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

Es. André Augusto Sak* e Marcos da Silva Correia†
Orientadora: Prof[a]. Me. Natália Ribeiro de Souza Evangelista‡

18 de junho de 2022

Abstract

Climate phenomena may have an important impact on different human activities. Understanding the behavior of these phenomena makes it possible for one to plan actions that can mitigate problems and increase profits. The study of a distribution of maximum daily temperatures in a given place can contribute to that purpose. This paper aims to find, among the Gamma, Gumbel, Log-normal and Normal distributions, which one has better adherence to the distribution of maximum daily temperatures recorded in the Brazilian capital, Brasília, between 1980 and 2021.

Key-words: Temperature, Probability Distributions, Brasília, Brazil

Resumo

Fenômenos climáticos podem ter importante impacto em diversas atividades humanas. A compreensão de diferentes aspectos desses fenômenos possibilitam o planejamento de ações que possam mitigar problemas e potencializar lucros. Entre as análises que podem ser feitas nesse sentido está a da distribuição das temperaturas máximas diárias. Propõe-se realizar esse estudo para encontrar, dentre as distribuições de Gama, Gumbel, Log-normal e Normal, qual possui melhor aderência à distribuição de temperaturas máximas registradas em Brasília (DF) durante os anos de 1980 e 2021.

Palavras-chaves: Temperatura, Distribuições de Probabilidade, Brasília

*Graduando em Ciência de Dados e Inteligência Artificial no Centro Universitário Instituto de Educação Superior de Brasília (IESB), Bacharel em Direito pelo Centro de Ensino Unificado de Brasília (Uniceub) e pós-graduado em direito legislativo pelo Instituto Legislativo Brasileiro (ILB-Senado Federal), *e-mail*: andre.sak@iesb.edu.br.

†Graduando em Ciência de Dados e Inteligência Artificial no Centro Universitário Instituto de Educação Superior de Brasília (IESB) e Técnico em Informática pela Escola Técnica de Ceilândia (ETC), *e-mail*: marcos.correia@iesb.edu.br.

‡Bacharel em Estatística pela Universidade de Brasília (UNB) e Mestre em Estatística pela Universidade de São Paulo (USP), *e-mail*: natalia.evangelista@iesb.edu.br.

Introdução

A questão climática está entre os principais desafios enfrentados atualmente pela humanidade. Como aponta o filósofo e historiador israelense Yuval Harari, diminuir a possibilidade de um colapso ecológico é uma tarefa que transcende preocupações nacionais e que, por isso, depende de soluções globais (HARARI, 2018, págs. 150-156). Assim, cientistas do mundo inteiro têm atuado colaborativamente para identificar padrões de identificação dessas mudanças para que soluções possam ser propostas e validadas.

Um aumento nos registros da média da temperatura máxima diária pode ter, entre outros efeitos, um impacto negativo no crescimento e desenvolvimento das plantas (ESTEFANEL; SCHNEIDER; BURIOL, 1994) e também pode reduzir a produtividade agrícola em até 40% (COSTA; SEDIYAMA, 1999).

Buscando identificar a probabilidade de ocorrência de dias com temperaturas iguais ou superiores a 35°C no florescimento do arroz no Rio Grande do Sul, pesquisadores realizaram um estudo específico sobre o assunto. Segundo eles, temperaturas máximas acima desse patamar, mesmo que por períodos curtos de tempo (uma hora), durante a floração, podem resultar em elevadas taxas de esterilidade no arroz, comprometendo sobremaneira a produção da cultura (MOTA; ROSKOFF; SILVA, 1999).

Além disso, análises estatísticas relacionadas ao comportamento do clima ao longo de determinado período podem ser uma importante ferramenta para o planejamento e organização não só na agricultura, mas também em outras atividades humanas, como o turismo (ASSIS et al., 2013). A noção de como esses fenômenos se comportam probabilisticamente pode ajudar a evitar prejuízos e também a alavancar lucros.

A presente pesquisa tem como objetivo contribuir com a questão ao realizar uma análise dos registros de temperaturas diárrias máximas em Brasília, no Distrito Federal, durante um período de 41 anos (de 1º de janeiro de 1980 até 31 de dezembro de 2021).

Conforme a metodologia de classificação climática proposta por Köpen-Geiger (1900) e adaptada por Stezer (1966), o clima em Brasília pode ser caracterizado como *Aw* (CARDOSO; MARCUZZO; BARROS, 2014), ou seja, tropical, com estação chuvosa no verão, de novembro a abril, e uma estação seca no inverno, entre os meses de maio a outubro.

Aos dados disponibilizados foram aplicadas quatro importantes funções de distribuição de probabilidade, com vistas a verificar se elas se ajustam às informações ora em análise: *Gama*, *Gumbel* ou valor extremo, *Log-normal* e *Normal*.

