Distribuição Inversa Weighted Lindley: propriedades, estimadores e uma aplicação em dados de tempo de falha

Aline Oliva Luiz

Dissertação de Mestrado do Programa de Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria (MECAI)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Aline Oliva Luiz

Distribuição Inversa Weighted Lindley: propriedades, estimadores e uma aplicação em dados de tempo de falha

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestra – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria.
VERSÃO REVISADA

Área de Concentração: Matemática, Estatística e Computação

Orientador: Prof. Dr. Francisco Louzada Neto

USP – São Carlos
Julho de 2018

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

048d Oliva Luiz, Aline
 Distribuição Inversa Weighted Lindley:
propriedades, estimadores e uma aplicação em dados
de tempo de falha / Aline Oliva Luiz; orientador
Francisco Louzada Neto. -- São Carlos, 2018.
 70 p.

 Dissertação (Mestrado - Programa de Pós-Graduação
em Mestrado Profissional em Matemática, Estatística
e Computação Aplicadas à Indústria) -- Instituto de
Ciências Matemáticas e de Computação, Universidade
de São Paulo, 2018.

 1. 1. Inversa Lindley Ponderada.. 2. 2. Análise
de Sobrevivência.. 3. 3. Aeronaves.. 4. 4. Dados
Censurados. I. Louzada Neto, Francisco , orient.
II. Título.

Aline Oliva Luiz

**The Inverse Weighted Lindley Distribution: properties,
estimation and an application on a failure time data**

Master dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC- USP, in partial fulfillment of the requirements for the degree of the Master – Professional Masters in Mathematics, Statistics and Computing Applied to Industry. *FINAL VERSION*

Concentration Area: Mathematics, Statistics and Computing

Advisor: Prof. Dr. Francisco Louzada Neto

**USP – São Carlos
July 2018**

Este trabalho é dedicado aqueles que sonham grande, e que buscam na sabedoria a maior riqueza de suas vidas.

Dedico aos meus avós Jose Oliva (em memória) e Josephina Oliva (em memória) por me ensinarem que o amor, a família e a sabedoria são os maiores tesouros que podemos conquistar nessa vida.

AGRADECIMENTOS

Inicio meus agradecimentos por DEUS, já que Ele colocou pessoas tão especiais a meu lado, sem as quais certamente não teria dado conta!

A meus pais, Venâncio e Neusa, meu infinito agradecimento, a cada expressão de orgulho por parte deles gera em mim forças para seguir. Obrigada pelo amor incondicional!

Muitos amigos participaram e acompanharam essa etapa da minha vida, sou grata a cada amigo que entendeu a minha ausência e aos colegas de sala no apoio aos estudos e trabalhos em equipe. Em especial, não poderia deixar de agradecer um amigo irmão, que todas as vezes que pensei em desistir, que todas as vezes em que eu disse "não vai dar, vou parar", ele sempre me encorajou dizendo "para de mimimi, eu confio em você", obrigada Julio por não desistir de mim até mesmo quando eu já havia desistido.

Às minhas irmãs, cunhados e sobrinhos por todo carinho e compreensão nessa fase, por vibrarem comigo a cada semestre concluído.

Agradeço a Marli Lanson, minha gerente na época do ingresso ao programa, por todo apoio e incentivo, por ter flexibilizado meu horário de trabalho para que essa conquista fosse possível. Obrigada aos colegas de trabalho Leandro e Rafael.

Jamais poderia deixar de agradecer a Pedro Luiz Ramos, por toda dedicação e contribuição a este trabalho, pelas horas empenhadas em gerarmos os resultados aqui apresentados.

Agradeço ao meu orientador Francisco Louzada Neto, por ter aceito esse desafio e por toda orientação e apoio não somente neste trabalho, mas no mestrado como um todo, meu muito obrigada!

Agradeço aos membros da banca de examinadora, Adriano K. Suzuki, Pedro L. Ramos e Paulo Henrique Ferreira pela dedicação em avaliar este trabalho e pelas recomendações pontuadas. Finalmente, gostaria de agradecer à Universidade de São Paulo Instituto de Ciências Matemáticas e de Computação por abrirem as portas para que eu pudesse realizar esta conquista.

“Eis o meu segredo: só se vê bem com o coração. O essencial é invisível aos olhos. Os homens esqueceram essa verdade, mas tu não a deves esquecer. Tu te tornas eternamente responsável por aquilo que cativas.”
(Antoine de Saint-Exupéry)

RESUMO

LUIZ, A. O. **Distribuição Inversa Lindley Ponderada: propriedades, estimadores e uma aplicação em dados de tempo de falha.** 2018. 70 p. Dissertação (Mestrado – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2018.

Quando considerado o funcionamento de uma aeronave em operação de voo para transporte de passageiros civis, há a ocorrência de falhas que podem afetar a aeronavegabilidade ou não e, dentro das falhas que não afetam, encontram-se, por exemplo, as falhas de recursos utilizados durante o voo pelos passageiros, como poltronas ou sistema de entretenimento de bordo. É notória a importância de uma análise em relação ao diagnóstico de falhas que possuem maior impacto na satisfação do cliente, assim como de uma tomada de medidas de manutenção preventiva para que tais ocorrências sejam minimizadas. Como resultado, entre as 16 principais falhas que ocorrem, apenas 4 falhas caracterizam-se como sendo de maior impacto. É nesse sentido que o principal objetivo deste trabalho é descrever o tempo de falha desses principais componentes. Para tanto, ajustamos diferentes distribuições de probabilidade que descrevem o comportamento dessas falhas. Levando-se em conta que em um tipo de falha os principais modelos utilizados na literatura como Weibull, Gama e Log-normal não são os mais adequados para descrever tal característica, propomos um novo modelo de confiabilidade intitulado Inversa Lindley Ponderada. As propriedades matemáticas e os métodos de estimação dos parâmetros dessa nova distribuição são exploradas ao longo dessa dissertação. O modelo é utilizado para descrever o tempo de vida de um tipo de falha, retornando melhores ajustes que as principais distribuições de confiabilidade. Com base nesses novos resultados, propomos uma política de manutenção preventiva para os principais componentes.

Palavras-chave: Inversa Lindley Ponderada, Análise de Sobrevivência, Aeronaves, Dados Censurados.

ABSTRACT

LUIZ, A. O. **The Inverse Weighted Lindley Distribution: properties, estimation and an application on a failure time data.** 2018. 70 p. Dissertação (Mestrado – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2018.

During the operation of an aircraft for commercial flight of passengers, there are occurrences of failures that can compromise the satisfaction of passengers. For instance, the malfunction of seats or the entertainment system. This problem is very important for the operation in order to maintain customer satisfaction and customer loyalty. Indeed, it is of main interesting analysis for diagnosis of failures, which have greater impact on customer satisfaction. The results of such analysis can be used to propose preventive maintenance schemes in order to minimize costs. Based on our analysis, among the 16 major types of failures that occur, 4 types showed to have greatest impacts in the role system. In this sense, the main objective of this work is to describe the failure time of these main components. To achieve that, we set different probability distributions to describe the random behavior of these failures. For one type of failure we showed that there is space for improvement of its probability distribution as the Weibull, Gamma and Log-normal are not the most adequate to describe the random behaviour of such characteristic. Therefore, we propose a new reliability distribution, the so called inverse Weighted Lindley distribution. The mathematical properties and the parameter estimation of the parameters for the proposed distribution are discussed. The new distribution is used to describe the lifetime of one of types of failure. Based on these new results, we propose a preventive maintenance policy to maintain the quality of the main components.

Keywords: Inverse Weighted Lindley, Survival Analysis, Aircraft, Censorship Data.

LISTA DE ILUSTRAÇÕES

Figura 1 – Porcentagem de falhas por ocorrências analisadas de 11 aeronaves no período aproximado de 2 anos. Fonte: Própria (2018).	27
Figura 2 – Porcentagem de falhas do Sistema 1 por posição, em ocorrências analisadas de 11 aeronaves no período aproximado de 2 anos. Fonte: Própria (2018).	27
Figura 3 – Porcentagem de falhas do Sistema 3 por posição, em ocorrências analisadas de 11 aeronaves no período aproximado de 2 anos. Fonte: Própria (2018).	27
Figura 4 – Porcentagem de falhas do Subsistema 1, em ocorrências analisadas de 11 aeronaves no período aproximado de 2 anos. Fonte: Própria (2018).	28
Figura 5 – Porcentagem de falhas do Subsistema 3, em ocorrências analisadas de 11 aeronaves no período aproximado de 2 anos. Fonte: Própria (2018).	28
Figura 6 – Gráfico de Pareto de falhas por subsistemas do Sistema 1, em ocorrências analisadas de 11 aeronaves no período aproximado de 2 anos. Fonte: Própria (2018).	29
Figura 7 – Gráfico de Pareto de ações corretivas por subsistemas do Sistema 1, em ocorrências analisadas de 11 aeronaves no período aproximado de 2 anos. Fonte: Própria (2018).	29
Figura 8 – Gráfico de Pareto de falhas por subsistemas do Sistema 3, em ocorrências analisadas de 11 aeronaves no período aproximado de 2 anos. Fonte: Própria (2018).	30
Figura 9 – Gráfico de Pareto de ações corretivas por subsistemas do Sistema 3, em ocorrências analisadas de 11 aeronaves no período aproximado de 2 anos. Fonte: Própria (2018).	30
Figura 10 – Formas da função de risco. Fonte: Ramos (2014).	32
Figura 11 – TTT plot de diferentes distribuições que indicam a forma da função de risco. Fonte: Ramos (2014).	33
Figura 12 – Funções densidade de probabilidade e acumulada da distribuição gama para diferentes valores de α e β . Fonte: Ramos (2014).	35
Figura 13 – Funções de sobrevivência e de risco da distribuição Gama com diferentes valores de α e β . Fonte: Ramos (2014).	36
Figura 14 – Funções densidade de probabilidade e acumulada da distribuição Weibull para diferentes valores de α e β . Fonte: Ramos (2014).	36
Figura 15 – Funções de sobrevivência e de risco da distribuição Weibull com diferentes valores de α e β . Fonte: Ramos (2014).	37

Figura 16 – Densidade e acumulada da distribuição Log-normal para vários valores de μ e σ . Fonte: Ramos (2014).	38
Figura 17 – Funções de sobrevivência e de risco da distribuição Log-normal para vários valores de μ e σ . Fonte: Ramos (2014).	38
Figura 18 – TTT-plot, função densidade de probabilidade e a função de risco, ambas ajustadas através da distribuição Log-normal, utilizando-se os dados de falha do subsistema1I.	40
Figura 19 – TTT-plot, função densidade de probabilidade e a função de risco, ambas ajustadas através da distribuição Log-normal, utilizando-se os dados do subsistema1J.	41
Figura 20 – TTT-plot, função densidade de probabilidade e a função de risco, ambas ajustadas através da distribuição Weibull, utilizando os dados de falha do sistema3E.	42
Figura 21 – TTT-plot, função densidade de probabilidade e a função de risco, ambas ajustadas através da distribuição Log-normal, utilizando-se os dados de falha gerado pelo subsistema 3F.	44
Figura 22 – Função densidade para distribuição ILP considerando diferentes valores de ϕ and λ	47
Figura 23 – Curva de risco para a distribuição ILP considerando diferentes valores de ϕ e λ	50
Figura 24 – MREs, MSEs relacionado para os estimadores de $\lambda = 2$ e $\phi = 4$ para $N = 10000$ amostras simuladas, considerando diferentes valores de n	58
Figura 25 – MREs, MSEs relacionado para estimadores de $\lambda = 0,5$ e $\phi = 2$ para $N = 10000$ amostras simuladas, considerando diferentes valores de n	58
Figura 26 – MREs, RMSEs relacionado para os estimadores de $\lambda = 0,5$ e $\phi = 2$ para $N = 10000$ amostras simuladas, considerando diferentes valores de n e 30% de censura.	60
Figura 27 – MREs, RMSEs relacionado para os estimadores de $\lambda = 0,5$ e $\phi = 2$ para $N = 10000$ amostras simuladas, considerando diferentes valores de n e 50% de censura.	61
Figura 28 – TTT-plot, função de sobrevivência ajustada por diferentes distribuições e o estimador Kaplan–Meier e a função de risco ajustada pela distribuição ILP considerando dados relacionados a tempo de falha de 194 ocorrências.	62

LISTA DE TABELAS

Tabela 1 – Resultados dos critérios AIC, AICC, HQIC e CAIC de diferentes distribuições de probabilidade considerando os dados do subsistema1I.	40
Tabela 2 – Tabela de frequência de ocorrência de falhas do subsistema1I por quartil determinado.	40
Tabela 3 – Resultados dos Critérios AIC, AICC, HQIC e CAIC de diferentes distribuições de probabilidade considerando os dados do subsistema1J.	41
Tabela 4 – Tabela de frequência de ocorrência de falhas do subsistema 1J por quartil determinado.	42
Tabela 5 – Resultados dos Critérios AIC, AICC, HQIC e CAIC de diferentes distribuições de probabilidade, considerando os dados de falha do subsistema 3E.	42
Tabela 6 – Tabela de frequência de ocorrência de falhas subsistema 3E por quartil determinado.	43
Tabela 7 – Resultados dos Critérios AIC, AICC, HQIC e CAIC de diferentes distribuições de probabilidade, considerando os dados do subsistema 3F.	43
Tabela 8 – Tabela de frequência de ocorrência de falhas geradas pelo subsistema 3F do sistema 3 por quartil determinado.	43
Tabela 9 – Dados relacionados a tempo de falha de 194 ocorrências de falha em uma aeronave.	60
Tabela 10 – MLE, erros-padrão e intervalo de confiança de 95% para ϕ e λ	61
Tabela 11 – Resultados dos critérios AIC, AICC, HQIC, CAIC para diferentes distribuições de probabilidade considerando dados relacionados para tempo de falha de 194 ocorrências de falha de uma aeronave.	62
Tabela 12 – Tabela de frequência de ocorrência de falhas do subsistema1J por quartil determinado considerando a distribuição Inversa Lindley Ponderada.	62

