

UNIVERSIDADE FEDERAL DE ALFENAS

NALVA APARECIDA MARTINS BATISTA

**MOMENTOS-L NA ESTIMAÇÃO DOS PARÂMETROS DA DISTRIBUIÇÃO
GENERALIZADA DE VALORES EXTREMOS**

**Alfenas/MG
2020**

NALVA APARECIDA MARTINS BATISTA

**MOMENTOS-L NA ESTIMAÇÃO DOS PARÂMETROS DA DISTRIBUIÇÃO
GENERALIZADA DE VALORES EXTREMOS**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística Aplicada e Biometria, pela Universidade Federal de Alfenas - MG, como parte dos requisitos para a obtenção do título de Mestre. Linha de Pesquisa: Modelagem Estatística e Estatística Computacional

Orientador: Prof. Dr. Fabricio Goecking Avelar

Coorientador: Prof. Dr. Luiz Alberto Beijo

**Alfenas/ MG
2020**

Dados Internacionais de Catalogação-na-Publicação (CIP)
Sistema de Bibliotecas da Universidade Federal de Alfenas
Biblioteca Central – Campus Sede

Batista, Nalva Aparecida Martins
B333m Momentos-L na estimação dos parâmetros da distribuição generalizada de
valores extremos / Nalva Aparecida Martins Batista – Alfenas, MG, 2021.
67 f.: il. –

Orientador: Fabricio Goecking Avelar.
Dissertação (Mestrado em Estatística Aplicada e Biometria) – Universidade
Federal de Alfenas, 2020.
Bibliografia.

1. Estatística. 2. Estatística matemática. 3. Amostragem (Estatística).
I. Avelar, Fabricio Goecking. II. Título.

CDD- 519.2

Ficha Catalográfica elaborada por Marlom Cesar da Silva
Bibliotecário-Documentalista CRB6/2735

MOMENTOS-L NA ESTIMAÇÃO DOS PARÂMETROS DA DISTRIBUIÇÃO GENERALIZADA DE VALORES EXTREMOS

A Banca examinadora abaixo-assinada aprova a Dissertação apresentada como parte dos requisitos para a obtenção do título de Mestre em Estatística Aplicada e Biometria pela Universidade Federal de Alfenas. Área de concentração: Estatística Aplicada e Biometria.

Aprovada em: 18 de dezembro de 2020.

Prof. Dr. Fabrício Goecking Avelar

Instituição: Universidade Federal de Alfenas - UNIFAL-MG

Prof. Dr. Gilberto Rodrigues Liska

Instituição: Universidade Federal de São Carlos - UFSCar

Prof. Dr. Thiago Taglialegna Salles

Instituição: Varacel Celulose



Documento assinado eletronicamente por **Fabrício Goecking Avelar, Professor do Magistério Superior**, em 21/12/2020, às 11:07, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Gilberto Rodrigues Liska, Usuário Externo**, em 21/12/2020, às 11:17, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Thiago Taglialegna Salles, Usuário Externo**, em 21/12/2020, às 11:22, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://sei.unifal-mg.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0439351** e o código CRC **27A337AA**.

Dedico primeiramente a Deus, senhor dos impossíveis, aos meus filhos Ryan e Renan fontes perseverança, ao meu esposo Nilson pela paciência e compreensão das minhas ausências durante os estudos, aos meus pais Joaquim e Lucia, aos meus irmãos, Afrânio, Enilda, Ivone, Sirlei, Janete, Adilson e todas as pessoas queridas que fazem parte de minha vida.

AGRADECIMENTOS

Primeiramente a Deus que me deu a oportunidade da vida e colocou em meu caminho pessoas maravilhosas que fizeram com esse projeto tornasse realidade.

À Universidade Federal de Alfenas e ao PPGAB pela oportunidade na realização do mestrado.

À toda minha família, especialmente, aos meus filhos e meu marido pela paciência e compreensão.

Em particular ao meu orientador, Prof. Dr. Fabricio Goecking Avelar, pelas orientações, incentivos e apoio em todos os momentos, pela sua amizade e por todo aprendizado vivenciado no dia a dia ao longo de todo curso.

Ao meu coorientador, Prof. Dr. Luiz Alberto Beijo, por compartilhar seus conhecimentos e incentivos, por fazer parte da construção desse trabalho.

Aos docentes do programa PPGAB pelos ensinamentos proporcionados e a todos os meus professores ao longo da vida acadêmica que propiciaram para esta conquista.

A todos meus companheiros (as) de mestrado pelo carinho e apoio nas horas difíceis e pelos momentos bons que passamos juntos.

Agradeço de coração a todos que de alguma maneira contribuíram para o êxito dessa conquista. Que Deus abençoe infinitamente a todos nós.

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

RESUMO

A Teoria de Valores Extremos (TVE) é uma importante ferramenta utilizada na análise de eventos extremos. A análise de tais eventos permite que haja um melhor planejamento nas ações humanas no sentido de minimizar os efeitos causados por eles. Uma distribuição proveniente da TVE utilizada na modelagem de eventos extremos é a distribuição generalizada de valores extremos (GEV). Vários métodos de estimação podem ser utilizados na estimação dos parâmetros da GEV. Dentre os mais utilizados são os de máxima verossimilhança. Contudo, esses estimadores não apresentam boas propriedades, ou não existem, quando o parâmetro forma da distribuição GEV é menor que $-0,5$. Nessas situações é necessário o uso de outro estimador. Nesse sentido o objetivo deste trabalho foi avaliar a adequabilidade do uso dos estimadores dos momentos-L no ajuste da distribuição GEV, comparando a precisão e a acurácia desses estimadores com a precisão e acurácia dos estimadores de máxima verossimilhança e utilizar os dois estimadores para ajustar a distribuição GEV a um conjunto de dados reais que contenha a estimativa do parâmetro forma entre -1 e $-0,5$. Verificou-se que para valores do parâmetro forma da GEV negativos o método dos momentos-L é indicado para tamanhos amostrais menores ou iguais a 25, com o parâmetro forma com valores próximos de zero o momentos-L é indicado para tamanhos amostrais menores ou iguais a 35 e para valores positivos do parâmetro forma, o momentos-L deve ser utilizado em todos os tamanhos amostrais.

Palavras-chave: Estimadores de Máxima Verossimilhança. Erro quadrático médio. Viés médio relativo. Tamanho amostral.

ABSTRACT

The Extreme Value Theory (EVT) is an important tool used in the analysis of extreme events. The analysis of such events allows for better planning in human actions in order to minimize the effects caused by them. A distribution from EVT that is widely used in extreme events modeling is the generalized extreme value distribution (GEV). Several estimation methods can be used to estimate the GEV distribution parameters. Among the most widely used estimators are the maximum likelihood estimators. However, these estimators are not suitable, or do not exist, when the shape parameter of the GEV distribution is less than -0.5. In these situations it is necessary to use another estimator. In this sense, the objective of this work was to evaluate the adequacy of the use of the L-moment estimators in the adjustment of the GEV distribution, comparing the precision and accuracy of these estimators with the precision and accuracy of the maximum likelihood estimators and to use these two estimators to adjust the GEV distribution to a set of real data that contains the estimate of the parameter form between -1 and - 0.5. It was found that for negative values of the shape parameter of the GEV the L-moment method is indicated for sample sizes less than or equal to 25, with the shape parameter with values close to zero the L-moments is indicated for sample sizes less than or equal to 35 and for positive values of the form parameter the L-moments should be used in all sample sizes.

Keywords: Maximum Likelihood Estimators. Mean squared error. Relative mean bias. Sample size.

LISTA DE FIGURAS

- Figura 1- Erro quadrático médio (EQM) dos estimadores de máxima verossimilhança e de momentos-L dos parâmetros posição (μ) e escala (σ) da distribuição GEV, considerando diferentes cenários e os tamanhos amostrais com 13, 15, 17, 20, 25, 35, 45, 60, 100, 150 e 200 observações.....46
- Figura 2-Viés médio relativo (VMR) dos estimadores de máxima verossimilhança e de momentos-L dos parâmetros posição (μ) e escala (σ) da distribuição GEV para os cenários, posição é igual a 2; escala 5; forma -0,75 e posição 19; escala 0,9; forma -0,46 e os tamanhos amostrais com 13, 15, 17, 20, 25, 35, 45, 60, 100, 150 e 200.....47
- Figura 3- Erro quadrático médio (EQM) e viés médio relativo (VMR) dos estimadores de máxima verossimilhança e de momentos-L do parâmetro forma (ξ), da distribuição GEV para diferentes cenários, e os tamanhos amostrais estudados.....48
- Figura 4- Viés médio relativo dos níveis de retorno obtidos pelo ajuste da distribuição GEV via MV e ML nos tempos de retorno de 10; 25; 50; 100; 200; 500 e 1000 anos, e com os tamanhos amostrais 17 e 13 para diferentes cenários.....49
- Figura 5-Teste Binomial e a proporção das amostras simuladas nos quais os estimadores de momentos-L apresentam erros quadráticos menores que os apresentados pelos estimadores de máxima verossimilhança, considerando os tamanhos amostrais estudado.....50

Figura 6-Erro quadrático médio dos estimadores de MV e de momentos-L dos parâmetros posição (μ) e escala (σ) da distribuição generalizada de valores extremos em diferentes cenários, para diferentes tamanhos amostrais.....	51
Figura 7-Viés médio relativo (VMR) dos estimadores de máxima verossimilhança e de momentos-L dos parâmetros posição (μ) e escala (σ) da distribuição GEV para cenários diferentes, nos respectivos tamanhos amostrais 13, 15, 17, 20, 25, 35, 45, 60, 100, 150 e 200	52
Figura 8-Erro quadrático médio (EQM) e viés médio relativo (VMR) dos estimadores de máxima verossimilhança e de momentos-L do parâmetro forma (ξ), da distribuição GEV para cenários diferentes e os tamanhos amostrais estudados.....	53
Figura 9- Viés médio relativo do nível de retorno obtidos no ajuste da GVE por meio dos estimadores de MV e ML para os tempos de retorno de 10; 25; 50; 100; 200; 500 e 1000 anos, no cenário ($\mu=37;\sigma=9;\xi=-0,16$) para os tamanhos amostrais $n=15$ e $n=25$	54
Figura10-Viés médio relativo do nível de retorno obtido no ajuste da distribuição GEV por meio dos métodos de MV e ML nos tempos de retorno 10, 25, 50, 100, 200, 500 e 1000 anos, em diferentes cenários nos tamanhos amostrais $n=45$ e $n=60$	55
Figura11-Teste binomial e a proporção das amostras simuladas nos quais os estimadores de momentos-L apresentam erros quadráticos menores que os apresentados pelos estimadores de máxima verossimilhança, considerando os diferentes tamanhos amostrais estudados.....	56

Figura12-Erro quadrático médio (EQM) dos estimadores de máxima verossimilhança e de momentos-L dos parâmetros posição (μ) e escala (σ) da distribuição GEV nos diferentes cenários, nos respectivos tamanhos amostrais 13, 15, 17, 20, 25, 35, 45, 60, 100, 150 e 200....57

Figura13-Viés médio relativo (VQM) dos estimadores de máxima verossimilhança e de momentos-L dos parâmetros posição (μ) e escala (σ) da distribuição GEV para os cenários, posição é igual a 19; escala 13; forma 0,4 e posição 19; escala 9; forma 0,5.....58

Figura14- Erro quadrático médio (EQM) e viés médio relativo (VMR) dos estimadores de MV e de momentos-L do parâmetro forma (ξ), da distribuição GEV para os cenários, posição é igual a 37; escala 13; forma 0,4 e posição 55; escala 9; forma 0,5 para os tamanhos amostrais 13, 15, 17, 20, 25, 35, 45, 60, 100, 150 e 200.....59

Figura15-Viés médio relativo do nível de retorno apresentado pelo ajuste da GEV por meio dos estimadores de MV e ML para os tempos de retorno de 10; 25; 50; 100; 200; 500 e 1000 anos, no cenário ($\mu=19; \sigma=13; \xi=0,4$) e ($\mu=19; \sigma=9; \xi=0,5$) para os tamanhos amostrais 150 e 200.....60

Figura16-Proporção das amostras simuladas nos quais os estimadores de momentos-L apresentam erros quadráticos menores que os apresentados pelos estimadores de máxima verossimilhança, considerando tamanhos amostrais estudados61

LISTA DE TABELAS

Tabela1-Parâmetros utilizados na composição dos cenários e os tamanhos amostrais utilizados na simulação.....	43
Tabela2-Estimativa dos parâmetros da distribuição GEV obtida pelos métodos de MV e momentos-L no ajuste da distribuição GEV aos dados de temperatura máxima de Franca-SP, no período de 1980 a 2012 e o resultado do teste de Kolmogorov-Smirnov (KS).....	62
Tabela3-Níveis de retorno (°C) para os tempos de retorno de 2, 4 e 6 anos da temperatura máxima anual de Franca-SP, calculadas via distribuição GEV no período de 1980 a 2012 e os erros percentual absoluto médio (EPAM).....	63

SUMÁRIO

1	INTRODUÇÃO GERAL.....	13
2	REFERENCIAL TEÓRICO	15
2.1	TEORIA DE VALORES EXTREMOS (TVE).....	15
2.2	DISTRIBUIÇÃO GENERALIZADA DE VALORES EXTREMOS	17
2.3	MÉTODO DA MÁXIMA VEROSSIMILHANÇA	19
2.4	MOMENTOS PONDERADOS POR PROBABILIDADE	23
2.5	MÉTODO MOMENTOS-L	27
2.6	PROPRIEDADES DOS ESTIMADORES	33
2.7	MOMENTOS-L NA ESTIMAÇÃO DOS PARÂMETROS DA DISTRIBUIÇÃO GENERALIZADA DE VALORES EXTREMOS	36
3	CONSIDERAÇÕES FINAIS.....	66
	REFERÊNCIAS.....	67

1 INTRODUÇÃO GERAL

Com aumento de fenômenos naturais extremos climáticos nos últimos tempos, as atividades humanas podem ser severamente afetadas causando prejuízos tanto financeiros quanto físicos e até mesmo com a perda de vidas humanas. São exemplos de eventos extremos: as precipitações máximas; as temperaturas mínimas e máximas; as inundações; os níveis elevados da poluição atmosférica; a velocidade máxima de ventos dentre outras. Neste sentido a teoria de valores extremos (TVE) visa auxiliar na análise desses fenômenos descrevendo e quantificando os seus comportamentos, buscando modelos que se ajustem a uma possível distribuição para os máximos (ou mínimos) de valores extremos a partir de uma amostra. A partir das informações obtidas das análises desses fenômenos extremos é possível ter a previsão de ocorrências dos mesmos, possibilitando a implementação de ações no sentido de amenizar os efeitos que estes podem causar.

Um modelo da TVE comumente utilizado na análise do comportamento de variáveis extremas é a distribuição generalizada de valores extremos do inglês “*Generalized Extreme Value*” (GEV), que unifica, em uma única forma paramétrica, as distribuições Gumbel, Fréchet e Weibull. A GEV tem sido amplamente utilizada, e mostrado resultado satisfatório em diversas áreas como hidrologia, engenharia, climatologia, economia, fenômenos meteorológicos, entre outros. A distribuição GEV possui três parâmetros μ, σ e ξ sendo μ parâmetro de posição, σ o parâmetro de escala e ξ o parâmetro de forma. Para que a distribuição GEV seja utilizada para modelar valores extremo é necessário estimar os seus parâmetros.

Para estimar os parâmetros da distribuição generalizada de valores extremos existem diversos métodos como, por exemplo, os estimadores de momentos, momentos-L e os estimadores de máxima verossimilhança entre outros. Como o parâmetro ξ determina o comportamento da cauda da distribuição é importante uma precisão na estimativa do mesmo, uma vez que, em determinados casos pode influenciar nas condições de regularidades dos estimadores. O método mais utilizado para estimar os parâmetros da GEV é o método da máxima verossimilhança, por possuir boas propriedades assintóticas. Porém em algumas

situações os estimadores de máxima verossimilhança não mantem as condições de regularidade. Mais especificamente, quando $-1 < \xi < -0,5$ os estimadores existem, mas não mantêm as condições de regularidade e quando $\xi < -1$, os estimadores de máxima verossimilhança não existem. Visto que, os estimadores de máxima verossimilhança são bons, mas como em alguns casos não mantêm regularidade, deve-se buscar alternativas para encontrar estimadores mais representativos possíveis e um estimador cujo comportamento busca-se avaliar são os estimadores dos momentos-L.

Neste contexto o objetivo desse trabalho foi avaliar e comparar a precisão e a acurácia dos estimadores de máxima verossimilhança e dos momentos-L dos parâmetros da distribuição generalizada de valores extremos para diferentes tamanhos amostrais. Também objetivou utilizar os estimadores para ajustar a distribuição GEV a um conjunto de dados reais que contenha a estimativa do parâmetro forma ente -1 e -0,5.

2 REFERENCIAL TEÓRICO

Nessa seção serão apresentadas as principais teorias desenvolvidas neste trabalho, compreendida nas seguintes subseções: 2.1 Teoria de Valores Extremos, 2.2 Distribuição Generalizada de Valores Extremos, 2.3 Método da máxima verossimilhança, 2.4 Método dos momentos ponderados por probabilidade, 2.5 Método momentos-L, 2.6 Propriedades dos estimadores e, na subseção 2.7, o artigo submetido à revista Ciência e Natura.

2.1 TEORIA DE VALORES EXTREMOS (TVE)

A TVE é um importante ramo da Estatística utilizada no estudo de eventos que apresentam baixa frequência relativa de ocorrência e podem ocasionar vários impactos (SILVA, 2008). Segundo Beijo (2002), a TVE é uma metodologia utilizada para descrever o comportamento dos pontos mais distantes da média de uma distribuição, as denominadas observações extremas. Como exemplo, o autor cita as precipitações máximas, temperaturas mínimas e máximas, inundações, poluição atmosférica extrema, velocidade máxima de ventos, dentre outros.

O desenvolvimento da TVE teve impulso em 1953 com o desastre ocorrido na Holanda. Por ser um país de delta baixo, boa parte de seu território está situado abaixo do nível do Mar. Para que a água do oceano não invadisse o território Holandês, foram construídos diques para impedir o avanço da água do mar. Entretanto, na madrugada de 31 de janeiro e 1º de fevereiro de 1953 a altura das ondas fez com os diques se rompessem causando destruição e mortes, esse fato levou o governo Holandês a entender que as previsões de eventos extremos tinham que ser melhoradas, pois os diques tinham que suportar as maiores ondas que pudessem acontecer. Foi então que, pensando em prever de forma confiável quais são as maiores ondas que poderiam acontecer, o governo criou um comitê para estudar o aquecimento global e suas consequências, e desenvolveu um sistema

anti-inundações visando maior segurança no planejamento da resistência desses diques (NOGUEIRA, 2016).

Os primeiros trabalhos referentes à teoria de valores extremos foram publicados há mais de 70 anos de acordo com Silva (2008). No artigo de 1927 Fréchet desenvolveu a distribuição assintótica do máximo de uma amostra aleatória, mas segundo Bautista (2002) os princípios referentes ao assunto da TVE inicialmente foram apresentados por Fischer e Tippett (1928), ao verificarem que as distribuições dos extremos para máximos em blocos convergiam para uma das três distribuições não degeneradas, definidas como distribuição assintótica dos valores extremos, conhecidas respectivamente como Gumbel (tipo I), Fréchet (tipo II) e de Weibull (tipo III). Além dos máximos em blocos uma variável aleatória extrema pode ser obtida utilizando a metodologia denominada picos acima de um valor limiar (*Peaks Over Threshold*) que consiste no ajuste de modelos probabilísticos baseados em séries históricas de dados, constituída somente por observações excedentes acima de um valor limiar (SILVA, 2008).

Em 1943 surge uma expressiva contribuição para o desenvolvimento da TVE com Gnedenko ao apresentar uma rigorosa fundamentação teórica necessária e suficiente para a convergência em distribuição do máximo ou do mínimo, caso estes não convirjam para uma distribuição degenerada¹ e determinou que essas distribuições poderiam ser representadas por distribuições contínuas.

Mas de acordo com Coles (2001) havia a princípio dificuldade em identificar qual das três famílias de distribuição de valores extremos seria mais apropriada na aplicação em determinada série de dados. Propondo resolver tal impasse surge outra contribuição de extrema relevância no desenvolvimento inicial da teoria de valores extremos com o trabalho de Jenkinson (1955), no qual o autor propõe a distribuição generalizada de valores extremos GEV. Neste artigo ele mostrou que as distribuições desenvolvidas por Fischer e Tippett (1928) podem ser unificadas como sendo casos particulares da GEV, ou seja, partindo das três formas assintóticas ele as unificou em uma única generalizada.

¹ De acordo com Nogueira (2016) uma função de distribuição é degenerada se assumir somente os valores 0 ou 1.

2.2 DISTRIBUIÇÃO GENERALIZADA DE VALORES EXTREMOS

A distribuição GEV tem sido utilizada em diversos estudos modelando os mais variados fenômenos e com isso, possibilitando prever e solucionar vários problemas nas mais diferentes áreas como, por exemplo, na hidrologia, na meteorologia, na engenharia, na economia e finanças, entre outras. Bautista (2002) aplicou a GEV na análise das velocidades máximas dos ventos. Arraes e Rocha (2006) estimaram perdas máximas esperadas para séries financeiras, empregando-se métodos tradicionais e metodologia dos Valores Extremos. Quadros, Queiroz e Boas (2011), analisaram séries históricas de precipitações máximas no município de Cascavel - PR e ajustaram as distribuições GEV e Gumbel, dentre as varias aplicabilidade da GEV.

De acordo com Jenkinson (1955) a GEV possui função densidade de probabilidade (f.d.p) expressa por:

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

definida em $-\infty < x < \mu - \sigma / \xi$ para $\xi < 0$, $\mu - \sigma / \xi < x < +\infty$ para $\xi > 0$ e $-\infty < x < +\infty$ quando $\xi \rightarrow 0$. Os parâmetros μ, σ, ξ representam os parametros de posição, escala e forma respectivamente, tal que $\mu \in \mathbb{R}, \sigma > 0$ e $\xi \in \mathbb{R}^*$. A GEV generaliza as três famílias de distribuições de valores extremos. Se $\xi > 0$ corresponde à distribuição Fréchet, se $\xi < 0$ corresponde à distribuição Weibull e por fim se $\xi \rightarrow 0$, a GEV é a distribuição de Gumbel (BEIJO; AVELAR, 2011).