1 Referencial Teórico

Em trabalho com objetivo semelhante (ARAUJO et al., 2010) foram analisadas a aplicação de distribuições de probabilidade a séries de temperatura máxima na cidade de Iguatu, no Ceará. Naquela ocasião, as distribuições que melhor se ajustaram foram as funções Normal e a Gama, quando sua forma foi aproximada da Normal.

Em artigo já mencionado, que estudou as séries históricas de umidade relativa mensal em Mossoró, no Rio Grande do Norte (ASSIS et al., 2013), foi possível verificar que as distribuições Gama, Gumbel e Normal apresentaram melhor ajuste em relação à Log-normal, à Beta, à Log-Pearson (Tipo III) e à Weibull. Nesse mesmo estudo, os autores apontam que a mera construção de um histograma para a visualização dos dados relacionados ao clima não é suficiente para inferir qual função de distribuição melhor adere aos dados. Sendo assim, é imprescindível a realização de testes de aderência "para verificar se a distribuição de probabilidade dos dados de uma variável em análise pode ser representada por uma determinada função de distribuição de probabilidade conhecida."(ASSIS et al., 2013, pág. 2).

Em outra pesquisa, que buscou a melhor aderência à distribuição das temperaturas máximas anuais em 17 estações meteorológicas na Malásia, as que se saíram melhor foram as distribuições Log-normal, Normal, Weibull ou a Logística Generalizada, conforme as características de cada estação (SUPIAN; HASAN, 2021).

2 Metodologia

Os dados foram obtidos a partir do site do Instituto Nacional de Meteorologia (INMET)[1], órgão ligado ao Ministério da Agricultura, Pecuária e Abastecimento (MAPA). Entre as diversas possibilidades de especificação, foram selecionados os dados diários das temperaturas máximas registradas na estação meteorológica de Brasília (nº 83377), no período de 01 de janeiro de 1980 até 31 de dezembro de 2021.

A referida estação possui a seguinte localização georreferencial, que pode ser visualizada pela Figura 1:

- Latitude: -15,78972221;
- Longitude: -47,92583332 e
- Altitude: 1161,42 metros.

Figura 1 – Localização da estação meteorológica nº 83377.

As informações solicitadas foram fornecidas em arquivo com extensão ".csv", com 252 KB de tamanho, duas colunas ("Data da Medição"e "Temperatura Máxima

[1] https://bdmep.inmet.gov.br/

Diária") e 15341 registros.

Os dados foram, a partir daí, tratados com o *software R*, no ambiente do *RStudio*.

Na base de dados, foram identificadas 51 datas sem o respectivo registro de temperatura máxima. Isso possivelmente ocorreu em razão de alguma falha ou necessidade de ajuste dos equipamentos de medição, já que, em outras tentativas de *download* dos dados, as mesmas ausências foram constatadas.

Como não se trata de uma análise de série temporal e por representarem apenas 0,3% dos dados, a solução adotada para os registros faltantes foi a sua exclusão da base.

As principais medidas extraídas dos dados foram:

- Mínimo: 15,00;
- 1º quartil: 25,40;
- Moda: 26,60;
- Mediana: 26,80;
- Média: 26,87;
- 3º quartil: 28,40 e
- Máximo: 36,40.

A proximidade das medidas de tendência central (média, mediana e moda) já permitem indicar que a distribuição provavelmente seguirá uma distribuição simétrica, como a Normal, por exemplo.

Essa indicação é corroborada pelo cálculo de assimetria, que retornou $-0,037$ e que, apesar de negativo, está muito próximo a 0, demonstrando que os dados são bastante simétricos. Por sua vez, o cálculo de curtose retornou $3,67$. Como $\gamma_1 > 3$, temos uma curva com um pico mais acentuado que seria o de uma Normal e, por isso, ela é descrita como leptocúrtica.

A Figura 2 ilustra a densidade da distribuição e ajuda a demonstrar a tendência simétrica dos dados:

