



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

Samara Rilda de Sousa Bezerra

Modelagem estatística de valores extremos aplicados a dados de temperatura máxima em São Gonçalo-PB.

Campina Grande
Maio de 2013

Samara Rilda de Sousa Bezerra

Modelagem estatística de valores extremos aplicados a dados de temperatura máxima em São Gonçalo-PB.

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de 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:

Ricardo Alves de Olinda

Campina Grande

Abril de 2013

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL – UEPB

B574m Bezerra, Samara Rilda de Sousa.
Modelagem estatística de valores extremos aplicados a dados de temperatura máxima em São Gonçalo-PB [manuscrito] / Samara Rilda de Sousa Bezerra. – 2013.
35 f. : il.

Trabalho de Conclusão de Curso (Graduação em Estatística) – Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2014.

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

1. Valores extremos. 2. Distribuição assintótica. 3. Método de máxima verossimilhança. I. Título.

21. ed. CDD 519.5

Samara Rilda de Sousa Bezerra

Modelagem estatística de valores extremos aplicados a dados de temperatura máxima em São Gonçalo-PB.

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de 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: 03/05/2013

Banca Examinadora:

Ricardo Alves de Olinda

Prof. Dr. Ricardo Alves de Olinda
Orientador

DMEsteves

Prof^a. Dr^a. Divanilda Maia Esteves
Universidade Estadual da Paraíba

GA 27

Prof. Dr. Gustavo Henrique Esteves
Universidade Estadual da Paraíba

Dedicatória

Aos meus pais;

Maria de Lourdes Sousa e João Galdino Bezerra.

Às minhas irmãs;

Carliméria Rita e Carolayne Ruth.

Às minhas avós;

Rita Sousa e Joana Gabriel Gonçalves(*in memoriam*).

Dedico com amor e gratidão.

Agradecimentos

À minha família, em especial a minha mãe Maria de Lourdes que sempre acreditou nos meus esforços.

À minha irmã Carolayne Ruth pela compreensão e carinho.

À minha namorada e amiga Vitória Dantas, por toda compreensão e imensurável amor. Com você meus dias se tornaram mais felizes!

Ao Ricardo Soares, pela a ajuda que foi essencial.

Às minhas amigas de graduação, Wanessa Luana, Adriana, Natielly, Arielly e Valnely que estavam presentes nos momentos de estudos dividindo conhecimentos.

Aos meus Amigos, Moises Moureira de Sá, Carla Rayanne e Ana Paula Xavier.

Ao professor Ricardo Alves de Olinda pela oportunidade, por sua orientação e incentivo.

Ao professor Tiago Almeida pela sua disponibilidade de ajuda.

Aos docentes do curso de Estatística da Universidade Estadual da Paraíba, pela transmissão de conhecimento.

Resumo

Os fundamentos básicos da teoria dos valores extremos foram inicialmente expostos por Fisher e Tippett, que estabeleceram os três tipos de distribuição assintótica de valores extremos, como Gumbel (Tipo I), Fréchet (Tipo II) e Weibull (Tipo III). Os dados provenientes para este estudo relatam uma série histórica de temperatura máxima no período de 39 anos (1970 a 2009), no município de São Gonçalo, localizado no sertão paraibano. As estimativas dos parâmetros da distribuição generalizada de valores foram obtidas pelo método de máxima verossimilhança, seguido do teste de Kolmogorov-Sminorv, gráficos de probabilidade-probabilidade e o quantil-quantil, aplicados para verificar o ajuste do modelo aos dados. Na sequência, obteve-se níveis de retorno para os períodos de retornos de 34, 35, 36 e 37 anos, construindo-se seus respectivos intervalos de confiança com 95% de significância. Verificou-se que as distribuições Gumbel e Weibull são adequadas para representar os dados de temperatura máxima dos meses em estudo.

Palavras-chave: Distribuição Assintótica, Níveis de Retorno, Valores Extremos.

Abstract

The basics theory of extreme values were initially exposed by Fisher and Tippett, that established the three types of asymptotic distribution of extreme values, as Gumbel (Type I), Fréchet (Type II) and Weibull (Type III). Data from this study relates to a series of historical maximum temperature during the period of 39 years (1970 to 2009), in the of São Gonçalo, located in the backlands of Paraíba. The estimatives the parameters of the widespread distribution of values were obtained by the maximum likelihood method, followed by Kolmogorov-Smirnov, probability-probability plots and quantile-quantile applied to check the fit the model to the data. As a result, we obtained the return levels for return periods of 34, 35, 36 and 37 years, building up their confidence intervals with 95% significance. It was found that the Gumbel and Weibull distributions are suitable for adjusting the temperature data maximum month study.

Keywords: Asymptotic Distribution, Levels of Return, Extreme Value .

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 11
2	Fundamentação Teórica	p. 12
2.1	Teoria dos Valores Extremos	p. 13
2.2	Teste de aleatoriedade	p. 15
2.3	Estimação dos parâmetros por máxima verossimilhança	p. 16
2.4	Seleção da distribuição de valores extremos	p. 19
2.5	Diagnóstico do ajuste da distribuição da GVE	p. 20
2.6	Estimações dos níveis de retorno da distribuição GVE	p. 21
2.7	Obtenção dos intervalos de confiança	p. 22
3	Resultados e discussão	p. 24
4	Conclusão	p. 32
5	Referências Bibliográficas	p. 33

Lista de Figuras

- 1 Gráfico de caixa (Box-Plot) referente a temperatura máxima no período de 1970 a 2009 no município de São Gonçalo. p. 25
- 2 Gráficos de quantil-quantil, diagnóstico das distribuições para os dados de temperatura máxima mensal dos meses, janeiro, fevereiro, março, abril, maio, junho, julho, agosto, setembro, outubro, novembro e dezembro. p. 30
- 3 Gráficos do teste de Kolmogorov-Smirnov da função de distribuição acumulada empírica (representada pelo os tracinhos) e teórica (representada pela a curva) para diagnóstico dos modelos ajustados aos dados de temperatura máxima mensal. p. 31

Lista de Tabelas

1	Estatística descritiva da variável aleatória temperatura máxima ($^{\circ}\text{C}$) mensal no período entre 1970 a 2009, do município de São Gonçalo - PB.	p. 24
2	Teste de chorrilho sob a pressuposição de independência dos dados entre os meses de janeiro a dezembro e o nível descritivo dos respectivos meses.	p. 26
3	Valores observados das estimativas dos estimadores dos parâmetros da GVE e suas variâncias e covariâncias estimadas para os dados de temperatura máxima ($^{\circ}\text{C}$).	p. 26
4	Intervalos de 95% para o parâmetro de forma e os respectivos valores da estatística de verossimilhança modificada para os meses.	p. 27
5	Estimativa dos estimadores dos parâmetros $\hat{\mu}$ e $\hat{\sigma}$ da distribuição Gumbel com suas variâncias e covariâncias estimadas para os respectivos meses em análise.	p. 28
6	Resultados do teste de Kolmogorov-Smirnov para verificar o qualidade do ajuste da distribuição aos dados de temperatura máxima dos meses em estudo.	p. 28
7	Probabilidades de ocorrência de temperatura máxima mensal acima de 34, 35, 36 e 37 anos, correspondentes aos 12 meses de estudo no município de São Gonçalo-PB.	p. 29
8	Níveis de retorno ($\hat{x}_p^{\circ}\text{C}$) estimados e limites inferior (LI) e superior (LS) e seus respectivos intervalos de 95% confiança, para os períodos de retorno de 34, 35, 36 e 37 período de retorno, obtidos pelo o método delta	p. 29

1 Introdução

Temperatura extremamente alta, associada às ondas de calor prolongadas prejudicam a produção agrícola, elevando-se a quantidade de energia e consumo de água. Na região nordeste todos esses problemas são visíveis e difíceis de contornar. Conforme Medeiros (2011) a ocorrência cíclica das secas e seus efeitos catastróficos são por demais conhecidos e remontam aos primórdios da história do Brasil.

