



Universidade Estadual da Paraíba
Centro de Ciências e Tecnologia
Departamento de Estatística

Elias Silva de Medeiros

Distribuição generalizada de valores extremos aplicada a dados de precipitação máxima na região de Moreilândia-PE

Campina Grande
Dezembro - 2011

Elias Silva de Medeiros

Distribuição generalizada de valores extremos aplicada a dados de precipitação máxima na região de Moreilândia-PE

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Orientador:

Msc. Ricardo Alves de Olinda

Campina Grande

Dezembro - 2011

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL – UEPB

M488d

Medeiros, Elias Silva de.

Distribuição generalizada de valores extremos aplicada a dados de precipitação máxima na região de Moreilândia-PE [manuscrito] / Elias Silva de Medeiros. – 2011.

39 f.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) – Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2011.

“Orientação: Prof. Me. Ricardo Alves de Olinda, Departamento de Estatística”.

1. Estatística. 2. Distribuições assintóticas. 3. Valores extremos . I. Título.

21. ed. CDD 519.5

Elias Silva de Medeiros

Distribuição generalizada de valores extremos aplicada a dados de precipitação máxima na região de Moreilândia-PE

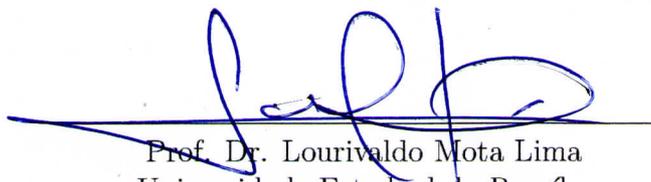
Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Aprovado em: 05/12/2017

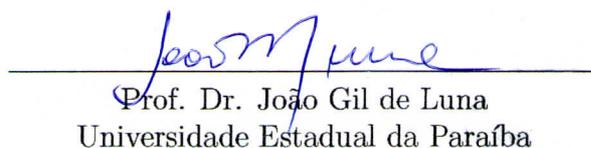
Banca Examinadora:



Prof. Msc. Ricardo Alves de Olinda
Orientador



Prof. Dr. Lourivaldo Mota Lima
Universidade Estadual da Paraíba



Prof. Dr. João Gil de Luna
Universidade Estadual da Paraíba

Dedicatória

Aos meus pais,

José Amauri de Medeiros e

Maria Joziene da Silva Medeiros

Ao meus irmãos,

Levi Medeiros e Elizeu Medeiros.

É com muito amor e gratidão, DEDICO.

Agradecimentos

A DEUS.

Ao Professor Ricardo Alves de Olinda, pela orientação e incentivo, colaborando de maneira extraordinária para as minhas conquistas acadêmicas.

Aos docentes do Curso de Bacharelado em Estatística da Universidade Estadual da Paraíba, que auxiliaram na minha formação acadêmica.

Ao Professor Dr. Lourivaldo Mota Lima pelo fornecimento dos dados e sua disponibilidade em me ajudar.

Aos meus amigos e colegas do curso, Cleber, Erivaldo e Isaac, pelo convívio e amizade.

À querida Leide Pricila da Silva, pelo carinho e amizade durante toda a minha graduação.

A igreja a qual eu congrego e todos os irmãos que estiveram sempre orando pra que vinhesse a se cumprir as promessas de Deus em minha vida

DEUS É BOM!!!

Resumo

A base conceitual da teoria de valores extremos foi proposta por Fisher-Tippett(1928) que estabeleceram os três tipos possíveis de distribuições assintóticas de valores extremos, conhecidas como de Gumbel (tipo I), Fréchet (tipo II) e Weibull (tipo III). Este trabalho é composto por uma série histórica de precipitação máxima pluviométrica registrada durante 47 anos (1962 a 2008), no Estado de Pernambuco, precisamente no Município de Moreilândia, sendo este estudo baseado nos valores máximos de precipitação de chuvas nessa região, que se dá nos meses de dezembro, janeiro, fevereiro, março, abril e maio. O presente trabalho consiste no ajuste da distribuição generalizada de valores extremos (GEV), que inclui como casos particulares, as distribuições Gumbel, Fréchet e de Weibull. O método de máxima verossimilhança foi utilizado para obter as estimativas dos parâmetros da GEV. O teste de Kolmogorov-sminov aplicado para o ajuste dos dados, assim como gráficos de quantil-quantil. Através desta análise verifica-se que a distribuição Gumbel é a mais adequada para representar os dados de precipitação máxima de chuvas para os meses em estudo. Por último, foram obtidos os níveis de retorno para os períodos de retorno de 5, 10 e 50 anos, e construídos seus respectivos intervalos de 95% de confiança, através do método delta.

Palavras-chave: assintóticas, níveis de retorno, intervalos de confiança.

Abstract

The conceptual basis of extreme value theory was proposed by Fisher-Tippett (1928) who established the three possible types of asymptotic distributions of extreme values, known as the Gumbel (type I), Fréchet (type II) and Weibull (type III). This work consists of a series of maximum precipitation rainfall recorded during 47 years (1962 to 2008), in Pernambuco State, precisely in the City of Moreilandia, this study maximum values was based on the rainfall in this region, which occurs in the months of december, january, february, march, april and may. This work consists of adjusting the generalized extreme value distribution (GEV), which includes as particular cases , the distributions Gumbel, Fréchet and Weibull. The maximum likelihood method was used to obtain the estimates of the parameters of GEV. The Kolmogorov-smminorv test applied to fit the data, as well as quantile-quantile plots. Through this analysis there is that the Gumbel distribution is more appropriate to represent the data of maximum precipitation of rain for the months under study. Finally, we obtained the return levels for return periods of 5, 10 and 50 years, and built their respective ranges of 95% confidence, the Delta Method.

Keywords: asymptotic, levels of return, confidence intervals.

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 10
2	Desenvolvimento da teoria	p. 12
2.1	Teoria dos valores extremos	p. 13
2.2	Teste de aleatoriedade	p. 17
2.3	Método de estimação	p. 18
2.3.1	Método de máxima verossimilhança	p. 18
2.4	Seleção da distribuição de valores extremos	p. 21
2.5	Diagnóstico do ajuste da GEV	p. 22
2.6	Período de retorno e nível de retorno	p. 24
2.7	Intervalos de confiança para os níveis de retorno	p. 25
3	Resultados e discussão	p. 28
4	Considerações finais	p. 35
	Referências	p. 36
	Anexo A	p. 39

Lista de Figuras

1	Função de densidade de probabilidade (a) e função de distribuição acumulada (b) da distribuição generalizada de valores extremos para $\xi = -0,30$ (<i>Weibull</i>), $\xi \rightarrow 0,00$ (<i>Gumbel</i>) e $\xi = 0,30$ (<i>Fréchet</i>), com $\mu = 10$ e $\sigma = 2,6$	p. 16
2	Gráficos de caixa (<i>box-plot</i>) para a variável precipitação máxima de chuva para cada um dos meses observados.	p. 29
3	Gráficos de probabilidade-probabilidade para diagnóstico da distribuição Gumbel aos dados de precipitação máxima mensal nos meses, dezembro, janeiro, fevereiro, março, abril e maio.	p. 32
4	Gráficos de quantil-quantil para diagnóstico da distribuição Gumbel aos dados de precipitação máxima mensal nos meses, dezembro, janeiro, fevereiro, março, abril e maio.	p. 33
5	Gráficos do teste de Kolgomorov-Kminorv da função de distribuição acumulada empírica (na cor cinza) e teórica (na cor vermelha) para diagnóstico da distribuição Gumbel aos dados de precipitação máxima mensal nos meses, dezembro, janeiro, fevereiro, março, abril e maio. . .	p. 34

Lista de Tabelas

1	Estatística descritiva da variável aleatória precipitação máxima mensal de chuvas nos períodos de 1962 a 2008, em Moreilandia-Pe.	p. 28
2	Teste de chorrilho sob a pressuposição de independência dos dados nos meses em estudo, determinando, respectivamente, o valor-p destes meses.	p. 29
3	Estimativas dos parâmetros da distribuição generalizada de valores extremos e respectivas variâncias e covariâncias estimadas, para os meses em análise.	p. 30
4	Intervalos de 95% de confiança para o parâmetro de forma (ξ) e valores da estatística de razão de verossimilhança modificado (T_{LR}^*).	p. 30
5	Estimativas dos parâmetros μ e σ da distribuição Gumbel e respectivas variâncias e covariâncias estimadas, para os meses em análises.	p. 31
6	Resultados do teste de Kolmogorov-Smirnov, e seus respectivos valor-p, para verificação da qualidade do ajuste da distribuição Gumbel aos dados de precipitação máxima mensal nos meses, dezembro, janeiro, fevereiro, março, abril e maio.	p. 31
7	Probabilidades de ocorrência de precipitações máximas mensais de chuvas acima de 50, 75, 100 e 125 mm, para os 6 meses do ano, Moreilândia, Pe. Dados de precipitação máxima mensal nos meses, dezembro, janeiro, fevereiro, março, abril e maio.	p. 31
8	Níveis de retorno (\hat{x}_p - mm) estimados e limites inferior (LI) e superior (LS) de seus respectivos intervalos de 95% de confiança, para os períodos de retorno 5, 10 e 50 anos, obtidos através do método delta.	p. 33
9	Precipitações pluviais máximas mensal (mm) na cidade de Moreilandia-Pe, nos períodos de 1962 a 2008	p. 39

1 Introdução

A precipitação máxima pluviométrica é uma das variáveis meteorológicas mais importantes quando se avalia estudos climáticos para os setores produtivos da sociedade tanto econômicos quanto social (agricultura, transporte, hidrologia, por exemplo), causando-se enchentes, assoreamento dos rios, quedas de barreiras, etc. Dentre os muitos aspectos apresentados pela Região Nordeste, o que mais se destaca é a seca, causada pela escassez de chuvas, o qual provoca problemas econômicos e sociais. De acordo com Freitas (1997), o Sertão nordestino apresenta as menores incidências de chuvas, isso em âmbito nacional. No Sertão, as chuvas se apresentam entre dezembro e maio, no entanto, em determinados anos isso não acontece, ocasionando-se um longo período sem chuvas, originando assim, a seca. A longa estiagem provoca uma série de prejuízos aos agricultores, como perda de plantações e animais sendo que a falta de produtividade causada pela seca provoca a fome.

Segundo Beltrão (2005), o Polígono das Secas apresenta um regime pluviométrico marcado por extrema irregularidade de chuvas, no tempo e no espaço. Neste cenário, a escassez de água constitui um forte entrave ao desenvolvimento socioeconômico e, até mesmo, à “subsistência” da população. A ocorrência cíclica das secas e seus efeitos catastróficos são por demais conhecidos e remontam aos primórdios da história do Brasil. A uma dada série hidrológica histórica, tem como meta determinar uma distribuição de probabilidade capaz de representar e extrapolar medições para ocorrências futuras de um fenômeno relacionado com esta série, como por exemplo, precipitação máxima mensal de uma região ou até mesmo de uma cidade, a qual consiste este estudo. Uma maneira de modelar estes eventos é através da teoria dos valores extremos proposta por Fisher e Tippett (1928). De acordo com esta teoria, existem três tipos de distribuições assintóticas de valores extremos. Em 1955, Jenkinson propôs, a distribuição generalizada de valores extremos (GEV), considerada como uma família de distribuições, pois esta é capaz de representar os três tipos de distribuições assintóticas de valores extremos, como caso particulares.

Através da evolução tecnológica digital e algoritmos computacionais, o estudo envolvendo valores extremos começou a ser mais bem investigado, principalmente, nesta última década. Os estudos envolvendo extremos de precipitação pluviométrica vêm sendo cada vez mais debatido entre os órgãos representantes.

