



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Faculdade de Ciências e Tecnologia
Câmpus de Presidente Prudente

Estimação de energia solar gerada por placas fotovoltaicas utilizando modelagem Estatística

Matheus Inacio Souza Santos

Orientador: Prof. Dr. Mário Hissamitsu Tarumoto

Programa: Matemática Aplicada e Computacional

Presidente Prudente, Março de 2021

UNIVERSIDADE ESTADUAL PAULISTA

Faculdade de Ciências e Tecnologia de Presidente Prudente

Programa de Pós-Graduação em Matemática Aplicada e Computacional

Estimação de energia solar gerada por placas fotovoltaicas utilizando modelagem Estatística

Matheus Inacio Souza Santos

Orientador: Prof. Dr. Mario Hissamitsu Tarumoto

Dissertação apresentada ao Programa de Pós-Graduação em Matemática Aplicada e Computacional da Faculdade de Ciências e Tecnologia da UNESP para obtenção do título de Mestre em Matemática Aplicada e Computacional.

Presidente Prudente, Março de 2021

S237e

Santos, Matheus Inacio Souza

Estimação de energia solar gerada por placas fotovoltaicas utilizando modelagem Estatística / Matheus Inacio Souza Santos. -- Presidente Prudente, 2021
95 p. : tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências e Tecnologia, Presidente Prudente

Orientador: Mario Hissamitsu Tarumoto

1. Modelagem Estatística. 2. Distribuição Bivariada. 3. Modelos markovianos. 4. Energia Limpa. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências e Tecnologia, Presidente Prudente. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: Estimacão de energia solar gerada por placas fotovoltaicas utilizando modelagem Estatística

AUTOR: MATHEUS INACIO SOUZA SANTOS

ORIENTADOR: MÁRIO HISSAMITSU TARUMOTO

Aprovado como parte das exigências para obtenção do Título de Mestre em MATEMÁTICA APLICADA E COMPUTACIONAL, pela Comissão Examinadora:



Prof. Dr. MÁRIO HISSAMITSU TARUMOTO (Participação Virtual)
Departamento de Estatística / Faculdade de Ciências e Tecnologia de Presidente Prudente

Profa. Dra. ELIZABETH MIE HASHIMOTO (Participação Virtual)
Campus Londrina / Universidade Tecnológica Federal do Paraná

Prof. Dr. FERNANDO ANTONIO MOALA (Participação Virtual)
Departamento de Estatística / Faculdade de Ciências e Tecnologia de Presidente Prudente

Presidente Prudente, 08 de março de 2021

*Dedico a minha mãe Maria Auxiliadora,
ao meu pai Jaime e aos meus irmãos Bruno e Ana Vitória.*

Agradecimentos

Existem inúmeras pessoas que contribuíram direta e indiretamente para a conclusão desta dissertação, deixo aqui minha gratidão a todos, em especial:

Ao meu pai Jaime Souza e minha mãe Maria Auxiliadora, por sempre estarem comigo independentemente da situação, sem eles isso nunca seria possível. Ao meu irmão Bruno Inacio por ser minha inspiração diária e minha irmã Ana Vitória, simplesmente por sua existência, que me alegra todos os dias. Aos meus colegas e amigos da graduação e pós graduação, por sempre deixarem meu dia mais alegre e descontraído, em especial à Beatriz Montini, Cleidison Souza, Erika Maeda, Gilberto Alcantara, Giovanni Piccirilli, Rebeca Santos, Vinícius Otalora e Vitória Oliveira. Aos meus amigos de longa data, Beatriz Pereira e Bruno Ricieri, que foram muito importantes e presentes durante todo o meu processo na pós graduação. Aos meus professores do curso de estatística e do programa PósMAC da FCT-UNESP. Ao meu orientador Prof. Dr. Mário Hissamitsu Tarumoto, por toda paciência que tive e todo conhecimento transmitido para o término da pesquisa, sou muito grato pela ajuda. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Só tenho gratidão a todos vocês!!!

*"A única coisa que separa
as mulheres negras de qualquer outra pessoa é oportunidade"*
Viola Davis

Resumo

A conversão direta de energia solar em elétrica é feita através da radiação sobre determinados materiais fotovoltaicos, onde os fótons contidos na luz solar são convertidos em energia elétrica através do uso de células solares. A radiação solar é a energia radiante emitida pelo sol, em particular aquela que é transmitida sob a forma da radiação eletromagnética, onde as matérias fotovoltaicos estão ligados diretamente. Existem diversas variáveis que influenciam para a quantidade de energia gerada ao longo do dia, como a nebulosidade, temperatura, entre outros. O tema abordado no trabalho aqui presente é a criação de modelos probabilísticos que ajustam a energia solar gerada por placas fotovoltaicas levando em consideração dados históricos de geração de energia solar, sugerindo dependência ao longo do tempo utilizando um processo de Markov. Neste trabalho o objetivo foi o de construir modelos usando diferentes funções de distribuição de probabilidade e verificar qual obteve o melhor resultado perante os dados. A base de dados é composta por dados de energia solar gerada por placas fotovoltaicas da cidade de São José do Rio Preto-SP no período de Fev/2019 a Jul/2020, sem utilizar as covariáveis, e como resultado, o Modelo Markoviano Weibull via a Distribuição Weibull Bivariada de Marshall e Olkin, foi o que apresentou o melhor resultado.

Palavras-Chave: *Modelagem Estatística, Distribuição Bivariada, Modelos markovianos, Energia Limpa, Energia Fotovoltaicas.*

Abstract

The direct conversion of solar energy into electricity is done through radiation on certain photovoltaic materials, where the photons contained in sunlight are converted into electrical energy through the use of solar cells. Solar radiation is the radiant energy emitted by the sun which is transmitted in the form of electromagnetic radiation, where photovoltaic materials are directly attached. Several variables influence the amount of energy generated throughout the day, such as cloudiness, temperature, rain, among others. In the present work, is the creation of probabilistic models that adjust the solar energy generated by photovoltaic plates will be approached taking into account historical data of solar energy generation, suggesting dependence over time using a Markov process. In this work, the objective was to build models using different probability distribution functions and verify which one obtained the best result in the face of the data. The database consists of solar energy data generated by photovoltaic plates from the city of São José do Rio Preto-SP in the period from Feb / 2019 to Jul / 2020, without using the covariates, and as a result, the Markoviano Weibull Model via the Marshall and Olkin Bivariate Weibull Distribution, was the one that presented the best result.

Keywords: *Statistical Modeling, Bivariate Distribution, Markovian Models, Clean Energy, Photovoltaic Energy.*

Lista de Figuras

1.1	Processo de Translação	20
2.1	Tipos de Radiação	24
3.1	Energia gerada pelas placas fotovoltaicas	30
3.2	Gráfico de linha da Energia Solar em KWH	31
4.1	Gráfico de linha da Energia solar em KWHs. Dados Imputados	36
5.1	Formas típica da função de densidade de probabilidade e da distribuição acumulada, da distribuição Weibull para alguns valores dos parâmetros α e β	38
5.2	Formas típica da função de densidade de probabilidade e da distribuição acumulada, da distribuição Gama para alguns valores dos parâmetros α e β	41
5.3	Formas típica da função de densidade de probabilidade e da distribuição acumulada, da distribuição Exponencial Generalizada para alguns valores dos parâmetros α e β	44
5.4	Formas típica da função de densidade de probabilidade e da distribuição acumulada, da distribuição Normal para alguns valores dos parâmetros μ e σ	46
5.5	Formas típica da função de densidade de probabilidade e da distribuição acumulada, da distribuição Weibull Exponenciada para alguns valores dos parâmetros α , β e a	48
5.6	Formas típica da função de densidade de probabilidade e da distribuição acumulada, da distribuição Weibull Kumaraswamy para alguns valores dos parâmetros α , β , a e b	51
5.7	Formas típica da função de densidade de probabilidade e da distribuição acumulada, da distribuição log-EK com diferentes valores de α , β , γ e σ	53
5.8	Forma típica da função de densidade de probabilidade e da distribuição acumulada, da distribuição Weibull exponenciada Kumaraswamy com diferentes valores de α , β , a , b e c	55
5.9	Curva das funções de densidade da Weibull e Gama versus a curva de suas funções de distribuições Empírica para seus respectivos parâmetros estimados	58
5.10	Curva das funções de densidade da Exponencial Generalizada e Normal versus a curva de suas funções de distribuições Empírica para seus respectivos parâmetros estimados	58
5.11	Curva das funções de densidade da Weibull Exponenciada e Weibull Kumaraswamy versus a curva de suas funções de distribuições Empírica para seus respectivos parâmetros estimados	59
5.12	Curva das funções de densidade da Log-exponenciada de Kumaraswamy e Weibull Exponenciada Kumaraswamy versus a curva de suas funções de distribuições Empírica para seus respectivos parâmetros estimados	59

5.13	Curva das funções de distribuições acumuladas da Weibull e Gama versus a curva de suas funções de distribuições acumuladas Empírica para seus respectivos parâmetros estimados	60
5.14	Curva das funções de distribuições acumuladas da Exponencial Generalizada e Normal versus a curva de suas funções de distribuições acumuladas Empírica para seus respectivos parâmetros estimados	60
5.15	Curva das funções de distribuições acumuladas da Weibull Exponenciada e Weibull Kumaraswamy versus a curva de suas funções de distribuições acumuladas Empírica para seus respectivos parâmetros estimados	61
5.16	Curva das funções de distribuições acumuladas da Log-exponenciada de Kumaraswamy e Weibull Exponenciada Kumaraswamy versus a curva de suas funções de distribuições acumuladas Empírica para seus respectivos parâmetros estimados	61
6.1	Curva da função de distribuição acumulada do Modelo markoviano via a distribuição bivariada Weibull de Freund versus a curva de sua função de distribuição acumulada Empírica para seu respectivos parâmetros estimados	71
6.2	Curva da função de distribuição acumulada do Modelo markoviano via a distribuição bivariada Weibull de Marshall e Olkin versus a curva de sua função de distribuição acumulada Empírica para seu respectivos parâmetros estimados	78
6.3	Curva da função de distribuição acumulada do Modelo markoviano via a Distribuição Bivariada Weibull Exponenciada Kumaraswamy versus a curva de sua função de distribuição acumulada Empírica para seu respectivos parâmetros estimados	85

Lista de Tabelas

5.1	Parâmetros estimados pelo EMV para cada distribuição univariada	57
6.1	Medidas de dependência para os tempos T_i e T_{i+i}	64
6.2	Parâmetros estimados pelo EMV para os parâmetros do Modelo markoviano via a distribuição bivariada Weibull de Freund	71
6.3	Parâmetros estimados pelo EMV para os parâmetros do Modelo markoviano via a distribuição bivariada Weibull de Marshall e Olkin	78
6.4	Parâmetros estimados pelo EMV para os parâmetros do Modelo markoviano via a Distribuição Bivariada Weibull Exponenciada Kumaraswamy . .	84
6.5	Seleção dos modelos	86
6.6	Estimativa da média esperada de energia solar gerada por placas fotovoltaicas nos próximos dias em função do número observado entre o dia anterior e o dia atual	87

Lista de Siglas

- AIC: Critério de Informação Akaike.
- AICc: Critério de Informação Akaike Corrigido.
- BIC: Critério de Informação Bayesiana.
- DBW: Distribuição Bivariada Weibull.
- DEB: Distribuição Exponencial Bivariada.
- EKwW: Weibull Exponenciada Kumaraswamy.
- EKwWB: Bivariada Weibull Exponenciada Kumaraswamy.
- EMV: Estimador de Máxima Verossimilhança.
- EW: Weibull Exponenciada.
- FDA: Função Densidade Acumulada.
- FDP: Função Densidade de Probabilidade.
- GE: Exponencial Generalizada.
- IM: Imputação Múltipla.
- INMET: Instituto Nacional de Meteorologia.
- IU: Imputação Única.
- KWH: Quilowatt-Hora.
- KwW: Weibull Kumaraswamy.
- Log-EK: Log- Exponenciada Kumaraswamy .

Sumário

Resumo	7
Abstract	9
Lista de Figuras	10
Lista de Tabelas	12
Lista de Siglas	15
Capítulos	
1 Introdução	19
2 Conceitos Básicos	23
2.1 Radiação Solar	23
2.2 Método de Máxima Verossimilhança	24
2.3 Modelos de Markov	25
2.4 Correlação entre variáveis	26
2.4.1 Coeficiente de correlação de Pearson	26
2.4.2 Coeficiente de correlação de Spearman	26
2.4.3 Coeficiente de correlação de Kendall	27
2.5 Método de Seleção de modelos	27
3 Descrição do Problema	29
4 Imputação de Dados	33
4.1 Métodos de abordagem a dados faltantes	33
4.2 Imputação Única	34
4.2.1 Imputação por constantes	34
4.2.2 Imputação de médias condicionadas	34
4.3 Imputação Múltipla	35
4.4 Seleção e aplicação do método de imputação apropriado	35
5 Modelos Univariados	37
5.1 Distribuição Weibull Univariada	37
5.1.1 Estimação de Máxima Verossimilhança da distribuição Weibull univariada	39
5.2 Distribuição Gama Univariada	40
5.2.1 Estimação de Máxima Verossimilhança da distribuição Gama univariada	41
5.3 Distribuição Exponencial Generalizada Univariada	42

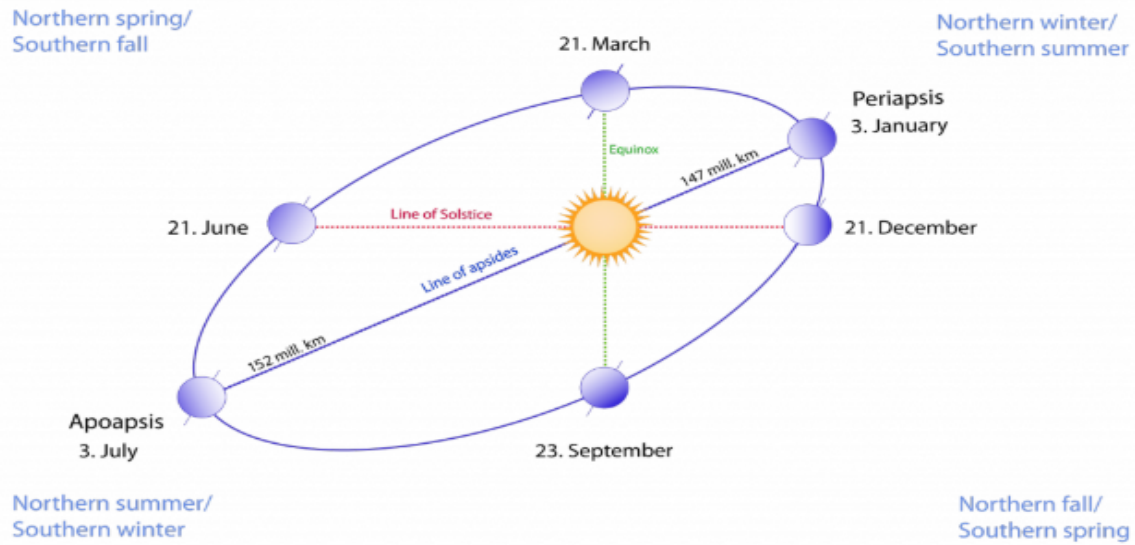
5.3.1	Estimação de Máxima Verossimilhança da distribuição Exponencial Generalizada univariada	44
5.4	Distribuição Normal Univariada	45
5.4.1	Estimação de Máxima Verossimilhança da distribuição Normal univariada	46
5.5	Distribuição Weibull Exponenciada	48
5.5.1	Estimação de Máxima Verossimilhança da distribuição Weibull Exponenciada	49
5.6	Generalizações da Distribuição Kumaraswamy	49
5.6.1	Distribuição Weibull Kumaraswamy	50
5.6.2	Distribuição Log-exponenciada Kumaraswamy	52
5.6.3	Distribuição Weibull Exponenciada Kumaraswamy	54
5.7	Resultado e discussão dos modelos univariados	56
6	Modelos Markovianos baseados em distribuições Bivariadas	63
6.1	Modelo Markoviano Weibull via Distribuição Bivariada Weibull de Freund	63
6.1.1	Distribuição Weibull Bivariada de Freund	64
6.1.2	Modelo Markoviano Weibull via a Distribuição Weibull Bivariada de Freund	66
6.1.3	Estimação de Máxima Verossimilhança do Modelo Markoviano Weibull via a Distribuição Weibull Bivariada de Freund	69
6.2	Modelo Markoviano Weibull via Distribuição Bivariada Weibull de Marshall-Olkin	73
6.2.1	Distribuição Weibull Bivariada de Marshall-Olkin	73
6.2.2	Modelo Markoviano Weibull via a Distribuição Weibull Bivariada de Marshall e Olkin	75
6.2.3	Estimação de Máxima Verossimilhança do Modelo Markoviano Weibull via a Distribuição Weibull Bivariada de Marshall e Olkin	76
6.3	Modelo Markoviano Weibull via Distribuição Bivariada Weibull Exponenciada Kumaraswamy	79
6.3.1	Bivariada Weibull Exponenciada Kumaraswamy	79
6.3.2	Modelo Markoviano Weibull via a Distribuição Bivariada Weibull Exponenciada Kumaraswamy	81
6.3.3	Estimação de Máxima Verossimilhança do Modelo Markoviano Weibull via a Distribuição Bivariada Weibull Exponenciada Kumaraswamy	83
6.4	Comparação dos modelos univariados e markovianos Weibull via as distribuições bivariada Weibull de Freund, Marshall-Olkin e Weibull Exponenciada Kumaraswamy	85
6.5	Escolha do melhor modelo	86
7	Considerações Finais	89
7.1	Perspectivas futuras	90
	Referências	90

Introdução

Atualmente é impossível viver sem o uso da energia elétrica, mas tamanha necessidade também traz imensos impactos ambientais e financeiros, pois existem diversas maneiras de se gerar energia, cada uma trazendo vantagens e desvantagens. Chamamos de Energia limpa, as fontes que são renováveis e que para sua geração não lançam poluentes (gases do efeito estufa) na atmosfera, causando impactos reduzidos na natureza. As energias que se enquadram nessas condições são: Energia eólica (gerada a partir da força do vento), Energia maremotriz (gerada a partir das ondas dos mares), Energia geotérmica (gerada a partir do calor da terra), Energia termoelétrica que é toda energia gerada a partir de gerações de calor resultante da queima de combustíveis sólidos, líquidos ou gasosos, Energia gerada pelas usinas hidrelétricas(gerada a partir de um potencial hidráulico de um rio) e Energia solar. A Energia Solar tem por característica única como o próprio nome diz, transformar as luzes dos raios solar em eletricidade, trazendo consigo um baixo impacto ambiental, dependendo apenas do clima da região onde está localizado.(Pítsica, 2015) [26]

O sol transmite sua energia gerada para o espaço através de ondas eletromagnéticas de amplo espectro de frequências. A distribuição da energia pelas diferentes frequências apresenta um ponto máximo, que depende da temperatura do sol. A terra é um planeta que orbita o sol e gira ao redor do eixo definido pelos polos Norte e Sul. Os movimentos da terra são: Rotação, Translação, Precessão, Nutação e Movimento no espaço. O conjunto destes movimentos afeta a energia recebida na terra como um todo e nos pontos específicos da superfície. Portanto, conhecê-los é indispensáveis para projetar usinas de energia solar. A energia solar recebida na terra depende da distância entre o sol e a terra e da área da terra exposta ao sol. Pode-se observar na Figura (1.1), diferentes estações do ano e a distância que a terra fica do sol.(Garcia, 2018) [14]

Figura 1.1: Processo de Translação



Fonte: (Garcia, Geração de energia elétrica. 2018) [14]

O nível de radiação na atmosfera é uma variável importante quando se fala em geração de energia solar, uma vez que, caso tenha um obstáculo entre o raio e as placas, a quantidade de energia gerada será menor. Radiação solar é a designação dada à energia radiante emitida pelo sol, sendo que existe três tipos: direta, difusa e refletida/Albedo. A direta é quando os raios refletidos nas placas solares são paralelos e vem diretamente do sol, a difusa são raios espalhados pela atmosfera terrestre reproduzido pelas nuvens e a Albedo são raios refletidos pelas superfícies adjacentes terrestres, tais como prédios, árvores, solos ou qualquer outro obstáculo. (Magarreiro, 2014) [20]

Além dos tipos de radiação, o clima e a localidade das placas também são fatores de grande relevância para se poder estimar a quantidade de energia solar gerada. Nesse contexto, tendo em mãos as informações dos dados históricos de energia solar é possível criar modelos estatísticos que ajustam a quantidade de energia que uma placa vai gerar. No entanto, para que isso seja possível é de grande importância, através de distribuições de probabilidades, enxergar como estão se comportando os dados ao longo do tempo, além de permitir uma perspectiva razoável do comportamento de gerações de energia fotovoltaicas de uma região, configurando assim uma ferramenta de grande valor para o planejamento e gestão de inúmeras atividades, até mesmo para o Governo (Santos, 2020) [32]. A criação de modelos através dessas funções de distribuição de probabilidade é muito importante para que se possa melhor explicar a geração de energia, como forma de planejamento para se investir em placas solares, levando em consideração que seu uso é caro comparado com outras fontes de energia.

A utilização de modelos estatísticos permite além de realizar previsões, realizar o estudo da dependência entre os tempos, ou seja, é possível observar se existe algum tipo de dependência entre os tempos com as variáveis de interesse. Diversas possibilidades de construção de modelos com estas características podem ser consideradas. Uma das suposições utilizadas nesse estudo, foi construção dos modelos através de um processo de Markov, ou seja, cada tempo observado depende somente do anterior. Para a construção

destes modelos foi necessário escolher as funções de distribuições de probabilidade que melhor se ajustam a dados de energia solar gerada por placas fotovoltaicas e essa escolha foi baseada em um estudo teste realizado por, Santos(2020) [32], onde foram constatados que as distribuições Weibull e Normal, apresentaram melhor ajuste aos dados. A partir dos possíveis modelos iniciais foi realizado estudos de distribuições bivariadas afim de comparar os modelos independentes e os markovianos.

A aplicação dos modelos foi realizada com base em um conjunto de dados da cidade de São José do Rio Preto-SP disponibilizada pela empresa (Estilo Solar) [2] detentora da quantidade de energia solar gerada por placas fotovoltaicas no período de Janeiro/2019 a Julho/2020. A comparação dos modelos pode ser feita através de análises gráficas ou por meio de estatísticas que podem ser utilizadas, entre elas os métodos de AIC, AICc e BIC.

Este trabalho está estruturado da seguinte forma: no capítulo 2 são apresentadas as revisões de conceitos básicos que são explorados ao longo do trabalho, no capítulo 3 é apresentada a descrição dos problemas, o conjunto de dados e o objetivo do trabalho, no capítulo 4 é estudada a imputação de dados, pois foi possível verificar anteriormente que há presença de dados faltantes no conjunto de dados, no capítulo 5 a construção de modelos univariados e análise gráfica com os dados ajustados. A seguir é dado o início do estudo de modelos markovianos Weibull e no capítulo 6 é apresentado a construção dos modelos markovianos baseados em distribuições bivariadas sendo elas a distribuição Weibull bivariada de Freund, a distribuição Weibull bivariada de Marshall e Olkin, e a distribuição Bivariada Weibull Exponenciada Kumaraswamy, ambos com a apresentação das estimações dos parâmetros e a análise gráfica. No capítulo 7 por fim são apresentados as considerações finais.

Tem-se como perspectivas futuras, a realização de análises com dados de outras regiões do estado ou do país com o objetivo de realizar comparações e a inclusão de covariáveis que possam ser utilizadas para capturar diferenças ao longo do tempo.

Conceitos Básicos

Este capítulo tem a função de mostrar uma breve explicação de alguns conceitos explorados ao longo da pesquisa.

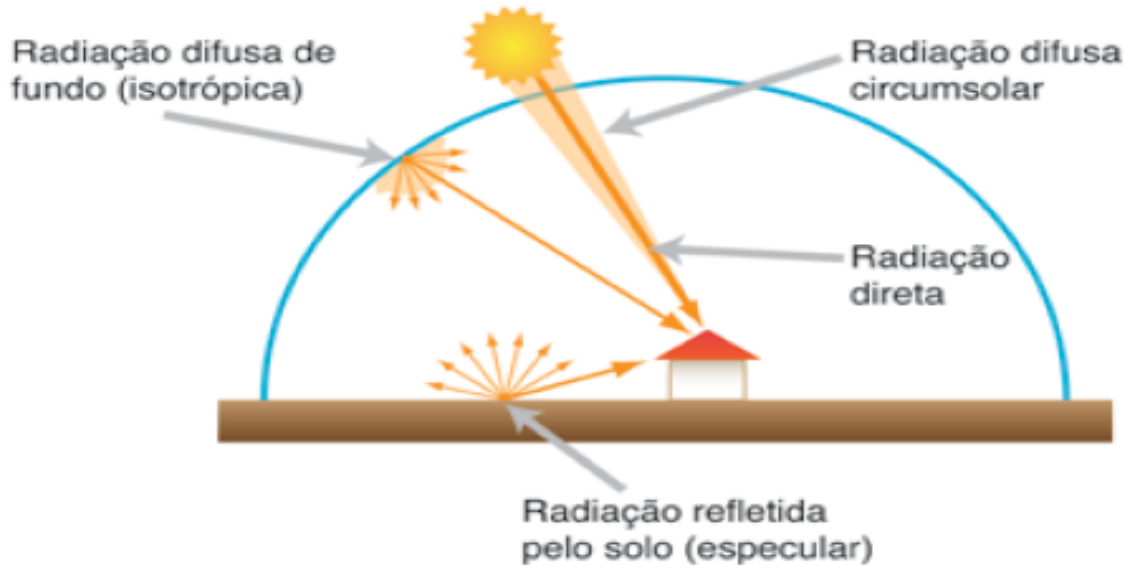
2.1 Radiação Solar

Quase toda a luz que ilumina o nosso planeta vem direta ou indiretamente do sol. A luz do dia e a iluminação natural são radiações eletromagnéticas e a radiação solar pode ser aproveitada para produzir eletricidade e alimentar o mundo em que vivemos. Ela se divide em três componentes principais, com características diferentes:

- A radiação direta (que projeta sombras) vem diretamente da direção do disco solar;
- A radiação difusa (que inclui a radiação solar de todas as direções do céu, com a exceção do disco solar)
- a radiação refletida (que resulta da reflexão da radiação solar na superfície), também denominada Albedo.

Na Figura (2.1) podemos observar como é cada representação.

Figura 2.1: Tipos de Radiação



Fonte: (Magarreiro, Radiação e energia solar. 2014) [20]

Essas características influenciam no potencial de energia solar fotovoltaica numa determinada região ou cidade em determinados instantes, além que os efeitos dos processos atmosféricos de absorção e dispersão da luz solar também dependem de condições meteorológicas. (Magarreiro, 2014) [20]

2.2 Método de Máxima Verossimilhança

Este método possibilita a obtenção de estimadores a partir da maximização da função de verossimilhança. A definição é dada por: Sejam X_1, \dots, X_n uma amostra aleatória de tamanho n da variável aleatória X com função de densidade (ou de probabilidade) $f(x|\theta)$, com $\theta \in \Theta$. A função de verossimilhança de θ corresponde a amostra aleatória observada, a qual é dada por:

$$L(\theta; x) = \prod_{i=1}^n f(x_i|\theta). \quad (2.1)$$

O estimador de máxima verossimilhança de θ é o valor $\hat{\theta} \in \Theta$ que maximiza a função de verossimilhança $L(\theta; x)$.

A função logarítmica é uma função monótona estritamente crescente e sendo assim, afim de facilitar o processo de obtenção dos estimadores de máxima verossimilhança (EMV), ela pode-se substituir pelo logaritmo da função verossimilhança de θ , que é denotado por:

$$l(\theta; x) = \ln L(\theta; x) = \sum_{i=1}^n \ln f(x_i|\theta). \quad (2.2)$$

Desta forma a função score é dada por:

$$l'(\boldsymbol{\theta}; x) = \frac{\partial l(\boldsymbol{\theta}; x)}{\partial \boldsymbol{\theta}} = 0. \quad (2.3)$$

É possível verificar que o valor de θ que maximiza a função $l(\boldsymbol{\theta}; x)$ é o mesmo valor que maximiza a função $L(\boldsymbol{\theta}; x)$.

O EMV $\hat{\theta}$ pode ser encontrado como a raiz da equação de verossimilhança. (Bolfarine, et al. 2010) [7]

2.3 Modelos de Markov

O estudo de Modelos Markovianos parte da teoria de cadeias de Markov que é um processo estocástico caracterizado por dependência entre os estados onde o futuro depende do seu estado atual. A garantia da existência de uma cadeia de markov, tem-se na construção de um espaço de probabilidade para tal processo estocástico.

Seja um vetor de variáveis aleatórias $T = (T_1, T_2, \dots, T_n)$, sendo que T_1, T_2, \dots, T_n são variáveis aleatórias dependentes, ou seja, a ocorrência de T_{i+1} é dependente da ocorrência de T_i , assim existe probabilidade de transição entre as variáveis. Logo, é possível ser representada pela probabilidade condicionada.

$$P(T_{i+1}|T_i) = \frac{P(T_{i+1}, T_i)}{P(T_i)}$$

ou seja,

$$P(T_{i+1}, T_i) = P(T_{i+1}|T_i)P(T_i).$$

Portando como é conhecida a probabilidade de transição de um estado para outro, define a probabilidade sobre o vetor aleatório como: (Chung, 1960) [9]

$$P(T_1 = i_1, T_2 = i_2, \dots, T_n = i_n) = P(T_1 = i_1)P(T_2 = i_2|T_1 = i_1) \dots P(T_n = i_n|T_{n-1} = i_{n-1}).$$

Para as realizações de inferências estatísticas em modelos markovianos a ideia segue sendo a mesma, seja T_i , $i = 1, 2, \dots, n$, variáveis aleatórias dependentes, a função de verossimilhança fica da seguinte forma:

$$L(\boldsymbol{\theta}, t) = f(t_1)f(t_2/t_1)f(t_3/t_1, t_2) \dots f(t_n/t_1, t_2, \dots, t_{n-1})$$

e por seguir um processo de Markov, a função de verossimilhança é dada por:

$$L(\boldsymbol{\theta}; t) = f(t_1) \prod_{i=1}^{n-1} f(t_{i+1}/t_i). \quad (2.4)$$

No Capítulo 6 será introduzido a construção de distribuições bivariada que segue o um processo de Markov.

2.4 Correlação entre variáveis

Quando se fala em dependência entre duas variáveis, se supõe que ambas estão correlacionadas entre si, ou seja, uma interfere no desenvolvimento da outra. Existem métodos para que se possa mensurar o quão dependente uma variável está da outra, e eles são chamados de coeficiente de correlação. Um coeficiente de correlação mede o grau pelo qual duas variáveis tendem a mudar juntas, além de precisar a direção e a força da relação entre as variáveis. Existem diferentes tipos de correlação, algumas delas são apresentadas a seguir.