Por meio de uma série hidrológica¹, tem-se como objetivo determinar uma distribuição de probabilidade que represente ocorrências futuras de um fenômeno relacionado com esta série hidrológica, no caso deste estudo, tem-se as ocorrências de temperaturas máximas registradas no período de 1970 a 2009. A teoria de valores extremos proposta por Fisher e Tippett (1928) é uma metodologia capaz de modelar esse tipo de evento. Conforme essa teoria, existem três tipos possíveis de distribuição assintótica de valores extremos, Gumbel, Fréchet e Weibull respectivamente Tipo I, Tipo II e Tipo III.

Por meio dessas distribuições pode-se estimar as probabilidades de ocorrência de temperaturas máximas e mínimas em um tempo pré-determinado de observações acerca dessa variável. Em meados da década de 50, Jenkinson (1955) reformulou a distribuição Generalizada de Valores Extremos (GVE), levando-se em consideração uma família de distribuições, pois esta seria capaz de representar os três tipos de distribuições assintóticas de valores extremos, como casos particulares.

Este trabalho foi desenvolvido com o objetivo principal de apresentar e implementar a metodologia para ajuste a distribuição GVE aos dados de temperatura máxima mensal do Município de São Gonçalo no Sertão Paraibano; verificar o ajuste do modelo aos dados por meio do teste de Kolmogorov-Sminorv, gráfico quantil-quantil e verificar qual das três distribuições assintóticas de valores extremos se ajustam melhor aos dados de temperatura máxima.

¹As variáveis hidrológicas são registradas por meio das chamadas séries hidrológicas, que constituem as observações organizadas no modo sequencial de sua ocorrência no tempo (ou espaço).

2 Fundamentação Teórica

Em 1700, Nicolas Bernoulli já se deparava com um problema de valor extremo, ao tratar da média da distância máxima à origem, de pontos aleatoriamente posicionados em uma linha reta de tamanho fixo. De acordo com Pereira e Pinto (2012) a Teoria de Valores Extremos (TVE) originou-se da necessidade dos astrônomos de utilizar ou rejeitar observações discrepantes e os primeiros artigos sobre o assunto datam de 1914 e 1920.

O desenvolvimento dessa teoria pode ser encontrada no artigo de Bortkiewicz (1922), onde o referido autor abordava distribuição do tamanho do intervalo entre o máximo e o mínimo, em uma amostra com distribuição normal. Um ano depois Von Mises (1923) calculou o valor esperado da distribuição e Dodd (1923) determinou a mediana. Tippett em (1925) estudou sobre a função de distribuição acumulada, os momentos da estatística de ordem extrema em uma amostra de população normal, Fréchet (1927), obteve grande relevância no artigo sobre as distribuições assintóticas de valores extremos.

Os primeiros fundamentos básicos dessa teoria foram inicialmente expostos por Fisher e Tippett (1928), que introduziram os três tipos possíveis de distribuição assintótica dos valores extremos, hoje conhecidas como Gumbel (tipo I), Fréchet (tipo II) e Weibull (tipo III), sendo Fisher e Tippett os primeiros formalizadores dessa aplicação obtendo assim importantes resultados para casos univariados. Posteriormente foram estudados os casos bivariados.

O termo *valor extremo* tem dois significados: máximo ou mínimo, ou seja, o maior ou menor valor de uma série que representa os limiares, ou seja, os maiores valores de um conjunto de dados a partir de um limiar suficientemente satisfatório. Dessa forma, a TVE modela os extremos utilizando-se a distribuição do máximo ou do mínimo.

2.1 Teoria dos Valores Extremos

Segundo Sanfins (2009), a teoria dos valores extremos é um ramo da probabilidade que estuda o comportamento assintótico de extremos associados a uma sequência de variáveis aleatórias. Segundo Ruben (2006), a abrangência de suas aplicações é ampla, incluindo-se uma variedade de fenômenos naturais tais como inundações, poluição atmosférica, correntes oceânicas e problemas oriundos de outras áreas tais como da engenharia, atuária e finanças. De acordo com Silva_a (2008), os valores extremos podem ser considerados aqueles eventos raros que ocorrem nas caudas das distribuições (fenômenos caudais), isto é, distante dos aglomerados ou da aglomeração (média, moda e mediana) do amontoado da distribuição.

Os fundamentos da teoria dos valores extremos foram inicialmente expostos por Fisher e Tippet (1928), que definiram os três tipos possíveis (I, II e III) de distribuição assintótica de valores extremos, respectivamente conhecidas como Gumbel, Fréchet e Weibull. Por exemplo, as caudas da distribuição de Fréchet seguem distribuição de Cauchy, Pareto ou *t* de Student, já as da distribuição Gumbel expressam distribuições exponencial, gama, normal ou log-normal e as da distribuição Weibull seguem uma distribuição uniforme.

Denotando-se o máximo amostral por $M_n = \max\{X_1, X_2, \dots, X_n\}$ e supondo-se que X_n são independentes e identicamente distribuídas (*i.i.d*), tem-se a distribuição do máximo, definida 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 Medeiros (2011), para n muito grande, a função de distribuição M_n pode ser degenerada¹. Com isso, seria de grande utilidade algum resultado assintótico para o máximo. Segundo Coles (2001), a TVE assegura a existência de uma distribuição assintótica não degenerada F , para uma transformação linear de M_n , isto é, sequências numéricas apropriadas $a_n > 0$ e $b_n \in \mathbb{R}$ e uma distribuição não degenerada de F tal que

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

Como $F_X(x)$ é uma função de distribuição não degenerada, então F pertence a uma

¹Em matemática, uma distribuição degenerada é a distribuição de probabilidade de uma variável aleatória discreta cujo suporte consiste em adotar de somente um valor.

das seguintes famílias

$$I : F_X(x) = \exp \left\{ \exp \left[- \left(\frac{x - \mu}{\sigma} \right) \right] \right\}, -\infty < x < \infty,$$

$$II : F_X(x) = \begin{cases} 0, & x \leq \mu \\ \exp \left\{ - \left(\frac{x - \mu}{\sigma} \right)^{-\xi} \right\}, & x > \mu, \end{cases}$$

$$III : F_X(x) = \begin{cases} \exp \left\{ - \left[- \frac{x - \mu}{\sigma} \right]^{-\xi} \right\}, & x < \mu \\ 1, & x \geq \mu, \end{cases}$$

para os parâmetros $\sigma > 0$, $\mu \in \mathbb{R}$ e no caso das famílias I e II $\xi > 0$, denotadas respectivamente com, I : Gumbel, II : Fréchet, III : Weibull.

Conforme Beijo (2010), em 1955 Jenkison propôs que os três tipos de distribuição de valores extremos (Gumbel, Fréchet e Weibull) poderiam ser representados em uma forma paramétrica única, denominada distribuição Generalizada de Valores Extremos (GVE) que tem distribuição de função acumulada definida pela seguinte expressão

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

Definida no conjunto $\{x : x > \mu - \frac{\sigma}{\xi}\}$ em que os parâmetros satisfazem $-\infty < \mu < \infty$, $\sigma > 0$ e $-\infty < \xi < \infty$. O modelo consiste de três parâmetros: um parâmetro de locação, μ ; um parâmetro de escala, σ e um parâmetro de forma, ξ . As famílias de distribuições Tipo I, Tipo II e Tipo III de valores extremos correspondem, aos casos, $\xi = 0$, $\xi > 0$, $\xi < 0$ respectivamente.

Para o caso particular em que $\xi = 0$ a função de distribuição de distribuição acumulada Gumbel é definida da seguinte forma

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

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

Segundo Safins (2009), a unificação das três famílias originais de valores extremos simplifica bastante a implementação da estatística. Por meio da inferência em ξ os dados por si só determinam o tipo mais apropriado de cauda, e não há necessidade de se fazer julgamentos subjetivos *a priori* sobre qual tipo de família de valores extremos adotar.

