

# Distribuições de Probabilidade das Temperaturas Diárias Máximas entre 1980 e 2021 em Brasília - DF

André Augusto Sak e Marcos da Silva Correia

Orientador: Me. Natália Ribeiro de Souza Evangelista

Brasília, 18 de junho de 2022.



# Sumário

- 1 Introdução
- 2 Referencial Teórico
- 3 Metodologia
- 4 Considerações Finais

# Motivação

- A questão climática está entre os principais desafios enfrentados atualmente pela humanidade (Harari, 2018);
- Impacto negativo no crescimento e desenvolvimento das plantas (Estefanel et al., 1994) e
- Redução da produtividade agrícola em até 40% (Costa e Sedyama, 1999).

## Estudo conduzido no Rio Grande do Sul:

- Temperaturas máximas acima de 35 °C resultaram em elevadas taxas de esterilidade.

# Motivação

- Importante ferramenta para o planejamento e organização de outras atividade humanas (de Assis et al., 2013) e
- Evitar prejuízos e alavancar lucros.

# Objetivos

- Analisar os registros de temperaturas máximas em Brasília, no Distrito Federal, durante um período de 41 anos (de 1° de janeiro de 1980 até 31 de dezembro de 2021) e
- Aplicar quatro importantes funções de distribuição de probabilidade: *Gama*, *Gumbel*, *Log-normal* e *Normal*.

# Trabalhos Semelhantes

## Araújo et al., 2010:

- Aplicação de distribuições de probabilidade a séries de temperatura máxima em Iguatu, Ceará.
- Normal e a Gama quando sua forma foi aproximada da Normal.

## de Assis et al., 2013:

- Aplicação de distribuições de probabilidade a séries de umidade relativa mensal em Mossoró, Rio Grande do Norte.
- Gama, Gumbel e a Normal.

## Supian; Hasan, 2013:

- Aplicação de distribuições de probabilidade a 17 estações meteorológicas na Malásia.
- Log-normal, Normal, Weibull ou Logística Generalizada.

# Obtenção dos Dados

- Os dados diários das temperaturas máximas foram obtidos a partir do site do Instituto Nacional de Meteorologia (Inmet)<sup>1</sup>.
- Estação meteorológica de Brasília n° 83377.
- Período: 01 de janeiro de 1980 a 31 de dezembro de 2021.



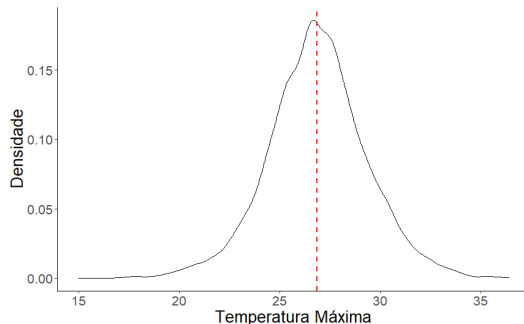
**Figura 1:** Localização da estação meteorológica n° 83377.

<sup>1</sup><https://bdmep.inmet.gov.br/>

# Descrição dos Dados

**Tabela 1:** Medidas estatísticas:

Medida	°C
Mínimo	15,00
1° quartil	25,40
Moda	26,60
Mediana	26,80
Média	26,87
3° quartil	28,40
Máximo	36,30