O presente trabalho foi desenvolvido com o objetivo principal de apresentar e implementar a metodologia para ajustar a distribuição generalizada de valores extremos aos dados de precipitação máxima mensal de chuvas em Moreilândia, Pernambuco, visando assim, investigar qual dos três tipos de distribuição melhor se ajusta aos dados. Este trabalho está organizado da seguinte forma, o Capítulo 2 traz o marco histórico, citando os principais artigos que utilizaram o método de valores extremos, e abordando a metodologia utilizada no capítulo 3 são apresentados os resultados e discussões, por fim, no Capítulo 4 são apresentadas as considerações finais.

2 Desenvolvimento da teoria

De acordo com Kotz e Nadarajah (2000), Nicolas Bernoulli, em 1700, já se deparava com um problema envolvendo valores extremos, ao tratar da média da distância máxima à origem, de pontos aleatoriamente posicionados em uma linha reta de tamanho fixo. A teoria de valores extremos teve sua origem, principalmente, a partir das necessidades dos astrônomos em rejeitar ou não observações remotas. Os primeiros artigos sobre este assunto foram datados por Fuller (1914) e Griffith (1920), nos quais discutiram as utilidades desta teoria tanto no campo das aplicação como também estudaram os métodos de análise matemática. Um desenvolvimento sistemático englobando esta teoria pode ser encontrado no artigo de Von Bortkiewicz (1922), que o referido autor abordou a distribuição de valores extremos em amostras aleatórias de uma distribuição normal. A importância do artigo de Bortkiewicz deve-se ao fato, de que o conceito dessa teoria tenha sido claramente introduzida pela primeira vez. No ano seguinte, von Mises (1923) calculou o valor esperado desta distribuição, e Dodd (1923) determinou a mediana. O referido autor discutiu algumas distribuições não-normais.

Em 1925, Tippett estudou a função de distribuição acumulada (*f.d.*), os momentos da estatística de ordem extrema em amostra de uma população normal. O artigo de maior relevância nesta época, sobre as distribuições assintóticas de valores extremos, foi o de Fréchet (1927). No ano seguinte, Fisher e Tippett (1928) publicaram os resultados de uma investigação independente sobre esse mesmo assunto, eles mostraram que as caudas da distribuição generalizada de valores extremos seguiam uma dos três tipos de distribuições: Gumbel (Tipo *I*), Fréchet (Tipo *II*) e Weibull (tipo *III*). Von Mises (1936) apresentou algumas condições simples e úteis para a convergência quase certa¹ da estatística de ordem superior para cada uma dos três tipos de distribuições citadas anteriormente por Fisher e Tippett (1928).

Os desenvolvimentos teóricos entre as décadas de 20 a 30 foram seguidos por uma

¹Convergência Quase-Certa: **Definição** Seja Y_n uma sequência de variáveis aleatórias. Diz que Y_n converge quase certamente para Y , e escreve-se $Y_n \xrightarrow{q.c.} Y$, se $P[w \in \Omega : Y_n(w) \xrightarrow{n} Y(w)] = 1$. Essa forma de convergência é a mais forte de todas.

série de artigos, com aplicações práticas da estatística de valores extremos, em problemas relacionados ao tempo de vida humano, as emissões radioativas Gumbel (1937a, b), entre outros. De acordo com Silva (2008), foi Gumbel em 1941 o primeiro a propor uma metodologia de análise estatística envolvendo a teoria dos valores extremos. Em seu artigo, foram estimados períodos de retornos dos dados de vazão máxima anual do rio Missipi em Vicksburg, Missipi, EUA e do rio Rhône em Lyon, França, por meio da distribuição de Gumbel.

Em 1943, outras contribuições importantes para o estudo de valores extremos foram dadas por Gnedenko², citado por Bautista (2002), que mostrou as condições necessárias e suficientes para a existência das distribuições assintóticas dos valores extremos e determinou que as caudas dessas distribuições, ou seja, a parte que trata dos valores máximos ou mínimos menos frequentes, podem ser modeladas por alguns tipos de distribuições contínuas. Um artigo importante, mas muito negligenciado por Juncosa (1949), Gnedenko estende os resultados para o caso em que as distribuições não sejam necessariamente variáveis aleatórias independentes e identicamente distribuídas. Apesar do forte valor teórico, os resultados de Juncosa não parecem ter muita utilidade. O fato de que as distribuições assintóticas possam ocorrer, em geral, não fornecem muitas indicações para aplicações práticas.

2.1 Teoria dos valores extremos

Segundo Bautista (2002), a teoria dos valores extremos desempenha um papel fundamental na modelagem de eventos associados a probabilidade muito pequenas ou eventos raros, por exemplo, na estimação de fenômenos climáticos, cálculos de seguros e casos pouco comuns no mercado financeiro. De acordo com Júnior (2010), os valores extremos trazem em si duas características: intensidade e frequência. São eventos intensos e representam os valores críticos de uma série, ao mesmo tempo estes valores críticos são muito pouco frequentes, ressaltando o aspecto caudal dos valores extremos. De maneira simbólica, é como ver através de uma lupa a calda da distribuição de uma dada variável aleatória. Segundo Sansigolo (2008), eventos raros ou extremos têm grande relevância na climatologia e hidrologia, e suas estimativas probabilísticas são imprescindíveis para o planejamento e desenvolvimento das atividades sujeitas a seus efeitos adversos. Os fundamentos desta teoria foram expressos por Fisher-Tippett (1928) que estabeleceram

²GNEDENKO, B. (1943). Sur la distribution limite du terme maximum d'une serie alatoire, *Ann. Math.* 44, 423-453. Translated and reprinted in: *Breakthroughs in Statistics*, Vol. I, 1992, eds. S. Kotz and N. L. Johnson, Springer-Verlag, pp. 195-225.

os três tipos possíveis de distribuições assintóticas de valores extremos, conhecidas como de Gumbel (tipo I), Fréchet (tipo II) e Weibull (tipo III). Por exemplo, as caudas da distribuição Weibull seguem uma distribuição uniforme, já as da Gumbel expressam uma distribuição exponencial, gama, normal ou log-normal, e as da distribuição Fréchet seguem uma distribuição de Cauchy, Pareto ou t de Student.

Seja X_1, X_2, \dots uma sequência de variáveis aleatórias independentes e identicamente distribuídas (*i.i.d.*) tendo função de distribuição F_X , e consideremos M_n como sendo o máximo das primeiras n variáveis, isto é, $M_n = \max(X_1, \dots, X_n)$.

A função de distribuição de variável aleatória $M_n, F_{M_n}(x)$, é dada por:

$$F_{M_n}(x) = \Pr(M_n \leq x) = \Pr(X_1 \leq x, \dots, X_n \leq x) = (F_X(x))^n$$

De acordo com Viola (2006), para n muito grande, a função de distribuição de M_n pode ser degenerada. Com isso, seria de grande utilidade algum resultado assintótico para o máximo. Segundo Coles (2001), a teoria de Valores Extremos assegura a existência de uma distribuição assintótica não-degenerada, F , para uma transformação linear de M_n , isto é, para sequências numéricas apropriadas $a_n > 0$ e $b_n \in \mathfrak{R}$

Teorema 2.1 (*Fisher-Tippett 1928*). *Seja X_1, X_2, \dots, X_n uma sequência de variáveis aleatórias e identicamente distribuídas. Se existem sequências de constantes normalizadas $a_n > 0$, $b_n \in \mathfrak{R}$ e uma função de distribuição não-degenerada F tal que,*

$$P \left\{ \frac{M_n - b_n}{a_n} \leq x \right\} \rightarrow F_1(x)$$

então $G_1(x)$ pertence a algum dos três tipos de funções de valores extremos:

$$\begin{aligned} I : F_1(x) &= \exp \left\{ \exp \left[- \left(\frac{x - \mu}{\sigma} \right) \right] \right\}, \quad -\infty < x < \infty \\ II : F_1(x) &= \begin{cases} 0, & x \leq \mu \\ \exp \left\{ - \left(\frac{x - \mu}{\sigma} \right)^{-\xi} \right\}, & x > \mu \end{cases} \\ III : F_1(x) &= \begin{cases} \exp \left\{ - \left[- \left(\frac{x - \mu}{\sigma} \right)^{-\xi} \right] \right\}, & x < \mu \\ 1, & x \geq \mu \end{cases} \end{aligned}$$

para $\sigma > 0$, $\mu \in \mathfrak{R}$, e no caso das famílias II e III, $\xi > 0$. Sendo denotados respectivamente como, I: Gumbel, II: Fréchet e III: Weibull.

Segundo Ferrari (2011), uma reformulação no Teorema de Fisher-Tippet combinou

os três tipos de distribuição (*I*, *II*, *III*) em uma única família, de um único parâmetro ξ , definida como distribuição generalizada de valores extremos que tem função de distribuição acumulada de probabilidade dada por:

$$F_X(x) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \quad (2.1)$$

definida no conjunto $\left\{ x : x > \mu - \frac{\sigma}{\xi} \right\}$, em que os parâmetros satisfazem $-\infty < \mu < \infty$, $\sigma > 0$ e $-\infty < \xi < \infty$. Esta é a distribuição generalizada de valores extremos (GEV). Em que μ é o parâmetro de locação; σ , é o parâmetro de escala; e ξ , o parâmetro de forma. Nos casos, em que, $\xi = 0$, $\xi > 0$ e $\xi < 0$, correspondem, respectivamente, as famílias de distribuição do Tipo I, Tipo II e do Tipo III.

Para o caso particular em que $\xi \rightarrow 0$, a função de distribuição acumulada de Gumbel é escrita da forma:

$$F_X(x) = \exp \left[- \exp \left(- \frac{x - \mu}{\sigma} \right) \right] \quad (2.2)$$

definida em $-\infty < x < \infty$, em que μ é o parâmetro de locação, e σ , o parâmetro de escala, com $\mu \in \mathfrak{R}$ e $\sigma > 0$.

Segundo Coles (2001), a unificação das três famílias originais das distribuições de valores extremos, simplifica bastante a implementação estatística. Ao se fazer inferência sobre ξ é possível verificar e determinar qual o tipo mais regular de comportamento da distribuição, com isso, não há necessidade de fazer levantamentos a priori sobre qual tipo de família adotar para os valores extremos.

Se a função de distribuição F é diferenciável, então pode-se definir a função densidade de probabilidade de X como

$$f_X(x) = \frac{\partial F_X(x)}{\partial x},$$

neste caso

$$F_X(x) = \int_{-\infty}^x f_X(x) \partial x$$

A função de densidade de probabilidade da Distribuição Generalizada de Valores Extremos é dada a seguir.

$$f_X(x) = - \left\{ - \frac{1}{\xi} \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\xi}{\xi}\right)} \frac{\xi}{\sigma} \right\} \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}$$

daí,

$$f_X(x) = \frac{1}{\sigma} \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\xi}{\xi}\right)} \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}. \quad (2.3)$$

Derivando-se a expressão 2.2 em relação a x , no caso particular em que $\xi \rightarrow 0$, obtém-se a função densidade de probabilidade *Gumbel*, dada por:

$$f_X(x) = \frac{1}{\sigma} \left\{ \exp \left(-\frac{x - \mu}{\sigma} \right) \exp \left[-\exp \left(-\frac{x - \mu}{\sigma} \right) \right] \right\}, \quad (2.4)$$

definida em $-\infty < x < \infty$ para $\mu \in \mathfrak{R}$ e $\sigma > 0$

De acordo com Ferrari (2011), as características e propriedades das distribuições de extremos são determinadas pelas caudas extremas (inferior e superior) da distribuição. Para visualizar o comportamento dos três casos particulares, foram construídos os gráficos da função densidade de probabilidade e da função acumulada de probabilidade.

