^{(\alpha-1)} (1-z_i)^{-1}}{[(1-z_i)^{-1} - 1]^\alpha + 1} \right] - 1 \right], \end{aligned}$$

$$\begin{aligned} \frac{\partial \ell}{\partial k} &= \frac{r}{k} - \alpha \sum_{i \in F} \log(1-z_i)^{(-1/k)} + (\alpha-1) \sum_{i \in F} \frac{\log(1-z_i)^{(-1/k)}}{z_i(1-z_i)^{-1}} \\ &\quad - 2\alpha \sum_{i \in F} \log(1-z_i)^{(-1/k)} \left[\left[\frac{[(1-z_i)^{-1} - 1]^{(\alpha-1)} (1-z_i)^{-1}}{[(1-z_i)^{-1} - 1]^\alpha + 1} \right] - 1 \right] \\ &\quad + \alpha \sum_{i \in C} \log(1-z_i)^{(-1/k)} \left[\left[\left[\frac{[(1-z_i)^{-1} - 1]^{(\alpha-1)} (1-z_i)^{-1}}{[(1-z_i)^{-1} - 1]^\alpha + 1} \right] - 1 \right] - 1 \right], \end{aligned}$$

$$\begin{aligned} \frac{\partial \ell}{\partial s} &= -\frac{rc}{s} + \frac{kc}{s} \sum_{i \in F} \left(\frac{x_i}{s}\right)^c (1-z_i)^{(-1/k)} \left[\left(\alpha + \frac{1}{k}\right) - \frac{(\alpha+1)}{z_i(1-z_i)^{-1}} \right] \\ &\quad + \frac{2kc\alpha}{s} \sum_{i \in F} \left(\frac{x_i}{s}\right)^c (1-z_i)^{(1/k)} \left[\frac{(1-z_i)^{-1} [(1-z_i)^{-1} - 1]^{(\alpha-1)}}{[(1-z_i)^{-1} - 1]^\alpha + 1} - 1 \right] \\ &\quad + \frac{kc\alpha}{s} \sum_{i \in C} \left(\frac{x_i}{s}\right)^c (1-z_i)^{(1/k)} \left[\frac{(1-z_i)^{-1} [(1-z_i)^{-1} - 1]^{(\alpha-1)}}{[(1-z_i)^{-1} - 1]^\alpha + 1} - 1 \right], \end{aligned}$$

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha} &= \frac{r}{\alpha} + \sum_{i \in F} \log\left(\frac{z_i}{(1-z_i)^{-1}}\right) \\ &\quad - 2 \sum_{i \in F} \log(1-z_i) + \frac{[(1-z_i)^{-1} - 1]^\alpha \log[(1-z_i)^{-1} - 1]}{[(1-z_i)^{-1} - 1]^\alpha + 1} \\ &\quad + 2 \sum_{i \in C} \log(1-z_i) + \frac{1}{2} \frac{[(1-z_i)^{-1} - 1]^\alpha \log[(1-z_i)^{-1} - 1]}{[(1-z_i)^{-1} - 1]^\alpha + 1}. \end{aligned}$$

O EMV $\hat{\gamma}$ da γ é obtida por meio da solução da equação não linear $U_c = 0$, $U_s = 0$, $U_k = 0$ e $U_\alpha = 0$. Pode-se notar que estas equações não podem ser resolvidos analiticamente e um software estatístico pode ser usado para resolve-las numericamente. Estimaco de mxima verossimilhana para o vetor parmetro γ pode ser implementado numericamente atravs da maximizao do logaritmo da funo de verossimilhana $\ell(\gamma) = \log L(\gamma)$ dada por (5.1) utilizando-se o software estatístico R.

Para a estimação do intervalo de confiança e do teste de hipóteses dos parâmetros em γ , requiere-se a matriz de informação da unidade observada $\mathbf{J} = \mathbf{J}(\gamma) = \{j_{r,s}\}$, em que os elementos são $j_{r,s}$ (para $r, s = c, k, s, \alpha$). Sob condições de regularidade o valor normal multivariado aproximado $N_4(0, n^{-1}\mathbf{J}(\hat{\gamma})^{-1})$ pode ser usado para construir o intervalo de confiança aproximado para os parâmetros do modelo.

O teste de razão de verossimilhança (RV) é utilizado para comparar a nova distribuição com alguns modelos especiais. Por exemplo, pode-se usar a estatística RV para checar se o ajuste usando a distribuição OLLBXII é estatisticamente "superior" para ajustar a distribuição BXII para um conjunto de dados. Considerando a partição $\gamma = (\gamma_1^T, \gamma_2^T)^T$, o teste de hipótese do tipo $H_0 : \gamma_1 = \gamma_1^{(0)}$ versus $H_1 : \gamma_1 \neq \gamma_1^{(0)}$ pode ser realizado usando a estatística RV $w = 2\{\ell(\hat{\gamma}) - \ell(\tilde{\gamma})\}$, em que $\hat{\gamma}$ e $\tilde{\gamma}$ são estimados de γ sob H_1 e H_0 , respectivamente. Sob a hipótese nula H_0 , $w \xrightarrow{d} \chi_q^2$, em que q é a dimensão do vetor parâmetro de interesse γ_1 . O teste RV rejeita H_0 se $w > \xi_\gamma$, sendo que ξ_γ denota o valor superior 100 γ % da distribuição χ_q^2 .

6 ESTUDOS DE SIMULAÇÕES

Quando novas distribuições são propostas, é importante estudar o comportamento dos estimadores dos parâmetros, e uma alternativa passa a ser o estudo de valores simulados da distribuição. Desta forma, um dos métodos utilizados é o método de Monte Carlo, estudado por [1, 18, 28, 33, 51, 60].

Neste capítulo realiza-se um estudo de simulação de Monte Carlo para avaliar o comportamento da amostra finita das estimativas de s , k , c e α . Os resultados obtidos a partir de 3.000 replicações de Monte Carlo das simulações são calculados usando o software estatístico R. Em cada replicação, uma amostra aleatória de tamanho n é retirada da distribuição $OLLBXII(s, k, c, \alpha)$ e os parâmetros são estimados pelo método da máxima verossimilhança. A variável aleatória X é gerada usando o método de inversão. As estimativas das médias dos quatro parâmetros do modelo, a raiz quadrada dos erros quadráticos médios correspondentes (REQMs) e os erros padrões (EP) para os tamanhos de amostra $n = 30, 100, 200$ e 500 estão listados nas Tabelas 6.1, 6.2, 6.3 e 6.4.

Tabela 6.1: Resultado da simulação de Monte Carlo: Tamanho da amostra, média estimada dos parâmetros, EP e REQMs dos parâmetros $s = 4, k = 12, c = 4, 2$ e $\alpha = 5, 5$.

	$n = 30$			$n = 100$			$n = 200$			$n = 500$		
	Média	EP	REQM	Média	EP	REQM	Média	EP	REQM	Média	EP	REQM
s	3,33	0,39	2,26	3,51	0,17	1,73	3,71	0,09	1,29	3,91	0,04	0,85
k	10,82	1,47	8,12	10,48	0,57	5,94	11,04	0,39	5,58	11,52	0,22	4,94
c	14,36	4,70	27,66	6,92	0,83	8,78	5,34	0,24	3,55	4,57	0,09	2,10
α	8,26	1,60	9,18	7,71	0,64	6,81	7,10	0,35	5,19	6,66	0,17	4,00

Tabela 6.2: Resultado da simulação de Monte Carlo: Tamanho da amostra, média estimada dos parâmetros, EP e REQMs dos parâmetros $s = 6, k = 1, c = 1$ e $\alpha = 0, 5$.

	$n = 30$			$n = 100$			$n = 200$			$n = 500$		
	Média	EP	REQM	Média	EP	REQM	Média	EP	REQM	Média	EP	REQM
s	13,62	3,68	21,46	6,30	1,18	12,3	8,04	0,54	7,83	6,97	0,18	4,17
k	1,22	0,17	0,94	1,00	0,04	0,42	1,02	0,02	0,27	1,01	0,01	0,17
c	1,98	0,42	2,53	1,00	0,10	1,10	1,23	0,06	0,85	1,10	0,03	0,61
α	0,95	0,20	1,21	0,50	0,08	0,84	0,71	0,05	0,69	0,66	0,02	0,56

Tabela 6.3: Resultado da simulação de Monte Carlo: Tamanho da amostra, média estimada dos parâmetros, EP e REQMs dos parâmetros $s = 3$, $k = 10$, $c = 0, 2$ e $\alpha = 15$.