2.4.1 Coeficiente de correlação de Pearson

O coeficiente de correlação de Pearson avalia a relação linear entre duas variáveis contínuas, ou seja, verificar se o aumento ou o decréscimo de algum valor X é associado a um aumento ou o decréscimo de algum valor Y . O coeficiente de correlação de Pearson é representado pela letra ρ e assume valores de -1 a 1 . Quando $\rho = 1$, representa a correlação perfeita e positiva entre duas variáveis, quando $\rho = -1$, representa correlação perfeita negativa entre duas variáveis, ou seja, enquanto uma aumenta a outra diminui, à medida que se aproxima do 1 vai ficando perfeita a correlação e quando $\rho = 0$ não existe correlação entre as variáveis. Assim, ρ é dada pela expressão (2.5) (Rufato, et al. 2010) [29]

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (2.5)$$

onde x_i e y_i podem ser qualquer variável aleatória contínua, $i=1, 2, \dots, n$.

2.4.2 Coeficiente de correlação de Spearman

O coeficiente de correlação de Spearman é conhecido como coeficiente de correlação por postos de Spearman, por calcular as correlações entre variáveis mensuradas em nível ordinal. Inicialmente as correlações ordinais não mostram necessariamente tendência linear, mas podem ser consideradas como índices de monotonicidade, ou seja, para aumentos positivos da correlação, aumentos no valor de X correspondem a aumentos no valor de Y , e para coeficientes negativos ocorre o oposto (Bunchaft e Kellner, 1999) [8]. O coeficiente de correlação de Spearman é representado pela letra ρ_s e é dada da seguinte expressão derivada de Pearson (Neto e Lira, 2006) [24]

$$\rho_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}, \quad (2.6)$$

onde, d_i é a diferença entre as ordenações de x_i e y_i e n é o número de pares de ordenações. Este coeficiente também assume valores de -1 a 1 , para detectar a correlação.

2.4.3 Coeficiente de correlação de Kendall

O coeficiente de correlação de Kendall, semelhante ao de Spearman, também utiliza os dados ordenados em postos, porém possui interpretações diferentes. Enquanto Kendall, representa uma probabilidade, em que as variáveis ordenadas estão na mesma ordem, Spearman nos mostra a proporção da variância explicada. Assim ρ_k é dada pela seguinte expressão:

$$\rho_k = \sum_{\text{pares}(i,j), i < j} \frac{2 \text{sign}(x_i - x_j) \text{sign}(y_i - y_j)}{n(n^2 - 1)}, \quad (2.7)$$

onde $\text{sign}(\cdot)$ representa o sinal da diferença entre as ordenadas, quando $x_i > x_j$ $\text{sign}(x_i - x_j) = +1$ e quando $x_i < x_j$ $\text{sign}(x_i - x_j) = -1$, o n é o número de pares de ordenações. (Origuela, 2018) [25]

2.5 Método de Seleção de modelos

Utilizar critérios que nos ajude na comparação e seleção de modelos estatísticos é de extrema importância em uma análise de dados. Existem diversas maneiras que nos auxiliam para tal escolha, a seguir serão apresentados alguns critérios de comparação.

O primeiro é o Critério de Informação Akaike (AIC) (Akaike, 1974) [3]. Este critério propõe um método baseado na distância de Kullback-Leibler (a informação de Kullback-Leibler é considerada como uma medida de discriminação entre o modelo proposto $g(x)$ e o modelo verdadeiro $f(x)$). Assim, seja k o número de parâmetros a serem estimados, n o número de observações de x e $\hat{\theta}$ um critério de informação de θ . O critério de informação de Akaike (AIC) é dado por:

$$AIC = -2 \ln(L(\hat{\theta}; x)) + 2k. \quad (2.8)$$

Dado um conjunto de modelos candidatos para x , com os dados ajustados, o melhor modelo será aquele que fornecer o menor AIC. Além de selecionar um ótimo ajuste, o critério penaliza a adição de parâmetros, não permitindo assim *overfitting*, ou seja, a seleção de um modelo que se adapta muito bem aos dados com os quais está sendo treinado.

O Critério de Informação de Akaike Corrigido (AICc) pode ser usado quando o número do parâmetro k é grande em relação ao tamanho da amostra n , situação em que os resultados que proporciona AIC podem ser ruins. Neste contexto, deve-se utilizar o critério de segunda ordem, a menos que $\frac{n}{k} > 40$ aproximadamente. Como o AICc converge para o AIC, quando n é grande, na prática, deve ser usado o AICc que é dado por:

$$AICc = -2 \ln(L(\hat{\theta})) + 2k + \frac{2k(k+1)}{n-k-1}. \quad (2.9)$$

O Critério de Informação Bayesiana (BIC), foi desenvolvido por Schwarz(1978) [33]. Este critério possui estrutura semelhante ao AIC, mas é adotado sobre o método Bayesiano. O BIC é obtido através de resultados assintóticos e da suposição de que os dados pertencem à família exponencial. A equação é dada por:

$$BIC = -2\ln(L(\hat{\theta}; x)) + k \ln(n). \quad (2.10)$$

Semelhante ao AIC, o critério penaliza a adição de parâmetros. Dos modelos candidatos, o adequado será o que possuir o menor BIC.

Descrição do Problema

Atualmente, um dos assuntos que necessitam de visibilidade em debates é o da energia limpa, que com o aumento do aquecimento global seu uso é de extrema importância, pois em geral possui um baixo impacto ambiental, uma vez que não produz dióxido de carbono ou outros gases com efeito de estufa. Constituem energias renováveis, por exemplo, a solar, a eólica, as maremotriz e a hidráulica. Todas elas, para além da sua presença ilimitada sobre a terra e da sua inocuidade ao meio ambiente, têm uma origem comum: o sol. Ele é o responsável direto pela energia solar, mas está também na origem da energia eólica, ao provocar as diferenças de pressão que dão origem ao vento e à energia das ondas. Pensando na energia solar, os fatores climáticos são covariáveis que implicam diretamente no desempenho da energia gerada, além dos locais e inclinação de instalações das placas fotovoltaicas. Levando em consideração o teor de comparação, placas instaladas em duas regiões tropicais, mas uma com muitos obstáculos para o contato do sol, como prédios, árvores, entre outros e a outra sem nenhum obstáculo, é de se esperar que a segunda tenha uma maior quantidade de energia gerada ao longo do tempo. Regiões com um nível baixo de radiação ou muitas nuvens também tende a ter um baixo desempenho de energia gerada. Segundo o sistema de classificação climática "SCC" de Köppen e Geiger (Köppen, et al.1928) [16], o clima de São José do Rio Preto é caracterizado como tropical chuvoso, com verões chuvosos e quentes, invernos secos e mais frios, sendo a temperatura média normal do mês mais quente ($>= 22.8^{\circ}\text{C}$) (Rolim, et al. 2007) [28].

Construção de modelos podem ajudar a entender o consumo de energia. Ou seja, ela pode explicar a relação entre a geração de energia solar em tais localidades e os problemas gerados por fenômenos naturais, como altas temperaturas, ondas de calor, ou mudanças de temperatura em determinadas épocas. Pode também direcionar formas de se gerar energia limpa com um menor custo financeiro e ambiental, proporcionando bom proveito a sociedade.

Para este trabalho tem-se em mãos um conjunto de dados de energia solar, fornecida pela empresa (estilo Solar) [2] localizada na cidade de São José do Rio Preto-SP. No conjunto de dados, são fornecidos a quantidade de energia gerada pelas placas solares em determinadas residências em tempo real, através do sistema **APsystems Alteryg power** [1]. O sistema nos mostra a quantidade de placas que tem uma residência, que é um fator importante para a estimação de energia. Pode-se também realizar filtros de interesse para visualização dos dados, sendo eles a quantidade de energia gerada por hora, por dia, semana, mês ou ano. A Figura (3.1) nos mostra uma tabela que contém uma exemplificação de como estão armazenados os dados filtrados em relação a quantidade de

energia gerada em cada dia dos meses de análise. (Nesse caso a residência é localizada na cidade de São José do Rio Preto-SP, onde se encontra 6 placas fotovoltaicas e os dados são fornecidos pela empresa Estilo Solar Engenharia.) [2]

O estudo será feito a partir do dia 21 de Fevereiro de 2019 até o final de Julho 2020, com o interesse de criar um modelo que nos dê a quantidade de energia solar que pode ser gerada por placas fotovoltaicas com diferentes estações e climas do ano.

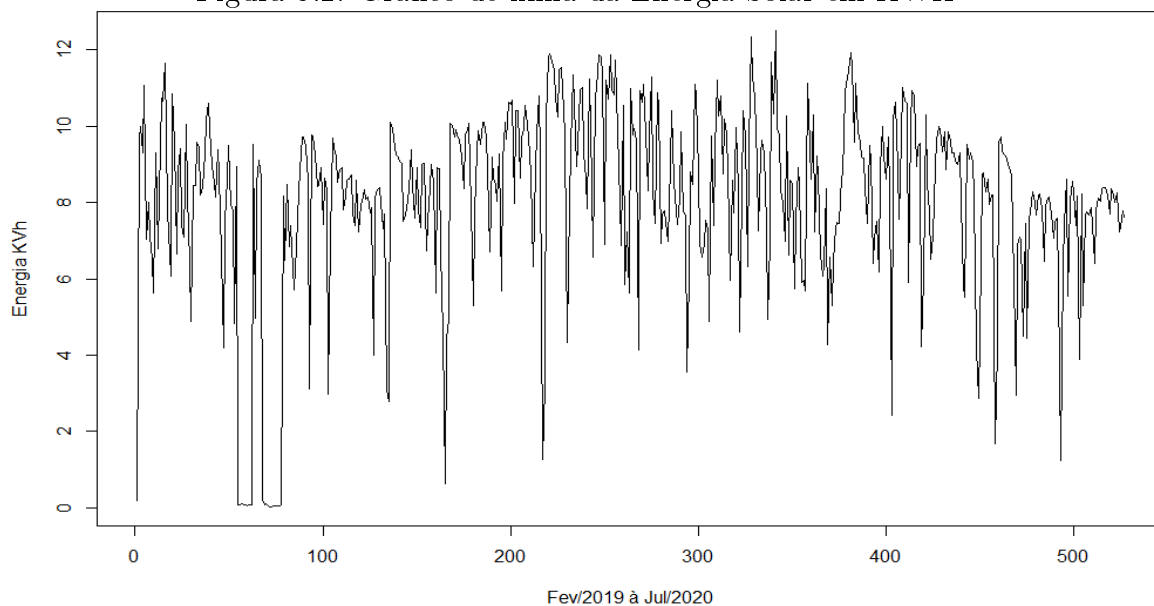
Figura 3.1: Energia gerada pelas placas fotovoltaicas

1	Data	energiaKWH	Dia
2	Fevereiro	0,18	21
3	Fevereiro	9,69	22
4	Fevereiro	9,98	23
5	Fevereiro	9,30	24
6	Fevereiro	11,06	25
7	Fevereiro	7,03	26
8	Fevereiro	8,00	27
9	Fevereiro	7,11	28
10	Março	6,50	1
11	Março	5,62	2
12	Março	9,30	3
13	Março	6,79	4
14	Março	8,23	5
15	Março	10,73	6
16	Março	10,65	7
17	Março	11,65	8
18	Março	7,80	9
19	Março	7,19	10
20	Março	6,06	11
21	Março	10,83	12
22	Março	9,55	13
23	Março	6,66	14
24	Março	8,84	15
25	Março	9,40	16
26	Março	7,42	17
27	Março	7,10	18

Fonte: (Estilo Solar Engenharia. São José do Rio Preto-SP 2019) [2]

É possível apresentar uma serie temporal da quantidade de Energia gerada no período de Fev/2019 até Jul/2020 do conjunto de dados usado. Podendo ser observadas as oscilações ao longo do tempo.

Figura 3.2: Gráfico de linha da Energia Solar em KWH



Fonte: (Estilo Solar Engenharia. São José do Rio Preto-SP 2020) [2]

O gráfico apresentado na Figura (3.2) é uma forma de se visualizar ao longo do tempo como é a quantidade de energia solar por residência.

Também é possível verificar alguns resultados em que houve sequências de dias com uma quantidade quase zero de energia gerada. Isso pode ter ocorrido por problemas técnicos envolvendo a medição de energia solar naquele dia, ou problemas com o aparelho. Nesta situação pode ser sugerido uma técnica de imputação de dados para substituir essas falhas, para que assim na hora da estimação dos modelos não se tenha problemas com as observações com valores de medições faltantes. Ao longo do estudo, esta técnica será apresentada em detalhes.

Imputação de Dados

Neste capítulo será apresentado métodos de imputação para as informações faltantes no conjunto de dados, onde também é discutido quais critérios que se devem levar em consideração para a utilização de diferentes técnicas. Após as devidas análises para o conjunto de dados de energia solar gerada por placas fotovoltaicas da cidade de São José do Rio Preto optou-se por utilizar a imputação única, utilizando o método de médias condicionadas. Como já mencionado no capítulo anterior ao longo das medições foi possível observar falhas, onde não é possível ter a quantidade exata de energia gerada em alguns dias, portanto os dados faltantes serão preenchidos por meio de imputação para que assim se possa construir os modelos.

4.1 Métodos de abordagem a dados faltantes

Quando trabalhamos com conjunto de dados reais, inicialmente é necessário verificar a consistência dos dados. Em algumas situações é necessário fazer a limpeza dos dados, em outras é necessário tratar os casos com dados faltantes. Quando uma informação de uma variável está vazia ou incompleta, ela acaba por se tornar um dado faltante. Isso pode ocorrer por alguma razão em que fez-se não possível incluí-la. Essa informação de algum modo deveria estar presente na variável, mas por algum motivo não foi possível incluí-la, e conseqüentemente quando não se tem uma informação completa o resultado que esperamos dessa variável ou conjunto de dados acaba sendo comprometido comparado com o que diria se tivesse com as informações completas. Dados faltantes também podem ser classificados por dados ruidosos, que são aqueles que apresentam um desvio significativo do valor real, representando uma inconsistência na informação. (Myrtveit et al., 2001) [23]

É importante saber o mecanismo que gerou a omissão dos dados, pois esse conhecimento auxilia na escolha do melhor tratamento dos dados faltantes. Existem três principais tipos de mecanismos que são:

MCAR (missing completely at random): ocorre quando a causa que gerou os dados faltantes é um evento completamente aleatório. (Veroneze, 2011) [35]

MAR (missing at random): ocorre quando os dados faltantes presentes em uma atributo dependem dos valores observados em outro(s) atributo(s). (Veroneze, 2011) [35]

MNAR (missing not at random): ocorre quando os dados faltantes presentes em um atributo dependem de seu próprio valor. (Veroneze, 2011) [35]

No presente trabalho trata-se do mecanismo MCAR, pelo fato das observações faltantes estarem sendo apresentadas de forma aleatória, assim o próximo passo é discutir as técnicas a serem utilizadas em diferentes comportamentos de dados faltantes. Existem diversos métodos para se trabalhar com essas informações, porém a seguir serão apresentadas as duas mais comuns e importantes na literatura, sendo elas Imputação Única (IU) e Imputação Múltipla (IM)

4.2 Imputação Única

O método de IU consiste em apenas preencher cada dado faltante do conjunto de dados e realizar as análises desejáveis. Portanto na etapa de IU, o único interesse é em prever os valores faltantes. Existem diversas técnicas dentro do método de IU. A seguir serão apresentados algumas deles:

4.2.1 Imputação por constantes

A imputação por constantes são as mais comuns e simples quando se fala em IU, pois se trata apenas de substituir os atributos faltantes por uma única informação, sendo ela considerada uma constante. Os valores mais utilizados neste método é a imputação por zeros, média ou mediana. (Veroneze, 2011) [35]

Quando se fala em imputação de zeros se deve tomar alguns cuidados, pois esse método só é recomendável quando nosso conjunto de dados tem valores próximos a zero e esse número represente de alguma forma falhas ou maus resultados. Outra questão muito importante é o número de dados faltantes em uma variável, pois com esse valor muito elevado a imputação de zeros, pode gerar resultados contraditórios com a verdade.

A imputação da média é o método mais comum se falando de IU, pois além de ser de fácil implementação apresenta resultados bons nas análises. Porém deve-se tomar cuidado novamente com a quantidade de dados faltantes presentes no conjunto de dados, pois com um número grande a média pode gerar a perda de variabilidade nas variáveis. Essa imputação apresenta bons resultados em variáveis que são normalmente distribuídas.

Na imputação da mediana, segue um procedimento semelhante a imputação da média, a única diferença é que ela apresenta um melhor desempenho em distribuições que não são normalmente distribuídas, pois a mediana tem melhor tendência central de uma distribuição que possui grandes desvios da distribuição normal (Veroneze, 2011) [35]

4.2.2 Imputação de médias condicionadas

A imputação de médias condicionadas ou CMI (Conditional Mean Imputation), diferentemente da imputação de médias, ao invés de substituir todos os dados faltantes por uma constante, ela adiciona cada informação da média em diferentes grupos, de acordo com a escolha do pesquisador. Por exemplo, em um conjunto de dados que contém uma variável "Estado Civil" e tem-se um dado faltante para uma pessoa solteira, este será subs-

tituído pela média calculada dentro do grupo de pessoas solteiras. (Veroneze, 2011) [35]

IU apresenta outros métodos de imputação, como Hot Deck, Cold Deck, NVCB(Value Carried Backward), LVCF(Last Value Carried Forward) entre outros, cada um apresentando suas vantagens e desvantagens. Cabe ao pesquisador se atentar a todas elas e escolher a que melhor se encaixe ao seu conjunto de dados e sua necessidade.

4.3 Imputação Múltipla

Os métodos de IM foram propostos com o intuito de ser uma alternativa em relação ao métodos de imputação por Máxima Verossimilhança, (que visa realizar imputação nos dados faltantes tendo como o objetivo, a estimação dos parâmetros da função de verossimilhança. Assim seu objetivo não é prever os valores faltantes (Allison, 2001) [4]), para uma grande demanda de problemas envolvendo os dados faltantes.

IM apresenta diversas vantagens em relação aos outros métodos de imputação. Entre todas as estimativas dos dados faltantes, é a mais utilizada, pois são mais confiáveis que os demais métodos, e pode ser aplicado em diferentes situações de dados faltantes (Veroneze, 2011) [35]. Uma das principais diferenças entre IU e IM é que ao invés de substituir os dados faltantes por um único valor, como é feito em IU, nos métodos de IM os atributos faltantes recebem x valores, $x \geq 2$. Assim são formados x bases de dados completas, que podem ser analisadas e seus resultados são ligados uns nos outros formando no fim valores que preenchem os dados faltantes. (Veroneze, 2011) [35]

Como acontece em métodos de IU, dentro de IM existem diversos métodos que são adequados para diferentes tipos de dados, assim deve-se se atentar na melhor escolha do método a ser utilizado. Existem etapas para a realização da técnicas de IM, que podem ser divididas em : Imputação, Análise, Agregação dos resultados e Cálculo da informação faltante. Em (Veroneze, 2011) é possível ter mais detalhes sobre cada uma dessas etapas que são de extrema importância para a construção dos métodos de IM.

4.4 Seleção e aplicação do método de imputação apropriado

Antes de realizar a imputação no seu conjunto de dados, deve-se saber quais métodos serão utilizados, para que os resultados obtidos sejam os melhores possíveis. Como já foi mostrado neste capítulo há diferenças significativas entre cada técnica de imputação, assim a escolha do melhor método de imputação de dados, neste estudo baseia-se na técnica sugerida por (Frank, 2001) [12], onde ele utiliza os seguintes critérios para a escolha dos métodos:

- Quando a proporção de dados faltantes é menor que 0,05 aplica-se a IU ou apenas analisa os dados completos;
- Quando a proporção de dados faltantes esta entre 0,05 e 0,15, pode-se utilizar a IU, mas é recomendável utilizar a IM;

- Quando a proporção de dados faltantes é maior que $\geq 0,15$: aplica-se a IM na grande maioria dos casos.

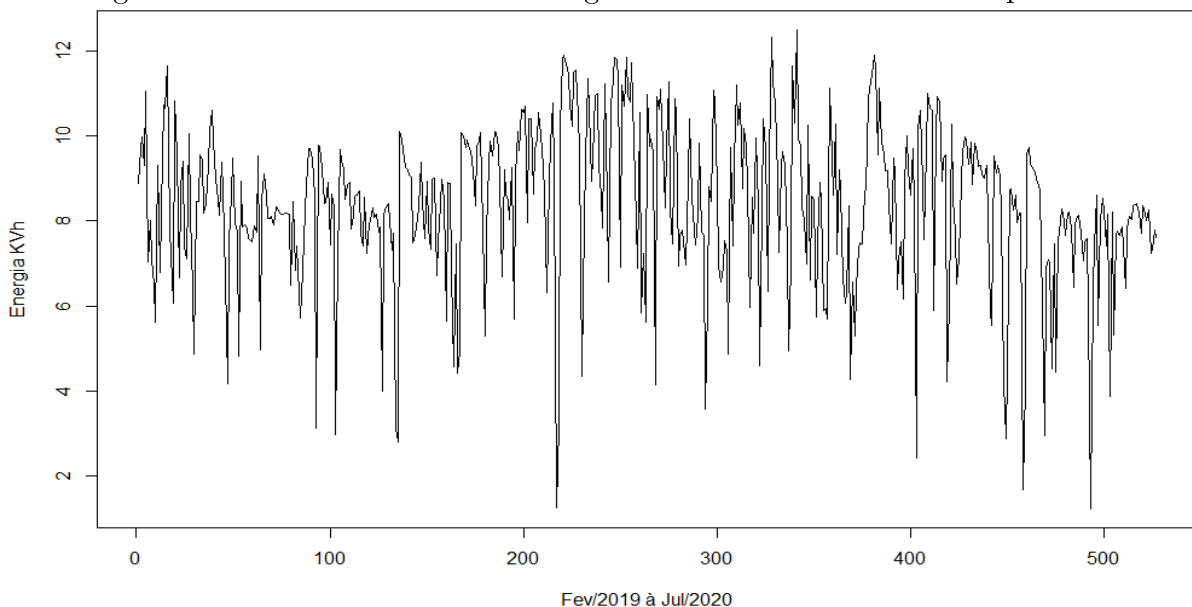
No conjunto de dados desta pesquisa, a proporção de dados faltantes é 0,039, assim seguindo o que sugere (Frank, 2001), será utilizado o método de IU. Como já foi mostrado existem diversas técnicas de imputação dentro no método de IU. Nesse caso a escolha é pelo método de imputação de médias condicionadas.

Na Figura (3.2) é possível observar como estão esses dados e visualizar onde estão as informações dadas como faltantes, sendo elas aquelas que apresentaram valores de geração de energia 0 KWHs ou um valor bem próximo. Os critérios escolhidos são da seguinte forma:

- A média de sete dias anterior a ocorrência dos dados faltantes;
- Essas médias serão preenchidas de acordo com o dia que contém dados faltantes, ou seja, se um dia tem um dado faltante, ele é preenchido pela média dos sete dias anteriores a ele;
- Como a primeira informação do conjunto de dados está sem preenchimento, ele será substituído pela média dos sete dias posteriores a ele.

Na Figura (4.1) é possível verificar como estão distribuídos o conjunto de dados após a realização de da imputação dos dados.

Figura 4.1: Gráfico de linha da Energia solar em KWHs. Dados Imputados



Fonte: (Estilo Solar Engenharia. São José do Rio Preto-SP 2020)) [2]

Assim, com o processo de imputação de dados realizado, nos próximos capítulos teremos o início da construção de modelos que ajustem esses dados.

Modelos Univariados

Antes de apresentar a construção das distribuições bivariadas foi necessário escolher as funções de distribuições de probabilidade que melhor se ajustam a dados de energia solar gerada por placas fotovoltaicas. Essa escolha se deu a partir de um estudo realizado em (Santos, 2020) [32]. Neste capítulo será apresentado um estudo mais aprofundado sobre essas distribuições e como elas se ajustaram em relação aos dados da cidade de São José do Rio Preto-SP.

5.1 Distribuição Weibull Univariada

A distribuição Weibull teve um início remoto em meados dos anos de 1920 proposta por Waloddi Weibull, onde ele realizou uma ampla aplicabilidade para a comunidade científica. Em 1951, ele realizou estudos relacionados ao tempo de falha devido a fadiga de metais, onde marcou o início do progresso triunfante desta distribuição na teoria estatística, bem como na estatística aplicada. Isso aconteceu por ela ter uma imensa variabilidade de formas, ambas com propriedades básicas, por exemplo, a sua taxa de falha é monótona, isto é, ela é estritamente crescente, decrescente ou constante. Grande parte das aplicações da distribuição Weibull se refere a dados da vida útil de unidades técnicas ou biológicas. No presente estudo a distribuição Weibull será ajustada a dados de energia solar, onde T_i , com $i = 1, 2, \dots, n$ denota a quantidade de energia solar gerada em cada dia i . A princípio neste capítulo foi considerado independência entre as observações T_i , $i = 1, 2, \dots, n$. A distribuição Weibull(α, β), tem sua função densidade de probabilidade dada por:

$$f(t) = \frac{\alpha}{\beta^\alpha} t^{\alpha-1} \exp \left\{ - \left(\frac{t}{\beta} \right)^\alpha \right\}, t \geq 0, \quad (5.1)$$

onde $\alpha > 0$ é o parâmetro de forma e $\beta > 0$ o parâmetro de escala. A sua função distribuição acumulada é dada por:

$$F(t) = 1 - \exp \left\{ - \left(\frac{t}{\beta} \right)^\alpha \right\} \quad (5.2)$$

para $t \geq 0$, α e $\beta > 0$. A função de sobrevivência e de risco são, respectivamente, dadas por:

$$S(t) = \exp \left\{ - \left(\frac{t}{\beta} \right)^\alpha \right\}, \quad (5.3)$$

$$\lambda(t) = \frac{\alpha}{\beta^\alpha} t^{\alpha-1}, \quad (5.4)$$

para $t \geq 0$, α e $\beta > 0$.

Caso se tenha o interesse em obter as expressões da média e variância da Weibull, deve-se trabalhar com a função gama, pelo fato que:

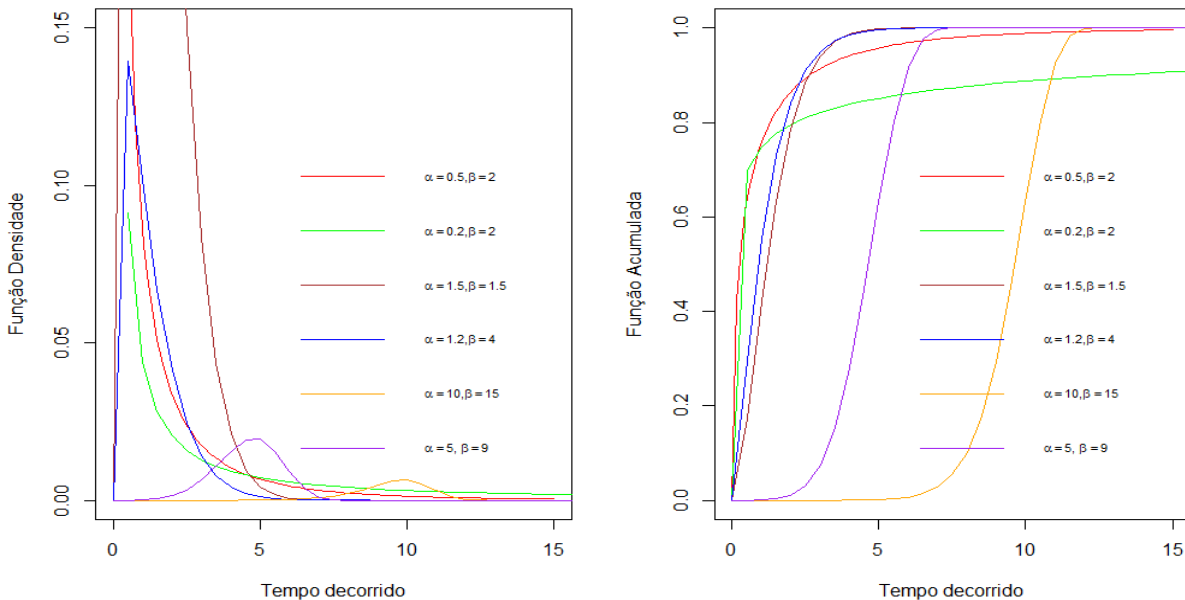
$$E(T) = \beta \Gamma \left[1 + \frac{1}{\alpha} \right],$$

$$Var(T) = \alpha^2 \left[\Gamma \left[1 + \frac{2}{\alpha} \right] - \Gamma \left[1 + \frac{1}{\alpha} \right]^2 \right],$$

sendo a função gama, $\Gamma(r)$, definida por $\Gamma(k) = \int_0^\infty x^{k-1} \exp\{-x\} dx$.

Na Figura (5.1) são apresentadas as formas da função de densidade de probabilidade e da distribuição acumulada, com diferentes valores de α e β .

Figura 5.1: Formas típica da função de densidade de probabilidade e da distribuição acumulada, da distribuição Weibull para alguns valores dos parâmetros α e β



Em seguida será realizada a estimação dos parâmetros da distribuição Weibull, afim de comparar com outras distribuições se realmente ela apresenta um melhor ajuste para que se tenha assim a mais perfeita explicação dos dados T_i (Quantidade de Energia solar gerada por placas fotovoltaicas de São José do Rio Preto).

5.1.1 Estimação de Máxima Verossimilhança da distribuição Weibull univariada

O Estimador de Máxima Verossimilhança (EMV) como introduzido na seção (2.2), trata do problema de estimação dos parâmetros a partir de resultados de uma amostra, escolhendo os estimadores que melhor explique a amostra observada. É possível obter esse resultado para dados de energia solar resultante da distribuição Weibull substituindo a expressão (5.1) em (2.1).

$$\begin{aligned} L(\alpha, \beta) &= \prod_{i=1}^n [f(t_i; \alpha, \beta)] \\ &= \prod_{i=1}^n \frac{\alpha}{\beta^\alpha} t_i^{\alpha-1} \exp \left\{ - \left(\frac{t_i}{\beta} \right)^\alpha \right\}. \end{aligned} \quad (5.5)$$

Aplicando o logaritmo na função para facilitar os cálculos temos:

$$\begin{aligned} \ln(L(\alpha, \beta)) &= \ln \left[\prod_{i=1}^n \frac{\alpha}{\beta^\alpha} t_i^{\alpha-1} \exp \left\{ - \left(\frac{t_i}{\beta} \right)^\alpha \right\} \right] \\ &= \sum_{i=1}^n \ln(\alpha) - \sum_{i=1}^n \alpha \ln(\beta) + (\alpha - 1) \sum_{i=1}^n \ln(t_i) - \beta^{-\alpha} \sum_{i=1}^n t_i^\alpha \\ &= n \ln(\alpha) - n\alpha \ln(\beta) + (\alpha - 1) \sum_{i=1}^n \ln(t_i) - \sum_{i=1}^n \left(\frac{t_i}{\beta} \right)^\alpha. \end{aligned} \quad (5.6)$$

Agora, fazendo as derivadas parciais da expressão $\ln(L(\alpha, \beta))$ em relação aos parâmetros α e β e igualando o resultado a zero, se tem os seguintes sistemas de equações:

$$\frac{\partial(\ln(L(\alpha, \beta)))}{\partial \alpha} = n \frac{1}{\alpha} - n \ln(\beta) + \sum_{i=1}^n \ln(t_i) - \sum_{i=1}^n \left(\frac{t_i}{\beta} \right)^\alpha \ln \left(\frac{t_i}{\beta} \right) = 0 \quad (5.7)$$

$$\frac{\partial(\ln(L(\alpha, \beta)))}{\partial \beta} = -n\alpha \frac{1}{\beta} + \sum_{i=1}^n \alpha t_i^\alpha \left(\frac{1}{\beta} \right)^{\alpha+1} = 0. \quad (5.8)$$

Os estimadores de máxima verossimilhança são os valores de $\hat{\alpha}$ e $\hat{\beta}$ que satisfazem as equações (5.7) e (5.8) respectivamente ou equivalentemente maximizam o logaritmo da função de verossimilhança. Para concluir que as soluções das equações (5.7) e (5.8) são pontos de máximo, tem-se que verificar se o sinal da segunda derivada das mesmas é positivo. A solução destas equações para o conjunto de dados não possui forma fechada, então pode ser obtida por meio de um método numérico iterativo, como por exemplo, os de Quase Newton, Newton-Raphson entre outros. O programa R tem o pacote Optim que minimiza uma função. Para o seu uso, temos que declarar como função objetivo o negativo do logaritmo da função de verossimilhança e especificar qual método utilizar. Neste caso utilizamos o método de Quase Newton. Pode-se ver os resultados obtidos na Tabela

(5.1) juntamente com as estimações dos parâmetros de outras funções de distribuição de probabilidade e logo em seguida a partir da Figura (5.9) tem-se as comparações de qual modelo univariado obteve o melhor ajuste ao nosso conjunto de dados.