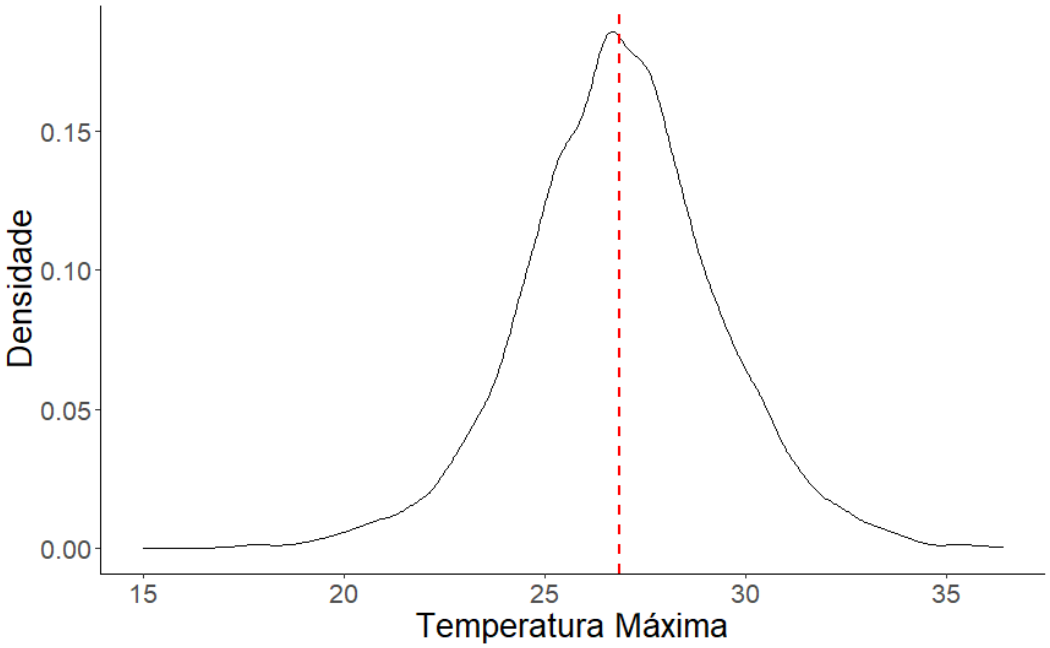

Figura 2 – Gráfico de densidade e linha média das temperaturas máximas de Brasília entre 1980 e 2021.

Com o objetivo de se verificar a homogeneidade dos dados, foi realizado o teste de variância. Para esse processo, a população foi dividida em dois grupos: um abaixo e outro acima da mediana e suas variâncias foram comparadas.

Para esse teste, foi considerada, como hipótese nula (H_0), a de que há igualdade entre as variâncias dos grupos e, como alternativa (H_1), a de que essa razão é diferente de 1. Obteve-se $p - valor = 0,9239$ e, portanto, como foi consideravelmente maior que $0,05$, não é possível rejeitar a hipótese nula. Nesse caso, deve-se concluir que há homogeneidade nos dados.

Para que se pudesse verificar qual o melhor critério de agrupamento dos dados, foram testados os seguintes procedimentos (DOGAN; DOGAN, 2010):

- Sturges (1926): 15 classes;
- Doane (1976): 10 classes;
- Larson (1975): 10 classes;
- Scott (1979): 62 classes;
- Rice: 50 classes e
- Freedman e Diaconis (1981): 89 classes.

Conforme é possível visualizar na Figura 3, os agrupamentos resultantes das fórmulas de *Sturges*, *Doane* e *Larson* suprimem alguns detalhes da distribuição, que são claramente perceptíveis quando adotadas os critérios de *Scott*, *Rice* e, principalmente, *Freedman e Diaconis*:

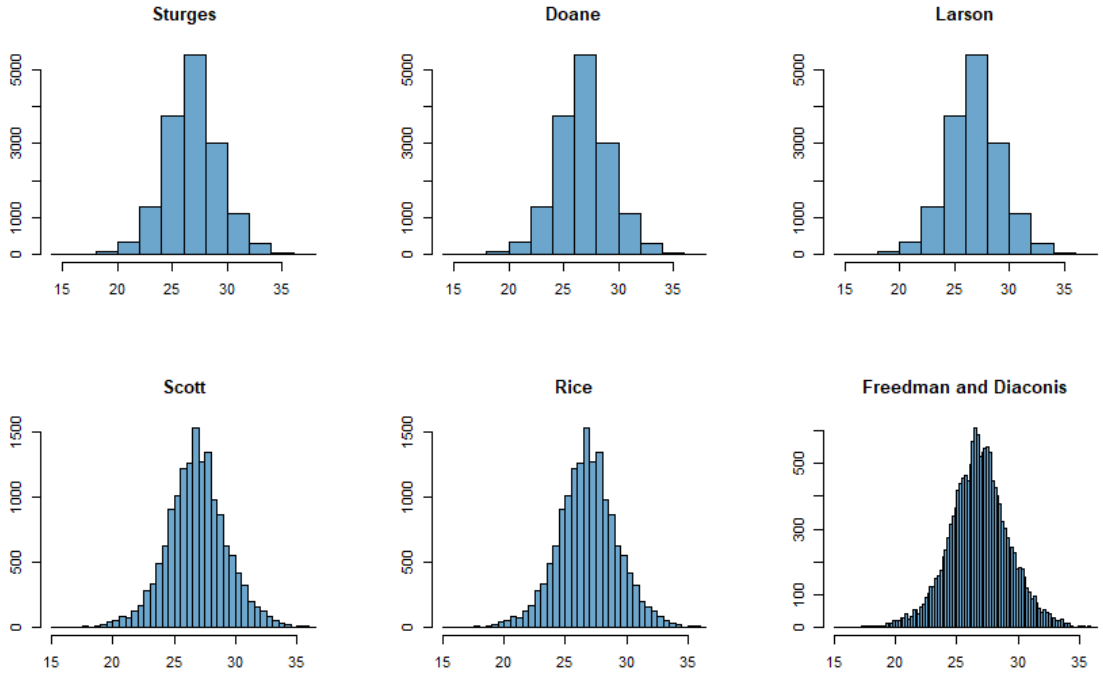

Figura 3 – Agrupamentos de classes conforme *Sturgers*, *Doane*, *Larson*, *Scott*, *Rice* e *Freedman e Diaconis*.

