



Universidade de Brasília
Departamento de Estatística

Modelo de Regressão bimodal para valores extremos

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
2023

Mathews de Noronha Silveira Lisboa




Modelo de Regressão bimodal para valores extremos

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
2023**

Resumo

Neste trabalho é introduzido o modelo regressão bimodal  GEV ao  modificar o modelo proposto por Otiniano et al. (2023).  A flexibilidade do modelo foi ilustrada graficamente uma vez que a funções de densidade de probabilidade e de risco apresentam uma variedade de formas. O método escolhido para a estimação dos parâmetros do modelo é o de máxima verossimilhança.

Palavras-chave: GEV; Bimodal; Regressão; Máxima verossimilhança

Sumário

1 Introdução	4
2 Revisão de Literatura	6
2.1 Distribuições extremas e GEV	6
2.2 GEV bimodal	9
3 Metodologia	17
3.1 Regressão	17
3.2 Estimação	18
4 Possíveis Aplicações	20
5 Considerações Finais	22

1 Introdução

A teoria dos valores extremos (TVE) trata do estudo assintótico das distribuições de estatísticas extremas, isto é, trata das 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 podem ser Gumbel, Fréchet ou Weibull. Referências clássicas sobre TVE são os livros de Kotz e Nadarajah (2000), Haan e Ferreira (2006), e Resnick (1987). A família de distribuições de probabilidade contínua desenvolvida para combinar essas três famílias foi denominada, por Jenkinson (1955), de GEV do inglês "generalized extreme value". Desde então a distribuição GEV tem sido amplamente aceita em diversas áreas como em finanças (Embrechts, 1997), confiabilidade (Samorodnitsky, 2014), e ciências da terra e meio ambiente (Raynal Villaseñor, 2021), entre outras.

A distribuição GEV acomoda distribuições com cauda pesada, cauda leve e é unimodal. Diversas generalizações da distribuição GEV foram propostas com a finalidade de capturar maior assimetria e possuir caudas mais pesadas, entre elas temos GEV transmutada de Nascimento et al. (2016) e estudada por Otiniano et al. (2019). Eventos extremos combinados com eventos raros, por exemplo velocidade do vento, despertaram o interesse de obter uma distribuição bimodal. Recentemente, Otiniano et al. (2023) propuseram uma nova generalização da distribuição GEV que pode ser bimodal e a denominaram de distribuição GEV bimodal.

Por outro lado, a regressão é uma ferramenta essencial para a estatística, pois possibilita encontrar relações entre variáveis explicativas e variáveis respostas. Poucas propostas de modelos de regressão no contexto de dados bimodais ainda são conhecidas. Ao usar a técnica Marshall-Olkin-odd-loglogistic-G (MOOLL-G) é possível obter uma nova distribuição bimodal. Cordeiro et al. (2021) apresentaram uma nova regressão log-Weibull estendida com base no modelo MOOLL-Weibull. Outros modelos de regressão para distribuições bimodais são de Vasconcelos et al. (2021a), Vasconcelos et al. (2021b), e Vila et al. (2021).

Neste trabalho é introduzido uma nova regressão GEV bimodal (EGEVB) útil para análise de dados de eventos extremos. O *software* estatístico R (R Core Team, 2022) em sua versão 4.3.3 de 2022 é utilizado para implementar as funções do modelo EGEVB e a estimação de seus parâmetros.

Este trabalho está organizado da seguinte forma. No Capítulo 2 é apresentado a revisão de literatura na qual é feito um levantamento das principais técnicas e teorias que serão utilizadas no decorrer do trabalho. No Capítulo 3 é apresentado a metodologia que é utilizada no trabalho, como o modelo de regressão EGEVB, a função de verossimilhança.

Por fim, **é feito** o Capítulo 4 é feito as considerações finais **que** é apresentado os passos que ainda serão realizados.

2 Revisão de Literatura

Há um interesse na estatística do estudo dos valores de máximo e mínimo de alguma determinada amostra ou população, esse ramo do conhecimento na teoria de estatística é denominado como teoria dos valores extremos. Nessa seção será apresentado a teoria de valores extremos e alguns de seus principais resultados como as distribuições extremas padrão e distribuição generalizada de valores extremos (GEV) em 2.1, além da função de uma transformação realizada na distribuição GEV introduzida na sub-seção 2.2 que será utilizada no modelo de regressão mais adiante.

2.1 Distribuições extremas e GEV

A primeira vez que distribuições para valores extremos foi proposta por Fisher e Tippett (1928) que apresentou o teorema do limite para distribuições extremas que foi provado pela primeira vez por Gnedenko (1943). O principal resultado do teorema foi a definição das três distribuições de valores extremo padrão que são escritas como:

$$\text{Weibull} : \Psi_{\alpha} = \begin{cases} \exp(-(-y)^{\alpha}), & \text{se } y \leq 0, \alpha > 0, \\ 1, & \text{se } y < 0 \end{cases}$$

$$\text{Gumbel} : \nu_{\alpha} = \exp(-\exp(-y)), \text{ para } y \in \mathbb{R},$$

$$\text{Fréchet} : \Phi_{\alpha} = \begin{cases} 0, & \text{se } y \leq 0, \\ \exp(-y^{-\alpha}), & \text{se } y > 0, \alpha > 0. \end{cases}$$

Na Figura 1 pode-se observar a forma que densidade de probabilidade das distribuições se comportam. É notável a diferença entre as três distribuições, considerando que a cauda da Gumbel possui um decaimento leve, ou suave, semelhante a normal, a Weibull é finita e por fim a Fréchet tem o decaimento de cauda pesada, sendo essa última mais amplamente utilizada em aplicações financeiras conforme Jondeau, Poon e Rockinger (2007). Por outro lado, observa-se que a Weibull é uma função monótona para valores acima de zero enquanto esse comportamento é invertido para a Fréchet.