Por sua vez, a distribuição de probabilidade acumulada (GEV) é expressa matematicamente por:

$$F(x) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\left(\frac{1}{\xi} \right)} \right\}, \quad (2)$$

com os parâmetros $-\infty < \mu < +\infty, \sigma > 0$ e $\xi \in \mathbb{R}^*$.

A função quantil da GEV é obtida a partir da inversa da expressão (2) e pode ser expressa por:

$$Q(y) = \mu + \frac{\sigma}{\xi} \left\{ -1 + [-\ln(y)]^{-\xi} \right\}, \quad 0 < y < 1. \quad (3)$$

Para calcular as probabilidades de ocorrência de uma determinada variável e de um fenômeno como, por exemplo, temperaturas máximas, velocidades máximas de ventos, precipitações máxima, dentre outros, acima de um valor x pré-estabelecido utiliza-se da equação:

$$P(X > x) = 1 - F(x) = 1 - \exp \left\{ - \left[1 + \xi \left(\frac{x - \hat{\mu}}{\hat{\sigma}} \right) \right]^{\frac{1}{\xi}} \right\}, \quad (4)$$

em que x é um valor da variável do período.

Bautista (2002) define período de retorno da seguinte maneira: seja E um evento e T o tempo aleatório entre as ocorrências consecutivas de eventos E . O valor médio da variável T é chamado de período de retorno do evento E . Logo, o período de retorno para esse evento, é dado por:

$$T = \frac{1}{P(E)} = \frac{1}{1 - F(x)}, \quad (5)$$

sendo T expresso geralmente em anos.

Assim de acordo como autor nível de retorno x_p da distribuição GEV, ou seja, o nível x_p , associado ao período de retorno T , é obtido a partir da solução da expressão:

$$\int_{-\infty}^{x_p} f(x | \xi, \sigma, \mu) dx = 1 - p \quad (6)$$

com $p = 1/T$, isto é,

$$F(x_p) = (1 - p) \quad (7)$$

Aplicando a função inversa F em ambos membros da equação (7) e relacionando com a probabilidade de ocorrências (da variável em estudo) (4), tem-se que:

$$\hat{x}_p = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}} \left\{ -1 + [-\ln(1-p)]^{-\hat{\xi}} \right\}, \quad (8)$$

Assim a estimativa \hat{x}_p do nível de retorno x_p para períodos de retorno $T = 1/p$ é obtido por substituição das estimativas de μ, σ , e ξ .

Vários métodos foram desenvolvidos propondo sugestões para a estimação dos parâmetros da GEV. Tais como máxima verossimilhança e método dos momentos, mas segundo Coles (2001) o método da máxima verossimilhança (MV) é mais aplicado devido a sua capacidade de se adaptar a distribuições mais complexas e por possuir boas propriedades assintóticas.

2.3 MÉTODO DA MÁXIMA VEROSSIMILHANÇA

Segundo Bolfarine e Sandoval (2001) a função de verossimilhança de uma variável aleatória pode ser definida como:

Definição 1. Sejam X_1, X_2, \dots, X_n uma sequência de variáveis aleatórias independentes e identicamente distribuídas (iid) de uma variável aleatória X e x_1, x_2, \dots, x_n uma amostra de tamanho n da variável aleatória X com função de densidade de probabilidade $f(x; \theta)$, com $\theta \in \Theta$, em que Θ é o espaço paramétrico. A função de verossimilhança de X é definida por:

$$L(\theta; x_1, x_2, \dots, x_n) = f(x_1; \theta) f(x_2; \theta) \dots f(x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) \quad (9)$$

O valor de θ que torna máxima $L(\theta; x_1, x_2, \dots, x_n)$ é a estimativa de máxima verossimilhança de θ . A função de verossimilhança de uma distribuição depende apenas dos seus parâmetros. Como os estimadores de máxima verossimilhança consistem em determinar os valores de θ que maximizam $L(\theta; \mathbf{x})$, em geral, esse processo é feito da seguinte forma: aplica-se o logaritmo natural na função de verossimilhança, encontrando-se assim a função suporte. Deriva-se parcialmente a função suporte em relação a cada parâmetro e resolve-se o sistema obtido ao se igualar cada derivada a zero (NOGUEIRA, 2016).

Aplicando-se esses procedimentos na função de verossimilhança da distribuição GEV, obtém-se:

$$\begin{aligned}
 L(\mu, \sigma, \xi | \mathbf{x}) &= \prod_{i=1}^n \frac{1}{\sigma} \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\xi}{\sigma} \right)} \exp \left\{ - \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1}{\xi} \right)} \right\} \\
 &= \prod_{i=1}^n \frac{1}{\sigma^n} \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\xi}{\sigma} \right)} \prod_{i=1}^n \exp \left\{ - \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1}{\xi} \right)} \right\} \\
 &= \frac{1}{\sigma^n} \prod_{i=1}^n \left\{ \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\xi}{\sigma} \right)} \right\} \prod_{i=1}^n \exp \left\{ - \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1}{\xi} \right)} \right\} \quad (10)
 \end{aligned}$$

Aplicando-se o logaritmo natural em ambos os membros da equação (10), obtém-se:

$$\begin{aligned}
 l(\mu, \sigma, \xi | \mathbf{x}) &= \ln \frac{1}{\sigma^n} \ln \left(\prod_{i=1}^n \left\{ \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\xi}{\sigma} \right)} \right\} \right) \ln \prod_{i=1}^n \exp \left\{ - \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1}{\xi} \right)} \right\} \\
 &= -n \ln \sigma - \sum_{i=1}^n \left(\frac{1+\xi}{\xi} \right) \ln \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^n \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1}{\xi} \right)} \quad (11)
 \end{aligned}$$

Derivando-se a expressão (11) em relação aos parâmetros μ, σ e ξ têm-se respectivamente que:

$$\begin{aligned}
 \frac{\partial l(\mu, \sigma, \xi | \mathbf{x})}{\partial \mu} &= \sum_{i=1}^n \frac{1+\xi}{\sigma \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]} - \sum_{i=1}^n \frac{1}{\sigma} \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\xi}{\xi} \right)} \\
 &= \sum_{i=1}^n \frac{1+\xi}{\sigma \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]} - \frac{1}{\sigma} \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\xi}{\xi} \right)}
 \end{aligned}$$

$$= \frac{1}{\sigma} \sum_{i=1}^n \left\{ \frac{1+\xi - \left[1+\xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{\left(\frac{1}{\xi} \right)}}{1+\xi \left(\frac{x_i - \mu}{\sigma} \right)} \right\}. \quad (12)$$

$$\begin{aligned} \frac{\partial l(\mu, \sigma, \xi | \mathbf{x})}{\partial \sigma} &= -\frac{n}{\sigma} + \sum_{i=1}^n \frac{(1+\xi)(x_i - \mu)}{\sigma^2 \left[1+\xi \left(\frac{x_i - \mu}{\sigma} \right) \right]} - \sum_{i=1}^n \frac{(x_i - \mu) \left[1+\xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{\left(\frac{1}{\xi} - 1 \right)}}{\sigma^2} \\ &= -\frac{n}{\sigma} + \frac{1}{\sigma^2} \sum_{i=1}^n \left\{ \frac{(1+\xi)(x_i - \mu) - (x_i - \mu) \left[1+\xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{\left(\frac{1}{\xi} \right)}}{\left[1+\xi \left(\frac{x_i - \mu}{\sigma} \right) \right]} \right\} \\ &= \frac{n}{\sigma} + \frac{1}{\sigma^2} \sum_{i=1}^n \left\{ \frac{(x_i - \mu) \left[(1+\xi) - \left[1+\xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{\left(\frac{1}{\xi} \right)} \right]}{\left[1+\xi \left(\frac{x_i - \mu}{\sigma} \right) \right]} \right\}. \quad (13) \end{aligned}$$

Agora, considerando $w_i = 1+\xi \left(\frac{x_i - \mu}{\sigma} \right)$ tem-se:

$$l(\mu, \sigma, \xi | \mathbf{x}) = -n \ln \sigma - \sum_{i=1}^n \left(\frac{1+\xi}{\xi} \right) \ln w_i - \sum_{i=1}^n w_i^{\left(\frac{1}{\xi} \right)} \quad (14)$$

$$\begin{aligned} \frac{\partial l(\mu, \sigma, \xi | \mathbf{x})}{\partial \xi} &= \sum_{i=1}^n \frac{1}{\xi^2} \ln w_i - \frac{1+\xi}{\xi} \left(\frac{x_i - \mu}{w_i} \right) - \sum_{i=1}^n \left(-\frac{1}{\xi} \right) w_i^{\left(\frac{1}{\xi} - 1 \right)} \left(\frac{x_i - \mu}{\sigma} \right) \\ &= \sum_{i=1}^n \frac{1}{\xi^2} \ln w_i - \frac{1+\xi}{\xi \sigma} \left(\frac{x_i - \mu}{w_i} \right) + \frac{1}{\xi \sigma} w_i^{\frac{1}{\xi} - 1} (x_i - \mu) \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n \frac{1}{\xi^2} \ln w_i - \frac{1}{\xi \sigma} \left(\frac{x_i - \mu}{w_i} \right) - \frac{1}{\sigma} \left(\frac{x_i - \mu}{w_i} \right) + \frac{1}{\xi \sigma} w_i^{-\frac{1}{\xi} - 1} (x_i - \mu) \\
&= \sum_{i=1}^n \frac{1}{\xi^2} \ln w_i - \frac{1}{\sigma} \left(\frac{x_i - \mu}{w_i} \right) - \frac{1}{\xi \sigma} \left(\frac{x_i - \mu}{w_i} \right) + \frac{1}{\xi \sigma} \left(\frac{x_i - \mu}{w_i} \right) w_i^{-\frac{1}{\xi}} \\
&= \sum_{i=1}^n \frac{1}{\xi^2} \ln w_i - \frac{1}{\sigma} \left(\frac{x_i - \mu}{w_i} \right) - \frac{1}{\xi \sigma} \left(\frac{x_i - \mu}{w_i} \right) \left[+1 - w_i^{-\frac{1}{\xi}} \right] \tag{15}
\end{aligned}$$

A partir das derivadas parciais em relação a cada um dos parâmetros da GEV obtém-se o seguinte sistema de equações:

$$\left\{ \begin{aligned}
&\frac{1}{\hat{\sigma}} \sum_{i=1}^n \left\{ \frac{1 + \hat{\xi} - [\hat{w}_i]^{-\left(\frac{1}{\hat{\xi}}\right)}}{\hat{w}_i} \right\} = 0 \\
&-\frac{n}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^2} \sum_{i=1}^n \left\{ \frac{(x_i - \hat{\mu}) \left[(1 + \hat{\xi}) - [\hat{w}_i]^{-\left(\frac{1}{\hat{\xi}}\right)} \right]}{\hat{w}_i} \right\} = 0 \\
&\sum_{i=1}^n \frac{1}{\hat{\xi}^2} \ln \hat{w}_i - \frac{1}{\hat{\sigma}} \left(\frac{x_i - \hat{\mu}}{\hat{w}_i} \right) - \frac{1}{\hat{\xi} \hat{\sigma}} \left(\frac{x_i - \hat{\mu}}{\hat{w}_i} \right) \left[+1 - \hat{w}_i^{-\left(\frac{1}{\hat{\xi}}\right)} \right] = 0
\end{aligned} \right. \tag{16}$$

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

Como o sistema (16) não possui soluções analíticas, assim os estimadores de máxima verossimilhança da distribuição GEV $(\mu, \sigma \text{ e } \xi)$, podem ser calculados utilizando métodos numéricos iterativos, tais como o método de Newton-Raphson recomendado por Vivanco (1994), o método “Nelder-Mead” proposto por Nelder e Mead (1965) e o método desenvolvido por Kirkpatrick, Gelatt e Vecchi em 1983, o *Simulated Annealing* (SANN).

Segundo Šimková e Picek (2017), os métodos baseados em verossimilhança geralmente são utilizados na estimação dos parâmetros da distribuição GEV. Contudo, as condições de regularidade para estimação dos parâmetros da distribuição generalizada de valores extremos pelo método da máxima

verossimilhança não são fundamentalmente satisfeitas. Existe regularidade nos casos em que $\xi > -0,5$. Quando $-1 < \xi < -0,5$ os estimadores existem, mas não satisfazem às condições de regularidades e para $\xi < -1$, os estimadores de máxima verossimilhança não existem (SMITH, 1985). Quando $-1 < \xi < -0,5$ e $\xi < -1$ pode haver necessidade do uso de outros estimadores para os parâmetros da distribuição GEV. Naghettini e Pinto (2007) afirmam que, além do MV e o método dos momentos (MM), os momentos-L (ML) e abordagens Bayesianas podem ser utilizados na estimação dos parâmetros da distribuição GEV.

Ainda de acordo com os autores o MV é considerado mais eficiente, pois produz os estimadores de menor variância. Mas em alguns casos a eficiência do MV é apenas assintótica, o que dificulta sua aplicação em amostras pequenas, ocasionando estimadores de qualidade comparável ou inferior a outros métodos. Neste caso, os estimadores ML mostram, com alguma frequência, mais acurados do que os de MV.

Comparando ML e MM, Hosking (1990) concluiu ML são menos sujeitos a viés nas estimativas. Pois, mesmo que o método dos momentos seja conceitualmente simples, e os cálculos sejam diretos, verifica-se que os valores numéricos da amostra podem ser muito diferentes daqueles da população da qual a amostra foi retirada, especialmente quando o tamanho da amostra é pequeno (SANKARASUBRAMANIAN; SRINIVASAN, 1999).

Os ML são baseados nos momentos ponderados por probabilidade (MPP), definidos por α_s e β_r determinados por $\alpha_s = E\{X[1-F(X)]^s\}$ ou $\beta_r = E\{X[F(X)]^r\}$, os momentos-L de ordem r , são combinações lineares dos correspondentes momentos ponderados por probabilidade (WOLFF, 2013).

2.4 MÉTODO DOS MOMENTOS PONDERADOS POR PROBABILIDADE (MPP)

Greenwood *et al.* (1979) introduziram os momentos ponderados por probabilidade (MPP). Definidos pela expressão:

$$M_{p,r,s} = E\left\{X^p [F_X(x)]^r [1-F_X(x)]^s\right\}, \quad (17)$$

em que X é uma variável aleatória contínua, p, r e s são números reais e F_X é a função de distribuição da variável aleatória X .

Aplicando-se a definição de Esperança na expressão (17), obtém-se:

$$M_{p,r,s} = \int_{-\infty}^{+\infty} x^p F_X(x)^r [1 - F_X(x)]^s f(x) dx. \quad (18)$$

Definindo-se $y = F_X(x)$, tem-se que $dy = f(x) dx$ e $F_X^{-1}(y) = x = Q(y)$. Se $x \rightarrow -\infty$, então $y \rightarrow 0$ e, se $x \rightarrow \infty$, então $y \rightarrow 1$. Portanto, realizando-se a mudança de variáveis $y = F_X(x)$ ao se aplicar o método da substituição na integral (18), obtém-se:

$$M_{p,r,s} = \int_0^1 Q(y)^p y^r (1-y)^s dy. \quad (19)$$

Em particular, segundo Naghettini e Pinto (2007), os momentos ponderados por probabilidade $M_{1,0,s}$ e $M_{1,r,0}$ são os mais utilizados na caracterização de distribuições de probabilidade, eles podem ser expressos por:

$$M_{1,0,s} = \alpha_s = \int_0^1 Q(y)(1-y)^s dy \quad (20)$$

$$M_{1,r,0} = \beta_r = \int_0^1 Q(y)y^r dy, \quad (21)$$

em que $Q(y) = F_X^{-1}(y)$ é a função quantil de uma variável X .

Hosking (1986) demonstrou que os métodos dos momentos ponderados por probabilidade MPP, possuem generalidade suficiente para a estimação de parâmetros de distribuições de probabilidade, estão menos sujeitos a flutuações amostrais e são mais robustos do que os correspondentes momentos convencionais (NAGHETTINI; PINTO, 2007).

Greenwood *et al.* (1979), a partir de uma amostra aleatória $(x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)})$ ordenada crescente de tamanho n , retirada de uma determinada população, definiram os momentos ponderados por probabilidade amostrais por:

$$\hat{a}_s = \frac{1}{n} \sum_{i=1}^n \frac{\binom{n-i}{s}}{\binom{n-1}{s}} x_i, \quad (22)$$

$$\hat{b}_r = \frac{1}{n} \sum_{i=1}^n \frac{\binom{n-i}{r}}{\binom{n-1}{r}} x_i, \quad (23)$$

sendo $i=1,2,\dots,n$; $s=0,\dots,n-1$.

Mesmo os métodos ponderados por probabilidades α_s e β_r , ser passíveis de serem utilizados na estimação de parâmetros, não possuem uma maneira fácil de interpretação dos descritores de escala e forma das distribuições de probabilidades (NAGHETTINI; PINTO, 2007).

Buscando resolver tal problema Hosking (1990) apresentou o conceito de momentos-L, que são funções lineares dos momentos ponderados por probabilidade (MPP), ou seja, combinações lineares dos α_s e β_r os quais possuem grandezas diretamente interpretáveis como descritores de escala e forma das distribuições de probabilidade. Assim os momentos-L populacionais λ_r são definidos formalmente por:

$$\lambda_r = (-1)^{r-1} \sum_{k=0}^{r-1} \rho_{r-1,k} \alpha_k = \sum_{k=0}^{r-1} \rho_{r-1,k} \beta_k, \quad (24)$$

em que

$$\rho_{r-1,k} = (-1)^{r-k-1} \binom{r-1}{k} \binom{r+k-1}{k}. \quad (25)$$

Aplicando-se a equação (24) tem-se, para $r=1,2,3$, que:

$$\lambda_1 = (-1)^0 \sum_{k=0}^0 \rho_{0,k} \alpha_k = \rho_{0,0} \alpha_0; \quad (26)$$

$$\lambda_2 = (-1)^1 \sum_{k=0}^1 \rho_{1,k} \alpha_k = (-1) [\rho_{1,0} \alpha_0 + \rho_{1,1} \alpha_1]; \quad (27)$$

$$\lambda_3 = (-1)^2 \sum_{k=0}^2 \rho_{2,k} \alpha_k = (-1)^2 [\rho_{2,0} \alpha_0 + \rho_{2,1} \alpha_1 + \rho_{2,2} \alpha_2]. \quad (28)$$

Procedendo aos cálculos utilizando a fórmula (25) tem-se que:

$$\rho_{0,0} = (-1)^0 \binom{0}{0} \binom{0}{0} = 1; \quad (29)$$

$$\rho_{1,0} = (-1)^1 \binom{1}{0} \binom{1}{0} = -1; \quad (30)$$

$$\rho_{1,1} = (-1)^0 \binom{1}{1} \binom{2}{1} = 2. \quad (31)$$

De modo análogo, obtêm-se os valores $\rho_{2,0} = 1$, $\rho_{2,1} = -6$, $\rho_{2,2} = 6$.

Logo, os Momentos-L populacionais de ordem inferior a quatro são dados por:

$$\lambda_1 = \alpha_0; \quad (32)$$

$$\lambda_2 = \alpha_0 - 2\alpha_1; \quad (33)$$

$$\lambda_3 = \alpha_0 - 6\alpha_1 + 6\alpha_2. \quad (34)$$

Os momentos-L amostrais são denotados por l_r e são calculados pela substituição de α_s ou β_r nas equações (32) a (34), pelas suas estimativas a_s ou b_r .

$$l_1 = a_0; \quad (35)$$

$$l_2 = a_0 - 2a_1; \quad (36)$$

$$l_3 = a_0 - 6a_1 + 6a_2. \quad (37)$$

De acordo com Gubareva e Gartsman (2010) o primeiro momento- L se equipara uma medida de tendência central, as demais medidas da distribuição, tais como variância, assimetria e curtose, são avaliados por meio de razões de momentos-L de ordem superior, τ é uma medida equivalente ao coeficiente de variação dado pelo quociente de λ_2 e λ_1 e as demais razões do momento-L são definidos por: $\tau_r = \lambda_r / \lambda_2$, $r = 3,4...$ Segundo Wolff (2013) para obter as estimativas dos parâmetros da distribuição probabilidade em estudo pelo método dos momentos-L, deve-se igualar os momentos-L populacionais aos momentos-L amostrais, ou seja,

$$\begin{cases} \lambda_1 = l_1 \\ \lambda_2 = l_2 \\ \tau_3 = t_3 \end{cases} \Rightarrow \begin{cases} \lambda_1 = a_0 \\ \lambda_2 = a_0 - 2a_1 \\ \tau_3 = \frac{a_0 - 6a_1 + 6a_2}{a_0 - 2a_1} \end{cases} . \quad (38)$$

Desse modo tem-se que λ_1 é equivalente à média da distribuição e, portanto, uma medida populacional de posição, λ_2 uma medida de dispersão (análogo ao desvio padrão ordinário), a razão $CV-L = \tau = \lambda_2 / \lambda_1$ é análogo ao coeficiente de variação convencional (o qual pode ser interpretado como uma medida populacional de dispersão ou de escala) e a razão $\tau_3 = \lambda_3 / \lambda_2$ representa uma medida de assimetria (NAGHETTINI; PINTO, 2007).

