PROBABILIDADE II

Precipitação máxima esperada na cidade de Lavras-MG via distribuição generalizada de valores extremos

Lucas Pereira Belo
Jonas Firmiano da Silva
Rodrigo da Cruz Nunes

15 de agosto de 2025

Estatística de ordem

Definição

Sejam $X_1, X_2, \ldots, X_n$ uma amostra aleatória de tamanho n de uma função de distribuição acumulada $F(\cdot)$. Então $Y_1 \leq Y_2 \leq \cdots \leq Y_n$, em que os Y_i são os X_i organizados em ordem de magnitudes crescentes, são definidos como as **estatísticas de ordem** correspondentes à amostra aleatória $X_1, \ldots, X_n$ (MOOD *et al.*, 1974).

Máximos e mínimos

Definição

- Y_1 é a primeira estatística de ordem e representa o **valor mínimo** da amostra aleatória.

$$Y_1 = \min\{X_1, X_2, \ldots, X_n\}$$

- Y_n é a n-ésima estatística de ordem e representa o **valor máximo** da amostra aleatória.

$$Y_n = \max\{X_1, X_2, \ldots, X_n\}$$

Qual a distribuição de Y_n?

Distribuição de Y_n

Exemplo

Imagine que o nosso experimento consiste em lançar um dado comum de seis faces 3 vezes.

Neste cenário, cada X_i representa o resultado de *um único lançamento* do dado. Como vamos lançar o dado 3 vezes ($n = 3$), teremos três variáveis aleatórias:

- X_1: O resultado do **primeiro** lançamento.
- X_2: O resultado do **segundo** lançamento.
- X_3: O resultado do **terceiro** lançamento.

Distribuição de Y_n

Exemplo

Você lança os dados e obtém a sequência: **4, 1, 5**. Neste caso, teríamos: $x_1 = 4$, $x_2 = 1$ e $x_3 = 5$.

As estatísticas de ordem, y_i, são os mesmos valores da nossa amostra, mas **colocados em ordem crescente**. Usando o mesmo resultado do exemplo acima (4, 1, 5):

- y_1: O **menor** valor que obtivemos. $y_1 = \min\{4, 1, 5\} = \mathbf{1}$.
- y_2: O valor do **meio**. $y_2 = \mathbf{4}$.
- y_3: O **maior** valor que obtivemos. $y_3 = \max\{4, 1, 5\} = \mathbf{5}$.

A distribuição de X_i vs. a distribuição de Y_n

Exemplo

A probabilidade de cada resultado para um único lançamento de um dado justo é uma **distribuição uniforme discreta**:

- $P(X_i = 1) = 1/6$
- $P(X_i = 2) = 1/6$
- $\dots$
- $P(X_i = 6) = 1/6$

Qual a distribuição de Y_n? (o máximo de 3 lançamentos)

Exemplo

- **Qual a probabilidade do máximo ser 1?** Para que o valor máximo dos três lançamentos seja 1, você precisa tirar obrigatoriamente a sequência $(1, 1, 1)$.
- **Qual a probabilidade do máximo ser 6?** Para que o valor máximo seja 6, basta que *pelo menos um* dos dados seja 6. Existem muito mais combinações que resultam em um máximo de 6 (por exemplo, $(6, 1, 2)$, $(3, 6, 4)$, $(6, 6, 1)$, etc.).

Qual a distribuição de Y_n? (o máximo de 3 lançamentos)

Exemplo

- A probabilidade do máximo ser 1 é:

$$P(Y_3 = 1) = P(X_1 = 1; X_2 = 1; X_3 = 1) = \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{216}$$

- Para calcular a probabilidade do máximo ser 6, usamos o evento complementar. A probabilidade do máximo ser menor ou igual a 5 ($Y_3 \leq 5$) ocorre se, e somente se, todos os três lançamentos forem menores ou iguais a 5.

$$P(Y_3 \leq 5) = P(X_1 \leq 5; X_2 \leq 5; X_3 \leq 5) = \left(\frac{5}{6}\right)^3 = \frac{125}{216}.$$

Portanto, a probabilidade do máximo ser exatamente 6 é:

$$P(Y_3 = 6) = 1 - P(Y_3 \leq 5) = 1 - \frac{125}{216} = \frac{91}{216}$$

Qual a distribuição de Y_n e Y_1?

