

Distribuição Generalizada de Valores Extremos (GVE): Um estudo aplicado a valores de temperatura mínima da Cidade de Viçosa-MG

Eduardo Campana Barbosa¹²

Carlos Henrique Osório Silva³

Rômulo César Manuli²

Ricardo Gonçalves Tavares⁴

Tiago Bittencourt Nazare⁵

1 Introdução

A modelagem de eventos extremos, principalmente os que envolvem variáveis climatológicas, como por exemplo, a ocorrência de valores mínimos de temperatura, é de fundamental importância para o gerenciamento de setores e atividades realizadas por empresas, agricultores e pessoas comuns. A ocorrência de valores mínimos de temperatura influencia negativamente na construção civil, turismo, lazer, agricultura, transporte, saúde pública etc. Para Bakonyi et al. (2004) a saúde pública merece cuidado especial, visto que baixos níveis de temperatura implicam no aumento de atendimentos ambulatoriais, causados por doenças respiratórias.

O presente trabalho tem por objetivo a modelagem de valores mínimos de temperatura da cidade de Viçosa (MG) para os meses de Junho, Julho, Agosto e Setembro. Para tal utilizou-se a distribuição Generalizada de Valores Extremos (GVE), Gumbel e Weibull. Como Viçosa é uma cidade universitária diversos estudantes residem no município, fato que requer atenção das autoridades em relação à saúde pública no período de baixas temperaturas. Destaca-se ainda que as principais linhas de pesquisa da Universidade Federal de Viçosa (UFV) estão relacionadas a ciências agrárias, e, portanto, existem diversos experimentos/pesquisas de alunos de mestrado e doutorado que podem ser afetados pela ocorrência destes eventos extremos.

¹ Agradecimentos à Fapemig, Capes e CNPq pelo apoio financeiro.

² Mestrando em Estatística Aplicada e Biometria (DET-UFV); duducampana@hotmail.com; romulomanuli@ig.com.br

³ Professor Associado III do Departamento de Estatística da UFV (DET-UFV); chos@ufv.br

⁴ Aluno de Ciências da Computação da UFV (DPI-UFV); ricardo.tavares@ufv.br

⁵ Professor das Faculdades Integradas de Cataguases (FIC-UNIS); tiago@unis.edu.br

2 Materiais e Métodos

Os dados foram obtidos no Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP) do Instituto Nacional de Meteorologia (INMET). As informações referem-se à série de temperatura (em °C) da cidade de Viçosa-MG, desde 03/01/1961 à 01/10/2013, coletadas diariamente no horário de 12:00. Destaca-se que nos períodos entre Janeiro de 1984 à Setembro de 1990 não há informações disponíveis no site. O valor mínimo de temperatura, obtido nos meses de Junho, Julho, Agosto e Setembro e em cada ano, foram agrupados constituindo 4 séries temporais.

Procedeu-se o *Runs Test* (Wald; Wolfowitz, 1940) e Ljung-Box (Ljung & Box, 1978) para verificar, respectivamente, a aleatoriedade e independência das séries, pressuposições fundamentais para aplicação da teoria de valores extremos.

Utilizou-se a distribuição GVE para o ajuste dos dados, porém, conforme o valor estimado do parâmetro de forma (ξ), outras distribuições de valores extremos foram sugeridas para critério de comparação. A estimação dos parâmetros da distribuição GVE e demais ocorreu por máxima verossimilhança. Para verificar se $\xi = 0$ procedeu-se o Teste da Razão de Verossimilhanças Modificado, conforme Hosking (1984). A escolha da distribuição que melhor se ajustou a cada conjunto de dados foi pelo teste Kolmogorov-Smirnov. Posteriormente, foram calculadas as probabilidades $P[X \leq x]$, onde x representa um determinado valor de temperatura mínima.

Utilizou-se o *software* livre R (R Development Core Team, 2012) para manipulação e análise estatística dos dados. O nível de significância adotado nos testes foi $\alpha = 5\%$.