**Figura 2:** Gráfico de densidade.

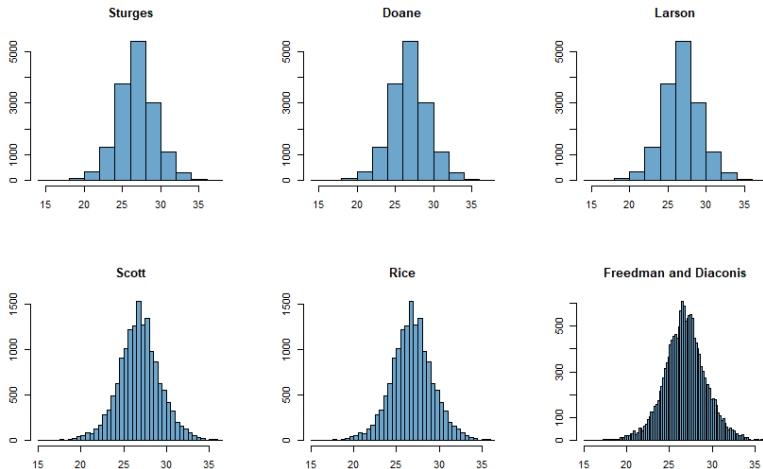


# Teste de Homogeneidade

## Teste de Variância

- A população foi dividida em dois grupos: um abaixo e outro acima da mediana.
- Hipótese nula ( $H_0$ ): há igualdade entre as variâncias dos grupos.
- Hipótese alternativa ( $H_1$ ): a razão entre as variâncias é menor do que 1.
- $p - \text{valor} = 0,9239$ .
- Os dados são homogêneos.

# Agrupamento dos Dados



**Figura 3:** Procedimentos trazidas por (Dogan and Dogan, 2010):

# Modelagem

## Gama

$$f(x; k, \theta) = \begin{cases} \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)}, & \text{se } x > 0 \text{ e } k, \theta > 0 \\ 0, & \text{caso contrário.} \end{cases} \quad (1)$$

## Parametrização

$$\theta = \frac{S^2}{\bar{X}}, \quad (2)$$

• e

$$k = \frac{\bar{X}}{\theta}. \quad (3)$$

# Modelagem

## Gumbel

$$F(x; \alpha, \beta) = \exp \left( -e^{-(x-\alpha)/\beta} \right), -\infty < \alpha < +\infty \text{ e } \beta > 0. \quad (4)$$

## Parametrização

$$\beta = \sqrt{6 \frac{S^2}{\pi^2}} \quad (5)$$

• e

$$\alpha = \beta \frac{\Gamma'(1)}{\Gamma(1)} + \bar{x}. \quad (6)$$

# Modelagem

## Log-normal

$$f(x; \mu, \sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right), -\infty < \mu < +\infty \text{ e } \sigma^2 > 0. \quad (7)$$

## Parametrização

$$\sigma^2 = \ln\left(\left(\frac{S^2}{\exp(2\ln(\bar{x}))}\right) + 1\right), \quad (8)$$

• e

$$\mu = \frac{2\ln(\bar{x}) - \sigma^2}{2}. \quad (9)$$

# Modelagem

## Normal

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right\}, \quad -\infty < x < \infty. \quad (10)$$

## Parametrização

$$\mu = \bar{x}, \quad (11)$$

• e

$$\sigma^2 = S^2. \quad (12)$$

# Método da Máxima Verossimilhança

## Média populacional

$$E[\bar{x}] = \mu = 26,86551. \quad (13)$$

## Variância populacional

$$E[S^2] = \sigma^2 = 6,019498. \quad (14)$$

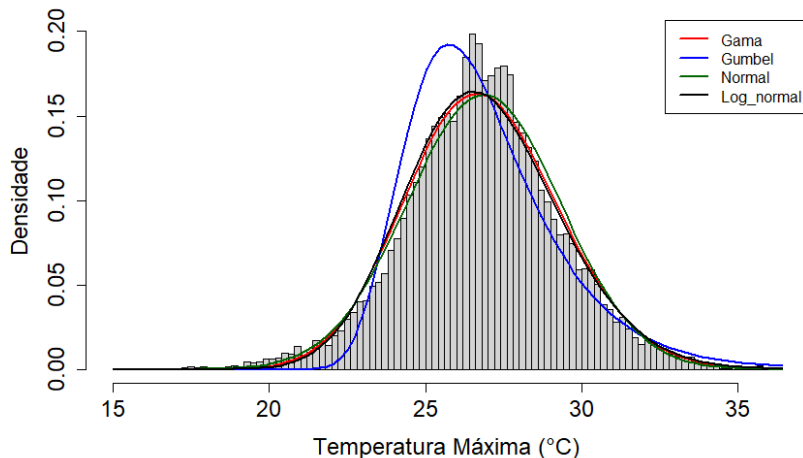
## Parâmetros calculados

**Tabela 2:** Parâmetros das distribuições Gama, Gumbel, Log-normal e Normal.

Distribuição	$\alpha$	$\beta$
Gama	119,903	4,463082
Gumbel	25,76132	1,91296
	$\mu$	$\sigma^2$
Normal	26,86551	6,019498
Log-normal	3,2866913	0,008305489

*Fonte:* Elaborado pelos autores (2022).

