



JOSÉ EDUARDO COLLE

**MODELAGEM E AJUSTE DE FUNÇÕES DENSIDADE DE  
PROBABILIDADE PARA TAMANHOS DE PARTÍCULAS ENCONTRADOS  
EM SÍTIOS ARQUEOLÓGICOS NO VALE DO RIO PARANAPANEMA**

Santo André, 2019





**UNIVERSIDADE FEDERAL DO ABC**

**CENTRO DE MATEMÁTICA, COMPUTAÇÃO E COGNIÇÃO**

**JOSÉ EDUARDO COLLE**

**MODELAGEM E AJUSTE DE FUNÇÕES DENSIDADE DE  
PROBABILIDADE PARA TAMANHOS DE PARTÍCULAS ENCONTRADOS  
EM SÍTIOS ARQUEOLÓGICOS NO VALE DO RIO PARANAPANEMA**

**Orientador: Prof. Dr. Valdecir Marvulle**

Dissertação de mestrado apresentada ao Centro de  
Matemática, Computação e Cognição para  
obtenção do título de Mestre

ESTE EXEMPLAR CORRESPONDE A VERSÃO FINAL DA DISSERTAÇÃO  
DEFENDIDA PELO ALUNO JOSÉ EDUARDO COLLE,  
E ORIENTADA PELO PROF. DR. VALDECIR MARVULLE.

**SANTO ANDRÉ, 2019**

Sistema de Bibliotecas da Universidade Federal do ABC  
Elaborada pelo Sistema de Geração de Ficha Catalográfica da UFABC  
com os dados fornecidos pelo(a) autor(a).

Colle, José Eduardo  
Modelagem e Ajuste de Funções Densidade de  
Probabilidade para Tamanhos de Partículas Encontradas em  
Sítios Arqueológicos no Vale do Rio Paranapanema / José  
Eduardo Colle. — 2019.

102 fls. : il.

Orientador: Valdecir Marvulle

Dissertação (Mestrado) — Universidade Federal do ABC,  
Mestrado Profissional em Matemática em Rede Nacional -  
PROFMAT, Santo André, 2019.

1. Comparação e Ajuste de Modelos. 2. Fdp Gama. 3. Fdp  
Log Normal. 4. Fdp Weibull. 5. Critérios AIC e BIC. I.  
Marvulle, Valdecir. II. Mestrado Profissional em Matemática  
em Rede Nacional - PROFMAT, 2019. III. Título.

**Este exemplar foi revisado e alterado em relação à versão original, de acordo com as observações levantadas pela banca no dia da defesa, sob responsabilidade única do autor e com a anuência de seu orientador.**

Santo André, 10 de MAIO de 2019.

Assinatura do autor: Jose Eduardo Blee...

Assinatura do orientador: [Assinatura]



**MINISTÉRIO DA EDUCAÇÃO**  
**Fundação Universidade Federal do ABC**  
**Programa de Pós-Graduação em Mestrado Profissional em Matemática**  
**em Rede Nacional**

Avenida dos Estados, 5001 – Bairro Santa Terezinha – Santo André – SP  
CEP 09210-580 · Fone: (11) 4996-0017  
profmat@ufabc.edu.br

**FOLHA DE ASSINATURAS**

Assinaturas dos membros da Banca Examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato José Eduardo Colle, realizada em 22 de fevereiro de 2019:

---

Prof.(a) Dr.(a) **Valdecir Marvulle** (Universidade Federal do ABC) – Presidente

---

Prof.(a) Dr.(a) **Rafael de Mattos Grisi** (Universidade Federal do ABC) – Membro Titular

---

Prof.(a) Dr.(a) **Paulo Teotonio Sobrinho** (Universidade de São Paulo) – Membro Titular

---

Prof.(a) Dr.(a) **Daniel Miranda Machado** (Universidade Federal do ABC) – Membro Suplente

---

Prof.(a) Dr.(a) **Itana Stiubiener** (Universidade Federal do ABC) – Membro Suplente

---

Dedico este trabalho

Aos meus queridos pais, Vicente (in memorian) e Cristina, pelo amor, carinho e por sempre acreditarem, investirem e apoiarem nos meus estudos e dos meus irmãos.

À minha querida esposa Lilian e à minha filha, princesa Manuela, pela doçura e amor que recebo todos os dias, apoio e compreensão. Nada disso teria sentido se vocês não existissem na minha vida.

---

## AGRADECIMENTOS

---

Agradeço, em primeiro lugar, a Deus por ter me dado a vida, minha família, minha esposa e à Manuela. Também, por tudo o que faz em minha vida: proteção, direção nos meus caminhos, inspiração e força para persistir até aqui.

Ao meu pai Vicente de Colle (in Memoriam) que sempre acreditou e investiu nos meus estudos desde a infância. Neste período de trabalho, sempre estive em pensamentos.

A minha mãe Cristina Paredes de Colle que, igualmente ao meu pai, sempre me incentivou e apoiou em meus estudos. Agradeço pela compreensão, por entender as ausências que muitas vezes foram necessárias.

A minha esposa, Lilian Kelly Alves de Matos Colle, pelo apoio, paciência, compreensão, parceria e por compartilhar todo o percurso deste trabalho. Por compreender que nem sempre podia estar presente em certos momentos.

Aos meus irmãos Rejane Aparecida Colle e Rafael de Colle pela compreensão, apoio e pelas minhas ausências.

A minha sobrinha Nathali Colle pelas dicas de Inglês.

A minha tia Rosely de Colle Abate pelo incentivo e pelo carinho que tem comigo desde a minha infância.

Ao meu orientador professor Doutor Valdecir Marvulle, pela paciência, tranquilidade com que sempre me orientou. Também por acreditar em mim de forma muito positiva, com disposição e presteza nos retornos do trabalho.

À doutoranda Tatiane de Souza pelo fornecimento dos dados para realização deste trabalho.

Aos professores doutores da UFABC que contribuíram com excelentes aulas para o meu crescimento profissional e realização deste trabalho: Ana Carolina Boero, Rafael de Matos Grisi, Daniel Miranda Machado, Marcus Antônio Mendonça Marrocos, Maurício Firmino Silva Lima, Jair Donadelli Júnior e Sinuê Dayan Barbero Lodovici.



Aos membros da banca, professor Dr. Rafael de Matos Grisi e professor Dr. Paulo Teotônio Sobrinho que contribuíram muito com suas críticas e sugestões no dia da defesa deste trabalho.

Ao grande amigo de turma de graduação professor Doutor Walter Martins Rodrigues, por sempre incentivar e apoiar para a continuidade dos meus estudos em Matemática no nível de Pós-Graduação.

Aos colegas de turma: Marilda, Simone, Juliana, Raphael, Alisson, Bruno, Marcelo, Paulo e Douglas. Pelo companheirismo, ajuda mútua e os momentos de estudo.

Ao mestre Thiago Lima, pela competência e presteza com que sempre me atendeu na secretaria de Pós-Graduação da UFABC.

À Sociedade Brasileira de Matemática (SBM) e à Universidade Federal do ABC (UFABC) pela realização do programa.

Por fim, a todos aqueles que contribuíram, direta ou indiretamente, para a realização deste trabalho.

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

---

*“A razão cardeal de toda superioridade humana é sem dúvida a vontade. O poder nasce do querer. Sempre que o homem aplique a veemência e perseverante energia de sua alma a um fim, ele vencerá os obstáculos, e se não atingir o alvo, fará pelo menos coisas admiráveis”*

(José de Alencar)

*“Mestre é aquele que às vezes para, para aprender”*

(Guimarães Rosa)

---

## RESUMO

---

O objetivo deste trabalho foi comparar a aderência das distribuições gama, log-normal e Weibull na modelagem e ajuste de curvas de densidade de probabilidade aos dados experimentais do tamanho de partículas de sítios arqueológicos na bacia do rio Paranapanema em Itapeva, São Paulo. O ajuste e aderência das distribuições dos tamanhos de partículas das escavações foram verificadas via Método de Informação de dois critérios: Akaike (AIC) e no Bayesiano (BIC). Os resultados obtidos indicaram que o modelo Log Normal ajustou-se mais adequadamente à distribuição dos tamanhos das partículas.

**Palavras-chave:** Comparação de Modelos. Distribuição Gama. Distribuição Log Normal. Distribuição Weibull. Critério AIC. Critério BIC.

---

## ABSTRACT

---

The objective of this science assignment was to compare the goodness of fit of distributions Gamma, Log-normal and Weibull in modeling and fitting of density and probability curves to experimental data of size particles of archaeological sites in Parapanema based in Itapeva, São Paulo. The adjustments and adherence of the size distributions of particles excavations was verified by information method of two criteria: Akaike (AIC) and Bayesian (BIC). The obtained results indicated that the model Log-normal adjusted more properly to the particle size distribution.

**Keywords:** Comparison of Models. Distribution Range. Normal Log Distribution. Weibull distribution. AIC criterion. BIC criterion.

---

## LISTA DE FIGURAS

---

Figura 1.1	Gráfico: Função de Densidade de Probabilidade Gama . . . . .	7
Figura 1.2	Gráfico: Função de Densidade de Probabilidade Log Normal . . .	10
Figura 1.3	Gráfico: Função de Densidade de Probabilidade Weibull . . . . .	12
Figura 1.4	Gráfico: Densidade Teórica e Empírica . . . . .	25
Figura 1.5	Gráfico Q-Q Plot . . . . .	25
Figura 1.6	Gráfico: Distribuição Acumulada Empírica (CDF) . . . . .	26
Figura 1.7	Gráfico P-P Plot . . . . .	27
Figura 2.1	Identificação das áreas D18 e G6 - Croqui: Astolfo Araújo, 2017.	29
Figura 2.2	Técnica: Passo 1 . . . . .	30
Figura 2.3	Técnica: Passo 2 . . . . .	31
Figura 2.4	Técnica: Passo 3 . . . . .	31
Figura 2.5	Resultados da Escavação G6 . . . . .	32
Figura 3.1	Tabela: amostra 1 . . . . .	39
Figura 3.2	Construção da tabela de probabilidade em massa: amostra 1 . . .	40
Figura 3.3	Construção da tabela . . . . .	40
Figura 3.4	Parâmetros da Gama, AIC e BIC para a amostra 1 . . . . .	41
Figura 3.5	Gráficos da Gama para a amostra 1 . . . . .	42
Figura 3.6	Exemplo gráfico com as 3 (fdp)s . . . . .	43
Figura 4.1	AMOSTRA 1: Curva de Distribuição GAMA . . . . .	45
Figura 4.2	AMOSTRA 1: Curva de Distribuição LOG NORMAL . . . . .	45
Figura 4.3	AMOSTRA 1: Curva de Distribuição WEIBULL . . . . .	46
Figura 4.4	AMOSTRA 1: Histograma de frequências dos dados observados e das 3 fdps ajustadas . . . . .	46

Figura 4.5	AMOSTRA 2: Curva de Distribuição GAMA . . . . .	48
Figura 4.6	AMOSTRA 2: Curva de Distribuição LOG NORMAL . . . . .	48
Figura 4.7	AMOSTRA 2: Curva de Distribuição WEIBULL . . . . .	49
Figura 4.8	AMOSTRA 2: Histograma de frequências dos dados observados e das 3 fdps ajustadas . . . . .	49
Figura 4.9	AMOSTRA 3: Curva de Distribuição GAMA . . . . .	51
Figura 4.10	AMOSTRA 3: Curva de Distribuição LOG NORMAL . . . . .	51
Figura 4.11	AMOSTRA3: Curva de Distribuição WEIBULL . . . . .	52
Figura 4.12	AMOSTRA 3: Histograma de frequências dos dados observados e das 3 fdps ajustadas . . . . .	52
Figura 4.13	AMOSTRA 4: Curva de Distribuição GAMA . . . . .	54
Figura 4.14	AMOSTRA 4: Curva de Distribuição LOG NORMAL . . . . .	54
Figura 4.15	AMOSTRA 4: Curva de Distribuição WEIBULL . . . . .	55
Figura 4.16	AMOSTRA 4: Histograma de frequências dos dados observados e das 3 fdps ajustadas . . . . .	55
Figura 4.17	AMOSTRA 5: Curva de Distribuição GAMA . . . . .	57
Figura 4.18	AMOSTRA 5: Curva de Distribuição LOG NORMAL . . . . .	57
Figura 4.19	AMOSTRA 5: Curva de Distribuição WEIBULL . . . . .	58
Figura 4.20	AMOSTRA 5: Histograma de frequências dos dados observados e das 3 fdps ajustadas . . . . .	58
Figura 4.21	AMOSTRA 6: Curva de Distribuição GAMA . . . . .	60
Figura 4.22	AMOSTRA 6: Curva de Distribuição LOG NORMAL . . . . .	60
Figura 4.23	AMOSTRA 6: Curva de Distribuição WEIBULL . . . . .	61
Figura 4.24	AMOSTRA 6: Histograma de frequências dos dados observados e das 3 fdps ajustadas . . . . .	61
Figura 4.25	AMOSTRA 7: Curva de Distribuição GAMA . . . . .	63
Figura 4.26	AMOSTRA 7: Curva de Distribuição LOG NORMAL . . . . .	63
Figura 4.27	AMOSTRA 7: Curva de Distribuição WEIBULL . . . . .	64

Figura 4.28	AMOSTRA 7: Histograma de frequências dos dados observados e das 3 fdps ajustadas . . . . .	64
Figura 4.29	AMOSTRA 8: Curva de Distribuição GAMA . . . . .	66
Figura 4.30	AMOSTRA 8: Curva de Distribuição LOG NORMAL . . . . .	66
Figura 4.31	AMOSTRA 8: Curva de Distribuição WEIBULL . . . . .	67
Figura 4.32	AMOSTRA 8: Histograma de frequências dos dados observados e das 3 fdps ajustadas . . . . .	67
Figura 4.33	AMOSTRA 9: Curva de Distribuição GAMA . . . . .	69
Figura 4.34	AMOSTRA 9: Curva de Distribuição LOG NORMAL . . . . .	69
Figura 4.35	AMOSTRA 9: Curva de Distribuição WEIBULL . . . . .	70
Figura 4.36	AMOSTRA 9: Histograma de frequências dos dados observados e das 3 fdps ajustadas . . . . .	70
Figura 4.37	AMOSTRA 10: Curva de Distribuição GAMA . . . . .	72
Figura 4.38	AMOSTRA 10: Curva de Distribuição LOG NORMAL . . . . .	72
Figura 4.39	AMOSTRA 10: Curva de Distribuição WEIBULL . . . . .	73
Figura 4.40	AMOSTRA 10: Histograma de frequências dos dados observados e das 3 fdps ajustadas . . . . .	73

---

## LISTA DE TABELAS

---

Tabela 2.1	Amostra 1: G6N4F3 . . . . .	34
Tabela 2.2	Amostra 2: G6N4F5 . . . . .	34
Tabela 2.3	Amostra 3: G6N4F4 . . . . .	35
Tabela 2.4	Amostra 4: G6N4F2 . . . . .	35
Tabela 2.5	Amostra 5: G6N3F2 . . . . .	36
Tabela 2.6	Amostra 6: G6N3F3 . . . . .	36
Tabela 2.7	Amostra 7: G6N1S/F . . . . .	37
Tabela 2.8	Amostra 8: G6N2S/F . . . . .	37
Tabela 2.9	Amostra 9: G6N4F6 . . . . .	38
Tabela 2.10	Amostra 10: G6N5F2 . . . . .	38
Tabela 4.1	AMOSTRA 1: ESTIMATIVAS DOS PARÂMETROS das 3 distri- buições . . . . .	46
Tabela 4.2	AMOSTRA 2: ESTIMATIVAS DOS PARÂMETROS das 3 distri- buições . . . . .	49
Tabela 4.3	AMOSTRA 3: ESTIMATIVAS DOS PARÂMETROS das 3 distri- buições . . . . .	52
Tabela 4.4	AMOSTRA 4: ESTIMATIVAS DOS PARÂMETROS das 3 distri- buições . . . . .	55
Tabela 4.5	AMOSTRA 5: ESTIMATIVAS DOS PARÂMETROS das 3 distri- buições . . . . .	58
Tabela 4.6	AMOSTRA 6: ESTIMATIVAS DOS PARÂMETROS das 3 distri- buições . . . . .	61
Tabela 4.7	AMOSTRA 7: ESTIMATIVAS DOS PARÂMETROS das 3 distri- buições . . . . .	64



Tabela 4.8	AMOSTRA 8: ESTIMATIVAS DOS PARÂMETROS das 3 distribuições . . . . .	67
Tabela 4.9	AMOSTRA 9: ESTIMATIVAS DOS PARÂMETROS das 3 distribuições . . . . .	70
Tabela 4.10	AMOSTRA 10: ESTIMATIVAS DOS PARÂMETROS das 3 distribuições . . . . .	73
Tabela A.1	Amostra 1: G6N4F3 . . . . .	79
Tabela A.2	Amostra 2:G6N4F5 . . . . .	79
Tabela A.3	Amostra 3:G6N4F4 . . . . .	80
Tabela A.4	Amostra 4: G6N4F2 . . . . .	80
Tabela A.5	Amostra 5: G6N3F2 . . . . .	81
Tabela A.6	Amostra 6: G6N3F3 . . . . .	81
Tabela A.7	Amostra 7: G6N1S/F . . . . .	82
Tabela A.8	Amostra 8: G6N2S/F . . . . .	82
Tabela A.9	Amostra 9: G6N4F6 . . . . .	83
Tabela A.10	Amostra 10: G6N5F2 . . . . .	83

---

# CONTEÚDO

---

Lista de Figuras	viii
Lista de Tabelas	xi
INTRODUÇÃO E OBJETIVOS	1
1 REVISÃO TEÓRICA	4
1.1 Função Densidade de Probabilidade	4
1.1.1 Modelo Gama	5
1.1.2 Modelo Log-Normal	8
1.1.3 Modelo Weibull	10
1.2 Estimação de Máxima Verossimilhança	12
1.2.1 Obtenção dos parâmetros de Verossimilhança da (fdp) Gama	13
1.2.2 Obtenção dos parâmetros de Verossimilhança da (fdp) Log Normal	15
1.2.3 Obtenção dos parâmetros de Verossimilhança da (fdp) Weibull	18
1.3 Testes de Aderência e Critérios de Melhor Ajuste	21
1.3.1 Testes de Aderência	21
1.3.2 Critérios de Melhor Ajuste	21
1.4 Software R	23
1.5 Gráficos de Análise e Qualidade dos Ajustes de Curvas das (fdp)s	24
1.5.1 Gráfico da Densidade Empírica e Teórica	24
1.5.2 Gráfico Q-Q plot	25
1.5.3 Gráfico Função Distribuição Acumulada Empírica (CDF)s	26
1.5.4 Gráfico P-P plot	26
2 DADOS COLETADOS	28
3 MÉTODOS NUMÉRICOS PARA ESTIMATIVA DOS PARÂMETROS	39
4 RESULTADOS	44
4.1 Resultados obtidos para a amostra 1	45
4.2 Resultados obtidos para a amostra 2	48

4.3 Resultados obtidos para a amostra 3 . . . . .	51
4.4 Resultados obtidos para a amostra 4 . . . . .	54
4.5 Resultados obtidos para a amostra 5 . . . . .	57
4.6 Resultados obtidos para a amostra 6 . . . . .	60
4.7 Resultados obtidos para a amostra 7 . . . . .	63
4.8 Resultados obtidos para a amostra 8 . . . . .	66
4.9 Resultados obtidos para a amostra 9 . . . . .	69
4.10 Resultados obtidos para a amostra 10 . . . . .	72
5 CONSIDERAÇÕES FINAIS	76
A APÊNDICE A	78
A.1 Tabelas de dados . . . . .	78
Bibliografia	84
Índice	86

---

## INTRODUÇÃO E OBJETIVOS

---

Nos dias de hoje, o estudo da Probabilidade e Estatística está cada vez mais presente na vida dos profissionais das mais diversas áreas do conhecimento humano, pesquisadores e, é claro, dos estudantes, desde o Ensino Fundamental até o Nível Superior. É um ramo muito importante da matemática, pois se tornou essencial para resolver problemas em que a Análise de Dados dá ótimas respostas às indagações científicas.

"Estamos rodeados de dados coletados a partir de levantamentos e experimentos, as duas formas de pesquisa empírica. A pesquisa empírica é qualquer investigação em que os dados são recolhidos a partir do mundo externo e, em seguida, conclusões são tiradas a partir deles"(ver em [11]).

E sobre pesquisas estatísticas,

"... os dados qualitativos significa 'trabalhar' todo o material obtido durante a pesquisa, ou seja, os relatos de observação, as transcrições de entrevista, as análises de documentos e as demais informações disponíveis. A tarefa de análise implica, num primeiro momento, a organização de todo material, dividindo-o em partes, relacionando essas partes e procurando identificar nele tendências e padrões relevantes. Num segundo momento essas tendências e padrões são reavaliados, buscando-se relações e inferências num nível de abstração mais elevado."( ver em [12])

Sobre a definição de estatística,

"A definição de estatística mais utilizada atualmente descreve que essa é uma área do conhecimento que utiliza teorias probabilísticas para explicação de eventos, estudos e experimentos, com o objetivo de obter, organizar e analisar dados, determinar e interpretar correlações para descrição, explicação e previsão."(ver em [18])

Neste trabalho, usaremos as Distribuições de Probabilidade, em particular as Distribuições de Probabilidades Contínuas, as chamadas funções de densidade de probabilidades (fdp): Gama, Log Normal e Weibull e verificaremos qual melhor se ajusta aos dados dos tamanhos das partículas coletadas em sítios arqueológicos.

Faremos as estimativas dos parâmetros das Distribuições pelo Método da Máxima Verossimilhança numericamente.

Na área acadêmica, em geral, é dada muita ênfase na distribuição Normal de Probabilidades. Sempre há tentativas iniciais de ajuste com a distribuição normal e por se aplicar em muitos problemas do cotidiano e também nas aulas os professores procuram dar maior importância.

Notamos pouco uso e apresentação das funções de distribuição de probabilidades Gama, Log Normal e Weibull em situações reais de aplicabilidade. Aqui desejamos mostrar uma aplicação real e com isso explorar a Inferência Estatística, de modo que possamos fazer previsões por meio das funções de densidade de probabilidade (fdp). Também pretendemos utilizar o trabalho como exemplo didático para o ensino em nível de graduação e pós-graduação.

Para Ben-Zvi e Biehler, em [9]: "Os educadores acreditam que o uso de dados reais em tópicos de interesse dos alunos, o que não acontece apenas em Estatística, contribui para a sua motivação para aprenderem Estatística"

Em [4] afirma-se que "a utilização de funções de distribuições de probabilidade requer o uso de testes para provar a adaptação de dados ou da série de dados às funções. Estes são os testes de aderência."

Para o ajuste dos dados à função de densidade de probabilidade usaremos os métodos de Informação de Akaike (AIC) e Bayesiano (BIC). Também nos auxiliará na análise dos dados os gráficos: Densidade Empírica e Teórica, Q-Q plot, Função Distribuição Acumulada Empírica e Teórica e P-P plot.

As estimativas dos parâmetros e os gráficos serão feitos pelo software R Studio com a versão 3.3.2 (2016-10-31) – "Sincere Pumpkin Patch" Copyright (C) 2016 – The R Foundation for Statistical Computing – Plataforma: x86\_64-w64-mingw32/x64 (64-bit)

A dissertação está organizada em cinco capítulos:

1. Revisão Teórica
2. Dados Coletados
3. Métodos Numéricos para Estimativa dos Parâmetros
4. Resultados
5. Considerações Finais

No primeiro capítulo apresentamos uma revisão teórica das funções de densidade de probabilidade usadas na dissertação, o Método da Máxima Verossimilhança, Testes de Aderência, os Critérios de Informação de Akaike (AIC) e o Bayesiano (BIC) e Tópicos do Software R utilizados na pesquisa.

No segundo capítulo, descreveremos de onde foram retirados os dados da pesquisa e a construção inicial da tabela de tamanhos das partículas em massa para cada uma das 10 amostras.