Por essa razão, para a presente pesquisa, optou-se pelo procedimento de *Freedman e Diaconis* para o agrupamento dos dados, cuja fórmula é a seguinte:

$$h = 2\frac{IQR(x)}{\sqrt[3]{n}}, \tag{1}$$

onde $IQR(x)$ = intervalo interquartílico.

A fim de modelar o comportamento desta variável climática, foram consideradas as seguintes distribuições de probabilidade: Gama, Gumbel, Log-normal e Normal:

a) Gama: uma variável aleatória X tem essa distribuição com parâmetros

$k > 0$ e $\theta > 0$, se sua função densidade de probabilidade é dada por:

$$f(x;k,\theta) = \begin{cases} \dfrac{x^{k-1}e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)}, \text{se } x > 0 \ e \ k, \theta > 0 \\ 0, \text{ caso contrário,} \end{cases} \tag{2}$$

em que:

θ = parâmetro de escala;

k = parâmetro de forma;

e = 2,718282;

$\Gamma(k) = \int_0^\infty x^{k-1}e^{-x}dx, \text{se } k > 0.$

Notação: $X \sim$ Gama(k, θ).

Sua esperança é dada por:

$$E(X) = k\theta, \tag{3}$$

onde $-\infty < k\theta < +\infty$,

e a sua variância é dada por:

$$Var(X) = k\theta^2, \tag{4}$$

onde $0 < k\theta^2 < +\infty$.

A distribuição Gama pode ser parametrizada em termos de $\alpha = k$ (parâmetro de forma) e $\beta = 1/\theta$ (parâmetro de escala inversa), em que:

$$\theta = \frac{\sigma^2}{\mu}, \tag{5}$$

onde $\theta > 0$,

e,

$$k = \frac{\mu}{\theta}, \tag{6}$$

onde $k > 0$.

b) Gumbel: uma variável aleatória X tem essa distribuição com parâmetros α e β, com $-\infty < \alpha < +\infty$ e $\beta > 0$, se sua função de distribuição acumulada é dada

por:

$$F(x) = \exp\left(-e^{-(x-\alpha)/\beta}\right), \tag{7}$$

em que:

α = parâmetro de localização;

β = parâmetro de escala;

exp = $e = 2,718282$

Notação: $X \sim$ Gum(α, β).

Sua esperança é dada por:

$$E(X) = \alpha + \beta\gamma, \tag{8}$$

onde $\gamma \approx 0,577216$ (constante de Euler–Mascheroni),

e a sua variância é dada por:

$$Var(X) = \frac{\pi^2\beta^2}{6}. \tag{9}$$

c) Log-normal: uma variável aleatória X tem essa distribuição com parâmetros μ e σ^2, com $-\infty < \mu < +\infty$ e $\sigma^2 > 0$, se sua função densidade de probabilidade é dada por:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(a) - \mu)^2}{2\sigma^2}\right), \tag{10}$$

em que:

σ = desvio-padrão da distribuição;

π = 3,141593;

exp = 2,718282;

μ = média da distribuição;

σ^2 = variância da distribuição.

Notação: $X \sim$ Lognormal(μ, σ^2).

Sua esperança é dada por:

$$E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right), \tag{11}$$

e a sua variância é dada por:

$$Var(X) = \left[\exp\left(\sigma^2\right) - 1\right] \exp\left(2\mu + \sigma^2\right). \tag{12}$$

d) Normal: uma variável aleatória X tem essa distribuição com parâmetros μ e σ^2, com $-\infty < \mu < +\infty$ e $0 < \sigma^2 < +\infty$, se sua função densidade de probabilidade é dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \quad -\infty < x < \infty, \tag{13}$$

em que:

μ = média da distribuição;

σ = desvio-padrão da distribuição;

π = 3,141593;

exp = 2,718282.

Notação: $X \sim$ N(μ, σ^2).

Sua esperança é dada por:

$$E(X) = \mu, \tag{14}$$

onde $-\infty < \mu < +\infty$,

e a sua variância é dada por:

$$Var(X) = \sigma^2, \tag{15}$$

onde $0 < \sigma^2 < +\infty$.