Se a função de distribuição F é diferenciável, logo se define como uma função de densidade de probabilidade de X como

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

Derivando-se a expressão (2.1) em relação a x obtém-se a função de densidade de probabilidade da GVE. Assim tem-se a função de densidade de probabilidade da GVE é a seguinte

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

Derivando a Equação (2.2) em relação a x , no caso particular em $\xi \rightarrow 0$ que assim tem-se a função de densidade de probabilidade de Gumbel, é a seguinte

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

definida para $-\infty < x < \infty$ com $\mu \in \mathbb{R}$ e $\sigma > 0$.

2.2 Teste de aleatoriedade

Conforme Coles (2001), os métodos estatísticos na sua maioria, baseiam-se no pressuposto da aleatoriedade da amostra ou exigem que as observações tal como são recolhidas sejam independentes.

Conforme Silva_b (2008), verifica-se a hipótese da independência dos dados observados por meio do teste de chorrilho (*run test*) descrito por Zar (1999). Esse teste consiste na definição de uma sequência dicotômica de tamanho n , a partir de uma amostra aleatória X_1, \dots, X_n , aplicando a cada X_i a função indicadora $A(x_i) = I_{X_i > M_d}(X_i)$, omitindo-se os casos em que $X_i = M_d$, M_d sendo a mediana dos dados.

O valor da mediana, M_d , é dada a seguir

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

Os valores da 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 , com o valor observado r , definindo-se as variáveis aleatórias, N_1 o número total de ocorrências de $X_i > M_d$ e N_2 o número total de ocorrências de $X_i < M_d$, respectivamente, os valores observados são n_1 e n_2 . Com $n_1 < 30$ e $n_2 < 30$, seguindo-se pares de valores críticos exatos $(r_{1,\alpha,n_1,n_2}; r_{2,\alpha,n_1,n_2})$ para um nível de significância α , conforme os procedimentos descritos por Eisenhat e Swed (1943) e Browlee (1965). A hipótese nula é rejeitada se $r \leq r_{1,\alpha,n_1,n_2}$, ou se $r \leq r_{2,\alpha,n_1,n_2}$. Se $n_1 < 30$ e $n_2 < 30$, sob a hipótese H_0 de independência segue que, assintoticamente R segue uma distribuição normal com esperança definida por

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

com variância descrita na forma

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

e suas estimativas definidas por

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

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

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

2.3 Estimação dos parâmetros por máxima verossimilhança

Segundo Rocha (2004), o método da máxima verossimilhança consiste em adotar como estimador a estatística que maximiza a probabilidade, ou a densidade de probabilidade de ser encontrada a amostra observada. Dada uma distribuição de probabilidade com parâmetros desconhecidos, tem-se por finalidade, a partir uma determinada amostra estimar o valor do parâmetro da distribuição de probabilidade que representa toda a

população.

Segundo Beijo e Avelar (2010), a partir dos valores extremos de amostras, vários métodos podem ser utilizados para obtenção dos estimadores dos parâmetros da distribuição GVE. Inúmeras sugestões foram propostas, entre elas, técnicas gráficas, estimadores baseados no método dos momentos, métodos de estatísticas de ordem, método dos momentos de probabilidade ponderada, método de máxima verossimilhança, método de máxima verossimilhança penalizada e métodos Bayesianos. Nos casos regulares, que acontecem quando a distribuição em estudo depende de parâmetros desconhecidos, os estimadores de máxima verossimilhança são consistentes, assintoticamente normais e eficientes. Nos casos não regulares que ocorrem quando o suporte da distribuição depende de parâmetros desconhecidos. Conforme FERRARI (2011), isso ocorre com as distribuições GEV pois os limites dessas distribuições dependem dos parâmetros μ e σ e, dessa forma, as condições de regularidade para estimação pelo método da máxima verossimilhança não são satisfeitas e o método não deve ser aplicado automaticamente. Smith (1985) estudou cuidadosamente o comportamento assintótico dos estimadores de máxima verossimilhança para a distribuição GEV e obteve os seguintes resultados:

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

De acordo com Smith (1985), os métodos baseados em verossimilhança são preferidos devido a teoria dos estimadores de máxima verossimilhança ser bem compreendida e as inferências serem facilmente modificadas ao incorporarem-se modelos com estruturas mais complexas.

Sob a hipótese de que X_1, X_2, \dots, X_n são variáveis aleatórias independentes e identicamente distribuídas (*i.i.d*), os estimadores, por meio do método de máxima verossimi-

lhança, são apresentados da seguinte forma

$$L(\mu, \sigma, \xi|x) = \prod_{i=1}^n f_X(x_i|\mu, \sigma, \xi) =$$

$$= \sigma^{-n} \prod_{i=1}^n \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\xi}{\xi}\right)} \exp \left\{ - \sum_{i=1}^n \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}. \quad (2.5)$$

calculando-se o logaritmo da função de verossimilhança, que é definida por

$$l(\boldsymbol{\theta}|x) = \ln [L(\mu, \sigma, \xi|x)] =$$

$$= -n \ln(\sigma) - \left(\frac{1+\xi}{\xi} \right) \sum_{i=1}^n \ln \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^n \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}}$$

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

para $\xi < 0$, e $x_i < \mu - \frac{\sigma}{\xi}$ (ou seja, $\mu - \frac{\sigma}{\xi} > M_{(n)}$), ou para $\xi > 0$ $x_i > \mu - \frac{\sigma}{\xi}$ (ou seja, $\mu - \frac{\sigma}{\xi} > M_{(1)}$). Maximizando-se a Equação (2.6), com relação ao vetor de parâmetros $\boldsymbol{\theta} = (\mu, \sigma, \xi)$, leva a estimativa de máxima verossimilhança para toda família da GVE, conduzindo-se ao sistema de equações não lineares, definida a seguir

$$\frac{\partial l(\boldsymbol{\theta}|x)}{\partial \mu} = \frac{1}{\hat{\sigma}} \sum_{i=1}^n \left(\frac{1 + \hat{\xi} - \omega_i^{-\frac{1}{\hat{\xi}}}}{\omega_i} \right) = 0.$$

$$\frac{\partial l(\boldsymbol{\theta}|x)}{\partial \sigma} = -\frac{n}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^2} \sum_{i=1}^n \left\{ \frac{\left(1 + \hat{\xi} - \omega_i^{-\frac{1}{\hat{\xi}}} \right) (x_i - \hat{\mu})}{\omega_i} \right\} = 0. \quad (2.7)$$

$$\frac{\partial l(\boldsymbol{\theta}|x)}{\partial \xi} = \sum_{i=1}^n \left\{ \left(1 - \omega_i^{-\frac{1}{\hat{\xi}}} \right) - \left[\frac{\ln(\omega_i)}{\hat{\xi}} - \left(\frac{x_i - \hat{\mu}}{\hat{\xi} \hat{\sigma} \omega_i} \right) \right] - \frac{x_i - \hat{\mu}}{\hat{\sigma} \omega_i} \right\} = 0,$$

em que $\omega_i = 1 + \hat{\xi} \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right)$. No caso particular da distribuição Gumbel a função de verossimilhança, definida pela a Equação (2.4) em que $\xi = 0$, conduz ao logaritmo da função

é definida por

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

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

derivando-se a Equação (2.8), em que μ e σ são os estimadores de máxima verossimilhança, obtidos pelo o sistema de equações não lineares, ou seja,

$$\frac{\partial l(\boldsymbol{\theta}|x)}{\partial \mu} = \frac{1}{\hat{\sigma}} \left\{ \left[\sum_{i=1}^n \exp \left(-\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) \right] - n \right\} = 0.$$

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

Conforme Ferrari (2011), os sistemas de Equações (2.9), em geral, não possuem soluções exatas pois são equações não-lineares. Uma solução aproximada é calculada pelo método iterativo 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-se o critério de convergência do método.