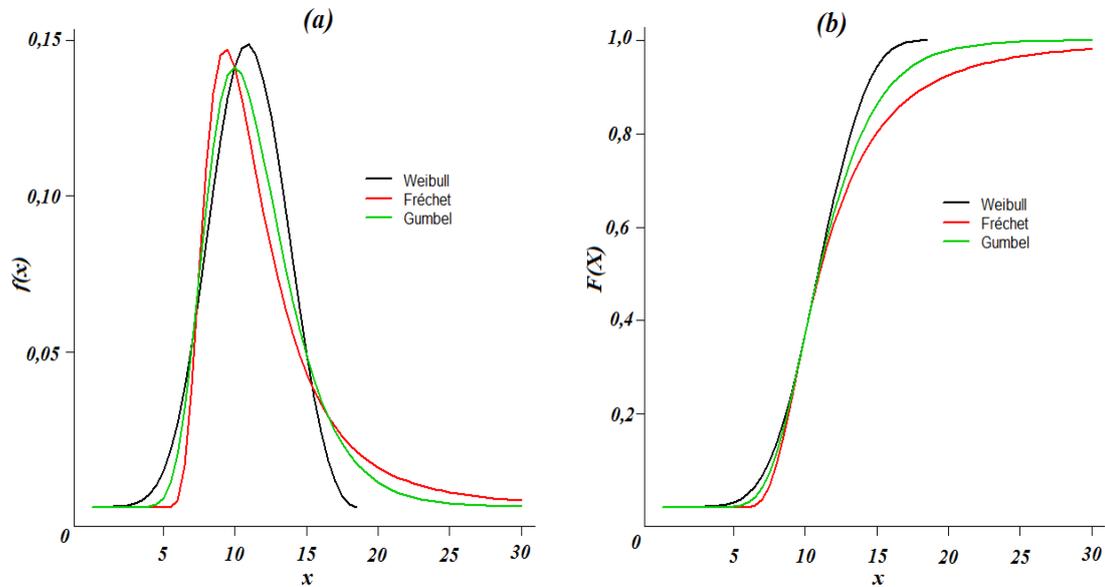


Figura 1: Função de densidade de probabilidade (a) e função de distribuição acumulada (b) da distribuição generalizada de valores extremos para $\xi = -0,30$ (*Weibull*), $\xi \rightarrow 0,00$ (*Gumbel*) e $\xi = 0,30$ (*Fréchet*), com $\mu = 10$ e $\sigma = 2,6$

Por meio da Figura 1 e do suporte da função (2.3) observa-se, ainda, que a distribuição Fréchet corresponde a um modelo com cauda inferior ($x > 1,33$) e cauda superior infinita. Por outro lado, no caso da distribuição Weibull, a cauda superior é finita ($x < 18,67$).

2.2 Teste de aleatoriedade

Conforme Bautista et al. (2004), a primeira etapa da análise é verificar a hipótese de independência dos dados observados por meio do teste de chorrilho (“run test”) descrito por Zar (1999). De acordo com Silva (2008), esse teste consiste em, inicialmente, definir uma sequência dicotômica de tamanho n , a partir da amostra aleatória X_1, \dots, X_n aplicando-se a cada X_i a função indicadora $A(x_i) = I_{X_i > M_d}(X_i)$ e omitindo-se os casos em que $X_i = M_d$ sendo M_d a mediana dos dados.

O valor da mediana, M_d , é dado por:

$$M_d = \begin{cases} x_{(\frac{n+1}{2})}, & \text{se } n \text{ for ímpar,} \\ \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}, & \text{se } n \text{ for par,} \end{cases}$$

Já a variável indicadora A_i , com $(i = 1, 2, \dots, n)$, é definida por:

$$A_i = \begin{cases} 1, & \text{se } x_i > M_d \\ 0, & \text{se } x_i < M_d \end{cases}$$

Segundo Bautista (2002), a variável aleatória número total de sequências de zeros e uns ao longo da amostra R e, r seu valor observado, define-se as variáveis aleatórias N_1 como sendo o número total de ocorrências de $X_i > M_d$ e N_2 como sendo o número total de ocorrências de $X_i < M_d$, cujos valores observados são, respectivamente, n_1 e n_2 .

Conforme os procedimentos adotados por Silva (2008), para $n_1 < 30$ e $n_2 < 30$, Zar (1999) apresenta os pares de valores críticos exatos $(r_{1,\alpha,n_1,n_2}; r_{2,\alpha,n_1,n_2})$ ao nível de significância α , usando-se o procedimento descrito por Eisenhat e Swed (1943) e Browlee (1965). Assim, rejeita-se a hipótese nula se $r \leq r_{1,\alpha,n_1,n_2}$, ou se $r \leq r_{2,\alpha,n_1,n_2}$. Caso $n_1 \geq 30$ ou $n_2 \geq 30$, sob a hipótese H_0 de independência tem-se que, assintoticamente R segue uma distribuição normal com esperança dada por

$$E(R) = \frac{2N_1N_2}{N} + 1,$$

e variância é dada por

$$Var(R) = \frac{2N_1N_2(2N_1N_2 - n)}{n^2(n - 1)}.$$

Cuja estimativa são dadas por

$$\hat{E}(R) = \frac{2n_1n_2}{n} + 1$$

e

$$\hat{V}ar(R) = \frac{2n_1n_2(2n_1n_2 - n)}{n^2(n - 1)}$$

em que n_1 e n_2 são os valores observados de N_1 e N_2 .

Para efetuar o teste deve-se calcular a mediana da amostra observada de precipitação máxima de chuva x_1, \dots, x_n , obter uma sequência dicotômica dessa mesma amostra $A(x_1), \dots, A(x_n)$, $n_1, n_2, r, \hat{E}(R)$ e $\hat{V}ar(R)$ e, por fim, calcular o valor-p por meio de

$$P \left(\frac{|r - \hat{E}(R)| - 0,5}{\sqrt{\hat{V}ar(R)}} \geq q_r \right)$$

em que, q_r é o quantil de ordem $\alpha/2$ da normal padrão e, α é o nível de significância adotado para o teste. Uma vez testada a independência das observações seguiu-se à estimação dos parâmetros.

2.3 Método de estimação

Este é o meio pelo qual os parâmetros desconhecidos do modelo são inferidos com base em dados históricos. Embora diferentes abordagens tenham sido propostas para as estimativas dos modelos de valor extremo, adota-se uma visão singular que restringi à atenção para técnicas baseadas na função de verossimilhança. Segundo Coles (2001), todas as técnicas de estimação têm os seus prós e contras, mas a probabilidade baseada em técnicas são únicas em sua capacidade de adaptação ao modelo de mudança. Neste trabalho adota-se o método da **máxima verossimilhança**, que tem um conjunto de propriedades convenientes na inferência estatística.

2.3.1 Método de máxima verossimilhança

De acordo com Viola (2006), dada uma distribuição de probabilidade cujos parâmetros são desconhecidos, tem-se como objetivo, a partir de uma amostra de uma determinada população, estimar o valor do parâmetro da distribuição de probabilidade que representa

toda a população. Seja $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ uma amostra aleatória *i.i.d.*, com distribuição de probabilidade $p(x | \boldsymbol{\theta})$, em que $\boldsymbol{\theta}$ é o vetor de parâmetros a ser estimado. Deseja-se dar uma estimativa do valor de $\boldsymbol{\theta}$ ou de uma função $p(\boldsymbol{\theta})$ a partir de valores observados x_1, x_2, \dots, x_n . O Método de Máxima Verossimilhança, por ser um método de estimação pontual, nos fornecerá um valor numérico de uma estatística $T(\mathbf{X})$ para estimar $p(\boldsymbol{\theta})$.

Ainda em Viola (2006) dada uma amostra $\underline{x} = (x_1, x_2, \dots, x_n)'$, o método de Máxima Verossimilhança consiste em maximizar a função de verossimilhança definida por $L(\boldsymbol{\theta} | \underline{x}) = p(\underline{x} | \boldsymbol{\theta}) = p(x_1 | \boldsymbol{\theta}) \times p(x_2 | \boldsymbol{\theta}) \times \dots \times p(x_n | \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$ em que Θ representa o espaço paramétrico de $\boldsymbol{\theta}$, isto é, o método consiste em encontrar $\hat{\boldsymbol{\theta}}$ tal que $\hat{\boldsymbol{\theta}} = \operatorname{argmax}_{\hat{\boldsymbol{\theta}} \in \Theta} L(\boldsymbol{\theta} | \underline{x})$, em que $\hat{\boldsymbol{\theta}}$ é o vetor de estimativas de $\boldsymbol{\theta}$ e $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(X)$ é o vetor dos estimadores de máxima verossimilhança de $\boldsymbol{\theta}$. Como o valor de $\boldsymbol{\theta}$ que maximiza $L(\boldsymbol{\theta} | \underline{x})$ é o mesmo que maximiza $l(\boldsymbol{\theta} | \underline{x}) = \ln(L(\boldsymbol{\theta} | \underline{x}))$ e $l(\boldsymbol{\theta} | \underline{x})$ sempre está definida pois $L(\boldsymbol{\theta} | \underline{x})$ é positiva (produto de funções de probabilidade) então, por simplicidade, maximiza-se $l(\boldsymbol{\theta} | \underline{x})$.

O Método da Máxima Verossimilhança pode apresentar algumas dificuldades práticas, tais como:

1. O cálculo de derivadas é muito complicado;
2. A resolução $\frac{\partial}{\partial \boldsymbol{\theta}} L(\boldsymbol{\theta} | \underline{x}) = 0$ não possui solução explícita;
3. $L(\boldsymbol{\theta} | \underline{x})$ não possui ponto de máximo ou $L(\boldsymbol{\theta} | \underline{x})$ possui máximos locais.

Segundo Ferrari (2011), muitas técnicas têm sido propostas para fazer inferências sobre os parâmetros da distribuição GEV. Estas incluem técnicas gráficas baseadas em gráficos de probabilidade, estimadores baseados no método dos momentos, método de regressão, método do L-momentos e no método da máxima verossimilhança. Em casos normais, os estimadores de máxima verossimilhança são consistentes, assintoticamente normal e eficientes. Já os casos não regulares, acontecem quando a distribuição em estudo depende de parâmetros desconhecidos. Smith (1985) estudou cuidadosamente o comportamento assintótico dos estimadores de máxima verossimilhança para a distribuição GEV e obteve os seguintes resultados, para o parâmetro de forma:

1. quando $\xi > -0,5$, os estimadores de máxima verossimilhança são regulares no sentido de ter as propriedades assintóticas habituais;
2. quando $-1 < \xi < -0,5$, os estimadores de máxima verossimilhança existem mas não são regulares;

3. quando $\xi < -1$, esses estimadores provavelmente não existem.

O caso em que $\xi < -0,5$ corresponde a uma distribuição com uma cauda superior muito curta e leve e, segundo Smith (1985), essa situação raramente é encontrada em aplicações de modelagem de valores extremos sendo que, as limitações teóricas da abordagem de máxima verossimilhança geralmente não são obstáculos na prática.

