



Universidade de Brasília
Departamento de Estatística

Modelo de regressão para valores extremos bimodais

Mathews de Noronha Silveira Lisboa

Dissertação apresentada ao Departamento
de Estatística da Universidade de Brasília,
como requisito parcial para obtenção do
título de Mestre em Estatística

Brasília
2024

Mathews de Noronha Silveira Lisboa

Modelo de regressão para valores extremos bimodais

Orientadora: Prof^ª. Dr^ª. Cira Etheowalda Guevara Otiniano

Dissertação apresentada ao Departamento
de Estatística da Universidade de Brasília,
como requisito parcial para obtenção do
título de Mestre em Estatística.

**Brasília
2024**

Agradecimentos

Primeiramente, agradeço à Prof. Dr. Cira E. G. Otiniano, por sua orientação incansável, paciência e valiosas contribuições ao longo deste processo. Sua expertise e dedicação foram fundamentais para o desenvolvimento deste trabalho.

Agradeço também à Prof. Dr. Terezinha K. A. Ribeiro, por seu apoio constante, conselhos e incentivo contínuo. Sua orientação foi essencial para a conclusão deste projeto.

Não poderia deixar de agradecer aos meus pais por seu amor incondicional, apoio e compreensão durante toda a minha jornada acadêmica. Vocês foram minha fonte de força e inspiração, e este trabalho é dedicado a vocês.

A todos, meu sincero muito obrigado.

Resumo

Eventos climáticos extremos resultam da combinação de valores extremos de duas ou mais variáveis aleatórias climáticas. Por outro lado, quando pelo menos uma dessas variáveis apresenta características multimodais, os sistemas climáticos tornam-se mais complexos. A distribuição GEV bimodal, proposta por Otiniano (2023), surge como uma alternativa para a modelagem univariada de extremos em sistemas climáticos complexos. Neste trabalho, propomos um novo modelo de regressão que descreve a relação entre uma variável resposta de valores extremos bimodais com outras variáveis aleatórias. Esse modelo de regressão é baseado na distribuição GEV bimodal (BGEV). Uma das inovações do novo modelo de regressão BGEV é a reparametrização do parâmetro de localização em termos de sua mediana, o que facilita a interpretação dos resultados. A estimação dos parâmetros do modelo de regressão BGEV foi desenvolvida com base no método de máxima verossimilhança. O desempenho desse estimador foi avaliado a partir de estudos de simulação de Monte Carlo. Por fim, o modelo de regressão BGEV foi aplicado a dados climáticos da estação meteorológica A002, localizado em Goiânia, esses dados são inéditos sendo essa uma novidade apresentada neste trabalho. A adequabilidade deste ajuste foi validada com base nos resíduos quantílicos.

Palavras-chave: BGEV; Modelo de regressão; Resíduos quantílicos; dados climáticos.

Abstract

Extreme climate events result from the combination of extreme values of two or more random climate variables. On the other hand, when at least one of these variables exhibits multimodal characteristics, climate systems become more complex. The bimodal GEV distribution, proposed by Otiniano (2023), emerges as an alternative for the univariate modeling of extremes in complex climate systems.

In this work, we propose a new regression model that describes the relationship between a bimodal extreme value response variable and other random variables. This regression model is based on the bimodal GEV distribution (BGEV).

One of the innovations of the new BGEV regression model is the reparameterization of the location parameter in terms of its median, which facilitates the interpretation of the results.

The estimation of the parameters of the BGEV regression model was developed based on the maximum likelihood method. The performance of this estimator was evaluated through Monte Carlo simulation studies. Finally, the BGEV regression model was applied to climate data from the A002 meteorological station located in Goiânia. These data are unprecedented, making this a novel contribution presented in this work. The adequacy of this fit was validated based on quantile residuals.

Keywords: GEV, Bimodal, Regression, MLE, Quantile residuals, Dew point, Relative humidity, Climate, Meteorology.

Lista de Figuras

1	Curvas para a função de densidade de probabilidade para as distribuições de valores extremos (com $\alpha = 1$)	12
2	$Y \sim F_{BG_{0,\mu,1,1}}$ com μ variando. Gráfico superior FDP, gráfico inferior FDA.	16
3	$Y \sim F_{BG_{0,0,\sigma,1}}$ com σ variando. Gráfico superior FDP, gráfico inferior FDA.	16
4	$Y \sim F_{BG_{0,0,1,\delta}}$ com δ variando. Gráfico superior FDP, gráfico inferior FDA.	17
5	$Y \sim F_{BG_{\xi,0,1,1}}$ com ξ variando. Gráfico superior FDP, gráfico inferior FDA.	17
6	$Y \sim F_{BG_{-0.5,0,1,\delta}}$ e δ variando. Gráfico superior da FDP. Gráfico inferior da FDA.	18
7	$Y \sim F_{BG_{0.5,0,1,\delta}}$ e δ variando. Gráfico superior da FDP. Gráfico inferior da FDA.	18
8	Boxplots dos parâmetros estimados sob o Cenário 1. A linha tracejada vermelha representa o verdadeiro valor do parâmetro	35
9	Boxplots dos parâmetros estimados sob o Cenário 2. A linha tracejada vermelha representa o verdadeiro valor do parâmetro	36
10	Boxplots dos parâmetros estimados sob o Cenário 3. A linha tracejada vermelha representa o verdadeiro valor do parâmetro	37
11	Boxplots dos parâmetros estimados sob o Cenário 4. A linha tracejada vermelha representa o verdadeiro valor do parâmetro	38
12	Boxplots dos parâmetros estimados sob o Cenário 5. A linha tracejada vermelha representa o verdadeiro valor do parâmetro	39
13	Boxplots dos parâmetros estimados sob o Cenário 6. A linha tracejada vermelha representa o verdadeiro valor do parâmetro	40
14	Histogramas dos dados originais	43
15	Comparativos entre GEV e BGEV	45
16	Análise de resíduos para os modelos para temperatura do ponto orvalho	48
17	Análise de resíduos para os modelos para umidade relativa do ar.	53

Lista de Tabelas

1	Viés, RMSE e ASE sob os Cenários 1, 3 e 5	33
2	Viés, RMSE e ASE sob os Cenários 2, 4 e 6	34
3	Estatísticas descritivas de TO e U para os dados originais	43
4	Estatísticas descritivas de TO e U com blocos mínimos N=60	44
5	Estimativas e erros padrão dos parâmetros da distribuição BGEV e GEV com EMV	45
6	Medidas de comparação entre modelos BGEV para temperatura do ponto de orvalho	47
7	Estimativas, erros padrão, estatística z e p -valores para os modelos RBGEV aplicados em temperatura do ponto de orvalho.	50
8	Medidas de comparação entre modelos BGEV para umidade relativa do ar	53
9	Estimativas, erros padrão, estatística z e p -valores para os modelos RBGEV aplicados em umidade relativa do ar.	54

Sumário

1 Introdução	8
2 Revisão de Literatura	11
2.1 Distribuições extremas e distribuição GEV	11
2.2 Distribuição GEV bimodal	14
2.2.1 Quantis	18
3 Algumas propriedades da distribuição BGEV	19
3.1 Momentos	19
3.2 Função geratriz de momentos	23
4 Modelo de regressão BGEV	25
4.1 Reparametrização e regressão	25
4.2 Inferência, diagnósticos e seleção de modelo	27
5 Simulação	31
6 Aplicação	41
6.1 Descrição dos dados	41
6.2 Ajuste dos dados extremos de TO e U	43
6.3 Modelo de regressão temperatura do ponto de orvalho	46
6.4 Modelo de regressão umidade relativa do ar	51
7 Conclusão	55
A Apêndices	59
A.1 Apêndice A	59

1 Introdução

A teoria dos valores extremos (TVE) trata do estudo assintótico das distribuições de estatísticas extremas. Isto é, estuda as distribuições-limite das estatísticas do máximo ou do mínimo normalizado de variáveis aleatórias independentes e identicamente distribuídas. Conforme Fisher e Tippet (1928) e Gnedenko (1943), essas distribuições limite (extremas) podem ser Gumbel, Fréchet, ou Weibull. As principais referências clássicas sobre TVE incluem os livros de Kotz e Nadarajah (2000), Haan e Ferreira (2006) e Resnick (1987). A distribuição de probabilidade contínua, desenvolvida para combinar as três distribuições extremas, foi denominada por Jenkinson (1955) de GEV (*generalized extreme value*). Desde então, a distribuição GEV tem sido amplamente utilizada em diversas áreas, como, por exemplo, em finanças (Embrechts, 1997), confiabilidade (Samorodnitsky, 2014), e ciências da terra e meio ambiente (Raynal Villaseñor, 2021), entre outras.

A família de distribuições GEV acomoda distribuições unimodais com cauda pesada e cauda leve. Diversas generalizações da distribuição GEV foram propostas com a finalidade de capturar maior assimetria e acomodar caudas mais pesadas. Entre as generalizações feitas, pode-se citar a distribuição GEV transmutada proposta por Aryal e Tsokos (2009) e estudada por Otiniano et al. (2019).

Em sistemas climáticos complexos os eventos raros se combinam com eventos extremos (caudas pesadas), assim a distribuição de probabilidade adequada para ajustar esses dados deve ser multimodal. Estudos de caso relativos à velocidade do vento mostram que sua distribuição de probabilidade é bimodal.

Em todos esses trabalhos, a mistura de diversas distribuições de probabilidade foi utilizada para modelar a velocidade do vento. Recentemente, Otiniano et al. (2023) propuseram uma nova generalização da distribuição GEV que admite bimodalidade e a denominaram distribuição BGEV. A distribuição BGEV é uma alternativa aos modelos de mistura, pois é identificável e possui inferência e parâmetros mais simples. Devido às boas propriedades da distribuição BGEV, neste trabalho, utilizamos essa distribuição como base para o desenvolvimento de uma nova classe de modelos de regressão.

A regressão é uma ferramenta essencial para a modelagem estatística, permitindo identificar relações entre variáveis explicativas e variáveis resposta. Ainda são poucas as propostas de modelos de regressão para dados bimodais. Ao utilizar a técnica Marshall-Olkin-odd-loglogistic-G (MOOLL-G), é possível obter uma nova distribuição bimodal. Com base nessa técnica, Cordeiro et al. (2021) propuseram uma nova regressão log-Weibull estendida, fundamentada no modelo MOOLL-Weibull. Outros modelos de regressão baseados em distribuições bimodais foram propostos por Vasconcelos et al. (2021a), Vascon-

celos et al. (2021b), e Vila et al. (2021).

Neste trabalho, propomos um modelo de regressão BGEV, baseado na distribuição BGEV (OTINIANO et al., 2023), útil para a análise de dados de eventos extremos de sistemas complexos, os quais podem ser bimodais. Estudaram-se algumas propriedades da distribuição BGEV, como quantis, momentos e função geratriz de momentos. A performance dos estimadores de máxima verossimilhança dos parâmetros do modelo de regressão BGEV foi testada através de experimentos de simulação Monte Carlo. A aplicabilidade do modelo foi ilustrada por meio de dados climáticos da estação meteorológica de Goiânia (A002), fornecidos pelo Instituto Nacional de Meteorologia (INMET).

O restante do trabalho está organizado da seguinte forma. No Capítulo 2, apresentamos uma revisão de literatura que aborda questões teóricas e técnicas relevantes para a análise subsequente, como a forma da distribuição GEV, a apresentação da GEV bimodal, suas respectivas funções de distribuição acumulada e de probabilidade, bem como seus quantis. A partir do Capítulo 3, estão os novos resultados deste trabalho. No Capítulo 3, mostramos algumas propriedades da distribuição BGEV, como os momentos e a função geradora de momentos. No Capítulo 4, apresentamos o modelo de regressão BGEV, começando pela reparametrização da distribuição GEV bimodal, que acomoda a estrutura de regressão na mediana m . Em seguida, apresentamos a função de regressão e a função de ligação escolhida. Discutimos a inferência dos parâmetros pelo método de máxima verossimilhança e calculamos a função score. Por fim, são apresentadas as expressões relativas aos resíduos quantílicos, que serão utilizados para a avaliação do ajuste do modelo, bem como os critérios de informação (AIC e BIC) para a seleção de modelos. O Capítulo 5 é dedicado a mostrar os resultados das estimativas de máxima verossimilhança, via simulações de Monte Carlo. Foram realizadas 10.000 réplicas de Monte Carlo e avaliados 6 diferentes cenários, calculando o viés, o erro quadrático médio e o erro padrão assintótico para cada parâmetro.

A aplicação do modelo de regressão BGEV foi realizada com dados da estação meteorológica A002 de Goiânia, conforme descrito no Capítulo 6. Primeiramente, são apresentados os dados utilizados e a definição de todas as variáveis climáticas envolvidas. Dada a importância dos períodos de estiagem na região de Brasília e a característica bimodal das variáveis, escolhemos como variáveis resposta a temperatura do ponto de orvalho e a umidade relativa do ar. As demais variáveis foram incluídas no modelo como variáveis explicativas. Para as aplicações, utilizou-se a técnica de blocos máximos e mínimos para extrair subamostras de extremos, com o objetivo de verificar a independência serial dos dados extremos. A primeira aplicação corresponde à análise da temperatura do ponto de orvalho. Com base em critérios de informação e na análise dos resíduos, foi realizada a seleção do melhor modelo entre diversos cenários que diferem nas variáveis explicativas. Um procedimento semelhante foi adotado para a umidade relativa do ar. Concluímos am-

bas as aplicações com a interpretação detalhada dos resultados. Finalmente, no Capítulo 7, apresentamos a conclusão deste trabalho e as considerações finais, assim como as perspectivas para trabalhos futuros.

O desenvolvimento computacional do modelo de regressão BGEV foi realizado utilizando a linguagem R, devido à sua flexibilidade e ampla gama de pacotes estatísticos para modelagem de valores extremos. Todas as etapas da implementação, incluindo a estimação dos parâmetros por máxima verossimilhança, simulações de Monte Carlo e validação do modelo, foram conduzidas nessa plataforma. Os códigos utilizados para a replicação dos resultados estão disponíveis no GitHub, proporcionando transparência e facilitando futuras aplicações e extensões do modelo.

2 Revisão de Literatura

A área da Estatística dedicada ao estudo das distribuições de máximos, mínimos e quantis extremos é denominada Teoria de Valores Extremos (TVE). Neste capítulo, apresentaremos alguns resultados da TVE, como as distribuições extremas e a distribuição generalizada de valores extremos (GEV). Em seguida, abordaremos a distribuição bimodal GEV, que será a base para nosso principal resultado.

2.1 Distribuições extremas e distribuição GEV

A primeira proposta para modelar valores extremos foi apresentada por Fisher e Tippett (1928) por meio do Teorema do Limite para distribuições extremas que foi provado pela primeira vez por Gnedenko (1943). Distribuições de valores extremos são distribuições limite de estatísticas extremas adequadamente normalizadas. Mais precisamente, sejam X_1, \dots, X_n variáveis aleatórias independentes e identicamente distribuídas (iid) com função de distribuição F e sejam

$$X_{(n)} = \max\{X_1, \dots, X_n\} \text{ e } X_{(1)} = \min\{X_1, \dots, X_n\}$$

as estatísticas extremas do máximo e mínimo, respectivamente. Se existirem sequências de constantes de normalização $a_n > 0$, b_n tais que

$$\lim_{n \rightarrow \infty} P\left(\frac{X_{(n)} - b_n}{a_n} \leq x\right) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x), \quad (2.1.1)$$

onde G é uma função de distribuição não degenerada, então G deve ser uma das três distribuições de valores extremos, Weibull, Gumbel, ou Fréchet. Essas distribuições são caracterizadas por sua função de distribuição cumulativa (FDA) padrão, que é definida por:

$$\begin{aligned} \text{Weibull} & : \Psi_\alpha(y) = \begin{cases} \exp(-(-y)^\alpha), & \text{se } y < 0, \alpha > 0, \\ 1, & \text{se } y \geq 0; \end{cases} \\ \text{Gumbel} & : \Upsilon(y) = \exp(-\exp(-y)), \text{ para } y \in \mathbb{R}; \\ \text{Fréchet} & : \Phi_\alpha(y) = \begin{cases} 0, & \text{se } y \leq 0, \\ \exp(-y^{-\alpha}), & \text{se } y > 0, \alpha > 0. \end{cases} \end{aligned}$$

Consequentemente, as respectivas funções densidade de probabilidade (FDP) são dadas por:

$$\begin{aligned} \text{Weibull : } \psi_{\alpha}(y) &= \begin{cases} \alpha \exp(-(-y)^{\alpha})(-y)^{\alpha-1}, & \text{se } y < 0, \alpha > 0; \\ 0, & \text{se } y \geq 0; \end{cases} \\ \text{Gumbel : } v(y) &= \exp(-(\exp(-y) + y)), \text{ para } y \in \mathbb{R}; \\ \text{Fréchet : } \phi_{\alpha}(y) &= \begin{cases} 0, & \text{se } y \leq 0; \\ \alpha \exp(-y^{-\alpha})y^{-(\alpha+1)}, & \text{se } y > 0, \alpha > 0. \end{cases} \end{aligned}$$

Na Figura 1 são apresentadas as curvas das funções densidade de probabilidade das três distribuições extremais. É notável a diferença entre as três curvas e que todas as curvas são assimétricas em relação ao eixo vertical. Considerando que a cauda produzida pela distribuição Gumbel possui um decaimento leve, ou suave, semelhante a distribuição normal. Considerando a distribuição Fréchet, a cauda produzida é pesada, sendo esta mais amplamente utilizada para aplicações financeiras conforme Jondeau, Poon e Rockinger (2007).

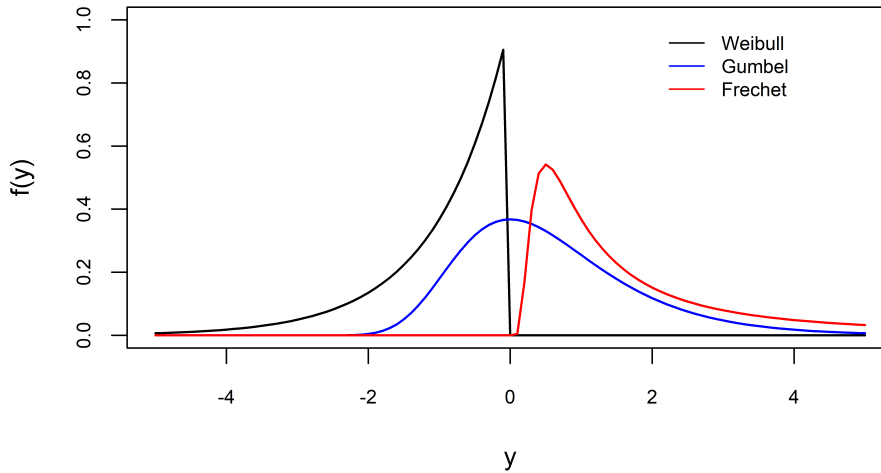


Figura 1: Curvas para a função de densidade de probabilidade para as distribuições de valores extremos (com $\alpha = 1$)

As três distribuições anteriores podem ser acomodadas em uma única família de distribuição mais geral. A família de distribuições generalizada de valores extremos, conhecida como GEV, que foi primeiramente apresentada por Jenkinson (1955), possui

FDP e FDA, respectivamente, definidas por

$$f(y) = \begin{cases} \frac{1}{\sigma} \left[1 + \xi \left(\frac{y - \mu}{\sigma} \right) \right]^{(-1/\xi)-1} \exp \left\{ - \left[1 + \xi \left(\frac{y - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}, & \text{se } \xi \neq 0 \\ \frac{1}{\sigma} \exp \left\{ - \left(\frac{y - \mu}{\sigma} \right) \right\} - \exp \left[- \left(\frac{y - \mu}{\sigma} \right) \right], & \text{se } \xi = 0, \end{cases} \quad (2.1.2)$$

e

$$F(y) = \begin{cases} \exp \left\{ - \left[1 + \xi \left(\frac{y - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}, & \text{se } \xi \neq 0 \\ \exp \left\{ - \exp \left[- \left(\frac{y - \mu}{\sigma} \right) \right] \right\}, & \text{se } \xi = 0, \end{cases} \quad (2.1.3)$$

em que $\mu \in \mathbb{R}$ é o parâmetro de localização, $\sigma > 0$ é o parâmetro de escala, e ξ é o parâmetro de forma. O suporte de F é $\left\{ y : 1 + \xi \left(\frac{y - \mu}{\sigma} \right) > 0 \right\}$ e é toda a reta real (\mathbb{R}) quando $\xi \neq 0$.

O parâmetro ξ é conhecido como índice caudal e possui relação direta com o parâmetro de forma das três distribuições extremas. Para obter a distribuição Gumbel não padronizada partindo da GEV é uma relação direta, uma vez que $\xi = 0$ considerando que a distribuição Gumbel não depende de um parâmetro α . Em contrapartida para a distribuição Gumbel padronizada basta considerar que os parâmetros μ e σ assumem valores 0 e 1 respectivamente. Para se obter a distribuição Fréchet partindo da distribuição GEV, utiliza-se da relação $\xi = \alpha^{-1} > 0$, substituindo na distribuição Fréchet não padronizada de tal maneira que $F(y) = \Phi(\alpha^{-1}y + (\mu + \sigma - \alpha^{-1}\mu))$. A transformação para a distribuição Weibull é feita de maneira similar, porém a relação é $\xi = -\alpha^{-1} > 0$.

Desde que foi formalizada pela primeira vez, a distribuição GEV tem sido aplicada em diversas áreas. Pode-se observar a aplicação teórica da teoria dos valores extremos e também da distribuição GEV em livros como Haan e Ferreira (2006), Kotz e Nadarajah (2000), Rudd et al. (2017) e entre outros. Apesar da flexibilidade da distribuição GEV para modelar eventos extremos, esta distribuição não contempla eventos raros sem serem extremos (dados heterogêneos multimodais). Em populações heterogêneas pode-se observar dentro de uma mesma população dois ou mais grupos que podem ser considerados distintos, sendo assim, é possível que esses grupos formem mais de uma moda, sendo que o caso mais comum é o comportamento bimodal. Dados com populações que apresentem o comportamento bimodal aparecem em diversas áreas como hidrologia, marketing, finanças, entre outros. Nesse contexto, uma generalização da distribuição GEV foi introduzida por Otiniano et al. (2021). Essa generalização não apresenta parâmetro de localização, então Otiniano et al. (2023) redefiniram esse modelo a fim de contemplar o parâmetro de localização.

2.2 Distribuição GEV bimodal

Para definir novos modelos probabilísticos bimodais existem pelo menos duas metodologias utilizando uma distribuição base, isto é, uma distribuição já conhecida que será utilizada de ponto de partida para a obtenção de uma distribuição mais geral. A primeira metodologia é baseada em utilizar uma transformação quadrática em uma distribuição base F . Alguns trabalhos que usaram esta metodologia considerando como base a distribuição normal (Elal-Olivero (2010)), Birnbaum-Saunders (Fonseca e Cribari-Neto (2018); Vila et al. (2021)) e Gamma (Vila et al. (2020)). A segunda abordagem é usar a transformação do tipo $T(x) = (x - \mu)|x - \mu|^\delta$ que foi aplicada sob o modelo log-logístico no trabalho Swamee e Rathie (2007) e também aplicado para a distribuição GEV Otiniano et al. (2021).

