



Universidade Federal da Paraíba
Centro de Informática
Programa de Pós-Graduação em Modelagem Matemática e Computacional

CRITÉRIOS DE SELEÇÃO DE MODELOS: UM ESTUDO COMPARATIVO

Adriana Ribeiro Moura

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Modelagem Matemática e Computacional, UFPB, da Universidade Federal da Paraíba, como parte dos requisitos necessários à obtenção do título de Mestre em Modelagem Matemática e Computacional.

Orientadores: Claudio Javier Tablada
Renilma Pereira da Silva

João Pessoa
Julho de 2021

CRITÉRIOS DE SELEÇÃO DE MODELOS: UM ESTUDO COMPARATIVO

Adriana Ribeiro Moura

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL (PPGMMC) DO CENTRO DE INFORMÁTICA DA UNIVERSIDADE FEDERAL DA PARAÍBA COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM MODELAGEM MATEMÁTICA E COMPUTACIONAL.

Examinada por:



Prof. Claudio Javier Tablada, Dr.



Prof.ª Renilma Pereira da Silva, Dra.



Prof.ª Tarciana Liberal Pereira de Araújo, Dra.



Prof. Rodrigo Bernardo da Silva, Dr.

JOÃO PESSOA, PB – BRASIL

JULHO DE 2021

Catálogo na publicação
Seção de Catalogação e Classificação

M929c Moura, Adriana Ribeiro.

Critérios de seleção de modelos : um estudo comparativo
/ Adriana Ribeiro Moura. - João Pessoa, 2021.
94 f.

Orientação: Claudio Javier Tablada, Orientação: Renilma
Pereira da Silva.

Dissertação (Mestrado) - UFPB/CI.

1. Matemática computacional. 2. Modelagem
computacional. 3. Probabilidade - distribuições. I.
Tablada, Claudio Javier. II. Silva, Renilma Pereira da.
III. Título.

UFPB/BC

CDU 519.6(043)

*Desde o nascer até o pôr do sol,
que o nome do Senhor seja
louvado! Salmos 113:3*

Agradecimentos

Agradeço primeiramente a Deus, que me proporcionou força, saúde e coragem para conquistar mais uma etapa da minha vida acadêmica, me mostrando que sou capaz de superar cada dificuldade encontrada no caminho.

Aos meus pais, José Carlos e Josinete, principalmente a você mãe exemplo de força e garra, energia viva das minhas motivações.

Ao meu esposo, Reginaldo, essencial na minha vida e que tanto contribui com apoio e compreensão nos momentos mais difíceis.

A minha irmã, Andréa, que contribui de todas as formas e sempre acreditou e depositou confiança de que tudo seria possível porque a vontade de Deus era maior.

Não poderia de deixar de agradecer a você, Bárbara (minha amiga), que tanto torceu para essa realização e que foi de grande importância na minha graduação, serei eternamente grata.

Aos meus orientadores, Dr.Sc. Claudio Javier Tablada e Dra.Sc. Renilma Pereira da Silva pela paciência, disponibilidade e sugestões.

À UFPB que me proporcionou a oportunidade de aprimoramento acadêmico.

A todos os professores do PPGMMC que contribuíram com seus conhecimentos ao longo do caminho.

A colega Emília pelo companheirismo, estudo, amizade e momentos de alegria que foram essenciais para mantermos firmes.

A todos, que direta ou indiretamente, contribuíram para a realização deste trabalho, muito obrigada!

Resumo da Dissertação apresentada ao PPGMMC/CI/UFPB como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

CRITÉRIOS DE SELEÇÃO DE MODELOS: UM ESTUDO COMPARATIVO

Adriana Ribeiro Moura

Julho/2021

Orientadores: Claudio Javier Tablada
Renilma Pereira da Silva

Programa: Modelagem Matemática e Computacional

Diversas pesquisas científicas em várias áreas, inclusive em Estatística, têm seus problemas de estudo ligados a situações práticas, que normalmente podem ser explicadas através de modelos, sendo comum o pesquisador se deparar com mais de um modelo descrevendo um mesmo fenômeno. Diante desse fato, autores defendem a necessidade de um critério padrão baseado em princípios científicos para a escolha do modelo que “melhor” explique o fenômeno e a literatura já dispõe de vários critérios de seleção de modelos com esse objetivo. Assim, considerando as famílias de distribuições generalizadas Sup e Inf, o presente trabalho tem como objetivo propor um novo critério de seleção de modelos para modelos não encaixados, baseado nessas famílias de distribuições e suas propriedades e comparar seu desempenho com os critérios: critério de informação de Akaike (AIC), critério de informação de Akaike corrigido (AICc), critério de informação bayesiano (BIC), critério de informação de Hannan-Quinn (HQIC) e os critérios de adequação de ajuste modificados de Crámer-Von Mises (W^*) e Anderson-Darling (A^*) através de diferentes cenários de simulação. Também para fins de comparação, foi ilustrado sua aplicabilidade por meio de conjuntos de dados reais. Adicionalmente, foram apresentados os resultados mais importantes sobre regressão linear múltipla e realizadas simulações com o objetivo de comparar o desempenho do novo critério proposto com os critérios AIC, AICc, BIC e HQIC na seleção de modelos de regressão, bem como uma aplicação a um conjunto de dados reais.

Abstract of Dissertation presented to PPGMMC/CI/UFPB as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

CRITÉRIOS DE SELEÇÃO DE MODELOS: UM ESTUDO COMPARATIVO

Adriana Ribeiro Moura

July/2021

Advisors: Claudio Javier Tablada
Renilma Pereira da Silva

Program: Computational Mathematical Modelling

Several scientific researches in various areas, including Statistics, have their study problems linked to practical situations, which can usually be explained through models, and it is common for researchers to come across more than one model describing the same phenomenon. Given this fact, authors defend the need for a standard criterion based on scientific principles for choosing the model that “best” explains the phenomenon and the literature already has several criteria for selecting models with this objective. Thus, considering the families of generalized distributions Sup and Inf, the present work aims to propose a new model selection criterion for non-embedded models, based on these families of distributions and their properties and to compare their performance with the criteria: information criterion of Akaike (AIC), corrected Akaike information criterion (AICc), Bayesian information criterion (BIC), Hannan information criterion- Quinn (HQIC) and the modified fit criteria of Crámer-Von Mises (W^*) and Anderson-Darling (A^*) through different simulation scenarios. Also for comparison purposes, its applicability was illustrated using real datasets. Additionally, the most important results on multiple linear regression were presented and simulations were performed in order to compare the performance of the new proposed criterion with the AIC, AICc, BIC and HQIC criteria in the selection of regression models, as well as an application to a set of real data.

Sumário

Lista de Figuras	xi
Lista de Tabelas	xii
Lista de Símbolos	xiii
Lista de Abreviaturas	xiv
1 Introdução	1
2 Preliminares	4
2.1 Conceitos Básicos na Teoria da Probabilidade	4
2.2 Conceitos Básicos na Teoria da Inferência Estatística	7
2.3 Análise de regressão linear múltipla	10
2.3.1 Modelo teórico	11
2.3.2 Pressupostos do modelo	12
2.3.3 Representação matricial do método de regressão linear múltipla	12
2.3.4 Estimação dos parâmetros do modelo	13
2.3.5 Propriedades dos estimadores de mínimos quadrados	14
2.3.6 Estimador de σ^2	15
2.3.7 Estimadores de Máxima Verossimilhança	16
2.4 Medidas para avaliar modelos de regressão	16
2.4.1 Coeficiente de determinação múltiplo - R^2	16
2.4.2 Coeficiente de determinação múltiplo ajustado - \bar{R}^2	17
3 Critérios de seleção de modelos: Revisão	18
3.0.1 Critério de Informação de Akaike - AIC	18
3.0.2 Critério de Informação de Akaike corrigido - AICc	19
3.0.3 Critério de Informação Bayesiano - BIC	19
3.0.4 Critério de Informação de Hannan-Quinn - HQIC	19
3.0.5 Critérios de Cramér-von Mises e Anderson-Darling Modificados	20

4	Um Novo Critério de Seleção de Modelos Baseado nas Propriedades das Famílias Sup e Inf	22
4.1	Famílias de distribuições Sup e Inf	22
4.2	Função de log-verossimilhança das famílias Sup e Inf	25
4.3	Proposta do novo critério de seleção de modelos	28
5	Resultados Numéricos	30
5.1	Cenário 1: Distribuição Weibull Exponencializada \times Distribuição Weibull	31
5.2	Cenário 2: Distribuição Lindley Ponderada \times Distribuição Weibull	34
5.3	Cenário 3: Distribuição Weibull Exponencializada \times Distribuição Weibull	37
5.4	Cenário 4: Distribuição Normal \times Distribuição Normal	39
6	Aplicações	43
6.1	Aplicação 1	43
6.2	Aplicação 2	45
6.3	Aplicação 3	47
6.4	Aplicação 4	49
6.5	Aplicação 5	51
7	Regressão Linear: Estudo de Simulação e Aplicação	54
7.1	Cenário 1	54
7.2	Cenário 2	57
7.3	Aplicação	60
8	Conclusões	67
	Referências Bibliográficas	69
A	Algumas Distribuições Contínuas	73
A.0.1	Distribuição Normal	73
A.0.2	Distribuição Weibull	74
A.0.3	Distribuição Weibull Exponencializada	74
A.0.4	Distribuição Lindley Ponderada	75
A.0.5	Distribuição Fréchet	76
A.0.6	Distribuição Fréchet Modificada	76
A.0.7	Distribuição Fréchet Exponencializada	77
A.0.8	Distribuição Lindley estendida de Marshall-Olkin	78
A.0.9	Distribuição Weibull estendida de Marshall-Olkin	78
A.0.10	Distribuição Fréchet tipo II	79

A.0.11 Distribuição Burr XII	80
A.0.12 Distribuição Gama	81
A.0.13 Distribuição Gompertz	81
A.0.14 Distribuição Kumaraswamy	82

Lista de Figuras

4.1	Plots da f.d.a da distribuição Weibull, Weibull-sup e Weibull-inf . . .	25
5.1	Densidades do modelo Weibull versus WE.	31
5.2	Taxas de acertos para o cenário 1.	34
5.3	Densidades do modelo Weibull versus LP.	34
5.4	Taxas de acertos para o cenário 2.	36
5.5	Densidades do modelo Weibull versus WE.	37
5.6	Taxas de acertos para o cenário 3.	39
5.7	Densidades do modelo Normal 1 versus Normal 2.	40
5.8	Taxas de acertos para o cenário 4.	42
6.1	Densidades estimadas de FE, FM e LP para Dados I.	44
6.2	Gráficos Q-Q plots para as distribuições FE, FM e LP.	45
6.3	Densidades estimadas de W e WE para Dados II.	46
6.4	Gráficos Q-Q plots para as distribuições W e WE.	47
6.5	Densidades estimadas de FE, LP e WE para Dados III.	48
6.6	Gráficos Q-Q plots para as distribuições FE, LP e WE.	49
6.7	Densidades estimadas de Go e Kw para Dados IV.	50
6.8	Gráficos Q-Q plots para as distribuições Go e Kw.	51
6.9	Densidades estimadas de LP, WE e FE para Dados V.	52
6.10	Gráficos Q-Q plots para as distribuições LP, WE e FE.	53
7.1	Taxas de acertos para o Cenário 1.	56
7.2	Taxas de acertos para o cenário 2.	59
7.3	Histograma das variáveis.	61
7.4	Boxplot.	62
7.5	Correlograma.	63
7.6	Gráfico de ajuste do Modelo 3.	65
7.7	Gráfico dos resíduos do Modelo 3.	65

Lista de Tabelas

2.1	Tipos de erros em testes de hipóteses.	9
5.1	Taxas de acertos para o cenário 1.	32
5.2	Taxas de acertos para o cenário 2.	35
5.3	Taxas de acertos para o cenário 3.	38
5.4	Taxas de acertos para o cenário 4.	41
6.1	Dados I: Vinte itens testados até a falha.	43
6.2	Critérios de adequação de ajuste para Dados I.	44
6.3	Dados II: Tempo de vida (em milhões) de rolamentos de esferas de revoluções.	45
6.4	Critérios de adequação para Dados II	46
6.5	Dados III: Fluxos anuais do rio Weldon em Mill Grove, Missouri (1930- 59).	47
6.6	Critérios de adequação para Dados III	48
6.7	Dados IV: Rebarbas (em milímetros).	49
6.8	Critérios de adequação para Dados IV	50
6.9	Dados V: resistência de fibras de carbono (em GPA).	51
6.10	Critérios de adequação para Dados V	52
7.1	Taxas de acertos para o cenário 1.	55
7.2	Taxas de acertos para o cenário 2.	58
7.3	Dados de tempo de entrega.	60
7.4	Resumo dos dados.	61
7.5	Critérios de adequação para os modelos.	64

Lista de Símbolos

$(\Omega, \mathcal{F}, \mathcal{P})$	Espaço de probabilidade, p. 2
G_c^{inf}	Distribuição generalizada inf com distribuição de base G, p. 12
G_c^{sup}	Distribuição generalizada sup com distribuição de base G, p. 12
Θ	Espaço paramétrico, p. 7
Θ_0	Espaço paramétrico sob a hipótese nula, p. 7
Θ_1	Espaço paramétrico sob a hipótese alternativa, p. 7
α	Nível de significância ou Nível nominal ou Erro tipo I, p. 7
β	Erro tipo II, p. 7
$\lambda^*(\mathbf{x})$	Estatística da razão de verossimilhanças, p. 8
\mathcal{F}	Álgebra de eventos, p. 2
\mathbf{A}_1^*	Região crítica da verdadeira distribuição de $\lambda^*(\mathbf{x})$, p. 7
$L(\cdot)$	Função de verossimilhança, p. 6
Ω	Espaço amostral, p. 2
$\ell(\cdot)$	Função de log-verossimilhança, p. 6
\mathcal{H}_0	Hipótese nula, p. 7
\mathcal{H}_1	Hipótese alternativa, p. 7
\mathcal{P}	Medida de probabilidade, p. 2

Lista de Abreviaturas

f.d.a	função de distribuição acumulada, p. 3
f.d.p	função densidade de probabilidade, p. 3
v.a	variável aleatória, p. 2
EMV	Estimador de Máxima Verossimilhança, p. 6
MMV	Método de Máxima Verossimilhança, p. 6
TRV	Teste da Razão de Verossimilhanças, p. 8

Capítulo 1

Introdução

Diversas pesquisas científicas realizadas em Estatística, Economia, Biologia, Engenharia e várias outras áreas têm seus problemas de estudo ligados a situações práticas, que normalmente podem ser explicadas através de modelos, sendo que estes são os principais instrumentos utilizados e constituem uma versão simplificada de tais situações. Em particular, na estatística, é comum a representação de um fenômeno em estudo por um modelo probabilístico.

Na maioria das situações, o pesquisador se depara com mais de um modelo descrevendo um mesmo fenômeno. Portanto, é necessário ter um critério padrão para escolher o modelo mais adequado. De acordo com [Burnham e Anderson \(2004\)](#), é importante que a seleção de modelos seja baseada em princípios científicos. O princípio científico fundamentado na parcimônia, requer que, dentre todos os modelos que expliquem bem os dados, o mais simples deva ser escolhido.

Dessa forma, a seleção de modelos tem ganhado destaque nas pesquisas e a literatura já dispõe de vários critérios de seleção de modelos, cujo objetivo é selecionar um modelo de um conjunto de modelos concorrentes que melhor explique o fenômeno em estudo.

Na literatura, há diversas abordagens com esse objetivo. Algumas, são baseadas no máximo da função de verossimilhança, tais como critério de informação de Akaike (AIC) ([AKAIKE, 1973](#)), critério de informação bayesiano (BIC) ([SCHWARZ, 1978](#)), dentre outros.

Outras abordagens envolvem problemas que são tratados com a hipótese que os dados são retirados de uma população com uma distribuição de probabilidade específica, tal como o critério de Anderson-Darling (A^2) ([ANDERSON e DARLING, 1954](#)), em que é considerado uma amostra aleatória X_1, X_2, \dots, X_n de uma população com função de distribuição acumulada contínua F_X . O objetivo é testar a hipótese nula $\mathcal{H}_0 : F_X = G$ contra a hipótese alternativa $\mathcal{H}_1 : F_X \neq G$, em que G é uma função de distribuição acumulada imposta e assim conhecida e F_X é a distribuição real dos dados que é geralmente desconhecida.

Esses critérios têm sido estudados e aplicados extensivamente na literatura, como: [KIM *et al.* \(2019\)](#) fornecem uma visão geral sobre o desempenho dos critérios AIC, BIC e Vuong ([VUONG, 1989](#)) na seleção de modelos para os dados de violação de tráfego em Taiwan; [Emiliano \(2013\)](#) utiliza e avalia o desempenho dos critérios AIC, AICc e BIC na seleção de modelos normais, séries temporais e modelos de crescimento; [Yamanouchi \(2017\)](#) avaliou o desempenho dos critérios de informação, AIC, BIC e de Kullback - KIC ([CAVANAUGH, 1999](#)) na seleção de modelos mistos (efeitos fixos, aleatórios e a estrutura de covariâncias) a partir de um conjunto de dados de pesos semanais de frangos num período de 7 semanas e [ABREU *et al.* \(2018\)](#) utilizaram os testes de Kolmogorov-Smirnov (KS), Qui-quadrado (χ^2), Filliben (Fi) e Anderson-Darling (A^2) para selecionar a função de distribuição de probabilidade (fdp) que melhor se ajusta em séries de precipitação diária máxima anual de 11 estações localizadas na bacia hidrográfica do rio Sapucaí.

Outros estudos também realizados na inferência clássica impõem famílias de distribuições de probabilidades como possíveis candidatas para a modelagem do fenômeno de interesse e deseja-se decidir entre distribuições pertencentes a essas famílias, qual a que melhor se ajusta aos dados através de um critério de seleção de modelos.

Sendo assim, esta pesquisa tem como objetivo propor um novo critério de seleção de modelos para modelos não encaixados, baseado nas famílias de distribuições generalizadas G^{sup} e G^{inf} e suas propriedades e comparar seu desempenho com os seguintes critérios: critério de informação de Akaike (AIC), critério de informação de Akaike corrigido (AICc), critério de informação bayesiano (BIC), critério de informação de Hannan-Quinn (HQIC) e os critérios de adequação de ajuste modificados de Crámer-Von Mises (W^*) e Anderson-Darling (A^*).

Este trabalho está organizado em oito capítulos. No Capítulo 2 apresentamos uma revisão sobre alguns conceitos necessários para o desenvolvimento do tema. Na Seção 2.1 abordamos conceitos básicos na teoria das probabilidades. Na Seção 2.2 realizamos um breve estudo sobre resultados na inferência estatística. Na Seção 2.3 apresentamos um breve estudo sobre análise de regressão linear múltipla e na Seção 2.4 são apresentadas as medidas mais utilizadas para avaliar um modelo de regressão.

No Capítulo 3 fizemos uma revisão de alguns dos critérios de seleção de modelos mais usados na literatura, os quais foram utilizados para comparar o seu desempenho com o novo critério proposto.

No Capítulo 4 abordamos as famílias de distribuições sup e inf. Na Seção 4.1 apresentamos as famílias de distribuições sup e inf e suas propriedades ([SOUSA, 2020](#); [TABLADA, 2017](#)) as quais darão suporte a nossa proposta de trabalho. Na Seção 4.2 apresentamos a função de log-verossimilhança das famílias sup e inf e os

resultados limites de tais funções. Na Seção 4.3 apresentamos a proposta do novo critério de seleção de modelos baseado nessas famílias de distribuições.

No Capítulo 5, tendo em vista validar nossa proposta, objetivamos avaliar o desempenho do critério proposto na seleção do modelo mais adequado para representar os dados e comparar com os critérios clássicos apresentados no Capítulo 3. Para isso, realizamos simulações de Monte Carlo utilizando a plataforma `0x`.

No Capítulo 6, comparamos o desempenho do critério proposto com os critérios apresentados no Capítulo 3 através de conjuntos de dados reais. São apresentadas cinco aplicações com conjuntos de dados considerados pequenos ou moderados.

Tendo em vista a grande importância do estudo de Regressão Linear e sua aplicabilidade em diversas áreas, no Capítulo 7 realizamos simulações de Monte Carlo com o objetivo de avaliar o desempenho do critério proposto na seleção de modelos de regressão linear múltipla e comparar com o desempenho de outros critérios apresentados nesse trabalho. Em seguida, comparamos o desempenho do novo critério na seleção de modelos de regressão linear múltipla aplicados a um conjunto de dados reais.

Finalmente, no Capítulo 8 são dadas as considerações finais deste trabalho.

Capítulo 2

Preliminares

2.1 Conceitos Básicos na Teoria da Probabilidade

A teoria da probabilidade consiste em estudar fenômenos do cotidiano através de experimentos aleatórios. Os resultados de um experimento aleatório são caracterizados pelas seguintes componentes:

1. Espaço amostral: O conjunto de resultados possíveis, denotado por Ω ;
2. σ -álgebra de eventos: A coleção de conjuntos de resultados de interesse, representada por \mathcal{F} ;
3. Medida de probabilidade: Uma função $P : \mathcal{F} \rightarrow [0, 1]$ que mede a verossimilhança ou probabilidade de ocorrência de cada um dos conjuntos de resultados de interesse.

A terna (Ω, \mathcal{F}, P) é chamada de espaço de probabilidade. Para detalhes sobre propriedades e axiomas consultar [Magalhães \(2006\)](#).

Antes da realização de um experimento aleatório, não sabemos seu resultado. No entanto, seu espaço de probabilidade pode ser estabelecido de forma conveniente, de modo a avaliar a probabilidade de qualquer evento de interesse da σ -álgebra. Sendo assim, podemos também atribuir probabilidades às funções desses eventos dando origem ao conceito de variável aleatória.

Definição 2.1 (Variável Aleatória). *Seja (Ω, \mathcal{F}, P) um espaço de probabilidade. Denominamos de variável aleatória (v.a) qualquer função $X : \Omega \rightarrow \mathbb{R}$ tal que*

$$X^{-1}(I) = \{\omega \in \Omega : X(\omega) \in I\} \in \mathcal{F},$$

para todo intervalo $I \subset \mathbb{R}$. É comum representar o evento $\{\omega \in \Omega : X(\omega) \in I\}$ por $[X \in I]$.

Em outras palavras, vemos que uma variável aleatória é uma função do espaço amostral Ω nos reais, para a qual é possível calcular a probabilidade de ocorrência dos eventos $[X \in I]$, para todo intervalo I .

Definição 2.2 (Função de Distribuição). *Seja X uma v.a em (Ω, \mathcal{F}, P) , sua função de distribuição é definida por*

$$F_X(x) = P(X \in (-\infty, x]) = P(X \leq x),$$

com x percorrendo todos os reais. Por simplicidade de notação, denotaremos F_X apenas por F .

Segundo Magalhães (2006), é importante conhecer a função de distribuição, pois ela nos permite obter qualquer informação sobre uma v.a. Mesmo que a variável só assuma valores num subconjunto dos reais, a função de distribuição é definida em toda a reta. Alguns autores se referem a ela como função de distribuição acumulada, por acumular as probabilidades dos valores inferiores a x .

Proposição 2.1 (Propriedades da Função de Distribuição). *Uma função de distribuição de uma v.a X obedece às seguintes propriedades:*

$$(F1) \lim_{x \rightarrow -\infty} F(x) = 0 \text{ e } \lim_{x \rightarrow \infty} F(x) = 1;$$

(F2) F é contínua à direita;

(F3) F é não decrescente, isto é, $F(x) \leq F(y)$, sempre que $x \leq y, \forall x, y \in \mathbb{R}$.

Observação 2.1. *Uma função $F : \mathbb{R} \rightarrow \mathbb{R}$ que satisfaz (F1), (F2), (F3) é a função de distribuição de alguma variável aleatória X (JAMES, 1981).*

A classificação das variáveis aleatórias é feita de acordo com os valores que assumem. As que assumem valores em um conjunto finito ou enumerável são chamadas discretas e aquelas que assumem valores em um intervalo da reta real são chamadas contínuas. As últimas são as de interesse nesse trabalho.

Definição 2.3 (Função Densidade de Probabilidade). *Uma v.a X com função de distribuição F , será classificada como contínua, se existir uma função não negativa f tal que:*

$$F(x) = \int_{-\infty}^x f(\omega) d\omega, \text{ para todo } x \in \mathbb{R}.$$

A função f é denominada função densidade de probabilidade (f.d.p).

Proposição 2.2 (Propriedades da Função Densidade). *A função densidade de X satisfaz:*

(fd1) $f(x) \geq 0, \forall x \in \mathbb{R};$

(fd2) $\int_{-\infty}^{\infty} f(\omega) d\omega = 1.$

Dada a relação entre a função de distribuição e a função de densidade, expressa na definição 2.3 temos que, pelo teorema fundamental do cálculo, obtém-se a função de densidade derivando F em relação à x e dada a f.d.p, a função de distribuição segue por integração.

Definição 2.4 (Função Quantílica). *Seja F uma função de distribuição qualquer. A função quantílica da distribuição F é definida da seguinte forma:*

$$Q(y) = F^{-1}(y) = \inf\{x \in \mathbb{R} : F(x) \geq y\}.$$

Observe que, se a inversa de F existe no sentido usual, então ela coincide com a função quantílica.

Proposição 2.3. *Seja X uma variável aleatória contínua com função de distribuição F . Então $Y = F(X)$ tem distribuição Uniforme Contínua em $(U [0, 1])$. Vale a recíproca, sendo $U \sim U [0, 1]$, então $X = F^{-1}(U)$ tem função de distribuição F .*

Se X é uma v.a com função quantílica conhecida, poderemos fazer uso dela para gerar observações de X , bastando, assim, gerar observações de uma v.a. Uniforme. Quando possível, gerar observações de X por meio da função quantílica geralmente tornará as simulações menos custosas quando comparadas a outros métodos computacionais para geração de números pseudo-aleatórios.

A realização de um experimento aleatório pode envolver mais de uma variável de interesse. A seguinte definição generaliza a ideia de v.a. para esses casos.

Definição 2.5 (Vetor Aleatório). *Um vetor $\mathbf{X} = (X_1, X_2, \dots, X_n)$ cujos componentes são variáveis aleatórias definidas no mesmo espaço de probabilidade (Ω, \mathcal{F}, P) , é chamado vetor aleatório (ou variável aleatória n -dimensional).*

Sua função de distribuição é denotada por $F = F_{\mathbf{X}}$ e definida por

$$F_{\mathbf{X}}(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n),$$

para todo (x_1, x_2, \dots, x_n) pertencente a \mathbb{R}^n . $F_{\mathbf{X}}$ é chamada de função de distribuição conjunta das variáveis aleatórias X_1, X_2, \dots, X_n .