SUMÁRIO

1	INTRODUÇÃO	21
2	O PROBLEMA REAL	25
3	MÉTODOS PARAMÉTRICOS PARA ANÁLISE DE SOBREVIVÊNCIA	31
3.1	Função de Sobrevivência	31
3.2	Gráfico TTT-plot	32
3.3	Censura	33
3.4	Modelos Paramétricos	34
3.4.1	<i>Distribuição Gama</i>	35
3.4.2	<i>Distribuição Weibull</i>	36
3.4.3	<i>Distribuição Log-normal</i>	37
3.4.4	<i>CrITÉrios de Seleção de Modelos</i>	38
3.5	Análise de Sobrevivência das principais falhas	39
3.5.1	<i>Tempo de Falha do Sistema 1I</i>	39
3.5.2	<i>Tempo de falha do subsistema 1J</i>	41
3.5.3	<i>Tempo de falha do subsistema 3E</i>	42
3.5.4	<i>Tempo de falha do subsistema 3F</i>	43
4	DISTRIBUIÇÃO INVERSA LINDLEY PONDERADA	45
4.1	Distribuição Inversa Lindley Ponderada	46
4.2	Propriedades da Distribuição ILP	47
4.2.1	<i>Momentos</i>	47
4.2.2	<i>Propriedades de Sobrevivência</i>	49
4.2.3	<i>Entropia</i>	50
4.3	Inferência	51
4.3.1	<i>Estimadores de Máxima Verossimilhança</i>	51
4.3.2	<i>Censura Aleatória</i>	52
4.4	Correção de vício para estimadores de máxima verossimilhança	53
4.4.1	<i>Aproximação corretiva</i>	54
4.4.2	<i>Correção de vício por bootstrap</i>	56
4.5	Simulações Numéricas	57
4.5.1	<i>Dados Completos</i>	57

4.5.2	<i>Dados Censurados</i>	59
4.6	Análise de Sobrevivência das falhas do subsistema 1J	59
5	CONCLUSÃO	65
	REFERÊNCIAS	67

INTRODUÇÃO

A aviação civil no Brasil, a partir de seu voo inaugural em 1927, tem passado por muitas transformações e vem crescendo ao longo dos anos. Segundo os dados da Infraero (Empresa Brasileira de Infraestrutura Aeroportuária), nos últimos seis anos (de 2012 a 2017), os pousos e decolagens nos aeroportos brasileiros somam mais de 7 milhões no chamado transporte regular (voo efetuado com a existência de HOTRAN, assim definido pelo órgão regulador de acordo com a Portaria Nº 569/GC5 de 05/09/2000).

De janeiro de 2015 até setembro do mesmo ano, o Brasil transportou em voos internacionais 4.368.922 passageiros (Anuário do Transporte Aéreo, 2015). Considerando o modelo de aeronave *Boeing B767*, 13,6% desses passageiros voaram de *business class* a um custo quase que dobrado quando comparado com valor da passagem para o mesmo destino e no mesmo período, mas pela classe econômica.

Um dos grandes desafios das companhias aéreas é manter o avanço da aviação no Brasil, considerando o equilíbrio entre o menor custo de operação e a máxima excelência na qualidade dos serviços oferecidos aos passageiros. Assim, identificar as principais falhas que possuem os maiores impactos nos passageiros da categoria *business* significa corroborar com um ambiente que acarrete a melhoria da qualidade do serviço prestado pela empresa ao cliente. Além disso, dado o caráter preventivo, há a redução de custos de manutenção (SIM; ENDRENYI, 1988; GERTSBAKH, 2013).

Verificadas as falhas de maior impacto na satisfação dos passageiros, o objetivo é descrever o tempo de vida destes componentes por meio da análise de confiabilidade. Do ponto de vista paramétrico, diferentes distribuições de probabilidade podem ser utilizadas para descrever o tempo de vida desses componentes. Assim, inicialmente, consideramos os modelos mais comuns utilizados em análise de confiabilidade, tais como: as distribuições Weibull, Gama e Log-normal (LOUZADA-NETO; MAZUCHELI; ACHCAR, 2001; LAWLESS, 2011). Em seguida, mostramos que, para um tipo de falha, tais distribuições podem não ser as mais adequadas. Assim,

propomos uma nova distribuição denominada Inversa Lindley Ponderada (ILP). A distribuição ILP é uma generalização da distribuição Inversa Lindley (SHARMA *et al.*, 2015), e pode ser representada como um modelo de mistura de duas distribuições Inversa Gama. Como modelo de mistura, esta nova distribuição pode ser utilizada para descrever a variabilidade entre dois grupos latentes (não-observáveis) distintos. Por exemplo, podemos estar interessados em estudar o tempo de vida de componentes que são compostos por itens novos e reparados, ou seja, nesse caso apenas o tempo de falha é observado e os grupos são variáveis latentes dado que não se sabe se o item é novo ou reparado. O uso de novos modelos de probabilidade tem sido extensivamente discutido na literatura nas últimas décadas. Tahir e Cordeiro (2016) revisaram uma grande quantidade de modelos de probabilidade recentemente propostos.

Neste trabalho, apresentamos diversas propriedades matemáticas da distribuição ILP, entre elas: os momentos, a média, a variância, as funções de sobrevivência e de risco, a medida de entropia e os estimadores de máxima verossimilhança (EMVs) dos parâmetros, bem como suas propriedades assintóticas. Mostramos que o EMV possui um vício sistemático para pequenas amostras. Nesse sentido, uma correção de vício é proposta para que os EMVs modificados possuam correção de vício de segunda ordem. Esta correção tem sido considerada para outras distribuições por muitos autores (CORDEIRO *et al.*, 1997; LEMONTE, 2011; TEIMOURI; NADARAJAH, 2013; GILES; FENG; GODWIN, 2013; RAMOS; LOUZADA; RAMOS, 2016; MAZUCHELI; MENEZES; DEY, 2017; MAZUCHELI; DEY, 2018). Em seguida, são apresentados os EMVs na presença de dados com censura aleatória. Para esses estimadores uma correção de vício também é discutida. Um estudo de simulação numérica é realizado para examinar o efeito da correção do vício nos EMVs, para dados completos e censurados.

Por fim, com base nos modelos ajustados, propomos um tempo ideal de manutenção preventiva, baseado nos quantis das distribuições ajustadas. Dessa forma, esperamos evitar uma grande quantidade de ocorrências de falhas acima do limite pré-estabelecido. Tal estudo deve ter grande impacto na forma de se realizar a manutenção preventiva dos itens estudados. Com base nesses resultados, esperamos que a política de manutenção adotada melhore a qualidade do produto, aumentando assim, a satisfação dos clientes da linha aérea, melhorando as margens de lucro da empresa.

O trabalho proposto é organizado da seguinte forma: no Capítulo 2 é apresentado o problema estudado, assim como uma análise estatística descritiva do mesmo, onde apresentamos os dados quantitativos dos tipos de falhas, a fim de escolhermos as falhas de maior impacto. Para tanto, é realizada a estratificação dos dados com a utilização da técnica de diagrama de Pareto. Esta técnica estatística defende que, para qualquer circunstância, 80% das consequências estão concentradas em 20% das causas (JURAN; GRZYNA, 1988). Então, uma vez que identificamos as causas que compõem esses 20%, estaremos resolvendo 80% dos problemas de falha. No Capítulo 3 é apresentada uma revisão dos métodos utilizados no estudo, assim como as distribuições consideradas: a Gama, a Weibull e a Log-normal. Ainda no Capítulo 3, apresentaremos a análise

paramétrica dos subsistemas selecionados, propondo tempos de manutenção preventiva para cada um deles. No Capítulo 4 é apresentada a nova distribuição de probabilidade Inversa Lindley Ponderada. Tal distribuição é utilizada para descrever o tempo de vida de um dos subsistemas, possibilitando propor o tempo de manutenção ideal. Por fim, no Capítulo 5, apresentamos alguns comentários finais e perspectivas futuras.

O PROBLEMA REAL

Para o presente trabalho, foram disponibilizadas as ocorrências de falhas na *business class* de 11 aeronaves de mesmo modelo, com voos realizados de origens e destinos diversos, em um período aproximado de 2 anos. Tais falhas se relacionam somente a falhas de sistemas utilizados pelos próprios passageiros, como, por exemplo, recline de poltronas e acesso ao entretenimento que, por sua vez, não impactam a aeronavegabilidade e a segurança do voo.

Sendo assim, para descrição adequada do banco de dados (não foi autorizada a publicação do mesmo na íntegra), dividimos o mesmo de acordo com quatro variáveis:

- Tipo de sistema da falha;
- Tipo de subsistema da falha;
- Posição da falha;
- Tipo de ação corretiva.

A variável **Tipo de sistema da falha** foi classificada em cinco sistemas chamados de:

- Sistema 1 - falhas relacionadas ao sistema 1;
- Sistema 2 - falhas relacionadas ao sistema 2;
- Sistema 3 - falhas relacionadas ao sistema 3;
- Sistema 4 - falhas relacionadas ao sistema 4;
- Outros - falhas pontuais agrupadas devido a baixa ocorrência.

A variável Subsistema consiste na combinação da variável sistema com uma reclassificação da mesma. Sendo assim, para cada possível sistema temos um subsistema. Assim, realizamos

o primeiro filtro para a escolha de uma falha de maior impacto na satisfação do cliente, resultando nos Sistemas 1 e 3 e os seus subsistemas:

- Subsistema 1A;
- Subsistema 1B;
- Subsistema 1C;
- Subsistema 1D;
- Subsistema 1E;
- Subsistema 1F;
- Subsistema 1G;
- Subsistema 1H;
- Subsistema 1I;
- Subsistema 1J;

- Subsistema 3A;
- Subsistema 3B;
- Subsistema 3C;
- Subsistema 3D;
- Subsistema 3E;
- Subsistema 3F.

A variável **Posição** foi classificada em posição de 1 a 5, enquanto a variável **Tipo de ação corretiva** foi classificada em:

- Ação 1;
- Ação 2;
- Ação 3;
- Ação 4.

A partir da classificação do banco de dados, realizamos a análise quantitativa descritiva do problema e o resultado dessa análise é condensado nas Figuras 1, 2 e 3.

Figura 1 – Porcentagem de falhas por ocorrências analisadas de 11 aeronaves no período aproximado de 2 anos. Fonte: Própria (2018).

Figura 2 – Porcentagem de falhas do Sistema 1 por posição, em ocorrências analisadas de 11 aeronaves no período aproximado de 2 anos. Fonte: Própria (2018).

Figura 3 – Porcentagem de falhas do Sistema 3 por posição, em ocorrências analisadas de 11 aeronaves no período aproximado de 2 anos. Fonte: Própria (2018).

Pelas Figuras 1,2 e 3, concluímos que existe uma maior incidência de falha nos Sistemas 1 e 3. Eles representam 40% e 24% das falhas, respectivamente. Verifica-se também que as falhas tendem a ocorrer uniformemente nas posições e, por isso, a análise por posição foi descartada. Assim, escolhemos o Sistema 1 e o Sistema 3 para seguirmos com a análise de seus subsistemas.

Os resultados da análise descritiva dos subsistemas do Sistema 1 e do Sistema 3 estão condensados nas Figuras 4 e 5.

Figura 4 – Porcentagem de falhas do Subsistema 1, em ocorrências analisadas de 11 aeronaves no período aproximado de 2 anos. Fonte: Própria (2018).

Figura 5 – Porcentagem de falhas do Subsistema 3, em ocorrências analisadas de 11 aeronaves no período aproximado de 2 anos. Fonte: Própria (2018).

Para o Sistema 1, 84% das falhas ocorrem nos subsistemas 1I e 1J, conforme ilustrado no gráfico da Figura 6. E, a ação corretiva para esses, em 77%, consiste na ação 1 e na ação 2, conforme a Figura 7.

Conforme podemos observar na Figura 8, 89% das falhas do Sistema 3 ocorrem nos subsistemas 3E e 3F, e de acordo com a Figura 9, em 93% dos casos, a ação corretiva consiste na ação 1 e na ação 2.

Figura 6 – Gráfico de Pareto de falhas por subsistemas do Sistema 1, em ocorrências analisadas de 11 aeronaves no período aproximado de 2 anos. Fonte: Própria (2018).

Figura 7 – Gráfico de Pareto de ações corretivas por subsistemas do Sistema 1, em ocorrências analisadas de 11 aeronaves no período aproximado de 2 anos. Fonte: Própria (2018).

Assim, as falhas com maiores incidências foram: subsistema 1J e subsistema 1I do Sistema 1, subsistema 3E e subsistema 3F do Sistema 3. Considerando um custo médio por trecho de voo de R\$3.000,00 (utilizando como referência um voo São Paulo (GRU) / Nova Iorque (EWR)), ao final de um ano, o custo dessas falhas será de, aproximadamente, R\$ 1.300.000,00. Para compor esse custo, consideramos R\$ 3.000,00 para os casos de assento restrito, ou seja, o assento fica impossibilitado de ocupação, e 30% desse valor para os casos de reclamação durante o voo (nesse caso a passagem não deixa de ser vendida, mas um custo de insatisfação do cliente é gerado. O marketing atual defende que a conquista de um cliente novo custará 5 vezes o custo de um cliente já fidelizado, desconsiderando a repercussão negativa nas mídias sociais, na qual a insatisfação propagada é estimada de 1 para 10 (KOTLER, 2000).

Figura 8 – Gráfico de Pareto de falhas por subsistemas do Sistema 3, em ocorrências analisadas de 11 aeronaves no período aproximado de 2 anos. Fonte: Própria (2018).

Figura 9 – Gráfico de Pareto de ações corretivas por subsistemas do Sistema 3, em ocorrências analisadas de 11 aeronaves no período aproximado de 2 anos. Fonte: Própria (2018).

Com base nesse alto custo, as Falhas 1I, 1J, 3E e 3F foram escolhidas para realizarmos a análise de confiabilidade, a fim de propormos um tempo de manutenção preventiva ótimo para a redução dessas falhas.

MÉTODOS PARAMÉTRICOS PARA ANÁLISE DE SOBREVIVÊNCIA

3.1 Função de Sobrevivência

A análise de sobrevivência tem como principal característica o uso de funções que descrevam diferentes problemas em que a variável resposta é dada pelo tempo de falha. Uma característica que também é muito encontrada em dados de sobrevivência é a presença de observações parcialmente conhecidas, geralmente denotada de observações censuradas.