No terceiro, descreveremos as etapas para a formação da tabela geral de dados dos tamanhos das partículas em massa: a probabilidade (em termos de massa), o diâmetro médio e o número de partículas em massa. Também como obter pelo software R Studio as estimativas dos parâmetros, gráficos e os valores dos Critérios de Informação: AIC e BIC.

No quarto capítulo, apresentaremos os resultados para futura análise dos dados: as estimativas dos parâmetros, valores de AIC e BIC calculados de cada distribuição e os gráficos Densidade Empírica e Teórica, Q-Q plot, Função Distribuição Acumulada Empírica e Teórica (CDFs) e P-P plot.

No último, serão feitas as considerações finais e discutidas as contribuições deste estudo.

R

---

## REVISÃO TEÓRICA

---

Neste capítulo, iremos apresentar conceitos, definições e propriedades relativas aos modelos de funções densidade de probabilidade, as chamadas (fdp). No trabalho usaremos a sigla (fdp) para o termo função densidade de probabilidade.

Trabalharemos com o conceito de Função de Máxima Verossimilhança para a estimação dos parâmetros dos modelos de (fdp)s.

Também estudaremos os métodos de aderência às curvas de distribuições e métodos de escolha do melhor modelo.

Estes conceitos serão posteriormente importantes para aplicação, ajustes de funções e seleção do melhor modelo para os dados.

Também será apresentado o software R, o qual será utilizado no cálculo das estimativas numéricas dos parâmetros e a confecção dos gráficos.

### 1.1 FUNÇÃO DENSIDADE DE PROBABILIDADE

No estudo das variáveis aleatórias contínuas, cujo valor da variável aleatória assume valores em um intervalo de números reais, também podemos fazer o cálculo de probabilidades para as mais diversas situações que podem assumir uma infinidade de valores. Para isso, precisamos da sua função de distribuição de probabilidade.

Dizemos que  $f(x)$  é uma função contínua de probabilidade ou função densidade de probabilidade (fdp), para uma variável aleatória contínua  $X$ , se são satisfeitas duas condições:

- i)  $f(x) \geq 0$ , para todo  $x \in (-\infty, \infty)$

ii) A área definida pela  $\int_{-\infty}^{\infty} f(x)dx$  é igual a 1.

Neste estudo, utilizaremos três funções densidade de probabilidade : Gama, Log Normal e Weibull. Elas foram escolhidas devido à grande possibilidade de se adequarem a dados experimentais já que, dependendo do valor dos parâmetros podem adquirir infinitas formas.

### 1.1.1 Modelo Gama

Uma variável aleatória  $X$  possui distribuição Gama de parâmetros  $\alpha > 0$  e  $\beta > 0$  quando sua função densidade de probabilidade é:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \quad (1.1)$$

com  $x > 0$  e  $\Gamma(\alpha)$  a função Gama definida pela integral

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt \quad (1.2)$$

A função  $\Gamma(\alpha)$  definida em 1.2 possui a seguinte propriedade recursiva:

$$\Gamma(\alpha) = (\alpha - 1) \cdot \Gamma(\alpha - 1) \quad (1.3)$$

*Demonstração.* Para demonstrar que  $\Gamma(\alpha) = (\alpha - 1) \cdot \Gamma(\alpha - 1)$ , podemos escrever a propriedade de uma forma equivalente, isto é,  $\Gamma(\alpha + 1) = \alpha \cdot \Gamma(\alpha)$  e assim:

$$\Gamma(\alpha + 1) = \int_0^{\infty} t^\alpha e^{-t} dt.$$

Utilizando o método da integração por partes em  $\int_0^{\infty} t^\alpha e^{-t} dt$ , e fazendo:

$$u = t^\alpha \Rightarrow du = \alpha \cdot t^{\alpha-1} \text{ e } dv = e^{-t} \Rightarrow v = -e^{-t}$$

obtemos:

$$\begin{aligned} \Gamma(\alpha + 1) &= -t\alpha \cdot e^{-t} \Big|_0^{\infty} - \int_0^{\infty} -e^{-t} \alpha t^{\alpha-1} dt \\ &= 0 + \alpha \cdot \int_0^{\infty} e^{-t} t^{\alpha-1} dt \\ &= \alpha \cdot \Gamma(\alpha) \end{aligned} \quad (1.4)$$

assim,  $\Gamma(\alpha + 1) = \alpha \cdot \Gamma(\alpha)$



Logo, fica demonstrado 1.3, isto é,  $\Gamma(\alpha) = (\alpha - 1) \cdot \Gamma(\alpha - 1)$  □

Para  $\alpha$  inteiro:  $\alpha = n$ , temos:  $\Gamma(n) = (n - 1)!$

Conforme definição de função densidade de probabilidade (fdp) apresentada em 1.1, isto é,

$$\int_0^{\infty} f(x)dx = 1$$

com  $f(x) > 0$ .

Faremos a seguir a demonstração de que a Distribuição Gama é uma (fdp),

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

*Demonstração.* Inicialmente demonstraremos um resultado importante e usaremos para demonstrar que a função de distribuição Gama é uma (fdp). O resultado é chamado de Função Gama Modificada, cuja expressão é dada por

$$\int_0^{\infty} x^{\alpha-1} e^{-\beta x} dx = \frac{\Gamma(\alpha)}{\beta^\alpha} \quad (1.5)$$

de 1.5 precisamos mostrar que  $\Gamma(\alpha) = \beta^\alpha \cdot \int_0^{\infty} x^{\alpha-1} e^{-\beta x} dx$  e também lembrando que a função Gama é definida por:  $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$

segue a demonstração da Função Gama Modificada,

fazendo a substituição na integral da Função Gama,  $y = \frac{x}{\beta} \Rightarrow x = \beta y$ . Neste caso,  $dx = \beta dy$ , portanto

$$\begin{aligned} \Gamma(\alpha) &= \int_0^{\infty} x^{\alpha-1} e^{-x} dx \\ &= \int_0^{\infty} (\beta y)^{\alpha-1} e^{-\beta y} \beta dy \\ &= \beta^\alpha \int_0^{\infty} y^{\alpha-1} e^{-\beta y} dy \end{aligned}$$

O resultado é uma função Gama Modificada em  $y$  e voltando para a variável  $x$ , temos:  $\Gamma(\alpha) = \beta^\alpha \cdot \int_0^{\infty} x^{\alpha-1} e^{-\beta x} dx$

Segue que  $\int_0^{\infty} x^{\alpha-1} e^{-\beta x} dx = \frac{\Gamma(\alpha)}{\beta^\alpha}$

Mas queremos provar que  $\int_0^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx = 1$ , daí

$$\begin{aligned} \int_0^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha-1} e^{-\beta x} dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha)}{\beta^\alpha} \\ &= 1 \end{aligned}$$

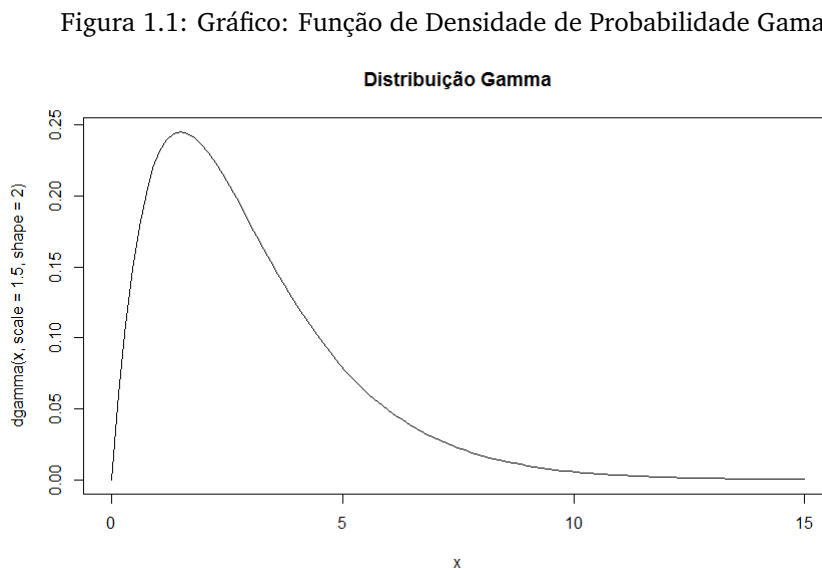
assim, fica demonstrado que  $\int_0^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx = 1$

Portanto, a função de distribuição Gama é uma (fdp).

□

"A distribuição gama é uma das mais gerais distribuições, pois diversas distribuições são caso particular dela como por exemplo a exponencial, a qui-quadrado, entre outras. Essa distribuição tem como suas principais aplicações a análise de tempo de vida de produtos."(ver em [1])

Na figura 1.1 o gráfico de uma distribuição de probabilidade Gama com parâmetros  $\alpha = 1,5$ ,  $\beta = 2$  e  $0 < x < 15$



Fonte: Elaborada pelo autor.

## 1.1.2 Modelo Log-Normal

Uma variável aleatória  $X$  possui distribuição log normal de parâmetros  $\mu$  e  $\sigma$  com  $-\infty < \mu < \infty$  e  $\sigma > 0$  quando sua função densidade de probabilidade é dada por:

$$f(x) = \frac{1}{x \cdot \sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} \quad (1.6)$$

com  $x > 0$

Conforme definição de função densidade de probabilidade (fdp) apresentada em 1.1, isto é,

$$\int_0^{\infty} f(x) dx = 1$$

com  $f(x) > 0$ .

Faremos a seguir a demonstração de que a Distribuição Log Normal é uma (fdp), isto é:

$$\int_0^{\infty} \frac{1}{x \cdot \sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} dx = 1 \quad (1.7)$$

com  $x > 0$

*Demonstração.* Quermos provar que  $\int_0^{\infty} \frac{1}{x \cdot \sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} dx = 1$

Na integral  $\int_0^{\infty} \frac{1}{x \cdot \sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} dx$  fazendo a substituição da variável  $x$

por  $y$ , com  $y = \frac{\log x - \mu}{\sigma}$  e conseqüentemente  $dy = \frac{1}{\sigma x} dx$ .

Neste caso, se  $x \rightarrow 0$ , temos que  $y \rightarrow -\infty$  e se  $x \rightarrow +\infty$ , temos que  $y \rightarrow +\infty$

Assim,

$$\begin{aligned} \int_0^{\infty} \frac{1}{x \cdot \sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} dx &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \frac{1}{x} \cdot e^{-\frac{y^2}{2}} \cdot \sigma x dy \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \end{aligned}$$

Para provar 1.7, basta provar que:  $\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \sqrt{2\pi}$  e assim teremos:

$$\int_0^{\infty} \frac{1}{x \cdot \sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} dx = 1$$

Seja  $I$  a integral após a mudança de variável,  $I = \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy$

elevando ao quadrado  $I$ , obtemos:  $I^2 = \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \cdot \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx$

transformando o produto das integrais em integral dupla, temos:

$$I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\left(\frac{y^2 + x^2}{2}\right)} dy dx$$

Para o cálculo da integral dupla, faremos a mudança de variáveis para coordenadas polares com  $x = r \cos\theta$  e  $y = r \sin\theta$  e com essa mudança, temos:  $dy dx = r d\theta dr$ , e segue que,

$$\begin{aligned} I^2 &= \int_0^{+\infty} \int_0^{2\pi} e^{-\frac{r^2}{2}} r d\theta dr \\ &= \pi \int_0^{+\infty} r e^{-\frac{r^2}{2}} dr \\ &= -2\pi e^{-\frac{r^2}{2}} \Big|_0^{+\infty} \\ &= 2\pi \end{aligned}$$

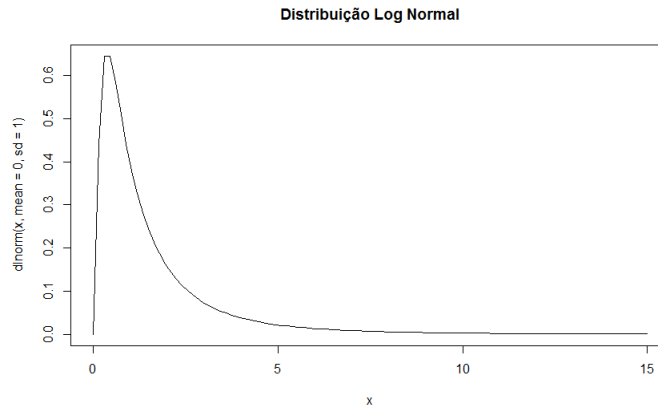
Como  $I^2 = \sqrt{2\pi} \Rightarrow I = \sqrt{2\pi}$ , segue que  $\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \sqrt{2\pi}$  e assim fica demonstrada a expressão 1.7.

Portanto, a distribuição Log Normal é uma (fdp).

□

"A distribuição Log-Normal é muito usada para caracterizar tempo de vida de produtos e materiais. Isto inclui fadiga de metal, semicondutores, diodos e isolamento elétrica."(ver em [2])

Figura 1.2: Gráfico: Função de Densidade de Probabilidade Log Normal



Fonte: Elaborada pelo autor.

Na figura 1.2, o gráfico de uma distribuição de probabilidade Log Normal com parâmetros  $\mu = 0,5$ ,  $\sigma = 1$  e  $0 < x < 15$

### 1.1.3 Modelo Weibull

Uma variável aleatória  $X$  possui distribuição Weibull com parâmetros  $\alpha > 0$  e  $\beta > 0$  quando sua função densidade de probabilidade é dada por:

$$f(x) = \frac{\alpha}{\beta} \cdot \left(\frac{x}{\beta}\right)^{\alpha-1} \cdot e^{-\left(\frac{x}{\beta}\right)^\alpha} \quad (1.8)$$

com  $x > 0$

Conforme definição de função densidade de probabilidade (fdp) apresentada em 1.1, isto é,

$$\int_0^{\infty} f(x) dx = 1$$

com  $f(x) > 0$ .

Faremos a seguir a demonstração de que a Distribuição de Weibull é uma (fdp), isto é:

$$\int_0^{\infty} \frac{\alpha}{\beta} \cdot \left(\frac{x}{\beta}\right)^{\alpha-1} \cdot e^{-\left(\frac{x}{\beta}\right)^\alpha} dx = 1$$

*Demonstração.* Queremos provar que  $\int_0^{\infty} \frac{\alpha}{\beta} \cdot \left(\frac{x}{\beta}\right)^{\alpha-1} \cdot e^{-\left(\frac{x}{\beta}\right)^{\alpha}} dx = 1$

Na integral  $\int_0^{\infty} \frac{\alpha}{\beta} \cdot \left(\frac{x}{\beta}\right)^{\alpha-1} \cdot e^{-\left(\frac{x}{\beta}\right)^{\alpha}} dx$

fazendo a substituição:  $u = \left(\frac{x}{\beta}\right)^{\alpha}$ , temos que  $du = \frac{\alpha}{\beta} \cdot \left(\frac{x}{\beta}\right)^{\alpha-1} dx$

Neste caso, se  $x \rightarrow 0$ , temos que  $u \rightarrow 0$  e  $x \rightarrow +\infty$ , temos que  $u \rightarrow +\infty$

Deste modo,

$$\begin{aligned} \int_0^{\infty} \frac{\alpha}{\beta} \cdot \left(\frac{x}{\beta}\right)^{\alpha-1} \cdot e^{-\left(\frac{x}{\beta}\right)^{\alpha}} dx &= \int_0^{\infty} e^{-u} du \\ &= -\frac{1}{e^u} \Big|_0^{\infty} \\ &= -\frac{1}{e^{\infty}} - \left(-\frac{1}{e^0}\right) \\ &= 1 \end{aligned}$$

assim, fica demonstrado que  $\int_0^{\infty} \frac{\alpha}{\beta} \cdot \left(\frac{x}{\beta}\right)^{\alpha-1} \cdot e^{-\left(\frac{x}{\beta}\right)^{\alpha}} dx = 1$

Portanto, a função de distribuição Weibull é uma (fdp).

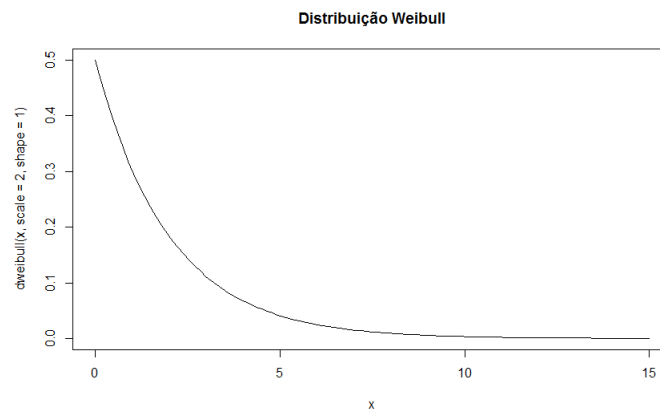
□

"A distribuição Weibull foi proposta originalmente por W. Weibull (1954) em estudos relacionados ao tempo de falha devido a fadiga de metais. Ela é frequentemente usada para descrever o tempo de vida de produtos industriais. A sua popularidade em aplicações práticas deve-se ao fato dela apresentar uma grande variedade de formas, todas com uma propriedade básica: a sua função de taxa de falha é monótona. Isto é, ou ela é crescente ou decrescente ou constante. Ela descreve adequadamente a vida de mananciais, componentes eletrônicos, cerâmicas, capacitores e dielétricos."(ver em [3])

Também verificamos aplicabilidade dessas distribuições de probabilidades (fdp)s em estudos de avaliação de regeneração natural na Engenharia Florestal, na Ecologia, teores de carbono do solo, séries de temperatura máxima e precipitação mensal.

Na figura 1.3 o gráfico de uma distribuição de probabilidade da Weibull com parâmetros  $\alpha = 2,5$ ,  $\beta = 1$  e  $0 < x < 15$

Figura 1.3: Gráfico: Função de Densidade de Probabilidade Weibull



Fonte: Elaborada pelo autor.

## 1.2 ESTIMAÇÃO DE MÁXIMA VEROSSIMILHANÇA

O Método da Máxima Verossimilhança é utilizado para obtenção de estimadores de parâmetros a partir de dados coletados de amostras.

"O Novo Dicionário da Língua Portuguesa [10], define verossímil (ou verossimilhança) aquilo que é semelhante à verdade, provável, e verossimilhança (ou verossimilidade, ou ainda verossimilitude), à qualidade ou caráter verossímil. O que seria uma amostra verossímil? Seria uma amostra que fornecesse a melhor informação possível sobre um parâmetro de interesse da população, desconhecido, e que desejamos estimar." (ver em [13])

Ainda segundo os mesmos autores, "o princípio da verossimilhança afirma que devemos escolher aquele valor do parâmetro desconhecido que maximiza a probabilidade de obter a amostra particular observada, ou seja, o valor que torna aquela amostra "mais provável". O uso desse princípio conduz a um método de estimação pelo qual se obtêm os chamados estimadores de máxima verossimilhança que, em geral têm propriedades muito boas."

Conforme [6], "O conceito de função de verossimilhança a seguir é central na Teoria de Verossimilhança."

**Definição 1.1.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da variável aleatória  $X$ , com função de densidade (ou de probabilidade)  $f(x | \theta)$ , com  $\theta \in \Theta$ , onde

$\Theta$  é o espaço paramétrico. A função de verossimilhança de  $\theta$  correspondente à amostra aleatória observada é dada por:

$$L(\theta; x) = \prod_{i=1}^n f(x_i | \theta) \quad (1.9)$$

**Definição 1.2.** O estimador de máxima verossimilhança de  $\theta$  é o valor de  $\hat{\theta} \in \Theta$  que maximiza a função de verossimilhança  $L(\theta; x)$

O logaritmo natural da função de verossimilhança de  $\theta$  é denotado por  $l(\theta; x) = \log L(\theta; x)$

Nesta seção apresentaremos o método da Máxima Verossimilhança para os três modelos de função de densidade de probabilidade (fdp), apresentadas neste capítulo, ou seja, as distribuições: Gama, Log Normal e de Weibull.

### 1.2.1 Obtenção dos parâmetros de Verossimilhança da (fdp) Gama

Após estudo das propriedades da função gama apresentadas no item (1.1.1), iremos determinar as equações de verossimilhança para estimação dos parâmetros  $\alpha$  e  $\beta$  da função de distribuição Gama.

Supondo que  $X_1, X_2, \dots, X_n$  seja uma amostra aleatória da variável  $X$  cuja fdp é dada por 1.1, as estimativas de máxima verossimilhança de  $\alpha$  e  $\beta$  são dadas por:

$$\hat{\beta} = \frac{n\hat{\alpha}}{\sum_{i=1}^n x_i} \quad (1.10)$$

e

$$n \cdot \log \left( \frac{n\hat{\alpha}}{\sum_{i=1}^n x_i} \right) - n \cdot \psi(\hat{\alpha}) + \sum_{i=1}^n \log x_i = 0 \quad (1.11)$$

$$\text{com } \psi(\hat{\alpha}) = \frac{\partial \log \Gamma(\alpha)}{\partial \alpha}$$



*Demonstração.* Seja  $L(\alpha, \gamma, x)$  a função de verossimilhança da fdp da distribuição Gama, segue que

$$\begin{aligned}
 L(\alpha, \beta, x) &= \prod_{i=1}^n \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot x_i^{\alpha-1} \cdot e^{-\beta x_i} \\
 &= \left( \frac{\beta^\alpha}{\Gamma(\alpha)} x_1^{\alpha-1} e^{-\beta x_1} \right) \cdot \left( \frac{\beta^\alpha}{\Gamma(\alpha)} x_2^{\alpha-1} e^{-\beta x_2} \right) \cdots \left( \frac{\beta^\alpha}{\Gamma(\alpha)} x_n^{\alpha-1} e^{-\beta x_n} \right) \\
 &= \frac{\beta^{n \cdot \alpha} (x_1 x_2 \cdots x_n)^{\alpha-1}}{[\Gamma(\alpha)]^n} \cdot e^{-\beta(x_1+x_2+\cdots+x_n)}
 \end{aligned} \tag{1.12}$$

Portanto, após a aplicação da definição de verossimilhança para a função de distribuição Gama, temos:

$$L(\alpha, \beta, x) = \frac{\beta^{n \cdot \alpha} (x_1 x_2 \cdots x_n)^{\alpha-1}}{[\Gamma(\alpha)]^n} \cdot e^{-\beta(x_1+x_2+\cdots+x_n)} \tag{1.13}$$

Uma alternativa mais simples para encontrarmos o valor da estimativa  $\hat{\alpha}$  e  $\hat{\beta}$  é trabalharmos com a função logaritmo da verossimilhança. Portanto

$$\log L = n \cdot \alpha \cdot \log \beta + (\alpha - 1) \cdot [\log x_1 + \log x_2 + \cdots + \log x_n] - \beta(x_1 + x_2 + \cdots + x_n) - n \log \Gamma(\alpha) \tag{1.14}$$

Para obter os estimadores de máxima verossimilhança,  $\hat{\alpha}$  e  $\hat{\beta}$ , devemos igualar as derivadas primeira da equação 1.14 a zero.

Inicialmente faremos a demonstração da equação 1.10, isto é, a estimativa de  $\beta$ .

Assim derivando 1.14 em relação a  $\beta$  e igualando à zero, temos:

$$\begin{aligned}
 \frac{\partial \log L}{\partial \beta} &= 0 \\
 \Rightarrow \frac{n \cdot \alpha}{\beta} - (x_1 + x_2 + \cdots + x_n) &= 0 \\
 \Rightarrow \frac{n \cdot \alpha}{\beta} &= (x_1 + x_2 + \cdots + x_n) \\
 \Rightarrow \frac{\alpha}{\beta} &= \frac{x_1 + x_2 + \cdots + x_n}{n}
 \end{aligned}$$

$$\Rightarrow \hat{\beta} = \frac{n \cdot \hat{\alpha}}{\sum_{i=1}^n x_i}$$

e assim concluída a demonstração da equação 1.10

A seguir, demonstraremos a equação 1.11 que serve para estimativa do parâmetro  $\alpha$ .