# Análise Visual



**Figura 4:** Ajustamento visual das distribuições Gama, Gumbel, Log-normal e Normal.



# Testes de Aderência

## Kolmogorov-Smirnov

$$KS = \max_{1 \leq i \leq n} |F(x_i) - F_e(x_i)|. \quad (15)$$

## Qui-Quadrado

$$\chi^2 = \sum_{i=1}^S \frac{(O_i - E_i)^2}{E_i}. \quad (16)$$

## Lilliefors (Normal)

$$LF = \max\{D^+, D^-\}, \quad (17)$$

$$D^+ = \max_{i=1, \dots, n} \{S(X) - E(X)\}, \quad (18)$$

$$D^- = \max_{i=1, \dots, n} \{F(X) - E(X)\}. \quad (19)$$

## Anderson-Darling

$$AD = -n - \frac{1}{n} \sum_{i=1}^n \{(2i-1) \times \log_e[S(X)] + [2(n-i)+1] \times \log_e[1-S(X)]\}. \quad (20)$$

# Testes de Aderência

- Hipótese nula ( $H_0$ ): o conjunto de dados  $X$  tem distribuição  $f$ .
- Hipótese alternativa ( $H_1$ ): o conjunto de dados  $X$  **não** tem distribuição  $f$ .

## Problema encontrado:

- Para  $n$  pequeno (menor que 30) são pouco úteis para e
- Para  $n$  grande (maior que 1000) são demasiadamente sensíveis (HAIR et al., 2009).

# Testes de Aderência

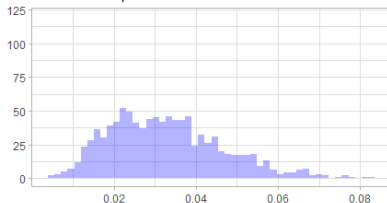
## Solução:

- *Bootstrappings* com 1000 reamostragens;
- $n = 500$ ;
- Calcular a média para a estatística de cada teste e
- O  $p$  – *valor* respectivo.

# Testes de Aderência *Kolmogorov-Smirnov* com *bootstrapping*

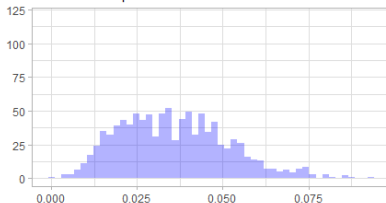
Distribuição KS para Reamostragem Log-normal

Média: 0.0326, p-value: 0.3463



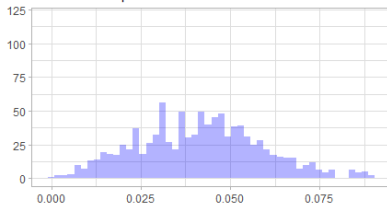
Distribuição KS para Reamostragem Gama

Média: 0.0358, p-value: 0.2792



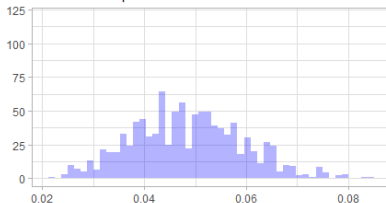
Distribuição KS para Reamostragem Normal

Média: 0.0415, p-value: 0.1804



Distribuição KS para Reamostragem Gumbel

Média: 0.0485, p-value: 0.0954



**Figura 5:** Distribuição das estatísticas do teste de aderência de *Kolmogorov-Smirnov* provenientes de *bootstrapping* com 1000 reamostragens e  $n = 500$ .

# Testes de Aderência com *bootstrapping*

**Tabela 3:** Resultados dos testes não paramétricos para a distribuição **Log-normal**.

Teste	Estatística	P-valor
Kolmogorov-Smirnov Test	0,020816	0,000001759
Bootstrap Kolmogorov-Smirnov Test (1.000 de n=500)	0,03339355	0,3462551
Chi-Square Test	4637,652	0
Bootstrap Chi-Square Test (1.000 de n=500)	15,76	0,0274018
Anderson Darling Test	29,76	0,00000003924
Bootstrap Anderson Darling Test (1.000 de n=500)	2,068904	0,08422735

Fonte: Elaborado pelos autores (2022).