Considerando a função densidade de probabilidade (f.d.p), com a probabilidade de ocorrência do evento de interesse no tempo t , sua expressão é dada por:

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t}, \quad f(t) \geq 0. \quad (3.1)$$

A variável aleatória $T > 0$ também pode ser expressa através de outras, tais como: a função acumulada (f.d.a) $F(t)$, a função de sobrevivência $S(t)$ e a função de risco $h(t)$. A função de sobrevivência de uma variável aleatória T positiva, com a probabilidade de um componente não falhar até o tempo t , é obtida através das seguintes relações ([LOUZADA-NETO; MAZUCHELI; ACHCAR, 2001](#); [LAWLESS, 2011](#)),

$$S(t) = P[T > t] = 1 - P[T \leq t] = 1 - \int_0^t f(t)dt = 1 - F(t), \quad (3.2)$$

em que $0 < S(t) < 1$, $F(t)$ é a probabilidade de ocorrência do evento de interesse até o tempo t . A $F(t)$ também é conhecida como função de distribuição acumulada (f.d.a) da variável aleatória T . A função de risco $h(t)$ apresenta a probabilidade de falha aleatória de um componente que sobreviveu até o tempo t , podendo ser obtida através dos seguintes resultados:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}. \quad (3.3)$$

A função de risco é utilizada para descrever o risco instantâneo de falha do evento de interesse. A Figura 10 mostra diferentes formas que a função de risco pode assumir, entre elas: constante, decrescente, crescente, forma de banheira e forma de banheira inversa.

Figura 10 – Formas da função de risco. Fonte: Ramos (2014).

Dentre as relações matemáticas importantes que podemos obter das funções anteriores segue:

$$h(t) = \frac{S'(t)}{S(t)} = \frac{\delta}{\delta t} \log S(t), \quad (3.4)$$

$$S(t) = \exp\left[-\int_0^t h(t)dt\right], \quad (3.5)$$

$$f(t) = h(t)S(t). \quad (3.6)$$

Assim, conhecer a função $h(t)$ implica conhecer as outras funções: $f(t)$, $F(t)$ e $S(t)$.

3.2 Gráfico TTT-plot

Barlow e Campo (1975) propuseram uma técnica gráfica chamada TTT-plot (*tempo total em teste*) para verificar o comportamento da função de risco. O gráfico é construído com o plot

consecutivo das quantidades $[r/n, G(r/n)]$, sendo $G(r/n)$ a função dada por:

$$G(r/n) = \frac{\sum_{i=1}^r t_i + (n-r)t_{(r)}}{\sum_{i=1}^n t_i} \quad (3.7)$$

para $r = 1, \dots, n$, t_i as estatísticas de ordem da amostra.

Por meio do TTT-plot, conseguimos identificar o comportamento da função de risco relacionada aos dados. Ao obtermos uma reta diagonal, concluímos que os dados possuem uma função de risco constante. No caso desta ser convexa, então o risco é decrescente e, no caso de ser côncava, o risco é crescente. Por outro lado, se a forma for convexa e depois côncava, então a função de risco possui forma de banheira e, se for côncava e depois convexa, dizemos que é a forma Inversa de banheira. Na Figura 11 apresenta-se o comportamento da função de risco, utilizando-se o TTT-plot.

Figura 11 – TTT plot de diferentes distribuições que indicam a forma da função de risco. Fonte: [Ramos \(2014\)](#).

[Nair et al. \(2008\)](#) mostram que algumas transformações TTT podem ser utilizadas em diferentes aplicações de análise de sobrevivência.

3.3 Censura

Em estudos em que a variável resposta é dada por tempo, é comum a não-observação do evento de interesse. Tal evento ocorre devido a diversos tipos de problemas durante o experimento, por exemplo: limitação do tempo, custo e perda de acompanhamento devido a diversos fatores (dados médicos). Tais dados, embora parcialmente conhecidos, possuem uma rica fonte de

informação, sendo que estes, em hipótese alguma, podem ser descartados, uma vez que fornecem informações importantes em relação ao tempo de sobrevivência até o momento da sua interrupção (censura). Descartar esse tipo de informação implica na obtenção de resultados viciados que podem comprometer todo o estudo.

Os dados de sobrevivência podem ser representados pelo par (t_i, δ_i) , sendo t_i o tempo de falha ou censura do i -ésimo componente e δ_i a variável indicadora de censura, isto é: $\delta_i = 1$ se t_i é uma observação completa e 0 se t_i é uma observação censurada, $i = 1, \dots, n$.

Os mecanismos de censura são diferenciados. Entre eles, iremos citar quatro tipos de censura mais utilizados (LOUZADA-NETO; MAZUCHELI; ACHCAR, 2001; LAWLESS, 2011):

- Censura tipo I: fixamos um tempo limite e o experimento termina após esse período. Nesse tipo de censura removemos os componentes que não falharam no tempo determinado, ou seja, o tempo de falha de todos os componentes que não falharam devido à interrupção são considerados censurados.
- Censura tipo II: fixamos um número de ocorrência de falhas e o experimento termina quando ocorrer a n -ésima falha pré-estabelecida. Nesse tipo de censura, supondo que temos n componentes no estudo e estabelecendo encerrar o experimento após r componentes falharem, então, ao final, teremos $n-r$ dados censurados.
- Censura aleatória: esse tipo de censura ocorre normalmente em estudos clínicos em que ao longo do experimento o paciente abandona o estudo. Tal tipo de censura pode ocorrer em estudos de confiabilidade em que os itens são colocados sob teste em diferentes tempos. No entanto, devido à sua estrutura geral de formulação, a mesma consegue englobar os tipos de censura I e II.
- Censura intervalar: quando não sabemos exatamente o tempo de ocorrência do evento de interesse, mas temos a informação sobre o intervalo em que ocorreu o evento.

3.4 Modelos Paramétricos

Diferentes métodos podem ser utilizados para analisar dados de sobrevivência. Neste trabalho, nosso foco será em modelar os tempos de falha por meio da abordagem paramétrica (LAWLESS, 2011). As distribuições de probabilidade escolhidas para ajustar os dados serão: a Gama, a Weibull e a Log-normal, devido à sua flexibilidade e popularidade em análise de sobrevivência.

3.4.1 Distribuição Gama

A distribuição Gama é uma importante distribuição de probabilidade para descrever dados não-negativos e não-nulos. Não limitada à análise de sobrevivência, a mesma é utilizada para solucionar diversos problemas em meteorologia, processamento de sinais, dados sismológicos, dentre outros. (SOBRINHO *et al.*, 2014)

Seja T uma variável aleatória com uma distribuição de probabilidade gama denotado por $Gama(\alpha, \beta)$, cuja função de densidade de probabilidade é dada por:

$$f(t|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} \exp(-\beta t), \quad (3.8)$$

para todo $t > 0$, $\alpha > 0$ e $\beta > 0$, sendo $\Gamma(\alpha)$ a função gama definida como:

$$\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx \quad (3.9)$$

Podemos ver na Figura 12 diferentes formas para a f.d.p. e f.d.a. da distribuição Gama.

Figura 12 – Funções densidade de probabilidade e acumulada da distribuição gama para diferentes valores de α e β . Fonte: Ramos (2014).

Na Figura 13, tem-se diferentes formas para a função de sobrevivência e função de risco da distribuição Gama.

Figura 13 – Funções de sobrevivência e de risco da distribuição Gama com diferentes valores de α e β .
Fonte: Ramos (2014).

3.4.2 Distribuição Weibull

Proposta por Waloddi Weibull em 1954 e nomeada em homenagem ao autor, a distribuição Weibull possui um papel fundamental em análise de sobrevivência. Comumente utilizada para descrever tempo de falha de produtos industriais, a mesma ficou popularizada em aplicações práticas devido a sua grande variedade de formas. (BRAILE; ANDRADE, 2013)

Figura 14 – Funções densidade de probabilidade e acumulada da distribuição Weibull para diferentes valores de α e β . Fonte: Ramos (2014).

Suponha que T segue uma distribuição Weibull denotada por $\text{Weibull}(\alpha, \beta)$, então sua

f.d.p. é dada por:

$$f(t|\alpha, \beta) = \frac{\alpha}{\beta} \left(\frac{t}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{t}{\beta}\right)^{\alpha}\right), \quad (3.10)$$

para todo $x > 0$, $\alpha > 0$ e $\beta > 0$.

Outra parametrização muito comum é considerar que $\mu = 1/\beta$. Notamos na Figura 14 as diferentes formas para a f.d.p. e f.d.a. da distribuição Weibull.

Na Figura 15 temos as diferentes formas para a função de sobrevivência e a função de risco da distribuição Weibull.

Figura 15 – Funções de sobrevivência e de risco da distribuição Weibull com diferentes valores de α e β .
Fonte: Ramos (2014).

3.4.3 Distribuição Log-normal

A distribuição Log-normal, assim como a distribuição Weibull, é normalmente utilizada em estudos de tempo de vida de produtos e materiais como: fadiga de metais, semicondutores, diodos e isolamento elétrica, sendo também conhecida por suas aplicações em engenharia e medicina. (NASCIMENTO, 2016)

Suponha que Y possui uma distribuição normal com média μ e desvio-padrão σ e $T = \exp(Y)$. Então, dizemos que T possui uma distribuição Log-normal em que sua f.d.p. é dada por:

$$f(t|\mu, \sigma) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma t} \exp\left(-\frac{1}{2}\left(\frac{\log(t) - \mu}{\sigma}\right)^2\right), \quad (3.11)$$

para todo $t > 0$, $\mu > 0$ e $\sigma > 0$.

Vemos na Figura 16 diferentes formas para a f.d.p. e f.d.a. da distribuição Log-normal. Na Figura 17 temos diferentes formas para a função de sobrevivência e para a função de risco da distribuição Log-normal.

Figura 16 – Densidade e acumulada da distribuição Log-normal para vários valores de μ e σ . Fonte: Ramos (2014).

Figura 17 – Funções de sobrevivência e de risco da distribuição Log-normal para vários valores de μ e σ . Fonte: Ramos (2014).

3.4.4 Critérios de Seleção de Modelos

O grande número de distribuições de probabilidade pode trazer dúvida sobre qual se mostra mais adequada a determinado tipo de situação. Nesse contexto, o uso de critérios que auxiliam na comparação e seleção de modelos paramétricos em casos relacionados a problemas de incerteza é de extrema importância. Aqui iremos discutir alguns dos critérios mais utilizados na literatura, sendo esses: AIC (critério de informação de Akaike), AICC (AIC corrigido), HQIC (critério de informação de Hannan–Quinn) e CAIC (AIC consistente). (SOBRAL; BARRETO,

2011).

Akaike (1974) propõe um método baseado na medida de Informação de Kullback-Leibler. Seja k o número de parâmetros a serem estimados, n o número de observações de $\mathbf{x}$ e $\hat{\theta}$ uma estimativa de θ , o critério de informação de Akaike (AIC) é obtido por:

$$AIC = -2\log(L(\hat{\theta}; \mathbf{x})) + 2k. \quad (3.12)$$

Onde L é a função de máxima verossimilhança do modelo.

Considerando um conjunto de modelos para $\mathbf{x}$, este ajustado aos dados, temos que o melhor modelo é aquele que obter o menor AIC. O critério não somente seleciona um ótimo ajuste, mas também penaliza a adição de parâmetros, desencorajando *overfitting*, ou seja, ele evita a seleção de um modelo extremamente complexo e com muitos parâmetros que tenham um desempenho preditivo ruim. Outros critérios utilizados similares a este foram:

$$AICC = -2\log(L(\hat{\theta}; \mathbf{x})) + 2k + \frac{2k(k+1)}{n-k-1}, \quad (3.13)$$

$$HQIC = -2\log(L(\hat{\theta}; \mathbf{x})) + 2k\log(\log(n)), \quad (3.14)$$

$$CAIC = -2\log(L(\hat{\theta}; \mathbf{x})) + k(\log(n) + 1). \quad (3.15)$$

É importante ressaltar que os métodos aqui discutidos não dizem o quão bom é o ajuste, apenas que, das distribuições consideradas, o modelo que possuir o melhor ajuste será o que possuir os menores valores.

3.5 Análise de Sobrevivência das principais falhas

Objetivando-se descrever os tempos de vida dos componentes citados no capítulo anterior, utilizamos as distribuições Log-normal, Weibull e Gama para descrever as principais falhas, sendo estas, as falhas do sistema 1, subsistema I e J, e do sistema 3, subsistema E e F.

3.5.1 Tempo de Falha do Sistema 1I

A Tabela 1 apresenta os resultados dos critérios AIC, AICC, HQIC e CAIC de diferentes distribuições de probabilidade, considerando os dados subsistema 1I.

Com base nos critérios utilizados, podemos concluir, através da Tabela 1, que a distribuição Log-normal foi a que melhor se ajustou aos dados de falha do subsistema 1I. Os EMVs obtidos foram $\mu = 2,79293$ e $\sigma = 1,45546$.

Tabela 1 – Resultados dos critérios AIC, AICC, HQIC e CAIC de diferentes distribuições de probabilidade considerando os dados do subsistemaII.

Crítérios	Log-normal	Weibull	Gama
AIC	1076,555	1088,914	1094,773
AICC	1072,649	1085,008	1090,867
HQIC	1078,892	1091,251	1097,110
CAIC	1084,306	1096,665	1102,524

Pode-se observar, através da Figura 18, a curva Kaplan Meier (K-M) (LAWLESS, 2003), a função de sobrevivência e a função de risco, ambas funções ajustadas através da distribuição Log-normal.

Figura 18 – TTT-plot, função densidade de probabilidade e a função de risco, ambas ajustadas através da distribuição Log-normal, utilizando-se os dados de falha do subsistemaII.