	$n = 30$			$n = 100$			$n = 200$			$n = 500$		
	Média	EP	REQM	Média	EP	REQM	Média	EP	REQM	Média	EP	REQM
s	2,30	0,19	1,24	2,19	0,11	1,40	2,26	0,07	1,28	2,26	0,04	1,19
k	11,31	0,40	2,55	12,17	0,35	4,10	12,39	0,27	4,53	12,67	0,21	5,43
c	0,22	0,01	0,04	0,23	0,00	0,06	0,23	0,00	0,05	0,23	0,00	0,05
α	14,40	0,49	2,77	13,67	0,25	2,80	13,66	0,15	2,56	13,66	0,09	2,45

Tabela 6.4: Resultado da simulação de Monte Carlo: Tamanho da amostra, média estimada dos parâmetros, EP e REQMs dos parâmetros $s = 4$, $k = 12$, $c = 4, 2$ e $\alpha = 5, 5$.

	$n = 30$			$n = 100$			$n = 200$			$n = 500$		
	Média	EP	REQM	Média	EP	REQM	Média	EP	REQM	Média	EP	REQM
s	3,33	0,39	2,26	3,51	0,17	1,73	3,71	0,09	1,29	3,91	0,04	0,85
k	10,82	1,47	8,12	10,48	0,57	5,94	11,04	0,39	5,58	11,52	0,22	4,94
c	14,36	4,70	27,66	6,92	0,83	8,78	5,34	0,24	3,55	4,57	0,09	2,10
α	8,26	1,60	9,18	7,71	0,64	6,81	7,10	0,35	5,19	6,66	0,17	4,00

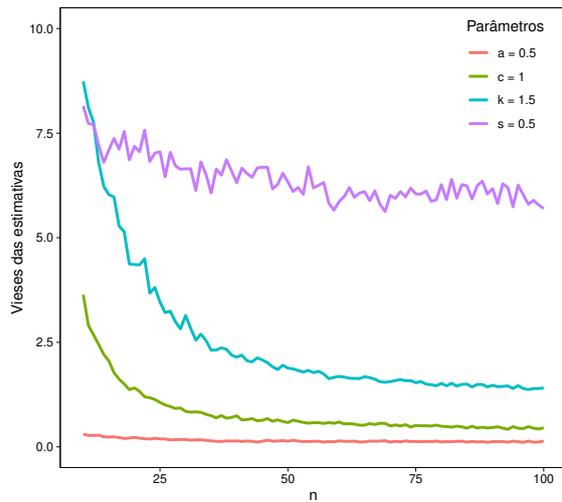
Para calcular os vieses das estimativas dos parâmetros, utilizou-se a expressão

$$\text{vieses}_h(n) = \frac{1}{3000} \sum_{i=1}^n (\hat{h}_i - h),$$

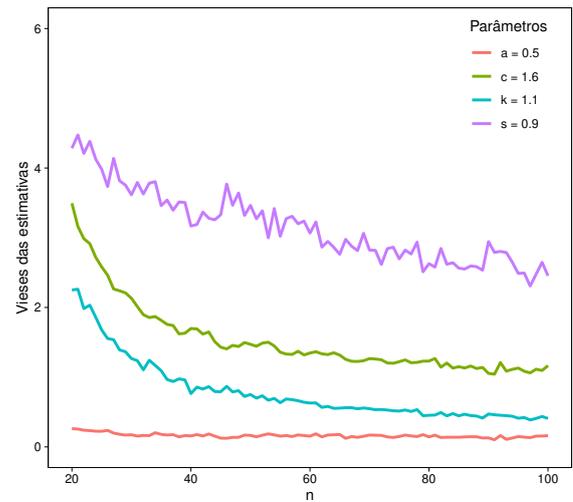
em que $h = s, k, c, \alpha$. A Figura 6.1 apresenta o comportamento dos vieses das estimativas para diferentes valores dos parâmetros da distribuição OLLBXII, conforme n aumenta, observa-se que todos os vieses decrescem, em grandezas diferentes.

Nota-se que, em todos os casos, os vieses e os EPs das estimativas de s , k , c e α tendem a decair para zero quando o tamanho da amostra aumenta, como o esperado. Pesquisas futuras devem ser conduzidas para obter correções de viés para esses estimadores.

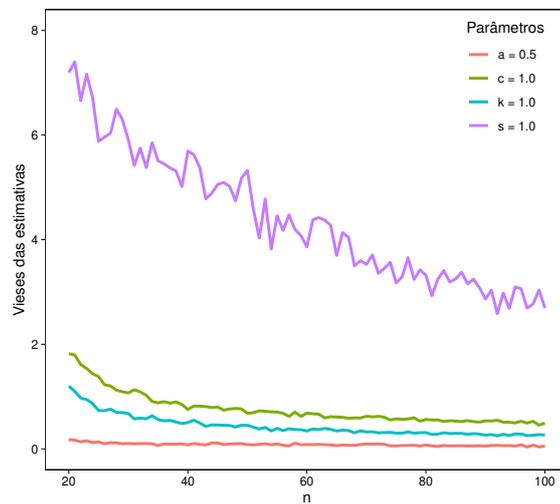
Figura 6.1: Vieses dos EMVs para os parâmetros s, k, c, α versus diferentes tamanhos da amostra.



(a) Vieses para $s = 0,5, k = 1,5, c = 1$ e $\alpha = 0,5$



(b) Vieses para $s = 0,9, k = 1,1, c = 1,6$ e $\alpha = 0,5$



(c) Vieses para $s = 1,0, k = 1,0, c = 1,0$ e $\alpha = 0,5$

Fonte: Autora.

7 MODELO DE REGRESSÃO LOLLBXII

O modelo de regressão como descrito na Seção 2.6, pode ser obtido para $X \sim \text{OLLBXII}(x, c, k, s, \alpha)$, considerando-se a transformação $Y = \log(X)$. Assim, como o método do Jacobiano é dado por,

$$f_Y(y) = \left| \frac{h^{-1}(y)}{y} \right| f_X(h^{-1}(y)),$$

então, $X = e^y$ e $\left| \frac{h^{-1}(y)}{y} \right| = (e^y)' = e^y$. Logo, tem-se que

$$f_Y(y) = \frac{e^y \alpha c k s^{-c} \left(1 + \frac{e^y c}{s^c}\right)^{-(k+1)} e^{y(c-1)} \left\{ \left[1 - \left(1 + \frac{e^y c}{s^c}\right)^{-k}\right] \left(1 + \frac{e^y c}{s^c}\right)^{-k} \right\}^{\alpha-1}}{\left\{ \left[1 - \left(1 + \frac{e^y c}{s^c}\right)^{-k}\right]^\alpha + \left(1 + \frac{e^y c}{s^c}\right)^{-k\alpha} \right\}^2}.$$

Realizando-se as reparametrizações $c = 1/\sigma$ e $s = \exp(\mu)$, tem-se

$$f_Y(y) = \frac{k \alpha}{\sigma} \exp\left(\frac{y - \mu}{\sigma}\right) \times \frac{\left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-(k+1)} \left\{ \left[1 - \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-k}\right] \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-k} \right\}^{\alpha-1}}{\left\{ \left[1 - \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-k}\right]^\alpha + \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-k\alpha} \right\}^2} \quad (7.1)$$

Logo se $X \sim \text{OLLBXII}(x; s, k, c, \alpha)$ então $Y = \log(X) \sim \text{LOLLBXII}(y; k, \mu, \sigma, \alpha)$. A Figura 7.1 apresenta algumas formas para diferentes parâmetros da distribuição LOLLBXII, mostrando sua flexibilidade. Observa-se ainda que para $\alpha = 1$ tem-se a distribuição $Y \sim \text{LBXII}(y; k, \mu, \sigma)$ e considerando-se $\alpha = k = 1$, tem-se $Y \sim \text{log-logística}(y; \mu, \sigma)$.