Teorema

Sejam $Y_1 \leq Y_2 \leq \cdots \leq Y_n$ as estatísticas de ordem de uma função de distribuição acumulada $F(\cdot)$. As funções de distribuição acumulada para a maior e a menor estatística de ordem são, respectivamente:(MOOD; GRAYBILL; BOES, 1974)

$$F_{Y_n}(y) = \sum_{j=n}^{n} \binom{n}{j} [F(y)]^j [1 - F(y)]^{n-j} = [F(y)]^n.$$

e

$$F_{Y_1}(y) = \sum_{j=1}^{n} \binom{n}{j} [F(y)]^j [1 - F(y)]^{n-j} = 1 - [1 - F(y)]^n.$$

O teorema de tipos extremais

Também conhecido como teorema de Fisher-Tippett-Gnedenko

Distribuição assintótica

A questão fundamental: Qual é a distribuição do valor máximo de uma amostra, $Y_n = \max\{X_1, \ldots, X_n\}$, quando o tamanho da amostra $n \to \infty$?

Teorema central do limite (TCL)

Para a média amostral $\bar{X}_n$, o TCL evita a convergência degenerada para μ através de uma normalização linear (desde que certas condições sejam atendidas):

$$\frac{\bar{X}_n - \mathbf{b_n}}{\mathbf{a_n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1)$$

Distribuição assintótica

De forma análoga, buscamos sequências de constantes de normalização, $\mathbf{a_n} > \mathbf{0}$ e $\mathbf{b_n}$, tais que:

$$\frac{Y_n - \mathbf{b_n}}{\mathbf{a_n}} \xrightarrow{d} G(x),$$

em que $G(x)$ é uma distribuição **não-degenerada**.

O teorema de tipos extremais

Teorema

Se existem sequências de constantes de normalização $a_n > 0$ e b_n tais que, para $n \to \infty$, a distribuição do máximo normalizado Y_n converge para uma distribuição não-degenerada $G(x)$:

$$\Pr\left\{\frac{Y_n - b_n}{a_n} \leq x\right\} \to G(x)$$

Então, $G(x)$ deve pertencer a uma das três famílias de distribuições a seguir:

I: Gumbel: $G(x) = \exp\{-\exp(-x)\} \quad -\infty < x < \infty;$

II: Fréchet: $G(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0, \alpha > 0; \end{cases}$

III: Weibull: $G(x) = \begin{cases} \exp\{-(-x)^{\alpha}\} & x < 0, \alpha > 0 \\ 1 & x \geq 0. \end{cases}$

A distribuição generalizada de valores extremos (GEV)

Definição

A Distribuição Generalizada de Valor Extremo (GEV) unifica os três tipos (Gumbel, Fréchet e Weibull) em uma única família. Sua função de distribuição acumulada é dada por:

- Para $\xi \neq 0$:

$$G(x; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_{+}^{-1/\xi}\right\}$$

- Para $\xi = 0$:

$$G(x; \mu, \sigma) = \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}$$

em que μ, σ, ξ são os parâmetros de localização, escala e forma respectivamente.

Parâmetros da GEV

O valor do parâmetro de forma, ξ, determina a qual das três distribuições de valor extremo a GEV corresponde:

- $\xi = 0$: Corresponde ao **Tipo I (Gumbel)**.
- $\xi > 0$: Corresponde ao **Tipo II (Fréchet)**.
- $\xi < 0$: Corresponde ao **Tipo III (Weibull)**.

Teoricamente, a convergência para a GEV ocorre com a normalização:

$$\frac{Y_n - b_n}{a_n} \xrightarrow{d} G(\mu, \sigma, \xi)$$

No entanto, pode-se mostrar que as constantes de normalização a_n e b_n são "absorvidas" pelos parâmetros de escala e localização, resultando em:

$$Y_n \xrightarrow{d} G(\mu^*, \sigma^*, \xi).$$

Estimação dos parâmetros da GEV

Estimação

1. Utilizamos a função de verossimilhança $L(\mu, \sigma, \xi; \boldsymbol{x}) = L(\mu, \sigma, \xi; x_1, \cdots, x_n)$;
2. Determinamos a função de log-verossimilhança (função suporte) da GEV;
3. Procuramos as estimativas de μ, σ, ξ que maximizam a função suporte.