Supondo que X_1, X_2, \dots, X_n são variáveis aleatórias *i.i.d.*. Os estimadores são apresentados da seguinte forma:

A verossimilhança é dada por:

$$L(\mu, \sigma, \xi | \underline{x}) = \prod_{i=1}^n f_X(x_i | \mu, \sigma, \xi) = \sigma^{-n} \prod_{i=1}^n \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\xi}{\xi}\right)} \exp \left\{ - \sum_{i=1}^n \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \quad (2.5)$$

Calcula-se a log-verossimilhança em 2.5

$$T = \ln [L(\mu, \sigma, \xi | \underline{x})] = -n \ln(\sigma) - \left(\frac{1+\xi}{\xi} \right) \sum_{i=1}^n \left\{ \ln \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right] \right\} - \sum_{i=1}^n \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \quad (2.6)$$

para $\xi < 0$, e $x_i < \mu - \frac{\sigma}{\xi}$ (ou seja, $\mu - \frac{\sigma}{\xi} > x_{(n)}$) ou para $\xi > 0$, e $x_i > \mu - \frac{\sigma}{\xi}$ (ou seja, $\mu - \frac{\sigma}{\xi} < x_{(1)}$)

Os estimadores de máxima verossimilhança de μ , σ , e ξ são obtidos derivando a Equação 2.6 com relação ao parâmetro de interesse, obtendo-se assim, o seguinte sistema de Equações não-lineares,

$$\begin{aligned} \frac{\partial T}{\partial \mu} &= \frac{1}{\hat{\sigma}} \sum_{i=1}^n \left(\frac{1 + \hat{\xi} - w_i^{-\frac{1}{\hat{\xi}}}}{w_i} \right) = 0 \\ \frac{\partial T}{\partial \sigma} &= -\frac{n}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^2} \sum_{i=1}^n \left\{ \frac{\left(1 + \hat{\xi} - w_i^{-\frac{1}{\hat{\xi}}} \right) (x_i - \hat{\mu})}{w_i} \right\} = 0 \\ \frac{\partial T}{\partial \xi} &= \sum_{i=1}^n \left\{ \left(1 - w_i^{-\frac{1}{\hat{\xi}}} \right) \left[\frac{\ln(w_i)}{\hat{\xi}^2} - \left(\frac{x_i - \hat{\mu}}{\hat{\xi} \hat{\sigma} w_i} \right) \right] - \frac{x_i - \hat{\mu}}{\hat{\sigma} w_i} \right\} = 0 \end{aligned} \quad (2.7)$$

em que $w_i = 1 + \hat{\xi} \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right)$.

A função de máxima verossimilhança escrita em 2.4, para o caso particular da distribuição *Gumbel* em que $\xi \rightarrow 0$ é calculada da seguinte forma:

$$L(\mu, \sigma, | x) = \prod_{i=1}^n f_X(x_i | \mu, \sigma) = \sigma^{-n} \exp \left\{ \sum_{i=1}^n \left(-\frac{x_i - \mu}{\sigma} \right) \right\} \exp \left\{ \sum_{i=1}^n \exp \left(-\frac{x_i - \mu}{\sigma} \right) \right\}$$

$$T = \ln(L(\mu, \sigma, | x)) = \sum_{i=1}^n \left\{ -\ln(\sigma) - \left(\frac{x_i - \mu}{\sigma} \right) - \exp \left(-\frac{x_i - \mu}{\sigma} \right) \right\} \quad (2.8)$$

Maximizando-se a expressão 2.8, os estimadores para os parâmetros μ e σ através do método de máxima verossimilhança são obtidos pelo seguinte sistema de equações não-lineares,

$$\frac{\partial T}{\partial \mu} = \frac{1}{\hat{\sigma}} \left\{ \left[\sum_{i=1}^n \exp \left(-\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) \right] - n \right\} = 0,$$

$$\frac{\partial T}{\partial \sigma} = -\frac{n}{\hat{\sigma}} + \sum_{i=1}^n \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}^2} \right) \left[1 - \exp \left(-\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) \right] = 0 \quad (2.9)$$

De acordo com Ferrari (2011), os sistemas de equações 2.8 e 2.9, em geral, não possuem soluções exatas pois são equações não-lineares. Uma solução aproximada é calculada pelo método iterativo de quasi-Newton que, para iniciar o algoritmo, especifica uma estimativa inicial para μ , σ e ξ . Neste trabalho, o software R é utilizado para calcular as estimativas de máxima verossimilhança por meio do pacote *evd* satisfazendo o critério de convergência do método.

2.4 Seleção da distribuição de valores extremos

Segundo Hoskink (1984), um dos procedimentos para testar se as observações seguem uma distribuição de valores extremos tipo *I*, *II* ou *III*, basta testar se ξ é igual a zero na distribuição GEV, o que pode ser feito através do teste da razão de verossimilhança modificado, descrito a seguir.

Tomando uma série de n observações (x_1, x_2, \dots, x_n) , $l(\hat{\theta}_{GEV})$ e $l(\hat{\theta}_G)$, os máximos do logaritmo da função de máxima verossimilhança das distribuições GEV 2.1 e de Gumbel 2.2 em que, $\hat{\theta}_{GEV} = (\hat{\mu}, \hat{\sigma}, \hat{\xi})'$ e $\hat{\theta}_G = (\hat{\mu}, \hat{\sigma})'$ são os vetores de estimativas de máxima verossimilhança.

A estatística de razão de verossimilhança (T_{LR}) é dado por:

$$T_{LR} = -2 \left[l(\hat{\theta}_G) - l(\hat{\theta}_{GEV}) \right] = 2 \left[l(\hat{\theta}_{GEV}) - l(\hat{\theta}_G) \right] \quad (2.10)$$

que tem distribuição assintótica χ^2 com 1 grau de liberdade.

Hosking (1984) sugere a utilização da estatística modificada cujo objetivo é de atender uma melhor aproximação à distribuição assintótica para 2.10:

$$T_{LR}^* = \left(1 - \frac{2,8}{n} \right) \times T_{LR}, \quad (2.11)$$

sendo n o tamanho da amostra.

Para testar a hipótese $H_0 : \xi = 0$ versus $H_1 : \xi \neq 0$, basta comparar o valor da estatística do teste T_{LR}^* com o valor tabelado da distribuição χ^2 com 1 grau de liberdade e um certo nível de significância (α), $\chi_{[\alpha,1]}^2$. Se $T_{LR}^* \geq \chi_{[\alpha,1]}^2$, rejeitá-se H_0 , ou seja, há fortes evidências de que as observações não são de uma distribuição do tipo I (*Gumbel*).

2.5 Diagnóstico do ajuste da GEV

Ajustes de distribuições são geralmente validadas por teste de aderência como por exemplo, Kolmogorov-Smirnov e qui-quadrado. O uso desta metodologia pode ser verificada em diversos autores como Bautista (2002), Beijo (2003), Sansigolo (2008) e Ferrari (2011). O teste de qui-quadrado é baseado em diferenças entre frequências observadas e frequências esperadas. Segundo Júnior (2010), diversas críticas são feitas a este teste no que diz respeito à natureza da hipótese a ser testada. Ele afirma ainda que, o número de classes bem como o tamanho da classe influenciam diretamente no poder desse teste. Uma alternatina no teste de qualidade de ajuste para distribuição de frequência é o método de Kolmogorov.

De acordo com Ferrari (2011), ao se ajustar uma distribuição de probabilidade a um conjunto de dados, trabalha-se com a hipótese de que a distribuição representa adequadamente aquele conjunto de informações. Seja X_1, X_2, \dots, X_n a série de dados observados ordenados de forma crescente, a função de distribuição acumulada assumida para os dados é definida por $F(x_{(i)})$ e a função de distribuição acumulada empírica de X é descrita da seguinte maneira,

$$\hat{F}(x_{(i)}) = \frac{i}{n+1}, \quad i = 1, 2, \dots, n. \quad (2.12)$$

Para se testar a suposição de que os dados seguem a distribuição GEV selecionada, pode-se utilizar a estatística D do teste Kolmogorov-Smirnov. Para uma população de função de densidade acumulada F , espera-se que F_n esteja bem próximo de F mas quão próxima deve estar é uma grandeza que merece atenção (Júnior, 2010). Tem-se então duas funções de x e defini-se como a distância entre elas como a maior das distâncias, a qual é conhecida como, a estatística D do **teste de Kolmogorov-Smirnov**:

$$D = \max | F(x_{(i)}) - \hat{F}(x_{(i)}) |, \quad i = 1, 2, \dots, n, \quad (2.13)$$

em que, $F(x_{(i)})$ é a distribuição teórica GEV com as estimativas obtidas e $\hat{F}(x_{(i)})$ é a distribuição empírica pela função 2.12. Rejeita-se a hipótese (H_0) de que os dados seguem uma distribuição GEV se a estatística do teste $D \geq D_{n,\alpha}$, em que $D_{n,\alpha}$ é o valor crítico para os valores de n a um nível de significância predeterminado.

De acordo com Júnior (2010), assumindo-se que X_i é uma amostra *i.i.d.* de uma distribuição contínua com função de distribuição acumulada F , e toma-se F_n a partir de uma distribuição empírica, então a significância do teste é dada por:

$$\begin{cases} H_0 : F(x) = F_n(x) \\ H_1 : F(x) \neq F_n(x). \end{cases} \quad (2.14)$$

Além do teste estatístico, o ajuste da distribuição pode ser avaliado graficamente por meio da construção dos gráficos *qq-plot* (gráfico quantil-quantil) e *pp-plot* (gráfico de probabilidade-probabilidade).

O gráfico *pp-plot* é formado com pontos dados pelas coordenadas,

$$\left[\hat{F}(x_{(i)}), F(x_{(i)}) \mid_{\theta=\hat{\theta}} \right], \quad i = 1, 2, \dots, n,$$

em que $\hat{\theta}$ são as estimativas de $\theta = (\mu, \sigma, \xi)'$, $F(x_{(i)})$ é a função de distribuição acumulada GEV em 2.1 com as estimativas obtidas e $\hat{F}(x_{(i)})$ é a distribuição empírica definida na expressão 2.12, se a função de distribuição GEV é um modelo razoável para a distribuição dos dados, os pontos estarão alinhados na reta pelos pontos (0,0) e (1,1), Assim, uma forma de interpretar o gráfico é observar o quão distantes esses pontos estão da reta. Quando mais distantes, menos adequada é a distribuição.

Segundo Bautista (2002), outra forma de avaliar graficamente o ajuste da distribuição é a utilização do gráfico *qq-plot*, formado pelos pontos de coordenadas:

$$\left[\hat{F}^{(-1)} \left(\frac{i}{n+1} \right), x_i \right], \quad i = 1, 2, \dots, n,$$

em que $\hat{F}^{(-1)}(\cdot)$ é a função inversa da expressão 2.12.

De acordo com Ferrari (2011), neste caso, também sob a hipótese de que os dados apresentam distribuição GEV, os pontos do gráfico estarão alinhados em uma reta. Quando mais afastados de uma reta, menos adequada é a distribuição. Os procedimentos utilizados por Bautista (2002), como forma de testar a hipótese (H_0) da distribuição dos dados seguir uma distribuição GEV, é utilizado o teste de Kolmogorov-Smirnov, utilizando-se os seguintes passos:

1. colocar a série de dados em ordem crescente,
2. obter os valores de probabilidade da distribuição teórica $F(x_{(i)})$ através da substituição das estimativas de máxima verossimilhança de μ , σ e ξ na equação (2.1) e os valores de probabilidade da distribuição empírica 2.12,
3. calcular a estatística D através da expressão 2.13,
4. rejeitar a hipótese (H_0) se $D \geq D_{n,\alpha}$, sendo $D_{n,\alpha}$ o valor crítico para valores de n e um determinado nível de significância (α). Os valores de $D_{n,\alpha}$ podem ser consultados na tabela construída por Massey (1951).