Para a realização dos cálculos das distribuições de probabilidade, foram utilizadas as respectivas funções provenientes do *software R*: *dgamma()*, *dlnorm()* e *dnorm()*[2]. Contudo, foi necessário criar uma função denominada *dgumbel()*, que efetuasse o cálculo da função de distribuição acumulada da função de Gumbel.

[2]As funções *dgamma()*, *dlnorm()* e *dnorm()* estão contidas no pacote *Stats* cuja documentação pode ser consultada aqui.

3 Resultados

Considerando que, pelo método da máxima verossimilhança (MVS), tem-se os seguintes estimadores para média e variância populacional (UFSC, 2012):

$$E[\overline{x}] = \mu = 26,86551 \tag{16}$$

$$E[S^2] = \sigma^2 = 6,019498, \tag{17}$$

pode-se determinar os parâmetros de cada uma das distribuições anteriormente descritas:

Tabela 1 – Parâmetros das distribuições Gama, Gumbel, Log-normal e Normal calculados.

Distribuição	α	β
Gama	119,903	4,463082
Gumbel	25,76132	1,91296
	μ	σ^2
Normal	26,86551	6,019498
Log-normal	3,2866913	0,008305489

Fonte: Elaborado pelos autores (2022).

A Figura 4 mostra como, a partir dos parâmetros calculados acima, a distribuição dos dados e as respectivas curvas dos modelos se ajustam. Nota-se que as distribuições Gama, Log-normal e Normal apresentam um comportamento similar, aderindo visualmente bem aos dados, diferentemente da curva relativa à função de Gumbel, que aparentemente não demonstra o mesmo nível de adequação:

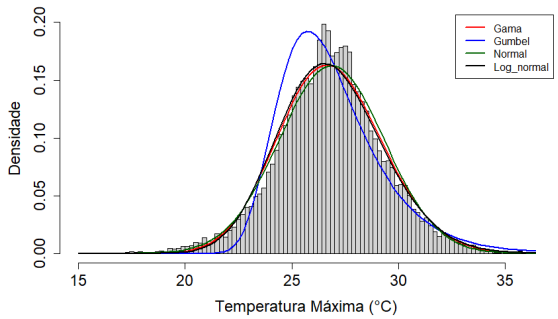

Figura 4 – Ajustamento visual das distribuições Gama, Gumbel, Log-normal e Normal à distribuição de temperaturas máximas diária em Brasília entre 1980 e 2021.

Após a etapa visual, é necessário aplicar testes de aderência que sejam capazes de quantificar a capacidade de ajustamento de cada tipo de distribuição à série de dados em análise. Para tanto, as respectivas aderências foram avaliadas utilizando os seguintes testes não paramétricos:

a) Kolmogorov-Smirnov, cuja estatística pode ser obtida por (BUSSAB; MORETTIN, 2017, pág. 430):

$$KS = max_{1 \leq i \leq n} |F(x_i) - F_e(x_i)|, \tag{18}$$

em que é calculada a discrepância entre a função da distribuição empírica e a função de distribuição esperada.

b) Qui-Quadrado, cuja estatística pode ser obtida por (BUSSAB; MORETTIN, 2017, pág. 418):

$$\chi^2 = \sum_{i=1}^{S} \frac{(O_i - E_i)^2}{E_i}, \tag{19}$$

em que O_i é o valor que foi observado para determinada classe e E_i o valor que era esperado.

c) Lilliefors, cuja estatística pode ser obtida por (UEL, 2014):

$$LF = max\{D^+, D^-\}, \tag{20}$$

para o que,

$$D^+ = max_{i=1,\ldots,n}\{S(X) - E(X)\}, \tag{21}$$

e

$$D^+ = max_{i=1,\ldots,n}\{F(X) - E(X)\}, \tag{22}$$

em que S(X) corresponde à função de distribuição empírica e E(X) à função de distribuição acumulada da função Normal.

d) Anderson-Darling, cuja estatística pode ser obtida por (UEL, 2012):

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}\{(2i-1) \times log_e[S(X)] + [2(n-i)+1] \times log_e[1 - S(X)]\}, \tag{23}$$

em que S(X) corresponde à função de distribuição empírica.

A realização dos testes de aderência levou em conta, para todos os casos, um nível de significânica de 5% e as seguintes hipóteses:

- H_0: o conjunto de dados se ajusta à distribuição (ou Gama ou Gumbel ou Log-normal ou Normal).
- H_1: o conjunto de dados **não** se ajusta a determinada distribuição (ou Gama ou Gumbel ou Log-normal ou Normal).

Conforme se pode vislumbrar pelas tabelas de resultados que serão apresentadas, quando foram realizados os testes a partir da integralidade da base de dados, os respectivos p-valores foram sempre muito baixos, rejeitando todas as hipóteses nulas, de forma que se teria que concluir que nenhuma das distribuições em análise aderia suficientemente aos dados, a despeito do que parecia indicar a Figura 4.