Um conceito importante em Probabilidade e Estatística é o de independência entre variáveis aleatórias, já que muitas técnicas estatísticas só podem ser aplicadas se o dados forem independentes.

Definição 2.6 (Independência de Variáveis Aleatórias). *Dizemos que as variáveis X_1, X_2, \dots, X_n , definidas em um mesmo espaço de probabilidade (Ω, \mathcal{F}, P) , são independentes, se tivermos:*

$$F_{\mathbf{X}}(x_1, x_2, \dots, x_n) = \prod_{j=1}^n F_{\mathbf{X}_j}(x_j), \quad \forall (x_1, x_2, \dots, x_n) \in \mathbb{R}^n,$$

em que $\mathbf{X} = (X_1, X_2, \dots, X_n)$ e para $j = 1, \dots, n$, $F_{\mathbf{X}_j}$ é uma função de distribuição de \mathbf{X}_j , chamada também de função de distribuição marginal da variável \mathbf{X}_j .

Em outras palavras, ocorre que as variáveis aleatórias X_1, X_2, \dots, X_n são independentes se sua função de distribuição conjunta fatora no produto das funções de distribuição marginais.

Proposição 2.4. *Sejam X_1, X_2, \dots, X_n variáveis aleatórias i.i.d (independentes e identicamente distribuídas), isto é, além de independentes, possuem a mesma função de distribuição $F(x)$. As expressões da função de distribuição de $X_{(1)} = \min(X_1, X_2, \dots, X_n)$ e $X_{(n)} = \max(X_1, X_2, \dots, X_n)$ são dadas, nessa ordem, por*

$$F_{X_{(1)}}(x) = P(X_{(1)} \leq x) = 1 - [1 - F(x)]^n$$

e

$$F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = [F(x)]^n.$$

2.2 Conceitos Básicos na Teoria da Inferência Estatística

No estudo da inferência estatística, o objetivo principal é obter informações a respeito de uma população a partir das informações de uma amostra. Isso pode ser realizado através de: estimação pontual, estimação intervalar e teste de hipóteses. Dentre os métodos de estimação pontual propostos na literatura, o que mais se destaca é o Método de Máxima Verossimilhança (MMV). O MMV permite obter estimadores para os parâmetros que maximizam a função de verossimilhança, ou seja, de maneira a maximizar a probabilidade que uma dada amostra venha ser obtida. Esses estimadores possuem propriedades atrativas na construção de intervalos de confiança e testes de hipóteses.

Definição 2.7. *Sejam X_1, X_2, \dots, X_n variáveis aleatórias i.i.d em que cada variável tem função de densidade de probabilidade $f(x; \theta)$, sendo θ um vetor p -dimensional de parâmetros, ou seja, $\theta = (\theta_1, \dots, \theta_p)^\top$ pertencente ao espaço paramétrico Θ . Sejam x_1, x_2, \dots, x_n os valores observados de X_1, X_2, \dots, X_n . A função de verossimilhança é uma função do vetor de parâmetros θ , correspondente à amostra observada dada por*

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n f(x_i; \theta) \tag{2.1}$$

em que $\mathbf{x} = (x_1, x_2, \dots, x_n)^\top$.

Definição 2.8. O Estimador de máxima verossimilhança (EMV) do parâmetro θ é o vetor $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_p)^\top \in \Theta$ que maximiza a função de verossimilhança $L(\theta; \mathbf{x})$, se esse vetor existir.

O EMV pode ser obtido como solução da equação

$$\frac{\partial L(\theta; \mathbf{x})}{\partial \theta} = 0. \quad (2.2)$$

A função de log-verossimilhança é definida por

$$\ell(\theta; \mathbf{x}) = \log(L(\theta; \mathbf{x})).$$

Observe que o EMV também maximiza a função de log-verossimilhança. Em alguns exemplos simples, a solução da equação 2.2 pode ser obtida explicitamente. Em situações mais complicadas, a solução será obtida por procedimentos numéricos (BOLFARINE e SANDOVAL, 2000).

Em muitas situações, existe um particular interesse em verificar determinada afirmação a respeito de um parâmetro populacional desconhecido. Essa afirmação é denominada de hipótese estatística. A ideia é que a partir de uma amostra retirada da população, será estabelecida uma regra de decisão segundo a qual rejeita-se ou aceita-se a hipótese proposta. Esta regra de decisão é estabelecida por um teste de hipóteses. Normalmente existe uma hipótese que é a mais importante para o pesquisador, a qual é denotada por \mathcal{H}_0 e é denominada de hipótese nula. Qualquer outra hipótese diferente de \mathcal{H}_0 será chamada de hipótese alternativa, sendo denotada por \mathcal{H}_1 .

Seja $\mathbf{X} = (X_1, \dots, X_n)^\top$ uma amostra aleatória tomada de uma distribuição que envolve um parâmetro θ desconhecido, definido em um espaço paramétrico Θ . Assim, as hipóteses podem ser definidas como $\mathcal{H}_0 : \theta \in \Theta_0$ e $\mathcal{H}_1 : \theta \in \Theta_1$ em que Θ_0 e Θ_1 são subconjuntos disjuntos de Θ , ou seja, $\Theta_0 \cup \Theta_1 = \Theta$, $\Theta_0 \cap \Theta_1 = \emptyset$, em que $\Theta_0 \neq \emptyset$ e $\Theta_1 \neq \emptyset$. O teste é especificado particionando o espaço amostral em dois subconjuntos. Segundo Casella e Berger (2002), o subconjunto que contem os valores de \mathbf{X} para os quais \mathcal{H}_0 será rejeitada é chamado de região de rejeição ou *região crítica*. O complemento da região de rejeição é chamado de região de aceitação. Com isso, um teste fica bem determinado quando especificado sua região crítica. Se o subconjunto Θ_i , $i = 0, 1$ contem um único valor, então \mathcal{H}_i é uma hipótese simples. Caso contrário, \mathcal{H}_i é uma hipótese composta.

Em todo procedimento de teste de hipóteses, cometemos dois tipos de erros, sendo eles erro tipo I e erro tipo II, denotados por α e β , respectivamente. O erro tipo I ocorre quando rejeitamos \mathcal{H}_0 , sendo \mathcal{H}_0 verdadeira, enquanto o erro tipo II ocorre quando não rejeitamos \mathcal{H}_0 , sendo \mathcal{H}_0 falsa. A Tabela 2.1 abaixo ilustra a situação descrita,

Tabela 2.1: Tipos de erros em testes de hipóteses.

Decisão	\mathcal{H}_0 é verdadeira	\mathcal{H}_0 é falsa
Não rejeitar \mathcal{H}_0	Decisão Correta	Erro tipo II (β)
rejeitar \mathcal{H}_0	Erro tipo I (α)	Decisão Correta

Além disso, α é chamado de nível de significância do teste ou nível nominal, o qual deve ser especificado antes de iniciar o procedimento do teste. Ao especificar α , o pesquisador indica o quanto está disposto a cometer o erro tipo I.

Definido o nível de significância α , podemos especificar a região crítica do teste.

Definição 2.9 (Poder do teste). *O poder do teste é a probabilidade de rejeitar corretamente a hipótese nula quando esta é falsa, ou seja*

$$\pi(\theta) = 1 - \beta = 1 - P(\text{erro tipo II}).$$

Definição 2.10 (Função do Poder). *A função $\Pi(\theta)$ é chamada função de poder do teste. Assim, denotada por \mathbf{A}_1^* a região crítica do teste, a função de poder é definida como*

$$\Pi(\theta) = P_\theta(\mathbf{x} \in \mathbf{A}_1^* | \theta), \forall \theta \in \Theta,$$

Note que $\Pi(\theta_0) = \alpha$, em que θ_0 é vetor de parâmetros especificado em \mathcal{H}_0 .

Um procedimento utilizado em muitos casos, sem muita dificuldade, é o Teste da Razão de Verossimilhanças (TRV) (LEHMANN e ROMANO, 2006; ROUSSAS, 2003).

Definição 2.11 (TRV). *O TRV para testar $\mathcal{H}_0 : \theta \in \Theta_0$ versus $\mathcal{H}_1 : \theta \in \Theta_1$ pode ser definido como o teste com região crítica dada por*

$$\mathbf{A}_1^* = \left\{ \mathbf{x}; \lambda^*(\mathbf{x}) = \frac{\sup_{\theta \in \Theta_0} L(\theta; \mathbf{x})}{\sup_{\theta \in \Theta} L(\theta; \mathbf{x})} \leq c \right\}, \quad (2.3)$$

em que $\lambda^*(\mathbf{x})$ é conhecida como a estatística da razão de verossimilhanças.

Note que $0 \leq \lambda^*(\mathbf{x}) \leq 1$, pois o numerador é o valor atingido pela função de verossimilhança quando θ pertence a um subconjunto de Θ_0 , enquanto que o denominador é o valor máximo atingido quando θ pertence a todo espaço paramétrico Θ . Se \mathcal{H}_0 for “verdadeira” é esperado que $\lambda^*(\mathbf{x})$ esteja “próximo” de 1 e se \mathcal{H}_0 for “falsa” espera-se que o denominador seja de ordem maior que o numerador, e $\lambda^*(\mathbf{x})$ seja “próximo” de zero.

O valor de c em (2.3) pode ser obtido resolvendo a equação

$$\sup_{\theta \in \Theta_0} P(\lambda^*(\mathbf{x}) \leq c) = \alpha.$$

Para isso é preciso conhecer a distribuição da estatística $\lambda^*(\mathbf{x})$ que, em geral, não é fácil de ser obtida. Note que, c é o quantil α com base na distribuição de $\lambda^*(\mathbf{x})$. Logo, a hipótese nula é rejeitada se o valor da estatística $\lambda^*(\mathbf{x})$ baseado nos dados amostrais \mathbf{x} for menor ou igual ao quantil α obtido através da distribuição de $\lambda^*(\mathbf{x})$.

O seguinte resultado, é de ajuda nos casos em que não conhecemos a verdadeira distribuição de $\lambda^*(\mathbf{x})$. É aplicado quando o teste de hipóteses é feito com base em uma amostra de tamanho grande.

Teorema 2.1. *Suponha que o vetor paramétrico está particionado como $\theta = (\psi_1^T, \psi_2^T)^T$, em que $\dim(\psi_1) + \dim(\psi_2) = \dim(\theta)$. Aqui, ψ_1 representa o vetor de parâmetros de interesse e ψ_2 o vetor de parâmetros de perturbação. A estatística da razão de verossimilhanças (RV) para testar a hipótese nula $H_0 : \psi_1 = \psi_1^{(0)} \times H_1 : \psi_1 \neq \psi_1^{(0)}$ é dada por*

$$\begin{aligned} RV = -2 \log(\lambda^*(\mathbf{x})) &= -2 \log \left(\frac{L(\tilde{\theta})}{L(\hat{\theta})} \right) = -2 \log(L(\tilde{\theta})) - \log(L(\hat{\theta})) \\ &= 2\ell(\hat{\theta}) - \ell(\tilde{\theta}), \end{aligned}$$

em que $\tilde{\theta} = (\psi_1^{(0)\top}, \tilde{\psi}_2^\top)^\top$ e $\hat{\theta} = (\hat{\psi}_1^\top, \hat{\psi}_2^\top)^\top$ são os estimadores de máxima verossimilhança sob a hipótese nula considerando todo o espaço paramétrico, respectivamente. Baseado na teoria assintótica de primeira ordem, sabe-se que a estatística RV segue distribuição assintoticamente qui-quadrado \mathcal{X}_k^2 com k graus de liberdade, em que $k = \dim(\psi_1)$.

A prova deste resultado pode ser consultada em Severini (2000, p. 114-115).

Desta forma, através do TRV, dado na definição 2.11, a região crítica para testar $H_0 : \theta \in \Theta_0 \times H_1 : \theta \in \Theta_1$ pode ser dada por

$$\mathbf{B}_1^* = \{\mathbf{x}; RV > q\}.$$

Assim, dado um nível de significância α , é plausível decidir rejeitar H_0 se $RV > q$ em que a constante q é obtida de modo que

$$P(RV > q) = \alpha.$$

Observe que, fixado uma probabilidade α , o número q é o quantil $1 - \alpha$ obtido da distribuição qui-quadrado com k graus de liberdade, em que k é a dimensão do vetor de parâmetros sob a hipótese nula.

2.3 Análise de regressão linear múltipla

Ao criar o termo regressão, Francis Galton verificou que a altura dos filhos de pais mais altos ou mais baixos que o padrão tende a mover-se ou “regredir” no sentido da

altura média da população. Essa lei foi confirmada por Karl Pearson, que constatou que filhos de pais altos e baixos “regrediam” igualmente à altura média de todos os homens.

De um modo geral, segundo [Gujarati e Porter \(2011\)](#) a análise de regressão linear diz respeito ao estudo da dependência de uma variável, a variável dependente, em relação a uma ou mais variáveis, as variáveis explicativas, visando estimar e ou prever o valor médio da variável dependente em termos dos valores conhecidos ou fixados das variáveis explicativas.

Quando estudamos a dependência de uma variável em relação a uma única variável explicativa, como é o caso descrito das alturas dos filhos em relação à altura dos pais, esse estudo é conhecido como análise de regressão linear simples. Quando estudamos a dependência de uma variável em relação a mais de uma variável explicativa, temos o estudo da análise de regressão linear múltipla.

Nessa seção, será apresentado um resumo sobre análise de regressão linear múltipla. Os resultados são extensões do estudo da análise de regressão linear simples.

2.3.1 Modelo teórico

No modelo de regressão linear múltipla, a variável dependente y está relacionada a k variáveis explicativas na seguinte forma:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n. \quad (2.4)$$

em que

- y_i representa o valor da variável dependente (ou variável resposta) na observação i , $i = 1, 2, \dots, n$;
- $x_{i1}, x_{i2}, \dots, x_{ik}$ são os valores da i -ésima observação das k variáveis explicativas;
- $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ são os parâmetros ou coeficientes de regressão.
- ε_i , $i = 1, 2, \dots, n$ correspondem aos erros aleatórios.

Esse modelo descreve um hiperplano no espaço k -dimensional referente às variáveis explicativas.

Os parâmetros β_j , $j = 1, \dots, k$, representam o impacto na média esperada da variável dependente, y , quando a variável x_j , $j = 1, \dots, k$ sofre uma mudança unitária enquanto as variáveis restantes x_i ($i \neq j$) são mantidas constantes.

O parâmetro β_0 no modelo de regressão linear simples corresponde ao intercepto da linha de regressão, aqui, ele corresponde ao intercepto do plano de regressão. Se a abrangência do modelo incluir $x_j = 0$, $j = 1, \dots, k$, então β_0 será a média de y nesse ponto. Caso contrário, não existe uma interpretação prática para β_0 .

2.3.2 Pressupostos do modelo

Os pressupostos para o modelo de regressão linear múltipla são análogos ao do modelo de regressão linear simples. Assim, tem-se:

- a) Os erros aleatórios são independentes, ou seja, a covariância entre cada variável explicativa e ε é igual a zero,

$$\text{cov}(X_i, \varepsilon_i) = 0, \quad i = 1, 2, \dots, n. \quad (2.5)$$

- b) O valor médio ou esperado, dos erros aleatórios é zero,

$$E(\varepsilon_i) = 0, \quad i = 1, 2, \dots, n. \quad (2.6)$$

- c) A variância dos erros aleatórios é constante,

$$\text{var}(\varepsilon_i) = \sigma^2, \quad i = 1, \dots, n. \quad (2.7)$$

- d) Dos pressupostos anteriores, tem-se que o erro aleatório têm distribuição normal, isto é,

$$\varepsilon_i \sim N(0, \sigma^2), \quad i = 1, \dots, n.$$

Desses pressupostos, temos por consequência que y tem distribuição normal com variância σ^2 e, para o caso de modelo definido em 2.4,

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k.$$

2.3.3 Representação matricial do método de regressão linear múltipla

Segundo [Montgomery, Peck, e Vining \(2012\)](#) é mais conveniente lidar com modelos de regressão múltipla se eles forem expressos em notação matricial. Isso permite uma exibição muito compacta do modelo, dados e resultados.

Para isso, consideremos a expressão geral da i -ésima observação no modelo de regressão linear dada por

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n.$$

Em notação matricial esse modelo é dado por

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.8)$$

em que

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

sendo que

- \mathbf{y} é um vetor de dimensão $n \times 1$ cujas componentes correspondem às n respostas, $y_1, y_2 \dots y_n$, constituído pelas observações da variável dependente.
- \mathbf{X} é a matriz de dimensão $n \times (k + 1)$ denominada matriz do modelo.
- $\boldsymbol{\beta}$ é um vetor coluna de dimensão $(k + 1) \times 1$ cujos elementos são os coeficientes de regressão.
- $\boldsymbol{\varepsilon}$ é um vetor coluna de dimensão $n \times 1$ cujas componentes são os erros aleatórios.

Como $\boldsymbol{\varepsilon}$ segue distribuição normal n -variada, tem-se que $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I_n)$, em que $\mathbf{0}$ representa o vetor nulo e I_n a matriz identidade de ordem n , assim \mathbf{y} tem distribuição normal com $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ e a matriz de variâncias-covariâncias $cov(\mathbf{y}) = \sigma^2 I_n$, ou seja,

$$\mathbf{y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 I_n).$$

2.3.4 Estimação dos parâmetros do modelo

Na análise de regressão linear, para encontrar o vetor de estimadores $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$, os métodos mais usados são os de mínimos quadrados ordinários (MQO) e o de máxima verossimilhança (MV).

Para encontrar o vetor de estimadores $\hat{\boldsymbol{\beta}}$ pelo método MQO, deve-se maximizar, em relação a $\boldsymbol{\beta}$, a função

$$\begin{aligned} S(\boldsymbol{\beta}) &= \sum_{i=1}^n \varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}'\mathbf{y} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}. \end{aligned} \tag{2.9}$$

Dessa forma, o estimador de MQO

$$\left. \frac{\partial S}{\partial \boldsymbol{\beta}} \right|_{\hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = 0.$$

Simplificando temos

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}. \quad (2.10)$$

As equações 2.10 são as equações normais de mínimos quadrados. Para resolver as equações normais, multipliquemos ambos os lados pelo inverso de $\mathbf{X}'\mathbf{X}$. Assim, o vetor de estimadores de mínimos quadrados de $\boldsymbol{\beta}$ é

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (2.11)$$

Vale ressaltar, que a matriz $(\mathbf{X}'\mathbf{X})^{-1}$ sempre existirá se as variáveis explicativas forem linearmente independentes.

Com isso, de 2.8 podemos concluir que o modelo de regressão linear ajustado é

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}.$$

A diferença entre o valor observado y_i e o valor ajustado correspondente \hat{y}_i é o resíduo $e_i = y_i - \hat{y}_i$. Os resíduos podem ser escritos em notação matricial como

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}.$$

2.3.5 Propriedades dos estimadores de mínimos quadrados

Temos as seguintes propriedades estatísticas do estimador de mínimos quadrados $\hat{\boldsymbol{\beta}}$:

a) O valor esperado,

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}] \\ &= \boldsymbol{\beta} \end{aligned}$$

pois $E(\boldsymbol{\varepsilon}) = 0$.

b) A matriz de covariância é expressa por

$$Cov(\hat{\boldsymbol{\beta}}) = E \left\{ [\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})][\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})]' \right\}$$

que é uma matriz simétrica $(k+1) \times (k+1)$ cujo j -ésimo elemento da diagonal é a variância de $\hat{\beta}_j$ e cujo (ij) -ésimo elemento fora da diagonal é a covariância entre $\hat{\beta}_i$ e $\hat{\beta}_j$. A matriz de covariância de $\hat{\beta}$ é encontrada aplicando um operador de variância a $\hat{\beta}$ da seguinte forma

$$Cov(\hat{\beta}) = Var(\hat{\beta})$$

assim

$$\begin{aligned} Var(\hat{\beta}) &= Var[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Var(\mathbf{y})[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']' \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

2.3.6 Estimador de σ^2

Podemos obter um estimador de σ^2 considerando a soma do quadrado dos resíduos, que é definido por

$$SS_{Res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2 = \mathbf{e}'\mathbf{e}$$

substituindo $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\beta}$ e sabendo que $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, obtemos

$$\begin{aligned} SS_{Res} &= (\mathbf{y} - \mathbf{X}\hat{\beta})' (\mathbf{y} - \mathbf{X}\hat{\beta}) \\ &= [\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}]' [\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] \\ &= \mathbf{y}' [I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \mathbf{y} \end{aligned}$$

é possível mostrar que a matriz $[I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']$ é simétrica idempotente. Dessa forma,

$$\frac{SS_{Res}}{\sigma^2} = \frac{1}{\sigma^2} \mathbf{y}' [I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \mathbf{y}$$

segue uma distribuição qui-quadrado com $n - (k+1)$ graus de liberdade ($\chi_{n-(k+1)}^2$) associados a ela,

$$\frac{SS_{Res}}{\sigma^2} \sim \chi_{n-(k+1)}^2.$$

Portanto, um estimador não viesado para σ^2 é dado por

$$\hat{\sigma}^2 = \frac{SS_{Res}}{n - k - 1}.$$

e conforme pode ser observado no caso de regressão linear simples, este estimador

de σ^2 é dependente do modelo.

2.3.7 Estimadores de Máxima Verossimilhança

Para obter os estimadores de máxima verossimilhança de $\boldsymbol{\beta}$ e σ^2 , são considerados os mesmos pressupostos para os estimadores MQO e o modelo 2.8. Assim, tem-se a função de densidade normal para os erros

$$f(\varepsilon_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}\varepsilon_i^2\right],$$

portanto, a função de verossimilhança é

$$L(\boldsymbol{\varepsilon}, \boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^n f(\varepsilon_i) = \frac{1}{\sigma^n(2\pi)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\right].$$

Substituindo $\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ na função de verossimilhança, obtemos

$$L(\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^n f(\varepsilon_i) = \frac{1}{\sigma^n(2\pi)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right],$$

Assim, segue a função de log-verossimilhança dada por,

$$\ell(\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Minimizando a função de log-verossimilhança, obtemos $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ equivalente ao estimador de MQO. O estimador de máxima verossimilhança de σ^2 é

$$\tilde{\sigma}^2 = \frac{SS_{Res}}{n}$$

que é um estimador viesado para σ^2 .

2.4 Medidas para avaliar modelos de regressão

A seleção de variável na análise de regressão envolve gerar modelos de determinados subconjuntos de variáveis explicativas e decidir se um modelo é melhor que outro. A seguir serão apresentadas as duas medidas mais utilizadas.

2.4.1 Coeficiente de determinação múltiplo - R^2

Essa é uma medida bastante utilizada e que fornece a proporção da variação total da variável dependente y que é explicada pelo modelo. É denotado por

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_{Res}}{SS_T}$$

em que $SS_R = \sum_{i=1}^n \hat{y}_i^2$ é a soma dos quadrados explicados e $SS_T = \sum_{i=1}^n y_i^2$ a soma total dos quadrados.

Note que R^2 apresenta as seguintes propriedades:

1. O coeficiente R sempre será positivo.
2. Tem-se que $0 \leq R^2 \leq 1$. Se $R^2 = 0$, diz-se que não há relação entre a variável dependente e as variáveis explicativas, se $R^2 = 1$ indica que a variável dependente é totalmente explicada pelo modelo.

Com relação ao fato de que quanto mais próximo R^2 situar-se de 1 a variável dependente é explicada pelo modelo, é necessário atenção, pois R^2 é uma função não decrescente em relação ao número de variáveis explicativas, ou seja, a medida que o número de variáveis explicativas (mesmo com pouco poder explicativo sobre y) aumenta, R^2 também aumenta. Dessa forma, ao compararmos dois modelos com a mesma variável dependente, mas com o número de variáveis explicativas diferentes, podemos ser levados a escolher o modelo com o R^2 maior, independente da qualidade de ajuste.

2.4.2 Coeficiente de determinação múltiplo ajustado - \bar{R}^2

Vimos que quanto maior o número de variáveis independentes no modelo, maior será o valor de R^2 . Logo, uma medida que corrige esta limitação se faz necessária. Portanto, o coeficiente de determinação múltiplo ajustado, \bar{R}^2 , corrige este problema, pois “penaliza” o R^2 à medida que as variáveis explicativas adicionais são incorporadas ao modelo. É definido por

$$\bar{R}^2 = 1 - \left(\frac{n-1}{n-(k+1)} \right) (1 - R^2).$$

Note que, ao contrário de R^2 , o coeficiente de determinação ajustado aumenta menos quando o número de variáveis X aumenta. Aliás, se forem adicionadas variáveis com pouco poder explicativo, \bar{R}^2 tende a decrescer. Assim, quando existir uma diferença significativa entre os dois coeficientes, provavelmente foram incluídas no modelo variáveis estatisticamente não significativas.

Embora o \bar{R}^2 não seja um critério de seleção de modelos, ele é usado juntamente com outros critérios para avaliar a adequação de um modelo de regressão. Alguns desses critérios estão descritos no Capítulo 3 e são utilizados nesse trabalho.

Capítulo 3

Critérios de seleção de modelos: Revisão

É fato, do ponto de vista estatístico, que a escolha do modelo mais adequado para a representação de um fenômeno em estudo é um tópico extremamente importante. Sendo assim, com esta finalidade, diversos critérios para seleção de modelos são encontrados na literatura e baseiam-se no Máximo da Função de Verossimilhança.

Tais critérios, definem diferentes penalidades no que diz respeito ao número de parâmetros contidos no modelo. Ressaltando, que o modelo selecionado, dentre um conjunto de modelos concorrentes, é o que minimizar o valor calculado do critério.

3.0.1 Critério de Informação de Akaike - AIC

O Critério de Informação de Akaike (AIC) foi desenvolvido por Hirotugu Akaike, pioneiro no campo da modelagem de dados estatísticos e identificação de modelos estatísticos. Originado da minimização da informação de Kullback-Leibler ([KULLBACK e LEIBLER, 1951](#)) e proposto em [Akaike \(1973\)](#), o AIC é um critério que avalia a qualidade do ajuste do modelo paramétrico, estimado pelo método da máxima verossimilhança e definido como:

$$\begin{aligned} \text{AIC} &= -2 \sum_{i=1}^n \log f(x_i | \hat{\theta}) + 2k \\ &= -2 \log L(\hat{\theta}) + 2k, \end{aligned}$$

em que $L(\hat{\theta})$ é a função de verossimilhança maximizada e k é o número de parâmetros livres no modelo. O termo $2k$ é o termo de penalidade e atua como uma compensação pelo viés na falta de ajuste quando os estimadores de máxima verossimilhança são usados.

O AIC é um critério assintoticamente eficiente. Contudo, por ser uma aproxi-

mação assintótica, pode não apresentar bom desempenho em pequenas amostras.