O procedimento para obter a distribuição GEV bimodal (BGEV) consiste em compor a distribuição de uma variável aleatória $Y \sim \text{GEV}(\xi, \mu = 0, \sigma)$, que é unimodal com a transformação T . Assim, a função de distribuição acumulada do modelo BGEV é dada por

$$F_{BG_{\xi, \mu, \sigma, \delta}}(y) = F_{\xi, 0, \sigma}(T_{\mu, \delta}(y)), \quad (2.2.1)$$

em que a função $T_{\mu, \delta}(y)$ é definida por:

$$T_{\mu, \delta}(y) = (y - \mu) |y - \mu|^\delta, \quad \delta > -1, \mu \in \mathbb{R}.$$

A função T é invertível e derivável, com a inversa e derivada representadas respectivamente por

$$T_{\mu, \delta}^{-1}(y) = \text{sinal}(y) |y| \frac{1}{(\delta + 1)} + \mu, \quad (2.2.2)$$

em que a $\text{sinal}(y)$ representa a função sinal de y .

$$\frac{T_{\mu, \delta}(y)}{dy} = (\delta + 1) |y - \mu|^\delta. \quad (2.2.3)$$

A partir de (2.2.1), (2.1.3) e (2.2.2) a função de distribuição acumulada da distribuição BGEV é dada por

$$F_{BG_{\xi, \mu, \sigma, \delta}}(y) = \begin{cases} \exp \left[- \left[1 + \xi \left(\frac{T_{\mu, \delta}(y)}{\sigma} \right) \right]^{-1/\xi} \right], & \xi \neq 0 \\ \exp \left[- \exp \left[- \frac{T_{\mu, \delta}(y)}{\sigma} \right] \right], & \xi = 0, \end{cases} \quad (2.2.4)$$

e sua função de densidade de probabilidade correspondente é dada por

$$f_{BG_{\xi,\mu,\sigma,\delta}}(y) = \begin{cases} \frac{1}{\sigma} \left[1 + \xi \left(\frac{T_{\mu,\delta}(y)}{\sigma} \right) \right]^{(-1/\xi)-1} \cdot \exp \left[- \left[1 + \xi \left(\frac{T_{\mu,\delta}(y)}{\sigma} \right) \right]^{-1/\xi} \right] \frac{T_{\mu,\delta}(y)}{dy}, & \xi \neq 0 \\ \frac{1}{\sigma} \exp \left(-\frac{T_{\mu,\delta}(y)}{\sigma} \right) \exp \left[-\exp \left(-\frac{T_{\mu,\delta}(y)}{\sigma} \right) \right] \frac{T_{\mu,\delta}(y)}{dy}, & \xi = 0, \end{cases} \quad (2.2.5)$$

em que $\frac{T_{\mu,\delta}(y)}{dy}$ está definido em (2.2.3). Nota-se que o parâmetro μ é considerado parâmetro de localização. A prova disso tem origem no fato de que $T_{\mu,\delta} = T_{0,\delta}(x - \mu)$ que pode ser verificada de forma direta substituindo os valores em (2.2.2). Logo, em consequência disso $f_{BG_{\xi,\mu,\sigma,\delta}}(x) = f_{BG_{\xi,0,\sigma,\delta}}(x - \mu)$. O suporte da distribuição BGEV é

$$Suporte(f_{BG_{\xi,\mu,\sigma,\delta}}) = \begin{cases} \left[\mu - \left(\frac{\sigma}{\xi} \right)^{\frac{1}{\delta+1}}, +\infty \right), & \text{se } \xi > 0, \\ \left(-\infty, \mu + \left| \frac{\sigma}{\xi} \right|^{\frac{1}{\delta+1}} \right], & \text{se } \xi < 0, \\ (-\infty, +\infty), & \text{se } \xi = 0. \end{cases} \quad (2.2.6)$$

Para melhor entendimento de como parâmetros afetam as possíveis formas da distribuição BGEV foram construídos gráficos variando valores para cada dos parâmetros individualmente enquanto os demais parâmetros permanecem fixos para a FDP e FDA da distribuição BGEV. Nas Figuras 2,3 e 4 estão apresentados os gráficos para a função de densidade e função de distribuição acumulada. Para a construção destes gráficos foi fixado $\xi = 0$, enquanto os demais parâmetros μ, σ e δ variam.

Observa-se na Figura 2 que o valor de μ está de fato, ligado diretamente com a posição da curva de distribuição no eixo das abscissas decorrente do fato de ser o parâmetro de localização. Já considerando a Figura 3 nota-se que existe um maior espaçamento de dados conforme os valores de σ aumentam, ou seja, σ influencia na dispersão dos dados. E por fim, na Figura 4 observa-se que caso $\delta = 0$ a distribuição tem formato unimodal, enquanto que, para valores de $\delta > 0$, tem formato bimodal.

Para o caso de $\xi \neq 0$ é preciso fazer uma análise prévia do impacto dos valores de ξ na forma das funções como está representado na Figura 5, no qual ξ varia enquanto $\mu = 0$, $\sigma = 1$ e $\delta = 1$, pois como dito anteriormente para valores de $\xi \neq 0$ o suporte da distribuição BGEV não compreende toda a reta real. Fica claro ao observar a Figura 5,

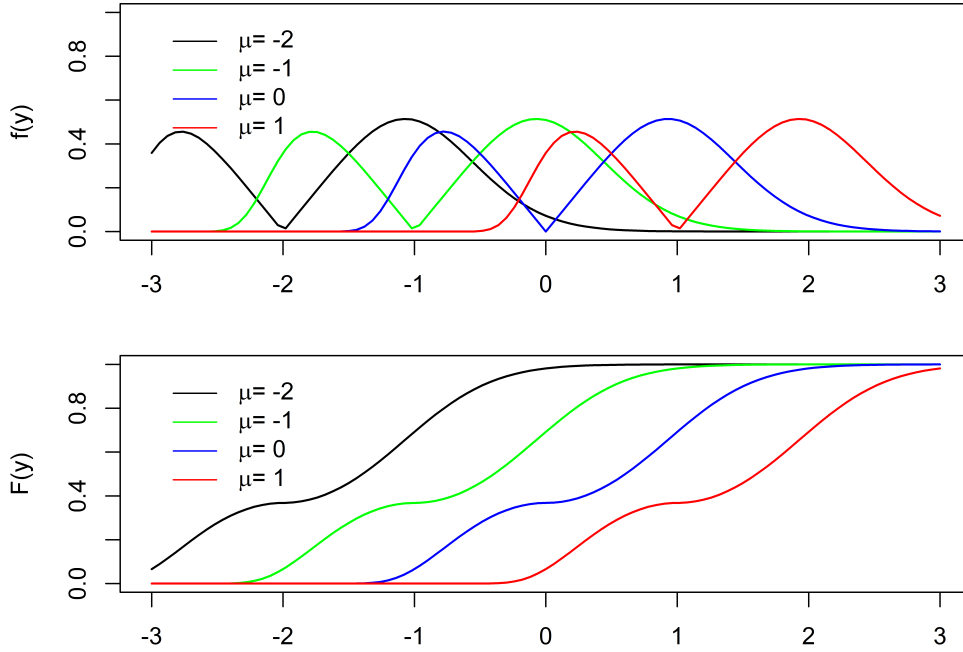


Figura 2: $Y \sim F_{BG_0, \mu, 1, 1}$ com μ variando. Gráfico superior FDP, gráfico inferior FDA.

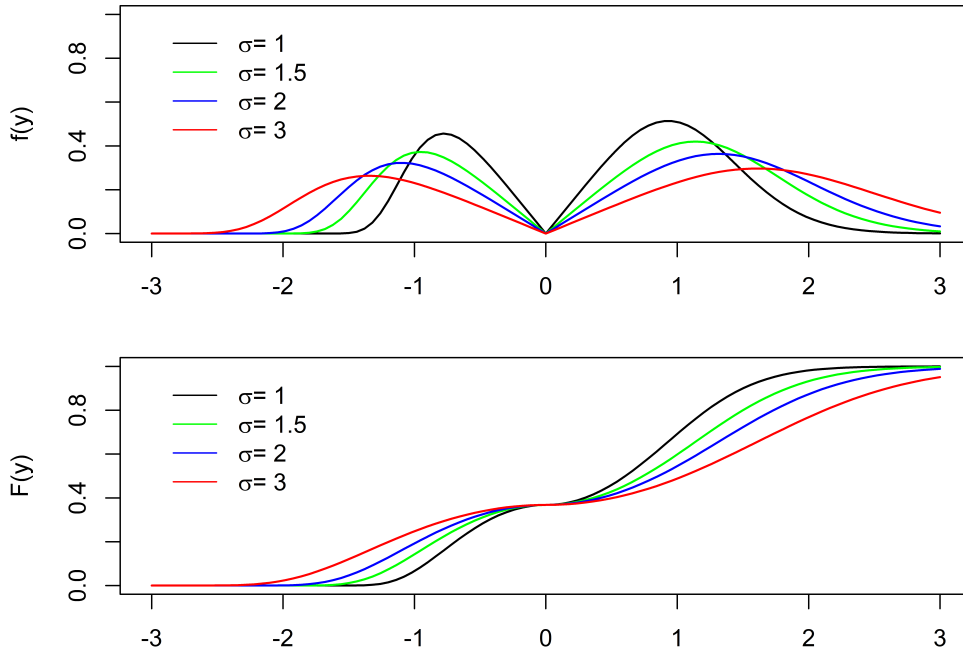


Figura 3: $Y \sim F_{BG_0, 0, \sigma, 1}$ com σ variando. Gráfico superior FDP, gráfico inferior FDA.

que o parâmetro ξ influencia na forma da distribuição BGEV.

Também é de interesse analisar os casos separadamente para $\xi < 0$ e $\xi > 0$, para diferentes valores de δ com o objetivo de observar o comportamento das funções quando os dois parâmetros de forma variam simultaneamente. Como observado nas Figuras 6 e 7 que representam ξ negativo e positivo respectivamente. É interessante observar que para $\xi > 0$ as curvas da distribuição BGEV parecem terminar com uma cauda mais pesada se

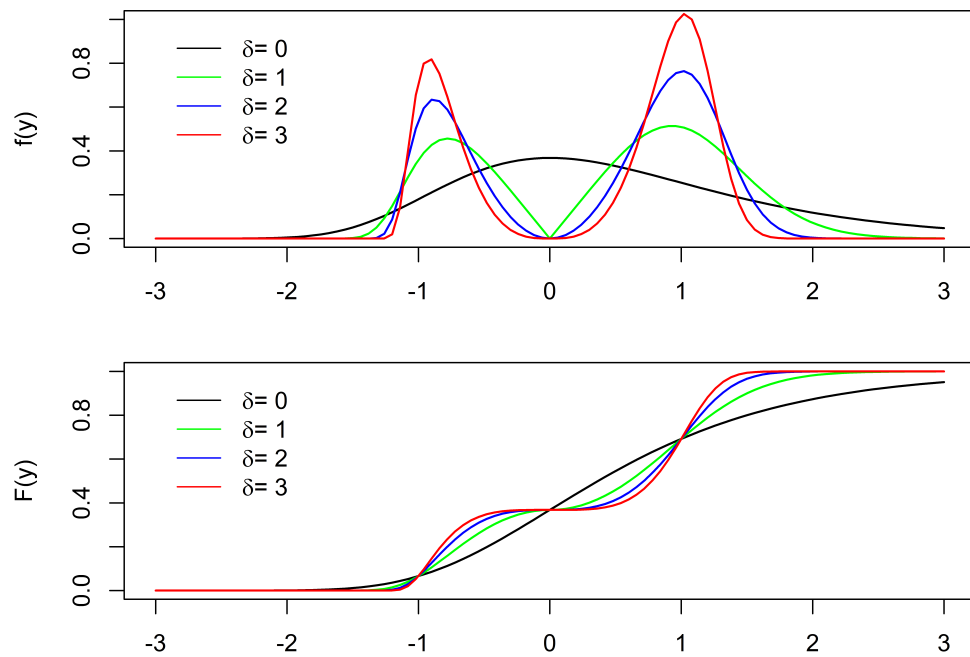


Figura 4: $Y \sim F_{BG_{0,0,1,\delta}}$ com δ variando. Gráfico superior FDP, gráfico inferior FDA.

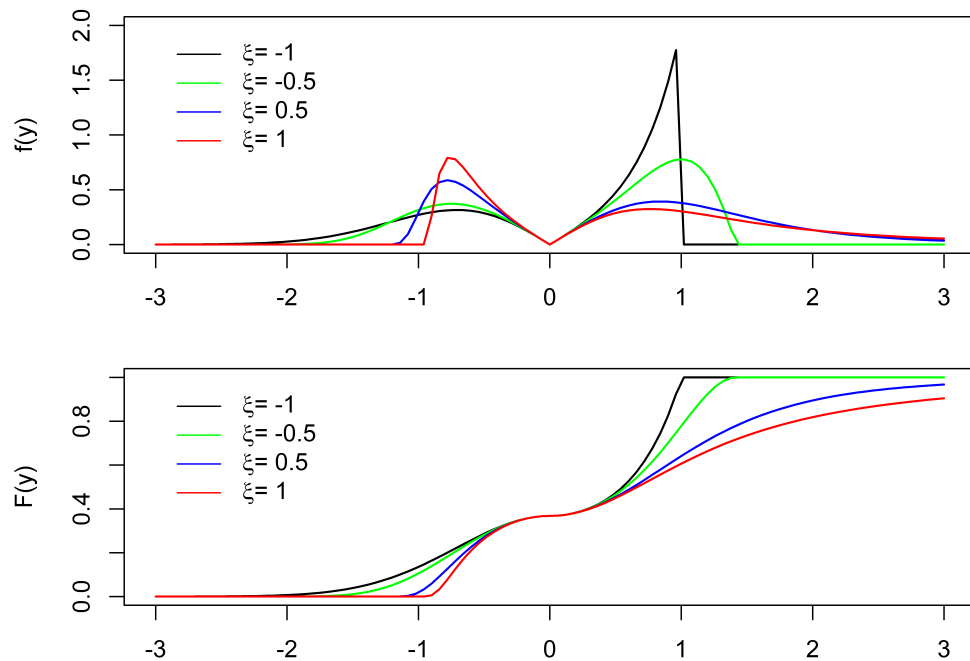


Figura 5: $Y \sim F_{BG_{\xi,0,1,1}}$ com ξ variando. Gráfico superior FDP, gráfico inferior FDA.

em comparação com $\xi < 0$.

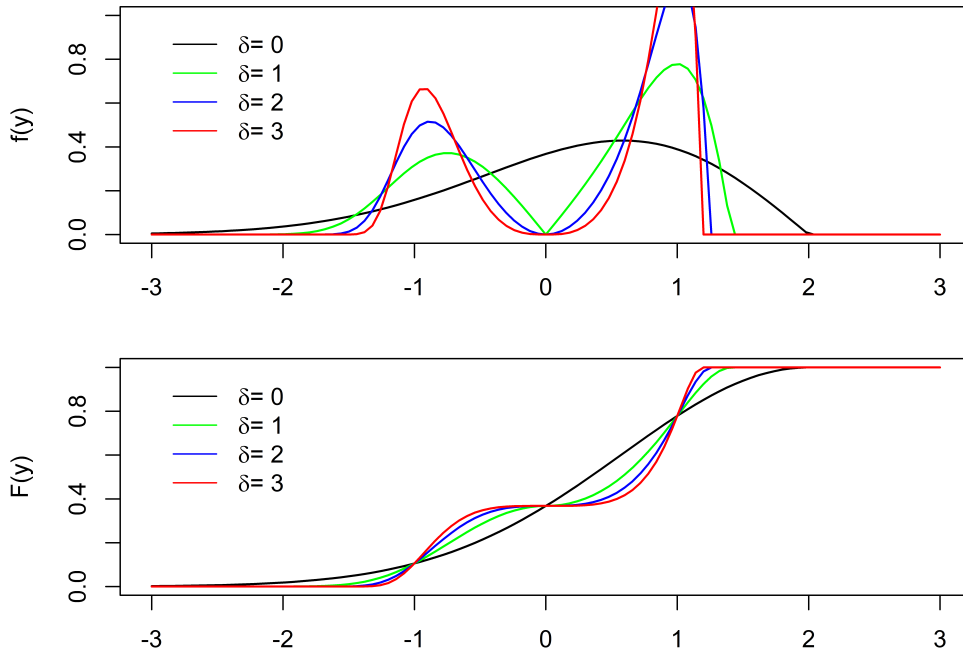


Figura 6: $Y \sim F_{BG_{-0.5,0,1,\delta}}$ e δ variando. Gráfico superior da FDP. Gráfico inferior da FDA.

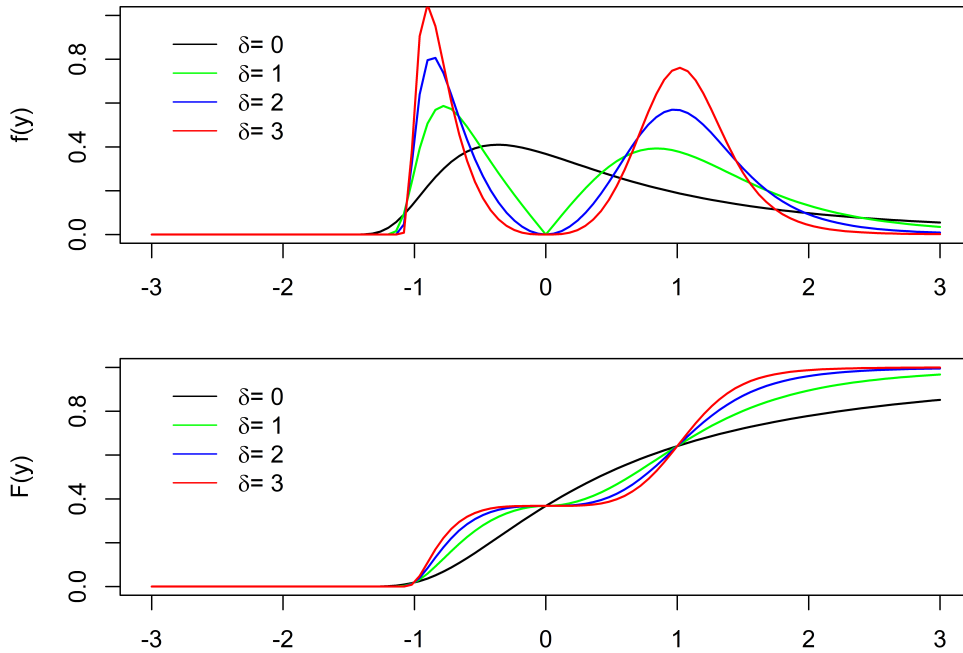


Figura 7: $Y \sim F_{BG_{0.5,0,1,\delta}}$ e δ variando. Gráfico superior da FDP. Gráfico inferior da FDA.

2.2.1 Quantis

Outra função associada a distribuição BGEV que será importante para este trabalho é a função quantil. A função quantil pode ser encontrada através da inversa da função de distribuição acumulada. Então, a função quantílica da distribuição BGEV é

definida por

$$Q_{\xi,\mu,\sigma,\delta}(x) = F_{BG_{\xi,\mu,\sigma,\delta}}^{-1}(y) = T^{-1}(G^{-1}(y)),$$

em que T^{-1} é a função inversa da função de transformação definida em (2.2.2) e $G^{-1}(\cdot)$ é a função inversa da distribuição acumulada da distribuição GEV unimodal. Considerando $\xi = 0$ temos que

$$G^{-1}(y) = -\sigma \log(\log(y^{-1})), \xi = 0. \quad (2.2.7)$$

Dessa forma, considerando (2.2.2) e (2.2.7), pode-se escrever a função quantil para a BGEV para $\xi = 0$ através de:

$$Q_{\xi,\mu,\sigma,\delta}(y) = \text{sinal}(-\sigma \log(\log(y^{-1}))) \left| -\sigma \log(\log(y^{-1})) \right|^{\frac{1}{\delta+1}} + \mu. \quad (2.2.8)$$

De forma análoga considerando $\xi \neq 0$ tem-se que

$$G^{-1}(y) = \frac{\sigma}{\xi} \left[(-\log(y))^{-\xi} - 1 \right]. \quad (2.2.9)$$

Sendo assim, considerando as funções (2.2.9) e (2.2.2), a função quantil no caso de $\xi \neq 0$ pode ser descrita como

$$Q_{\xi,\mu,\sigma,\delta}(y) = \text{sinal} \left(\frac{\sigma}{\xi} [(-\log(y))^{-\xi} - 1] \right) \left| \frac{\sigma}{\xi} [(-\log(y))^{-\xi} - 1] \right|^{\frac{1}{\delta+1}} + \mu. \quad (2.2.10)$$

3 Algumas propriedades da distribuição BGEV

3.1 Momentos

Conforme Abramowitz, Stegun e Miller (1965), para todo número $a \in \mathbb{R}^+$, as funções gama, gama incompleta superior e gama incompleta inferior são definidas, respectivamente, por

$$\Gamma(a) := \int_0^{\infty} t^{a-1} e^{-t} dt \quad (3.1.1)$$

$$\Gamma(a, x) := \int_x^{\infty} t^{a-1} e^{-t} dt \quad (3.1.2)$$

e

$$\gamma(a, x) := \int_0^x t^{a-1} e^{-t} dt. \quad (3.1.3)$$

Estas funções são algumas das ferramentas que serão utilizadas na seguinte proposição. Em todos os casos a é um parâmetro complexo, tal que a parte real de a é positiva.

Proposição 1. Seja $X \sim F_{BG_{\xi,\mu,\sigma,\delta}}$. Então o k -ésimo momento inteiro de X é dado por

$$\begin{aligned} E(X^k) &= \sum_{j=0}^k \binom{k}{j} (-1)^{\frac{(k-j)(\delta+2)}{\delta+1}} \left(\frac{\sigma}{\xi}\right)^{\frac{k-j}{\delta+1}} \\ &\quad \cdot \left[\sum_{i=0}^{\lfloor \frac{k-j}{\delta+1} \rfloor} \binom{\lfloor \frac{k-j}{\delta+1} \rfloor}{i} (-1)^i \gamma(1 - \xi(\lfloor \frac{k-j}{\delta+1} \rfloor - i), 1) \right] \\ &\quad + \sum_{j=0}^k \binom{k}{j} \left(\frac{\sigma}{\xi}\right)^{\frac{k-j}{\delta+1}} \left[\sum_{i=0}^{\lfloor \frac{k-j}{\delta+1} \rfloor} \binom{\lfloor \frac{k-j}{\delta+1} \rfloor}{i} (-1)^i \Gamma(1 - \xi(\lfloor \frac{k-j}{\delta+1} \rfloor - i), 1) \right], \end{aligned} \quad (3.1.4)$$

para $\xi > 0$, sempre que $\xi < \frac{\delta+1}{k}$ e

$$\begin{aligned} E(X^k) &= \sum_{j=0}^k \binom{k}{j} (-1)^{\frac{(k-j)(\delta+2)}{\delta+1}} \left(\frac{\sigma}{\xi}\right)^{\frac{k-j}{\delta+1}} \left[\sum_{i=0}^{\lfloor \frac{k-j}{\delta+1} \rfloor} \binom{\lfloor \frac{k-j}{\delta+1} \rfloor}{i} (-1)^i \Gamma\left(1 - \xi\left(\lfloor \frac{k-j}{\delta+1} \rfloor - i\right), 1\right) \right] \\ &\quad + \sum_{j=0}^k \binom{k}{j} \left(\frac{\sigma}{\xi}\right)^{\frac{k-j}{\delta+1}} \left[\sum_{i=0}^{\lfloor \frac{k-j}{\delta+1} \rfloor} \binom{\lfloor \frac{k-j}{\delta+1} \rfloor}{i} (-1)^i \gamma\left(1 - \xi\left(\lfloor \frac{k-j}{\delta+1} \rfloor - i\right), 1\right) \right], \end{aligned} \quad (3.1.5)$$

para $\xi < 0$, sempre que $\xi < \frac{\delta+1}{k}$.