A fda da distribuição LOLLBXII descrita em (2.9) é dada por

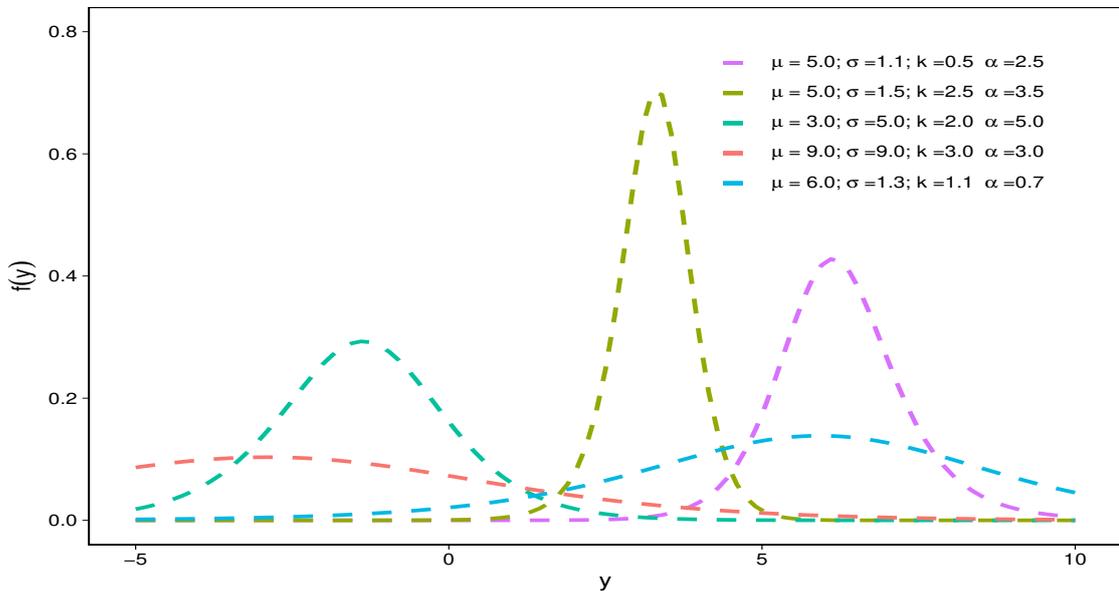
$$F(y) = \frac{\left[\left(1 - \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-k}\right) \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-k} \right]^\alpha}{\left(1 - \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-k}\right)^\alpha + \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-k\alpha}}.$$

Agora considere Z um variável aleatória padronizada dada por $Z = (Y - \mu)/\sigma$, a fdp (7.1) pode ser reescrita assim,

$$f_Z(z) = \frac{k \alpha}{\sigma} \exp(z) \frac{\left[1 + \exp(z)\right]^{-(k+1)} \left\{ \left[1 - \left[1 + \exp(z)\right]^{-k}\right] \left[1 + \exp(z)\right]^{-k} \right\}^{\alpha-1}}{\left\{ \left[1 - \left[1 + \exp(z)\right]^{-k}\right]^\alpha + \left[1 + \exp(z)\right]^{-k\alpha} \right\}^2},$$

em que $-\infty < z < \infty$. Com o propósito de construir um modelo de regressão log-linear, as

Figura 7.1: Função densidade de probabilidade para o modelo de regressão LOLLBXII.



Fonte: Autora.

observações y_i da variável resposta Y podem ser escritas da seguinte forma

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \sigma z_i = \boldsymbol{\mu}(x) + \sigma z_i,$$

em que $i = 1, \dots, n$, $\mathbf{x} = (x_1, \dots, x_p)^T$ é o vetor de covariáveis, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ é o vetor de parâmetros, $\sigma > 0$, $\alpha > 0$ são os parâmetros desconhecidos, e por fim $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ é o vetor de parâmetros de locação. Assim, a função de sobrevivência para o modelo de regressão de locação pode ser escrita como

$$S(y_i|\mathbf{x}) = \frac{\left[1 + \exp\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}\right)\right]^{-k\alpha}}{\left(1 - \left[1 + \exp\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}\right)\right]^{-k}\right)^\alpha + \left[1 + \exp\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}\right)\right]^{-k\alpha}}.$$

Ressalta-se que, o cálculo da máxima verossimilhança para dados com e sem censuras é similar a metodologia utilizada para calcular a máxima verossimilhança para a distribuição OLLBXII.

8 APLICAÇÕES

Neste Capítulo será utilizado três conjuntos de dados que demonstram a flexibilidade e a viabilidade aos quais o modelo OLLBXII pode ser aplicado, e um conjunto de dados em que estuda o comportamento do modelo de regressão LOLLBXII. Todas as análises foram desenvolvidas no software estatístico R [66]. Construiu-se um pacote no R disponível em (<https://github.com/MaraTorres>). Para tanto, a nova distribuição será comparada com diferentes distribuições propostas na literatura.

- **Frechét (Fr):** A distribuição Frechét também estudada em aplicações de análise de sobrevivência, [46] tem sua fdp e fda dadas por

$$f(x) = \left(\frac{\lambda}{\sigma}\right) \left(\frac{\sigma}{x}\right)^{(\lambda+1)} \exp\left[-\left(\frac{\sigma}{x}\right)^\lambda\right] \quad (8.1)$$

e

$$F(x) = \exp\left[-\left(\frac{\sigma}{x}\right)^\lambda\right], \quad (8.2)$$

em que $x > 0$, $\lambda > 0$ e $\sigma > 0$.

- **Log-logística (LL):** A distribuição Log-logística muito utilizada para modelar dados com função taxa de falha monótona, tem fdp e fda dadas por

$$f(x) = \left(\frac{\gamma}{\lambda}\right) \left(\frac{x}{\lambda}\right)^{\gamma-1} \left[1 + \left(\frac{x}{\lambda}\right)^\gamma\right]^{-2} \quad (8.3)$$

e

$$F(x) = \left[1 + \left(\frac{x}{\lambda}\right)^{-\gamma}\right]^{-1}, \quad (8.4)$$

sendo $x > 0$, $\lambda > 0$ e $\gamma > 0$.

- **Weibull (W):** A distribuição Weibull foi amplamente estudada por diversos autores. Logo, sua fdp e fda são dadas por

$$f(x) = \frac{\gamma}{\lambda^\gamma} x^{\gamma-1} \exp\left[-\left(\frac{x}{\lambda}\right)^\gamma\right] \quad (8.5)$$

e

$$F(x) = 1 - \exp\left[-\left(\frac{x}{\lambda}\right)^\gamma\right], \quad (8.6)$$

em que $\gamma > 0$ e $\lambda > 0$ são os parâmetros de forma e escala, respectivamente.

- **Odd Log Logística Frechét (OLLFr):** A distribuição OLLFr é uma extensão da distribuição Frechét [18], e sua fdp e fda são dadas por

$$f(x) = \frac{\gamma}{\lambda} \left(\frac{x}{\lambda}\right)^{\gamma-1} \left[1 + \left(\frac{x}{\lambda}\right)^\gamma\right]^{-2} \quad (8.7)$$

e

$$F(x) = \frac{[1 + (\frac{x}{\lambda})^{-\gamma}]^{-\alpha}}{[1 + (\frac{x}{\lambda})^{-\gamma}]^{-\alpha} + \{1 - [1 + (\frac{x}{\lambda})^{-\gamma}]^{-1}\}^\alpha}, \quad (8.8)$$

em que $\alpha > 0$ é o parâmetro adicional de forma.

- **Odd Log-logística Log-logística (OLLLL):** A distribuição OLLLL também estudada em aplicações de análise de sobrevivência [18], tem sua fdp e fda dadas por

$$f(x) = \frac{\alpha \{(\frac{\gamma}{\lambda})(\frac{x}{\lambda})^{\gamma-1} [1 + (\frac{x}{\lambda})^\gamma]^{-2}\} \{[1 + (\frac{x}{\lambda})^{-\gamma}]^{-1} \{1 - [1 + (\frac{x}{\lambda})^{-\gamma}]^{-1}\}\}^{\alpha-1}}{\{[1 + (\frac{x}{\lambda})^{-\gamma}]^{-\alpha} + \{1 - [1 + (\frac{x}{\lambda})^{-\gamma}]^{-1}\}^\alpha\}^2} \quad (8.9)$$

e

$$F(x) = \frac{[1 + (\frac{x}{\lambda})^{-\gamma}]^{-\alpha}}{[1 + (\frac{x}{\lambda})^{-\gamma}]^{-\alpha} + \{1 - [1 + (\frac{x}{\lambda})^{-\gamma}]^{-1}\}^\alpha}, \quad (8.10)$$

em que $\alpha > 0$ é o parâmetro adicional de forma.

- **Odd Log-logística Weibull (OLLW):** A distribuição OLLW é uma extensão da distribuição W, também foi estudada em aplicações de análise de sobrevivência [15, 18], tem sua fdp e fda dadas por

$$f(x) = \frac{\alpha \gamma x^{\gamma-1} [\exp(-(\frac{x}{\lambda})^\gamma)]^\alpha [1 - \exp(-(\frac{x}{\lambda})^\gamma)]^{\alpha-1}}{\lambda^\gamma \{[1 - \exp(-(\frac{x}{\lambda})^\gamma)]^\alpha + [\exp(-(\frac{x}{\lambda})^\gamma)]^\alpha\}^2} \quad (8.11)$$

e

$$F(x) = \frac{\{1 - \exp[-(\frac{x}{\lambda})^\gamma]\}^\alpha}{\{1 - \exp[-(\frac{x}{\lambda})^\gamma]\}^\alpha + \exp[-(\frac{x}{\lambda})^\gamma]^\alpha} \quad (8.12)$$

em que $\alpha > 0$ é o parâmetro adicional de forma.