3.0.2 Critério de Informação de Akaike corrigido - AICc

Levando em consideração a preocupação que o AIC pode ter um mau desempenho se houver muitos parâmetros em relação ao tamanho da amostra, diversos autores propuseram algumas variações para o AIC com o objetivo de ajustar o viés (distorção entre a medida de uma variável estatística e o valor real da grandeza a estimar). Uma dessas variações é o Critério de Informação de Akaike corrigido, AICc, proposto por [Sugiura \(1978\)](#) como uma variante de segunda ordem do AIC, dado por:

$$\text{AICc} = -2 \log L(\hat{\theta}) + 2k + \left(\frac{2k^2 + 2k}{n - k - 1} \right),$$

em que k é o número de parâmetros e n é o número de observações da amostra.

De acordo com [Burnham e Anderson \(2002\)](#), se o tamanho da amostra for grande com relação a quantidade de parâmetros, essa correção é desprezível e o AIC deve ter um bom desempenho. Geralmente o AICc é aconselhado quando $\frac{n}{k} < 40$. Por outro lado, se $\frac{n}{k}$ for suficientemente grande, os critérios AICc e AIC apresentam resultados semelhantes, tendendo ambos a selecionar o mesmo modelo.

3.0.3 Critério de Informação Bayesiano - BIC

O Critério de Informação Bayesiano (BIC) proposto por [Schwarz \(1978\)](#) é um critério baseado na avaliação de modelos definidos em termos da probabilidade a posteriori. Assim, seja $F(x_n | \hat{\theta}_i)$ um modelo estatístico estimado através do método de máxima verossimilhança. Então o critério BIC é definido por

$$\begin{aligned} \text{BIC} &= -2 \sum_{i=1}^n \log f(x_n | \hat{\theta}) + k \log n \\ &= -2 \log L(\hat{\theta}) + k \log n, \end{aligned}$$

em que k é o número de parâmetros a serem escolhidos e n é o número de observações da amostra.

Segundo [Burnham e Anderson \(2002\)](#), o AIC e o BIC são critérios assintóticos para comparar modelos encaixados, mas também podem ser aplicados em modelos não encaixados.

3.0.4 Critério de Informação de Hannan-Quinn - HQIC

O critério de Informação de Hannan-Quinn (HQIC) é uma medida de adequação de um modelo estatístico e é frequentemente usado como uma alternativa ao AIC e

BIC para selecionar um modelo a partir de um conjunto finito de modelos. Proposto por [Hannan e Quinn \(1979\)](#), é definido como sendo

$$\text{HQIC} = -2k \log(\log(n)) - 2 \log(L(\hat{\theta})),$$

em que $L(\hat{\theta})$ é a função de verossimilhança maximizada, k é o número de parâmetros e n é o número de observações da amostra. É importante ressaltar, que o HQIC assim como o BIC não é um estimador originado da minimização da informação de Kullback-Leibler.

A literatura, também dispõe de critérios de seleção de modelos não paramétricos. Em [Eubank e Hart \(1993\)](#) é fornecida uma visão geral da metodologia. Aqui, dentre os vários critérios destacaremos os seguintes.

3.0.5 Critérios de Cramér-von Mises e Anderson-Darling Modificados

Alguns critérios são baseados na função de distribuição empírica dada pela fórmula

$$F_n(x) = n^{-1} \sum_{i=1}^n I_{(-\infty, x]}(x_{(i)}),$$

sendo $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ a amostra ordenada e

$$I_{(-\infty, x]}(x_{(i)}) = \begin{cases} 1, & \text{se } x_{(i)} \leq x, \\ 0, & \text{se } x_{(i)} > x. \end{cases}$$

Dois critérios muito usados que baseiam-se em $F_n(x)$ são o critério de Cramér-von Mises, dado por

$$W^2 = n \int_{-\infty}^{+\infty} \{F_n(x) - F(x; \hat{\theta})\}^2 dF(x; \hat{\theta})$$

e expressado por

$$W^2 = \sum_{i=1}^n [u_i - \{(2i-1)/(2n)\}]^2 + 1/(12n),$$

e o critério de Anderson-Darling ([ANDERSON e DARLING, 1954](#)) dado por

$$A^2 = n \int_{-\infty}^{+\infty} \frac{\{F_n(x) - F(x; \hat{\theta})\}^2}{\{F(x; \hat{\theta})(1 - F(x; \hat{\theta}))\}} dF(x; \hat{\theta}),$$

e expressado por

$$A^2 = -n - n^{-1} \sum_{i=1}^n \{(2i-1) \log(u_i) + (2n+1-2i) \log(1-u_i)\}.$$

Em ambos critérios, $\hat{\theta}$ é uma estimativa de θ .

De acordo com [Chen e Balakrishnan \(1995\)](#), critérios baseados na função de distribuição empírica para o caso de parâmetros desconhecidos, são principalmente critérios assintóticos, onde não apenas os pontos críticos limitantes devem ser calculados sendo que é necessário um grande esforço, por meio de simulações, para lidar com situações realistas de tamanho de amostra finita resultando em um número grande de tabelas ([D' AGOSTINO e STEPHENS, 1986](#)). Dessa forma, os autores propõem uma tabela que possa servir para a maioria das distribuições de maneira simples e prática e explica como esse procedimento pode ser feito. Também, detalham os critérios de teste de normalidade modificados de W^2 e A^2 denotados, respectivamente, como:

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

e

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

sendo n o tamanho da amostra.

Geralmente, considera-se a seguinte sequência para os cálculos de W^* e A^* :

- Estima-se θ por $\hat{\theta}_n$ e ordena-se as observações para computar $v_i = F(x_i; \hat{\theta}_n)$;
- Computa-se $y_i = \Phi^{-1}(v_i)$, em que Φ^{-1} é a função quantil normal padrão;
- Computa-se $u_i = \Phi(y_i - \bar{y})/s_y$, em que $\bar{y} = n^{-1} \sum_{i=1}^n y_i$ e $s_y^2 = (n-1)^{-1} \sum_{i=1}^n (y_i - \bar{y})^2$;
- Computa-se W^2 e A^2 e finalmente computa-se W^* e A^* .

Vale ressaltar, que os critérios de seleção de modelos mencionados neste capítulo serão usados para fins de comparação com o novo critério proposto apresentado no [Capítulo 4](#).

Capítulo 4

Um Novo Critério de Seleção de Modelos Baseado nas Propriedades das Famílias Sup e Inf

4.1 Famílias de distribuições Sup e Inf

Com o objetivo de tornar uma distribuição mais flexível e melhorar a qualidade do ajuste nas caudas, [Tablada \(2017\)](#) propôs duas novas famílias de distribuições generalizadas denominadas de G_c^{sup} e G_c^{inf} , induzindo um parâmetro $c > 0$ a uma distribuição G (denominada de base). Essas famílias serão apresentadas nesta seção, bem como algumas das suas propriedades, as quais serão de fundamental importância neste trabalho.

Para a obtenção das famílias G_c^{sup} e G_c^{inf} , inicialmente, foram definidas as funções:

$$R_c^{\text{sup}}(z) = \begin{cases} 0, & \text{se } z < 0, \\ z + z^c - z^{c+1}, & \text{se } 0 \leq z \leq 1, c > 0, \\ 1 & \text{se } z > 1 \end{cases} \quad (4.1)$$

e

$$R_c^{\text{inf}}(z) = \begin{cases} 0, & \text{se } z < 0, \\ z[1 - (1 - z)^c], & \text{se } 0 \leq z \leq 1, c > 0, \\ 1 & \text{se } z > 1. \end{cases} \quad (4.2)$$

As funções (4.1) e (4.2) satisfazem as propriedades descritas na proposição 2.1, e portanto, existem v.a's T^{sup} e T^{inf} definidas em algum espaço de probabilidade (Ω, \mathcal{F}, P) , tais que $R_c^{\text{sup}}(z)$ e $R_c^{\text{inf}}(z)$ são as funções de distribuição acumuladas dessas v.a's. As funções de densidade são dadas, respectivamente, por:

$$r_c^{\text{sup}}(z) = \begin{cases} 0, & \text{se } z < 0 \text{ ou } z > 1, \\ 1 + cz^{c-1} - (c+1)z^c, & \text{se } 0 \leq z \leq 1, c > 0, \end{cases} \quad (4.3)$$

e

$$r_c^{\text{inf}}(z) = \begin{cases} 0, & \text{se } z < 0 \text{ ou } z > 1, \\ 1 - (1-z)^c + cz(1-z)^{c-1}, & \text{se } 0 \leq z \leq 1, c > 0. \end{cases} \quad (4.4)$$

Considerando que as funções (4.3) e (4.4) têm como suporte o intervalo $[0, 1]$ então, para uma distribuição de base G , as famílias de distribuições generalizadas G_c^{sup} e G_c^{inf} são dadas por

$$G_c^{\text{sup}}(x) = R_c^{\text{sup}}(G(x)) = G(x) + G^c(x) - G^{c+1}(x) \quad (4.5)$$

e

$$G_c^{\text{inf}}(x) = R_c^{\text{inf}}(G(x)) = G(x)[1 - (1 - G(x))^c]. \quad (4.6)$$

Para o caso em que G é contínua, derivando as funções de distribuição G_c^{sup} e G_c^{inf} obtemos as funções de densidade de probabilidade que são, respectivamente

$$g_c^{\text{sup}}(x) = g(x)[1 + cG^{c-1}(x) - (c+1)G^c(x)] \quad (4.7)$$

e

$$g_c^{\text{inf}}(x) = g(x)[1 - (1 - G(x))^c + cG(x)(1 - G(x))^{c-1}]. \quad (4.8)$$

Note que a função $g(x)$ é a f.d.p da v.a X que segue distribuição $G(x)$. A seguir, serão apresentadas algumas propriedades das famílias G_c^{sup} e G_c^{inf} que contribuirão para os resultados teóricos deste trabalho. As provas podem ser encontradas em [Tablada \(2017\)](#) e [Sousa \(2020\)](#).

Proposição 4.1. *Seja X uma v.a que segue distribuição G . Então,*

$$G_c^{\text{inf}}(x) \leq G(x) \leq G_c^{\text{sup}}(x), \quad \forall x \in \mathbb{R}, c > 0.$$

Proposição 4.2. *Seja X uma v.a que segue distribuição G . Então para todo $j > 0$,*

$$G_{c+j}^{\text{sup}}(x) \leq G_c^{\text{sup}}(x)$$

e

$$G_{c+j}^{\text{inf}}(x) \geq G_c^{\text{inf}}(x),$$

para todo $x \in \mathbb{R}$. Isto é, G_c^{sup} e G_c^{inf} são, nessa ordem, famílias monótonas não-crescentes e não-decrescentes em relação ao parâmetro c .

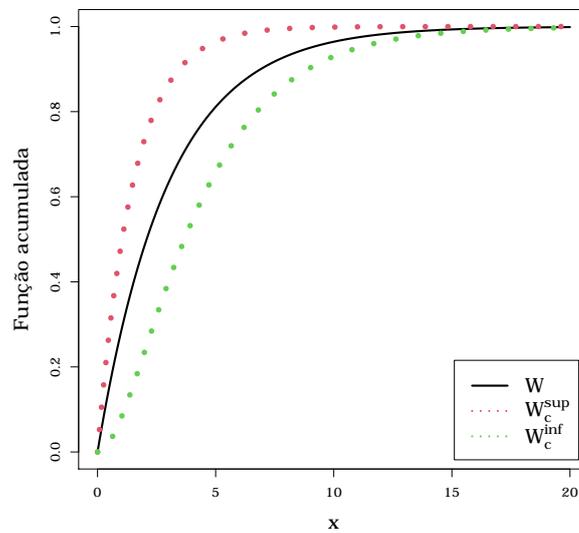
Proposição 4.3. *Seja X uma v.a que segue distribuição G . Então*

$$\lim_{c \rightarrow \infty} G_c^{\text{sup}}(x) = G(x)$$

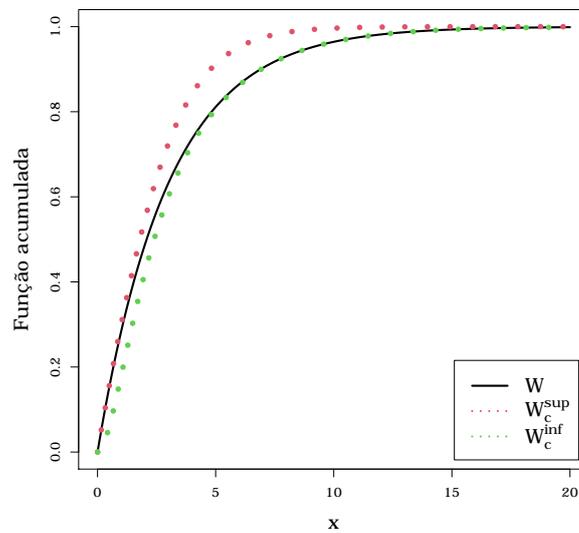
e

$$\lim_{c \rightarrow \infty} G_c^{\text{inf}}(x) = G(x).$$

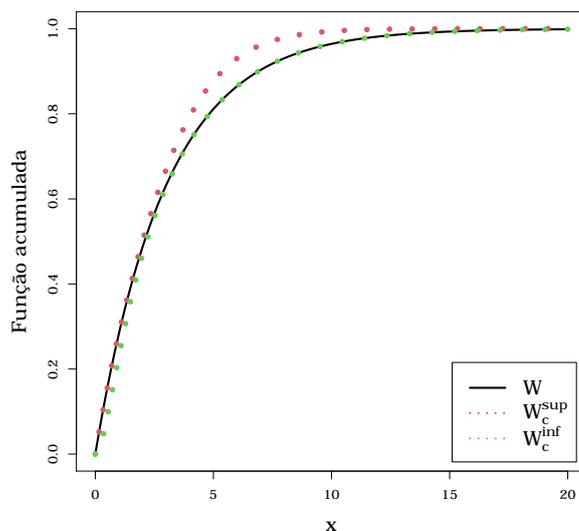
A Figura 4.1 ilustra as proposições 4.1 a 4.3 para as funções de distribuições acumuladas Weibull, Weibull-sup e Weibull-inf. Ressaltando, que o fato de G ser uma distribuição qualquer, teríamos um comportamento semelhante das curvas para outras distribuições.



(a) $c = 1$



(b) $c = 3$



(c) $c = 5$

Figura 4.1: Plots da f.d.a da distribuição Weibull, Weibull-sup e Weibull-inf .

Com relação a proposição 4.1 podemos observar, de fato, que a curva da distribuição acumulada Weibull-sup sempre fica por cima da curva da distribuição acumulada Weibull e a curva da distribuição acumulada Weibull-inf sempre fica por baixo. Já, com relação as proposições 4.2 e 4.3 é notável que à medida que o parâmetro c cresce as curvas das distribuições Weibull-sup e Weibull-inf tendem a coincidir progressivamente com a curva da distribuição Weibull.

Proposição 4.4. *Seja X uma v.a contínua com f.d.a $G(x)$ e f.d.p $g(x)$. Para $0 < G(x) < 1$, existe $A > 0$ tal que $g_c^{\text{sup}}(x) < g(x) < g_c^{\text{inf}}(x)$, sempre que $x > A$.*

Proposição 4.5. *Seja X uma v.a contínua com f.d.a $G(x)$ e f.d.p $g(x)$. Para $0 < G(x) < 1$, existe $A < 0$ tal que $g_c^{\text{inf}}(x) < g(x) < g_c^{\text{sup}}(x)$, sempre que $x < A$.*

Observação 4.1. *As proposições (4.4) e (4.5) foram provadas em Sousa (2020) para $g_c^{\text{inf}}(x) \leq g(x) \leq g_c^{\text{sup}}(x)$, aqui iremos considerá-las apenas com as desigualdades estritas para as quais suas demonstrações também são válidas.*

4.2 Função de log-verossimilhança das famílias Sup e Inf

Considere X_1, X_2, \dots, X_n uma amostra aleatória da v.a X contínua que segue distribuição $G_c^{\text{sup}}(x, \theta)$ ou $G_c^{\text{inf}}(x, \theta)$, em que $\theta = (c, \delta^T)^T$ é o vetor de parâmetros da distribuição G_c^{sup} ou G_c^{inf} e δ é o vetor de parâmetros da distribuição de base G .

Denotamos por $\ell_{g_c^{\text{sup}}}(\theta; \mathbf{x})$, $\ell_{g_c^{\text{inf}}}(\theta; \mathbf{x})$ e $\ell_g(\delta; \mathbf{x})$ a função de log-verossimilhança de G_c^{sup} , G_c^{inf} e G , respectivamente, e por $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ a amostra observada.

A função de verossimilhança de G_c^{sup} é dada por

$$\begin{aligned} L_{\text{sup}}(\theta; \mathbf{x}) &= \prod_{i=1}^n g_c^{\text{sup}}(x_i; \theta) \\ &= \prod_{i=1}^n g(x_i; \delta) [1 + cG^{c-1}(x_i; \delta) - (c+1)G^c(x_i; \delta)] \\ &= \prod_{i=1}^n g(x_i; \delta) h_1(x_i; \theta), \end{aligned}$$

em que $h_1(x; \theta) = [1 + cG^{c-1}(x; \delta) - (c+1)G^c(x; \delta)]$.

Com isso, segue que a função de log-verossimilhança de G_c^{sup} é dada por

$$\begin{aligned} \ell_{g_c^{\text{sup}}}(\theta; \mathbf{x}) &= \log[L_{\text{sup}}(\theta; \mathbf{x})] \\ &= \sum_{i=1}^n \log[g(x_i; \delta) h_1(x_i; \theta)] \\ &= \sum_{i=1}^n \log(g(x_i; \delta)) + \sum_{i=1}^n \log(h_1(x_i; \theta)) \\ &= \ell_g(\delta; \mathbf{x}) + \sum_{i=1}^n \log(h_1(x_i; \theta)). \end{aligned} \tag{4.9}$$

Considerando $0 \leq G(x; \delta) < 1$, é possível provar que $\lim_{c \rightarrow \infty} \sum_{i=1}^n \log(h_1(x_i; \theta)) = 0$. Para isso, é suficiente provar que $\lim_{c \rightarrow \infty} h_1(x; \theta) = 1$. De fato, se $G(x; \delta) = 0$, segue que $h_1(x; \theta) = 1$. Para $0 < G(x; \delta) < 1$, temos que

$$\begin{aligned} \lim_{c \rightarrow \infty} h_1(x; \theta) &= \lim_{c \rightarrow \infty} [1 + cG^{c-1}(x; \delta) - (c+1)G^c(x; \delta)] \\ &= 1 + \lim_{c \rightarrow \infty} \frac{c}{[G(x; \delta)^{-1}]^{c-1}} - \lim_{c \rightarrow \infty} \frac{c+1}{[G(x; \delta)^{-1}]^c}. \end{aligned}$$

Note que, os dois limites acima, na última igualdade são do tipo $\frac{+\infty}{+\infty}$, sendo necessário utilizar a regra de L'Hôpital, derivando o numerador e o denominador de ambos os limites em relação a c . Daí,

$$\lim_{c \rightarrow \infty} \frac{1}{[G(x; \delta)^{-1}]^{c-1} \log(G(x; \delta)^{-1})} = \lim_{c \rightarrow \infty} \frac{1}{[G(x; \delta)^{-1}]^c \log(G(x; \delta)^{-1})} = 0,$$

uma vez que $G(x; \delta)^{-1} > 1$, já que $0 < G(x; \delta) < 1$. Logo, tem-se

$$\lim_{c \rightarrow \infty} h_1(x; \theta) = 1 + 0 - 0 = 1.$$

Dessa forma, aplicando o limite em 4.9 segue que para todo x_1, x_2, \dots, x_n no interior do suporte de G ,

$$\lim_{c \rightarrow \infty} \ell_{g_c^{\text{sup}}}(\theta; \mathbf{x}) = \ell_g(\delta; \mathbf{x}). \quad (4.10)$$

Para G_c^{inf} , temos que a função de verossimilhança é dada por

$$\begin{aligned} L_{\text{inf}}(\theta; \mathbf{x}) &= \prod_{i=1}^n g_c^{\text{inf}}(x_i; \theta) \\ &= \prod_{i=1}^n g(x_i; \delta) [1 - (1 - G(x_i; \delta))^c + c G(x_i; \delta) (1 - G(x_i; \delta))^{c-1}] \\ &= \prod_{i=1}^n g(x_i; \delta) h_2(x_i; \theta), \end{aligned}$$

em que $h_2(x; \theta) = [1 - (1 - G(x; \delta))^c + c G(x; \delta) (1 - G(x; \delta))^{c-1}]$.

Com isso, segue que a função de log-verossimilhança é dada da seguinte forma

$$\begin{aligned} \ell_{g_c^{\text{inf}}}(\theta; \mathbf{x}) &= \log[L_{\text{inf}}(\theta; \mathbf{x})] \\ &= \sum_{i=1}^n \log[g(x_i; \delta) h_2(x_i; \theta)] \\ &= \sum_{i=1}^n \log(g(x_i; \delta)) + \sum_{i=1}^n \log(h_2(x_i; \theta)) \\ &= \ell_g(\delta; \mathbf{x}) + \sum_{i=1}^n \log(h_2(x_i; \theta)). \end{aligned} \quad (4.11)$$

Aqui, também será provado que considerando $0 < G(x; \delta) \leq 1$, $\lim_{c \rightarrow \infty} \sum_{i=1}^n \log(h_2(x_i; \theta)) = 0$. Pra isso, iremos provar que $\lim_{c \rightarrow \infty} h_2(x; \theta) = 1$. De fato, se $G(x; \delta) = 1$, segue que $h_2(x; \theta) = 1$, satisfazendo o limite. Para $0 < G(x; \delta) < 1$, temos que

$$\begin{aligned} \lim_{c \rightarrow \infty} h_2(x; \theta) &= \lim_{c \rightarrow \infty} [1 - (1 - G(x; \delta))^c + c G(x; \delta) (1 - G(x; \delta))^{c-1}] \\ &= 1 - \lim_{c \rightarrow \infty} (1 - G(x; \delta))^c + G(x; \delta) \lim_{c \rightarrow \infty} \frac{c}{[(1 - G(x; \delta))^{-1}]^{c-1}}. \end{aligned}$$

Note que $\lim_{c \rightarrow \infty} (1 - G(x; \delta))^c = 0$, pois $0 < (1 - G(x; \delta)) < 1$ e utilizando mais uma vez a regra de L'Hôpital, segue

$$\lim_{c \rightarrow \infty} \frac{c}{[(1 - G(x; \delta))^{-1}]^{c-1}} = \lim_{c \rightarrow \infty} \frac{1}{[(1 - G(x; \delta))^{-1}]^{c-1} \log(1 - G(x; \delta))^{-1}} = 0,$$

uma vez que $[(1 - G(x_i; \delta))^{-1}]^{c-1}$ se aproxima de $+\infty$ quando c tende a $+\infty$. Logo,

$$\lim_{c \rightarrow \infty} h_2(x; \theta) = 1 - 0 + 0 = 1.$$

Portanto, aplicando o limite em 4.11 segue que para todo x_1, x_2, \dots, x_n no interior do suporte de G ,

$$\lim_{c \rightarrow \infty} \ell_{g_c^{\text{inf}}}(\theta; \mathbf{x}) = \ell_g(\delta; \mathbf{x}). \quad (4.12)$$

4.3 Proposta do novo critério de seleção de modelos

Nesta pesquisa propomos um novo critério de seleção para modelos não encaixados, baseado nas famílias de distribuições G_c^{sup} e G_c^{inf} e suas propriedades apresentadas nesse Capítulo..

Seja $G(x; \delta)$ a função de distribuição acumulada do modelo de interesse, com vetor de parâmetros δ , função densidade denotada $g(x; \delta)$ e log-verossimilhança ℓ_g . Em seguida, considere as famílias $G_{\theta_1}^{\text{sup}} = \{g^{\text{sup}}(x; \theta_1) : \theta_1 \in \Theta\}$, e $G_{\theta_2}^{\text{inf}} = \{g^{\text{inf}}(x; \theta_2) : \theta_2 \in \Theta\}$, em que $\theta_1 = (\delta, c_1)$ e $\theta_2 = (\delta, c_2)$, sendo $c_1 > 0$ e $c_2 > 0$. Pelas Proposições 4.4 e 4.5, podemos garantir que $g^{\text{sup}}(x; \theta_1) \neq g^{\text{inf}}(x; \theta_2)$ para quase todo x , o qual por sua vez implica que $G_{\theta_1}^{\text{sup}} \cap G_{\theta_2}^{\text{inf}} = \emptyset$, isto é, os modelos sup e inf são não encaixados.

Dado um conjunto de dados $\mathbf{x} = (x_1, x_2, \dots, x_n)$ temos que o EMV de δ é dado por

$$\hat{\delta} = \arg \max_{\delta} \ell_g(\delta).$$

Assim, considerando os vetores paramétricos $\theta_1 = (\delta, c_1)$ e $\theta_2 = (\delta, c_2)$ dos modelos sup e inf, respectivamente, e sendo $\ell_{g_c^{\text{sup}}}$ e $\ell_{g_c^{\text{inf}}}$ suas respectivas log-verossimilhanças, definimos

$$\hat{c}_1^* = \arg \max_{c_1} \ell_{g_c^{\text{sup}}}(\hat{\delta}, c_1) \quad \text{e} \quad \hat{c}_2^* = \arg \max_{c_2} \ell_{g_c^{\text{inf}}}(\hat{\delta}, c_2)$$

Com isso, propomos o critério de seleção de modelos V dado por

$$V = \left[\sum_{i=1}^n \log(h_1(x_i; \hat{\delta}, \hat{c}_1^*)) \right]^2 + \left[\sum_{i=1}^n \log(h_2(x_i; \hat{\delta}, \hat{c}_2^*)) \right]^2. \quad (4.13)$$

De acordo com a Seção 4.2, se $G(x; \delta)$ é um bom modelo para ajustar os dados, então esperamos que $\sum_{i=1}^n \log(h_1(x_i; \hat{c}_1^*)) \approx \sum_{i=1}^n \log(h_2(x_i; \hat{c}_2^*)) \approx 0$. Com esse resultado, temos que $\ell_{g_c^{\text{sup}}}(\hat{\delta}, c_1; \mathbf{x}) \approx \ell_{g_c^{\text{inf}}}(\hat{\delta}, c_2; \mathbf{x}) \approx \ell_g(\hat{\delta}; \mathbf{x})$. Considerando as proposições 4.1 e 4.3, isso implicaria na equivalência entre $G_{\theta_1}^{\text{sup}}$ e $G_{\theta_2}^{\text{inf}}$ o que por sua vez

implicaria que os dados poderão ser modelados por $G(x; \delta)$. Dessa forma, o melhor modelo para ajustar os dados é aquele para o qual o critério V for aproximadamente igual a zero.