Métodos numéricos

- Newton-Raphson (NEWTON, 1711; NEWTON, 1774; RAPHSON, 1690);
- Gradiente decrescente (CAUCHY et al., 1847);

Níveis de Retorno

Exemplos

O foco não está nas estimativas dos parâmetros da distribuição GEV em si, mas na aplicação do modelo ajustado para prever quantidades de interesse prático. Por exemplo:

- Qual deve ser a altura ideal de um muro de contenção para suportar a maré mais alta esperada a cada cem anos;
- Qual é a maior velocidade do vento esperada em um intervalo de cinquenta anos, a fim de projetar estruturas capazes de resistir a esse tipo de evento extremo.

Definição

Essas quantidades são conhecidas, no contexto da teoria dos valores extremos, como níveis de retorno.

LAVRAS-MG

Figura: Imagem da cidade de Lavras.

Por que escolhemos Lavras?

Razões para a escolha

- **Destaque na agricultura:** Forte presença no cultivo de **café**, soja, milho e feijão;
- **Pecuária leiteira:** Reconhecida por um dos melhores rebanhos de gado leiteiro do estado;
- **Agricultura familiar:** Importante setor de produção local;
- **Condições climáticas favoráveis:** Clima propício para atividades agrícolas e estudos;
- **Projetos de extensão da UFLA:** Apoio e desenvolvimento de iniciativas na agricultura e pecuária.

Produção em larga escala: Café e Leite

Figura: Fazenda Faria

Figura: Fazenda Palmital

Agricultura familiar

É um pilar essencial na produção de alimentos, com atuação destacada em duas frentes:

- A participação em chamadas públicas para a merenda escolar por meio do PNAE, garantindo alimentos frescos e de qualidade para os estudantes;
- Comercialização direta de seus produtos ao consumidor, fortalecendo a economia local.

Principais produtos da agricultura familiar.

Segundo Lage(2019), esses são os principais produtos da agricultura familiar:

- Leite (30%);
- Hortaliças (13%);
- Café (12%);
- Milho (12%);
- Ovo caipira (7%);
- Queijo (5%);
- Gado para corte (4%);
- Frutas (4%);
- Feijão (4%).

Por que o estudo sobre precipitação em Lavras?

Eventos recentes e impactos

- **Transtornos significativos (Início de 2025):** Apesar do histórico de poucas enchentes, Lavras enfrentou sérios problemas devido a fortes chuvas.
- **Impactos noticiados (G1 Sul de Minas, 2025):**
  - Ruas alagadas.
  - Residências parcialmente submersas.
  - Quedas de pontes.
  - Interrupção no fornecimento de energia elétrica.

Fatores agravantes

- **Crescimento urbano desordenado:** Contribui para a vulnerabilidade da cidade.
- **Mudanças climáticas constantes:** Aumentam a frequência e intensidade dos eventos extremos.

Fontes e período dos dados

Origem e abrangência dos dados

- **Fontes principais:**
  - Banco de Dados Meteorológicos para Ensino e Pesquisa (**BDMEP**) do INMET.
  - Informações complementares do trabalho de Beijo *et al.* (2005).
- **Tipo de informação:** Registros diários de **precipitação pluvial (mm)**.
- **Local de coleta:** Cidade de Lavras.
- **Período abrangido: 01/01/1961 a 02/05/2025**.

Organização dos dados de precipitação

Processamento das máximas anuais

- Os dados foram agrupados anualmente:
  - Grupos de **365 dias** para anos comuns.
  - Grupos de **366 dias** para anos bissextos.
- Dentro de cada grupo, foram extraídas as **maiores precipitações diárias** observadas;
- Isso resultou em um conjunto de dados com **65 observações** das precipitações máximas anuais.

Adequação da base amostral para GEV

Recomendações e fundamentação

- A **Organização Mundial de Meteorologia** sugere análises com séries históricas de **pelo menos 30 anos** (BADDOUR; KONTONGOMDE, 2007).
- Cai e Hames (2011) indicam um mínimo de **40 observações** para a validade das inferências estatísticas fundamentais da GEV.
- Roslan *et al.* (2020) consideram que o número mínimo de observações deve ser **50**, pois isso:
  - Garante a **normalidade assintótica** dos estimadores de máxima verossimilhança.
  - Torna as estimativas dos níveis de retorno **mais confiáveis**.

