UNIVERSIDADE ESTADUAL DA PARAÍBA
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE MATEMÁTICA
CURSO DE ESPECIALIZAÇÃO EM MATEMÁTICA
PURA E APLICADA

Dois modelos de distribuição de probabilidade para tratar dados de tempo de vida: O Exponencial e o de Weibull

Fábio Azevedo de Souza

Campina Grande - PB

Julho de 2011

Fábio Azevedo de Souza

Dois modelos de distribuição de probabilidade para tratar dados de tempo de vida: O Exponencial e o de Weibull

Monografia apresentada ao Curso de Especialização em Matemática Pura e Aplicada do Departamento de Matemática do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de Especialista em Matemática Pura e Aplicada.

Profa. Dra. Divanilda Maia Esteves

Campina Grande-PB

Julho de 2011

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL-UEPB

S729d Souza, Fábio Azevedo de.

Dois modelos probabilísticos para tratar dados de tempo de vida [manuscrito]: O exponencial e o de Weibull / Fábio Azevedo de Souza. - 2011.

38 f.

Monografia (Especialização em Matemática Pura e Aplicada) - Universidade Estadual da Paraíba, Centro de Ciências Tecnologias, 2011.

“Orientação: Profa. Dra. Divanilda Maia Esteves, Departamento de Estatística”.

1. Probabilidade. 2. Distribuição de Weibull. 3. Teste de Aderência. I. Título.

22. ed. CDD 519.2

Fábio Azevedo de Souza

Dois modelos de distribuição de probabilidade para tratar dados de tempo de vida: O Exponencial e o de Weibull

Aprovado em: 28 / 07 / 2011

COMISSÃO EXAMINADORA

DMEsteves

Profa. Dra. Divanilda Maia Esteves
Departamento de Estatística - CCT/UEPB
Orientadora

Prof. Dr. Gustavo Henrique Esteves
Departamento de Estatística - CCT/UEPB
Examinador

Prof. Dr. Dr. João Gil de Luna
Departamento de Estatística - CCT/UEPB
Examinador

Dedicatória

Dedico este trabalho aos meus pais (in memorian), a minha esposa Fátima, as minhas filhas Tamiris e Priscila, aos meus irmãos Rildo, Adiêr e Paizinha e a todos que contribuíram de alguma forma para esta conquista.

Agradecimentos

Em primeiro lugar, Deus, que em sua infinita bondade me fez existir, agradeço-te por ter me dado a capacidade para alcançar mais uma meta, realizar mais um sonho e sempre estar ao meu lado dando-me força, coragem e companhia nos momentos mais difíceis;

À minha esposa Fátima por seu amor, incentivo e compreensão, filhas, irmãos, por compartilharem sempre todos os momentos da minha vida, emfim agradeço a toda minha família;

Ao meu amigo Aldo Trajano, que foi de fundamental importância me dando incentivo e oportunidade de alcançar um novo caminho na vida profissional e, que esteve sempre do meu lado apoiando nas horas que mais precisei.

À professora Divanilda Maia, por ter aceito o convite para ser orientadora neste trabalho, pela paciência e dedicação.

Aos professores João Gil de Luna e Gustavo Henriques, por gentilmente aceitarem ao convite de avaliar este trabalho;

A todos os professores da Especialização, Aldo louredo, Aldo Maciel, Osmundo, Abigail, Ernesto e Wandenberg;

Aos meus colegas Luciano, Elisângela, Edna, André,Edleuza pela amizade e companherismo de todos durante o período do curso.

Resumo

Neste trabalho, serão apresentados dois modelos de distribuição de probabilidade, o Exponencial e o de Weibull, tendo como objetivo principal apresentar suas características principais, tais como: função de densidade de probabilidade, função de distribuição acumulada, esperança, variância, função geratriz de momentos, estimação de parâmetros. Para complementação dos estudos, fizemos uma apliacação utilizando dados de precipitação pluviométrica, para podermos comparar qual delas ajusta melhor tais dados.

Palavras chave: Distribuição de Weibull, Probabilidade, Teste de Aderência.

Abstract

In this work, we will present two probability models often used to model time data: Exponential and Weibull distributions. We will show their main characteristics, such as: probability density function, mean and variance, moment generating function, estimation of parameters. We also present an aplication utilizing rainfall amounts data.

Key word: Weibull Distribution , Exponential distribution, Aderence Test.

Sumário

Capítulo 1

Introdução

Quando a Estatística é empregada para estudar fenômenos observáveis, faz-se necessário a construção de um modelo matemático para explicar tal fenômeno. Existem dois tipos de modelos matemáticos: os determinísticos e os probabilísticos. Esse segundo tipo é usado para aqueles experimentos que são aleatórios, ou seja, aqueles cujo resultado não se conhece antes que o experimento seja realizado e além disso, quando reproduzido em condições idênticas pode gerar um resultado diferente daquele observado inicialmente. O uso da probabilidade tem se difundido desde as suas primeiras aplicações nos jogos de azar, até as amplas utilização em áreas de alta complexidade das ciências. Os modelos probabilísticos são úteis para descrever o comportamento de dados coletados em experimentos. Estes dados são modelados através de funções de densidade de probabildade, levando-se em consideração sua natureza. Distribuições de probabilidade podem ser usadas por exemplo, para descrever o comportamento de determinadas variáveis aleatórias num dado intervalo de tempo, em particular nos casos em que o interesse é o tempo de vida de um componente, precipitação pluviométrica de uma região, taxas de falha. As variáveis podem ser classificados em dois tipos para os casos acima citados: aqueles em que a taxa de fala é constante e o que as taxas de falhas são aleatórias. Ambos os casos podem ser modelados através de distribuições de probalidades contínuas, visto que a variável pode assumir valores num intervalo de número reais. O uso da distribuição Exponencial é utilizada nos casos em que a taxa de falha é constante ao longo de um intervalo, sendo esta um caso particular da distribução de Weibull quando $\alpha = 1$. Quando a taxa de falhas é aleatória ao longo do tempo, faz-se o uso da distribuição de Weibull, sendo esta bastante versátil, pelo motivo de se ajustar a algumas outras distribuições de acordo com os valores de seus parâmetros. Segundo Werner (1996), nos tempos atuais, face as mudanças que vem ocorrendo e a complexidade apresentada pelos equipamentos e produtos, passou a ser vital a produção de equipamentos e instrumentos altamente confiáveis. Como exemplo, podemos citar os equipamentos cirúrgicos utilizados em hospitais, as transmissões feitas via satélite, e a exploração do espaço em sondas ou ônibus espaciais. A confiabilidade é de fundamental importância e vem nos auxiliar para fazer previsões sobre quando e que equipamentos, peças e instrumentos irão falhar. Sendo assim, é possível uma substituição prévia de peças ou equipamentos para que vidas sejam poupadas em acidentes aéreos, para que missões de pesquisa espaciais não sejam abortadas e nem transmissões de lazer e cultura sejam bruscamente interrompidas (WERNER, 1996). Segundo Werner (1996), devido ao aumento da concorrência e às alterações no mercado consumidor nas últimas décadas, as empresas necessitam gerar esforços cada vez maiores para sobreviverem. A obtenção de prazos e preços competitivos, a flexibilidade produtiva ou ainda o aumento na qualidade dos produtos, são alguns dos modos de sobrevivência diante dos competidores. Para pode garantir a confiabilidade de um produto, é necessário que a empresa possua um programa de confiabilidade, este programa inclui os procedimentos a serem utilizados na fase de projeto, na fase de manufatura e no pós-venda. Entretanto um programa de confiabilidade pode ser implementado somente após o entendimento do significado de confiabilidade, e somente compreendendo o que é confiabilidade é que poderemos atingi-la. Segundo Halpern (1978) a confiabilidade está embasada em quatro elementos principais: a probabilidade, que busca mensurar a confiabilidade, através da distribuição das falhas; o desempenho, que é o conjunto de requisitos de uso que definem uma função a ser executada, de preferência sem falha; o tempo de operação que está vinculado a operar, sem falhas, num período previamente definido; as condições de operação que são as circunstâncias ambientais e operacionais a qual o produto é submetido.