- **Beta Burr XII (BBXII):** A distribuição Beta Burr XII é uma extensão da distribuição BXII [51], tem sua fdp e fda são dadas por,

$$f(x) = \frac{ckx^{c-1}}{s^c B(a, b)} [1 + (x/s)^c]^{-(kb+1)} \left\{1 - [1 + (x/s)^c]^{-k}\right\}^{a-1} \quad (8.13)$$

e

$$F(x) = I_{[1 - 1 + (x/s)^c]^{-k}}(a, b), \quad (8.14)$$

em que $s, k, c, a, b > 0$ são os parâmetros da distribuição, $I_y(a, b) = B_y(a, b)/B(a, b)$ é taxa da função beta incompleta, e $B_y(a, b) = \int_0^y \omega^{a-1} (1 - \omega)^{b-1} d\omega$ é a função beta incompleta.

- **Kumaraswamy Burr XII (KwBXII):** A distribuição KwBXII também é uma extensão para a distribuição BXII [50], tem sua fdp e fda dadas por

$$f(x) = abck s^{-c} x^{c-1} z^{-k-1} (1 - z^{-k})^{a-1} (1 - (1 - z^{-k})^a)^{b-1} \quad (8.15)$$

e

$$F(x) = 1 - \left[1 - \left\{ 1 - \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k} \right\}^a \right]^b, \quad (8.16)$$

sendo $z = z(x) = 1 + \left(\frac{x}{s} \right)^c$ e $s, k, c, a, b > 0$ são os parâmetros da distribuição.

Compara-se também com a distribuição BXII descrita no Capítulo 2 na Seção 2.5, assim para cada modelo estima-se os parâmetros usando o método da máxima verossimilhança.

8.1 CONJUNTO DE DADOS 1: ATUARIA

Considerando-se conjuntos de dados sem censuras, ou seja, será estudado um conjunto de dados em que o evento de interesse ocorreu em toda a amostra.

O primeiro conjunto de dados foi estudado e analisado por [8, 17]. Observa-se que a ciência atuarial é um tema de estudo muito abrangente [35], e que objetiva estudar o risco de um evento de interesse ocorrer, sendo muito aplicada ao determinar prazos de seguros, planos de saúde, previdência complementar, dentre outros, pois possibilita o cálculo de estimativas financeiras de longo e curto prazo. Em particular, o conjunto de dados estudado refere-se ao comportamento distributivo da mortalidade de aposentados por invalidez, mais especificamente correspondem a 280 vidas (em anos) de mulheres aposentadas com deficiência temporária, incorporadas ao sistema público de seguros mexicano e que faleceram em 2004. Assim, o tempo de sobrevivência X corresponde ao tempo dessas 280 mulheres serem aposentadas por invalidez até falecerem em 2004.

A Tabela 8.1 fornece um resumo descritivo da variável tempo e sugere distribuições assimétricas positivamente, em que a média é de 47,99 anos, além disso, o desvio padrão do tempo de vida é de 10,37 anos.

Tabela 8.1: Principais medidas descritivas da variável tempo até a morte de mulheres aposentadas.

Mínimo	1° Quartil	Mediana	Média	3° Quartil	Máximo
22,00	40,00	49,00	47,99	56,00	86,00

Fonte: Autora.

Para estimar os parâmetros dos modelos, considera-se como valores iniciais iguais a um para o procedimento iterativo numérico. A Tabela 8.2 lista as EMVs dos parâmetros e seus respectivos erros-padrão (entre parênteses), além dos valores das estatísticas baseadas nos

critérios AIC, CAIC e BIC. Os resultados indicam que o modelo OLLBXII possui os menores valores dessas estatísticas entre todos os modelos ajustados.

Tabela 8.2: EMVs, erro padrão em parenteses e as estatísticas AIC, CAIC e BIC para a variável tempo até a morte de mulheres aposentadas.

Modelos	Estimativas					Estatísticas		
						AIC	CAIC	BIC
Fr(λ, σ)	3,96 (0,16)	41,52 (0,67)				2213,54	2213,69	2220,81
LL(λ, γ)	47,61 (0,65)	7,70 (0,38)				2128,52	2128,67	2135,79
W(λ, γ)	52,11 (0,65)	5,05 (0,22)				2114,36	2114,51	2121,63
BXII(s, k, c)	65,16 (6,28)	4,49 (1,94)	5,95 (0,41)			2106,78	2106,99	2117,68
OLLFr(λ, σ, α)	0,16 (0,07)	5,10 (4,69)	33,74 (14,00)			2131,89	2132,11	2142,80
OLLLL(λ, γ, α)	47,61 (0,65)	7,70 (562,34)	1,00 (73,04)			2130,52	2130,74	2141,43
OLLW(λ, γ, α)	53,17 (0,96)	3,75 (0,62)	1,46 (0,27)			2110,19	2110,41	2121,10
OLLBXII(s, k, c, α)	51,49 (0,75)	2,49 (0,34)	30,62 (0,68)	0,19 (0,01)		2093,14	2093,45	2107,68
BBXII(s, k, c, a, b)	57,30 (0,23)	0,01 (0,01)	24,84 (0,23)	0,17 (0,01)	41,51 (37,76)	2106,51	2106,93	2124,69
KwBXII(s, k, c, a, b)	60,52 (0,10)	0,09 (0,06)	26,27 (0,10)	0,19 (0,01)	2,78 (0,73)	2101,95	2102,36	2120,12

Fonte: Autora

Comparando-se as distribuições OLLBXII e BXII usando um teste de RV, rejeita-se a hipótese nula, ou seja, a distribuição OLLBXII pode ser escolhida como melhor modelo. A Tabela 8.3 apresenta resultado do teste RV. Observa-se que existe evidências do potencial do parâmetro extra ao modelar dados reais. Pode-se concluir que a distribuição OLLBXII é um modelo adequado para ajustar esses dados.

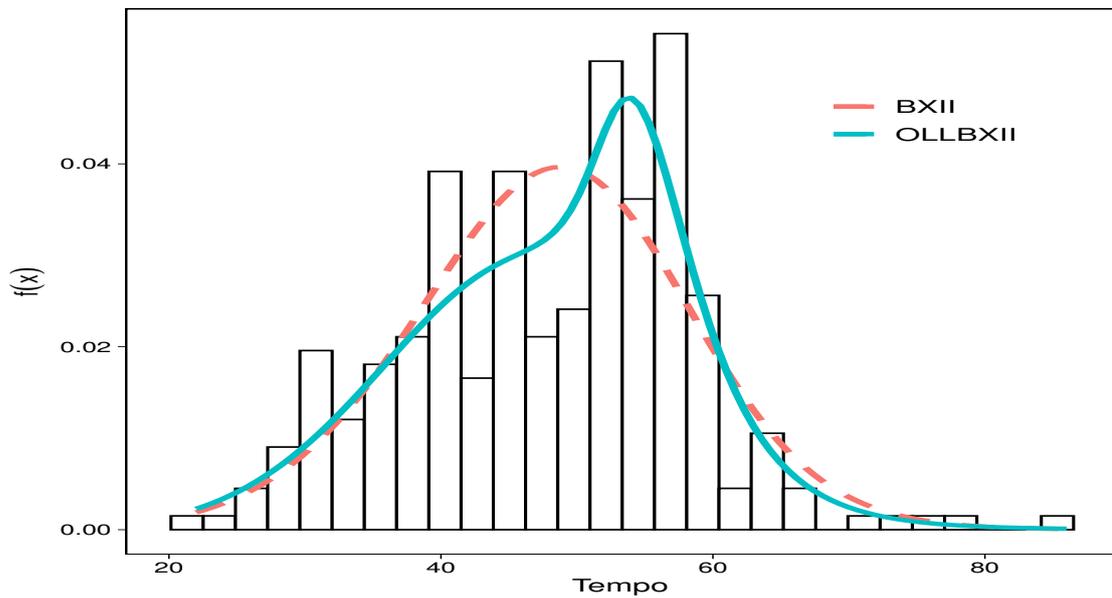
Tabela 8.3: Teste RV para a variável tempo até a morte de mulheres aposentadas.