Prova. Por definição

$$E(X^k) = \int_{-\infty}^{+\infty} x^k f_{GEV_{\xi,0,\sigma}}(T_{\mu,\delta}(x)) \left(\frac{T_{\mu,\delta}(x)}{dx}\right) dx, \quad (3.1.6)$$

sendo $f_{GEV_{\xi,0,\sigma}}$ como definida em (2.1.2), T como em (2.2.2), e $\frac{T}{dx}$ em (2.2.3). Ao substituir $y = T_{\mu,\delta}(x)$ em (3.1.6), os momentos são expressos por

$$E(X^k) = \int_{-\infty}^{+\infty} \left[(\text{sign}(y)) |y|^{\frac{1}{\delta+1}} + \mu \right]^k f_{GEV_{\xi,0,\sigma}}(y) dy. \quad (3.1.7)$$

Como $k \in \mathbb{Z}^+$, utiliza-se a fórmula do Binómio de Newton, então (3.1.7) é atualizada pela integral

$$\begin{aligned}
E(X^k) &= \sum_{j=0}^k \binom{k}{j} \mu^j \left[\int_{-\infty}^{+\infty} [\text{sinal}(y)]^{k-j} |y|^{\frac{k-j}{\delta+1}} f_{GEV_{\xi,0,\sigma}}(y) dy \right] \\
&= \sum_{j=0}^k \binom{k}{j} \mu^j (-1)^{\frac{(k-j)(\delta+2)}{\delta+1}} E\left(Y^{\frac{k-j}{\delta+1}} I_{[Y<0]}\right) + \sum_{j=0}^k \binom{k}{j} \mu^j E\left(Y^{\frac{k-j}{\delta+1}} I_{[Y\geq 0]}\right),
\end{aligned}$$

sendo $Y \sim F_{GEV_{\xi,0,\sigma}}$ e I_A é a função indicadora do conjunto A ; $I_A(\omega) = 1$ se $\omega \in A$ e $I_A(\omega) = 0$ caso contrário. Agora precisam-se analisar os casos $\xi > 0$ e $\xi < 0$.

Caso $\xi > 0$. Ao utilizar a expressão (2.1.2) em (3.1.8) e a substituição $t = \left[1 + \frac{\xi}{\sigma}y\right]^{-\frac{1}{\xi}}$, para $r = \lfloor \frac{k-j}{\delta+1} \rfloor$, segue que

$$\begin{aligned}
E(Y^k I_{[Y\geq 0]}) &= \int_0^{+\infty} y^r \frac{1}{\sigma} \left[1 + \frac{\xi}{\sigma}y\right]^{-\frac{1}{\xi}-1} \exp\left\{-\left[1 + \frac{\xi}{\sigma}y\right]^{-\frac{1}{\xi}}\right\} dy \\
&= \int_1^{+\infty} \left(\frac{\sigma}{\xi}t^{-\xi} - \frac{\sigma}{\xi}\right)^r e^{-t} dt.
\end{aligned} \tag{3.1.8}$$

Em (3.1.8), novamente, utiliza-se o Binômio de Newton e obtêm-se

$$E(Y^k I_{[Y\geq 0]}) = \sum_{i=0}^r \binom{r}{i} (-1)^i \left(\frac{\sigma}{\xi}\right)^r \int_1^{+\infty} t^{-\xi(r-i)} e^{-t} dt. \tag{3.1.9}$$

Da mesma forma é obtido o k -ésimo momento de Y truncado na parte negativa

$$\begin{aligned}
E(Y^k I_{[Y<0]}) &= \int_{-\frac{\sigma}{\xi}}^0 y^r \frac{1}{\sigma} \left[1 + \frac{\xi}{\sigma}y\right]^{-\frac{1}{\xi}-1} \exp\left\{-\left[1 + \frac{\xi}{\sigma}y\right]^{-\frac{1}{\xi}}\right\} dy \\
&= \int_0^1 \left(\frac{\sigma}{\xi}t^{-\xi} - \frac{\sigma}{\xi}\right)^r e^{-t} dt \\
&= \sum_{i=0}^r \binom{r}{i} (-1)^i \left(\frac{\sigma}{\xi}\right)^r \int_0^1 t^{-\xi(r-i)} e^{-t} dt.
\end{aligned}$$

A função gama incompleta inferior (3.1.3) e gama incompleta superior (3.1.2) são utilizadas para representar as integrais de (3.1.9) e (3.1.10), respectivamente. Com isto, a prova de (3.1.4) segue ao substituir essas atualizações na equação (3.1.7). **Caso $\xi < 0$.** Repete-se o mesmo procedimento que para o caso $\xi > 0$ respeitando o suporte da densidade de $Y \sim F_{GEV_{\xi,0,\sigma}}$ que é dada por (2.1.2) para os valores $\{y : y \in (-\infty - \frac{\sigma}{\xi}]\}$.

Corolário 1. Seja $X \sim F_{BG_{\xi,\mu,\sigma,0}} = F_{\xi,\mu,\sigma}$. Então o k -ésimo momento inteiro de X é

dado por:

$$E(X^k) = \sum_{j=0}^k \binom{k}{j} \mu^j \left(\frac{\sigma}{\xi}\right)^{k-j} \Gamma(1 - \xi(k-j)). \quad (3.1.10)$$

Prova. De (3.1.8), quando $\delta = 0$, obtêm-se

$$\begin{aligned} E(X^k) &= \sum_{j=0}^k \binom{k}{j} \mu^j E(Y^{k-j}) \\ &= E\left(\sum_{j=0}^k \binom{k}{j} \mu^j Y^{k-j}\right) \\ &= E(Y + \mu)^k, \end{aligned} \quad (3.1.11)$$

em que $Y + \mu \sim F_{\xi, \mu, \sigma}$, pois $Y \sim F_{xi, 0, \sigma}$. A prova é finalizada com as expressões (3.1.9) e (3.1.10) e o fato que $\Gamma(x, s) + \gamma(x, s) = \Gamma(x)$.

Corolário 2. Seja $X \sim F_{BG_{\xi, \mu, \sigma, \delta}}$. Então, para $\xi > 0$, a esperança de X é dado por:

$$\begin{aligned} E(X) &= (-1)^{\frac{\delta+2}{\delta+1}} \left(\frac{\sigma}{\xi}\right)^{\frac{1}{\delta+1}} \left[\sum_{i=0}^1 \binom{1}{i} (-1)^i \gamma(1 - \xi(1-i), 1) \right] \\ &+ \left(\frac{\sigma}{\xi}\right)^{\frac{1}{\delta+1}} \left[\sum_{i=0}^1 \binom{1}{i} (-1)^i \Gamma(1 - \xi(1-i), 1) \right] \end{aligned} \quad (3.1.12)$$

e para $\xi < 0$ a esperança de X é dado por:

$$\begin{aligned} E(X) &= (-1)^{\frac{\delta+2}{\delta+1}} \left(\frac{\sigma}{\xi}\right)^{\frac{1}{\delta+1}} \left[\sum_{i=0}^1 \binom{1}{i} (-1)^i \Gamma(1 - \xi(1-i), 1) \right] \\ &+ \left(\frac{\sigma}{\xi}\right)^{\frac{1}{\delta+1}} \left[\sum_{i=0}^1 \binom{1}{i} (-1)^i \gamma(1 - \xi(1-i), 1) \right]. \end{aligned} \quad (3.1.13)$$

3.2 Função geratriz de momentos

Proposição 2. Seja $X \sim F_{BG_{\xi,\mu,\sigma,\delta}}$, $\xi = 0$. Então a função geratriz de momentos de X é dado por:

$$M_X(t) = e^{\mu t} \sum_{k=0}^{\infty} \frac{t^k}{k!} \left[(-1)^{\frac{k(\delta+2)}{\delta+1}} E \left(Y^{\frac{k}{\delta+1}} I_{[Y < 0]} \right) + E \left(Y^{\frac{k}{\delta+1}} I_{[Y \geq 0]} \right) \right], \quad (3.2.1)$$

$Y \sim F_{\xi,0,\sigma}$, $\xi = 0$.

Prova. Por definição tem-se que

$$M_X(t) = \int_{-\infty}^{+\infty} \frac{1}{\sigma} \exp\{tx\} \exp \left\{ -\frac{T_{\mu,\delta}(x)}{\sigma} - \exp \left[-\frac{T_{\mu,\delta}(x)}{\sigma} \right] \right\} dx. \quad (3.2.2)$$

Ao usar a substituição $y = T_{\mu,\delta}(x)$ em (3.2.2) e o fato que $x = \text{sgn}(\ln y^{-\sigma}) |\ln y^{-\sigma}|^{\frac{1}{\delta+1}} + \mu$ obtem-se

$$M_X(t) = e^{\mu t} \int_0^{+\infty} \exp \left\{ \text{sgn}(\ln y^{-\sigma}) |\ln y^{-\sigma}|^{\frac{1}{\delta+1}} t \right\} \exp\{-y\} dy. \quad (3.2.3)$$

A nova substituição $s = \ln(y^{-\sigma})$ permite atualizar (3.2.3) por

$$M_X(t) = \frac{e^{\mu t}}{\sigma} \int_{-\infty}^{+\infty} \exp \left\{ \text{sgn}(s) |s|^{\frac{1}{\delta+1}} t \right\} \exp \left\{ -\frac{s}{\sigma} - \exp \left[-\frac{s}{\sigma} \right] \right\} dy. \quad (3.2.4)$$

Para finalizar a prova utiliza-se a representação em série da função exponencial. Assim (3.2.4) é re-escrita pela equação

$$\begin{aligned} M_X(t) &= e^{\mu t} \sum_{k=0}^{\infty} \frac{t^k}{k!} (-1)^{\frac{k(\delta+2)}{\delta+1}} \int_{-\infty}^0 \frac{1}{\sigma} s^{\frac{k}{\delta+1}} \exp \left\{ -\frac{s}{\sigma} - \exp \left[-\frac{s}{\sigma} \right] \right\} ds \\ &+ e^{\mu t} \sum_{k=0}^{\infty} \frac{t^k}{k!} \int_0^{+\infty} \frac{1}{\sigma} s^{\frac{k}{\delta+1}} \exp \left\{ -\frac{s}{\sigma} - \exp \left[-\frac{s}{\sigma} \right] \right\} ds. \end{aligned}$$

Corolário 3. Seja $X \sim F_{BG_{\xi,\mu,\sigma,0}} = F_{\xi,\mu,\sigma}$. Então a função geratriz de momentos de X é dada por:

$$M_X(t) = e^{\mu t} \Gamma(1 - \sigma t). \quad (3.2.5)$$

Prova. Quando $\delta = 0$ a expressão (3.2.1) se reduz a

$$\begin{aligned}
 M_X(t) &= e^{\mu t} \sum_{k=0}^{\infty} \frac{t^k}{k!} E\left(Y^{\frac{k}{\delta+1}}\right) \\
 &= e^{\mu t} E\left(\sum_{k=0}^{\infty} \frac{(tY)^k}{k!}\right) \\
 &= e^{\mu t} E\left(e^{tY}\right) \\
 &= e^{\mu t} \Gamma(1 - \sigma t).
 \end{aligned}$$

Corolário 4. Seja $X \sim F_{BG_{\xi, \mu, \sigma, \delta}}$, com $\xi > 0$. Então, a esperança de X é dado por:

$$E(X) = \mu E\left(Y^{\frac{1}{\delta+1}}\right) + (-1)^{\frac{\delta+2}{\delta+1}} E\left(Y^{\frac{1}{\delta+1}} I_{[Y < 0]}\right) + E\left(Y^{\frac{1}{\delta+1}} I_{Y \geq 0}\right), \quad (3.2.6)$$

$Y \sim F_{\xi, 0, \sigma}$, $\xi = 0$.

Prova. A prova é direta ao avaliar a derivada de (3.2.1) em $t = 0$.

4 Modelo de regressão BGEV

Neste capítulo será apresentado o novo modelo de regressão baseado em uma reparametrização da distribuição BGEV.

4.1 Reparametrização e regressão

A modelagem de dados usando modelos de regressão é uma das abordagens mais amplamente utilizadas por pesquisadores em diversas áreas. Esses modelos permitem que os pesquisadores identifiquem covariáveis que influenciam o comportamento de uma variável de resposta. Por exemplo, modelos de regressão podem ser usados para explicar os salários dos executivos de uma empresa com base em fatores como experiência e gênero. No entanto, a maioria dos modelos de regressão é baseada em distribuições unimodais, que não conseguem acomodar dados que apresentam comportamento bimodal. Isso é particularmente comum em fenômenos aleatórios que produzem dados extremos, como extremos de temperatura.

De acordo com os resultados de Otiniano et al. (2023), a distribuição BGEV serve como um modelo natural da teoria dos valores extremos e acomoda efetivamente dados que apresentam duas modas. Nesta seção, apresentamos uma nova classe de modelos de regressão baseada na distribuição BGEV, especificamente projetada para acomodar dados que exibem bimodalidade. Essa nova classe de modelos de regressão será chamada de modelo de regressão BGEV, ou então, RBGEV.

Na formulação da distribuição BGEV; $Y \sim F_{BG}(\cdot; \xi, \mu, \sigma, \delta)$, μ é um parâmetro de localização. No entanto, μ não é a média nem qualquer outra característica direta de Y , e, portanto, não possui uma interpretação direta. Na modelagem de regressão, é mais comum modelar diretamente o parâmetro da média ou da mediana da distribuição. Assim, neste trabalho, introduzimos uma reparametrização baseada na mediana da distribuição BGEV.

Da função quantílica da distribuição BGEV em (2.2.10), obtemos

$$\mu = \begin{cases} q_\tau - \operatorname{sgn} \left(\frac{\sigma}{\xi} [(-\log(\tau))^{-\xi} - 1] \right) \left| \frac{\sigma}{\xi} [(-\log(\tau))^{-\xi} - 1] \right|^{1/(\delta+1)}, & \text{se } \xi \neq 0, \\ q_\tau - \operatorname{sgn} (-\log(-\log(\tau))) |-\sigma \log(-\log(\tau))|^{1/(\delta+1)}, & \text{se } \xi = 0. \end{cases}$$

em que $q_\tau = Q(\tau)$ representa o τ -ésimo quantil da distribuição BGEV. Note que se definirmos $\tau = 0,5$, q_τ corresponde para a mediana (m) de uma distribuição BGEV, que

possui como mediana $m = q_{0,5}$. Fixando $\tau = 0,5$ temos

$$\mu^* = \begin{cases} m - \left[\frac{\sigma}{\xi} (c_1^{-\xi} - 1) \right]^{1/(\delta+1)}, & \xi \neq 0, \\ m - (c_2\sigma)^{1/(\delta+1)}, & \xi = 0, \end{cases} \quad (4.1.1)$$

em que $c_1 = -\log(0.5)$ e $c_2 = -\log(-\log(0.5))$ são constantes.

A ideia proposta aqui é substituir μ na FDA $F_{BG_{\xi,\mu,\sigma,\delta}}(\cdot)$ por μ^* apresentado em (4.1.1), e reparametrizar a distribuição BGEV indexando pelos parâmetros ξ , m , σ , e δ , sendo que m denota a mediana. Essa abordagem resultará em um modelo mais simples e mais interpretável baseado na distribuição BGEV, com a estrutura de regressão atribuída ao parâmetro da mediana m .

A FDA reparametrizada da distribuição BGEV é dada por

$$F_{BG_{\xi,m,\sigma,\delta}}^*(y) = \begin{cases} \exp \left\{ - \left[1 + \xi \left(\frac{T}{\sigma} \right) \right]^{-1/\xi} \right\}, & \xi \neq 0, \\ \exp \left\{ - \exp \left[-\frac{T}{\sigma} \right] \right\}, & \xi = 0, \end{cases} \quad (4.1.2)$$

em que

$$T = (y - \mu^*) |y - \mu^*|^\delta, \quad (4.1.3)$$

cujos suporte é dado por (2.2.6) substituindo μ por μ^* . Denotamos por $Y \sim F_{BG}^*(\cdot; \xi, m, \sigma, \delta)$.

Note que μ^* não é mais um parâmetro que indexa a distribuição BGEV. De agora em diante, μ^* é somente uma função que depende de ξ , m , σ , e δ , estes sim, são parâmetros que indexam a distribuição BGEV. Além disso, nessa reparametrização, T em (4.1.3) depende de todos os parâmetros da distribuição BGEV, o que diferencia da parametrização original apresentada na Subseção 2.2. A FDP da distribuição reparametrizada da BGEV é dada pela expressão (2.2.5) substituindo T pela expressão em (4.1.3).

Para estabelecer a estrutura de regressão, considere o seguinte. Sejam Y_1, \dots, Y_n , n variáveis aleatórias independentes representando n variáveis respostas, em que $Y_i \sim F_{BG}^*(\cdot; \xi, m_i, \sigma, \delta)$. O modelo é obtido assumido que a mediana de Y_i pode ser escrita

como

$$g(m_i) = \eta_i = \sum_{j=1}^p x_{ij}\beta_j = \mathbf{X}_i^\top \boldsymbol{\beta}, \quad (4.1.4)$$

em que $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top \in \mathbb{R}^p$ é um vetor de parâmetros de regressão desconhecidos, $\mathbf{X}_i = (x_{i1}, \dots, x_{ip})^\top$ é um vetor de covariáveis de tamanho p ($p < n$), que são assumidos como conhecidos e fixos. Além disso, $g(\cdot)$ é uma função de ligação estritamente monótona e duas vezes diferenciável. Existem várias escolhas possíveis para a função de ligação $g(\cdot)$. Por exemplo, pode-se usar a função identidade, a função inversa e a função logarítmica definidas como $g(x) = x$, $g(x) = 1/x$, e $g(x) = \log(x)$, respectivamente. No modelo de regressão BGEV, a mediana é o único parâmetro que pode variar com as observações. Portanto, os parâmetros ξ , σ , e δ são constantes desconhecidas uma vez que não atribuímos estruturas de regressão a estes.

4.2 Inferência, diagnósticos e seleção de modelo

O vetor de parâmetros desconhecidos para o modelo de regressão BGEV é representado como $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \xi, \sigma, \delta)^\top \in \mathbb{R}^{p+3}$ quando $\xi \neq 0$ e como $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma, \delta)^\top \in \mathbb{R}^{p+2}$ quando $\xi = 0$. Portanto, $\boldsymbol{\theta}$ deve ser estimado. Seja k o número de parâmetros no modelo. O procedimento de inferência adotado aqui será baseado no estimador de máxima verossimilhança. O logaritmo da função de verossimilhança baseada em uma amostra de n observações independentes é dada por

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}), \quad (4.2.1)$$

em que $\ell_i(\boldsymbol{\theta}) = \log(f_{BG_{\xi, m_i, \sigma, \delta}}^*(y_i))$ com $f_{BG_{\xi, m_i, \sigma, \delta}}^*(\cdot)$ denotando a FDP da distribuição BGEV reparametrizada. O estimador de máxima verossimilhança (EMV) é obtido maximizando a função de log-verossimilhança em (4.2.1) com respeito a $\boldsymbol{\theta}$. O EMV para $\boldsymbol{\theta}$ será denotado por $\hat{\boldsymbol{\theta}}$.

A contribuição da i -ésima observação no logaritmo da função de verossimilhança é dada por

$$\ell_i(\boldsymbol{\theta}) = \begin{cases} -\log(\sigma) + \log(\delta + 1) + \delta \log |y_i - \mu_i^*| - \left(1 + \frac{1}{\xi}\right) \log(\Psi_i(\boldsymbol{\theta})) - \Psi_i^{-1/\xi}(\boldsymbol{\theta}), & \xi \neq 0, \\ -\log(\sigma) + \log(\delta + 1) + \delta \log |y_i - \mu_i^*| - \frac{T_i}{\sigma} - e^{-T_i/\sigma}, & \xi = 0, \end{cases} \quad (4.2.2)$$

em que

$$\Psi_i(\boldsymbol{\theta}) = 1 + \frac{\xi T_i}{\sigma}, \quad T_i = (y_i - \mu_i^*) |y_i - \mu_i^*|^\delta.$$

A função escore, obtida pela diferenciação do logaritmo da função de verossimilhança apresentada em (4.2.1) com respeito a $\boldsymbol{\theta}$ (observe o apêndice A), é dada por

$$\mathbf{U}(\boldsymbol{\theta}) = \begin{cases} (\mathbf{U}_\beta^\top, U_\xi, U_\sigma, U_\delta)^\top, & \text{if } \xi \neq 0, \\ (\mathbf{U}_\beta^\top, U_\sigma, U_\delta)^\top, & \text{if } \xi = 0, \end{cases}$$

com

$$\mathbf{U}_\beta = X^\top H \mathbf{a},$$

$$U_\xi = \mathbf{b}^\top \mathbf{1}_n,$$

$$U_\sigma = \mathbf{d}^\top \mathbf{1}_n,$$

$$U_\delta = \mathbf{e}^\top \mathbf{1}_n,$$

em que X é uma matriz $n \times p$ a qual i -ésima linha é \mathbf{X}_i^\top , $H = \text{diag}\{h_1, \dots, h_n\}$ é uma matriz diagonal $n \times n$ com $h_i = 1/g'(m_i)$, e $\mathbf{a} = (a_1, \dots, a_n)^\top$, $\mathbf{b} = (b_1, \dots, b_n)^\top$, $\mathbf{d} = (d_1, \dots, d_n)^\top$, $\mathbf{e} = (e_1, \dots, e_n)^\top$, $\mathbf{1}_n = (1, \dots, 1)^\top$ são n vetores dimensionais com

$$\begin{aligned} a_i &= \frac{1}{\sigma(y_i - \mu_i^*)} [(\delta + 1)T_i\Omega_i(\boldsymbol{\theta}) - \delta\sigma]; \\ b_i &= \frac{a_i}{\xi(\delta + 1)}(m_i - \mu_i^*) \left(1 + \frac{\xi c_2 c_1^{-\xi}}{1 - c_1^{-\xi}} \right) + \frac{1}{\xi^2} \log(\Psi_i(\boldsymbol{\theta})) [1 - \Psi_i^{-1/\xi}(\boldsymbol{\theta})] - \frac{T_i\Omega_i(\boldsymbol{\theta})}{\xi\sigma}; \\ d_i &= -\frac{1}{\sigma} + \frac{\delta}{\sigma(\delta + 1)} \frac{(m_i - \mu_i^*)}{(y_i - \mu_i^*)} + \frac{1}{\sigma^2} \Omega_i(\boldsymbol{\theta}) (y_i - m_i) |y_i - \mu_i^*|^\delta; \\ e_i &= \log |y_i - \mu_i^*| \left(1 - \frac{T_i\Omega_i(\boldsymbol{\theta})}{\sigma} \right) + \frac{1}{\delta + 1} [1 + a_i(m_i - \mu_i^*) \log(m_i - \mu_i^*)]; \\ \Omega_i(\boldsymbol{\theta}) &= \begin{cases} \Psi_i^{-1}(\boldsymbol{\theta}) [1 + \xi - \Psi_i^{-1/\xi}(\boldsymbol{\theta})], & \text{se } \xi \neq 0, \\ 1 - e^{-T_i/\sigma}, & \text{se } \xi = 0. \end{cases} \end{aligned}$$

O EMV de $\boldsymbol{\theta}$ é obtido resolvendo o sistema $\mathbf{U}(\boldsymbol{\theta}) = \mathbf{0}$, em que $\mathbf{0}$ representa o vetor nulo em \mathbb{R}^k . No entanto, esse sistema não pode ser resolvido analiticamente, e métodos numéricos iterativos devem ser empregados para obter uma solução aproximada. Neste trabalho, usamos o método de Broyden–Fletcher–Goldfarb–Shanno (BFGS) (NOCEDAL; WRIGHT, 1999).