Na área de Engenharia, em particular na engenharia agrícola, muitas vezes o interesse do pesquisados é determinar previamente o comportamento da variação dos elementos meteorológicos ao longo do ano, de forma à possibilitar um estudo para planejar uma melhora nas diversas atividades do ramo. Por considerar a precipitação pluvial um dos elementos meteorológico de suma importância, é feito coleta de dados mensais e baseado em análise desses dados são tomadas algumas decisões. Sendo o regime pluviométrico de grande influência nos mais diversos setores que abrange o homem, tais como a economia, o meio ambiente e a sociedade. Na agricultura, o conhecimento antecipado das condições locais de solo, radiação solar e precipitação pluvial, e sua variação ao longo de um ciclo de cultivo, são significativos para a obtenção de rendimentos satisfatórios, visto que esses fatores são determinantes para o sucesso nos cultivos. Ribeiro & Lunardi (1997) salientam a importância da caracterização da precipitação em um local para o planejamento de atividades agrícolas, sendo imprescindível também no dimensionamento de reservatórios de água, na elaboração de projetos de proteção e conservação de solos e em atividades de lazer e esportivas. Segundo Souza (2010) O estudo climatológico das diversas variáveis do tempo é de extrema importância, tendo em vista o impacto ambiental que a anomalia dessas componentes provoca no clima regional. A precipitação pluvial é um dos elementos meteorológicos que exerce mais influência sobre condições ambientais. Além do efeito direto sobre o balanço hídrico, exerce influência indiretamente sobre outras variáveis, tais como: temperatura do ar e do solo, a umidade relativa do ar e a radiação solar que, no conjunto atuam como fatores básicos para o crescimento e desenvolvimento das plantas. Em estudos para caracterizar a variabilidade da precipitação pluvial, faz-se uma análise da distribuição dessa variável. Para isso, verifica-se qual distribuição de probabilidade ajusta melhor os dados e testes estatísticos são efetuados para determinar qual função de distribuição de probabilidade é mais adequada para calcular a probabilidade de ocorrer determinado fenômeno.

O objetivo desta monografia é estudar os modelos de distribuição de probabilidade para tempo de vida, o Exponencial e o de Webull, suas características, tais como a função densidade de probabilidade, a função de sobrevivência, a função de risco, demonstrando-as matematicamente Neste trabalho, faremos a comparação das distribuições de Weibull e Exponencial, utilizando um banco de dados da precipitação pluviométrica da cidade de Campina Grande entre os anos de 1978 à 1997, com o objetivo de testar qual das duas distribuições se ajusta melhor aos dados.

Capítulo 2

Fundamentação Teórica

2.1 Conceitos Básicos

A seguir, serão apresentados alguns aspectos teóricos importantes para a compreensão do estudo desenvolvido ao longo deste trabalho. Para proceder à análise, faz-se necessário o conhecimento prévio dos seguintes conteúdos: experimento aleatório, espaço amostral e evento, variável aleatória, confiabildade, censura, independência, outros conteúdos e diversidade de exercícios.

Um experimento é dito aleatório, quando mesmo repetido várias vezes sob as mesmas condições, pode apresentar resultados diferentes, isto é, embora conheça os resultados possíveis do experimento, não há como dizer um resultado particular desse experimento antecipadamente. O conjunto de todos os resultados possíveis de um experimento aleatório, é denominado Espaço amostral e será representado pela letra S. Todo resultado ou subconjunto de resultados um experimento é chamado de evento e, em geral, é denotado por letras maiúsculas do nosso alfabeto.

Muitas vezes, os resultados possíveis de um experimento não são numéricos. Para que seja possível usar uma modelagem matemática para tais experimentos, faz-se necessário associar números aos resultados do experimento.

Definição 2.1.1 *Seja S o espaço amostral associado a um experimento. Uma* **variável aleatória** *é uma função*

$$\begin{array}{rccc} X: & S & \to & \mathbb{R} \\ & s & \mapsto & X(s). \end{array}$$

Ou seja, uma variável aleatória é uma função que associa valores reais aos resultados de um experimento. Se uma variável X assume valores em um conjunto enumerável, ela é dita discreta. Por outro lado, se X assume valores em um intervalo da reta ou toda a reta, então X é dita uma variável aleatória contínua. As duas distribuições que serão alvo deste estudo são contínuas.

Definição 2.1.2 *Considere*

$$\begin{array}{rccc} f: & X(S) & \to & \mathbb{R} \\ & x & \mapsto & f(x) \end{array}$$

uma função que satisfaz as condições:

i) $f(x) \geq 0, \ \forall x \in X(S)$;

ii) $\displaystyle\int_{\infty}^{-\infty} f(x)dx = 1.$

é dita **função de densidade de probabilidade** *(f.d.p.) da variável aleatória X.*

Definição 2.1.3 *Seja X uma variável aleatória contínua com fdp $f(x)$. Dizemos que a Função de Distribuição Acumulada de X é a função F definida por:*

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(s)ds.$$

2.2 Inferência em Modelos Paramétricos

Os modelos probabilísticos são caracterizados por quantidade (valores) desconhecidos, chamados de parâmetros. Em cada estudo envolvendo análise de confiabilidade, os parâmetros devem ser estimados a partir de observações amostrais, para que o modelo fique determinado e, assim, seja possível responder as perguntas de interesse. Entre os vários métodos de estimação conhecidos na literatura estatística, está o método da máxima verossimilhança (Dantas, 2008).

2.2.1 Estimadores de Máxima Verossimilhança

Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória da variável aleatória X contínua com f.d.p. $f(x;\theta)$, onde destacamos o parâmetro θ desconhecido. Quando retiramos a amostra, observamos os valores obtidos $(x_1, x_2, \ldots, x_n)$.

Definição 2.2.1 *Definimos a função de verossimilhança por*

$$L(\theta, x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i;\theta) = f(x_1;\theta)f(x_2;\theta)\ldots f(x_n;\theta). \tag{2.1}$$

O estimador de máxima verossimilhança de θ, isto é, $\hat{\theta}$ é o valor de θ que maximiza

$$L(\theta, x_1, x_2, \ldots, x_n).$$

Em geral, se usa maximizar o logaritmo da verossimilhança l dado por:

$$l(\theta, x_1, x_2, \ldots, x_n) = ln(\theta, x_1, x_2, \ldots, x_n),$$

em lugar da função de verossimilhança. Isso pode ser feito porque a função logaritmo é uma função estritamente crescente. A vantagem de se fazer isso é que ao aplicarmos a função ln, transformamos o produtório em somatório e isso facilita os cálculos.

2.2.2 Confiabilidade

O termo confiabilidade é muito usado na manutenção, e teve origem na década de 50 nos Estados Unidos para análise de falha em equipamentos eletrônicos de uso militar. Confiabilidade é a probabilidade que um item possa desempenhar sua função, por um intervalo de tempo $[0, t]$, sob condições definidas de uso. O valor t não pode ser previsto a partir de um modelo determinístico, isto é, componentes "idênticos"sujeitos a "idênticos"esforços falharão em diferentes e imprevistos instantes. Deste modo o emprego de um modelo probabilístico, considerando T uma variável aleatória, constitui-se no único tratamento realista do assunto (GONDIM e DUARTE 2005).

Definição 2.2.2 *A confiabilidade de um sistema cuja função de distribuição do tempo de vida é $F(x)$ é definida por*

$$R(x) = 1 - F(x). \tag{2.2}$$

Um outro conceito importante no estudo da confiabilidade de sistemas é o de taxa de falha, cuja definição damos a seguir.

Definição 2.2.3 *A taxa de falha de um sistema cujo tempo de vida tem função de distribuição $F(x)$, com densidade de probabilidade $f(x)$ é dada por:*

$$h(x) = \frac{f(x)}{1 - F(x)} \tag{2.3}$$

A taxa de falha, que é uma função de tempo, é aproximadamente igual à probabilidade de que ocorra falha num intervalo de tempo, pequeno em relação a x, após o instante x, dado que não houve falha até esse instante.

De fato:

$$\lim_{\triangle u \to 0} \frac{P[x < X \leq x + \triangle x \mid X > x]}{\triangle x} = \lim_{\triangle u \to 0} \frac{P[x < X \leq x + \triangle x]}{(\triangle x) P[x > X]} =$$

$$= \frac{1}{1 - F(x)} \lim_{\triangle u \to 0} \frac{F(x + \triangle x) - F(x)}{\triangle x} = \frac{f(x)}{1 - F(x)}$$

Uma distribuição de vida pode ser melhor compreendida através da função taxa de falhas, pois esta reflete o comportamento dos itens sobreviventes. (WERNER 1996).

2.2.3 Censura

Observações incompletas freqüentemente ocorrem nos estudos de sobrevivência. Dados Censurados ocorrem, quando alguns sujeitos em estudo não "terminam" o evento de interesse, ou seja, falham até o fim do estudo ou tempo de análise. Por exemplo, em alguns estudos, pacientes abandonam o tratamento ou continuam vivos depois do final dos estudos, resultando em algumas observações incompletas, ditas censuradas. É válido salientar que, mesmo censurados, todos os resultados provenientes de um estudo de sobrevivência devem ser usados na análise estatística. Na aplicação abordada neste trabalho, isso ocorre quando o acompanhamento dos indíces pluviométricos são interrompidos, geralmente por mudanças climatológicas.