Em resumo, para calcular V considera-se a seguinte sequência:

- Seja $G(x; \delta)$ a função de distribuição acumulada do modelo de interesse, $g(x; \delta)$ a sua função de densidade, $\ell_g(\delta; x)$, $\ell_{g_c^{\text{sup}}}(\theta_1; x)$ e $\ell_{g_c^{\text{inf}}}(\theta_2; x)$, em que $\theta_1 = (\delta, c_1)$ e $\theta_2 = (\delta, c_2)$ as respectivas, função de log-verossimilhança do modelo de interesse, a função de log-verossimilhança sup e a função de log-verossimilhança inf;
- Obtêm-se o EMV $\hat{\delta}$, maximizando a função $\ell_g(\delta; x)$;
- Obtêm-se \hat{c}_1^* e \hat{c}_2^* , maximizando $\ell_{g_c^{\text{sup}}}(\hat{\delta}, c_1; x)$ e $\ell_{g_c^{\text{inf}}}(\hat{\delta}, c_2; x)$;
- Por fim, computa-se o valor de V através da equação 4.13.

Capítulo 5

Resultados Numéricos

Nesta pesquisa propomos um novo critério de seleção de modelos a saber, V , baseado nas famílias de distribuições G_c^{sup} e G_c^{inf} , e suas propriedades. Com isso, neste capítulo objetivamos comparar o desempenho do novo critério em relação aos critérios AIC, AICc, BIC, HQIC, W^* e A^* . Os diferentes critérios foram avaliados com base nas taxas ou porcentagens de acerto, isto é, a porcentagem de vezes em que o modelo correto foi selecionado.

Para avaliar o desempenho do novo critério, foram realizadas simulações de Monte Carlo utilizando distribuições encaixadas e/ou concorrentes nos diferentes cenários. As distribuições foram: Weibull Exponencializada (MUDHOLKAR *et al.*, 1993), Weibull (WEIBULL, 1939), Lindley Ponderada (MAZUCHELI *et al.*, 2016) e Normal. Essas distribuições são apresentadas no Apêndice.

Para essas simulações foram considerados os seguintes cenários:

- **Cenário 1:** comparação dos modelos encaixados/concorrentes Weibull Exponencializado, $WE(\theta, \alpha, \sigma)$, com o modelo Weibull(α, σ) considerando o parâmetro θ conhecido;
- **Cenário 2:** comparação dos modelos concorrentes Lindley Ponderada, $LP(\mu, \lambda)$, com o modelo Weibull(α, σ);
- **Cenário 3:** comparação do modelo $WE(\theta, \alpha, \sigma)$ com o modelo Weibull(α, σ) considerando o parâmetro θ desconhecido;
- **Cenário 4:** comparação de dois modelos normais, $N(\mu_1, \sigma_1^2)$ com o modelo $N(\mu_2, \sigma_2^2)$, considerando o parâmetro σ conhecido.

Para a obtenção dos resultados numéricos considerou-se 20.000 réplicas de Monte Carlo. O código foi implementado utilizando a linguagem de programação `0x` (DOORNIK, 2018), o qual mostrou um ótimo desempenho para a quantidade de réplicas consideradas. Os tamanhos amostrais utilizados no estudo foram

$n = 20, 30, 50, 100, 200$.

5.1 Cenário 1: Distribuição Weibull Exponencializada \times Distribuição Weibull

Neste cenário, o objetivo é comparar o modelo $WE(\theta, \alpha, \sigma)$, em que θ é conhecido, com o modelo $Weibull(\alpha, \sigma)$. É importante ressaltar que o modelo Weibull é um caso particular do modelo Weibull exponencializado, quando $\theta = 1$. A suposição de que θ é conhecido foi necessária para comparar modelos com o mesmo número de parâmetros e, além disso, estabelecer uma relação de proximidade entre os dois modelos considerados. Sendo assim, os valores de θ foram selecionados com o objetivo de estudar o desempenho dos critérios para detectar pequenas diferenças entre os dois modelos ajustados.

Para este cenário de simulação, os dados foram obtidos a partir do modelo $WE(\theta, \alpha, \sigma)$, isto é, o modelo Weibull exponencializado é o correto. O valor do parâmetro θ variou em 1.1, 1.3, 1.5 e 3.0, enquanto que os valores dos parâmetros α e σ foram fixados em 3.5.

Para uma visualização das diferenças entre os modelos a Figura 5.1 mostra as densidades do modelo Weibull e do modelo Weibull exponencializado para as variações do parâmetro θ considerados neste cenário.

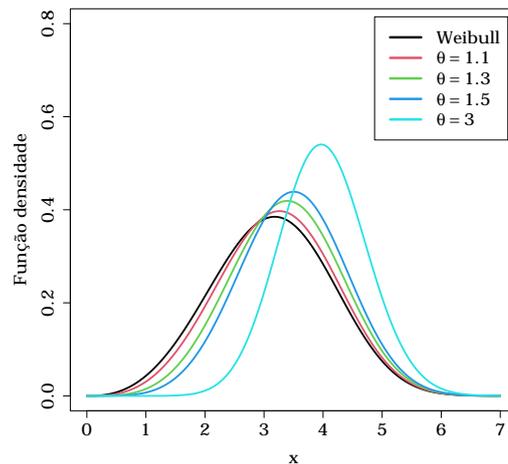


Figura 5.1: Densidades do modelo Weibull versus WE.

A Tabela 5.1 é referente às taxas de acertos para este cenário. Ao analisar a tabela, conclui-se que o critério V apresenta taxa de acerto superior a 50% para todos os tamanhos amostrais. Para $\theta = 1.1$, note que a taxa de acerto de V diminui

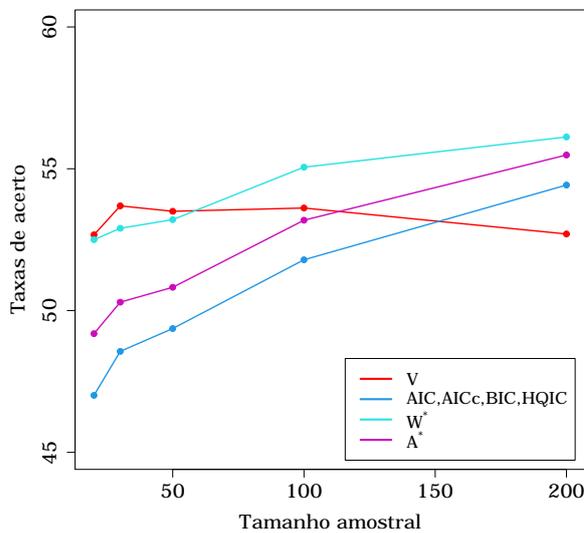
Tabela 5.1: Taxas de acertos para o cenário 1.

θ	n	V	AIC	AICc	BIC	HQIC	W^*	A^*
1.1	20	52.670	47.010	47.010	47.010	47.010	52.500	49.185
	30	53.690	48.560	48.560	48.560	48.560	52.900	50.295
	50	53.500	49.365	49.365	49.365	49.365	53.205	50.820
	100	53.615	51.790	51.790	51.790	51.790	55.055	53.185
	200	52.700	54.425	54.425	54.425	54.425	56.120	55.485
1.3	20	55.285	49.955	49.955	49.955	49.955	54.320	51.405
	30	56.690	52.230	52.230	52.230	52.230	55.475	53.300
	50	57.925	54.495	54.495	54.495	54.495	57.160	55.240
	100	59.135	58.855	58.855	58.855	58.855	60.635	59.790
	200	59.670	64.340	64.340	64.340	64.340	64.495	64.535
1.5	20	57.300	52.415	52.415	52.415	52.415	55.895	53.445
	30	59.470	55.805	55.805	55.805	55.805	58.040	56.330
	50	61.065	59.020	59.020	59.020	59.020	60.465	59.065
	100	63.950	65.050	65.050	65.050	65.050	65.305	65.130
	200	65.840	72.275	72.275	72.275	72.275	70.605	71.510
3.0	20	65.790	62.780	62.780	62.780	62.780	63.290	62.590
	30	69.940	67.940	67.940	67.940	67.940	67.530	67.315
	50	75.055	75.005	75.005	75.005	75.005	73.005	73.815
	100	82.180	84.840	84.840	84.840	84.840	81.860	83.160
	200	88.600	92.970	92.970	92.970	92.970	89.880	91.655

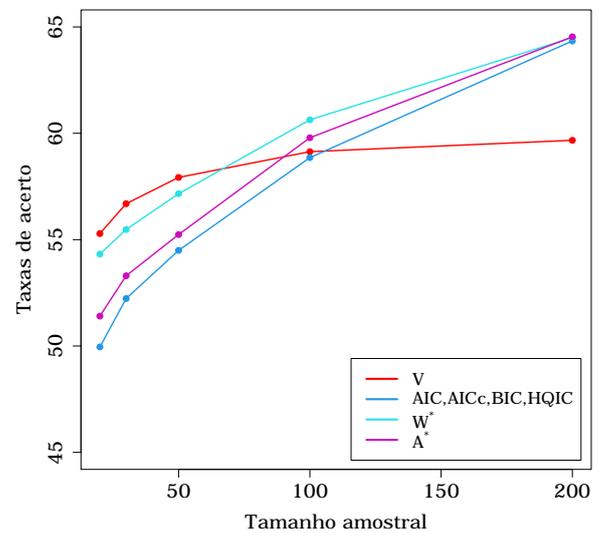
quando o tamanho da amostra aumenta. No entanto, para os valores de θ maiores que 1.1 a taxa de acerto aumenta conforme o tamanho amostral cresce. Observe também que para $\theta = 1.1$, apesar do modelo Weibull exponencializado se aproximar do modelo Weibull, a taxa de acerto do critério V é 53.615%, sendo inferior apenas com relação a taxa do critério W^* com 55.055% para o tamanho da amostra $n = 100$.

Ainda analisando a Tabela 5.1, observamos que o critério V mostra um desempenho superior para os tamanhos amostrais $n \leq 50$ quando comparado aos demais critérios. Dessa forma o critério proposto apresenta um melhor desempenho em relação aos outros critérios quando o tamanho amostral é pequeno ou moderado. Ainda com relação às taxas de acertos, é possível notar que os critérios AIC, AICc, BIC e HQIC apresentam desempenho equivalentes em todos os casos, sendo esse desempenho superior aos demais critérios para o parâmetro θ igual a 1.5 e 3.0 e o tamanho amostral igual a 200. Para o tamanho amostral n igual a 100 e θ igual a 1.3 e 1.5 os critérios W^* e A^* apresentam desempenho superior, bem como para n igual a 200 e θ igual a 1.1 e 1.3.

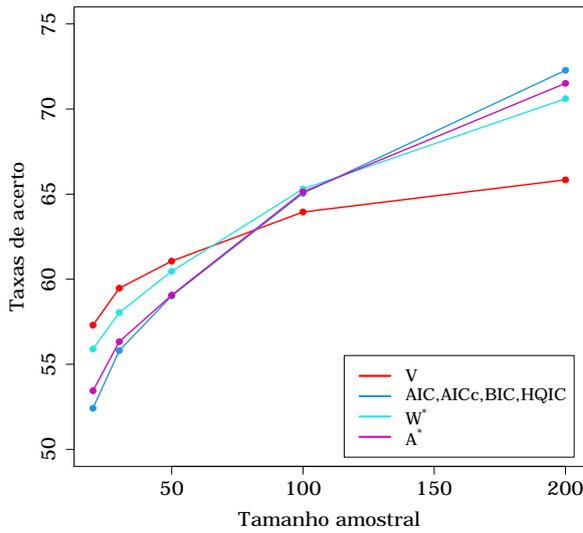
A Figura 5.2 ilustra a situação apresentada na Tabela 5.1. Os gráficos, mostram claramente um desempenho superior do critério V para amostras pequenas e um desempenho ruim dos critérios AIC, AICc, BIC e HQIC para amostras pequenas.



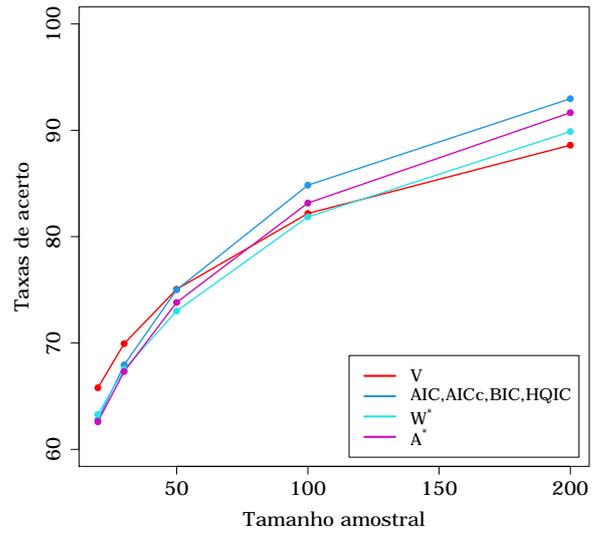
(a) $\theta = 1.1$



(b) $\theta = 1.3$



(c) $\theta = 1.5$



(d) $\theta = 3.0$

Figura 5.2: Taxas de acertos para o cenário 1.

5.2 Cenário 2: Distribuição Lindley Ponderada \times Distribuição Weibull

Nesse cenário, o objetivo é comparar o modelo $LP(\mu, \lambda)$ com o modelo $Weibull(\alpha, \sigma)$, ambos com a mesma quantidade de parâmetros. Os dados foram obtidos a partir do modelo $LP(\mu, \lambda)$, isto é, o modelo LP é o correto. O valor do parâmetro λ variou em 1.5, 2.0, 2.5, enquanto que o valor de μ foi fixado em 1.5.

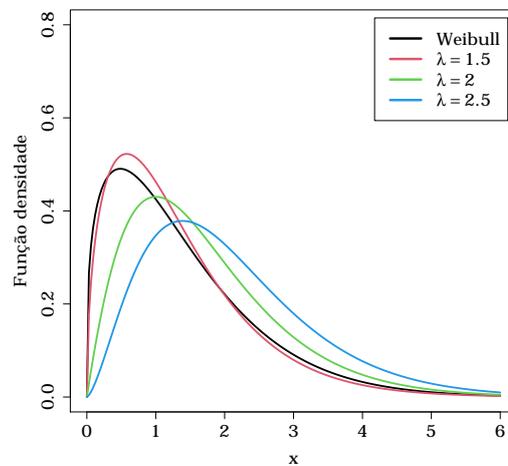


Figura 5.3: Densidades do modelo Weibull versus LP.

Para uma visualização das diferenças entre os modelos a Figura 5.3 mostra as densidades do modelo Weibull e do modelo LP para as variações do parâmetro λ considerados neste cenário.

Tabela 5.2: Taxas de acertos para o cenário 2.

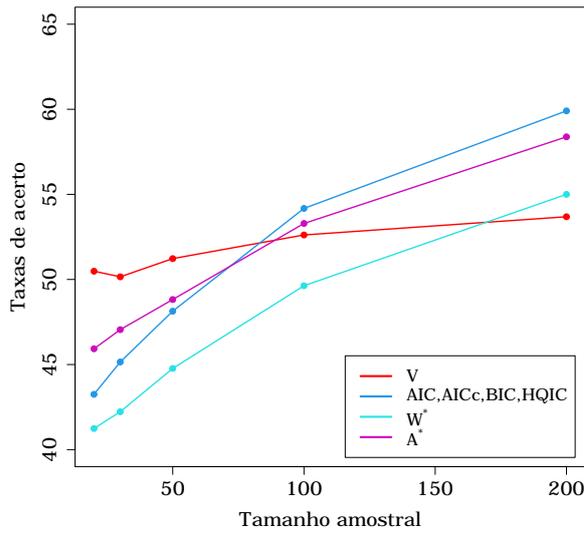
λ	n	V	AIC	AICc	BIC	HQIC	W^*	A^*
1.5	20	50.485	43.250	43.250	43.250	43.250	41.240	45.925
	30	50.155	45.150	45.150	45.150	45.150	42.225	47.050
	50	51.225	48.135	48.135	48.135	48.135	44.770	48.820
	100	52.615	54.175	54.175	54.175	54.175	49.630	53.290
	200	53.685	59.905	59.905	59.905	59.905	55.000	58.380
2.0	20	55.385	50.555	50.555	50.555	50.555	47.080	49.545
	30	57.480	54.540	54.540	54.540	54.540	51.185	52.970
	50	60.195	59.190	59.190	59.190	59.190	56.330	58.090
	100	63.585	66.825	66.825	66.825	66.825	64.145	66.320
	200	66.650	74.955	74.955	74.955	74.955	71.735	74.375
2.5	20	59.550	56.775	56.775	56.775	56.775	53.810	54.870
	30	61.220	59.715	59.715	59.715	59.715	57.615	58.750
	50	65.020	65.885	65.885	65.885	65.885	63.360	65.125
	100	70.125	73.750	73.750	73.750	73.750	70.800	73.060
	200	74.335	82.305	82.305	82.305	82.305	78.145	80.920

A Tabela 5.2 é referente às taxas de acertos para este cenário. Analisando essa tabela, observamos que, o critério V apresenta taxa de acerto maior que 50% para todos os tamanhos amostrais e aumentam conforme n e λ aumentam.

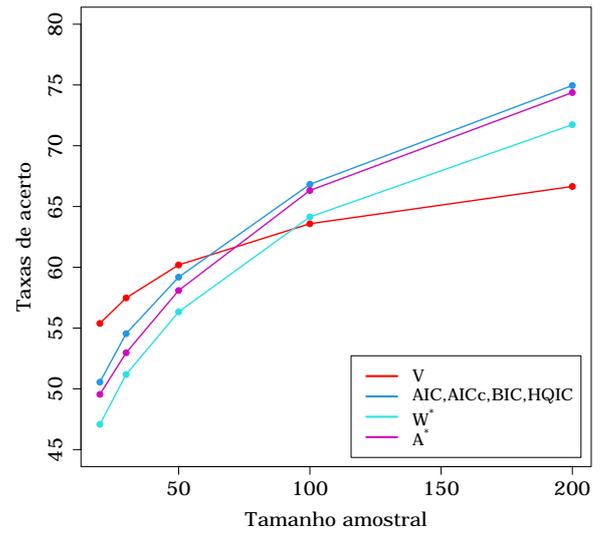
Ao observar a Tabela 5.2, notamos que o critério V apresenta desempenho superior quando comparados aos demais critérios para os tamanhos de amostras $n \leq 50$ nos casos $\lambda = 1.5, 2.0$. Por exemplo, para $n = 20$ e $\lambda = 1.5$ a taxa de acerto do critério V foi de 50.485% e a taxa de acerto do critério A^* foi de 45.925%. Dessa forma, podemos constatar, novamente, que o critério proposto tem melhor desempenho quando o tamanho amostral é pequeno ou moderado.

Ainda analisando a Tabela 5.2, podemos destacar que os critérios AIC, AICc, BIC e HQIC, assim como no cenário 1, apresentam desempenho equivalentes em todos os casos. Ressaltando, que para λ igual a 2.5 e tamanho da amostra $n \geq 50$ esses critérios mostram desempenho superior.

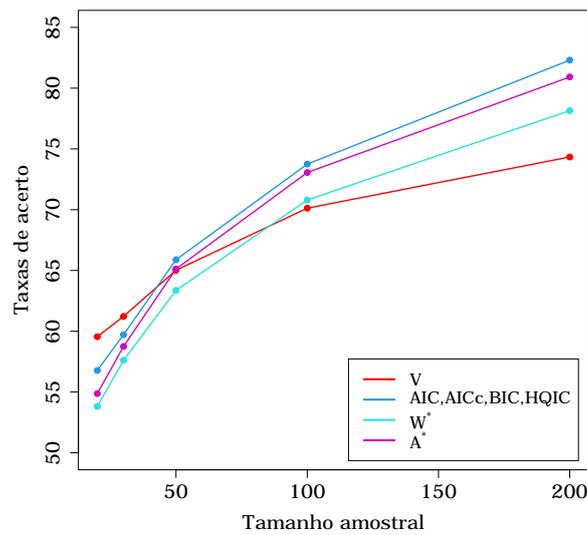
A Figura 5.4 ilustra a situação apresentada na Tabela 5.2. Os gráficos, mostram um desempenho superior do critério V para amostras pequenas.



(a) $\lambda = 1.5$



(b) $\lambda = 2.0$



(c) $\lambda = 2.5$

Figura 5.4: Taxas de acertos para o cenário 2.

5.3 Cenário 3: Distribuição Weibull Exponencializada \times Distribuição Weibull

Neste cenário, o objetivo é comparar o modelo $WE(\theta, \alpha, \sigma)$, em que θ é desconhecido, com o modelo $Weibull(\alpha, \sigma)$. Sendo assim, os valores de θ foram selecionados com o objetivo de estudar o desempenho dos critérios para modelos com o número de parâmetros diferentes.

Para este cenário de simulação, os dados foram obtidos a partir do modelo $WE(\theta, \alpha, \sigma)$, isto é, o modelo WE é o correto. O valor do parâmetro θ variou em 2.0, 2.5 e 3.0, enquanto que os valores dos parâmetros α e σ foram fixados em 12.5 e 0.5, respectivamente. Neste cenário, não consideramos a amostra de tamanho n igual a 20, por conter pouca informação para uma distribuição de três parâmetros, o que causou um grande número de não convergências.

Para uma visualização das diferenças entre os modelos a Figura 5.5 mostra as densidades do modelo Weibull e do modelo Weibull exponencializado para as variações do parâmetro θ considerados neste cenário.

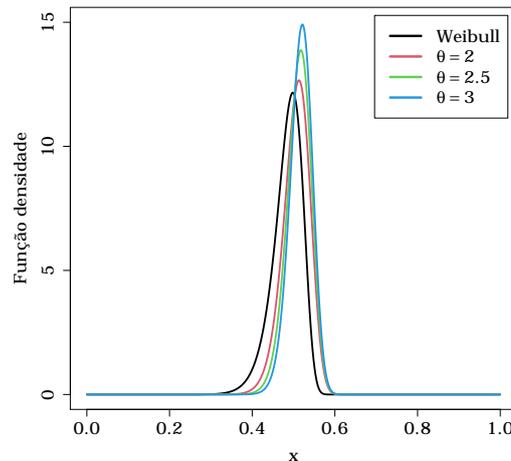


Figura 5.5: Densidades do modelo Weibull versus WE.

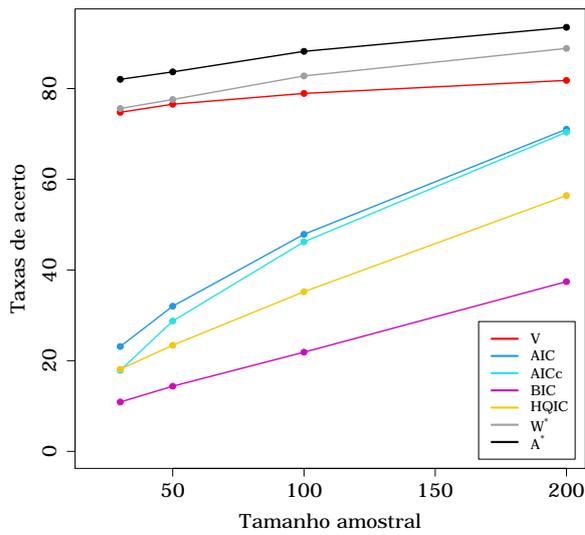
Na Tabela 5.3 apresentamos as taxas de acerto para este cenário. Avaliando essa tabela, observamos que para todos os tamanhos amostrais o critério V continua apresentando taxa de acerto maior que 50%, além disso, essas taxas aumentam à medida que os tamanhos amostrais e o parâmetro θ aumentam. Por outro lado, o critério A^* apresenta desempenho superior em todo o cenário, enquanto os critérios AIC, AICc, BIC e HQIC apresentaram o menor desempenho.

Tabela 5.3: Taxas de acertos para o cenário 3.

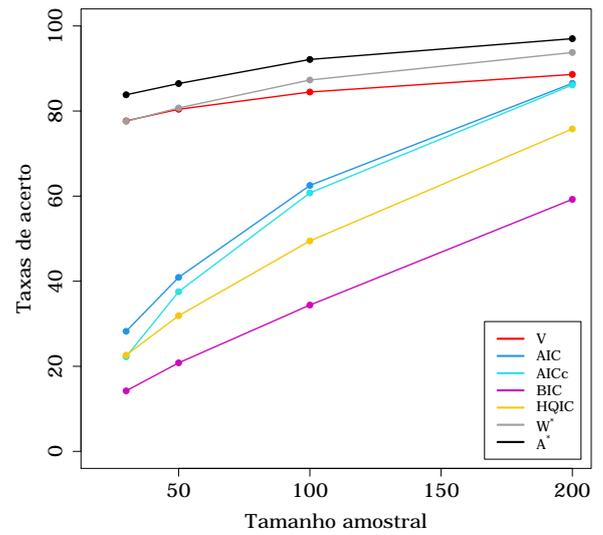
θ	n	V	AIC	AICc	BIC	HQIC	W^*	A^*
2.0	30	74.760	23.120	17.850	10.880	18.115	75.595	82.035
	50	76.540	32.010	28.740	14.365	23.400	77.585	83.670
	100	78.925	47.860	46.200	21.870	35.200	82.775	88.200
	200	81.800	71.030	70.370	37.425	56.410	88.850	93.500
2.5	30	77.690	28.225	22.250	14.220	22.620	77.575	83.825
	50	80.410	40.900	37.530	20.805	31.880	80.720	86.460
	100	84.470	62.515	60.750	34.390	49.465	87.290	92.125
	200	88.605	86.515	86.120	59.240	75.800	93.760	97.010
3.0	30	79.680	32.800	26.410	17.180	26.735	79.290	85.090
	50	83.405	48.140	44.460	26.765	38.760	83.310	88.625
	100	87.895	73.045	71.535	45.295	60.800	90.130	94.420
	200	92.620	93.305	93.005	73.815	86.625	96.165	98.510

Ainda com relação a Tabela 5.3, notamos que V apresentou o segundo desempenho superior para amostras de tamanho n igual a 30, 50 e θ igual a 3.0. Vale ressaltar, que para esses tamanhos amostrais os critérios AIC, AICc, BIC e HQIC apresentaram baixo desempenho, sendo o critério BIC o que apresenta menor desempenho, pois, por exemplo, para n igual a 30 e θ igual a 2.0 a taxa de acerto foi de 10.880%. Por outro lado, percebemos um crescimento significativo das taxas de acerto desses critérios quando o tamanho da amostra n é maior que 50 e θ aumenta, principalmente os critérios AIC e AICc que apresentam taxas de acerto cada vez mais próximas à medida que o tamanho amostral aumenta.