Modelos	Hipóteses	Estatística w	p -valor
OLLBXII vs BXII	$H_0 : a = 1$ vs $H_1 : H_0$ é falso	15,64	$< 0,0001$

Fonte: Autora.

Para a avaliação da qualidade do modelo, a Figura 8.1 mostra os ajustes das funções densidades das distribuições BXII e OLLBXII, observando-se que o melhor ajuste corresponde a distribuição OLLBXII.

Figura 8.1: Estimação da densidade para a variável tempo até a morte de mulheres aposentadas.



Fonte: Autora.

A Tabela 8.4 mostra o teste de qualidade de ajuste do modelo ao tempo de aposentadoria até a morte de mulheres no México em 2004. Assim, realiza-se os testes formais de adequação para verificar qual distribuição se ajusta melhor aos dados. Considera-se as estatísticas de Cramér-Von Mises (W^*) e Anderson-Darling (A^*) [10]. Em geral, comparando-se as medidas desses testes de adequação formal, conclui-se que a distribuição OLLBXII supera todos os modelos considerados e, portanto, pode ser uma alternativa interessante para a modelagem de dados de tempo de vida. Esses resultados ilustram a importância da nova distribuição e a necessidade do parâmetro de forma adicional para analisar dados reais.

Tabela 8.4: Teste qualidade do ajuste para a variável tempo até a morte de mulheres aposentadas.

Modelo	Estatísticas	
	A^*	W^*
Fr	9,54	1,65
LL	3,81	0,67
W	1,20	0,16
OLLFr	3,92	0,69
OLLLL	3,81	0,67
OLLW	1,58	0,27
BXII	1,46	0,25
OLLBXII	0,59	0,09
BBXII	1,29	0,21
KwBXII	1,31	0,22

Fonte: Autora.

8.2 CONJUNTO DE DADOS 2: MELANOMA

Nesta seção e na seguinte serão considerados conjuntos de dados censurados, ou seja, não são em todas as observações que o evento de interesse ocorreu, ao menos no período de estudo em questão.

A primeira análise propõe uma aplicação da distribuição OLLBXII a um conjunto de dados de melanoma. Os dados fazem parte de um estudo sobre melanoma cutâneo (um tipo de câncer maligno) para a avaliação do desempenho de um tratamento pós-operatório com uma alta dose de um determinado medicamento (interferon alfa-2b) administrado a fim de prevenir a recorrência. Os pacientes foram incluídos no estudo de 1991 a 1995, com $n = 417$ de tamanho de amostra, e o acompanhamento foi realizado até 1998. Os dados foram coletados e analisados por [51]. O tempo de sobrevivência X é referido como o tempo até a morte do paciente. A porcentagem de observações censuradas foi de 56%. A Tabela 8.5 apresenta as EMVs e os erros-padrão (entre parênteses) dos parâmetros, para isso considerou-se como valores iniciais iguais a um para cada parâmetro das distribuições base. Além disso, para as extensões das distribuições base foram utilizados os valores estimados dos parâmetros das distribuições base e $\alpha = 1$. Os resultados indicam que a distribuição OLLBXII possui os menores valores das estatísticas (AIC e CAIC) em relação ao seu submodelo.

Um teste para a necessidade do parâmetro extra na distribuição OLLBXII pode ser baseado na estatística RV conforme descrito anteriormente. Aplicando-se as estatísticas RV a esses dados, os resultados são listados na Tabela 8.6. O p-valor mostra que a distribuição proposta produz o melhor ajuste aos dados.

Tabela 8.5: EMVs, erro padrão (entre parênteses) e as estatísticas AIC, CAIC e BIC para a variável tempo do tratamento pós-operatório até a morte de pessoas com melanoma.

Modelo	Estimativas					Estatísticas		
						AIC	CAIC	BIC
Fr(λ, σ)	0,66 (0,04)	3,00 (0,27)				1062,13	1062,23	1070,20
LL(λ, γ)	4,74 (0,39)	1,27 (0,08)				1086,32	1086,42	1094,38
W(λ, γ)	6,94 (0,56)	1,05 (0,07)				1102,11	1102,21	1110,17
BXII(s, k, c)	0,87 (0,12)	0,15 (0,03)	2,57 (0,35)			1058,35	1058,49	1070,45
OLLFr(λ, σ, α)	0,71 (0,09)	3,21 (0,49)	0,88 (0,20)			1063,84	1063,98	1075,93
OLLLL(λ, γ, α)	4,74 (0,39)	1,13 (35,56)	1,13 (35,56)			1088,32	1088,46	1100,42
OLLW(λ, γ, α)	57,96 (40,36)	0,15 (0,04)	6,57 (1,84)			1092,66	1092,81	1104,76
OLLBXII(s, k, c, α)	1,89 (0,29)	0,10 (0,04)	5,39 (1,50)	0,36 (0,11)		1054,97	1055,18	1071,11
BBXII(s, k, c, a, b)	1,77 (0,06)	0,34 (0,08)	11,46 (0,08)	0,15 (0,01)	0,07 (0,01)	1055,53	1055,81	1075,70
KwBXII(s, k, c, a, b)	1,69 (0,22)	0,02 (0,03)	6,23 (2,72)	0,28 (0,15)	0,85 (1,00)	1057,24	1057,51	1077,40

Tabela 8.6: Teste RV para a variável tempo do tratamento pós-operatório até a morte de pessoas com melanoma.

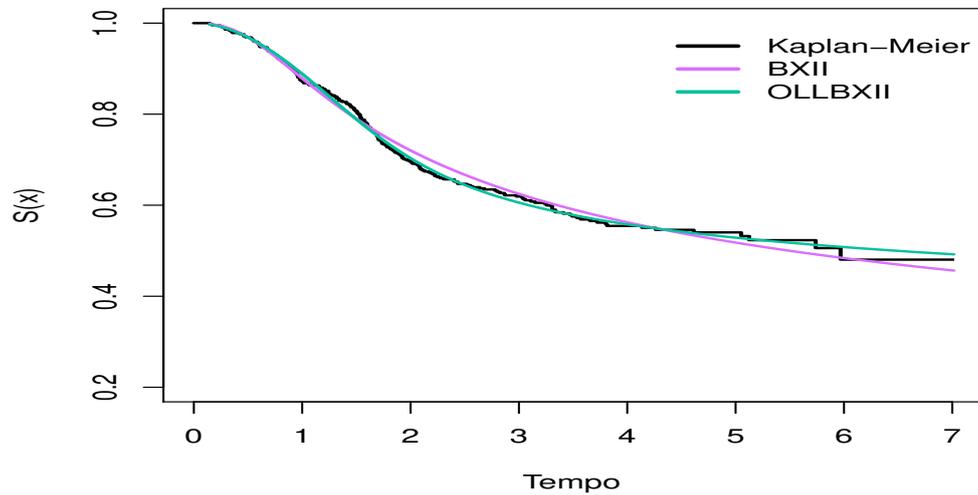
Modelo	Hipótese	Estatística w	p-valor
OLLBXII vs BXII	$H_0 : \alpha = 1$ vs $H_1 : H_0$ é falso	162	< 0,0001

Na Figura 8.2 exibe-se gráficos das funções de sobrevivência empíricas e estimadas da distribuição OLLBXII e seu submodelo. Nota-se que a distribuição OLLBXII fornece um ajuste satisfatório.

8.3 CONJUNTO DE DADOS 3: LINFOMA

O conjunto de dados consiste em informações sobre o tempo de vida ou recidiva (ou cura) de 43 pacientes com transplante de medula óssea que foram diagnosticados com linfoma de Hodgkin ou não Hodgkin e tratados com um transplante de medula óssea alogênico (irmão compatível com HLA) ou autogênico na Universidade Estadual de Ohio. A porcentagem de dados censurados foi de 39 %. O tempo de sobrevivência X é o tempo até a morte do paciente. A Tabela 8.7 mostra as EMVs dos parâmetros e seus respectivos erros padrão (entre

Figura 8.2: Função de sobrevivência estimada e empírica das distribuições BXII e OLLBXII para a variável tempo do tratamento pós-operatório até a morte de pacientes com melanoma.



Fonte: Autora.

parênteses) dos modelos ajustados acima e os valores do AIC, CAIC e BIC que indicam que a distribuição OLLBXII possui os menores valores dessas estatísticas.

Uma comparação da distribuição proposta com o modelo BXII usando estatísticas RV é fornecida na Tabela 8.8. Rejeita-se a hipótese nula do teste RV em favor da distribuição OLLBXII. A rejeição é extremamente significativa. Isso indica uma evidência clara do potencial do parâmetro extra α ao modelar dados reais.