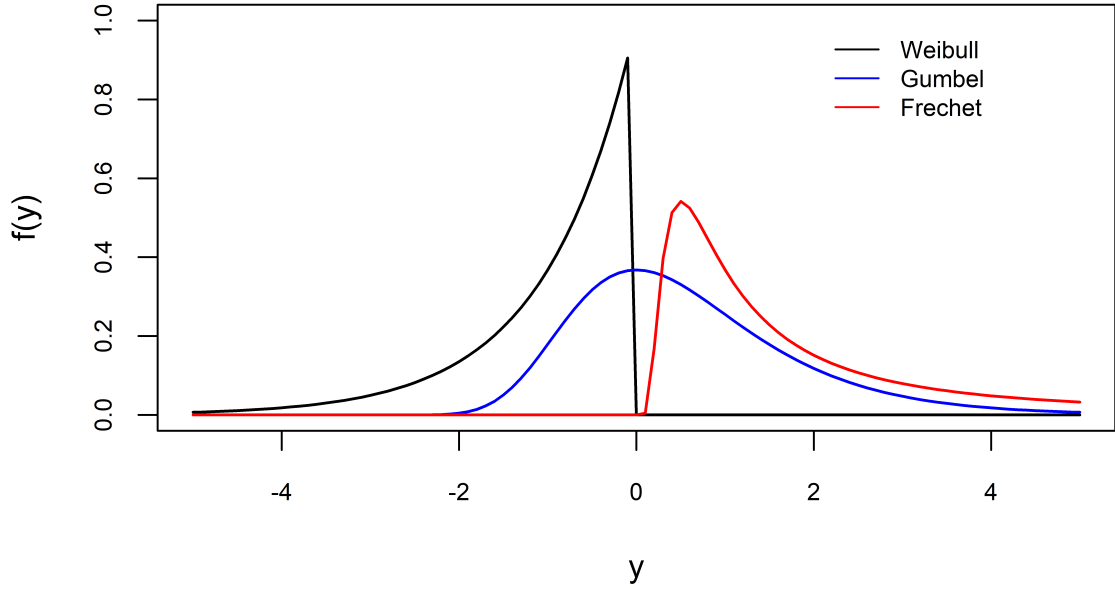


Figura 1: Função de densidade de probabilidade para as distribuições de valores extremos (com $\alpha = 1$)

Depois do desenvolvimento das distribuições extremas padrão, foi notado a necessidade de se condensar as três distribuições. Esse feito ocorreu com o desenvolvimento da distribuição generaliza de valores extremos, conhecida como GEV (*Generalized Extreme Value*) que foi primeiramente apresentada por Jenkinson (1955). A forma da função densidade de probabilidade GEV está escrita na equação 2.1.1.

$$f(y, \xi, \sigma, \mu) = \begin{cases} \frac{1}{\sigma} \left[1 + \xi \left(\frac{y - \mu}{\sigma} \right) \right]^{(-1/\xi)-1} \exp \left\{ - \left[1 + \xi \left(\frac{y - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}, & \text{se } \xi \neq 0 \\ \frac{1}{\sigma} \exp \left\{ - \left(\frac{y - \mu}{\sigma} \right) - \exp \left[- \left(\frac{y - \mu}{\sigma} \right) \right] \right\}, & \text{se } \xi = 0, \end{cases} \quad (2.1.1)$$

e a função de distribuição acumulada é escrita da seguinte forma:

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

que quando $\xi \neq 0$ é definida para y tal que $1 + \xi \left(\frac{y - \mu}{\sigma} \right) > 0$ e quando $\xi = 0$ definida em toda a reta real (\mathbb{R}). O parâmetro de forma ξ conhecido como índice caudal tem relação

direta com o parâmetro de forma das três distribuições extremas. A relação é a seguinte

$$\xi = \alpha^{-1} > 0 \text{ Para a Fréchet}$$

$$\xi = 0 \text{ Para Gumbel}$$

$$\xi = -\alpha^{-1} < 0 \text{ Para a Weibull}$$

Os demais parâmetros são σ de escala, σ que quanto mais o valor aumenta maior é a comprimento da densidade, assemelhando-se muito assim a variância. Em seguida, o parâmetro de localização μ , faz uma mudança lateral na densidade pelo mesmo valor do parâmetro. Na Figura 2 mostra-se a mudança do parâmetro de forma ξ para diferentes valores é notável que conforme os valores de forma mudam entre os valores de -1, 0 e 1 as curvas tornam-se similares as da Figura 1 na imagem superior a direita, que representa a densidade da GEV, na imagem superior a esquerda encontra-se a função de probabilidade acumulada, na imagem inferior a direita a função de sobrevivência e na imagem inferior a esquerda a função de risco.

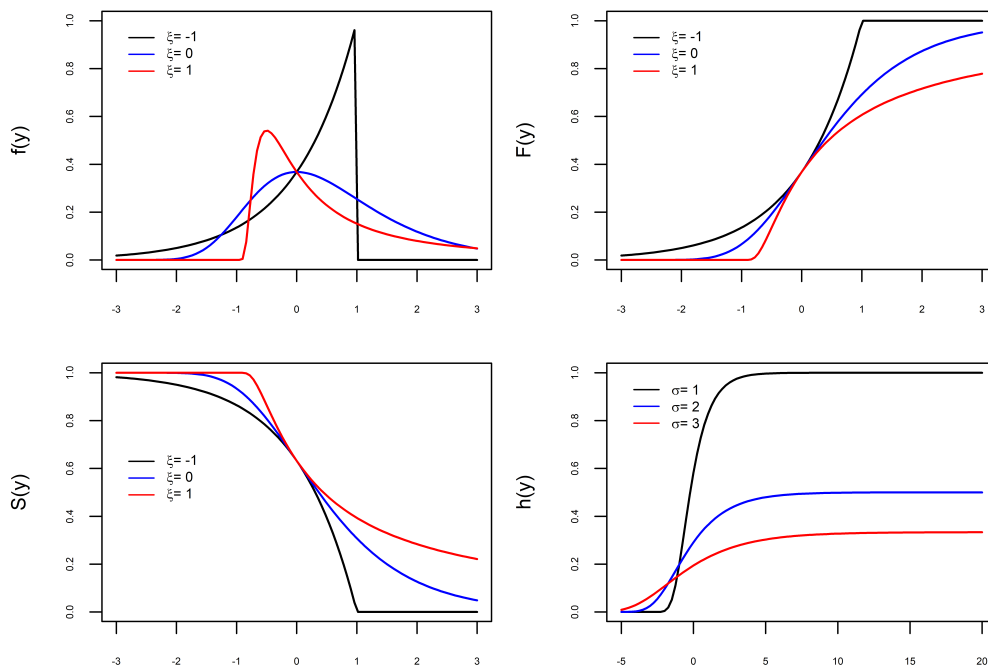


Figura 2: $Y \sim GEV$ e ξ variando. Painéis superiores FDP (esquerda) e FDA (direita), painéis inferiores função e sobrevivência (esquerda) e Função de risco (direita).

Desde que foi formalizada pela primeira vez a GEV tem sido aplicada em diversas áreas pode-se observar a aplicação teórica da teoria dos valores extremos e também da GEV em livros como Haan e Ferreira (2006), Kotz e Nadarajah (2000), Rudd et al. (2017) e entre outros mais. Apesar da flexibilidade da distribuição GEV para modelar eventos extremos, ela não contempla comportamentos heterogêneos (bimodais) de dados

que aparecem em **diversa** áreas como hidrologia, marketing, finanças, entre outros. Neste sentido, algumas generalizações da GEV tem surgido. **na** seguinte subseção apresentamos a recente generalização de Otiniano et al. (2021), chamada de distribuição GEV bimodal.

2.2 GEV bimodal

Para definir novos modelos bimodais há pelo menos duas metodologias utilizando uma **distribuição base**. Uma delas é utilizar uma transformação quadrática em uma distribuição base F . Alguns trabalhos que usaram essa metodologia são para **Normal** (Elal-Olivero (2010)), Birnbaum-Saunders (Fonseca e Cribari-Neto (2018); Vila et al. (2021)) e Gamma (Vila et al. (2020)). **Já usando** uma transformação do tipo $T(x) = (x - \mu)|x - \mu|^\delta$ **te-se** os trabalhos para log-logística (Swamee e Rathie (2007)) e para **GEV** (Otiniano et al. (2021)).