5.2 Distribuição Gama Univariada

A distribuição Gama usada por Brown e Flood(1947) é bastante utilizada para dados de sobrevivência. Seja T com distribuição $Gama(\alpha, \beta)$, onde $\alpha > 0$ é chamado parâmetro de forma e $\beta > 0$ de escala, e a sua função de densidade de probabilidade é dada por:

$$f(t) = \frac{\beta^\alpha t^{\alpha-1}}{\Gamma(\alpha)} \exp\{-\beta t\}, t \geq 0, \quad (5.9)$$

onde a função gama, é definida por $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp\{-x\} dx$. A função distribuição acumulada é dada por:

$$F(t) = \frac{\gamma(\alpha, t\beta)}{\Gamma(\alpha)}, \quad (5.10)$$

para $t \geq 0$, $\alpha > 0$, $\beta \geq 0$, onde $\gamma(\alpha, t\beta)$ é uma função gama incompleta e tem a forma dada por $\gamma(\alpha, t\beta) = \int_0^{t\beta} x^{\alpha-1} \exp\{-x\} dx$. As funções de sobrevivência e de risco da distribuição Gama são dadas, respectivamente, por:

$$S(t) = 1 - \frac{\gamma(\alpha, t\beta)}{\Gamma(\alpha)}, \quad (5.11)$$

$$\lambda(t) = \frac{\beta^\alpha t^{\alpha-1} \exp\{-\beta t\}}{\Gamma(\alpha) - \gamma(\alpha, t\beta)}, \quad (5.12)$$

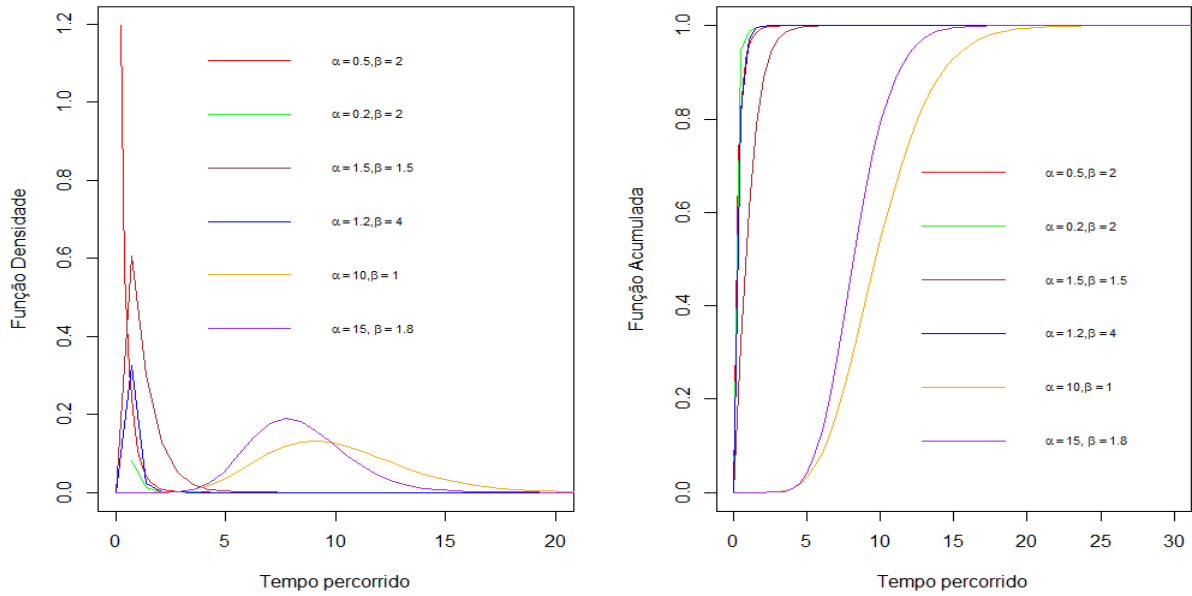
para $t \geq 0$, $\alpha > 0$, $\beta \geq 0$.

É possível observar algumas formas da função de densidade de probabilidade e da distribuição acumulada, de uma variável T com distribuição Gama na Figura (5.2) com diferentes valores de α e β .

A expressão da média e variância da distribuição Gama é dada por:

$$E(T) = \frac{\alpha}{\beta}; Var(T) = \frac{\alpha}{\beta^2}.$$

Figura 5.2: Formas típicas da função de densidade de probabilidade e da distribuição acumulada, da distribuição Gama para alguns valores dos parâmetros α e β



A seguir será realizado a estimação dos parâmetros da distribuição Gama pelo EMV com os dados da Quantidade de Energia solar gerada por placas fotovoltaicas de São José do Rio Preto.

5.2.1 Estimação de Máxima Verossimilhança da distribuição Gama univariada

Para esta distribuição de probabilidade, substituindo a expressão (5.9) em (2.1), a função de verossimilhança pode ser escrita como:

$$\begin{aligned}
 L(\alpha, \beta) &= \prod_{i=1}^n [f(t_i; \alpha, \beta)] \\
 &= \prod_{i=1}^n \left[\frac{\beta^\alpha t_i^{\alpha-1}}{\Gamma(\alpha)} \exp\{-\beta t_i\} \right].
 \end{aligned} \tag{5.13}$$

Aplicando o logaritmo na função para facilitar os cálculos temos:

$$\begin{aligned}
 \ln(L(\alpha, \beta)) &= \ln \left[\prod_{i=1}^n \frac{\beta^\alpha t_i^{\alpha-1}}{\Gamma(\alpha)} \exp\{-\beta t_i\} \right] \\
 &= \sum_{i=1}^n \alpha \ln(\beta) + (\alpha - 1) \sum_{i=1}^n \ln(t_i) - \sum_{i=1}^n \ln \Gamma(\alpha) - \beta \sum_{i=1}^n t_i \\
 &= n\alpha \ln(\beta) + (\alpha - 1) \sum_{i=1}^n \ln(t_i) - n(\ln \Gamma(\alpha)) - \beta \sum_{i=1}^n t_i.
 \end{aligned} \tag{5.14}$$

Agora fazendo as derivadas parciais da expressão $\ln(L(\alpha, \beta))$ em relação aos parâmetros α e β e igualando o resultado a zero, se tem os seguintes sistemas de equações:

$$\frac{\partial(\ln(L(\alpha, \beta)))}{\partial\alpha} = n \ln(\beta) + \sum_{i=1}^n \ln(t_i) - \frac{n\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} = 0 \quad (5.15)$$

$$\frac{\partial(\ln(L(\alpha, \beta)))}{\partial\beta} = \frac{n\alpha}{\hat{\beta}} - \sum_{i=1}^n t_i = 0. \quad (5.16)$$

Os estimadores de máxima verossimilhança da distribuição Gama são os valores de $\hat{\alpha}$ e $\hat{\beta}$ que satisfazem as equações (5.15) e (5.16), respectivamente, ou equivalentemente maximizam o logaritmo da função de verossimilhança. Para concluir que as soluções das equações (5.15) e (5.16) são pontos de máximo, tem-se que verificar se o sinal da segunda derivada das mesmas é positivo. A solução destas equações para o conjunto de dados não possui forma fechada, então pode ser obtida por meio de um método numérico iterativo. O programa R tem o pacote Optim que minimiza uma função. Para seu uso, temos que declarar como função objetivo o negativo do logaritmo da função de verossimilhança e especificar qual método utilizar. Neste caso, utilizamos o método de Quase Newton. Pode-se, ver os resultados obtidos na Tabela (5.1) juntamente com as estimações dos parâmetros de outras funções de distribuição de probabilidade e a partir da Figura (5.9) tem-se as comparações de qual modelo univariado obteve o melhor ajuste ao nosso conjunto de dados.

5.3 Distribuição Exponencial Generalizada Univariada

Outra distribuição muito utilizada em dados de sobrevivência é a Exponencial Generalizada estudada por Gupta e Kundu (1999) [15]. Ela é uma alternativa ao uso das distribuições Weibull e Gama. Muitas vezes é a partir da distribuição Exponencial Generalizada Bivariada que temos construções de modelos Weibull bivariados, vamos ver melhor nos próximos capítulos.

No caso univariado, seja T com distribuição $GE(\alpha, \beta)$, onde $\alpha \geq 0$ é chamado parâmetro de forma e $\beta > 0$ de escala. A sua função de densidade de probabilidade é dado por:

$$f(t) = \alpha\beta(1 - \exp\{-t\beta\})^{\alpha-1} \exp\{-t\beta\}, t \geq 0. \quad (5.17)$$

Se $\alpha = 1$ na expressão (5.17), temos a distribuição exponencial com parâmetro β . A função distribuição acumulada é dada por:

$$F(t) = (1 - \exp\{-t\beta\})^\alpha, t \geq 0. \quad (5.18)$$

As funções de sobrevivência e de risco da distribuição Exponencial Generalizada são dadas, respectivamente, por:

$$S(t) = 1 - (1 - \exp\{-t\beta\})^\alpha, \quad (5.19)$$

$$\lambda(t) = \frac{\alpha\beta(1 - \exp\{t\beta\})^{\alpha-1} \exp\{-t\beta\}}{1 - (1 - \exp\{-t\beta\})^\alpha}. \quad (5.20)$$

para $t \geq 0$, $\alpha > 0$, $\beta \geq 0$.

Se tem o interesse em obter as expressões da média e variância da Exponencial Generalizada, deve-se trabalhar com a função gama, pelo fato que:

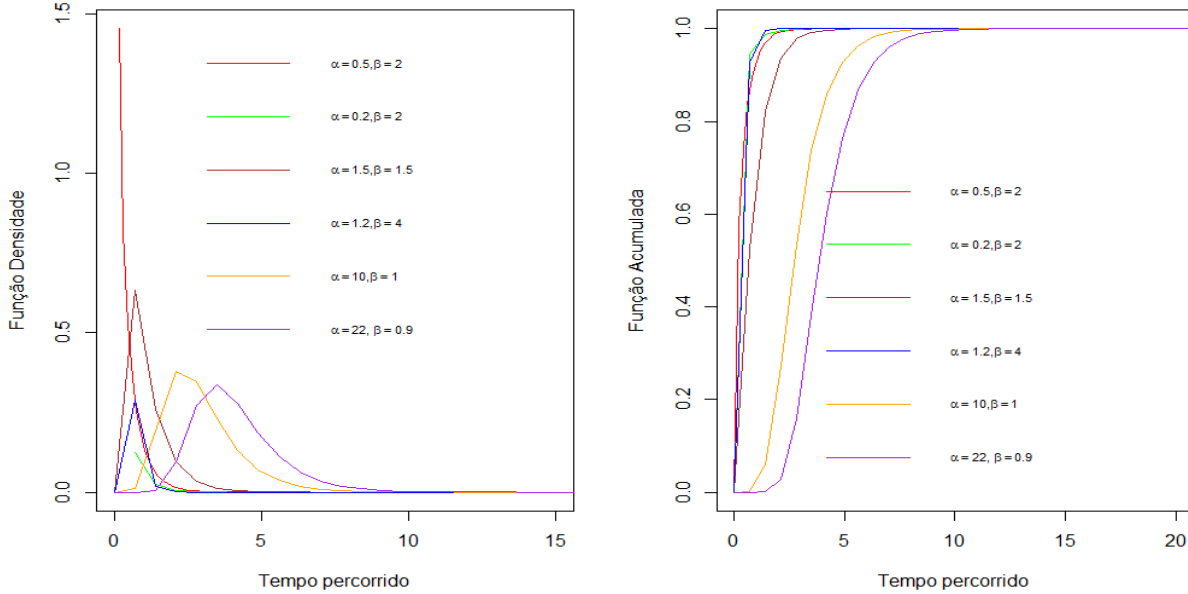
$$E(T) = \frac{1}{\beta}[\psi(\alpha + 1)\psi(1)],$$

$$Var(T) = \frac{1}{\beta^2}[\psi'(1) - \psi'(\alpha + 1)],$$

onde $\psi(\cdot)$ é uma função digamma dada por $\psi(x) = \frac{d}{dx} \ln \Gamma(x)$ e $\Gamma(x)$ é a função gama. (Boleta, 2012) [6]

É possível observar algumas formas da função de densidade de probabilidade e da distribuição acumulada, de uma variável T com distribuição Exponencial Generalizada na Figura (5.3) com diferentes valores de α e β .

Figura 5.3: Formas típicas da função de densidade de probabilidade e da distribuição acumulada, da distribuição Exponencial Generalizada para alguns valores dos parâmetros α e β



A seguir será realizado a estimação dos parâmetros da distribuição Exponencial Generalizada pelo EMV com os dados da Quantidade de Energia solar gerada por placas fotovoltaicas de São José do Rio Preto.

5.3.1 Estimação de Máxima Verossimilhança da distribuição Exponencial Generalizada univariada

Para esta distribuição de probabilidade, substituindo a expressão (5.17) em (2.1), a função de verossimilhança pode ser escrita como:

$$\begin{aligned}
 L(\alpha, \beta) &= \prod_{i=1}^n [f(t_i; \alpha, \beta)] \\
 &= \prod_{i=1}^n [\alpha \beta (1 - \exp\{-t_i \beta\})^{\alpha-1} \exp\{-t_i \beta\}].
 \end{aligned} \tag{5.21}$$

Aplicando o logaritmo na função para facilitar os cálculos temos:

$$\begin{aligned}
 \ln(L(\alpha, \beta)) &= \ln\left[\prod_{i=1}^n [\alpha \beta (1 - \exp\{-t_i \beta\})^{\alpha-1} \exp\{-t_i \beta\}]\right] \\
 &= \sum_{i=1}^n \ln(\alpha) + \sum_{i=1}^n \ln(\beta) + (\alpha - 1) \sum_{i=1}^n \ln(1 - \exp\{-t_i \beta\}) - \sum_{i=1}^n t_i \beta \quad (5.22) \\
 &= n \ln(\alpha) + n \ln(\beta) + (\alpha - 1) \sum_{i=1}^n \ln(1 - \exp\{-t_i \beta\}) - \beta \sum_{i=1}^n t_i.
 \end{aligned}$$

Agora fazendo as derivadas parciais da expressão $\ln(L(\alpha, \beta))$ em relação aos parâmetros α e β e igualando o resultado a zero, se tem os seguintes sistemas de equações:

$$\frac{\partial(\ln(L(\alpha, \beta)))}{\partial\alpha} = \frac{n}{\hat{\alpha}} + \sum_{i=1}^n \ln(1 - \exp\{-t_i\hat{\beta}\}) = 0 \quad (5.23)$$

$$\frac{\partial(\ln(L(\alpha, \beta)))}{\partial\beta} = \frac{n}{\hat{\beta}} + (\hat{\alpha} - 1) \sum_{i=1}^n t_i \frac{\exp\{-t_i\hat{\beta}\}}{1 - \exp\{-t_i\hat{\beta}\}} - \sum_{i=1}^n t_i = 0 \quad (5.24)$$

Os estimadores de máxima verossimilhança da distribuição Exponencial Generalizada são os valores de $\hat{\alpha}$ e $\hat{\beta}$ que satisfazem as equações (5.23) e (5.24) respectivamente ou equivalentemente maximizam o logaritmo da função de verossimilhança. Para concluir que as soluções das equações (5.23) e (5.24) são pontos de máximo, tem-se que verificar se o sinal da segunda derivada das mesmas é positivo. A solução destas equações para o conjunto de dados não possui forma fechada, então pode ser obtida por meio de um método numérico iterativo. O programa R tem o pacote Optim que minimiza uma função. Para o seu uso, temos que declarar como função objetivo o negativo do logaritmo da função de verossimilhança e especificar qual método utilizar. Neste caso, utilizamos o método de Quase Newton. Pode-se ver os resultados obtidos na Tabela (5.1) juntamente com as estimações dos parâmetros de outras funções de distribuição de probabilidade e a partir da Figura (5.9) tem-se as comparações de qual modelo univariado obteve o melhor ajuste ao nosso conjunto de dados.

5.4 Distribuição Normal Univariada

A distribuição Normal em sua forma usual foi introduzida por De Moivre em um artigo em 1733, mais tarde em 1812 seu resultado foi estendido por Laplace, também tivemos Legendre em 1805 e Gauss em 1809 utilizando a distribuição desde os princípios.

No caso univariado seja T com distribuição $N(\mu, \sigma^2)$, onde $-\infty < \mu < \infty$ é o parâmetro de localização e $\sigma > 0$ parâmetro de forma, e a sua função de densidade de probabilidade é dada por:

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right\}, t \in \Re. \quad (5.25)$$

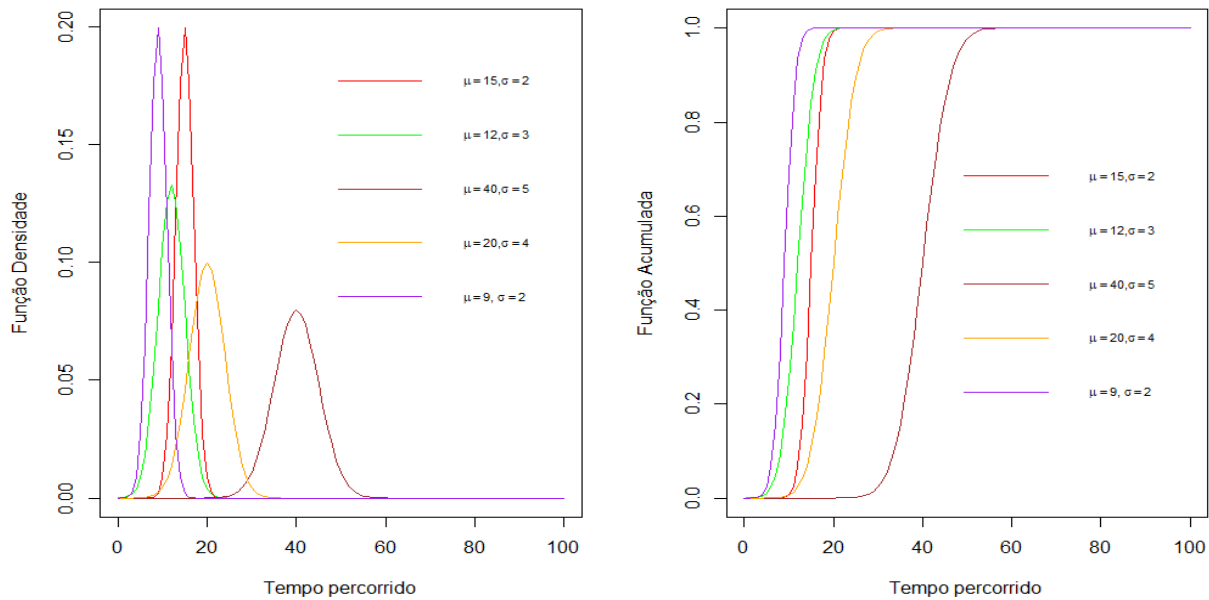
A expressão da média e variância da distribuição Normal é dada por:

$$E(T) = \mu$$

$$\text{Var}(T) = \sigma^2.$$

É possível observar algumas formas da função de densidade de probabilidade e da distribuição acumulada, de uma variável T com distribuição Normal na Figura (5.4) com diferentes valores de μ e σ . O espaço amostral da distribuição Normal é $-\infty < t < \infty$, no entanto, a escolha de alguns parâmetros particulares, obtemos $P(T \leq 0) \cong 0$. Deve-se tomar cuidado se a opção for a escolha desta distribuição. Por esse motivo, não é comum sua utilização para dados de tempos de vida.

Figura 5.4: Formas típica da função de densidade de probabilidade e da distribuição acumulada, da distribuição Normal para alguns valores dos parâmetros μ e σ



A seguir será realizado a estimação dos parâmetros da distribuição Normal pelo EMV com os dados da Quantidade de Energia solar gerada por placas fotovoltaicas de São José do Rio Preto.

5.4.1 Estimação de Máxima Verossimilhança da distribuição Normal univariada

Para esta distribuição de probabilidade, substituindo a expressão (5.25) em (2.1), a função de verossimilhança pode ser escrita como:.

$$\begin{aligned}
L(\mu, \sigma^2) &= \prod_{i=1}^n f(t_i | \mu, \sigma^2) \\
&= \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2} \left(\frac{t_i - \mu}{\sigma}\right)^2\right\} \right].
\end{aligned} \tag{5.26}$$

Aplicando o logaritmo na função para facilitar os cálculos temos:

$$\begin{aligned}
\ln(L(\mu, \sigma^2)) &= \ln \left[\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2} \left(\frac{t_i - \mu}{\sigma}\right)^2\right\} \right] \\
&= -\sum_{i=1}^n \ln(2\pi)^{\frac{1}{2}} - \sum_{i=1}^n \ln(\sigma^2)^{\frac{1}{2}} - \frac{1}{2} \sum_{i=1}^n \left(\frac{t_i - \mu}{\sigma}\right)^2 \\
&= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2} \sum_{i=1}^n \frac{(t_i - \mu)^2}{\sigma^2}.
\end{aligned} \tag{5.27}$$

Agora fazendo as derivadas parciais da expressão $\ln(L(\mu, \sigma^2))$ em relação aos parâmetros μ e σ^2 e igualando o resultado a zero, se tem os seguintes sistemas de equações:

$$\begin{aligned}
\frac{\partial(\ln(L(\mu, \sigma^2)))}{\partial\mu} &= \sum_{i=1}^n \frac{t_i - \hat{\mu}}{\sigma^2} = 0 \\
&\Leftrightarrow \sum_{i=1}^n (t_i - \hat{\mu}) = 0 \\
n\hat{\mu} &= \sum_{i=1}^n t_i \\
&\Leftrightarrow \hat{\mu} = \bar{t}
\end{aligned} \tag{5.28}$$

$$\begin{aligned}
\frac{\partial(\ln(L(\mu, \sigma^2)))}{\partial\sigma^2} &= -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2} \sum_{i=1}^n \left(\frac{t_i - \hat{\mu}}{\hat{\sigma}^2}\right)^2 = 0 \\
&\Leftrightarrow -n + \sum_{i=1}^n \frac{(t_i - \hat{\mu})^2}{\hat{\sigma}^2} = 0 \\
&\Leftrightarrow \hat{\sigma}^2 = \frac{(n-1)s^2}{n},
\end{aligned} \tag{5.29}$$

onde $s^2 = \sum_{i=1}^n \frac{(t_i - \hat{\mu})^2}{n-1}$. Os estimadores de máxima verossimilhança da distribuição Normal são os valores de $\hat{\mu}$ e $\hat{\sigma}^2$ que satisfazem as equações (5.28) e (5.29), respectivamente, ou equivalentemente maximizam o logaritmo da função de verossimilhança.

5.5 Distribuição Weibull Exponenciada

A distribuição Weibull Exponenciada (EW) foi introduzida por Mudholkar e Srivastava (1993) [22] ela possui um parâmetro de escala e dois parâmetros de forma. Na distribuição EW à diversas propriedades que são semelhantes às da família Weibull ou gama, assim, ela pode ser usada como uma alternativa à uma distribuição Weibull ou gama. (Woo. et al, 2003) [36]

Seja T com distribuição $EW(\alpha, \beta, a)$, onde $\beta > 0$ é o parâmetro de escala e $\alpha > 0$ e $a > 0$ os parâmetros de forma, sua função de densidade de probabilidade é dada por:

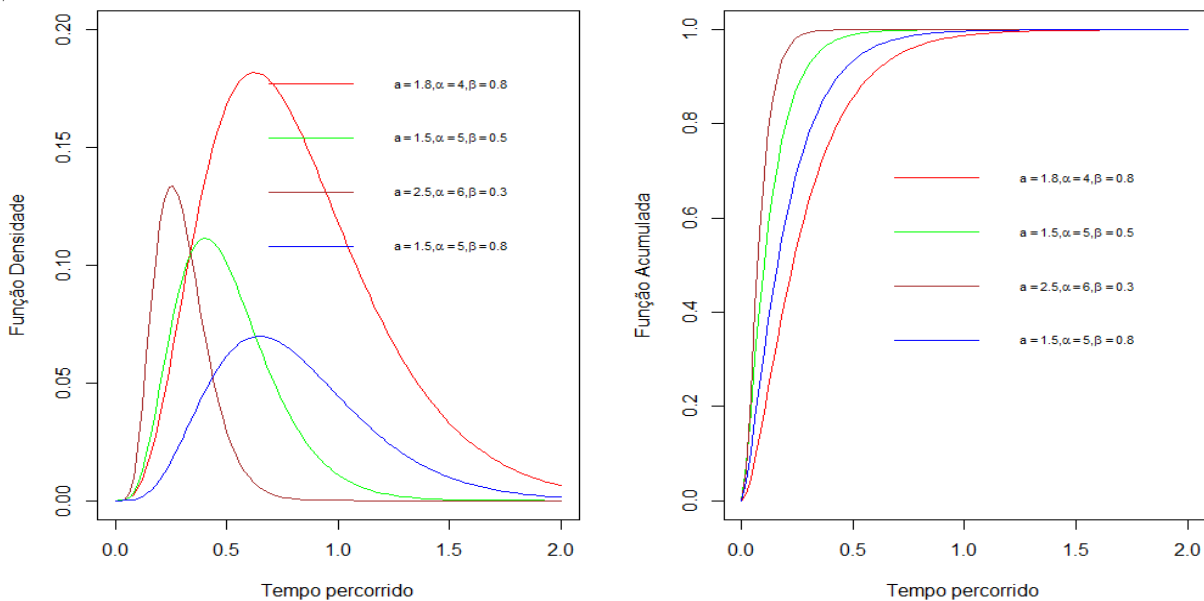
$$f(t) = a \frac{\alpha}{\beta^\alpha} t^{\alpha-1} \exp \left\{ -\frac{t}{\beta} \right\}^\alpha \left[1 - \exp \left\{ -\frac{t}{\beta} \right\}^\alpha \right]^{a-1}, t > 0. \quad (5.30)$$

A sua Função distribuição acumulada é dada por:

$$F(t) = \left[1 - \exp \left\{ -\frac{t}{\beta} \right\}^\alpha \right]^a, t > 0. \quad (5.31)$$

É possível observar algumas formas da função de densidade de probabilidade e da distribuição acumulada, de uma variável T com distribuição EW na Figura (5.5) com diferentes valores de α , β e a

Figura 5.5: Formas típica da função de densidade de probabilidade e da distribuição acumulada, da distribuição Weibull Exponenciada para alguns valores dos parâmetros α , β e a



A seguir será realizado a estimação dos parâmetros da distribuição EW pelo EMV com os dados da Quantidade de Energia solar gerada por placas fotovoltaicas de São José do Rio Preto.

5.5.1 Estimação de Máxima Verossimilhança da distribuição Weibull Exponenciada

O método de EMV como já foi introduzido na seção (2.2) trata do problema de estimação dos parâmetros a partir de resultados de uma amostra, escolhendo os estimadores que melhor explique a amostra observada. É possível de se obter esse resultado para dados de energia solar resultante da distribuição EW substituindo a expressão (5.30) em (2.1).

$$\begin{aligned} L(\alpha, \beta, a) &= \prod_{i=1}^n [f(t_i; \alpha, \beta, a)] \\ &= \prod_{i=1}^n \left[a \frac{\alpha}{\beta^\alpha} t^{\alpha-1} \exp \left\{ -\frac{t}{\beta} \right\}^\alpha \left[1 - \exp \left\{ -\frac{t}{\beta} \right\}^\alpha \right]^{a-1} \right]. \end{aligned} \quad (5.32)$$

Aplicando o logaritmo na função para facilitar os cálculos temos:

$$\begin{aligned} \ln(L(\alpha, \beta, a)) &= \ln \left[\prod_{i=1}^n \left[a \frac{\alpha}{\beta^\alpha} t^{\alpha-1} \exp \left\{ -\frac{t}{\beta} \right\}^\alpha \left[1 - \exp \left\{ -\frac{t}{\beta} \right\}^\alpha \right]^{a-1} \right] \right] \\ &= n \ln(a) + n \ln(\alpha) - n\alpha \ln(\beta) + (\alpha - 1) \sum_{i=1}^n \ln(t_i) \\ &\quad - \sum_{i=1}^n \left(\frac{t_i}{\beta} \right)^\alpha + (a - 1) \sum_{i=1}^n \ln \left[1 - \exp \left\{ -\frac{t_i}{\beta} \right\}^\alpha \right]. \end{aligned} \quad (5.33)$$

Para encontrar os devidos resultados dos estimadores será usado o programa R que tem o pacote Optim que minimiza uma função, então temos que declarar como função objetivo o negativo do logaritmo da função de verossimilhança e especificar qual método utilizar. Neste caso, utilizamos o método de Quase Newton. Pode-se ver os resultados obtidos na Tabela (5.1) juntamente com as estimações dos parâmetros de outras funções de distribuição de probabilidade e a partir da Figura (5.9) tem-se as comparações de qual modelo univariado obteve o melhor ajuste ao nosso conjunto de dados.

5.6 Generalizações da Distribuição Kumaraswamy

Além das diversas distribuições usuais que podem ser vista com grande frequência em sala de aula ou em literaturas, existem novas distribuições que são generalizadas de outras distribuições que também nos apresentam resultados muito interessante ao se ajustar a dados reais. Muitos pesquisadores nos oferecem caminhos diferentes para que possamos ainda mais aperfeiçoar o estudo aos dados. Uma delas é generalizar distribuição a Kumaraswamy proposta por Kumaraswamy (1980) [17] com distribuições já conhecida por muitos como por exemplo a Exponencial e a Weibull.