2.4 Seleção da distribuição de valores extremos

Conforme Hosking (1984), uma das formas para testar se as observações seguem uma distribuição de valores extremos tipo I (*Gumbel*), é testar se $\xi = 0$ na distribuição GVE, utilizando-se o teste da razão de verossimilhança modificado descrito a seguir.

Uma série de n observações (x_1, x_2, \dots, x_n) , $l(\hat{\boldsymbol{\theta}}_{GVE})$ e $l(\hat{\boldsymbol{\theta}}_{Gumbel})$ os máximos dos logaritmos das funções de verossimilhança das distribuições GVE (2.1) e Gumbel (2.2) em que, $\hat{\boldsymbol{\theta}}_{GVE} = (\hat{\mu}, \hat{\sigma}, \hat{\xi})^T$ e $\hat{\boldsymbol{\theta}}_{Gumbel} = (\hat{\mu}, \hat{\theta})^T$, são vetores das estimativas de máxima verossimilhança.

A estatística de razão verossimilhança é descrita a seguir

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

com distribuição assintótica χ^2 com 1 grau de liberdade. Hosking (1984) propôs a utilização da estatística modificada obtendo-se uma aproximação mais precisa à distribuição assintótica para definida em (2.10), ou seja,

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

Desse modo, para testar a hipótese $H_0 : \xi = 0$ versus $H_1 : \xi \neq 0$, o valor da estatística teste T_{LR}^* deve ser comparado com o valor tabelado da distribuição do quadrado χ^2 , a um grau de liberdade e um nível de significância (α), $\chi_{[\alpha,1]}^2$. Se $T_{LR}^* \geq \chi^2$, rejeita-se a hipótese H_0 , ou seja, há indícios de que as observações não são de uma distribuição do tipo I (*Gumbel*).

O teste rejeitando a hipótese H_0 , resta partir para analisar por meio das estatísticas das seções posteriores duas outras distribuições de valores extremos tipo II (*Fréchet*) e tipo III (*Weibull*).

2.5 Diagnóstico do ajuste da distribuição da GVE

Conforme Ferrari (2011), para se testar a suposição de que os dados seguem uma distribuição GVE selecionada, pode-se recorrer à estatística D do teste de Kolmogorov-Smirnov, que é definida por

$$D = \max \left| F(x_{(i)}) - \hat{F}(x_{(i)}) \right|, i = 1, 2, \dots, n, \quad (2.12)$$

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

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

O gráfico qq-plot é constituído com os pontos dados pelas as coordenadas,

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

em que $\hat{\boldsymbol{\theta}}$ são as estimativas de $\boldsymbol{\theta} = (\mu, \sigma, \xi)'$, $F(x_{(i)})$ é a função (2.1) com suas estimativas

obtidas e $\hat{F}(x_{(i)})$ é uma distribuição empírica definida pela a função (2.12). Uma das formas de interpretar esse gráfico é observar o quanto esses pontos estão distantes da reta. Quanto maior a distância, menos adequada é a distribuição.

2.6 Estimações dos níveis de retorno da distribuição GVE

Conforme FERRARI (2011), o período de retorno (τ) é o intervalo de tempo estimado para ocorrência de um determinado evento e é definido como o inverso da probabilidade de um evento a ser igualado ou superado, ou seja,

$$\tau = \frac{1}{p},$$

em que p é a probabilidade do evento ser igualado ou ultrapassado ($P(X \geq x)$).

No caso em estudo, o período de retorno é o intervalo de tempo estimado para a ocorrência de temperatura máximas em São Gonçalo no sertão paraibano e é dado por

$$\tau = \frac{1}{1 - F(x)},$$

O nível de retorno (x_p), está associado ao período do retorno τ e a sua função é obtida por meio da solução da Equação abaixo:

$$\int_{i=1}^{x_p} f(\boldsymbol{\theta}) dx = 1 - p, \quad (2.13)$$

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

$$F(x_p) = (1 - p), \quad (2.14)$$

ao inverter a Equação (2.14), tem-se a solução

$$x_p = F^{-1}(1 - p) = \mu - \frac{\sigma}{\xi} \left\{ 1 - [-\ln(1 - p)]^{-\xi} \right\}. \quad (2.15)$$

Para $\xi \neq 0$, do qual o limite $\xi \rightarrow 0$ é definido a seguir

$$x_p = F^{-1}(1 - p) = \mu - \sigma \{ \ln[-\ln(1 - p)] \}. \quad (2.16)$$

De acordo com Medeiros (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 (2.15) e de μ e σ em (2.16).

2.7 Obtenção dos intervalos de confiança

Os Intervalos de Confiança (*I.C.*) com nível de $(1 - \alpha)100\%$ para os níveis de retorno x_p foram construídos e baseados no método delta, logo depois, no método estatístico de razão de verossimilhança. O intervalo de confiança para x_p com $(1 - \alpha)100\%$ de confiança é descrito a seguir:

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

em que α é o nível de significância, $z_{\alpha/2}$ o valor tal que $P(|Z| < z_{\alpha/2}) = 1 - \alpha$, Z uma variável com distribuição normal padronizada e $\text{Var}(\hat{x}_p)$ é a variância associada ao nível de retorno \hat{x}_p calculada através do método delta. Esse método é baseado no fato de que uma distribuição de $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}, \hat{\xi})'$ ser assintoticamente normal com média $\boldsymbol{\theta} = (\mu, \sigma, \xi)'$ e as matrizes de variâncias e covariância dado $\mathbf{I}(\boldsymbol{\theta})^{-1}$. Como a equação (2.15) é uma função não linear de μ, σ e ξ , pode-se linearizá-la por meio da expansão da primeira ordem de Taylor em torno do ponto inicial correspondente ao vetor das estimativas dos parâmetros.

Conforme Ferrari (2011), o método delta descrito por Rao e Toutenburg (1999), é realizado da seguinte forma: calcula-se $\text{Var}(\hat{x}_p)$ por meio da matriz de variâncias e covariâncias de μ, σ e ξ , estimada pela inversa da matriz de segundas derivadas da função log-verossimilhança (matriz hessiana calculada em μ, σ , e ξ). Assim o método delta estima a variância de \hat{x}_p por meio da expressão

$$\text{Var}(\hat{x}_p) \approx \nabla x_p' V \nabla x_p. \quad (2.17)$$

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

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

e

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

a matriz de derivadas parciais de x_p avaliada em μ, σ , e ξ .

Logo, a variância do nível de retorno estimado $x_{\hat{p}}$ para $\xi \neq 0$ ser calculada a seguir:

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

em que

$$\frac{\partial \hat{x}_p}{\partial \mu} = 1$$

$$\frac{\partial \hat{x}_p}{\partial \sigma} = -\frac{1}{\hat{\xi}} \left\{ 1 - [-\ln(1-p)]^{-\hat{\xi}} \right\},$$

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

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

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

com que $(\hat{\theta} = \hat{\mu}, \hat{\sigma})$ são as estimativas de máxima verossimilhança.

3 Resultados e discussão

Os dados utilizados neste estudo são referentes as temperaturas máximas mensais (expressos em °C) em uma série histórica no período de janeiro 1970 à dezembro de 2009 totalizando-se 39 anos, no município de São Gonçalo localizado no sertão paraibano. Pode-se observar por meio da Tabela 1 as estatísticas descritivas referentes a temperatura máxima mensal desse município.

Tabela 1: Estatística descritiva da variável aleatória temperatura máxima (°C) mensal no período entre 1970 a 2009, do município de São Gonçalo - PB.