2.5 MÉTODO MOMENTOS-L

Para encontrar os estimadores de momentos-L dos parâmetros da GEV é necessário encontrar os momentos ponderados por probabilidade populacionais α_s que, pela expressão (20) são obtidos pelo cálculo das integrais,

$$\alpha_s = \int_0^1 Q(y)(1-y)^s dy \quad (39)$$

Como $Q(y) = F_X^{-1}(y) = \frac{\sigma}{\xi} \left\{ -1 + [-\ln(y)]^{-\xi} \right\} + \mu$. Desse modo o primeiro momento ponderado por probabilidade populacional α_0 é dado pelo cálculo da integral:

$$\alpha_0 = \int_0^1 \left\{ \frac{\sigma}{\xi} \left\{ -1 + [-\ln(y)]^{-\xi} \right\} + \mu \right\} dy \quad (40)$$

Pela propriedade da soma tem-se que:

$$\alpha_0 = \int_0^1 \frac{\sigma}{\xi} \left\{ -1 + [-\ln(y)]^{-\xi} \right\} dy + \int_0^1 \mu dy \quad (41)$$

Definindo-se $v = -\ln(y)$, então $y = \exp(-v)$ e $dv = -dy/y$, conseqüentemente $dv = -ydv$. Aplicando a regra de substituição no primeiro membro da expressão (41) obtém-se o resultado:

$$\begin{aligned}
 \int_{\infty}^0 \left\{ \frac{\sigma}{\xi} [-1 + (v^{-\xi})] \right\} \exp(-v) dv &= \int_{\infty}^0 \frac{\sigma}{\xi} \exp(-v) dv + \int_{\infty}^0 -\frac{\sigma}{\xi} v^{-\xi} \exp(-v) dv \\
 &= \int_0^{\infty} \frac{\sigma}{\xi} \exp(-v) dv + \int_0^{\infty} \frac{\sigma}{\xi} v^{-\xi} \exp(-v) dv \\
 &= -\frac{\sigma}{\xi} + \frac{\sigma}{\xi} \int_0^{\infty} v^{-\xi+1} \exp(-v) dv \\
 &= -\frac{\sigma}{\xi} + \frac{\sigma}{\xi} \Gamma(-\xi+1) \\
 &= \frac{\sigma}{\xi} [-1 + \Gamma(-\xi+1)]^2 \tag{42}
 \end{aligned}$$

Acrescentando ao resultado (42) o resultado da integral $\int_0^1 \mu dy$ obtém-se:

$$\alpha_0 = \frac{\sigma}{\xi} [-1 + \Gamma(-\xi+1)] + \mu. \tag{43}$$

O segundo momento ponderado por probabilidade populacional, ou seja, α_1 é dado por:

$$\alpha_1 = \int_0^1 \left\{ \frac{\sigma}{\xi} [-1 + (-\ln(y))^{-\xi}] + \mu \right\} (1-y) dy \tag{44}$$

Da expressão (44) tem-se que:

$$\alpha_1 = \int_0^1 \left\{ \frac{\sigma}{\xi} [-1 + (-\ln(y))^{-\xi}] + \mu \right\} dy - \int_0^1 \left\{ \frac{\sigma}{\xi} [-1 + (-\ln(y))^{-\xi}] + \mu \right\} y dy \tag{45}$$

Note que a solução da primeira parcela da soma da expressão (45) foi calculada anteriormente e o resultado é expresso pela equação (43). Logo falta obter o resultado da expressão:

$$-\int_0^1 \left\{ \frac{\sigma}{\xi} [-1 + (-\ln(y))^{-\xi}] + \mu \right\} y dy \tag{46}$$

² $\Gamma(\cdot)$ Função gama denotada por $\Gamma(y) = \int_0^{\infty} t^{y-1} \exp(-t) dt$.

Definindo-se $v = -\ln(y)$, então $y = \exp(-v)$ e $dv = -dy/y$. Consequentemente $dy = ydv$, assim $ydy = -y^2dv$ e, portanto $ydy = -\exp(-2v)$. Realizando as devidas substituições na expressão (46) tem-se

$$\begin{aligned} -\int_{\infty}^0 \left\{ \frac{\sigma}{\xi} [-1+v^{-\xi}] + \mu \right\} \exp(-2v) dv &= \int_0^{\infty} \left\{ \left[\frac{\sigma}{\xi} [-1+v^{-\xi}] + \mu \right] \right\} \exp(-2v) dv \\ &= \int_0^{\infty} \frac{\sigma}{\xi} \exp(-2v) dv - \int_0^{\infty} \frac{\sigma}{\xi} v^{-\xi} \exp(-2v) dv - \int_0^{\infty} \mu \exp(-2v) dv \end{aligned} \quad (47)$$

Da expressão (47) tem-se que $\int_0^{\infty} \frac{\sigma}{\xi} \exp(-2v) dv = \frac{\sigma}{2\xi}$ e $-\int_0^{\infty} \mu \exp(-2v) dv = -\frac{\mu}{2}$, e

para o cálculo de:

$$-\int_0^{\infty} \frac{\sigma}{\xi} v^{-\xi} \exp(-2v) dv. \quad (48)$$

Defina $w = 2v$ então $w = 2dv \Rightarrow dv = \frac{dw}{2}$, isolando $v = \frac{w}{2}$ então

$v^{-\xi} = \left(\frac{w}{2}\right)^{-\xi} = \frac{w^{-\xi}}{2^{-\xi}}$. Realizando as substituições necessárias na equação (48) obtém-se:

se:

$$-\int_0^{\infty} \frac{\sigma}{\xi} v^{-\xi} \exp(-w) \frac{dw}{2} = -\frac{\sigma}{2\xi} \int_0^{\infty} v^{-\xi+1-1} \exp(-w) dw. \quad (49)$$

$$\begin{aligned} -\int_0^{\infty} \frac{\sigma}{\xi} \frac{w^{-\xi+1-1}}{2^{-\xi}} \exp(-w) \frac{dw}{2} &= -\frac{\sigma}{\xi} \frac{1}{2^{-\xi+1}} \int_0^{\infty} w^{-\xi+1-1} \exp(-w) dw \\ &= -\frac{\sigma}{\xi} \frac{1}{2^{-\xi+1}} \Gamma(-\xi+1). \end{aligned} \quad (50)$$

Portanto o resultado da expressão (47) é a solução da integral expressa na equação (46) cujo resultado é dado por:

$$\frac{\sigma}{2\xi} - \frac{\sigma}{\xi} \frac{1}{2^{-\xi+1}} \Gamma(-\xi+1) - \frac{\mu}{2}. \quad (51)$$

Logo o segundo Momento Ponderado por Probabilidade Populacional é obtido pela soma dos resultados das expressões (43) e (51), ou seja,

$$\alpha_1 = -\frac{\sigma}{\xi} + \frac{\sigma}{\xi} \Gamma(-\xi+1) + \mu + \frac{\sigma}{2\xi} - \frac{\sigma}{\xi} \frac{1}{2^{-\xi+1}} \Gamma(-\xi+1) - \frac{\mu}{2}$$

$$\begin{aligned}
&= -\frac{\sigma}{2\xi} + \frac{\mu}{2} + \frac{\sigma}{\xi} \Gamma(-\xi+1) \left(1 - \frac{1}{2^{-\xi+1}}\right) \\
&= -\frac{\sigma}{2\xi} + \frac{\mu}{2} + \frac{\sigma}{\xi} \Gamma(-\xi+1) \left[\frac{2^{-\xi+1}-1}{2^{-\xi+1}}\right].
\end{aligned} \tag{52}$$

Por fim α_2 é dado por:

$$\alpha_2 = \int_0^1 \left\{ \frac{\sigma}{\xi} \left[-1 + (-\ln(y))^{-\xi} \right] + \mu \right\} (1-y)^2 dy, \tag{53}$$

ou seja,

$$\int_0^1 \left\{ \frac{\sigma}{\xi} \left[-1 + (-\ln(y))^{-\xi} \right] + \mu \right\} (1-2y+y^2) dy. \tag{54}$$

Assim tem-se que a expressão (54) pode ser descrita por:

$$\int_0^1 Q(y) dy - 2 \int_0^1 Q(y) y dy + \int_0^1 Q(y) y^2 dy \tag{55}$$

Desse modo como já foram apresentadas as soluções da primeira e da segunda soma da expressão (55), na qual os resultados foram expressos nas soluções (43) e (51) respectivamente. Então para obter o α_2 , deve-se multiplicar o resultado da segunda expressão na equação (55) por -2 , em seguida calcular a terceira integral da expressão (55) e somar todos os resultados dessa expressão. Procedendo ao cálculo da terceira parte da expressão (55) tem-se que:

$$\int_0^1 \left\{ \frac{\sigma}{\xi} \left[-1 + (-\ln(y))^{-\xi} \right] + \mu \right\} (y^2) dy \tag{56}$$

Definindo-se $v = -\ln(y)$, então $y = \exp(-v)$ e $dv = -dy/y$. Conseqüentemente $dy = -ydv$, assim $y^2 dy = -y^3 dv$. Aplicando as substituições na expressão (56) tem-se:

$$\begin{aligned}
& \int_0^{\infty} \left\{ \frac{\sigma}{\xi} [-1 + (-v)^{-\xi} + \mu] \right\} \exp(-3v) dv = \\
& = \int_{\infty}^0 \frac{\sigma}{\xi} \exp(-3v) dv + \frac{\sigma}{\xi} \int_{\infty}^0 -v^{-\xi} \exp(-3v) dv + \int_{\infty}^0 -\mu \exp(-3v) dv \\
& = -\int_0^{\infty} \frac{\sigma}{\xi} \exp(-3v) dv + \frac{\sigma}{\xi} \int_0^{\infty} v^{-\xi} \exp(-3v) dv + \int_0^{\infty} \mu \exp(-3v) dv \quad (57)
\end{aligned}$$

Da expressão (57) tem-se

$$-\int_0^{\infty} \frac{\sigma}{\xi} \exp(-3v) dv = -\frac{\sigma}{3\xi} \text{ e } \int_0^{\infty} \mu \exp(-3v) dv = \frac{\mu}{3}, \quad (58)$$

para o cálculo de $\frac{\sigma}{\xi} \int_0^{\infty} v^{-\xi} \exp(-3v) dv$, Defina-se $w = 3v$ então $dw = 3dv$,

Conseqüentemente $dv = dw/3$, isolando v tem-se $v = w/3$, de modo que

$v^{-\xi} = \left(\frac{w}{3}\right)^{-\xi} = \frac{w^{-\xi}}{3^{-\xi}}$. Realizando devidas substituições em $\frac{\sigma}{\xi} \int_0^{\infty} v^{-\xi} \exp(-3v) dv$ obtém-se:

$$\frac{\sigma}{\xi} \int_0^{\infty} \frac{w^{-\xi}}{3^{-\xi}} \exp(-w) dw = \frac{\sigma}{\xi} \frac{1}{3^{-\xi}} \int_0^{\infty} w^{-\xi+1-1} \exp(-w) dw = \frac{\sigma}{\xi} \frac{1}{3^{-\xi}} \Gamma(-\xi+1) \quad (59)$$

Logo o resultado da expressão (56) é dado por:

$$-\frac{\sigma}{3\xi} + \frac{\mu}{3} + \frac{\sigma}{\xi} \frac{1}{3^{-\xi}} \Gamma(-\xi+1) \quad (60)$$

Desse modo tem-se que:

$$\alpha_2 = -\frac{\sigma}{\xi} + \frac{\sigma}{\xi} \Gamma(-\xi+1) + \mu - 2 \left(\frac{\sigma}{2\xi} - \frac{\mu}{2} - \frac{\sigma}{\xi} \frac{1}{2^{-\xi+1}} \Gamma(-\xi+1) \right) - \frac{\sigma}{3\xi} + \frac{\mu}{3} + \frac{\sigma}{\xi} \frac{1}{3^{-\xi}} \Gamma(-\xi+1) \quad (61)$$

Portanto,

$$\alpha_2 = -\frac{\sigma}{3\xi} + \frac{\mu}{3} + \frac{\sigma}{\xi} \Gamma(-\xi+1) [1 - 2^{-\xi} + 3^{-\xi+1}] \quad (62)$$

Por definição de momentos-L tem-se que os λ_r , são combinações lineares dos α_s , assim tem que $\lambda_1 = \alpha_0$ logo,

$$\lambda_1 = \frac{\sigma}{\xi} [-1 + \Gamma(-\xi+1)] + \mu. \quad (63)$$

Como $\lambda_2 = \alpha_0 - 2\alpha_1$ então segue que:

$$\begin{aligned}
\lambda_2 &= \mu - \frac{\sigma}{\xi} \left[+1 - \Gamma(-\xi + 1) \right] - 2 \left[-\frac{\sigma}{2\xi} + \frac{\mu}{2} + \frac{\sigma}{\xi} \Gamma(-\xi + 1) \left(1 - \frac{1}{2^{\xi+1}} \right) \right] \\
&= \mu - \frac{\sigma}{\xi} + \frac{\sigma}{\xi} \Gamma(-\xi + 1) + \frac{\sigma}{\xi} - \mu - \frac{\sigma}{\xi} \Gamma(-\xi + 1) \left[\frac{2^{\xi+1} - 1}{2^\xi} \right] \\
&= \frac{\sigma}{\xi} \Gamma(-\xi + 1) \left[1 - \frac{2^{\xi+1} - 1}{2^\xi} \right] \\
&= \frac{\sigma}{\xi} \Gamma(-\xi + 1) \left[\frac{2^\xi - 2^{\xi+1} + 1}{2^\xi} \right] \\
&= \frac{\sigma}{\xi} \Gamma(-\xi + 1) \left[\frac{2^\xi + 1}{2^\xi} \right] \\
&= \frac{\sigma}{\xi} \Gamma(-\xi + 1) [1 + 2^\xi].
\end{aligned} \tag{64}$$

Para $\lambda_3 = \alpha_0 - 6\alpha_1 + 6\alpha_2$ temos que:

$$\begin{aligned}
\lambda_3 &= -\frac{\sigma}{\xi} [1 - \Gamma(-\xi + 1)] + \mu - 6 \left(-\frac{\sigma}{2\xi} + \frac{\mu}{2} + \frac{\sigma}{\xi} \Gamma(-\xi + 1) \left[1 - \frac{1}{2^{\xi+1}} \right] \right) + \\
&\quad + 6 \left(-\frac{\sigma}{3\xi} + \frac{\mu}{3} + \frac{\sigma}{\xi} \Gamma(-\xi + 1) [1 - 2^\xi + 3^{\xi+1}] \right) \\
&= -\frac{\sigma}{\xi} + \frac{\sigma}{\xi} \Gamma(-\xi + 1) + \mu + \frac{3\sigma}{\xi} - 3\mu - 6 \frac{\sigma}{\xi} \Gamma(-\xi + 1) \left[1 - \frac{1}{2^{\xi+1}} \right] - \frac{2\sigma}{\xi} + 2\mu \\
&\quad + 6 \frac{\sigma}{\xi} \Gamma(-\xi + 1) [1 - 2^\xi + 3^{\xi-1}] \\
&= \frac{\sigma}{\xi} \Gamma(-\xi + 1) \left[1 - 6 \left(1 - \frac{1}{2^{\xi+1}} \right) + 6(1 - 2^\xi + 3^{\xi-1}) \right] \\
&= \frac{\sigma}{\xi} \Gamma(-\xi + 1) \left[1 - 6 + \frac{6}{2^{\xi+1}} + 6 - 6 \times 2^\xi + 2 \times 3^\xi \right] \\
&= \frac{\sigma}{\xi} \Gamma(-\xi + 1) \left[\frac{2^{\xi+1} + 6 - 12 + 2 \times 2^{\xi+1} \times 3^\xi}{2^{\xi+1}} \right] \\
&= \frac{\sigma}{\xi} \Gamma(-\xi + 1) [1 - 6 \times 2^{\xi-1} + 2 \times 3^\xi] \\
&= \frac{\sigma}{\xi} \Gamma(-\xi + 1) [1 - 3 \times 2^\xi + 2 \times 3^\xi].
\end{aligned} \tag{65}$$

Por fim temos que τ_3 é obtido através do quociente entre λ_3 e λ_2 , ou seja,

$$\begin{aligned}
 \tau_3 &= \frac{\frac{\sigma}{\xi} \Gamma(-\xi+1) [1-3 \times 2^\xi + 2 \times 3^\xi]}{\frac{\sigma}{\xi} \Gamma(-\xi+1) [1+2^\xi]} \\
 &= \frac{1-3 \times 2^\xi + 2 \times 3^\xi + 2 - 2}{[1+2^\xi]} \\
 &= \frac{3-3 \times 2^\xi}{-1+2^\xi} + \frac{2 \times 3^\xi - 2}{-1+2^\xi} \\
 &= -3 + \frac{2(1-3^\xi)}{-1+2^\xi}.
 \end{aligned} \tag{66}$$

Assim o sistema de equações que fornecem os estimadores de momentos-L da GEV, pode ser expresso por:

$$\begin{cases}
 \frac{\hat{\sigma}}{\hat{\xi}} [-1 + (-\hat{\xi} + 1)] + \hat{\mu} = a_0 \\
 \frac{\hat{\sigma}}{\hat{\xi}} (-\hat{\xi} + 1) [1 - 2^{\hat{\xi}}] = a_0 - 2a_1 \\
 -3 + \frac{2[1 - 3^{\hat{\xi}}]}{1 - 2^{\hat{\xi}}} = \frac{a_0 - 6a_1 + 6a_2}{a_0 - 2a_1}
 \end{cases} \tag{67}$$

As estimativas de momentos-L dos parâmetros da GEV podem ser obtidas encontrando-se primeiramente a estimativa de $\hat{\xi}$ na terceira equação do sistema (67), por meio de um método numérico iterativo e, posteriormente, resolvendo-se analiticamente as outras duas equações do sistema.

2.6 PROPRIEDADES DOS ESTIMADORES

Devido à existência dos vários métodos de estimação para obtenção dos estimadores, é relevante indagar como se comparar dois ou mais estimadores. A resposta está relacionada às propriedades dos estimadores (NAGHETTINI; PINTO, 2007), pois a escolha de estimador de um parâmetro θ qualquer em detrimento de

outro depende de uma criteriosa avaliação dessas propriedades (FERREIRA, 2009). Duas das propriedades mais importantes para um estimador são a tendenciosidade e a eficiência (MORETTIN, 2010).

a) Tendenciosidade: Um estimador $\hat{\theta}$ é dito sem viés (ou não enviesado) do parâmetro populacional de θ se $E(\hat{\theta}) = \theta$.

Assim um estimador é não viesado se, em média, o valor estimado para o parâmetro for igual ao valor real do parâmetro. Caso $E(\hat{\theta}) \neq \theta$ então $\hat{\theta}$ é dito viesado, e o viés é dado pela diferença $v(\hat{\theta}) = E(\hat{\theta}) - \theta$ em que $E(\hat{\theta})$ é o valor esperado do estimador de θ e θ o real valor do parâmetro. Caso $\lim_{n \rightarrow +\infty} E(\hat{\theta}) = \theta$, diz que o estimador é assintoticamente não tendencioso.

b) Eficiência ou precisão: Dados dois estimadores θ_1 e θ_2 , define-se eficiência de um parâmetro em relação a outro, para um mesmo tamanho de amostra como:

$$Ef = \frac{\text{Var}(\theta_1)}{\text{Var}(\theta_2)}. \quad (68)$$

Um estimador é mais eficiente que o outro se sua variância for menor que a do outro. Se θ_1 é menos eficiente que θ_2 , então $Ef > 1$. Caso θ_1 seja mais eficiente que θ_2 , então $Ef < 1$.

O procedimento mais comum utilizado para se avaliar a eficiência de um estimador é o cálculo do seu erro quadrático médio, definido por Bolfarine e Sandoval (2001) por:

Definição 2 : O erro quadrático médio (EQM) de um estimador $\hat{\theta}$ do parâmetro θ é dado por $\text{EQM}[\hat{\theta}] = E\left[\left(\hat{\theta} - \theta\right)^2\right]$.

Aplicando as propriedades da esperança na definição 2 e realizando operações algébricas, obtém-se:

$$\text{EQM}[\hat{\theta}] = E\left[\left(\hat{\theta} - \theta\right)^2\right] = \text{VAR}[\hat{\theta}] + \underbrace{\left[E(\hat{\theta}) - \theta\right]^2}_{\text{Viés}} = \text{VAR}[\hat{\theta}] + v^2[\hat{\theta}]. \quad (69)$$

O erro quadrático médio incorpora os componentes de variância, que mede a variabilidade (precisão) do estimador, e o viés, que mede a exatidão do estimador (GAJO, 2016). Caso o estimador for não viesado, então o EQM será igual à variância do estimador (BOLFARINE; SANDOVAL, 2001). Assim, dados dois estimadores quaisquer não viesados de um mesmo parâmetro, o que possuir menor erro quadrático médio é o mais eficiente.

Šimková e Pícek (2017) utilizaram erro quadrático médio da amostra na comparação do momento-L, LQ, TL e máxima verossimilhança nas estimativas quantílicas altas da distribuição Generalizada de Pareto (GP) e GEV e concluíram que o momento-L e o método da máxima verossimilhança fornecem as melhores estimativas de quantil alto para tais distribuições.

2.7 Momentos-L na estimação dos parâmetros da distribuição generalizada de valores extremos

L-moments in the parameters estimation of the generalized extreme value distribution

Nalva Aparecida Martins Batista¹, Mesek Felipe de Souza², Fabricio Goecking Avelar³ e Luiz Alberto Beijo⁴

¹Mestranda do Programa de Pós-Graduação Estatística Aplicada e Biometria- PPGEAB, Universidade Federal de Alfenas, Brasil.
nalvaaf@hotmail.com

²Discente de Licenciatura em Matemática da Universidade Federal de Alfenas, Brasil.
mekvemon@gmail.com

^{3,4}Doutores, Docentes do Programa de Pós-Graduação Estatística Aplicada e Biometria- PPGEAB, Universidade Federal de Alfenas, Brasil.
fabricio@unifal-mg.edu.br; luiz.beijo@unifal-mg.edu.br

RESUMO

A Teoria de Valores Extremos (TVE) é uma importante ferramenta utilizada na análise de eventos extremos. A análise de tais eventos permite que haja um melhor planejamento nas ações humanas no sentido de minimizar os efeitos causados por eles. Uma distribuição proveniente da TVE utilizada na modelagem de eventos extremos é a distribuição generalizada de valores extremos (GEV). Vários métodos de estimação podem ser utilizados na estimação dos parâmetros da GEV. Dentre os mais utilizados são os de máxima verossimilhança. Contudo, esses estimadores não apresentam boas propriedades, ou não existem, quando o parâmetro forma da distribuição GEV é menor que $-0,5$. Nessas situações é necessário o uso de outro estimador. Nesse sentido o objetivo deste trabalho foi avaliar a adequabilidade do uso dos estimadores dos momentos-L no ajuste da distribuição GEV, comparando a precisão e a acurácia desses estimadores com a precisão e acurácia dos estimadores de máxima verossimilhança e utilizar os dois estimadores para ajustar a distribuição GEV a um conjunto de dados reais que contenha a estimativa do parâmetro forma ente -1 e $-0,5$. Verificou-se que para valores do parâmetro forma da GEV negativos o método dos momentos-L é indicado para tamanhos amostrais menores ou iguais a 25, com o parâmetro forma com valores próximos de zero o momentos-L é indicado para tamanhos amostrais menores ou iguais a 35 e para valores positivos do parâmetro forma o momentos-L deve ser utilizado em todos os tamanhos amostrais.