O presente estudo

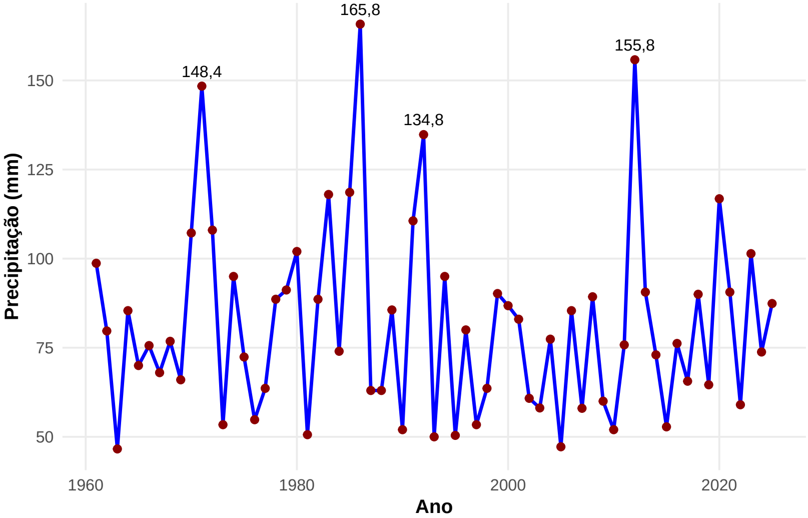
- O presente estudo, com **65 observações**, **atende a todos esses requisitos mínimos sugeridos**.

Valores de precipitação máxima diária anual (mm) 1961-2025.

98,7	79,7	46,6	85,4	70,0	75,6	68,0
76,8	66,0	107,2	148,4	108,0	53,4	95,0
72,4	54,8	63,6	88,6	91,2	102,0	50,6
88,6	118,0	74,0	118,6	165,8	63,0	63,0
85,6	52,0	110,6	134,8	50,0	95,0	50,4
80,0	53,4	63,6	90,2	86,8	83,0	60,8
58,1	77,4	47,2	85,4	58,0	89,3	60,0
52,0	75,8	155,8	90,6	73,0	52,8	76,2
65,6	90,0	64,6	116,8	90,6	59,0	101,4
73,8	87,4					

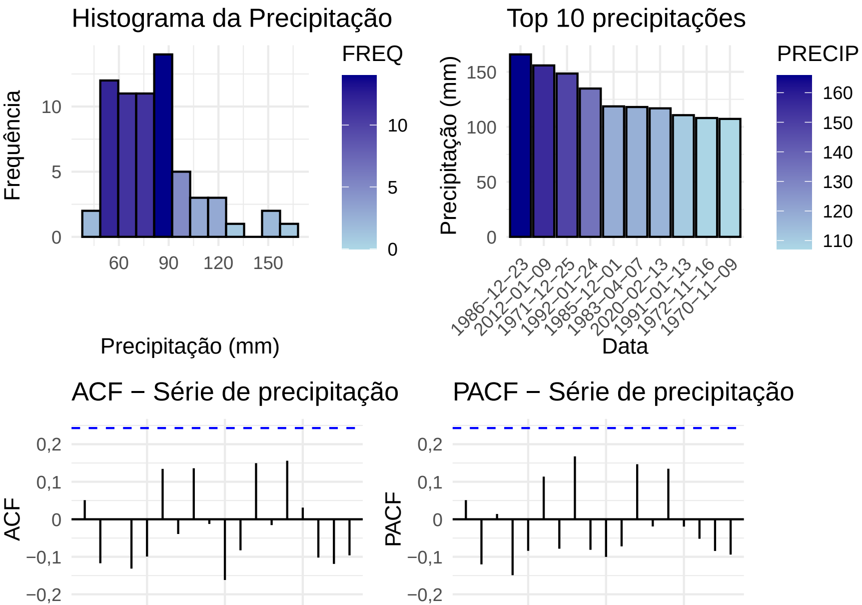
Valores em azul: 10 maiores precipitações

Série temporal de máximas diárias anuais.