Nesta seção detalhamos o modelo GEV bimodal obtido por Otiniano et al. (2023). O procedimento consiste em compor a distribuição de uma variável **aleatório** $X \sim GEV(x, \mu, 1, \xi)$, que é unimodal com a transformação T . Assim a distribuição BGEV é dada por



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

sendo que a função $T_{\mu, \sigma}(y)$ é definida **como**:

$$T_{\mu, \sigma}(y) = (y - \mu) |y - \mu|^\delta, \delta > -1, \mu \in \mathbb{R} \quad (2.2.2)$$

além disso a função T é inversível, com a inversa dada por:

$$T^{-1}(y)_{\mu, \delta} = \text{sgn}(y) |y|^{\frac{1}{\delta+1}} + \mu \quad (2.2.3)$$

E também derivável sendo que a derivada tem a seguinte forma:

$$T'_{\mu, \delta}(y) = (\delta + 1) |y - \mu|^\delta \quad (2.2.4)$$

Considere então uma união da equação 2.1.2 e 2.2.2 conforme o indicado em 2.2.1, **desse modo temos** uma forma explícita da função de distribuição acumulada da BGEV **de forma explícita como** :

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

e sua função de densidade de probabilidade correspondente como:

$$f_{BG_{\xi,\mu,\sigma,\delta}}(y) = \begin{cases} \frac{1}{\sigma} \left[1 + \xi \left(\frac{T_{\mu,\delta}(y)}{\sigma} \right) \right]^{(-1/\xi)-1} \exp \left[- \left[1 + \xi \left(\frac{T_{\mu,\delta}(y)}{\sigma} \right) \right]^{-1/\xi} \right] T'_{\mu,\delta}(y), & \xi \neq 0 \\ \frac{1}{\sigma} \exp \left(- \frac{T_{\mu,\delta}(y)}{\sigma} \right) \exp \left[- \exp \left(- \frac{T_{\mu,\delta}(y)}{\sigma} \right) \right] T'_{\mu,\delta}(y), & \xi = 0, \end{cases} \quad (2.2.6)$$

sendo que $T_{\mu,\delta}(y)$ definido em 2.2.2, $T'_{\mu,\delta}$ como definido em 2.2.4 e o suporte de f_{BG} é:

$$\text{Suporte}(f_{BG_{\xi,\mu,\sigma,\delta}}) = \begin{cases} \left[\mu - \left(\frac{\sigma}{\xi} \right)^{\frac{1}{\delta+1}}, +\infty \right), & \text{se } \xi > 0, \\ \left(-\infty, \mu + \left| \frac{\sigma}{\xi} \right|^{\frac{1}{\delta+1}} \right], & \text{se } \xi < 0, \\ (-\infty, +\infty), & \text{se } \xi = 0. \end{cases} \quad (2.2.7)$$

Com as funções de densidade de probabilidade e distribuição acumulada pode-se obter por meio de relações simples outras duas funções que são de interesse **nesse trabalho para entender** as possibilidades de aplicação do modelo que será proposto com BGEV definida em 2.2.6.

Sendo assim, calcula-se a função de sobrevivência e também a função de risco para o caso particular da BGEV considerando a função de distribuição acumulada 2.2.5 encontra-se a função de sobrevivência por meio da **propriedade:**

$$\begin{aligned} S(y) &= P(Y > y), \text{ então} \\ S(y) &= \begin{cases} 1 - \exp \left[- \left[1 + \xi \left(\frac{T_{\mu,\delta}(y)}{\sigma} \right) \right]^{-1/\xi} \right], & \xi \neq 0 \\ 1 - \exp \left[- \exp \left[- \frac{T_{\mu,\delta}(y)}{\sigma} \right] \right], & \xi = 0 \end{cases} \end{aligned}$$

Além disso, pode-se encontrar a função de risco com a relação entre função de sobrevivência e função de densidade acumulada com a seguinte relação:

$$h(y) = \frac{f(y)}{S(y)}$$

Desse modo separando para o caso de $\xi \neq 0$ podemos ter:

$$h(y) = \frac{\frac{1}{\sigma} \left[1 + \xi \left(\frac{T_{\mu,\delta}(y)}{\sigma} \right) \right]^{(-1/\xi)-1} \exp \left[- \left[1 + \xi \left(\frac{T_{\mu,\delta}(y)}{\sigma} \right) \right]^{-1/\xi} \right] T'_{\mu,\delta}(y)}{1 - \exp \left[- \left[1 + \xi \left(\frac{T_{\mu,\delta}(y)}{\sigma} \right) \right]^{-1/\xi} \right]}$$

de forma análoga, pode escrever para o caso de $\xi = 0$ da seguinte forma:

$$h(y) = \frac{\exp \left(-\frac{T_{\mu,\delta}(y)}{\sigma} \right) \exp \left[-\exp \left(-\frac{T_{\mu,\delta}(y)}{\sigma} \right) \right] T'_{\mu,\delta}(y)}{1 - \exp \left[-\exp \left[-\frac{T_{\mu,\delta}(y)}{\sigma} \right] \right]}$$

Para melhor entendimento da influência dos parâmetros foi feito gráficos variando cada um deles individualmente enquanto os demais permanecem fixos, dessa forma pode-se definir o comportamento da função definida em 2.2.6 e observar casos nos quais esse comportamento ocorre.

Nas Figuras 3, 4 e 5 foi feito os gráficos da função de densidade que foi definida na equação, função de distribuição acumulada, função de sobrevivência e função de risco da BGEV tal qual definidas na equação 2.2.6. No caso dessas Figuras em específico foi utilizado o parâmetro $\xi = 0$ para todas as 3, enquanto que os demais parâmetros variam em diferentes imagens. No caso da imagem, 3 o que ocorre é a variação de μ enquanto $\sigma = 1$ e $\delta = 1$, no caso da Figura 4 o σ varia enquanto $\mu = 0$ e $\delta = 1$ e por fim no caso da Figura 5 o δ varia enquanto $\mu = 0$ e $\sigma = 1$.

Para o caso de valores de $\xi \neq 0$ é preciso antes fazer uma análise primeiramente do impacto dos valores de ξ na forma das funções que está representado na Figura 6 no qual ξ varia enquanto $\mu = 0$, $\sigma = 1$ e $\delta = 1$. É de interesse então analisar os casos em que $\xi < 0$ e $\xi > 0$ para diferentes valores de δ para observar o comportamento que as funções tem quando dois parâmetros de forma agem.