Mês	Média	Mediana	Variância	D.P.	C.V.%	C.A.	Curtose
Janeiro	36,10	36,25	1,2152	1,1023	3,0409	-0,6003	0,5686
Fevereiro	34,76	34,70	2,6559	1,6297	4,6965	0,0163	-1,0821
Março	34,07	34,00	2,3087	1,5194	4,4689	0,9138	0,0388
Abril	33,38	33,00	2,2633	1,5044	4,5589	0,3084	-0,6353
Mai	33,13	32,50	2,6061	1,6143	4,9672	0,8155	-0,4531
Junho	33,32	33,25	2,4073	1,5515	4,6663	0,2824	-1,0304
Julho	34,05	34,15	2,7721	1,6649	4,8755	1,0145	2,9055
Agosto	35,00	35,00	0,8253	0,9084	2,5956	-0,4653	0,6074
Setembro	36,10	36,00	0,8323	0,9123	2,5343	0,7630	0,6326
Outubro	36,79	36,80	0,3623	0,6019	1,6357	-0,6649	1,1907
Novembro	36,60	36,60	0,5422	0,7363	2,0119	-0,5481	-0,1984
Dezembro	36,42	36,40	0,5286	0,7271	1,9975	-0,3413	0,1303

Com base nos dados referentes a Tabela 1, tem-se que os meses de agosto, setembro, outubro, novembro, dezembro e janeiro, apresentam em média os valores mais altos de temperatura máxima, no qual o mês de outubro apresenta-se, em média, o maior valor. Para os meses de agosto e novembro, a média e mediana são próximas, enquanto que nos meses de fevereiro a julho e setembro a mediana é menor que a média, este fato pode indicar uma possível assimetria à direita. Por outro lado, para os demais meses os dados aparentemente seguem uma distribuição assimétrica à esquerda. Por meio da Tabela 1 observa-se a curtose, grau de achatamento de uma curva de distribuição de frequência, do qual os meses de janeiro e de julho a outubro os dados estão em torno da moda, obtendo-se uma curva leptocúrtica, os demais resultados aproximam-se de uma curva platicúrtica.

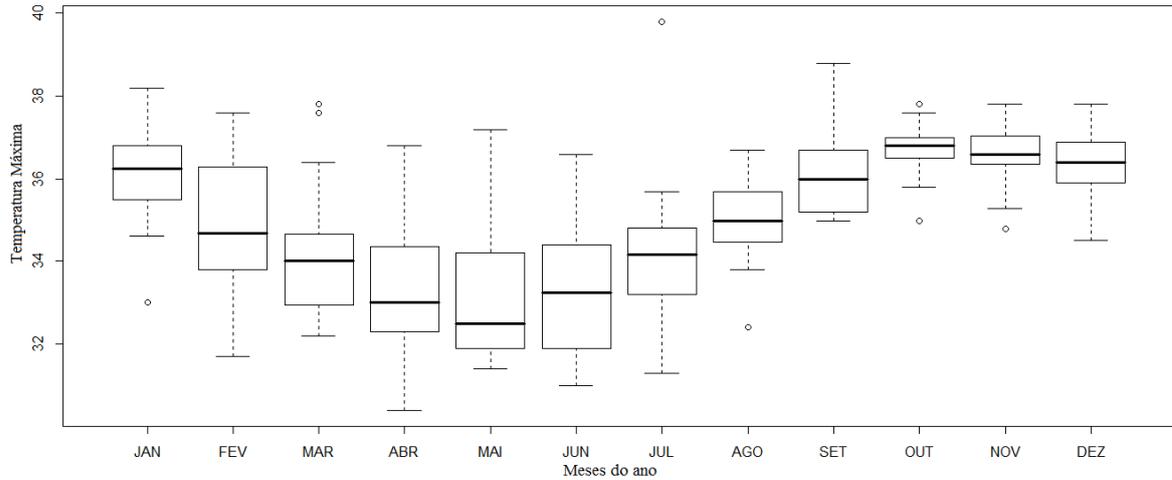


Figura 1: Gráfico de caixa (Box-Plot) referente a temperatura máxima no período de 1970 a 2009 no município de São Gonçalo.

Observa-se por meio da Figura 1, o gráfico box-plot, que por sua vez, refere-se aos dados da variável temperatura máxima de cada mês analisado, no qual pode-se perceber presença de alguns valores discrepantes (atípicos) representado pelo símbolo (o), nos respectivos meses de janeiro, março, julho, agosto, outubro e novembro, os demais não apresentam valores atípicos. Percebe-se que os meses de março, maio e julho tem uma considerável assimetria em relação os demais meses analisados.

O teste de chorrilho foi realizado para verificar a pressuposição de independência entre os dados de temperatura máxima, ao nível de 0,05 de significância, sendo o mesmo comparado com nível descritivo. Por meio da Tabela 2 pode-se observar que apenas no mês de outubro o nível descritivo é menor que o nível de significância, concluindo-se que a hipótese de independência entre os dados foi rejeitada. Nos demais meses não houve rejeição da hipótese, ou seja, a hipótese de independência não foi rejeitada ao nível de significância de 0,05.

Verificando-se a Tabela 2, em 11 dos 12 meses observados a hipótese de independência dos dados não foi rejeitada com nível de 5% de probabilidade. Conforme Medeiros (2011) o cumprimento dessa pressuposição garante a obtenção de inferências estatísticas satisfatórias a partir dos modelos de valores extremos. Após obter esses resultados calculou-se as estimativas dos estimadores pontuais dos três parâmetros localização, escala e forma respectivamente μ , σ e ξ da distribuição GVE, obtidos por meio do método de máxima verossimilhança, e suas variâncias e covariâncias estimadas para os respectivos meses es-

Tabela 2: Teste de chorrilho sob a pressuposição de independência dos dados entre os meses de janeiro a dezembro e o nível descritivo dos respectivos meses.

Meses	Valor-P
Janeiro	0,2298
Fevereiro	0,1093
Março	0,3293
Abril	0,5603
Maió	0,3293
Junho	0,7001
Julho	0,2479
Agosto	0,0848
Setembro	0,5840
Outubro	0,0052
Novembro	0,0781
Dezembro	0,6764

tudados.

Tabela 3: Valores observados das estimativas dos estimadores dos parâmetros da GVE e suas variâncias e covariâncias estimadas para os dados de temperatura máxima ($^{\circ}\text{C}$).

Mês	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$	$\text{Vâr}(\hat{\mu})$	$\text{Vâr}(\hat{\sigma})$	$\text{Vâr}(\hat{\xi})$	$\text{Côv}(\hat{\mu}, \hat{\sigma})$	$\text{Côv}(\hat{\mu}, \hat{\xi})$	$\text{Côv}(\hat{\sigma}, \hat{\xi})$
Janeiro	35,7900	1,1605	-0,4184	0,0605	0,0300	0,0113	-0,0093	-0,0095	-0,0114
Fevereiro	34,2694	1,6326	-0,3697	0,1369	0,0826	0,0369	0,0065	-0,0342	-0,0385
Março	33,3072	1,0034	0,1796	0,0549	0,0357	0,0477	0,0272	-0,0234	-0,0139
Abril	32,8077	1,3694	-0,1868	0,0879	0,0446	0,0213	0,0112	-0,0175	-0,0158
Maió	32,1377	0,8282	0,5423	0,0413	0,0390	0,0795	0,0320	-0,0261	-0,0105
Junho	32,7073	1,3791	-0,1615	0,0970	0,0561	0,0404	0,0243	-0,0315	-0,0297
Julho	33,3671	1,4270	-0,0837	0,0870	0,0405	0,0090	0,0144	-0,0085	-0,0057
Agosto	34,7530	0,9523	-0,4192	0,0397	0,0203	0,0135	-0,0056	-0,0086	-0,0102
Setembro	35,6918	0,7195	-0,0087	0,0275	0,0157	0,0342	0,0092	-0,0136	-0,0106
Outubro	36,6389	0,6404	-0,4665	0,0188	0,0099	0,0168	-0,0033	-0,0069	-0,0080
Novembro	36,4415	0,7895	-0,5278	0,0276	0,0167	0,0195	-0,0057	-0,0092	-0,0119
Dezembro	36,2087	0,7522	-0,3944	0,0280	0,0142	0,0155	-0,0031	-0,0078	-0,0091

Analisando-se o parâmetro de forma ($\hat{\xi}$), apresentado na Tabela 3, que indica quais das três distribuições (Gumbel, Fréchet e Weibull) os dados ajustam-se melhor, acerca deste parâmetro, percebe-se que as estimativas pontuais apresentam valores negativos e positivos, podendo-se assim testar o ajuste das três distribuições. Conforme Bautista (2002) observando-se as estimativas pontuais menores do que zero, há indícios de que os dados se ajustam a distribuição de Weibull, no entanto, este fato só poderá ser confirmado mediante análises posteriores.