A explicação para esse fenômeno pode estar no alerta de que, via de regra, os testes de significância são pouco úteis para n pequeno (menor que 30) e demasiadamente sensíveis para n grande (maior que 1000) (HAIR et al., 2009, pág. 84).

A solução para contornar essa dificuldade foi a utilização de *bootstrappings*, uma abordagem em que se valida determinado modelo a partir da testagem de uma grande quantidade de amostras (HAIR et al., 2009, pág. 84). Essa abordagem foi adotada para todos os testes de aderência, definindo-se, para isso, 1000 reamostragens para $n = 500$. Posteriormente, foi calculada a média para a estatística de cada teste e o $p - valor$ respectivo para a realização do teste de hipóteses.

Com vistas à exemplificação da abordagem de *bootstrappings*, a Figura 5 contém a distribuição das estatísticas do teste de Kolmogorov-Smirnov aplicado a cada uma das distribuições em análise:

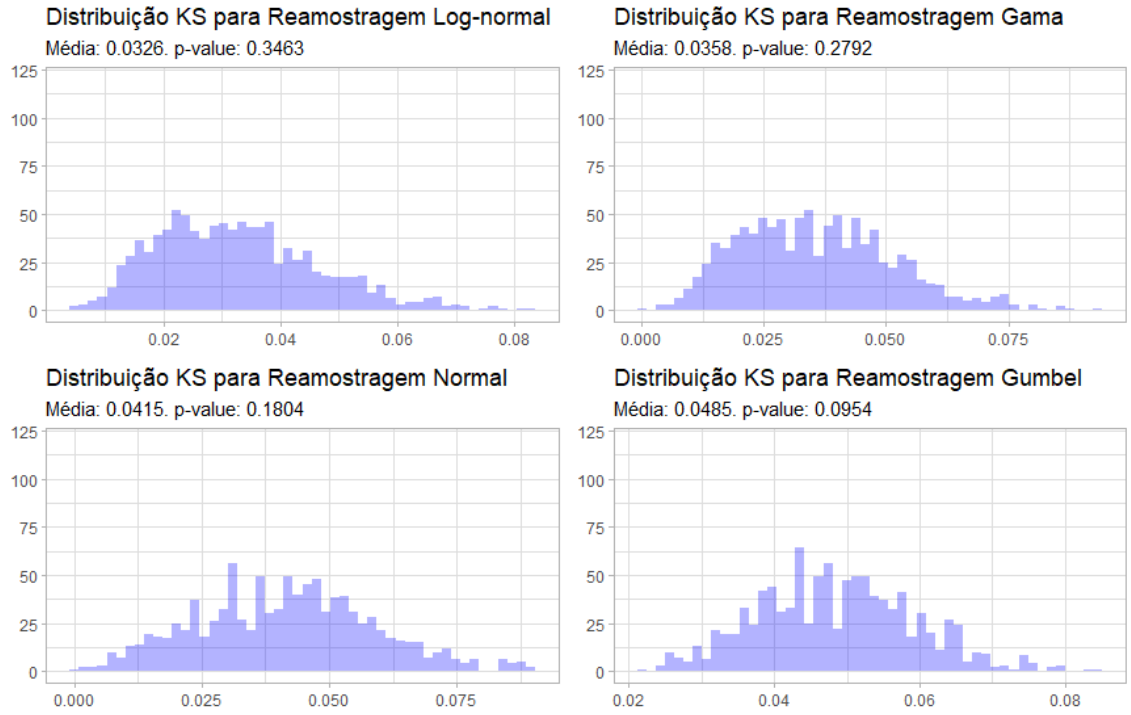

Figura 5 – Distribuição das estatísticas do teste de aderência de *Kolmogorov-Smirnov* provenientes de *bootstrapping* com 1000 reamostragens e $n = 500$.

A Figura 5 revela, a partir da abordagem de *bootstrapping*, que nenhuma H_0 (hipótese de que determinada distribuição teórica adere à distribuição empírica) foi rejeitada e o maior p-valor foi encontrado para a distribuição Log-normal.

Realizados todos os demais testes, foram obtidos os seguintes resultados para cada distribuição:

- Log-normal:

Tabela 2 – 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).

- Normal:

Tabela 3 – Resultados dos testes não paramétricos para a distribuição Normal.

Teste	**Estatística**	**P-valor**
Kolmogorov-Smirnov Test	0,032155	0,00000000000001857
Bootstrap Kolmogorov-Smirnov Test (1.000 de n=500)	0,04239254	0,1804151
Lilliefors Test	0,032155	0,00000000000000022
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	0,00000000000000022
Bootstrap Anderson Darling Test (1.000 de n=500)	1,65203	0,144019

Fonte: Elaborado pelos autores (2022).

- Gama:

Tabela 4 – 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).

- Gumbel:

Tabela 5 – 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).