2.1 Distribuição GVE

Proposta por Jenkinson (1955) a Distribuição Generalizada de Valores Extremos (*Generalized Extreme Value*) possui função densidade de probabilidade (f.d.p) conforme em (1) e Função de Probabilidade $F(x)$ dada em (2). Os termos ξ , μ e σ representam, respectivamente, os parâmetros de forma, posição e escala.

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

$$F(x|\xi, \mu, \sigma) = P[X \leq x] = \int_0^x f(x|\xi, \mu, \sigma) \cdot dx = \exp \left\{ - \left[1 - \xi \left(\frac{x - \mu}{\sigma} \right)^{\frac{1}{\xi}} \right] \right\} \quad (2)$$

Definida em $-\infty < x < \mu - \frac{\sigma}{\xi}$ se $\xi < 0$, $-\infty < x < \infty$ se $\xi \rightarrow 0$ e por fim $\mu - \frac{\sigma}{\xi} < x < \infty$ se $\xi > 0$ a distribuição GVE combina os três tipos de distribuições de valores extremos, ou seja, Gumbel (Tipo I), Frechet (Tipo II) e Weibull (Tipo III). Em outras palavras, estas 3 distribuições são casos particulares da distribuição GVE, pois se $\xi = 0$ tem-se distribuição Gumbel, se $\xi < 0$ Weibull e $\xi > 0$ Frechet. Maiores detalhes em Liska e Beijo (2012) e Beijo e Avelar (2010)

3 Resultados e Discussão

A Tabela 1 apresenta os resultados dos testes Runs e Ljung-Box.

Tabela 1: *p-value* para os testes de aleatoriedade e independência

Séries	Runs Test	Ljung-Box
Junho	0,0672	0,0988
Julho	0,0207	0,5966
Agosto	0,1657	0,9091
Setembro	1,0000	0,9817

O Ljung-Box demonstra que as 4 séries são independentes, pois para todas o *p-value* foi superior a 0,05 (séries não autocorrelacionadas). Porém, a pressuposição de aleatoriedade foi violada para a série de Julho (*p-value* < 0,05), o que implica que a mesma não é identicamente distribuída. Logo, procedeu-se a estimação dos parâmetros da distribuição GVE apenas para as séries aleatórias, conforme visualizado na Tabela 2.

Tabela 2: EMV e os respectivos [Erros-Padrão] para a Distribuição GVE

Parâmetros	Junho	Agosto	Setembro
$\hat{\xi}$	-0,3676 [0,1078]	-0,6654 [0,1432]	-0,2136 [0,1055]
$\hat{\sigma}$	2,4466 [0,2985]	2,0853 [0,3136]	1,8392 [0,2185]
$\hat{\mu}$	5,7646 [0,4085]	6,4037 [0,3523]	7,8584 [0,3094]

O Teste da Razão de Verossimilhança Modificado indicou que para as séries de Junho e Agosto, $\xi \neq 0$ (*p-value* < 0,05). Como as respectivas estimativas foram negativas, sugere-se, além da distribuição GVE, a Weibull. Para a série de Setembro, $\xi = 0$ (*p-value* > 0,05), o que implica que a distribuição Gumbel pode também ser utilizada.

Tabela 3: Resultados do Teste da Razão de Verossimilhanças Modificado

TRV	Junho	Agosto	Setembro
<i>Deviance</i>	8,0651	15,8657	2,9304
χ_1^2	3,8414	3,8414	3,8414
<i>p-value</i>	0,0045	0,0000	0,0869

Na Tabela 4 os EMV para os modelos Weibull (Junho e Agosto) e Gumbel (Setembro). Posteriormente, na Tabela 5, o resultado do teste de aderência de Kolmogorov-Smirnov para comparar o ajuste das distribuições.