Utilizando os parâmetros ajustados, conseguimos estimar que o tempo médio de falha é de 47 dias. A Tabela 2 apresenta a quantidade de falhas esperada, com base em diferentes quantis relacionados ao tempo de falha. Desta forma, podemos afirmar que, se a manutenção preventiva for feita assumindo que a porcentagem aceitável de falhas do sistema II é de 25%, então serão 8 dias para ocorrer essa quantidade. Assim, para que não haja uma quantidade acima de 25%, as manutenções devem ser feitas após 8 dias da última manutenção.

Tabela 2 – Tabela de frequência de ocorrência de falhas do subsistemaII por quartil determinado.

Quantil	10%	25%	50%	75%
Em dias	2	8	18	43

3.5.2 Tempo de falha do subsistema 1J

De forma análoga à seção anterior, a Tabela 3 apresenta os resultados dos critérios AIC, AICC, HQIC e CAIC para as diferentes distribuições de probabilidade ajustadas, considerando os dados do subsistema 1J.

Tabela 3 – Resultados dos Critérios AIC, AICC, HQIC e CAIC de diferentes distribuições de probabilidade considerando os dados do subsistema 1J.

Critérios	Log-normal	Weibull	Gama
AIC	1408,443	1452,372	1474,438
AICC	1408,505	1452,434	1474,501
HQIC	1411,089	1455,018	1477,084
CAIC	1416,978	1460,907	1482,974

Por meio dos resultados acima verificamos que todos critérios nos levam a afirmar que, dentre os modelos considerados, a distribuição Log-normal foi a que melhor se ajustou aos dados de falha do subsistema 1J. Os EMVs obtidos foram $\mu = 2,09352$ e $\sigma = 1,53036$.

Temos, através da Figura 19 o gráfico TTT-plot, a curva K-M, a função de sobrevivência e a função de risco, ambas ajustadas através da distribuição Log-normal.

Figura 19 – TTT-plot, função densidade de probabilidade e a função de risco, ambas ajustadas através da distribuição Log-normal, utilizando-se os dados do subsistema 1J.

Com base nos parâmetros ajustados da distribuição Log-normal, conseguimos estimar que o tempo médio de falha é de 26 dias. A Tabela 4 apresenta a quantidade de falhas esperadas, com base em diferentes quantis relacionados ao tempo de falha do subsistema 1J. Por meio dos resultados obtidos verificamos que, se a manutenção preventiva for feita assumindo que porcentagem aceitável de falhas do sistema 1J é de 25%, então serão necessários apenas 2 dias para ocorrer esta quantidade de falhas. Assim, para que não haja uma quantidade acima de 25%, as manutenções devem ser feitas logo após 2 dias da última manutenção.

Tabela 4 – Tabela de frequência de ocorrência de falhas do subsistema 1J por quartil determinado.

Quantil	10%	25%	50%	75%
Em dias	1	2	7	20

3.5.3 Tempo de falha do subsistema 3E

Considerando os tempos de falha do subsistema 3E a Tabela ?? apresenta os resultados dos critérios AIC, AICC, HQIC e CAIC para os dados ajustados sob a suposição que seguem as distribuições Gamma, Weibull e Lognormal.

Tabela 5 – Resultados dos Critérios AIC, AICC, HQIC e CAIC de diferentes distribuições de probabilidade, considerando os dados de falha do subsistema 3E.

Critérios	Log-normal	Weibull	Gama
AIC	943,729	941,548	942,220
AICC	939,842	937,661	938,333
HQIC	945,912	943,731	944,403
CAIC	951,112	948,930	949,603

Neste caso, com base em qualquer um dos critérios utilizados, podemos concluir que, através da Tabela 5, temos a indicação de que a distribuição Weibull foi a que melhor se ajustou aos dados de falha do subsistema 3E. Os EMVs para o conjunto de dados proposto são $\beta = 42,75045$ e $\alpha = 0,876715$.

Pode-se observar, através da Figura 20 gráfico TTT-plot, a curva K-M, a função de sobrevivência e a função de risco, ambas ajustadas através da distribuição Weibull.

Figura 20 – TTT-plot, função densidade de probabilidade e a função de risco, ambas ajustadas através da distribuição Weibull, utilizando os dados de falha do sistema 3E.

Utilizando os parâmetros ajustados para a distribuição Weibull, temos que o tempo médio

de falha é de 46 dias. A Tabela 6 apresenta a quantidade de falhas esperada, com base em diferentes quantis relacionados ao tempo de falha do subsistema 3E.

Tabela 6 – Tabela de frequência de ocorrência de falhas subsistema 3E por quantil determinado.

Quantil	10%	25%	50%	75%
Em dias	3	9	27	58

Assumindo novamente que 25% é a porcentagem aceitável de falhas para o sistema proposto, temos que serão necessários 9 dias para ocorrer esta quantidade de falhas. Logo, as manutenções devem ser feitas logo após 9 dias da última manutenção.

3.5.4 Tempo de falha do subsistema 3F

Por fim, para o conjunto de dados relacionados ao subsistema 3F, a Tabela 7 apresenta os resultados dos critérios AIC, AICC, HQIC e CAIC de diferentes distribuições de probabilidade, considerando os dados subsistema 3F.

Tabela 7 – Resultados dos Critérios AIC, AICC, HQIC e CAIC de diferentes distribuições de probabilidade, considerando os dados do subsistema 3F.

Critérios	Log-normal	Weibull	Gama
AIC	1209,648	1222,990	1226,400
AICC	1205,731	1219,073	1222,483
HQIC	1212,084	1225,425	1228,836
CAIC	1217,643	1230,984	1234,395

Baseado em qualquer um dos critérios utilizados, podemos concluir, através da Tabela 7, que a distribuição Log-normal foi a que melhor se ajustou aos dados de falha do subsistema 3F. Os EMVs obtidos foram $\mu = 2,72770$ e $\sigma = 1,36243$.

Pode-se observar, através da Figura 21 o gráfico TTT-plot, a curva K-M, a função de sobrevivência e a função de risco, ambas ajustadas através da distribuição Log-normal.

Por meio dos parâmetros ajustados da distribuição Log-normal, estimamos que o tempo médio de falha é de 39 dias. A Tabela 8 apresenta a quantidade de falhas esperadas, com base em diferentes quantis relacionados ao tempo de falha do subsistema 3F.

Tabela 8 – Tabela de frequência de ocorrência de falhas geradas pelo subsistema 3F do sistema 3 por quantil determinado.

Quantil	10%	25%	50%	75%
Em dias	2	5	16	44

Supondo que 25% é uma porcentagem aceitável de falhas para o sistema proposto, temos que serão necessários 5 dias para ocorrer esta quantidade de falhas. Assim, as manutenções devem ser realizadas após 5 dias da última manutenção.

Figura 21 – TTT-plot, função densidade de probabilidade e a função de risco, ambas ajustadas através da distribuição Log-normal, utilizando-se os dados de falha gerado pelo subsistema 3F.

A partir dos resultados apresentados, notamos que as distribuições Log-normal e Weibull podem ser utilizadas para descrever o tempo de vida dos subsistemas estudados. Nesse caso, a distribuição Log-normal destacou-se em três dos quatro casos. Por outro lado, para os tempos da falha do subsistema 1I, discutido no C capítulo 2, verificamos que uma outra distribuição se mostra adequada para representar esse conjunto de dados. Tal distribuição não foi proposta ainda na literatura e nomeamos a mesma de distribuição Inversa Lindley Ponderada. As propriedades matemáticas, métodos de inferência e aplicação serão apresentadas no capítulo a seguir.

DISTRIBUIÇÃO INVERSA LINDLEY PONDERADA

Uma importante distribuição para a modelagem de diversos tipos de fenômenos foi introduzida por [Lindley \(1958\)](#), comumente conhecida por distribuição Lindley e introduzida no contexto de estatística fiducial e Teorema de Bayes. [Ghitany, Atieh e Nadarajah \(2008\)](#) mostraram que a distribuição Lindley fornece ajustes mais flexíveis, superando a distribuição Exponencial em diversos tipos de situação. Desde então, muitas generalizações da distribuição Lindley têm sido propostas, dentre elas podemos citar a distribuição Lindley Generalizada, Lindley Estendida ([BAKOUCH *et al.*, 2012](#)) e distribuição Power Lindley ([GHITANY *et al.*, 2013](#)). O estudo de novas distribuições de probabilidade fornece melhor compreensão das distribuições básicas e possibilita adicionar mais flexibilidade aos ajustes dos dados ([PATIL; RAO, 1978](#)). [Ghitany *et al.* \(2011\)](#) propõem uma distribuição de probabilidade Lindley Ponderada (LP) com dois parâmetros, com forma de banheira e taxa de risco crescente. A distribuição LP possui a função densidade de probabilidade dada por:

$$f(t|\phi, \lambda) = \frac{\lambda^{\phi+1}}{(\phi + \lambda)\Gamma(\phi)} t^{\phi-1} (1+t) e^{-\lambda t}, \quad (4.1)$$

para todo $t > 0$, $\lambda > 0$, $\phi > 0$ em que $\Gamma(\phi) = \int_0^\infty e^{-x} x^{\phi-1} dx$ é a função Gama.

[Mazucheli, Louzada e Ghitany \(2013\)](#) comparou diferentes métodos de estimação para os parâmetros da distribuição LP e realizou um estudo de simulação numérica para encontrar o melhor estimador. Em seguida, [Wang e Wang \(2017\)](#) apresentou uma correção de vício para os estimadores de máxima verossimilhança (EMVs) e demonstrou, através de simulação, que os estimadores propostos são superiores aos outros estimadores sem a correção de vício. Do ponto de vista Bayesiano, [Ali \(2015\)](#) considerou e derivou diferentes prioris informativas e não-informativas sob diferentes funções de perdas. Por fim, [Ramos e Louzada \(2016\)](#) introduziram um parâmetro adicional na distribuição LP e demonstraram que ela possui melhores ajustes, em muitas situações, do que distribuições de três parâmetros comumente utilizadas na literatura.

Neste capítulo introduzimos uma nova distribuição de dois parâmetros, em que a função de risco possui comportamento unimodal. Tal distribuição foi denominada de distribuição Inversa Lindley Ponderada (ILP), já que o novo modelo pode ser reescrito como a Inversa da distribuição Lindley Ponderada. Além disso, tal distribuição é uma generalização da distribuição Inversa Lindley (veja [Sharma *et al.* \(2015\)](#) para mais detalhes) e pode ser representada como um modelo de mistura. Desta forma, essa nova distribuição também pode ser utilizada para descrever a variabilidade entre dois grupos distintos com uma variável latente. Por exemplo, nós podemos estar interessados no tempo de vida de componentes que são compostos por itens novos e reparados, porém não há informação de quais são novos ou não. Assim, modelos com esta característica podem, geralmente, ser representados como modelos de mistura.

Para o modelo proposto, apresentamos diferentes expressões matemáticas, entre elas: os momentos, a média e a variância. Apresentamos também suas respectivas funções de sobrevivência e risco. Os estimadores dos parâmetros são obtidos por meio do método da máxima verossimilhança. Como os estimadores são viciados para amostras pequenas, derivamos as expressões analíticas dos vícios para obtenção de estimadores que são livres de vício até a segunda ordem. Por fim, uma aplicação real relativa ao subsistema 1J é apresentada, para demais subsistemas a distribuição não apresentou um bom ajuste.

4.1 Distribuição Inversa Lindley Ponderada

Uma variável aleatória não negativa T segue uma distribuição ILP com parâmetros $\phi > 0$ e $\lambda > 0$ se sua f.d.p é dada por

$$f(t|\phi, \lambda) = \frac{\lambda^{\phi+1}}{(\phi + \lambda)\Gamma(\phi)} t^{-\phi-1} \left(1 + \frac{1}{t}\right) e^{-\lambda t^{-1}}. \quad (4.2)$$

A distribuição ILP pode ser expressa como um modelo de mistura representado por:

$$f(t|\phi, \lambda) = p f_1(t|\phi, \lambda) + (1 - p) f_2(t|\phi, \lambda),$$

em que $p = \lambda/(\lambda + \phi)$ e $T_j \sim \text{IG}(\phi + j - 1, \lambda)$, para $j = 1, 2$, ou seja, $f_j(t|\lambda, \phi)$ possui distribuição Inversa Gama, dada por

$$f_j(t|\phi, \lambda) = \frac{\lambda^{\phi+j-1}}{\Gamma(\phi + j - 1)} t^{-\phi-j} e^{-\lambda t^{-1}}.$$

Desta forma, a distribuição ILP é um modelo de mistura podendo descrever melhor a heterogeneidade nos dados.

Proposição 4.1.1. Considerando $T \sim \text{ILP}(\phi, \lambda)$, segue $X = 1/T$ temos a distribuição Lindley Ponderada ([GHITANY *et al.*, 2011](#)).

Demonstração. Considere a transformação $X = g(T) = \frac{1}{T}$ em que T segue uma distribuição ILP. Desta forma, temos que a f.d.p de X é dada por:

$$\begin{aligned} f_X(x) &= f_T(g^{-1}(x)) \left| \frac{d}{dx} g^{-1}(x) \right| = \frac{\lambda^{\phi+1}}{(\phi + \lambda)\Gamma(\phi)} x^{\phi+1} (1+x) e^{-\lambda x} \frac{1}{x^2} \\ &= \frac{\lambda^{\phi+1}}{(\phi + \lambda)\Gamma(\phi)} x^{\phi-1} (1+x) e^{-\lambda x}. \end{aligned}$$

□

Note que, quando $\phi = 1$, a distribuição ILP se reduz para a Inversa da distribuição Lindley (SHARMA *et al.*, 2015) A Figura 22 apresenta o comportamento da f.d.p. para diferentes valores de ϕ e λ .

Figura 22 – Função densidade para distribuição ILP considerando diferentes valores de ϕ and λ .

A função de distribuição acumulada da ILP é dada por

$$F(t|\phi, \lambda) = \frac{\Gamma(\phi, \lambda t^{-1}) (\lambda + \phi) + (\lambda t^{-1})^\phi e^{-\lambda t^{-1}}}{(\lambda + \phi)\Gamma(\phi)},$$

em que $\Gamma(x, y) = \int_x^\infty w^{y-1} e^{-w} dw$ é conhecida como função gama incompleta superior.

4.2 Propriedades da Distribuição ILP

Nesta seção são apresentadas diferentes propriedades matemáticas para a distribuição proposta.