Observação 1 *Existem três tipos de censuras*

A Censura Tipo I *é aquela, onde o teste será terminado após um período pré-estabelecido de tempo, ou seja, Observações são acompanhadas até um período pré-estabelecido de tempo.*

A Censura Tipo II *é aquela, onde o teste será terminado após ter ocorrido a falha em um número pré-estabelecido de elementos sob teste,ou seja, Observações são acompanhadas até obter-se um número pré-determinado de falhas.*

Na Censura Tipo III *o período de estudo é fixado e os elementos entram no estudo em diferentes tempos durante aquele período, oe seja, acontece quando um elemento é retirado no decorrer do estudo sem ter ocorrido falha, ou se por exemplo, o elemento falhar por uma razão diferente da estudada.*

Capítulo 3

Distribuição Exponencial

Sendo um caso particular da distribuição de Weibull, a distribuição Exponencial caracteriza-se por desempenhar importante papel na descrição de fenômenos na área de confiabilidade, pois a probabilidade de falha não se altera ao longo do uso (taxa de falha constante). Segundo Meyer (1983), é bastante razoável admitir que um fusível ou um rolamento de rubis sejam "tão bons quanto novos", enquanto estiverem ainda funcionando. Isto é, se um fusível não tiver fundido, estará praticamente em estado de novo; nem o rolamento se alterará (muito) devido ao desgaste. A distribuição Exponencial apresenta-se como um modelo adequado de análise para esses casos.

i) É utilizada normalmente para representar a duração de um determinado serviço;

ii) Intervalo de tempo até a falha de uma peça de um equipamento;

iii) Fadiga.

Definição 3.0.4 *Diz-se que uma variável X tem distribuição exponencial com parâmetro λ se sua função de densidade de probabilidade é dada por:*

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & c.c. \end{cases} \tag{3.1}$$

Sendo f uma função de distribuição de densidade da distribuição exponencial, então om gráfico f é a curva esboçada na Figura 3.1.

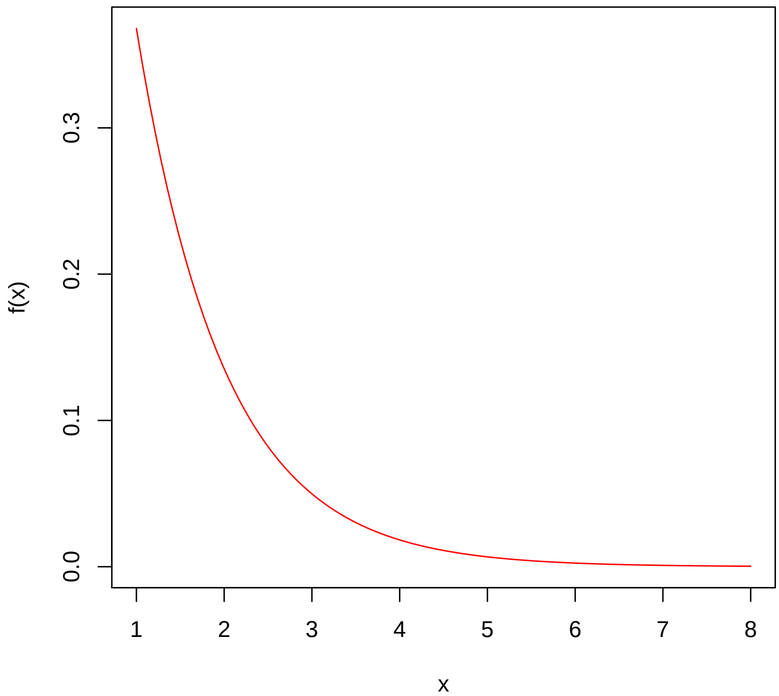

Figura 3.1: Curva da função de densidade da distribuição Exponencial.

Alternativamente, a distribuição exponencial pode ser caracterizada por sua função de distribuição acumulada a qual é obtida como:

$$F(x) = P(X \leq x) = \int_0^x f(t)dt = \int_0^x \lambda e^{-\lambda t} dt = -e^{-\lambda t}|_0^x = -(e^{-\lambda x} - 1)$$

Ou seja,

$$F(x) = \begin{cases} 1 - \lambda e^{-\lambda x}, & x > 0 \\ 0, & c.c. \end{cases} \tag{3.2}$$

Se F é um função de distribuição acumulada (f.d.a.) de uma variável aleatória contínua (v.a.c.) com distribuição exponencial, então o gráfico de F é a curva esboçada na Figura 3.2.

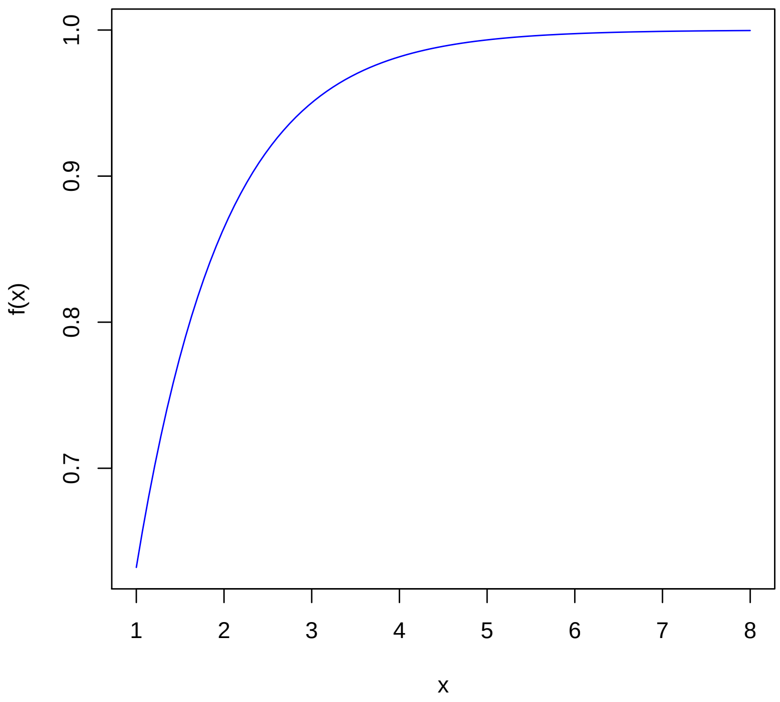

Figura 3.2: Curva da função de distribuição acumulada da Exponencial.

3.1 Propriedades da Distribuição Exponencial

A seguir, serão apresentadas algumas propriedades importantes que caracterizam a distribuição em questão.

3.1.1 Propriedades da falta de memória

A propriedade da falta de memória significa por exemplo, que a probabilidade de que seja necessário esperar, por exemplo, mais que 30 segundos até que o evento aconteça, dado que esse evento não aconteceu antes de 20 segundos, é a mesma de que esse evento ocorra depois dos 10 segundos iniciais. A única distribuição contínua com essa característica, é a exponencial. Uma forma equivalente de expressar essa propriedade consiste em mostrar que a distribuição exponencial não sofre desgaste ou, em outra linguagem, não tem memória. Considere um componente que tem distribuição de tempo de vida exponencial. Se ele durou até o instante t, então a probabilidade condicional dele durar mais s unidades de tempo além do instante t, é a mesma que um componente novo venha durar s unidades de tempo (Dantas, 2008). A expressão é dada por:

$$P(X > t+s|X > t) = \frac{P(X > t+s, X > t)}{P(X > t)} = \frac{P(X > t+s)}{P(x > t)} = P(X > s) \tag{3.3}$$

Demonstração:

De fato,

$$P(X > t+s) = 1 - P(X \leq t+s) = 1 - \int_0^{t+s} \lambda e^{-\lambda x} dx = 1 - \lambda \int_0^{t+s} e^{-\lambda x} dx$$

integrando por substituição e fazendo

$$u = \lambda x \Rightarrow du = \lambda dx$$

obtém-se

$$\begin{aligned} 1 - \lambda \int_0^{t+s} e^{-\lambda x} dx &= 1 - \lambda \int_0^{\lambda(t+s)} e^{-u} \frac{du}{\lambda} \\ &= 1 - \lambda \int_0^{\lambda(t+s)} e^{-u} du = 1 - \left[-e^{-u}\right]_0^{\lambda(t+s)} = \\ &= 1 - \left[-e^{-\lambda(t+s)} + e^0\right] = \\ &= 1 + e^{-\lambda(t+s)} - 1 = e^{-\lambda(t+s)}. \end{aligned} \tag{3.4}$$

Logo, $P(X > t+s) = e^{-\lambda(t+s)}$.