Tabela 8.8: Teste RV para a variável tempo até a morte de pacientes com linfoma que foram transplantados.

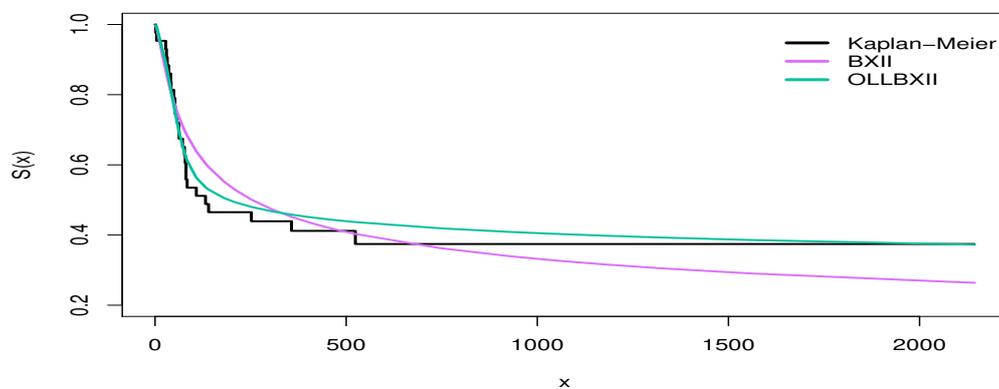
Modelo	Hipótese	Estatística w	p-valor
OLLBXII vs BXII	$H_0 : \alpha = 1$ vs $H_1 : H_0$ é falso	20,69	< 0,0001

A fim de avaliar se o modelo é apropriado, a Figura 8.3 exibe funções empíricas e estimadas de sobrevivência OLLBXII para os dados de linfoma. Pode-se concluir que a nova distribuição é um modelo muito adequado para ajustar esses conjuntos de dados.

Tabela 8.7: EMVs, erro padrão (entre parênteses) e as estatísticas AIC, BIC e CAIC para a variável tempo até a morte de pacientes com linfoma que foram transplantados.

Modelo	Estimativas					Estatística		
						AIC	CAIC	BIC
Fr(λ, σ)	0,42 (0,06)	110,66 (43,45)				359,34	360,40	362,87
LL(λ, γ)	247,93 (96,08)	0,70 (0,12)				364,42	365,47	367,94
W(λ, γ)	676,60 (246,41)	0,49 (0,08)				370,70	371,75	374,22
BXII(s, k, c)	25,74 (10,62)	0,18 (0,07)	1,64 (0,47)			357,85	359,47	363,13
OLLFr(λ, σ, α)	0,43 (0,13)	113,63 (60,72)	0,98 (0,46)			361,33	362,96	366,62
OLLLL(λ, γ, α)	247,64 (95,89)	1,33 (30,77)	0,53 (12,30)			366,42	368,04	371,71
OLLW(λ, γ, α)	676,63 (273,32)	0,33 (0,13)	1,56 (0,68)			371,56	373,19	376,85
OLLBXII(s, k, c, α)	80,02 (0,02)	0,10 (0,07)	8,29 (0,02)	0,19 (0,03)		349,24	351,57	356,29
BBXII(s, k, c, a, b)	62,24 (0,04)	0,01 (0,03)	7,64 (0,04)	0,19 (0,04)	0,54 (0,87)	350,96	354,16	359,77
KwBXII(s, k, c, a, b)	43,28 (0,03)	0,18 (0,03)	9,75 (0,02)	0,09 (0,04)	0,13 (0,03)	354,05	357,25	362,85

Figura 8.3: Função de sobrevivência estimada e empírica para os modelos BXII e OLLBXII para a variável tempo até a morte de pacientes com linfoma que foram transplantados.



Fonte: Autora.

8.4 CONJUNTO DE DADOS 4: TEOR ALCOÓLICO

O último conjunto de dados estudado nessa dissertação trata-se do teor alcoólico no sangue e na expiração entre homens e mulheres, em uma amostra de 59 austríacos, e foi descrita em [52]. A variável resposta consiste em uma taxa de eliminação de álcool no sangue (g por litro por hora), com média de 0,1695 e desvio padrão de 0,0328. As covariáveis consideradas são: x_1 : representa a taxa de eliminação de álcool no ar expirado (mg por litro por hora) e x_2 : indica se o austríaco estudado é do gênero masculino, indicado por 1, ou feminino, indicado por 2. A Tabela 8.9 exibe a comparação entre as respectivas EMVs, seus erros padrões, p-valor e as estatísticas aplicadas.

Tabela 8.9: EMVs, erro padrão (entre parênteses), p-valor [.] e as estatísticas AIC, CAIC e BIC para os dados de teor alcoólico.

Modelo	Estimativas						Estatística		
	β_0	β_1	β_2	σ	k	α	AIC	CAIC	BIC
LL	0,0215 (0,0106) [0,0466]	1,7934 (0,1260) [< 0,0001]	0,0005 (0,0041) [0,9068]	0,0087 (0,0009) -	-	-	-315,7995	-314,1841	-307,4893
LBXII	0,0260 (0,0124) [0,0409]	1,6511 (0,1855) [< 0,0001]	-0,0008 (0,0039) [0,8460]	0,0062 (0,0017) -	0,4706 (0,2503) -	-	-316,3555	-314,1595	-305,9678
LOLLBXII	0,0214 (0,0148) [0,1549]	1,6534 (0,1901) [< 0,0001]	-0,0008 (0,0039) [0,8354]	0,0076 (0,0031) -	0,3920 (0,3578) -	1,3667 (0,8834) -	-314,6129	-311,7329	-302,1476

A Tabela 8.9, mostra que o modelo de regressão que mais acomoda o conjunto de dados por meio dos critérios de informação é o modelo LOLLBXII. Além disso, a Tabela 8.10 apresenta o intervalo de confiança para os parâmetros do modelo de regressão LOLLBXII.

Tabela 8.10: Intervalo de confiança de 95% para os parâmetros da regressão LOLLBXII da variável explicativa teor alcoólico.

Parâmetros	IC(95%)
β_0	(-0,0083; 0,0511)
β_1	(1,2729; 2,0339)
β_2	(-0,0085; 0,0069)
σ	(0,0014; 0,0138)
k	(-0,3240; 1,1080)
α	(-0,4010; 3,1345)

Observa-se que a variável x_1 demonstrou ser significativa a um nível de significância de 5% como exibem as Tabelas 8.9 e 8.10, ou seja, é possível ao mensurar o ar expirado,

determinar se alguém apresentou um alto consumo de bebida alcoólica. Por essa razão a Tabela 8.11 apresenta as novas estatísticas para cada modelo sem a covariável x_2 , o modelo LOLLBXII continua sendo a melhor opção entre os modelos estudados, sendo o modelo escolhido.

Tabela 8.11: EMVs, erro padrão (entre parênteses), p-valor [.] e as estatísticas AIC, CAIC e BIC para os dados de teor alcoólico, sem a covariável x_2 .

Modelo	Estimativas					Estatística		
	β_0	β_1	σ	k	α	AIC	CAIC	BIC
LL	0,0219 (0.0099) [0.0313]	1,7965 (0.1234) [< 0,0001]	0,0087 (0.0009) —	7 — —	— — —	-317.7862	-316.6541	-311.5536
LBXII	0,0275 (0.0127) [0.0351]	1,5965 (0.2060) [< 0,0001]	0,0055 (0.0017) —	0,3855 (0.2155) —	— — —	-318.5851	-316.9697	-310.2749
LOLLBXII	0,0279 (0.0175) [0.1162]	1,5952 (0.5863) [0.0087]	0,0053 (0.0018) —	0,3887 (1.0232) —	0,9538 (1.3185) —	-316.5529	-314.3568	-306.1652

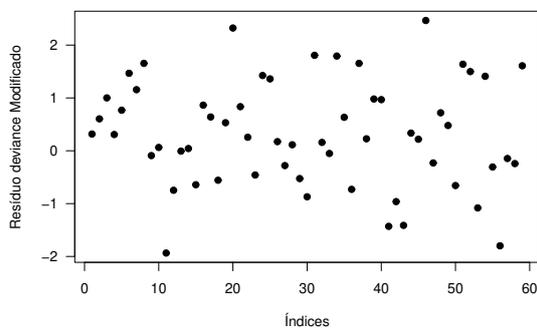
As Figura 8.4a e 8.4b apresentam a análise de resíduos para o modelo LOLLBXII, utilizando o resíduo da deviance modificada para verificar se o modelo LOLLBXII é adequado para o ajuste dos dados de teor alcoólico. Observa-se que é importante que os resíduos estejam dispersos em torno do zero de forma aleatória e que não encontrem-se muito distantes de zero, o que acontece em ambos os gráficos. O envelope simulado é apresentado na Figura 8.4c em que as bandas do envelope contém todos os pontos. Assim, por meio da análise de resíduos o modelo de regressão LOLLBXII é adequado para o ajuste dos dados.