Kumaraswamy introduziu uma distribuição de probabilidade para processos aleatórios de dupla ligação que tem aplicações com hidrologia e áreas relacionadas. Como já foi dito, utilizando uma distribuição contínua é capaz de descrever uma nova família de

distribuições generalizadas (denotadas com o prefixo "Kw") para estender as distribuições como a Weibull, Exponencial, entre várias distribuições conhecidas (Das, 2015) [11]. Sua função de distribuição acumulada com os parâmetros $Kw(\alpha, \beta)$ é dada por:

$$F_K(t) = 1 - (1 - t^\alpha)^\beta, t \in (0, 1), \quad (5.34)$$

onde $\alpha > 0$ e $\beta > 0$ que são parâmetros de forma. Sua função densidade de probabilidade é dada por:

$$f_K(t) = \alpha\beta t^{\alpha-1}(1 - t^\alpha)^{\beta-1}, t \in (0, 1). \quad (5.35)$$

A distribuição K como é chamada é muito usada em aplicações hidrológicas e muitos autores dizem que ela é uma boa alternativa em relação a distribuição Beta. As seguir serão apresentadas algumas generalizações da distribuição K e aplicá-las referente aos dados da presente pesquisa

5.6.1 Distribuição Weibull Kumaraswamy

A partir da apresentação da distribuição de Kumaraswamy, temos as suas generalizações, a primeira que será mostrada é a Weibull Kumaraswamy (KwW), onde Cordeiro e Castro (2009) [10] apresentam sua construção.

Seja uma variável aleatória T com distribuição $KwW(\alpha, \beta, a, b)$, onde $\alpha > 0$, $a > 0$ e $b > 0$ são parâmetros de forma e $\beta > 0$ o parâmetro de escala. Sua função de densidade de probabilidade é dada por:

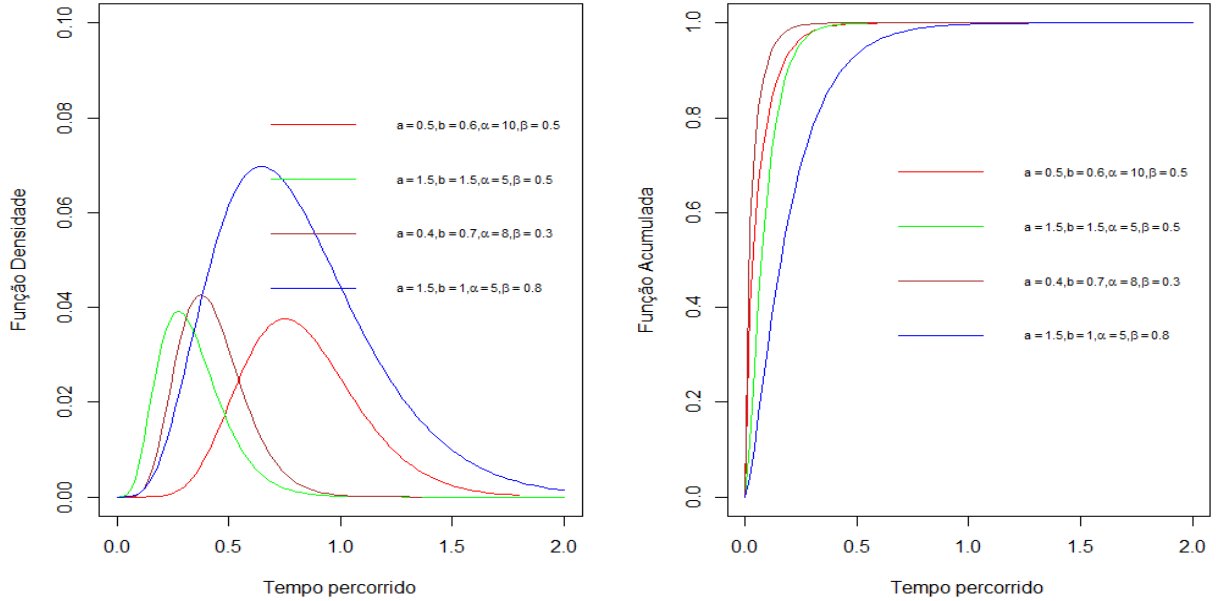
$$f(t) = ab \frac{\alpha}{\beta^\alpha} t^{\alpha-1} \exp\left\{-\frac{t}{\beta}\right\}^\alpha \left[1 - \exp\left\{-\frac{t}{\beta}\right\}^\alpha\right]^{a-1} \left\{1 - \left[1 - \exp\left\{-\frac{t}{\beta}\right\}^\alpha\right]^a\right\}^{b-1}, t > 0. \quad (5.36)$$

A sua função distribuição acumulada é dada por:

$$F(t) = 1 - \left\{1 - \left[1 - \exp\left\{-\frac{t}{\beta}\right\}^\alpha\right]^a\right\}^b, t > 0. \quad (5.37)$$

É possível observar algumas formas da função de densidade de probabilidade e da distribuição acumulada, de uma variável T com distribuição KwW na Figura (5.6) com diferentes valores de α , β , a e b .

Figura 5.6: Formas típica da função de densidade de probabilidade e da distribuição acumulada, da distribuição Weibull Kumaraswamy para alguns valores dos parâmetros α , β , a e b



A seguir será realizado a estimação dos parâmetros da distribuição KwW pelo EMV com os dados da Quantidade de Energia solar gerada por placas fotovoltaicas de São José do Rio Preto.

5.6.1.1 Estimação de Máxima Verossimilhança da Distribuição Weibull Kumaraswamy

Para esta distribuição de probabilidade, substituindo a expressão (5.36) em (2.1), a função de verossimilhança pode ser escrita como:

$$\begin{aligned}
 L(\alpha, \beta, a, b) &= \prod_{i=1}^n [f(t_i; \alpha, \beta, a, b)] \\
 &= \prod_{i=1}^n \left[ab \frac{\alpha}{\beta^\alpha} t^{\alpha-1} \exp \left\{ -\frac{t}{\beta} \right\}^\alpha \left[1 - \exp \left\{ -\frac{t}{\beta} \right\}^\alpha \right]^{a-1} \left\{ 1 - \left[1 - \exp \left\{ -\frac{t}{\beta} \right\}^\alpha \right]^a \right\}^{b-1} \right].
 \end{aligned}
 \tag{5.38}$$

Aplicando o logaritmo na função para facilitar os cálculos temos:

$$\begin{aligned}
\ln(L(\sigma, \alpha, a, b)) &= \ln \left[\prod_{i=1}^n \left[ab \frac{\alpha}{\beta^\alpha} t^{\alpha-1} \exp \left\{ -\frac{t}{\beta} \right\}^\alpha \left[1 - \exp \left\{ -\frac{t}{\beta} \right\}^\alpha \right]^{a-1} \left\{ 1 - \left[1 - \exp \left\{ -\frac{t}{\beta} \right\}^\alpha \right]^a \right\}^{b-1} \right] \right] \\
&= n \ln(a) + n \ln(b) + n \ln(\alpha) - n\alpha \ln(\beta) + (\alpha - 1) \sum_{i=1}^n \ln(t_i) \\
&\quad - \sum_{i=1}^n \left(\frac{t_i}{\beta} \right)^\alpha + (a - 1) \sum_{i=1}^n \ln \left[1 - \exp \left\{ -\frac{t_i}{\beta} \right\}^\alpha \right] + (b - 1) \sum_{i=1}^n \ln \left\{ 1 - \left[1 - \exp \left\{ -\frac{t_i}{\beta} \right\}^\alpha \right]^a \right\}
\end{aligned} \tag{5.39}$$

Para encontrar os devidos resultados dos estimadores será usado o programa R que tem o pacote Optim que minimiza uma função então temos que declarar como função objetivo o negativo do logaritmo da função de verossimilhança e especificar qual método utilizar. Neste caso utilizamos o método de Quase Newton. Pode-se ver os resultados obtidos na Tabela (5.1) juntamente com as estimações dos parâmetros de outras funções de distribuição de probabilidade e a partir da Figura (5.9) tem-se as comparações de qual modelo univariado obteve o melhor ajuste ao nosso conjunto de dados.

5.6.2 Distribuição Log-exponenciada Kumaraswamy

A próxima distribuição apresentada é a log-exponenciada Kumaraswamy (log-EK), onde Lemonte, et al. (2013) [18] apresenta sua construção. Antes de apresentar a distribuição log-EK é preciso ter a construção da exponenciada Kumaraswamy(EK) e segundo (Lemonte) a construção da distribuição EK se baseia na observação de que elevando uma função de densidade acumulada $G(x)$ para um poder arbitrário $\gamma > 0$ tem-se uma nova função de densidade cumulada, onde $F(x) = G(x)^\gamma$ nos sugerindo como um parâmetro adicional, onde γ é o parâmetro de assimetria, curtose e cauda da distribuição F. Na construção da exponenciada $G(x)$ se caracteriza como a distribuição da linha de base e $F(x)$ pode ser referido como a distribuição G exponenciada.

Assim seja X uma variável aleatória com distribuição $EK(\alpha, \beta, \gamma)$, onde $\alpha \geq 0$, $\beta \geq 0$ e $\gamma \geq 0$, sua função de densidade de probabilidade é dada por:

$$f_{EK}(x) = \alpha\beta\gamma x^{\alpha-1}(1-x^\alpha)^{\beta-1}[1-(1-x^\alpha)^\beta]^{\gamma-1}, x \in (0, 1). \tag{5.40}$$

É possível observar que se $\gamma = 1$ temos a distribuição Kw com parâmetros α e β . A função distribuição acumulada de EK é dada por:

$$F_{EK}(x) = [1 - (1 - x^\alpha)^\beta]^\gamma, x \in (0, 1). \tag{5.41}$$

Percebe-se que a distribuição (5.40) se restringe em valores no intervalo de (0, 1), assim não se ajusta aos dados da pesquisa, portanto tem-se logo em seguida a transformação para transformação para log-EK.

A distribuição log-EK é bem útil para modelar dados da vida útil. Seja uma variável X com distribuição $EK(\alpha, \beta, \gamma)$, pode-se definir a distribuição log-EK padrão fazendo

$Y = -\log(1 - X)$. Pode-se também adicionar um parâmetro de escala $\sigma > 0$ utilizando a transformação de escala $T = \sigma Y$, assim tem-se que T é uma variável com distribuição $\log\text{-EK}(\sigma, \alpha, \beta, \gamma)$ Lemonte, et al. (2013) [18]. Sua função de densidade de probabilidade é dada por:

$$f_{\log\text{-EK}}(t) = \frac{\alpha\beta\gamma \exp\{-\frac{t}{\sigma}\} (1 - \exp\{-\frac{t}{\sigma}\})^{\alpha-1} [1 - (1 - \exp\{-\frac{t}{\sigma}\})^{\alpha}]^{\beta-1}}{\sigma \{1 - [1 - (1 - \exp\{-\frac{t}{\sigma}\})^{\alpha}]^{\beta}\}^{1-\gamma}}, t > 0. \quad (5.42)$$

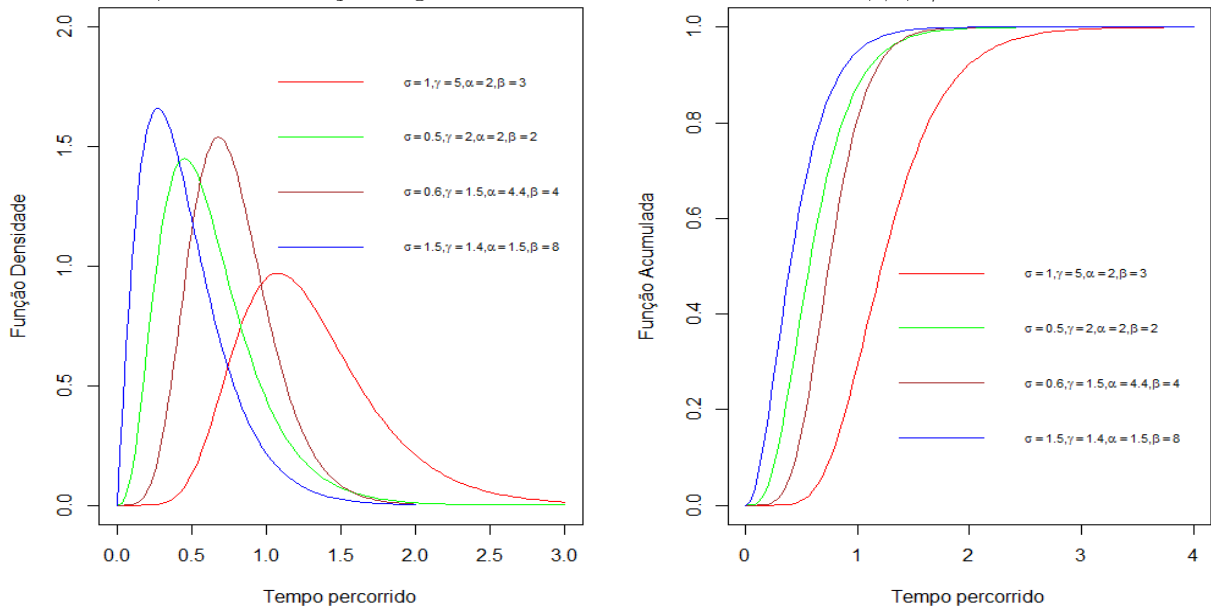
A função distribuição acumulada é dada por:

$$F_{\log\text{-EK}}(t) = \left\{ 1 - \left[1 - \left(1 - \exp\left\{-\frac{t}{\sigma}\right\}\right)^{\alpha} \right]^{\beta} \right\}^{\gamma}, t > 0. \quad (5.43)$$

A distribuição $\log\text{-EK}$ pode ser reduzida a uma distribuição GE, quando $\gamma = 1$ tem-se uma $GE(\sigma^{-1}, \alpha\beta)$, quando $\beta = 1$ tem-se $GE(\sigma^{-1}, \gamma\alpha)$ e quando $\beta = \gamma = 1$ tem-se $GE(\sigma^{-1}, \alpha)$.

É possível observar algumas formas da função de densidade de probabilidade e da distribuição acumulada, de uma variável T com distribuição $\log\text{-EK}$ na Figura (5.7) com diferentes valores de α , β , γ e σ .

Figura 5.7: Formas típica da função de densidade de probabilidade e da distribuição acumulada, da distribuição $\log\text{-EK}$ com diferentes valores de α , β , γ e σ



A seguir será realizado a estimação dos parâmetros da distribuição $\log\text{-EK}$ pelo EMV com os dados da Quantidade de Energia solar gerada por placas fotovoltaicas de São José do Rio Preto.

5.6.2.1 Estimação de Máxima Verossimilhança da Distribuição Log-exponenciada de Kumaraswamy

O método de EMV como já foi introduzido na seção(2.2) trata do problema de estimação dos parâmetros a partir de resultados de uma amostra, escolhendo os estimadores que melhor explique a amostra observada. É possível de se obter esse resultado para dados de energia solar resultante da distribuição log-EK substituindo a expressão (5.42) em (2.1).

$$\begin{aligned} L(\sigma, \alpha, \beta, \gamma) &= \prod_{i=1}^n [f(t_i; \sigma, \alpha, \beta, \gamma)] \\ &= \prod_{i=1}^n \left[\frac{\alpha\beta\gamma \exp\left\{-\frac{t_i}{\sigma}\right\} (1 - \exp\left\{-\frac{t_i}{\sigma}\right\})^{\alpha-1} [1 - (1 - \exp\left\{-\frac{t_i}{\sigma}\right\})^\alpha]^{\beta-1}}{\sigma \{1 - [1 - (1 - \exp\left\{-\frac{t_i}{\sigma}\right\})^\alpha]^\beta\}^{1-\gamma}} \right]. \end{aligned} \quad (5.44)$$

Aplicando o logaritmo na função para facilitar os cálculos temos:

$$\begin{aligned} \ln(L(\sigma, \alpha, \beta, \gamma)) &= \ln \left[\prod_{i=1}^n \left[\frac{\alpha\beta\gamma \exp\left\{-\frac{t_i}{\sigma}\right\} (1 - \exp\left\{-\frac{t_i}{\sigma}\right\})^{\alpha-1} [1 - (1 - \exp\left\{-\frac{t_i}{\sigma}\right\})^\alpha]^{\beta-1}}{\sigma \{1 - [1 - (1 - \exp\left\{-\frac{t_i}{\sigma}\right\})^\alpha]^\beta\}^{1-\gamma}} \right] \right] \\ &= n \ln(\alpha) + n \ln(\beta) + n \ln(\gamma) - n \ln(\sigma) - \frac{1}{\sigma} \sum_{i=1}^n t_i + (\alpha - 1) \sum_{i=1}^n \ln \left(1 - \exp \left\{ -\frac{t_i}{\sigma} \right\} \right) \\ &\quad + (\beta - 1) \sum_{i=1}^n \ln \left(1 - \left(1 - \exp \left\{ -\frac{t_i}{\sigma} \right\} \right)^\alpha \right) \\ &\quad + (\gamma - 1) \sum_{i=1}^n \ln \left\{ 1 - \left(1 - \left(1 - \exp \left\{ -\frac{t_i}{\sigma} \right\} \right)^\alpha \right)^\beta \right\}. \end{aligned} \quad (5.45)$$

Para encontrar os devidos resultados dos estimadores será usado o programa R que tem o pacote Optim que minimiza uma função então temos que declarar como função objetivo o negativo do logaritmo da função de verossimilhança e especificar qual método utilizar. Neste caso, utilizamos o método de Quase Newton. Pode-se ver os resultados obtidos na Tabela (5.1) juntamente com as estimações dos parâmetros de outras funções de distribuição de probabilidade e a partir da Figura (5.9) tem-se as comparações de qual modelo univariado obteve o melhor ajuste ao nosso conjunto de dados.

5.6.3 Distribuição Weibull Exponenciada Kumaraswamy

Em Silva, et al. (2019) [34] é apresentada a Distribuição Weibull exponenciada Kumaraswamy (EKwW) com a proposta de melhor o ajuste que a distribuição Weibull e outras distribuição generalizadas dela. Silva, et al. (2019) diz que a partir da função distribuição acumulada da Weibull dada na expressão (5.2) uma variável aleatória T com distribuição $EKwW(a, b, c, \alpha, \beta)$ onde $\alpha > 0$, $a > 0$, $b > 0$ e $c > 0$ são parâmetros de forma e $\beta > 0$ o parâmetro de escala, pode ser apresentada como:

$$F_{EKwW}(t) = \left\{ 1 - \left[1 - \left(1 - \exp \left\{ -\frac{t}{\beta} \right\}^\alpha \right)^a \right]^b \right\}^c, t > 0. \quad (5.46)$$

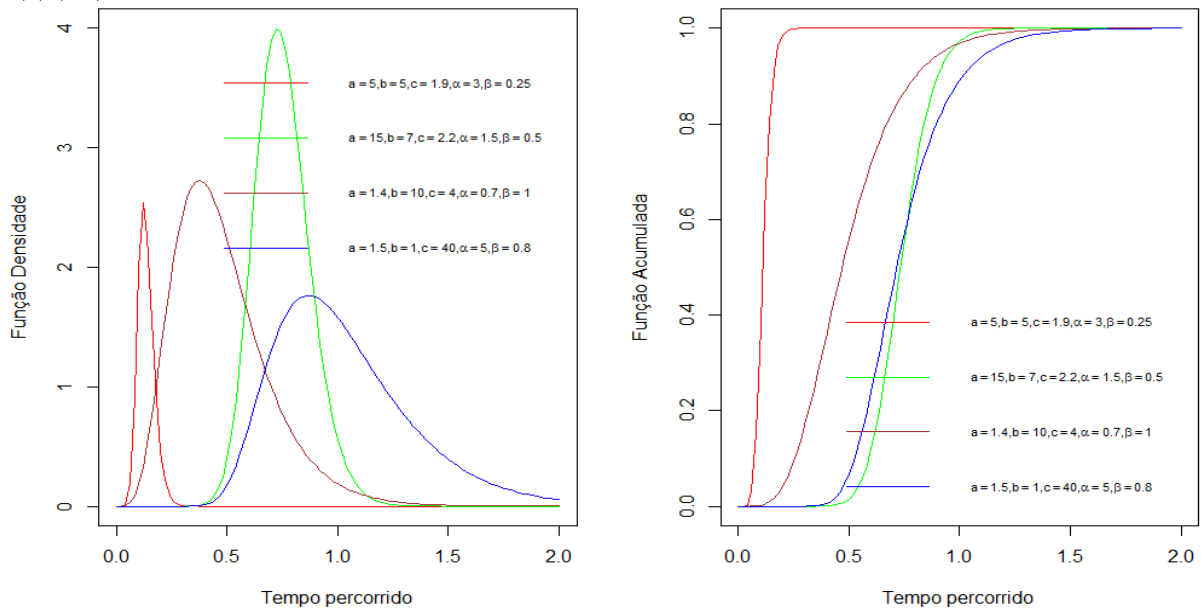
Assim a função densidade de probabilidade pode ser expressa por:

$$f_{EKwW}(t) = abc \frac{\alpha}{\beta^\alpha} t^{\alpha-1} \exp \left\{ -\frac{t}{\beta} \right\}^\alpha \left[1 - \exp \left\{ -\frac{t}{\beta} \right\}^\alpha \right]^{a-1} \left\{ 1 - \left[1 - \exp \left\{ -\frac{t}{\beta} \right\}^\alpha \right]^a \right\}^{b-1} \\ \times \left\{ 1 - \left[1 - \left(1 - \exp \left\{ -\frac{t}{\beta} \right\}^\alpha \right)^a \right]^b \right\}^{c-1}, t > 0. \quad (5.47)$$

A distribuição EKwW pode ser reduzida a uma distribuição KwW, quando $c = 1$. Ela também pode ser reduzida a uma distribuição EW quando $b = c = 1$

É possível observar algumas formas da função de densidade de probabilidade e da distribuição acumulada, de uma variável T com distribuição EKwW na Figura (5.8) com diferentes valores de α , β , a , b e c

Figura 5.8: Forma típica da função de densidade de probabilidade e da distribuição acumulada, da distribuição Weibull exponenciada Kumaraswamy com diferentes valores de α , β , a , b e c



A seguir será realizado a estimação dos parâmetros da distribuição EKwW pelo EMV com os dados da Quantidade de Energia solar gerada por placas fotovoltaicas de São José do Rio Preto.

5.6.3.1 Estimação de Máxima Verossimilhança da Distribuição Weibull Exponenciada Kumaraswamy

O método de EMV como já foi introduzido na seção(2.2) trata do problema de estimação dos parâmetros a partir de resultados de uma amostra, escolhendo os estimadores que melhor explique a amostra observada. É possível de se obter esse resultado para dados de energia solar resultante da distribuição EKwW substituindo a expressão (5.47) em (2.1).

$$\begin{aligned}
 L(\alpha, \beta, a, b, c) &= \prod_{i=1}^n [f(t_i; \alpha, \beta, a, b, c)] \\
 &= \prod_{i=1}^n [abc \frac{\alpha}{\beta^\alpha} t_i^{\alpha-1} \exp\left\{-\frac{t_i}{\beta}\right\}^\alpha \left[1 - \exp\left\{-\frac{t_i}{\beta}\right\}^\alpha\right]^{a-1} \left\{1 - \left[1 - \exp\left\{-\frac{t_i}{\beta}\right\}^\alpha\right]^a\right\}^{b-1} \\
 &\quad \times \left\{1 - \left[1 - \left(1 - \exp\left\{-\frac{t_i}{\beta}\right\}^\alpha\right)^a\right]^b\right\}^{c-1}].
 \end{aligned} \tag{5.48}$$

Aplicando o logaritmo na função para facilitar os cálculos temos:

$$\begin{aligned}
 \ln(L(\sigma, \alpha, a, b, c)) &= \ln\left[\prod_{i=1}^n [abc \frac{\alpha}{\beta^\alpha} t_i^{\alpha-1} \exp\left\{-\frac{t_i}{\beta}\right\}^\alpha [1 - \exp\left\{-\frac{t_i}{\beta}\right\}^\alpha]^{a-1} \{1 - [1 - \exp\left\{-\frac{t_i}{\beta}\right\}^\alpha]^a\}^{b-1} \right. \\
 &\quad \left. \times \{1 - [1 - (1 - \exp\left\{-\frac{t_i}{\beta}\right\}^\alpha)^a]^b\}^{c-1}\right] \\
 &= n \ln(a) + n \ln(b) + n \ln(c) + n \ln(\alpha) - n\alpha \ln(\beta) + (\alpha - 1) \sum_{i=1}^n \ln(t_i) \\
 &\quad - \sum_{i=1}^n \left(\frac{t_i}{\beta}\right)^\alpha + (a - 1) \sum_{i=1}^n \ln\left[1 - \exp\left\{-\frac{t_i}{\beta}\right\}^\alpha\right] + (b - 1) \sum_{i=1}^n \ln\left\{1 - \left[1 - \exp\left\{-\frac{t_i}{\beta}\right\}^\alpha\right]^a\right\} \\
 &\quad + (c - 1) \sum_{i=1}^n \ln\left\{1 - \left[1 - \left(1 - \exp\left\{-\frac{t_i}{\beta}\right\}^\alpha\right)^a\right]^b\right\}.
 \end{aligned} \tag{5.49}$$

Para encontrar os devidos resultados dos estimadores será usado o programa R que tem o pacote Optim que minimiza uma função então temos que declarar como função objetivo o negativo do logaritmo da função de verossimilhança e especificar qual método utilizar. Neste caso, utilizamos o método de Quase Newton. Pode-se ver os resultados obtidos na Tabela (5.1) juntamente com as estimações dos parâmetros de outras funções de distribuição de probabilidade e a partir da Figura (5.9) tem-se as comparações de qual modelo univariado obteve o melhor ajuste ao nosso conjunto de dados.

5.7 Resultado e discussão dos modelos univariados

Como já foi visto anteriormente o programa R possui o pacote Optim que pode ser utilizado para minimizar uma função. Para sua utilização devemos declarar como função

objetivo, o negativo do logaritmo da função de verossimilhança e especificar qual método a ser utilizado. O método numérico foi o de Quase Newton. Na Tabela (5.1) tem-se os resultados dos valores dos parâmetros estimados para os modelos univariados com seus respectivos Erros Padrões nos indicando a precisão do estimador:

Tabela 5.1: Parâmetros estimados pelo EMV para cada distribuição univariada

Distribuição	Parâmetros				
Weibull	$\hat{\alpha} = 5.394$	$\hat{\beta} = 9.081$	—	—	—
(Erro Padrão)	(0.186)	(0.076)	—	—	—
Gama	$\hat{\alpha} = 15.340$	$\hat{\beta} = 1.830$	—	—	—
(Erro Padrão)	(0.934)	(0.113)	—	—	—
EG	$\hat{\alpha} = 22.487$	$\hat{\beta} = 0.429$	—	—	—
(Erro Padrão)	(2.130)	(0.013)	—	—	—
Normal	$\hat{\mu} = 8.381$	$\hat{\sigma} = 1.861$	—	—	—
(Erro Padrão)	(0.081)	(0.057)	—	—	—
EW	$\hat{\alpha} = 2.856$	$\hat{\beta} = 4.240$	$\hat{a} = 77.876$	—	—
(Erro Padrão)	(5.873)	(0.078)	(0.097)	—	—
KwW	$\hat{\alpha} = 1.130$	$\hat{\beta} = 7.241$	$\hat{a} = 11.040$	$\hat{b} = 22.678$	—
(Erro Padrão)	(0.334)	(0.847)	(0.011)	(0.117)	—
log-EK	$\hat{\alpha} = 15.871$	$\hat{\beta} = 107.912$	$\hat{\gamma} = 0.474$	$\hat{\sigma} = 7.298$	—
(Erro Padrão)	(0.058)	(0.027)	(0.079)	(8.430)	—
EKwW	$\hat{\alpha} = 1.270$	$\hat{\beta} = 6.424$	$\hat{a} = 23.304$	$\hat{b} = 30.811$	$\hat{c} = 0.424$
(Erro Padrão)	(0.005)	(0.010)	(0.019)	(0.005)	(0.005)

Com os valores dos estimadores de cada distribuição é possível então realizar os ajuste e verificar qual distribuição se ajustou melhor aos dados de energia solar gerada por placas fotovoltaicas, na cidade de São José do Rio Preto-SP.

As análises gráficas propostas para verificar o melhor ajuste foram as curva da densidade e distribuição acumulada de cada distribuição univariada em relação a curva empírica. Nas Figura (5.9) até (5.12) tem-se os quatro gráficos das curvas das funções de densidade *versus* suas empíricas, onde já é possível de se observar que a distribuição Weibull e a Weibull Exponenciada Kumaraswamy obtiveram um melhor ajuste comparado com as demais distribuições.