Palavras-chave: Estimadores de Máxima Verossimilhança. Erro quadrático médio. Viés médio relativo. Tamanho amostral.

ABSTRACT

The Extreme Value Theory (EVT) is an important tool used in the analysis of extreme events. The analysis of such events allows for better planning in human actions in order to minimize the effects caused by them. A distribution from EVT that is widely used in extreme events modeling is the generalized extreme value distribution (GEV). Several estimation methods can be used to estimate the GEV distribution parameters. Among the most widely used estimators are the maximum likelihood estimators. However, these estimators are not suitable, or do not exist, when the shape parameter of the GEV

distribution is less than -0.5 . In these situations it is necessary to use another estimator. In this sense, the objective of this work was to evaluate the adequacy of the use of the L-moment estimators in the adjustment of the GEV distribution, comparing the precision and accuracy of these estimators with the precision and accuracy of the maximum likelihood estimators and to use these two estimators to adjust the GEV distribution to a set of real data that contains the estimate of the parameter form between -1 and -0.5 . It was found that for negative values of the shape parameter of the GEV the L-moment method is indicated for sample sizes less than or equal to 25, with the shape parameter with values close to zero the L-moments is indicated for sample sizes less than or equal to 35 and for positive values of the form parameter the L-moments should be used in all sample sizes.

Keywords: Maximum Likelihood Estimators. Mean squared error. Relative mean bias. Sample size.

1 INTRODUÇÃO

A frequência de fenômenos naturais climáticos extremos tem aumentado nos últimos anos (MARENGO *et al.*, 2014). Esses eventos podem afetar severamente diversas atividades humanas, causando prejuízos financeiros, físicos e até mesmo a perda de vidas humanas (ŠIMKOVÁ; PICEK, 2017). São exemplos de eventos extremos: as precipitações máximas; as temperaturas mínimas e máximas; a poluição atmosférica extrema; a velocidade máxima de ventos; dentre outros. A teoria de valores extremos (TVE) visa auxiliar na análise desses fenômenos descrevendo e quantificando os seus comportamentos, buscando modelos que se ajustem a uma possível distribuição para os máximos (ou mínimos) de valores extremos a partir de uma amostra (BEIJO; AVELAR, 2011; ŠIMKOVÁ; PICEK, 2017).

Os princípios teóricos referentes a TVE foram inicialmente apresentados por Fisher e Tippett (1928), ao verificarem que as distribuições assintóticas dos extremos para máximos em blocos convergiam para uma das três distribuições não degeneradas, conhecidas respectivamente como Gumbel (Tipo I), Fréchet (tipo II) e de Weibull (tipo III). Em 1943, Gnedenko apresentou rigorosa fundamentação teórica das condições necessárias e suficientes para que a distribuição dos máximos, ou dos mínimos, convirja para uma distribuição não degenerada. Jenkinson (1955) propôs a distribuição generalizada de valores extremos (GEV, do inglês “Generalized Extreme Value”). Ele mostrou que as três distribuições assintóticas dos valores extremos podem ser unificadas como casos particulares da GEV. Essa distribuição possui três parâmetros: posição, escala e forma e é amplamente utilizada e apresenta resultados satisfatórios na modelagem do comportamento de variáveis extremas em diversas áreas do conhecimento. (ARRAES; ROCHA, 2006; BLAIN, 2011; OLIVEIRA *et al.*, 2014).

Há vários estimadores que podem ser utilizados na estimação dos parâmetros da GEV. Os estimadores mais utilizados são os de máxima verossimilhança por possuírem boas propriedades assintóticas. Mas segundo Smith (1985) esses estimadores não apresentam boas propriedades, ou não existem, quando o parâmetro forma da distribuição GEV é menor que $-0,5$. Nesses casos pode ser adequado ou até mesmo necessário o uso de outro estimador como, por exemplo, os estimadores de momentos-L.

Neste sentido o objetivo desse trabalho foi avaliar e comparar a precisão e a acurácia dos estimadores de máxima verossimilhança e dos momentos-L dos parâmetros da distribuição GEV. Também objetivou utilizar os estimadores para ajustar a distribuição GEV a um conjunto de dados reais que contenha a estimativa do parâmetro forma ente -1 e $-0,5$.

2 Distribuição Generalizada de Valores Extremos

A distribuição GVE tem sido utilizada em diversos estudos solucionando vários problemas nas mais diferentes áreas como, por exemplo, na hidrologia, na meteorologia, na engenharia, na economia e finanças, entre outras. Bautista (2002) aplicou a distribuição GEV na análise das velocidades máximas dos ventos. Arraes e Rocha (2006) estimaram perdas máximas esperadas para séries financeiras, empregando-se métodos tradicionais e metodologia dos Valores Extremos. Quadros, Queiroz e Boas (2011), analisaram séries históricas de precipitações máximas no município de Cascavel - PR e ajustaram as distribuições GEV e Gumbel. De acordo com Jenkinson (1955) a GEV possui função densidade de probabilidade expressa por:

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

definida em $-\infty < x < \mu - \sigma / \xi$ para $\xi < 0$, $\mu - \sigma / \xi < x < +\infty$ para $\xi > 0$. Os parâmetros μ, σ, ξ representam os parâmetros de posição, escala e forma respectivamente, com $\mu \in \mathbb{R}$, $\sigma > 0$ e $\xi \in \mathbb{R}^*$. A GEV generaliza as três famílias de distribuições de valores extremos, correspondendo às distribuições Fréchet, Weibull e Gumbel, respectivamente, para $\xi > 0$, $\xi < 0$ e $\xi \rightarrow 0$. (BEIJO; AVELAR, 2011). A distribuição de probabilidade acumulada da GEV é expressa por:

$$F(x) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\left(\frac{1}{\xi} \right)} \right\}, \quad (2)$$

A função quantil da GEV é obtida a partir da inversa da expressão (2) e pode ser expressa por:

$$Q(y) = \mu + \frac{\sigma}{\xi} \left\{ -1 + [-\ln(y)]^{-\xi} \right\}, \quad 0 < y < 1. \quad (3)$$

Para calcular as probabilidades de ocorrência de uma determinada variável e de um fenômeno como, por exemplo, temperaturas máximas, velocidades máximas de ventos, precipitações máxima, dentre outros, acima de um valor x pré-estabelecido utiliza-se da equação:

$$P(X > x) = 1 - F(x) = 1 - \exp \left\{ - \left[1 + \hat{\xi} \left(\frac{x - \hat{\mu}}{\hat{\sigma}} \right) \right]^{-\left(\frac{1}{\hat{\xi}} \right)} \right\}, \quad (4)$$

em que x é um valor da variável do período.

Bautista (2002) define período de retorno da seguinte maneira: seja E um evento e T o tempo aleatório entre as ocorrências consecutivas de eventos E . O valor médio da variável T é chamado de período de retorno do evento E . Logo, o período de retorno para esse evento, é dado por:

$$T = \frac{1}{P(E)} = \frac{1}{1-F(x)}, \quad (5)$$

sendo T expresso geralmente em anos.

Assim de acordo como autor nível de retorno x_p da distribuição GEV, ou seja, o nível x_p , associado ao período de retorno T , é obtido a partir da solução da expressão:

$$\int_{-\infty}^{x_p} f(x | \xi, \sigma, \mu) dx = 1-p \quad (6)$$

com $p = 1/T$, isto é,

$$F(x_p) = (1-p) \quad (7)$$

Aplicando a função inversa F em ambos os membros da equação (7) e relacionando com a probabilidade de ocorrências (da variável em estudo) (4), tem-se que:

$$\hat{x}_p = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}} \left\{ -1 + [-\ln(1-p)]^{-\hat{\xi}} \right\}, \quad (8)$$

Assim a estimativa \hat{x}_p do nível de retorno x_p para períodos de retorno $T = 1/p$ é obtido por substituição das estimativas de μ , σ , e ξ .

Vários métodos foram desenvolvidos propondo sugestões para a estimação dos parâmetros da GEV. Mas, segundo Coles (2001) o método da máxima verossimilhança (MV) é mais aplicado devido a sua capacidade de se adaptar a distribuições mais complexas e por possuir boas propriedades em situações assintóticas.

2.1 Método da Máxima Verossimilhança

Segundo Bolfarine e Sandoval (2001), dada X_1, X_2, \dots, X_n uma sequência de variáveis aleatórias independentes e identicamente distribuídas (iid) de uma variável aleatória X e x_1, x_2, \dots, x_n uma amostra de tamanho n da variável aleatória X com função de densidade de probabilidade $f(x; \theta)$, com $\theta \in \Theta$, em que Θ é o espaço paramétrico, a função de verossimilhança de X é definida por:

$$L(\theta; x_1, x_2, \dots, x_n) = f(x_1; \theta) f(x_2; \theta) \dots f(x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) \quad (9)$$

Os estimadores de máxima verossimilhança consistem em encontrar o vetor $\hat{\theta}$ que maximiza a função (9). A partir das expressões (1) e (9), verifica-se que a função de verossimilhança da distribuição GEV, pode ser expressa por:

$$L(\mu, \sigma, \xi | \mathbf{x}) = \frac{1}{\sigma^n} \prod_{i=1}^n \left\{ \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\xi}{\sigma} \right)} \right\} \prod_{i=1}^n \exp \left\{ - \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1}{\xi} \right)} \right\} \quad (10)$$

Derivando-se parcialmente a expressão (10) em relação aos parâmetros μ , σ e ξ e igualando-se as derivadas resultantes a zero obtém-se o sistema de equações cuja solução são os estimadores de máxima verossimilhança dos parâmetros da distribuição GEV.

$$\left\{ \begin{array}{l} \frac{1}{\hat{\sigma}} \sum_{i=1}^n \left\{ \frac{1 + \hat{\xi} - [\hat{w}_i]^{-\left(\frac{1}{\hat{\xi}}\right)}}{\hat{w}_i} \right\} = 0 \\ -\frac{n}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^2} \sum_{i=1}^n \left\{ \frac{(x_i - \hat{\mu}) \left[(1 + \hat{\xi}) - [\hat{w}_i]^{-\left(\frac{1}{\hat{\xi}}\right)} \right]}{\hat{w}_i} \right\} = 0 \\ \sum_{i=1}^n \frac{1}{\hat{\xi}^2} \ln \hat{w}_i - \frac{1}{\hat{\sigma}} \left(\frac{x_i - \hat{\mu}}{\hat{w}_i} \right) - \frac{1}{\hat{\xi} \hat{\sigma}} \left(\frac{x_i - \hat{\mu}}{\hat{w}_i} \right) \left[+1 - \hat{w}_i^{-\left(\frac{1}{\hat{\xi}}\right)} \right] = 0 \end{array} \right. \quad (11)$$

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

Como o sistema (11) não possui soluções analíticas, métodos numéricos iterativos devem ser utilizados na obtenção da solução.

Apesar do método de MV ser comumente utilizado, as condições de regularidade para estimação dos parâmetros da distribuição GEV pelo método MV não são fundamentalmente satisfeitas em algumas situações. Existe regularidade no caso em que $\xi > -0,5$. Quando $-1 < \xi < -0,5$ os estimadores existem, mas não satisfazem às condições de regularidades e para $\xi < -1$, os estimadores de MV não existem (SMITH, 1985). Desse modo em situação nos quais os valores do parâmetro forma são menores $-0,5$, pode vir a exigir o uso de outros estimadores para os parâmetros da distribuição GVE. Além disso, em alguns casos a eficiência do método da máxima verossimilhança é apenas assintótica, o que dificulta sua aplicação em amostras pequenas, produzindo estimadores de qualidade análogo ou inferior a outros métodos. Nestes casos, os estimadores ML mostram-se, com alguma frequência, mais acurados do que os de MV por produzir estimadores de parâmetros comparáveis, em qualidade, àqueles gerados por esse método, mas com a vantagem de exigirem um menor esforço computacional (NAGHETTINI; PINTO, 2007).

2.2 Método dos momentos-L

Greenwood *et al.* (1979), introduziram os momentos ponderados por probabilidade (MPP), que podem ser definidos por:

$$M_{p,r,s} = E \left\{ X^p [F_X(x)]^r [1-F_X(x)]^s \right\} = \int_{-\infty}^{+\infty} x^p F_X(x)^r [1-F_X(x)]^s f(x) dx = \int_0^1 Q(y)^p y^r (1-y)^s dy \quad (12)$$

em que X é uma variável aleatória contínua, p , r e s são números reais, F_X é a função de distribuição da variável aleatória X e $Q(y)$ denota a função de quantis. Caso r e s sejam nulos e p um número inteiro não negativo os momentos ponderados por probabilidade são iguais aos momentos convencionais.

Segundo Naghettini e Pinto (2007), os momentos ponderados por probabilidade $M_{1,0,s} = \alpha_s = \int_0^1 Q(y)(1-y)^s dy$ e $M_{1,r,0} = \beta_r = \int_0^1 Q(y)y^r dy$, são os mais utilizados na caracterização de distribuições de probabilidade.

Hosking (1986) demonstrou que os métodos dos momentos ponderados por probabilidade MPP, possuem generalidade suficiente para a estimação de parâmetros de distribuições de probabilidade, estão menos sujeitos a flutuações amostrais e são mais

robustos do que os correspondentes momentos convencionais (NAGHETTINI; PINTO, 2007).

Greenwood *et al.* (1979), a partir de uma amostra aleatória $(x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)})$ ordenada crescente de tamanho n , retirada de uma determinada população, definiram os momentos ponderados por probabilidade amostrais por:

$$\hat{a}_s = \frac{1}{n} \sum_{i=1}^n \frac{\binom{n-i}{s}}{\binom{n-1}{s}} x_i, \quad (13)$$

$$\hat{b}_r = \frac{1}{n} \sum_{i=1}^n \frac{\binom{n-i}{r}}{\binom{n-1}{r}} x_i, \quad (14)$$

sendo $i=1,2,\dots,n$; s e $r=0,\dots,n-1$.

Apesar dos métodos ponderados por probabilidade poderem ser utilizados na estimação dos parâmetros de um modelo, seus descritores α_s e β_r , não possibilitam uma interpretação natural de escala e forma das distribuições de probabilidades. Visando resolver tal problema Hosking (1990) apresentou o conceito de momentos-L, que nada mais são do que funções lineares dos momentos ponderados por probabilidade (MPP), ou seja, combinações lineares dos α_s e β_r , os quais possuem grandezas diretamente interpretáveis como descritores de escala e forma das distribuições de probabilidade. Formalmente, os momentos-L populacionais, denotados por λ_r , podem ser definidos por:

$$\lambda_r = (-1)^{r-1} \sum_{k=0}^{r-1} \rho_{r-1,k} \alpha_k = \sum_{k=0}^{r-1} \rho_{r-1,k} \beta_k, \quad (15)$$

em que

$$\rho_{r-1,k} = (-1)^{r-k-1} \binom{r-1}{k} \binom{r+k-1}{k}. \quad (16)$$

Logo, os momentos-L populacionais, de ordem inferior a quatro, são dados por:

$$\lambda_1 = \alpha_0; \quad (17)$$

$$\lambda_2 = \alpha_0 - 2\alpha_1; \quad (18)$$

$$\lambda_3 = \alpha_0 - 6\alpha_1 + 6\alpha_2. \quad (19)$$

Os momentos-L amostrais são denotados por l_r e são calculados pela substituição de α_s ou β_r nas equações (17) a (19), pelas suas estimativas a_s ou b_r .

$$l_1 = a_0; \quad (20)$$

$$l_2 = a_0 - 2a_1; \quad (21)$$

$$l_3 = a_0 - 6a_1 + 6a_2. \quad (22)$$

De acordo com Gubareva e Gartsman (2010) o primeiro momento-L se equipara uma medida de tendência central, as demais medidas da distribuição, tais como variância, assimetria e curtose, são avaliadas por meio de razões de momentos-L de ordem superiores a 1. τ é uma medida equivalente ao coeficiente de variação dado pelo

quociente de λ_2 e λ_1 e as demais razões do momento-L são definidos por: $\tau_r = \lambda_r / \lambda_2$, $r = 3, 4, \dots$, determinam a forma das distribuições de probabilidades. Segundo Wolff (2013) para obter as estimativas dos parâmetros da distribuição probabilidade em estudo pelo método dos momentos-L, deve-se igualar os momentos-L populacionais aos momentos-L amostrais, ou seja,

$$\begin{cases} \lambda_1 = l_1 \\ \lambda_2 = l_2 \\ \tau_3 = t_3 \end{cases} \Rightarrow \begin{cases} \lambda_1 = a_0 \\ \lambda_2 = a_0 - 2a_1 \\ \tau_3 = \frac{a_0 - 6a_1 + 6a_2}{a_0 - 2a_1} \end{cases} . \quad (23)$$

Desse modo tem-se que λ_1 é equivalente à média da distribuição e, portanto, uma medida populacional de posição, λ_2 uma medida de dispersão (análogo ao desvio padrão ordinário), a razão $CV-L = \tau = \lambda_2 / \lambda_1$ é análogo ao coeficiente de variação convencional (o qual pode ser interpretado como uma medida populacional de dispersão ou de escala) e a razão $\tau_3 = \lambda_3 / \lambda_2$ representa uma medida de assimetria (NAGHETTINI; PINTO, 2007).

2.3 Estimativa dos parâmetros da GEV via MML

Para encontrar os estimadores de momentos-L dos parâmetros da GEV é necessário encontrar os momentos ponderados por probabilidade populacionais α_s que, são obtidos pelo cálculo das integrais,

$$\alpha_s = \int_0^1 Q(y)(1-y)^s dy. \quad (24)$$

em que $Q(y) = F_X^{-1}(y)$ é a função quantil de uma variável X .

Desse modo o primeiro momento ponderado por probabilidade populacional α_0 é dado pelo cálculo da integral:

$$\alpha_0 = \int_0^1 \left\{ \frac{\sigma}{\xi} \left\{ -1 + [-\ln(y)]^{-\xi} \right\} dy + \mu \right\} dy = \frac{\sigma}{\xi} [-1 + \Gamma(-\xi + 1)] + \mu.$$

em que $\Gamma(\cdot)$ representa a função gama denotada por $\Gamma(y) = \int_0^\infty t^{y-1} \exp(-t) dt$.

O segundo momento ponderado por probabilidade populacional, ou seja, α_1 é dado por:

$$\alpha_1 = \int_0^1 \left\{ \frac{\sigma}{\xi} [-1 + (-\ln(y))^{-\xi}] + \mu \right\} (1-y) dy = -\frac{\sigma}{2\xi} + \frac{\mu}{2} + \frac{\sigma}{\xi} \Gamma(-\xi + 1) \left[\frac{2^{-\xi+1} - 1}{2^{-\xi+1}} \right].$$

E por fim α_2 é dado por:

$$\alpha_2 = \int_0^1 \left\{ \frac{\sigma}{\xi} [-1 + (-\ln(y))^{-\xi}] + \mu \right\} (1-y)^2 dy = -\frac{\sigma}{3\xi} + \frac{\mu}{3} + \frac{\sigma}{\xi} \Gamma(-\xi + 1) [1 - 2^{-\xi} + 3^{-\xi+1}].$$

Por definição de momentos-L tem-se que os λ_r , são combinações lineares dos α_s , assim tem-se os respectivos momentos-L populacional da GEV:

$$\begin{aligned}\lambda_1 &= \frac{\sigma}{\xi} [-1 + \Gamma(-\xi + 1)] + \mu; \\ \lambda_2 &= \frac{\sigma}{\xi} \Gamma(-\xi + 1) [1 + 2^\xi]; \\ \lambda_3 &= \frac{\sigma}{\xi} \Gamma(-\xi + 1) [1 - 3 \times 2^\xi + 2 \times 3^\xi]; \\ \tau_3 &= -3 + \frac{2(1 - 3^\xi)}{1 - 2^\xi}.\end{aligned}$$

Assim o sistema de equações que fornecem os estimadores de momentos-L da GEV, pode ser expresso por:

$$\begin{cases} \frac{\hat{\sigma}}{\hat{\xi}} [-1 + (-\hat{\xi} + 1)] + \hat{\mu} = a_0 \\ \frac{\hat{\sigma}}{\hat{\xi}} (-\hat{\xi} + 1) [1 - 2^{\hat{\xi}}] = a_0 - 2a_1 \\ -3 + \frac{2[1 - 3^{\hat{\xi}}]}{1 - 2^{\hat{\xi}}} = \frac{a_0 - 6a_1 + 6a_2}{a_0 - 2a_1} \end{cases} \quad (25)$$

As estimativas de momentos-L dos parâmetros da GEV podem ser obtidas encontrando-se primeiramente a estimativa de $\hat{\xi}$ na terceira equação do sistema (25), por meio de um método numérico iterativo e, posteriormente, resolvendo-se analiticamente as outras duas equações do sistema.

3 MATERIAL E MÉTODOS

3.1 Simulações e ajustes da GEV

A comparação da precisão e acurácia dos estimadores de máxima verossimilhança e de momentos-L na estimação dos parâmetros da distribuição GEV foi realizada por meio de simulação de dados. A definição dos cenários simulados foi baseada em artigos referentes à aplicação da GEV utilizada na modelagem de eventos extremos.

A tabela 1 apresenta os valores utilizados para os parâmetros da GEV e os tamanhos amostrais utilizados na composição dos cenários.

Tabela 1 – Parâmetros utilizados na composição dos cenários e os tamanhos amostrais utilizados na simulação

Parâmetro			Tamanho amostral
Posição (μ)	Escala (σ)	Forma (ξ)	
2	0,9; 5	-0,75; -0,52; -0,46	13; 15; 17; 20; 25; 35; 45; 60; 100; 150; 200
19	0,9; 5; 9; 13		
37	5; 9; 13	-0,16; 0,16	
55	9; 13		
70	13	0,4; 0,5	

Fonte: Elaboração dos autores.

Os cenários analisados compreendem um total 924. Os cenários são compostos pelas combinações possíveis dos valores dos parâmetros posição, escala em cada linha com todos os valores do parâmetro forma para todos tamanhos amostrais apresentados

na Tabela 1. Foram realizados 3000 simulações para cada cenário das séries de dados da distribuição generalizada de valores Extremos.