Portanto, o modelo de regressão após uma seleção de variáveis como mais apropriado para ajustar os dados teor alcoólico é o LOLLBXII, e é dado por,

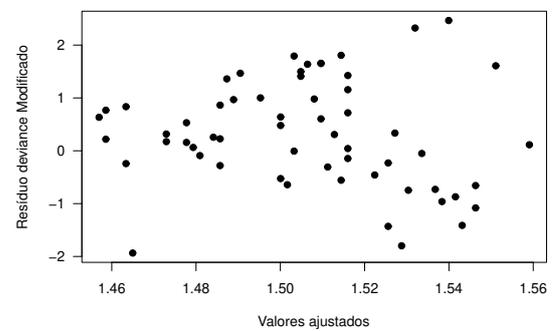
$$y_i = \beta_0 + \beta_1 x_1 + \sigma z_i \quad \text{em que } i = 1, \dots, 59.$$

Observa-se que, conforme o teor alcoólico eliminado pelo ar expirado aumenta a cada mg/l/h em uma unidade, o teor alcoólico eliminado no sangue aumenta em média $\exp(1,6534)$, ou seja 5,225 g/l/h.

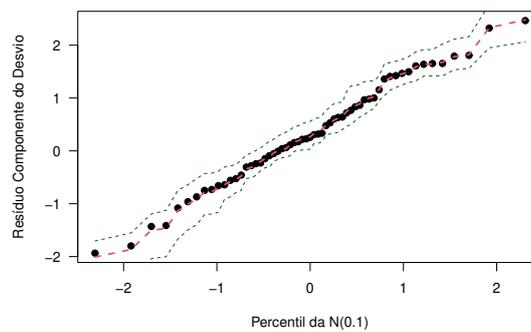
Figura 8.4: Resíduos para a regressão LOLLBXII dos dados teor alcoólico.



(a) Resíduo deviance modificado versus ordens dos índices das observações.



(b) Resíduos deviance modificado versus valores ajustados.



(c) Envelope simulado

Fonte: Autora.

9 CONSIDERAÇÕES FINAIS

Verificou-se neste trabalho que a análise de sobrevivência é extremamente abrangente, e tem uma gama de possíveis aplicações, assim, famílias que permitam obter extensões de distribuições existentes são importantes.

Neste sentido, propôs-se a distribuição Odd Log-Logística Burr XII (OLLB-XII) que representa uma extensão da distribuição Burr XII (BXII) em que é adicionado um parâmetro de forma.

Foram obtidas, diferentes propriedades matemáticas da distribuição OLLB-XII ao representá-la como uma combinação linear da distribuição BXII, o que possibilita calcular os momentos ordinários e incompletos e a função geradora de momentos.

Além disso, foram realizadas simulações de Monte Carlo para diferentes parâmetros e tamanhos de amostras. Para os conjuntos de dados estudados foi possível observar que ao comparar com sub-modelos e modelos não encaixados a distribuição OLLBXII obteve um bom desempenho. Assim é possível afirmar que a nova distribuição pode ser considerada uma alternativa para os modelos comparados neste estudo.

Foi proposto também um modelo de regressão de locação baseado na distribuição OLLBXII e uma aplicação a um conjunto de dados reais. Assim, comparou-se as estatísticas AIC, CAIC e BIC do modelo proposto LOLLBXII com os modelos log-logístico, e LBXII. Realizou-se análise de resíduos e construiu-se o envelope simulado para o modelo proposto, demonstrando que o modelo LOLLBXII obteve um bom desempenho ao ajustar os dados.

Para trabalhos futuros, pode-se propor outros tipos de modelos de regressão utilizando a distribuição OLLBXII, como por exemplo, modelos de fração de cura e modelos GAMLSS, além de aplicar conjuntos de dados de diferentes áreas de pesquisa.

REFERÊNCIAS

- [1] AFIFY, A; ABDELLATIF, A. The extended burr xii distribution: properties and applications. *J. of Nonlinear Sci. and Appl* 13 (2020), 133–146.
- [2] AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 6 (1974), 716–723.
- [3] AL-HUSSAINI, E. K., AND HUSSEIN, M. Estimation using censored data from exponentiated burr type xii population. *American Open Journal of Statistics* 1 (2011), 33–45.
- [4] AL-SAIARI, A. Y., BAHARITH, L. A., AND MOUSA, S. A. Marshall-olkin extended burr type xii distribution. *International Journal of Statistics and probability* 3, 1 (2014), 78.
- [5] ALEXANDER, C., CORDEIRO, G. M., ORTEGA, E. M., AND SARABIA, J. M. Generalized beta-generated distributions. *Computational Statistics & Data Analysis* 56, 6 (2012), 1880–1897.
- [6] ANDERSON, T W; DARLING, D. A. A test of goodness of fit. *Journal of the American statistical association* 49, 268 (1954), 765–769.
- [7] ATKINSON, A. C. Two graphical displays for outlying and influential observations in regression. *Biometrika* 68, 1 (1981), 13–20.
- [8] BALAKRISHNAN, N., LEIVA, V., SANHUEZA, A., AND CABRERA, E. Mixture inverse gaussian distributions and its transformations, moments and applications. *Statistics* 43, 1 (2009), 91–104.
- [9] BURR, I. W. Cumulative frequency functions. *The Annals of mathematical statistics* 13, 2 (1942), 215–232.
- [10] CHEN, G., AND BALAKRISHNAN, N. A general purpose approximate goodness-of-fit test. *Journal of Quality Technology* 27, 2 (1995), 154–161.
- [11] CHOULAKIAN, V., AND STEPHENS, M. A. Goodness-of-fit tests for the generalized pareto distribution. *Technometrics* 43, 4 (2001), 478–484.
- [12] COLLETT, D. *Modelling survival data in medical research*. CRC press, 2015.
- [13] COLOSIMO, E. A., AND GIOLO, S. R. *Análise de sobrevivência aplicada*. Editora Blucher, 2006.

- [14] COOK, R. D., AND WEISBERG, S. *Residuals and influence in regression*. New York: Chapman and Hall, 1982.
- [15] COORAY, K. Generalization of the weibull distribution: the odd weibull family. *Statistical Modelling* 6, 3 (2006), 265–277.
- [16] CORDEIRO, G. M., ALIZADEH, M., PESCIM, R. R., AND ORTEGA, E. M. M. The odd log-logistic generalized half-normal lifetime distribution: Properties and applications. *Communications in Statistics-Theory and Methods* 46, 9 (2017), 4195–4214.
- [17] CORDEIRO, G. M., PESCIM, R. R., DEMÉTRIO, C. G., AND ORTEGA, E. M. The kummer beta generalized gamma distribution. *Journal of Data Science* 12, 4 (2014), 661–697.
- [18] CRUZ, J. N. D. *A nova família de distribuições odd log-logística: teoria e aplicações*. PhD thesis, Universidade de São Paulo, 2017.
- [19] DA SILVA, R. V., GOMES-SILVA, F., RAMOS, M. W. A., AND CORDEIRO, G. M. The exponentiated burr xii poisson distribution with application to lifetime data. *International Journal of Statistics and Probability* 4, 4 (2015), 112.
- [20] DR, C., AND EJ, S. A general definition of residuals. *J R Stat Soc Ser B* 30 (1968), 248–275.
- [21] EL-BASSIOUNY, A. H., AND ABDO, N. F. Reliability properties of seven parameters burr xii distribution. *Comput Methods Sci Technol* 16 (2010), 127–133.
- [22] EUGENE, N., LEE, C., AND FAMOYE, F. Beta-normal distribution and its applications. *Communications in Statistics-Theory and methods* 31, 4 (2002), 497–512.
- [23] FERNANDES, V. V. *Contribuições sobre o envelope simulado na análise de diagnóstico em modelos de regressão*. PhD thesis, Universidade Federal de São Carlos, 2019.
- [24] FLEMING, T. R., AND HARRINGTON, D. P. *Counting processes and survival analysis*, vol. 169. John Wiley & Sons, 2011.
- [25] GAD, A. M., HAMEDANI, G. G., SALEHABADI, S. M., AND YOUSOF, H. M. The burr xii-burr xii distribution: mathematical properties and characterizations. *Pakistan Journal of Statistics* (2019).
- [26] GLEATON, J. U., AND LYNCH, J. D. Extended generalized log-logistic families of lifetime distributions with an application. *J. Probab. Stat. Sci* 8, 1 (2010), 1–17.
- [27] GOMES, A. E., DA SILVA, C. Q., AND CORDEIRO, G. M. Two extended burr models: Theory and practice. *Communications in Statistics-Theory and Methods* 44, 8 (2015), 1706–1734.