Sob as condições de regularidade habituais para o estimador de máxima verossimilhança, quando o tamanho da amostra é grande,

$$\hat{\boldsymbol{\theta}} \stackrel{a}{\sim} N_k(\boldsymbol{\theta}, J^{-1}(\boldsymbol{\theta})),$$

em que $\stackrel{a}{\sim}$ denota distribuição aproximada, $N_k(\boldsymbol{\theta}, J^{-1}(\boldsymbol{\theta}))$ denota a distribuição normal k -variada com média $\boldsymbol{\theta}$ e matriz de variância-covariância $J^{-1}(\boldsymbol{\theta})$. A matriz $J^{-1}(\boldsymbol{\theta})$ é a inversa da matriz de informação de Fisher observada. Neste trabalho, a matriz de informação observada é obtida numericamente.

Considere o interesse em testar $H_0: \theta_j = \theta_j^0$ contra $H_1: \theta_j \neq \theta_j^0$, onde θ_j^0 é um valor específico para o parâmetro desconhecido θ_j , $j = 1, \dots, k$. Sob H_0 ,

$$z_j = \frac{\hat{\theta}_j - \theta_j^0}{\text{se}(\hat{\theta}_j)} \stackrel{a}{\sim} N(0, 1),$$

em que $\hat{\theta}_j$ é o EMV de θ_j , e $\text{se}(\hat{\theta}_j) = \sqrt{J_{jj}}$ é o erro padrão assintótico de $\hat{\theta}_j$, com J_{jj} sendo o j -ésimo elemento da diagonal de $J^{-1}(\boldsymbol{\theta})$.

A estatística do teste z_j é a raiz quadrada da estatística de Wald (Wald, 1943), que é amplamente utilizada em aplicações práticas com modelos de regressão. No Capítulo 6, este teste será empregado para avaliar a significância dos parâmetros no modelo de regressão BGEV, com $\theta_j^0 = 0$.

Para avaliar a adequação do ajuste sob modelos de regressão BGEV, os resíduos introduzidos por Dunn e Smyth (1996) podem ser empregados. O resíduo quantílico r_i , $i = 1, \dots, n$, é dado por

$$r_i = \begin{cases} \Phi^{-1}(F_{BG_{\hat{\xi}, \hat{m}_i, \hat{\sigma}, \hat{\delta}}}^*(y_i)), & \xi \neq 0, \\ \Phi^{-1}(F_{BG_{0, \hat{m}_i, \hat{\sigma}, \hat{\delta}}}^*(y_i)), & \xi = 0, \end{cases}$$

em que $\Phi^{-1}(\cdot)$ é a FDP da distribuição normal padrão. Se o modelo de regressão BGEV estiver bem ajustado, r_i é aproximadamente distribuído como uma distribuição normal padrão. Para avaliar a distribuição empírica dos resíduos, uma prática comum é utilizar gráficos de probabilidade normal dos resíduos com envelope simulado (Pereira (2019); Espinheira e Silva (2020); Tsuyuguchi et al., 2020). Esta é a abordagem que adotaremos no Capítulo 6 para avaliar a adequação do ajuste dos modelos de regressão BGEV.

Para a seleção do modelo, utilizaremos o Critério de Informação de Akaike (AIC) (Akaike (1974)) e o Critério de Informação Bayesiano (BIC) (Schwarz (1978)), definidos respectivamente por

$$\text{AIC} = -2\ell(\hat{\boldsymbol{\theta}}) + 2k, \quad \text{BIC} = -2\ell(\hat{\boldsymbol{\theta}}) + k \log(n).$$

Dado um conjunto de modelos candidatos para os dados, o modelo de regressão BGEV preferido é aquele com os menores valores de AIC e BIC.

5 Simulação

Para avaliar o desempenho das estimativas de máxima verossimilhança do modelo de regressão BGEV, proposto no capítulo anterior, foram realizados estudos de simulação de Monte Carlo para vários cenários.

Os tamanhos de amostra considerados são $n = 50, 100, 500$. Para todos os cenários, consideramos a função de ligação identidade na estrutura de regressão para a mediana em (4.1.4). Os modelos incluem interceptos, ou seja, $x_{i1} = 1$, para todos $i = 1, 2, \dots, n$, e as outras covariáveis são geradas aleatoriamente de uma distribuição uniforme padrão e de uma distribuição Bernoulli com probabilidade de sucesso 0.4. As covariáveis foram mantidas constantes ao longo das amostras simuladas.

Todos os procedimentos computacionais foram realizados no ambiente do R Project, versão 4.33 (R Core Team, 2022).

Com 10000 repetições de Monte Carlo, foram obtidas as estimativas médias de θ_t , $\hat{\theta}_t$, para $t = 1, \dots, k, \dots, 10000$ bem como o viés $b(\hat{\theta}_t)$ e a raiz quadrada do erro quadrático médio $\text{RMSE}(\hat{\theta}_t)$, definidos, respectivamente, por

$$\begin{aligned} b(\hat{\theta}_t) &= \frac{1}{10000} \sum_{j=1}^{10000} (\hat{\theta}_t^j - \theta_t), \\ \text{RMSE}(\hat{\theta}_t) &= \sqrt{\frac{1}{10000} \sum_{j=1}^{10000} (\hat{\theta}_t^j - \theta_t)^2}, \end{aligned}$$

em que $\hat{\theta}_t^j$ é o EMV do t -ésimo elemento do vetor de parâmetros $\boldsymbol{\theta}$ na j -ésima repetição de Monte Carlo, com $j \in 1, \dots, 10000$. Além disso, também obtemos a média do erro padrão assintótico (ASE) para cada θ_t , definida por

$$\text{ASE}(\hat{\theta}_t) = \frac{1}{10000} \sum_{j=1}^{10000} \text{se}(\hat{\theta}_t^j),$$

em que $\text{se}(\hat{\theta}_t^j)$ é o erro padrão assintótico de t -ésimo elemento de $\boldsymbol{\theta}$ na j -ésima replica.

Diferentes configurações de parâmetros foram consideradas como descritas abaixo.

Cenário 1: regressão BGEV com ξ igual a zero e $m \approx 2.8$

Os valores dos parâmetros foram definidos como $\beta_1 = 1.2, \beta_2 = 2.0, \beta_3 = 1.5$, que produzem $m_i \in (1.20, 4.68)$ com $\text{mediana}(m) \approx 2.8$, $\xi = 0$, $\sigma = 1.0$, e $\delta = 0.5$.

Cenário 2: Regressão BGEV com ξ igual a zero e $m \approx 5.6$

Os valores dos parâmetros foram definidos como $\beta_1 = 4.0, \beta_2 = 2.0, \beta_3 = 3.0$, que produzem $m_i \in (4.00, 8.99)$ com $\text{mediana}(m) \approx 5.6$, $\xi = 0$, $\sigma = 1.0$, e $\delta = 1.0$.

Cenário 3: Regressão BGEV com ξ positivo e $m \approx 2.3$

Os valores dos parâmetros foram definidos como $\beta_1 = 1.2, \beta_2 = 2.0, \beta_3 = 1.5$, que produzem $m_i \in (-4.65, 7.77)$ com $\text{mediana}(m) \approx 2.3$, $\xi = 0.25$, $\sigma = 1.0$, e $\delta = 0.5$.

Cenário 4: Regressão BGEV com ξ negativo e $m \approx 5.6$

Os valores dos parâmetros foram definidos como $\beta_1 = 4.0, \beta_2 = 2.0, \beta_3 = 3.0$, que produzem $m_i \in (4.00, 8.99)$ com $\text{mediana}(m) \approx 5.6$, $\xi = -0.25$, $\sigma = 1.0$, e $\delta = 1.0$.

Cenário 5: regressão BGEV com ξ negativo e $m \approx 1.67$

Os valores dos parâmetros foram definidos como $\beta_1 = 1.2, \beta_2 = 2.0, \beta_3 = 1.5$, que produzem $m_i \in (-4.025, 6.957)$ com $\text{mediana}(m) \approx 1.67$, $\xi = -0.25$, $\sigma = 1.0$, e $\delta = 0.5$.

Cenário 6: Regressão BGEV com ξ positivo e $m \approx 5.6$

Os valores dos parâmetros foram definidos como $\beta_1 = 4.0, \beta_2 = 2.0, \beta_3 = 3.0$, que produzem $m_i \in (4.00, 8.99)$ com $\text{mediana}(m) \approx 5.6$, $\xi = 0.25$, $\sigma = 1.0$, e $\delta = 1.0$.

De modo análogo, os cenários 2, 4 e 6 também são semelhantes entre si, variando apenas os valores de ξ , enquanto mantêm os mesmos valores para os demais parâmetros.

Tabela 1: Viés, RMSE e ASE sob os Cenários 1, 3 e 5

Cenário 1									
parâmetros	$n = 50$			$n = 100$			$n = 500$		
	Viés	RMSE	ASE	Viés	RMSE	ASE	Viés	RMSE	ASE
β_1	-0.048	0.273	0.136	-0.024	0.192	0.090	-0.029	0.098	0.035
β_2	0.047	0.376	0.184	0.055	0.283	0.126	0.044	0.147	0.046
β_3	0.038	0.236	0.119	0.036	0.183	0.081	0.018	0.086	0.028
σ	-0.025	0.137	0.117	-0.005	0.092	0.082	0.000	0.040	0.036
δ	0.047	0.219	0.177	0.015	0.143	0.122	-0.004	0.060	0.053

Cenário 3									
parâmetros	$n = 50$			$n = 100$			$n = 500$		
	Viés	RMSE	ASE	Viés	RMSE	ASE	Viés	RMSE	ASE
β_1	-0.066	0.239	0.095	-0.044	0.185	0.066	0.013	0.079	0.027
β_2	0.003	0.121	0.049	0.006	0.108	0.041	0.001	0.042	0.013
β_3	0.018	0.274	0.122	0.034	0.185	0.074	0.022	0.089	0.027
ξ	-0.086	0.236	0.151	-0.033	0.130	0.087	-0.001	0.043	0.030
σ	-0.032	0.170	0.123	-0.014	0.103	0.085	-0.001	0.039	0.036
δ	0.117	0.311	0.234	0.028	0.182	0.148	-0.003	0.073	0.061

Cenário 5									
parâmetros	$n = 50$			$n = 100$			$n = 500$		
	Viés	RMSE	ASE	Viés	RMSE	ASE	Viés	RMSE	ASE
β_1	-0.035	0.196	0.104	-0.024	0.150	0.071	0.013	0.079	0.027
β_2	0.003	0.107	0.049	0.002	0.076	0.037	0.001	0.042	0.013
β_3	0.036	0.247	0.119	0.036	0.160	0.076	0.022	0.089	0.027
ξ	-0.049	0.220	0.168	-0.033	0.130	0.087	-0.001	0.043	0.030
σ	-0.038	0.159	0.121	-0.014	0.103	0.085	-0.001	0.039	0.036
δ	0.135	0.330	0.229	0.028	0.182	0.148	-0.003	0.073	0.061

Os resultados das simulações para os Cenários 1,3 e 5 estão apresentados na Tabela 1 e pode-se observar que, tal qual o esperado do viés diminui conforme o tamanho de amostra n aumenta. Também é possível notar que o erro quadrático médio diminui para todos os parâmetros conforme o tamanho de amostra aumenta e que o mesmo ocorre para o erro padrão assintótico. Esses resultados indicam a consistência e a eficiência assintótica dos estimadores de máxima verossimilhança para todos os cenários considerados. Ocorre porém um efeito incomum quando se observa o viés para o parâmetro β_2 que em 2 cenários o viés aumenta quando o tamanho de amostra aumenta de $n = 50$ para $n = 100$.

Tabela 2: Viés, RMSE e ASE sob os Cenários 2, 4 e 6

Cenário 2									
<i>parâmetros</i>	<i>n</i> = 50			<i>n</i> = 100			<i>n</i> = 500		
	Viés	RMSE	ASE	Viés	RMSE	ASE	Viés	RMSE	ASE
β_1	-0.035	0.225	0.116	-0.019	0.147	0.083	-0.007	0.060	0.036
β_2	0.033	0.288	0.151	0.032	0.211	0.115	0.010	0.082	0.048
β_3	0.022	0.176	0.098	0.015	0.124	0.070	0.003	0.050	0.029
σ	-0.021	0.136	0.118	-0.007	0.089	0.083	-0.001	0.038	0.036
δ	0.048	0.315	0.235	0.023	0.198	0.163	0.003	0.078	0.071

Cenário 4									
<i>parâmetros</i>	<i>n</i> = 50			<i>n</i> = 100			<i>n</i> = 500		
	Viés	RMSE	ASE	Viés	RMSE	ASE	Viés	RMSE	ASE
β_1	-0.060	0.241	0.119	-0.050	0.165	0.084	-0.036	0.070	0.037
β_2	0.025	0.309	0.144	0.030	0.215	0.113	0.013	0.079	0.047
β_3	0.017	0.188	0.094	0.013	0.120	0.068	0.004	0.045	0.027
ξ	-0.053	0.264	0.185	-0.021	0.140	0.112	-0.003	0.052	0.044
σ	-0.039	0.161	0.126	-0.016	0.104	0.086	-0.002	0.043	0.038
δ	0.115	0.452	0.299	0.048	0.256	0.195	0.008	0.090	0.082

Cenário 6									
<i>parâmetros</i>	<i>n</i> = 50			<i>n</i> = 100			<i>n</i> = 500		
	Viés	RMSE	ASE	Viés	RMSE	ASE	Viés	RMSE	ASE
β_1	-0.049	0.261	0.111	-0.022	0.183	0.078	0.013	0.073	0.033
β_2	0.023	0.303	0.1423	0.026	0.225	0.111	0.019	0.091	0.045
β_3	0.011	0.199	0.093	0.014	0.135	0.064	0.006	0.055	0.027
ξ	-0.040	0.264	0.158	-0.013	0.124	0.090	0.001	0.042	0.032
σ	-0.028	0.155	0.126	-0.010	0.096	0.085	-0.001	0.037	0.036
δ	0.060	0.420	0.294	0.017	0.264	0.193	-0.001	0.098	0.082

Nota-se, na Tabela 2, que o viés, o erro quadrático médio e o erro padrão assintótico diminuem à medida que o tamanho da amostra aumenta. Esses resultados indicam a eficiência e consistência dos estimadores de máxima verossimilhança nos casos apresentados. Além disso, não foi possível identificar nenhum padrão para o viés nos casos em que $\xi = 0$, $\xi > 0$ e $\xi < 0$.

As Figuras 8, 9, 10, 11, 12 e 13 ilustram a convergência do estimador de máxima verossimilhança para os cenários 1, 2, 3, 4, 5 e 6 respectivamente. Observa-se que as estimativas dos parâmetros, obtidas por meio do EMV para cada cenário, estão concentradas em torno dos valores verdadeiros dos parâmetros. Adicionalmente, verifica-se que a variabilidade das estimativas diminui à medida que o tamanho da amostra aumenta.

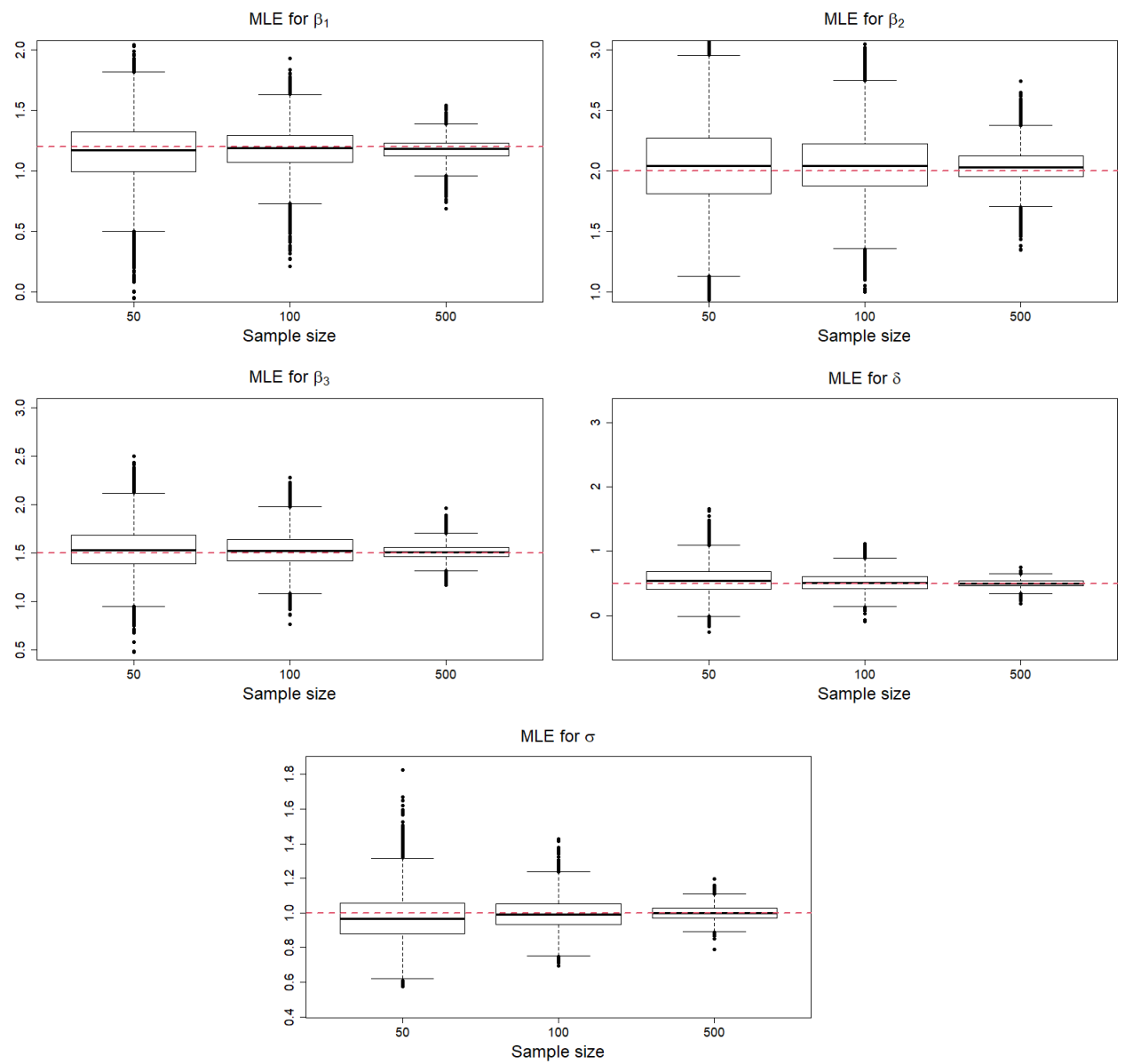


Figura 8: Boxplots dos parâmetros estimados sob o Cenário 1. A linha tracejada vermelha representa o verdadeiro valor do parâmetro

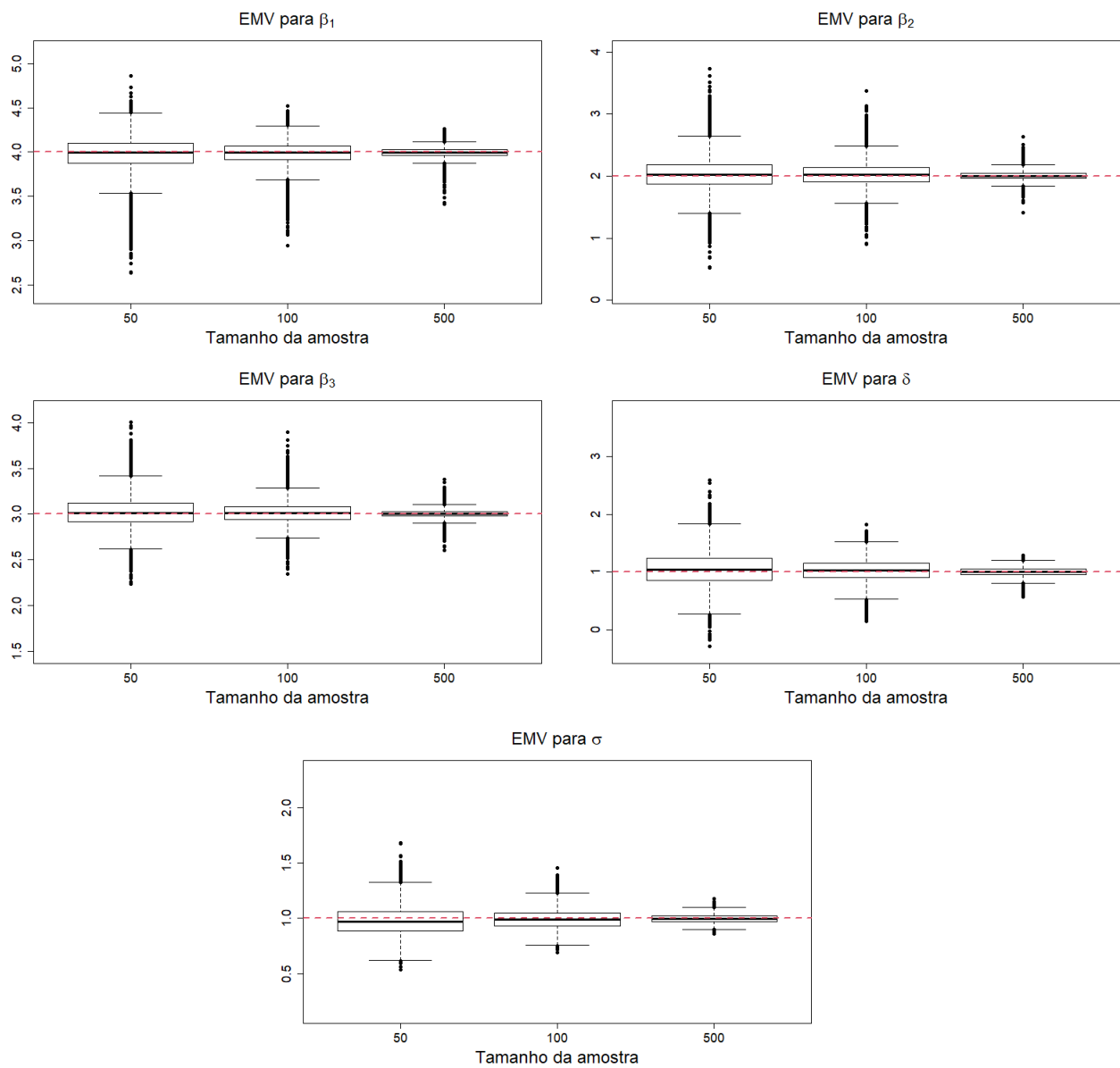


Figura 9: Boxplots dos parâmetros estimados sob o Cenário 2. A linha tracejada vermelha representa o verdadeiro valor do parâmetro