Figura 5.9: Curva das funções de densidade da Weibull e Gama versus a curva de suas funções de distribuições Empírica para seus respectivos parâmetros estimados

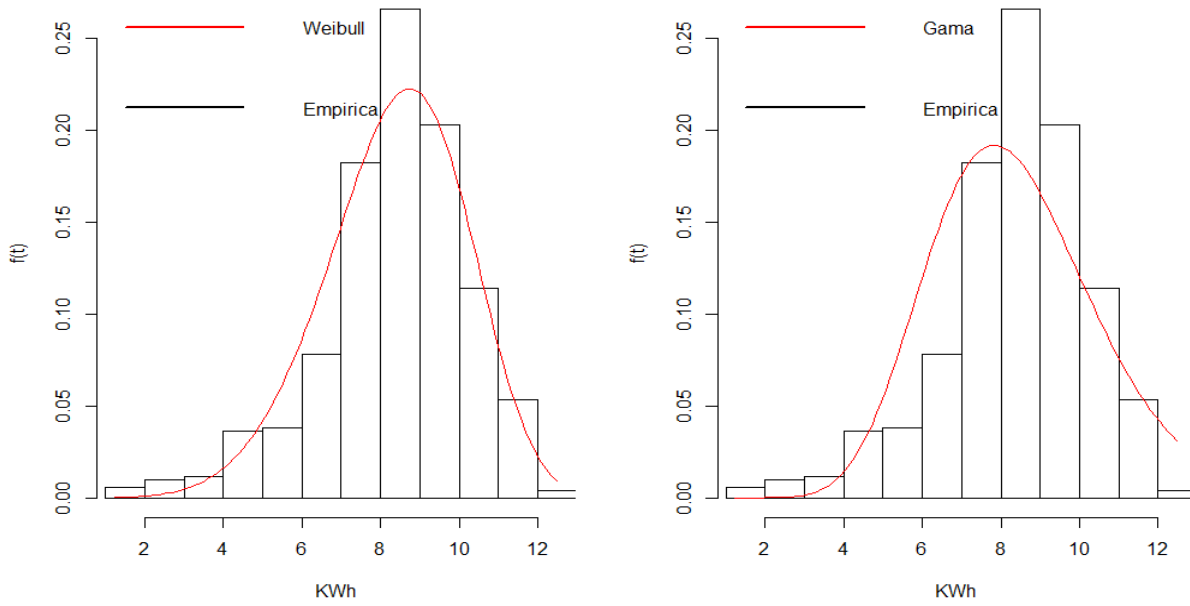


Figura 5.10: Curva das funções de densidade da Exponencial Generalizada e Normal versus a curva de suas funções de distribuições Empírica para seus respectivos parâmetros estimados

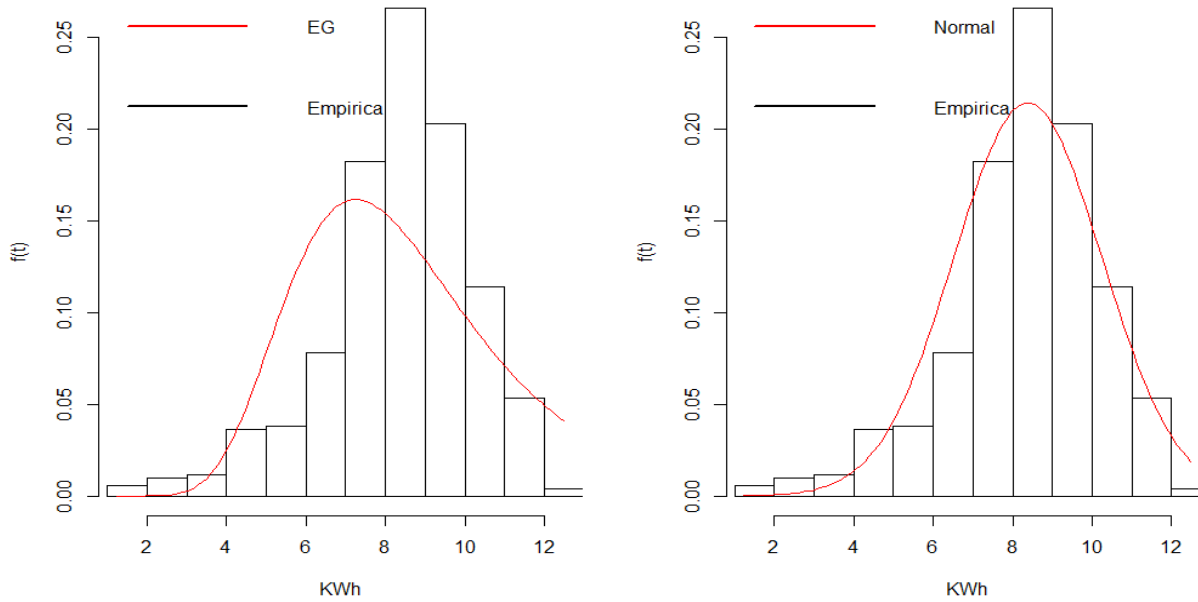


Figura 5.11: Curva das funções de densidade da Weibull Exponenciada e Weibull Kumaraswamy versus a curva de suas funções de distribuições Empírica para seus respectivos parâmetros estimados

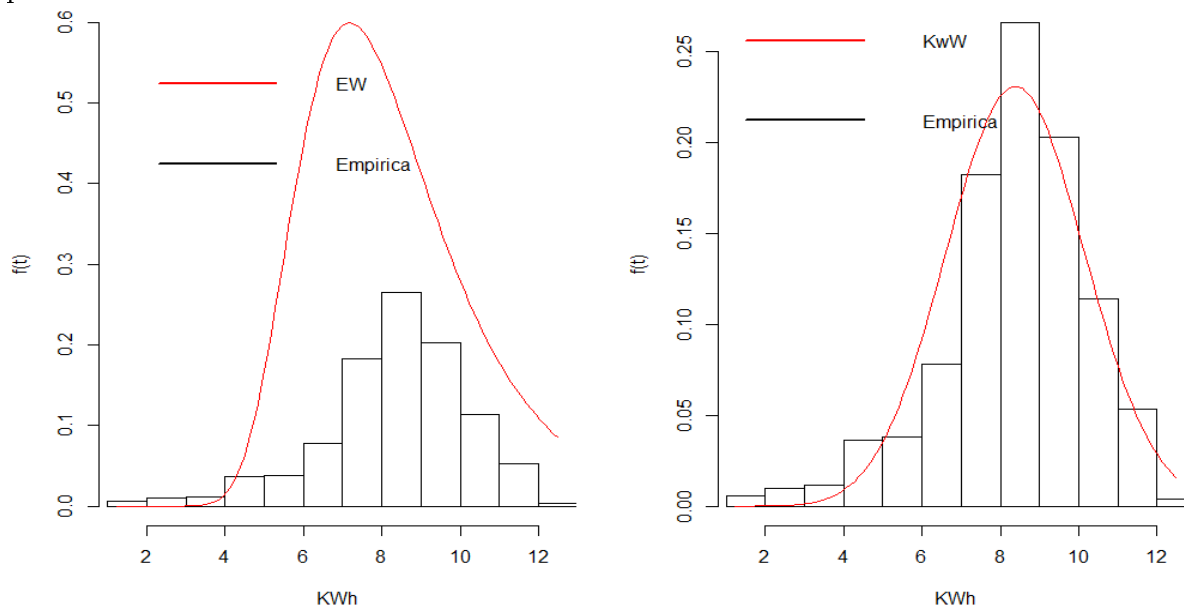
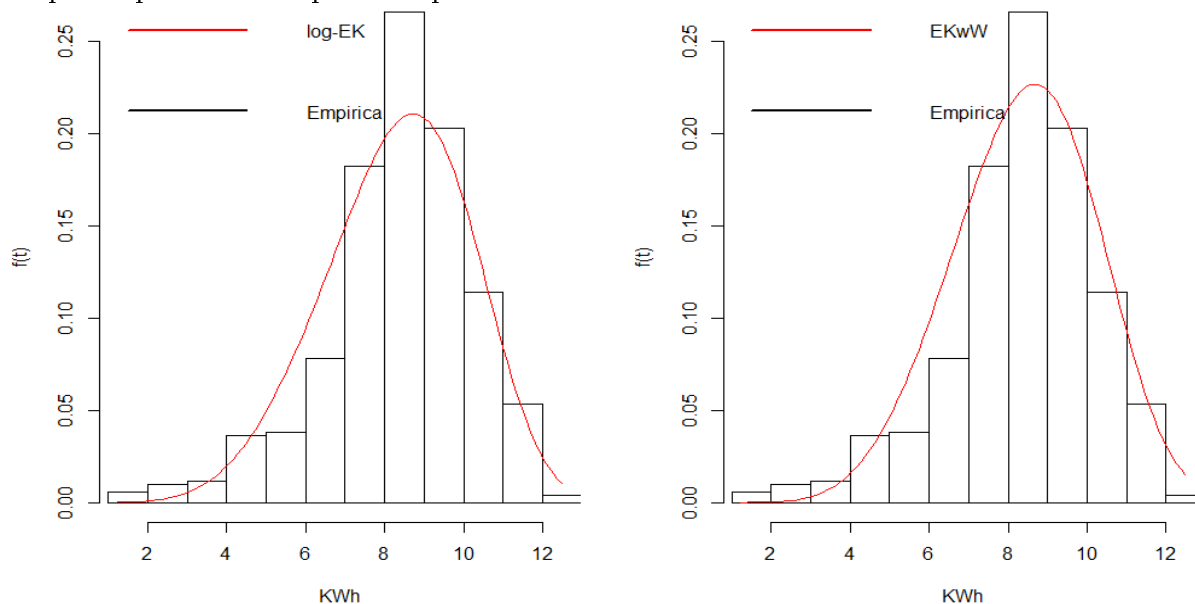


Figura 5.12: Curva das funções de densidade da Log-exponenciada de Kumaraswamy e Weibull Exponenciada Kumaraswamy versus a curva de suas funções de distribuições Empírica para seus respectivos parâmetros estimados



Nas Figuras (5.13) até (5.16) tem-se os plot de 4 gráficos das curvas da função distribuição acumulada versus suas empíricas. Em todas as visualizações gráficas tem-se uma predominância das distribuições Weibull, Weibull Exponenciada Kumaraswamy e Log-exponenciada de Kumaraswamy sendo assim as melhores alternativas em se estudar os dados mais afundo.

Figura 5.13: Curva das funções de distribuições acumuladas da Weibull e Gama versus a curva de suas funções de distribuições acumuladas Empírica para seus respectivos parâmetros estimados

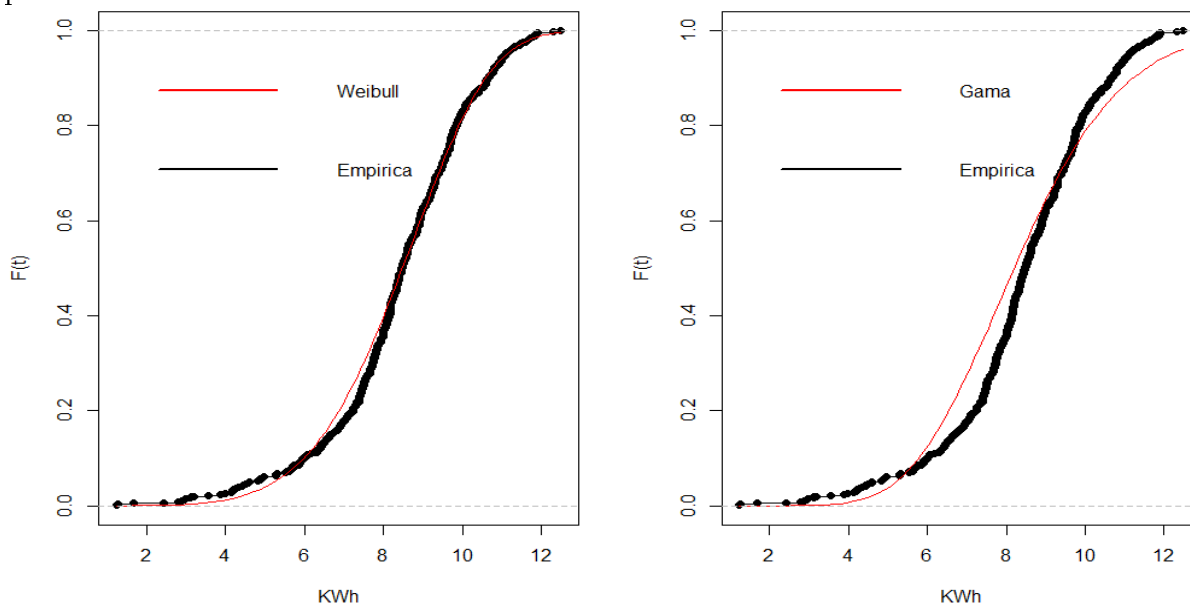


Figura 5.14: Curva das funções de distribuições acumuladas da Exponencial Generalizada e Normal versus a curva de suas funções de distribuições acumuladas Empírica para seus respectivos parâmetros estimados

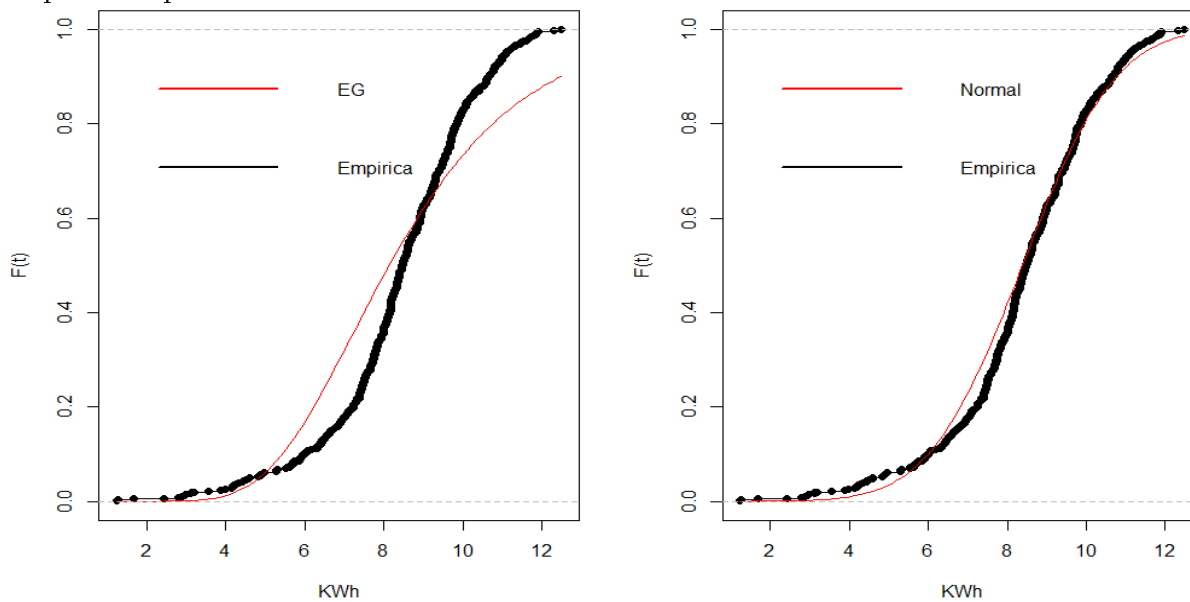


Figura 5.15: Curva das funções de distribuições acumuladas da Weibull Exponenciada e Weibull Kumaraswamy versus a curva de suas funções de distribuições acumuladas Empírica para seus respectivos parâmetros estimados

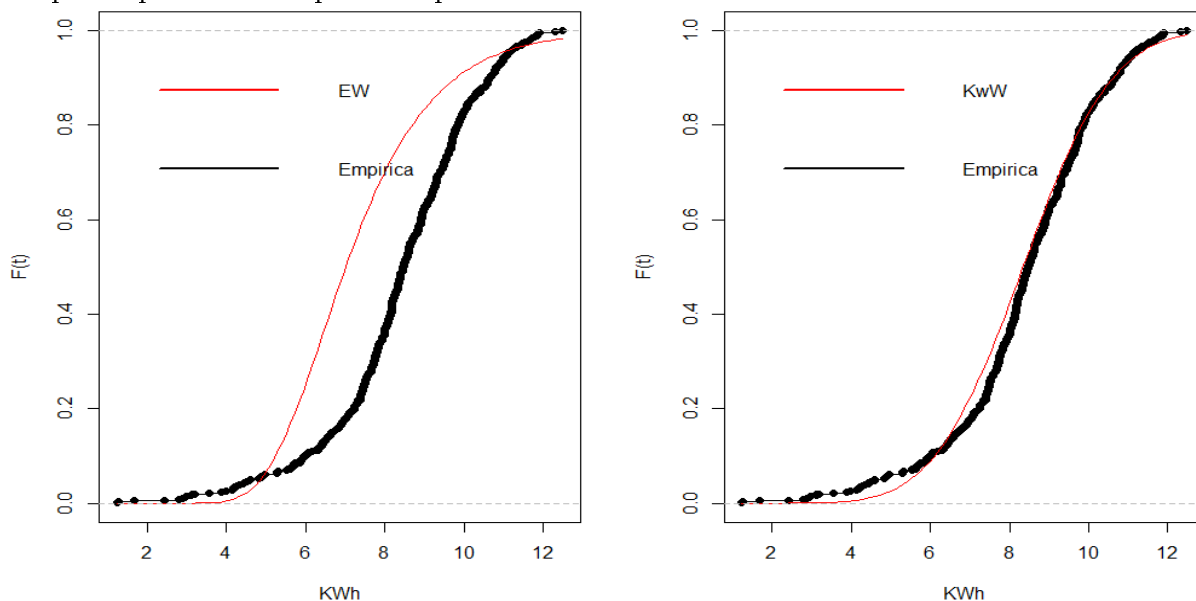
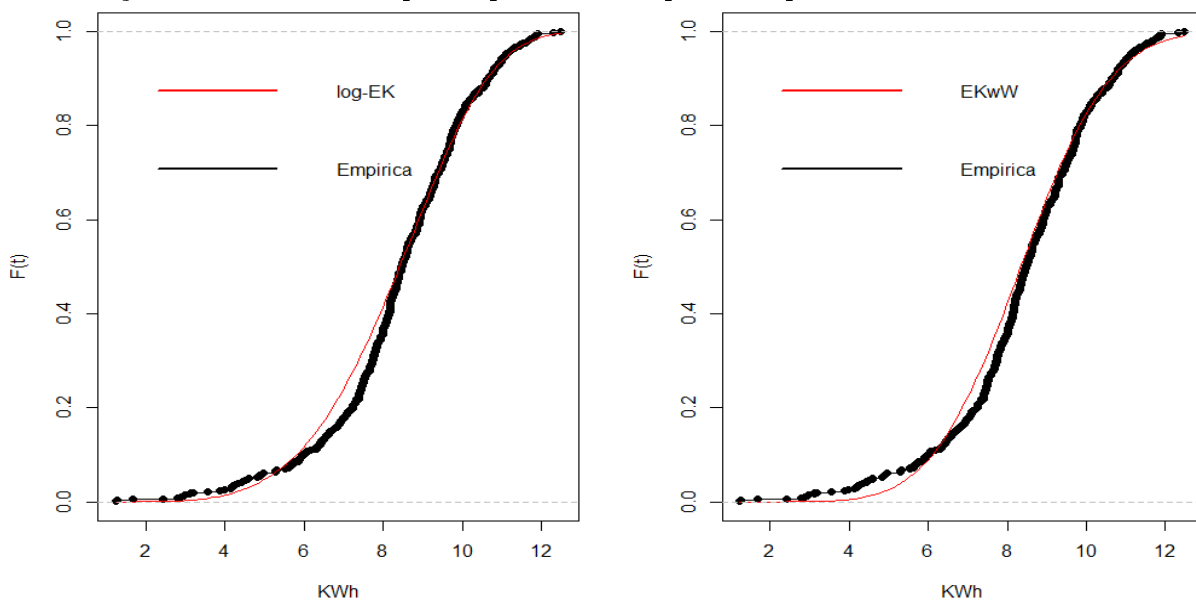


Figura 5.16: Curva das funções de distribuições acumuladas da Log-exponenciada de Kumaraswamy e Weibull Exponenciada Kumaraswamy versus a curva de suas funções de distribuições acumuladas Empírica para seus respectivos parâmetros estimados



Graficamente é possível verificar que os melhores ajustes foram visto na sequência pelas distribuições EKwW, Weibull, Log-EK, Normal e KwW. As outras três não citadas, Gama, EG e EW não tiveram bons resultados comparados com as demais. Este resultado será confirmado no próximo capítulo com a apresentação dos critérios de seleção de modelos.

Será feito no próximo capítulo a suposição de dependência entre os tempos de gerações de energia solar.

Modelos Markovianos baseados em distribuições Bivariadas

No ajuste de modelos univariados, houve a suposição de independência entre as observações. No entanto, em se tratando de observações sequenciais, esta suposição nem sempre é válida. Portanto, neste capítulo, será construído modelos sob a suposição markoviana, ou seja, a distribuição T_{i+1} depende somente do valor de T_i . Desta forma, a função de verossimilhança pode ser construída baseando-se em dados condicionais.

6.1 Modelo Markoviano Weibull via Distribuição Bivariada Weibull de Freund

É necessário definir a função de densidade conjunta de T_i e T_{i+i} , para isso foram estudadas três formulações de modelos Weibull Bivariados, sendo eles de Freund (1961) [13], Marshall e Olkin (1967) [21] e Ryu (1993) [30]. O modelo de Ryu foi descartado, pelo fato de não se encaixar aos dados de estudo, pois é um modelo absolutamente contínuo e assim não permite que $P(T_i = T_{i+1}) > 0$. Por exemplo, pode-se ter sequências de dias com uma mesma quantidade de energia gerada, pois a temperatura está muito alta ou muito baixa e assim a chance dessa probabilidade acontecer é alta. Portanto, o estudo seguiu com as outras duas formulações que serão apresentadas a seguir com intuito de comparar qual apresentou um melhor ajuste aos dados de interesse. As distribuições bivariadas Weibull (DBW) a seguir são todas obtidas de uma distribuição exponencial bivariada (DEB).

Antes de se iniciar a construção das distribuições bivariadas com um processo de Markov é interessante verificar se realmente existe dependência entre os tempos T_i e T_{i+i} . Assim na Tabela (6.1) é apresentada algumas medidas de dependência para os dados (Quantidade de energia solar gerada por placas fotovoltaicas na cidade de São José do Rio Preto -SP) T_i e T_{i+i} .

Observa-se na Tabela (6.1) que existe uma dependência aceitável entre os tempos, assim é uma boa alternativa à criação de modelos markovianos.

Tabela 6.1: Medidas de dependência para os tempos T_i e T_{i+i}

Método	Correlação
Pearson	0.4127
Kendall	0.3433
Spearman	0.4704

6.1.1 Distribuição Weibull Bivariada de Freund

Distribuição Weibull Bivariada (DBW) de Freund foi proposta por Freund (1961) [13] com a suposição que o mecanismo de falha de um sistema de dois componentes inicialmente têm taxas de falha constantes, λ_1 e λ_2 , com funções de densidades independentes quando ambos estão em operação:

$$f_i(t_i|\lambda_i) = \lambda_i \exp(-\lambda_i t_i); t_i \geq 0, \lambda_i > 0; i = 1, 2. \quad (6.1)$$

Neste modelo, os tempos T_1 e T_2 são dependentes porque uma falha de um componente não resulta em uma substituição, mas altera o parâmetro da distribuição da vida útil do outro componente para λ_i^* , principalmente para $\lambda_i^* > \lambda_i$ pois o componente sem falha tem uma carga de trabalho maior.

O tempo até a primeira falha é distribuído exponencialmente com o parâmetro:

$$\lambda = \lambda_1 + \lambda_2. \quad (6.2)$$

Levando assim a consideração de dependência, e a probabilidade de que o componente i seja o primeiro a falhar é $\frac{\lambda_i}{\lambda}$, sempre que ocorrer a primeira falha. A distribuição do tempo da primeira falha à falha do outro componente é, portanto, uma mistura de duas distribuições exponenciais com os parâmetros λ_1^* ou λ_2^* nas proporções $\frac{\lambda_2}{\lambda}$ e $\frac{\lambda_1}{\lambda}$, respectivamente. E assim, a função distribuição T_1 e T_2 é dada por:

$$f_{1,2}(t_1, t_2) = \begin{cases} f_1(t_1|\lambda_1)S_2(t_1|\lambda_2)f_2(t_2 - t_1|\lambda_2^*) & \text{para } 0 \leq t_1 < t_2 \\ f_2(t_2|\lambda_2)S_1(t_2|\lambda_1)f_1(t_1 - t_2|\lambda_1^*) & \text{para } 0 \leq t_2 < t_1, \end{cases} \quad (6.3)$$

ou seja,

$$f_{1,2}(t_1, t_2) = \begin{cases} \lambda_1 \lambda_2^* \exp\{-\lambda_2^* t_2 - (-\lambda_1 + \lambda_2 - \lambda_2^*) t_1\} & \text{para } 0 \leq t_1 < t_2 \\ \lambda_2 \lambda_1^* \exp\{-\lambda_1^* t_2 - (-\lambda_1 + \lambda_2 - \lambda_1^*) t_2\} & \text{para } 0 \leq t_2 < t_1, \end{cases} \quad (6.4)$$

onde:

$$S_i(t_i|\lambda_i) = \exp\{-\lambda_i t_i\}.$$

Esta função bivariada é absolutamente contínua, pois $P(T_1 = T_2) \geq 0$
A função de distribuição acumulada conjunta é dada por:

$$F_{1,2}(t_1, t_2) = \begin{cases} \frac{1}{\lambda_1 + \lambda_2 - \lambda_2^*} [\lambda_1 \exp\{-(\lambda_1 + \lambda_2 - \lambda_2^*)t_1 - \lambda_2^*t_2\} + \\ (\lambda_2 - \lambda_2^*) \exp\{-(\lambda_1 + \lambda_2)t_2\}] \text{para } 0 \leq t_1 < t_2 \\ \frac{1}{\lambda_1 + \lambda_2 - \lambda_2^*} [\lambda_1 \exp\{-(\lambda_1 + \lambda_2 - \lambda_2^*)t_1 - \lambda_2^*t_2\} + \\ (\lambda_2 - \lambda_2^*) \exp\{-(\lambda_1 + \lambda_2)t_2\}] \text{para } 0 \leq t_1 < t_2. \end{cases} \quad (6.5)$$

Desde que $\lambda_1 + \lambda_2 \neq \lambda_i^*$, a marginal da função densidade de t_i é:

$$f_i(t_i) = \frac{1}{\lambda_1 + \lambda_2 + \lambda_i^*} \{(\lambda_i - \lambda_i^*)(\lambda_1 + \lambda_2) \exp\{-(\lambda_1 + \lambda_2)t_i\} + \lambda_i^* \lambda_{3-1} \exp\{-\lambda_i^*t_i\}\}. \quad (6.6)$$

Desta forma, para a obtenção da distribuição Weibull Bivariada, considere a distribuição Weibull sob a seguinte parametrização:

$$f_i(t_i | \alpha_i, \lambda_i) = \alpha_i \lambda_i (\lambda_i t_i)^{\alpha_i - 1} \exp\{-(\lambda_i t_i)^{\alpha_i}\}; t_i \geq 0; \lambda_i, \alpha_i > 0; i = 1, 2. \quad (6.7)$$

Quando um componente falha, a vida útil restante do outro componente ainda é Weibull, mas com os parâmetros possivelmente alterados λ_i^* e α^* . O tempo é redefinido no componente sobrevivente. O parâmetro de forma α_i^* pode ser igual a α_i , mas o parâmetro de escala λ_i^* não é igual a λ_i em geral, o que faz com que a taxa de risco do componente não seja alterada com a falha.

Como o tempo é redefinido no componente sobrevivente, ele tem uma distribuição Weibull deslocada e (6.3) pode ser escrito como:

$$f_{1,2}(t_1, t_2) = \begin{cases} f_1(t_1 | \alpha_1, \lambda_1) S_2(t_1 | \alpha_2, \lambda_2) f_2(t_2 - t_1 | \alpha_2^*, \lambda_2^*) & \text{para } 0 \leq t_1 < t_2 \\ f_2(t_2 | \alpha_2, \lambda_2) S_1(t_2 | \alpha_1, \lambda_1) f_1(t_1 - t_2 | \alpha_1^*, \lambda_1^*) & \text{para } 0 \leq t_2 < t_1, \end{cases} \quad (6.8)$$

ou seja,

$$f_{1,2}(t_1, t_2) = \begin{cases} \alpha_1 \lambda_1 (\lambda_1 t_1)^{\alpha_1 - 1} \exp\{-(\lambda_1 t_1)^{\alpha_1}\} \exp\{-(\lambda_2 t_1)^{\alpha_2}\} \times \\ \alpha_2^* \lambda_2^* [\lambda_2^* (t_2 - t_1)]^{\alpha_2^* - 1} \exp\{-[\lambda_2^* (t_2 - t_1)]^{\alpha_2^*}\} & \text{para } 0 \leq t_1 < t_2 \\ \alpha_2 \lambda_2 (\lambda_2 t_2)^{\alpha_2 - 1} \exp\{-(\lambda_2 t_2)^{\alpha_2}\} \exp\{-(\lambda_1 t_2)^{\alpha_1}\} \times \\ \alpha_1^* \lambda_1^* [\lambda_1^* (t_1 - t_2)]^{\alpha_1^* - 1} \exp\{-[\lambda_1^* (t_1 - t_2)]^{\alpha_1^*}\} & \text{para } 0 \leq t_2 < t_1. \end{cases} \quad (6.9)$$

Contudo em Lu (1989) [19] é mostrado que a extensão Weibull de Freund não é igual a um modelo Weibull bivariada obtido usando uma transformação de potência direta das marginais de Freund, resultando uma nova transformação em:

$$f_{1,2}^*(t_1, t_2) = \begin{cases} f_1(t_1|\alpha_1, \lambda_1)S_1(t_1|(\lambda_2 - \lambda_2^*), \alpha_1)f_2(t_2|\lambda_2^*, \alpha_2) & \text{para } 0 \leq t_1^{\alpha_1} < t_2^{\alpha_2} \\ f_2(t_2|\alpha_2, \lambda_2)S_2(t_2|(\lambda_1 - \lambda_1^*), \alpha_2)f_1(t_1|\lambda_1^*, \alpha_1) & \text{para } 0 \leq t_2^{\alpha_2} < t_1^{\alpha_1}. \end{cases} \quad (6.10)$$

Tem-se assim a DWB de Freund, mais detalhes pode ser encontrado em Rinne (2009) [27]. A seguir temos a construção do modelo Markoviano Weibull para que posteriormente se tenha a estimação dos parâmetros.