Por outro lado, temos

$$P(X > t) = 1 - P(X \leq t) = 1 - \int_0^t \lambda e^{-\lambda x} dx = 1 - [-e^{\lambda x}]_0^t = e^{-\lambda t} \tag{3.5}$$

Note que,

$$P(X > t+s) = e^{-\lambda(t+s)}$$

e

$$P(X > t) = e^{-\lambda t}$$

Fazendo o quociente de (3.4) e (3.5), obtemos

$$P(X > t+s|X > t) = \frac{e^{\lambda}(t+s)}{e^{\lambda x}} = 1 - \left[-e^{-\lambda x}\right]_0^s = e^{-\lambda s} \tag{3.6}$$

Encontremos agora $P(X > s)$

$$P(X > s) = 1 - P(X \leq s) = 1 - \int_0^s \lambda e^{-\lambda x} dx = 1 - \left[-e^{-\lambda x}\right]_0^s = e^{-\lambda s} \tag{3.7}$$

Portanto, de (3.6) e (3.7), obtemos

$$P(X > t+s|X > t) = P(X > s).$$

3.2 Esperança e Variância da Exponencial

Seja X uma variável aleatória contínua com distribuição exponencial com parâmetro λ. A esperança e a variância são dadas respectivamente por:

$$E(X) = \frac{1}{\lambda} \quad \text{e} \quad Var(X) = \frac{1}{\lambda^2}$$

Demonstrações:

Inicialmente será calculada a esperança da distribuição. Por definição tem-se

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{0}^{\infty} x\lambda e^{-\lambda x}dx. \tag{3.8}$$

Integrando por partes e fazendo $u = x$, $dv = \lambda e^{-\lambda x}$, obtemos $du = dx, v = -e^{\lambda x}$. Daí,

$$E(X) = \left[-xe^{-\lambda x}\right]|_0^\infty + \int_0^\infty -e^{-\lambda x}dx \tag{3.9}$$

Integrando por substituição o 2º membro do lado direito da expressão (3.9) e fazendo $w = -\lambda$, obtemos $du = -\lambda dx$. Logo,

$$E(X) = [-xe^{-\lambda x}]|_0^\infty + \int_0^\infty \frac{e^u}{\lambda}du = \frac{1}{\lambda} \tag{3.10}$$

Portanto,

$$E(X) = \frac{1}{\lambda}.$$

Agora será calculada a variância. Usaremos a seguinte fórmula para encontar a variância:

$$Var(X) = E(X^2) - [E(X)]^2 \tag{3.11}$$

Onde

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^\infty x^2\lambda e^{-\lambda x}dx.$$

Integrando por partes e fazendo $u = x^2$, $dv = \lambda e^{-\lambda x}$, obtemos $du = 2xdx, v = -e^{-\lambda x}$. Logo,

$$E(X) = [-x^2 e^{-\lambda x}]|_0^\infty + 2\int_0^\infty e^{-\lambda x}dx = \frac{2}{\lambda^2} \tag{3.12}$$

Substituindo (3.10) e (3.12) em (3.11), obtemos

$$Var)x) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda 2}$$

Portanto,

$$Var(X) = \frac{1}{\lambda^2}.$$

3.3 Função geratriz de momentos

Definição 3.3.1 *A função geratriz de momentos de uma variável aleatória X é dada por*

$$M_x(t) = E(e^{tX}), \tag{3.13}$$

desde que tal esperança exista. Em particular, se X é contínua com f.d.p $f(x)$, tem-se

$$M_x(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx. \tag{3.14}$$

Suponha que X tem distribuição exponencial, com parâmetro λ, então

$$M_x(t) = \frac{\lambda}{\lambda - t} \tag{3.15}$$

Demonstração:

Observe que,

$$M_x(t) = E(e^{tX}) = \int_0^{\infty} e^{tx}\lambda e^{-\lambda x}dx = \lambda \int_0^{\infty} e^{-(\lambda-t)x}dx.$$

A convergência dessa integral acontece somente se $t < \lambda$. Por isso, a $f.g.m.$ existe somente para aqueles valores de t. Admitamos que a condição seja satisfeita, portanto podemos prosseguir. Daí, integrando por substituição e fazendo $u = (\lambda - t)$, obtemos $du = (\lambda - t)dx$. Portanto,

$$\begin{aligned}
&\int_0^{\infty} e^{tx}\lambda e^{-\lambda x}dx = \lambda \int_0^{\infty} e^{-u}\frac{du}{\lambda - t} = \\
&= \frac{\lambda}{\lambda - t}\int_0^{\infty} e^{-u}du = \\
&= \frac{\lambda}{\lambda - t}[-e^{-u}]|_0^{\infty} = \\
&= \frac{\lambda}{\lambda - t}, \quad t < \lambda.
\end{aligned} \tag{3.16}$$

3.4 Parametrização alternativa

É possível parametrizar a densidade exponencial em termos de um parâmetro.

Seja

$$\beta = \frac{1}{\lambda} \tag{3.17}$$

Neste caso, $f(x) = \frac{1}{\beta}e^{\frac{-x}{\beta}}, \quad x > 0; \quad \beta > 0.$

$E(X) = \beta,$

$E(X^2) = 2\beta^2,$

$Var(X) = \beta^2.$

Essa parametrização alternativa é bastante interessante, pois o valor médio é igual ao parâmetro.

3.5 Estimação de parâmetro

Uma vez determinado qual modelo probabilístico é mais adequado para representar a variável de interesse na População em estudo, resta obter os seus parâmetros. Nos estudos feitos com base em amostras é preciso escolher qual das estatísticas da amostra será o melhor estimador para cada parâmetro do modelo. Como os parâmetros serão estimados através das estatísticas (estimadores) de uma amostra aleatória, e como para cada amostra aleatória as estatísticas apresentarão diferentes valores, os estimadores também terão valores aleatórios. Em outras palavras um Estimador é uma variável aleatória que segue uma distribuição de probabilidade. Existem técnicas para estimação de parâmatros que são bastante difundidas nos estudos envolvendo estatísticas, como por exemplo, estimação por máxima verossimilhança, método dos momentos, método da regresão, etc. No estudo em questão, utilizaremos o método da máxima verossimilhança para encontrar o estimador da distribuição exponencial.

3.5.1 Método da máxima verossimilhança

Em estudos inferenciais, quando se tem um conjunto de dados e um modelo estatístico, usa-se o método de máxima verossimilhança para estimar os valores dos diferentes parâmetros do modelo estatístico de forma a maximizar a probabilidade dos dados observados (ou seja, busca parâmetros que maximizem a função de verossimilhança). O método de máxima verossimilhança apresenta-se como um método geral para estimação de parâmetros. Porém, deve-se observar que em alguns casos a estimativa por máxima verossimilhança pode ser inadequada. Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória com distribuição exponencial. O estimador de máxima verossimilhança de λ é aquele para o qual a função

$$L(\lambda; x_1, x_2, \ldots, x_n) = \prod_{i=1}^{\infty} f(x_i),$$

é máxima, sendo $X_1, X_2, \ldots, X_n$ uma amostra independente e identicamente distribuída (i.i.d.) de uma distribuição exponencial com parâmetro λ e a parametrização considerada será aquela dada no item acima.

Neste caso,

$$f(x) = \begin{cases} \dfrac{1}{\lambda} e^{\frac{-x}{\lambda}}, & x > 0 \\ 0, & c.c. \end{cases} \tag{3.18}$$

Daí,

$$\begin{aligned} & L(\lambda, x_1, x_2, \ldots, x_n) = \textstyle\prod_{i=1}^{\infty} f(x_i;\theta) = \\ & = \frac{1}{\lambda} e^{\frac{-x_1}{\lambda}} \frac{1}{\lambda} e^{\frac{-x_2}{\lambda}} \ldots \frac{1}{\lambda} e^{\frac{-x_n}{\lambda}} = \\ & = \frac{1}{\lambda^n} e^{\frac{-1}{\lambda}} \sum_{i=1}^{n} x_i. \end{aligned} \tag{3.19}$$

Como se sabe, maximizar a função (3.20), é equivalente a maximizar $l(\lambda, x_1, x_2, \ldots, x_n)$ definida a seguir, visto que a função $ln(.)$ é uma função estritamente crescente.

$$\begin{aligned} & l(\lambda, x_1, x_2, \ldots, x_n) = lnL(\lambda, x_1, x_2, \ldots, x_n) = \\ & = -nln\lambda - \frac{1}{\lambda} \sum_{i=1}^{n} xi \end{aligned} \tag{3.20}$$

Assim, o candidato a ponto de máximo da função $l(\lambda, x_1, x_2, \ldots, x_n)$ é o valor $\hat{\lambda}$ tal que:

$$\frac{\partial l}{\partial \lambda} = \frac{-n}{\hat{\lambda}} + \frac{\sum_{i=1}^{n} x_i}{\hat{\lambda}^2} = 0$$

ou seja, $-n\hat{\lambda} + \sum_{i=1}^{n} x_i = 0$, o que implica

$$\hat{\lambda} = \frac{\sum_{i=1}^{n}}{n} x_i = \bar{X} \tag{3.21}$$

Para que este seja realmente o máximo, é preciso observar se

$$\frac{\partial^2 l}{\partial \lambda^2} < 0 \tag{3.22}$$

De fato,

$$\frac{\partial^2 l}{\partial \lambda^2} = \frac{n}{\hat{\lambda^2}} - \frac{2\sum_{i=1}^{n} x_i}{\hat{\lambda^4}} < 0$$

Vale salientar que no caso em questão, optou-se por usar a parametrização (3.17), tem-se

$$\lambda^* = \frac{1}{\lambda} \tag{3.23}$$

e então, pelo princípio da invariância da estimação por máxima verossimilhança $(E.M.V.)$,

$$\hat{\lambda^*} = \frac{1}{\hat{\lambda}} = \frac{1}{\bar{X}}. \tag{3.24}$$

.