No caso de se desejar testar a hipótese da distribuição dos dados ser uma distribuição Gumbel, os valores de $F(x_{(i)})$, no passo (2), são obtidos através da substituição das estimativas de máxima verossimilhança de μ e σ em 2.2.

2.6 Período de retorno e nível de retorno

Considere um determinado experimento ε , e seja V um evento particular ocorrido em ε , e T o tempo aleatório entre ocorrências consecutivas do evento V sobre o experimento. O período de retorno do evento V é definido como sendo o valor médio τ da variável T . No caso em estudo, V é o evento: “precipitação máxima excede um determinado valor x ” cuja probabilidade $P(V)$ é dada por $1 - F_X(x)$, (Bautista, 2002).

Portanto, uma expressão que represente o período de retorno para esse evento é:

$$\tau = \frac{1}{P(V)} = \frac{1}{1 - F_X(x)},$$

sendo τ geralmente expresso em anos.

O nível de retorno (x_p), relacionado ao período de retorno τ , é calculado a partir da solução da equação:

$$\int_{i=1}^{x_p} f(\theta) \partial x = 1 - p$$

para $p = \frac{1}{\tau}$, ou seja,

$$F(x_p) = (1 - p) \quad (2.15)$$

Invertendo 2.15, obtem-se a solução:

$$x_p = F^{-1}(1 - p) = \mu - \frac{\sigma}{\xi} \left\{ 1 - [-\ln(1 - p)]^{-\xi} \right\}, \quad (2.16)$$

para $\xi \neq 0$, cujo limite para $\xi \rightarrow 0$ é dada por:

$$x_p = F^{-1}(1 - p) = \mu - \sigma \{ \ln[-\ln(1 - p)] \} \quad (2.17)$$

De acordo com Ferrari (2011), o nível x_p deverá ser excedido em média uma vez a cada $\frac{1}{p}$ anos. Mais precisamente, x_p é excedido pelo máximo anual em algum ano particular com probabilidade p . A estimativa de \hat{x}_p do nível de retorno x_p para períodos de retorno τ é obtida pela substituição das estimativas de máxima verossimilhança de μ , σ , e ξ em 2.16 e de μ , e σ em 2.17.

2.7 Intervalos de confiança para os níveis de retorno

Intervalos de confiança (I.C.) ao nível de $(1 - \alpha)100\%$ para os níveis de retorno x_p foram construídos, baseando-se no método delta e, em seguida, no método baseado na estatística de razão de verossimilhança. O intervalo de confiança para x_p com $(1 - \alpha)100\%$ de confiança é dado por:

$$I.C.(x_p) = \left[\hat{x}_p \pm z_{\alpha/2} \sqrt{Var(\hat{x}_p)} \right],$$

em que, α é o nível de significância, $z_{\alpha/2}$ o valor tal que $P(|Z| < z_{\alpha/2}) = 1 - \alpha$, e Z , uma variável com distribuição normal padronizada e $Var(\hat{x}_p)$ é a variância associada ao nível de retorno \hat{x}_p calculada a partir da aplicação do método Delta. Este método baseia-se no fato de que a distribuição de $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}, \hat{\xi})'$ ser assintoticamente normal com média $\boldsymbol{\theta} = (\mu, \sigma, \xi)'$ e matriz de variâncias e covariâncias dada por $\mathbf{I}(\boldsymbol{\theta})^{-1}$. Como a eq. 2.16 é uma função não linear de μ , σ e ξ , pode-se linearizá-la através da expansão de primeira ordem em série de Taylor em torno do ponto inicial correspondente ao vetor de estimativas dos parâmetros $\hat{\mu}$, $\hat{\sigma}$ e $\hat{\xi}$. De acordo com os procedimentos utilizados por MAZECHLI

et al. (2006), Ferrari (2011) e Bautista (2002), o método Delta (Rao e Tputenburg, 1999) calcula $Var(\hat{x}_p)$ a partir da matriz de variâncias-covariâncias de $\hat{\mu}$, $\hat{\sigma}$ e $\hat{\xi}$, estimada pela inversa da matriz de segundas derivadas da função log-verossimilhança (a matriz hessiana calculada localmente em $\hat{\mu}$, $\hat{\sigma}$ e $\hat{\xi}$).

O método Delta estima a variância de \hat{x}_p a partir da expressão

$$Var(\hat{x}_p) \approx \nabla x_p' V \nabla x_p \quad (2.18)$$

Sendo que, para o caso em que $\xi \neq 0$, J é uma matriz de variâncias e covariâncias de $\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\xi})$ obtidos da inversa da matriz de informação dada por

$$J = \begin{bmatrix} \frac{\partial^2}{\partial \mu \partial \mu} l(\theta) & \frac{\partial^2}{\partial \mu \partial \sigma} l(\theta) & \frac{\partial^2}{\partial \mu \partial \xi} l(\theta) \\ \frac{\partial^2}{\partial \sigma \partial \mu} l(\theta) & \frac{\partial^2}{\partial \sigma \partial \sigma} l(\theta) & \frac{\partial^2}{\partial \sigma \partial \xi} l(\theta) \\ \frac{\partial^2}{\partial \xi \partial \mu} l(\theta) & \frac{\partial^2}{\partial \xi \partial \sigma} l(\theta) & \frac{\partial^2}{\partial \xi \partial \xi} l(\theta) \end{bmatrix}_{\theta=\hat{\theta}}^{-1} = \begin{bmatrix} Var(\hat{\mu}) & Cov(\hat{\mu}, \hat{\sigma}) & Cov(\hat{\mu}, \hat{\xi}) \\ Cov(\hat{\mu}, \hat{\sigma}) & Var(\hat{\sigma}) & Cov(\hat{\sigma}, \hat{\xi}) \\ Cov(\hat{\mu}, \hat{\xi}) & Cov(\hat{\sigma}, \hat{\xi}) & Var(\hat{\xi}) \end{bmatrix},$$

e

$$\nabla x_p' = \left[\frac{\partial x_p}{\partial \mu}, \frac{\partial x_p}{\partial \sigma}, \frac{\partial x_p}{\partial \xi} \right],$$

a matriz de derivadas parciais de x_p avaliadas em $\hat{\mu}$, $\hat{\sigma}$ e $\hat{\xi}$.

Quando o parâmetro ξ na distribuição GEV é diferente de zero, a variância do nível de retorno \hat{x}_p pode ser calculada por

$$\begin{aligned} Var(\hat{x}_p) &= \left(\frac{\partial \hat{x}_p}{\partial \mu} \right)^2 Var(\hat{\mu}) + \left(\frac{\partial \hat{x}_p}{\partial \sigma} \right)^2 Var(\hat{\sigma}) + \left(\frac{\partial \hat{x}_p}{\partial \xi} \right)^2 Var(\hat{\xi}) \\ &+ 2 \frac{\partial \hat{x}_p}{\partial \mu} \frac{\partial \hat{x}_p}{\partial \sigma} Cov(\hat{\mu}, \hat{\sigma}) + 2 \frac{\partial \hat{x}_p}{\partial \mu} \frac{\partial \hat{x}_p}{\partial \xi} Cov(\hat{\mu}, \hat{\xi}) + 2 \frac{\partial \hat{x}_p}{\partial \sigma} \frac{\partial \hat{x}_p}{\partial \xi} Cov(\hat{\sigma}, \hat{\xi}) \end{aligned}$$

em que,

$$\begin{aligned} \frac{\partial \hat{x}_p}{\partial \mu} &= 1, \\ \frac{\partial \hat{x}_p}{\partial \sigma} &= -\frac{1}{\hat{\xi}} \left\{ 1 - [-\ln(1-p)]^{-\hat{\xi}} \right\}, \\ \frac{\partial \hat{x}_p}{\partial \xi} &= \frac{\hat{\sigma}}{\hat{\xi}^2} = -\frac{1}{\hat{\xi}} \left\{ 1 - [-\ln(1-p)]^{-\hat{\xi}} \right\} - \frac{\hat{\sigma}}{\hat{\xi}} [-\ln(1-p)]^{-\hat{\xi}} \ln[\ln(1-p)]. \end{aligned}$$

Para o caso em que $\xi \rightarrow 0$, tem-se a matriz de variâncias e covariâncias de $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$ obtidos da inversa da matriz de informação dada por

$$J_G = \left[\begin{array}{cc} \frac{\partial^2}{\partial \mu \partial \mu} l(\boldsymbol{\theta}) & \frac{\partial^2}{\partial \mu \partial \sigma} l(\boldsymbol{\theta}) \\ \frac{\partial^2}{\partial \sigma \partial \mu} l(\boldsymbol{\theta}) & \frac{\partial^2}{\partial \sigma \partial \sigma} l(\boldsymbol{\theta}) \end{array} \right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}^{-1} = \left[\begin{array}{cc} Var(\hat{\mu}) & Cov(\hat{\mu}, \hat{\sigma}) \\ Cov(\hat{\mu}, \hat{\sigma}) & Var(\hat{\sigma}) \end{array} \right]$$

em que $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma})$ são estimativas de máxima verossimilhança de $\boldsymbol{\theta} = (\mu, \sigma)$. E

$$\begin{aligned} \frac{\partial \hat{x}_p}{\partial \mu} &= 1, \\ \frac{\partial \hat{x}_p}{\partial \sigma} &= \ln[-\ln(1-p)]. \end{aligned}$$

Assim, a variância do nível de retorno estimado \hat{x}_p é dada por

$$Var(\hat{x}_p) = Var(\hat{\mu}) + \ln[-\ln(1-p)]^2 Var(\hat{\sigma}) + 2\ln[-\ln(1-p)] Cov(\hat{\mu}, \hat{\sigma}).$$

3 Resultados e discussão

Os dados de precipitações máximas mensais de chuva (expressos em mm) foram obtidos no período de 1962 a 2008, referentes ao Município de Moreilândia localizado na região central do Polígono das Secas, no Estado do Pernambuco. Os meses analisados foram dezembro, janeiro, fevereiro, março, abril e maio, período em que as chuvas ocorrem com maior incidência para esta região.

Para que possamos iniciar as análises da variável aleatória precipitação máxima das chuvas, se faz necessário realizar um estudo descritivo para cada mês do ano, cujos resultados se encontram na Tabela 1.

Tabela 1: Estatística descritiva da variável aleatória precipitação máxima mensal de chuvas nos períodos de 1962 a 2008, em Moreilandia-Pe.

Mês	Média	Mediana	Variância	Desvio Padrão	Coefficiente de Variação (%)	Coefficiente de Assimetria
Dezembro	28,57	21,10	553,01	23,52	111,45	0,99
Janeiro	38,03	33,25	629,49	25,09	75,46	0,73
Fevereiro	40,94	40,50	665,55	25,80	63,70	1,28
Março	45,35	44,50	329,06	18,14	40,76	0,83
Abril	33,47	35,00	629,65	25,09	71,69	1,38
Maio	20,82	19,30	229,53	15,15	78,50	0,35

Com base na Tabela 1, observa-se que, em média, o período que se dá entre dezembro a abril, apresenta precipitação máxima de chuva acima de 25,00 mm, sendo março, o mês que mostra, em média, o maior valor (45,35 mm). Segundo Freitas (1997), nessa região as chuvas se apresentam entre dezembro e maio, no entanto, em determinados anos isso não acontece, ocasionando um longo período sem chuvas, originando assim, a seca.

A Figura 2 apresenta os gráficos de caixa (*box-plot*) para a variável precipitação máxima de chuva, para cada mês analisado, que sugerem a presença de alguns valores atípicos (representados pelo símbolo *), o mês de maio, é o único que não apresenta valores discrepantes. Os meses de dezembro, março e abril, visualmente, demonstram ter uma distribuição bem assimétrica. Segundo Bussab e Morettin (2002), o *box-plot* dá uma idéia da posição, dispersão, assimetria, caudas e dados discrepantes.