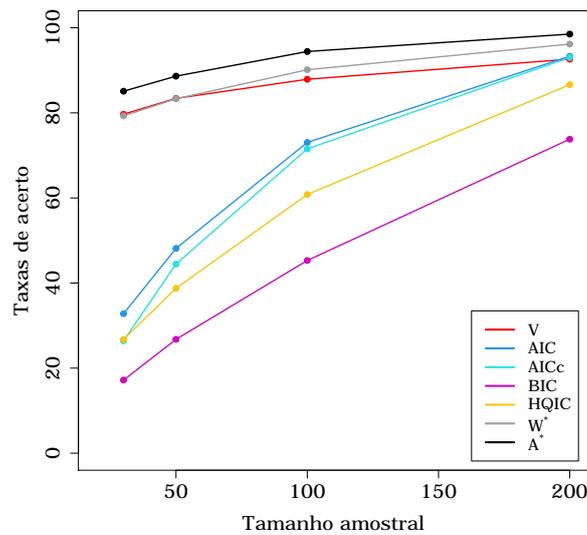
A Figura 5.6 ilustra a situação apresentada na Tabela 5.3. Os gráficos, mostram a comparação dos critérios confirmando a análise feita nesse cenário.



(a) $\theta = 2.0$



(b) $\theta = 2.5$



(c) $\theta = 3.0$

Figura 5.6: Taxas de acertos para o cenário 3.

5.4 Cenário 4: Distribuição Normal \times Distribuição Normal

Neste cenário, o objetivo é comparar o modelo $N(\mu_1, \sigma_1^2)$ com o modelo $N(\mu_2, \sigma_2^2)$, em que o parâmetro σ_2 é conhecido. Sendo assim, os valores de σ_2 foram selecionados

com o objetivo de estudar o desempenho dos critérios e detectar diferenças entre os dois modelos.

Para este cenário de simulação, os dados foram obtidos a partir do modelo $N(\mu_1, \sigma_1^2)$, isto é, o modelo Normal 1 é o correto. O valor do parâmetro σ_2 variou em 2.1, 2.3 e 2.5, enquanto que os valores dos parâmetros μ_1 e σ_1 foram fixados em 12.0 e 2.0, respectivamente.

Para uma visualização das diferenças entre os modelos a Figura 5.7 mostra as densidades do modelo Normal 1 e do modelo Normal 2, $N(\mu_2, \sigma_2^2)$, para as variações do parâmetro σ_2 considerados neste cenário.

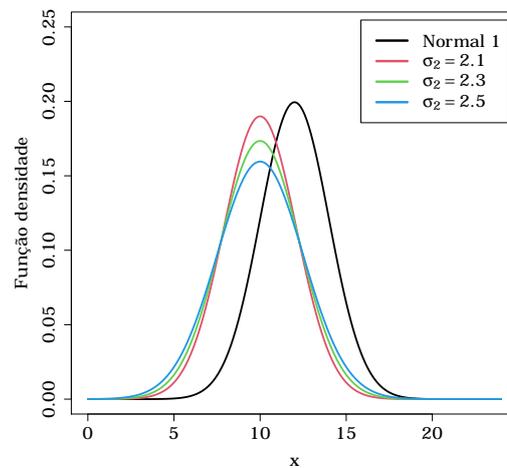


Figura 5.7: Densidades do modelo Normal 1 versus Normal 2.

Na Tabela 5.4 apresentamos as taxas de acerto para este cenário. Avaliando essa tabela, observamos que o critério V mostra um desempenho superior em todo o cenário quando comparado aos demais critérios. Observe também, que as taxas de acerto aumentam à medida que os tamanhos amostrais e o parâmetro σ_2 aumentam.

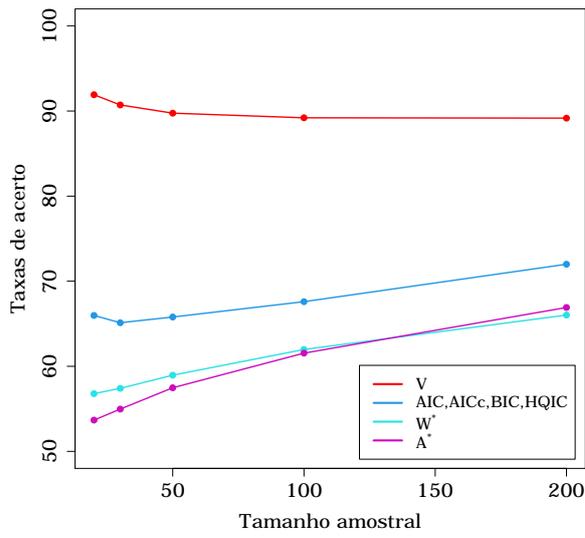
Ainda analisando a Tabela 5.4, podemos notar que os critérios AIC, AICc, BIC e HQIC, apresentam desempenhos equivalentes em todos os casos. Vale destacar, que, especialmente para $\sigma_2 = 2.3, 2.5$, esses critérios mostram as segundas maiores taxas de acerto em todos os tamanhos amostrais.

Tabela 5.4: Taxas de acertos para o cenário 4.

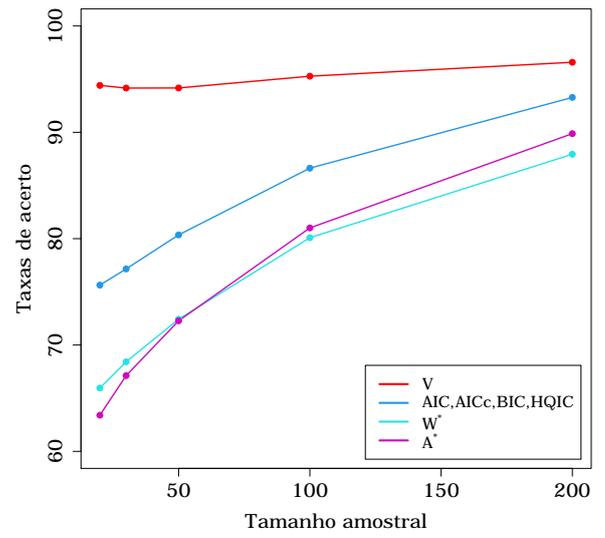
σ_2	n	V	AIC	AICc	BIC	HQIC	W^*	A^*
2.1	20	91.915	65.975	65.975	65.975	65.975	56.785	53.675
	30	90.715	65.120	65.120	65.120	65.120	57.405	54.970
	50	89.750	65.790	65.790	65.790	65.790	58.955	57.475
	100	89.205	67.595	67.595	67.595	67.595	61.975	61.545
	200	89.165	71.995	71.995	71.995	71.995	66.020	66.910
2.3	20	94.410	75.635	75.635	75.635	75.635	65.955	63.400
	30	94.160	77.155	77.155	77.155	77.155	68.420	67.125
	50	94.170	80.355	80.355	80.355	80.355	72.430	72.275
	100	95.275	86.630	86.630	86.630	86.630	80.090	81.010
	200	96.585	93.280	93.280	93.280	93.280	87.940	89.875
2.5	20	96.345	82.940	82.940	82.940	82.940	73.595	72.010
	30	96.565	85.615	85.615	85.615	85.615	77.575	76.880
	50	97.345	89.600	89.600	89.600	89.600	82.900	83.250
	100	98.695	95.740	95.740	95.740	95.740	90.875	92.165
	200	99.625	98.835	98.835	98.835	98.835	97.115	97.920

A Figura 5.8 ilustra a situação na Tabela 5.4. Os gráficos, confirmam a análise feita da tabela com relação ao desempenho do critério V .

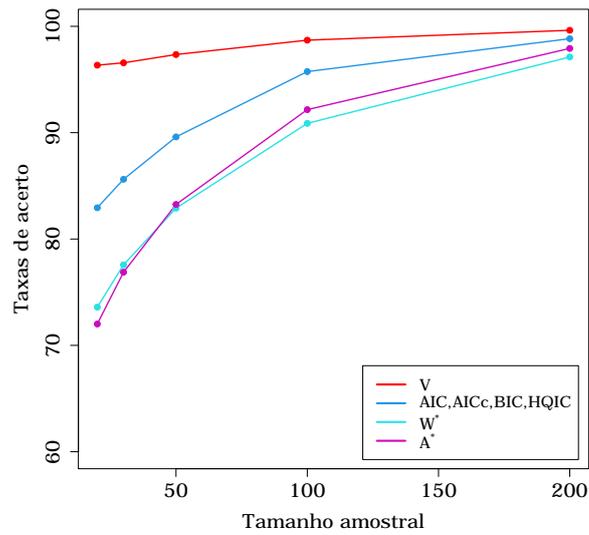
De modo geral, verificamos nessas simulações que o critério proposto mostrou um melhor desempenho em relação aos demais critérios para um tamanho amostral pequeno ou moderado quando consideradas pequenas diferenças entre as distribuições. Além disso, houve um aumento nas taxas de acerto à medida que os parâmetros e o tamanho amostral aumentaram. Outro fato, foi que na última simulação, o critério proposto mostrou um desempenho superior em todos os tamanhos amostrais quando considerada a distribuição normal.



(a) $\sigma_2 = 2.1$



(b) $\sigma_2 = 2.3$



(c) $\sigma_2 = 2.5$

Figura 5.8: Taxas de acertos para o cenário 4.

Capítulo 6

Aplicações

Neste capítulo ilustraremos a aplicabilidade do critério proposto através de conjuntos de dados reais. Para fins de comparação, obtemos o valor do novo critério V , e dos critérios AIC, AICc, BIC, HQIC, W^* e A^* para verificarmos, dentre um conjunto de modelos concorrentes, qual que melhor se ajusta aos dados. Todos os cálculos foram realizados usando o software `0x`.

Todas as distribuições consideradas nesse capítulo estão mencionadas no Apêndice.

6.1 Aplicação 1

Como uma primeira aplicação, consideramos um conjunto de dados fornecido em [Murthy, Xie, e Jiang \(2004\)](#) correspondente a 20 observações referente a vinte itens testados até a falha, os quais são apresentados na Tabela 6.1 e referenciados como Dados I.

Tabela 6.1: Dados I: Vinte itens testados até a falha.

11.24	11.50	30.42	38.14	1.92	8.86	9.17	2.99	12.74	7.75
10.20	16.58	22.48	5.73	5.52	18.92	9.60	9.37	5.85	13.36

Os valores calculados dos critérios para Dados I são fornecidos na Tabela 6.2 . Para ajustar os dados, consideramos a distribuição Fréchet (Fr) ([FRÉCHET, 1928](#)), Weibull (W), Lindley Ponderada (LP), Fréchet Modificada (FM) ([TABLADA, 2017](#)) e a Fréchet Exponencializada (FE) ([NADARAJAH e KOTZ, 2001](#)).

Tabela 6.2: Critérios de adequação de ajuste para Dados I.

Distribuição	Critérios						
	V	AIC	AICc	BIC	HQIC	W^*	A^*
$Fr(\hat{\alpha}, \hat{\beta})$	9.6084	143.81	144.52	145.81	144.20	1.8345	0.80957
$W(\hat{\alpha}, \hat{\beta})$	0.13903	139.88	140.58	141.87	140.27	1.7822	0.44412
$LP(\hat{\sigma}, \hat{\mu})$	0.052741	138.94	139.64	140.93	139.33	1.7627	0.34222
$FM(\hat{\alpha}, \hat{\beta}, \hat{\lambda})$	0.046165	140.33	141.83	143.31	140.91	1.7492	0.27470
$FE(\hat{\sigma}, \hat{\eta}, \hat{\rho})$	0.049950	140.34	141.84	143.33	140.92	1.7480	0.26703

Os gráficos na Figura 6.1 mostram as densidades estimadas de FE, FM e LP.

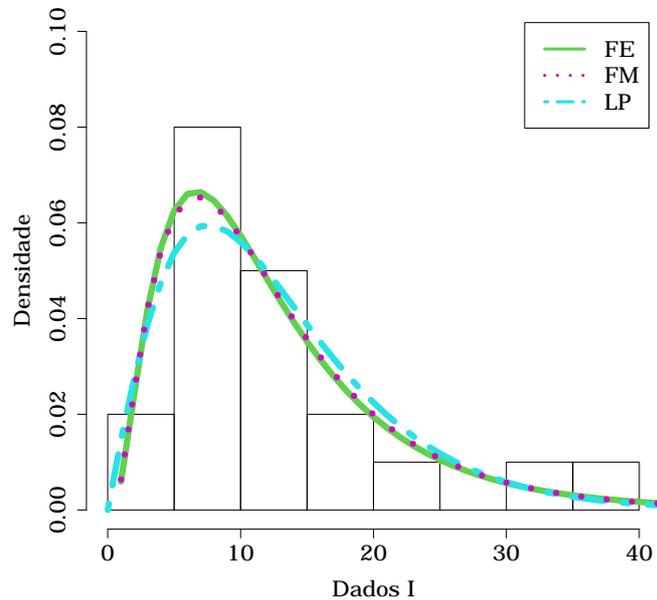


Figura 6.1: Densidades estimadas de FE, FM e LP para Dados I.

Os valores da Tabela 6.2 revelam que a distribuição FM apresenta o menor valor do critério V entre os modelos ajustados, enquanto que a distribuição FE apresenta os menores valores de W^* e A^* . Já os critérios AIC, AICc, BIC e HQIC apresentaram o menor valor para a distribuição LP, pois eles tendem a selecionar, dentre os modelos concorrentes, aqueles com menos parâmetros. Com isso, temos que, as distribuições FM, FE e LP apresentam o melhor ajuste aos dados, segundo os critérios utilizados.

Note que se considerarmos somente os modelos com dois parâmetros, os critérios V , W^* e A^* também selecionariam o modelo LP.

Uma análise gráfica da qualidade do ajuste pode ser avaliada por meio de Q-Q plots. Dessa forma, na Figura 6.2 são dados os Q-Q plots para as distribuições apresentadas na Figura 6.1.

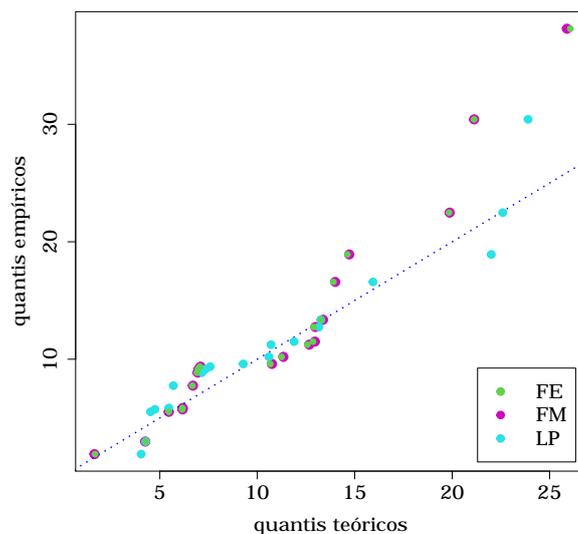


Figura 6.2: Gráficos Q-Q plots para as distribuições FE, FM e LP.

A partir desses gráficos, podemos observar que, em relação as distribuições FM e FE elas mostram um ajuste semelhante, como mostrado no histograma. Já, com relação a LP, podemos perceber que ela ajusta melhor na cauda do que a FM e FE. Esse ajuste pode ser constatado ao compararmos com a análise quantitativa fornecida pelo critério V , pois essas distribuições foram as que apresentaram os menores valores calculados para o conjunto de dados.

6.2 Aplicação 2

Nessa aplicação consideramos o conjunto de Dados II, o qual consiste em 23 observações sobre tempo de vida (em milhões) de rolamentos de esferas de revoluções (LIEBLEIN e ZELEN, 1956).

Tabela 6.3: Dados II: Tempo de vida (em milhões) de rolamentos de esferas de revoluções.

17.88	28.92	33.00	41.52	42.12	45.60	48.48	51.84	51.96	54.12
55.56	67.80	68.64	68.64	68.88	84.12	93.12	98.64	105.12	105.84
127.92	128.04	173.40							

Os valores calculados dos critérios para o conjunto de Dados II são fornecidos na Tabela 6.4. Para ajustar os dados, consideramos a distribuição Weibull (W), Lindley estendida de Marshall-Olkin (LEMO) (MARSHALL e OLKIN, 1997) ou Lindley Geométrica (ZAKERZADEH e MAHMOUDI, 2012), Weibull estendida de Marshall-Olkin (WEMO) (MARSHALL e OLKIN, 1997), Weibull Exponencializada (WE) e

Fréchet tipo II (Fr tipo II) (JENKINSON, 1955).

Tabela 6.4: Critérios de adequação para Dados II

Distribuição	Critérios						
	V	AIC	AICc	BIC	HQIC	W^*	A^*
$W(\hat{\alpha}, \hat{\beta})$	0.35159	231.38	231.98	233.65	231.95	2.0139	0.34072
LEMO($\hat{\theta}, \hat{\alpha}$)	0.79891	232.01	232.61	234.28	232.58	2.0094	0.34928
WEMO($\hat{\alpha}, \hat{\beta}, \hat{\lambda}$)	0.033905	231.86	233.12	235.27	232.72	1.9895	0.20402
WE($\hat{\theta}, \hat{\alpha}, \hat{\sigma}$)	0.033859	231.95	233.21	235.36	232.81	1.9883	0.19685
Fr tipo II($\hat{\alpha}, \hat{\beta}, \hat{\lambda}$)	0.036053	232.20	233.47	235.61	233.06	1.9893	0.20152

Os valores da Tabela 6.4 revelam que a distribuição WE apresenta os menores valores para os critérios V , W^* e A^* . Como os critérios AIC, AICc, BIC e HQIC penalizam a quantidade de parâmetros, o modelo que apresentou o menor valor calculado foi W, desse fato, veja que, se considerados os modelos com dois parâmetros o critério V também selecionaria esse modelo.

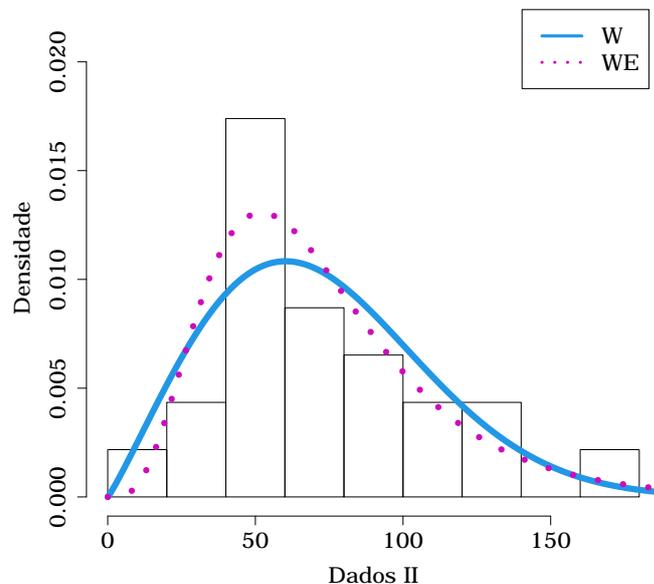


Figura 6.3: Densidades estimadas de W e WE para Dados II.

Os gráficos na Figura 6.3 mostram as densidades estimadas das distribuições que apresentaram o melhor ajuste aos dados atuais, segundo os critérios considerados. Com base nesses gráficos, é possível avaliar um ajuste geral das distribuições.

Uma análise gráfica da qualidade do ajuste pode ser avaliada por meio de gráficos Q-Q plots (Figura 6.4) para as distribuições apresentadas na Figura 6.3.

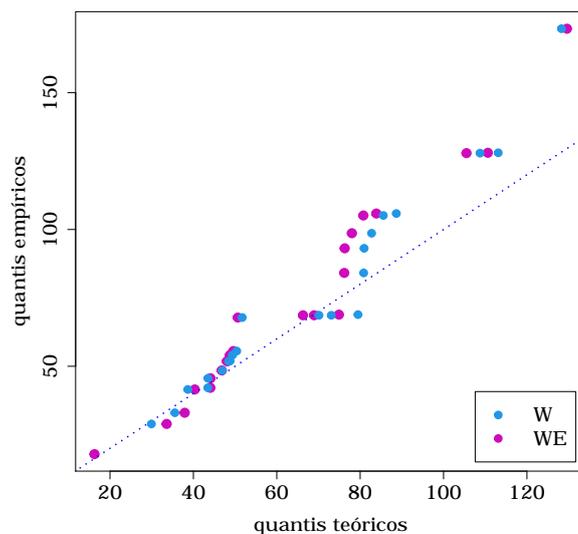


Figura 6.4: Gráficos Q-Q plots para as distribuições W e WE.

Observe que esses gráficos indicam que as distribuições W e WE apresentam um ajuste semelhante aos dados. Analisando os valores calculados dos critérios, podemos perceber que AIC, AICc, BIC e HQIC e W^* não mostram uma diferença significativa na seleção de uma distribuição e outra, dentre o conjunto de modelos considerados. Porém se observamos os critérios V e A^* essa diferença é maior.

6.3 Aplicação 3

Nessa aplicação consideramos o conjunto de Dados III, o qual consiste em 30 observações referentes aos fluxos anuais do rio Weldon em Mill Grove, Missouri (1930-1959) (VUJICA, 1972).

Tabela 6.5: Dados III: Fluxos anuais do rio Weldon em Mill Grove, Missouri (1930-59).

108.0	472.0	143.0	441.0	244.0	132.0	53.6	96.5	93.7	386.0
400.0	44.0	585.0	217.0	398.0	567.0	245.0	72.5	98.1	42.7
298.0	122.0	114.0	135.0	40.6	208.0	248.0	151.0	659.0	635.0

Os valores calculados dos critérios para o conjunto de Dados III são fornecidos na Tabela 6.6. Para ajustar os dados, consideramos a distribuição Fréchet (Fr), Weibull (W), Lindley Ponderada (LP), Weibull Exponencializada (WE) e Fréchet Exponencializada (FE).

Tabela 6.6: Critérios de adequação para Dados III

Distribuição	Critérios						
	V	AIC	AICc	BIC	HQIC	W^*	A^*
$Fr(\hat{\alpha}, \hat{\beta})$	2.5952	393.73	394.17	396.53	394.63	2.6156	0.62168
$W(\hat{\alpha}, \hat{\beta})$	1.2934	390.78	391.23	393.58	391.68	2.6239	0.53150
$LP(\hat{\sigma}, \hat{\mu})$	0.85703	390.38	390.83	393.19	391.28	2.6221	0.52250
$WE(\hat{\theta}, \hat{\alpha}, \hat{\sigma})$	0.88672	391.99	392.91	396.19	393.33	2.6015	0.43012
$FE(\hat{\sigma}, \hat{\eta}, \hat{\rho})$	0.83003	391.83	392.75	396.03	393.18	2.6044	0.43851

Segundo a Tabela 6.6, o critério V indicou o modelo FE como o que melhor se ajusta aos dados, dentre os modelos considerados. Enquanto os critérios AIC, AICc, BIC e HQIC selecionaram o modelo LP, e os critérios W^* e A^* selecionaram o modelo WE. Analisando os valores calculados, podemos destacar que se considerados os modelos com dois parâmetros, o critério V também selecionaria o modelo LP. Analisando a tabela de modo geral, podemos destacar que os modelos selecionados pelos critérios AIC, AICc, BIC, HQIC, W^* e A^* foram os que, para o critério V , apresentaram os menores valores.

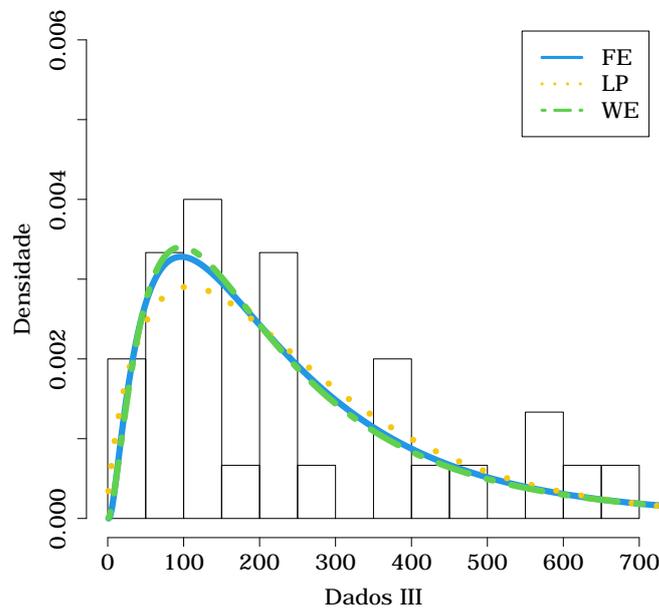


Figura 6.5: Densidades estimadas de FE, LP e WE para Dados III.

Os gráficos na Figura 6.5 mostram as densidades estimadas das distribuições que apresentaram o melhor ajuste aos dados atuais, segundo os critérios considerados. Com base nesses gráficos, é possível avaliar um ajuste geral das distribuições.

Uma análise gráfica da qualidade do ajuste pode ser avaliada por meio de gráficos Q-Q plots na Figura 6.6, para as distribuições apresentadas na Figura 6.5.

A partir desses gráficos, podemos observar que as distribuições FE e WE mostram um ajuste semelhante, como mostrado no histograma. Com relação a distribuição LP, podemos observar que o ajuste é melhor na cauda, comparada a FE e WE. Dessa forma, essa análise gráfica é consistente com a análise da tabela, feita anteriormente, de que os modelos apresentam um bom ajuste aos dados e bem próximo um do outro.

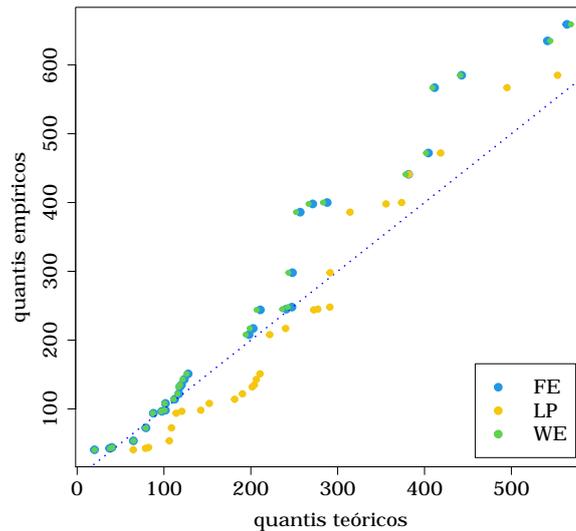


Figura 6.6: Gráficos Q-Q plots para as distribuições FE, LP e WE.

6.4 Aplicação 4

Nessa aplicação consideramos o conjunto Dados IV, o qual consiste em 50 observações sobre rebarbas (em milímetros) (DASGUPTA, 2011).

Tabela 6.7: Dados IV: Rebarbas (em milímetros).