Capítulo 4

Distribuição de Weibull

4.1 Distribuição de Weibull

Uma generalização da distribuição exponencial foi feita por Waloddi Weibull (1951) para atender aos problemas não resolvidos pela distribuição exponencial. Segundo Filho (2006) em setembro de 1951, foi publicado um artigo intitulado "Uma Função de Distribuição Estatística de Larga Aplicação"pelo "Jornal de Mecânica Aplicada"no qual Ernest Hjalmar Wallodi Weibull apresentou o estudo feito a respeito da resistência mecânica de aços, estudos de tração em correntes construídas com estes aços e a apresentação de seu modelo semi-empírico, sendo que este permite a representação de: falhas típicas de partida, falhas aleatórias e falhas devido ao desgaste. Esta distribuição é uma generalização da distribuição exponencial, para os casos em que a taxa de falha seja decrescente (típico de falhas de juventude) ou crescente (típico dos mecanismos de desgaste). Pode-se dizer que distribuição de Weibull é uma função de densidade de probabilidade usada na análise de dados de vida, por exemplo, em engenharia de confiabilidade (Meyer, 1995), análise de sobrevivência e em outras áreas devido à versatilidade desta distribuição.

Aplica-se na análise de confiabilidade, pois permite:

- Reperesntar falhas típicas de partida (mortalidade infantil;
- Fahas aleatórias;
- Obtenção de parâmetros significativos da configuração de falhas;
- Representação gráfica simples.

Definição 4.1.1 *Uma variável aleatória X tem distribuição de Weibull com parâmetros $\alpha > 0$ e $\beta > 0$ se sua função de densidade de probabilidade é dada por:*

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{\left(-\frac{x}{\beta}\right)^{\alpha}}, \quad x > 0, \tag{4.1}$$

.

Devido a sua flexibilidae, pode-se citar dois casos especiais na distribuição de Weibull, um deles é quando o parâmetro de forma $\alpha = 2$ a distribuição de Weibull se transforma na distribuição de Rayleigh, e quando o parâmetro de forma $\alpha = 1$ a distribuição de Weibull se reduz a distribuição exponecial.

A Figura 4.1 ilustra o comportamento da função de densidade de probabilidade (f.d.p.) de Weibull para diferentes valores do parâmetro α.

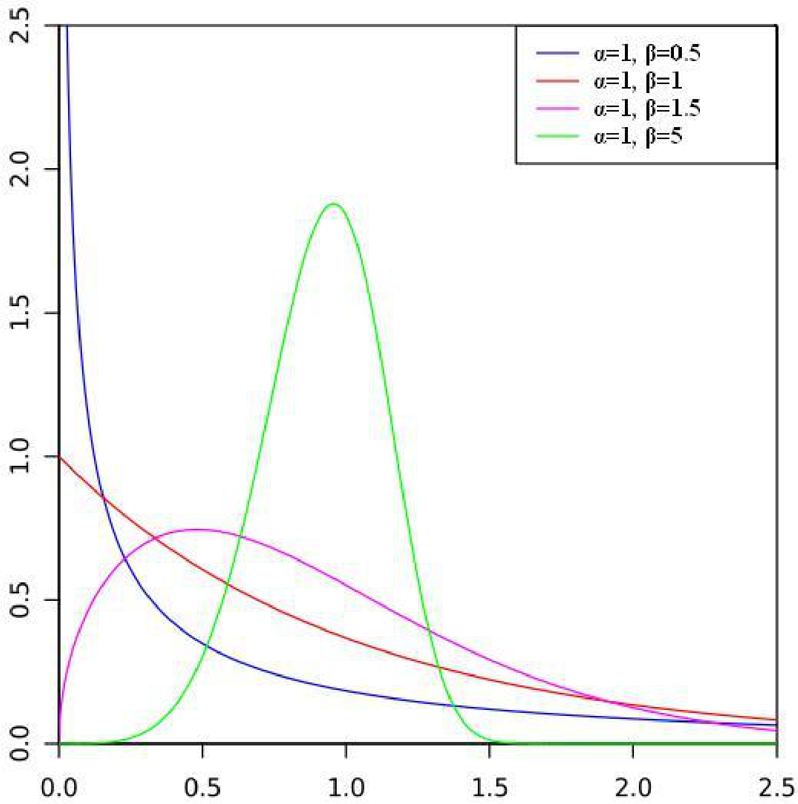

Figura 4.1: Curva da função de densidade de probabilidade da weibull.

A seguir, será demonstrado que a função $f(X)$ é de fato, uma função densidade de probabilidade. Incialmente, o que se quer mostrar é que

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\infty} \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{\left(-\frac{x}{\beta}\right)^{\alpha}} dx = 1. \tag{4.2}$$

De fato, fazendo a transformação de variáveis $u = (\dfrac{x}{\beta})^{\alpha}$, obtém-se $du = (\dfrac{\alpha}{\beta})^{\alpha-1}$. Desse modo,

$$\int_0^{\infty} \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{\left(-\frac{x}{\beta}\right)^{\alpha}} dx = = \int_0^{\infty} e^{-u} du = 1. \tag{4.3}$$

4.1.1 Função de distribuição acumulada

Se X tem distribuição de Weibull, sua função de distribuição acumulda ($f.d.a.$) é dada por:

$$F(x) = 1 - e^{\left(\frac{x}{\beta}\right)^{\alpha}}. \tag{4.4}$$

Na Figura 4.2, pode-se observar o comportamento do gráfico da f.d.a. para valores variados dos parâmetros α, e β.

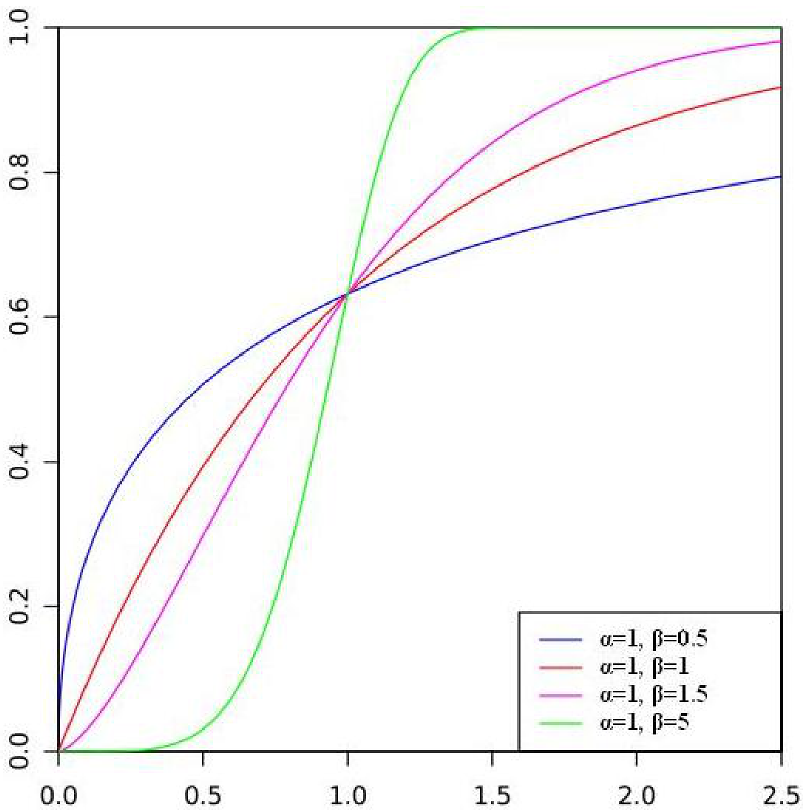

Figura 4.2: Curva da função de distribuição acumulada da weibull.