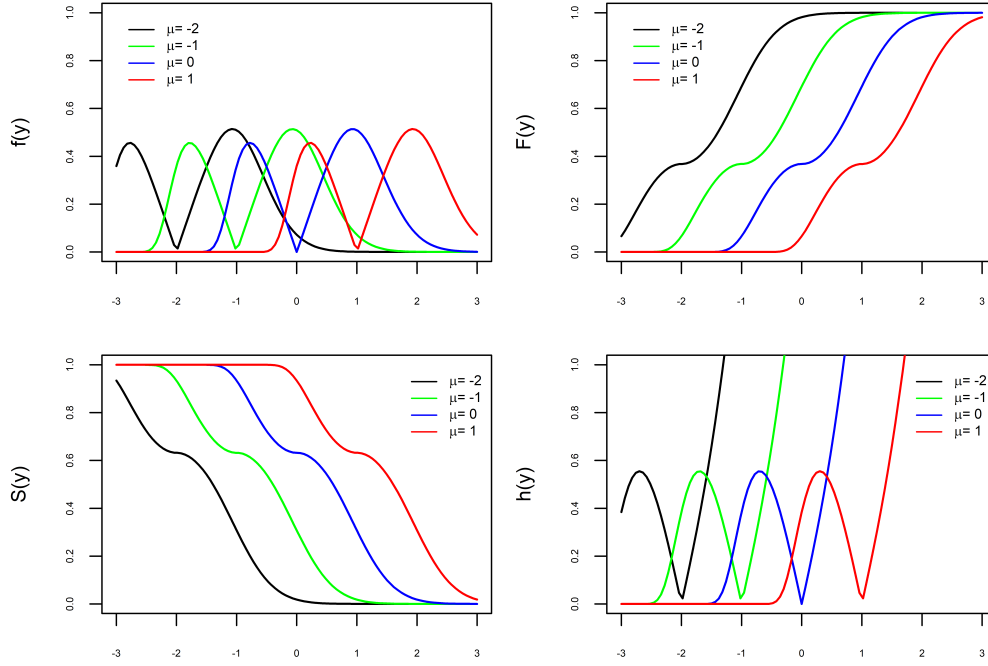


Figura 3: $Y \sim F_{BG_{0,\mu,1,1}}$ e μ variando. Painéis superiores FDP (esquerda) e FDA (direita), painéis inferiores função e sobrevivência (esquerda) e Função de risco (direita).

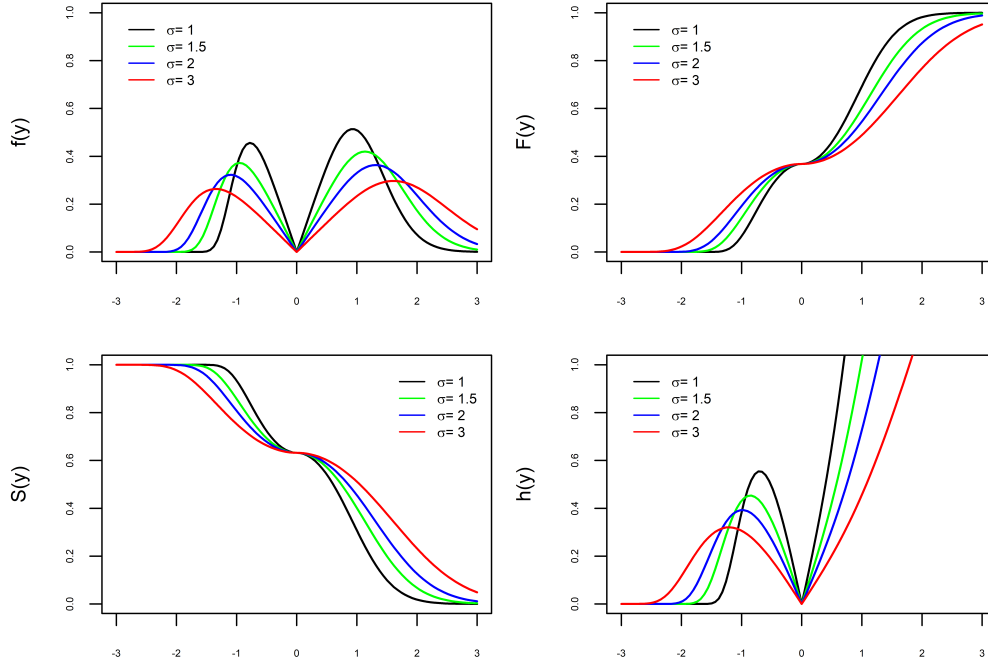


Figura 4: $Y \sim F_{BG_{0,0,\sigma,1}}$ e σ variando. Painéis superiores FDP (esquerda) e FDA (direita), painéis inferiores função e sobrevivência (esquerda) e Função de risco (direita).

Variando-se o parâmetro μ na Figura 3 percebe-se como um parâmetro de **locação** muito similar ao caso que ocorre com a distribuição GEV unimodal, uma vez que a densidade **realiza um deslocamento lateral** de mesma intensidade que o valor do parâmetro. Enquanto **o** para o parâmetro σ observa-se na Figura 4 que o a amplitude entre os valores

aumenta, similar a uma variância de dos dados.

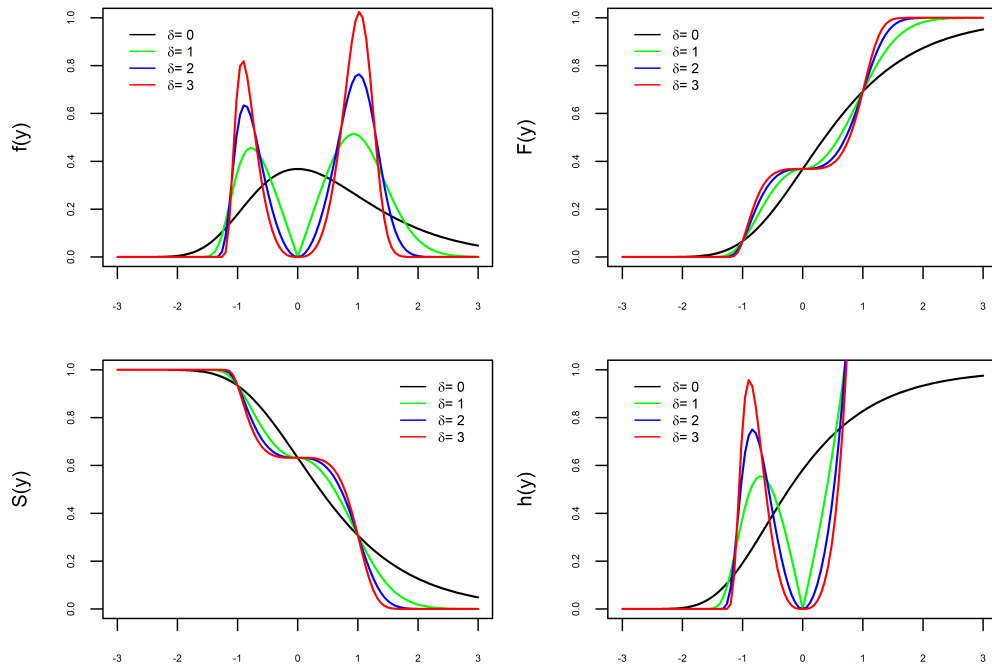


Figura 5: $Y \sim F_{BG_{0,0,1,\delta}}$ e δ variando. Painéis superiores FDP (esquerda) e FDA (direita), painéis inferiores função e sobrevivência (esquerda) e Função de risco (direita).