6.1.2 Modelo Markoviano Weibull via a Distribuição Weibull Bivariada de Freund

O primeiro passo para definir a função de densidade condicional de T_{i+1} dado T_i será utilizar o resultado da equação (6.10), assim temos:

$$f_{1,2}^*(t_1, t_2) = \begin{cases} \alpha_1 \lambda_1 (\lambda_1 t_1)^{\alpha_1 - 1} \exp\{-(\lambda_1 t_1)^{\alpha_1}\} \exp\{-[(\lambda_2^{\alpha_1} - \lambda_2^{*\alpha_1})t_1^{\alpha_1}]\} \times \\ \alpha_2 \lambda_2^* [\lambda_2^* t_2]^{\alpha_2 - 1} \exp\{-(\lambda_2^* t_2)^{\alpha_2}\} & \text{para } 0 \leq t_1^{\alpha_1} < t_2^{\alpha_2} \\ \alpha_2 \lambda_2 (\lambda_2 t_2)^{\alpha_2 - 1} \exp\{-(\lambda_2 t_2)^{\alpha_2}\} \exp\{-[(\lambda_1^{\alpha_2} - \lambda_1^{*\alpha_2})t_2^{\alpha_2}]\} \times \\ \alpha_1 \lambda_1^* [\lambda_1^* t_1]^{\alpha_1 - 1} \exp\{-(\lambda_1^* t_1)^{\alpha_1}\} & \text{para } 0 \leq t_2^{\alpha_2} < t_1^{\alpha_1}, \end{cases} \quad (6.11)$$

que pode ser reduzido em:

$$f_{1,2}^*(t_1, t_2) = \begin{cases} \alpha_1 \lambda_1 (\lambda_1 t_1)^{\alpha_1 - 1} \alpha_2 \lambda_2^* [\lambda_2^* t_2]^{\alpha_2 - 1} \times \\ \exp\{-(\lambda_1 t_1)^{\alpha_1} - [(\lambda_2^{\alpha_1} - \lambda_2^{*\alpha_1})t_1^{\alpha_1}] - (\lambda_2^* t_2)^{\alpha_2}\} & \text{para } 0 \leq t_1^{\alpha_1} < t_2^{\alpha_2} \\ \alpha_2 \lambda_2 (\lambda_2 t_2)^{\alpha_2 - 1} \alpha_1 \lambda_1^* [\lambda_1^* t_1]^{\alpha_1 - 1} \times \\ \exp\{-(\lambda_2 t_2)^{\alpha_2} - [(\lambda_1^{\alpha_2} - \lambda_1^{*\alpha_2})t_2^{\alpha_2}] - (\lambda_1^* t_1)^{\alpha_1}\} & \text{para } 0 \leq t_2^{\alpha_2} < t_1^{\alpha_1}. \end{cases} \quad (6.12)$$

Além da equação (6.12) considere as seguintes reparametrizações: $\lambda_1 = \frac{1}{\beta_1}$, $\lambda_2 = \frac{1}{\beta_2}$, $\lambda_1^* = \frac{1}{\beta_1^*}$ e $\lambda_2^* = \frac{1}{\beta_2^*}$, com isso tem-se a função de densidade conjunta de (T_{i+1}, T_i) , dada por:

$$f(t_i, t_{i+1}) = \begin{cases} \alpha_1 \frac{1}{\beta_1} \left(\frac{1}{\beta_1} t_i \right)^{\alpha_1 - 1} \alpha_2 \frac{1}{\beta_2^*} \left[\frac{1}{\beta_2^*} t_{i+1} \right]^{\alpha_2 - 1} \times \\ \exp \left\{ - \left(\frac{1}{\beta_1} t_i \right)^{\alpha_1} - \left[\left(\frac{1}{\beta_2^{\alpha_1}} - \frac{1}{\beta_2^{*\alpha_1}} \right) t_i^{\alpha_1} \right] - \left[\frac{1}{\beta_2^*} t_{i+1} \right]^{\alpha_2} \right\} & \text{para } 0 \leq t_i^{\alpha_1} < t_{i+1}^{\alpha_2} \\ \alpha_2 \frac{1}{\beta_2} \left(\frac{1}{\beta_2} t_{i+1} \right)^{\alpha_2 - 1} \alpha_1 \frac{1}{\beta_1^*} \left[\frac{1}{\beta_1^*} t_i \right]^{\alpha_1 - 1} \times \\ \exp \left\{ - \left(\frac{1}{\beta_2} t_{i+1} \right)^{\alpha_2} - \left[\left(\frac{1}{\beta_1^{\alpha_2}} - \frac{1}{\beta_1^{*\alpha_2}} \right) t_{i+1}^{\alpha_2} \right] - \left[\frac{1}{\beta_1^*} t_i \right]^{\alpha_1} \right\} & \text{para } 0 \leq t_{i+1}^{\alpha_2} < t_i^{\alpha_1}, \end{cases} \quad (6.13)$$

que é equivalente à:

$$f(t_i, t_{i+1}) = \begin{cases} \frac{\alpha_1}{\beta_1^{\alpha_1}} t_i^{\alpha_1 - 1} \frac{\alpha_2}{\beta_2^{*\alpha_2}} t_{i+1}^{\alpha_2 - 1} \times \\ \exp \left\{ - \left(\frac{t_i}{\beta_1} \right)^{\alpha_1} - \left(\left(\frac{t_i}{\beta_2} \right)^{\alpha_1} - \left(\frac{t_i}{\beta_2^*} \right)^{\alpha_1} \right) - \left[\frac{t_{i+1}}{\beta_2^*} \right]^{\alpha_2} \right\} & \text{para } 0 \leq t_i^{\alpha_1} < t_{i+1}^{\alpha_2} \\ \frac{\alpha_2}{\beta_2^{\alpha_2}} t_{i+1}^{\alpha_2 - 1} \frac{\alpha_1}{\beta_1^{*\alpha_1}} t_i^{\alpha_1 - 1} \times \\ \exp \left\{ - \left(\frac{t_{i+1}}{\beta_2} \right)^{\alpha_2} - \left(\left(\frac{t_{i+1}}{\beta_1} \right)^{\alpha_2} - \left(\frac{t_{i+1}}{\beta_1^*} \right)^{\alpha_2} \right) - \left[\frac{t_i}{\beta_1^*} \right]^{\alpha_1} \right\} & \text{para } 0 \leq t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases} \quad (6.14)$$

Portanto a função de densidade condicional de T_{i+1} dado T_i é dada pela expressão:

$$f(t_{i+1}/t_i) = \frac{f(t_i, t_{i+1})}{f(t_i)},$$

e pelas equações (5.1) e (6.14) tem-se o seguinte resultado para a densidade condicional:

$$f(t_{i+1}/t_i) = \begin{cases} \frac{\alpha_2}{\beta_2^{*\alpha_2}} t_{i+1}^{\alpha_2 - 1} \exp \left\{ - \left(\left(\frac{t_i}{\beta_2} \right)^{\alpha_1} - \left(\frac{t_i}{\beta_2^*} \right)^{\alpha_1} \right) - \left[\frac{t_{i+1}}{\beta_2^*} \right]^{\alpha_2} \right\} & \text{para } 0 \leq t_i^{\alpha_1} < t_{i+1}^{\alpha_2} \\ \frac{\alpha_2}{\beta_2^{\alpha_2}} t_{i+1}^{\alpha_2 - 1} \frac{1}{\beta_1^{*\alpha_1}} \left(\frac{1}{\beta_1^{\alpha_1}} \right)^{-1} \times \\ \exp \left\{ - \left(\frac{t_{i+1}}{\beta_2} \right)^{\alpha_2} - \left(\left(\frac{t_{i+1}}{\beta_1} \right)^{\alpha_2} - \left(\frac{t_{i+1}}{\beta_1^*} \right)^{\alpha_2} \right) - \left[\frac{t_i}{\beta_1^*} \right]^{\alpha_1} + \left[\frac{t_i}{\beta_1} \right]^{\alpha_1} \right\} & \text{para } 0 \leq t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases} \quad (6.15)$$

A função de distribuição acumulada condicional T_{i+1} dado T_i é dada pela expressão:

$$F(t_{i+1}/t_i) = \int_0^{t_{i+1}} f(t_{i+1}/t_i) dt_{i+1}$$

$$F(t_{i+1}/t_i) = \begin{cases} \int_0^{t_{i+1}} \frac{\alpha_2}{\beta_2^{*\alpha_2}} t_{i+1}^{\alpha_2-1} \exp \left\{ - \left(\left(\frac{t_i}{\beta_2} \right)^{\alpha_1} - \left(\frac{t_i}{\beta_2^*} \right)^{\alpha_1} \right) - \left[\frac{t_{i+1}}{\beta_2^*} \right]^{\alpha_2} \right\} dt_{i+1} & \text{para } 0 \leq t_i^{\alpha_1} < t_{i+1}^{\alpha_2} \\ \int_{t_{i+1}}^{\infty} \frac{\alpha_2}{\beta_2^{\alpha_2}} t_{i+1}^{\alpha_2-1} \frac{1}{\beta_1^{*\alpha_1}} \left(\frac{1}{\beta_1^{\alpha_1}} \right)^{-1} \times \\ \exp \left\{ - \left(\frac{t_{i+1}}{\beta_2} \right)^{\alpha_2} - \left(\left(\frac{t_{i+1}}{\beta_1} \right)^{\alpha_2} - \left(\frac{t_{i+1}}{\beta_1^*} \right)^{\alpha_2} \right) - \left[\frac{t_i}{\beta_1^*} \right]^{\alpha_1} + \left[\frac{t_i}{\beta_1} \right]^{\alpha_1} \right\} dt_{i+1} & \text{para } 0 \leq t_{i+1}^{\alpha_2} < t_i^{\alpha_1} \end{cases}$$

$$= \begin{cases} \exp \left\{ - \left(\left(\frac{t_i}{\beta_2} \right)^{\alpha_1} - \left(\frac{t_i}{\beta_2^*} \right)^{\alpha_1} \right) \right\} \int_0^{t_{i+1}} \frac{\alpha_2}{\beta_2^{*\alpha_2}} t_{i+1}^{\alpha_2-1} \exp \left\{ - \left[\frac{t_{i+1}}{\beta_2^*} \right]^{\alpha_2} \right\} dt_{i+1} & \text{para } 0 \leq t_i^{\alpha_1} < t_{i+1}^{\alpha_2} \\ \frac{\beta_1^{\alpha_1}}{\beta_1^{*\alpha_1}} \exp \left\{ - \left[\frac{t_i}{\beta_1^*} \right]^{\alpha_1} + \left[\frac{t_i}{\beta_1} \right]^{\alpha_1} \right\} \times \\ \int_{t_{i+1}}^{\infty} \frac{\alpha_2}{\beta_2^{\alpha_2}} t_{i+1}^{\alpha_2-1} \exp \left\{ - \left(\frac{1}{\beta_1^{\alpha_2}} + \frac{1}{\beta_2^{\alpha_2}} - \frac{1}{\beta_1^{*\alpha_2}} \right) t_{i+1}^{\alpha_2} \right\} dt_{i+1} & \text{para } 0 \leq t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases}$$

Assim para $0 \leq t_i^{\alpha_1} < t_{i+1}^{\alpha_2}$ temos uma integral conhecida, sendo a acumulada da distribuição Weibull

$$= \begin{cases} \exp \left\{ - \left(\left(\frac{t_i}{\beta_2} \right)^{\alpha_1} - \left(\frac{t_i}{\beta_2^*} \right)^{\alpha_1} \right) \right\} \left[1 - \exp \left\{ - \left[\frac{t_{i+1}}{\beta_2^*} \right]^{\alpha_2} \right\} \right] & \text{para } 0 \leq t_i^{\alpha_1} < t_{i+1}^{\alpha_2} \\ \frac{\beta_1^{\alpha_1}}{\beta_1^{*\alpha_1} \beta_2^{\alpha_2}} \exp \left\{ - \left[\frac{t_i}{\beta_1^*} \right]^{\alpha_1} + \left[\frac{t_i}{\beta_1} \right]^{\alpha_1} \right\} \times \\ \int_{t_{i+1}}^{\infty} \left(\frac{1}{\beta_1^{\alpha_2}} + \frac{1}{\beta_2^{\alpha_2}} - \frac{1}{\beta_1^{*\alpha_2}} \right)^{-1} \left(\frac{1}{\beta_1^{\alpha_2}} + \frac{1}{\beta_2^{\alpha_2}} - \frac{1}{\beta_1^{*\alpha_2}} \right) \alpha_2 t_{i+1}^{\alpha_2-1} \times \\ \exp \left\{ - \left(\frac{1}{\beta_1^{\alpha_2}} + \frac{1}{\beta_2^{\alpha_2}} - \frac{1}{\beta_1^{*\alpha_2}} \right) t_{i+1}^{\alpha_2} \right\} dt_{i+1} & \text{para } 0 \leq t_{i+1}^{\alpha_2} < t_i^{\alpha_1} \end{cases}$$

$$= \begin{cases} \exp \left\{ - \left(\left(\frac{t_i}{\beta_2} \right)^{\alpha_1} - \left(\frac{t_i}{\beta_2^*} \right)^{\alpha_1} \right) \right\} \left[1 - \exp \left\{ - \left[\frac{t_{i+1}}{\beta_2^*} \right]^{\alpha_2} \right\} \right] & \text{para } 0 \leq t_i^{\alpha_1} < t_{i+1}^{\alpha_2} \\ \frac{\beta_1^{\alpha_1}}{\beta_1^{*\alpha_1} \beta_2^{\alpha_2}} \left(\frac{1}{\beta_1^{\alpha_2}} + \frac{1}{\beta_2^{\alpha_2}} - \frac{1}{\beta_1^{*\alpha_2}} \right)^{-1} \exp \left\{ - \left[\frac{t_i}{\beta_1^*} \right]^{\alpha_1} + \left[\frac{t_i}{\beta_1} \right]^{\alpha_1} \right\} \times \\ \int_{t_{i+1}}^{\infty} \left(\frac{1}{\beta_1^{\alpha_2}} + \frac{1}{\beta_2^{\alpha_2}} - \frac{1}{\beta_1^{*\alpha_2}} \right) \alpha_2 t_{i+1}^{\alpha_2-1} \exp \left\{ - \left(\frac{1}{\beta_1^{\alpha_2}} + \frac{1}{\beta_2^{\alpha_2}} - \frac{1}{\beta_1^{*\alpha_2}} \right) t_{i+1}^{\alpha_2} \right\} dt_{i+1} & \text{para } 0 \leq t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases}$$

Novamente após algumas transformações, para $0 \leq t_{i+1}^{\alpha_2} < t_i^{\alpha_1}$ a integral é conhecida, sendo a acumulada da distribuição Weibull, portando a função acumulada condicional de T_{i+1} dado T_i é dada por:

$$F(t_{i+1}/t_i) = \begin{cases} \exp \left\{ - \left(\left(\frac{t_i}{\beta_2} \right)^{\alpha_1} - \left(\frac{t_i}{\beta_2^*} \right)^{\alpha_1} \right) \right\} \left[1 - \exp \left\{ - \left[\frac{t_{i+1}}{\beta_2^*} \right]^{\alpha_2} \right\} \right] & \text{para } 0 \leq t_i^{\alpha_1} < t_{i+1}^{\alpha_2} \\ \frac{\beta_1^{\alpha_1}}{\beta_1^{*\alpha_1} \beta_2^{\alpha_2}} \left(\frac{1}{\beta_1^{\alpha_2}} + \frac{1}{\beta_2^{\alpha_2}} - \frac{1}{\beta_1^{*\alpha_2}} \right)^{-1} \exp \left\{ - \left[\frac{t_i}{\beta_1^*} \right]^{\alpha_1} + \left[\frac{t_i}{\beta_1} \right]^{\alpha_1} \right\} \times \\ \left[1 - \exp \left\{ - \left(\frac{1}{\beta_1^{\alpha_2}} + \frac{1}{\beta_2^{\alpha_2}} - \frac{1}{\beta_1^{*\alpha_2}} \right) t_{i+1}^{\alpha_2} \right\} \right] & \text{para } 0 \leq t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases} \quad (6.16)$$

Com as funções definidas tem-se a seguir as estimações de seus devidos parâmetros

6.1.3 Estimação de Máxima Verossimilhança do Modelo Markoviano Weibull via a Distribuição Weibull Bivariada de Freund

Com a suposição de que os tempos seguem um processo de Markov, o que indica que a função de verossimilhança é apresentada de uma maneira diferente do que a vista no capítulo 5, pois agora T_i , $i = 1, 2, \dots, n$, são dependentes. Assim, a expressão (2.4) para um processo de Markov, pode ser escrita como:

$$L(\boldsymbol{\theta}; T) = f(t_1) \prod_{i=1}^{n-1} f(t_{i+1}/t_i | \boldsymbol{\theta}). \quad (6.17)$$

Assim substituindo as expressões (5.1) e (6.15), em (6.17), temos a função de verossimilhança dado por $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_1^*, \beta_2^*)$:

$$\begin{aligned} L(\boldsymbol{\theta}) &= \frac{\alpha_1}{\beta_1^{\alpha_1}} t_1^{\alpha_1-1} \exp \left\{ - \left(\frac{t_1}{\beta_1} \right)^{\alpha_1} \right\} \prod_{i=1}^{n-1} [f(t_{i+1}/t_i | \boldsymbol{\theta})] \\ &= \frac{\alpha_1}{\beta_1^{\alpha_1}} t_1^{\alpha_1-1} \exp \left\{ - \left(\frac{t_1}{\beta_1} \right)^{\alpha_1} \right\} \\ &\times \begin{cases} \prod_{i=1}^{n-1} \frac{\alpha_2}{\beta_2^{*\alpha_2}} t_{i+1}^{\alpha_2-1} \exp \left\{ - \left(\left(\frac{t_i}{\beta_2} \right)^{\alpha_1} - \left(\frac{t_i}{\beta_2^*} \right)^{\alpha_1} \right) - \left[\frac{t_{i+1}}{\beta_2^*} \right]^{\alpha_2} \right\} & \text{para } 0 \leq t_i^{\alpha_1} < t_{i+1}^{\alpha_2} \\ \prod_{i=1}^{n-1} \frac{\alpha_2}{\beta_2^{\alpha_2}} t_{i+1}^{\alpha_2-1} \frac{1}{\beta_1^{*\alpha_1}} \left(\frac{1}{\beta_1^{\alpha_1}} \right)^{-1} \times \\ \exp \left\{ - \left(\frac{t_{i+1}}{\beta_2} \right)^{\alpha_2} - \left(\left(\frac{t_{i+1}}{\beta_1} \right)^{\alpha_2} - \left(\frac{t_{i+1}}{\beta_1^*} \right)^{\alpha_2} \right) - \left[\frac{t_i}{\beta_1^*} \right]^{\alpha_1} + \left[\frac{t_i}{\beta_1} \right]^{\alpha_1} \right\} & \text{para } 0 \leq t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases} \end{aligned}$$

Aplicando o logaritmo na função acima para facilitar os cálculos temos:

$$\ln(L(\boldsymbol{\theta})) = \ln(\alpha_1) - \alpha_1 \ln(\beta_1) + (\alpha_1 - 1) \ln(t_1) - \left(\frac{t_1}{\beta_1} \right)^{\alpha_1} +$$

$$\begin{aligned}
& + \begin{cases} \sum_{i=1}^{n-1} \ln(\alpha_2) - \sum_{i=1}^{n-1} (\alpha_2) \ln(\beta_2^*) + \sum_{i=1}^{n-1} (\alpha_2 - 1) \ln(t_{i+1}) & \text{para } 0 \leq t_i^{\alpha_1} < t_{i+1}^{\alpha_2} \\ \sum_{i=1}^{n-1} \ln(\alpha_2) + \sum_{i=1}^{n-1} (\alpha_2 - 1) \ln(t_{i+1}) - \sum_{i=1}^{n-1} (\alpha_2) \ln(\beta_2) & \text{para } 0 \leq t_{i+1}^{\alpha_2} < t_i^{\alpha_1} \end{cases} \\
& + \begin{cases} \sum_{i=1}^{n-1} \left[- \left(\left(\frac{t_i}{\beta_2} \right)^{\alpha_1} - \left(\frac{t_i}{\beta_2^*} \right)^{\alpha_1} \right) - \left[\frac{t_{i+1}}{\beta_2^*} \right]^{\alpha_2} \right] & \text{para } 0 \leq t_i^{\alpha_1} < t_{i+1}^{\alpha_2} \\ \sum_{i=1}^{n-1} (\alpha_1) \ln(\beta_1) - \sum_{i=1}^{n-1} (\alpha_1) \ln(\beta_1^*) + \\ \sum_{i=1}^{n-1} \left[- \left(\frac{t_{i+1}}{\beta_2} \right)^{\alpha_2} - \left(\left(\frac{t_{i+1}}{\beta_1} \right)^{\alpha_2} - \left(\frac{t_{i+1}}{\beta_1^*} \right)^{\alpha_2} \right) - \left[\frac{t_i}{\beta_1^*} \right]^{\alpha_1} + \left[\frac{t_i}{\beta_1} \right]^{\alpha_1} \right] & \text{para } 0 \leq t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases}
\end{aligned}$$

Simplificando a expressão do logaritmo na função temos a seguinte equação

$$\begin{aligned}
\ln(L(\boldsymbol{\theta})) &= \ln(\alpha_1) - \alpha_1 \ln(\beta_1) + (\alpha_1 - 1) \ln(t_1) - \left(\frac{t_1}{\beta_1} \right)^{\alpha_1} + \\
& + \begin{cases} (n-1) \ln(\alpha_2) - (n-1) (\alpha_2) \ln(\beta_2^*) + (\alpha_2 - 1) \sum_{i=1}^{n-1} \ln(t_{i+1}) & \text{para } 0 \leq t_i^{\alpha_1} < t_{i+1}^{\alpha_2} \\ (n-1) \ln(\alpha_2) - (n-1) (\alpha_2) \ln(\beta_2) + (\alpha_2 - 1) \sum_{i=1}^{n-1} \ln(t_{i+1}) & \text{para } 0 \leq t_{i+1}^{\alpha_2} < t_i^{\alpha_1} \end{cases} \\
& + \begin{cases} \sum_{i=1}^{n-1} \left[- \left(\left(\frac{t_i}{\beta_2} \right)^{\alpha_1} - \left(\frac{t_i}{\beta_2^*} \right)^{\alpha_1} \right) - \left[\frac{t_{i+1}}{\beta_2^*} \right]^{\alpha_2} \right] & \text{para } 0 \leq t_i^{\alpha_1} < t_{i+1}^{\alpha_2} \\ (n-1) (\alpha_1) \ln(\beta_1) - (n-1) (\alpha_1) \ln(\beta_1^*) + \\ \sum_{i=1}^{n-1} \left[- \left(\frac{t_{i+1}}{\beta_2} \right)^{\alpha_2} - \left(\left(\frac{t_{i+1}}{\beta_1} \right)^{\alpha_2} - \left(\frac{t_{i+1}}{\beta_1^*} \right)^{\alpha_2} \right) - \left[\frac{t_i}{\beta_1^*} \right]^{\alpha_1} + \left[\frac{t_i}{\beta_1} \right]^{\alpha_1} \right] & \text{para } 0 \leq t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases} \tag{6.18}
\end{aligned}$$

Para encontrar os estimadores de máxima verossimilhança para $(\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_1^*, \beta_2^*)$, será usado o programa R com a utilização do pacote Optim. Pode-se ver os resultados obtidos na Tabela (6.2) e na Figura (6.1) tem-se o ajuste da função acumulada ao conjunto de dados.

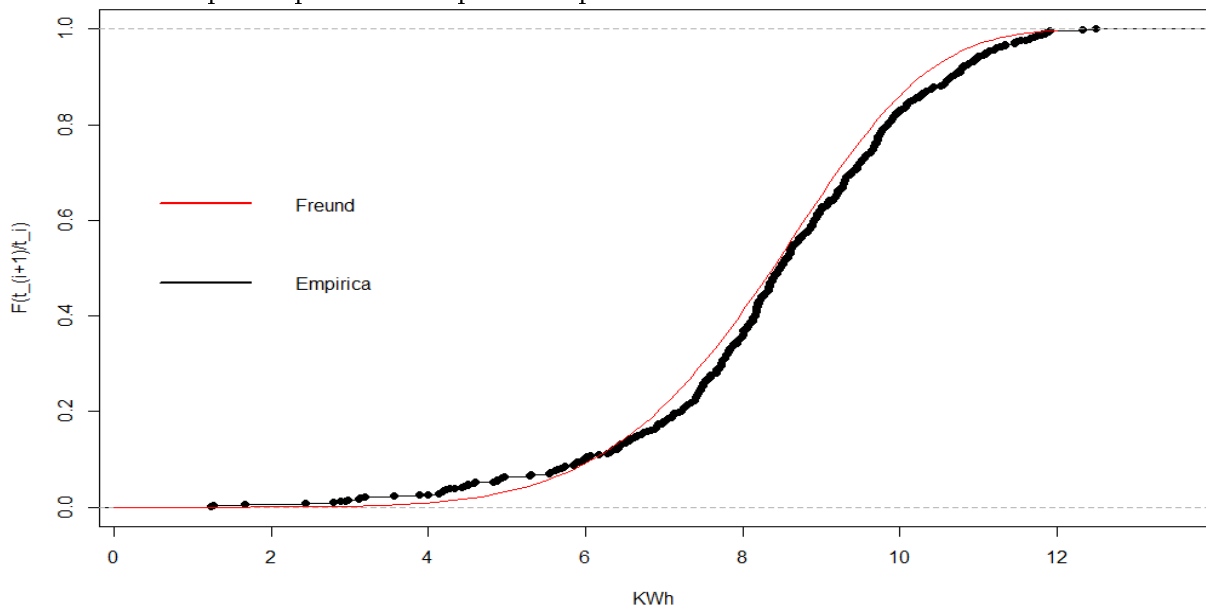
Tabela 6.2: Parâmetros estimados pelo EMV para os parâmetros do Modelo markoviano via a distribuição bivariada Weibull de Freund

Parâmetros	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1^*$	$\hat{\beta}_2^*$
EMV	5.071	5.922	8.751	8.911	8.750	8.913
(Erro Padrão)	(1.246)	(0.071)	(0.084)	(0.004)	(0.041)	(0.106)

Após a obtenção dos estimadores e dos seus respectivos Erros Padrões, é possível então verificar o ajuste do modelo comparando se a função acumulada do modelo e os dados empíricos. Estas curvas são apresentadas na Figura (6.1).

Pela análise gráfica é possível observar que a EMV obteve um bom resultado para os parâmetros da função de distribuição acumulada no Modelo markoviano via a distribuição bivariada Weibull de Freund, resultando assim em um ajuste satisfatório aos dados reais.

Figura 6.1: Curva da função de distribuição acumulada do Modelo markoviano via a distribuição bivariada Weibull de Freund versus a curva de sua função de distribuição acumulada Empírica para seu respectivos parâmetros estimados



Posteriormente será apresentado os resultados de AIC, BIC e AICc, confirmando se realmente esse ajuste foi bom e qual modelo é o melhor a se utilizar para os dados em questão.

Outro resultado possível de se obter é a estimação da média esperada de energia solar gerada por placas fotovoltaicas nos próximos dias em função do número observado entre o dia anterior e o dia atual, isto é, a esperança de T_{i+1} dado $T_i = t_i$.

$$E[T_{i+1}/T_i = t_i] = \int_0^{\infty} t_{i+1} [f(t_{i+1}/t_i|\boldsymbol{\theta})] dt_{i+1}$$

$$\begin{aligned}
&= \begin{cases} \int_0^\infty t_{i+1} \frac{\alpha_2}{\beta_2^{*\alpha_2}} t_{i+1}^{\alpha_2-1} \exp \left\{ - \left(\left(\frac{t_i}{\beta_2} \right)^{\alpha_1} - \left(\frac{t_i}{\beta_2^*} \right)^{\alpha_1} \right) - \left[\frac{t_{i+1}}{\beta_2^*} \right]^{\alpha_2} \right\} dt_{i+1} & \text{para } 0 \leq t_i^{\alpha_1} < t_{i+1}^{\alpha_2} \\ \int_0^\infty t_{i+1} \frac{\alpha_2}{\beta_2^{\alpha_2}} t_{i+1}^{\alpha_2-1} \frac{1}{\beta_1^{*\alpha_1}} \left(\frac{1}{\beta_1^{\alpha_1}} \right)^{-1} \times \\ \exp \left\{ - \left(\frac{1}{\beta_2} t_{i+1} \right)^{\alpha_2} - \left[\left(\frac{1}{\beta_1^{\alpha_2}} - \frac{1}{\beta_1^{\alpha_2}} \right) t_{i+1}^{\alpha_2} \right] - \left[\frac{1}{\beta_1^*} t_i \right]^{\alpha_1} \right\} dt_{i+1} & \text{para } 0 \leq t_{i+1}^{\alpha_2} < t_i^{\alpha_1} \end{cases} \\
&= \begin{cases} \exp \left\{ - \left(\left(\frac{t_i}{\beta_2} \right)^{\alpha_1} - \left(\frac{t_i}{\beta_2^*} \right)^{\alpha_1} \right) \right\} \int_0^\infty \frac{\alpha_2}{\beta_2^{*\alpha_2}} t_{i+1}^{\alpha_2-1} \exp \left\{ - \left[\frac{t_{i+1}}{\beta_2^*} \right]^{\alpha_2} \right\} dt_{i+1} & \text{para } 0 \leq t_i^{\alpha_1} < t_{i+1}^{\alpha_2} \\ \frac{\beta_1^{\alpha_1}}{\beta_1^{*\alpha_1} \beta_2^{\alpha_2}} \exp \left\{ - \left[\frac{t_i}{\beta_1^*} \right]^{\alpha_1} + \left[\frac{t_i}{\beta_1} \right]^{\alpha_1} \right\} \times \\ \int_0^\infty \left(\frac{1}{\beta_1^{\alpha_2}} + \frac{1}{\beta_2^{\alpha_2}} - \frac{1}{\beta_1^{*\alpha_2}} \right)^{-1} \left(\frac{1}{\beta_1^{\alpha_2}} + \frac{1}{\beta_2^{\alpha_2}} - \frac{1}{\beta_1^{*\alpha_2}} \right) \alpha_2 t_{i+1}^{\alpha_2-1} \times \\ \exp \left\{ - \left(\frac{1}{\beta_1^{\alpha_2}} + \frac{1}{\beta_2^{\alpha_2}} - \frac{1}{\beta_1^{*\alpha_2}} \right) t_{i+1}^{\alpha_2} \right\} dt_{i+1} & \text{para } 0 \leq t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases}
\end{aligned}$$

Pode-se observar que quando $t_i^{\alpha_1} < t_{i+1}^{\alpha_2}$ temos uma esperança conhecida da distribuição Weibull. Assim tem-se o seguinte resultado:

$$\begin{aligned}
&= \begin{cases} \exp \left\{ - \left(\left(\frac{t_i}{\beta_2} \right)^{\alpha_1} - \left(\frac{t_i}{\beta_2^*} \right)^{\alpha_1} \right) \right\} \beta_2^* \Gamma \left(1 + \left(\frac{1}{\alpha_2} \right) \right) & \text{para } 0 \leq t_i^{\alpha_1} < t_{i+1}^{\alpha_2} \\ \frac{\beta_1^{\alpha_1}}{\beta_1^{*\alpha_1} \beta_2^{\alpha_2}} \left(\frac{1}{\beta_1^{\alpha_2}} + \frac{1}{\beta_2^{\alpha_2}} - \frac{1}{\beta_1^{*\alpha_2}} \right)^{-1} \exp \left\{ - \left[\frac{t_i}{\beta_1^*} \right]^{\alpha_1} + \left[\frac{t_i}{\beta_1} \right]^{\alpha_1} \right\} \times \\ \int_0^\infty \frac{\alpha_2}{\left(\alpha_2 \sqrt{\left(\frac{1}{\beta_1^{\alpha_2}} + \frac{1}{\beta_2^{\alpha_2}} - \frac{1}{\beta_1^{*\alpha_2}} \right)^{-1}} \right)^{\alpha_2}} t_{i+1}^{\alpha_2-1} \times \\ \exp \left\{ - \left(\frac{t_{i+1}}{\alpha_2 \sqrt{\left(\frac{1}{\beta_1^{\alpha_2}} + \frac{1}{\beta_2^{\alpha_2}} - \frac{1}{\beta_1^{*\alpha_2}} \right)^{-1}}} \right)^{\alpha_2} \right\} dt_{i+1} & \text{para } 0 \leq t_{i+1}^{\alpha_2} < t_i^{\alpha_1} \end{cases}
\end{aligned}$$

Agora também é possível observar que quando $t_{i+1}^{\alpha_2} < t_i^{\alpha_1}$ também temos uma esperança conhecida da distribuição Weibull. Assim tem-se o seguinte resultado:

$$\begin{aligned}
&= \begin{cases} \exp \left\{ - \left(\left(\frac{t_i}{\beta_2} \right)^{\alpha_1} - \left(\frac{t_i}{\beta_2^*} \right)^{\alpha_1} \right) \right\} \beta_2^* \Gamma \left(1 + \left(\frac{1}{\alpha_2} \right) \right) & \text{para } 0 \leq t_i^{\alpha_1} < t_{i+1}^{\alpha_2} \\ \frac{\beta_1^{\alpha_1}}{\beta_1^{*\alpha_1} \beta_2^{\alpha_2}} \left(\frac{1}{\beta_1^{\alpha_2}} + \frac{1}{\beta_2^{\alpha_2}} - \frac{1}{\beta_1^{*\alpha_2}} \right)^{-1} \exp \left\{ - \left[\frac{t_i}{\beta_1^*} \right]^{\alpha_1} + \left[\frac{t_i}{\beta_1} \right]^{\alpha_1} \right\} \times \\ \alpha_2 \sqrt{\left(\frac{1}{\beta_1^{\alpha_2}} + \frac{1}{\beta_2^{\alpha_2}} - \frac{1}{\beta_1^{*\alpha_2}} \right)^{-1}} \Gamma \left(1 + \left(\frac{1}{\alpha_2} \right) \right) & \text{para } 0 \leq t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases} \quad (6.19)
\end{aligned}$$

Portanto conhecendo os parâmetros pela estimação feita, pode-se calcular essas esperanças. Na próxima seção será construído um novo modelo markoviano, considerando distribuição bivariada Weibull de Marshal-Olkin.

6.2 Modelo Markoviano Weibull via Distribuição Bivariada Weibull de Marshall-Olkin

O próximo modelo a ser construído é o modelo markoviano Weibull, construído a partir da distribuição bivariada Weibull proposta por Marshall e Olkin (1967) [21]. Será mantida a suposição de que os dados de geração de energia solar por placas fotovoltaica são identicamente distribuídos com distribuição $Weibull(\alpha, \beta)$, e sob a suposição de que eles seguem um processo de Markov, ou seja, é proposto a criação de funções de densidades condicionais de T_{i+1} dado T_i . A construção a seguir é baseada na formulação do modelo Weibull Bivariado de Marshall e Olkin visto em Alvaro (2019) [5].