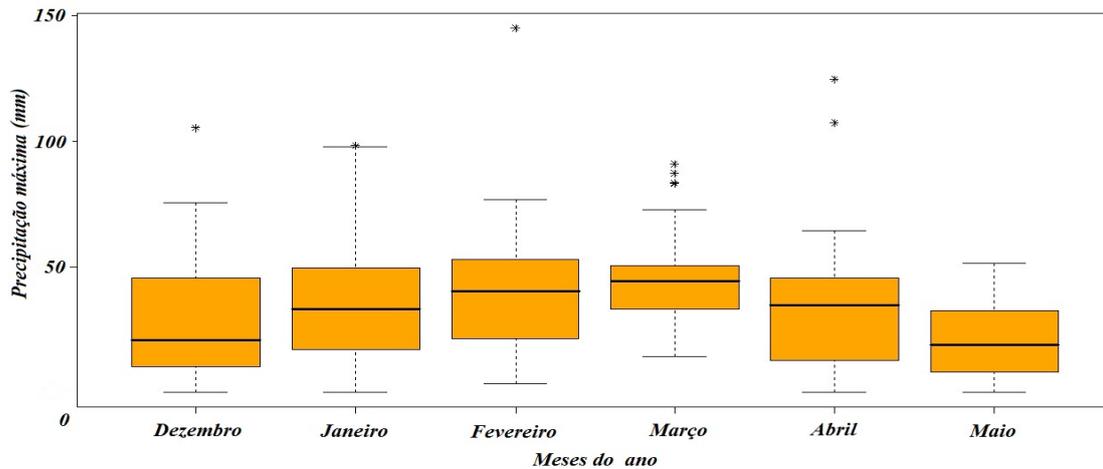


Figura 2: Gráficos de caixa (*box-plot*) para a variável precipitação máxima de chuva para cada um dos meses observados.

Com o objetivo de verificar a pressuposição de independência dos dados, realizou-se o teste de chorrilho, ao nível de 0,05 de significância, sendo este nível comparado com o valor-p. A Tabela 2 resume estas informações, em que, o valor-p maior que o nível de significância nos leva a concluir a não rejeição da hipótese de independência dos dados em estudo.

Tabela 2: Teste de chorrilho sob a pressuposição de independência dos dados nos meses em estudo, determinando, respectivamente, o valor-p destes meses.

Mês	valor-p
Dezembro	0,4527
Janeiro	1,0000
Fevereiro	0,6600
Março	0,2925
Abril	0,2329
Maio	0,3710

Analisando a Tabela 2, pode-se afirmar ao nível de 5% de probabilidade que para os meses em observação a hipótese de independência não deve ser rejeitada. De acordo com Bautista (2002), o cumprimento dessa pressuposição garante a obtenção de inferências estatísticas satisfatórias a partir dos modelos de valores extremos. Em seguida, calcula-se as estimativas pontuais dos parâmetros da GEV, μ , σ , e ξ , obtidos através do método de máxima verossimilhança, com suas respectivas variâncias e covariâncias estimadas, para os meses dezembro, janeiro, fevereiro, março, abril e maio.

Observa-se por meio da Tabela 3, que as estimativas pontuais do parâmetro de forma (ξ) há valores negativos e positivos, mas próximo de zero. De acordo com Ferrari (2010),

Tabela 3: Estimativas dos parâmetros da distribuição generalizada de valores extremos e respectivas variâncias e covariâncias estimadas, para os meses em análise.

Mês	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$	$V\hat{ar}(\hat{\mu})$	$V\hat{ar}(\hat{\sigma})$	$V\hat{ar}(\hat{\xi})$	$C\hat{ov}(\hat{\mu}, \hat{\sigma})$	$C\hat{ov}(\hat{\mu}, \hat{\xi})$	$C\hat{ov}(\hat{\sigma}, \hat{\xi})$
Dezembro	16,02	15,00	0,24	7,96	5,58	0,04	4,54	-0,28	-0,17
Janeiro	26,02	18,79	0,06	10,78	6,31	0,02	4,15	-0,22	-0,15
Fevereiro	29,40	19,75	0,01	10,61	5,73	0,01	3,16	-0,12	-0,08
Março	37,32	14,35	-0,02	5,77	2,97	0,01	1,46	-0,09	-0,06
Abril	22,09	18,19	0,05	9,32	5,16	0,01	3,11	-0,13	-0,07
Mai	14,22	12,85	-0,09	5,68	3,48	0,04	2,08	-0,25	-0,23

considerando-se que o parâmetro ξ define o tipo de distribuição de valores extremos a ser utilizado, então, para verificar qual a distribuição melhor se ajusta aos dados, foram construídos intervalos com 95% de confiança para o parâmetro ξ (Tabela 4). Analisando-se os intervalos de confiança na Tabela 4 para o parâmetro ξ , em todos os meses estudados, o valor zero encontra-se dentro do intervalo de confiança (I.C.) de 95%. Portanto há fortes evidências para não se rejeitar a hipótese de que a distribuição Gumbel é a mais adequada para modelar os dados em estudo. Esta conclusão é reforçada pela estatística T_{LR}^* , calculada em 2.11, comparando-se o valor que se encontra na Tabela 4, com o valor tabelado de χ_1^2 ao nível de 0,05 de significância ($\chi_{1;0,05}^2 = 3,84$), este mesmo permite tirar as mesmas conclusões. Segundo Sansigolo (2008), a distribuição Gumbel é a mais adequada para estimar as probabilidades de ocorrência de eventos extremos de precipitação de diversas durações.

Tabela 4: Intervalos de 95% de confiança para o parâmetro de forma (ξ) e valores da estatística de razão de verossimilhança modificado (T_{LR}^*).

Mês	Limites de 95% de confiança para ξ		T_{LR}^*
	Inferior	Superior	
Dezembro	-0,14	0,62	3,12
Janeiro	-0,24	0,35	0,20
Fevereiro	-0,20	0,22	0,01
Março	-0,23	0,19	0,04
Abril	-0,18	0,27	0,22
Mai	-0,46	0,29	0,31

Prosseguindo com as análises observa-se que na Tabela 5, encontram-se as estimativas dos parâmetros μ e σ , da distribuição Gumbel, com suas respectivas variâncias e covariâncias, por meio do método da máxima verossimilhança para os meses dezembro, janeiro, fevereiro, março, abril e maio.

Para verificar a qualidade do ajuste da distribuição, foram esboçados os gráficos *pp-plot* e *qq-plot*, localizados nas Figuras 3 e 4, respectivamente, e que sugerem um bom ajuste da distribuição Gumbel aos dados de precipitação máxima mensal dos meses em estudo. Visualmente, pode-se observar que houve um bom ajuste à distribuição Gumbel.

Tabela 5: Estimativas dos parâmetros μ e σ da distribuição Gumbel e respectivas variâncias e covariâncias estimadas, para os meses em análises.

Mês	$\hat{\mu}$	$\hat{\sigma}$	$V\hat{ar}(\hat{\mu})$	$V\hat{ar}(\hat{\sigma})$	$C\hat{ov}(\hat{\mu}, \hat{\sigma})$
Dezembro	18,13	17,01	7,08	4,38	1,69
Janeiro	26,60	19,25	8,92	5,22	2,12
Fevereiro	29,51	19,82	9,29	5,16	2,18
Março	37,16	14,26	5,01	2,72	1,16
Abril	22,59	18,54	8,27	4,66	1,92
Maió	13,64	12,39	3,71	2,14	0,90

Em seguida, é apresentada na Tabela 6, um resumo das informações na aplicação do teste de Kolmogorov-Smirnov ao nível de 0,05 de significância. Encontram-se na Tabela 6 as diferenças máximas absolutas observadas entre os valores de probabilidade das funções de distribuição empírica e Gumbel (teórica) para os meses em análise e, seus respectivos valores-p. De acordo com este teste, a distribuição Gumbel ajusta-se bem aos dados pois, $D < D_{n,0,05} \cong 0,20$ para todos os meses especificados na tabela 6, esta mesma conclusão pode ser vista por meio da estatística do valor-p pois, para todos os meses estudados ele foi maior que o nível de significância especificado. A seguir é apresentado na Figura 5, o esboço do teste de Kolmogorov-Smirnov, o qual proporciona, visualmente, as mesmas conclusões citadas anteriormente.

Tabela 6: Resultados do teste de Kolmogorov-Smirnov, e seus respectivos valor-p, para verificação da qualidade do ajuste da distribuição Gumbel aos dados de precipitação máxima mensal nos meses, dezembro, janeiro, fevereiro, março, abril e maio.

Mês	Diferença máxima absoluta (D)	valor-p
Dezembro	0,13	0,46
Janeiro	0,09	0,79
Fevereiro	0,09	0,74
Março	0,08	0,91
Abril	0,12	0,50
Maió	0,08	0,88

Tabela 7: Probabilidades de ocorrência de precipitações máximas mensais de chuvas acima de 50, 75, 100 e 125 mm, para os 6 meses do ano, Moreilândia, Pe. Dados de precipitação máxima mensal nos meses, dezembro, janeiro, fevereiro, março, abril e maio.

Mês	> 50	> 75	> 100	> 125
Dezembro	0,1423	0,0347	0,0081	0,0019
Janeiro	0,2566	0,0777	0,0218	0,0060
Fevereiro	0,2993	0,0959	0,0282	0,0081
Março	0,3339	0,0680	0,0121	0,0021
Abril	0,2038	0,0574	0,0152	0,0040
Maió	0,0517	0,0070	0,0009	0,0001