A distribuição GEV foi ajustada a cada uma das 3000 amostras em cada cenário, por meio dos estimadores de máxima verossimilhança e de momentos-L. A qualidade dos ajustes foi verificada utilizando o teste de Kolmogorov-Smirnov (KS) ao nível de 5% de significância. O método numérico utilizado para obter as estimativas de máxima verossimilhança dos parâmetros da GEV foi *simulated annealing* (SANN) e para momentos-L o Método BFGS.

Foram calculados, em cada um dos cenários, o erro quadrático médio (EQM) e do viés médio relativo (VMR) obtidos por meio das respectivas expressões:

$$EQM = \frac{\sum_{i=1}^{3000} (\hat{\theta}_i - \theta)^2}{3000} \quad (26)$$

$$VMR = \frac{\sum_{i=1}^{3000} |\hat{\theta}_i - \theta|}{3000 \times \theta}, \quad (27)$$

em que $\hat{\theta}_i$ é o valor estimado para o parâmetro do modelo na i -ésima amostra simulada, e θ o valor real do parâmetro e 3000 o total de amostras.

Foram realizados os calculados da proporção do número de vezes em que o método dos momentos-L apresentou erro quadrático menor que o apresentado pelo método da máxima verossimilhança. Para testar se a proporção de vezes que o método dos momentos-L apresentou erro quadrático menor que o método da máxima verossimilhança é maior que 0,5; foi aplicado o teste binomial para proporção (π) considerando o nível de 5% de significância. Para tanto definiu-se a seguinte proporção:

π_{ML} : Proporção do número de vezes em que o erro quadrático dos estimadores de momentos-L foi menor que os estimadores de máxima verossimilhança. Foi realizado o seguinte teste de hipóteses:

$$\begin{cases} H_0 : \pi_{ML} \leq 0,5 \\ H_1 : \pi_{ML} > 0,5 \end{cases}$$

Foram calculados ainda, os vieses dos níveis de retorno para os tempos de retorno de 10, 25, 50, 100, 200, 500 e 1.000 anos.

Os cálculos e simulações foram todos realizados no software R (R Core Team, 2019), utilizando os pacotes *evd* e *extRemes*.

3.2 Aplicação aos dados de temperatura máxima de Franca-SP

3.2.1. Fonte dos dados

As séries históricas de temperatura máximas (°C) foram obtidas junto ao Instituto Nacional de Meteorologia - INMET - do Ministério da Agricultura, Pecuária e Abastecimento para a cidade de Franca-SP. A escolha da cidade se deu devido ao fato que dentre as séries de dados observados ao realizar o ajuste da distribuição pelo método da máxima verossimilhança foi a que apresentou a estimativa do parâmetro forma da GEV entre -1 e -0,5. A série de dados corresponde ao período de 1980 a 2018, para a realização das análises, os dados foram agrupados por blocos em período anual, retirando-se a temperatura máxima ocorrida em cada ano, as observações correspondem

um total de 39 anos. Segundo Jurca (2005) o clima de Franca corresponde ao Aw- clima tropical com inverno seco.

3.2.2 Ajustes da GEV e avaliação

A independência e a aleatoriedade das séries de dados foram verificadas utilizando-se respectivamente, o teste de Ljung-Box (LJUNG; BOX, 1978) e o Runs Test (SIEGEL; CASTELLAN JR, 2006) ao nível de 5% de significância. Para o ajuste da distribuição aos dados, a série em estudo foi dividida em duas partes sendo a primeira parte compreendida no período de 1980 a 2012, utilizada para obter as estimativas dos parâmetros pelos métodos MV e ML e calcular a temperatura máxima provável, para os tempos de retorno de 2,4, 6 anos. A segunda parte correspondente ao período de 2012 a 2018 na qual, extraiu-se a maior temperatura máxima observada para cada tempo de retorno. A qualidade do ajuste da distribuição GEV aos dados foi avaliada pelo teste de Kolmogorov-Smirnov (KS), o qual foi adotado o nível de significância de 5%.

E para comparar qual método é mais acurado em relação ao nível de retorno dos estimadores foi utilizado o erro percentual absoluto médio (EPAM) dado pela expressão:

$$EPAM = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - \hat{x}_i}{x_i} \right| \times 100, \quad (28)$$

na qual, x_i é a temperatura máxima observada e \hat{x}_i é a temperatura máxima predita para a i -ésimo tempo de retorno.

4 RESULTADOS E DISCUSSÃO

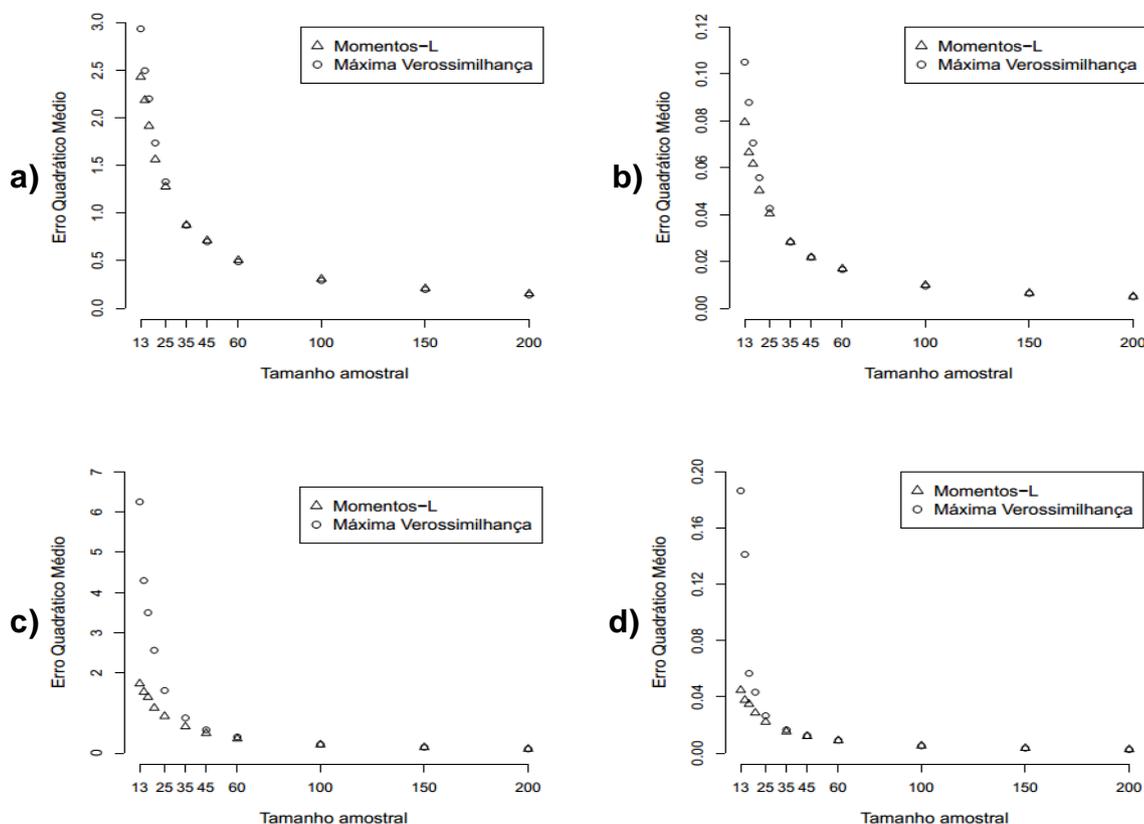
Os resultados das análises podem ser divididos, devido às suas similitudes, em três grupos: cenários nos quais o parâmetro forma é negativo, próximo de zero e positivo.

4.1 Cenários com parâmetro forma negativo

Os resultados dos cenários simulados nos quais os valores do parâmetro forma foram -0,75, -0,52 e -0,46 são apresentados nesta seção.

Na Figura-1 é apresentado o erro quadrático médio (EQM) dos estimadores de momentos-L e de máxima verossimilhança dos parâmetros posição (μ) e escala (σ) da distribuição GEV, nos cenários: posição 2, escala 5 e forma -0,75 e posição 19, escala 0,9 e forma -0,46 nos tamanhos amostrais estudados.

Figura 1 - Erro quadrático médio (EQM) dos estimadores de máxima verossimilhança e de momentos-L dos parâmetros posição (μ) e escala (σ) da distribuição GEV, considerando diferentes cenários e os tamanhos amostrais com 13, 15, 17, 20, 25, 35, 45, 60, 100, 150 e 200 observações

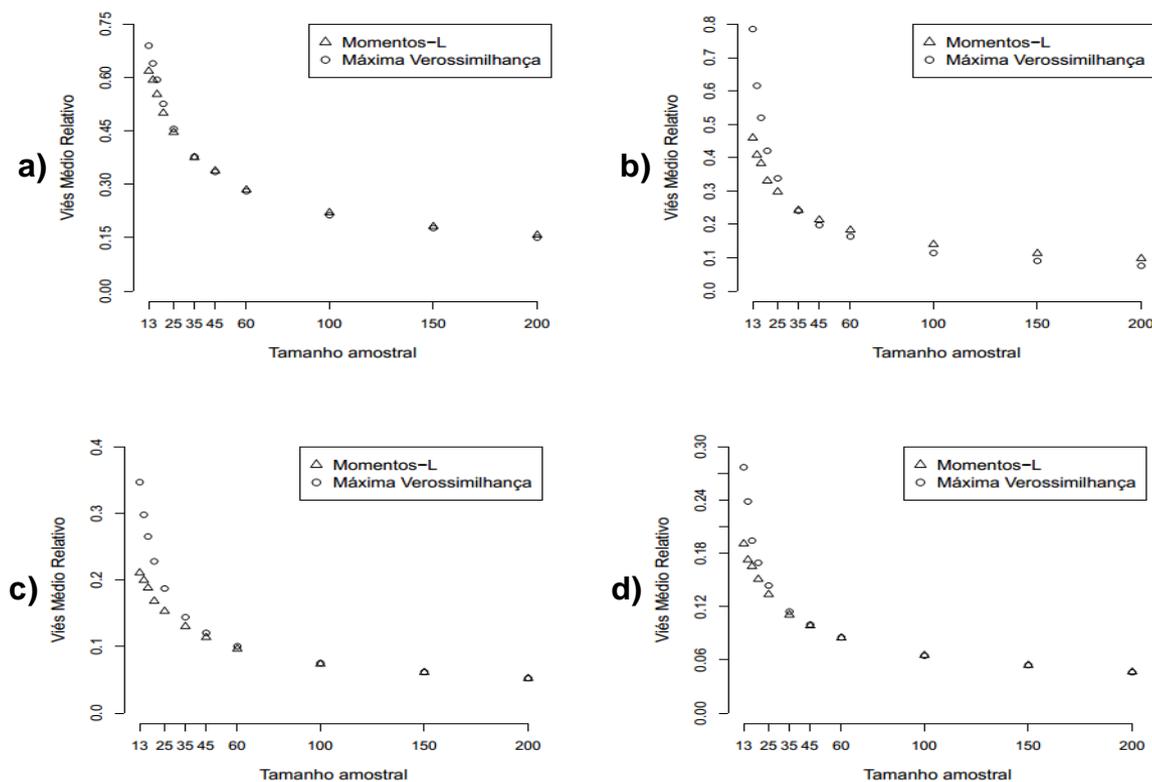


Legenda: (a) EQM parâmetro posição; cenário ($\mu = 2; \sigma = 5; \xi = -0,75$)
 (b) EQM parâmetro posição; cenário ($\mu = 19; \sigma = 0,9; \xi = -0,46$)
 (c) EQM parâmetro escala; cenário ($\mu = 2; \sigma = 5; \xi = -0,75$)
 (d) EQM parâmetro escala; cenário ($\mu = 19; \sigma = 0,9; \xi = -0,46$)

Observando-se a Figura 1, é possível perceber por meio dos gráficos 1(a) e 1(b) que os estimadores de momentos-L apresentaram menor EQM que os estimadores de máxima verossimilhança para o parâmetro posição até o tamanho amostral 35. Após esse tamanho amostral, os dois estimadores apresentaram valores similares para o EQM. A partir dos gráficos 1(c) e 1(d) é possível perceber que os estimadores de momentos-L apresentaram menores valores de EQM para o parâmetro escala até o tamanho amostral 60. A partir desse tamanho amostral, os dois estimadores apresentaram valores de EQM semelhantes. Observa-se também que os valores dos EQM diminuem à medida que o tamanho amostral aumenta em todos os cenários apresentados. Os resultados obtidos para o comportamento do EQM dos parâmetros posição e escala nos demais cenários, considerando-se o parâmetro forma com valores iguais a -0,75, -0,52 e -0,46 foram semelhantes e, por essa razão, foram omitidos.

Na Figura 2 é apresentado o viés médio relativo dos estimadores de máxima verossimilhança e de momentos-L obtido para os parâmetros posição e escala nos seguintes cenários: parâmetro posição 2, escala 5 e forma -0,75 e posição 19, escala 0,9 e forma -0,46.

Figura 2 – Viés médio relativo (VMR) dos estimadores de máxima verossimilhança e de momentos-L dos parâmetros posição (μ) e escala (σ) da distribuição GEV para os cenários, posição é igual a 2; escala 5; forma -0,75 e posição 19; escala 0,9; forma -0,46 e os tamanhos amostrais com 13, 15, 17, 20, 25, 35, 45, 60, 100, 150 e 200 observações

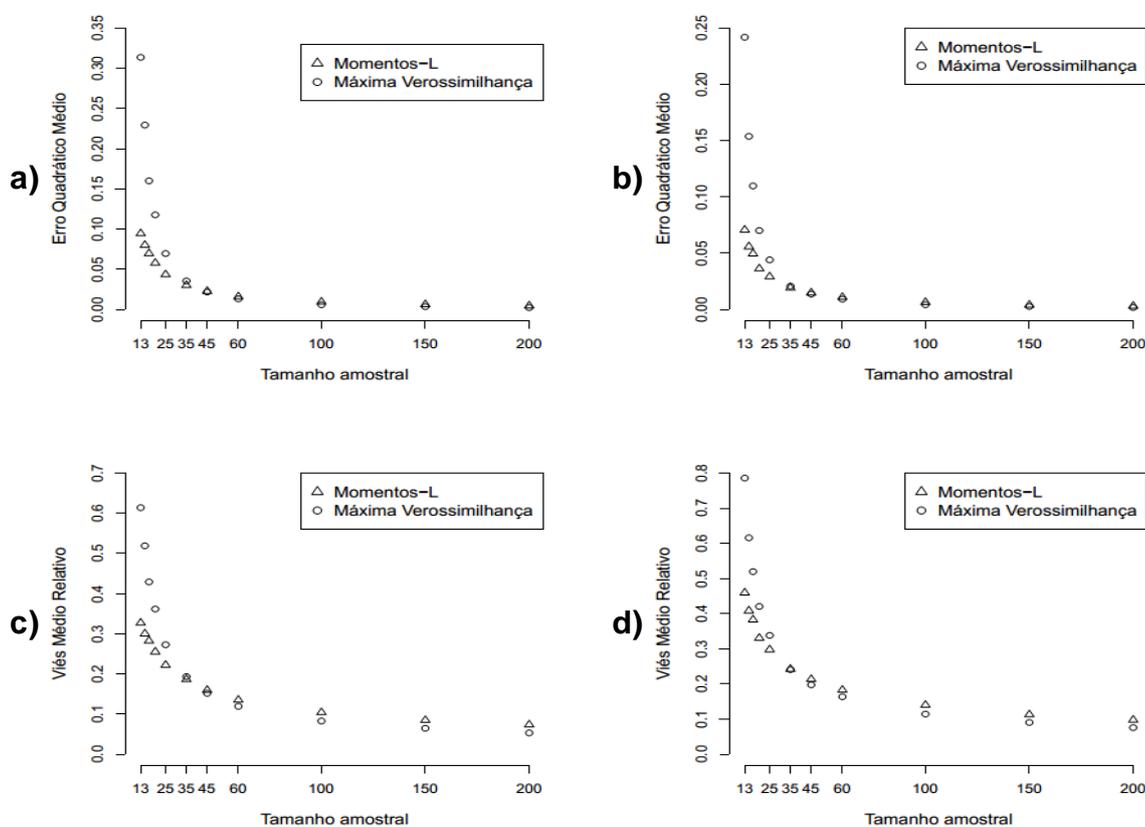


Legenda: (a) VMR parâmetro posição; cenário ($\mu = 2; \sigma = 5; \xi = -0,75$)
 (b) VMR parâmetro posição; cenário ($\mu = 19; \sigma = 0,9; \xi = -0,46$)
 (c) VMR parâmetro escala; cenário ($\mu = 2; \sigma = 5; \xi = -0,75$)
 (d) VMR parâmetro escala; cenário ($\mu = 19; \sigma = 0,9; \xi = -0,46$)

Na Figura 2, é possível perceber por meio dos gráficos 2(a) e 2(b) que os estimadores de momentos-L apresentam menor VMR que os de MV para o parâmetro posição até o tamanho amostral 35. Após esse tamanho amostral, os dois estimadores apresentam valores semelhantes para VMR. A partir dos gráficos 2(c) e 2(d) é possível observar que os estimadores de momentos-L apresentaram menores valores de VMR para o parâmetro escala até o tamanho amostral 60. A partir desse tamanho amostral, os dois estimadores apresentam valores de VMR análogos. Observa-se ainda que os valores dos VMR diminuem à medida que o tamanho amostral aumenta em todos cenários apresentados. Os resultados obtidos para o comportamento do VMR dos parâmetros posição e escala nos demais cenários, considerando-se os parâmetros forma com os valores iguais a -0,75, -0,52 e -0,46 foram semelhantes e, por isso, foram omitidos.

Na Figura 3 são apresentados o erro quadrático médio e o viés médio relativo dos estimadores de MV e de momentos-L obtido para o parâmetro forma nos seguintes cenários: parâmetro posição 2, escala 5 e forma -0,75 e posição 19, escala 0,9 e forma -0,46.

Figura 3 - Erro quadrático médio (EQM) e viés médio relativo (VMR) dos estimadores de máxima verossimilhança e de momentos-L do parâmetro forma (ξ), da distribuição GEV para diferentes cenários, e os tamanhos amostrais estudados

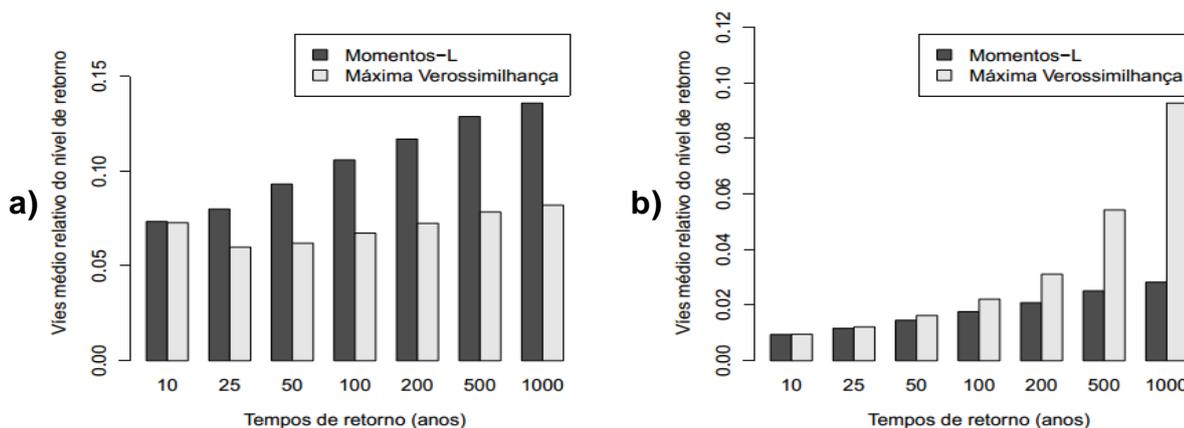


- Legenda: (a) EQM parâmetro forma; cenário ($\mu = 2; \sigma = 5; \xi = -0,75$)
 (b) EQM parâmetro forma; cenário ($\mu = 19; \sigma = 0,9; \xi = -0,46$)
 (c) VMR parâmetro forma; cenário ($\mu = 2; \sigma = 5; \xi = -0,75$)
 (d) VMR parâmetro forma; cenário ($\mu = 19; \sigma = 0,9; \xi = -0,46$)

Pelos gráficos (a) e (b) da Figura 3 é possível perceber que o erro quadrático médio do parâmetro forma da distribuição GEV do estimador de momentos-L é menor que o obtido pelo estimador de máxima verossimilhança até o tamanho amostral 35. A partir desse tamanho amostral, os dois estimadores possuem erros quadráticos médios semelhantes. Também pode-se observar que os valores dos erros quadráticos médios diminuem à medida que o tamanho da amostra aumenta. Avaliando-se os gráficos (c) e (d) da Figura 3 percebe-se que o viés médio relativo do parâmetro forma do estimador de momentos-L foi menor que o do de máxima verossimilhança até o tamanho amostral 35. Do tamanho amostral 35 até 60, os dois estimadores apresentaram vieses médios relativos semelhantes e, a partir do tamanho 60 o estimador de máxima verossimilhança apresentou viés médio relativo ligeiramente menor que o apresentado pelo estimador de momentos-L. Os resultados obtidos para o comportamento do EQM e do VMR do parâmetro forma nos demais cenários, considerando-se o parâmetro forma com valores iguais a -0,75, -0,52 e -0,46 foram semelhantes e, por isso, foram omitidos.

Na Figura 4 é apresentado o viés médio relativo dos níveis de retorno obtido com ajuste da distribuição GEV via MV e ML para os tempos de retorno de 10, 25, 50, 100, 200, 500 e 1000 anos nos seguintes cenários: parâmetro posição 2; escala 5; forma -0,75 e posição 19; escala 0,9; forma -0,46 nos tamanhos amostrais 17 e 13.