6.2.1 Distribuição Weibull Bivariada de Marshall-Olkin

A DBW de Marshall e Olkin é obtida por meio de uma transformação de variáveis da DEB, onde esta distribuição não é absolutamente contínua sendo $P(T_1 = T_2) > 0$, ou seja, a ocorrência do evento de interesse T_{i+1} e T_i pode ser por causas simultâneas. Primeiramente na construção desta distribuição bivariada exponencial, Marshall e Olkin (1967) consideram um modelo de choque fatal e outro de choque não fatal. No caso do modelo de choque fatal, suponha que os componentes de um sistema com dois componentes falham após receber um choque que é sempre fatal. Suponha ainda que processos de Poisson independentes $Z_1(t; \lambda_1)$; $Z_2(t; \lambda_2)$ e $Z_{12}(t; \lambda_{12})$ governam a ocorrência de choques, onde $Z(t; \lambda) \equiv \{z(t) \geq 0, \lambda\}$ representa um processo de Poisson homogêneo com parâmetro λ . Os eventos no processo $Z_1(t; \lambda_1)$ são os choques no componente 1, eventos no processo $Z_2(t; \lambda_2)$ são os choques no componente 2 e $Z_{12}(t; \lambda_{12})$ são os choques em ambos os componentes. Portanto, se T_1 e T_2 denotam as variáveis aleatórias que representam os tempos de vida do primeiro e do segundo componentes, respectivamente, e levando em consideração que $Z_1(t; \lambda_1)$; $Z_2(t; \lambda_2)$ e $Z_{12}(t; \lambda_{12})$ são independentes, a função de sobrevivência conjunta entre T_1 e T_2 é dada como:

$$\begin{aligned} S(t_1, t_2) &= P(T_1 > t_1, T_2 > t_2) = P\{Z_1(s; \lambda_1) = 0, Z_2(t, \lambda_2) = 0, Z_{12}(\max(s, t); \lambda_{12}) = 0\}, \\ &= P\{Z_1(s; \lambda_1) = 0\} \{Z_2(t, \lambda_2) = 0\} \{Z_{12}(\max(s, t) = 0)\}, \\ &= \exp\{-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_{12} \max(t_1, t_2)\}. \end{aligned}$$

No segundo caso, no modelo de choque não fatal, considere novamente três processos de Poisson independentes, $Z_1(t; \delta_1)$; $Z_2(t; \delta_2)$ e $Z_{12}(t; \delta_{12})$, governando a ocorrência de choques com a modificação de que estes choques não necessitam ser fatais. Os estados dos sistemas podem ser descritos pelos pares ordenados, $(0, 0)$; $(0, 1)$; $(1, 0)$ e $(1, 1)$, onde 1 no primeiro elemento do par ordenado indica que o componente está operante e 0 que ele já falhou, e analogamente para o segundo elemento do par ordenado. Os eventos no processo $Z_1(t; \delta_1)$ são choques no primeiro componente que causam uma transição do estado $(1, 1)$ para $(0, 1)$ com probabilidade p_1 e de $(1, 1)$ para $(1, 1)$ com probabilidade $(1 - p_1)$. Similarmente, os eventos no processo $Z_2(t; \delta_2)$ causam uma transição do estado $(1, 1)$ para $(0, 1)$ ou $(1, 1)$ com probabilidades p_2 e $(1 - p_2)$, respectivamente. Eventos no processo $Z_{12}(t; \delta_{12})$ causam uma transição do estado $(1, 1)$ para $(0, 0)$; $(1, 0)$; $(0, 1)$ ou $(1, 1)$ com probabilidades p_{00} ; p_{10} ; p_{01} e p_{11} respectivamente. Como $Z_1(t; \delta_1)$; $Z_2(t; \delta_2)$ e $Z_{12}(t; \delta_{12})$ são independentes e com incrementos independentes, para $t_2 \geq t_1 \geq 0$, a função de sobrevivência conjunta pode ser escrita como:

$$\begin{aligned}
S(t_1, t_2) &= P(T_1 > t_1, T_2 > t_2) = P\{Z_1(s; \delta_1) = 0, Z_2(t, \delta_2) = 0, Z_{12}(\max(s, t); \delta_{12}) = 0\}, \\
&= \left\{ \sum_{k=0}^{\infty} e^{-\delta_1 t_1} \frac{(\delta_1 t_1)^k}{k!} (1 - p_1)^k \right\} \left\{ \sum_{l=0}^{\infty} e^{-\delta_2 t_2} \frac{(\delta_2 t_2)^l}{l!} (1 - p_2)^l \right\} \\
&\quad \left\{ \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} [e^{-\delta_{12} t_1} \frac{(\delta_{12} t_1)^m}{m!} (p_{11})^k] [e^{-\delta_{12} (t_2 - t_1)} \frac{(\delta_{12} (t_2 - t_1))^n}{n!} (p_{11} + p_{01})^k] \right\} \\
&= \exp\{-t_1[\delta_1 p_1 + \delta_{12} p_{01}] - t_2[\delta_2 p_2 + \delta_{12}(1 - p_{11} - p_{01})]\}
\end{aligned}$$

por simetria, para $t_1 \geq t_2 \geq 0$

$$S(t_1, t_2) = P(T_1 > t_1, T_2 > t_2) = \exp\{-t_1[\delta_1 p_1 + \delta_{12}(1 - p_{11} - p_{10})] - t_2[\delta_2 p_2 + \delta_{12} p_{10}]\}$$

Portanto: $S(t_1, t_2) = \exp\{-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_{12} \max(t_1, t_2)\}$,

onde: $\lambda_1 = \delta_1 p_1 + \delta_{12} p_{01}$, $\lambda_2 = \delta_2 p_2 + \delta_{12} p_{10}$ e $\lambda_{12} = \delta_{12} p_{00}$.

As funções de sobrevivência marginais de T_1 e de T_2 são dadas por:

$$S_{T_1}(t_1) = \exp\{-(\lambda_1 + \lambda_{12} t_1)\} \text{ e } S_{T_2}(t_2) = \exp\{-(\lambda_2 + \lambda_{12} t_2)\},$$

a função de densidade da DEB é:

$$f(t_1, t_2) = \frac{\partial^2 S(t_1, t_2)}{\partial t_1 \partial t_2} = \begin{cases} (\lambda_1 + \lambda_{12}) \lambda_2 \exp\{-(\lambda_1 + \lambda_{12}) t_1 - \lambda_2 t_2\}, & \text{para } t_1 \geq t_2 \\ (\lambda_2 + \lambda_{12}) \lambda_2 \exp\{-\lambda_1 t_1 - (\lambda_2 + \lambda_{12}) t_2\}, & \text{para } t_1 < t_2, \end{cases}$$

e as densidades marginais exponenciais são: $f_{T_1}(t_1) = (\lambda_1 + \lambda_{12}) \exp\{-(\lambda_1 + \lambda_{12} t_1)\}$ e $f_{T_2}(t_2) = (\lambda_2 + \lambda_{12}) \exp\{-(\lambda_2 + \lambda_{12} t_2)\}$

Esta distribuição não é absolutamente contínua, ou seja, sua função de sobrevivência pode ser decomposta em uma parte singular e a outra parte absolutamente contínua.

Seja T com distribuição Weibull dada pela seguinte forma:

$$f(t; \alpha, \lambda) = \alpha \lambda t^{\alpha-1} \exp\{-\lambda t^\alpha\}.$$

A construção do modelo Weibull bivaridada de Marshall e Olkin é obtida por meio de uma transformação de variáveis, ou seja se (T_1, T_2) é DEB então $(T_1^{\frac{1}{\alpha_1}}, T_2^{\frac{1}{\alpha_2}})$ é uma DBW.

Assim a função de densidade de probabilidade é dada por:

$$f(t_1, t_2) = \begin{cases} (\lambda_1 + \lambda_{12}) \lambda_2 \alpha_1 t_1^{\alpha_1-1} \alpha_2 t_2^{\alpha_2-1} \exp\{-(\lambda_1 + \lambda_{12}) t_1^{\alpha_1} - \lambda_2 t_2^{\alpha_2}\}, & \text{para } t_1^{\alpha_1} \geq t_2^{\alpha_2} \\ (\lambda_2 + \lambda_{12}) \lambda_1 \alpha_1 t_1^{\alpha_1-1} \alpha_2 t_2^{\alpha_2-1} \exp\{-(\lambda_2 + \lambda_{12}) t_2^{\alpha_2} - \lambda_1 t_1^{\alpha_1}\}, & \text{para } t_1^{\alpha_1} < t_2^{\alpha_2}. \end{cases}$$

Para facilitar o uso deste último resultado temos:

$$f(t_2, t_1) = \begin{cases} (\lambda_2 + \lambda_{12})\lambda_1\alpha_1 t_1^{\alpha_1-1} \alpha_2 t_2^{\alpha_2-1} \exp\{-(\lambda_2 + \lambda_{12})t_2^{\alpha_2} - \lambda_1 t_1^{\alpha_1}\}, & \text{para } t_2^{\alpha_2} \geq t_1^{\alpha_1} \\ (\lambda_1 + \lambda_{12})\lambda_2\alpha_1 t_1^{\alpha_1-1} \alpha_2 t_2^{\alpha_2-1} \exp\{-(\lambda_1 + \lambda_{12})t_1^{\alpha_1} - \lambda_2 t_2^{\alpha_2}\}, & \text{para } t_2^{\alpha_2} < t_1^{\alpha_1}. \end{cases} \quad (6.20)$$

Desta forma, a equação (6.20) é a DWB de Marshall e Olkin. A seguir temos a construção do modelo Markoviano Weibull para que posteriormente se tenha a estimação dos parâmetros.

6.2.2 Modelo Markoviano Weibull via a Distribuição Weibull Bivariada de Marshall e Olkin

O primeiro passo para definir a função de densidade condicional de T_{i+1} dado T_i é utilizar o resultado da equação (6.20), assim será considerada a seguinte reparametrização: $\lambda_1 = \frac{1}{\beta_1^{\alpha_1}}$, $\lambda_2 = \frac{1}{\beta_2^{\alpha_2}}$ e $\lambda_{12} = \theta$, assim:

$$f(t_{i+1}, t_i) = \begin{cases} \left(\frac{1}{\beta_2^{\alpha_2}} + \theta\right) \alpha_2 t_{i+1}^{\alpha_2-1} \frac{1}{\beta_1^{\alpha_1}} \alpha_1 t_i^{\alpha_1-1} \exp\left\{-\frac{1}{\beta_1^{\alpha_1}} t_i^{\alpha_1} - \left(\frac{1}{\beta_2^{\alpha_2}} + \theta\right) t_{i+1}^{\alpha_2}\right\}, & \text{para } t_{i+1}^{\alpha_2} \geq t_i^{\alpha_1} \\ \left(\frac{1}{\beta_1^{\alpha_1}} + \theta\right) \alpha_1 t_i^{\alpha_1-1} \frac{1}{\beta_2^{\alpha_2}} \alpha_2 t_{i+1}^{\alpha_2-1} \exp\left\{-\frac{1}{\beta_2^{\alpha_2}} t_{i+1}^{\alpha_2} - \left(\frac{1}{\beta_1^{\alpha_1}} + \theta\right) t_i^{\alpha_1}\right\}, & \text{para } t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases} \quad (6.21)$$

Portanto a função de densidade condicional de T_{i+1} dado T_i é dada pela expressão:

$$f(t_{i+1}/t_i) = \frac{f(t_{i+1}, t_i)}{f(t_i)}$$

e pelas equações (5.1) e (6.21) tem-se o seguinte resultado para a densidade condicional:

$$f(t_{i+1}/t_i) = \begin{cases} \left(\frac{1}{\beta_2^{\alpha_2}} + \theta\right) \alpha_2 t_{i+1}^{\alpha_2-1} \exp\left\{-\left(\frac{1}{\beta_2^{\alpha_2}} + \theta\right) t_{i+1}^{\alpha_2}\right\}, & \text{para } t_{i+1}^{\alpha_2} \geq t_i^{\alpha_1} \\ \left(\frac{1}{\beta_1^{\alpha_1}} + \theta\right) \frac{1}{\beta_2^{\alpha_2}} \alpha_2 t_{i+1}^{\alpha_2-1} \exp\left\{-\frac{1}{\beta_2^{\alpha_2}} t_{i+1}^{\alpha_2} - \theta t_i^{\alpha_1}\right\} \left(\frac{1}{\beta_1^{\alpha_1}}\right)^{-1}, & \text{para } t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases} \quad (6.22)$$

A função de distribuição acumulada condicional T_{i+1} dado T_i é dada pela expressão:

$$F(t_{i+1}/t_i) = \int_0^{t_{i+1}} f(t_{i+1}/t_i) dt_{i+1}$$

$$F(t_{i+1}/t_i) = \begin{cases} \int_0^{t_{i+1}} \left(\frac{1}{\beta_2^{\alpha_2}} + \theta \right) \alpha_2 t_{i+1}^{\alpha_2-1} \exp \left\{ - \left(\frac{1}{\beta_2^{\alpha_2}} + \theta \right) t_{i+1}^{\alpha_2} \right\} dt_{i+1}, & \text{para } t_{i+1}^{\alpha_2} \geq t_i^{\alpha_1} \\ \int_{t_{i+1}}^{\infty} \left(\frac{1}{\beta_1^{\alpha_1}} + \theta \right) \frac{1}{\beta_2^{\alpha_2}} \alpha_2 t_{i+1}^{\alpha_2-1} \exp \left\{ - \frac{1}{\beta_2^{\alpha_2}} t_{i+1}^{\alpha_2} - \theta t_i^{\alpha_1} \right\} \left(\frac{1}{\beta_1^{\alpha_1}} \right)^{-1} dt_{i+1}, & \text{para } t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases}$$

Assim para $t_{i+1}^{\alpha_2} \geq t_i^{\alpha_1}$ temos uma integral conhecida, sendo a acumulada da distribuição Weibull,

$$= \begin{cases} 1 - \exp \left\{ - \left(\frac{1}{\beta_2^{\alpha_2}} + \theta \right) t_{i+1}^{\alpha_2} \right\}, & \text{para } t_{i+1}^{\alpha_2} \geq t_i^{\alpha_1} \\ \left(\frac{1}{\beta_1^{\alpha_1}} + \theta \right) \exp \{ -\theta t_i^{\alpha_1} \} \beta_1^{\alpha_1} \int_{t_{i+1}}^{\infty} \frac{1}{\beta_2^{\alpha_2}} \alpha_2 t_{i+1}^{\alpha_2-1} \exp \left\{ - \left(\frac{t_{i+1}}{\beta_2} \right)^{\alpha_2} \right\} dt_{i+1}, & \text{para } t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases}$$

Assim para $t_{i+1}^{\alpha_2} < t_i^{\alpha_1}$ uma integral conhecida, sendo a acumulada da distribuição Weibull, portando a função acumulada condicional de T_{i+1} dado T_i é dada por:

$$F(t_{i+1}/t_i) = \begin{cases} 1 - \exp \left\{ - \left(\frac{1}{\beta_2^{\alpha_2}} + \theta \right) t_{i+1}^{\alpha_2} \right\}, & \text{para } t_{i+1}^{\alpha_2} \geq t_i^{\alpha_1} \\ \left(\frac{1}{\beta_1^{\alpha_1}} + \theta \right) \exp \{ -\theta t_i^{\alpha_1} \} \beta_1^{\alpha_1} \left(1 - \exp \left\{ - \left(\frac{t_{i+1}}{\beta_2} \right)^{\alpha_2} \right\} \right), & \text{para } t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases} \quad (6.23)$$

Com as funções definidas tem-se a seguir as estimações de seus devidos parâmetros.

6.2.3 Estimação de Máxima Verossimilhança do Modelo Markoviano Weibull via a Distribuição Weibull Bivariada de Marshall e Olkin

Com a suposição de que os tempos seguem um processo de Markov o que indica que o tempo atual só depende do anterior portanto T_i , $i = 1, 2, \dots, n$ são dependentes. Assim, tem-se que a expressão (2.4) para um processo de Markov, a função de verossimilhança pode ser escrita como:

$$L(\boldsymbol{\theta}; T) = f(t_1) \prod_{i=1}^{n-1} f(t_{i+1}/t_i | \boldsymbol{\theta}). \quad (6.24)$$

Assim substituindo as expressões (5.1) e (6.22), em (6.24,) temos a função de verossimilhança dado por $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_1^*, \beta_2^*)$:

$$\begin{aligned}
L(\boldsymbol{\theta}) &= \frac{\alpha_1}{\beta_1^{\alpha_1}} t_1^{\alpha_1-1} \exp \left\{ - \left(\frac{t_1}{\beta_1} \right)^{\alpha_1} \right\} \prod_{i=1}^{n-1} [f(t_{i+1}/t_i | \boldsymbol{\theta})] \\
&= \frac{\alpha_1}{\beta_1^{\alpha_1}} t_1^{\alpha_1-1} \exp \left\{ - \left(\frac{t_1}{\beta_1} \right)^{\alpha_1} \right\} \\
&\times \begin{cases} \prod_{i=1}^{n-1} \left(\frac{1}{\beta_2^{\alpha_2}} + \theta \right) \alpha_2 t_{i+1}^{\alpha_2-1} \exp \left\{ - \left(\frac{1}{\beta_2^{\alpha_2}} + \theta \right) t_{i+1}^{\alpha_2} \right\}, & \text{para } t_{i+1}^{\alpha_2} \geq t_i^{\alpha_1} \\ \prod_{i=1}^{n-1} \left(\frac{1}{\beta_1^{\alpha_1}} + \theta \right) \frac{1}{\beta_2^{\alpha_2}} \alpha_2 t_{i+1}^{\alpha_2-1} \exp \left\{ - \frac{1}{\beta_2^{\alpha_2}} t_{i+1}^{\alpha_2} - \theta t_i^{\alpha_1} \right\} \left(\frac{1}{\beta_1^{\alpha_1}} \right)^{-1}, & \text{para } t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases}
\end{aligned}$$

Simplificando a expressão do logaritmo na função temos a seguinte equação:

$$\begin{aligned}
\ln(L(\boldsymbol{\theta})) &= \ln(\alpha_1) - \alpha_1 \ln(\beta_1) + (\alpha_1 - 1) \ln(t_1) - \left(\frac{t_1}{\beta_1} \right)^{\alpha_1} + \\
&+ \begin{cases} (n-1) \ln \left(\frac{1}{\beta_2^{\alpha_2}} + \theta \right) + (n-1) \ln(\alpha_2) + (\alpha_2 - 1) \sum_{i=1}^{n-1} \ln(t_{i+1}) + \\ \sum_{i=1}^{n-1} \left(- \left(\frac{1}{\beta_2^{\alpha_2}} + \theta \right) t_{i+1}^{\alpha_2} \right), & \text{para } t_{i+1}^{\alpha_2} \geq t_i^{\alpha_1} \\ (n-1) \ln \left(\frac{1}{\beta_1^{\alpha_1}} + \theta \right) - (n-1)(\alpha_2) \ln(\beta_2) + (n-1) \ln(\alpha_2) + \\ (\alpha_2 - 1) \sum_{i=1}^{n-1} \ln(t_{i+1}) + (n-1) \alpha_1 \ln(\beta_1) + \sum_{i=1}^{n-1} \left(- \frac{1}{\beta_2^{\alpha_2}} t_{i+1}^{\alpha_2} - \theta t_i^{\alpha_1} \right) & \text{para } t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases}
\end{aligned} \tag{6.25}$$

Para encontrar os estimadores de máxima verossimilhança para $(\alpha_1, \alpha_2, \beta_1, \beta_2$ e $\theta)$, será usado o programa R com a utilização do pacote Optim. Pode-se ver os resultados obtidos na Tabela (6.3) e na Figura (6.2) tem-se o ajuste da função acumulada ao nosso conjunto de dados.

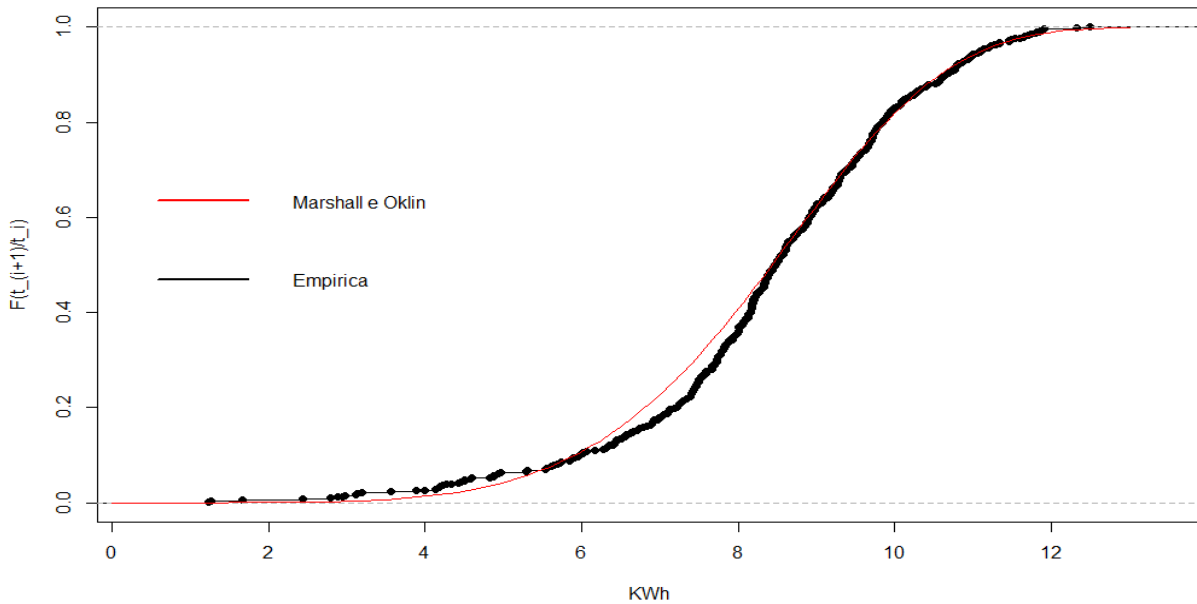
Tabela 6.3: Parâmetros estimados pelo EMV para os parâmetros do Modelo markoviano via a distribuição bivariada Weibull de Marshall e Olkin

Parâmetros	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\theta}$
EMV	0.094	5.310	11.409	11.476	$6.03 * 10^{-6}$
(Erro Padrão)	(0.046)	(0.041)	(0.144)	(0.457)	(0.001)

Após a obtenção dos estimadores é possível então verificar o ajuste do modelo comparando se a função acumulada do modelo e os dados empíricos. Estas curvas são apresentadas na Figura (6.2).

Pela análise gráfica é possível observar que a EMV obteve um bom resultado para os parâmetros da função de distribuição acumulada no Modelo markoviano via a distribuição bivariada Weibull de Marshall e Olkin, resultando assim em um bom ajuste aos dados reais.

Figura 6.2: Curva da função de distribuição acumulada do Modelo markoviano via a distribuição bivariada Weibull de Marshall e Olkin versus a curva de sua função de distribuição acumulada Empírica para seu respectivos parâmetros estimados



Outro resultado possível de se obter é a estimação da média esperada de energia solar gerada por placas fotovoltaicas nos próximos dias em função do número observado entre o dia anterior e o dia atual, isto é, a esperança de T_{i+1} dado $T_i = t_i$.

$$E[T_{i+1}/T_i = t_i] = \int_0^{\infty} t_{i+1} [f(t_{i+1}/t_i|\boldsymbol{\theta})] dt_{i+1}$$

$$= \begin{cases} \int_0^{\infty} t_{i+1} \left(\frac{1}{\beta_2^{\alpha_2}} + \theta \right) \alpha_2 t_{i+1}^{\alpha_2-1} \exp \left\{ - \left(\frac{1}{\beta_2^{\alpha_2}} + \theta \right) t_{i+1}^{\alpha_2} \right\} dt_{i+1} & \text{para } t_{i+1}^{\alpha_2} \geq t_i^{\alpha_1} \\ \int_0^{\infty} t_{i+1} \left(\frac{1}{\beta_1^{\alpha_1}} + \theta \right) \frac{1}{\beta_2^{\alpha_2}} \alpha_2 t_{i+1}^{\alpha_2-1} \exp \left\{ - \frac{1}{\beta_2^{\alpha_2}} t_{i+1}^{\alpha_2} - \theta t_i^{\alpha_1} \right\} \left(\frac{1}{\beta_1^{\alpha_1}} \right)^{-1} dt_{i+1} & \text{para } t_{i+1}^{\alpha_2} < t_i^{\alpha_1} \end{cases}$$

$$= \begin{cases} \int_0^\infty \frac{\alpha_2}{\left(\sqrt[\alpha_2]{\left(\frac{1}{\beta_2^{\alpha_2}} + \theta\right)^{-1}}\right)^{\alpha_2}} t_{i+1}^{\alpha_2} \exp\left\{-\left(\frac{t_{i+1}}{\sqrt[\alpha_2]{\left(\frac{1}{\beta_2^{\alpha_2}} + \theta\right)^{-1}}}\right)^{\alpha_2}\right\} dt_{i+1} & \text{para } t_{i+1}^{\alpha_2} \geq t_i^{\alpha_1} \\ \left(\frac{1}{\beta_1^{\alpha_1}} + \theta\right) \exp\{-\theta t_i^{\alpha_1}\} \beta_1^{\alpha_1} \int_0^\infty \frac{\alpha_2}{\beta_2^{\alpha_2}} t_{i+1}^{\alpha_2} \exp\left\{-\left(\frac{t_{i+1}}{\beta_2}\right)^{\alpha_2}\right\} dt_{i+1} & \text{para } t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases}$$

Pode-se observar que em ambas as expressões tem-se uma esperança conhecida da distribuição Weibull. Assim o resultado é dado por:

$$= \begin{cases} \sqrt[\alpha_2]{\left(\frac{1}{\beta_2^{\alpha_2}} + \theta\right)^{-1}} \Gamma\left(1 + \left(\frac{1}{\alpha_2}\right)\right) & \text{para } t_{i+1}^{\alpha_2} \geq t_i^{\alpha_1} \\ \left(\frac{1}{\beta_1^{\alpha_1}} + \theta\right) \exp\{-\theta t_i^{\alpha_1}\} \beta_1^{\alpha_1} \beta_2 \Gamma\left(1 + \left(\frac{1}{\alpha_2}\right)\right) & \text{para } t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases} \quad (6.26)$$

Portanto conhecendo os parâmetros pela estimação feita, pode-se calcular essas esperanças.

6.3 Modelo Markoviano Weibull via Distribuição Bivariada Weibull Exponenciada Kumaraswamy

O terceiro modelo markoviano Weibull a ser construído é a partir da distribuição bivariada Weibull Exponenciada Kumaraswamy apresentada em Saboor, et al. (2016) [31].

6.3.1 Bivariada Weibull Exponenciada Kumaraswamy

Em Saboor, et al.(2016) a construção é feita diferente do que vimos em Freund e Marshall-Olkin, aqui são definidas três distribuições EKwW independentes, onde que a partir delas é usadas as definições de máximo e mínimo para que se possa encontrar a função acumulada conjunta. Assim, seja $U_1 \sim \text{EKwW}(\alpha, \beta, a, b, c_1)$, $U_2 \sim \text{EKwW}(\alpha, \beta, a, b, c_2)$, e $U_3 \sim \text{EKwW}(\alpha, \beta, a, b, c_3)$ identicamente distribuídas, definimos: $Y_1 = \max(U_1, U_2)$ e $Y_2 = \max(U_2, U_3)$. Portanto um vetor bivariado de $(Y_1, Y_2) \sim \text{EKwW}(\alpha, \beta, a, b, c_1, c_2, c_3)$

Com base nos resultados apresentados, seguindo a suposição que $U_i, i = 1, 2, 3$ são independentes, logo:

$$\begin{aligned} F(y_1, y_2) &= P(Y_1 \leq y_1, Y_2 \leq y_2) \\ &= P(\max(U_1, U_2) \leq y_1, \max(U_2, U_3) \leq y_2) \\ &= P(U_1 \leq y_1, U_3 \leq y_1, U_2 \leq y_2, U_3 \leq \min(y_1, y_2)) \end{aligned}$$

Como $U_i, i = 1, 2, 3$ são independentes e pela equação (5.46) segue então:

$$\begin{aligned} F(y_1, y_2) &= (U_1 \leq y_1, U_2 \leq y_2, U_3 \leq \min(y_1, y_2)) \\ &= F(\alpha, \beta, a, b, c_1)F(\alpha, \beta, a, b, c_2)F(\alpha, \beta, a, b, c_3) \\ &= \{1 - [1 - (1 - \exp\{-\frac{y_1}{\beta}\}^\alpha)^a]^b\}^{c_1} \{1 - [1 - (1 - \exp\{-\frac{y_2}{\beta}\}^\alpha)^a]^b\}^{c_2} \{1 - [1 - (1 - \exp\{-\frac{z}{\beta}\}^\alpha)^a]^b\}^{c_3}, \end{aligned}$$

onde $z = \min(y_1, y_2)$.

Combinando o resultado de (5.46), com o obtido na expressão a cima tem-se a Função acumulada conjunta de EKwW dada por:

$$F(y_1, y_2) = \begin{cases} 1 - [1 - (1 - \exp\{-\frac{y_1}{\beta}\}^\alpha)^a]^b\}^{(c_1+c_3)} \times \\ \quad 1 - [1 - (1 - \exp\{-\frac{y_2}{\beta}\}^\alpha)^a]^b\}^{c_2} \text{ para } y_1 \leq y_2 \\ 1 - [1 - (1 - \exp\{-\frac{y_2}{\beta}\}^\alpha)^a]^b\}^{(c_2+c_3)} \times \\ \quad 1 - [1 - (1 - \exp\{-\frac{y_1}{\beta}\}^\alpha)^a]^b\}^{c_1} \text{ para } y_2 \leq y_1 \\ 1 - [1 - (1 - \exp\{-\frac{y_1}{\beta}\}^\alpha)^a]^b\}^{(c_1+c_2+c_3)} \text{ para } y_2 = y_1 \end{cases} \quad (6.27)$$

Com a Função Acumulada dada na equação (6.27) é possível obter a densidade conjunta de EKwW, realizando as derivadas parciais de segunda ordem.

$$f(y_1, y_2) = \frac{\partial^2 F(y_1, y_2)}{\partial y_1 \partial y_2}$$

$$f(y_1, y_2) = \begin{cases} \beta^{-2} \alpha^2 a^2 b^2 c_2 (c_1 + c_3) \exp\{-\frac{y_1}{\beta}\}^\alpha \exp\{-\frac{y_2}{\beta}\}^\alpha (\frac{y_1}{\beta})^{\alpha-1} \times \\ \quad (\frac{y_2}{\beta})^{\alpha-1} (1 - \exp\{-\frac{y_1}{\beta}\}^\alpha)^{a-1} (1 - \exp\{-\frac{y_2}{\beta}\}^\alpha)^{a-1} \times \\ \quad (1 - (1 - \exp\{-\frac{y_1}{\beta}\}^\alpha)^a)^{b-1} (1 - (1 - \exp\{-\frac{y_2}{\beta}\}^\alpha)^a)^{b-1} \times \\ \quad (1 - (1 - (1 - \exp\{-\frac{y_2}{\beta}\}^\alpha)^a)^b)^{c_2-1} (1 - (1 - (1 - \exp\{-\frac{y_1}{\beta}\}^\alpha)^a)^b)^{c_1+c_3-1} \text{ para } y_1 \leq y_2 \\ \beta^{-2} \alpha^2 a^2 b^2 c_1 (c_2 + c_3) \exp\{-\frac{y_2}{\beta}\}^\alpha \exp\{-\frac{y_1}{\beta}\}^\alpha (\frac{y_2}{\beta})^{\alpha-1} \times \\ \quad (\frac{y_1}{\beta})^{\alpha-1} (1 - \exp\{-\frac{y_2}{\beta}\}^\alpha)^{a-1} (1 - \exp\{-\frac{y_1}{\beta}\}^\alpha)^{a-1} \times \\ \quad (1 - (1 - \exp\{-\frac{y_2}{\beta}\}^\alpha)^a)^{b-1} (1 - (1 - \exp\{-\frac{y_1}{\beta}\}^\alpha)^a)^{b-1} \times \\ \quad (1 - (1 - (1 - \exp\{-\frac{y_1}{\beta}\}^\alpha)^a)^b)^{c_1-1} (1 - (1 - (1 - \exp\{-\frac{y_2}{\beta}\}^\alpha)^a)^b)^{c_2+c_3-1} \text{ para } y_2 \leq y_1 \\ abc_3 \frac{\alpha}{\beta^\alpha} y^{\alpha-1} \exp\{-\frac{y}{\beta}\} [1 - \exp\{-\frac{y}{\beta}\}^\alpha]^{\alpha-1} \times \\ \quad (1 - (1 - \exp\{-\frac{y}{\beta}\}^\alpha)^a)^{b-1} (1 - (1 - (1 - \exp\{-\frac{y}{\beta}\}^\alpha)^a)^b)^{c_1+c_2+c_3-1} \text{ para } y_2 = y_1 \end{cases} \quad (6.28)$$

Assim, a equação (6.28) é a Distribuição Bivariada Weibull Exponenciada Kumaraswamy. A seguir temos a construção do modelo Markoviano Weibull para que posteriormente se tenha a estimação dos parâmetros.