Tabela 4: EMV e os respectivos [Erros-Padrão] para distribuição Weibull e Gumbel

Distribuição	Junho	Agosto	Setembro
Weibull	3,0334 [0,3681] 7,2686 [0,3786]	4,3846 [0,5526] 7,3697 [0,2686]	-
Gumbel	-	-	1,7641 [0,1993] 7,6534 [0,2816]

Tabela 5: *p-value* para o teste de Kolmogorov-Smirnov

Distribuições	Junho	Agosto	Setembro
GVE	0,8107	0,9811	0,9636
Weibull	0,9495	0,8216	-
Gumbel	-	-	0,7206

Para a série de Junho a distribuição Weibull apresentou melhor ajuste (maior *p-value*). Para as séries de Agosto e Setembro, a distribuição GVE foi superior em relação à Weibull e Gumbel, respectivamente. Neste sentido, tais distribuições foram utilizadas para estimar $P[X \leq x]$ nos meses de Junho, Agosto e Setembro para a cidade de Viçosa-MG, conforme apresentado na Tabela 6.

Tabela 6: Probabilidade (%) de ocorrência de Temperaturas menores que um dado x

Meses	x	0	1	2	3	4	5	6	8	10
Junho	$P[X \leq x]$	0,00	0,24	1,98	6,60	15,07	27,49	42,81	73,75	92,81
Agosto	$P[X \leq x]$	0,49	1,10	2,38	4,88	9,51	17,48	30,12	70,96	100
Setembro	$P[X \leq x]$	0,00	0,00	0,00	0,03	0,35	2,18	8,24	39,64	76,94

4 Conclusões

Pode-se concluir que a Distribuição Generalizada de Valores Extremos (GVE) obteve um bom ajuste aos dados analisados, sendo inferior apenas ao modelo Weibull para a série de temperatura mínima do mês de Junho. Além disso, sua utilização simplifica a escolha de

distribuições de valores extremos. De acordo com os resultados, em Junho e Agosto a probabilidade de ocorrência de temperaturas menores ou iguais a 0, 1, 2 ou 3 °C é baixa, porém, ainda existe esta possibilidade. Nestes mesmos meses, a probabilidade de ocorrência de temperaturas menores ou iguais a 4, 5 ou 6 °C é significativa. O mês de Setembro apresenta baixas probabilidades para a ocorrência de temperaturas extremamente mínimas, porém, com 76,94% de probabilidade, é provável que ocorra em Viçosa temperaturas inferiores a 10 °C. Logo, a atenção quanto à agricultura, experimentos e saúde pública devem concentrar-se principalmente nos meses de Junho e Agosto. Destaca-se que a série de Julho não foi aleatória e por isso não foi possível concluir sobre a mesma por esta abordagem. Sugere-se outros tipos de metodologias para sua análise, como por exemplo, séries temporais via modelo Box & Jenkins.

5 Referências

- [1] BAKONYI, S. M. C.; DANNI-OLIVEIRA, I. M.; MARTINS, L. C.; BRAGA, L. F. Poluição atmosférica e doenças respiratórias em crianças na cidade de Curitiba, PR. **Rev. Saúde Pública, São Paulo, v.38, n.5, p.695-700, 2004.**
- [2] BEIJO, L.A.; AVELAR, F. G. Distribuição Generalizada de Valores Extremos no estudo de dados climáticos: uma breve revisão e aplicação. **Revista da Estatística da Universidade Federal de Ouro Preto, v. 1, p. 10-16, 2011.**
- [3] HOSKING, J. R. M. Algorithm AS 215: maximum-likelihood estimation of the parameters of the generalized extreme-value distribution. **Journal of the Royal Statistical Society . Series C . Applied statistics, London, v. 34, p. 301-310, 1985.**
- [4] JENKINSON, A. F. The frequency distribution of the annual maximum (or minimum) of meteorological elements. **Quart. Journal R. Meteorological Soc, v. 81, p. 158-171. 1955.**
- [5] LISKA, G. R. ; BEIJO, L. A. Distribuições de Probabilidade Aplicadas na Análise de Níveis Máximos de MP10 e O3 nas cidades de Cubatão-SP e Paulínia-SP. **Revista de Estudos Ambientais (Online), v. 14, p. 35-47, 2012.**
- [6] LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. **Biometrika, v.65, p 297–303, 1978.**
- [7] **R DEVELOPMENT CORE TEAM. R: a language and environment for statistical computing.** Vienna: R Foundation for Statistical Computing.
- [8] WALD, A.; WOLFOWITZ, J. On a test whether two samples are from the same population. **Ann. Math Statist., v. 11, p. 147-162, 1940.**