Derivando 1.14 em relação a  $\alpha$  e igualando a zero, temos:

$$\frac{\partial \log L}{\partial \alpha} = 0$$

$$\Rightarrow n \log \beta + \log x_1 + \log x_2 + \dots + \log x_n - n \cdot \frac{d[\log \Gamma(\alpha)]}{d\alpha} = 0$$

$$\Rightarrow n \log \beta + \sum_{i=1}^n \log x_i - n \cdot \frac{d[\log \Gamma(\alpha)]}{d\alpha} = 0$$

$$\Rightarrow \log \beta + \frac{\sum_{i=1}^n \log x_i}{n} - \frac{d[\log \Gamma(\alpha)]}{d\alpha} = 0$$

como  $\hat{\beta} = \frac{n \cdot \hat{\alpha}}{\sum_{i=1}^n x_i}$  e  $\psi(\alpha) = \frac{\partial \log[\Gamma(\alpha)]}{\partial \alpha}$ , temos:

$$n \cdot \log \left( \frac{n \hat{\alpha}}{\sum_{i=1}^n x_i} \right) - n \cdot \psi(\hat{\alpha}) + \sum_{i=1}^n \log x_i = 0$$

e assim concluída a demonstração da equação 1.11, isto é, a estimativa de  $\alpha$ .

Portanto, com a demonstração das duas equações 1.10 e 1.11 ficam determinadas as estimativas de  $\beta$  e  $\alpha$  da função de distribuição Gama.  $\square$

### 1.2.2 Obtenção dos parâmetros de Verossimilhança da (fdp) Log Normal

Aqui iremos determinar as equações de verossimilhança para estimação dos parâmetros  $\mu$  e  $\sigma^2$  da função de distribuição Log Normal.

Supondo que  $X_1, X_2, \dots, X_n$  seja uma amostra aleatória da variável  $X$  cuja (fdp) é dada por 1.6, as estimativas de máxima verossimilhança de  $\mu$  e  $\sigma^2$  são dadas por:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \log x_i \quad (1.15)$$

e

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left( \log x_i - \frac{1}{n} \sum_{i=1}^n \log x_i \right)^2 \quad (1.16)$$

*Demonstração.* Seja  $L(\mu, \sigma, x)$  a função de verossimilhança da fdp da distribuição Log Normal, dada por:

$$L(\mu, \sigma, x) = \prod_{i=1}^n \frac{1}{x_i \sigma \sqrt{2\pi}} \cdot e^{\left[ \frac{-(\log x_i - \mu)^2}{2\sigma^2} \right]}$$

Escrevendo  $L(\mu, \sigma, x)$  com a aplicação das operações: produtório e propriedades dos logaritmos, temos:

$$\begin{aligned} L(\mu, \sigma, x) &= \prod_{i=1}^n \frac{1}{x_i \sigma \sqrt{2\pi}} \cdot e^{\left[ \frac{-(\log x_i - \mu)^2}{2\sigma^2} \right]} \\ &= \frac{1}{x_1 \sigma \sqrt{2\pi}} \cdot e^{\left[ \frac{-(\log x_1 - \mu)^2}{2\sigma^2} \right]} \cdot \frac{1}{x_2 \sigma \sqrt{2\pi}} \cdot e^{\left[ \frac{-(\log x_2 - \mu)^2}{2\sigma^2} \right]} \dots \\ &\dots \cdot \frac{1}{x_n \sigma \sqrt{2\pi}} \cdot e^{\left[ \frac{-(\log x_n - \mu)^2}{2\sigma^2} \right]} \\ &= \frac{1}{(x_1 x_2 \dots x_n) \sigma^n (\sqrt{2\pi})^n} \cdot e^{\left[ \frac{-[(\log x_1 - \mu)^2 + (\log x_2 - \mu)^2 + \dots + (\log x_n - \mu)^2]}{2\sigma^2} \right]} \end{aligned} \quad (1.17)$$

Portanto, após a aplicação da definição de verossimilhança para a função de distribuição Log Normal, temos:

$$L(\mu, \sigma, x) = \frac{1}{(x_1 x_2 \dots x_n) \sigma^n (\sqrt{2\pi})^n} \cdot e^{\left[ \frac{-[(\log x_1 - \mu)^2 + (\log x_2 - \mu)^2 + \dots + (\log x_n - \mu)^2]}{2\sigma^2} \right]} \quad (1.18)$$

Uma alternativa mais simples para encontrarmos o valor da estimativa  $\hat{\mu}$  e  $\hat{\sigma}^2$  é trabalharmos com a função logaritmo da verossimilhança.

Portanto,

$$\log L = -\log(x_1 x_2 \cdots x_n) - n \log \sigma - \frac{n}{2} \log 2\pi - \frac{1}{2\sigma^2} [(\log x_1 - \mu)^2 + (\log x_2 - \mu)^2 + \cdots + (\log x_n - \mu)^2] \quad (1.19)$$

Para obter os estimadores de máxima verossimilhança  $\hat{\mu}$  e  $\hat{\sigma}^2$  devemos igualar as derivadas primeira da equação 1.19 a zero.

Primeiro derivando 1.19 em relação a  $\mu$  e igualando à zero, temos:

$$\frac{\partial \log L}{\partial \mu} = 0$$

$$\Rightarrow -\frac{1}{2\sigma^2} \cdot [-2(\log x_1 - \mu) - 2(\log x_2 - \mu) - \cdots - 2(\log x_n - \mu)] = 0$$

$$\Rightarrow \frac{1}{\sigma^2} \sum_{i=1}^n (\log x_i - \mu) = 0$$

$$\Rightarrow \sum_{i=1}^n \log(x_i - \mu) = 0$$

$$\Rightarrow \sum_{i=1}^n \log x_i - \mu \cdot n = 0$$

$$\Rightarrow \hat{\mu} = \frac{\sum_{i=1}^n \log x_i}{n}$$

e assim concluída a demonstração da equação 1.15, isto é, a estimativa de  $\hat{\mu}$

Agora derivando 1.19 em relação a  $\sigma$  e igualando a zero, demonstraremos a estimativa de  $\sigma$ .

$$\frac{\partial \log L}{\partial \sigma} = 0$$

$$\Rightarrow -\frac{n}{\sigma} - \frac{1}{2} [(\log x_1 - \mu)^2 + (\log x_2 - \mu)^2 + \cdots + (\log x_n - \mu)^2] \cdot \left(-\frac{2}{\sigma^3}\right) = 0$$

$$\Rightarrow -\frac{n}{\sigma} + \frac{1}{\sigma^3} [(\log x_1 - \mu)^2 + (\log x_2 - \mu)^2 + \cdots + (\log x_n - \mu)^2] = 0$$

$$\Rightarrow \frac{n}{\sigma} = \frac{1}{\sigma^3} [(\log x_1 - \mu)^2 + (\log x_2 - \mu)^2 + \cdots + (\log x_n - \mu)^2]$$

$$\Rightarrow \sigma^2 = \frac{\sum_{i=1}^n \log(x_i - \mu)^2}{n}$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{n} \cdot \sum_{i=1}^n \left( \log x_i - \frac{1}{n} \sum_{i=1}^n \log x_i \right)^2$$

e assim concluída a demonstração da equação 1.16, isto é, a estimativa de  $\sigma^2$ .

Portanto, com a demonstração das duas equações 1.15 e 1.16 ficam determinadas as estimativas de  $\mu$  e  $\sigma$  da função de distribuição Log Normal.  $\square$

### 1.2.3 Obtenção dos parâmetros de Verossimilhança da (fdp) Weibull

Aqui iremos determinar as equações de verossimilhança para estimação dos parâmetros  $\beta$  e  $\alpha$  da função de distribuição Weibull.

Supondo que  $X_1, X_2, \dots, X_n$  seja uma amostra aleatória da variável aleatória  $X$  cuja fdp é dada por 1.8, as estimativas de máxima verossimilhança de  $\beta$  e  $\alpha$  são dadas por:

$$\hat{\beta} = \left( \frac{1}{n} \sum_{i=1}^n x_i^{\hat{\alpha}} \right)^{\frac{1}{\hat{\alpha}}} \quad (1.20)$$

e

$$\hat{\alpha} = \left[ \left( \sum_{i=1}^n x_i^{\hat{\alpha}} \cdot \log x_i \right) \left( \sum_{i=1}^n x_i^{\hat{\alpha}} \right)^{-1} - \frac{1}{n} \sum_{i=1}^n \log x_i \right]^{-1} \quad (1.21)$$

*Demonstração.* Seja  $L(\alpha, \beta, x)$  a função de verossimilhança da fdp da distribuição de Weibull dada por:

$$L(\alpha, \beta, x) = \prod_{i=1}^n \frac{\alpha}{\beta} \cdot \left( \frac{x_i}{\beta} \right)^{\alpha-1} \cdot e^{-\left( \frac{x_i}{\beta} \right)^\alpha}$$

Escrevendo  $L(\alpha, \beta, x)$  com a execução das operações do produtório e das propriedades da potenciação, temos:

$$\begin{aligned}
 L(\alpha, \beta, x) &= \prod_{i=1}^n \frac{\alpha}{\beta} \cdot \left(\frac{x_i}{\beta}\right)^{\alpha-1} \cdot e^{-\left(\frac{x_i}{\beta}\right)^\alpha} \\
 &= \frac{\alpha}{\beta} \cdot \left(\frac{x_1}{\beta}\right)^{\alpha-1} \cdot e^{-\left(\frac{x_1}{\beta}\right)^\alpha} \cdot \frac{\alpha}{\beta} \cdot \left(\frac{x_2}{\beta}\right)^{\alpha-1} \cdot e^{-\left(\frac{x_2}{\beta}\right)^\alpha} \cdots \frac{\alpha}{\beta} \left(\frac{x_n}{\beta}\right)^{\alpha-1} \cdot e^{-\left(\frac{x_n}{\beta}\right)^\alpha} \\
 &= \frac{\alpha^n (x_1 x_2 \cdots x_n)^{\alpha-1}}{\beta^n \cdot \beta^{(\alpha-1) \cdot n}} \cdot e^{\left[-\left(\frac{x_1}{\beta}\right)^\alpha - \left(\frac{x_2}{\beta}\right)^\alpha \cdots - \left(\frac{x_n}{\beta}\right)^\alpha\right]} \\
 &= \left(\frac{\alpha}{\beta^\alpha}\right)^n \cdot (x_1 x_2 \cdots x_n)^{\alpha-1} \cdot e^{\left[-\left(\frac{x_1}{\beta}\right)^\alpha - \left(\frac{x_2}{\beta}\right)^\alpha \cdots - \left(\frac{x_n}{\beta}\right)^\alpha\right]} \quad (1.22)
 \end{aligned}$$

Portanto, após a aplicação da definição de verossimilhança para a função de distribuição Weibull, temos:

$$L(\alpha, \beta, x) = \left(\frac{\alpha}{\beta^\alpha}\right)^n \cdot (x_1 x_2 \cdots x_n)^{\alpha-1} \cdot e^{\left[-\left(\frac{x_1}{\beta}\right)^\alpha - \left(\frac{x_2}{\beta}\right)^\alpha \cdots - \left(\frac{x_n}{\beta}\right)^\alpha\right]} \quad (1.23)$$

Uma alternativa mais simples para encontrarmos o valor da estimativa de  $\alpha$  e  $\beta$  é trabalharmos com a função logaritmo da verossimilhança. Portanto:

$$\log L = n \cdot (\log \alpha - \alpha \cdot \log \beta) + (\alpha - 1) \cdot \log (x_1 x_2 \cdots x_n) - \left[ \left(\frac{x_1}{\beta}\right)^\alpha + \left(\frac{x_2}{\beta}\right)^\alpha \cdots + \left(\frac{x_n}{\beta}\right)^\alpha \right] \quad (1.24)$$

Para obter os estimadores de máxima verossimilhança  $\hat{\alpha}$  e  $\hat{\beta}$  devemos igualar as derivadas primeira da equação 1.24 a zero.

Derivando 1.24 em relação a  $\beta$  e igualando à zero, temos:

$$\begin{aligned}
 \frac{\partial \log L}{\partial \beta} &= 0 \\
 \Rightarrow n \cdot \left(-\frac{\alpha}{\beta}\right) + \left(\frac{\alpha x_1^\alpha}{\beta^{\alpha+1}} + \frac{\alpha x_2^\alpha}{\beta^{\alpha+1}} + \cdots + \frac{\alpha x_n^\alpha}{\beta^{\alpha+1}}\right) &= 0 \\
 \Rightarrow n \cdot \frac{\alpha}{\beta} &= \alpha \cdot \left(\frac{x_1^\alpha}{\beta^{\alpha+1}} + \frac{x_2^\alpha}{\beta^{\alpha+1}} + \cdots + \frac{x_n^\alpha}{\beta^{\alpha+1}}\right) \\
 \Rightarrow \frac{n}{\beta} &= \frac{1}{\beta^{\alpha+1}} \sum_{i=1}^n x_i^\alpha
 \end{aligned}$$

$$\Rightarrow \frac{\beta^{\alpha+1}}{\beta} = \frac{\sum_{i=1}^n x_i^\alpha}{n}$$

$$\Rightarrow \beta^\alpha = \frac{1}{n} \cdot \sum_{i=1}^n x_i^\alpha$$

$$\Rightarrow \hat{\beta} = \left( \frac{1}{n} \sum_{i=1}^n x_i^{\hat{\alpha}} \right)^{\frac{1}{\hat{\alpha}}}$$

e assim concluída a demonstração da equação 1.20, isto é, a estimativa de  $\beta$

Agora derivando 1.24 em relação a  $\alpha$  e igualando a zero, demonstraremos a estimativa de  $\alpha$ .

$$\frac{\partial \log L}{\partial \alpha} = 0$$

$$\Rightarrow \frac{n}{\alpha} - n \cdot \log \beta + \sum_{i=1}^n \log x_i - \sum_{i=1}^n \left( \frac{x_i}{\beta} \right)^\alpha \log \left( \frac{x_i}{\beta} \right) = 0$$

$$\Rightarrow \frac{n}{\alpha} - n \cdot \log \beta + \sum_{i=1}^n \log x_i - \frac{1}{\beta^\alpha} \left[ \sum_{i=1}^n x_i^\alpha (\log x_i - \log \beta) \right] = 0$$

Da equação 1.20, podemos escrever que:  $\beta^\alpha = \frac{1}{n} \cdot \sum_{i=1}^n x_i^\alpha$ , logo

$$\frac{n}{\alpha} - n \cdot \log \beta + \sum_{i=1}^n \log x_i - \frac{1}{\beta^\alpha} \cdot \sum_{i=1}^n x_i^\alpha \log x_i + \frac{\log \beta}{\frac{1}{n} \cdot \sum_{i=1}^n x_i^\alpha} \cdot \sum_{i=1}^n x_i^\alpha = 0$$

$$\Rightarrow \frac{n}{\alpha} - n \cdot \log \beta + \sum_{i=1}^n \log x_i - \frac{1}{\beta^\alpha} \cdot \sum_{i=1}^n x_i^\alpha \log x_i + n \cdot \log \beta = 0$$

$$\Rightarrow \frac{n}{\alpha} + \sum_{i=1}^n \log x_i - \frac{1}{\beta^\alpha} \cdot \sum_{i=1}^n x_i^\alpha \log x_i = 0$$

$$\Rightarrow \frac{1}{\alpha} = \frac{1}{\sum_{i=1}^n x_i^\alpha} \cdot \sum_{i=1}^n x_i^\alpha \log x_i - \frac{1}{n} \sum_{i=1}^n \log x_i$$

$$\Rightarrow \hat{\alpha} = \left[ \left( \sum_{i=1}^n x_i^{\hat{\alpha}} \cdot \log x_i \right) \left( \sum_{i=1}^n x_i^{\hat{\alpha}} \right)^{-1} - \frac{1}{n} \sum_{i=1}^n \log x_i \right]^{-1}$$

e assim concluída a demonstração da equação 1.21, isto é, a estimativa de  $\alpha$ .

Portanto, com a demonstração das duas equações 1.20 e 1.21 ficam determinadas as estimativas de  $\beta$  e  $\alpha$  da função de distribuição Weibull.  $\square$

### 1.3 TESTES DE ADERÊNCIA E CRITÉRIOS DE MELHOR AJUSTE

#### 1.3.1 Testes de Aderência

"A utilização de funções de probabilidade requer o uso de testes para provar a adaptação dos dados às funções, conhecidos como testes de aderência, os quais visam verificar forma de uma distribuição, através da análise de adequação dos dados à curva de um modelo de distribuição hipotética."(ver em [4])

Alguns testes de aderência bem conhecidos e importantes: Qui-quadrado, Kolmogorov-Smirnov, Lilliefors, Shapiro-Wilk, Cramer-von Mises.

Os Testes podem ser calculados pelos softwares estatísticos, como o R, R Studio, Minitab, SPSS entre outros.

Segundo Assis et al., em [5], "os testes de aderência servem para comparar as probabilidades empíricas de uma variável com as probabilidades teóricas estimadas pela função de distribuição em teste, verificando-se os valores da amostra podem ser provenientes de uma população com aquela distribuição teórica."

#### 1.3.2 Critérios de Melhor Ajuste

"A modelagem matemática é de grande importância no estudo de fenômenos da natureza, assim como nas áreas de psicologia, botânica, entre outros. Ao definir um modelo matemático para explicar um conjunto de dados em estudo, é possível fazer inferências sobre o fenômeno de interesse. Um modelo é dado como uma estrutura de parâmetros, que são definidos com base nos dados coletados, através de processos de estimação como os Mínimos Quadrados ou o Estimador de Máxima Verossimilhança. O objetivo destes métodos é estimar os parâmetros do modelo de forma a acolher melhor os dados utilizados para a modelagem. Estas técnicas são reconhecidas e largamente utilizadas, de forma que não há muita discussão sobre sua eficácia. Entretanto, tal paradigma não existe na hora da definição do tipo do modelo, nem da sua ordem:



Não existe uma metodologia única, tampouco uma metodologia que funcione bem em todas as situações."(ver em [15])

"Ao trabalharmos com modelos é necessário que internalizemos sua condição de aproximação da realidade, ou seja, não existem modelos probabilísticos exatos, sempre haverá perda de informações, fato inerente à presença da componente de aleatoriedade. Dessa forma, é necessário selecionar o melhor modelo dentre aqueles utilizados no ajuste de dados."(ver em [14])

"Se uma boa estimativa para a log verossimilhança esperada puder ser obtida através dos dados observados, esta estimativa poderá ser utilizada como um critério para comparar modelos. Assim um modo de comparar  $n$  modelos, é simplesmente comparar as magnitudes da função suporte maximizada  $L(\hat{\theta}_i)$ ."(ver em [8])

Para escolha do melhor modelo vamos utilizar o Critério de Informação de Akaike (AIC), a mais usada e criada por Akaike em 1983, baseado na Teoria de Decisão,

$$AIC = -2 \cdot \log L(\hat{\theta}) + 2p \quad (1.25)$$

onde  $\log L(\hat{\theta})$  representa o valor máximo da função de verossimilhança e  $p$  é o número de parâmetros.

O melhor modelo é o que apresenta o menor valor de AIC de acordo com esse critério.

Na estatística Bayesiana, o Critério de Informação Bayesiano (BIC) proposto por Schwarz (1978), é dado por:

$$BIC = -2 \log L(\hat{\theta}) + p \cdot \log n \quad (1.26)$$

onde  $L(\hat{\theta})$  é o modelo de verossimilhança escolhido,  $p$  é o número de parâmetros a serem estimados e  $n$  é o número de observações da amostra.

De maneira análoga ao critério (AIC), o melhor modelo é aquele que apresenta menor valor de BIC.

A diferença entre os dois critérios é que o AIC penaliza sempre com fator constante  $2p$ , enquanto que o critério BIC penaliza com o logaritmo do número de dados  $n$ , mostrando que quanto maior o número de parâmetros, um número maior de dados é necessário, e neste caso o valor de AIC será menos penalizado.

Portanto o critério BIC é mais conservador que o AIC.

## 1.4 SOFTWARE R

"A influência da tecnologia na Estatística e no seu ensino tem sido reconhecida internacionalmente, designadamente pela International Association for Statistical Education (IASE), nos muitos congressos e publicações centradas na Educação Estatística. Nestas instâncias tem-se discutido sobre o software disponível para o ensino, as mudanças ao nível do conteúdo e das metodologias de ensino que o seu uso tem implicado e o seu impacto na aprendizagem e nas atitudes dos alunos."(ver em [9])

O R, um software de uso Livre, foi criado por Ross Ihaka e por Robert Gentleman na Universidade Auckland, Nova Zelândia, e é desenvolvido pelas pessoas em vários lugares do mundo. O nome R vem das iniciais dos nomes dos criadores.

O software R é uma linguagem muito bem estruturada para Computação Estatística e produz excelentes cálculos e gráficos.

O R é uma linguagem de programação específica bem poderosa para a análise de dados através da computação. É instalado em vários sistemas operacionais, como por exemplo no Windows e Linux. Também o R possui os pacotes específicos para cada área da matemática ou estatística que ficam disponíveis no R (em inglês CRAN) no seguinte endereço:

<https://cran.r-project.org/>

No CRAN, encontraremos o download R for Linux, download R for (Mac) OS X e download R for Windows além da documentação e os pacotes.

Também existe o R Studio, uma versão mais moderna do R que é encontrada para download no seguinte endereço:

<https://www.rstudio.com/>

No R ou R Studio podemos fazer operações matemáticas simples, funções, vetores, matrizes e gráficos. Na Estatística Descritiva, determinação das medidas descritivas e gráficos como por exemplo: histograma, box plot, ramo e folhas e pizza. Na Inferência Estatística, determina cálculos probabilísticos das distribuições discretas e contínuas e seus respectivos gráficos. Faz cálculo de Correlação e Regressão Linear entre muitos modelos probabilísticos e estatísticos unidimensional, bidimensional e multivariado.

Neste trabalho, os cálculos e gráficos foram feitos com o R Studio, exceto no início, que em para a formação das tabelas, foi utilizado o Microsoft Excel.

"Na estimação dos parâmetros das (fdp)s e respectivos gráficos foi utilizado o pacote do R Studio 'fitdistrplus'".

"Ajustar distribuições aos dados é uma tarefa muito comum nas estatísticas e consiste na escolha de uma distribuição de probabilidade modelar a variável aleatória, bem como encontrar estimativas de parâmetros para essa distribuição. Isso requer julgamento especial e geralmente necessita de um processo iterativo de escolha de distribuição, estimativa de parâmetros e avaliação de ajuste."(ver em [16] e [17])

"... o pacote 'fitdistrplus' faz a estimação dos parâmetros das distribuições de probabilidade (fdp)s pelo Método da Máxima Verossimilhança". (ver em [7])

O pacote 'fitdistrplus', após carregamento dos dados, nos devolve os parâmetros da distribuição e os gráficos Densidade Empírica e Teórica, Q-Q plot, Função Distribuição Acumulada Empírica e Teórica e P-P plot.

Foi utilizado o comando `curve` para o traçado dos gráficos de exemplo das (fdp)s e o comando `denscomp` para o traçado dos gráficos das 3 distribuições de 1 amostra em apenas 1 gráfico.

## 1.5 GRÁFICOS DE ANÁLISE E QUALIDADE DOS AJUSTES DE CURVAS DAS (FDP)S

Aqui apresentaremos a construção e aplicabilidade dos gráficos usados neste trabalho: Densidade Empírica e Teórica, Q-Q plot, Função Distribuição Acumulada Empírica e Teórica (CDF)s e P-P plot.

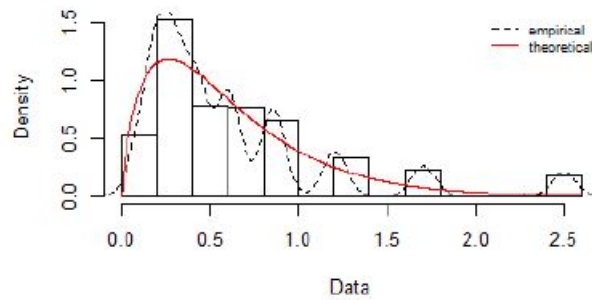
A análise desses gráficos nos dá uma ótima visualização da forma e da qualidade do ajuste das curvas das (fdp)s.