Fonte: Autores.

Análise descritiva



Resultados dos testes estatísticos

Teste de aleatoriedade

O *teste de sequência* não rejeitou a hipótese nula de que a sequência de dados é aleatória, com um valor$-p = 0,0604$.

Teste de independência

O *teste de Ljung-Box* também não rejeitou a hipótese nula de independência dos dados, valor$-p = 0,6739$ corroborando com os gráficos ACF e PACF da Figura 29.

Teste de estacionariedade

O *teste de Dickey-Fuller* rejeitou a hipótese nula de existência de raiz unitária, valor$-p = 0,01$ indicando que a série é estacionária.

Estimativas dos parâmetros do modelo GEV

Método de estimação

Estimativas dos parâmetros do modelo GEV foram obtidas pelo método da **máxima verossimilhança**, utilizando o algoritmo de **Newton-Raphson**.

Parâmetros estimados

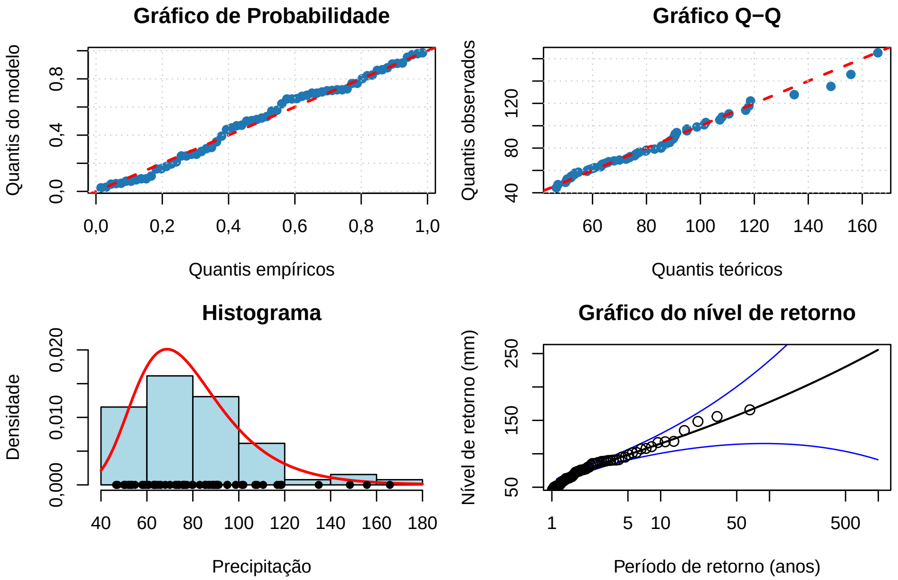
Os parâmetros de locação (μ), escala (α) e forma (ξ) foram estimados como:

- **Locação ($\hat{\mu}$):** 68,7578 (com erro padrão de 2,6719)
- **Escala ($\hat{\alpha}$):** 18,2955 (com erro padrão de 2,0695)
- **Forma ($\hat{\xi}$):** 0,1071 (com erro padrão de 0,1209)

Adequação a GEV

O teste de Kolmogorov-Smirnov não rejeitou a hipótese nula, com um valor$-p = 0,6963$ indicando adequação da GEV aos dados.

Adequação do modelo



Níveis de retorno e intervalos de confiança

Estimativas e intervalos de confiança

Os níveis de retorno estimados pelo Método da Máxima Verossimilhança (EMV), juntamente com seus respectivos intervalos de confiança (IC), são apresentados a seguir:

- Para $\hat{\mathbf{z}}_{10}$: **115,323** mm (IC: 103,6287; 137,334)
- Para $\hat{\mathbf{z}}_{25}$: **138,546** mm (IC: 120,0185; 186,4295)
- Para $\hat{\mathbf{z}}_{50}$: **157,358** mm (IC: 131,6001; 236,7152)
- Para $\hat{\mathbf{z}}_{100}$: **177,483** mm (IC: 142,5628; 302,3116)

Implicações e monitoramento

Em média, espera-se um nível de precipitação diária de **115,323 mm em um período de 10 anos**. Para os períodos de retorno de 25 e 50 anos, os níveis estimados são de **138,546 mm** e **157,358 mm**, respectivamente.