0.04	0.02	0.06	0.12	0.14	0.08	0.22	0.12	0.08	0.26
0.24	0.04	0.14	0.16	0.08	0.267	0.32	0.28	0.14	0.16
0.24	0.22	0.12	0.18	0.24	0.32	0.16	0.14	0.08	0.16
0.24	0.16	0.32	0.18	0.24	0.22	0.16	0.12	0.24	0.06
0.02	0.18	0.22	0.14	0.06	0.04	0.14	0.26	0.18	0.16

Nessa aplicação, consideramos distribuições com o mesmo número de parâmetros com o objetivo de comparar todos os critérios de forma mais justa. Os valores calculados dos critérios para Dados IV são fornecidos na Tabela 6.8. Para ajustar os dados, consideramos a distribuição Burr XII (BXII) (BURR, 1942), Gama, Gompertz (Go) (GOMPERTZ, 1825), Weibull (W) e a Kumaraswamy (Kw) (KUMARASWAMY, 1980).

Tabela 6.8: Critérios de adequação para Dados IV

Distribuição	Critérios						
	V	AIC	AICc	BIC	HQIC	W^*	A^*
BXII($\hat{\alpha}, \hat{\beta}$)	0.80294	-107.47	-107.22	-103.65	-106.02	4.3165	0.74740
Gama($\hat{\sigma}, \hat{\mu}$)	3.5786	-102.74	-102.49	-98.919	-101.29	4.3972	1.1290
Go($\hat{\alpha}, \hat{\beta}$)	0.68557	-110.15	-109.90	-106.33	-108.70	4.2805	0.44665
W($\hat{\alpha}, \hat{\beta}$)	0.73312	-107.78	-107.53	-103.96	-106.33	4.3143	0.71962
Kw($\hat{\theta}, \hat{\alpha}$)	0.65527	-108.14	-107.88	-104.31	-106.68	4.3119	0.68841

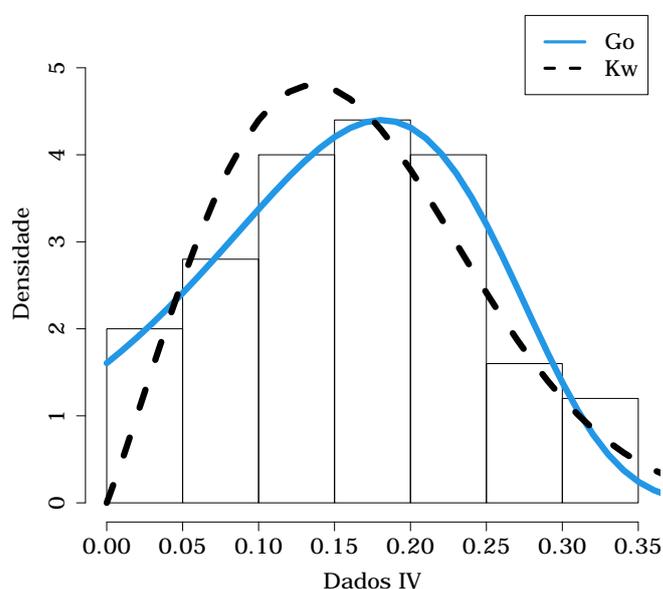


Figura 6.7: Densidades estimadas de Go e Kw para Dados IV.

Fazendo uma comparação geral dos critérios, os números da Tabela 6.8 revelam que a distribuição Go apresenta o menor valor para os critérios AIC, AICc, BIC, HQIC, W^* e A^* . Enquanto a distribuição Kw apresenta o menor valor para o critério V , entre os modelos ajustados. Podemos notar também, que esses dois modelos são os que, de acordo com o critério V , apresentaram os menores valores calculados.

Os gráficos na Figura 6.7 mostram as densidades estimadas das distribuições que apresentaram o melhor ajuste aos dados atuais, segundo os critérios considerados.

Uma análise gráfica da qualidade do ajuste pode ser avaliada por meio de gráficos Q-Q plots na Figura 6.8 para as distribuições apresentadas na Figura 6.7.

A partir desses gráficos, podemos observar que ambas distribuições apresentam um ajuste similar para os dados reais atuais.

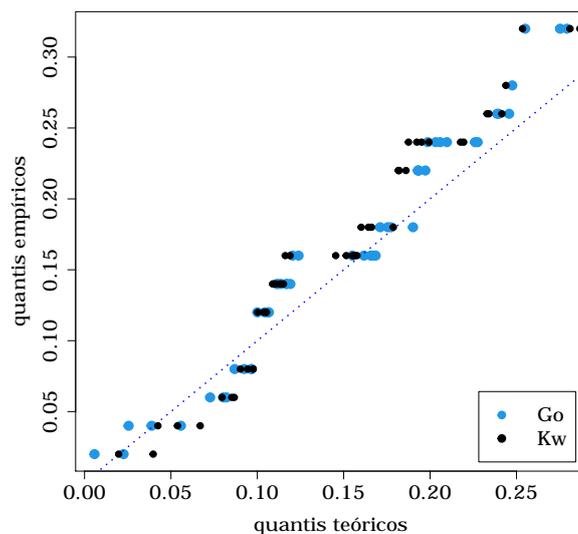


Figura 6.8: Gráficos Q-Q plots para as distribuições Go e Kw.

6.5 Aplicação 5

Nessa aplicação consideramos o conjunto Dados V , o qual consiste em 56 observações sobre resistências (medido em GPA) de fibras de carbono individuais, testadas sob tensão (RISTIĆ e KUNDU, 2015).

Tabela 6.9: Dados V : resistência de fibras de carbono (em GPA).

2.247	2.64	2.908	3.099	3.126	3.245	3.328	3.355	3.383	3.572
3.581	3.681	3.726	3.727	3.728	3.783	3.785	3.786	3.896	3.912
3.964	4.05	4.063	4.082	4.111	4.118	4.141	4.246	4.251	4.262
4.326	4.402	4.457	4.466	4.519	4.542	4.555	4.614	4.632	4.634
4.636	4.678	4.698	4.738	4.832	4.924	5.043	5.099	5.134	5.359
5.473	5.571	5.684	5.721	5.998	6.06				

Os valores calculados dos critérios para o conjunto Dados V são fornecidos na Tabela 6.10. Para ajustar os dados, consideramos a distribuição Weibull (W), Lindley Ponderada (LP), Gama, Weibull Exponencializada (WE) e Fréchet Exponencializada (FE).

Da Tabela 6.10 vemos que a distribuição FE apresenta o menor valor para o critério V . Por outro lado, a distribuição LP apresentou o menor valor para os critérios AIC, AICc, BIC, HQIC, enquanto isso, a distribuição WE apresentou o menor valor para os critérios W^* e A^* . Analisando quantitativamente, as distribuições WE e FE apresentaram valores muito similares para V , A^* e W^* .

Tabela 6.10: Critérios de adequação para Dados V

Distribuição	Critérios						
	V	AIC	AICc	BIC	HQIC	W*	A*
$W(\hat{\alpha}, \hat{\beta})$	0.20148	141.87	142.09	145.92	143.44	4.7755	0.44808
$LP(\hat{\sigma}, \hat{\mu})$	0.33474	140.75	140.97	144.80	142.32	4.7350	0.19669
Gama($\hat{\sigma}, \hat{\mu}$)	0.33581	140.76	140.99	144.81	142.33	4.7351	0.19730
$WE(\hat{\theta}, \hat{\alpha}, \hat{\sigma})$	0.17284	141.75	142.21	147.82	144.10	4.7310	0.17321
$FE(\hat{\sigma}, \hat{\eta}, \hat{\rho})$	0.16149	141.84	142.30	147.92	144.19	4.7333	0.18808

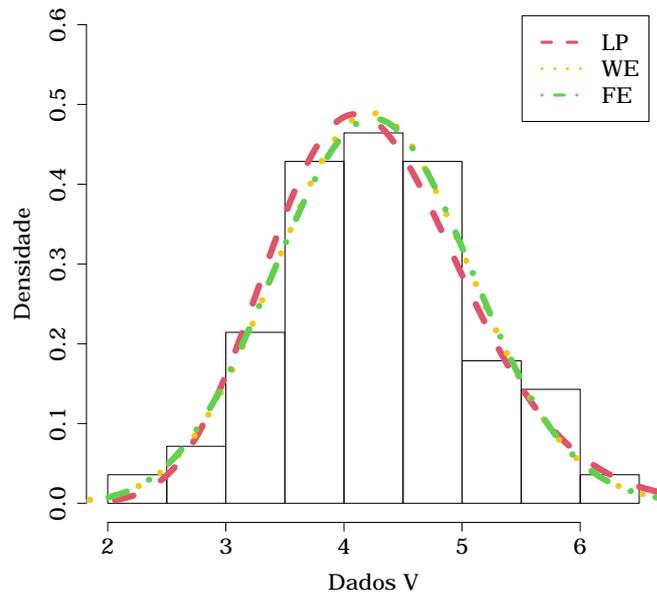


Figura 6.9: Densidades estimadas de LP, WE e FE para Dados V.

Os gráficos na Figura 6.9 mostram as densidades estimadas das distribuições que apresentaram o melhor ajuste aos dados atuais, segundo os critérios considerados. Com base nesses gráficos, é possível observarmos que as curvas das distribuições LP, WE e FE apresentam um ajustamento semelhante aos dados.

Uma análise gráfica da qualidade do ajuste pode ser avaliada por meio de gráficos Q-Q plots na Figura 6.10 para as distribuições apresentadas na Figura 6.9.

A partir dos gráficos da Figura 6.10, podemos observar que as distribuições LP, WE e FE mostram um ajuste semelhante, sendo que a LP parece mostrar um melhor ajuste na cauda.

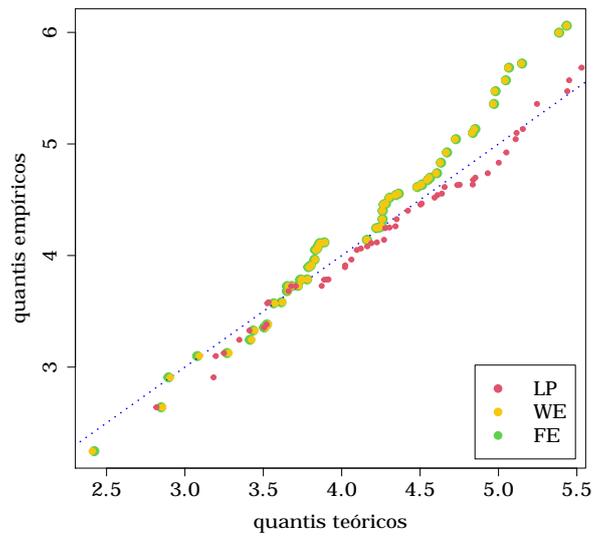


Figura 6.10: Gráficos Q-Q plots para as distribuições LP, WE e FE.

Os resultados das aplicações mostram, que o critério V teve um desempenho esperado em relação aos demais critérios. De forma geral, os dois ou três menores valores de V corresponderam aos melhores ajustes, segundo os outros critérios.

Capítulo 7

Regressão Linear: Estudo de Simulação e Aplicação

Neste capítulo, objetivamos comparar os desempenhos dos critérios V, AIC, AICc, BIC e HQIC na seleção de modelos de regressão em amostras finitas ($n = 20, 30, 50$ e 100). Para isso, realizamos um estudo de simulação de Monte Carlo (20 000 réplicas) utilizando a linguagem de programação `0x`. O desempenho de cada critério foi avaliado com base na porcentagem de acerto, isto é, no percentual de vezes que o modelo correto foi selecionado. Além disso, através de um conjunto de dados, ilustramos o uso dos critérios V, AIC, BIC e \bar{R}^2 ajustado para avaliar os modelos.

Para todos os cenários os valores observados das variáveis explicativas, x_{ij} , foram gerados em cada réplica independentemente com distribuição exponencial de parâmetro 3, $\text{Exp}(3)$, sendo $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, 4$. Os parâmetros de regressão, β 's, foram fixados em 1 e assumimos que o parâmetro σ é conhecido. Todos os modelos considerados no estudo incluem o intercepto β_0 . Os valores observados da variável dependente, $y_i = (y_1, \dots, y_n)^T$, foram gerados a partir da distribuição $N(\mu, \sigma^2)$, em que $\mu = \mathbf{X}\beta$.

O estudo de simulação foi realizado considerando dois cenários, os quais serão descritos a seguir.

7.1 Cenário 1

Neste cenário, o objetivo é comparar os modelos, M_1 e M_2 , ambos com a mesma quantidade de variáveis explicativas e estudar o desempenho dos critérios para detectar pequenas diferenças entre os dois modelos ajustados quando o parâmetro σ varia. Os modelos foram especificados da seguinte forma:

$$M_1 : y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i, \quad i = 1, 2, \dots, n$$

e

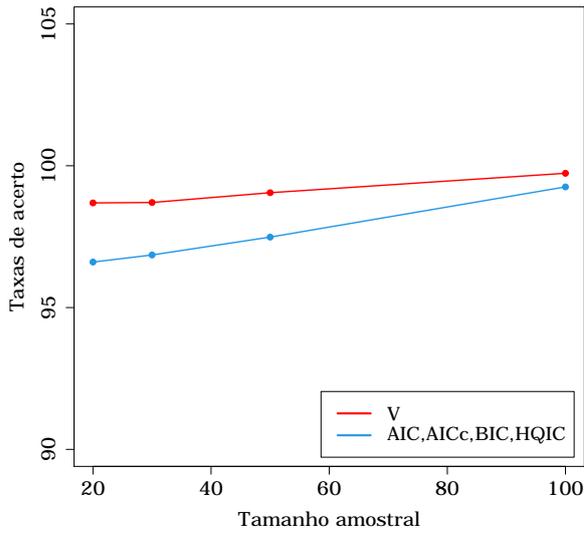
$$M_2 : y_i = \gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \gamma_3 x_{i3} + \gamma_4 x_{i4} + \bar{\varepsilon}_i, \quad i = 1, 2, \dots, n.$$

Considerando-se o modelo M_1 verdadeiro, isto é, o modelo correspondente ao processo gerador dos dados, o valor para σ_1 foi fixado em 1.5, enquanto para o modelo M_2 , alternativo, o valor de σ_2 variou em 2, 3 e 5.

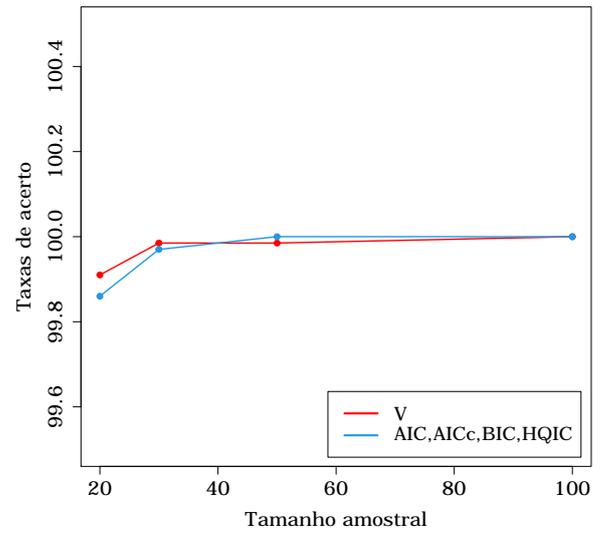
A Tabela 7.1 apresenta as taxas de acertos para o Cenário 1. De modo geral, todos os critérios apresentaram taxas de acerto próximas de 100%. No entanto, nota-se que o critério V tem desempenho ligeiramente superior aos demais para amostra de tamanho pequeno a moderado. Por exemplo, para $n = 20$ e $\sigma_2 = 2$, tem-se que as taxas de acerto dos critérios V , AIC, AICc, BIC e HQIC foram, respectivamente, 98.690, 96.605, 96.605, 96.605 e 96.605.

Tabela 7.1: Taxas de acertos para o cenário 1.

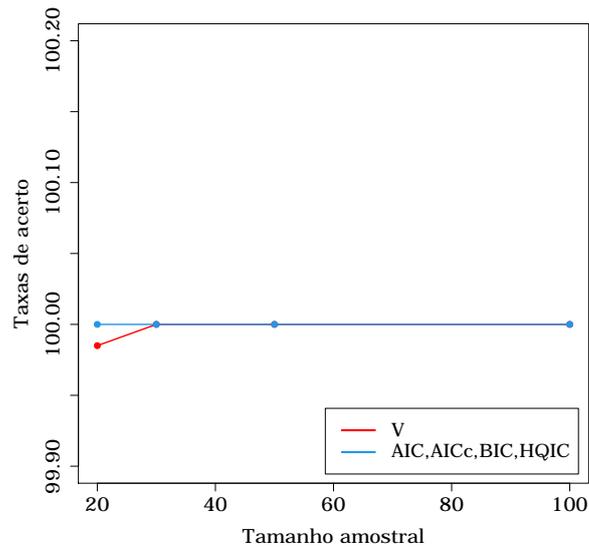
n	σ_2	V	AIC	AICc	BIC	HQIC
20	2	98.690	96.605	96.605	96.605	96.605
	3	99.910	99.860	99.860	99.860	99.860
	5	99.985	100.00	100.00	100.00	100.00
30	2	98.705	96.855	96.855	96.855	96.855
	3	99.985	99.970	99.970	99.970	99.970
	5	100.00	100.00	100.00	100.00	100.00
50	2	99.050	97.485	97.485	97.485	97.485
	3	99.985	100.00	100.00	100.00	100.00
	5	100.00	100.00	100.00	100.00	100.00
100	2	99.735	99.255	99.255	99.255	99.255
	3	100.00	100.00	100.00	100.00	100.00
	5	100.00	100.00	100.00	100.00	100.00



(a) $\sigma_2 = 2.0$



(b) $\sigma_2 = 3.0$



(c) $\sigma_2 = 5.0$

Figura 7.1: Taxas de acertos para o Cenário 1.

A Figura 7.1 ilustra a situação apresentada na Tabela 7.1. Os gráficos mostram que, para σ_2 fixo, a taxa de acerto de todos os critérios se aproximam de 100% conforme o tamanho da amostra aumenta. Ainda, também é possível perceber o desempenho superior do critério V nas situações em que a amostra é pequena.

7.2 Cenário 2

Neste cenário, o objetivo é comparar os modelos, M_1 e M_2 , com um controle na quantidade de parâmetros, e estudar o desempenho dos critérios entre os dois modelos ajustados. Os modelos foram especificados da seguinte forma:

$$M_1 : y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i, \quad i = 1, 2, \dots, n.$$

e

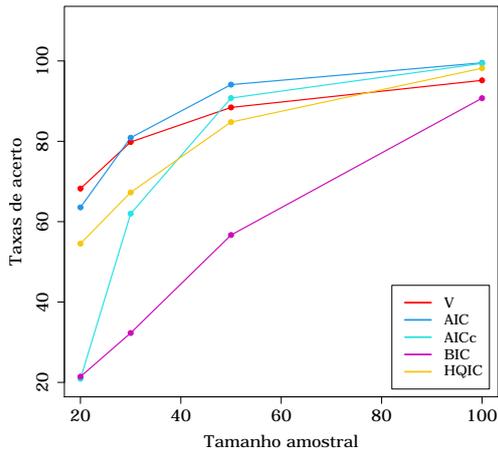
$$M_2 : \begin{cases} y_i = \gamma_0, & \text{se } k = 0, \\ y_i = \gamma_0 + \sum_{j=1}^k \gamma_j x_{ij} + \varepsilon'_i, & \text{se } k = 1, 2, 3, 4, \quad i = 1, 2, \dots, n. \end{cases}$$

As simulações foram realizadas assumindo que o modelo M_1 é verdadeiro, isto é, o modelo correspondente ao processo gerador dos dados, o valor para σ_1 foi fixado em 1.5, enquanto para o modelo M_2 , alternativo, o valor de σ_2 foi fixado em 2 e a quantidade de parâmetros no modelo, k , variou em 0, 1, 2, 3, 4.

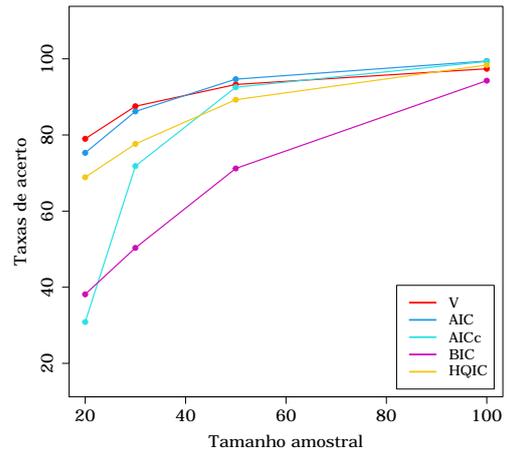
A Tabela 7.2 apresenta as taxas de acertos para este Cenário. Avaliando essa tabela, observamos que para todos os tamanhos amostrais, o critério V continua apresentando taxa de acerto maior que 50%, além disso, para $n = 20$ mostrou um desempenho superior para todo k , comparado aos demais critérios. Ainda com relação ao critério V , podemos notar que a porcentagem em que o modelo verdadeiro é selecionado pelo critério aumenta a medida que o tamanho amostral cresce e a quantidade de parâmetros do modelo alternativo se aproxima da quantidade de parâmetros do modelo verdadeiro. Analisando somente os critérios AIC, AICc, BIC e HQIC, a tabela mostra o critério AIC como o que mais seleciona o modelo verdadeiro para $k \leq 3$, e o critério BIC o que apresenta as menores taxas de acerto para $n \geq 30$ e $k \leq 3$. Ainda podemos observar que o critério AIC apresenta um desempenho superior comparado ao critério V em alguns casos em que $n \geq 50$ e $k \leq 2$. Por exemplo, para $n = 50$ e $k = 0$ as taxas de acerto dos critérios V e AIC foram, respectivamente, 88.430 e 94.105. Essas conclusões podem ser observadas graficamente na Figura 7.2.

Tabela 7.2: Taxas de acertos para o cenário 2.

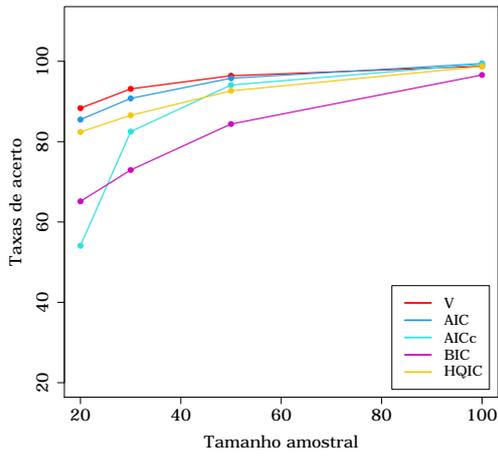
n	k	V	AIC	AICc	BIC	HQIC
20	0	68.235	63.555	20.895	21.425	54.530
	1	79.000	75.310	30.865	38.135	68.890
	2	88.330	85.490	54.090	65.130	82.400
	3	95.360	92.555	82.245	87.240	91.620
	4	98.690	96.605	96.605	96.605	96.605
30	0	79.835	80.895	61.995	32.290	67.275
	1	87.545	86.190	71.830	50.335	77.645
	2	93.130	90.765	82.490	72.940	86.555
	3	96.910	94.400	91.300	89.085	93.000
	4	98.705	96.855	96.855	96.855	96.855
50	0	88.430	94.105	90.770	56.670	84.790
	1	93.265	94.665	92.540	71.205	89.260
	2	96.400	95.780	94.090	84.380	92.635
	3	98.210	96.535	95.980	93.090	95.615
	4	99.050	97.485	97.485	97.485	97.485
100	0	95.170	99.535	99.395	90.735	98.190
	1	97.390	99.430	99.305	94.265	98.390
	2	98.770	99.440	99.290	96.590	98.700
	3	99.420	99.335	99.290	98.270	99.000
	4	99.735	99.255	99.255	99.255	99.255



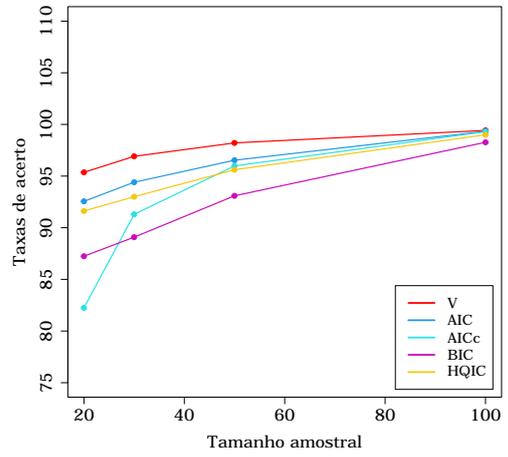
(a) $k = 0$



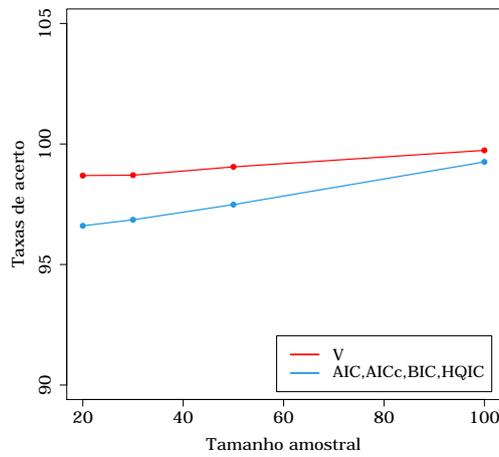
(b) $k = 1$



(c) $k = 2$



(d) $k = 3$



(e) $k = 4$

Figura 7.2: Taxas de acertos para o cenário 2.

Analisando os resultados obtidos nas simulações, verificamos um desempenho superior do critério proposto em relação aos demais critérios na seleção do modelo verdadeiro principalmente para pequenas amostras.

7.3 Aplicação

Consideramos um conjunto de dados reais para avaliar o desempenho do critério proposto na seleção de modelos de regressão.

O conjunto de dados considerado está apresentado na Tabela 7.3 e fornecido em [Montgomery, Peck, e Vining \(2012\)](#). As informações consistem em 25 observações coletadas por um engenheiro industrial no estudo em que um engarrafador de refrigerantes está analisando as rotas de serviço das máquinas de venda automática em seu sistema de distribuição, e está interessado em prever o tempo necessário para o motorista de trajeto atender as solicitações em um ponto de venda. Esta atividade de serviço inclui estocar a máquina com bebidas e pequenas manutenções ou serviços de limpeza. O estudo sugeriu que as duas variáveis mais importantes que afetam o tempo de entrega (y) são o número de caixas de produto estocadas (x_1) e a distância percorrida pelo motorista da rota (x_2).

Tabela 7.3: Dados de tempo de entrega.