- [28] GUERRA, R. R. Some generalized burr xii distributions with applications to income and lifetime data. *Biometrika* (2017).
- [29] GUPTA, R. C., GUPTA, P. L., AND GUPTA, R. D. Modeling failure time data by lehman alternatives. *Communications in Statistics-Theory and methods* 27, 4 (1998), 887–904.
- [30] HANDIQUE, L., AND CHAKRABORTY, S. A new four-parameter extension of burr-xii distribution: its properties and applications. *Japanese Journal of Statistics and Data Science* 1, 2 (2018), 271–296.
- [31] HASHIMOTO, E. M. *Modelo de regressão gama-G em análise de sobrevivência*. PhD thesis, Universidade de São Paulo, 2013.
- [32] HASHIMOTO, E. M., ORTEGA, E. M. M., CANCHO, V. G., AND CORDEIRO, G. M. The log-exponentiated weibull regression model for interval-censored data. *Computational statistics & data analysis* 54, 4 (2010), 1017–1035.
- [33] JAMAL, F., CHESNEAU, C., NASIR, M. A., SABOOR, A., ALTUN, E., AND KHAN, M. A. On a modified burr xii distribution having flexible hazard rate shapes. *Mathematica Slovaca* 70, 1 (2020), 193–212.
- [34] KAPLAN, E. L., AND MEIER, P. Nonparametric estimation from incomplete observations. *Journal of the American statistical association* 53, 282 (1958), 457–481.
- [35] KLEIBER, C., AND KOTZ, S. *Statistical size distributions in economics and actuarial sciences*, vol. 470. John Wiley & Sons, 2003.
- [36] KLEINBAUM, D. G., AND KLEIN, M. *Survival analysis*. Springer, 2012.
- [37] KUMAR, D. Ratio and inverse moments of marshall-olkin extended burr type xii distribution based on lower generalized order statistics. *Journal of Data Science* 14, 1 (2016), 53–66.
- [38] KUMAR, D. The burr type xii distribution with some statistical properties. *Journal of data science* 15, 3 (2017).
- [39] LAWLESS, J. F. *Statistical models and methods for lifetime data*, vol. 362. John Wiley & Sons, 2003.
- [40] MOOD, A., GRAYBILL, F., AND BOES, D. *Introduction to the theory of statistics*. Open University Pres, 1974.
- [41] MOORE, D. F. *Applied survival analysis using R*. Springer, 2016.
- [42] MUDHOLKAR, G. S., AND SRIVASTAVA, D. K. Exponentiated weibull family for analyzing bathtub failure-rate data. *IEEE transactions on reliability* 42, 2 (1993), 299–302.

- [43] NADARAJAH, S. The exponentiated gumbel distribution with climate application. *Environmetrics: The official journal of the International Environmetrics Society* 17, 1 (2006), 13–23.
- [44] NADARAJAH, S., AND GUPTA, A. K. The exponentiated gamma distribution with application to drought data. *Calcutta Statistical Association Bulletin* 59, 1-2 (2007), 29–54.
- [45] NADARAJAH, S., AND HAGHIGHI, F. An extension of the exponential distribution. *Statistics* 45, 6 (2011), 543–558.
- [46] NADARAJAH, S., AND KOTZ, S. The exponentiated fréchet distribution. *Interstat Electronic Journal* 14 (2003), 01–07.
- [47] NADARAJAH, S., AND KOTZ, S. The exponentiated type distributions. *Acta Applicandae Mathematica* 92, 2 (2006), 97–111.
- [48] ORTEGA, E. M., CORDEIRO, G. M., AND LEMONTE, A. J. A log-linear regression model for the β -birnbaum–saunders distribution with censored data. *Computational Statistics & Data Analysis* 56, 3 (2012), 698–718.
- [49] PARANAÍBA, P. F. *Caracterização e extensões da distribuição Burr XII: Propriedades e Aplicações*. PhD thesis, Universidade de São Paulo, 2012.
- [50] PARANAÍBA, P. F., ORTEGA, E. M., CORDEIRO, G. M., AND PASCOA, M. A. D. The kumaraswamy burr xii distribution: theory and practice. *Journal of Statistical Computation and Simulation* 83, 11 (2013), 2117–2143.
- [51] PARANAÍBA, P. F., ORTEGA, E. M., CORDEIRO, G. M., AND PESCIM, R. R. The beta burr xii distribution with application to lifetime data. *Computational Statistics & Data Analysis* 55, 2 (2011), 1118–1136.
- [52] PAVLIC, M., GRUBWEISER, P., LIBISILLER, K., AND RAB, W. Elimination rates of breath alcohol. *Forensic Science International* 171 (2007), 16–21.
- [53] PESCIM, R. R. *The new class of Kummer beta generalized distributions: theory and applications*. PhD thesis, Universidade de São Paulo, 2014.
- [54] PESCIM, R. R., CORDEIRO, G. M., DEMÉTRIO, C. G., ORTEGA, E. M., AND NADARAJAH, S. The new class of kummer beta generalized distributions. *SORT-Statistics and Operations Research Transactions* (2012), 153–180.
- [55] PESCIM, R. R., ORTEGA, E. M., CORDEIRO, G. M., AND ALIZADEH, M. A new log-location regression model: estimation, influence diagnostics and residual analysis. *Journal of Applied Statistics* 44, 2 (2017), 233–252.

- [56] PESCIM, R. R., ORTEGA, E. M. M., CORDEIRO, G. M., DEMTRIO, C. G. B., AND HAMEDANI, G. The log-beta generalized half-normal regression model. *Journal of Statistical Theory and Applications* (2013).
- [57] PRATAVIERA, F., ORTEGA, E. M., CORDEIRO, G. M., AND BRAGA, A. D. S. The heteroscedastic odd log-logistic generalized gamma regression model for censored data. *Communications in Statistics-Simulation and Computation* 48, 6 (2019), 1815–1839.
- [58] PRUDNIKOV, A. P., BRYCHKOV, Y. A., MARICHEV, O. I., AND ROMER, R. H. Integrals and series, 1988.
- [59] RAMINELLI, J. A. *Métodos de adequação e diagnóstico em modelos de sobrevivência dinâmicos*. PhD thesis, Universidade de São Paulo, 2015.
- [60] SABOOR, A., ALIZADEH, M., KHAN, M. N., GHOSH, I., AND CORDEIRO, G. M. Odd log-logistic modified weibull distribution. *Mediterranean Journal of Mathematics* 14, 2 (2017), 96.
- [61] SANTANA, T. V. F. D. *As distribuições Kumaraswamy-log-logística e Kumaraswamy-logística*. PhD thesis, Universidade de São Paulo, 2010.
- [62] SHAO, Q. Notes on maximum likelihood estimation for the three-parameter burr xii distribution. *Computational statistics & data analysis* 45, 3 (2004), 675–687.
- [63] SILVA, G. O., ORTEGA, E. M., CANCHO, V. G., AND BARRETO, M. L. Log-burr xii regression models with censored data. *Computational statistics & data analysis* 52, 7 (2008), 3820–3842.
- [64] SILVA, G. O., ORTEGA, E. M. M., AND CORDEIRO, G. M. A log-extended weibull regression model. *Computational statistics & data analysis* 53, 12 (2009), 4482–4489.
- [65] SILVA, R. B., AND CORDEIRO, G. M. The burr xii power series distributions: A new compounding family. *Brazilian Journal of Probability and Statistics* 29, 3 (2015), 565–589.
- [66] TEAM, R. C. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.
- [67] TERRY, M. T., PATRICIA, M. G., AND THOMAS, R. F. Martingale-based residuals for survival models. *Biometrika* 77, 1 (1990), 147–160.
- [68] WINGO, D. Maximum likelihood methods for fitting the burr type xii distribution to life test data. *Biometrical journal* 25, 1 (1983), 77–84.

- [69] WU, S.-J., CHEN, Y.-J., AND CHANG, C.-T. Statistical inference based on progressively censored samples with random removals from the burr type xii distribution. *Journal of Statistical Computation and Simulation* 77, 1 (2007), 19–27.
- [70] ZIMMER, W. J., KEATS, J. B., AND WANG, F. The burr xii distribution in reliability analysis. *Journal of quality technology* 30, 4 (1998), 386–394.