Figura 4 - Viés médio relativo dos níveis de retorno obtidos pelo ajuste da distribuição GEV via MV e ML nos tempos de retorno de 10; 25; 50; 100; 200; 500 e 1000 anos, e com os tamanhos amostrais 17 e 13 para diferentes cenários



Legenda: (a) cenário ($\mu = 2; \sigma = 5; \xi = -0,75$), $n=17$

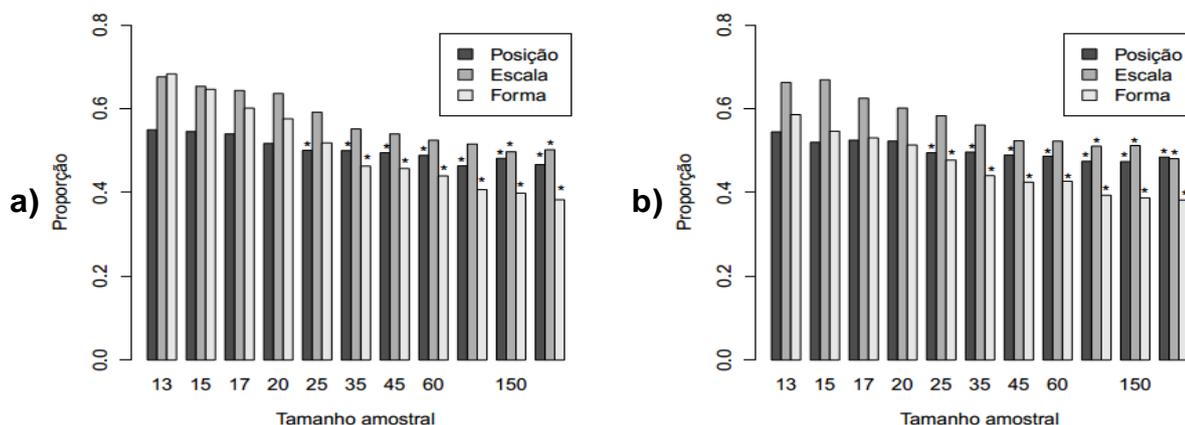
(b) cenário ($\mu = 19; \sigma = 0,9; \xi = -0,46$), $n=13$

Pela Figura 4(a) é possível observar que o viés médio do nível de retorno da distribuição GEV ajustada por meio dos dois estimadores é semelhante para o tempo de retorno de 10 anos para o cenário em que o parâmetro posição é igual a 2, o parâmetro escala é 5, o forma é $-0,75$ e o tamanho amostral é 17. Para os demais tempos de retorno, considerando-se esse cenário, o viés médio do nível de retorno da distribuição GEV, ajustada pelo método dos momentos-L é maior que o da ajustada pelo método da máxima verossimilhança. Também é possível observar que a diferença entre os vieses do nível de retorno aumenta à medida que o tempo de retorno aumenta. Resultados semelhantes foram obtidos para o viés médio do nível de retorno para todos os cenários em que o parâmetro forma é igual a $-0,75$. A maior diferença entre o viés do nível de retorno da distribuição GEV ajustada pelo método dos momentos-L e o viés do nível de retorno da distribuição GEV ajustada pelo método dos momentos-L foi de 5,40 pontos percentuais e ela aconteceu no tempo de retorno de 1.000 anos no cenário apresentado na Figura 4(a).

É possível observar na Figura 4(b) que o viés médio do nível de retorno da distribuição GEV, ajustada por meio dos dois estimadores é semelhante para os tempos de retorno de 10 e 25 anos para o cenário em que o tamanho amostral é 13, parâmetro posição é igual a 19, o parâmetro escala é 0,9 e o forma é $-0,46$. Para os demais tempos de retorno, considerando-se esse cenário, o viés médio do nível de retorno da distribuição GEV, ajustada pelo método dos momentos-L é menor que o da ajustada pelo método de MV. Também é possível observar que a diferença entre os vieses do nível de retorno aumenta à medida que o tempo de retorno aumenta. Resultados análogos foram obtidos para o viés médio do nível de retorno para todos os cenários em que o parâmetro forma é igual a $-0,46$. A maior diferença entre o viés do nível de retorno da distribuição GEV ajustada pelo método dos momentos-L e o viés do nível de retorno da distribuição GEV ajustada pelo método da MV foi de 1.627,727 e ela aconteceu no tempo de retorno de 1.000 anos no cenário cujo tamanho amostral é 13, o parâmetro posição é 19, escala é 0,9 e forma é $-0,46$. Portanto, a maior diferença obtida entre os níveis de retorno preditos a partir do ajuste da distribuição GEV pelo método da MV e o método dos momentos-L é de 162.772,7 pontos percentuais.

Na Figura 5 são apresentados o resultado do teste binomial e a proporção das amostras simuladas nos quais os estimadores de momentos-L apresentaram erros quadráticos menores que os apresentados pelos estimadores de máxima verossimilhança para os três parâmetros, considerando os seguintes cenários: parâmetro posição 2, escala 5 e forma -0,75 e posição 19, escala 0,9 e forma -0,46.

Figura 5 – Teste binomial e a proporção das amostras simuladas nos quais os estimadores de momentos-L apresentam erros quadráticos menores que os apresentados pelos estimadores de máxima verossimilhança, considerando os tamanhos amostrais estudados



Legenda: (a) cenário ($\mu = 2; \sigma = 5; \xi = -0,75$)

(b) cenário ($\mu = 19; \sigma = 0,9; \xi = -0,46$)

(c) *: Proporção não significativamente maior que 0,5 pelo teste binomial ao nível de 5% de significância

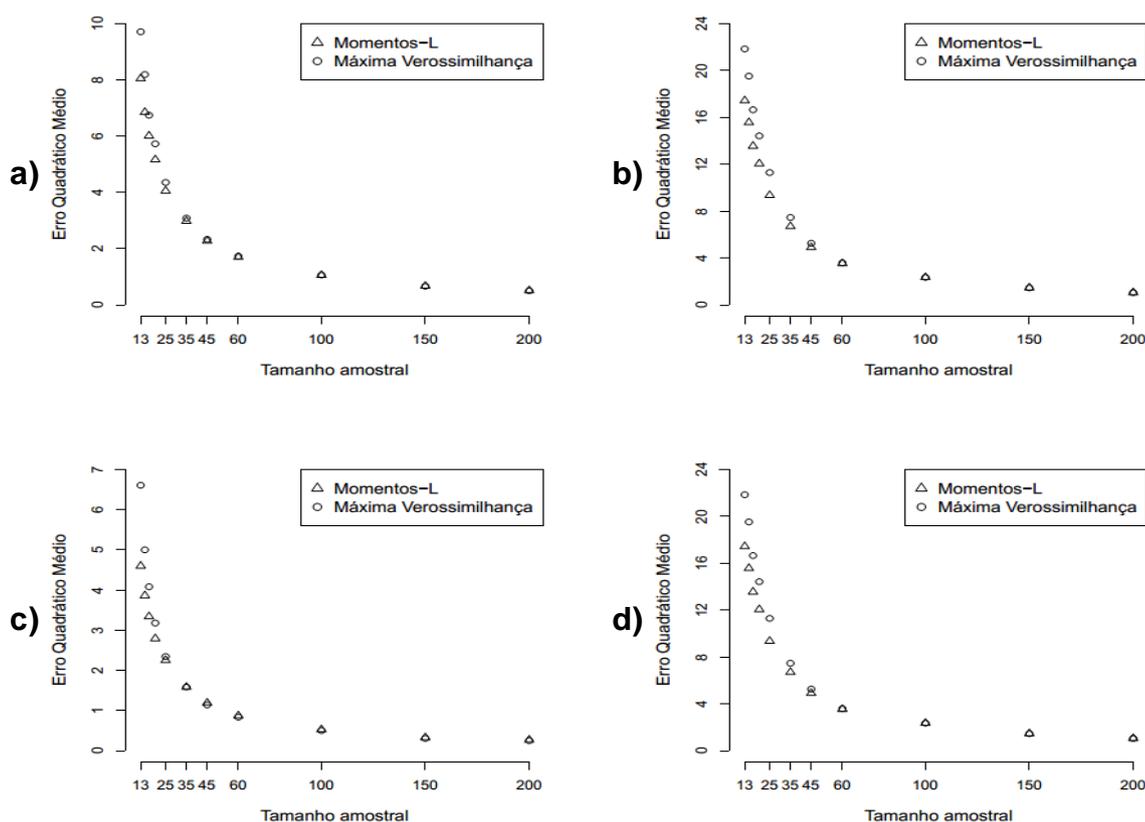
Observando-se as Figuras 5(a) e 5(b) é possível perceber que, até o tamanho amostral 20, mais da metade das amostras simuladas nos cenários apresentados tiveram menor erro quadrático para os três parâmetros da GEV quando foi utilizado o estimador de momentos-L na estimação dos parâmetros dessa distribuição. No cenário apresentado pela Figura 5(a), a proporção de amostras simuladas que obteve menor erro quadrático do parâmetro posição ao ser ajustada pelo método dos momentos-L deixa de ser significativamente maior que 0,5, pelo teste binomial ao nível de 5% de significância a partir do tamanho amostral 25. O mesmo ocorre para o parâmetro escala a partir do tamanho amostral 150 e, para o parâmetro forma, a partir de 35. Considerando-se o cenário apresentado pela Figura 5(b), a proporção de amostras simuladas que obteve menor erro quadrático dos parâmetros posição e forma ao ser ajustada pelo método dos momentos-L deixa de ser significativamente maior que 0,5, pelo teste binomial ao nível de 5% de significância a partir do tamanho amostral 25. O mesmo ocorre para o parâmetro escala a partir do tamanho amostral 100. Os demais cenários nos quais o parâmetro forma foi igual a -0,75, -0,52 e -0,46 apresentaram resultados semelhantes e, por essa razão, não foram apresentados. Portanto, mais da metade dos dados simulados com o parâmetro forma negativo e tamanho amostral menor que 25 tiveram erro quadrático dos três parâmetros menores com o uso dos estimadores de Momentos-L.

4.2 Cenários com o parâmetro forma próximo de zero.

Os resultados dos cenários simulados nos quais os valores do parâmetro forma foram -0,16 e 0,16 são apresentados nesta seção.

Na Figura 6 é apresentado o erro quadrático médio dos estimadores de máxima verossimilhança e de momentos-L obtido para os parâmetros posição e escala nos seguintes cenários: parâmetro posição igual a 37; escala 9; forma -0,16 e posição 55; escala 13; forma 0,16 respectivamente.

Figura 6 - Erro quadrático médio dos estimadores de MV e de momentos-L dos parâmetros posição (μ) e escala (σ) da distribuição generalizada de valores extremos em diferentes cenários, para diferentes tamanhos amostrais



Legenda: (a) EQM parâmetro posição; cenário ($\mu = 37; \sigma = 9; \xi = -0,16$)

(b) EQM parâmetro posição; cenário ($\mu = 55; \sigma = 13; \xi = 0,16$)

(c) EQM parâmetro escala; cenário ($\mu = 37; \sigma = 9; \xi = -0,16$)

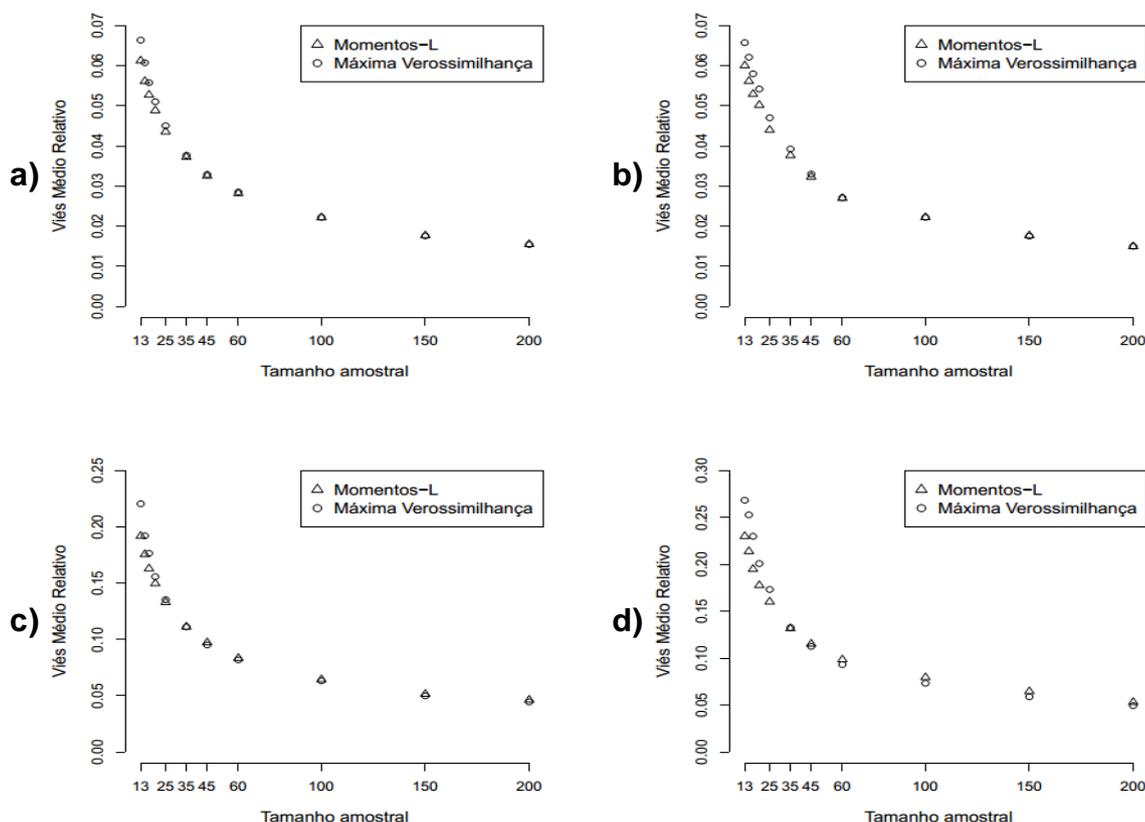
(d) EQM parâmetro escala; cenário ($\mu = 55; \sigma = 13; \xi = -0,16$)

É possível observar pela Figura 6 por meio dos gráficos 6(a) e 6(b) que os estimadores de momentos-L apresentam menor EQM que os estimadores de máxima verossimilhança para o parâmetro posição até o tamanho amostral 45. Depois desse tamanho amostral, os dois estimadores apresentaram valores similares para EQM. Observa-se pelos gráficos 6(c) e 6(d) que os estimadores de momentos-L apresentaram menores valores de EQM para o parâmetro escala até o tamanho amostral 35. Depois desse tamanho amostral, ambos estimadores apresentaram valores de EQM similares. Pode-se observar também que os valores dos EQM diminuem à medida que o tamanho amostral aumenta em todos os cenários apresentados. Os resultados obtidos para o comportamento do EQM dos parâmetros posição e escala nos demais cenários, considerando-se o parâmetro forma com valores iguais a -0,16 e 0,16 foram semelhantes e, por isso, foram omitidos.

Na Figura 7 é apresentado o viés médio relativo dos estimadores de máxima verossimilhança e de momentos-L obtido para os parâmetros posição e escala nos

seguintes cenários: parâmetro posição igual a 37; escala 9; forma -0,16 e posição 55; escala 13; forma 0,16.

Figura 7 - Viés médio relativo (VMR) dos estimadores de máxima verossimilhança e de momentos-L dos parâmetros posição (μ) e escala (σ) da distribuição GEV para cenários diferentes, tamanhos amostrais 13, 15, 17, 20, 25, 35, 45, 60, 100, 150 e 200



Legenda: (a) VMR parâmetro posição; cenário ($\mu = 37; \sigma = 9; \xi = -0,16$)

(b) VMR parâmetro posição; cenário ($\mu = 55; \sigma = 13; \xi = 0,16$)

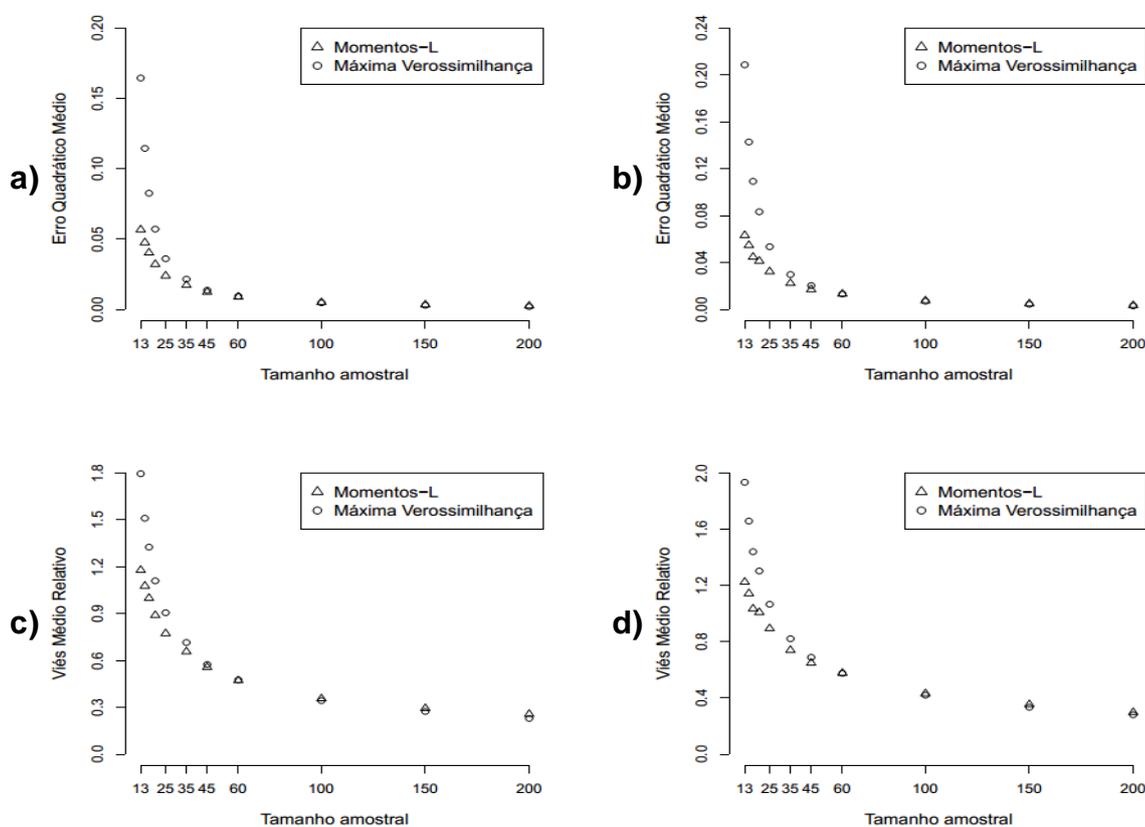
(c) VMR parâmetro escala; cenário ($\mu = 37; \sigma = 9; \xi = 0,16$)

(d) VMR parâmetro escala; cenário ($\mu = 55; \sigma = 13; \xi = -0,16$)

Pela Figura 7, observa-se por meio dos gráficos 7(a) e 7(b) que os estimadores de momentos-L apresentam menor VMR que os de máxima verossimilhança para o parâmetro posição até o tamanho de amostra igual 35. A partir desse tamanho amostral, os dois estimadores apresentam valores similares para VMR. A partir dos gráficos 7(c) e 7(d) é possível observar que os estimadores de momentos-L também apresentaram menores valores de VMR para o parâmetro escala até o tamanho amostral 35. Após esse tamanho amostral, os dois estimadores apresentam valores de VMR semelhantes. Além disso, observa ainda que os valores VMR diminuem com o aumento do tamanho amostral nos cenários apresentados. Os resultados obtidos para o comportamento do VMR dos parâmetros posição e escala nos demais cenários, considerando-se os parâmetros forma com os valores iguais a -0,16 e 0,16 foram análogos, por isso, foram omitidos.

Na Figura 8 são apresentados EQM e VMR dos estimadores de máxima verossimilhança e de momentos-L obtidos para o parâmetro forma nos cenários: posição igual a 37; escala 9; forma -0,16 e posição 55; escala 13; forma 0,16.

Figura 8 - Erro quadrático médio (EQM) e viés médio relativo (VMR) dos estimadores de máxima verossimilhança e de momentos-L do parâmetro forma (ξ), da distribuição GEV para cenários diferentes e tamanhos amostrais estudados



Legenda: (a) EQM parâmetro forma; cenário ($\mu = 37; \sigma = 9; \xi = -0,16$)

(b) EQM parâmetro forma; cenário ($\mu = 55; \sigma = 13; \xi = 0,16$)

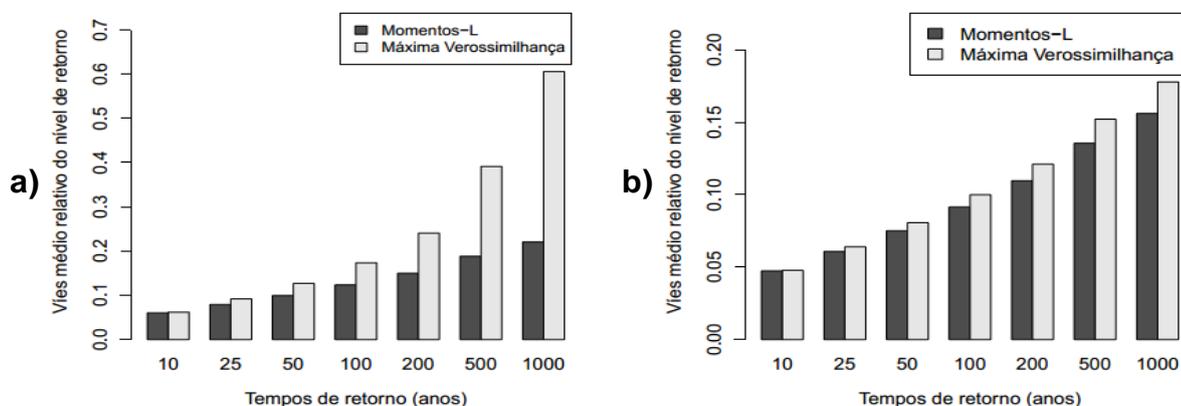
(c) VMR parâmetro forma; cenário ($\mu = 37; \sigma = 9; \xi = -0,16$)

(d) VMR parâmetro forma; cenário ($\mu = 55; \sigma = 13; \xi = 0,16$)

Observa-se pelos gráficos 8(a) e 8(b) que o erro quadrático médio do parâmetro forma da distribuição GEV do estimador de momentos-L é menor que o obtido pelo estimador de MV até o tamanho amostral 45. Após esse tamanho amostral, o erro EQM apresentado pelos dois estimadores são semelhantes. Observa-se também que o EQM do parâmetro forma da distribuição GEV diminui à medida que o tamanho amostral aumenta, considerando-se os dois estimadores. Ao analisar os gráficos 8(c) e 8(d) observa-se que o viés médio relativo do parâmetro forma do estimador de momentos-L foi menor que o do de máxima verossimilhança até o tamanho amostral 60. A partir desse tamanho amostral os dois estimadores apresentaram valores dos vieses médios relativos muito semelhantes e para ambos estimadores os valores do VMR diminuem a medida que o tamanho da amostra aumenta. Os resultados obtidos para o comportamento do EQM e do VMR do parâmetro forma nos demais cenários, considerando-se o parâmetro forma com valores iguais a -0.16 e 0,16 foram semelhantes e, portanto, foram omitidos.