**Tabela 4:** Resultados dos testes não paramétricos para a distribuição **Normal**.

Teste	Estatística	P-valor
Kolmogorov-Smirnov Test	0,032155	0,000000000000001857
Bootstrap Kolmogorov-Smirnov Test (1.000 de n=500)	0,04239254	0,1804151
Lilliefors Test	0,032155	0,000000000000000022
Bootstrap Lilliefors Test (1.000 de n=500)	0,04481782	0,01791657
Chi-Square Test	4398,152	0
Bootstrap Chi-Square Test (1.000 de n=500)	13,92	0,002732447
Anderson Darling Test	19,307	0,000000000000000022
Bootstrap Anderson Darling Test (1.000 de n=500)	1,65203	0,144019

Fonte: Elaborado pelos autores (2022).

# Testes de Aderência com *bootstrapping*

**Tabela 5:** Resultados dos testes não paramétricos para a distribuição **Gama**.

Teste	Estatística	P-valor
Kolmogorov-Smirnov Test	0,023812	0,00000002948
Bootstrap Kolmogorov-Smirnov Test	0,03544893	0,2792153
Chi-Square Test	4495,19	0
Bootstrap Chi-Square Test (1.000 de n=500)	13,52	0,003347909
Anderson Darling Test	23,827	0,00000003924
Bootstrap Anderson Darling Test (1.000 de n=500)	1,772153	0,1230396

Fonte: Elaborado pelos autores (2022).

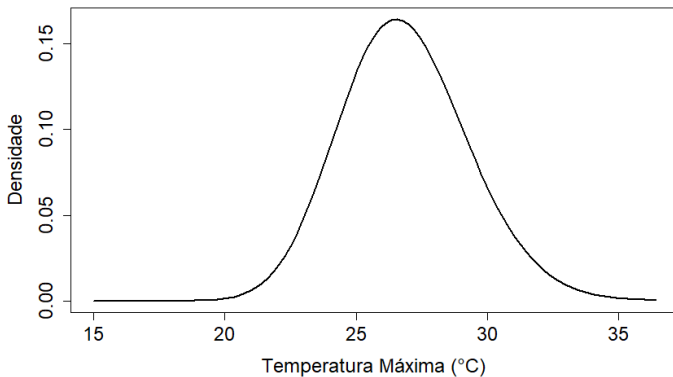
**Tabela 6:** Resultados dos testes não paramétricos para a distribuição **Gumbel**.

Teste	Estatística	P-valor
Kolmogorov-Smirnov Test	0,044149	0,00000000000000022
Bootstrap Kolmogorov-Smirnov Test (1.000 de n=500)	0,04804161	0,09544168
Chi-Square Test	4798,192	0
Bootstrap Chi-Square Test (1.000 de n=500)	11,52	0,0002555464
Anderson Darling Test	263,17	0,00000003924
Bootstrap Anderson Darling Test (1.000 de n=500)	10,0396	0,000007505501

Fonte: Elaborado pelos autores (2022).

# Conclusão

- **Critério:** maior  $p$  – *valor*.
- **Melhor aderência:** distribuição Log-normal.



**Figura 6:** Curva da distribuição Log-normal teórica.

# Momentos Estatísticos

## Momentos de ordem k

$$E(X^k) = \int_{-\infty}^{+\infty} x^k f(x) dx, \quad (21)$$

## Momentos centrais de ordem k

$$E[(X - \mu)^k] = \int_{-\infty}^{+\infty} (x - \mu)^k f(x) dx, \quad (22)$$

## Momentos da distribuição Log-normal

- 1º momento ( $m'_1$ ): 26,87089;
- 2º momento ( $m_2$ ): 6,01853;
- 3º momento ( $m_3$ ): 3,902572;
- 4º momento ( $m_4$ ): 113,0211.



# Comparações

## Assimetria

$$\frac{m_3}{(m_2)^{2/3}} \quad (23)$$

## Curtose

$$\frac{m_4}{(m_2)^2} \quad (24)$$

**Tabela 7:** Comparações da média, variância, assimetria e curtose entre a distribuição empírica e a teórica.

Distribuição	Média	Variância	Assimetria	Curtose
Empírica	26,86551	6,019498	-0,03713757	3,672045
Teórica	26,87089	6,01853	0,264311	3,120173

*Fonte:* Elaborado pelos autores (2022).

FIM