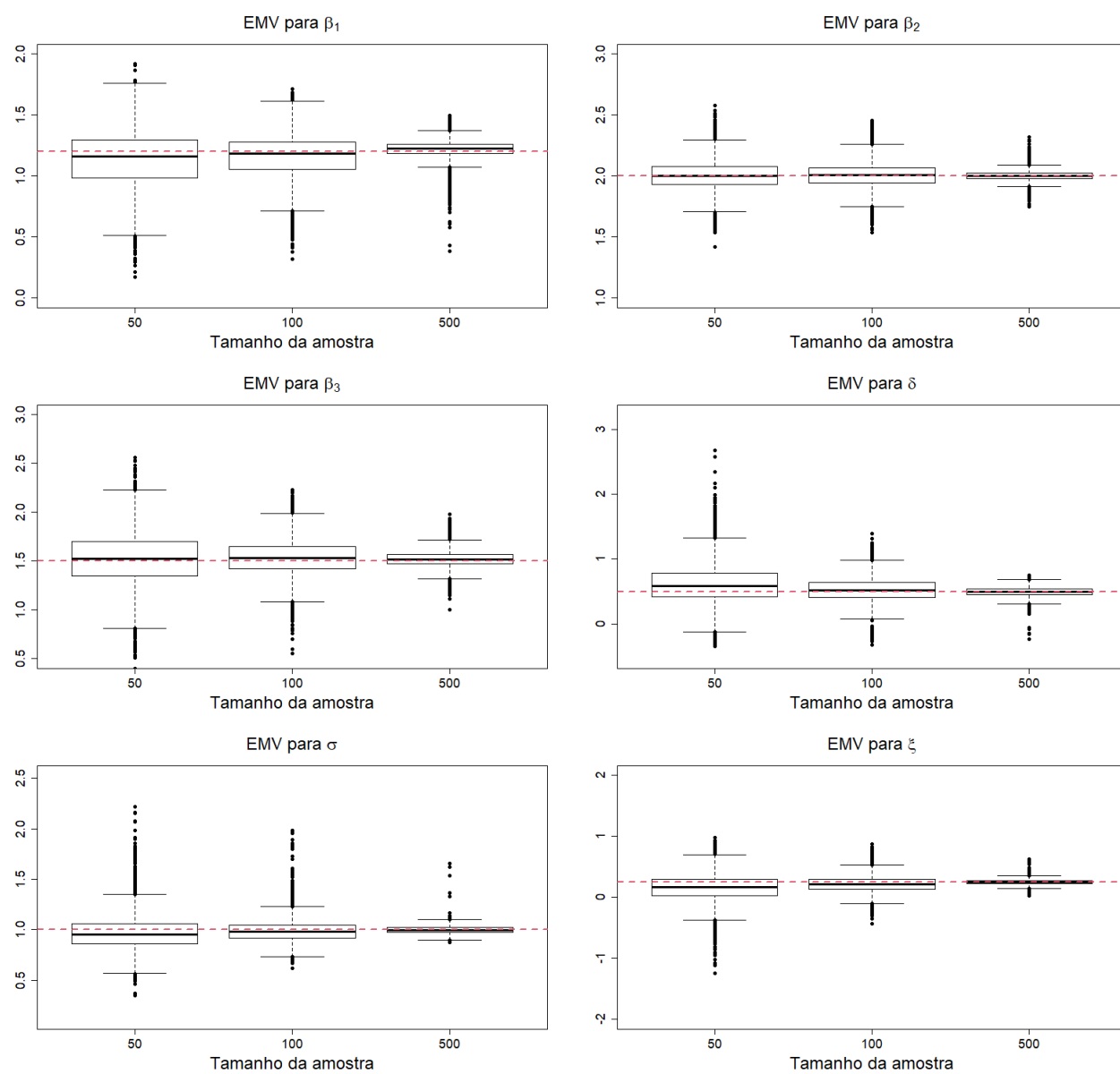


Figura 10: Boxplots dos parâmetros estimados sob o Cenário 3. A linha tracejada vermelha representa o verdadeiro valor do parâmetro

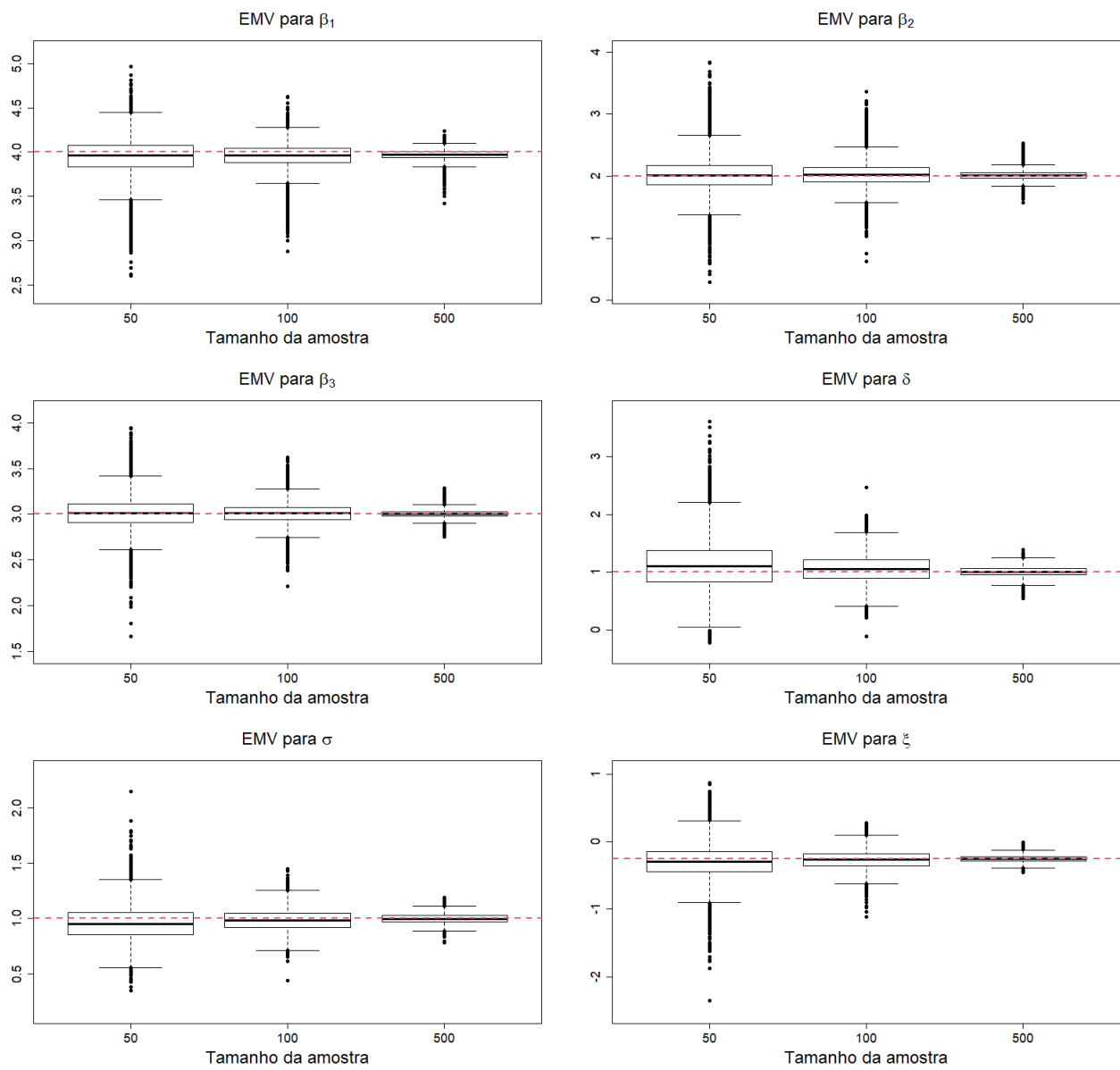


Figura 11: Boxplots dos parâmetros estimados sob o Cenário 4. A linha tracejada vermelha representa o verdadeiro valor do parâmetro

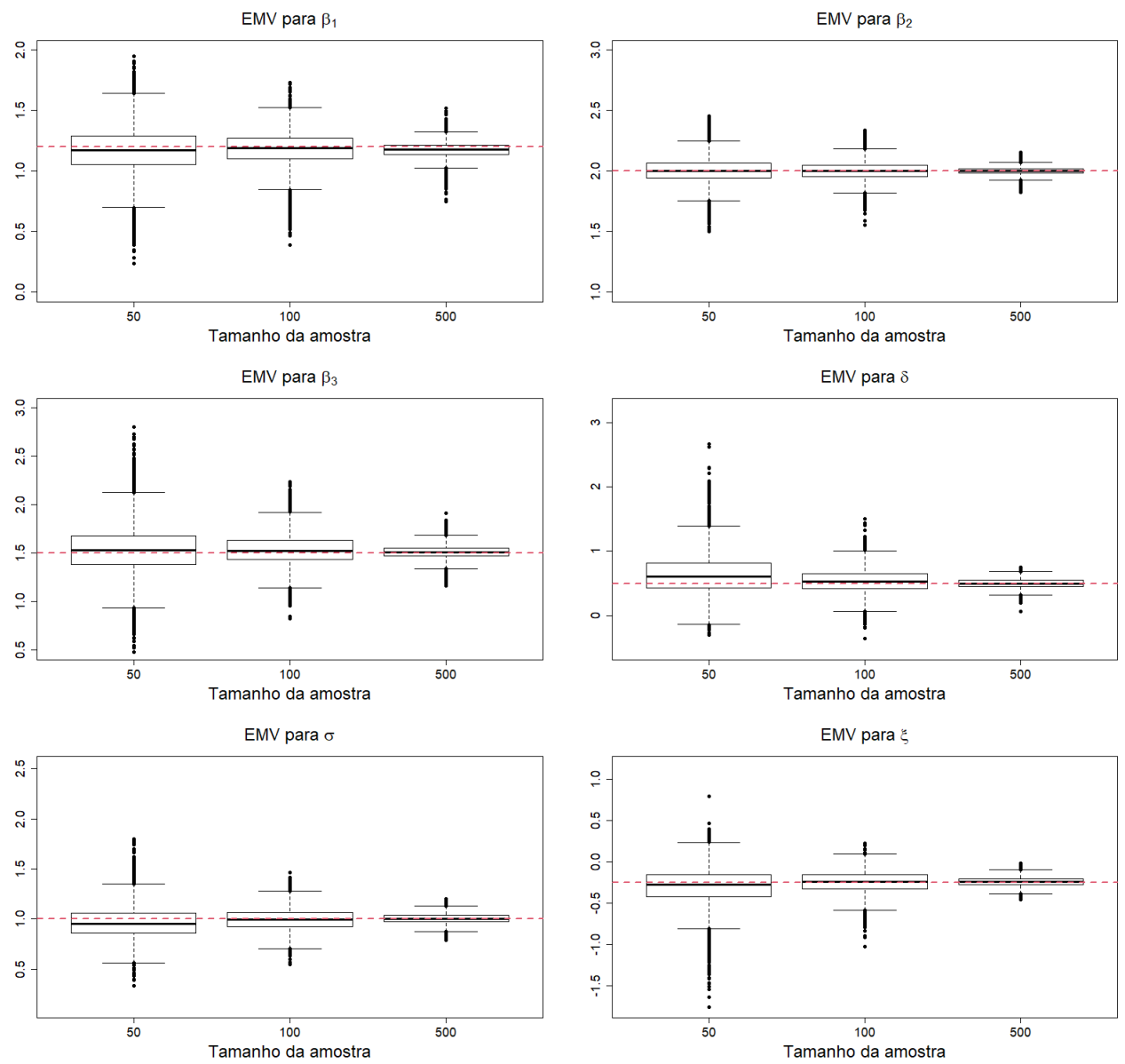


Figura 12: Boxplots dos parâmetros estimados sob o Cenário 5. A linha tracejada vermelha representa o verdadeiro valor do parâmetro

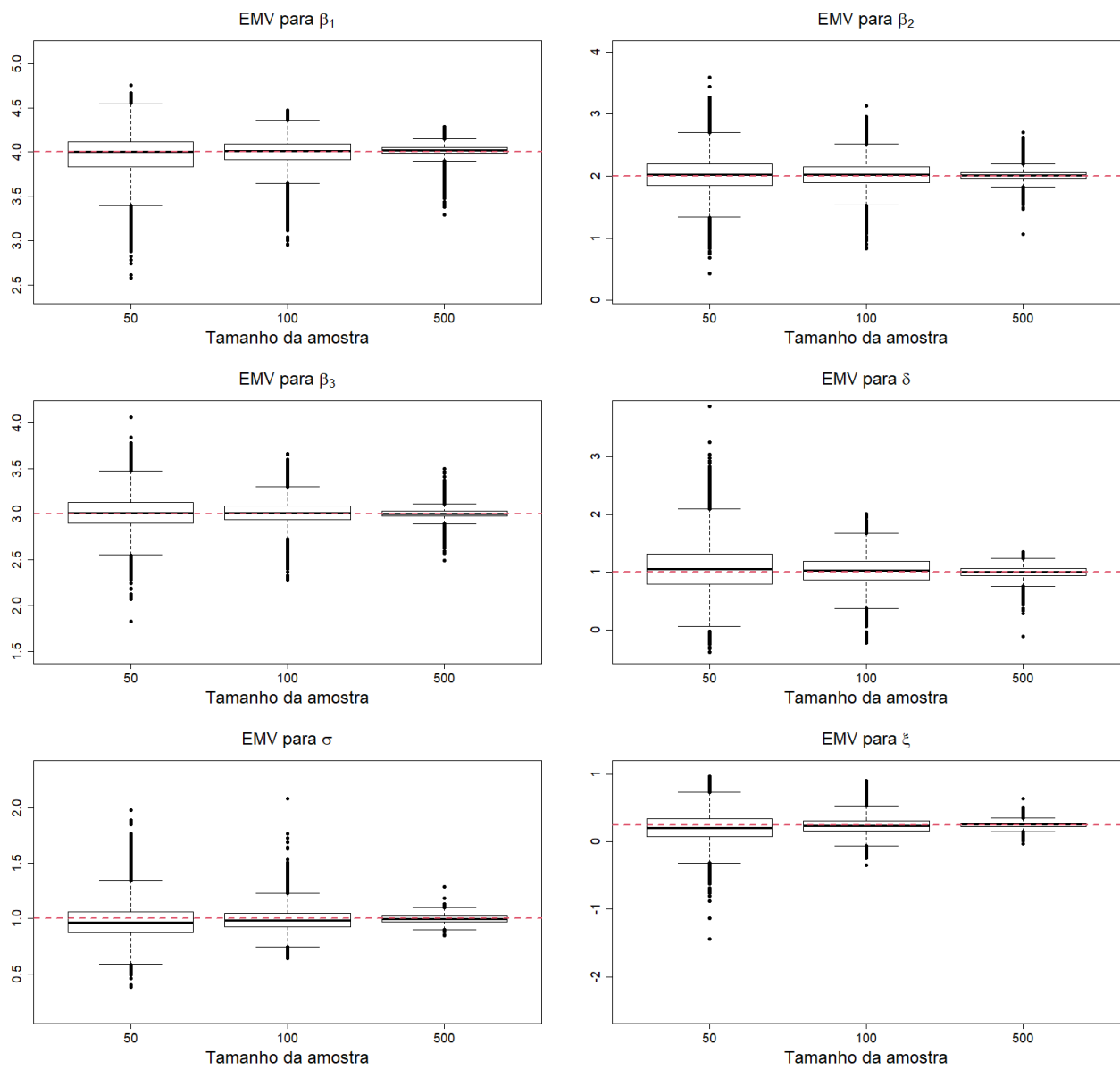


Figura 13: Boxplots dos parâmetros estimados sob o Cenário 6. A linha tracejada vermelha representa o verdadeiro valor do parâmetro

6 Aplicação

Neste capítulo, utilizamos dados climáticos de uma região do Brasil para demonstrar a aplicabilidade do modelo (4.1.2). Como mencionado anteriormente, uma das principais áreas de aplicação da teoria dos valores extremos é a climatologia. Esses dados são fundamentais para o monitoramento de biomas, a produção agrícola e o estudo das mudanças climáticas — temas que, atualmente, ocupam um lugar de destaque na mídia.

O Brasil, conhecido por sua vasta diversidade ecológica e por ser um dos maiores exportadores agrícolas do mundo, desperta grande interesse no monitoramento climático de suas diversas regiões. Neste capítulo, trabalhamos com os dados de Goiânia (GO).

Goiânia é um município brasileiro e a capital do estado de Goiás. A cidade está localizada a 209 km de Brasília, a capital do Brasil. Possui vegetação de cerrado, um dos biomas mais devastados do país. Goiânia e sua região metropolitana fazem parte de uma das áreas de Goiás com menor preservação de sua vegetação original.

6.1 Descrição dos dados

O Instituto Nacional de Meteorologia (INMET), vinculado ao Ministério da Agricultura e Pecuária do Brasil, desempenha um papel crucial nesse processo, sendo responsável pela criação e manutenção de bancos de dados meteorológicos alimentados por estações automáticas espalhadas por todo o país.

Os dados utilizados nesta seção são provenientes da estação automática A002, localizada em Goiânia (GO). Eles estão disponíveis gratuitamente no site do INMET (<https://portal.inmet.gov.br/>) e correspondem ao período de 2011 a 2022.

Do total de 19 variáveis que compõem o banco de dados, neste trabalho utilizamos 5. São elas:

- **TEMPERATURA DO AR BULBO SECO (T)** : Medida da temperatura do ar a partir do termômetro de bulbo seco ocorrida na última hora antes da mensagem de dados. É medida em graus Celsius ($^{\circ}\text{C}$);
- **TEMPERATURA DO PONTO DE ORVALHO (TO)**: O ponto de orvalho é definido como a temperatura até a qual o ar (ou gás) deve ser resfriado para que a condensação de água se inicie, ou seja, para que o ar fique saturado de vapor de água. É medida em graus Celsius ($^{\circ}\text{C}$).
- **UMIDADE RELATIVA DO AR (U)**: Medida de vapor de água contida na atmosfera, medida de umidade relativa do ar média ocorrida na última hora antes

de cada mensagem de dados. É calculada por porcentagem de vapor de água na atmosfera;

- **PRESSÃO ATMOSFÉRICA (P)**: Medida de pressão atmosférica média que foi da última hora antes da mensagem de dados. É medida em milibares (mb);
- **VELOCIDADE DO VENTO RAJADA MÁXIMA (VR)**: Medida máxima capturada da velocidade do vento na última hora. É medida por metros por segundo (m/s);

Além dessas 5 variáveis, consideramos a variável binária ESTAÇÕES (E). A criação dessa variável se justifica pelas características do clima de Goiânia. Segundo o INMET, Goiânia tem um clima tropical úmido, com duas estações bem definidas: seca e úmida. A estação chuvosa (1) ocorre de outubro a abril, e a estação seca (0) ocorre de maio a setembro.

De todos os eventos climáticos extremos que vêm acontecendo no mundo, a ESTIAGEM é o que mais vem castigando a população de Goiânia. Segundo notícias de sites como Agrolink (23/09/2024), Sociedade Nacional da Agricultura (23/09/2024) e Edição Brasil (04/10/2024), a ESTIAGEM prolongada em Goiás tem causado sérios impactos na produção agrícola em 2024. O fenômeno climático La Niña, que se intensificou neste ano, agravou a escassez de chuvas e aumentou as temperaturas, afetando cultivos essenciais como soja, milho e hortaliças. A pecuária também sofre com a degradação das pastagens e a necessidade de suplementação alimentar, o que eleva os custos de produção.

Das cinco variáveis climáticas citadas acima, a temperatura mínima do ponto de orvalho (TO) e umidade mínima (U) são as variáveis que permitem explicar melhor a estiagem e sua relação com a produção agrícola, Changnon, Sandstrom e Schaffer (2003).

A temperatura do ponto de orvalho está intimamente relacionada à umidade do ar refletindo a quantidade de vapor d'água presente na atmosfera (Talaia e Vigário (2016), Junior (2011)). Quanto maior a umidade relativa do ar, mais alto é o ponto de orvalho, pois há mais vapor d'água disponível para condensar. Em contrapartida, quando o ar é mais seco, o ponto de orvalho é mais baixo, indicando que a quantidade de vapor d'água é menor e a condensação ocorre a uma temperatura mais baixa. Esse cenário favorece a rápida evaporação de umidade das superfícies e da vegetação, tornando a biomassa mais suscetível a ignição e facilitando a propagação de incêndios.

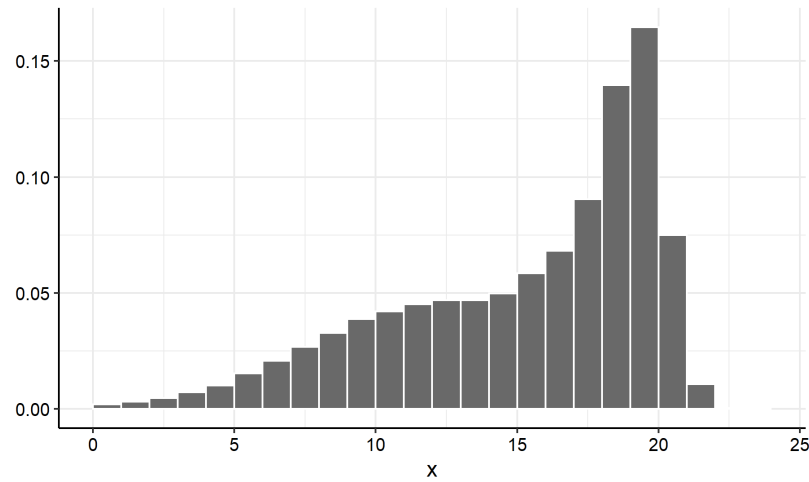
Por essa razão, o modelo de regressão BGEV foi aplicado aos valores mínimos dessas duas variáveis. Na primeira aplicação, a variável TO foi considerada como resposta, enquanto U, E, P e VR foram utilizadas como covariáveis. Na segunda aplicação, a variável U foi definida como resposta, e T, P e VR foram empregadas como covariáveis.

6.2 Ajuste dos dados extremos de TO e U

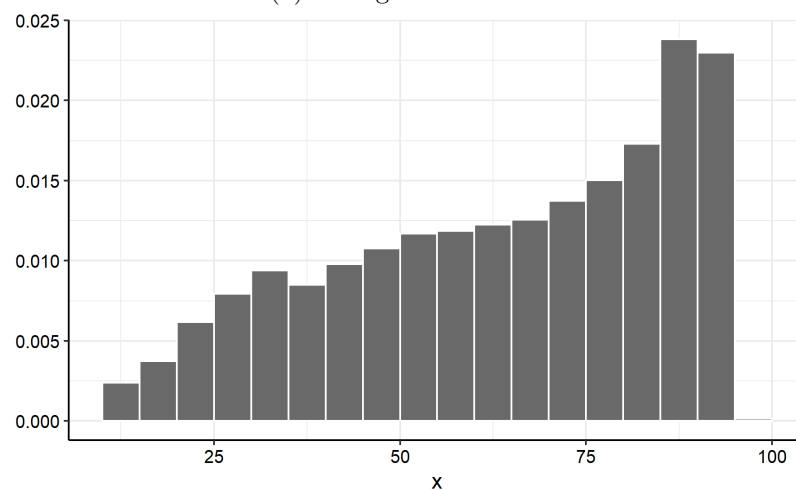
Nesta seção modelamos os dados da TO e U com a distribuição BGEV. Desconsideramos a dependência de TO e U com as variáveis E, P, e VR.