### 1.5.1 *Gráfico da Densidade Empírica e Teórica*

Segundo o Software Minitab, os gráficos de densidade são normalmente utilizados para se conhecer a forma da distribuição dos dados. É gerado como um histograma, com uma classe móvel, isto é, a classe que tem uma certa amplitude, move-se da esquerda para direita e em cada ponto estima a densidade probabilística da variável.

No software R Studio a função 'density' nos devolve a análise da densidade. Para a construção do gráfico utilizamos juntamente com a função 'plot' ou a função 'lines', para adicionar uma linha a um gráfico já criado.

Figura 1.4: Gráfico: Densidade Teórica e Empírica



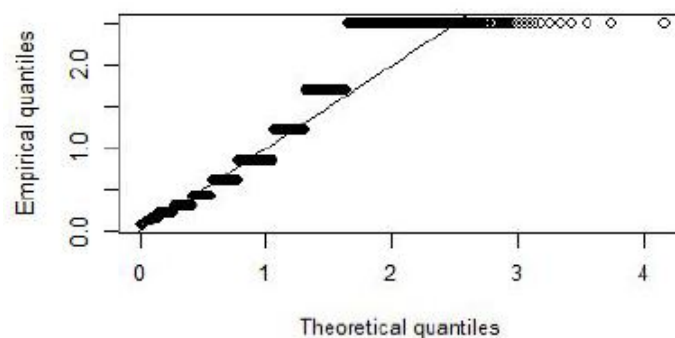
Fonte: Elaborada pelo autor.

### 1.5.2 Gráfico Q-Q plot

O Q-Q-Plot é o gráfico Quantil-Quantil, utilizado para verificar a validade de uma distribuição para um conjunto de dados. Também usado para comparação e ajuste das distribuições.

Neste caso, é feito o cálculo do valor teoricamente esperado para cada ponto de dados com base na distribuição estudada. Se os dados de fato seguirem a distribuição assumida, os pontos deste gráfico formarão aproximadamente uma linha reta.

Figura 1.5: Gráfico Q-Q Plot



Fonte: Elaborada pelo autor.

### 1.5.3 Gráfico Função Distribuição Acumulada Empírica (CDF)s

Função Distribuição Acumulada (fda) é a probabilidade que a variável aleatória  $X$  assume quando um valor é inferior ou igual a um determinado  $x$ , isto é

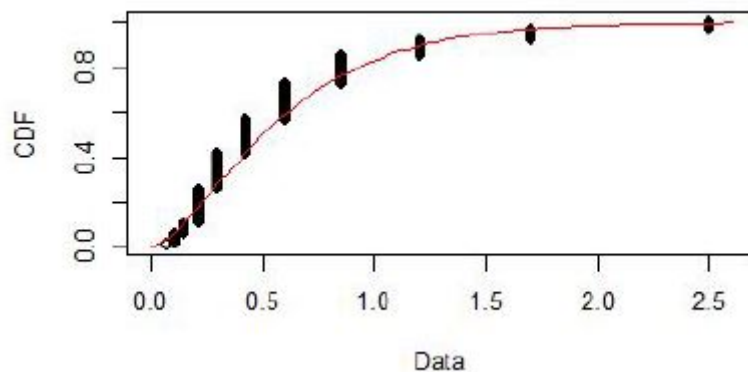
$$F(x) = P(X \leq x).$$

Ela descreve, de forma completa, a distribuição da probabilidade de uma variável aleatória de valor real  $X$ .

Conforme Software Minitab, o gráfico da Função de Distribuição Acumulada Empírica serve para avaliar a qualidade do ajuste de uma distribuição aos seus dados, também para verificação dos percentis estimados para a população e os percentis reais para os valores da amostra, e comparação das distribuições amostrais.

No R Studio utilizamos a função 'plot(ecdf)' para retorno do gráfico da Função de Distribuição Acumulada, onde *ecdf* é a sigla de Empirical Cumulative Distribution Function.

Figura 1.6: Gráfico: Distribuição Acumulada Empírica (CDF)



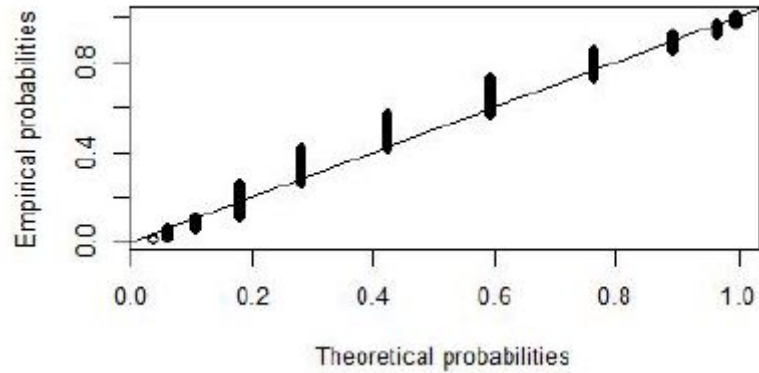
Fonte: Elaborada pelo autor.

### 1.5.4 Gráfico P-P plot

Os P-P plot são gráficos que representam no plano cartesiano, a probabilidade observada versus probabilidade estimada (esperada) de uma determinada distribuição.

É aplicada para verificação da qualidade do ajuste das distribuições de probabilidades.

Figura 1.7: Gráfico P-P Plot



Fonte: Elaborada pelo autor.

Neste trabalho iremos utilizar os gráficos: Densidade Empírica e Teórica, Q-Q Plot, Função Distribuição Acumulada Empírica e P-P Plot para verificação da forma da distribuição das (fdp)s Gama, Log Normal e Weibull e a qualidade do ajuste, com isso avaliaremos qual das (fdp)s se ajusta melhor aos dados.

I

---

## DADOS COLETADOS

---

Os dados foram produzidos por um trabalho de doutorado em andamento, de Tatiene de Souza, no Museu de Arqueologia e Etnologia (MAE-USP), cujo título é, Geoarqueologia no Abrigo de Itapeva, SP: Contribuições advindas da micromorfologia de solos.

Tem como objetivo o estudo da estratigrafia do Abrigo de Itapeva, do conjunto de montículos Areia Branca e da Estrutura Subterrânea.

Estratigrafia é uma Abordagem amostral e em microescala da sucessão das camadas de extrator que aparecem num corte geológico.

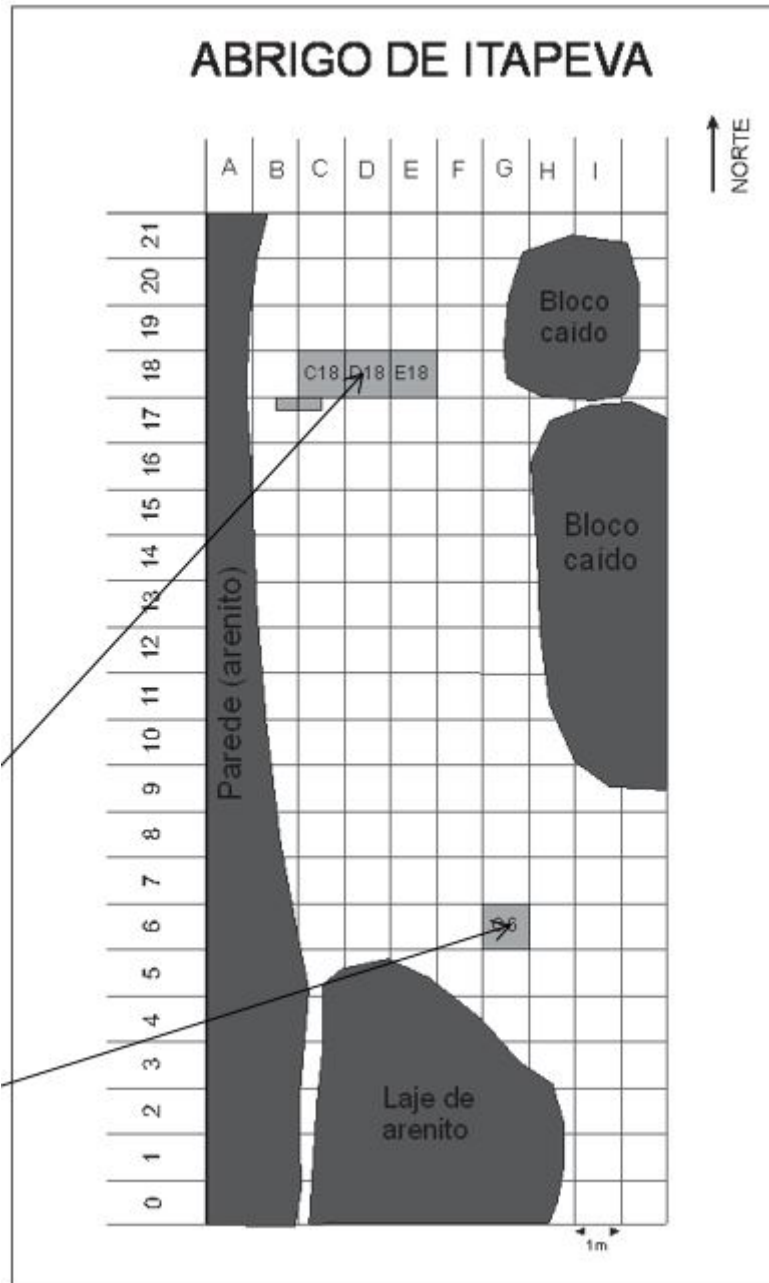
O método utilizado é da Micromorfologia de solos. O estudo de micromorfologia de solos é uma ferramenta útil para detectar materiais arqueológicos que passam despercebidos aos arqueólogos em suas escavações, tais como microartefatos, cinzas e fitólitos, além de apontar processos de formação e áreas de atividades dentro de um sítio arqueológico.

Os resultados parciais deste trabalho são derivados da produção de 14 secções delgadas a partir de coleta e descrição de materiais provenientes de duas áreas distintas do sítio, denominadas unidades de escavação, D18 e G6.

Para a modelagem e ajuste das funções que faremos neste trabalho utilizaremos dez amostras da região G6 a seguir: G6N4F3, G6N4F5, G6N4F4, G6N4F2, G6N3F2, G6N3F3, G6N1SF, G6N2SF, G6N4F6 e G6N5F2

Na figura [2.1](#) segue um croqui para melhor observar as duas áreas citadas: D18 e G6.

Figura 2.1: Identificação das áreas D18 e G6 - Croqui: Astolfo Araújo, 2017.



Fonte: Tatiane de Souza



A técnica de coleta de micromorfologia de solos no Abrigo de Itapeva é dada pelos itens a seguir:

**01.** Todas as amostras indeformadas foram recolhidas em saboneteiras com exceção de 1 amostra recolhida em caixa inoxidável;

**02.** Acondicionadas em embalagens e levadas ao laboratório de sedimentologia do Instituto de Geociência da USP;

**03.** Secadas por 48 horas em estufa;

**04.** Impregnadas com Araldite Ciba Geybi química S.A com a técnica de gotejamento;

**05.** Depois de revestidas com resinas, as amostras foram cortadas e examinadas em microscópio óptico de luz polarizada.

Abaixo os passos da técnica:

#### **Passo 1**

- Limpeza do perfil;
- Distinção de camadas ou horizontes;
- Descrição de textura

Figura 2.2: Técnica: Passo 1



Fonte: Tatiane de Souza

**Passo 2**

Escultura da amostra com faca do ponto de coleta até que esteja totalmente preenchido e retirado sem que a estrutura do solo seja desorganizada

Figura 2.3: Técnica: Passo 2



Fonte: Tatiane de Souza

**Passo 3**

Coleta da amostra, acondicionamento em embalagem especial, anotação de centímetros e orientação de topo

Figura 2.4: Técnica: Passo 3



Fonte: Tatiane de Souza

Segue a descrição dos resultados na unidade de escavação G6 – Lâminas - Observação em LEICA DM 750 P, conforme dados cedidos pela doutoranda Tatiane Souza.

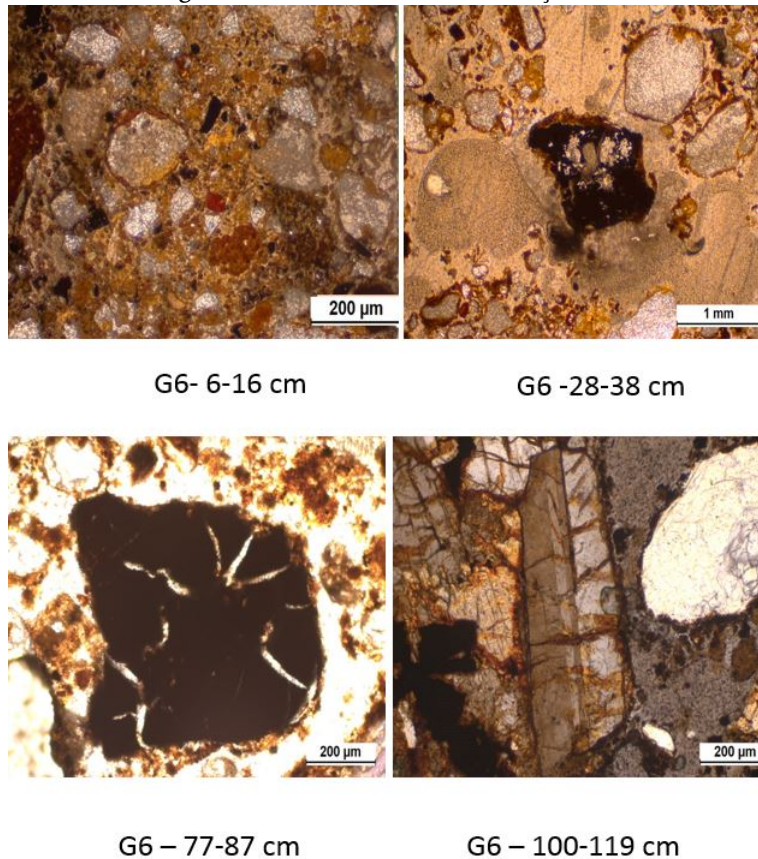
1) G6: 6-16 cm – Os grãos estão picotados Há presença de microartefato (cerâmica). Esta amostra apresenta carvões e nódulos ferruginosos desagregados.

2) G6: 28- 38 cm - A distribuição de base é concêntrica no centro da lâmina e agrupada na periferia. Os grãos estão picotados, formam pontes e películas no plasma.

3) G6: 77-87 cm – Presença de carvão em estado de craqueamento.

4) G6: 100-119 cm - Nível de base com presença, guardando semelhança com o que há na composição do próprio abrigo. Destaca-se a presença de microartefatos em lítico e osso na base..Alto nível de porosidade.

Figura 2.5: Resultados da Escavação G6



Fonte: Tatiane de Souza

Para cada uma das dez amostras selecionadas de partículas retiradas da escavação G6, foram medidos os pesos das partículas e associadas a cada malha de peneira.

Em cada tabela, a malha da peneira varia de 0,063 mm até 2,000 mm.

A quantidade em massa dos diferentes tamanhos de partículas possuem implicação arqueológicas importantes, que não nos cabe discutir neste trabalho. Por isso, a importância de determinarmos as probabilidades de encontrar os diferentes tamanhos de partículas nestes sítios arqueológicos. Daí surgiu a ideia de ajustarmos diferentes (fdp)s aos tamanhos das partículas, o que será feito a seguir.

A importância de se estudar o tamanho de partículas num sítio arqueológico está no fato de que indicam o grau de conhecimento de tecnologias de materiais, principalmente na confecção de cerâmicas, destas civilizações.

Quanto maior o tamanho das partículas encontradas, maior o conhecimento dos povos em relação à construção de materiais duráveis e resistentes, e portanto admite-se um maior grau de desenvolvimento geral destas civilizações.

A seguir, as tabelas das 10 amostras contruídas no Microsoft Excel.

Tabela 2.1: Amostra 1: G6N4F3

Malha da peneira (milímetros)	Peso (gramas)
2,000	3,1743
1,410	3,9608
1,000	5,8840
0,710	11,4827
0,500	13,4827
0,350	13,7357
0,250	13,6068
0,177	13,3120
0,125	4,7280
0,088	3,2336
0,063	1,2310
Total	87,8316

Fonte: Elaborada pelo autor.

Tabela 2.2: Amostra 2: G6N4F5

Malha da peneira (milímetros)	Peso (gramas)
2,000	5,3342
1,410	4,4119
1,000	5,9526
0,710	9,4391
0,500	20,2121
0,350	9,6164
0,250	17,7727
0,177	11,9446
0,125	3,4484
0,088	3,5170
0,063	0,7613
Total	92,4103

Fonte: Elaborada pelo autor.

Tabela 2.3: Amostra 3: G6N4F4

Malha da peneira (milímetros)	Peso (gramas)
2,000	3,7232
1,410	3,9579
1,000	5,1440
0,710	10,3688
0,500	12,4969
0,350	11,6134
0,250	16,5684
0,177	10,8369
0,125	5,2125
0,088	3,8751
0,063	2,4894
Total	86,2865

Fonte: Elaborada pelo autor.

Tabela 2.4: Amostra 4: G6N4F2

Malha da peneira (milímetros)	Peso (gramas)
2,000	5,1784
1,410	3,8709
1,000	5,1670
0,710	10,7812
0,500	11,4173
0,350	16,5693
0,250	17,1986
0,177	8,9492
0,125	2,1737
0,088	2,1975
0,063	0,6127
Total	84,1158

Fonte: Elaborada pelo autor.

Tabela 2.5: Amostra 5: G6N3F2

Malha da peneira (milímetros)	Peso (gramas)
2,000	2,9989
1,410	3,9721
1,000	5,9721
0,710	10,6556
0,500	12,5746
0,350	11,9566
0,250	18,5357
0,177	9,7877
0,125	3,8820
0,088	1,0802
0,063	0,8342
Total	82,2487

Fonte: Elaborada pelo autor.

Tabela 2.6: Amostra 6: G6N3F3

Malha da peneira (milímetros)	Peso (gramas)
2,000	6,0789
1,410	5,3545
1,000	7,4486
0,710	13,1959
0,500	18,1297
0,350	14,5960
0,250	14,7742
0,177	9,8872
0,125	3,4083
0,088	2,862
0,063	1,4770
Total	97,1823

Fonte: Elaborada pelo autor.

Tabela 2.7: Amostra 7: G6N1S/F

Malha da peneira (milímetros)	Peso (gramas)
2,000	3,7514
1,410	4,0166
1,000	5,7286
0,710	10,2362
0,500	11,2751
0,350	17,1223
0,250	10,5290
0,177	3,3771
0,125	10,4996
0,088	2,8124
0,063	1,0533
Total	80,4016

Fonte: Elaborada pelo autor.

Tabela 2.8: Amostra 8: G6N2S/F

Malha da peneira (milímetros)	Peso (gramas)
2,000	5,1815
1,410	4,5174
1,000	6,6531
0,710	11,2622
0,500	12,1651
0,350	12,4460
0,250	21,5461
0,177	12,5912
0,125	4,8062
0,088	5,1911
0,063	2,3037
Total	98,6636

Fonte: Elaborada pelo autor.



Tabela 2.9: Amostra 9: G6N4F6

Malha da peneira (milímetros)	Peso (gramas)
2,000	4,6879
1,410	7,3468
1,000	4,7764
0,710	12,4474
0,500	13,6426
0,350	13,3578
0,250	20,6012
0,177	11,1303
0,125	4,1890
0,088	2,3788
0,063	0,0000
Total	94,5582

Fonte: Elaborada pelo autor.

Tabela 2.10: Amostra 10: G6N5F2

Malha da peneira (milímetros)	Peso (gramas)
2,000	4,1489
1,410	4,6004
1,000	5,8718
0,710	12,3700
0,500	13,5670
0,350	12,0156
0,250	15,7736
0,177	10,8321
0,125	1,1598
0,088	0,7546
0,063	0,0000
Total	81,0938

Fonte: Elaborada pelo autor.

---

## MÉTODOS NUMÉRICOS PARA ESTIMATIVA DOS PARÂMETROS

---

Após a coleta dos pesos das partículas, de acordo com os seus diâmetros, os mesmos foram tabelados no Microsoft Excel com a finalidade de organizá-los e começar a aplicação de técnicas estatísticas. Foram selecionadas dez amostras do processo de escavação.

Para cada amostra foi construída uma tabela, os pesos (medidas em gramas) foram associados respectivamente aos valores das malhas das peneiras em ordem decrescente (medidas em milímetros), que correspondem aos diâmetros das partículas conforme mostra figura 3.1.

Figura 3.1: Tabela: amostra 1

Malha da peneira (milímetros)	Peso (gramas)
2,000	3,1743
1,410	3,9608
1,000	5,8840
0,710	11,4827
0,500	13,4827
0,350	13,7357
0,250	13,6068
0,177	13,3120
0,125	4,7280
0,088	3,2336
0,063	1,2310
Total	87,8316

Fonte: Elaborada pelo autor.

Em seguida, foi determinada a razão entre a medida do peso de cada classe da partícula pela soma dos pesos totais para a obtenção dos valores da probabilidade em massa. Os valores são mostrados na 3ª coluna da tabela na figura 3.2.

O diâmetro das partículas (medida em milímetros) foi estimado pelo ponto médio dos valores da malha das peneiras, mostradas na 4ª coluna da tabela, na figura 3.2.

Figura 3.2: Construção da tabela de probabilidade em massa: amostra 1

Malha da peneira(mm)	Peso(g)	Probabilidade(em massa)	Diâmetro(mm)
2,000	3,1743	0,0361	2,500
1,410	3,9608	0,0451	1,705
1,000	5,8840	0,0670	1,205
0,710	11,4827	0,1307	0,855
0,500	13,4827	0,1535	0,605
0,350	13,7357	0,1564	0,425
0,250	13,6068	0,1549	0,300
0,177	13,3120	0,1516	0,214
0,125	4,7280	0,0538	0,151
0,088	3,2336	0,0368	0,107
0,063	1,2310	0,0140	0,076
Peso Total	87,8316	1,000	

Fonte: Elaborada pelo autor.

Como o objetivo é encontrar a (fdp) que melhor modele a massa encontrada das partículas, estimamos o número de partículas proporcional à probabilidade em massa. Nos cálculos foram utilizadas quatro casas decimais após a vírgula nos valores de probabilidade em massa.

Figura 3.3: Construção da tabela

Malha da peneira(mm)	Peso(g)	Probabilidade(em massa)	Diâmetro(mm)	Nº partículas
2,000	3,1743	0,0361	2,500	361
1,410	3,9608	0,0451	1,705	451
1,000	5,8840	0,0670	1,205	670
0,710	11,4827	0,1307	0,855	1307
0,500	13,4827	0,1535	0,605	1535
0,350	13,7357	0,1564	0,425	1564
0,250	13,6068	0,1549	0,300	1549
0,177	13,3120	0,1516	0,214	1516
0,125	4,7280	0,0538	0,151	538
0,088	3,2336	0,0368	0,107	368
0,063	1,2310	0,0140	0,076	140
Peso Total	87,8316	1,000		

Fonte: Elaborada pelo autor.

Assim, obtemos uma estimativa do número de partículas relacionadas a massa para cada diâmetro médio da malha da peneira de cada amostra, como podemos ver nas duas últimas colunas na tabela da figura 3.3.

As 10 tabelas relativas as amostras estão no apêndice A.

Com estes dados, iremos verificar se existe alguma distribuição de probabilidade que possa estimar a massa de partículas de um dado diâmetro.

Deste modo foram determinados os valores dos parâmetros das três distribuições de probabilidades: Gama, Log Normal e Weibull e os seguintes gráficos: Densidade Empírica e Teórica, Q-Q plot, Função Distribuição Acumulada Empírica e Teórica, P-P plot para cada uma das dez amostras selecionadas. Também calculamos os valores AIC e BIC para cada ajuste.