4.2.1 Momentos

Os momentos desempenham um papel fundamental na Estatística. Eles podem ser usados em muitas aplicações. Por exemplo, o primeiro momento de uma f.d.p é usado para determinar a

média da função, enquanto o segundo momento, conjuntamente com o primeiro, é usado para obter a variância, assimetria e curtose. A seguir, derivamos os momentos da distribuição ILP.

Proposição 4.2.1. Para a variável aleatória T com distribuição ILP, o r -ésimo momento é dado por:

$$\mu_r = E[T^r] = \frac{\lambda^r(\phi + \lambda - r)}{(\lambda + \phi)(\phi - 1)(\phi - 2) \cdots (\phi - r)}, \quad \text{em que } \phi > r. \quad (4.3)$$

Demonstração. Suponha que $W \sim \text{IG}(\phi, \lambda)$ então o r -ésimo momento da variável aleatória W é dado por:

$$E_{(\phi, \lambda)}[W^r] = \frac{\lambda^r \Gamma(\phi - r)}{\Gamma(\phi)} = \frac{\lambda^r}{(\phi - 1)(\phi - 2) \cdots (\phi - r)}, \quad \text{em que } \phi > r.$$

Como supracitado, a distribuição ILP pode ser escrita como a mistura de dois componentes. Portanto, temos:

$$\begin{aligned} \mu_r = E[T^r] &= \int_0^\infty t^r f(t|\phi, \lambda) dt = pE_{(\phi, \lambda)}[W^r] + (1-p)E_{(\phi+1, \lambda)}[W^r] \\ &= \frac{\lambda}{(\lambda + \phi)} \frac{\Gamma(\phi - r)}{\Gamma(\phi)} + \frac{\phi}{(\lambda + \phi)} \frac{\Gamma(\phi + 1 - r)}{\Gamma(\phi + 1)} = \frac{\lambda^r(\lambda + \phi - r)\Gamma(\phi - r)}{(\lambda + \phi)\Gamma(\phi)} \\ &= \frac{\lambda^r(\phi + \lambda - r)}{(\lambda + \phi)(\phi - 1)(\phi - 2) \cdots (\phi - r)}, \quad \text{em que } \phi > r. \end{aligned}$$

□

Proposição 4.2.2. O r -ésimo momento central para uma variável aleatória T é dado por:

$$\begin{aligned} M_r = E[T - \mu]^r &= \sum_{i=0}^r \binom{r}{i} (-\mu)^{r-i} E[T^i] \\ &= \sum_{i=0}^r \binom{r}{i} \left(-\frac{\lambda(\phi + \lambda - 1)}{(\lambda + \phi)(\phi - 1)} \right)^{r-i} \left(\frac{\lambda^i(\phi + \lambda - i)}{(\lambda + \phi)(\phi - 1)(\phi - 2) \cdots (\phi - i)} \right). \end{aligned} \quad (4.4)$$

Demonstração. O resultado é obtido diretamente da proposição 4.2.1. □

Proposição 4.2.3. Dada uma variável aleatória T com distribuição ILP, sua média e variância são dadas por:

$$\begin{aligned} \mu &= \frac{\lambda(\phi + \lambda - 1)}{(\lambda + \phi)(\phi - 1)}, \\ \sigma^2 &= \frac{\lambda^2 \left((\phi + \lambda - 2)(\phi - 1) - (\phi + \lambda - 1)^2(\phi - 2) \right)}{(\lambda + \phi)(\phi - 2)(\phi - 1)^2}. \end{aligned}$$

Demonstração. A partir do resultado (4.3) e considerando $r = 1$, temos que $\mu_1 = \mu$. Fazendo a soma algébrica do segundo resultado a partir da equação (4.4) e considerando $r = 2$ a prova está completa. □

4.2.2 Propriedades de Sobrevivência

A função de sobrevivência da distribuição ILP que representa a probabilidade de uma observação não falhar em um determinado tempo t é dada por:

$$S(t|\phi, \lambda) = \frac{\gamma(\phi, \lambda t^{-1}) (\lambda + \phi) - (\lambda t^{-1})^\phi e^{-\lambda t^{-1}}}{(\lambda + \phi)\Gamma(\phi)},$$

em que $\gamma(y, x) = \int_0^x w^{y-1} e^{-w} dw$ é a função Gama incompleta inferior. A função de risco de T é dada por:

$$h(t|\phi, \lambda) = \frac{\lambda^{\phi+1} t^{-\phi-1} (1 + t^{-1}) e^{-\lambda t^{-1}}}{\gamma(\phi, \lambda t^{-1}) (\lambda + \phi) - (\lambda t^{-1})^\phi e^{-\lambda t^{-1}}}. \quad (4.5)$$

Este modelo possui a função de risco com curva de banheira invertida (unimodal). Para provar tal resultado considere o lema a seguir.

Lema 4.2.4. (GLASER, 1980): Seja T uma variável aleatória não negativa e contínua, em que sua f.d.p. é diferenciável de segunda ordem em t , e seja $\eta(t) = -\frac{\partial}{\partial t} \log f(t)$, então se $\eta(t)$ tem forma de banheira invertida, $h(t)$ também possui curva da banheira invertida.

Teorema 4.2.5. A função de risco da distribuição ILP possui forma de banheira invertida para todo $\phi > 0$ e $\lambda > 0$.

Demonstração. Para distribuição ILP, temos:

$$\eta(t) = \frac{\phi}{t} + \frac{2}{t} - \frac{1}{(t+1)} - \frac{\lambda}{t^2}.$$

Assim,

$$\eta'(t) = -\frac{\phi}{t^2} - \frac{2}{t^2} + \frac{1}{(t+1)^2} + \frac{2\lambda}{t^3}.$$

O estudo do comportamento de $\eta'(t)$ não é simples. No entanto, usando o Software Wolfram|Alpha, podemos verificar que para todo $\phi > 0$ e $\lambda > 0$, $\eta'(t)$ cresce em $(0, \xi(t|\phi, \lambda))$ e decresce em $(\xi(t|\phi, \lambda), \infty)$, ou seja, $\eta'(t) = 0$ para $\xi(t|\phi, \lambda)$, em que $\xi(t|\phi, \lambda)$ é uma função muito extensa calculada diretamente pelo Wolfram|Alpha. Portanto, $\eta(t)$ e, consequentemente $h(t)$, possuem curva em forma de banheira invertida. $\square$

Esta propriedade faz da distribuição ILP um modelo flexível para ajustar dados de confiabilidade. A Figura 23 apresenta exemplos de diferentes curvas para função de risco.

Proposição 4.2.6. A função de vida média residual, denotada por $r(t|\phi, \lambda)$, para a distribuição ILP é dada por:

$$r(t|\phi, \lambda) = \frac{1}{S(t)} \int_t^\infty y f(y|\lambda, \phi) dy - t = \frac{\lambda \gamma(\phi, \lambda t^{-1}) + \lambda^2 \gamma(\phi, \lambda t^{-1})}{\gamma(\phi, \lambda t^{-1}) (\lambda + \phi) - (\lambda t^{-1})^\phi e^{-\lambda t^{-1}}} - t.$$

Figura 23 – Curva de risco para a distribuição ILP considerando diferentes valores de ϕ e λ .

Demonstração. Note que, para a distribuição Inversa Gama, temos que:

$$\int_t^\infty y f_j(y|\phi, \lambda) dy = \frac{\lambda}{\Gamma(\phi + j - 1)} \gamma[\phi + j - 2, \lambda t^{-1}], \quad j = 1, 2.$$

Usando a equação a seguir,

$$r(t|\phi, \lambda) = \frac{1}{S(t)} \left[p \int_t^\infty y f_1(y|\lambda, \phi) dy + (1 - p) \int_t^\infty y f_2(y|\lambda, \phi) dy \right] - t,$$

e, após algumas somas algébricas, a prova está completa. $\square$

4.2.3 Entropia

Na teoria da informação, a medida de entropia desempenha um papel fundamental para mensurar a incerteza associada a uma variável aleatória. A entropia de Shannon é uma importante métrica utilizada na teoria da informação. Para a distribuição ILP, tal medida de informação é dada pela solução da equação a seguir:

$$H_S(\phi, \lambda) = - \int_0^\infty \log \left(\frac{\lambda^{\phi+1}}{(\phi + \lambda)\Gamma(\phi)} t^{-\phi-1} \left(1 + \frac{1}{t} \right) e^{-\lambda t^{-1}} \right) f(t|\phi, \lambda) dt. \quad (4.6)$$

Proposição 4.2.7. Uma variável aleatória T com distribuição ILP, tem a Entropia de Shannon dada por:

$$H_S(\phi, \lambda) = \log(\lambda + \phi) + \log \Gamma(\phi) + \frac{\phi(\lambda + \phi + 1)}{(\lambda + \phi)} - (\phi + 1) \left(\frac{1}{\lambda + \phi} + \lambda(\phi) \right) - \frac{\lambda^{\phi+1} \Omega(\phi, \lambda)}{(\lambda + \phi)\Gamma(\phi)},$$

em que $\Omega(\phi, \lambda) = \int_0^\infty (x + 1) \log(x + 1) x^{(\phi-1)} e^{-\lambda x} dx$.

Demonstração. A partir da equação (4.6), temos:

$$H_S(\phi, \lambda) = (\phi + 1) \log \lambda - \log(\lambda + \phi) - \log \Gamma(\phi) - \lambda E[t_i^{-1}] - (\phi + 1) E[\log(t_i)] + E[\log(1 + t_i^{-1})].$$

Note que:

$$\begin{aligned} E[\log(t)] &= \log(\lambda) - \frac{1}{\lambda + \phi} - \lambda(\phi), \\ E[t_i^{-1}] &= \frac{(\phi + 1)}{\lambda} - \frac{1}{\lambda + \phi} = \frac{\phi(\lambda + \phi + 1)}{\lambda(\lambda + \phi)}, \quad e \\ E[\log(1 + t_i^{-1})] &= \frac{\lambda^{\phi+1}}{(\lambda + \phi)\Gamma(\phi)} \int_0^\infty (x + 1) \log(x + 1) x^{(\phi-1)} e^{-\lambda x} dx. \end{aligned}$$

Então,

$$\begin{aligned} H_S(\phi, \lambda) &= (\phi + 1) \left(\frac{1}{\lambda + \phi} + \lambda(\phi) \right) + \log(\lambda + \phi) + \log \Gamma(\phi) + \frac{\phi(\lambda + \phi + 1)}{(\lambda + \phi)} \\ &\quad - \frac{\lambda^{\phi+1} \Omega(\phi, \lambda)}{(\lambda + \phi)\Gamma(\phi)}, \end{aligned}$$

como queríamos demonstrar. $\square$

4.3 Inferência

Nesta seção discutimos a estimação dos parâmetros ϕ e λ da distribuição ILP por meio do estimador de máxima verossimilhança. Adicionalmente, consideramos tal abordagem na presença de observações com censura aleatória.

4.3.1 Estimadores de Máxima Verossimilhança

Dentre os métodos de inferência estatística, o método de máxima verossimilhança é amplamente utilizado devido a suas boas propriedades assintóticas. Sob o método de máxima verossimilhança, os estimadores são obtidos a partir da maximização da função de verossimilhança. Considere uma amostra aleatória e identicamente distribuída $T_1, \dots, T_n$ em que $T \sim \text{ILP}(\phi, \lambda)$, a função de verossimilhança de (4.2) é dada por:

$$L(\theta; \mathbf{t}) = \frac{\lambda^{n(\phi+1)}}{(\phi + \lambda)^n \Gamma(\phi)^n} \left\{ \prod_{i=1}^n t_i^{-\phi-1} \right\} \prod_{i=1}^n \left(1 + \frac{1}{t_i} \right) \exp \left\{ -\lambda \sum_{i=1}^n \frac{1}{t_i} \right\}.$$

A função log-verossimilhança $l(\theta; \mathbf{t}) = \log L(\theta; \mathbf{t})$ é dada por:

$$l(\theta; \mathbf{t}) = n(\phi + 1) \log \lambda - n \log(\lambda + \phi) - n \log \Gamma(\phi) - \lambda \sum_{i=1}^n \frac{1}{t_i} - (\phi + 1) \sum_{i=1}^n \log(t_i). \quad (4.7)$$

Por meio das derivadas parciais $\frac{\partial}{\partial \phi} l(\theta; \mathbf{t}) = 0$, $\frac{\partial}{\partial \lambda} l(\theta; \mathbf{t}) = 0$, temos que as equações de verossimilhança são:

$$n \log(\lambda) - \sum_{i=1}^n \log(t_i) - \frac{n}{\lambda + \phi} - n\psi(\phi) = 0,$$

$$\frac{n(\phi + 1)}{\lambda} - \sum_{i=1}^n \frac{1}{t_i} - \frac{n}{\lambda + \phi} = 0,$$

em que $\psi(k) = \frac{\partial}{\partial k} \log \Gamma(k) = \frac{\Gamma'(k)}{\Gamma(k)}$ é a função Poligama. Após algumas somas algébrica,s a solução de λ_{MLE} é dada por:

$$\hat{\lambda}_{MLE} = \frac{-\hat{\phi}_{MLE} (\xi(t) - 1) + \sqrt{(\hat{\phi}_{MLE} (\xi(t) - 1))^2 + 4 \xi(t) (\hat{\phi}_{MLE}^2 + \hat{\phi}_{MLE})}}{2 \xi(t)},$$

em que $\xi(t) = \sum_{i=1}^n (nt_i)^{-1}$ e $\hat{\phi}_{MLE}$ podem ser obtidos solucionando o sistema não-linear:

$$n \log(\hat{\lambda}_{MLE}) - \sum_{i=1}^n \log(t_i) - \frac{n}{\hat{\lambda}_{MLE} + \hat{\phi}_{MLE}} - n \psi(\hat{\phi}_{MLE}) = 0. \quad (4.8)$$

Estes resultados são uma simples modificação dos resultados obtidos por [Ghitany, Atieh e Nadarajah \(2008\)](#) para a distribuição de LP.