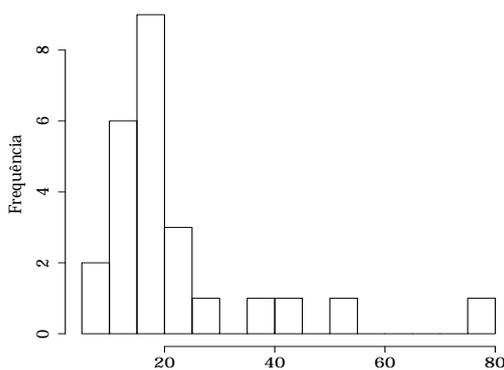
Número de Observação	Tempo de entrega, y (min)	Número de caixas, x_1	Distância, x_2 (pés)
1	16.68	7	560
2	11.50	3	220
3	12.03	3	340
4	14.88	4	80
5	13.75	6	150
6	18.11	7	330
7	8.00	2	110
8	17.83	7	210
9	79.24	30	1460
10	21.50	5	605
11	40.33	16	688
12	21.00	10	215
13	13.50	4	255
14	19.75	6	462
15	24.00	9	448
16	29.00	10	776
17	15.35	6	200
18	19.00	7	132
19	9.50	3	36
20	35.10	17	770
21	17.90	10	140
22	52.32	26	810
23	18.75	9	450
24	19.83	8	635
25	10.75	4	150

Antes de modelarmos os dados, realizaremos uma análise exploratória fazendo uso de técnicas estatísticas como histogramas, boxplot e correlogramas. Essa análise é essencial para a tomada de decisões.

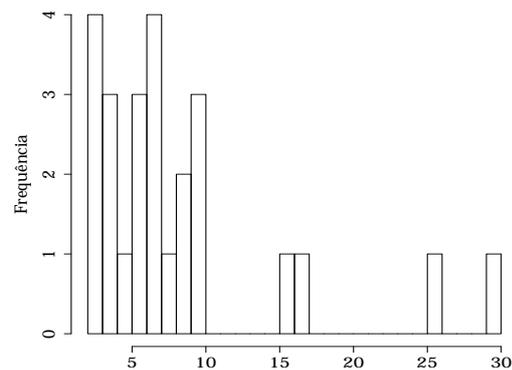
Tabela 7.4: Resumo dos dados.

	Min.	1° Qu.	Mediana	Média	3° Qu.	Max.
Tempo de entrega	8.00	13.75	18.11	22.38	21.50	79.24
Número de caixas	2.00	4.00	7.00	8.76	10.00	30.00
Distância	36.0	150.0	330.0	409.3	605.0	1460.0

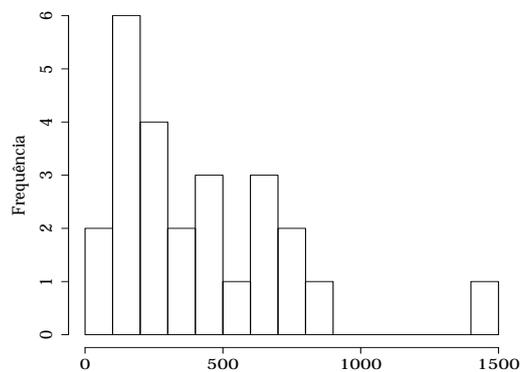
As medidas descritivas apresentadas na Tabela 7.4 mostram aparente assimetria no que diz respeito as variáveis tempo de entrega e distância. Nota-se também a grande variação da variável distância, com valores entre 36.00 e 1460.0.



(a) Tempo de entrega



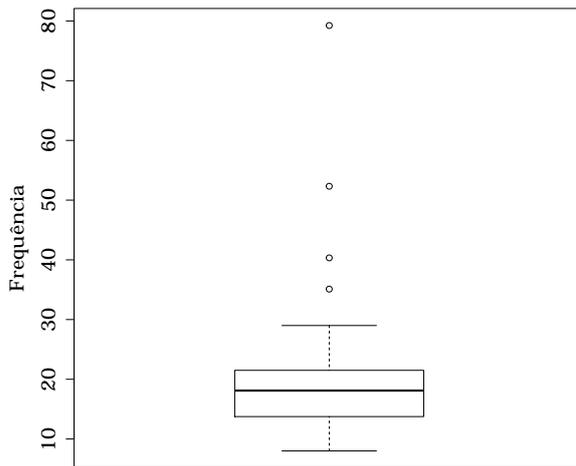
(b) Número de caixas



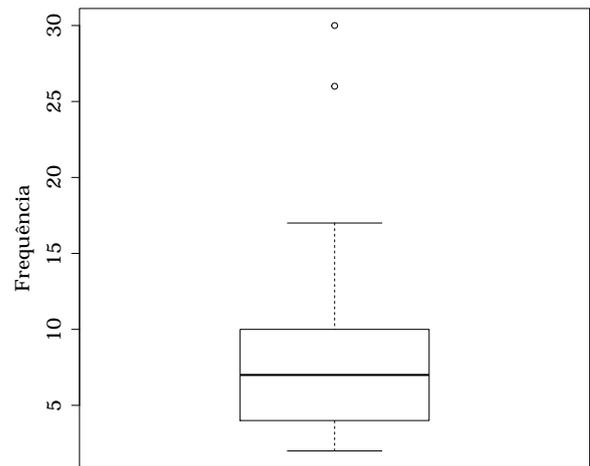
(c) Distância

Figura 7.3: Histograma das variáveis.

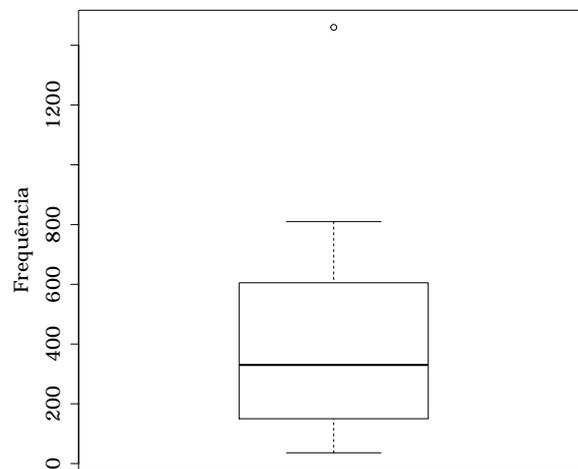
Na Figura 7.3 (a) observa-se que os dados estão mais concentrados ao lado direito onde o tempo de entrega tem a maior frequência, dando indícios de que existe uma violação da suposição de normalidade. Com isso, aplicamos o teste de normalidade de Shapiro-Wilk para essa variável que retornou um $p\text{-valor} < 0,05$. Logo, ao nível nominal de 5% rejeita-se a hipótese de que a variável y segue distribuição normal.



(a) Tempo de entrega



(b) Número de caixas



(c) Distância

Figura 7.4: Boxplot.

A Figura 7.4 apresenta o Boxplot para as variáveis utilizadas no estudo. Analisando a Tabela 7.4 e a Figura 7.4 (a), observamos que para a variável tempo de

entrega, 50% do serviço é realizado entre 13.35 min e 21.50 min. Vemos ainda, que as variáveis apresentam pontos atípicos (outliers). Na Figura 7.4 (c) vemos que a variabilidade da variável distância é alta, isto é, as distâncias percorridas pelo motorista apresentam uma significativa dispersão em torno da mediana. Notamos também, que há uma distância considerada atípica e que de acordo com a Tabela 7.4 75% do serviço tem mais de 36 (pés) de distância.

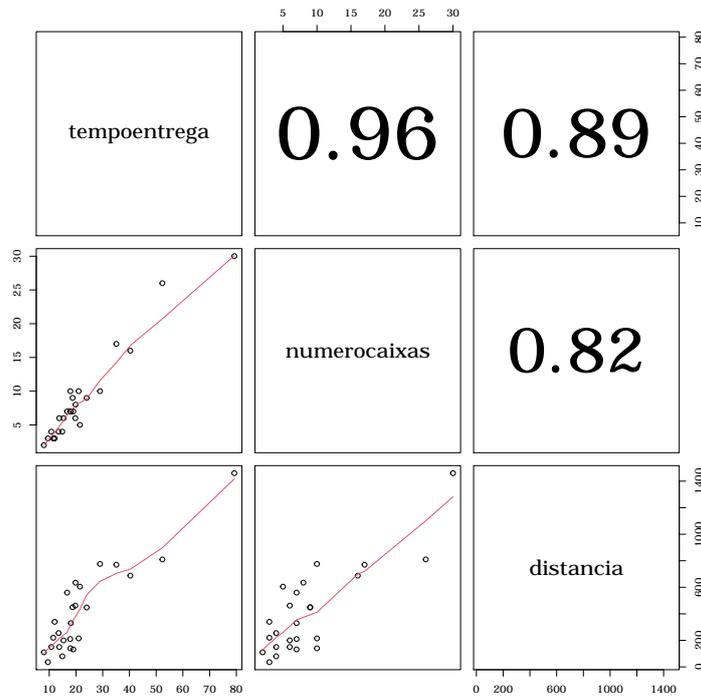


Figura 7.5: Correlograma.

De acordo com o correlograma apresentado na Figura 7.5, a variável tempo de entrega possui correlação forte e positiva com as variáveis número de caixas e distância.

Partindo da análise exploratória dos dados, o objetivo é modelar o tempo de entrega, baseando-se nas variáveis número de caixas e distância.

A fim de melhorar o ajuste do modelo de regressão linear múltipla $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, no sentido de minorar as violações dos pressupostos na Subseção 2.3.2 c) e d), aplicamos a transformação de Box e Cox. Em seguida, considerando as variáveis explicativas citadas, postulamos três modelos envolvendo essas variáveis, satisfazendo as seguintes condições:

- Os erros aleatórios apresentam normalidade (de acordo com o teste Jarque-Bera);
- Está corretamente especificado (de acordo com o teste Reset);

- As variâncias dos erros são constantes, ou seja, há homoscedasticidade (de acordo com o teste de Breush-Pagan);
- O ajuste é linear (de acordo com o teste Arco-Íris).
- Todos os parâmetros dos modelos são significativos.

Com isso, definimos os seguintes modelos:

- Modelo 1:

$$\log(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

- Modelo 2: .

$$\log(y) = \beta_0 + \beta_1 \log(x_1) + \beta_2 x_2^2$$

- Modelo 3:

$$\log(y) = \beta_0 + \beta_1 \sqrt{x_1} + \beta_2 x_2$$

Vale ressaltar, que o nosso interesse é a seleção do modelo que mais se ajuste aos dados baseando-se nos critérios V , AIC, BIC e \bar{R}^2 . Não é objetivo deste trabalho realizar diagnóstico de tais modelos. A Tabela 7.5 mostra os valores calculados dos critérios para os três modelos.

Tabela 7.5: Critérios de adequação para os modelos.

Modelo	Critérios			
	V	AIC	BIC	\bar{R}^2
1	0.93733	-15.205	-11.548	0.897
2	0.73410	-28.039	-24.382	0.9383
3	0.52891	-28.630	-24.973	0.9398

De acordo com a Tabela 7.5, os critérios V , AIC e BIC indicaram o Modelo 3 como o que melhor se ajusta aos dados, e o coeficiente \bar{R}^2 também indicou o Modelo 3 como o que melhor explica a variável tempo de entrega.

Para a verificação do ajuste do modelo 3, a Figura 7.6 mostra o gráfico para os resíduos com envelope simulado. No gráfico, podemos ver o padrão esperado para um modelo bem ajustado, em que aos pontos (resíduos) se apresentam de forma aleatória e entre os limites do envelope.

A Figura 7.7 mostra os gráficos dos resíduos para o Modelo 3, podemos verificar que os gráficos indicam possíveis padrões não aleatórios, heteroscedasticidade, e possíveis observações atípicas. O QQ-plot (canto superior direito) mostra os pontos dispersos aleatoriamente próximos à linha pontilhada, indicando normalidade.

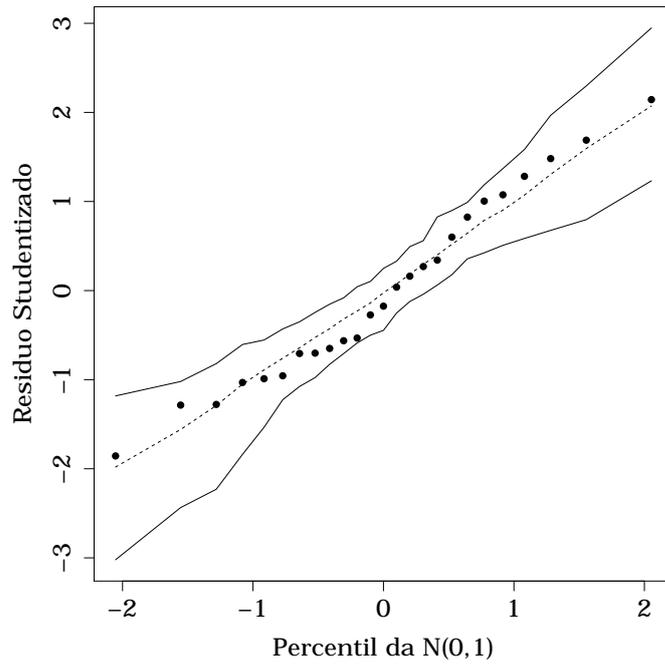


Figura 7.6: Gráfico de ajuste do Modelo 3.

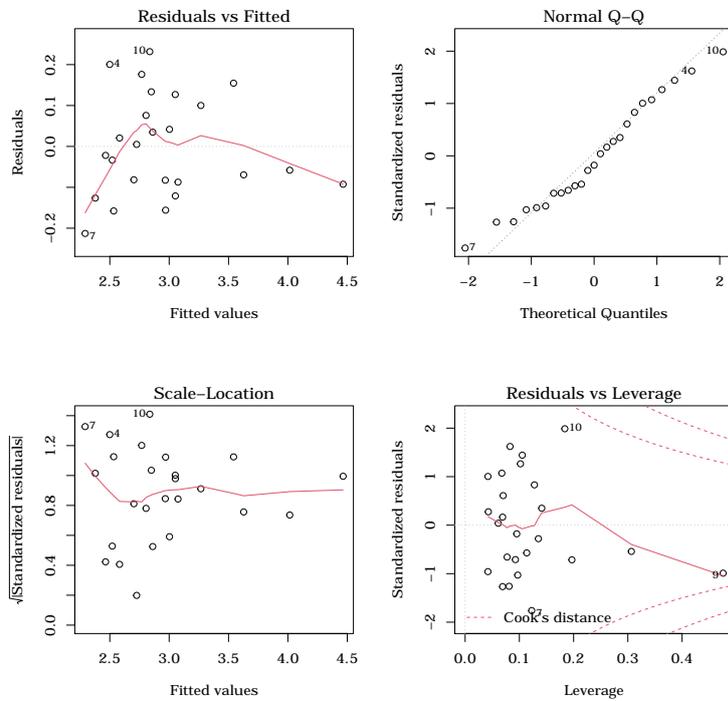


Figura 7.7: Gráfico dos resíduos do Modelo 3.

De modo geral, podemos ver que o desempenho do critério V para seleção de

modelos de regressão, aplicados a dados reais, condiz com os resultados obtidos nas simulações, mostrando um desempenho esperado quando comparado aos demais critérios.

Capítulo 8

Conclusões

É fato, do ponto de vista estatístico, que a escolha do modelo mais adequado para a representação de um fenômeno em estudo é um tópico extremamente importante. Dessa forma, diversos critérios para seleção de modelos são encontrados na literatura com esta finalidade.

Este trabalho teve como objetivo propor um novo critério de seleção de modelos, V , para modelos não encaixados baseado nas famílias de distribuições generalizadas e suas propriedades propostas por [Tablada \(2017\)](#), denominadas de G_c^{sup} e G_c^{inf} , e comparar seu desempenho com outros critérios conhecidos como o critério AIC, AICc, BIC, HQIC, W^* e A^* . Para avaliar o desempenho do novo critério, foram construídos diferentes cenários de simulação (utilizando a linguagem de programação `0x`) com o objetivo de avaliar a porcentagem de vezes em que o modelo correto foi selecionado.

Nos cenários de simulações, pudemos verificar que o critério proposto mostrou um desempenho superior em relação aos demais critérios quando realizadas pequenas perturbações no parâmetro da distribuição considerada para um tamanho amostral pequeno a moderado ($n = 20, 30, 50$), além disso, houve um aumento nas taxas de acerto à medida que os parâmetros e o tamanho amostral aumentaram. Outro fato mostrado, no Cenário 4, foi que o critério proposto mostrou um desempenho superior para todos os tamanhos amostrais quando considerada a distribuição normal.

Por meio de conjuntos de dados reais e considerando diferentes modelos, ilustramos a aplicabilidade do novo critério na tarefa de selecionar aquele modelo que fornecesse o melhor ajuste ao conjunto de dados. Uma característica que podemos destacar é que de acordo com as propriedades pelas quais o critério V foi proposto, temos que o modelo que melhor se ajusta aos dados é aquele cujo valor calculado pelo critério é mais próximo de zero. Essa característica permite que seja feito um julgamento “a priori” sobre a qualidade do ajuste proporcionado por um determinado modelo, mesmo antes de compará-lo com o ajuste proporcionado por outros modelos.

Tendo em vista a grande importância do estudo de Regressão Linear e sua aplicabilidade em diversas áreas, foi investigado o desempenho do critério V e comparado a outros critérios na seleção de modelos de regressão linear múltipla em amostra finitas, em que foram considerados dois cenários de simulações, nos quais também pudemos perceber um desempenho superior do critério proposto para pequenas amostras. Em seguida, ilustramos a aplicabilidade do critério a um conjunto de dados reais.

Portanto, neste trabalho, pudemos verificar que o critério de seleção de modelos proposto apresentou um desempenho satisfatório na seleção do modelo mais adequado quando comparado a alguns dos critérios mais utilizados na literatura, principalmente para amostras de tamanho pequeno a moderado.

Em estudos futuros, planeja-se o desenvolvimento de um pacote, na linguagem de programação R, que permita a implementação do critério V , o qual facilitará outros pesquisadores abordarem diversos cenários usando esse critério. Além disso, outras propriedades podem ser investigadas com relação ao critério, como por exemplo:

- a) estudar a adaptação do critério para modelos discretos;
- b) desempenho em amostras grandes considerando outros tipos de cenários,
- c) no contexto de regressão linear múltipla avaliar o impacto de observações atípicas no desempenho do critério, assim como a violação de suposições do modelo,
- d) aplicação do critério para outras classes de modelos de regressão, por exemplo, na seleção de modelos de regressão Beta.

Referências Bibliográficas

- ABREU, M. C., CECÍLIO, R. A., PRUSKI, F. F., et al., 2018, “Critérios para escolha de distribuições de probabilidades em estudos de eventos extremos de precipitação”, *Revista brasileira de meteorologia*, v. 33, pp. 601–613. doi: <http://dx.doi.org/10.1590/0102-778633400>.
- AKAIKE, H., 1973, “Information theory as an extension of the maximum likelihood principle”, in *B.N.Petrov, and F.Csaki, (eds.), Second International Symposium on Information Theory. Budapest: Academiai Kiado.*, pp. 267–281.
- ANDERSON, T. W., DARLING, D. A., 1954, “A Test of Goodness of Fit”, *Journal of the American Statistical Association*, v. 49, n. 268, pp. 765–769. doi: 10.1080/01621459.1954.10501232.
- BOLFARINE, H., SANDOVAL, M. C. A., 2000, *Introdução à Inferência Estatística*. São Paulo: Sociedade Brasileira de Matemática.
- BURNHAM, K. P., ANDERSON, D. R., 2004, “Multimodel Inference: Understanding AIC and BIC in Model Selection”, *Sociological Methods & Research*, v. 33, n. 2, pp. 261–304. doi: 10.1177/0049124104268644. Disponível em: <<https://doi.org/10.1177/0049124104268644>>.
- BURNHAM, K., ANDERSON, D., 2002, “Model selection and multimodel inference: a practical information-theoretic approach”, *New York: Springer*, p. 488.
- BURR, I. W., 1942, “Cumulative frequency functions”, *The Annals of Mathematical Statistics, JSTOR*, v. 13, n. 2, pp. 215–232.
- CASELLA, G., BERGER, R. L., 2002, *Statistical inference*. Duxbury Pacific Grove, CA.
- CAVANAUGH, J. E., 1999, “A large-sample model selection criterion based on Kullback’s symmetric divergence”, *Statistics and Probability Letters* 42, pp. 333–343.

- CHEN, G., BALAKRISHNAN, N., 1995, “A general purpose approximate goodness-of-fit test”, *Journal of Quality Technology*, v. 27, pp. 154–161.
- D’ AGOSTINO, R. B., STEPHENS, M. A., 1986, “Goodness-of-Fit Techniques”, *Marcel Dekker, New York, NY*.
- DASGUPTA, R., 2011, “On the distribution of Burr with applications”, *Sankhya B*, v. 73 (05), pp. 1–19. doi: 10.1007/s13571-011-0015-y.
- DOORNIK, J. A., 2018, *Developer’s manual for Ox 8*.
- EMILIANO, P. C., 2013, *Critérios de informação: como eles se comportam em diferentes modelos?*. Tese de Doutorado, Universidade Federal de Lavras.
- EUBANK, R. L., HART, J. D., 1993, “Testing Goodness-of-fit via Nonparametric Function Estimation Techniques”, *Communications in Statistics - Theory and Methods* 22, pp. 3327–3354.
- FRÉCHET, M., 1928, “Sur la loi de probabilité de l’écart maximum”. In: *Annales de la société Polonaise de Mathématique*. [sn].
- GOMPERTZ, B., 1825, “On the nature of the function expressive of the law of human mortality and on a new mode of determining life contingencies”, *Philosophical Transactions of the Royal Society, London*, v. 115, pp. 513–585.
- GUJARATI, D. N., PORTER, D. C., 2011, *Econometria Básica*. N. 5. Mc Graw Hill and Bookman.
- HANNAN, E. J., QUINN, B. G., 1979, “The Determination of the Order of an Autoregression”, *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 41, n. 2, pp. 190–195. ISSN: 00359246.
- JAMES, R. B., 1981, *Probabilidade: um curso em nível intermediário*. Rio de Janeiro: Instituto de Matemática Pura e Aplicada (Projeto Euclides).
- JENKINSON, A. F., 1955, “The frequency distribution of the annual maximum (or minimum) values of meteorological elements”, *Q. J. R. Meteorol. Soc., Brackneel*, v. 81, pp. 158–171.
- KIM, P. H., LÍ, S., LY, S., et al., 2019, “Comparison among Akaike Information Criterion, Bayesian Information Criterion and Vuong’s test in Model Selection: A Case Study of Violated Speed Regulation in Taiwan”, *Journal of Advanced Engineering and Computation*, v. 3 (03), pp. 293. doi: 10.25073/jaec.201931.220.

- KULLBACK, S., LEIBLER, R., 1951, “On information and sufficiency”, *Annals of Mathematical Statistics*, v. 22, n. 1, pp. 79–86. doi: 10.1214/aoms/1177729694.
- KUMARASWAMY, P., 1980, “Generalized probability density-function for double-bounded random-processes”, *J. Hydrol*, n. 46, pp. 79–88.
- LEHMANN, E. L., ROMANO, J. P., 2006, *Testing statistical hypotheses*. Springer Science & Business Media.
- LIEBLEIN, J., ZELEN, M., 1956, “Statistical investigation of the fatigue life of deep-groove ball bearing”, *J. Research National Bureau of Standards*, v. 57, pp. 273–316.
- MAGALHÃES, M. N., 2006, *Probabilidade e variáveis aleatórias*. São Paulo: Edusp.
- MARSHALL, A. W., OLKIN, I., 1997, “A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families”, *Biometrika*, v. 84, pp. 641–652.
- MAZUCHELI, J., BARROS, E. A. C., ACHCAR, J. A., 2016, “An alternative reparametrization for the weighted Lindley distribution”, *Brazilian Operation Research Society*, v. 36, pp. 345–353.
- MONTGOMERY, D. C., PECK, E. A., VINING, G. G., 2012, *Introduction to Linear Regression Analysis*. N. 5. John Wiley & Sons.
- MUDHOLKAR, G. S., SRIVASTAVA, D. K., FRAIMER, M., 1993, “Exponentiated weibull family for analyzing bathtub failure-rate data”, *IEEE Transactions on Reliability*, v. 42, n. 2, pp. 299–302.
- MURTHY, D. P., XIE, M., JIANG, R., 2004, *Weibull models*, v. 505. John Wiley & Sons.
- NADARAJAH, S., KOTZ, S., 2001, “The exponentiated Fréchet distribution”, *Interstat Electronic Journal*, v. 31, pp. 1–7.
- RISTIĆ, M. M., KUNDU, D., 2015, “Marshall-Olkin Generalized Exponential Distribution”, *Metron* 73, pp. 317–333. doi: <https://doi.org/10.1007/s40300-014-0056-x>.
- ROUSSAS, G. G., 2003, *An introduction to probability and statistical inference*. Elsevier.

- SCHWARZ, G., 1978, "Estimating the dimensional of a model." *Annals of Statistics*, *Hayward*, v. 6, n. 2 (Mar), pp. 461–464.
- SEVERINI, T. A., 2000, *Likelihood methods in statistics*. Oxford University Press.
- SOUSA, D. A. R., 2020, *Distribuição qui-quadrado inf: uma nova abordagem para o aperfeiçoamento do teste da razão de verossimilhanças*. Tese de Mestrado, Universidade Federal da Paraíba.
- SUGIURA, N., 1978, "Further analysts of the data by akaike's information criterion and the finite corrections", *Communications in Statistics - Theory and Methods*, *Ontario*, v. 7, n. 1, pp. 13–26.
- TABLADA, C. J., 2017, *Generalized probability distributions for lifetime applications*. Tese de Doutorado, Universidade Federal de Pernambuco.
- VUJICA, Y., 1972, *Probability and Statistics in Hydrology*. Water Resources Publication.
- VUONG, Q. H., 1989, "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses", *The Econometric Society*, v. 57, n. 2, pp. 307–333.
- WEIBULL, W. A., 1939, "A statistical theory of the strength of material", *Royal Swedish Institute for Engineering Research*, v. 151, pp. 1–45.
- YAMANOUCHI, T. K., 2017, *Seleção de modelos lineares mistos utilizando critérios de informação*. Tese de Mestrado, Universidade de São Paulo, Escola superior de agricultura "Luiz de Queiroz".
- ZAKERZADEH, H., MAHMOUDI, E., 2012, "A new two parameter lifetime distribution: model and properties", .

Apêndice A

Algumas Distribuições Contínuas

Neste apêndice são apresentadas as distribuições de probabilidade contínuas que foram utilizadas no decorrer da dissertação.