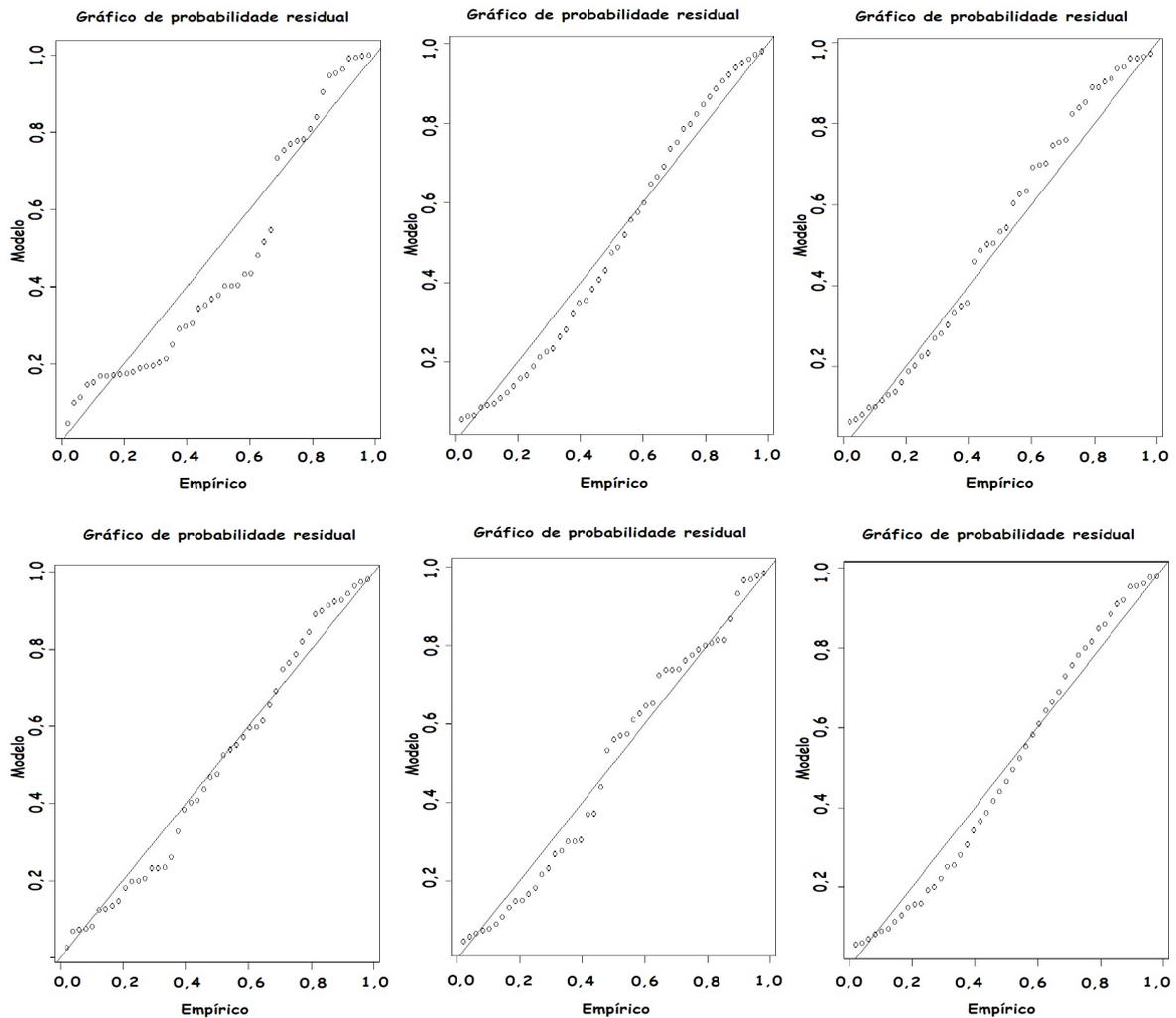


Figura 3: Gráficos de probabilidade-probabilidade para diagnóstico da distribuição Gumbel aos dados de precipitação máxima mensal nos meses, dezembro, janeiro, fevereiro, março, abril e maio.

Na Tabela 7, encontram-se as probabilidades de ocorrência de precipitação máxima de chuvas acima de 50, 75, 100 e 125 mm, para os meses de dezembro, janeiro, fevereiro, março, abril e maio. Observa-se nessa Tabela que os meses de janeiro, fevereiro e março registram maiores probabilidade para períodos acima de 50 milímetros de chuva.

As estimativas dos níveis de retorno mensais e intervalos de confiança associados aos períodos de retorno 5, 10, e 50 anos obtidos pelo método delta são apresentados na Tabela 8. Observa-se nessa Tabela que em fevereiro será o mês que terá os maiores registros pluviométrico nos períodos de retornos de 5, 10 e 50 anos, respectivamente. Já em maio será o mês que ocorrerá os menores índices de precipitação para os mesmos períodos de retorno.

Observa-se na Tabela 8, em um tempo médio de 10 anos, o valor de precipitação diária

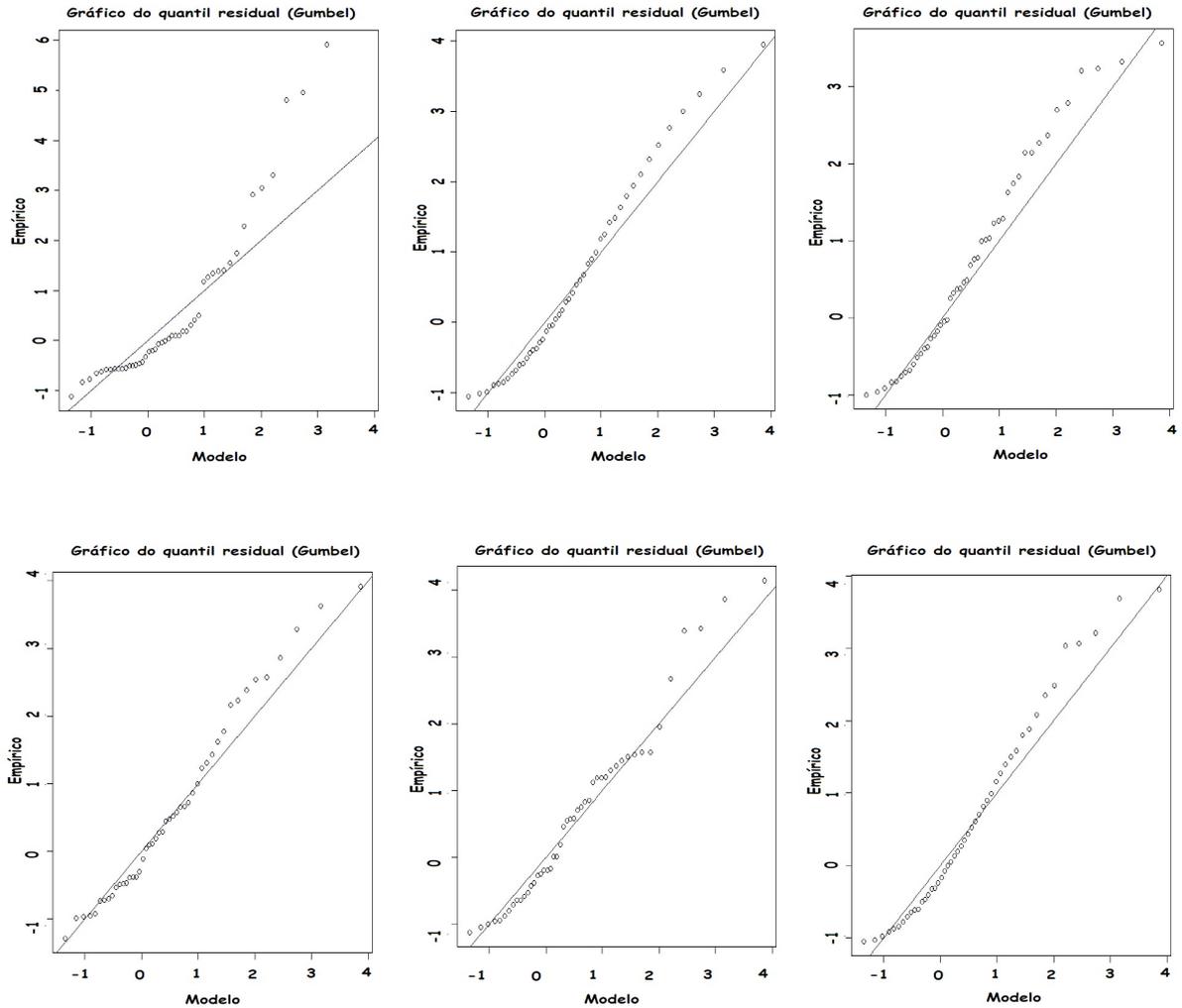


Figura 4: Gráficos de quantil-quantil para diagnóstico da distribuição Gumbel aos dados de precipitação máxima mensal nos meses, dezembro, janeiro, fevereiro, março, abril e maio.

Tabela 8: Níveis de retorno (\hat{x}_p - mm) estimados e limites inferior (LI) e superior (LS) de seus respectivos intervalos de 95% de confiança, para os períodos de retorno 5, 10 e 50 anos, obtidos através do método delta.

Mês	Período de retorno (anos)								
	5 anos			10 anos			50 anos		
	LI	\hat{x}_p	LS	LI	\hat{x}_p	LS	LI	\hat{x}_p	LS
Dezembro	34,44	43,64	52,84	44,50	56,41	68,32	66,21	84,50	102,79
Janeiro	45,29	55,48	65,67	56,79	69,92	83,05	81,63	101,72	121,81
Fevereiro	48,98	59,24	69,50	60,94	74,12	87,30	86,79	106,86	126,93
Março	51,06	58,55	66,04	59,65	69,25	78,85	78,20	92,80	107,40
Abril	40,68	50,39	60,10	51,82	64,30	76,78	75,89	94,92	113,95
Maio	8,79	16,85	24,91	7,48	18,45	29,42	4,27	21,99	39,71

máxima no mês de fevereiro esteja entre 60,94 e 87,30 mm, com 95% de confiabilidade. Já para o mês de março neste mesmo período, espera-se que o valor de precipitação diária

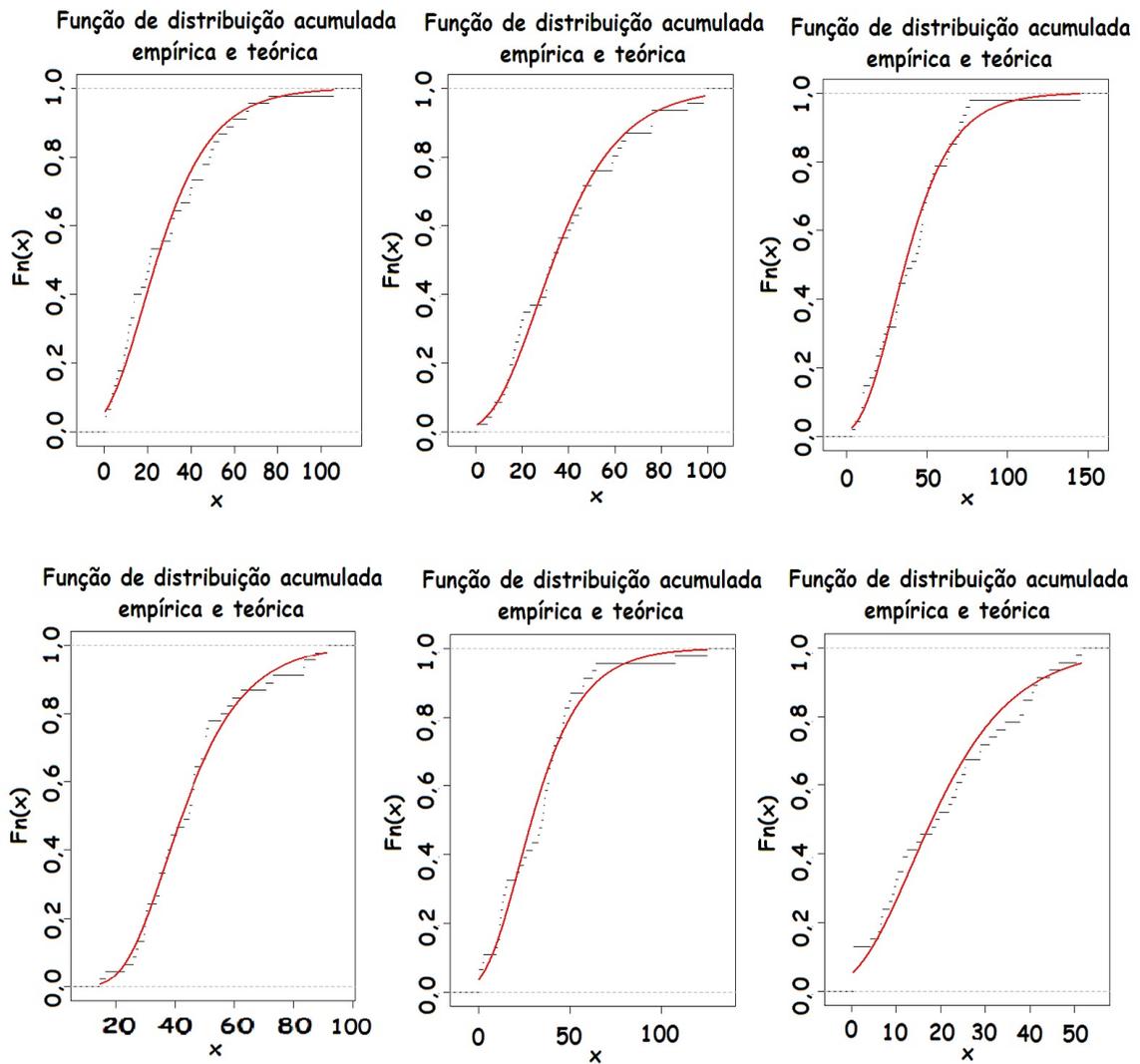


Figura 5: Gráficos do teste de Kolmogorov-Kminorv da função de distribuição acumulada empírica (na cor cinza) e teórica (na cor vermelha) para diagnóstico da distribuição Gumbel aos dados de precipitação máxima mensal nos meses, dezembro, janeiro, fevereiro, março, abril e maio.

máxima encontra-se entre 7,48 e 29,42 mm, com 95% de confiança.