Sob algumas condições de regularidade, os EMVs são consistentes e assintoticamente eficientes com distribuição normal bivariada expressa por meio de:

$$(\hat{\phi}, \hat{\lambda}) \sim N_2[(\phi, \lambda), I^{-1}(\phi, \lambda)] \text{ para } n \rightarrow \infty,$$

em que os elementos da matriz de informação de Fisher $I(\phi, \lambda)$ são dados por:

$$h_{11}(\phi, \lambda) = -\frac{n}{(\lambda + \phi)^2} + n \psi'(\phi),$$

$$h_{12}(\phi, \lambda) = h_{21}(\phi, \lambda) = -\frac{n}{\lambda} - \frac{n}{(\lambda + \phi)^2},$$

$$h_{22}(\phi, \lambda) = \frac{n(\phi + 1)}{\lambda^2} - \frac{n}{(\lambda + \phi)^2}$$

e $\psi'(k) = \frac{\partial}{\partial^2 k} \log \Gamma(k)$ é a função trigama. Uma propriedade interessante da distribuição ILP é que a matriz de informação observada é igual a matriz de informação esperada.

4.3.2 Censura Aleatória

Em análise de sobrevivência, uma característica comum é a presença de observações censuradas. Suponha que i -ésimo componente tenha um tempo de vida T_i e um tempo de censura C_i ; além disso os tempos de censura aleatória são independentes de T_i e sua distribuição não depende dos parâmetros; então o conjunto de dados é dado por (t_i, δ_i) , em que $t_i = \min(T_i, C_i)$ e $\delta_i = I(T_i \leq C_i)$. Este tipo de censura tem como caso especial os tipos I e II de mecanismo de censura. Neste caso, a função de verossimilhança para θ é dada por:

$$L(\theta, t) = \prod_{i=1}^n f(t_i | \theta)^{\delta_i} S(t_i | \theta)^{1-\delta_i}.$$

Seja $T_1, \dots, T_n$ uma amostra aleatória i.i.d. da distribuição ILP, a função de verossimilhança considerando censura aleatória é dada por:

$$L(\lambda, \phi | t) = \frac{\lambda^{n(\phi+1)}}{(\lambda + \phi)^n \Gamma(\phi)^n} \prod_{i=1}^n \left((\lambda + \phi) \gamma(\phi, \lambda t_i^{-1}) - (\lambda t_i^{-1})^\phi e^{-\lambda t_i^{-1}} \right)^{1-\delta_i} \times \left(t_i^{-\phi-1} (1 + t_i^{-1}) e^{-\lambda t_i^{-1}} \right)^{\delta_i}. \quad (4.9)$$

O logaritmo da função de verossimilhança é dado por:

$$\begin{aligned} l(\lambda, \phi | t) = & -(\phi + 1) \sum_{i=1}^n \delta_i \log(t_i) - \lambda \sum_{i=1}^n \delta_i t_i^{-1} + n(\phi + 1) \log(\lambda) - n \log(\phi + \lambda) \\ & + \sum_{i=1}^n (1 - \delta_i) \log \left((\lambda + \phi) \gamma(\phi, \lambda t_i^{-1}) - (\lambda t_i^{-1})^\phi e^{-\lambda t_i^{-1}} \right) - n \log(\Gamma(\phi)) \\ & + \sum_{i=1}^n \delta_i \log(1 + t_i^{-1}). \end{aligned} \quad (4.10)$$

Considerando as derivadas parciais $\partial l(\lambda, \phi | t) / \partial \lambda = 0$ e $\partial l(\lambda, \phi | t) / \partial \phi = 0$, as equações de verossimilhança são apresentadas a seguir:

$$\begin{aligned} \sum_{i=1}^n \frac{(1 - \delta_i) \left(\gamma(\phi, \lambda t_i^{-1}) + (\lambda + \phi) (\lambda t_i^{-1})^{\phi-1} e^{-\lambda t_i^{-1}} - \phi \lambda^{\phi-1} t_i^{-\phi} e^{-\lambda t_i^{-1}} - (\lambda t_i^{-1})^{\phi+1} e^{-\lambda t_i^{-1}} \right)}{((\lambda + \phi) \gamma(\phi, \lambda t_i^{-1})) - (\lambda t_i^{-1})^\phi e^{-\lambda t_i^{-1}}} = \\ \frac{n}{\lambda + \phi} - \frac{n(\phi + 1)}{\lambda} + \sum_{i=1}^n \delta_i t_i^{-1}, \end{aligned} \quad (4.11)$$

$$\begin{aligned} \sum_{i=1}^n \frac{(1 - \delta_i) \left(\gamma(\phi, \lambda t_i^{-1}) + (\lambda + \phi) \Psi(\phi, \lambda t_i^{-1}) - (\lambda t_i^{-1})^\phi \log(\lambda t_i^{-1}) e^{-\lambda t_i^{-1}} \right)}{((\lambda + \phi) \gamma(\phi, \lambda t_i^{-1})) - (\lambda t_i^{-1})^\phi e^{-\lambda t_i^{-1}}} = -n \log(\lambda) \\ + \frac{n}{\lambda + \phi} + n \psi(\phi) + \sum_{i=1}^n \delta_i \log(t_i^{-1}), \end{aligned} \quad (4.12)$$

em que $\Psi(k, x) = \partial \gamma(k, x) / \partial k$ pode ser calculado numericamente. Além disto, métodos numéricos são necessários para encontrar as soluções das equações não-lineares.

4.4 Correção de vício para estimadores de máxima verossimilhança

Nesta seção apresentamos EMVs modificados baseados em duas aproximações obtidas por meio de métodos de correção de vício. Primeiramente, uma correção analítica será apresentada com base na teoria apresentada por [Cox e Snell \(1968\)](#). Em seguida, iremos considerar uma correção por meio de técnicas *bootstrap* (veja [Efron \(1979\)](#)).

4.4.1 Aproximação corretiva

Considere a função de verossimilhança $L(\theta; \mathbf{t})$ com um vetor p -dimensional de parâmetros θ . Agora, seja as esperanças dadas pelas derivadas: $l(\theta; t)$ pode ser escrito por

$$h_{ij}(\theta) = E \left(\frac{\partial^2 l(\theta; t)}{\partial \theta_i \partial \theta_j} \right), \quad h_{ijl}(\theta) = E \left(\frac{\partial^3 l(\theta; t)}{\partial \theta_i \partial \theta_j \partial \theta_l} \right) e$$

$$h_{ij,l}(\theta) = E \left(\frac{\partial^2 l(\theta; t)}{\partial \theta_i \partial \theta_j} \cdot \frac{\partial l(\theta; t)}{\partial \theta_l} \right), \quad \text{para } i, j, l = 1, \dots, p.$$

assim como suas respectivas derivadas notadas por:

$$h_{ij}^{(l)}(\theta) = \frac{\partial h_{ij}(\theta)}{\partial \theta_l}, \quad \text{for } i, j, l = 1, \dots, p.$$

Uma expressão analítica para a correção de vício de θ_m foi discutida por [Cox e Snell \(1968\)](#), utilizando como suposição que os dados provém de uma amostra aleatória independente, mas sem necessariamente serem distribuídos identicamente. Tal expressão é dada por:

$$Vicio(\hat{\theta}_m) = \sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p s_{mi}(\theta) s_{jl}(\theta) (h_{ij,l}(\theta) + 0.5 h_{ijl}(\theta)) + O(n^{-2}), \quad (4.13)$$

onde s^{ij} é o (i, j) -ésimo elemento da matriz de informação de Fisher $\hat{\theta}$, $K = \{-h_{ij}\}$.

[Cordeiro e Klein \(1994\)](#) provaram que, mesmo se os dados são dependentes, a expressão (4.13) pode ser reescrita como:

$$Vicio(\hat{\theta}_m) = \sum_{i=1}^p s_{mi}(\theta) \sum_{j=1}^p \sum_{k=1}^p s_{jl}(\theta) (h_{ij}^{(l)}(\theta) - 0.5 h_{ijl}(\theta)) + O(n^{-2}). \quad (4.14)$$

Considerando $a_{ij}^{(l)} = h_{ij}^{(l)} - \frac{1}{2} h_{ijl}^{(l)}$ e definindo a matriz $A = [A^{(1)} | A^{(2)} | \dots | A^{(p)}]$ com $A^{(l)} = \{a_{ij}^{(l)}\}$, para $i, j, l = 1, \dots, p$, temos que a expressão para o vício de $\hat{\theta}$ pode ser reescrita como:

$$Vicio(\hat{\theta}_m) = K^{-1} A \cdot \text{vec}(K^{-1}) + O(n^{-2}), \quad (4.15)$$

em que vec é um operador que cria um vetor coluna a partir da matriz.

Desta forma, um EMVs com vício corrigido para $\hat{\theta}$ é dado por:

$$\hat{\theta}_{CMLE} = \hat{\theta} - \hat{K}^{-1} \hat{A} \cdot \text{vec}(\hat{K}^{-1}), \quad (4.16)$$

em que $\hat{\theta}$ é o EMV do parâmetro θ , $\hat{K} = K|_{\theta=\hat{\theta}}$ e $\hat{A} = A|_{\theta=\hat{\theta}}$. O vício de $\hat{\theta}_{CMLE}$ decresce com ordem de $O(n^{-2})$.

Para distribuição ILP, as derivadas de ordem superior podem ser facilmente obtidas, já que não dependem de t . Portanto, temos:

$$h_{111}(\theta) = h_{11}^{(1)}(\theta) = -\frac{2n}{(\lambda + \phi)^3} - n\psi''(\phi),$$

$$h_{122}(\theta) = h_{221}(\theta) = h_{212}(\theta) = h_{12}^{(2)}(\theta) = h_{22}^{(1)}(\theta) = -\frac{2n}{(\lambda + \phi)^3} - \frac{n}{\lambda^2},$$

$$h_{222}(\theta) = h_{22}^{(2)}(\theta) = -\frac{2n}{(\lambda + \phi)^3} - \frac{2n(\phi + 1)}{\lambda^3} e$$

$$h_{211}(\theta) = h_{112}(\theta) = h_{121}(\theta) = h_{12}^{(1)}(\theta) = h_{11}^{(2)}(\theta) = -\frac{2n}{(\lambda + \phi)^3},$$

em que $\psi''(k) = \frac{\partial}{\partial^3 k} \log \Gamma(k)$. A matriz K é dada por:

$$K = \begin{bmatrix} \frac{n}{(\lambda + \phi)^2} - n\psi'(\phi) & \frac{n}{\lambda} + \frac{n}{(\lambda + \phi)^2} \\ \frac{n}{\lambda} + \frac{n}{(\lambda + \phi)^2} & -\frac{n(\phi + 1)}{\lambda^2} + \frac{n}{(\lambda + \phi)^2} \end{bmatrix}.$$

Para obter a matriz A de (4.15), apresentamos os elementos de $A^{(1)}$:

$$\begin{aligned} a_{11}^{(1)} &= h_{11}^{(1)} - \frac{1}{2}h_{111} = -\frac{n}{(\lambda + \phi)^3} - \frac{n\psi''(\phi)}{2}, \\ a_{12}^{(1)} &= a_{21}^{(1)} = h_{12}^{(1)} - \frac{1}{2}h_{112} = -\frac{n}{(\lambda + \phi)^3}, \\ a_{22}^{(1)} &= h_{22}^{(1)} - \frac{1}{2}h_{221} = -\frac{n}{(\lambda + \phi)^3} - \frac{n}{2\lambda^2}, \end{aligned}$$

e os elementos de $A^{(2)}$ são:

$$\begin{aligned} a_{11}^{(2)} &= h_{11}^{(2)} - \frac{1}{2}h_{112} = -\frac{n}{(\lambda + \phi)^3}, \\ a_{12}^{(2)} &= a_{21}^{(2)} = h_{12}^{(2)} - \frac{1}{2}h_{122} = -\frac{n}{(\lambda + \phi)^3} - \frac{n}{2\lambda^2}, \\ a_{22}^{(2)} &= h_{22}^{(2)} - \frac{1}{2}h_{222} = -\frac{n}{(\lambda + \phi)^3} - \frac{n(\phi + 1)}{\lambda^3}. \end{aligned}$$

Portanto, a matriz $A = [A^{(1)} | A^{(2)}]$ é expressa por:

$$A = n \begin{pmatrix} -\frac{1}{(\lambda + \phi)^3} - \frac{\psi''(\phi)}{2} & -\frac{1}{(\lambda + \phi)^3} & -\frac{1}{(\lambda + \phi)^3} & -\frac{1}{(\lambda + \phi)^3} - \frac{1}{2\lambda^2} \\ -\frac{1}{(\lambda + \phi)^3} & -\frac{1}{(\lambda + \phi)^3} - \frac{1}{2\lambda^2} & -\frac{1}{(\lambda + \phi)^3} - \frac{1}{2\lambda^2} & -\frac{1}{(\lambda + \phi)^3} - \frac{(\phi + 1)}{\lambda^3} \end{pmatrix}.$$

Logo, os EMVs com vício corrigido podem ser escritos como:

$$\begin{pmatrix} \hat{\phi}_{CMLE} \\ \hat{\lambda}_{CMLE} \end{pmatrix} = \begin{pmatrix} \hat{\phi}_{MLE} \\ \hat{\lambda}_{MLE} \end{pmatrix} - \hat{K}^{-1} \hat{A} \cdot \text{vec}(\hat{K}^{-1}), \quad (4.17)$$

em que $\hat{K} = K|_{\phi=\hat{\phi}, \lambda=\hat{\lambda}}$ e $\hat{A} = A|_{\phi=\hat{\phi}, \lambda=\hat{\lambda}}$. É importante ressaltar que as derivadas de ordem superior não dependem de t . Logo, elas são as mesmas já derivadas para a distribuição Lindley Ponderada (WANG; WANG, 2017).