Na Figura 9 é apresentado o viés médio relativo do nível de retorno obtidos no ajuste da GVE por meio dos estimadores de MV e momentos-L nos tempos de retorno de 10, 25, 50, 100, 200, 500 e 1000 anos no seguinte cenário: parâmetro posição 37, escala 9 e forma -0,16 nos tamanhos amostrais 15 e 25.

Figura 9 - Víés médio relativo do nível de retorno obtidos no ajuste da GVE por meio dos estimadores de MV e momentos-L para os tempos de retorno de 10; 25; 50; 100; 200; 500 e 1000 anos, no cenário ($\mu=37, \sigma=9, \xi=-0,16$) para os tamanhos amostrais $n=15$ e $n=25$,



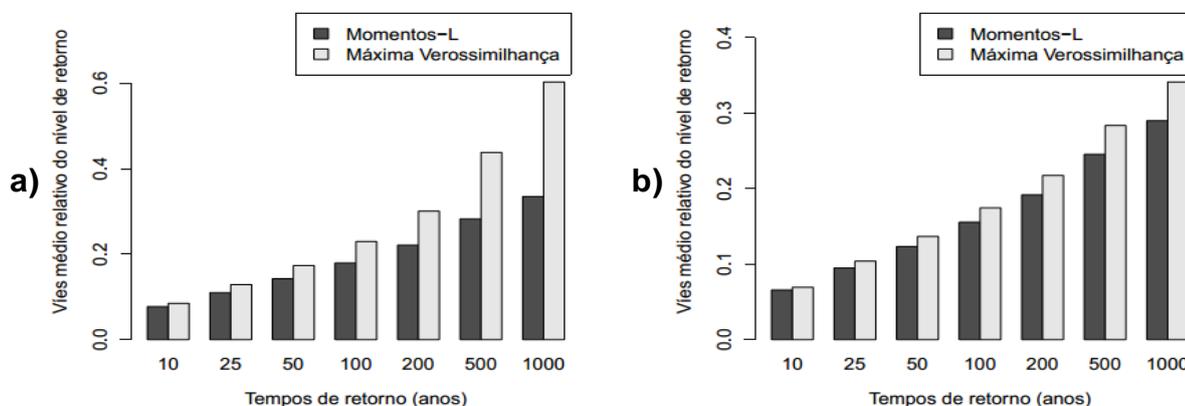
Legenda: (a) cenário ($\mu=37; \sigma=9; \xi=-0,16$), $n=15$

(b) cenário ($\mu=37; \sigma=9; \xi=-0,16$), $n=25$

Pode se observar pela Figura 9(a) que o viés médio do nível de retorno da distribuição GEV ajustada por meio dos dois estimadores é semelhante para o tempo de retorno de 10 anos para o cenário em que o parâmetro posição é igual a 37, o parâmetro escala é 9, o forma é -0,16 e o tamanho amostral é 15. Para os demais tempos de retorno, considerando-se esse cenário, o viés médio do nível de retorno da distribuição GEV, ajustada pelo método dos momentos-L é menor que o da ajustada pelo método da máxima verossimilhança. Pode-se observar na Figura 9(b) que o viés dos níveis de retorno da GEV para o momento-L vai aumentando à medida que o tempo de retorno aumenta, mas ainda apresenta viés para os níveis de retorno menores do que a GEV ajustada pelo método MV, para todos os tempos de retorno.

Na Figura 10 é apresentado o viés médio relativo do nível de retorno obtidos no ajuste da GEV aos métodos de MV e ML nos tempos de retorno de 10, 25, 50, 100, 200, 500 e 1000 anos no seguinte cenário: parâmetro posição 55, escala 13 e forma 0,16 nos tamanhos amostrais 45 e 60.

Figura 10 - Viés médio relativo do nível de retorno obtido no ajuste da distribuição GEV por meio dos métodos de MV e ML nos tempos de retorno 10, 25, 50, 100, 200, 500 e 1000 anos, no cenário: posição 55, escala 13 e forma 0,16 nos tamanhos amostrais $n=45$ e $n=60$



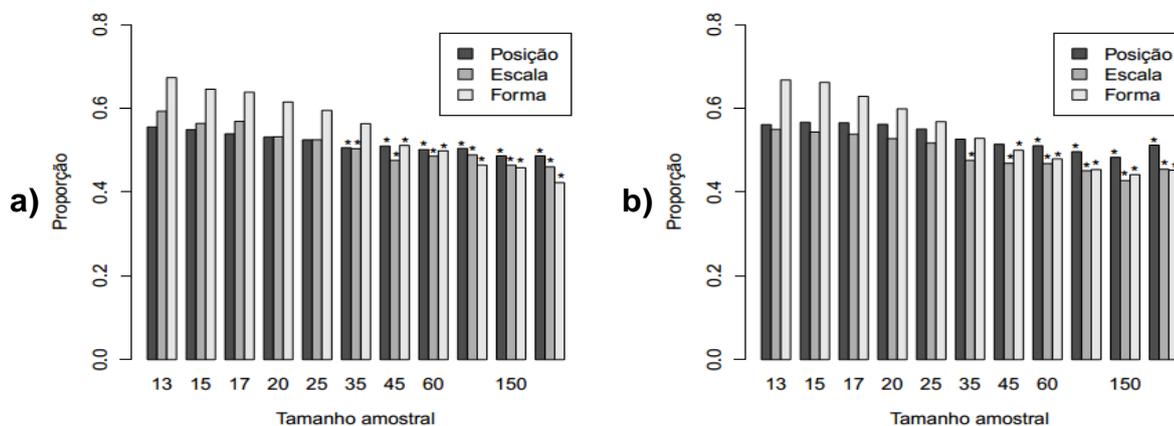
Legenda: (a) cenário ($\mu = 55; \sigma = 13; \xi = 0,16$), $n=45$

(b) cenário ($\mu = 55; \sigma = 13; \xi = 0,16$), $n=60$

Pode-se observar pela Figura 10(a) que o viés médio relativo do nível de retorno da distribuição GEV, ajustada por meio dos momentos-L, para o cenário em que o parâmetro posição é igual a 55, o parâmetro escala é 13, o parâmetro forma é 0,16 e o tamanho amostral 45 é menor que o apresentado pelo ajuste por meio do método máxima verossimilhança para todos os tempos de retorno. Ainda considerando esse mesmo cenário na Figura 10(b) para o tamanho amostral 60, pode-se observar que o viés médio relativo do nível de retorno da distribuição GEV, ajustada pelo método dos momentos-L é menor que o ajustado pelo método da máxima verossimilhança.

Na Figura 11 São apresentados o resultado do teste binomial e a proporção das amostras simuladas nos quais os estimadores de momentos-L apresentaram erros quadráticos menores que os apresentados pelos estimadores de máxima verossimilhança para os três parâmetros, considerando os seguintes cenários: parâmetro posição 37, escala 9 e forma -0,16 e posição 55, escala 13 e forma 0,16.

Figura 11 – Teste binomial e a proporção das amostras simuladas nos quais os estimadores de momentos-L apresentam erros quadráticos menores que os apresentados pelos estimadores de máxima verossimilhança, considerando os diferentes tamanhos amostrais estudados



Legenda: (a) cenário ($\mu = 37$; $\sigma = 9$; $\xi = -0,16$)

(b) cenário ($\mu = 55$; $\sigma = 13$; $\xi = 0,16$)

(c) *: Proporção não significativamente maior que 0,5 pelo teste binomial ao nível de 5% de significância

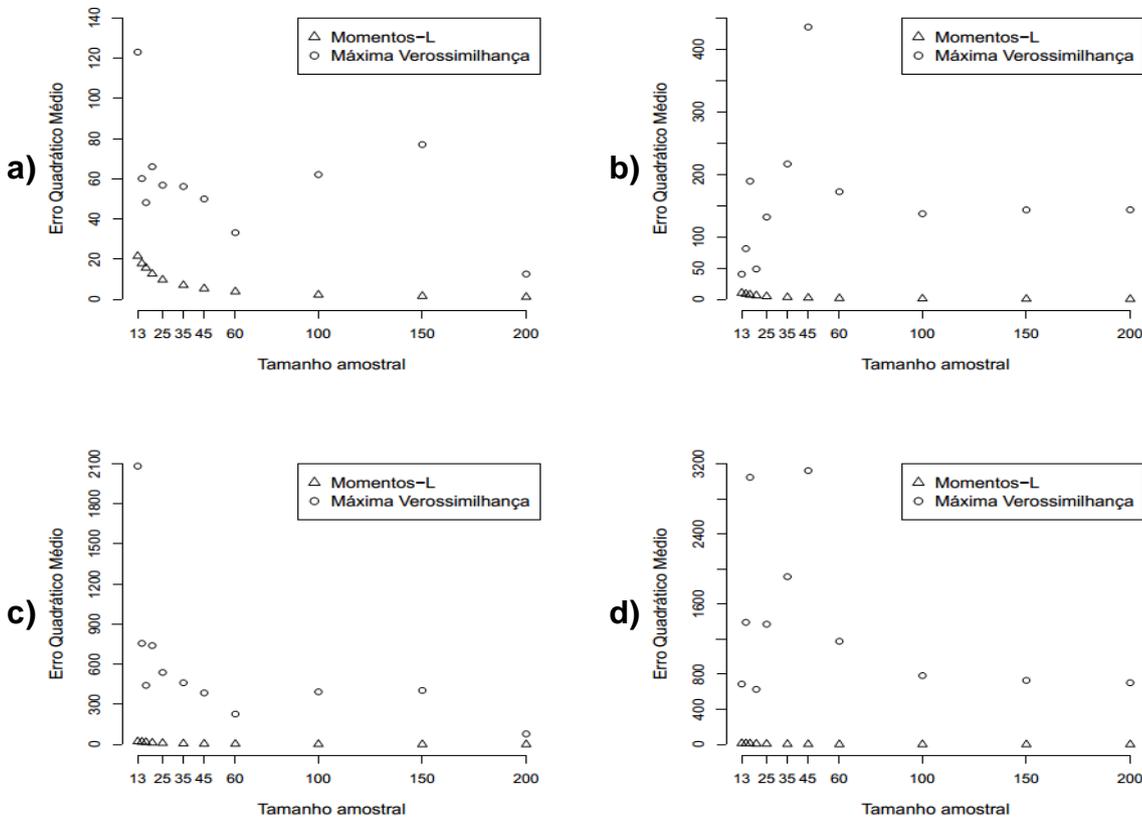
Para tamanho amostral 25 observando-se as Figuras 11(a) e 11(b) é possível perceber que, mais da metade das amostras simuladas nos cenários apresentados tiveram menor erro quadrático para os três parâmetros da GEV quando foi utilizado o estimador de momentos-L na estimação dos parâmetros dessa distribuição. No cenário apresentado pela Figura 11(a), a proporção de amostras simuladas que obteve menor erro quadrático dos parâmetros posição e escala ao ser ajustada pelo método dos momentos-L deixa de ser significativamente maior que 0,5, pelo teste binomial ao nível de 5% de significância a partir do tamanho amostral 25 e, para o parâmetro forma, a partir de 35. Considerando-se o cenário apresentado pela Figura 11(b), a proporção de amostras simuladas que obteve menor erro quadrático dos parâmetros posição e forma ao ser ajustada pelo método dos momentos-L deixa de ser significativamente maior que 0,5, pelo teste binomial ao nível de 5% de significância, a partir do tamanho amostral 45. O mesmo ocorre para o parâmetro escala a partir do tamanho amostral 35. Os demais cenários nos quais o parâmetro forma foi igual a -0,16 e 0,16 apresentaram resultados semelhantes e, por isso, não foram apresentados. Portanto, mais da metade dos dados simulados com tamanho amostral menor que 25 apresentaram erro quadrático dos três parâmetros menores com o uso dos estimadores de Momentos-L.

4.3 Cenários com o parâmetro forma positivo.

Os resultados dos cenários simulados nos quais os valores do parâmetro forma foram 0,4 e 0,5 são apresentados nesta seção.

Na Figura 12 é apresentado o erro quadrático médio (EQM) dos estimadores de máxima verossimilhança e de momentos-L dos parâmetros posição (μ) e escala (σ) da distribuição GEV nos cenários: parâmetro posição 19, escala 13 e forma 0,4 e posição 19, escala 9 e forma 0,5.

Figura 12 - Erro quadrático médio (EQM) dos estimadores de máxima verossimilhança e de momentos-L dos parâmetros posição (μ) e escala (σ) da distribuição GEV nos diferentes cenários, nos respectivos tamanhos amostrais 13, 15, 17, 20, 25, 35, 45, 60, 100, 150 e 200

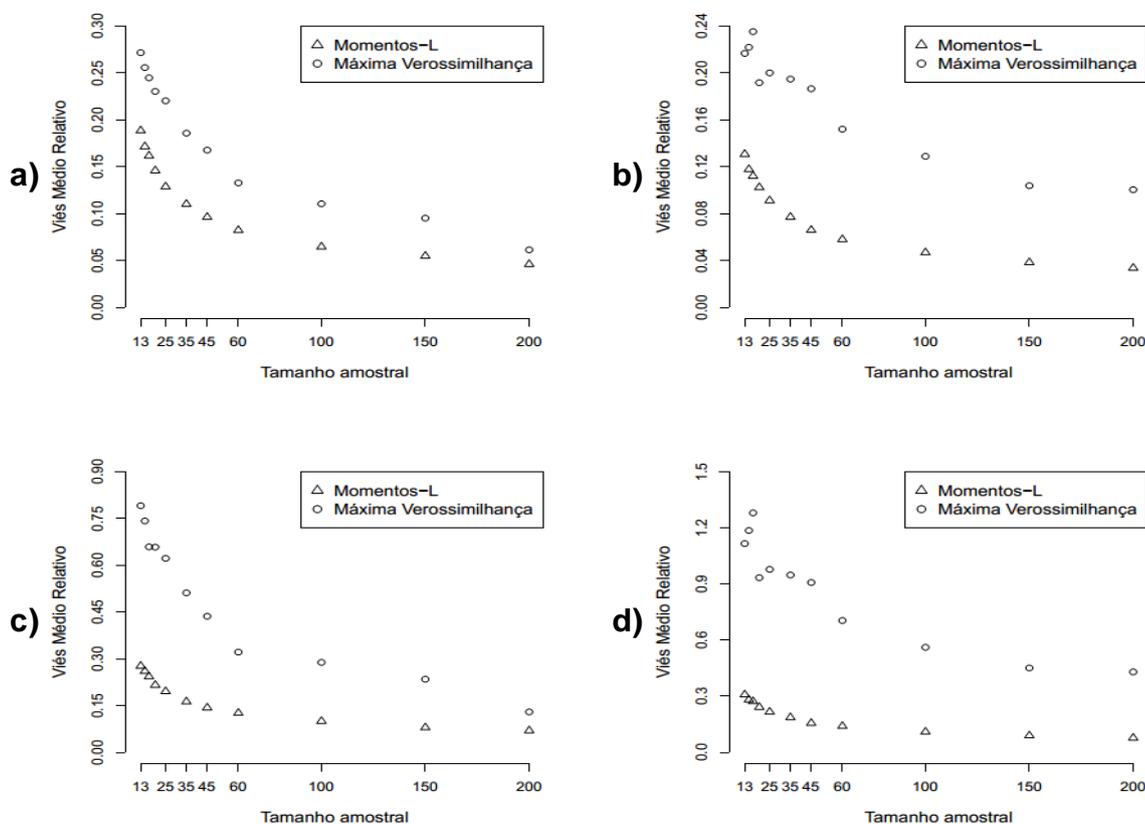


- Legenda: (a) EQM parâmetro posição; cenário ($\mu = 19; \sigma = 13; \xi = 0,4$)
 (b) EQM parâmetro posição; cenário ($\mu = 19; \sigma = 9; \xi = 0,5$)
 (c) EQM parâmetro escala; cenário ($\mu = 19; \sigma = 13; \xi = 0,4$)
 (d) EQM parâmetro escala; cenário ($\mu = 19; \sigma = 9; \xi = 0,5$)

Observando-se os gráficos da Figura 12 é possível perceber que o EQM dos estimadores de momentos-L dos parâmetros posição e escala é menor que o dos estimadores de máxima verossimilhança, para todos os tamanhos amostrais. À medida que o tamanho da amostra aumenta, o EQM dos estimadores de momentos-L diminuem. Tal comportamento não ocorre com os estimadores de máxima verossimilhança. Os resultados obtidos para o comportamento do EQM dos parâmetros posição e escala nos demais cenários, considerando-se o parâmetro forma com valores iguais a 0,4 e 0,5 foram semelhantes e, por essa razão, foram omitidos.

Na Figura 13 é apresentado o VMR dos estimadores de máxima verossimilhança e de momentos-L dos parâmetros posição (μ) e escala (σ) da distribuição GEV nos cenários: parâmetro posição 19, escala 13 e forma 0,4 e posição 19, escala 9 e forma 0,5.

Figura 13 - Apresenta os gráficos do viés médio relativo (VMR) dos estimadores de máxima verossimilhança e de momentos-L dos parâmetros posição (μ) e escala (σ) da distribuição GEV para os cenários, posição é igual a 19; escala 13; forma 0,4 e posição 19; escala 9; forma 0,5



Legenda: (a) VMR parâmetro posição; cenário ($\mu = 19; \sigma = 13; \xi = 0,4$)

(b) VMR parâmetro posição; cenário ($\mu = 19; \sigma = 9; \xi = 0,5$)

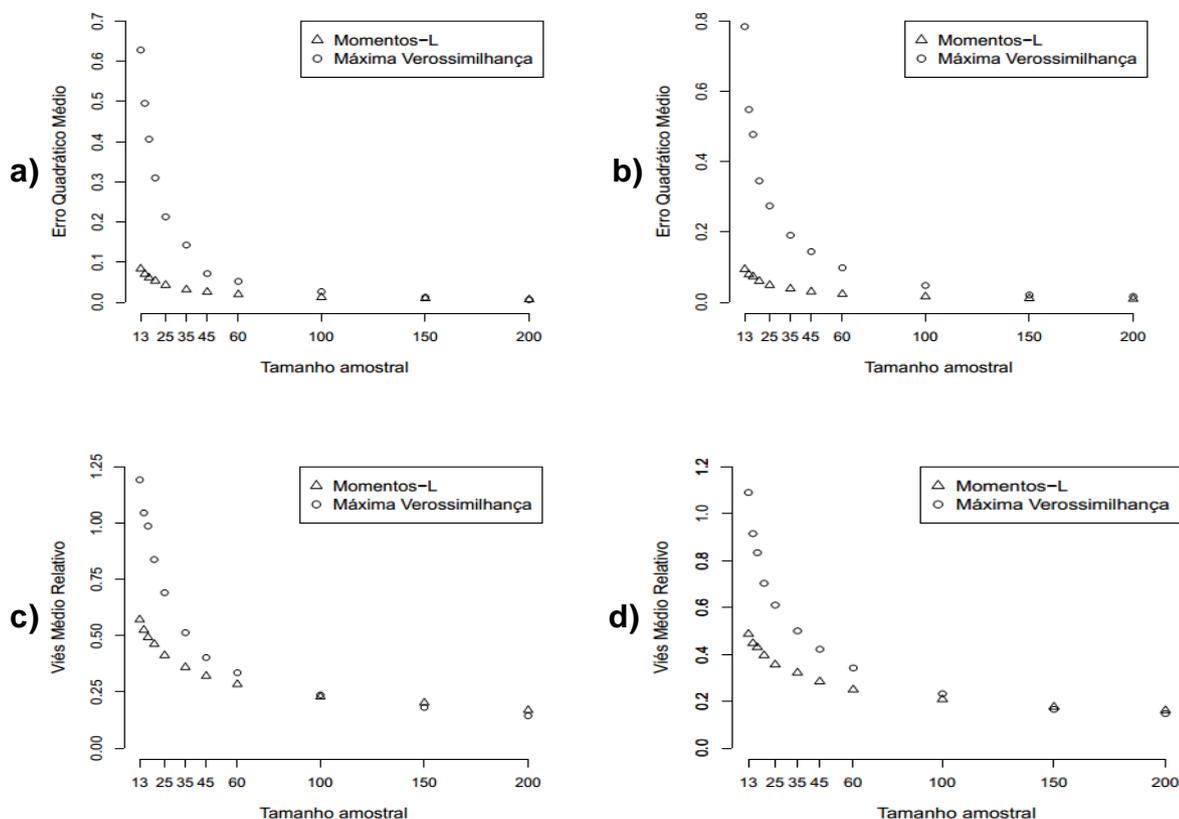
(c) VMR parâmetro escala; cenário ($\mu = 19; \sigma = 13; \xi = 0,4$)

(d) VMR parâmetro escala; cenário ($\mu = 19; \sigma = 9; \xi = 0,5$)

Observa-se pela Figura 13, por meio dos gráficos que os estimadores de momento-L apresentam menor VMR que os de máxima verossimilhança para os parâmetros posição e escala em todos tamanhos amostrais. Observa-se ainda que VMR diminuem a medida que o tamanho da amostra aumenta e que os estimadores de MV não apresenta o mesmo comportamento dos estimadores de ML. Os resultados obtidos para o comportamento do VMR dos parâmetros posição e escala nos demais cenários, considerando-se os parâmetros forma com os valores iguais a 0,4 e 0,5 foram semelhantes, por isso, foram omitidos.

Na Figura 14 são apresentados o erro quadrático médio (EQM) e (VMR) dos estimadores de máxima verossimilhança e de momentos-L do parâmetro forma da distribuição GEV nos cenários: parâmetro posição 19, escala 13 e forma 0,4 e posição 19, escala 9 e forma 0,5.