Os resultados mais significativos para os testes de aderência foram os das distribuições Log-normal, Normal e Gama. A distribuição de Gumbel, como já parecia indicar a Figura 4, obteve os resultados menos satisfatórios, sendo prontamente descartada.

A partir dos parâmetros calculados na Tabela 1, puderam ser desenhadas as curvas teóricas de Log-normal, Normal e Gama para a distribuição das temperaturas máximas em Brasília no período selecionado, conforme ilustra a Figura 6:

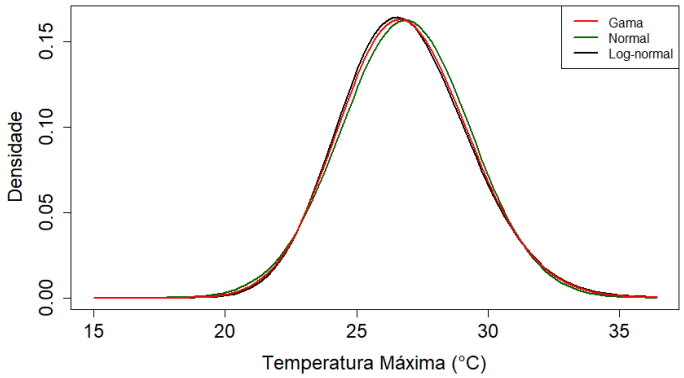

Figura 6 – Curvas das distribuições teóricas Log-normal, Normal e Gama.

É possível notar, conforme já mostrava a Figura 4, que, visualmente, a partir dos parâmetros utilizados, essas 3 curvas são bem semelhantes. Assim, para uma melhor compreensão dessas distribuições teóricas, é recomendável calcular os momentos relacionados a elas. Os quatro primeiros momentos estatísticos de uma distribuição revelam importantes medidas estatísticas (KUJAWSKA, 2021). Portanto, a partir disso, será possível verificar se uma dessas três distribuições reflete melhor os dados iniciais de temperaturas máximas em Brasília.

Para distribuições com variáveis contínuas, os momentos de ordem k são calculados pela equação:

$$E(X^k) = \int_{-\infty}^{+\infty} x^k f(x) dx, \tag{24}$$

em que $f(x)$ é a função densidade de probabilidade.

Por sua vez, os momentos centrais de ordem k são calculados pela equação:

$$E\left[(X-\mu)^k\right] = \int_{-\infty}^{+\infty} (x-\mu)^k f(x) dx, \tag{25}$$

em que $f(x)$ é a função densidade de probabilidade.

Por meio do pacote *moments*, no *R*, foi possível calcular com facilidade os quatro primeiros momentos relativos à distribuição em questão:

Tabela 6 – Comparações entre os 4 primeiros momentos das distribuições teóricas de Log-normal, Normal e Gama.

Distribuição	**1º Momento**	**2º Momento**	**3º Momento**	**4º Momento**
Log-normal	26,87089	6,018530	3,9025727	113,0211
Normal	26,83083	6,028461	0,6248341	107,2331
Gama	26,88609	5,995935	2,7405657	109,4008

Fonte: Elaborado pelos autores (2022).

Os primeiros dois momentos fornecem, respectivamente, o valor esperado (m'_1) e a variância (m_2) da distribuição teórica. Já os demais momentos são necessários para os cálculos da assimetria e da curtose:

$$assimetria = \frac{m_3}{(m_2)^{2/3}}, \tag{26}$$

assim tem-se assimetria = 0,264311.

$$curtose = \frac{m_4}{(m_2)^2}, \tag{27}$$

e tem-se curtose = 3,120173.

A Tabela 7 compara as medidas geradas por esses momentos, para cada distribuição teórica, com as que inicialmente descreviam a distribuição empírica, com destaque para as menores diferenças:

Tabela 7 – Comparações da média, variância, assimetria e curtose entre a distribuição empírica e as teóricas (Log-normal, Normal e Gama).

Distribuição	**Média**	**Variância**	**Assimetria**	**Curtose**
Empírica	26,86551	6,019498	-0,03713757	3,672045
Log-normal	26,87089	6,018530	0,26431102	3,120173
Normal	26,83083	6,028461	0,04221386	2,950638
Gama	26,88609	5,995935	0,18666156	3,043033

Fonte: Elaborado pelos autores (2022).

À exceção da medida para a assimetria, a distribuição de Log-normal apresentou as menores diferenças para com a distribuição empírica. Somando-se a isso o fato

de ela ter produzido o maior $p - valor$ quando da realização do teste de aderência de Kolmogorov-Smirnov (*Boostrapping*), pode-se afirmar que, dentre as distribuições consideradas, ela é a que melhor se ajusta aos dados estudados.