A.0.1 Distribuição Normal

Uma v.a X segue distribuição normal com média μ e variância σ^2 , ($X \sim N(\mu, \sigma^2)$), se sua função densidade de probabilidade é dada por

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right], \quad (\text{A.1})$$

em que $-\infty < x < \infty$ e o espaço paramétrico é definido como $\Theta = \{(\mu, \sigma^2); -\infty < \mu < \infty, \sigma^2 > 0\}$.

De (A.1), vemos que a função de verossimilhança correspondente à amostra observada $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ proveniente de uma população com distribuição $N(\mu, \sigma^2)$ é dada por,

$$\begin{aligned} L(\mu, \sigma^2; \mathbf{x}) &= \prod_{i=1}^n f(x_i; \mu, \sigma^2) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}\right]. \end{aligned} \quad (\text{A.2})$$

Assim, segue a função de log-verossimilhança dada por,

$$\ell(\mu, \sigma^2; \mathbf{x}) = -\frac{n}{2} \log(2\pi\sigma^2) - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}. \quad (\text{A.3})$$

A.0.2 Distribuição Weibull

Uma v.a X segue distribuição Weibull de parâmetros α e σ , ($X \sim \text{Weibull}(\alpha, \sigma)$), se sua função de distribuição acumulada é dada por

$$F(x; \alpha, \sigma) = 1 - \exp \left[- \left(\frac{x}{\sigma} \right)^\alpha \right]. \quad (\text{A.4})$$

A sua função densidade de probabilidade é

$$f(x; \alpha, \sigma) = \left(\frac{\alpha}{\sigma} \right) \left(\frac{x}{\sigma} \right)^{\alpha-1} \exp \left[- \left(\frac{x}{\sigma} \right)^\alpha \right], \quad (\text{A.5})$$

em que $x > 0$ e o espaço paramétrico é $\Theta = \{(\alpha, \sigma); \alpha > 0, \sigma > 0\}$.

De (A.5), vemos que a função verossimilhança correspondente à amostra observada $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ proveniente de uma população com distribuição Weibull(α, σ) é dada por,

$$\begin{aligned} L(\alpha, \sigma; \mathbf{x}) &= \prod_{i=1}^n f(x_i; \alpha, \sigma) \\ &= \left(\frac{\alpha}{\sigma} \right)^n \prod_{i=1}^n \left(\frac{x_i}{\sigma} \right)^{\alpha-1} \exp \left\{ \sum_{i=1}^n \left[- \left(\frac{x_i}{\sigma} \right)^\alpha \right] \right\}. \end{aligned} \quad (\text{A.6})$$

Assim, segue a função de log-verossimilhança dada por,

$$\ell(\alpha, \sigma; \mathbf{x}) = n \log \left(\frac{\alpha}{\sigma} \right) + (\alpha - 1) \sum_{i=1}^n \log \left(\frac{x_i}{\sigma} \right) - \sum_{i=1}^n \left(\frac{x_i}{\sigma} \right)^\alpha. \quad (\text{A.7})$$

A.0.3 Distribuição Weibull Exponencializada

Uma v.a X segue distribuição Weibull Exponencializada de parâmetros θ, α e σ ($X \sim \text{WE}(\theta, \alpha, \sigma)$) se sua função de distribuição acumulada é dada por

$$F(x; \theta, \alpha, \sigma) = \left\{ 1 - \exp \left[- \left(\frac{x}{\sigma} \right)^\alpha \right] \right\}^\theta. \quad (\text{A.8})$$

A sua função densidade de probabilidade é

$$f(x; \theta, \alpha, \sigma) = \left(\frac{\theta \alpha}{\sigma} \right) \left(\frac{x}{\sigma} \right)^{\alpha-1} \left\{ 1 - \exp \left[- \left(\frac{x}{\sigma} \right)^\alpha \right] \right\}^{\theta-1} \exp \left[- \left(\frac{x}{\sigma} \right)^\alpha \right], \quad (\text{A.9})$$

em que $x > 0$ e o espaço paramétrico é $\Theta = \{(\theta, \alpha, \sigma); \theta > 0, \alpha > 0, \sigma > 0\}$.

De (A.9) vemos que a função de verossimilhança correspondente à amostra observada $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ proveniente de uma população com distribuição WE(θ, α, σ) é dada por,

$$\begin{aligned}
L(\theta, \alpha, \sigma; \mathbf{x}) &= \prod_{i=1}^n f(x_i; \theta, \alpha, \sigma) \\
&= \left(\frac{\theta\alpha}{\sigma}\right)^n \prod_{i=1}^n \left\{ \left(\frac{x_i}{\sigma}\right)^{\alpha-1} \left\{ 1 - \exp \left[- \left(\frac{x_i}{\sigma}\right)^\alpha \right] \right\}^{\theta-1} \right\} \\
&\times \exp \left\{ \sum_{i=1}^n \left[- \left(\frac{x_i}{\sigma}\right)^\alpha \right] \right\}.
\end{aligned} \tag{A.10}$$

Assim, segue a função de log-verossimilhança dada por,

$$\begin{aligned}
\ell(\theta, \alpha, \sigma; \mathbf{x}) &= n \log \left(\frac{\theta\alpha}{\sigma} \right) + (\theta - 1) \sum_{i=1}^n \log \left\{ 1 - \exp \left[- \left(\frac{x_i}{\sigma}\right)^\alpha \right] \right\} \\
&- \sum_{i=1}^n \left(\frac{x_i}{\sigma}\right)^\alpha + (\alpha - 1) \sum_{i=1}^n \log \left(\frac{x_i}{\sigma}\right).
\end{aligned} \tag{A.11}$$

A distribuição Weibull Exponencializada é uma generalização da distribuição Weibull. De fato, podemos obter esta última da distribuição Weibull Exponencializada ao fazer $\theta = 1$.

A.0.4 Distribuição Lindley Ponderada

Uma v.a X segue distribuição Lindley Ponderada de parâmetros μ e λ , ($X \sim \text{LP}(\mu, \lambda)$), se sua função densidade de probabilidade é dada por

$$f(x; \mu, \lambda) = \frac{\mu^{\lambda+1}}{(\mu + \beta) \Gamma(\lambda)} x^{\lambda-1} (1 + x) e^{-\mu x}, \tag{A.12}$$

em que $x > 0$ e $\Gamma(\lambda) = \int_0^\infty x^{\lambda-1} e^{-x} dx$ é a função gama. O espaço paramétrico é $\Theta = \{(\mu, \lambda); \mu > 0, \lambda > 0\}$.

De (A.12), vemos que a função de verossimilhança correspondente à amostra observada $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ proveniente de uma população com distribuição $\text{LP}(\mu, \lambda)$ é dada por,

$$\begin{aligned}
L(\mu, \lambda; \mathbf{x}) &= \prod_{i=1}^n f(x_i; \mu, \lambda) \\
&= \left(\frac{\mu^{\lambda+1}}{(\mu + \lambda) \Gamma(\lambda)} \right)^n \exp \left[-\mu \sum_{i=1}^n x_i \right] \prod_{i=1}^n x_i^{\lambda-1} (1 + x_i).
\end{aligned} \tag{A.13}$$

Assim, segue a função de log-verossimilhança dada por,

$$\ell(\mu, \lambda; \mathbf{x}) = n \log \left(\frac{\mu^{\lambda+1}}{(\mu + \lambda)\Gamma(\lambda)} \right) - \mu \sum_{i=1}^n x_i + \sum_{i=1}^n \log [x_i^{\lambda-1}(1 + x_i)]. \quad (\text{A.14})$$

A.0.5 Distribuição Fréchet

Uma v.a X segue distribuição Fréchet de parâmetros α e β , ($X \sim \text{Fr}(\alpha, \beta)$), se sua função de distribuição acumulada é dada por

$$F(x; \alpha, \beta) = \exp \left[- \left(\frac{\alpha}{x} \right)^\beta \right]. \quad (\text{A.15})$$

A sua função densidade de probabilidade é

$$f(x; \alpha, \beta) = \left(\frac{\beta}{x} \right) \left(\frac{\alpha}{x} \right)^\beta \exp \left[- \left(\frac{\alpha}{x} \right)^\beta \right], \quad (\text{A.16})$$

em que $x > 0$ e o espaço paramétrico é $\Theta = \{(\alpha, \beta); \alpha > 0, \beta > 0\}$.

De (A.16), vemos que a função verossimilhança correspondente à amostra observada $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ proveniente de uma população com distribuição $\text{Fr}(\alpha, \beta)$ é dada por,

$$\begin{aligned} L(\alpha, \beta; \mathbf{x}) &= \prod_{i=1}^n f(x_i; \alpha, \beta) \\ &= \prod_{i=1}^n \left(\frac{\beta}{x_i} \right) \left(\frac{\alpha}{x_i} \right)^\beta \exp \left\{ \sum_{i=1}^n \left[- \left(\frac{\alpha}{x_i} \right)^\beta \right] \right\}. \end{aligned} \quad (\text{A.17})$$

Assim, segue a função de log-verossimilhança dada por,

$$\ell(\alpha, \beta; \mathbf{x}) = \sum_{i=1}^n \log \left(\frac{\beta}{x_i} \right) + (\beta) \sum_{i=1}^n \log \left(\frac{\alpha}{x_i} \right) - \sum_{i=1}^n \left(\frac{\alpha}{x_i} \right)^\beta. \quad (\text{A.18})$$

A.0.6 Distribuição Fréchet Modificada

Uma v.a X segue distribuição Fréchet Modificada de parâmetros α e β , ($X \sim \text{FM}(\alpha, \beta, \lambda)$), se sua função de distribuição acumulada é dada por

$$F(x; \alpha, \beta, \lambda) = \exp \left[- \left(\frac{\alpha}{x} \right)^\beta e^{-\lambda x} \right]. \quad (\text{A.19})$$

A sua função densidade de probabilidade é

$$f(x; \alpha, \beta, \lambda) = \frac{1}{x} (\beta + \lambda x) \left(\frac{\alpha}{x} \right)^\beta \exp \left[-\lambda x - \left(\frac{\alpha}{x} \right)^\beta e^{-\lambda x} \right], \quad (\text{A.20})$$

em que $x > 0$ e o espaço paramétrico é $\Theta = \{(\alpha, \beta, \lambda); \alpha > 0, \beta > 0, \lambda \geq 0\}$.

De (A.20), vemos que a função verossimilhança correspondente à amostra observada $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ proveniente de uma população com distribuição FM(α, β, λ) é dada por,

$$\begin{aligned} L(\alpha, \beta, \lambda; \mathbf{x}) &= \prod_{i=1}^n f(x_i; \alpha, \beta, \lambda) \\ &= \prod_{i=1}^n \frac{1}{x_i} (\beta + \lambda x_i) \left(\frac{\alpha}{x_i}\right)^\beta \exp \left[-\lambda x_i - \left(\frac{\alpha}{x_i}\right)^\beta e^{-\lambda x} \right]. \end{aligned} \quad (\text{A.21})$$

Assim, segue a função de log-verossimilhança dada por,

$$\begin{aligned} \ell(\alpha, \beta, \lambda; \mathbf{x}) &= \sum_{i=1}^n \log \left(\frac{1}{x_i} \right) + \sum_{i=1}^n \log (\beta + \lambda x_i) \\ &\quad + \beta \sum_{i=1}^n \log \left(\frac{\alpha}{x_i} \right) + \sum_{i=1}^n \left[-\lambda x_i - \left(\frac{\alpha}{x_i}\right)^\beta e^{-\lambda x} \right]. \end{aligned} \quad (\text{A.22})$$

A.0.7 Distribuição Fréchet Exponencializada

Uma v.a X segue distribuição Fréchet Exponencializada de parâmetros σ , η e ρ , ($X \sim \text{FE}(\sigma, \eta, \rho)$), se sua função de distribuição acumulada é dada por

$$F(x; \sigma, \eta, \rho) = 1 - \left\{ 1 - \exp \left[- \left(\frac{\sigma}{x} \right)^\eta \right] \right\}^\rho. \quad (\text{A.23})$$

A sua função densidade de probabilidade é

$$f(x; \sigma, \eta, \rho) = \rho \eta \sigma^\eta \left\{ 1 - \exp \left[- \left(\frac{\sigma}{x} \right)^\eta \right] \right\}^{\rho-1} x^{-(1+\eta)} \exp \left[- \left(\frac{\sigma}{x} \right)^\eta \right], \quad (\text{A.24})$$

em que $x > 0$ e o espaço paramétrico é $\Theta = \{(\sigma, \eta, \rho); \sigma > 0, \eta > 0, \rho > 0\}$.

De (A.24), vemos que a função verossimilhança correspondente à amostra observada $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ proveniente de uma população com distribuição FE(σ, η, ρ) é dada por,

$$\begin{aligned} L(\sigma, \eta, \rho; \mathbf{x}) &= \prod_{i=1}^n f(x_i; \sigma, \eta, \rho) = (\rho \eta \sigma^\eta)^n \\ &\quad \times \prod_{i=1}^n \left\{ \left\{ 1 - \exp \left[- \left(\frac{\sigma}{x_i} \right)^\eta \right] \right\}^{\rho-1} x_i^{-(1+\eta)} \exp \left[- \left(\frac{\sigma}{x_i} \right)^\eta \right] \right\}. \end{aligned} \quad (\text{A.25})$$

Assim, segue a função de log-verossimilhança dada por,

$$\begin{aligned} \ell(\sigma, \eta, \rho; \mathbf{x}) = & n \log(\rho \eta \sigma^n) + (\rho - 1) \sum_{i=1}^n \log \left\{ 1 - \exp \left[- \left(\frac{\sigma}{x_i} \right)^\eta \right] \right\} \\ & - (1 + \eta) \sum_{i=1}^n \log(x_i) - \sum_{i=1}^n \left(\frac{\sigma}{x_i} \right)^\eta \end{aligned} \quad (\text{A.26})$$

A.0.8 Distribuição Lindley estendida de Marshall-Olkin

Uma v.a X segue distribuição Lindley estendida de Marshall - Olkin de parâmetros θ e α , ($X \sim \text{LEMO}(\theta, \alpha)$), se sua função de distribuição acumulada é dada por

$$F(x; \theta, \alpha) = \frac{1 - \left(1 + \frac{\theta x}{\theta + 1}\right) e^{-\theta x}}{1 - (1 - \alpha) \left(1 + \frac{\theta x}{\theta + 1}\right) e^{-\theta x}}. \quad (\text{A.27})$$

A sua função densidade de probabilidade é

$$f(x; \theta, \alpha) = \frac{\alpha \theta^2}{\theta + 1} (1 + x) e^{-\theta x} \left[1 - (1 - \alpha) \left(1 + \frac{\theta x}{\theta + 1}\right) e^{-\theta x} \right]^{-2}, \quad (\text{A.28})$$

em que $x > 0$ e o espaço paramétrico é $\Theta = \{(\theta, \alpha); \theta > 0, \alpha > 0\}$.

De (A.28), vemos que a função verossimilhança correspondente à amostra observada $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ proveniente de uma população com distribuição $\text{LEMO}(\theta, \alpha)$ é dada por,

$$\begin{aligned} L(\theta, \alpha; \mathbf{x}) = & \prod_{i=1}^n f(x_i; \theta, \alpha) \\ = & \left(\frac{\alpha \theta^2}{\theta + 1} \right)^n \prod_{i=1}^n (1 + x_i) e^{-\theta x_i} \left[1 - (1 - \alpha) \left(1 + \frac{\theta x_i}{\theta + 1}\right) e^{-\theta x_i} \right]^{-2}. \end{aligned} \quad (\text{A.29})$$

Assim, segue a função de log-verossimilhança dada por,

$$\begin{aligned} \ell(\theta, \alpha; \mathbf{x}) = & n \log \left(\frac{\alpha \theta^2}{\theta + 1} \right) + \sum_{i=1}^n \log(1 + x_i) - \sum_{i=1}^n (\theta x_i) \\ & - 2 \sum_{i=1}^n \log \left[1 - (1 - \alpha) \left(1 + \frac{\theta x_i}{\theta + 1}\right) e^{-\theta x_i} \right]. \end{aligned} \quad (\text{A.30})$$

A.0.9 Distribuição Weibull estendida de Marshall-Olkin

Uma v.a X segue distribuição Weibull estendida de Marshall - Olkin de parâmetros α e β e λ , ($X \sim \text{WEMO}(\alpha, \beta, \lambda)$), se sua função de distribuição acumulada é dada

por

$$F(x; \alpha, \beta, \lambda) = 1 - \frac{\lambda \exp \left[- \left(\frac{x}{\alpha} \right)^\beta \right]}{1 - (1 - \lambda) \exp \left[- \left(\frac{x}{\alpha} \right)^\beta \right]}. \quad (\text{A.31})$$

A sua função densidade de probabilidade é

$$f(x; \alpha, \beta, \lambda) = \frac{\left(\frac{\lambda \beta}{\alpha} \right) \left(\frac{x}{\alpha} \right)^{\beta-1} \exp \left[- \left(\frac{x}{\alpha} \right)^\beta \right]}{\left[1 - (1 - \lambda) \exp \left[- \left(\frac{x}{\alpha} \right)^\beta \right] \right]^2}, \quad (\text{A.32})$$

em que $x > 0$ e o espaço paramétrico é $\Theta = \{(\alpha, \beta, \lambda); \alpha > 0, \beta > 0, \lambda > 0\}$.

De (A.32), vemos que a função verossimilhança correspondente à amostra observada $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ proveniente de uma população com distribuição WEMO(α, β, λ) é dada por,

$$\begin{aligned} L(\alpha, \beta, \lambda; \mathbf{x}) &= \prod_{i=1}^n f(x_i; \alpha, \beta, \lambda) \\ &= \prod_{i=1}^n \left[\frac{\left(\frac{\lambda \beta}{\alpha} \right) \left(\frac{x_i}{\alpha} \right)^{\beta-1} \exp \left[- \left(\frac{x_i}{\alpha} \right)^\beta \right]}{\left[1 - (1 - \lambda) \exp \left[- \left(\frac{x_i}{\alpha} \right)^\beta \right] \right]^2} \right]. \end{aligned} \quad (\text{A.33})$$

Assim, segue a função de log-verossimilhança dada por,

$$\begin{aligned} \ell(\alpha, \beta, \lambda; \mathbf{x}) &= n \log \left(\frac{\lambda \beta}{\alpha} \right) + (\beta - 1) \sum_{i=1}^n \log \left(\frac{x_i}{\alpha} \right) - \beta \sum_{i=1}^n \left(\frac{x_i}{\alpha} \right) \\ &\quad - 2 \sum_{i=1}^n \log \left[1 - (1 - \lambda) \exp \left[- \left(\frac{x_i}{\alpha} \right)^\beta \right] \right]. \end{aligned} \quad (\text{A.34})$$

A.0.10 Distribuição Fréchet tipo II

Uma v.a X segue distribuição Fréchet tipo II de parâmetros α , β e λ , ($X \sim \text{Fr tipo II}(\alpha, \beta, \lambda)$), se sua função de distribuição acumulada é dada por

$$F(x; \alpha, \beta, \lambda) = \exp \left[- \left(\frac{\beta}{x - \lambda} \right)^\alpha \right]. \quad (\text{A.35})$$

A sua função densidade de probabilidade é

$$f(x; \alpha, \beta, \lambda) = \left(\frac{\alpha}{\beta} \right) \left[\frac{\beta}{x - \lambda} \right]^{\alpha+1} \exp \left[- \left(\frac{\beta}{x - \lambda} \right)^\alpha \right], \quad (\text{A.36})$$

em que $x > 0$ e o espaço paramétrico é $\Theta = \{(\alpha, \beta, \lambda); \alpha > 0, \beta > 0, \lambda > 0\}$.

De (A.36), vemos que a função verossimilhança correspondente à amostra observada $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ proveniente de uma população com distribuição Fr tipo

$\Pi(\alpha, \beta, \lambda)$ é dada por,

$$\begin{aligned} L(\alpha, \beta, \lambda; \mathbf{x}) &= \prod_{i=1}^n f(x_i; \alpha, \beta, \lambda) \\ &= \left(\frac{\alpha}{\beta}\right)^n \prod_{i=1}^n \left[\frac{\beta}{x_i - \lambda}\right]^{\alpha+1} \exp\left[-\left(\frac{\beta}{x_i - \lambda}\right)^\alpha\right]. \end{aligned} \quad (\text{A.37})$$

Assim, segue a função de log-verossimilhança dada por,

$$\ell(\alpha, \beta, \lambda; \mathbf{x}) = n \log\left(\frac{\alpha}{\beta}\right) + (\alpha + 1) \sum_{i=1}^n \log\left(\frac{\beta}{x_i - \lambda}\right) - \alpha \sum_{i=1}^n \left(\frac{\beta}{x_i - \lambda}\right). \quad (\text{A.38})$$

A.0.11 Distribuição Burr XII

Uma v.a X segue distribuição Burr XII de parâmetros α e β , ($X \sim \text{BXII}(\alpha, \beta)$), se sua função de distribuição acumulada é dada por

$$F(x; \alpha, \beta) = 1 - (1 + x^\alpha)^{-\beta}. \quad (\text{A.39})$$

A sua função densidade de probabilidade é

$$f(x; \alpha, \beta) = \alpha \beta x^{\alpha-1} (1 + x^\alpha)^{-(\beta+1)}, \quad (\text{A.40})$$

em que $x > 0$ e o espaço paramétrico é $\Theta = \{(\alpha, \beta); \alpha > 0, \beta > 0\}$.

De (A.40), vemos que a função verossimilhança correspondente à amostra observada $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ proveniente de uma população com distribuição $\text{BXII}(\alpha, \beta)$ é dada por,

$$\begin{aligned} L(\alpha, \beta; \mathbf{x}) &= \prod_{i=1}^n f(x_i; \alpha, \beta) \\ &= (\alpha \beta)^n \prod_{i=1}^n x_i^{\alpha-1} (1 + x_i^\alpha)^{-(\beta+1)}. \end{aligned} \quad (\text{A.41})$$

Assim, segue a função de log-verossimilhança dada por,

$$\ell(\alpha, \beta; \mathbf{x}) = n \sum_{i=1}^n \log(\alpha \beta) + (\alpha - 1) \sum_{i=1}^n \log(x_i) - (\beta + 1) \sum_{i=1}^n \log(1 + x_i^\alpha). \quad (\text{A.42})$$

A.0.12 Distribuição Gama

Uma v.a X segue distribuição Gama de parâmetros σ e μ , ($X \sim \text{Gama}(\sigma, \mu)$), se sua função densidade de probabilidade é dada por

$$f(x; \sigma, \mu) = \frac{\mu^\sigma x^{\sigma-1} e^{-\mu x}}{\Gamma(\sigma)}, \quad (\text{A.43})$$

em que $x > 0$ e $\Gamma(\sigma) = \int_0^\infty x^{\sigma-1} e^{-x} dx$ é a função gama. O espaço paramétrico é $\Theta = \{(\sigma, \mu); \sigma > 0, \mu > 0\}$.

De (A.43), vemos que a função de verossimilhança correspondente à amostra observada $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ proveniente de uma população com distribuição Gama(σ, μ) é dada por,

$$\begin{aligned} L(\sigma, \mu; \mathbf{x}) &= \prod_{i=1}^n f(x_i; \sigma, \mu) \\ &= \left(\frac{\mu^\sigma}{\Gamma(\sigma)} \right)^n \exp \left[-\mu \sum_{i=1}^n x_i \right] \prod_{i=1}^n x_i^{\sigma-1}. \end{aligned} \quad (\text{A.44})$$

Assim, segue a função de log-verossimilhança dada por,

$$\ell(\sigma, \mu; \mathbf{x}) = n \log \left(\frac{\mu^\sigma}{\Gamma(\sigma)} \right) - \mu \sum_{i=1}^n x_i + \sum_{i=1}^n \log [x_i^{\sigma-1}]. \quad (\text{A.45})$$

A.0.13 Distribuição Gompertz

Uma v.a X segue distribuição Gompertz de parâmetros α e β , ($X \sim \text{Go}(\alpha, \beta)$), se sua função de distribuição acumulada é dada por

$$F(x; \alpha, \beta) = 1 - \exp \left[-\frac{\alpha}{\beta} (e^{\beta x} - 1) \right]. \quad (\text{A.46})$$

A sua função densidade de probabilidade é

$$f(x; \alpha, \beta) = \alpha \exp \left[\beta x - \frac{\alpha}{\beta} (e^{\beta x} - 1) \right], \quad (\text{A.47})$$

em que $x > 0$ e o espaço paramétrico é $\Theta = \{(\alpha, \beta); \alpha > 0, \beta > 0\}$.

De (A.47), vemos que a função verossimilhança correspondente à amostra observada $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ proveniente de uma população com distribuição Go(α, β)

é dada por,

$$\begin{aligned} L(\alpha, \beta; \mathbf{x}) &= \prod_{i=1}^n f(x_i; \alpha, \beta) \\ &= \alpha^n \exp \sum_{i=1}^n \left[\beta x_i - \frac{\alpha}{\beta} (e^{\beta x_i} - 1) \right]. \end{aligned} \quad (\text{A.48})$$

Assim, segue a função de log-verossimilhança dada por,

$$\ell(\alpha, \beta; \mathbf{x}) = n \log(\alpha) + \sum_{i=1}^n \left[\beta x_i - \frac{\alpha}{\beta} (e^{\beta x_i} - 1) \right]. \quad (\text{A.49})$$

A.0.14 Distribuição Kumaraswamy

Uma v.a X segue distribuição Kumaraswamy de parâmetros θ e α , ($X \sim \text{Kw}(\theta, \alpha)$), se sua função de distribuição acumulada é dada por

$$F(x; \theta, \alpha) = 1 - (1 - x^\theta)^\alpha. \quad (\text{A.50})$$

A sua função densidade de probabilidade é

$$f(x; \theta, \alpha) = \theta \alpha x^{\theta-1} (1 - x^\theta)^{\alpha-1}, \quad (\text{A.51})$$

em que $x > 0$ e o espaço paramétrico é $\Theta = \{(\theta, \alpha); \theta > 0, \alpha > 0\}$.

De (A.51), vemos que a função verossimilhança correspondente à amostra observada $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ proveniente de uma população com distribuição $\text{Kw}(\theta, \alpha)$ é dada por,

$$\begin{aligned} L(\theta, \alpha; \mathbf{x}) &= \prod_{i=1}^n f(x_i; \theta, \alpha) \\ &= (\theta \alpha)^n \prod_{i=1}^n [x_i^{\theta-1} (1 - x_i^\theta)^{\alpha-1}]. \end{aligned} \quad (\text{A.52})$$

Assim, segue a função de log-verossimilhança dada por,

$$\ell(\theta, \alpha; \mathbf{x}) = n \log(\theta \alpha) + \sum_{i=1}^n \log [x_i^{\theta-1} (1 - x_i^\theta)^{\alpha-1}]. \quad (\text{A.53})$$