A distribuição de Weibull, segundo Holmes e Moriarty (1999) é a mais adequada para representar fenômenos ambientais. No entanto para os meses de fevereiro, março, abril, junho, julho e setembro as estimativas pontuais estão próximas de zero correspondendo-se a distribuição Gumbel entranto este fato só sera consumado por meio do teste da razão de verossimilhança modificado T_{LR}^* . Enquanto que, no mês de maio, a estimativa

do parâmetro de forma é maior que zero, levando-se a distribuição Fréchet. Conforme Coles (2001), esta distribuição não é adequada para estudar o comportamento de alguns fenômenos ambientais, pois apresenta uma cauda superior com limite infinito.

O parâmetro de forma define qual tipo de distribuição de valores extremos se adequa melhor aos dados. A Tabela 4 apresenta os intervalos de confiança com 95% para o parâmetro de forma e a estatística da razão de verossimilhança modificada.

Tabela 4: Intervalos de 95% para o parâmetro de forma e os respectivos valores da estatística de verossimilhança modificada para os meses.

Mês	Limites de 95% de confiança para $\hat{\xi}$		T_{LR}^*
	Superior	Inferior	
Janeiro	-0,6275	-0,2093	11,5382
Fevereiro	-0,7466	0,0071	3,6763
Março	-0,2484	0,6077	1,1072
Abril	-0,4733	0,0996	1,4987
Maio	0,4325	1,0950	8,0265
Junho	-0,5558	0,2321	0,7227
Julho	-0,2703	0,1029	0,5810
Agosto	-0,6470	-0,1913	11,5072
Setembro	-0,3716	0,3541	0,0027
Outubro	-0,7210	-0,2119	14,1956
Novembro	-0,7210	-0,2119	11,6843
Dezembro	-0,6390	-0,1497	8,1557

Analisando-se os intervalos de confiança por meio da Tabela 4, para o parâmetro de forma (ξ), pode-se concluir que os meses de fevereiro, março, abril, junho, julho e setembro parecem seguir uma distribuição Gumbel, essa conclusão acerca dos dados é comprovada devido ao fato do valor nulo de ξ está contido no intervalo de confiança que corresponde a distribuição Gumbel, compreendido dentro dos limites do intervalo de 95% e também reforçadas pela estatística T_{LR}^* , calculada em (2.10), comparando-se o valor que se encontra na Tabela 4, com o valor tabelado da χ^2 ao nível de 0,05 significância ($\chi_{1;0,05}^2 = 3,85$).

Para os meses de janeiro, agosto, outubro, novembro e dezembro parecem seguir uma distribuição Weibull, pode-se observar que o intervalo de confiança com 95% significância contém todos os valores negativos para o parâmetro de forma, contemplando-se a confiabilidade desta conclusão acerca da distribuição ajustada aos dados de temperatura máxima, a distribuição de Fréchet ajustou-se apenas no mês de maio.

Dando sequência às análises, pode-se observar por meio da Tabela 5 as estimativas

dos estimadores dos dois parâmetros $\hat{\mu}$ e $\hat{\sigma}$, com suas respectivas variâncias e covariâncias. Baseando-se na distribuição Tipo I e na distribuição GVE, na Figura 2 foram construídos os gráfico de quantil-quantil para todos os meses em estudo. Analizando-se todos os meses, pode-se observar que o ajuste foi satisfatório, apenas nos meses de janeiro, outubro e dezembro percebe-se um maior desvio em relação ao ajuste dos quantis.

Tabela 5: Estimativa dos estimadores dos parâmetros $\hat{\mu}$ e $\hat{\sigma}$ da distribuição Gumbel com suas variâncias e covariâncias estimadas para os respectivos meses em análise.

Mês	$\hat{\mu}$	$\hat{\sigma}$	Vâr($\hat{\mu}$)	Vâr($\hat{\sigma}$)	Côv($\hat{\mu}, \hat{\sigma}$)
Fevereiro	35,528	1,247	0,067	0,014	0,045
Março	33,401	1,095	0,048	0,029	0,011
Abril	32,675	1,319	0,072	0,036	0,016
Junho	32,591	1,295	0,066	0,037	0,016
Julho	33,300	1,406	0,079	0,039	0,018
Setembro	35,688	0,716	0,021	0,012	0,005

Mediante resultados da Tabela 6 observa-se as informações do teste Kolmogorov-Smirnov ao nível de 5% de significância, encontra-se nesta Tabela as diferenças máximas absolutas observadas entre os valores probabilísticos das funções de distribuição empírica e teórica para cada mês observado, seguido dos níveis descritivos. Conforme o teste, as distribuições ajustam-se bem aos dados, pois, $D \geq D_{n,\alpha} = 0,20$ para todos os meses analisados. Na Figura 3 observa-se o esboço do teste de Kolmogorov Sminorv, que visualmente proporciona as mesmas conclusões citadas anteriormente.

Tabela 6: Resultados do teste de Kolmogorov-Smirnov para verificar o qualidade do ajuste da distribuição aos dados de temperatura máxima dos meses em estudo.

Mês	Diferença máxima absoluta (D)	Valor-p
Janeiro	0,1672	0,4656
Fevereiro	0,1231	0,8260
Março	0,1365	0,6954
Abril	0,1061	0,9214
Mai	0,1932	0,2658
Junho	0,1090	0,8932
Julho	0,1872	0,2802
Agosto	0,1736	0,3901
Setembro	0,1388	0,6987
Outubro	0,1986	0,0763
Novembro	0,1887	0,2916
Dezembro	0,1665	0,5191

A Tabela 7 apresenta as respectivas probabilidades de ocorrência de temperatura máxima acima de 34, 35, 36 e 37 °C entre os meses de janeiro a dezembro. Observa-se

que acima de 34 °C os meses de setembro, outubro, novembro e dezembro foram baixas a probabilidade de ocorrência, ou seja, não são tão expressivas. Por meio da Tabela 7 pode-se visualizar que acima de 37°C os meses de fevereiro a agosto foram registrados probabilidade bastante expressivas.

Tabela 7: Probabilidades de ocorrência de temperatura máxima mensal acima de 34, 35, 36 e 37 anos, correspondentes aos 12 meses de estudo no município de São Gonçalo-PB.

Mês	> 34	> 35	> 36	> 37
Janeiro	0,0373	0,1619	0,4367	0,7755
Fevereiro	0,3092	0,5416	0,7707	0,9287
Março	0,5583	0,7915	0,9105	0,9631
Abril	0,6930	0,8420	0,9226	0,9629
Mai	0,7945	0,8669	0,9068	0,9310
Junho	0,7138	0,8558	0,9305	0,9673
Julho	0,5443	0,7417	0,8635	0,9304
Agosto	0,1381	0,4677	0,8609	0,9999
Setembro	0,0002	0,0733	0,5233	0,8517
Outubro	0,0004	0,0045	0,1034	0,5946
Novembro	0,0019	0,0275	0,1954	0,6619
Dezembro	0,0008	0,0310	0,2722	0,7733

Na Tabela 8 encontram-se presentes as estimativas dos níveis de retorno mensais e os respectivos intervalos de confiança associados aos períodos de retorno 34, 35, 36 e 37 anos, obtidos pelo o método delta e por meio da equação 2.13. Por meio da Tabela 8 pode-se observar que o mês outubro terá os maiores registros de temperaturas nos períodos de retornos de 34, 35, 36 e 37 anos, respectivamente. Enquanto que em maio será o mês que ocorrerá os menores índices de temperaturas para os mesmos períodos de retorno.