Uma correção de vício para os estimadores na presença de censura também pode ser considerada. Apesar da matriz de informação de Fisher, na presença de dados censurados, não possuir forma fechada, podemos considerar as equações de correção de vício presentes para observações completas. Neste caso, os estimadores com vício corrigido aproximados dos EMVs (ACMLE) são dados por:

$$\begin{pmatrix} \hat{\phi}_{ACMLE} \\ \hat{\lambda}_{ACMLE} \end{pmatrix} = \begin{pmatrix} \hat{\phi}_{MLE} \\ \hat{\lambda}_{MLE} \end{pmatrix} - \hat{K}^{-1} \hat{A} \cdot \text{vec}(\hat{K}^{-1}),$$

em que $\hat{K} = K|_{\phi=\hat{\phi}_{MLE}, \lambda=\hat{\lambda}_{MLE}}$, $\hat{A} = A|_{\phi=\hat{\phi}_{MLE}, \lambda=\hat{\lambda}_{MLE}}$ e $\hat{\phi}_{MLE}$ and $\hat{\lambda}_{MLE}$ são a solução de (4.11) e (4.12). No entanto, como as expressões não foram derivadas diretamente da matriz de informação de Fisher, na presença de censura, a correção de vício de $\hat{\theta}_{ACMLE}$ não é um estimador de ordem $O(n^{-2})$.

4.4.2 Correção de vício por bootstrap

Utilizaremos o método *bootstrap* proposto por Efron (1979) para remover o vício dos EMVs em uma ordem $O(n^{-2})$. O vício corrigido dos EMVs é dado pela subtração do vício estimado em relação aos EMVs.

Seja $\mathbf{y} = (y_1, \dots, y_n)^\top$ uma amostra com n observações selecionadas aleatoriamente a partir da variável aleatória Y relacionada à função de distribuição acumulada $F = F_V(y)$. Temos aqui que o parâmetro v é uma função de F sendo desta forma dada por $v = t(F)$. Além disto, temos que $\hat{v}$ é um estimador de v baseado em $\mathbf{y}$, isto é, $\hat{v} = s(\mathbf{y})$. A pseudo-amostra $\mathbf{y}^* = (y_1^*, \dots, y_n^*)^\top$ é obtida a partir da amostra original $\mathbf{y}$ através do processo de amostragem com substituição. Com base nesta pseudo-amostra, o *bootstrap* replicado de $\hat{v}$ é calculado, sendo $\hat{v}^* = s(\mathbf{y}^*)$ e a distribuição acumulada empírica (ecdf) de $\hat{v}^*$ é usada para estimar $F_{\hat{v}}$ (cdf de $\hat{v}$). Definimos $B_F(\hat{v}, v)$ como o vício do estimador $\hat{v} = s(\mathbf{y})$, sendo este dado por:

$$B_F(\hat{v}, v) = E_F[\hat{v}, v] = E_F[s(\mathbf{y})] - v(F).$$

O vício estimado utilizando o *bootstrap* é obtido substituindo F por $F_{\hat{v}}$, em que F é gerado da amostra original. Portanto, o vício estimado de *bootstrap* é dado por:

$$\hat{B}_{F_{\hat{v}}}(\hat{v}, v) = E_{F_{\hat{v}}}[\hat{v}^*] - \hat{v}.$$

Se temos B amostras *bootstrap* $(\mathbf{y}^{*(1)}, \mathbf{y}^{*(2)}, \dots, \mathbf{y}^{*(B)})$, que são geradas independentes baseadas da amostra original $\mathbf{y}$, teremos, então, os respectivos estimadores *bootstrap* $(\hat{v}^{*(1)}, \hat{v}^{*(2)}, \dots, \hat{v}^{*(B)})$. Baseado nesses estimadores, é possível determinar o estimador *bootstrap* esperado $E_{F_{\hat{v}}}[\hat{v}^*]$, dado, aproximadamente, por:

$$\hat{v}^{*(.)} = \frac{1}{B} \sum_{i=1}^B \hat{v}^{*(i)}.$$

Portanto, o vício estimado via *bootstrap*, baseado em B replicações de $\hat{v}$ é $\hat{B}_F(\hat{v}, v) = \hat{v}^{*(.)} - \hat{v}$. Logo, os EMVs com vício corrigido por meio do *bootstrap* são dados por:

$$v^B = \hat{v} - \hat{B}_F(\hat{v}, v) = 2\hat{v} - \hat{v}^{*(.)}.$$

Neste caso, temos que v^B é denotado por $\hat{\theta}_{BOOT} = (\hat{\phi}_{BOOT}, \hat{\lambda}_{BOOT})^\top$.

4.5 Simulações Numéricas

Nesta seção, um estudo de simulação numérica é apresentado para comparar a eficiência do método de máxima verossimilhança (MLEs) em relação aos métodos de correção de vício (CMLEs) na presença de dados completos e censurados. A proposta da comparação é avaliada por média relativa de erros (MRE) e média relativa de erros quadrados (RMSE), dados por:

$$\text{MRE}_i = \frac{1}{N} \sum_{j=1}^N \frac{\hat{\theta}_{i,j}}{\theta_i}, \quad \text{RMSE}_i = \frac{1}{N} \sum_{j=1}^N \frac{(\hat{\theta}_{i,j} - \theta_i)^2}{\theta_i^2}, \quad \text{para } i = 1, 2,$$

em que $N = 10000$ é o número de estimativas obtidas através dos MLEs, CMLEs e do *bootstrap*. Estudos de comparação de métodos de estimação utilizando a mesma métrica têm sido considerados por outros autores (BAKOUCH *et al.*, 2017; RODRIGUES; LOUZADA; RAMOS, 2018; LOUZADA; RAMOS, 2018). A probabilidade de cobertura de 95% dos intervalos de confiança também é avaliada. Considerando essa abordagem, esperamos que o procedimento que retorne o melhor método de estimação produza MREs próximos de 1 e RMSE próximos de zero. Além disso, para um grande número de experimentos, usando um nível de confiança de 95%, a frequência de intervalos que cobrem os valores verdadeiros de θ deve ser próxima de 95%. Seguindo os mesmos passos de Reath, Dong e Wang (2016), utilizamos $B=1000$ para o método *bootstrap*. O programa foi desenvolvido em R pode ser obtido através de solicitação. A amostra aleatória de ILP foi gerada considerando o seguinte algoritmo:

1. Gere $U_i \sim \text{Uniform}(0, 1), i = 1, \dots, n$;
2. Gere $X_i \sim \text{IG}(\phi, \lambda), i = 1, \dots, n$;
3. Gere $Y_i \sim \text{IG}(\phi + 1, \lambda), i = 1, \dots, n$;
4. Se $U_i \leq p = \lambda / (\lambda + \phi)$, então $T_i = X_i$. Caso contrário, temos que $T_i = Y_i, i = 1, \dots, n$.

4.5.1 Dados Completos

O estudo de simulação é realizado considerando os valores: $\theta = ((0, 5; 2), (2; 4))$ e $n = (10, 20, \dots, 130)$. É importante ressaltar que resultados similares foram obtidos considerando diferentes escolhas de ϕ e λ . O procedimento *uniroot* disponível no software R foi considerado para encontrar a solução da equação não-linear (4.8). O vício corrigido é calculado através de (4.17) e por meio do bootstrap. As Figuras 24 e 25 apresentam os MREs, RMSEs e as probabilidades de cobertura com nível de confiança de 95% para os EMVs, CMLEs e *bootstrap* com diferentes valores de n .

Por meio das Figuras 24 e 25, observamos que os estimadores de ϕ e λ são assintoticamente não-viciados, isto é, conforme o tamanho da amostra cresce, os valores das MREs tendem a 1 e os RMSEs tendem a zero. O CMLE apresenta um desempenho superior em relação aos resultados obtidos por meio da técnica *bootstrap*, para ambos parâmetros, assim como

Figura 24 – MREs, MSEs relacionado para os estimadores de $\lambda = 2$ e $\phi = 4$ para $N = 10000$ amostras simuladas, considerando diferentes valores de n .

Figura 25 – MREs, MSEs relacionado para estimadores de $\lambda = 0,5$ e $\phi = 2$ para $N = 10000$ amostras simuladas, considerando diferentes valores de n .

para qualquer tamanho de amostra. Considerando os resultados obtidos nos estudos simulados, verificamos que o estimador de máxima verossimilhança combinado com o vício corrigido pela técnica discutida por [Cox e Snell \(1968\)](#) devem ser considerados para estimar os parâmetros da distribuição ILP.

4.5.2 Dados Censurados

Nesta seção, consideramos os EMVs na presença de dados censurados. Observações com censura do tipo aleatória ocorrem com muita frequência. Um algoritmo para gerar tal tipo de censura é descrito a seguir:

1. Defina o tamanho da amostra n .
2. Gere uma amostra da v.a. $X \sim IWL(\phi, \lambda)$ e $T \sim U(0, t_c)$ de tamanho n ;
3. Obtenha uma nova v.a. T em que

$$T_j = \min(X_j, W_j)$$

e $\delta_j = 0$ se $T_j = X_j$ ou $\delta_j = 1$ se $T_j = W_j$. Note que t_c tem que ser escolhido de forma que a proporção de censura seja a mesma de interesse durante o estudo de simulação.

Nesse caso, apresentamos dois cenários onde obtivemos, aproximadamente, as proporções de censura de 0,3 e 0,5, isto é, 30% e 50% de censura. O estudo de simulação é realizado considerando $\theta = (2, 4)$, $N = 10000$ e $n = (10, 15, \dots, 130)$.

Os estimadores de máxima verossimilhança foram calculados usando a função log-verossimilhança (4.10), com o pacote *maxLik*, disponível em R. A solução para o máximo foi única para diferentes valores iniciais considerados.

As Figuras 26 e 27 apresentam as MREs, os RMSEs e as probabilidades de cobertura dos intervalos de confiança com 95% de nível de confiança relacionados aos MLEs, CMLEs e os estimadores com vício corrigido pelo *bootstrap* para diferentes tamanhos de amostras.

Conforme apresentado nas Figuras 26 e 27, o estimador CMLE retornou melhores resultados se comparado com os outros métodos considerados nesse estudos. Sendo assim, tanto para amostras completas quanto para amostras censuradas, o vício corrigido por meio da abordagem de ([COX; SNELL, 1968](#)) retornou melhores estimativas e deve ser utilizado para estimar os parâmetros da distribuição ILP.

4.6 Análise de Sobrevivência das falhas do subsistema 1J

Nesta seção revistamos parte dos dados reais apresentados no capítulo 2. A análise da distribuição que melhor se ajusta aos dados inerentes ao problema é relevante para propor

Figura 26 – MREs, RMSEs relacionado para os estimadores de $\lambda = 0,5$ e $\phi = 2$ para $N = 10000$ amostras simuladas, considerando diferentes valores de n e 30% de censura.

melhores formas de manutenção e evitar custos desnecessários para a companhia. A Tabela 9 apresenta os dados relacionados a tempo de falha (em dias) de 194 ocorrências de falha do Subsistema 1J (+ indicações da presença de censura).

Tabela 9 – Dados relacionados a tempo de falha de 194 ocorrências de falha em uma aeronave.

43	29	37	88	5	14	9	43+	1	78	1	77	17	100
3	119+	22	3	8	80	1	19	157+	65	34	13	62+	2
1	1	2	3	6	1	2	5	7	6	1	1	4	1
1	1	2	7	2	1	1	2	1	1	7	1	1	4
1	4	2	4	5	5	4	3	2	2	2	3	3	9
1	6	9	2	5	7	4	2	1	2	2	3	11	8
3	1	2	2	2	2	2	1	3	20+	8	8	197	20
14	7	29	7	16	34	25	10	80	42	32	1	3	1
12	7	7	39+	60	53	32	9	8	1	1	27	2	4
8	13	7	7	1	19	7	12	19	5	18	1	4	18
20	9	14	13	70	18	3	7	20	3	11	10	3	38+
278	13	79	145+	19	2	18	2	65	14	31	10	19	5
9	45	13	5	1	1	31	35	34	4	3	5	12	140+
106	5	40	130+	21	19	7	10	91	193	64	85+		

O resultado obtido para a distribuição ILP foi comparado com os resultados de outras

Figura 27 – MREs, RMSEs relacionado para os estimadores de $\lambda = 0,5$ e $\phi = 2$ para $N = 10000$ amostras simuladas, considerando diferentes valores de n e 50% de censura.

distribuições, tais como: da distribuição Weibull, da Gamma e da Log-normal. Tais resultados foram sobrepostos sob a curva de sobrevivência não-paramétrica ajustada por meio do estimador de Kaplan Meier ((LAWLESS, 2003)). Inicialmente, para verificar o comportamento da função de risco, será considerado o TTT-plot (tempo total em teste) proposto por Barlow e Campo (1975).

Como temos a presença de censura nos dados, sendo esta censura do tipo aleatória, as equações (4.11) e (4.12) foram utilizadas para o calculo dos EMVs. A Tabela 10 apresenta os EMVs, erros-padrão e intervalos de confiança de 95% para ϕ e λ . A Tabela 11 apresenta o resultado dos critérios AIC, AICC, HQIC, CAIC, para diferentes funções densidades de probabilidades.

Tabela 10 – MLE, erros-padrão e intervalo de confiança de 95% para ϕ e λ .

θ	MLE	SE	$IC_{95\%}(\theta)$
ϕ	0,643	0,059	(0,527; 0,760)
λ	2,825	0,296	(2,245; 3,405)

A Figura 28 apresenta o TTT-plot, função de sobrevivência ajustada por diferentes distribuições e o estimador Kaplan-Meier e a função de risco ajustada pela distribuição de ILP.

Tabela 11 – Resultados dos critérios AIC, AICC, HQIC, CAIC para diferentes distribuições de probabilidade considerando dados relacionados para tempo de falha de 194 ocorrências de falha de uma aeronave.

Critério	ILP	Weibull	Gamma	Log-normal
AIC	1392,66	1452,37	1474,44	1408,44
AICC	1392,73	1452,43	1474,50	1408,51
HQIC	1395,31	1455,02	1477,08	1411,09
CAIC	1401,20	1460,91	1482,97	1416,98

Figura 28 – TTT-plot, função de sobrevivência ajustada por diferentes distribuições e o estimador Kaplan-Meier e a função de risco ajustada pela distribuição ILP considerando dados relacionados a tempo de falha de 194 ocorrências.