Nota-se na Figura 5 que o parâmetro de $\delta = 0$ torna a função a uma GEV unimodal e quanto maior o valor de δ mais distantes e proeminentes são as duas modas da função BGEV. Nota-se também a forma de "N" que a função de risco toma conforme valores positivos de delta.

Na Figura 6 é apresentado as funções de densidade, probabilidade acumulada, sobrevivência e risco da BGEV com valores negativos e positivos de ξ . Dessa forma, pode-se observar que mantendo constante os demais parâmetros, valores muito absolutos muito altos de ξ podem estourar as modas, tornando a distribuição concentrada em suas modas, principalmente no caso negativo, uma vez que a função é limitada por um suporte a direita. No caso de valores positivos de ξ a função é limitada por um valor de suporte que depende dos parâmetros a esquerda e parece ter que a primeira moda possui uma concentração maior que a segunda, além disso, nota-se a diferença de formato da função de risco.

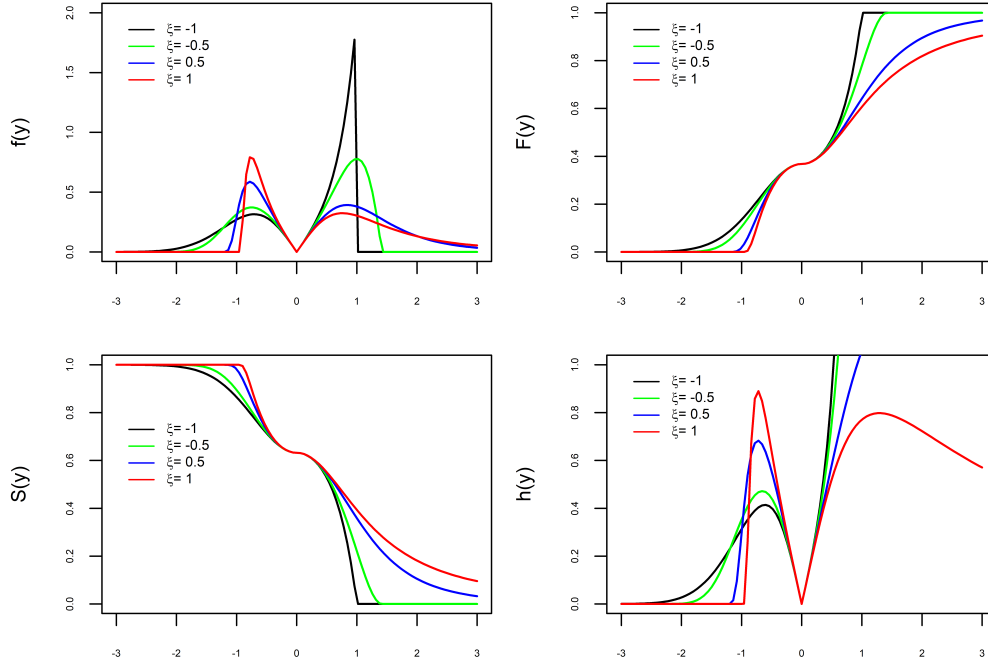


Figura 6: $Y \sim F_{BG_{\xi,0,1,1}}$ e ξ variando. Painéis superiores FDP (esquerda) e FDA (direita), painéis inferiores função e sobrevivência (esquerda) e Função de risco (direita).

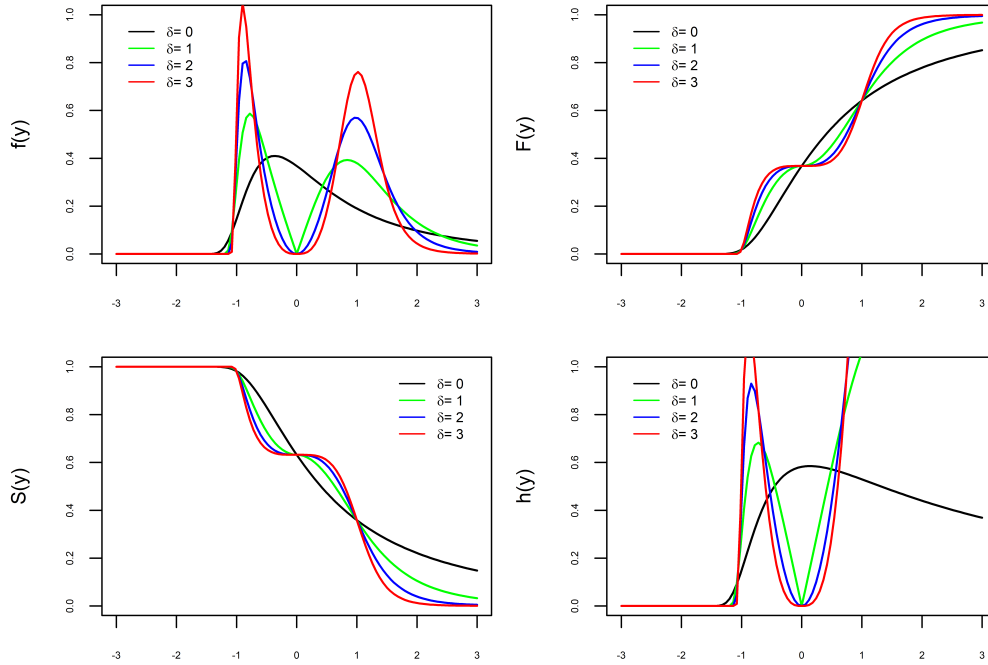


Figura 7: $Y \sim F_{BG_{0.5,0,1,\delta}}$ e δ variando. Painéis superiores FDP (esquerda) e FDA (direita), painéis inferiores função e sobrevivência (esquerda) e Função de risco (direita).

Como dito anteriormente, uma vez que visto o comportamento de diferentes valores de ξ é interessante analisar o comportamento de um $\xi \neq 0$ com diferentes valores de δ , desse modo, na Figura 7 é observado diferentes valores de δ para $\xi = 0.5$ enquanto que para 8 observa-se para $\xi = -0.5$. É interessante notar que para valores baixos de ξ a taxa

de risco tem o comportamento similar de "N", porém como visto na Figura 6 quando o valor positivo é cada vez mais alto percebe-se que a função de risco tende a zero.