Tabela 8: Níveis de retorno (\hat{x}_p °C) estimados e limites inferior (LI) e superior (LS) e seus respectivos intervalos de 95% confiança, para os períodos de retorno de 34, 35, 36 e 37 período de retorno, obtidos pelo o método delta

Mês	Período de retorno expresso em anos											
	34 anos			35 anos			36 anos			37 anos		
	LI	\hat{x}_p	LS	LI	\hat{x}_p	LS	LI	\hat{x}_p	LS	LI	\hat{x}_p	LS
Janeiro	36,91	37,53	38,15	37,59	38,40	39,21	39,01	40,31	41,62	41,51	43,80	46,09
Fevereiro	35,57	36,71	37,86	36,44	37,94	39,44	38,28	40,63	42,99	41,54	45,54	49,54
Março	34,28	35,05	35,81	34,88	35,87	36,85	36,17	37,67	39,18	38,46	40,96	43,47
Abril	33,77	34,65	35,53	34,51	35,64	36,77	36,12	37,82	39,52	39,00	41,79	44,58
Mai	32,44	33,38	34,31	32,78	34,00	35,21	33,52	35,36	37,21	34,85	37,85	40,86
Junho	33,65	34,53	35,40	34,38	35,50	36,63	35,93	37,64	39,35	38,71	41,53	44,36
Julho	34,48	35,41	36,33	31,87	36,46	41,05	32,67	38,78	44,90	34,71	43,01	53,32
Agosto	35,66	36,18	36,69	36,22	36,89	37,56	37,38	38,46	39,55	39,43	41,33	43,22
Setembro	36,26	36,76	37,26	36,65	37,30	37,94	37,49	38,48	39,47	9,00	40,63	42,43
Outubro	37,25	37,59	37,94	37,62	38,08	38,53	38,39	39,13	39,88	39,74	41,06	42,37
Novembro	37,19	37,62	38,05	37,64	38,21	38,79	38,56	39,52	40,47	40,20	41,89	43,58
Dezembro	36,89	37,33	37,77	37,32	37,90	38,47	38,22	39,14	40,06	39,80	41,40	43,00

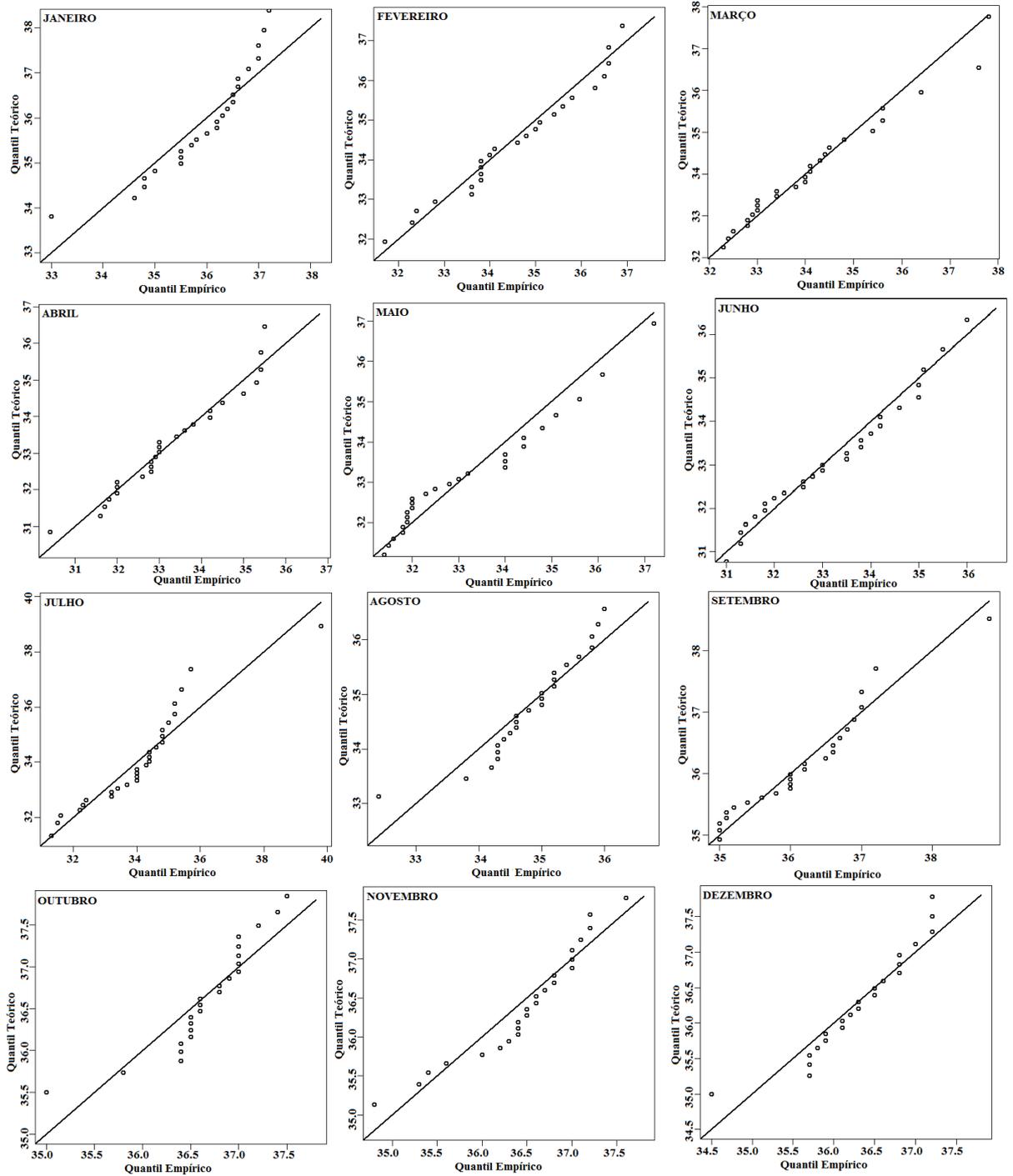


Figura 2: Gráficos de quantil-quantil, diagnóstico das distribuições para os dados de temperatura máxima mensal dos meses, janeiro, fevereiro, março, abril, maio, junho, julho, agosto, setembro, outubro, novembro e dezembro.