A seguir, será demonstrado que a f.d.a. de uma v.a. X com distribuição de Weibull com parâmetros α e β é dada pela fórmula apresentada na equação (4.4).

Integrando por substituição e fazendo $u = \left(\dfrac{x}{\beta}\right)^{\alpha}$, $t = 0$, $t = x$, obtemos $u = 0$, $u =$

$\left(\frac{x}{\beta}\right)^{\alpha}$, $du = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}$. Portanto,

$$\begin{aligned} F(x) &= \int_0^x \frac{\alpha}{\beta}\left(\frac{t}{\alpha}\right)^{\alpha-1} e^{\left(\frac{-t}{\beta}\right)}\alpha dt = \\ &= \int_0^{\left(\frac{x}{\beta}\right)^{\alpha}} e^{-u}du = e^{-u}\Big|_0^{\left(\frac{x}{\beta}\right)^{\alpha}} = 1 - e^{\left(\frac{x}{\beta}\right)^{\alpha}}. \end{aligned} \tag{4.5}$$

que é exatamente o que se queria demonstrar.

4.1.2 Esperança e Variância da distribuição de Weibull

Definição 4.1.2 *Se X é uma variável aleatória com distribuição de Weibull com parâmetro α e β, então*

$$E(X) = \beta\Gamma\left(\frac{\alpha+1}{\alpha}\right) \tag{4.6}$$

e

$$Var(x) = \beta^2[\Gamma\left(\frac{\alpha+2}{\alpha} - \Gamma^2\right) - \Gamma^2\left(\Gamma\frac{\alpha+1}{\alpha}\right) \tag{4.7}$$

Demonstração:

Provemos inicialmente (4.6).

Por definição a esperança é dada por (3.8). Sabe-se ainda que a esperança de uma variável aleatória X é o primeiro momento dessa variável. Portanto, usaremos esse fato para provar (4.6). Vamos calcular o momento de ordem r:

$$E(X^r) = \int_0^{\infty} \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{\left(\frac{-x}{\beta}\right)^{\alpha}} x^r dx. \tag{4.8}$$

Integrando por substiutição e fazendo $u = \frac{x}{\beta}$, $\beta^r = \frac{x^r}{u^r}$, obtemos $dx = \beta du$, e $x^r = \beta^r u^r$. Segue que

$$E(X^r) = \int_0^{\infty} \frac{\alpha}{\beta} u^{\alpha-1} e^{(-u)^{\alpha}} \beta^r u^r \beta du = \int_0^{\infty} \alpha u^{\alpha-1} e^{(-u)^{\alpha}} \beta^r u^r du$$

fazendo $u^{\alpha} = t$, e $\alpha u^{\alpha-1}du = dt$, obtemos $u = t^{\frac{1}{\alpha}}$. Logo,

$$\begin{aligned} E(X^r) &= \int_0^{\infty} e^{-t}\beta^r \left(t^{\frac{1}{\alpha}}\right)^r dt = \\ &= \beta^r \int_0^{\infty} t^{\frac{r}{\alpha}} e^{-t} dt = \\ &= \beta^r \int_0^{\infty} t^{\frac{r}{\alpha}+1-1} e^{-t} dt = \\ &= \beta^r \int_0^{\infty} t^{\frac{r+\alpha}{\alpha}-1} e^{-t} dt = \beta\Gamma\left(\frac{r+\alpha}{\alpha}\right). \end{aligned} \tag{4.9}$$

Tomando $r = 1$ em (4.9), obtemos

$$E(X) = \beta\Gamma\left(\frac{\alpha+1}{\alpha}\right).$$

Provemos agora (4.7).

Pode-se calcular a variância utilizando (3.11). Fazendo $r = 2$ em (4.9), obtemos

$$E(X^2) = \beta^2\Gamma\left(\frac{\alpha+2}{\alpha}\right). \tag{4.10}$$

Substituindo (4.9) e (4.10) em (3.11), obtemos

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

o que conclui a demonstração.

4.1.3 Função geratriz de momentos

Se X tem distribuição de Weibull com parâmentros α e β, então sua função geratriz de momentos ($f.g.m.$), $M_x(t)$, é dada por

$$M_x(t) = \sum_{r=0}^{\infty}\frac{t^r}{r!}\beta^r\Gamma\left(1+\frac{r}{\alpha}\right), \tag{4.11}$$

e tal função é obtida da seguinte forma:

$$M_x(t) = E\left(e^{tx}\right) = E\left(\sum_{r=0}^{\infty}\frac{(tx)^r}{r!}\right) = 1 + \sum_{r=1}^{\infty}\frac{t^r E\left(x^r\right)}{r!} =$$
$$= 1 + \sum_{r=1}^{\infty}\frac{t^r}{r!}\beta\Gamma\left(\frac{r+\alpha}{\alpha}\right) = \sum_{r=1}^{\infty}\frac{t^r}{r!}\beta\Gamma\left(1+\frac{r}{\alpha}\right).$$

4.1.4 Função de confiabilidade

A função de confiabilidade, denotada por $R(t)$ (do inglês Reliability), representa a probabilidade de não haver falha no intervalo $(0, t]$. Sua expressão matemática é a seguinte:

$$R(t) = e^{-\left(\frac{t}{\alpha}\right)^{\beta}}. \tag{4.12}$$

A Figura 4.3 esboça o gráfico de $R(t)$.

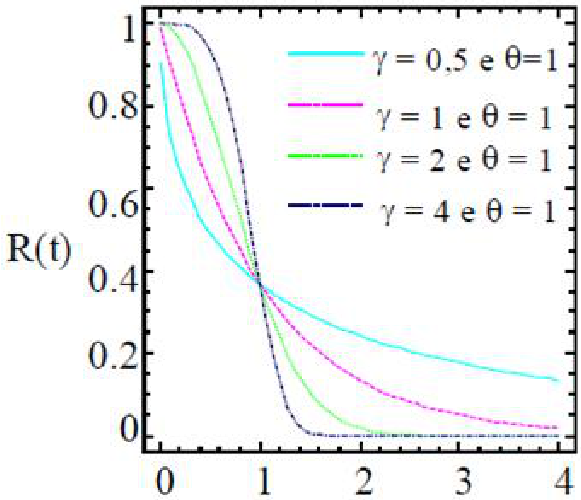

Figura 4.3: Gráfico da função de confiabilidade

4.1.5 Função de Sobrevivência

A função de sobrevivência, denotada por $S(t)$, representa a probabilidade de um indivíduo sobreviver pelo menos até certo tempo t. Sua expressão matemática é a seguinte:

$$S(t) = e^{\left(\frac{t}{\alpha}\right)^{r}}. \tag{4.13}$$

A partir de $f(x)$ e de $F(x)$ pode-se obter a taxa de falha da distribuição de Weibull da seguinte forma:

$$h(x) = \frac{f(x)}{1 - F(x)} = \frac{\frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}}{e^{-\left(\frac{x}{\beta}\right)^{\alpha}}} = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}, \quad se \quad x > 0, \tag{4.14}$$

Onde,

$$h(x) = \begin{cases} \text{Estritamente crescente para } \beta > 1 \\ \text{Estritamente decrescente para } \beta < 1 \\ \text{Constante para } \beta = 1 \end{cases}$$

4.1.6 Estimação dos parâmetros

Se a distribuição do tempo de vida T pertence a família Weibull com parâmetros de forma α e de escala β, as funções de densidade de probabilidade, de sobrevivência e de risco são dadas respectivamente por,(LEE, 1992, MILLER, 1981).

$$f_T(t) = \beta \left(\tfrac{1}{\alpha}\right)^{\beta} t^{\beta-1} exp\left[-\left(\tfrac{t}{\alpha}\right)^{\beta}\right], \quad se \quad t \geq 0, \quad \alpha, \beta > 0; \tag{4.15}$$

$$S_T(t) = exp\left[-\left(\tfrac{t}{\alpha}\right)^{\beta}\right]; \tag{4.16}$$

$$h(t) = \frac{1}{\alpha}\beta\left(\frac{1}{\alpha}t\right)^{\beta-1}, \quad para \quad t \geq 0, \quad \alpha, \beta > 0. \tag{4.17}$$

Quando algumas das n observações são censuradas em uma amostra de n indivíduos observados sob um esquema de censura aleatória tipo I, e os tempos de vida seguem a distrbuição de Weibull com função de densididade dada por $f(x)$, a função de verossimilhança dos parâmetros α e β é dada por:

$$L(\alpha, \beta) = \prod_{i=1}^{n} \left(f_T(t_i; \alpha, \beta)\right)^{\delta_i} \prod_{i=1}^{n} \left(S_T(t_i; \alpha, \beta)\right)^{1-\delta}. \tag{4.18}$$

Com relação a função de risco e a função de sobrevivência, a função de verossimilhança pode, ainda, ser escrita como:

$$L(\alpha, \beta) = \prod_{i=1}^{n} \left(h_T(t_i; \alpha, \beta)\right)^{\delta_i} \prod_{i=1}^{n} \left(S_T(t_i; \alpha, \beta)\right). \tag{4.19}$$

Portanto,

$$L(\alpha, \beta) = \prod_{i=1}^{n} \left(\beta(\frac{1}{\alpha})^{\beta}(t_i)^{\beta-1}\right)^{\delta_i} \prod_{i=1}^{n} exp\left[-(\frac{t_i}{\alpha})^{\beta}\right]. \tag{4.20}$$

O logarítmo da função de verossimilhança é:

$$ln\left[L(\alpha, \beta)\right] = fln(\beta) - r\beta ln(\alpha) + (\beta - 1)\sum_{i=1}^{n} \delta_i ln(t_i - \sum_{i=1}^{n} \left(\frac{t_i}{\alpha}\right)^{\beta}. \tag{4.21}$$

Onde, $r = \sum_{i=1}^{n} \delta_i$ é o número de observações não censuradas.