Utilizamos o software R Studio e com a função FITDIST foi determinada a estimação dos parâmetros das (fdp)s, valores de AIC e BIC e gráficos.

A sequência de comandos no R Studio é dada por (exemplo para a amostra 1)

```
x<- c(rep(2.5, each=361), rep(1.705, each=451), rep(1.205, each=670), rep(0.855,
each=1307), rep(0.605, each=1535), rep(0.425, each=1564), rep(0.3, each=1549),
rep(0.2135, each=15126), rep(0.151, each=538), rep(0.1065, each=368), rep(0.0755,
each=140))
```

```
fitg<-fitdist(x, "gamma")
```

```
summary(fitg)
```

Logo, a saída no software R Plus é mostrada na figura 3.4.

Figura 3.4: Parâmetros da Gama, AIC e BIC para a amostra 1

```
> x<- c(rep(2.5, each=361), rep(1.705, each=451), rep(1.205, each=670), rep(0.855, each=1307), rep(0.605, each=1535), rep(0.425
, each=1564), rep(0.3, each=1549), rep(0.2135, each=1516), rep(0.151, each=538), rep(0.1065, each=368), rep(0.0755, each=140))
> fitg<-fitdist(x, "gamma")
> summary(fitg)
Fitting of the distribution ' gamma ' by maximum likelihood
Parameters :
      estimate Std. Error
shape 1.760777 0.02293296
rate  2.881952 0.04336665
Loglikelihood: -4260.346  AIC: 8524.691  BIC: 8539.112
Correlation matrix:
      shape  rate
shape 1.000000 0.865539
rate  0.865539 1.000000
```

Fonte: Elaborada pelo autor.

Portanto, para este exemplo da amostra 1 e ajuste pela (fdp) Gama, temos:

$$\hat{\alpha} = 1,761 \pm 0,022$$

$$\hat{\beta} = 2,882 \pm 0,043$$

$$AIC = 8524,691$$

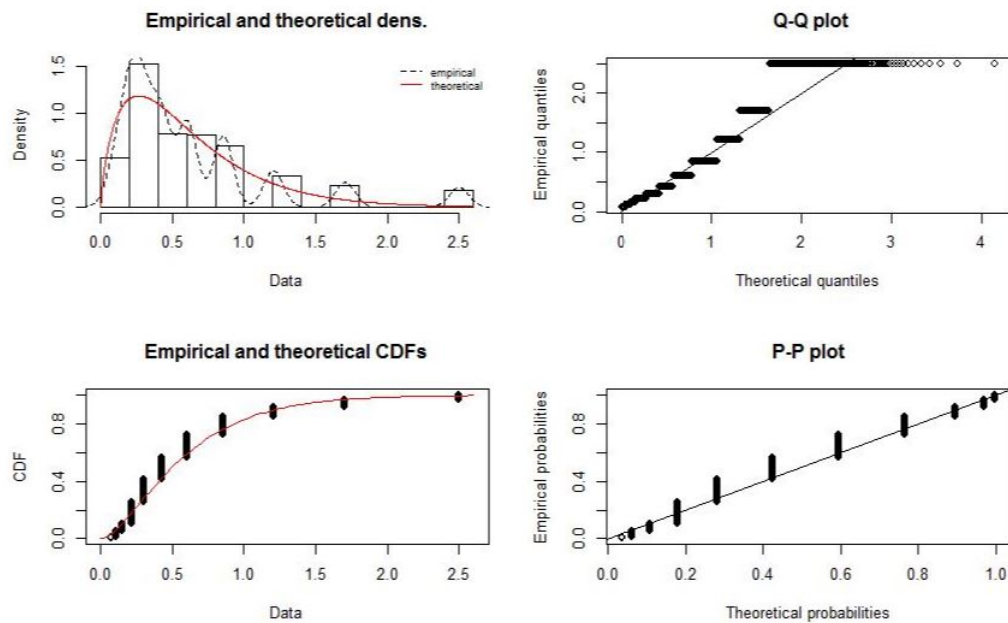
$$BIC = 8539,112$$

A obtenção dos gráficos: Densidade Empírica e Teórica, Q-Q plot, Função Distribuição Acumulada Empírica e Teórica e P-P plot é feita executando o comando abaixo:

```
plot(fitg, demp=TRUE)
```

Os gráficos são apresentados na figura 3.5 para a amostra 1.

Figura 3.5: Gráficos da Gama para a amostra 1



Fonte: Elaborada pelo autor.

Para as distribuições Log Normal e Weibull foi feito o processo análogo acima com a utilização das seguintes funções no R Studio:

\* Na (fdp) Log Normal utilizamos o comando: `fitln<-fitdist(x, "lnorm")`

\* No caso da (fdp) Weibull utilizamos o comando: `fitw<-fitdist(x, "weibull")`

\* Para os gráficos das (fdp)s Log Normal e Weibull utilizamos respectivamente os comandos: `plot(fitln, demp=TRUE)` e `plot(fitw, demp=TRUE)`.

\* Na plotagem do gráfico contendo as 3 (fdp)s foi utilizada a sequência de comandos do R Studio abaixo (exemplo para a amostra 1)

```
x<- c(rep(2.5, each=361), rep(1.705, each=451), rep(1.205, each=670), rep(0.855,
each=1307), rep(0.605, each=1535), rep(0.425, each=1564), rep(0.3, each=1549),
rep(0.2135, each=1516), rep(0.151, each=538), rep(0.1065, each=368), rep(0.0755,
each=140))
```

```
fit_g <- fitdist(x, "gamma")
```

```
fit_ln <- fitdist(x, "lnorm")
```

```
fit_w <- fitdist(x, "weibull")
```

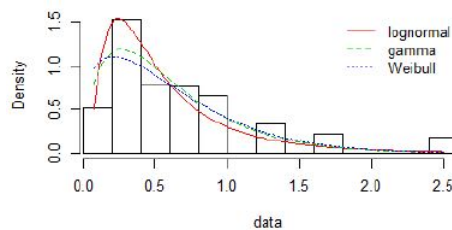
```
par(mfrow=c(2,2))
```

```
plot.legend <- c("lognormal", "gamma", "Weibull")
```

```
denscomp(list(fit_ln, fit_g, fit_w), legendtext = plot.legend)
```

Na figura 3.6 segue o gráfico produzido pela sequência dos comandos.

Figura 3.6: Exemplo gráfico com as 3 (fdp)s  
Histogram and theoretical densities



Fonte: Elaborada pelo autor.

Os resultados para todas as amostras com os 3 ajustes feitos para cada uma são apresentados no capítulo seguinte.

# 4

---

## RESULTADOS

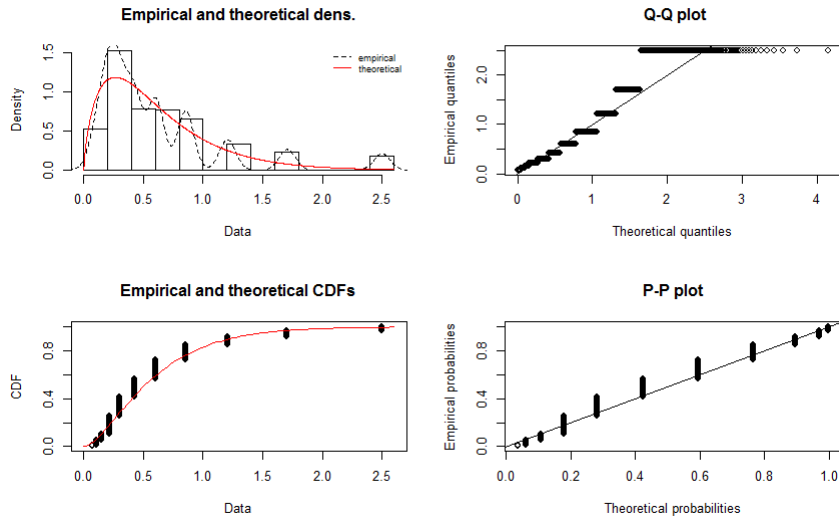
---

Neste capítulo apresentaremos os resultados dos ajustes feitos com as 10 amostras utilizadas, mostrando os gráficos e critérios AIC e BIC que permitirão selecionar qual a função densidade de probabilidade (fdp) melhor se adequam as amostras.

Após os resultados de todas as amostras, serão feitas as verificações e observações das 10 amostras estudadas.

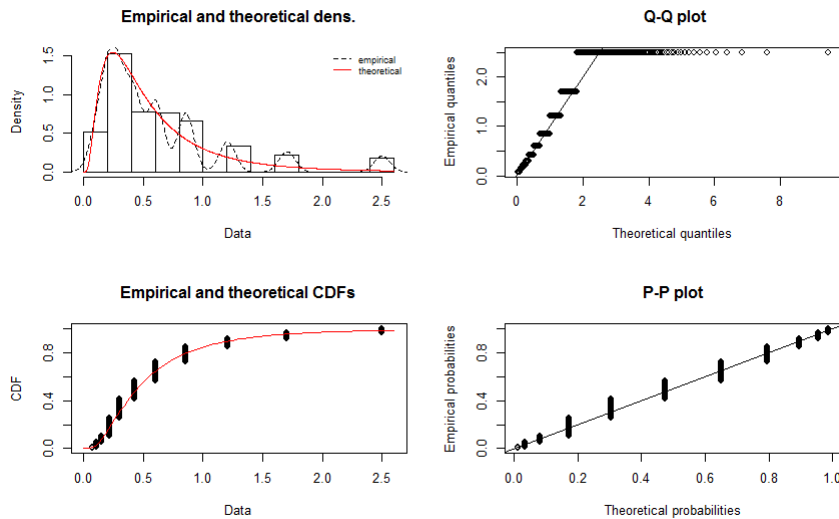
4.1 RESULTADOS OBTIDOS PARA A AMOSTRA 1

Figura 4.1: AMOSTRA 1: Curva de Distribuição GAMA



Fonte: Elaborada pelo autor.

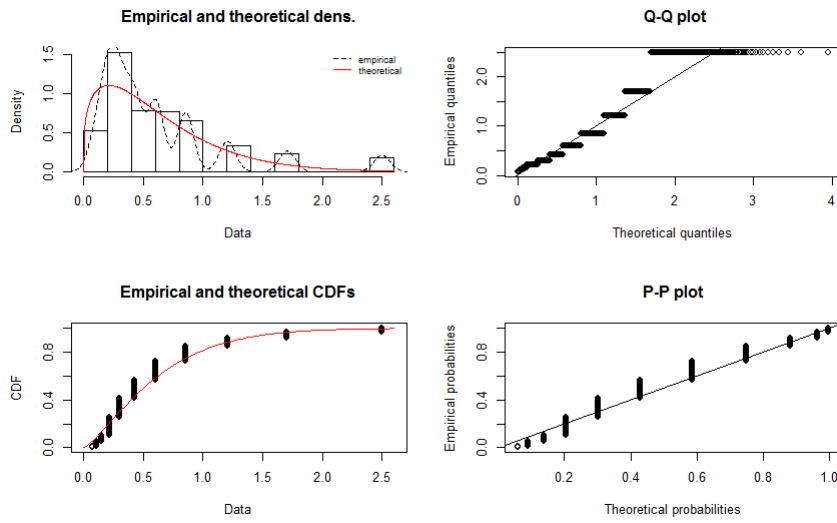
Figura 4.2: AMOSTRA 1: Curva de Distribuição LOG NORMAL



Fonte: Elaborada pelo autor.

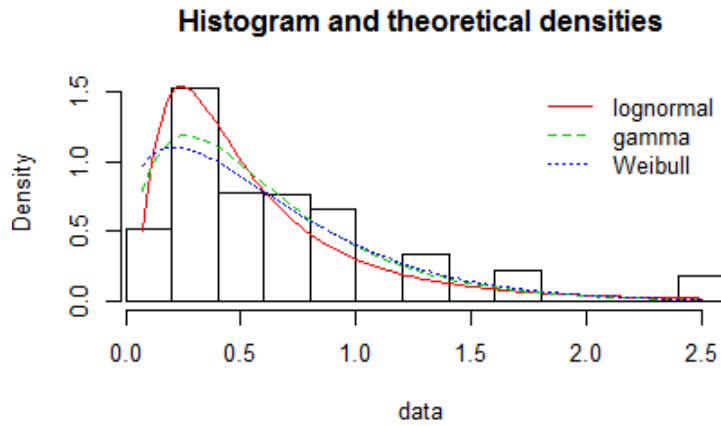


Figura 4.3: AMOSTRA 1: Curva de Distribuição WEIBULL



Fonte: Elaborada pelo autor.

Figura 4.4: AMOSTRA 1: Histograma de frequências dos dados observados e das 3 fdps ajustadas



Fonte: Elaborada pelo autor.

Tabela 4.1: AMOSTRA 1: ESTIMATIVAS DOS PARÂMETROS das 3 distribuições

Distribuições	$\alpha$	$\beta$	$\mu$	$\sigma^2$	AIC	BIC
<i>Gamma</i>	1,761 ± 0,023	2,882 ± 0,043	-	-	8524,69	8539,11
Log Normal	-	-	-0,803 ± 0,007	0,784 ± 0,005	7448,59	7463,01
Weibull	1,289 ± 0,009	0,666 ± 0,005	-	-	9094,34	9108,76

Fonte: Elaborada pelo autor.

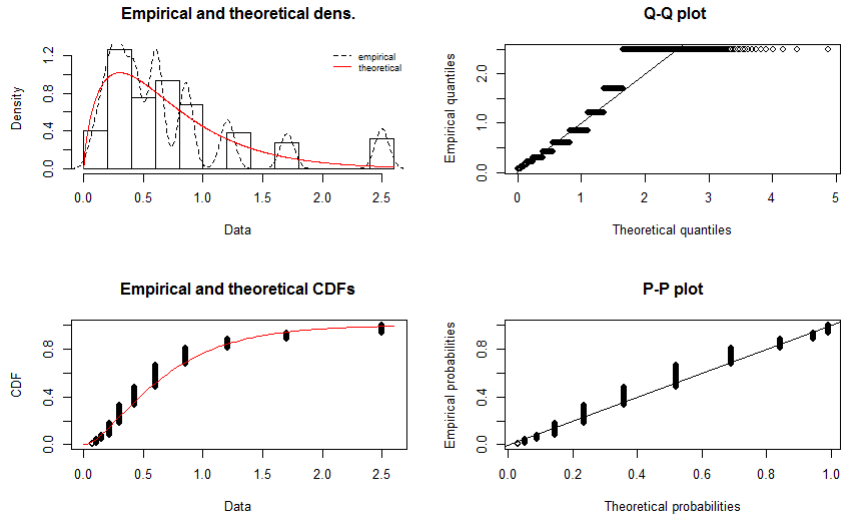
O menor valor de AIC e BIC ocorre para a (fdp)Log Normal, já o maior valor de AIC e BIC foi da (fdp) Weibull. Pelo critério de Akaike, a (fdp) Log Normal se ajusta melhor aos dados.

Nos gráficos Q-Q plot, P-P plot e da Função Distribuição Acumulada Empírica e Teórica (CDFs) verificamos uma sobreposição mais regular das linhas verticais sobre cada gráfico no caso da (fdp)Log Normal.

Pelo Histograma e a Densidade Teórica percebemos que a curva que se ajusta melhor é a (fdp)Log Normal e a (fdp)Weibull possui o pior ajuste.

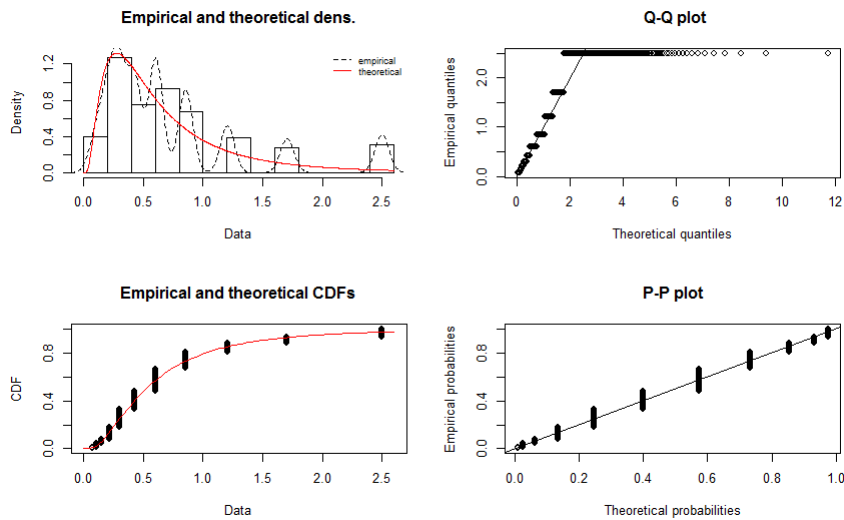
4.2 RESULTADOS OBTIDOS PARA A AMOSTRA 2

Figura 4.5: AMOSTRA 2: Curva de Distribuição GAMA



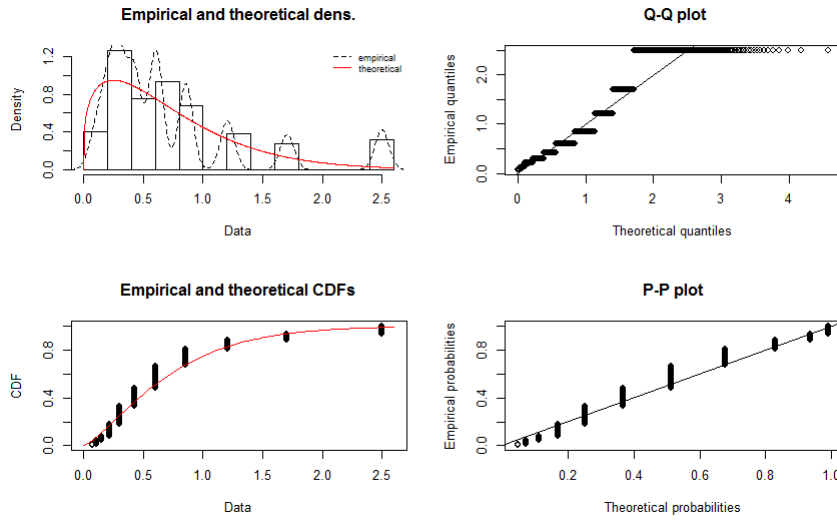
Fonte: Elaborada pelo autor.

Figura 4.6: AMOSTRA 2: Curva de Distribuição LOG NORMAL



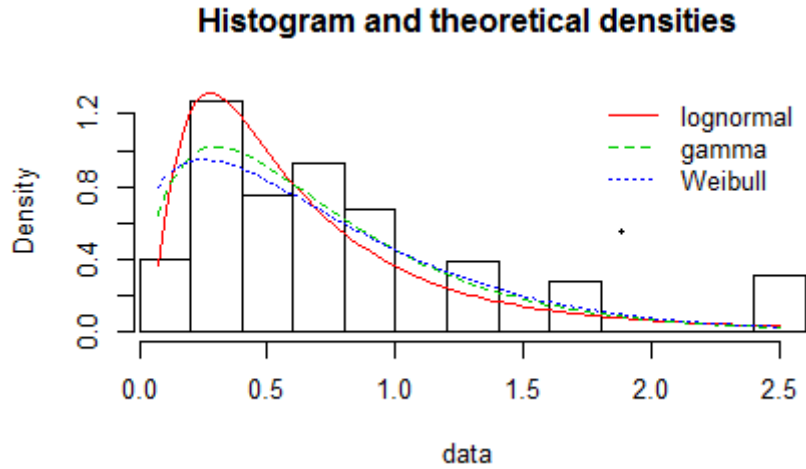
Fonte: Elaborada pelo autor.

Figura 4.7: AMOSTRA 2: Curva de Distribuição WEIBULL



Fonte: Elaborada pelo autor.

Figura 4.8: AMOSTRA 2: Histograma de frequências dos dados observados e das 3 fdps ajustadas



Fonte: Elaborada pelo autor.

Tabela 4.2: AMOSTRA 2: ESTIMATIVAS DOS PARÂMETROS das 3 distribuições

Distribuições	$\alpha$	$\beta$	$\mu$	$\sigma^2$	AIC	BIC
Gamma	$1,738 \pm 0,023$	$2,439 \pm 0,037$	-	-	11644,55	11678,97
Log Normal	-	-	$-0,654 \pm 0,008$	$0,800 \pm 0,006$	10858,31	10873,31
Weibull	$1,295 \pm 0,009$	$0,777 \pm 0,006$	-	-	12154,08	12168,50

Fonte: Elaborada pelo autor.

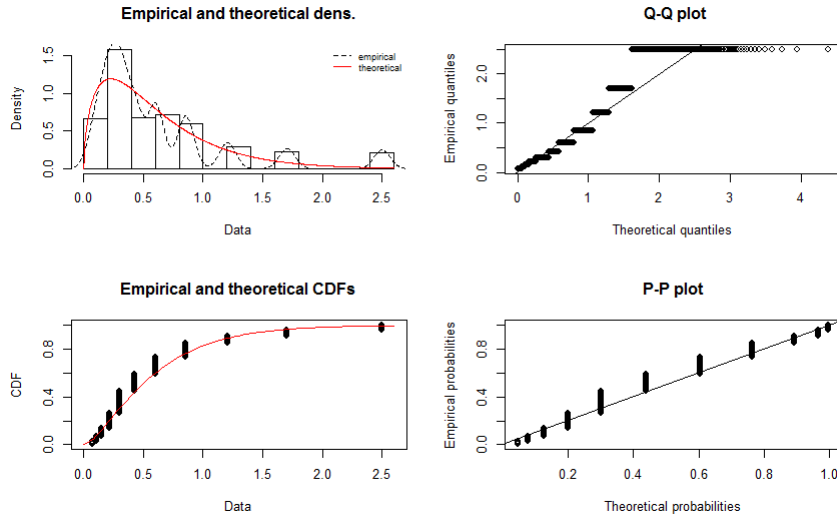
O menor valor de AIC e BIC ocorre para a (fdp)Log Normal, já o maior valor de AIC e BIC foi da (fdp) Weibull. Pelo critério de Akaike, a (fdp) Log Normal se ajusta melhor aos dados.

Nos gráficos Q-Q plot, P-P plot e da Função Distribuição Acumulada Empírica e Teórica (CDFs) verificamos uma sobreposição mais regular das linhas verticais sobre cada gráfico no caso da (fdp)Log Normal.

Pelo Histograma e a Densidade Teórica percebemos que a curva que se ajusta melhor é a (fdp)Log Normal e a (fdp)Weibull possui o pior ajuste.