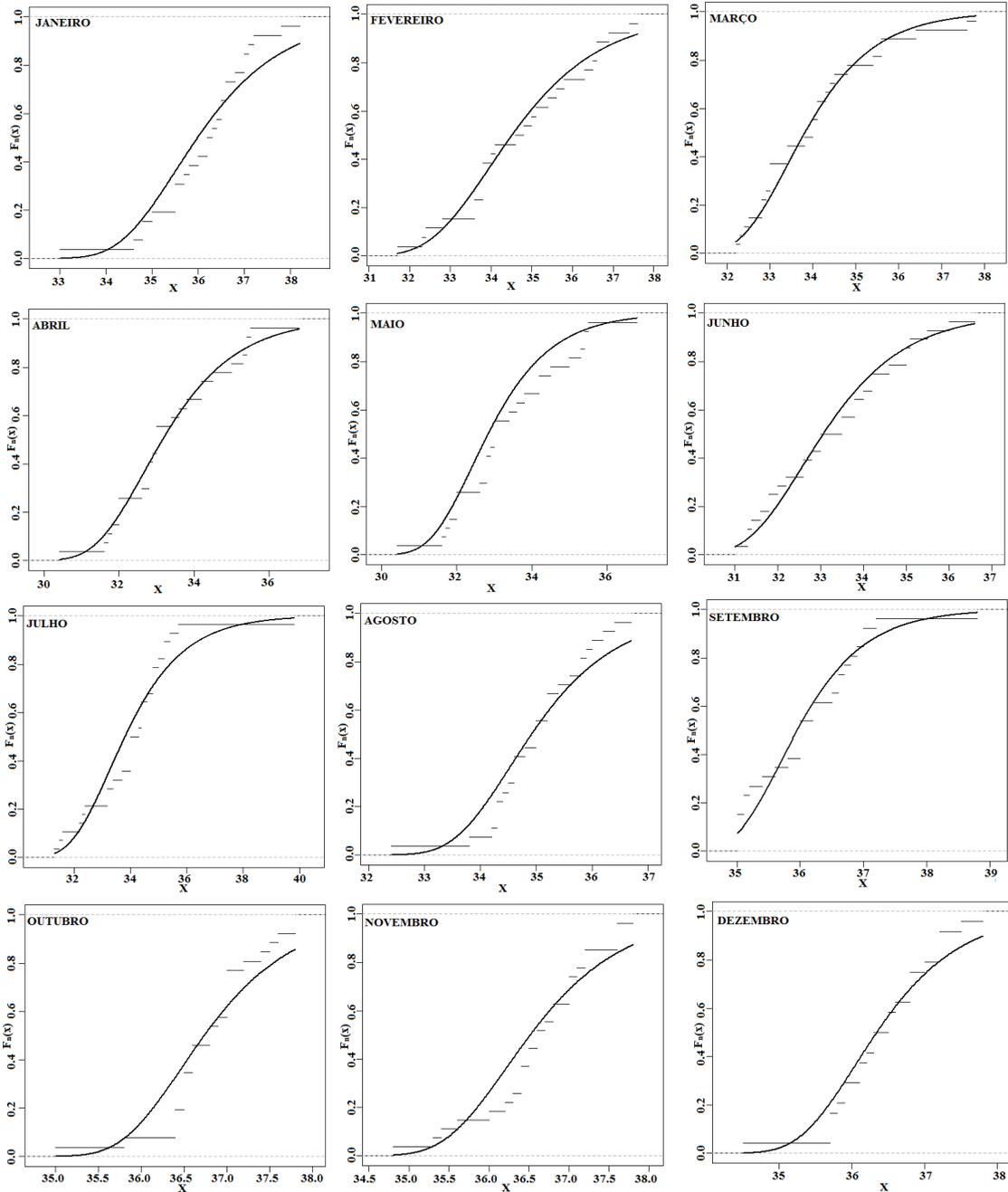


Figura 3: Gráficos do teste de Kolmogorov-Smirnov da função de distribuição acumulada empírica (representada pelo os tracinhos) e teórica (representada pela a curva) para diagnóstico dos modelos ajustados aos dados de temperatura máxima mensal.

4 Conclusão

A Teoria dos valores extremos, conforme foi abordada nesse estudo, pode solucionar vários problemas da sociedade nos dias atuais. Entretanto alguns problemas práticos em relação ao parâmetro de forma pode ocorrer, pois a estimativa pontual do mesmo não pode ultrapassar o limite inferior de $-0,5$ e superior de $0,5$, neste trabalho houve a ocorrência deste problema nos meses de maio e novembro. Para obtenção dos parâmetros das distribuições dos valores extremos, o método da máxima verossimilhança foi satisfatório quando aplicados aos dados de temperatura máxima.

As probabilidades de ocorrências de temperatura máxima, acima de 34°C nos meses de setembro a dezembro foram baixas, ou seja, não são tão expressivas. Por outro lado os meses de fevereiro a agosto foram registradas probabilidades bastante expressivas. Nos níveis de retorno, observou-se que a medida que o tempo médio do período de retorno aumenta a ocorrência dos máximos de temperatura tendem também a crescer.

5 Referências Bibliográficas

- BAUTISTA, E.L.B. A distribuição generalizada de valores extremos no estudo da velocidade máxima do vento em Piracicaba, SP. **Dissertação de Mestrado**, ESALQ, USP, Piracicaba estado de São Paulo. 2002. 61p.
- BEIJO, L.A.; AVELAR, F.G. A distribuição generalizada de valores extremos no estudo de dados climáticos extremos: uma breve revisão. **X Semana da Matemática e II Semana da Estatística**, Universidade Federal de Ouro Preto- UFOP. 2010. 70p.
- BORTKIEWICZ, L. Variationsbreite und mittlerer Fehler, Sitzungsber. **Berlin. Math. Ges.** v.21, p.3-11, 1922.
- BROWLEE, K.W. **Statical theory and methodology in science and engineering**. 2 ed. New York: John Wiley. 1965. 590p.
- COLES, S. **An Introduction to Statistical Modeling of Extreme Values**. Berlin: **Springer**, 2001. 208p.
- DODD, E. L. The greatest and least variate under general laws of error, **Trans. Amer. Math. SOC.** v.25, p.525-539, 1923.
- 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.
- FRÉCHET, M. Sur la loi de probilité de l' écart maximum, **Ann. Soc. Math.** v.6, p.93-116, 1927.
- FERRARI, G.T. Imputação de dados pluviométricos e sua aplicação na modelagem de eventos extremos de seca agrícola. **Dissertação de Mestrado**, ESALQ, Piracicaba estado de São Paulo. 2011. 70p.
- FISHER, R.A; TIPPETT, L.H.C. Limiting forms of the frenquency distribution of the largest or smallest member of a sample, **Procs. Cambridge Philos. SOC.** v.24, p.180-

190, 1928.

HOSKING, J. R. M. Testing whether the shape parameter is zero in the generalized extreme-value distribution. **Biometrika**. v.71, p.367-374, 1984.

HOLMES, J.D; MORIARTY, W.W. application of the generalized Pareto distribution to extreme value analysis in wind engineering. **Jour of Wind Engineering and Industrial Aerodynamics**, v.83, p.1-10, 1999.

JENKINSON, A.F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. **Quarterly Journal of the Royal Meteorological Society**, v.81, p.159-171, 1955.

MEDEIROS, E.S. Distribuição generalizada de valores extremos aplicada a dados de precipitação máxima na região de Morelândia- PE. **Trabalho de Conclusão de Curso de Bacharelado em Estatística**, UEPB, Campina Grande, Paraíba. 2011, 41p.

MISES, R., von. Über die Variationsbreite einer Beobachtungsreihe, Sitzungsber. **Berlin. Math. Ges.** v.22, p.3-8, 1923.

PEREIRA, V. P. P. L.; PINTO, F. C. Teoria de Valores Extremos: aplicações em Valor em Risco. Disponível em: <http://virtualbib.fgv.br/dspace/handle/>. Acesso em 16 set de 2012.

RAO, C.R.TOUTENBURG,H.**Linear models**. 2nd. ed. New York: Springer-Verlag, 1999. 443p.

ROCHA, A.L. Mercado de risco e a teoria dos valores extremos: estudo empírico dos casos. **Dissertação de Mestrado**, Universidade Federal do Ceará, Fortaleza. 2004, 89p.

RUBEN, A.P.S. Modelagem de extremos baseada nas r-maiores estatísticas de ordem: Uma aplicação do cálculo do valor em risco em mercados emergentes. **Dissertação de Mestrado**, Universidade Federal do Rio de Janeiro, Rio de Janeiro. 2006, 127p.

SAFINS, M.A.S. Cópulas para distribuição generalizadas de valores extremos multidimensionais. **Dissertação de Doutorado**, Universidade Federal do Rio de Janeiro, Rio de Janeiro. 2009, 108p.

SILVA_a, R.N.C. Caracterização estatística de extremos de processos sísmicos via distribuição generalizada de pareto. Estudo de caso: João Câmara-RN. **Dissertação de Mestrado**, Universidade Federal do Rio Grande do Norte, Rio Grande do Norte. 2008, 71p.

SILVA_b, R.R. A distribuição generalizada de Pareto e mistura de distribuição de Gumbel no estudo da razão e da velocidade máxima do vento em Piracicaba-SP. **Dissertação de Mestrado**, ESALQ, Piracicaba estado de São Paulo. 2008, 100p.

SMITH, R. L. Maximum likelihood estimation in class of nonregular cases. **Biometrika**, Cambridge, v.72, p.67-92, 1985.

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

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