Tabela 3: Estatísticas descritivas de TO e U para os dados originais

	TO	U
Média	15.27	64.10
Desvio Padrão	4.62	22.55
Mediana	16.72	68.00
max	23.60	97.00
min	-6.43	9.83



(a) Histograma de TO



(b) Histograma de U

Figura 14: Histogramas dos dados originais

É importante ressaltar que os dados originais são compostos por observações registradas a cada hora, o que equivale, para o período estudado, a 105192 observações.

Esses dados não poderiam ser utilizados diretamente para regressão, uma vez que apresentam dependência temporal. A primeira medida adotada para resolver esse problema foi resumir os dados em médias diárias, resultando em um novo conjunto de dados com uma observação por dia, totalizando 4383 observações.

Os dados resumidos a médias diárias ainda apresentam dependência temporal. Obtemos subamostras de valores mínimos de blocos de tamanho N de tal forma que os valores dessas subamostras são iid, para cada conjunto de dados.

Para obter cada subamostra de valores mínimos independentes a partir dos dados originais, utilizamos a técnica do bloco mínimo (Jondeau e Rockinger (1999)) combinado com o teste de Ljung-Box de tal forma que a um nível de 5% os valores de cada subamostra de tamanho $N = 60$ dias são independentes.

Com o uso dos pacotes EVD (Stephenson (2002)), obtivemos as estimativas de máxima verossimilhança para os as duas subamostras de mínimos da GEV, enquanto os valores da BGEV foram calculados utilizando as fórmulas apresentadas na Seção 4. Sendo assim, é apresentado na Tabela 5 estimativas e erros padrão dos parâmetros da distribuição BGEV utilizando o método de máxima verossimilhança em comparação com as estimativas e erros padrão dos parâmetros da distribuição GEV.

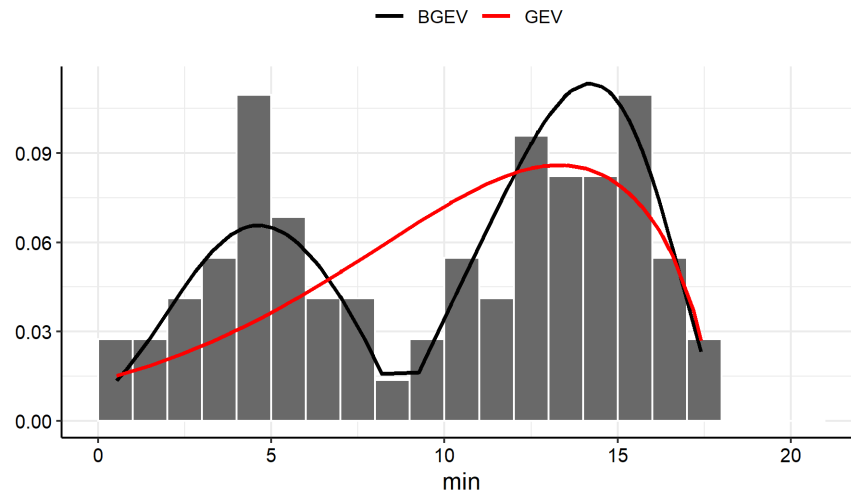
Tabela 4: Estatísticas descritivas de TO e U com blocos mínimos N=60

	TO	U
Média	9.78	45.76
Desvio Padrão	5.04	13.35
Mediana	10.44	49.00
max	17.40	67.20
min	0.52	16.83

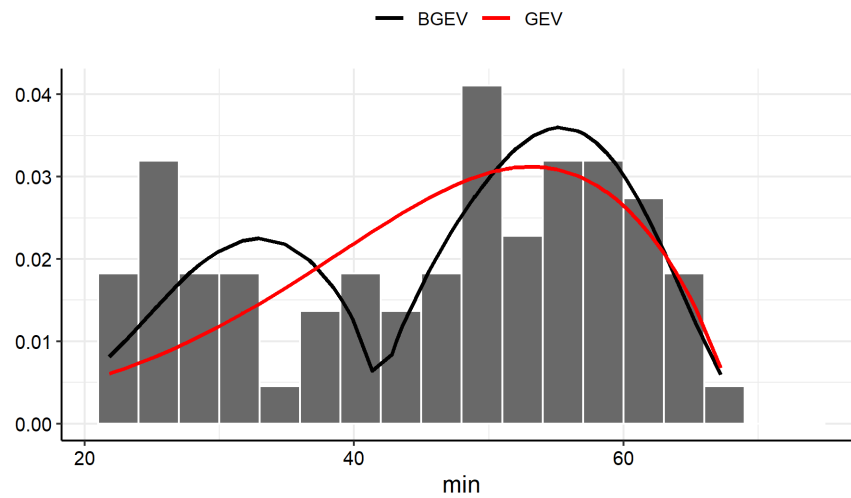
Os histogramas dos dados originais e sub-amostras de mínimos estão nas Figuras 14 e 15. Primeiramente, na Figura 14, estão representadas os histogramas das variáveis TO e U originais, ou seja, dos dados registrados pela estação meteorológica. Já na Figura 15, estão os histogramas dos dados obtidos dos blocos mínimos de tamanho 60. Os histogramas da Figura 16 indicam que a densidade adequada para seu ajuste é bimodal. Sendo assim, esses dados mínimos foram ajustados pela densidades da distribuições GEV e BGEV, para as duas variáveis climáticas.

Tabela 5: Estimativas e erros padrão dos parâmetros da distribuição BGEV e GEV com EMV

	BGEV		GEV	
	Estimativas	EP	Estimativas	EP
$\hat{\xi}$	-038	0.09	-0.65	0.9
$\hat{\mu}$	8.72	0.23	0.90	0.72
$\hat{\sigma}$	25.10	9.49	5.80	0.63
$\hat{\delta}$	0.86	0.20	-	-
$\hat{\xi}$	-0.37	0.09	-0.58	0.08
$\hat{\mu}$	41.88	0.49	43.07	1.85
$\hat{\sigma}$	64.31	32.92	14.66	1.53
$\hat{\delta}$	0.54	0.18	-	-



(a) Histograma comparativo entre GEV e BGEV TO



(b) Histograma comparativo entre GEV e BGEV U

Figura 15: Comparativos entre GEV e BGEV

Discutindo especificamente a variável TO, quando observado o histograma dos da-

dos originais na Figura 14a, nota-se que os dados estão concentrados em valores próximos de 20 °C e que não há, de forma clara, a formação de uma segunda moda. Por outro lado, ao considerar os dados tratados em blocos mínimos na Figura 15a, a formação de duas modas é nítida, e agora não há valores maiores que 18 °C na amostra.

Tratando-se da variável U , observa-se que, no caso dos dados originais apresentados na Figura 14b, a distribuição é unimodal. Usando como referência a Tabela 3, nota-se que a mediana está em 68, ou seja, 50% dos valores da amostra original são maiores que esse valor, e apresenta um valor máximo de 97. Quando se refere à Figura 15b, a distribuição já apresenta um comportamento bimodal e observa-se na Tabela 4 que a mediana é de 49.

A Figura 15 também apresenta uma comparação entre os ajustes das curvas de densidade das distribuições BGEV e GEV. Nesse caso, fica evidente que a BGEV leva vantagem ao se ajustar aos dados, uma vez que possui a capacidade de acomodar distribuições bimodais, uma característica que falta na GEV.

6.3 Modelo de regressão temperatura do ponto de orvalho

De acordo com Changnon, Sandstrom e Schaffer (2003) a temperatura do ponto de orvalho (TO) é a medida que indica a temperatura na qual o vapor d'água presente na atmosfera começa a se condensar, formando orvalho. Em meteorologia, ela é útil para previsões climáticas, identificação de nevoeiros e avaliação de condições de conforto térmico. Na indústria, o controle da temperatura do ponto de orvalho é crucial para evitar problemas como corrosão em equipamentos, danos a materiais sensíveis à umidade e falhas em processos que dependem de condições atmosféricas específicas. Além disso, na agricultura, conhecer essa temperatura ajuda a prever geadas e a gerenciar a umidade ideal para o cultivo de plantas.

Devido a importância de TO, nesta seção, consideramos como variável resposta (Y) e como covariáveis umidade U_i^* , estações E_i^* , pressão P_i^* e velocidade do vento rajada VR_i^* . No modelo de regressão BGEV, sejam Y_1, \dots, Y_n variáveis aleatórias independentes correspondentes à temperatura mínima do ponto de orvalho, em que $Y_i \sim F_{BG}^*(\cdot; \xi, m_i, \sigma, \delta)$, definida em (4.1.2).

Para comparar submodelos do modelo de regressão BGEV, foram definidas as seguintes especificações:

- Modelo 1 : $m_i = \beta_1 + \beta_2 U_i^*$,
- Modelo 2 : $m_i = \beta_1 + \beta_2 U_i^* + \beta_3 E_i^*$,
- Modelo 3 : $m_i = \beta_1 + \beta_2 U_i^* + \beta_3 E_i^* + \beta_4 P_i^*$,

- Modelo 4 : $m_i = \beta_1 + \beta_2 U_i^* + \beta_3 E_i^* + \beta_4 P_i^* + \beta_5 VR_i^*$;

em que m_i representa a mediana da temperatura mínima do ponto de orvalho na i -ésima posição. E em que

$$U_i^* = U_i - \bar{U},$$

$$P_i^* = P_i - \bar{P},$$

$$VR_i^* = VR_i - \bar{VR}$$

são, respectivamente, a umidade relativa centralizada pela média amostral, a pressão atmosférica centralizada pela média amostral e a velocidade do vento rajada centralizada pela média amostral, e E_i é a estação do ano.

As especificações dos modelos apresentadas foram selecionadas por meio da seleção *forward*, em que se acrescentam covariáveis ao modelo, avaliando se o p -valor da estimativa de β , correspondente à covariável, possui valor menor que um α determinado. Se essa condição for satisfeita, a variável deve ser incluída no modelo. Neste trabalho, foi definido que $\alpha = 0.05$.

Além disso, há a questão do parâmetro ξ , que implica duas formas diferentes para o modelo de regressão BGEV, $F_{BG}^*(\cdot; \xi, m_i, \sigma, \delta)$, conforme apresentado em 4.1.2. Dessa forma, é necessário definir $\xi = 0$ ou $\xi \neq 0$ antes de realizar as estimativas dos modelos. Por isso, cada um dos quatro submodelos especificados foi ajustado considerando separadamente os casos em que $\xi = 0$ e $\xi \neq 0$.

Uma vez ajustados os modelos, calculam-se os critérios de informação AIC e BIC apresentados em 4. Os resultados para todas as 4 especificações de modelo, considerando $\xi = 0$ e $\xi \neq 0$, estão apresentados na Tabela 6. Apesar da necessidade de se ajustar os dados de maneira separada para os dois casos de ξ , a comparação foi feita de maneira conjunta, a fim de encontrar o melhor modelo e, portanto, também definir se $\xi = 0$ ou $\xi \neq 0$. Considerando os resultados da Tabela 6, o melhor modelo é o Modelo 3 com $\xi \neq 0$, pois este apresenta os menores valores tanto para AIC quanto para BIC.

Tabela 6: Medidas de comparação entre modelos BGEV para temperatura do ponto de orvalho

Modelo	Parâmetro ξ	AIC	BIC
Modelo 1	$\xi = 0$	392.6367	401.7986
	$\xi \neq 0$	371.7461	383.1984
Modelo 2	$\xi = 0$	316.7746	328.2269
	$\xi \neq 0$	314.8961	328.6389
Modelo 3	$\xi = 0$	300.3726	314.1154
	$\xi \neq 0$	289.292	305.3252
Modelo 4	$\xi = 0$	298.3478	314.381
	$\xi \neq 0$	302.8041	321.1278

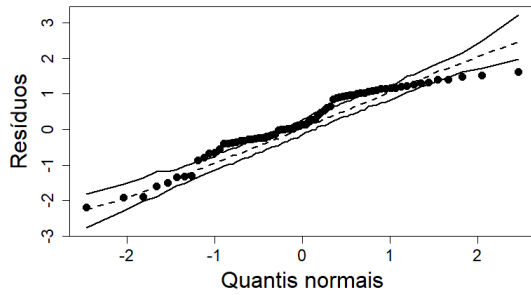
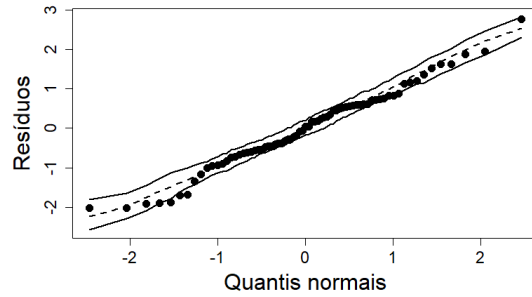
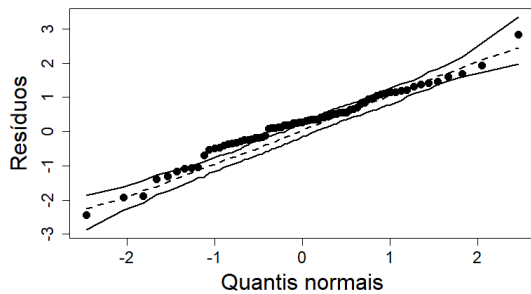
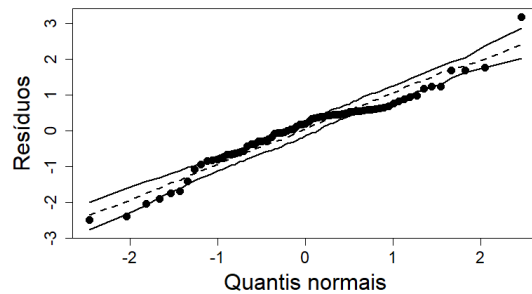
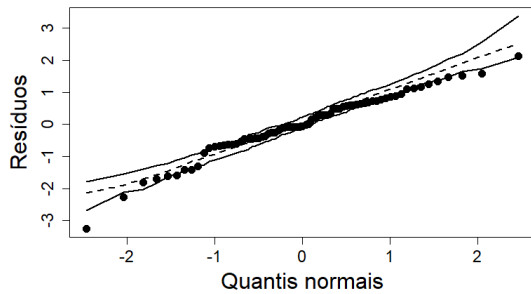
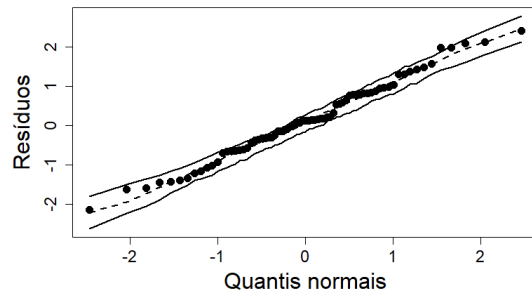
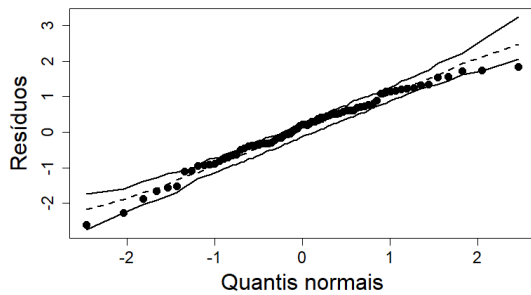
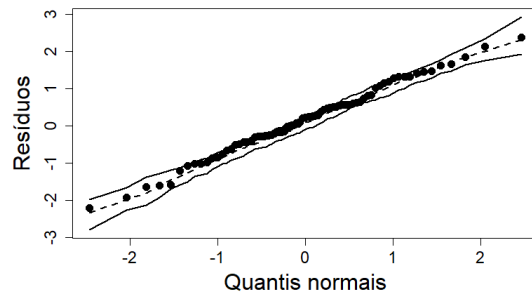
(a) Modelo 1 $\xi = 0$ (b) Modelo 1 $\xi \neq 0$ (c) Modelo 2 $\xi = 0$ (d) Modelo 2 $\xi \neq 0$ (e) Modelo 3 $\xi = 0$ (f) Modelo 3 $\xi \neq 0$ (g) Modelo 4 $\xi = 0$ (h) Modelo 4 $\xi \neq 0$

Figura 16: Análise de resíduos para os modelos para temperatura do ponto orvalho

Embora os critérios de informação AIC e BIC sejam ferramentas úteis para comparar modelos, eles não são suficientes por si só para garantir a escolha do modelo mais adequado. Ambos os critérios se baseiam em uma penalização da verossimilhança em função do número de parâmetros, favorecendo modelos com melhor ajuste e menor complexidade. A análise de resíduos é uma abordagem fundamental para complementar a seleção de modelos, pois permite avaliar a qualidade do ajuste de maneira mais detalhada. Para tanto, é apresentada a análise de resíduos para todas as especificações de modelos na Figura 16.

Ao observar então a Figura 16, é notável que as primeiras duas especificações de modelos, ou seja, o Modelo 1 e o Modelo 2, realmente não apresentam um bom ajuste aos dados, formando gráficos que fogem do comportamento esperado dos resíduos. O modelo apontado como o melhor pelos critérios de informação, isto é, o Modelo 3 com $\xi \neq 0$, apresenta resíduos quantílicos todos compreendidos pelo envelope, reforçando mais uma vez que essa deve ser a escolha adequada. Os gráficos para os modelos especificados no Modelo 4 também apresentam o comportamento esperado, seguindo a reta dos quantis normais, apesar de valores maiores para os critérios de informação.

Um modelo pode ter um bom ajuste estatístico, mas, se os parâmetros estimados não fizerem sentido teórico ou prático no contexto do problema estudado, sua utilidade fica comprometida. Por fim, resta avaliar a Tabela 7, na qual encontram-se as estimativas, erros padrão, z -stat e p -valor para os parâmetros de cada especificação dos modelos avaliados. A primeira coisa que chama a atenção é o fato de que, em todos os modelos, as estimativas dos parâmetros apresentam p -valor menor que 0.05, exceto pela estimativa de ξ para a especificação do Modelo 4 com $\xi \neq 0$, sugerindo então que, ao acrescentar a variável VR_i^* , o valor de ξ deveria ser igual a 0.

Outro fator a se considerar é que, para todos os modelos, a ação de U_i^* é positiva sobre a mediana da temperatura do ponto de orvalho, o que faz sentido, uma vez que, quanto maior a umidade, maior a temperatura do ponto de orvalho. A variável E_i também possui efeito positivo, o que é esperado, uma vez que toma-se como referência o efeito da estação quente e úmida. O efeito de P_i^* é negativo sobre m_i , o que também faz sentido, uma vez que, quanto maior a pressão, menor deve ser a temperatura para se formar orvalho. Por fim, o efeito de VR_i^* , quando inserida no modelo, é positivo.

Ponderando todas essas considerações, o modelo escolhido para ajustar a variável temperatura do ponto de orvalho é a especificação do Modelo 3 com $\xi \neq 0$. Sendo assim, uma das vantagens da reparametrização foi facilitar a interpretação dos resultados. Começamos com a estimativa do parâmetro β_1 , que representa o intercepto e possui valor de 8.672, ou seja, quando as demais covariáveis contínuas estiverem em sua média amostral, uma vez que estão centralizadas, m_i assume esse valor. Passando para a estimativa de β_2 , para cada unidade de umidade a mais, a mediana da temperatura do ponto de

orvalho aumenta em 0.234. O parâmetro vinculado a E_i é o β_3 , e, como a referência é a estação quente e chuvosa, essa possui mediana de m_i 3.262 maior que a estação árida e amena. Por fim, a estimativa de β_4 indica que, para cada unidade de pressão atmosférica, a mediana m_i diminui em 0.423 °C.

Tabela 7: Estimativas, erros padrão, estatística z e p -valores para os modelos RBGEV aplicados em temperatura do ponto de orvalho.

Modelo 1								
	$\xi = 0$				$\xi \neq 0$			
	Estimativa	Erro padrão	z -stat	p -valor	Estimativa	Erro padrão	z -stat	p -valor
β_1	9.399	0.244	38.423	< 0.001	11.091	0.288	38.482	< 0.001
β_2	0.255	0.010	25.239	< 0.001	0.331	0.011	29.516	< 0.001
σ	8.275	2.243	3.688	< 0.001	6.796	1.690	4.020	< 0.001
δ	0.598	0.163	3.663	< 0.001	0.388	0.153	2.529	0.011
ξ	-	-	-	-	0.579	0.072	8.003	< 0.001

Modelo 2								
	$\xi = 0$				$\xi \neq 0$			
	Estimativa	Erro padrão	z -stat	p -valor	Estimativa	Erro padrão	z -stat	p -valor
β_1	7.454	0.213	35.006	< 0.001	7.619	0.217	35.180	< 0.001
β_2	0.260	0.008	17.554	< 0.001	0.210	0.008	25.830	< 0.001
β_3	4.697	0.268	34.171	< 0.001	5.103	0.241	21.194	< 0.001
σ	6.065	1.562	3.882	< 0.001	3.891	0.836	2.333	< 0.001
δ	0.918	0.194	4.738	< 0.001	0.530	0.177	2.999	0.002
ξ	-	-	-	-	0.192	0.067	2.877	0.004

Modelo 3								
	$\xi = 0$				$\xi \neq 0$			
	Estimativa	Erro padrão	z -stat	p -valor	Estimativa	Erro padrão	z -stat	p -valor
β_1	8.309	0.184	45.093	< 0.001	8.672	0.181	47.814	< 0.001
β_2	0.213	0.007	29.226	< 0.001	0.234	0.006	36.500	< 0.001
β_3	3.396	0.250	13.531	< 0.001	3.262	0.175	18.667	< 0.001
β_4	-0.737	0.056	-13.207	< 0.001	-0.423	0.036	-11.766	< 0.001
σ	2.606	0.433	6.013	< 0.001	3.237	0.613	5.280	< 0.001
δ	0.465	0.148	3.139	0.001	0.442	0.182	2.420	0.015
ξ	-	-	-	-	0.504	0.114	4.435	< 0.001

Modelo 4								
	$\xi = 0$				$\xi \neq 0$			
	Estimativa	Erro padrão	z -stat	p -valor	Estimativa	Erro padrão	z -stat	p -valor
β_1	7.978	0.171	46.551	< 0.001	7.904	0.194	40.679	< 0.001
β_2	0.239	0.006	38.843	< 0.001	0.249	0.007	31.091	< 0.001
β_3	3.795	0.276	13.751	< 0.001	3.475	0.307	11.305	< 0.001
β_4	-0.690	0.050	-13.758	< 0.001	-0.954	0.068	-13.864	< 0.001
β_5	0.456	0.087	5.254	< 0.001	0.412	0.119	3.448	< 0.001
σ	2.450	0.380	6.435	< 0.001	2.154	0.341	6.315	< 0.001
δ	0.486	0.148	3.282	0.001	0.380	0.192	1.980	0.040
ξ	-	-	-	-	-0.006	0.144	-0.048	0.962

6.4 Modelo de regressão umidade relativa do ar

A umidade relativa do ar é especialmente importante em Goiânia, capital de Goiás, devido às características climáticas da região, que incluem períodos de seca prolongada e alta variação térmica ao longo do dia. Durante a estação seca, que vai de maio a setembro, a umidade relativa do ar pode atingir níveis críticos, muitas vezes abaixo de 30%, o que aumenta os riscos de problemas respiratórios, ressecamento da pele e incêndios florestais. Por outro lado, na estação chuvosa, a umidade elevada pode favorecer a proliferação de fungos e ácaros, afetando a saúde da população e a conservação de materiais. Além disso, o monitoramento da umidade relativa é importante para atividades econômicas locais, como a agricultura, que depende de condições climáticas adequadas para o cultivo de grãos e outras culturas típicas da região.