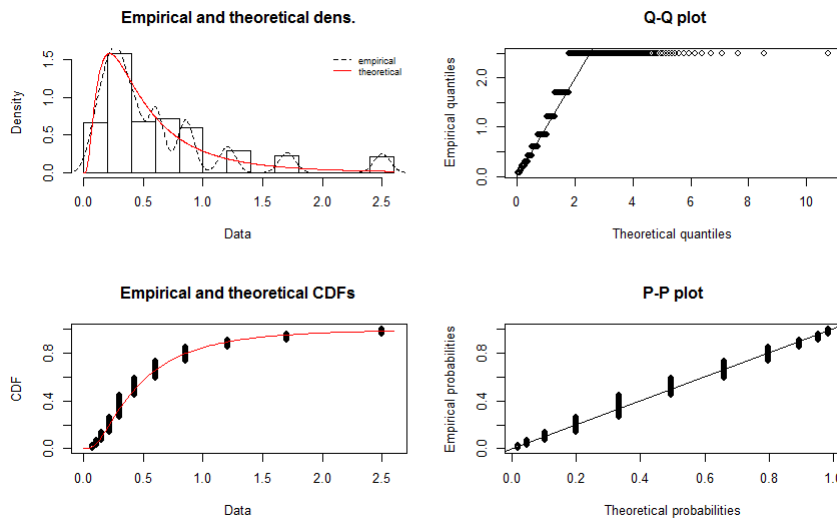
4.3 RESULTADOS OBTIDOS PARA A AMOSTRA 3

Figura 4.9: AMOSTRA 3: Curva de Distribuição GAMA



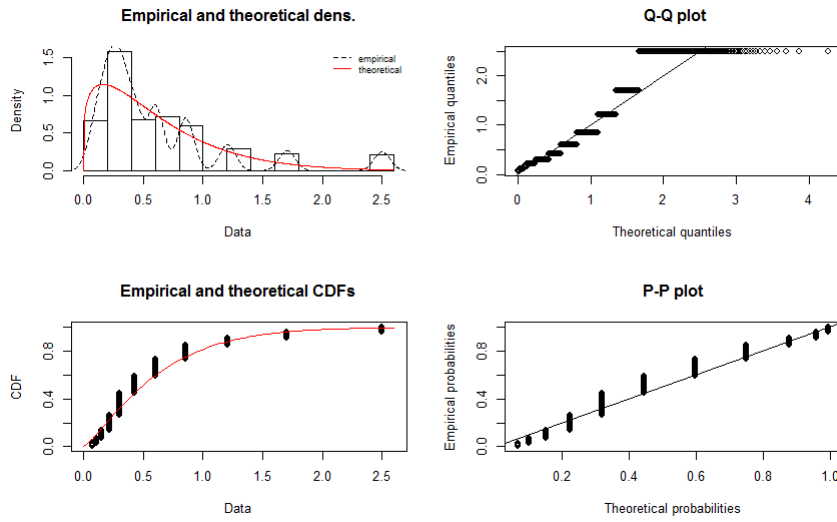
Fonte: Elaborada pelo autor.

Figura 4.10: AMOSTRA 3: Curva de Distribuição LOG NORMAL



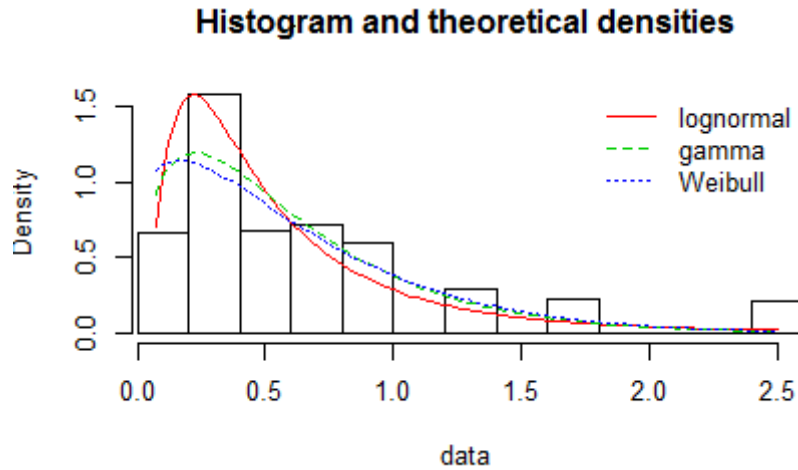
Fonte: Elaborada pelo autor.

Figura 4.11: AMOSTRA3: Curva de Distribuição WEIBULL



Fonte: Elaborada pelo autor.

Figura 4.12: AMOSTRA 3: Histograma de frequências dos dados observados e das 3 fdps ajustadas



Fonte: Elaborada pelo autor.

Tabela 4.3: AMOSTRA 3: ESTIMATIVAS DOS PARÂMETROS das 3 distribuições

Distribuições	$\alpha$	$\beta$	$\mu$	$\sigma^2$	AIC	BIC
Gamma	$1,595 \pm 0,020$	$2,632 \pm 0,040$	-	-	8833,31	8847,73
Log Normal	-	-	$-0,846 \pm 0,008$	$0,828 \pm 0,006$	7677,60	7692,02
Weibull	$1,223 \pm 0,009$	$0,652 \pm 0,006$	-	-	9306,46	9320,86

Fonte: Elaborada pelo autor.

O menor valor de AIC e BIC ocorre para a (fdp)Log Normal, já o maior valor de AIC e BIC foi da (fdp) Weibull. Pelo critério de Akaike, a (fdp) Log Normal se ajusta melhor aos dados.

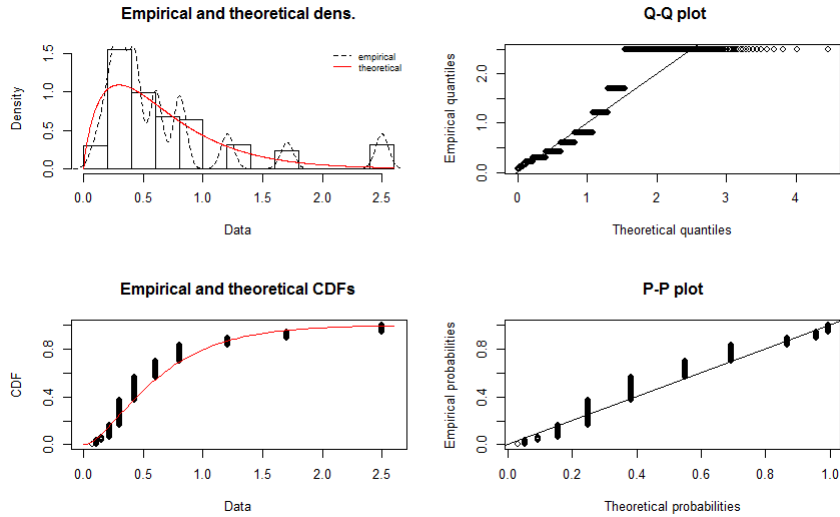
Nos gráficos Q-Q plot, P-P plot e da Função Distribuição Acumulada Empírica e Teórica (CDFs) verificamos uma sobreposição mais regular das linhas verticais sobre cada gráfico no caso da (fdp)Log Normal.

Pelo Histograma e a Densidade Teórica percebemos que a curva que se ajusta melhor é a (fdp)Log Normal e a (fdp)Weibull possui o pior ajuste.



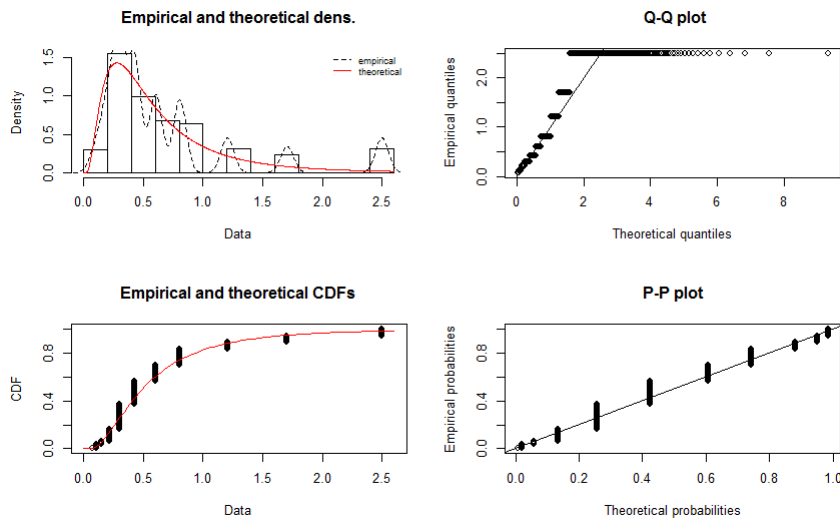
4.4 RESULTADOS OBTIDOS PARA A AMOSTRA 4

Figura 4.13: AMOSTRA 4: Curva de Distribuição GAMA



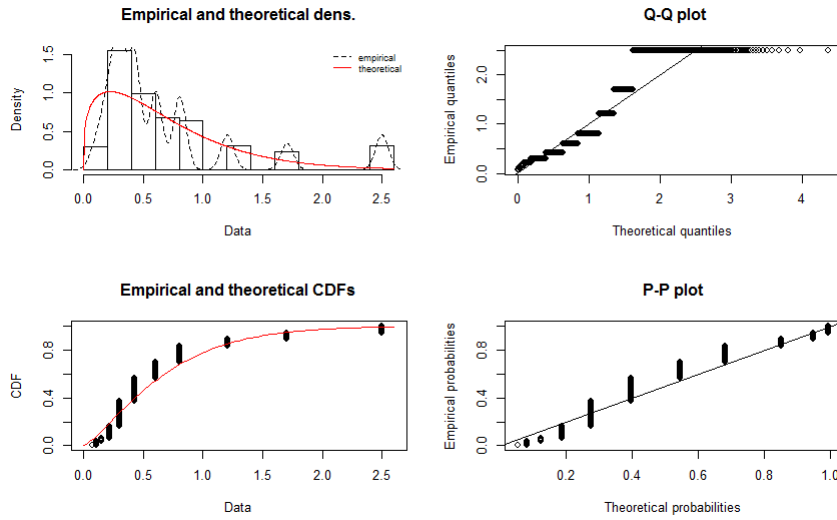
Fonte: Elaborada pelo autor.

Figura 4.14: AMOSTRA 4: Curva de Distribuição LOG NORMAL



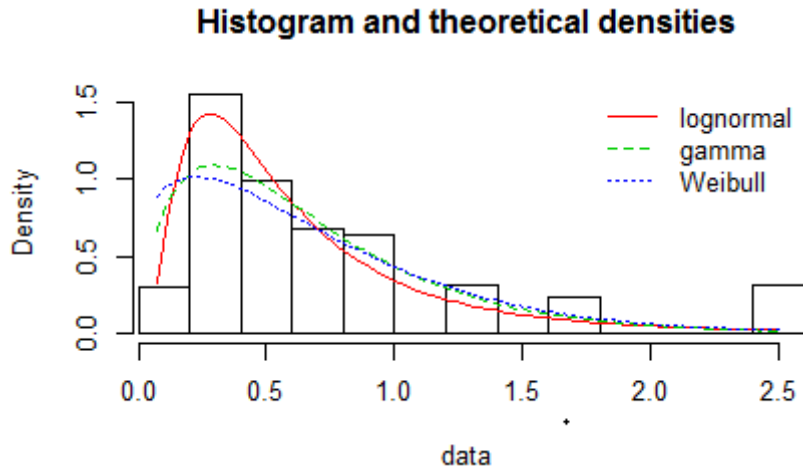
Fonte: Elaborada pelo autor.

Figura 4.15: AMOSTRA 4: Curva de Distribuição WEIBULL



Fonte: Elaborada pelo autor.

Figura 4.16: AMOSTRA 4: Histograma de frequências dos dados observados e das 3 fdps ajustadas



Fonte: Elaborada pelo autor.

Tabela 4.4: AMOSTRA 4: ESTIMATIVAS DOS PARÂMETROS das 3 distribuições

Distribuições	$\alpha$	$\beta$	$\mu$	$\sigma^2$	AIC	BIC
Gamma	$1,797 \pm 0,023$	$2,694 \pm 0,040$	-	-	10168,51	10182,93
Log Normal	-	-	$-0,709 \pm 0,005$	$0,756 \pm 0,005$	8607,53	8621,95
Weibull	$1,280 \pm 0,009$	$0,727 \pm 0,006$	-	-	10875,53	10889,95

Fonte: Elaborada pelo autor.

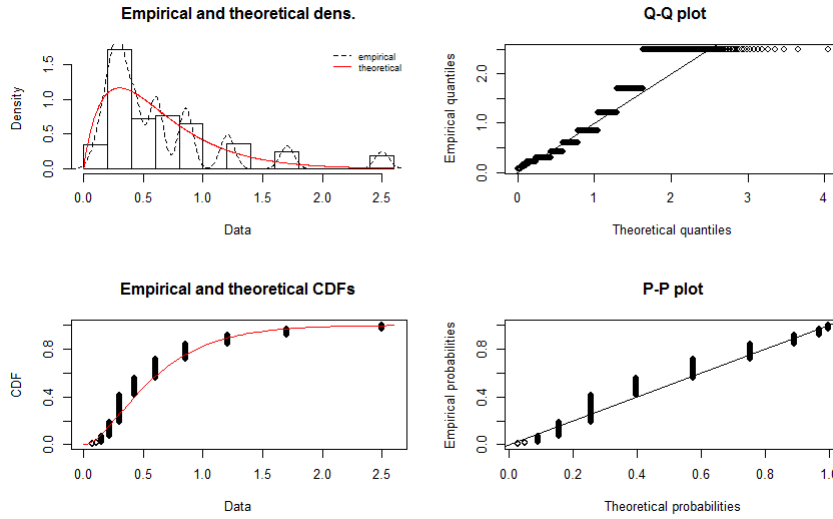
O menor valor de AIC e BIC ocorre para a (fdp)Log Normal, já o maior valor de AIC e BIC foi da (fdp) Weibull. Pelo critério de Akaike, a (fdp) Log Normal se ajusta melhor aos dados.

Nos gráficos Q-Q plot, P-P plot e da Função Distribuição Acumulada Empírica e Teórica (CDFs) verificamos uma sobreposição mais regular das linhas verticais sobre cada gráfico no caso da (fdp)Log Normal.

Pelo Histograma e a Densidade Teórica percebemos que a curva que se ajusta melhor é a (fdp)Log Normal e a (fdp)Weibull possui o pior ajuste.

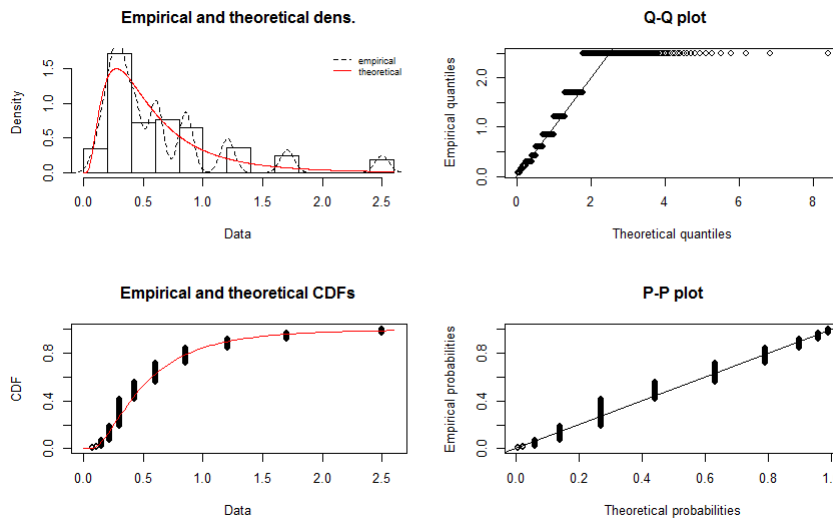
4.5 RESULTADOS OBTIDOS PARA A AMOSTRA 5

Figura 4.17: AMOSTRA 5: Curva de Distribuição GAMA



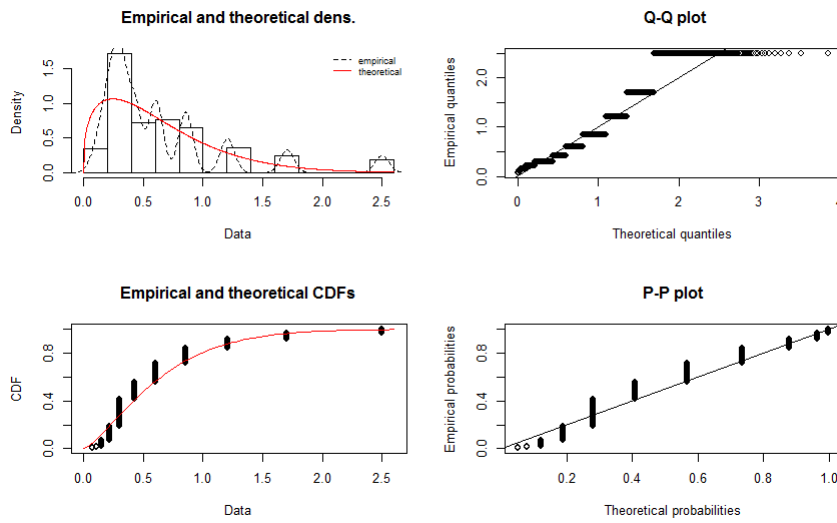
Fonte: Elaborada pelo autor.

Figura 4.18: AMOSTRA 5: Curva de Distribuição LOG NORMAL



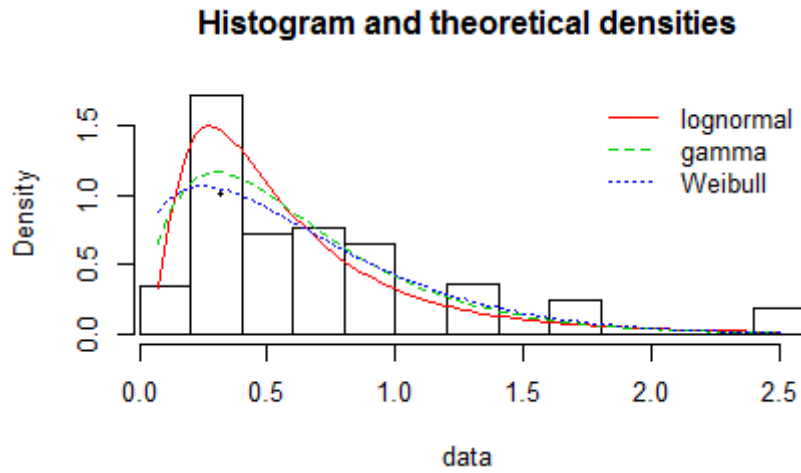
Fonte: Elaborada pelo autor.

Figura 4.19: AMOSTRA 5: Curva de Distribuição WEIBULL



Fonte: Elaborada pelo autor.

Figura 4.20: AMOSTRA 5: Histograma de frequências dos dados observados e das 3 fdps ajustadas



Fonte: Elaborada pelo autor.

Tabela 4.5: AMOSTRA 5: ESTIMATIVAS DOS PARÂMETROS das 3 distribuições

Distribuições	$\alpha$	$\beta$	$\mu$	$\sigma^2$	AIC	BIC
Gamma	$1,910 \pm 0,025$	$3,040 \pm 0,045$	-	-	8656,51	8670,69
Log Normal	-	-	$-0,749 \pm 0,007$	$0,740 \pm 0,005$	7381,32	7395,74
Weibull	$1,334 \pm 0,009$	$0,691 \pm 0,005$	-	-	9367,19	9381,61

Fonte: Elaborada pelo autor.

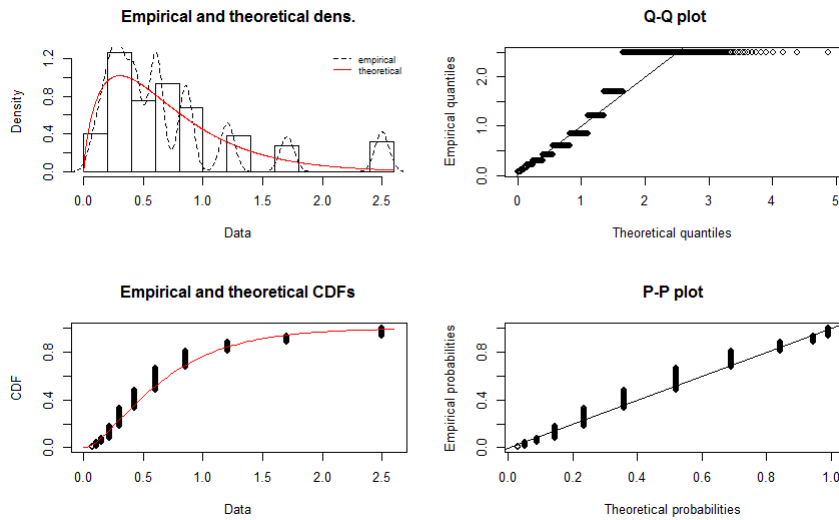
O menor valor de AIC e BIC ocorre para a (fdp)Log Normal, já o maior valor de AIC e BIC foi da (fdp) Weibull. Pelo critério de Akaike, a (fdp) Log Normal se ajusta melhor aos dados.

Nos gráficos Q-Q plot, P-P plot e da Função Distribuição Acumulada Empírica e Teórica (CDFs) verificamos uma sobreposição mais regular das linhas verticais sobre cada gráfico no caso da (fdp)Log Normal.

Pelo Histograma e a Densidade Teórica percebemos que a curva que se ajusta melhor é a (fdp)Log Normal e a (fdp)Weibull possui o pior ajuste.

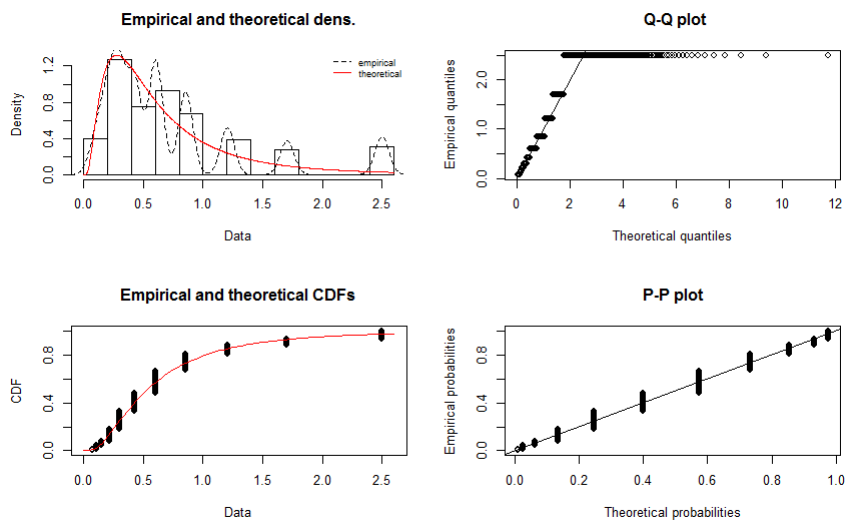
4.6 RESULTADOS OBTIDOS PARA A AMOSTRA 6

Figura 4.21: AMOSTRA 6: Curva de Distribuição GAMA



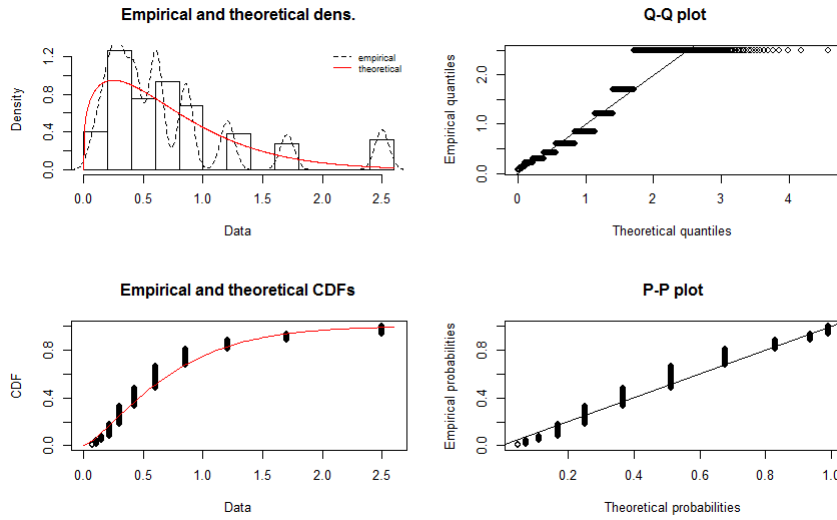
Fonte: Elaborada pelo autor.