Figura 14 - Erro quadrático médio (EQM) e viés médio relativo (VMR) dos estimadores de MV e de momentos-L do parâmetro forma (ξ), da distribuição GEV para os cenários, posição é igual a 19; escala 13; forma 0,4 e posição 19; escala 9; forma 0,5 para os tamanhos amostrais 13, 15, 17, 20, 25, 35, 45, 60, 100, 150 e 200



Legenda: (a) EQM parâmetro forma; cenário ($\mu = 19; \sigma = 13; \xi = 0,4$)

(b) EQM parâmetro forma; cenário ($\mu = 19; \sigma = 9; \xi = 0,5$)

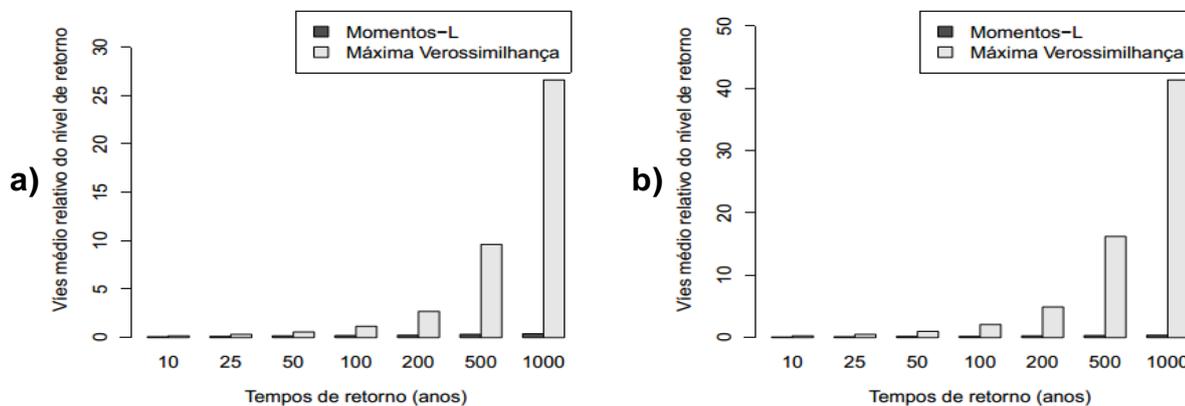
(c) VMR parâmetro forma; cenário ($\mu = 19; \sigma = 13; \xi = 0,4$)

(d) VMR parâmetro forma; cenário ($\mu = 19; \sigma = 9; \xi = 0,5$)

Dos gráficos (a) e (b) da Figura 14 é possível perceber que o erro quadrático médio do estimador de ML para o parâmetro forma da distribuição GEV é menor que o obtido pelo estimador de máxima verossimilhança para todos tamanhos amostrais até o tamanho amostral 150. A partir desse tamanho amostral os estimadores ML e MV possuem valores de EQM semelhantes. Também pode-se observar que os valores dos erros quadráticos médios diminuem para ambos estimadores à medida que o tamanho da amostra aumenta. Avaliando-se os gráficos (c) e (d) da Figura 14 nota-se que o viés médio relativo do parâmetro forma do estimador de momentos-L foi menor que o do de máxima verossimilhança até o tamanho amostral 100, a partir desse tamanho amostral o viés diminui para o MV quanto para ML. Os resultados obtidos para o comportamento do EQM e do VMR do parâmetro forma nos demais cenários, considerando-se o parâmetro forma com valores iguais a 0,4 e 0,5 foram semelhantes e, por essa razão, foram omitidos.

Na Figura 15 é apresentado o viés médio relativo do nível retorno do ajuste da GEV via MV e ML para os tempos de retorno de 10, 25, 50, 100, 200, 500 e 1000 anos nos seguintes cenários: parâmetro posição 19, escala 13, forma 0,4 e posição 19, escala 9, forma 0,5 nos tamanhos amostrais 150 e 200.

Figura 15 - Vies médio relativo do nível de retorno apresentado pelo ajuste da GEV por meio dos estimadores de MV e ML para os tempos de retorno de 10; 25; 50; 100; 200; 500 e 1000 anos, no cenário ($\mu=19; \sigma=13; \xi=0,4$) e ($\mu=19; \sigma=9; \xi=0,5$) para os tamanhos amostrais $n=150$ e $n=200$



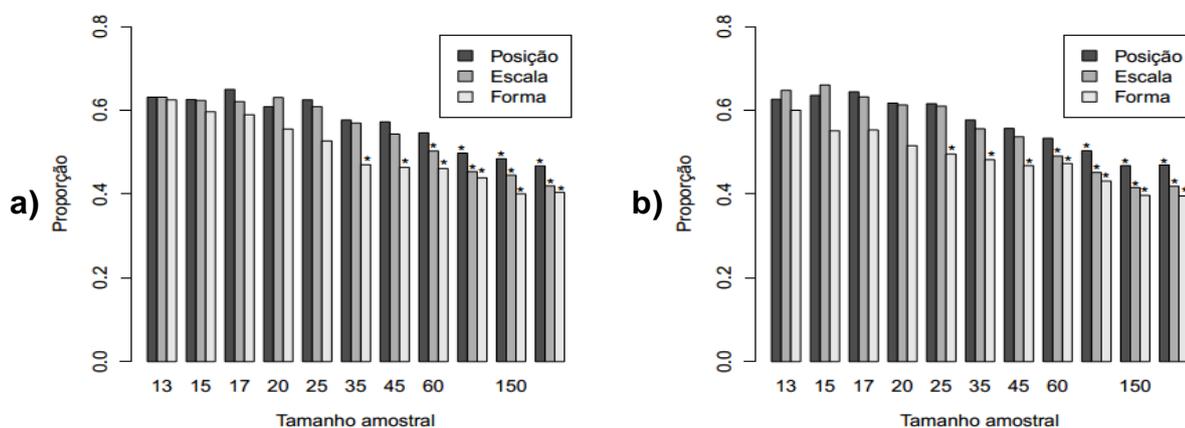
Legenda: (a) cenário ($\mu=19; \sigma=13; \xi=0,4$), $n=150$

(b) cenário ($\mu=19; \sigma=9; \xi=0,5$), $n=200$

A partir da Figura 15 é possível observar que o vies médio relativo do nível de retorno da distribuição GEV obtido por meio dos estimadores de momentos-L é menor que o obtido por máxima verossimilhança para todos os tempos de retorno observados. Observa-se também que, quanto maior o tempo de retorno, maior a diferença entre os vieses médio relativo dos níveis de retorno. Os demais cenários estudados com o parâmetro forma positivo apresentaram comportamentos semelhantes e, por essa razão, não serão apresentados.

Na Figura 16 é apresentado o resultado do teste binomial e a proporção das amostras simuladas nos quais os estimadores de momentos-L apresentaram erros quadráticos menores que os apresentados pelos estimadores de máxima verossimilhança para os três parâmetros, considerando os seguintes cenários: parâmetro posição 19, escala 13 e forma 0,4 e posição 19, escala 9 e forma 0,5.

Figura 16 – Teste binomial e a proporção das amostras simuladas nos quais os estimadores de momentos-L apresentam erros quadráticos menores que os apresentados pelos estimadores de máxima verossimilhança, considerando tamanhos amostrais estudados



Legenda: (a) cenário ($\mu = 19; \sigma = 13; \xi = 0,4$)

(b) cenário ($\mu = 19; \sigma = 9; \xi = 0,5$)

(c) *: Proporção não significativamente maior que 0,5 pelo teste binomial ao nível de 5% de significância

A partir das Figuras 16(a) e 16(b) pode-se observar que, até o tamanho amostral 20, mais da metade das amostras simuladas nos cenários apresentados tiveram menor erro quadrático para os três parâmetros da GEV quando utilizado o estimador de momentos-L no ajuste dessa distribuição. Também é possível observar que a proporção de amostras nas quais os estimadores de momentos-L apresentaram menor erro quadrático que os de máxima verossimilhança não é significativamente maior que 0,5 para os parâmetros a partir do tamanho amostral 100, ao nível de 5% de significância pelo teste binomial. Para os tamanhos amostrais intermediários, observa-se que, para pelo menos um parâmetro, a proporção citada não foi significativamente maior que 0,5, pelo teste binomial ao nível de 5% de significância. Comportamento semelhante foi observado nos demais cenários estudados com o parâmetro forma positivo. Por essa razão, a apresentação desses resultados foi omitida.

De maneira geral o EQM e o VMR encontrado a partir das estimativas do parâmetro forma por meio do ML foram menores que os encontrados por meio dos estimadores de MV para tamanhos amostrais menores ou iguais a 35. A partir desse ponto os estimadores demonstraram comportamentos semelhantes. Observou-se também que os erros quadráticos médios diminuem à medida que aumenta o tamanho da amostra. Resultado semelhante foi apresentado por Delicado e Glória (2007) ao comparar os métodos da máxima verossimilhança, momentos e de momentos-L para distribuição exponencial assimétrica de quatro parâmetros. Os autores concluíram que o método dos momentos-L possui bom desempenho perceptível para n pequeno ou moderado, e o método da máxima verossimilhança é preferível para n igual ou maior que 50.

Quanto aos parâmetros posição e escala nota-se que o método dos momentos-L se equipara ao método de máxima verossimilhança para tamanho amostral $n=35$, ou um pouco menor do que ocorre para o parâmetro forma.

O viés médio relativo do estimador de ML diminui à medida que o tamanho da amostra aumenta para todos os valores do parâmetro forma analisados, já os estimadores de MV quando os valores do parâmetro forma são positivos não apresenta o mesmo comportamento. Uma vantagem apresentada pelo método do momento-L em relação ao MV é que pode ser implementado por processo de computação menos complexo e, a

priori, não requer que se conheça lei de distribuição real dos dados, ou seja, o método momento-L permite que a partir dos momentos-L amostrais seja possível escolher a distribuição que melhor modela os dados a serem analisados (GUBAREVA; GARTSMAN, 2010).

Comparando ML e MM, Hosking (1990) concluiu que os ML são menos sujeitos a viés nas estimativas. Pois, mesmo que o método dos momentos seja conceitualmente simples, e os cálculos diretos, verifica-se que os valores numéricos da amostra podem ser muito diferentes daqueles da população da qual a amostra foi retirada, especialmente quando o tamanho da amostra é pequeno (SANKARASUBRAMANIAN; SRINIVASAN, 1999).

Valverde *et al.* (2004) e Oliveira *et al.* (2014) indicam o método momentos-L na estimação dos parâmetros da GEV principalmente em se tratando de variáveis hídricas, pois segundo Valverde *et al.* (2004) o estudo de frequência de precipitações ou de vazões extremas é dificultado, devido a baixa densidade de estações e também por curto período de tempo dos registros disponíveis.

4.4 Aplicação aos dados reais

Primeiramente, verificou-se as pressuposições de independência e aleatoriedade das observações, utilizando o teste de Ljung-Box ($p=0,2366$) e o Runs test ($p=0,8637$) respectivamente, ao nível de 5% de significância, logo conclui-se pelos resultados do valor p que os dados são independentes e aleatórios.

Na Tabela 2 são apresentadas as estimativas dos parâmetros da distribuição GEV e o resultado do teste de aderência de Kolmogorov-Smirnov.

Tabela 2- Estimativa dos parâmetros da distribuição GEV obtida pelos métodos de MV e momentos-L no ajuste da distribuição GEV aos dados de temperatura máxima de Franca-SP, no período de 1980 a 2012 e o resultado do teste de Kolmogorov-Smirnov (KS)

Métodos	Estimativas			Teste (KS) valor p
	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$	
MV	32,9069	1,3629	-0,7039	0,5006
ML	32,9087	1,4129	-0,7176	0,5921

Pelo resultado do teste de Kolmogorov-Smirnov apresentado na Tabela 2 verifica-se que a distribuição GEV é aderente aos dados. Ainda conforme a Tabela 2 nota-se que ambos estimadores podem ser utilizados para descrever o comportamento da temperatura máxima anual na cidade em análise para esse período. Com $n > 35$, este resultado está em acordo com o que foi visto no estudo de simulação, ou seja, os estimadores possuem comportamentos semelhantes a partir desse tamanho amostral.

A Tabela 3 apresenta os níveis de retorno nos tempos de retorno 2, 4 e 6 anos e o erro percentual absoluto médio (EPAM) da temperatura anual de Franca-SP.

Tabela 3 - Níveis de retorno (°C) para os tempos de retorno de 2, 4 e 6 anos da temperatura máxima anual de Franca-SP, calculadas via distribuição GEV no período de 1980 a 2012 e os erros percentual absoluto médio (EPAM)

Métodos	Tempos de retorno (em anos)			EPAM (%)
	2	4	6	
MV	33,3473	34,0376	34,2589	6,76
ML	33,3641	34,0724	34,2972	6,67
Valor observado	36,2	36,4	36,4	-

Observa-se pela Tabela 3 que os níveis máximos de temperatura °C preditos para Franca-SP, utilizando os dois métodos de estimação estão próximos aos valores observados. Quanto ao erro percentual absoluto médio o ML apresenta um percentual relativamente menor em relação ao MV. Ainda de acordo com a Tabela 3 uma interpretação dos resultados para os níveis de retorno da temperatura máxima de Franca, considerando o tempo de retorno de 6 anos é a seguinte: espera-se em um tempo médio 6 anos, que ocorra pelo menos um dia com temperatura máxima, em Franca-SP, maior ou igual a 34.3 °C.

5 CONCLUSÃO

A partir dos resultados conclui que para valores do parâmetro forma iguais a -0,75; -0,52 e -0,46 o momentos-L é indicado para tamanhos amostrais menores ou iguais a 25. Com o parâmetro forma iguais a -0,16; 0,16 o momentos-L é indicado para tamanhos amostrais menores ou iguais a 35 e para valores do parâmetro forma iguais a 0,4; 0,5 o momentos-L deve ser utilizado ao invés de utilizar o método da máxima verossimilhança em todos os tamanhos amostrais.

De modo geral pode-se concluir que para amostras menores até um tamanho amostral menor ou igual a 35 os estimadores via método dos momentos-L são mais precisos e acurados na estimação dos parâmetros da distribuição GEV. Para tamanhos amostrais maiores que 35 os dois métodos se equivalem em precisão e acurácia. Em quase todos os cenários o viés para ML diminuiu à medida que o tamanho da amostra aumentou.

REFERÊNCIAS

ARRAES, R. A.; ROCHA, A. S. Perdas extremas em mercados de risco. **Revista Contabilidade e Finanças**, São Paulo, v. 17, n. 42, p. 22-34, 2006.

BAUTISTA, E.A.L. **A Distribuição Generalizada de Valores Extremos no Estudo da Velocidade Máxima do Vento em Piracicaba, SP**. 2002. 61f. Dissertação (Mestrado em Agronomia) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2002.

BEIJO, L.; AVELAR, F.G. Distribuição Generalizada de Valores Extremos no estudo de dados climáticos: uma breve revisão e aplicação. **Revista da Estatística da UFOP**, Ouro Preto, v. 1, p. 10-16, 2011.

BOLFARINE, H.; SANDOVAL, M. C. **Introdução à Inferência Estatística**. Rio de Janeiro: Sociedade Brasileira de Matemática, 2001. 125 p.

BLAIN, G. C. Cento e vinte anos de totais extremos de precipitação pluvial máxima diária em Campinas, Estado de São Paulo: análises estatísticas. **Bragantia**, Campinas, v. 70, n. 3, p. 722-728, 2011.

COLES, S. G. **An introduction to statistical modeling of extreme values**. London: Springer, 2001. 208p.

DELICADO, P.; GORIA, M. N. A small sample comparison of maximum likelihood, moments and L-moments methods for the asymmetric exponential power distribution. **Computational Statistics & Data Analysis**.v.52, n.3,p. 1661-1673,2008.

FISHER, R. A.; TIPPETT, L. H. C. Limiting forms of the frequency distribution of largest or smallest member of a sample. **Proceedings of the Cambridge Philosophy Society**, Cambridge, v. 24, p. 180-190, 1928.

GNEDENKO, B. Sur la distribution limite du terme maximum d'une série aléatoire, **Annals of Mathematics**, v. 44, n. 3, p. 423-453, 1943.

GREENWOOD, J.A. *et al.* R. Probability weighted moments: definition and relation to parameters expressible in inverse form. **Water Resources Research**, v.15, n.5, p.1049-1054, 1979.

GUBAREVA, T.S.; GARTSMAN, B.I. Estimating distribution parameters of extreme hydrometeorological characteristics by L-moments method. **Water Resources**. v.37, n.4, p.437-445, 2010.

HOSKING J. R. M. The theory of probability weighted moments. Research Report RC 12210. **Yorktown Heights (NY): IBM Research**, 1986.

HOSKING, J. R. M. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. **Jornal da Royal Statistical Society**, v. 52, n.1, p. 105-124, 1990.

INMET, Instituto Nacional de Meteorologia do Brasil. Dados históricos anuais, 1980-2018. Disponível em: <https://portal.inmet.gov.br/>. Acesso em: 15 jun. 2020.

JURCA, J. **Classificações climáticas: variações temporo-especiais e suas aplicações nos livros didáticos e como subsídios ao zoneamento agroclimático**. 2005. 100f. Dissertação (mestrado)-Universidade Estadual Paulista, Faculdade de ciência e tecnologia, Presidente Prudente, 2005.

JENKINSON, A. F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. **Quartely Journal of the Royal Meteorological Society**, v. 81, n. 348, p. 158-171, 1955.

LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. **Biometrika**, v. 65, n. 2, p. 297-303, 1978.

MARENGO, J. A. O futuro clima do Brasil. **Revista USP**, n. 103, p. 25-32, 2014.

NAGHETTINI, M.; PINTO, E. J. A. **Hidrologia estatística**, 1.ed, Belo Horizonte: CPRM, 2007, 552p.

OLIVEIRA, A. S. de. *et al.* Aplicabilidade da distribuição GEV ao estudo da precipitação máxima diária anual na região sul de Minas Gerais. **Revista Agrogeoambiental**. v.6, n.1, p.31-44.2014.

QUADROS, L. E.; QUEIROZ, M. M. F.; BOAS, M. A. V. Distribuição de frequência etemporal de chuvas intensas. **Acta Scientiarum Agronomy**, v. 33, n.3, p. 401-410,2011.

R Core Team (2019). **R: a language and environment for statistical computing**. R Foundation for Statistical Computing, Viena, Austria. URL <https://www.R-project.org/>.

SANKARASUBRAMANIAN, A.; SRINIVASAN, K. Investigation and comparison of sampling properties of L-moments and conventional moments. **Journal of Hydrology**, v. 218, n. 1-2, p.13-34,1999.

SIEGEL, S.; N. J. CASTELLAN JÚNIOR. **Estatística não-paramétrica para ciências do comportamento**. 2. ed. Porto Alegre: Artmed, 2006. 448 p.

ŠIMKOVÁ, T.; PICEK, J. A. Comparison of L-, LQ-, TL-moment and maximum likelihood high quantile estimates of the GPD and GEV distribution. **Communications in Statistics - Simulation and Computation**, v.46, n.8, p.5991–6010,2017.

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

VALVERDE, A.E.L. *et al.* Momentos-l: teoria e aplicação em hidrologia. **Revista Árvore**. v.28, n.6, p.927-933,2004.

WOLFF, W. **Avaliação e nova proposta de regionalização hidrológica para o Estado de São Paulo**. 2013.112f. Dissertação (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2013.

3 CONSIDERAÇÕES FINAIS

O estudo realizado através de simulação mostrou que, para amostras pequenas, os estimadores dos momentos-L são mais adequados na estimação dos parâmetros da distribuição GEV. Para tamanhos amostrais grandes os dois métodos se equivalem. No caso particular do parâmetro forma ser positivo, os estimadores de momentos-L é mais adequado em todos os cenários. Assim, o método dos momentos-L pode ser utilizado em todos os cenários sem prejuízos em comparação com o método da máxima verossimilhança na estimação dos parâmetros da distribuição GEV. A análise da aplicação dos dois estimadores no ajuste da distribuição GEV aos dados reais mostrou que os dois possuem adequabilidade semelhante, corroborando com os resultados obtidos a partir de dados simulados.

REFERÊNCIAS

- BEIJO, L. A. **Distribuição de Gumbel: estudo de métodos de estimação dos parâmetros e ajuste aos dados de precipitação máxima de Lavras, Minas Gerais**. 2002. 91f. Dissertação (Mestrado em Estatística Aplicada) -Universidade Federal de Lavras, Lavras, MG, 2002.
- CASELLA, G.; BERGER, R. L. **Inferência estatística**. Tradução da 2. ed. Norte-Americana. São Paulo: Cengage Learning, 2010. 588p.
- FERREIRA, D.F. **Estatística Básica**. 2. ed. Lavras: UFLA, 2009. 664p.
- FRÉCHET, M. Sur la loi de probabilité de l'écart maximum. **Ann. Soc. Polon. Math.** Cracovie, v. 6, p. 93-116, 1927.
- GAJO, C. A. **Propriedades e aspectos geométricos de estimadores tipo James-Stein e do estimador de Hartigan**. 2016. 156f. Tese (Doutorado em Estatística e Experimentação Agropecuária)-Universidade Federal de Lavras, Lavras, 2016.
- KIRKPATRICK, S.; GELATT JR., C.D.;VECCHI, M.P. Optimization by simulated annealing, **Science**, v. 220, n. 4598,p.671-680, 1983.
- MORETTIN. L. G. **Estatística básica: Probabilidade e Inferência**. São Paulo: Pearson Prentice Hall, 2010, 378p.
- NELDER, J. A.; MEAD, R. A simplex algorithm for function minimization. **Computer Journal** 7, v. 7, n. 4, p. 308-313, 1965.
- NOGUEIRA, R. S. **Precisão e acurácia dos estimadores de máxima verossimilhança dos parâmetros da distribuição Gumbel não estacionária**. 2017. 54 f. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Alfenas, Alfenas, MG, 2017.
- SILVA, R. R. **A distribuição generalizada de Pareto e mistura de distribuições de Gumbel no estudo da vazão e da velocidade máxima do vento em Piracicaba, SP**. 2008. 101f. Dissertação (Mestrado em Agronomia) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2008.
- VIVANCO, M. J. F. **Análise de valores extremos no tratamento estatístico da corrosão de equipamentos**. 1994.107 f. Dissertação (Mestrado em Estatística) - Universidade Estadual de Campinas, Campinas,1994.