Pensando na relevância dessa variável, sejam Y_1, \dots, Y_n variáveis aleatórias independentes correspondentes à umidade relativa do ar mínima, em que $Y_i \sim F_{BG}^*(\cdot; \xi, m_i, \sigma, \delta)$, definida em (4.1.2). Para comparar submodelos do modelo de regressão BGEV, foram definidas as seguintes especificações:

- Modelo 1 : $m_i = \beta_1 + \beta_2 T_i^*$,
- Modelo 2 : $m_i = \beta_1 + \beta_2 T_i^* + \beta_3 VR_i^*$,
- Modelo 3 : $m_i = \beta_1 + \beta_2 T_i^* + \beta_3 VR_i^* + \beta_4 P_i^*$;

em que m_i representa a mediana da umidade relativa mínima do ar na i -ésima posição. E em que

$$\begin{aligned} T_i^* &= T_i - \bar{T}, \\ VR_i^* &= VR_i - \overline{VR}, \\ P_i^* &= P_i - \bar{P} \end{aligned}$$

são, respectivamente, a temperatura do ar seco centralizada pela média, a velocidade do vento rajada centralizada pela média amostral e a pressão atmosférica centralizada pela média amostral para cada i -ésima posição.

Assim como na aplicação com a temperatura do ponto de orvalho, as especificações dos modelos apresentadas foram selecionadas por meio da seleção *forward*, em que é definido $\alpha = 0.05$. Da mesma forma que anteriormente, o parâmetro ξ implica em duas formas diferentes para o modelo de regressão BGEV. Portanto, também foi definido $\xi = 0$ e $\xi \neq 0$ antes de realizar as estimativas do modelo. Assim, cada uma das 3 especificações foi considerada tanto na versão com $\xi = 0$ quanto na versão com $\xi \neq 0$.

Seguindo a mesma lógica aplicada anteriormente, apesar de ajustadas separadamente as especificações com $\xi = 0$ e $\xi \neq 0$, a avaliação é feita de forma conjunta. Considerando a Tabela 8, o modelo que apresenta, em conjunto, os menores valores para AIC e BIC é o Modelo 2 com $\xi \neq 0$.

Ainda é necessário realizar a avaliação dos resíduos quantílicos. Para tanto, a Figura 17 apresenta os gráficos de envelope de resíduos de todas as especificações de modelos para i (mediana da umidade relativa do ar) como variável resposta. Assim, nota-se que o gráfico da especificação escolhida por meio dos critérios de informação AIC e BIC, o Modelo 2 com $\xi \neq 0$, apresenta o comportamento esperado quando o modelo está bem ajustado aos dados, seguindo os quantis normais e permanecendo dentro das bandas de confiança do envelope.

Finalmente, resta avaliar a Tabela 9, na qual encontram-se as estimativas, erros padrão, z -stat e p -valor para os parâmetros de cada especificação dos modelos avaliados para a mediana m_i da umidade relativa do ar. Nota-se, primeiramente, que, em relação ao Modelo 1 com $\xi \neq 0$, a estimativa do parâmetro σ possui p -valor maior que 0.05, e o mesmo ocorre para ξ no mesmo modelo. Outro ponto a se destacar é que todas as covariáveis, em todos os modelos estudados, apresentam efeito negativo na mediana da umidade relativa do ar, o que faz sentido considerando a relação das covariáveis com a umidade relativa do ar.

Analisando agora especificamente o modelo escolhido, o Modelo 2 com $\xi \neq 0$ foi selecionado por apresentar os menores valores de AIC e BIC, um gráfico de resíduos que não descarta a hipótese de que o modelo está adequado e estimativas para os parâmetros coerentes com o que é estudado na prática. Sendo assim, a interpretação das estimativas dos parâmetros começa com β_1 , que representa o intercepto, ou seja, o valor assumido pela mediana da umidade relativa do ar quando as demais variáveis possuem valores iguais aos de suas médias amostrais, uma vez que as covariáveis estão centralizadas. Seguindo para β_2 , com valor de -1.842 , valor negativo indicando que, quanto maior a temperatura do ar, menor será a mediana da umidade relativa do ar. Em seguida, β_3 , com valor de -5.146 , também com valor negativo, indicando que, quanto maior for a velocidade da rajada de vento, menor será a mediana da umidade relativa do ar. Dado isso, é de se ressaltar que a estimativa de δ é positiva e $\xi > 0$, indicando bimodalidade e forma de cauda pesada, respectivamente.

Tabela 8: Medidas de comparação entre modelos BGEV para umidade relativa do ar

Modelo	Parâmetro ξ	AIC	BIC
Modelo 1	$\xi = 0$	560,0423	569,2041
	$\xi \neq 0$	553,6306	565,0829
Modelo 2	$\xi = 0$	556,3689	567,8212
	$\xi \neq 0$	544,1218	557,8646
Modelo 3	$\xi = 0$	552,9676	566,7104
	$\xi \neq 0$	551,1549	567,1881

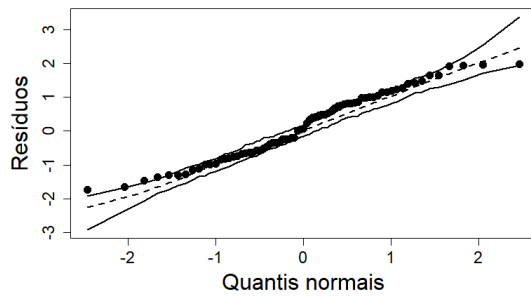
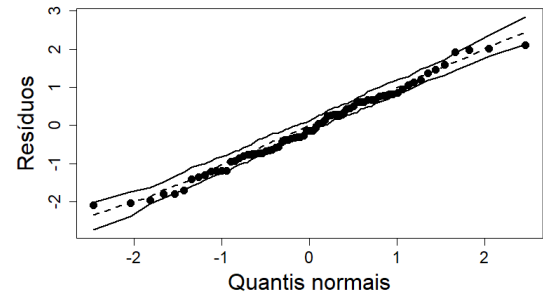
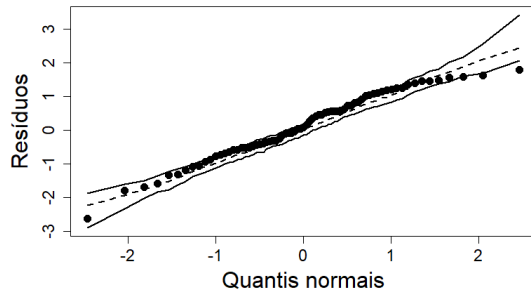
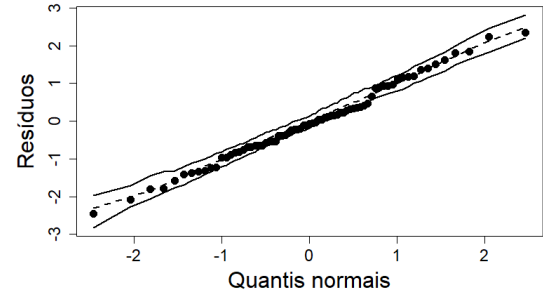
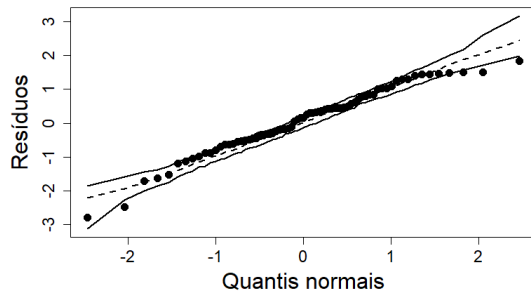
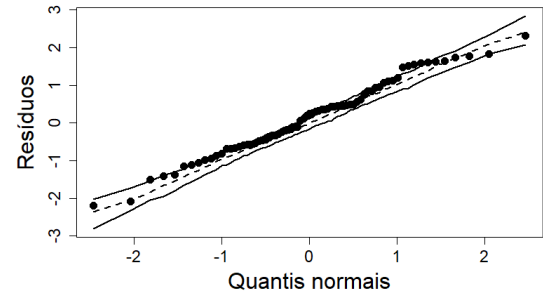
(a) Modelo 1 $\xi = 0$ (b) Modelo 1 $\xi \neq 0$ (c) Modelo 2 $\xi = 0$ (d) Modelo 2 $\xi \neq 0$ (e) Modelo 3 $\xi = 0$ (f) Modelo 3 $\xi \neq 0$

Figura 17: Análise de resíduos para os modelos para umidade relativa do ar.

Tabela 9: Estimativas, erros padrão, estatística z e p -valores para os modelos RBGEV aplicados em umidade relativa do ar.

Modelo 1								
	$\xi = 0$				$\xi \neq 0$			
	Estimativa	Erro padrão	z -stat	p -valor	Estimativa	Erro padrão	z -stat	p -valor
β_1	46.500	0.818	56.849	< 0.001	50.035	0.853	58.637	< 0.001
β_2	-1.522	0.159	-9.589	< 0.001	-2.132	0.148	-14.360	< 0.001
σ	75.089	33.917	2.214	0.027	131.886	72.104	1.829	0.067
δ	0.715	0.163	4.461	< 0.001	0.898	0.201	4.475	< 0.001
ξ	-	-	-	-	0.080	0.108	0.739	0.460

Modelo 2								
	$\xi = 0$				$\xi \neq 0$			
	Estimativa	Erro padrão	z -stat	p -valor	Estimativa	Erro padrão	z -stat	p -valor
β_1	45.389	0.704	64.398	< 0.001	49.241	0.886	55.585	< 0.001
β_2	-1.803	0.143	-12.632	< 0.001	-1.842	0.174	-10.581	< 0.001
β_3	-3.542	0.308	-11.477	< 0.001	-5.146	0.461	-11.158	< 0.001
σ	37.227	14.990	2.483	0.013	35.377	14.920	2.371	0.018
δ	0.507	0.147	3.448	< 0.001	0.452	0.157	2.873	0.004
ξ	-	-	-	-	0.377	0.089	4.224	< 0.001

Modelo 3								
	$\xi = 0$				$\xi \neq 0$			
	Estimativa	Erro padrão	z -stat	p -valor	Estimativa	Erro padrão	z -stat	p -valor
β_1	45.779	0.762	60.032	< 0.001	46.037	0.726	63.433	< 0.001
β_2	-2.372	0.306	-7.754	< 0.001	-2.275	0.349	-6.510	< 0.001
β_3	-3.315	0.567	-5.843	< 0.001	-3.671	0.664	-5.525	< 0.001
β_4	-1.888	0.406	-4.648	< 0.001	-1.787	0.426	-4.199	< 0.001
σ	36.658	14.968	2.449	0.014	27.831	11.685	2.382	0.017
δ	0.512	0.150	3.416	< 0.001	0.380	0.162	2.340	0.019
ξ	-	-	-	-	0.224	0.117	1.915	0.055

7 Conclusão

Neste trabalho, propõe-se uma nova classe de modelos de regressão para dados extremos que podem ser bimodais, o modelo de regressão BGEV, baseado na distribuição BGEV (*Bimodal Generalized Extreme Values*), proposto por Otiniano et al. (2023). Dada a formulação da distribuição BGEV, $Y \sim F_{BG}(\cdot; \xi, \mu, \sigma, \delta)$, μ é um parâmetro de localização. No entanto, μ não representa a média de Y , portanto, não possui uma interpretação direta no modelo de regressão. Dessa forma, a distribuição BGEV foi reparametrizada pela mediana (m), e a estrutura de regressão foi atribuída a esse parâmetro, o que torna mais simples a interpretação dos resultados. A inferência dos parâmetros do modelo de regressão BGEV foi realizada pelo método de máxima verossimilhança.

Considerando sempre o melhor modelo para ajuste de dados extremos, o modelo de regressão BGEV mostrou-se uma boa proposta para a modelagem da mínima da temperatura do ponto de orvalho e da umidade relativa do ar mínima para a estação meteorológica automática da região de Goiânia, região que historicamente sofre com secas. Para os dois conjuntos de dados, o modelo de regressão foi capaz de aderir aos dados de forma satisfatória, tornando possível a análise e interpretação da dependência entre a temperatura do ponto de orvalho e as variáveis estação, pressão e umidade. Da mesma forma, foi possível mostrar a dependência entre a umidade relativa do ar e outras variáveis, como temperatura do ar seco, velocidade do vento rajada e pressão atmosférica.

Portanto, este trabalho apresenta um novo modelo de regressão capaz de ajustar dados com comportamento bimodal, ampliando assim o conjunto de técnicas disponíveis para modelagem de dados extremos. Uma aplicação em dados relevantes para a região foi realizada, considerando que mudanças climáticas estão cada vez mais recorrentes e há a perspectiva de que eventos extremos ocorram com maior frequência. Daí a importância de ferramentas como a apresentada neste modelo para auxiliar na análise e previsão de eventos climáticos extremos.

O cálculo das estimativas foi satisfatório para os casos descritos nas simulações e aplicações. Porém, como o suporte da densidade da BGEV depende de alguns parâmetros, os estimadores dos parâmetros podem não convergir, especialmente para $\xi < 1/2$. Em um trabalho futuro, seria interessante utilizar outra abordagem para o cálculo das estimativas e compará-las com as de máxima verossimilhança.

Em síntese, este estudo traz uma contribuição relevante para a área da estatística ao propor uma nova distribuição com a capacidade de identificar duas modas, expandindo as opções metodológicas para a análise de dados extremos. As abordagens desenvolvidas neste trabalho têm o potencial de estimular pesquisas futuras e aplicações práticas em diversos campos que trabalham com dados bimodais.

Referências

ABRAMOWITZ, M.; STEGUN, I. A.; MILLER, D. Handbook of mathematical functions with formulas, graphs and mathematical tables (national bureau of standards applied mathematics series no. 55). *Journal of Applied Mechanics*, ASME International, v. 32, n. 1, p. 239–239, 1965. 19

AKAIKE, H. A new look at the statistical model identification. *IEEE transactions on automatic control*, Ieee, v. 19, n. 6, p. 716–723, 1974. 29

ARYAL, G. R.; TSOKOS, C. P. On the transmuted extreme value distribution with application. *Nonlinear Analysis: Theory, Methods & Applications*, Elsevier, v. 71, n. 12, p. e1401–e1407, 2009. 8

CHANGNON, D.; SANDSTROM, M.; SCHAFFER, C. Relating changes in agricultural practices to increasing dew points in extreme chicago heat waves. *Climate Research*, v. 24, n. 3, p. 243–254, 2003. 42, 46

CORDEIRO, G. M. et al. A new extended log-weibull regression: Simulations and applications. *Hacettepe Journal of Mathematics and Statistics*, v. 50, n. 3, p. 855–871, 2021. 8

DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. *Journal of Computational and graphical statistics*, Taylor & Francis, v. 5, n. 3, p. 236–244, 1996. 29

ELAL-OLIVERO, D. Alpha-skew-normal distribution. *Proyecciones (Antofagasta)*, SciELO Chile, v. 29, n. 3, p. 224–240, 2010. 14

EMBRECHTS, P. et al. Risk theory. *Modelling Extremal Events: for Insurance and Finance*, Springer, p. 21–57, 1997. 8

ESPINHEIRA, P. L.; SILVA, A. de O. Residual and influence analysis to a general class of simplex regression. *Test*, Springer, v. 29, n. 2, p. 523–552, 2020. 29

FISHER, R. A.; TIPPETT, L. H. C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In: CAMBRIDGE UNIVERSITY PRESS. *Mathematical proceedings of the Cambridge philosophical society*. [S.l.], 1928. v. 24, n. 2, p. 180–190. 8, 11

FONSECA, R. V.; CRIBARI-NETO, F. Bimodal birnbaum–saunders generalized autoregressive score model. *Journal of Applied Statistics*, Taylor & Francis, v. 45, n. 14, p. 2585–2606, 2018. 14

GNEDENKO, B. Sur la distribution limite du terme maximum d’une serie aleatoire. *Annals of mathematics*, JSTOR, p. 423–453, 1943. 8, 11

GRIGORIU, M.; SAMORODNITSKY, G. Reliability of dynamic systems in random environment by extreme value theory. *Probabilistic engineering mechanics*, Elsevier, v. 38, p. 54–69, 2014. 8

HAAN, L.; FERREIRA, A. *Extreme value theory: an introduction*. [S.l.]: Springer, 2006. v. 3, 8, 13

- JENKINSON, A. F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, Wiley Online Library, v. 81, n. 348, p. 158–171, 1955. 8, 12
- JONDEAU, E.; POON, S.-H.; ROCKINGER, M. *Financial modeling under non-Gaussian distributions*. [S.l.]: Springer Science & Business Media, 2007. 12
- JONDEAU, E.; ROCKINGER, M. The tail behavior of stock returns: Emerging versus mature markets. Banque de France Working Paper, 1999. 44
- JUNIOR, J. S. da C. *Sensor de umidade microeletrônico fabricado em substrato de alumina e baseado na detecção do ponto de orvalho*. Tese (Doutorado) — Universidade Estadual de Campinas, 2011. 42
- KOTZ, S.; NADARAJAH, S. *Extreme value distributions: theory and applications*. [S.l.]: world scientific, 2000. 8, 13
- NOCEDAL, J.; WRIGHT, S. J. *Numerical optimization*. [S.l.]: Springer, 1999. 28
- OTINIANO, C. E. et al. A bimodal model for extremes data. *Environmental and Ecological Statistics*, Springer, p. 1–28, 2023. 5, 6, 8, 9, 13, 25, 55
- OTINIANO, C. E. et al. The transmuted gev distribution: properties and application. *Communications for Statistical Applications and Methods*, Korean Statistical Society, v. 26, n. 3, p. 239–259, 2019. 8
- OTINIANO, C. E. et al. A bimodal model for extremes data. *arXiv preprint arXiv:2109.12738*, 2021. 13, 14
- PEREIRA, G. H. On quantile residuals in beta regression. *Communications in Statistics-Simulation and Computation*, Taylor & Francis, v. 48, n. 1, p. 302–316, 2019. 29
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2022. Disponível em: <https://www.R-project.org/>. 31
- RESNICK, S. Extreme values, regular variation and weak convergence. Springer Verlag, New York, 1987. 8
- RUDD, E. M. et al. The extreme value machine. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 40, n. 3, p. 762–768, 2017. 13
- SCHWARZ, G. Estimating the dimension of a model. *The annals of statistics*, JSTOR, p. 461–464, 1978. 29
- STEPHENSON, A. G. evd: Extreme value distributions. *R News*, v. 2, n. 2, p. 31–32, June 2002. Disponível em: <https://CRAN.R-project.org/doc/Rnews/>. 44
- SWAMEE, P. K.; RATHIE, P. N. Invertible alternatives to normal and lognormal distributions. *Journal of Hydrologic Engineering*, American Society of Civil Engineers, v. 12, n. 2, p. 218–221, 2007. 14
- TALAIA, M.; VIGÁRIO, C. Temperatura de ponto de orvalho: um risco ou uma necessidade. Imprensa da Universidade de Coimbra, 2016. 42

- VASCONCELOS, J. C. S. et al. A new regression model for bimodal data and applications in agriculture. *Journal of Applied Statistics*, Taylor & Francis, v. 48, n. 2, p. 349–372, 2021. 8
- VASCONCELOS, J. C. S. et al. A regression model for extreme events and the presence of bimodality with application to energy generation data. *IET Renewable Power Generation*, Wiley Online Library, v. 15, n. 2, p. 452–461, 2021. 9
- VILA, R. et al. A bimodal gamma distribution: Properties, regression model and applications. *Statistics*, Taylor & Francis, v. 54, n. 3, p. 469–493, 2020. 14
- VILA, R. et al. On a bimodal birnbaum–saunders distribution with applications to lifetime data. 2021. 9, 14
- VILLASEÑOR, J. A. R.; VILLASEÑOR, J. A. R. General extreme value distribution for the minima. *Frequency Analyses of Natural Extreme Events: A Spreadsheets Approach*, Springer, p. 363–396, 2021. 8
- WALD, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical society*, JSTOR, v. 54, n. 3, p. 426–482, 1943. 29

A Apêndices

A.1 Apêndice A

Diferenciando a função de log-verossimilhança $\ell(\boldsymbol{\theta})$ dada em (4.2.1) em relação a β_j , $j = 1, 2, \dots, p$:

$$U_{\beta_j} = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{i=1}^n \frac{d\ell_i(\boldsymbol{\theta})}{dm_i} \frac{dm_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \sum_{i=1}^n \frac{d\ell_i(\boldsymbol{\theta})}{dm_i} \frac{x_{ij}}{g'(m_i)},$$

em que $m_i = g^{-1}(\eta_i)$, $dm_i/d\eta_i = [g'(m_i)]^{-1}$ with $g'(z) = dg(z)/dz$, e $\partial \eta_i / \partial \beta_j = x_{ij}$, $i = 1, \dots, n$.

Para $\xi \in \mathbb{R}$, $d\mu_i^*/dm_i = 1, \forall i$, e

$$\frac{dT_i}{dm_i} = \frac{dT_i}{d\mu_i^*} \frac{d\mu_i^*}{dm_i} = \frac{dT_i}{d\mu_i^*} = -|y_i - \mu_i^*|^\delta - \delta(y_i - \mu_i^*)|y_i - \mu_i^*|^{\delta-1} \text{sgn}(y_i - \mu_i^*) = -|y_i - \mu_i^*|^\delta (\delta + 1).$$

Para $\xi \neq 0$,

$$\begin{aligned} \frac{d\ell_i(\boldsymbol{\theta})}{dm_i} &= \frac{\delta \text{sgn}(y_i - \mu_i^*)}{|y_i - \mu_i^*|} - \frac{d\mu_i^*}{dm_i} - \left(1 + \frac{1}{\xi}\right) \frac{1}{\Psi_i(\boldsymbol{\theta})} \frac{d\Psi_i(\boldsymbol{\theta})}{dm_i} + \frac{1}{\xi} \Psi_i^{-1/\xi-1}(\boldsymbol{\theta}) \frac{d\Psi_i(\boldsymbol{\theta})}{dm_i} \\ &= -\frac{\delta}{(y_i - \mu_i^*)} - \frac{d\Psi_i(\boldsymbol{\theta})}{dm_i} \frac{\Psi_i^{-1}(\boldsymbol{\theta})}{\xi} \left[1 + \xi - \Psi_i^{-1/\xi}(\boldsymbol{\theta})\right] \end{aligned}$$

com

$$\frac{d\Psi_i(\boldsymbol{\theta})}{dm_i} = \frac{\xi}{\sigma} \frac{dT_i}{dm_i} = -\frac{\xi}{\sigma} (\delta + 1) |y_i - \mu_i^*|^\delta.$$

Sendo assim,

$$\begin{aligned} \frac{d\ell_i(\boldsymbol{\theta})}{dm_i} &= -\frac{\delta}{(y_i - \mu_i^*)} + \frac{\xi}{\sigma} (\delta + 1) |y_i - \mu_i^*|^\delta \frac{\Psi_i^{-1}(\boldsymbol{\theta})}{\xi} \left[1 + \xi - \Psi_i^{-1/\xi}(\boldsymbol{\theta})\right] \\ &= -\frac{\delta}{(y_i - \mu_i^*)} + \frac{(\delta + 1)}{\sigma} |y_i - \mu_i^*|^\delta \Omega_i(\boldsymbol{\theta}) \\ &= \frac{1}{\sigma(y_i - \mu_i^*)} [(\delta + 1) T_i \Omega_i(\boldsymbol{\theta}) - \delta \sigma], \end{aligned}$$

em que $\Omega_i(\boldsymbol{\theta})$ é definida na Subseção 4.2 e varia dependendo de ser $\xi = 0$ ou $\xi \neq 0$.