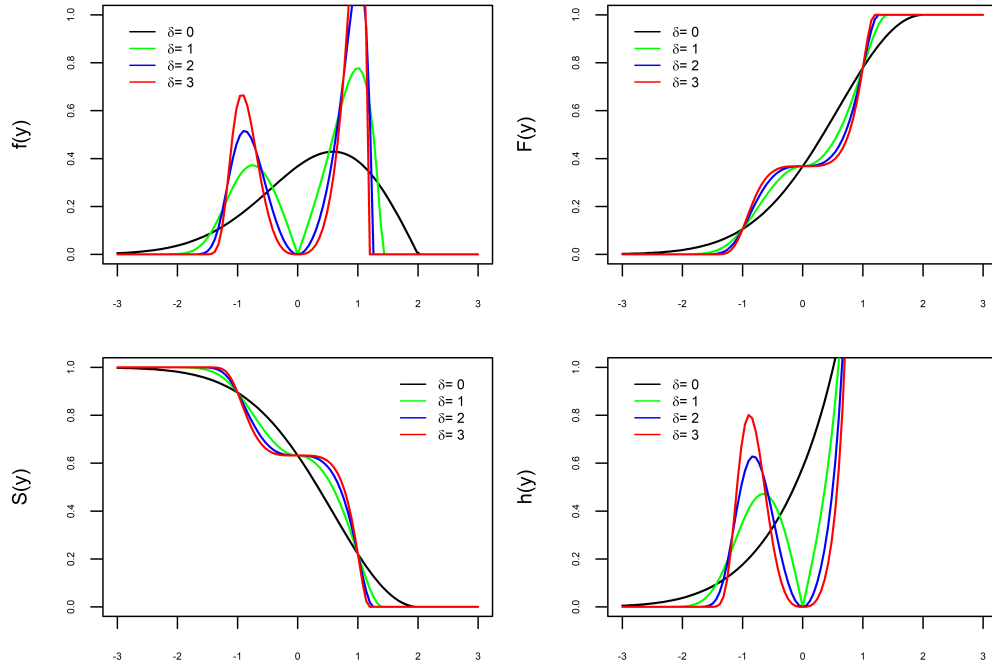


Figura 8: $Y \sim F_{BG_{-0.5,0,1,\delta}}$ e δ variando. Painéis superiores FDP (esquerda) e FDA (direita), painéis inferiores função e sobrevivência (esquerda) e Função de risco (direita).

Além das funções que foram apresentadas até então e exploradas nas imagens das Figuras de 3 a 8, também é importante explorar a função quantil pois **é fundamental para realizar a amostra pseudo aleatória da distribuição BGEV**. Para isso é preciso novamente separar os casos entre $\xi = 0$ e $\xi \neq 1$, por definição da função quantil pode ser encontrada pela inversa da função distribuição acumulada, então considere a função:

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

em que T^{-1} é a função inversa da função de transformação definida em 2.2.3 e G^{-1} é a função inversa da distribuição acumulada da GEV unimodal. Considerando então primeiramente para $\xi = 0$, é preciso primeiramente pensar na inversa da função inversa, sendo que pode ser escrita como:

$$G^{-1}(y) = -\sigma \ln(\ln(y^{-1})), \quad xi = 0. \quad (2.2.8)$$

Dessa forma, considerando 2.2.3 e 2.2.8, pode-se escrever a função quantil para a BGEV para $\xi = 0$ como:

$$Q_{\xi,\mu,\sigma,\delta}(x) = \text{sgn} \left(-\sigma \ln \left(\ln \left(y^{-1} \right) \right) \right) \left| -\sigma \ln \left(\ln \left(y^{-1} \right) \right) \right|^{\frac{1}{(\delta+1)}} + \mu \quad (2.2.9)$$

De forma análoga ao que foi feito para $\xi = 0$, pode-se pensar na função de distribuição acumulada inversa da GEV como:

$$G^{-1}(y) = \frac{\sigma}{\xi} (\ln y^{\xi} - 1), \xi \neq 0 \quad (2.2.10)$$

Sendo assim, pode-se considerar a função descrita em 2.2.10 e 2.2.3, para que a função quantil no caso de $\xi \neq 0$ pode ser descrita como:

$$Q_{\xi,\mu,\sigma,\delta}(y) = \text{sgn} \left(\frac{\sigma}{\xi} (\ln y^{\xi} - 1) \right) \left| \frac{\sigma}{\xi} (\ln y^{\xi} - 1) \right|^{\frac{1}{(\delta+1)}} + \mu \quad (2.2.11)$$

3 Metodologia

Nesta seção será apresentada os passos da metodologia aplicada para regressão, estimação dos parâmetros e simulação da distribuição apresentada na seção 2.

3.1 Regressão

Em diversos campos de estudo dentro da estatística é importante a verificação da relação entre duas ou mais variáveis. Esse tipo de relação pode ser encontrada por meio de uma regressão, a técnica no que é chamada dentro da teoria estatística consiste em determinar o quanto variáveis explicativas podem explicar a variação de uma variável resposta. A regressão é amplamente utilizada em diversos caso, porém, é mais comumente realizada em casos de distribuições unimodais, apesar disso estudos de regressão com distribuições bimodais podem ser encontrados na literatura, como é o caso de Vasconcelos et al. (2021b), que propôs um modelo de regressão sob distribuição de chances log-logística log-normal (OLLLN) com aplicação em geração de energia, Cordeiro et al. (2021) apresenta o modelo baseado na distribuição log-MOOLL-Weibull, Vasconcelos et al. (2021a) com um modelo OLLExGa aplicado em sementes na agricultura.

Considere que $\mathbf{x}^t = (1, x_1, \dots, x_p)$, seja o vetor de covariáveis e $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$ são os coeficientes de regressão, no caso da função de densidade da BGEV definida em 2.2.6 o parâmetro de localização é o μ tal qual é confirmado na Figura 3. Dessa forma, ao utilizar, como função de ligação a identidade $I(\cdot)$ no parâmetro μ , $\mu = \mathbf{x}^t \boldsymbol{\beta}$, em ambas as funções 2.2.2 e 2.2.4, tem-se que:

$$\begin{aligned} T_{\mu, \sigma, \beta}(y|\mathbf{x}) &= (y - \mathbf{x}^t \boldsymbol{\beta}) |y - \mathbf{x}^t \boldsymbol{\beta}|^\delta \\ T'_{\mu, \delta, \beta}(y|\mathbf{x}^t) &= (\delta + 1) |y - \mathbf{x}^t \boldsymbol{\beta}|^\delta. \end{aligned}$$

Sendo assim, a FDP dos modelos de regressão bimodal GEV (RBGEV) é definida por:

- Caso 1. Para $\xi = 0$,

$$f_{\mu, \sigma, \delta, \beta}(y|\mathbf{x}^t) = \frac{1}{\sigma} \exp\left(-\frac{T_{\mu, \delta, \beta}(y|\mathbf{x}^t)}{\sigma}\right) \exp\left[-\exp\left(-\frac{T_{\mu, \delta, \beta}(y|\mathbf{x}^t)}{\sigma}\right)\right] T'_{\mu, \delta, \beta}(y|\mathbf{x}^t). \quad (3.1.1)$$

- Caso 2. Para $\xi \neq 0$,

$$\begin{aligned}
f_{\xi,\mu,\sigma,\delta,\beta}(y|\mathbf{x}^t) &= \frac{1}{\sigma} \left[1 + \xi \left(\frac{T_{\mu,\delta,\beta}(y|\mathbf{x}^t)}{\sigma} \right) \right]^{-\frac{1}{\xi}-1} \\
&\times \exp \left\{ - \left[1 + \xi \left(\frac{T_{\mu,\delta,\beta}(y|\mathbf{x}^t)}{\sigma} \right) \right]^{-1/\xi} \right\} T'_{\mu,\delta}(y|\mathbf{x}^t). \quad (3.1.2)
\end{aligned}$$

3.2 Estimação

Dentro da área de conhecimento da estatística há diversas maneiras de se realizar a estimação de parâmetros, como por exemplo o estimador Bayesiano que já foi aplicado em casos com a GEV em Nascimento et al. (2016), Lima, Kwon e Kim (2018) e Lima et al. (2016). Porém ao se trabalhar com distribuições bimodais, nota-se que o mais usual é trabalhar com o estimador de máxima verossimilhança como visto em Vasconcelos et al. (2021a), Vasconcelos et al. (2021b) e Cordeiro et al. (2021). Por isso nesse trabalho foi feito a escolha de se realizar a estimação por meio do método de máxima verossimilhança.