Conclusões

Por fim, a partir dos resultados obtidos, pode-se afirmar que, para a distribuição de temperaturas máximas de Brasília durante 1980 e 2021:

- considerando todos os testes de aderência realizados (Kolmogorov-Smirnov, Qui-quadrado e Anderson Darling), a distribuição com o pior ajuste foi a de Gumbel I;
- as distribuições Log-normal, Normal e Gama obtiveram os melhores resultados, não tendo sua H_0 (os dados se ajustam a determinada distribuição) rejeitada para os testes de Kolmogorov-Smirnov (*bootstrapping*) e Anderson Darling (*bootstrapping*);
- A partir do cálculo dos momentos, pode-se observar que a distribuição teórica da Log-normal obteve as menores diferenças para a média (1º momento), variância (2º momento) e curtose se comparadas às mesmas medidas da distribuição empírica. Por esse critério, a distribuição Log-normal foi considerada a que melhor se ajustou aos dados originais.

Tendo em vista que o presente estudo objetivou analisar unicamente o melhor ajuste à distribuição das temperaturas máximas, indica-se, a futuros estudos, que se leve em conta também as datas em que essas temperaturas ocorreram. Análises de séries temporais, por exemplo, podem ser ferramentas bastante úteis para previsões de temperaturas futuras, considerando, inclusive, a sazonalidade a que o comportamento dessa variável está sujeita.

Referências

ARAUJO, E. M. et al. Aplicação de seis distribuições de probabilidade a séries de temperatura máxima em iguatu - ce. *Revista Ciência Agronômica*, Centro de Ciências Agrárias - Universidade Federal do Ceará, v. 41, n. 1, p. 36–45, 2010. Citado na página 3.

ASSIS, J. P. de et al. Ajuste de sete modelos de distribuições densidade de probabilidade às séries históricas de umidade relativa mensal em mossoró – rn. *Revista Verde de Agroecologia e Desenvolvimento Sustentável*, v. 8, n. 1, p. 01–10, 2013. Citado 2 vezes nas páginas 2 e 3.

BUSSAB, W. O.; MORETTIN, P. A. *Estatística Básica*. 9. ed. São Paulo: Saraiva, 2017. Citado na página 12.

CARDOSO, M. R. D.; MARCUZZO, F. F. N.; BARROS, J. R. Classificação climática de köpen-geiger para o estado de goiás e o distrito federal. *ACTA Geográfica*, v. 8, n. 16, p. 40–55, 2014. Citado na página 3.

COSTA, L. C.; SEDIYAMA, G. C. Elementos climáticos e produtividade agrícola. *Ação Ambiental*, v. 2, n. 7, p. 24–28, 1999. Citado na página 2.

DOGAN, N.; DOGAN, I. Determination of the number of bins / classes used in histograms and frequency tables: a short bibliography. *Journal of Statistical Research*, v. 07, n. 02, p. 77–86, 2010. Citado na página 6.

ESTEFANEL, V.; SCHNEIDER, F. M.; BURIOL, G. A. Probabilidade de ocorrência de temperaturas máximas do ar prejudiciais aos cultivos agrícolas em santa maria, rs. *Revista Brasileira de Agrometeorologia*, v. 2, n. 1, p. 57–63, 1994. Citado na página 2.

HAIR, J. F. et al. *Análise multivariada de dados. 6ª ed.* Rio de Janeiro: Bookman, 2009. Citado na página 13.

HARARI, Y. N. *21 Lições para o século 21*. 1. ed. São Paulo: Companhia das Letras, 2018. Citado na página 2.

KUJAWSKA, A. *Statistical Moments in Data Science interviews*. [S.l.], 2021. Disponível em: <https://towardsdatascience.com/statistical-moments-in-data-science-interviews-bfecd207843d>. Acesso em: 15 jun. 2022. Citado na página 16.

MOTA, F. S. da; ROSKOFF, J. L. da C.; SILVA, J. B. da. Probabilidade de ocorrência de dias com temperaturas iguais ou superiores a 35 graus no florescimento de arroz no rio grande do sul. *Revista Brasileira de Agrometeorologia*, v. 7, p. 147–149, 1999. Citado na página 2.

SUPIAN, N. M.; HASAN, H. B. Selecting the probability distribution of annual maximum temperature in malaysia. *ITM Web of Conferences*, v. 36, 2021. Citado na página 3.

UEL. *Anderson-Darling*. [S.l.], 2012. Disponível em: <http://www.uel.br/projetos/experimental/pages/arquivos/Anderson_Darling.html>. Acesso em: 12 jun. 2022. Citado na página 12.

UEL. *Teste de Lilliefors*. [S.l.], 2014. Disponível em: <http://www.uel.br/projetos/experimental/pages/arquivos/Lilliefors.html#>. Acesso em: 12 jun. 2022. Citado na página 12.

UFSC. *Estimação de Parâmetros*. [S.l.], 2012. Disponível em: <https://www.inf.ufsc.br/~andre.zibetti/probabilidade/estimacao-de-parametros.html>. Acesso em: 10 jun. 2022. Citado na página 11.