Para $\xi = 0$,

$$\frac{d\ell_i(\boldsymbol{\theta})}{dm_i} = -\frac{\delta}{(y_i - \mu_i^*)} + \frac{(\delta + 1)}{\sigma} |y_i - \mu_i^*|^\delta \left(1 - e^{-\frac{T_i}{\sigma}}\right) = \frac{1}{\sigma(y_i - \mu_i^*)} [(\delta + 1) T_i \Omega_i(\boldsymbol{\theta}) - \delta \sigma].$$

Fixando $a_i = d\ell_i(\boldsymbol{\theta})/dm_i$,

$$\mathbf{U}_\beta = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta} = X^\top H \mathbf{a},$$

em que X , H , e \mathbf{a} são definidos na subseção 4.2.

Diferenciando $\ell(\boldsymbol{\theta})$ em relação a ξ ($\xi \neq 0$):

$$U_\xi = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \xi} = \sum_{i=1}^n \frac{d\ell_i(\boldsymbol{\theta})}{d\xi},$$

com

$$\begin{aligned} \frac{d\ell_i(\boldsymbol{\theta})}{d\xi} &= -\frac{\delta \operatorname{sgn}(y_i - \mu_i^*)}{|y_i - \mu_i^*|} \frac{d\mu_i^*}{d\xi} + \frac{1}{\xi^2} \log(\Psi_i(\boldsymbol{\theta})) - \left(1 + \frac{1}{\xi}\right) \frac{1}{\Psi_i(\boldsymbol{\theta})} \frac{d\Psi_i(\boldsymbol{\theta})}{d\xi} \\ &\quad - \Psi_i^{-1/\xi}(\boldsymbol{\theta}) \left[\frac{1}{\xi^2} \log(\Psi_i(\boldsymbol{\theta})) - \frac{1}{\xi} \frac{1}{\Psi_i(\boldsymbol{\theta})} \frac{d\Psi_i(\boldsymbol{\theta})}{d\xi} \right] \\ &= -\frac{\delta}{(y_i - \mu_i^*)} \frac{d\mu_i^*}{d\xi} + \frac{1}{\xi^2} \log(\Psi_i(\boldsymbol{\theta})) [1 - \Psi_i^{-1/\xi}(\boldsymbol{\theta})] - \frac{1}{\xi} \frac{d\Psi_i(\boldsymbol{\theta})}{d\xi} \Omega_i(\boldsymbol{\theta}). \end{aligned}$$

De (3.1.6),

$$\mu_i^* = m_i - \left[\frac{\sigma}{\xi} (e^{c_2 \xi} - 1) \right]^{1/(\delta+1)}.$$

Assim,

$$\begin{aligned} \frac{d\mu_i^*}{d\xi} &= -\frac{\sigma}{(\delta+1)} \left[\frac{\sigma}{\xi} (e^{c_2 \xi} - 1) \right]^{1/(\delta+1)-1} \left[-\frac{1}{\xi^2} (e^{c_2 \xi} - 1) + \frac{c_2}{\xi} e^{c_2 \xi} \right] \\ &= \frac{1}{\xi(\delta+1)} \left[\frac{\sigma}{\xi} (c_1^{-\xi} - 1) \right]^{1/(\delta+1)-1} \frac{\sigma}{\xi} (c_1^{-\xi} - 1) - \frac{1}{\xi(\delta+1)} \left[\frac{\sigma}{\xi} (c_1^{-\xi} - 1) \right]^{1/(\delta+1)-1} \sigma c_2 c_1^{-\xi} \\ &= \frac{1}{\xi(\delta+1)} \left[\frac{\sigma}{\xi} (c_1^{-\xi} - 1) \right]^{1/(\delta+1)} - \frac{1}{\xi(\delta+1)} \left[\frac{\sigma}{\xi} (c_1^{-\xi} - 1) \right]^{1/(\delta+1)-1} \sigma c_2 c_1^{-\xi} \\ &= \frac{1}{\xi(\delta+1)} \left[\frac{\sigma}{\xi} (c_1^{-\xi} - 1) \right]^{1/(\delta+1)} \left\{ 1 - \sigma c_2 c_1^{-\xi} \left[\frac{\sigma}{\xi} (c_1^{-\xi} - 1) \right]^{-1} \right\} \\ &= \frac{1}{\xi(\delta+1)} (m_i - \mu_i^*) \left(1 + \frac{\xi c_2 c_1^{-\xi}}{1 - c_1^{-\xi}} \right). \end{aligned}$$

Adicionalmente,

$$\frac{dT_i}{d\xi} = \frac{dT_i}{d\mu_i^*} \frac{d\mu_i^*}{d\xi} = -\frac{1}{\xi} (m_i - \mu_i^*) |y_i - \mu_i^*|^\delta \left(1 + \frac{\xi c_2 c_1^{-\xi}}{1 - c_1^{-\xi}} \right),$$

e

$$\frac{d\Psi_i(\boldsymbol{\theta})}{d\xi} = \frac{1}{\sigma} \left[T_i + \xi \frac{dT_i}{d\xi} \right] = \frac{1}{\sigma} \left[T_i - (m_i - \mu_i^*) |y_i - \mu_i^*|^\delta \left(1 + \frac{\xi c_2 c_1^{-\xi}}{1 - c_1^{-\xi}} \right) \right].$$

Dessa forma,

$$\begin{aligned}
\frac{d\ell_i(\boldsymbol{\theta})}{d\xi} &= -\frac{\delta}{(y_i - \mu_i^*)} \frac{1}{\xi(\delta + 1)} (m_i - \mu_i^*) \left(1 + \frac{\xi c_2 c_1^{-\xi}}{1 - c_1^{-\xi}} \right) + \frac{1}{\xi^2} \log(\Psi_i(\boldsymbol{\theta})) [1 - \Psi_i^{-1/\xi}(\boldsymbol{\theta})] \\
&\quad - \frac{1}{\xi} \Omega_i(\boldsymbol{\theta}) \frac{1}{\sigma} \left[T_i - (m_i - \mu_i^*) |y_i - \mu_i^*|^\delta \left(1 + \frac{\xi c_2 c_1^{-\xi}}{1 - c_1^{-\xi}} \right) \right] \\
&= \frac{1}{\xi} (m_i - \mu_i^*) \left(1 + \frac{\xi c_2 c_1^{-\xi}}{1 - c_1^{-\xi}} \right) \left(\frac{|y_i - \mu_i^*|^\delta \Omega_i(\boldsymbol{\theta})}{\sigma} - \frac{\delta}{(\delta + 1)(y_i - \mu_i^*)} \right) \\
&\quad + \frac{1}{\xi^2} \log(\Psi_i(\boldsymbol{\theta})) [1 - \Psi_i^{-1/\xi}(\boldsymbol{\theta})] - \frac{T_i \Omega_i(\boldsymbol{\theta})}{\sigma \xi} \\
&= \frac{1}{\xi \sigma (\delta + 1)} \frac{(m_i - \mu_i^*)}{(y_i - \mu_i^*)} \left(1 + \frac{\xi c_2 c_1^{-\xi}}{1 - c_1^{-\xi}} \right) [(\delta + 1) T_i \Omega_i(\boldsymbol{\theta}) - \delta \sigma] \\
&\quad + \frac{1}{\xi^2} \log(\Psi_i(\boldsymbol{\theta})) [1 - \Psi_i^{-1/\xi}(\boldsymbol{\theta})] - \frac{T_i \Omega_i(\boldsymbol{\theta})}{\xi \sigma} \\
&= \frac{a_i}{\xi(\delta + 1)} (m_i - \mu_i^*) \left(1 + \frac{\xi c_2 c_1^{-\xi}}{1 - c_1^{-\xi}} \right) + \frac{1}{\xi^2} \log(\Psi_i(\boldsymbol{\theta})) [1 - \Psi_i^{-1/\xi}(\boldsymbol{\theta})] - \frac{T_i \Omega_i(\boldsymbol{\theta})}{\xi \sigma}
\end{aligned}$$

Portanto, fixando $b_i = d\ell_i(\boldsymbol{\theta})/d\xi$,

$$U_\xi = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \xi} = \mathbf{b}^\top \mathbf{1}_n,$$

em que \mathbf{b} e $\mathbf{1}_n$ são definidos na Subseção 4.2.

Diferenciando $\ell(\boldsymbol{\theta})$ com relação a σ :

$$U_\sigma = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma} = \sum_{i=1}^n \frac{d\ell_i(\boldsymbol{\theta})}{d\sigma}.$$

Para $\xi \neq 0$,

$$\begin{aligned}
\frac{d\ell_i(\boldsymbol{\theta})}{d\sigma} &= -\frac{1}{\sigma} - \frac{\delta \operatorname{sgn}(y_i - \mu_i^*)}{|y_i - \mu_i^*|} \frac{d\mu_i^*}{d\sigma} - \left(1 + \frac{1}{\xi} \right) \frac{1}{\Psi_i(\boldsymbol{\theta})} \frac{d\Psi_i(\boldsymbol{\theta})}{d\sigma} + \frac{1}{\xi} \Psi_i^{-1/\xi-1}(\boldsymbol{\theta}) \frac{d\Psi_i(\boldsymbol{\theta})}{d\sigma} \\
&= -\frac{1}{\sigma} - \frac{\delta}{(y_i - \mu_i^*)} \frac{d\mu_i^*}{d\sigma} - \frac{1}{\xi} \frac{d\Psi_i(\boldsymbol{\theta})}{d\sigma} \Omega_i(\boldsymbol{\theta}),
\end{aligned}$$

com

$$\frac{d\mu_i^*}{d\sigma} = -\frac{1}{(\delta + 1)} \left[\frac{\sigma}{\xi} (c_1^{-\xi} - 1) \right]^{1/(\delta+1)-1} \frac{1}{\xi} (c_1^{-\xi} - 1) = -\frac{1}{\sigma(\delta + 1)} \left[\frac{\sigma}{\xi} (c_1^{-\xi} - 1) \right]^{1/(\delta+1)} = -\frac{(m_i - \mu_i^*)}{\sigma(\delta + 1)},$$

desde que $\left[\frac{\sigma}{\xi}(c_1^{-\xi} - 1)\right]^{1/(\delta+1)} = m_i - \mu_i^*$, e

$$\frac{d\Psi_i(\boldsymbol{\theta})}{d\sigma} = \xi \left[-\frac{1}{\sigma^2} T_i + \frac{1}{\sigma} \frac{dT_i}{d\sigma} \right] = -\frac{\xi}{\sigma^2} (y_i - m_i) |y_i - \mu_i^*|^\delta.$$

Para $\xi = 0$,

$$\frac{d\ell_i(\boldsymbol{\theta})}{d\sigma} = -\frac{1}{\sigma} - \frac{\delta \operatorname{sgn}(y_i - \mu_i^*)}{|y_i - \mu_i^*|} \frac{d\mu_i^*}{d\sigma} + \frac{T_i}{\sigma^2} - \frac{1}{\sigma} \frac{dT_i}{d\sigma} - e^{-T_i/\sigma} \left(\frac{T_i}{\sigma^2} - \frac{1}{\sigma} \frac{dT_i}{d\sigma} \right),$$

com

$$\frac{d\mu_i^*}{d\sigma} = -\frac{c_2}{(\delta+1)} (c_2\sigma)^{1/(\delta+1)-1} = -\frac{(c_2\sigma)^{1/(\delta+1)}}{\sigma(\delta+1)} = -\frac{(m_i - \mu_i^*)}{\sigma(\delta+1)},$$

desde que $(c_2\sigma)^{1/(\delta+1)} = m_i - \mu_i^*$, e

$$\frac{dT_i}{d\sigma} = \frac{dT_i}{d\mu_i^*} \frac{d\mu_i^*}{d\sigma} = |y_i - \mu_i^*|^\delta (\delta+1) \frac{(m_i - \mu_i^*)}{\sigma(\delta+1)} = \frac{1}{\sigma} (m_i - \mu_i^*) |y_i - \mu_i^*|^\delta.$$

Assim,

$$\begin{aligned} \frac{d\ell_i(\boldsymbol{\theta})}{d\sigma} &= -\frac{1}{\sigma} + \frac{\delta}{\sigma(\delta+1)} \frac{(m_i - \mu_i^*)}{(y_i - \mu_i^*)} + \frac{T_i}{\sigma^2} - \frac{1}{\sigma^2} (m_i - \mu_i^*) |y_i - \mu_i^*|^\delta \\ &\quad - e^{-T_i/\sigma} \left[\frac{T_i}{\sigma^2} - \frac{1}{\sigma^2} (m_i - \mu_i^*) |y_i - \mu_i^*|^\delta \right] \\ &= -\frac{1}{\sigma} + \frac{\delta}{\sigma(\delta+1)} \frac{(m_i - \mu_i^*)}{(y_i - \mu_i^*)} + \frac{T_i}{\sigma^2} \frac{(1 - e^{-T_i/\sigma})}{(y_i - \mu_i^*)} [y_i - \mu_i^* - (m_i - \mu_i^*)] \\ &= -\frac{1}{\sigma} + \frac{\delta}{\sigma(\delta+1)} \frac{(m_i - \mu_i^*)}{(y_i - \mu_i^*)} + \frac{1}{\sigma^2} \Omega_i(\boldsymbol{\theta}) |y_i - \mu_i^*|^\delta (y_i - m_i). \end{aligned}$$

Fixando $d_i = d\ell_i(\boldsymbol{\theta})/d\sigma$,

$$U_\sigma = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma} = \mathbf{d}^\top \mathbf{1}_n,$$

em que \mathbf{d} é definido na Subseção 4.2.

Finalmente, diferenciando $\ell(\boldsymbol{\theta})$ em relação a δ :

$$U_\delta(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \delta} = \sum_{i=1}^n \frac{d\ell_i(\boldsymbol{\theta})}{d\delta}.$$

Para $\xi \neq 0$,

$$\begin{aligned}
 \frac{d\ell_i(\boldsymbol{\theta})}{d\delta} &= \frac{1}{\delta+1} + \log |y_i - \mu_i^*| - \frac{\delta \operatorname{sgn}(y_i - \mu_i^*)}{|y_i - \mu_i^*|} \frac{d\mu_i^*}{d\delta} - \left(1 + \frac{1}{\xi}\right) \frac{1}{\Psi_i(\boldsymbol{\theta})} \frac{d\Psi_i(\boldsymbol{\theta})}{d\delta} \\
 &+ \frac{1}{\xi} \Psi_i^{-1/\xi-1}(\boldsymbol{\theta}) \frac{d\Psi_i(\boldsymbol{\theta})}{d\delta} \\
 &= \frac{1}{\delta+1} + \log |y_i - \mu_i^*| - \frac{\delta}{(y_i - \mu_i^*)} \frac{d\mu_i^*}{d\delta} - \frac{1}{\xi} \frac{d\Psi_i(\boldsymbol{\theta})}{d\delta} \Omega_i(\boldsymbol{\theta}),
 \end{aligned}$$

com

$$\begin{aligned}
 \frac{d\mu_i^*}{d\delta} &= \frac{d}{d\delta} \left\{ m_i - \exp \left\{ \frac{1}{\delta+1} \log \left[\frac{\sigma}{\xi} (c_1^{-\xi} - 1) \right] \right\} \right\} \\
 &= - \left[\frac{\sigma}{\xi} (c_1^{-\xi} - 1) \right]^{1/(\delta+1)} \log \left[\frac{\sigma}{\xi} (c_1^{-\xi} - 1) \right] \\
 &= \frac{1}{\delta+1} (m_i - \mu_i^*) \log(m_i - \mu_i^*),
 \end{aligned}$$

$$\frac{d}{d\delta} [\Psi_i(\boldsymbol{\theta})] = \frac{\xi}{\sigma} \frac{dT_i}{d\delta},$$

e

$$\begin{aligned}
 \frac{dT_i}{d\delta} &= -|y_i - \mu_i^*|^\delta \frac{d\mu_i^*}{d\delta} + (y_i - \mu_i^*) \frac{d}{d\delta} \left\{ e^{\delta \log |y_i - \mu_i^*|} \right\} \\
 &= -|y_i - \mu_i^*|^\delta \frac{d\mu_i^*}{d\delta} + (y_i - \mu_i^*) \left\{ |y_i - \mu_i^*|^\delta \left[\log |y_i - \mu_i^*| - \delta \frac{\operatorname{sgn}(y_i - \mu_i^*)}{|y_i - \mu_i^*|} \frac{d\mu_i^*}{d\delta} \right] \right\} \\
 &= -|y_i - \mu_i^*|^\delta \frac{d\mu_i^*}{d\delta} + T_i \left[\log |y_i - \mu_i^*| - \frac{\delta}{(y_i - \mu_i^*)} \frac{d\mu_i^*}{d\delta} \right] \\
 &= T_i \log |y_i - \mu_i^*| - \frac{d\mu_i^*}{d\delta} \left(|y_i - \mu_i^*|^\delta + \frac{\delta T_i}{(y_i - \mu_i^*)} \right) \\
 &= T_i \log |y_i - \mu_i^*| - \frac{d\mu_i^*}{d\delta} \frac{T_i(\delta+1)}{(y_i - \mu_i^*)}.
 \end{aligned}$$

Isso implica em

$$\begin{aligned}
 \frac{d}{d\delta} [\Psi_i(\boldsymbol{\theta})] &= \frac{\xi}{\sigma} \left[T_i \log |y_i - \mu_i^*| - \frac{d\mu_i^*}{d\delta} \frac{T_i(\delta+1)}{(y_i - \mu_i^*)} \right] \\
 &= \frac{\xi}{\sigma} \left[T_i \log |y_i - \mu_i^*| - |y_i - \mu_i^*|^\delta (m_i - \mu_i^*) \log(m_i - \mu_i^*) \right].
 \end{aligned}$$

Assim,

$$\begin{aligned}
\frac{d\ell_i(\boldsymbol{\theta})}{d\delta} &= \frac{1}{\delta+1} + \log |y_i - \mu_i^*| - \frac{\delta}{(y_i - \mu_i^*)} \frac{d\mu_i^*}{d\delta} - \frac{1}{\xi} \frac{d\Psi_i(\boldsymbol{\theta})}{d\delta} \Omega_i(\boldsymbol{\theta}) \\
&= \frac{1}{\delta+1} + \log |y_i - \mu_i^*| - \frac{\delta}{(\delta+1)(y_i - \mu_i^*)} (m_i - \mu_i^*) \log(m_i - \mu_i^*) \\
&\quad - \frac{\Omega_i(\boldsymbol{\theta})}{\sigma} [T_i \log |y_i - \mu_i^*| - |y_i - \mu_i^*|^\delta (m_i - \mu_i^*) \log(m_i - \mu_i^*)] \\
&= \frac{1}{\delta+1} + \log |y_i - \mu_i^*| \left(1 - \frac{T_i \Omega_i(\boldsymbol{\theta})}{\sigma} \right) \\
&\quad - (m_i - \mu_i^*) \log(m_i - \mu_i^*) \left[\frac{\delta}{(\delta+1)(y_i - \mu_i^*)} - \frac{|y_i - \mu_i^*|^\delta \Omega_i(\boldsymbol{\theta})}{\sigma} \right] \\
&= \frac{1}{\delta+1} + \log |y_i - \mu_i^*| \left(1 - \frac{T_i \Omega_i(\boldsymbol{\theta})}{\sigma} \right) + \frac{(m_i - \mu_i^*) \log(m_i - \mu_i^*)}{\sigma(\delta+1)(y_i - \mu_i^*)} [(\delta+1)T_i \Omega_i(\boldsymbol{\theta}) - \delta\sigma] \\
&= \log |y_i - \mu_i^*| \left(1 - \frac{T_i \Omega_i(\boldsymbol{\theta})}{\sigma} \right) + \frac{1}{\delta+1} [1 + a_i(m_i - \mu_i^*) \log(m_i - \mu_i^*)].
\end{aligned}$$

Para $\xi = 0$,

$$\frac{d\ell_i(\boldsymbol{\theta})}{d\delta} = \frac{1}{\delta+1} + \log |y_i - \mu_i^*| - \frac{\delta}{(y_i - \mu_i^*)} \frac{d\mu_i^*}{d\delta} - \frac{1}{\sigma} \frac{dT_i}{d\delta} \Omega_i(\boldsymbol{\theta}),$$

com

$$\begin{aligned}
\frac{d\mu_i^*}{d\delta} &= \frac{d}{d\delta} \left\{ m_i - \exp \left\{ \frac{1}{\delta+1} \log(c_2\sigma) \right\} \right\} \\
&= \frac{1}{\delta+1} (c_2\sigma)^{1/(\delta+1)} \log [(c_2\sigma)^{1/(\delta+1)}] \\
&= \frac{1}{\delta+1} (m_i - \mu_i^*) \log(m_i - \mu_i^*).
\end{aligned}$$

Assim,

$$\begin{aligned}
\frac{d\ell_i(\boldsymbol{\theta})}{d\delta} &= \frac{1}{\delta+1} + \log |y_i - \mu_i^*| - \frac{\delta}{(y_i - \mu_i^*)} \frac{d\mu_i^*}{d\delta} - \frac{1}{\sigma} \left[T_i \log |y_i - \mu_i^*| - \frac{d\mu_i^*}{d\delta} \frac{T_i(\delta+1)}{(y_i - \mu_i^*)} \right] \Omega_i(\boldsymbol{\theta}) \\
&= \frac{1}{\delta+1} + \log |y_i - \mu_i^*| \left(1 - \frac{T_i \Omega_i(\boldsymbol{\theta})}{\sigma} \right) \\
&\quad - (m_i - \mu_i^*) \log(m_i - \mu_i^*) \left[\frac{\delta}{(\delta+1)(y_i - \mu_i^*)} - \frac{|y_i - \mu_i^*|^\delta \Omega_i(\boldsymbol{\theta})}{\sigma} \right] \\
&= \log |y_i - \mu_i^*| \left(1 - \frac{T_i \Omega_i(\boldsymbol{\theta})}{\sigma} \right) + \frac{1}{\delta+1} [1 + a_i(m_i - \mu_i^*) \log(m_i - \mu_i^*)].
\end{aligned}$$

Fixando $e_i = d\ell_i(\boldsymbol{\theta})/d\delta$,

$$U_\delta = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \delta} = \mathbf{e}^\top \mathbf{1}_n,$$

em que \mathbf{e} é definido na Subseção 4.2.