Comparando a função de sobrevivência empírica com o modelo ajustado, observamos um melhor ajuste para a distribuição Inversa Lindley Ponderada. Este resultado é também confirmado pelos diferentes critérios de discriminação. Para tais métodos, observamos que a distribuição ILP, que possui valor mínimo, possui o melhor ajuste. Baseado no TTT-plot a função de risco possui a forma de curva da banheira invertida.

Portanto, a partir da metodologia proposta, os dados relacionados a tempo de falha de 194 itens do subsistema 1J podem ser descritos pela distribuição Inversa Lindley Ponderada. O tempo mediano, assim como diferentes tempos de falha para os quantis, podem ser vistos na Tabela 12.

Tabela 12 – Tabela de frequência de ocorrência de falhas do subsistema 1J por quartil determinado considerando a distribuição Inversa Lindley Ponderada.

Quantil	10%	25%	50%	75%
Em dias	2	3	6	21

Por meio dos resultados obtidos na Tabela 12, podemos verificar que, se uma manutenção preventiva fosse feita assumindo uma quantidade de falhas menor ou igual a 25% dos itens, o tempo ideal para tal manutenção seria de 3 dias após a última manutenção. Diferentes valores

também podem ser obtidos assumindo diferentes níveis de falha. Tais resultados se mostram de grande importância para o planejamento da empresa.

CONCLUSÃO

Dada a dinâmica e a importância do setor de aviação, torna-se extremamente valiosa a garantia de fidelidade entre o cliente e a empresa, tornando o processo de satisfação do cliente uma poderosa ferramenta de gestão. Assim, um dos fatores que podem afetar a satisfação dos clientes durante um voo são as possíveis falhas, que não afetam a aeronavegabilidade, mas são perceptíveis e de grande valor para os passageiros como, por exemplo, defeito em aparelhos eletrônicos, defeito em suas poltronas, etc.

Portanto, este trabalho teve por objetivo a realização do estudo de potenciais falhas e a modelagem de seus tempos de falha. Para tanto, foi definida uma primeira etapa na qual foram classificadas as falhas que possuíam mais representatividade no sistema em geral, revelando que quatro falhas representam 84% das falhas totais. Após a análise, foram realizados ajustes por meio de diferentes distribuições de probabilidade que são comumente utilizadas para descrever dados de confiabilidade, sendo elas: Log-normal, Weibull e Gama. Através de vários critérios de seleção de modelos, verificamos que a distribuição Log-normal obteve melhores ajustes para três das quatro principais falhas e a distribuição Weibull se ajustou melhor em uma situação.

Por outro lado, para um dos tempos de falha, verificamos que uma nova distribuição retornou melhores resultados que os obtidos pelas distribuições mais comumente utilizadas em análise de sobrevivência. A distribuição proposta, distribuição ILP, e suas propriedades matemáticas foram estudadas, entre elas: momentos, média, variância, a função de sobrevivência e de risco, medida de entropia, os estimadores de máxima verossimilhança dos parâmetros e suas propriedades assintóticas. Mostramos que os estimadores de máxima verossimilhança possuem vício sistemático para pequenas amostras. Nesse sentido, uma correção de vício foi proposta para que os EMVs modificados possuam correção de vício de segunda ordem. Em seguida, foi realizado a aplicação desta distribuição considerando a principal falha encontrada na etapa 1, a falha do subsistema 1J. Tal distribuição retornou melhor ajuste do que as outras supracitadas anteriormente.

Há um grande número de extensões que pode ser considerado. Até o momento discutimos apenas a inferência dos parâmetros da distribuição ILP com base na inferência clássica. Tais métodos de inferência podem ser flexibilizados ao considerar a abordagem Bayesiana ([ALI, 2015](#); [RAMOS *et al.*, 2017](#)). A presença de covariáveis, assim como um componente de longa duração, ocorrem com grande frequência na prática. [Louzada e Ramos \(2017\)](#) apresentam um modelo de longa duração baseado na distribuição Lindley ponderada, tal abordagem pode ser investigada futuramente utilizando a distribuição ILP como distribuição de base.

REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. **IEEE transactions on automatic control**, Ieee, v. 19, n. 6, p. 716–723, 1974. Citado na página 39.

ALI, S. On the bayesian estimation of the weighted lindley distribution. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 85, n. 5, p. 855–880, 2015. Citado nas páginas 45 e 66.

BAKOUCH, H. S.; AL-ZAHRANI, B. M.; AL-SHOMRANI, A. A.; MARCHI, V. A.; LOUZADA, F. An extended lindley distribution. **Journal of the Korean Statistical Society**, Elsevier, v. 41, n. 1, p. 75–85, 2012. Citado na página 45.

BAKOUCH, H. S.; DEY, S.; RAMOS, P. L.; LOUZADA, F. Binomial-exponential 2 distribution: Different estimation methods with weather applications. **TEMA (São Carlos)**, SciELO Brasil, v. 18, n. 2, p. 233–251, 2017. Citado na página 57.

BARLOW, R. E.; CAMPO, R. A. **Total Time on Test Processes and Applications to Failure Data Analysis**. [S.l.], 1975. Citado nas páginas 32 e 61.

BRAILE, N.; ANDRADE, J. Estudo de falhas em equipamentos de costura industriais utilizando o fmea e a análise de confiabilidade. **XXXIII Encontro Nacional de Engenharia de Produção-ENEGEP, Salvador**, 2013. Citado na página 36.

CORDEIRO, G. M.; KLEIN, R. Bias correction in arma models. **Statistics & Probability Letters**, Elsevier, v. 19, n. 3, p. 169–176, 1994. Citado na página 54.

CORDEIRO, G. M.; ROCHA, E. C. D.; ROCHA, J. G. C. D.; CRIBARI-NETO, F. Bias-corrected maximum likelihood estimation for the beta distribution. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 58, n. 1, p. 21–35, 1997. Citado na página 22.

COX, D. R.; SNELL, E. J. A general definition of residuals. **Journal of the Royal Statistical Society. Series B (Methodological)**, JSTOR, p. 248–275, 1968. Citado nas páginas 53, 54 e 59.

EFRON, B. Bootstrap methods: another look at the jackknife. **The Annals of Statistics**, v. 7, n. 1, p. 1–26, 1979. Citado nas páginas 53 e 56.

GERTSBAKH, I. **Reliability theory: with applications to preventive maintenance**. [S.l.]: Springer, 2013. Citado na página 21.

GHITANY, M.; AL-MUTAIRI, D.; BALAKRISHNAN, N.; AL-ENEZI, L. Power lindley distribution and associated inference. **Computational Statistics & Data Analysis**, Elsevier, v. 64, p. 20–33, 2013. Citado na página 45.

GHITANY, M.; ALQALLAF, F.; AL-MUTAIRI, D.; HUSAIN, H. A two-parameter weighted lindley distribution and its applications to survival data. **Mathematics and Computers in Simulation**, Elsevier, v. 81, n. 6, p. 1190–1201, 2011. Citado nas páginas 45 e 46.

- GHITANY, M.; ATIEH, B.; NADARAJAH, S. Lindley distribution and its application. **Mathematics and computers in simulation**, Elsevier, v. 78, n. 4, p. 493–506, 2008. Citado nas páginas 45 e 52.
- GILES, D. E.; FENG, H.; GODWIN, R. T. On the bias of the maximum likelihood estimator for the two-parameter lomax distribution. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 42, n. 11, p. 1934–1950, 2013. Citado na página 22.
- GLASER, R. E. Bathtub and related failure rate characterizations. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 75, n. 371, p. 667–672, 1980. Citado na página 49.
- JURAN, J.; GRUNA, F. Quality control handbook. **New York: McGrawHill**, 1988. Citado na página 22.
- KOTLER, P. **Marketing para o século XXI: como criar, conquistar e dominar mercados**. [S.l.]: Futura, 2000. Citado na página 29.
- LAWLESS, J. Event history analysis and longitudinal surveys. **Analysis of Survey data**, Wiley Online Library, p. 221–243, 2003. Citado nas páginas 40 e 61.
- LAWLESS, J. F. **Statistical models and methods for lifetime data**. [S.l.]: John Wiley & Sons, 2011. v. 362. Citado nas páginas 21, 31 e 34.
- LEMONTE, A. J. Improved point estimation for the kumaraswamy distribution. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 81, n. 12, p. 1971–1982, 2011. Citado na página 22.
- LINDLEY, D. V. Fiducial distributions and bayes' theorem. **Journal of the Royal Statistical Society. Series B (Methodological)**, JSTOR, p. 102–107, 1958. Citado na página 45.
- LOUZADA, F.; RAMOS, P. L. A new long-term survival distribution. **Biostat Biometrics Open Acc J**, v. 1, n. 4, p. 1–6, 2017. Citado na página 66.
- _____. Efficient closed-form maximum a posteriori estimators for the gamma distribution. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 88, n. 6, p. 1134–1146, 2018. Citado na página 57.
- LOUZADA-NETO, F.; MAZUCHELI, J.; ACHCAR, J. A. **Uma introdução à análise de sobrevivência e confiabilidade**. [S.l.]: Sociedad Chilena de Estadística, 2001. Citado nas páginas 21, 31 e 34.
- MAZUCHELI, J.; DEY, S. Bias-corrected maximum likelihood estimation of the parameters of the generalized half-normal distribution. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 88, n. 6, p. 1027–1038, 2018. Disponível em: <<https://doi.org/10.1080/00949655.2017.1413649>>. Citado na página 22.
- MAZUCHELI, J.; LOUZADA, F.; GHITANY, M. Comparison of estimation methods for the parameters of the weighted lindley distribution. **Applied Mathematics and Computation**, Elsevier, v. 220, p. 463–471, 2013. Citado na página 45.
- MAZUCHELI, J.; MENEZES, A. F. B.; DEY, S. Improved maximum-likelihood estimators for the parameters of the unit-gamma distribution. **Communications in Statistics - Theory and Methods**, Taylor & Francis, v. 0, n. 0, p. 1–12, 2017. Disponível em: <<https://doi.org/10.1080/03610926.2017.1361993>>. Citado na página 22.

NAIR, R. R.; BLAKE, P.; GRIGORENKO, A. N.; NOVOSELOV, K. S.; BOOTH, T. J.; STAUBER, T.; PERES, N. M.; GEIM, A. K. Fine structure constant defines visual transparency of graphene. **Science**, American Association for the Advancement of Science, v. 320, n. 5881, p. 1308–1308, 2008. Citado na página 33.

NASCIMENTO, R. d. **Estudo do comportamento em fadiga de alto ciclo da liga Ti-13V-11Cr-3Al**. Tese (Doutorado) — Universidade de São Paulo, 2016. Citado na página 37.

PATIL, G. P.; RAO, C. R. Weighted distributions and size-biased sampling with applications to wildlife populations and human families. **Biometrics**, JSTOR, p. 179–189, 1978. Citado na página 45.

RAMOS, P.; LOUZADA, F. The generalized weighted lindley distribution: Properties, estimation, and applications. **Cogent Mathematics**, Cogent OA, v. 3, n. 1, p. 1256022, 2016. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/23311835.2016.1256022>>. Citado na página 45.

RAMOS, P. L. Aspectos computacionais para inferência na distribuição gama generalizada. Universidade Estadual Paulista (UNESP), 2014. Citado nas páginas 15, 16, 32, 33, 35, 36, 37 e 38.

RAMOS, P. L.; ACHCAR, J. A.; MOALA, F. A.; RAMOS, E.; LOUZADA, F. Bayesian analysis of the generalized gamma distribution using non-informative priors. **Statistics**, Taylor & Francis, v. 51, n. 4, p. 824–843, 2017. Citado na página 66.

RAMOS, P. L.; LOUZADA, F.; RAMOS, E. An efficient, closed-form map estimator for nakagami-m fading parameter. **IEEE Communications Letters**, IEEE, v. 20, n. 11, p. 2328–2331, 2016. Citado na página 22.

REATH, J.; DONG, J.; WANG, M. Improved parameter estimation of the log-logistic distribution with applications. **Computational Statistics**, Springer, p. 1–18, 2016. Citado na página 57.

RODRIGUES, G. C.; LOUZADA, F.; RAMOS, P. L. Poisson–exponential distribution: different methods of estimation. **Journal of Applied Statistics**, Taylor & Francis, v. 45, n. 1, p. 128–144, 2018. Citado na página 57.

SHARMA, V. K.; SINGH, S. K.; SINGH, U.; AGIWAL, V. The inverse lindley distribution: a stress-strength reliability model with application to head and neck cancer data. **Journal of Industrial and Production Engineering**, Taylor & Francis, v. 32, n. 3, p. 162–173, 2015. Citado nas páginas 22, 46 e 47.

SIM, S.; ENDRENYI, J. Optimal preventive maintenance with repair. **IEEE Transactions on Reliability**, IEEE, v. 37, n. 1, p. 92–96, 1988. Citado na página 21.

SOBRAL, T. E.; BARRETO, G. Análise dos critérios de informação para a seleção de ordem em modelos auto regressivos. In: **X Conferência Brasileira de Dinâmica, Controle e Aplicações**. [S.l.: s.n.], 2011. Citado na página 39.

SOBRINHO, S. d. P.; MATOS, V. A. T. d.; PEREIRA, A. P. M. S.; PIVETTA, F.; SEIXAS, G. B.; JUNIOR, J. H. C. Determination of parameters of gamma distribution and average rainfall 10 days stations for the mato grosso of state. **Revista Brasileira de Meteorologia**, SciELO Brasil, v. 29, n. 2, p. 183–196, 2014. Citado na página 35.

TAHIR, M. H.; CORDEIRO, G. M. Compounding of distributions: a survey and new generalized classes. **Journal of Statistical Distributions and Applications**, Springer, v. 3, n. 1, p. 1–35, 2016. Citado na página [22](#).

TEIMOURI, M.; NADARAJAH, S. Bias corrected mles for the weibull distribution based on records. **Statistical Methodology**, Elsevier, v. 13, p. 12–24, 2013. Citado na página [22](#).

WANG, M.; WANG, W. Bias-corrected maximum likelihood estimation of the parameters of the weighted lindley distribution. **Communications in Statistics-Simulation and Computation**, Taylor & Francis, v. 46, n. 1, p. 530–545, 2017. Citado nas páginas [45](#) e [55](#).