Figura 4.22: AMOSTRA 6: Curva de Distribuição LOG NORMAL



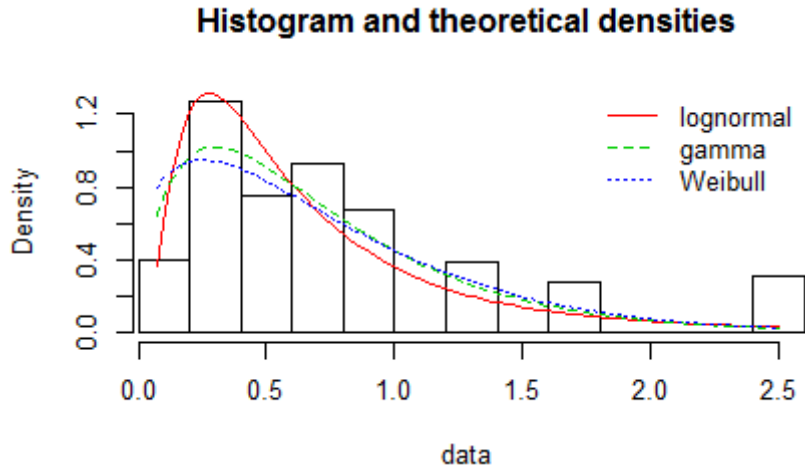
Fonte: Elaborada pelo autor.

Figura 4.23: AMOSTRA 6: Curva de Distribuição WEIBULL



Fonte: Elaborada pelo autor.

Figura 4.24: AMOSTRA 6: Histograma de frequências dos dados observados e das 3 fdps ajustadas



Fonte: Elaborada pelo autor.

Tabela 4.6: AMOSTRA 6: ESTIMATIVAS DOS PARÂMETROS das 3 distribuições

Distribuições	$\alpha$	$\beta$	$\mu$	$\sigma^2$	AIC	BIC
Gamma	$1,738 \pm 0,022$	$2,439 \pm 0,037$	-	-	11664,55	11678,97
Log Normal	-	-	$-0,654 \pm 0,008$	$0,800 \pm 0,005$	10858,89	1083,31
Weibull	$1,295 \pm 0,009$	$0,777 \pm 0,006$	-	-	12154,08	12168,50

Fonte: Elaborada pelo autor.



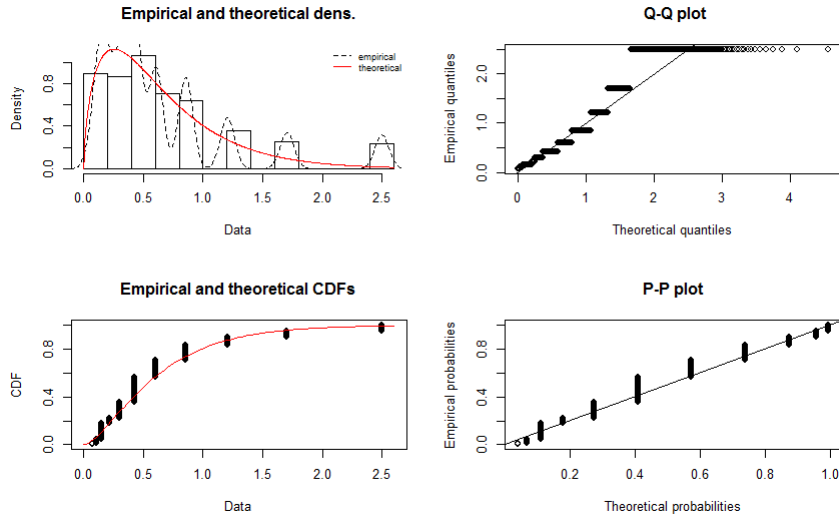
O menor valor de AIC e BIC ocorre para a (fdp)Log Normal, já o maior valor de AIC e BIC foi da (fdp) Weibull. Pelo critério de Akaike, a (fdp) Log Normal se ajusta melhor aos dados.

Nos gráficos Q-Q plot, P-P plot e da Função Distribuição Acumulada Empírica e Teórica (CDFs) verificamos uma sobreposição mais regular das linhas verticais sobre cada gráfico no caso da (fdp)Log Normal.

Pelo Histograma e a Densidade Teórica percebemos que a curva que se ajusta melhor é a (fdp)Log Normal e a (fdp)Weibull possui o pior ajuste.

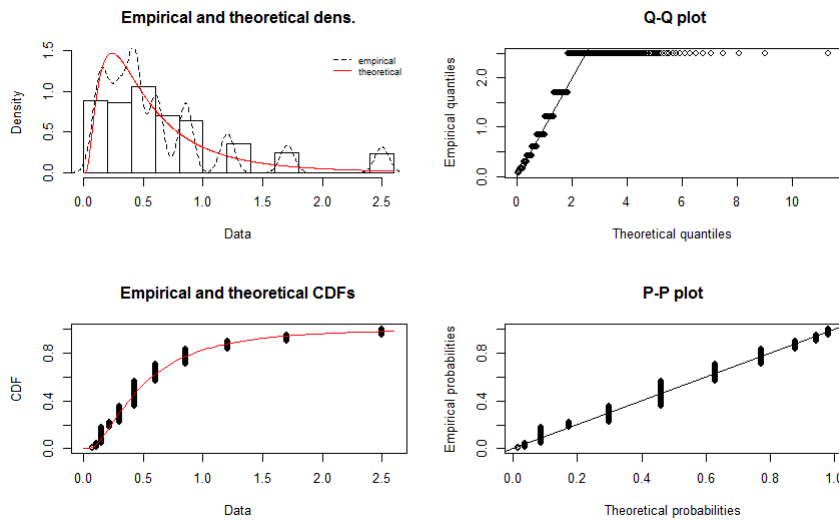
4.7 RESULTADOS OBTIDOS PARA A AMOSTRA 7

Figura 4.25: AMOSTRA 7: Curva de Distribuição GAMA



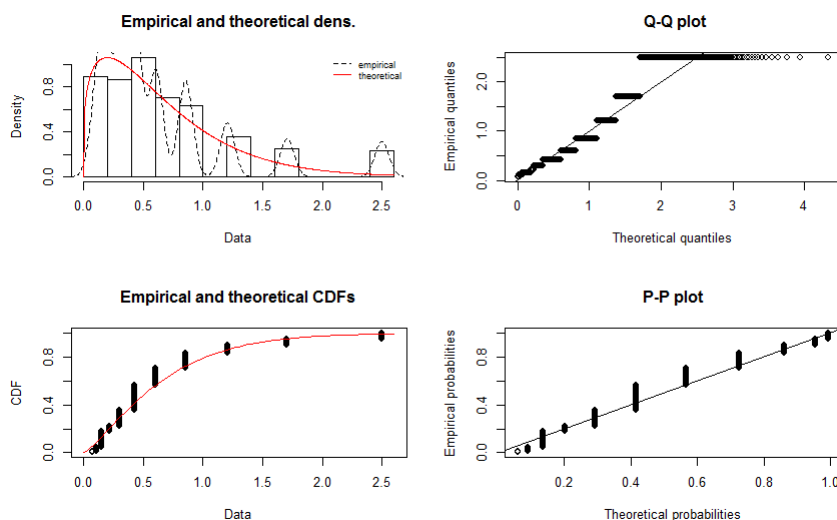
Fonte: Elaborada pelo autor.

Figura 4.26: AMOSTRA 7: Curva de Distribuição LOG NORMAL



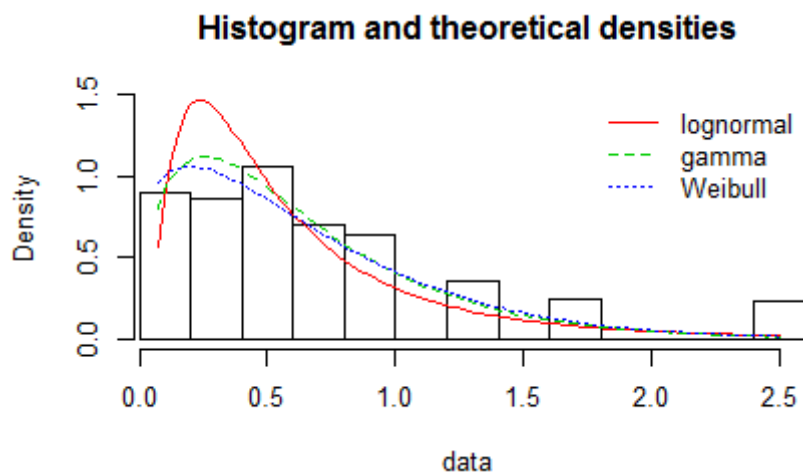
Fonte: Elaborada pelo autor.

Figura 4.27: AMOSTRA 7: Curva de Distribuição WEIBULL



Fonte: Elaborada pelo autor.

Figura 4.28: AMOSTRA 7: Histograma de frequências dos dados observados e das 3 fdps ajustadas



Fonte: Elaborada pelo autor.

Tabela 4.7: AMOSTRA 7: ESTIMATIVAS DOS PARÂMETROS das 3 distribuições

Distribuições	$\alpha$	$\beta$	$\mu$	$\sigma^2$	AIC	BIC
Gamma	$1,651 \pm 0,021$	$2,561 \pm 0,039$	-	-	9914,74	9929,17
Log Normal	-	-	$-0,772 \pm 0,008$	$0,821 \pm 0,006$	9919,58	8266,21
Weibull	$1,257 \pm 0,009$	$0,698 \pm 0,006$	-	-	10371,01	1052,69

Fonte: Elaborada pelo autor.

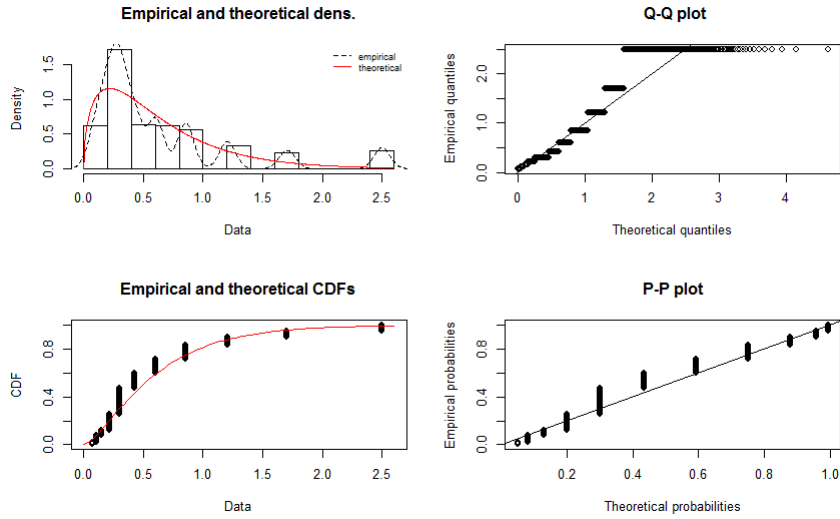
O menor valor de AIC ocorre para a (fdp)Gama e o menor BIC ocorre para a (fdp)Log Normal, já o maior valor de AIC e BIC foi da (fdp) Weibull. Pelo critério de Akaike, a (fdp) Gama se ajusta melhor aos dados.

Nos gráficos Q-Q plot, P-P plot e da Função Distribuição Acumulada Empírica e Teórica (CDFs) verificamos uma sobreposição mais regular das linhas verticais sobre cada gráfico no caso da (fdp)Log Normal.

Pelo Histograma e a Densidade Teórica percebemos que a curva que se ajusta melhor é a (fdp)Gama e a (fdp)Weibull possui o pior ajuste.

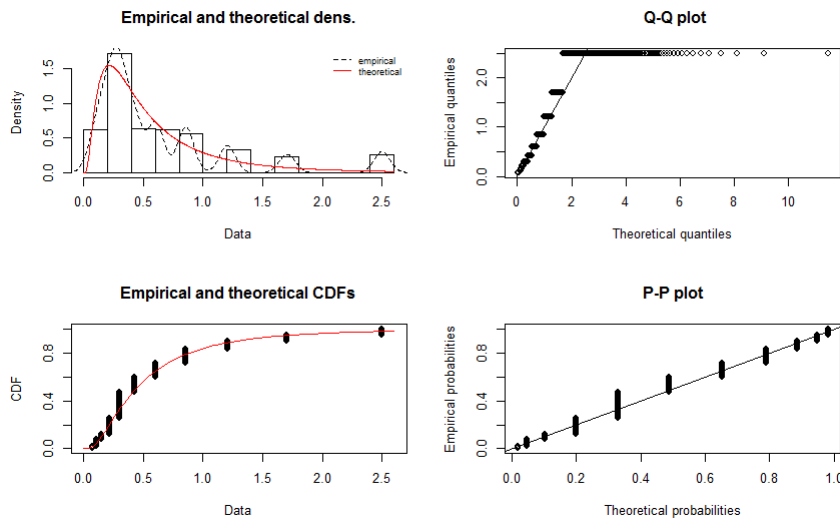
4.8 RESULTADOS OBTIDOS PARA A AMOSTRA 8

Figura 4.29: AMOSTRA 8: Curva de Distribuição GAMA



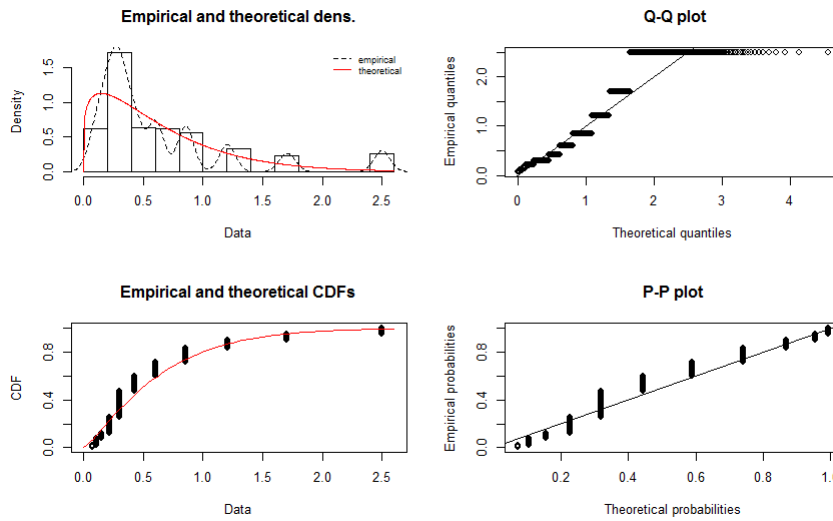
Fonte: Elaborada pelo autor.

Figura 4.30: AMOSTRA 8: Curva de Distribuição LOG NORMAL



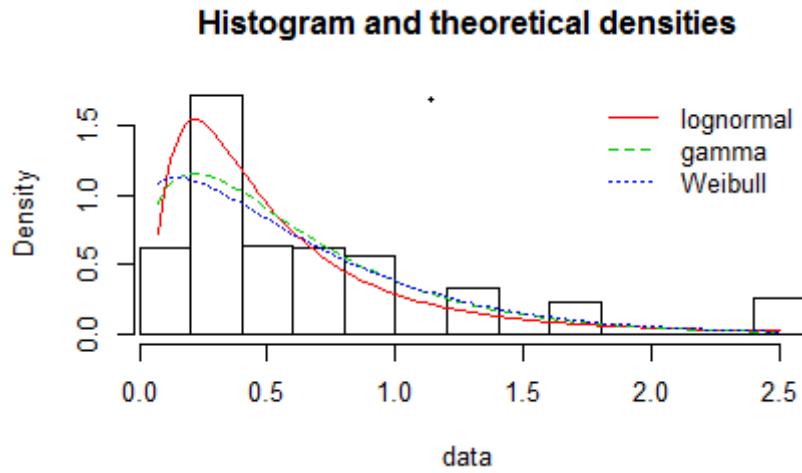
Fonte: Elaborada pelo autor.

Figura 4.31: AMOSTRA 8: Curva de Distribuição WEIBULL



Fonte: Elaborada pelo autor.

Figura 4.32: AMOSTRA 8: Histograma de frequências dos dados observados e das 3 fdps ajustadas



Fonte: Elaborada pelo autor.

Tabela 4.8: AMOSTRA 8: ESTIMATIVAS DOS PARÂMETROS das 3 distribuições

Distribuições	$\alpha$	$\beta$	$\mu$	$\sigma^2$	AIC	BIC
Gamma	1,530 ± 0,020	2,453 ± 0,037	-	-	9595,65	9610,07
Log Normal	-	-	-0,883 ± 0,008	0,841 ± 0,006	8251,79	9202,41
Weibull	1,193 ± 0,009	0,667 ± 0,006	-	-	10038,27	10052,68

Fonte: Elaborada pelo autor.

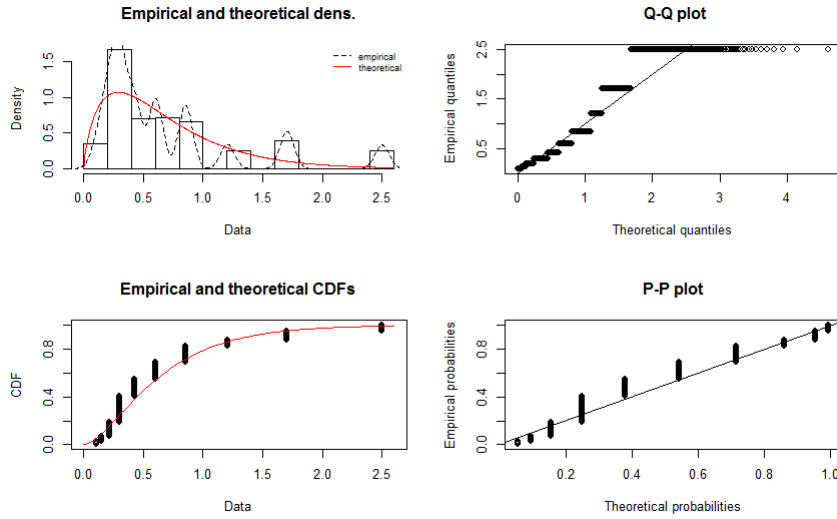
O menor valor de AIC e BIC ocorre para a (fdp)Log Normal, já o maior valor de AIC e BIC foi da (fdp) Weibull. Pelo critério de Akaike, a (fdp) Log Normal se ajusta melhor aos dados.

Nos gráficos Q-Q plot, P-P plot e da Função Distribuição Acumulada Empírica e Teórica (CDFs) verificamos uma sobreposição mais regular das linhas verticais sobre cada gráfico no caso da (fdp)Log Normal.

Pelo Histograma e a Densidade Teórica percebemos que a curva que se ajusta melhor é a (fdp)Log Normal e a (fdp)Weibull possui o pior ajuste.

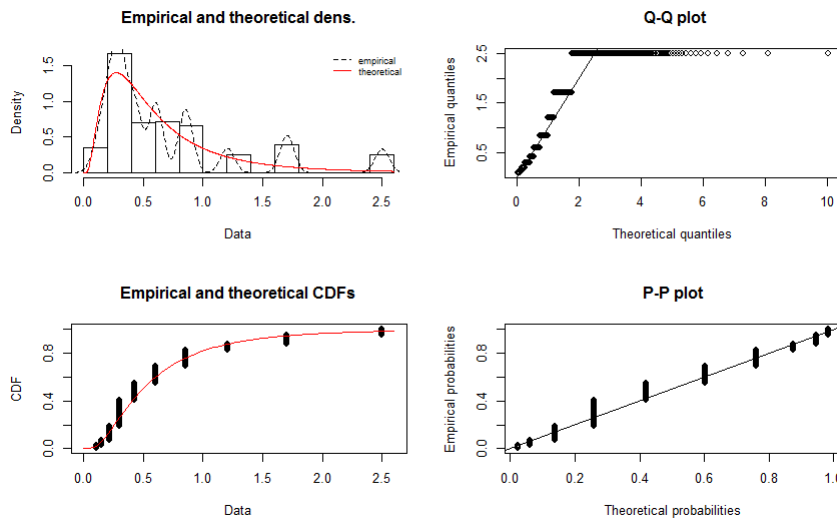
4.9 RESULTADOS OBTIDOS PARA A AMOSTRA 9

Figura 4.33: AMOSTRA 9: Curva de Distribuição GAMA



Fonte: Elaborada pelo autor.

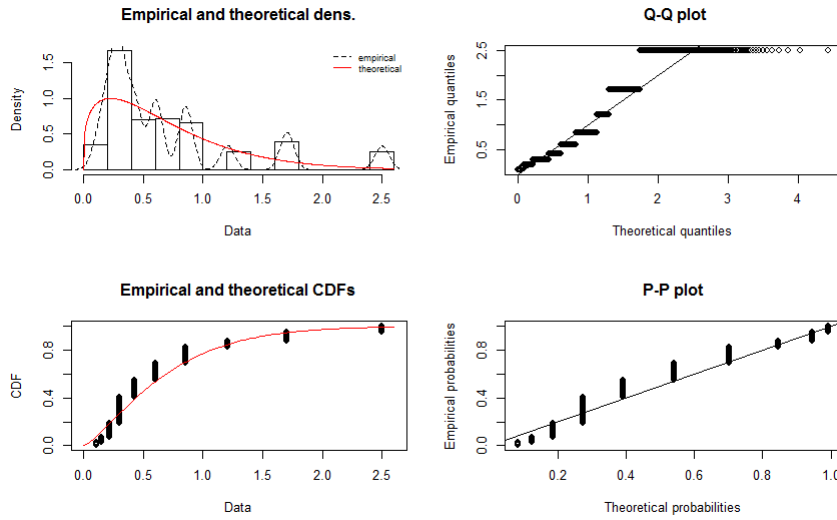
Figura 4.34: AMOSTRA 9: Curva de Distribuição LOG NORMAL



Fonte: Elaborada pelo autor.

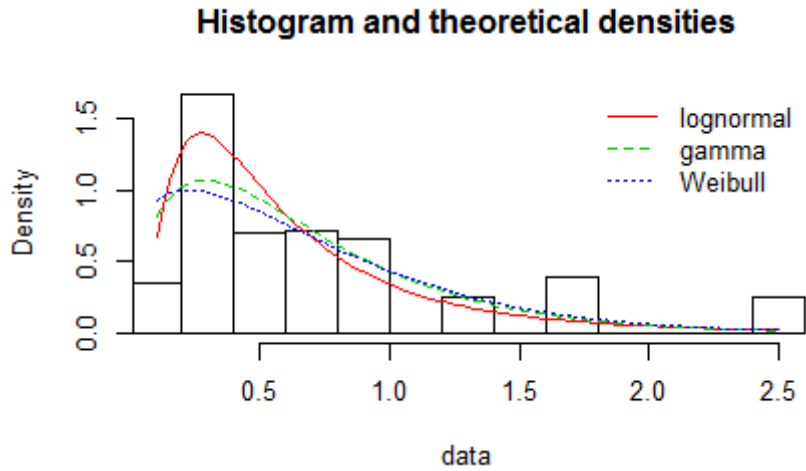


Figura 4.35: AMOSTRA 9: Curva de Distribuição WEIBULL



Fonte: Elaborada pelo autor.

Figura 4.36: AMOSTRA 9: Histograma de frequências dos dados observados e das 3 fdps ajustadas



Fonte: Elaborada pelo autor.

Tabela 4.9: AMOSTRA 9: ESTIMATIVAS DOS PARÂMETROS das 3 distribuições

Distribuições	$\alpha$	$\beta$	$\mu$	$\sigma^2$	AIC	BIC
Gamma	$1,752 \pm 0,023$	$2,588 \pm 0,039$	-	-	10606,68	10621,10
Log Normal	-	-	$-0,702 \pm 0,008$	$0,773 \pm 0,005$	9187,99	9202,41
Weibull	$1,280 \pm 0,009$	$0,738 \pm 0,006$	-	-	11205,08	11219,50

Fonte: Elaborada pelo autor.