Por definição, a função de máxima verossimilhança de uma amostra Y_1, \dots, Y_n que segue uma distribuição RBGEV com valores observados y_1, \dots, y_n e definido que θ o vetor de parâmetros da função 3.1.1 pode ser definida como:

$$L_{\theta,\beta} = \prod_{i=1}^n f(y_i|\mathbf{x}^t\beta), \quad (3.2.1)$$

por motivos computacionais e de cálculo é comum utilizar o logaritmo da função de verossimilhança, pois dessa forma muitas interações são simplificadas, essa função é usualmente chamada log-verossimilhança. Dessa forma, considerando a existência de duas funções de regressão considerando os valores de $\xi = 0$ e $\xi \neq 0$ as funções de log-verossimilhança para RBGEV é explícita nas funções 3.2.2 e 3.2.3.

- Caso 1. Para $\xi = 0$

$$l_{\theta,\beta} = \sum_{i=1}^n \log \left(\frac{T'_{\mu,\delta,\beta}(y_i|\mathbf{x}^t)}{\sigma} \right) - \sum_{i=1}^n \frac{T_{\mu,\delta,\beta}(y_i|\mathbf{x}^t)}{\sigma} - \sum_{i=1}^n \exp \left(- \frac{T_{\mu,\delta,\beta}(y_i|\mathbf{x}^t)}{\sigma} \right) \quad (3.2.2)$$

- Caso 2. Para $\xi \neq 0$

$$\begin{aligned}
l_{\theta,\beta} &= \sum_{i=1}^n \log \left(\frac{T'_{\mu,\delta,\beta}(y_i|\mathbf{x}^t)}{\sigma} \right) - \frac{\xi+1}{\xi} \sum_{i=1}^n \log \left[1 + \xi \left(\frac{T_{\mu,\delta,\beta}(y_i|\mathbf{x}^t)}{\sigma} \right) \right] \\
&\quad - \sum_{i=1}^n \left[1 + \xi \left(\frac{T_{\mu,\delta,\beta}(y_i|\mathbf{x}^t)}{\sigma} \right) \right]^{-\frac{1}{\xi}}
\end{aligned} \tag{3.2.3}$$

4 Possíveis Aplicações

Usualmente a GEV é muito aplicada nas áreas de confiabilidade, finanças e meio ambiente, além de outras com **menor expressão**. Nesse trabalho, por outro lado, foi pensado em se trabalhar no ambiente de dados educacionais. Pensando nisso, nessa seção é apresentado nas Figuras 9,10, 11 e 12 alguns possíveis **bancos** para aplicação.