O sistema de equações (4.22) não lineares, não tem solução analítica, sendo necessário métodos numéricos para solução. Um método iterativo é utilizado para resolver o sistema de equações não lineares para encontrar os estimadores de máxima verossilmilhança de α e β.

$$\left|\begin{array}{l} \dfrac{\partial l(\alpha, \beta)}{\partial \alpha} = -\dfrac{\beta}{\alpha}\displaystyle\sum_{i=1}^{n} \delta_i + \dfrac{\beta}{\alpha}\sum_{i=1}^{n}\left(\dfrac{t_i}{\alpha}\right)^{\beta} = 0 \\ \dfrac{\partial l(\alpha, \beta)}{\partial \beta} = \dfrac{1}{\beta}\displaystyle\sum_{i=1}^{n} \delta_i + \sum_{i=1}^{n} \delta_i log\left(\dfrac{t_i}{\alpha}\right) - \sum_{i=1}^{n}\left(\dfrac{t_i}{\alpha}\right)^{\beta} log\left(\dfrac{t_i}{\alpha}\right) = 0 \end{array}\right. \tag{4.22}$$

Através de software estatísticos como SAS e R, pode-se encontrar facilmente as estimativas de máxima verossimilhança, uma vez que tenha sido espcificada a função de logverossimilhança.

Capítulo 5

Aplicação

Vimos que as distribuições apresentadas são usadas amplamente para modelar dados de tempo de vida, mas essa não é a única aplicação. Devido a dificuldade de encontrar conjunto de dados para serem usados neste trabalho, não serão usados dados de tempo de vida.

Neste capítulo, será apresentado uma aplicação na área de Meteorologia utilizando dados pluviométricos coletados mensalmente na estação Bodocongó (PLU.01384) localizada na cidade de Campina Grande - PB no período de 1978 à 1997.

5.1 Material

Os dados utilizados neste trabalho, estão disponíveis no site da Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA, os quais podem ser verificados em [23]. Foram selecionados dados mensais de precipitação pluviométrica da cidade de Campina Grande - PB, no período de 1978 a 1997 num total de 18 anos, através da estação bodocongó (.PLU.01384). Foram feitos os tratamentos estatísticos iniciais do dados, os quais consistem em coleta, organização e tabulação. Em seguida,foram iniciadas as análises a partir das distribuições de probabildade Exponencial e Weibull, tendo como objetivo verificar o ajuste dos dados as respectivas distribuições. A análise dos dados foi feita através do software estatístico R V.2.13.0.

5.2 Resultado e discussão

A partir dos dados, construiu-se histogramas para cada um dos meses do período em estudo, totalizando 12 histogramas, e foram verificados os ajustes dos dados para cada distribuição. Os histogramas para cada mês estão apresentados nas Figuras 5.1 e 5.2 respectivamente. Analisando os histogramas abaixo, pode-se concluir que tipo de função se ajusta aos dados. No caso em questão ajustaremos as distrbuições exponencial e de Weibull.

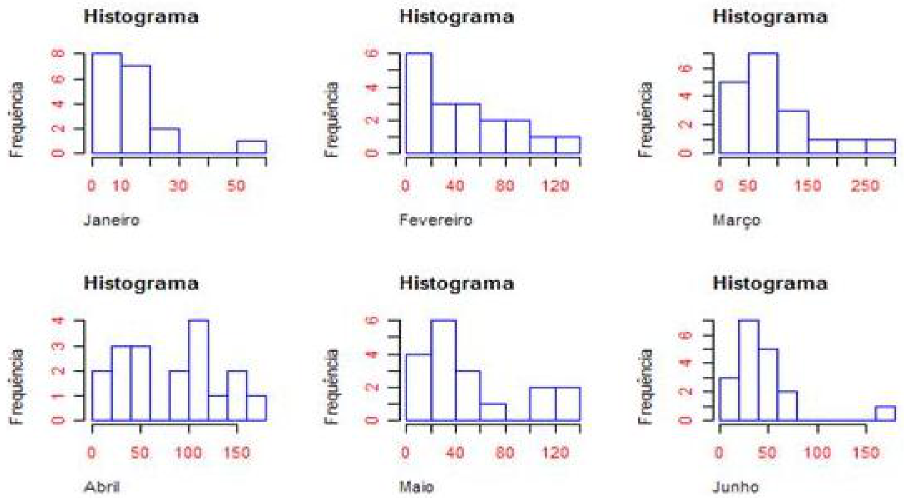

Figura 5.1: Histogramas dos meses de Jan, Fev, Març, Abr, Mai e Jun.

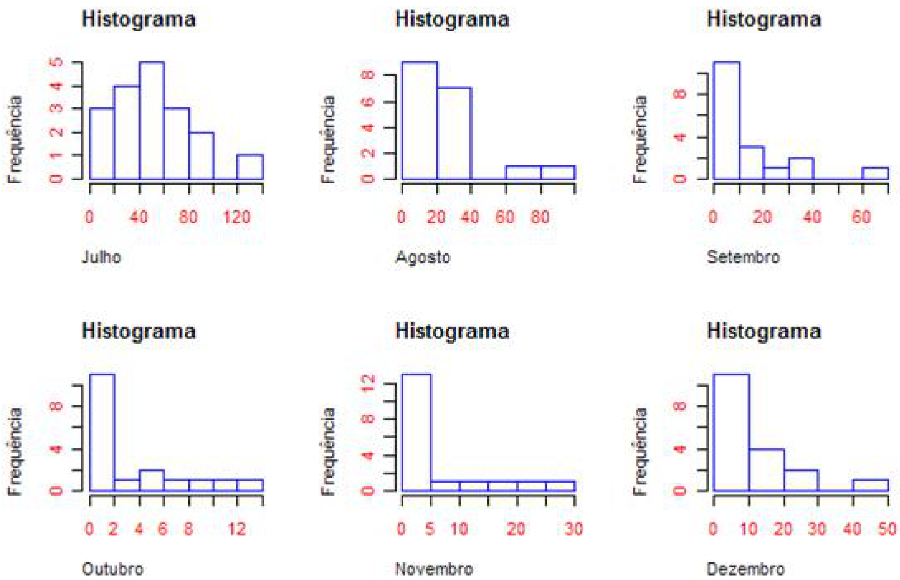

Figura 5.2: Histogramas dos meses de Jul, ago, Set, out, Nov e Dez

Após o tratamento inicial dos dados (levantamento da amostra e ordenação), foi observado dois tipos de distribuição e proposto um modelo para distribuição: Exponencial e Weibull. Foram estimados os parâmetros de que dependem essas distribuições propostas, conforme pode-se verificar na Tabela 1.

Tabela 1: Parâmetros das distribuições de probabilidade.

Mês	Distribuições de Probabilidade		
	Parâmetros		
	Weibull		Exponencial
	α	β	α
Janeiro	0,68956	9,91194	12,53529
Fevereiro	1,1189	51,4311	48,45292
Março	1,35389	99,90883	85,08742
Abril	1,57848	92,60079	81,17643
Maio	1,09148	49,43046	50,48233
Junho	1,18997	40,17237	39,58787
Julho	1,81586	59,9895	51,00586
Agosto	0,94976	23,79897	24,54116
Setembro	0,41348	6,30635	14,09999
Outubro	0,20059	0,2482	6,30635
Novembro	0,60067	4,28741	6,51764
Dezembro	0,22447	1,49261	10,98235

Fonte: Dados da pesquisa

Com as estimativas, foi executado o ajustamento, verificando quais seriam os valores esperados, com base na estimativa, isto é testa-se a aderência, verificando se é possível admitir que os valores seguem uma das distribuições proposta. Foi testada a hipótes de que os dados seguem uma distribuição exponencial e depois testada a hipótese de que os dados seguem uma distribuição de Weibull, ao nível de 5% de significância.