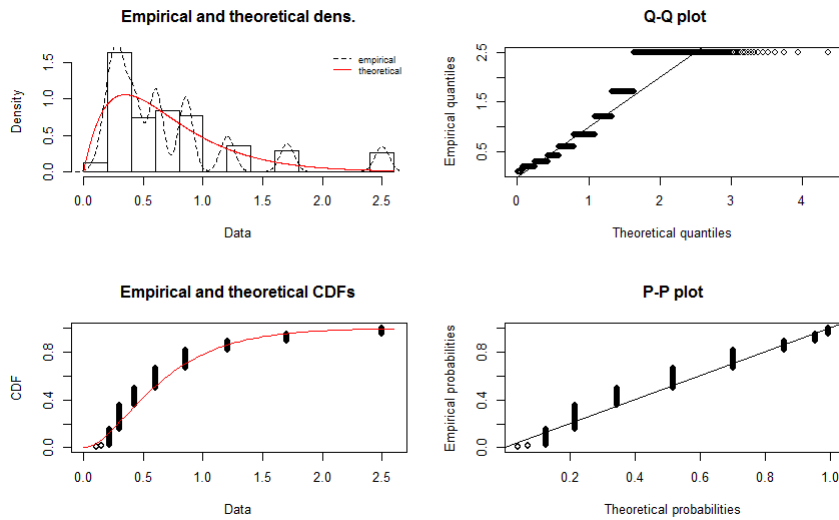
O menor valor de AIC e BIC ocorre para a (fdp)Log Normal, já o maior valor de AIC e BIC foi da (fdp) Weibull. Pelo critério de Akaike, a (fdp) Log Normal se ajusta melhor aos dados.

Nos gráficos Q-Q plot, P-P plot e da Função Distribuição Acumulada Empírica e Teórica (CDFs) verificamos uma sobreposição mais regular das linhas verticais sobre cada gráfico no caso da (fdp)Log Normal.

Pelo Histograma e a Densidade Teórica percebemos que a curva que se ajusta melhor é a (fdp)Log Normal e a (fdp)Weibull possui o pior ajuste.

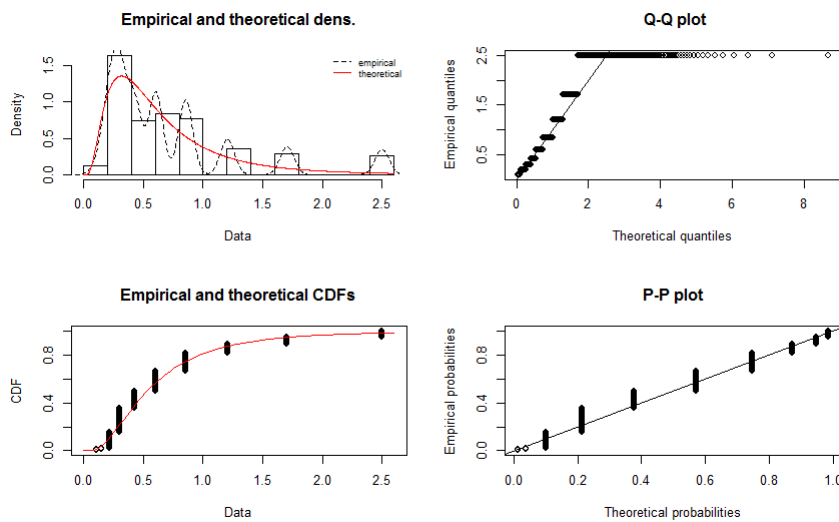
4.10 RESULTADOS OBTIDOS PARA A AMOSTRA 10

Figura 4.37: AMOSTRA 10: Curva de Distribuição GAMA



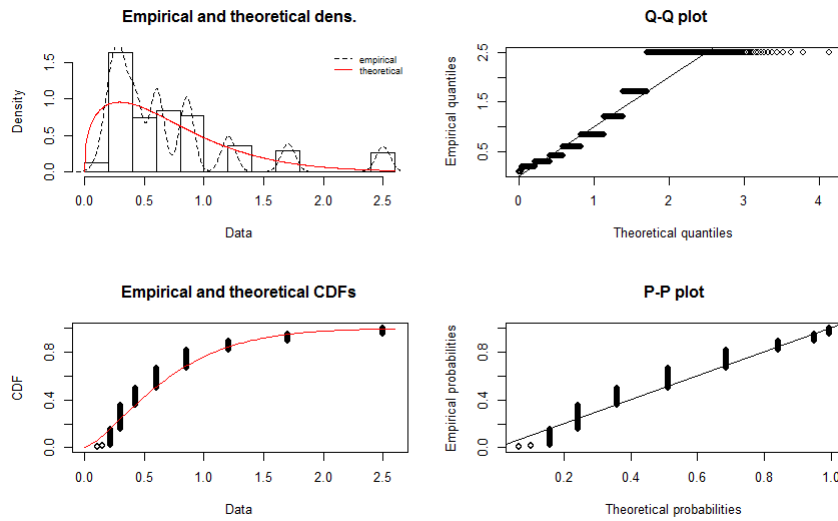
Fonte: Elaborada pelo autor.

Figura 4.38: AMOSTRA 10: Curva de Distribuição LOG NORMAL



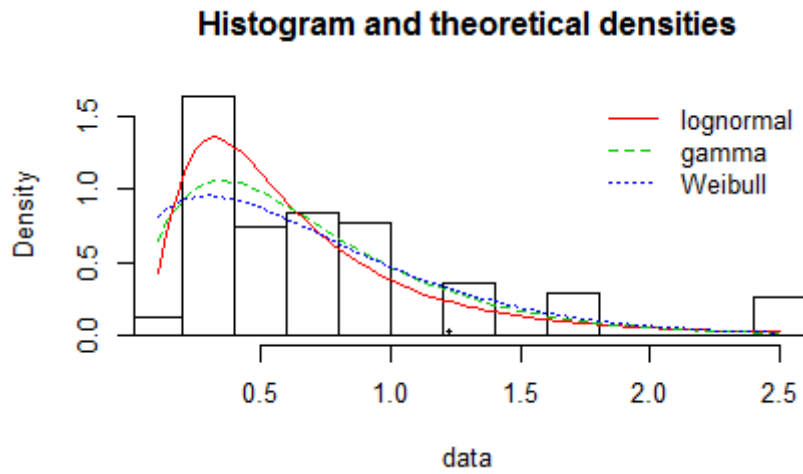
Fonte: Elaborada pelo autor.

Figura 4.39: AMOSTRA 10: Curva de Distribuição WEIBULL



Fonte: Elaborada pelo autor.

Figura 4.40: AMOSTRA 10: Histograma de frequências dos dados observados e das 3 fdps ajustadas



Fonte: Elaborada pelo autor.

Tabela 4.10: AMOSTRA 10: ESTIMATIVAS DOS PARÂMETROS das 3 distribuições

Distribuições	$\alpha$	$\beta$	$\mu$	$\sigma^2$	AIC	BIC
Gamma	$1,997 \pm 0,026$	$2,867 \pm 0,043$	-	-	10460,86	10475,28
Log Normal	-	-	$-0,632 \pm 0,008$	$0,717 \pm 0,005$	9086,72	9101,14
Weibull	$1,336 \pm 0,099$	$0,769 \pm 0,006$	-	-	11221,92	11236,34

Fonte: Elaborada pelo autor.

O menor valor de AIC e BIC ocorre para a (fdp)Log Normal, já o maior valor de AIC e BIC foi da (fdp) Weibull. Pelo critério de Akaike, a (fdp) Log Normal se ajusta melhor aos dados.

Nos gráficos Q-Q plot, P-P plot e da Função Distribuição Acumulada Empírica e Teórica (CDFs) verificamos uma sobreposição mais regular das linhas verticais sobre cada gráfico no caso da (fdp)Log Normal.

Pelo Histograma e a Densidade Teórica percebemos que a curva que se ajusta melhor é a (fdp)Log Normal e a (fdp)Weibull possui o pior ajuste.

Uma avaliação das 10 amostras, observamos que o ajuste com menor valor de AIC e BIC para os dados experimentais foi obtido para a (fdp) Log Normal, exceto na amostra 7, em que por uma diferença muito pequena, a (fdp) Gamma obteve menor AIC.

Também podemos verificar a qualidade do ajuste das distribuições pelos gráficos, e nos dá a (fdp) Log Normal (exceto a amostra 7), identificamos uma melhor aproximação dos valores teóricos com os empíricos por meio do histograma juntamente com a densidade teórica.

Nos gráficos Q-Q plot, P-P plot e da função distribuição acumulada empírica e teórica (CDFs) é possível a verificação de uma sobreposição mais regular das linhas verticais sobre cada gráfico.

Portanto das 10 amostras, a (fdp) Log Normal é a que melhor se ajusta aos dados em 9 amostras.

---

## CONSIDERAÇÕES FINAIS

---

Após a análise de dados feita nas 10 amostras deste trabalho, percebemos que o modelo que melhor se ajusta aos dados dos tamanhos das partículas em massa é o modelo da (fdp) Log Normal, conforme os resultados obtidos pelo Critério de Informação AIC e BIC e análise gráfica.

Pelo Critério de Informação de Akaike (AIC) e no Bayesiano (BIC), após o cálculo nas 10 amostras, o valor de AIC ficou menor para a (fdp) Log Normal em 9 amostras e o mesmo ocorrendo nas 10 amostras com o valor do BIC. O modelo de pior ajuste foi a (fdp) de Weibull, por apresentar maior valor de AIC e BIC em todas as amostras.

Os gráficos Densidade Empírica e Teórica, Q-Q plot, Função Distribuição Acumulada Empírica e Teórica e os P-P plot mostraram que a curva empírica que fica mais próxima da teórica foi a (fdp) Log Normal nas 10 amostras estudadas.

Dessa forma, sugerimos que o melhor ajuste para as 10 amostras fica determinado pela função de densidade de probabilidade Log Normal. Assim poderemos fazer previsões para o número de partículas em massa e para valores de diâmetro das partículas obtidas na escavação de cada sítio.

O método da Máxima Verossimilhança foi muito importante para a determinação dos parâmetros das distribuições e nos cálculos de AIC e BIC. O cálculo dos parâmetros só pode ser obtido numericamente pois são raízes de funções não triviais e assim havendo necessidade do uso de software para a resolução, no caso o R Studio.

O trabalho nos mostra, por meio de um exemplo prático, a aplicação de modelos estatísticos a problemas de situações reais. Isto é muito importante para os alunos entenderem e compreenderem, na prática, algumas aplicações de Probabilidade e Estatística.

Em geral, nos cursos de graduação, as funções de densidade de probabilidades: Gama, Log Normal e Weibull são poucas vezes trabalhadas em aulas, deixando mais para cursos de pós-graduação e pesquisas em nível de mestrado e doutorado . O objetivo foi apresentá-las e aplicá-las em um exemplo real para que possa ser compartilhada em sala de aula ou mesmo em futuras pesquisas.

Pelas colocações acima, acreditamos que o trabalho possa contribuir com alunos, professores e pesquisadores mostrando um caso real de aplicabilidade das funções de densidade de probabilidade (fdp) da Inferência Estatística, pelo desenvolvimento teórico e as técnicas utilizadas e testadas neste trabalho.



# A

---

## APÊNDICE A

---

### A.1 TABELAS DE DADOS

Nas tabelas são encontradas: a malha da peneira (em milímetros), peso (em gramas), probabilidade em massa, diâmetro médio (em milímetros) e o número de partículas em massa.

Todas as tabelas contidas neste anexo foram elaboradas no Microsoft Excel.

Tabela A.1: Amostra 1: G6N4F3

Malha da peneira(mm)	Peso(g)	Probabilidade(em massa)	Diâmetro(mm)	Nº partículas
2,000	3,1743	0,0361	2,500	361
1,410	3,9608	0,0451	1,705	451
1,000	5,8840	0,0670	1,205	670
0,710	11,4827	0,1307	0,855	1307
0,500	13,4827	0,1535	0,605	1535
0,350	13,7357	0,1564	0,425	1564
0,250	13,6068	0,1549	0,300	1549
0,177	13,3120	0,1516	0,214	1516
0,125	4,7280	0,0538	0,151	538
0,088	3,2336	0,0368	0,107	368
0,063	1,2310	0,0140	0,076	140
Peso Total	87,8316	1,000		

Fonte: Elaborada pelo autor.

Tabela A.2: Amostra 2:G6N4F5

Malha da peneira(mm)	Peso(g)	Probabilidade(em massa)	Diâmetro(mm)	Nº partículas
2,000	5,3342	0,0577	2,500	577
1,410	4,4119	0,0477	1,705	477
1,000	5,9526	0,0644	1,205	644
0,710	9,4391	0,1021	0,855	1021
0,500	20,2121	0,2187	0,605	2187
0,350	9,6164	0,1041	0,425	1041
0,250	17,7727	0,1923	0,300	1923
0,177	11,9446	0,1293	0,214	1293
0,125	3,4484	0,0373	0,151	373
0,088	3,5170	0,0381	0,107	381
0,063	0,7613	0,0082	0,076	82
Peso Total	92,4103	1,000		

Fonte: Elaborada pelo autor.

Tabela A.3: Amostra 3:G6N4F4

Malha da peneira(mm)	Peso(g)	Probabilidade(em massa)	Diâmetro(mm)	Nº partículas
2,000	3,7232	0,0431	2,500	431
1,410	3,9579	0,0459	1,705	459
1,000	5,1440	0,0596	1,205	596
0,710	10,3688	0,1202	0,855	1202
0,500	12,4969	0,1448	0,605	1448
0,350	11,6134	0,1346	0,425	1346
0,250	16,5684	0,1920	0,300	1920
0,177	10,8369	0,1256	0,214	1256
0,125	5,2125	0,0604	0,151	604
0,088	3,8751	0,0449	0,107	449
0,063	2,4894	0,0289	0,076	289
Peso Total	86,2865	1,000		

Fonte: Elaborada pelo autor.

Tabela A.4: Amostra 4: G6N4F2

Malha da peneira(mm)	Peso(g)	Probabilidade(em massa)	Diâmetro(mm)	Nº partículas
2,000	5,1784	0,0616	2,500	616
1,410	3,8709	0,0460	1,705	460
1,000	5,1670	0,0614	1,205	614
0,710	10,7812	0,1282	0,855	1282
0,500	11,4173	0,1357	0,605	1357
0,350	16,5693	0,1970	0,425	1970
0,250	17,1986	0,2045	0,300	2045
0,177	8,9492	0,1064	0,214	1064
0,125	2,1737	0,0258	0,151	258
0,088	2,1975	0,0261	0,107	261
0,063	0,6127	0,0073	0,076	73
Peso Total	84,1158	1,000		

Fonte: Elaborada pelo autor.

Tabela A.5: Amostra 5: G6N3F2

Malha da peneira(mm)	Peso(g)	Probabilidade(em massa)	Diâmetro(mm)	Nº partículas
2,000	2,9989	0,0365	2,500	365
1,410	3,9721	0,0483	1,705	483
1,000	5,9721	0,0726	1,205	726
0,710	10,6556	0,1296	0,855	1296
0,500	12,5746	0,1529	0,605	1529
0,350	11,9566	0,1454	0,425	1454
0,250	18,5357	0,2254	0,300	2254
0,177	9,7877	0,1190	0,214	1190
0,125	3,8820	0,0472	0,151	472
0,088	1,0802	0,0131	0,107	131
0,063	0,8342	0,0101	0,076	101
Peso Total	82,2497	1,000		

Fonte: Elaborada pelo autor.

Tabela A.6: Amostra 6: G6N3F3

Malha da peneira(mm)	Peso(g)	Probabilidade(em massa)	Diâmetro(mm)	Nº partículas
2,000	6,0789	0,0626	2,500	626
1,410	5,3545	0,0551	1,705	551
1,000	7,4486	0,0766	1,205	766
0,710	13,1959	0,1358	0,855	1358
0,500	18,1297	0,1866	0,605	1866
0,350	14,5960	0,1502	0,425	1502
0,250	14,7442	0,1517	0,300	1517
0,177	9,8872	0,1017	0,214	1017
0,125	3,4083	0,0351	0,151	351
0,088	2,8620	0,0294	0,107	294
0,063	1,4770	0,0152	0,076	152
Peso Total	97,1823	1,000		

Fonte: Elaborada pelo autor.

Tabela A.7: Amostra 7: G6N1S/F

Malha da peneira(mm)	Peso(g)	Probabilidade(em massa)	Diâmetro(mm)	Nº partículas
2,000	3,7514	0,0467	2,500	467
1,410	4,0166	0,0500	1,705	500
1,000	5,7286	0,0712	1,205	712
0,710	10,2362	0,1273	0,855	1273
0,500	11,2761	0,1402	0,605	1402
0,350	17,1223	0,2130	0,425	2130
0,250	10,5290	0,1310	0,300	1310
0,177	3,3771	0,0420	0,214	420
0,125	10,4996	0,1306	0,151	1306
0,088	2,8124	0,0350	0,107	350
0,063	1,0533	0,0131	0,076	131
Peso Total		1,000		

Fonte: Elaborada pelo autor.

Tabela A.8: Amostra 8: G6N2S/F

Malha da peneira(mm)	Peso(g)	Probabilidade(em massa)	Diâmetro(mm)	Nº partículas
2,000	5,1815	0,0525	2,500	525
1,410	4,5174	0,0458	1,705	458
1,000	6,6531	0,0674	1,205	674
0,710	11,2622	0,1141	0,855	1141
0,500	12,1651	0,1233	0,605	1233
0,350	12,4460	0,1261	0,425	1261
0,250	21,5461	0,2184	0,300	2184
0,177	12,5912	0,1276	0,214	1276
0,125	4,8062	0,0487	0,151	487
0,088	5,1911	0,0526	0,107	526
0,063	2,3037	0,0233	0,076	233
Peso Total	98,6636	1,000		

Fonte: Elaborada pelo autor.

Tabela A.9: Amostra 9: G6N4F6

Malha da peneira(mm)	Peso(g)	Probabilidade(em massa)	Diâmetro(mm)	Nº partículas
2,000	4,6879	0,0496	2,500	496
1,410	7,3468	0,9777	1,705	777
1,000	4,7764	0,0505	1,205	505
0,710	12,4474	0,1316	0,855	1316
0,500	13,6426	0,1443	0,605	1443
0,350	13,3578	0,1413	0,425	1413
0,250	20,6012	0,2179	0,300	2179
0,177	11,1303	0,1177	0,214	1177
0,125	4,1890	0,9443	0,151	443
0,088	2,3788	0,0252	0,107	252
0,063	0,0000	0,0000	0,076	0
Peso Total		1,000		

Fonte: Elaborada pelo autor.

Tabela A.10: Amostra 10: G6N5F2

Malha da peneira(mm)	Peso(g)	Probabilidade(em massa)	Diâmetro(mm)	Nº partículas
2,000	4,1489	0,0512	2,500	512
1,410	4,6004	0,0567	1,705	567
1,000	5,8718	0,0724	1,205	724
0,710	12,3700	0,1525	0,855	1525
0,500	13,5670	0,1673	0,605	1673
0,350	12,0156	0,1482	0,425	1482
0,250	15,7736	0,1945	0,300	1945
0,177	10,8321	0,1336	0,214	1336
0,125	1,1598	0,0143	0,151	143
0,088	0,7546	0,0093	0,107	93
0,063	0,0000	0,0000	0,076	0
Peso Total		1,000		

Fonte: Elaborada pelo autor.

---

## BIBLIOGRAFIA

---

- [1] ACTION, P.: *Ambiente Virtual de Aprendizagem*. <http://www.portalaction.com.br/probabilidades/69-distribuicao-gama>. [Acesso em: 27-Abril-2018].
- [2] ACTION, P.: *Ambiente Virtual de Aprendizagem*. <http://www.portalaction.com.br/probabilidades/615-distribuicao-log-normal>. [Acesso em: 27-Abril-2018].
- [3] ACTION, P.: *Ambiente Virtual de Aprendizagem*. <http://www.portalaction.com.br/probabilidades/613-distribuicao-weibull>. [Acesso em: 27-Abril-2018].
- [4] ARAÚJO, Eliakim Martins; SILVA, I. N. O. J. B. d. J. E. G. C. A. B. M.: *Aplicação de seis distribuições de probabilidade a séries de temperatura máxima em Iguatu - CE*. *Revista Ciência Agronômica*, 41(1):36–45, 2010.
- [5] ASSIS, Francisco Neto de; ARRUDA, H. V. d. P. A. R. d.: *Aplicações de estatística a climatologia*. UFPEL, Ed. Universitária, 1996.
- [6] BOLFARINE, Heleno; SANDOVAL, M. C.: *Introducao a Inferência*. SBM, Rio de janeiro, 2ª ed., 2010.
- [7] DELIGNETTE-MULLER, Marie Laure; DUTANG, C. o.: *Fitdistrplus: An R package for fitting distributions*. *Journal of Statistical Software*, 64(4):1–34, 2015.
- [8] EMILIANO, Paulo C; VEIGA, E. P. V. M. J. M. F. S.: *Critérios de informação de Akaike versus Bayesiano: análise comparativa*. 19º Simpósio Nacional de Probabilidade e Estatística, 2010.
- [9] FERNANDES, José António; BATANERO, C. C. J. D. C.: *A simulação em Probabilidades e Estatística: potencialidades e limitações*. *Quadrante*, XVIII, 1:161–183, 2009.
- [10] FERREIRA, A. B. d. H.: *Novo dicionário Aurélio da língua portuguesa*. 2004.
- [11] LOPES, C. E.: *Educação estatística no curso de licenciatura em matemática*. *Boletim de Educação Matemática*, 27(47), 2013.
- [12] LÜDKE, M; MEDA, A.: *Pesquisa em educação: abordagens qualitativas*. São Paulo:

- EPU. Mathematics Teaching in the Middle School*, 12(5):244–250, 1986.
- [13] MORETTIN, Pedro Albert; BUSSAB, W.d.O.: *Estatística Básica*. Saraiva, São Paulo, 6ª ed., 2010.
- [14] RODRIGUES, Jailson Araújo; FILHO, J. d. S. C. L. M.: *Probability density functions for the estimation of monthly rainfall*. Seminário: Ciências Exatas e Tecnológicas, 34(1):3–8, 2013.
- [15] SOBRAL, Thales Lima; BARRETO, G.: *Utilização dos critérios de informação na seleção de modelos de regressão linear*. Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, 4(1), 2016.
- [16] TEAM, R. C. o.: *R: A language and environment for statistical computing*. 2013.
- [17] VENABLE, W N; RIPLEY, B.: *Modern Applied Statistics with S*. Springer-Verlag. 4ª ed., 2010.
- [18] WEBER, S. H.: *Desenvolvimento de nova função densidade de probabilidade para avaliação de regeneração natural*. Universidade Federal do Paraná, Curitiba, p. 87, 2006.



---

## ÍNDICE

---

- aderência, 4
- análise dos dados, 76
  
- comando curve, 24
- comando denscomp, 43
- comando fitdist, 43
- comando plot, 42
- critério AIC, 22, 44
- critério BIC, 22, 44
- critérios de melhor ajuste, 21
  
- dados, 28
- distribuição gama, 7
  
- estimação de parâmetros dos modelos, 4
- estimativa do número de partículas, 41
- estratigrafia, 28
  
- função contínua de probabilidade, 4
- função de máxima verossimilhança, 4, 22
- função densidade de probabilidade, 5, 6, 8, 10
- função gama modificada, 6
- funções densidade de probabilidade, 4, 5
  
- Gráfico da Densidade Empírica e Teórica, 24, 42
- Gráfico Função Distribuição Acumulada Empírica (CDF)s, 26, 42
- Gráfico P-P plot, 26, 42
- Gráfico Q-Q plot, 25, 42
  
- inferência estatística, 77
  
- método da máxima verossimilhança, 12, 24, 76
- malha das peneiras, 39
- massa de partículas, 41
- micromorfologia, 28
- modelagem matemática, 21
- Modelo Gama
  - distribuição gama, 5, 7
  - propriedade recursiva, 5
- Modelo Log Normal
  - distribuição log normal, 8, 9
- Modelo Weibull
  - distribuição Weibull, 10, 11
- modelos estatísticos, 76
  
- pacote fitdistrplus, 24
- parâmetros da (fdp) gama, 13
- parâmetros da (fdp) log normal, 15
- parâmetros da (fdp) Weibull, 18
- parâmetros das 3 distribuições, 41
- probabilidade em massa, 40
  
- R Studio, 41
- resultados dos ajustes, 44
  
- software R, 4, 21, 23
- software R Studio, 21, 23, 41
- softwares estatísticos, 21
  
- testes de aderência, 21
  
- variável aleatória contínua, 4