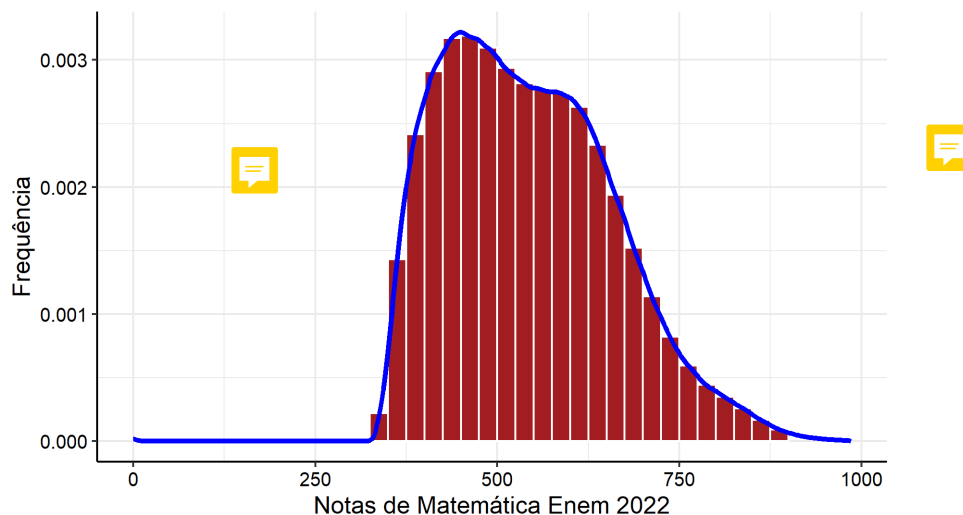


Figura 9: Distribuição das notas do Enem 2022 na proficiência de matemática

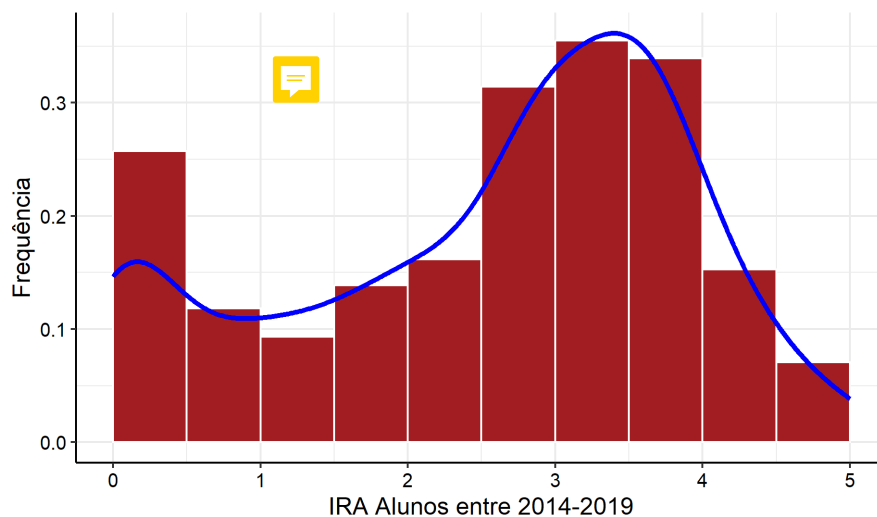


Figura 10: Distribuição do IRA de alunos do CIC entre 2014-2019

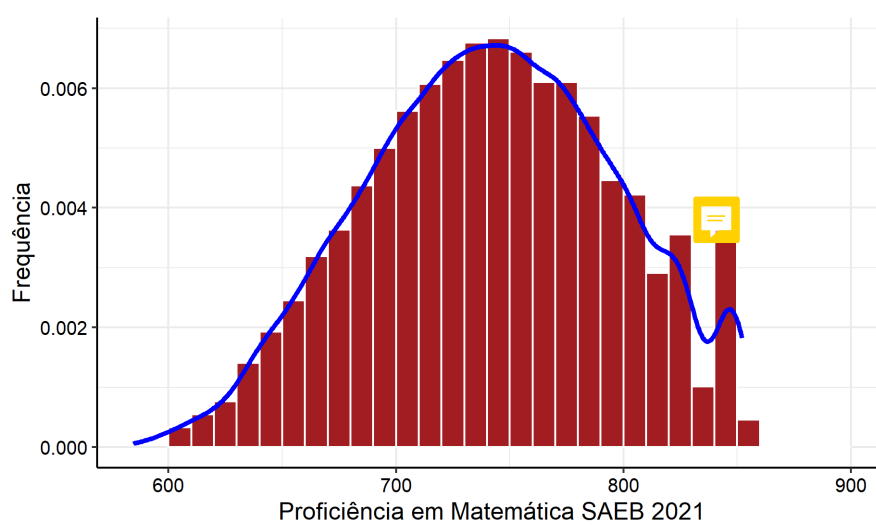


Figura 11: Distribuição da proficiência em matemática SAEB 2021

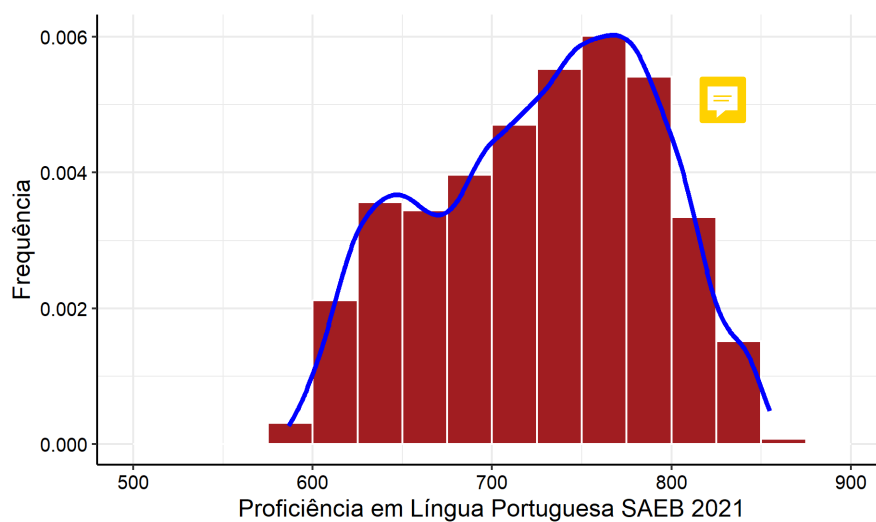


Figura 12: Distribuição da proficiência em língua portuguesa SAEB 2021

5 Considerações Finais

O trabalho ainda não está finalizado e é necessário considerar o que deve ser feito a seguir. Primeiramente, ainda é preciso realizar as simulações de desempenho dos estimadores. Nesse caso o planejamento é realizar amostras aleatórias de uma variável com distribuição BGEV por meio do método da transformada inversa. O método consiste na utilização da função quantil da distribuição apresentada na Equação 2.2.9 e 2.2.11 com domínio uma variável uniforme de intervalo 0 a 1 ($U \sim U(0, 1)$) gerando valores aleatório e aplicar na expressão $X = F^{-1}(U)$. E então definir um número de amostras de Monte Carlo, com diversas combinações de parâmetros e tamanhos de amostra. Em seguida, calcula-se o viés e o erro quadrático médio para avaliar a qualidade dos estimadores.



Além disso, um capítulo fundamental para a conclusão do trabalho é a aplicação. Esse tópico é sensível, pois para esse trabalho tem-se pensado além das aplicações clássicas da GEV, como o caso de dados climáticos, financeiros ou confiabilidade. O intuito inicial é trazer um olhar para as aplicações desse modelo nos casos de desempenho acadêmico, uma vez que a hipótese inicial da existência de grupos heterogêneos na educação.

Referências

- 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. 4, 17, 18
- ELAL-OLIVERO, D. Alpha-skew-normal distribution. *Proyecciones (Antofagasta)*, SciELO Chile, v. 29, n. 3, p. 224–240, 2010. 9
- EMBRECHTS, P. et al. Risk theory. *Modelling Extremal Events: for Insurance and Finance*, Springer, p. 21–57, 1997. 4
- 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. 4, 6
- 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. 9
- GNEDENKO, B. Sur la distribution limite du terme maximum d’une serie aleatoire. *Annals of mathematics*, JSTOR, p. 423–453, 1943. 4, 6
- 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. 4
- HAAN, L.; FERREIRA, A. *Extreme value theory: an introduction*. [S.l.]: Springer, 2006. v. 3. 4, 8
- 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. 4, 7
- JONDEAU, E.; POON, S.-H.; ROCKINGER, M. *Financial modeling under non-Gaussian distributions*. [S.l.]: Springer Science & Business Media, 2007. 6
- KOTZ, S.; NADARAJAH, S. *Extreme value distributions: theory and applications*. [S.l.]: world scientific, 2000. 4, 8
- LIMA, C. H.; KWON, H.-H.; KIM, Y.-T. A local-regional scaling-invariant bayesian gev model for estimating rainfall idf curves in a future climate. *Journal of Hydrology*, Elsevier, v. 566, p. 73–88, 2018. 18
- LIMA, C. H. et al. A hierarchical bayesian gev model for improving local and regional flood quantile estimates. *Journal of Hydrology*, Elsevier, v. 541, p. 816–823, 2016. 18
- NASCIMENTO, F. et al. Extended generalized extreme value distribution with applications in environmental data. *Hacettepe Journal of Mathematics and Statistics*, Hacettepe University, v. 45, n. 6, p. 1847–1864, 2016. 4, 18

- OTINIANO, C. E. et al. A bimodal model for extremes data. *Environmental and Ecological Statistics*, Springer, p. 1–28, 2023. 4, 9
- 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. 4
- OTINIANO, C. E. et al. A bimodal model for extremes data. *arXiv preprint arXiv:2109.12738*, 2021. 9
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2022. Disponível em: <https://www.R-project.org/>. 4
- RESNICK, S. Extreme values, regular variation and weak convergence. Springer Verlag, New York, 1987. 4
- 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. 8
- 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. 9
- 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. 4, 17, 18
- 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. 4, 17, 18
- VILA, R. et al. A bimodal gamma distribution: Properties, regression model and applications. *Statistics*, Taylor & Francis, v. 54, n. 3, p. 469–493, 2020. 9
- VILA, R. et al. On a bimodal birnbaum–saunders distribution with applications to lifetime data. 2021. 4, 9
- 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. 4