4 Considerações finais

A Teoria de Valores Extremos tem ajudado a sociedade na solução de muitos problemas práticos, no entanto, existe complexidade no ajuste e verificação dessa teoria, um dos problemas está no valor da estimativa do parâmetro de forma (ξ), que por sua vez, não poderá ultrapassar o intervalo de -0,5 a 0,5. Pode-se observar nesse trabalho a não ocorrência desse problema nos meses analisados. Uma situação que provavelmente venha a comprometer as análises na Teoria de Valores Extremos é a ocorrência de muitos dados faltantes ou até mesmo, a presença de muitos zeros em uma série histórica analisada, este problema poderá ser solucionado com a imputação de dados. O método iterativo quasi-Newton foi utilizado para determinar uma solução aproximada para o sistema de equações não lineares, pois observou-se no método Newton-Raphson vários máximos locais.

A distribuição generalizada de valores extremos com parâmetro $\xi = 0$, que corresponde à distribuição de valores extremos tipo I ou de Gumbel é adequada para estudar o comportamento da precipitação máxima de chuvas nos meses que foram analisados, no município de Moreilândia-PE.

Os maiores índices de máximos de precipitação pluviométrica verificam-se nos meses de fevereiro e março. Já com relação aos níveis de retorno observou-se, à medida que o tempo médio do período de retorno aumenta a ocorrências dos máximos de precipitação tendem também a crescer, sendo o mês de fevereiro responsável pelo maiores níveis de retorno e o mês de maio os menores níveis no que diz respeito a precipitação máxima.

Referências

- BAUTISTA, E. Z. A distribuição generalizada de valores extremos no estudo da velocidade máxima do vento em Piracicaba, SP. Dissertação de Mestrado, ESALQ, USP, Piracicaba. 2002.
- BAUTISTA, E. A. L.; ZOCCHI, S. S.; ANGELOCCI, L. R. A distribuição generalizada de valores extremos aplicada ao ajuste dos dados de velocidade máxima de vento em Piracicaba, SP. **Revista de Matemática e Estatística**, Marília, v.22, n.1, p. 95-111, 2004.
- BEIJO, L. A. Distribuição Gumbel: Estudo de métodos de estimação dos parâmetros e ajuste aos dados de precipitação máxima de Lavras, Minas Gerais. Dissertação de Mestrado, UFLA, Lavras. 2002.
- BORTKIEWICZ, L., von (1922). Variationsbreite und mittlerer Fehler, Sitzungsber. Berli. Math. Ges. 21, 3-11.
- BELTRÃO, B. A. Diagnóstico do Município de Moreilândia, Pe. Programa de Desenvolvimento Energético dos Estados e Municípios - PRODEEM. Serviço Geológico do Brasil, Moreilândia - Pe. 2005.
- BROWNLEE, K. A. **Statistical theory and methodology in science and engineering**. 2 ed. New York: John Wiley, 1965. 590 p.
- BUSSAB, W. O.; MORETTIN, P. A. Estatística básica. São Paulo: Editora Saraiva, 2002.
- COLES, S. **An Introduction to Statistical Modeling of Extreme Values**. Berlin: Springer, 2001. 208 p.
- DODD, E. L. (1923). The greatest and least variate under general laws of error, Trans. Amer. Math. SOC. 25, 525-539.
- EISENHART, C.; SWED, F. Tables for testing randomness of grouping in a sequence of alternatives. **Annals of Mathematical Statistics**, Ann Arbor, v. 14, n. 1, p. 66-87. 1941.
- FERRARI, G. T. Imputação de dados pluviométricos e sua aplicação na modelagem de eventos extremos de seca agrícola. Dissertação de Mestrado. ESALQ, USP, Piracicaba. 2011.
- FISHER, R. A. and TIPPETT, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample, Procs. Cambridge Philos. SOC. 24, 180-190.

- FRÉCHET, M. (1927). Sur la loi de probabilité de l'écart maximum, *Ann. Soc. polon. Math. Cracovie* 6, 93-116.
- FREITAS, M. A.S., 1997: Análise Estatística da Relação entre o Fenômeno do El-Niño e a Seca no Estado do Ceará, III Encontro de Iniciação à Pesquisa, de 15 a 17 de setembro de 1997, Universidade de Fortaleza - UNIFOR.
- FULLER, W. E. (1914). Flood flows, *Trans. Amer. SOC. Civil Engineers* 77, 564.
- GRIFFITH, A. A. (1920). The phenomena of rupture and flow in solids, *Philos. Trans. Roy. SOC. London A*, 221, 163-198.
- GUMBEL, E. J. (1937a). Les intervalles extrêmes entre les émissions radioactives, *J. Phys. Radium* 8, 446-452.
- GUMBEL, E. J. (1937b). La durée extrême de la vie humaine, *Actualités Scientifique et Industrielles*, Hermann et Cie.
- Gumbel, E. J. (1941). The return period of flood flows, *Ann. Math. Statist.* 12, 163-190.
- HOSKING, J. R. M. Testing whether the shape parameter is zero in the generalized extreme-value distribution. **Biometrika**. v. 71, p.367-374, 1984.
- JUNCOSA, M. L. (1949). The asymptotic behavior of the minimum in a sequence of random variables, *Duke Math. J.* 16, 609-618.
- JÚNIOR, H. M. Análise de valores extremos para dados de poluição atmosférica na cidade de São Paulo. Dissertação de Mestrado, Universidade Federal de Lavras, Lavras - MG, 27p. 2010.
- Kotz, S.; Nadarajah, S. **Extreme Value Distributions: Theory and Applications**. London: Imperial College Press, 2000. 195 p.
- MAZUCHELI, J.; COELHO-BARROS, E. A.; ACHCAR, J. A. Inferências para o máximo da função de risco da distribuição Log-Logística *Rev. Mat. Estat.*, São Paulo, v.24, n.4, p.73-87, 2006.
- MISES, R., von (1923). Uber die Variationsbreite einer Beobachtungsreihe, *Sitzungsber. Berlin. Math. Ges.* 22, 3-8.
- MISES, R., von (1936). La distribution de la plus grande de n valeurs, *Rev. Math. Union Interbalk.* 1, 141-160. Reproduced in *Selected Papers of Richard von Mises*, II(1954), pp. 271-294, *Amer. Math. SOC*
- SANSIGOLO, C. A. A distribuição de extremos e precipitação diária, temperatura máxima e mínima e velocidade do vento em Piracicaba, SP (1917 - 2006). *Rev. Bras. Met.*, v.23, n.3, p.341-346, 2008.
- SMITH, R. L. Maximum likelihood estimation in class of nonregular cases. **Biometrika**, Cambridge, v.72, p.67-92, 1958.
- SILVA, R. R. A distribuição generalizada de Pareto e mistura da distribuição de Gumbel no estudo da vazão e da velocidade máxima do vento em Piracicaba, SP. Dissertação de Mestrado, ESALQ, USP, Piracicaba, 07p. 2008

TIPPETT, L. H. C. (1925). On the extreme individuals and the range of samples taken from a normal population, *Biometrika* 17, 364-387.

VIOLA, M. L. L. Teoria de Valores Extremos e Cópulas: Distribuição Valor Extremo e Cópulas Arquimedianas Generalizadas Trivariadas. Dissertação de Mestrado, IMECC-UNICAMP, Campinas-SP, 27p. 2006.

ZAR , J. H. **Biostatistical analysis**. 4.ed. New Jersey: Prentice Hall, 1999. 911 p.

ANEXO A

Tabela 9: Precipitações pluviiais máximas mensal (mm) na cidade de Moreilandia-Pe, nos períodos de 1962 a 2008

Ano	Janeiro	Fevereiro	Março	Abril	Mai	Dezembro
1962	45,80	45,60		42,40	6,20	18,40
1963	62,30	76,40	70,20	15,00	37,20	75,20
1964		63,80	46,00	45,60	24,30	19,20
1965	97,40	18,00	42,40	47,00	17,60	11,60
1966	16,70	74,20	29,80	46,40	6,60	10,50
1967	13,20	54,60	72,60	45,20	44,00	58,80
1968	19,90	30,30	55,20	35,30	19,20	52,40
1969	32,20	10,60	83,20	12,70	10,00	16,50
1970	34,90	144,50	36,40	0,00	0,00	25,80
1971	58,60	30,20	90,70	37,20	14,20	5,10
1972	30,00	37,00	29,40	35,60	41,20	56,10
1973	41,50	22,90	50,20	56,60	9,10	20,60
1974	60,70	62,00	57,30	124,30	28,30	4,20
1975	45,00	32,00	30,20	11,80	21,30	10,80
1976	6,20	44,20	22,30	11,60	3,80	39,30
1977	33,30	46,60	26,20	38,90	40,30	21,30
1978	15,30	20,50	29,30	21,90	32,10	20,20
1979	75,20	62,30	48,60	24,30	28,50	48,60
1980	75,20	69,40	62,00	9,20	0,00	50,00
1981	18,00	8,00	87,00	38,00	0,00	40,00
1982	30,00	10,00	50,00	107,00	15,00	45,00
1983	23,00	18,00	50,00	13,00	0,00	30,00
1984	7,40	17,00	25,40	62,00	8,60	10,00
1985	58,00	50,00	40,00	49,00	23,00	
1986	10,80	43,00	51,00	57,00	5,60	13,20
1987	18,60	14,20	45,00	2,60	46,00	13,00
1988	49,00	25,00	46,00	36,00	50,00	105,00
1989	16,00	10,00	45,00	41,00	30,40	66,00
1990	0,00	70,00	16,00	40,00	8,00	65,00
1991	64,00	32,00	29,00	33,00	38,00	3,00
1992	28,00	47,00	34,00	50,00	11,00	
1993	14,00	31,00				0,50
1994	98,00	47,00	50,00	34,00	40,00	48,00
1995	45,00	52,00	14,00	29,00	38,00	9,00
1996	90,50	22,00	34,00	38,00	34,00	5,60
1997	14,00	5,00	33,00	64,00	18,40	45,00
1998	31,00	53,20	38,00	11,00	0,00	35,00
1999	39,00	47,20	83,00	9,80	22,00	32,00
2000	19,10	36,00	27,00	20,40	0,00	31,00
2001	12,00	32,00	44,00	14,50	9,50	0,00
2002	43,80	3,20	37,00	0,00	25,00	13,00
2003	35,00	9,00	44,30	2,00	12,00	8,50
2004	75,00	68,00	34,00	0,00	51,00	39,00
2005	49,00	49,00	38,00	11,10	25,10	31,00
2006	4,50	40,00	46,40	32,00	11,00	0,00
2007	40,50	72,00	59,00	35,00	23,10	8,30
2008	29,70	45,00	36,00	26,00	6,10	2,30