6.3.2 Modelo Markoviano Weibull via a Distribuição Bivariada Weibull Exponenciada Kumaraswamy

O primeiro passo para a definição da função de densidade condicional de T_{i+1} dado T_i é utilizar o resultado da equação (6.28), assim será considerada a reparametrização substituindo y_1 e y_2 para T_1 e T_{i+1} respectivamente.

$$f(t_i, t_{i+1}) = \left\{ \begin{array}{l} \beta^{-2} \alpha^2 a^2 b^2 c_2 (c_1 + c_3) \exp\{-\frac{t_i}{\beta}\}^\alpha \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha (\frac{t_i}{\beta})^{\alpha-1} \times \\ (\frac{t_{i+1}}{\beta})^{\alpha-1} (1 - \exp\{-\frac{t_i}{\beta}\}^\alpha)^{a-1} (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^{a-1} \times \\ (1 - (1 - \exp\{-\frac{t_i}{\beta}\}^\alpha)^a)^{b-1} (1 - (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^a)^{b-1} \times \\ (1 - (1 - (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^a)^b)^{c_2-1} (1 - (1 - (1 - \exp\{-\frac{t_i}{\beta}\}^\alpha)^a)^b)^{c_1+c_3-1} \text{para } t_i \leq t_{i+1} \\ \beta^{-2} \alpha^2 a^2 b^2 c_1 (c_2 + c_3) \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha \exp\{-\frac{t_i}{\beta}\}^\alpha (\frac{t_{i+1}}{\beta})^{\alpha-1} \times \\ (\frac{t_i}{\beta})^{\alpha-1} (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^{a-1} (1 - \exp\{-\frac{t_i}{\beta}\}^\alpha)^{a-1} \times \\ (1 - (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^a)^{b-1} (1 - (1 - \exp\{-\frac{t_i}{\beta}\}^\alpha)^a)^{b-1} \times \\ (1 - (1 - (1 - \exp\{-\frac{t_i}{\beta}\}^\alpha)^a)^b)^{c_1-1} (1 - (1 - (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^a)^b)^{c_2+c_3-1} \text{para } t_{i+1} \leq t_i \\ abc_3 \frac{\alpha}{\beta^\alpha} t_i^{\alpha-1} \exp\{-\frac{t_i}{\beta}\} [1 - \exp\{-\frac{t_i}{\beta}\}^\alpha]^{\alpha-1} \times \\ (1 - (1 - \exp\{-\frac{t_i}{\beta}\}^\alpha)^a)^{b-1} (1 - (1 - (1 - \exp\{-\frac{t_i}{\beta}\}^\alpha)^a)^b)^{c_1+c_2+c_3-1} \text{para } t_{i+1} = t_i \end{array} \right. \quad (6.29)$$

Portanto a função de densidade condicional de T_{i+1} dado T_i é dada pela expressão:

$$f(t_{i+1}/t_i) = \frac{f(t_{i+1}, t_i)}{f(t_i)},$$

e pelas equações (5.47) e (6.29) tem-se o seguinte resultado para a densidade condicional:

$$f(t_{i+1}/t_i) = \begin{cases} \beta^{-1} \alpha abc_2 (1 + c_3) \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha (\frac{1}{\beta})^{\alpha-1} (\frac{t_{i+1}}{\beta})^{\alpha-1} \times \\ (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^{a-1} (1 - (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^a)^{b-1} \times \\ (1 - (1 - (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^b)^{c_2-1} (1 - (1 - (1 - \exp\{-\frac{t_i}{\beta}\}^\alpha)^b)^{c_3}) \text{ para } t_i \leq t_{i+1} \\ \beta^{-1} \alpha ab (c_2 + c_3) \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha (\frac{t_{i+1}}{\beta})^{\alpha-1} (\frac{1}{\beta})^{\alpha-1} \times \\ (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^{a-1} (1 - (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^a)^{b-1} \times \\ (1 - (1 - (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^a)^b)^{c_2+c_3-1} \text{ para } t_{i+1} \leq t_i \end{cases} \quad (6.30)$$

A função de distribuição acumulada condicional T_{i+1} dado T_i é dada pela expressão:

$$F(t_{i+1}/t_i) = \int_0^{t_{i+1}} f(t_{i+1}/t_i) dt_{i+1}$$

$$= \begin{cases} (1 - (1 - (1 - \exp\{-\frac{t_i}{\beta}\}^\alpha)^b)^{c_3} (1 + c_3) (\frac{1}{\beta})^{\alpha-1} \times \\ \int_0^{t_{i+1}} abc_2 \frac{\alpha}{\beta^\alpha} t_{i+1}^{\alpha-1} \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^{a-1} (1 - (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^a)^{b-1} \times \\ (1 - (1 - (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^b)^{c_2-1} dt_{i+1} \quad \text{para } t_i \leq t_{i+1} \\ (\frac{1}{\beta})^{\alpha-1} \times \\ \int_0^{t_{i+1}} \frac{\alpha ab}{\beta^\alpha} (c_2 + c_3) t_{i+1}^{\alpha-1} \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^{a-1} (1 - (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^a)^{b-1} \times \\ (1 - (1 - (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^a)^b)^{c_2+c_3-1} dt_{i+1} \quad \text{para } t_{i+1} \leq t_i \end{cases}$$

Como as integrais são conhecidas temos que a Função Acumulada condicional de T_{i+1} dado T_i é dada por:

$$F(t_{i+1}/t_i) = \begin{cases} (1 - (1 - (1 - \exp\{-\frac{t_i}{\beta}\}^\alpha)^b)^{c_3} (1 + c_3) (\frac{1}{\beta})^{\alpha-1} \times \\ (1 - (1 - (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^b)^{c_2} \quad \text{para } t_i \leq t_{i+1} \\ (\frac{1}{\beta})^{\alpha-1} (1 - (1 - (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^b)^{c_2+c_3}) \quad \text{para } t_{i+1} \leq t_i \end{cases} \quad (6.31)$$

Com as funções definidas tem-se a seguir as estimações de seus devidos parâmetros.

6.3.3 Estimação de Máxima Verossimilhança do Modelo Markoviano Weibull via a Distribuição Bivariada Weibull Exponenciada Kumaraswamy

Com a suposição de que os tempos seguem um processo de Markov o que indica que o tempo atual só depende do anterior portanto $T_i, i = 1, 2, \dots, n$ são dependentes. Assim tem-se que da expressão (2.4) para um processo de Markov, a função de verossimilhança pode ser escrita como:

$$L(\boldsymbol{\theta}; T) = f(t_1) \prod_{i=1}^{n-1} f(t_{i+1}/t_i | \boldsymbol{\theta}). \quad (6.32)$$

Assim substituindo as expressões (5.47) e (6.30), em (6.32) temos a função de verossimilhança dado por $\boldsymbol{\theta} = (\alpha, \beta, a, b, c_2$ e $c_3)$:

$$\begin{aligned} L(\boldsymbol{\theta}) &= f(t_1) \prod_{i=1}^{n-1} [f(t_{i+1}/t_i | \boldsymbol{\theta})] \\ &= abc \frac{\alpha}{\beta^\alpha} t^{\alpha-1} \exp\left\{-\frac{t}{\beta}\right\}^\alpha [1 - \exp\left\{-\frac{t}{\beta}\right\}^\alpha]^{a-1} \{1 - [1 - \exp\left\{-\frac{t}{\beta}\right\}^\alpha]^a\}^{b-1} \\ &\quad \times \{1 - [1 - (1 - \exp\left\{-\frac{t}{\beta}\right\}^\alpha)^b]\}^{c-1} \\ &\times \left\{ \begin{array}{l} \prod_{i=1}^{n-1} \beta^{-1} \alpha abc_2 (1 + c_3) \exp\left\{-\frac{t_{i+1}}{\beta}\right\}^\alpha \left(\frac{1}{\beta}\right)^{\alpha-1} \left(\frac{t_{i+1}}{\beta}\right)^{\alpha-1} \times \\ (1 - \exp\left\{-\frac{t_{i+1}}{\beta}\right\}^\alpha)^{a-1} (1 - (1 - \exp\left\{-\frac{t_{i+1}}{\beta}\right\}^\alpha)^a)^{b-1} \times \\ (1 - (1 - (1 - \exp\left\{-\frac{t_{i+1}}{\beta}\right\}^\alpha)^a)^b)^{c_2-1} (1 - (1 - (1 - \exp\left\{-\frac{t_i}{\beta}\right\}^\alpha)^a)^b)^{c_3} \quad \text{para } t_i \leq t_{i+1} \\ \prod_{i=1}^{n-1} \beta^{-1} \alpha ab(c_2 + c_3) \exp\left\{-\frac{t_{i+1}}{\beta}\right\}^\alpha \left(\frac{t_{i+1}}{\beta}\right)^{\alpha-1} \left(\frac{1}{\beta}\right)^{\alpha-1} \times \\ (1 - \exp\left\{-\frac{t_{i+1}}{\beta}\right\}^\alpha)^{a-1} (1 - (1 - \exp\left\{-\frac{t_{i+1}}{\beta}\right\}^\alpha)^a)^{b-1} \times \\ (1 - (1 - (1 - \exp\left\{-\frac{t_{i+1}}{\beta}\right\}^\alpha)^a)^b)^{c_2+c_3-1} \quad \text{para } t_{i+1} \leq t_i \end{array} \right. \end{aligned}$$

Simplificando a expressão do logaritmo na função temos a seguinte equação:

$$\left\{ \begin{array}{l} \ln(L(\boldsymbol{\theta})) = \ln(a) + \ln(b) + \ln(c_1) + \ln(\alpha) - \alpha \ln(\beta) + (\alpha - 1) \ln(t_1) \\ - \left(\frac{t_1}{\beta}\right)^\alpha + (a - 1) \ln[1 - \exp\left\{-\frac{t_1}{\beta}\right\}^\alpha] + (b - 1) \ln\{1 - [1 - \exp\left\{-\frac{t_1}{\beta}\right\}^\alpha]^a\} \\ + (c - 1) \ln\{1 - [1 - (1 - \exp\left\{-\frac{t_1}{\beta}\right\}^\alpha)^a]^b\} \end{array} \right.$$

$$\begin{aligned}
& \left((n-1)(-\beta + \alpha + a + b + c_2 + (1 + c_3) - (\alpha - 1)\beta) \right. \\
& \quad - \sum_{i=1}^{n-1} \left(\frac{t_{i+1}}{\beta}\right)^\alpha + (\alpha - 1) \sum_{i=1}^{n-1} \ln\left(\frac{t_{i+1}}{\beta}\right) + \\
& \quad (a - 1) \sum_{i=1}^{n-1} \ln(1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha) + (b - 1) \sum_{i=1}^{n-1} \ln(1 - (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^a) \\
& \quad + (c_2 - 1) \sum_{i=1}^{n-1} \ln(1 - (1 - (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^a)^b) + \\
& \quad \left. + c_3 \sum_{i=1}^{n-1} \ln(1 - (1 - (1 - \exp\{-\frac{t_i}{\beta}\}^\alpha)^a)^b) \right) \quad \text{para } t_i \leq t_{i+1} \\
& \left((n-1)(-\beta + \alpha + a + b + (c_2 + c_3) - (\alpha - 1)\beta) \right. \\
& \quad - \sum_{i=1}^{n-1} \left(\frac{t_{i+1}}{\beta}\right)^\alpha + (\alpha - 1) \sum_{i=1}^{n-1} \ln\left(\frac{t_{i+1}}{\beta}\right) + \\
& \quad (a - 1) \sum_{i=1}^{n-1} \ln(1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha) + (b - 1) \sum_{i=1}^{n-1} \ln(1 - (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^a) \\
& \quad \left. + (c_2 + c_3 - 1) \sum_{i=1}^{n-1} \ln(1 - (1 - (1 - \exp\{-\frac{t_{i+1}}{\beta}\}^\alpha)^a)^b) \right) \quad \text{para } t_{i+1} \leq t_i.
\end{aligned} \tag{6.33}$$

Para encontrar os estimadores de máxima verossimilhança para $(\alpha, \beta, a, b, c_2$ e $c_3)$, será usado o programa R com a utilização do pacote Optim. Pode-se ver os resultados obtidos na Tabela (6.4) e na Figura (6.3) tem-se o ajuste da função acumulada ao nosso conjunto de dados.

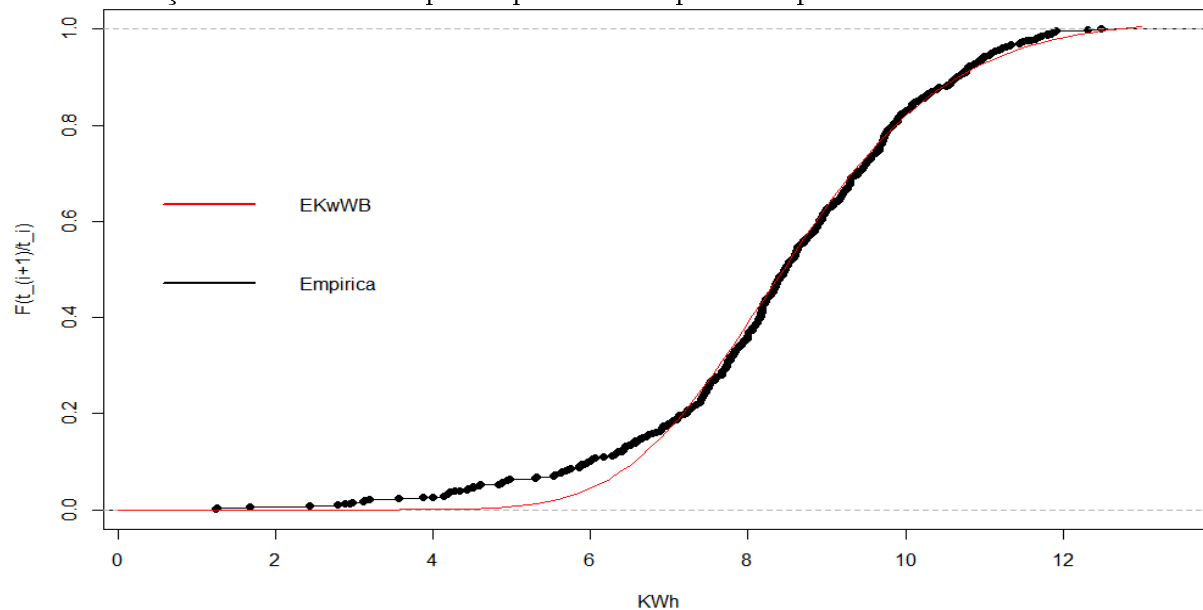
Tabela 6.4: Parâmetros estimados pelo EMV para os parâmetros do Modelo markoviano via a Distribuição Bivariada Weibull Exponenciada Kumaraswamy

Parâmetros	$\hat{\alpha}$	$\hat{\beta}$	\hat{a}	\hat{b}	\hat{c}_2	\hat{c}_3
EMV	0.095	0.509	7.276	7.734	2.202	0.872
(Erro Padrão)	(0.874)	(0.005)	(0.779)	(0.044)	(0.061)	(0.006)

Após a obtenção dos estimadores é possível então verificar o ajuste do modelo comparando a função acumulada do modelo e os dados empíricos. Estas curvas são apresentadas na Figura (6.3).

Pela análise gráfica é possível observar que a EMV obteve um bom resultado para os parâmetros da função de distribuição acumulada no Modelo markoviano via a distribuição Bivariada Weibull Exponenciada Kumaraswamy, resultando assim em um bom ajuste aos dados reais.

Figura 6.3: Curva da função de distribuição acumulada do Modelo markoviano via a Distribuição Bivariada Weibull Exponenciada Kumaraswamy versus a curva de sua função de distribuição acumulada Empírica para seu respectivos parâmetros estimados



Na tabela (6.5) serão apresentados os resultados de AIC, BIC e AICc, confirmando se realmente esse ajuste foi bom e qual modelo é o melhor a se utilizar para os dados da pesquisa.

6.4 Comparação dos modelos univariados e markovianos Weibull via as distribuições bivariada Weibull de Freund, Marshall-Olkin e Weibull Exponenciada Kumaraswamy

Nesta seção serão apresentados os valores de AIC, AICc e BIC, onde, como já visto no capítulo 2 eles apresentam critérios que nos auxiliam na comparação e seleção de modelos estatísticos. Ao se utilizar a função `optim` do R eles nos fornecem o valor do logaritmo da função de verossimilhança e assim dadas as expressões (2.8), (2.9) e (2.10), tem-se os respectivos valores de AIC, AICc e BIC. A escolha do melhor modelo é feita a partir do menor valor encontrado nas três expressões. Na Tabela (6.5) é apresentado os resultados.

Tabela 6.5: Seleção dos modelos

Distribuição	AIC	AIC _c	BIC
Weibull	2130.887	2130.909	2139.422
Gama	2278.409	2278.431	2286.943
EG	2412.200	2412.222	2420.734
Normal	2154.363	2154.385	2162.898
EW	2463.693	2463.738	2476.495
KwW	2181.697	2181.773	2198.766
log-EK	2146.546	2146.661	2163.615
EKwW	2101.399	2101.560	2122.735
Distribuição bivariada Weibull Freund	1302.423	1302.584	1340.0253
Distribuição bivariada Weibull de Marshall-Olkin	1098.851	1099.006	1130.187
Distribuição EKwWB	1105.241	1109.455	1130.843

Analisando a Tabela (6.5), tem-se que pelos os critérios estudados os melhores modelos a serem escolhidos são os modelos markovianos, onde a bivariada Weibull de Marshall-Olkin se destacou comparado aos outros modelos seguido pela Bivariada Weibull Exponenciada Kumaraswamy e bivariada Weibull Freund. Os modelos univariados ficaram com resultados de ajuste inferior a esses três modelos markovianos, apresentando uma sequencia de bons resultados as distribuições EKwW, Weibull, Log-EK, Normal e KwW respectivamente. As outras três não citadas, Gama, EG e EW não tiveram bons resultados comparados com as outras. Este resultado confirma o que já foi visto na análise gráfica nos capítulos 5 e 6.

6.5 Escolha do melhor modelo

Como já foi dito anteriormente, após os ajustes dos modelos tanto no caso univariado, como no caso markoviano é possível verificar graficamente que os dois melhores modelos para o conjunto de dados da pesquisa são o Markoviano via a Distribuição Bivariada Weibull Exponenciada Kumaraswamy e o Modelo Markoviano Weibull via a Distribuição Weibull Bivariada de Marshall e Olkin. Na presente pesquisa os dois modelos tiveram melhores resultados e assim é de escolha do pesquisador qual modelo seria mais útil.

Quando a escolha é feita utilizando os critérios de seleção dos modelos AIC, AIC_c e BIC, o resultado se confirma no que foi visto na análise gráfica. Porém, no BIC a diferença entre esses dois modelos é tão pequena que fica difícil a escolha. Assim, entre as duas opções qual é o melhor para o estudo de energia solar gerada por placas fotovoltaicas da cidade de São José do Rio Preto?

Quando analisamos os gráficos (6.2) e (6.3) em determinados momentos é possível observar uma falta significativa de ajustes, principalmente nos valores baixos em KWh de energia, esses valores são os que apresentam uma menor frequência no conjunto de dados, assim não seria um peso grande para a escolha de um melhor modelo. Já um fator muito importante que justifica a escolha de um determinado modelo levando em consideração todas essas situações impostas é o numero de parâmetros que cada um tem que estimar.

O modelo de Marshall e Olkin tem 5 parâmetros a serem estimados, um a menos em relação ao Weibull Exponenciada Kumaraswamy. Isso implica no tempo de execução do

programa, onde que um modelo com um número menor de parâmetros é mais rápido e eficiente em relação a performance a execução, assim com menos possibilidades de erros. Portanto, a escolha do modelo para a pesquisa é o Markoviano Weibull via a Distribuição Weibull Bivariada de Marshall e Olkin.

Com a escolha do modelo como já foi dito anteriormente é possível de se obter a estimação da média esperada de energia solar gerada por placas fotovoltaicas nos próximos dias em função do número observado entre o dia anterior e o dia atual, isto é, a esperança de T_{i+1} dado $T_i = t_i$, pois temos os valores dos parâmetros estimados. O resultado é apresentado na Tabela (6.6), com os valores obtidos pela equação (6.26)

Tabela 6.6: Estimativa da média esperada de energia solar gerada por placas fotovoltaicas nos próximos dias em função do número observado entre o dia anterior e o dia atual

$T_i = t_i$	1	2	3	4	...	526
$E[T_{i+1}/T_i = t_i]$	8.325	8.325	8.325	8.325	...	8.325

Na Tabela (6.6) podemos observar que mesmo conforme o tempo vai passando a média de energia solar é constante, pois nas condições impostas na equação (6.26) não mudamos a expressão para se calcular a média, assim o valor médio de energia solar gerada por placas fotovoltaicas ao longo desse período fica em 8.325 KWh.

Considerações Finais

O objetivo nesta dissertação foi a construção de modelos que poderiam ser utilizados para a estimação de energia solar gerada por placas fotovoltaicas. A escolha dos devidos modelos para o estudo teve como base as distribuições de probabilidades contínuas como por exemplo a Weibull, Gama e Normal, além das distribuições da família Kumaraswamy. Também foram considerados modelos markovianos para a construção da função de verossimilhança levando em consideração que a quantidade de energia gerada pelas placas podem apresentar dependência entre as observações. Foi utilizado três modelos para a construção markoviana, sendo eles o Weibull bivariado de Freund, Weibull bivariado de Marshall e Olkin e bivarida Weibull Exponenciada Kumaraswamy.

Tanto na análise gráfica de ajuste do melhor modelo e nos critérios de seleção AIC, AICc e BIC tem-se que os modelos markovianos foram melhores que os modelos univariados e entre eles o modelo markoviano via distribuição Weibull bivariada de Marshall e Olkin foi o escolhido para o conjunto de dados da pesquisa. Com a escolha de modelos que assumem dependência entre as variáveis de estudo é possível obter resultados bastante interessantes, um deles como foi mostrado no capítulo 6, é a média esperada dos valores de energia gerada em relação ao valor observado no dia anterior, além de outros resultados como a probabilidade de se obter uma quantidade " x " de energia em KWh dado o valor obtido no dia anterior.

Os resultados nos mostra que em alguns casos existem vantagens em construir modelos que exijam a dependência entre as observações que são sequenciais, neste caso modelos markovianos, onde assim é possível extrair mais informações que podem gerar insumos para empresas, órgãos públicos ou pesquisadores. Nem sempre essa construção é simples, porém os resultados alcançados valem o esforço do pesquisador.

7.1 Perspectivas futuras

- Realização de análises com dados de outras regiões do estado ou país, com o objetivo de realizar comparações;

- Inclusão de covariáveis que possam ser utilizadas para capturar diferenças ao longo do tempo.

Referências

- [1] APsystems Altery power . *Cidade de São José do Rio preto, de 2019 a 2020*, page Disponível em:<<https://apsystemsema.com/ema/index.action>>. Acesso em: 19 de maio de 2019.
- [2] Estilo Solar . *Cidade de São José do Rio Preto, de 2019 a 2020*, page Disponível em:<<https://www.estilosolarengenharia.com.br/>>. Acesso em: 19 de maio de 2019.
- [3] H. AKEIKE. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716- 723, 1974.
- [4] D. ALLISON. Missing data. *Sage University Papers Series on Quantitative Applications in the Social Sciences*, 07-136. Thousand Oaks, CA: Sage, 2001.
- [5] M. M. K. ALVARO. Comparação de Modelos Estatísticos para Estimação do Intervalo de Tempos entre Ultrapasses de um Limiar de Temperatura na Cidade de P. Prudente-SP. *Programa Matemática Aplicada e Computacional*, 2019.
- [6] J. BOLETA. Distribuição exponencial generalizada: uma análise bayesiana aplicada a dados de câncer. *Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo*, 2012.
- [7] H. BOLFARINE and M. C. SANDOVAL. Introdução à inferência estatística. 2. ed, *Rio de Janeiro, SBM*, 2010.
- [8] G. BUNCHAFT and S. R. O. KELLNER. Estatística sem mistérios. 2.ed. *Petrópolis: Vozes, v.2*, 1999.
- [9] K. L. CHUNG and et al. Markov chains with stationary transition. *New York: Springer*, 1960.
- [10] G. M. CORDEIRO and M. CASTRO. A new family of generalized distributions. *Journal of Statistical Computation Simulation Vol. 00, No. 00, August 2009, 1-17*, 2009.
- [11] P. DAS. Kumaraswamy distribution. *10.13140/RG.2.2.22579.73768*, 2015.
- [12] E. FRANK and Jr. HARRELL. Regression modeling strategies with applications to linear models, logistic regression and survival analysis. *Springer-Verlag, New York*, 2001.
- [13] J.E. FREUND. A bivariate extension of the exponential distribution. *Journal of the American Statistical Association*, 56:971–977, 1961.
- [14] A. GARCIA. Geração de Energia Elétrica . *Curso de Engenharia Elétrica 2018.*, pages Disponível em:<<https://www.antonioguilherme.web.br.com/blog/geracao-de-energia-eletrica/energia-solar/>>. Acesso em: 30 de set. de 2019.

- [15] R. D. GUPTA and D. KUNDU. Generalized exponential distributions. *Australian and New Zealand Journal of Statistics* 41, 173-188, 1999.
- [16] W. KOPPEN and R. GEIGER. *Klimate der erde. gotha: Verlag justus perthes. Wall-map 150cm x 200cm*, 1928.
- [17] P. KUMARASWAMY. Generalized probability density-function for double-bounded random processes. *Journal of Hydrology* 46, 79-88, 1980.
- [18] A. J. LEOMONTE, W. B. SOUZA, and G. M. CORDEIRO. The exponentiated kumaraswamy distribution and its log-transform. *Brazilian Journal of Probability and Statistics, Vol.27 No. 1, 31-53*, 2013.
- [19] J. C. LU. Weibull extensions on the freund and marshall-olkin bivariate exponential models. *IEEE Transactions on Reliability*, 38, 615-619, 1989.
- [20] C. MAGARREIRO, S. FREITAS, and M. C. BRITO. Radiação e Energia Solar. *Instituto Dom Luiz, Faculdade de Ciências da Universidade de Lisboa*, v. 39, n. 1/2:p. 57-59, 2014.
- [21] A.J. MARSHALL and I. OLKIN. A multivariate exponential distribution. *Journal of the American Statistical Association*, 62(317):30-44, 1967.
- [22] G. S. MUDHOLKAR and D.K. SRIVASTAVA. Exponentiated weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability*, vol. 42, no. 2, pp. 299-302, June, doi: 10.1109/24.229504, 1993.
- [23] I. MYRTVEIT, E. STENSRUD, and U. H. OLSSON. Analyzing data sets with missing data: an empirical evaluation of imputation methods and likelihood-based methods. *IEEE Transactions On Software Engineering*, vol. 27, no. 11, pp. 999-1013,, 2001.
- [24] A. C. NETO and S. A. LIRA. Coeficiente de correlação para variáveis ordinais e dicotômicas derivados do coeficiente linear de pearson. *RECIE, Uberlândia*, v. 15, n. 1/2, p. 45-53, jan.-dez., 2006.
- [25] L. A. ORIGUELA. Estudo da influência de eventos sobre a estrutura do mercado brasileiro de ações a partir de redes ponderadas por correlações de pearson, spearman e kendall. *Universidade de São Paulo, Faculdade de Economia, Administração e contabilidade de Ribeirão Preto*, 2018.
- [26] M. PÍTSICA. *Energias renováveis: O papel da irena. 1. ed, Curitiba, Appris*, 2015.
- [27] H. RAINNE. *The weibull distribution : a handbook. Chapman , Hall/CRC Taylor, Francis Group 6000 Broken Sound Parkwa*, 2009.
- [28] G. S. ROLIM and et. al. Classificação climática de köppen e de thornthwaite e sua aplicabilidade na determinação de zonas agroclimáticas para o estado de São Paulo. *Bragantia*, pages p. 711-720, 2007.
- [29] A. R. RUFATO and et al. Correlação de pearson e análise de trilha identificando variáveis para caracterizar porta-enxerto de pyrus communis l. *Ciênc. agrotec. vol.34 no.4 Lavras July/Aug*, 2010.
- [30] K. RYU. An extension of Marshall and Olkin bivariate exponential distribution. *Journal of the American Statistical Association*, 88,424:1458-1465, 1993.

- [31] A. SABOOR and et al. The kumaraswamy exponential-weibull distribution: theory and applications. *Hacetatepe Journal of Mathematics and Statistics*, 45(4), 1203 - 1229, 2016.
- [32] M. I. S. SANTOS and M. H. TARUMOTO. Ajuste de funções de distribuição de probabilidade à energia fotovoltaica da cidade de São José do Rio Peto-SP. *Programa Matemática Aplicada e Computacional*, 2020.
- [33] G. SCHWARZ. Estimating the Dimension of a Model. *The Annals of Statistics*, v. 6, n. 2:p. 461–464, Mar. 1978.
- [34] F. G. SILVA and et al. The exponentiated kumaraswamy-g class: General properties and application. *Revista Colombiana de Estadística*, January 2019, Volume 42, Issue 1, pp. 1 to 33 DOI: <http://dx.doi.org/10.15446/rce.v42n1.66205>, 2019.
- [35] R. VERONEZE. Tratamento de dados faltantes empregando biclusterização com imputação múltipla. *Dissertação de Mestrado UNICAMP à Faculdade de Engenharia Elétrica e de Computação*, 2011.
- [36] J. WOO and et al. Exponentiated weibull distribution. *Statistica*. 66. 139-147. 10.6092/issn.1973-2201/493, 2003.