Comparativo com estudos anteriores e implicações

Subestimação das estimativas anteriores

Ao comparar nossos resultados com os apresentados por Beijo *et al.* (2005), observa-se que as estimativas anteriores estavam **subestimadas**.

- Para um período de retorno de **10 anos** (a partir de 2003), esperava-se uma precipitação máxima diária de **129 mm)** (Beijo *et al.*, 2005).
- Contudo, esse valor foi **superado já em 2012**, com um registro de **155,8 mm em apenas uma hora**.
- De acordo com as estimativas de Beijo *et al.* (2005), um evento dessa magnitude era esperado apenas em um horizonte de **30 anos**.

Implicações e relevância do estudo

Por que isso importa?

A superação dos níveis de precipitação esperados em um período menor do que o previsto destaca a **urgência de atualizar e aprimorar** as estimativas de níveis de retorno.

Isso reforça a necessidade de:

- **Revisão e adequação** de planejamentos urbanos e infraestruturas hídricas.
- **Monitoramento contínuo** das condições climáticas e hidrológicas da região.
- Desenvolvimento de estratégias de **gestão de riscos** mais eficazes, especialmente para áreas de risco (alagamentos, deslizamentos, inundações).

Conclusão: Desafios e próximos passos

Principais pontos e implicações para Lavras

- A **distribuição de Gumbel** foi considerada adequada para modelar os dados de precipitação diária máxima em Lavras, alinhando-se a estudos anteriores (Beijo *et al.*, 2005).
- As estimativas dos tempos de retorno e seus intervalos de confiança são **ferramentas valiosas** para decisões preventivas no município.
- Um evento extremo como o de **165,8 mm** em 23 de dezembro de 1986 tem um tempo de retorno estimado em aproximadamente **25 anos**, considerando o intervalo de confiança do modelo.
- Contudo, é crucial reconhecer a **possível subestimação** dessas estimativas devido aos efeitos das **mudanças climáticas**, conforme já apontado por Beijo *et al.* (2005).
- Essa observação reforça a necessidade **de estudos contínuos** sobre eventos extremos, com revisões periódicas das estimativas de precipitação e atualização constante dos modelos de predição.

Referências bibliográficas

BEIJO, L. A.; MUNIZ, J. A.; NETO, P. C. Tempo de retorno das precipitações máximas em lavras (mg) pela distribuição de valores extremos do tipo i. **Ciência e agrotecnologia**, SciELO Brasil, v. 29, p. 657–667, 2005.

CAUCHY, A. et al. Méthode générale pour la résolution des systemes d'équations simultanées. **Comp. Rend. Sci. Paris**, v. 25, n. 1847, p. 536–538, 1847.

G1 – Sul de Minas. **Chuva causa transtornos neste domingo no Sul de Minas**. 2025. Acesso em: 22 jun. 2025. Disponível em: ⟨https://g1.globo.com/mg/sul-de-minas/noticia/2025/01/26/chuva-causa-transtornos-neste-domingo-no-sul-de-minas.ghtml⟩.

LAGE, B. G. P. **Acesso e Inserção da Agricultura Familiar Camponesa de Lavras-MG em Mercados de Cadeias Curtas**. Dissertação (Dissertação de Mestrado Profissional) — Universidade Federal de Lavras, Lavras, MG, 2019. Orientador(a): Thiago Rodrigo de Paula Assis; Coorientador(a): Nathalia de Fátima Joaquim.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the Theory of Statistics**. 3rd. ed. [S.l.]: McGraw-Hill, 1974.

NEWTON, I. **De analysi per aequationes numero terminorum infinitas**. [S.l.: s.n.], 1711.

NEWTON, I. Methodus fluxionum et serierum infinitarum. **Opuscula mathematica, philosophica et philologica**, v. 1, p. 1774, 1774.

RAPHSON, J. **Analysis Aequationum Universalis.** 1690.

ROSLAN, R.; NA, C. S.; GABDA, D. Parameter estimations of the generalized extreme value distributions for small sample size. **Mathematics and Statistics**, v. 8, n. 2, p. 47–51, 2020.