Foi utilizado o teste de Kolmogorv-Smirnov para verificação da aderência, onde o valor da estatística D máximo do teste de aderência de Kolmogorov-Smirnov (Campos, 1983; Assis et al., 1996; Bussab & Morettin, 2004) informa a máxima distância entre as probabilidades empíricas e as teóricas obtidas sob a função de distribuição de probabilidade em teste. Assim, quanto menores os valores da estatística D, maiores serão fornecidos os valores do p-valor e, consequentemente, maior evidência de não-rejeição da hipótese nula (H_0), ou seja, maior aderência dos dados à distribuição em teste (FILHO et al., 2004). Como se pode observar através dos p-valores da tabela 2, a distrinuição dos dados não se adéqua a distribuição exponencial para todos meses do período, ou seja, não há evidências estatísticas que nos leve a aceitação de que a distribuição dos dados seguem uma distribuição exponecial ao nível de 5% de significância.

Por outro lado, verifica-se que a distribuição dos dados se adéqua a distribuição de Weibull em todos os meses, isto é, a 5% de significância podemos aceitar que os dados seguem a distribuição de Weibull.

Portanto, através do teste de Kolmogorov-Smirnov, verifica-se que a distribuição de Weibull é a mais adequada para o estudo de precipitações pluviométricas na cidade de Campina Grande.

Tabela 2: P-valores e Estatistica de teste D.

Mês	Distribuições					
	Weibull	**D**	**p-valor**	**Exponencial**	**D**	**p-valor**
Janeiro	Aceita	0.194	0.451	Rejeita	1	$2,2 \times 10^{-16}$
Fevereiro	Aceita	0.1788	0.5535	Rejeita	1	$2,2 \times 10^{-16}$
Março	Aceita	0.2404	0.2118	Rejeita	1	$2,2 \times 10^{-16}$
Abril	Aceita	0.1291	**0.8884**	Rejeita	1	$2,2 \times 10^{-16}$
Maio	Aceita	0.1127	**0.9566**	Rejeita	1	$2,2 \times 10^{-16}$
unho	Aceita	0.2029	0.3956	Rejeita	1	$2,2 \times 10^{-16}$
Julho	Aceita	0.1422	0.8114	Rejeita	1	$2,2 \times 10^{-16}$
Agosto	Aceita	0.2534	0.166	Rejeita	1	$2,2 \times 10^{-16}$
Setembro	Aceita	0.1659	0.6453	Rejeita	1	$2,2 \times 10^{-16}$
Outubro	Aceita	0.2824	0.09204	Rejeita	1	$2,2 \times 10^{-16}$
Novembro	Aceita	0.2471	0.1872	Rejeita	1	$2,2 \times 10^{-16}$
Dezembro	Aceita	0.1475	0.7763	Rejeita	1	$2,2 \times 10^{-16}$

Fonte: Dados da pesquisa

Capítulo 6

Conclusão

Neste trabalho, foram apresentadas duas distribuições de probabilidade de larga aplicação em situações práticas: a Exponencial e a de Weibull. Apesar de ambas as distribuições serem frequentemente usadas para modelar dados relacionados a tempo, esse não é o único contexto em que se aplica tais distribuições. Aqui foram usados dados referentes à precipitação pluvial mensal observada na cidade e Campina Grande - PB, no período de 1978 a 1997. Feitos os histogramas, observou-se que sua forma para os meses de outubro, novembro e dezembro indicam que a distribuição exponencial poderia ser adequada para modelar os dados.

No entanto, quando se propõe tal ajuste, estimando os parâmetros do modelo com base na amostra e usando o teste de aderência de Kolmogorov-Smirnov para avaliar a adequacidade do modelo, viu-se que os dados não se ajusta ao modelo, da mesma forma para os demais meses. Assim, foi proposta uma nova abordagem usando a distribuição de Weibull, dada a a maior flexibilidade do modelo, por conta de seus parâmetros, Desta vez, estimados os parâmetros, o teste de aderência indicou que o modelo de Weibull se adequa bem para ajustar os dados com base na amostra que se tem.

Referências Bibliográficas

[1] BARROS, M. Probabilidade II. Rio de Janeiro: ENCE, 2009. 55 slides.

[2] BUSSAB, W.O.; MORETTIN, P.A. Estatística básica. 5.ed. São Paulo: Saraiva, 2004.526p.

[3] CAMPOS, H. Estatística experimental não-paramétrica. 4 ed. Piracicaba: Departamento de Matemática e Estatística - ESALQ,1983. 349p.

[4] CASELLA, G., BERGER, R.L. Inferência Estatística. Cengage Learning,2010. (Versão em português da 2nd edição em inglês).

[5] COLOSIMO, Enrico Antônio, Giolo, Suely Ruiz. Análise de Sobrevivência Aplicada. Editora Edgard Blucher, São Paulo, 2006.

[6] DANTAS, B. A. C. Probabilidade: Um curso Introdutório. 3. ed. rev. São Paulo: Editora da Universidade de São Paulo, 2008.

[7] DANTAS, M. 2004, "Modelagem de dados de falhas de equipamentos de sub-superfície em poços de petróleo da bacia Potiguar". Dissertação de Mestrado, Universidade Federal do Rio Grande do Norte, Natal - RN, Brasil, 118p.

[8] FARIAS, A. M. L. Variáveis Aleatórias Contínuas. Rio de Janeiro: Universidade Federal Fluminense, Instituto de Matemática, 2009, 14 p. (Notas de aula).

[9] FILHO, A.C.; MATZENAUER, R.; TRINDADE. J.K. Ajustes de funções de distribuição de probabilidade à radiação solar global no Estado do Rio Grande do Sul Pesquisa agropecuária brasileira, v.39, n.12, p. 1157-1166, 2004.

[10] Filho, G.B., 2006, "Distribuição de Weibull na Manutenção," Disponível em http://www.inpg.org/upload/entrance_quis/20060807102833.pdf, Acesso em 10/05/2007.

[11] GONDIM, M. R.; DUARTE, V. A M. Aplicação da Estatística na Manutenção preventiva, Uberlândia, p. 1-2, 2005.

[12] I. F. SOUZA, W. J. C. LUNDGREN, A. O. A. NETTO, Comparação entre distribuições de probabilidades da precipitação mensal no Estado de Pernambuco. Scientia Plena 6, 067001 (2010

[13] LEE, E. T. Statistical methods for survival data analysis. 2nd ed. New York: John Wiley, 1992. 482p.

[14] MANTOVANI, A.; FRANCO, P. A M. Estudo da Distribuição Assintótica dos Estimadores dos Parâmetros da Distribuição Weibull na Presença de dados Sujeitos a Censura Aleatória, v. 22, 12 n.3, p.7-20, 2004.

[15] MEYER, Paul L, "Probabilidade :Aplicações à Estatística", 2.ed. São Paulo: LTC, 1995.

[16] MILLER, R. G. Survival analysis. New Work: John Wiley, 1981. 238p.

[17] PEREIRA, N. E. 2008, "Distribuição de Pressão e Condutância Térmica em Contatos de Superfícies em Junções Aparafusadas". 2008. 138 f. Dissertação de Mestrado, Universidade Federal de Santa Catarina,Florianópolis - SC, Brasil, 106p.

[18] Reliasoft@Brasil, 2005, "Características da Distribuição de Weibull", http://www.reliasoft.com.br/hotwire/edicao3/conceito3.htm, Acesso em 10/05/2007.

[19] ROSS, S. Probabilidade: um curso moderno com aplicações. Tradutor: Alberto Resende De Conti. - 8 ed. Porto Alegre, Bookman, 2010.

[20] SANGIOLO, A. C. Distribuição de Extremos de Precipitação diária, Temperatura máxima e Mínima e Velocidade do Vento em Piracicaba , SP (1917-2006), Revista Brasileira de Meteorologia, v.23, n.3, 341-346, 2008.

[21] SILVA, C. J.; HELDWEIN, B. A.; MARTINS, B. F.; TRENTIN, G.; GRIMM,L. E. Análise de Distribuição de Chuva de Santa Maria, Rev. Bras. de Eng. Agríc. e Ambient. Campina Grande, vol.11 nº1, Jan./Feb. 2007.

[22] WERNER, L. 1996, "Modelagem de dados de falhas ao longo do calendário". Dissertação de Mestrado, Universidade Federal do Rio Grande do Sul, Porto Alegre - RS, Brasil, 107p.

[23] http://www.bdpa.cnptia.embrapa.br.