

Aplicação das Distribuições de Probabilidade T de Student e Normal a séries de temperaturas máximas em Brasília–DF, Brasil

Evaluation of T Student and Normal probability distributions at maximum temperature series in Brasília–DF, Brazil

Guilherme Rocha Duarte¹

Resumo - O estudo analisa as séries temporais de temperatura máxima em Brasília, entre 1980 e 2023, considerando distribuições Normal e T de Student. Aborda a importância da compreensão das variações climáticas e propõe determinar a distribuição mais adequada para modelar os dados climáticos da região.

Palavras-chave - Distribuição de frequência. Séries temporais. Teste de aderência. T de Student. Distribuição normal.

Abstract - The study analyzes the maximum temperature time series in Brasília, from 1980 to 2023, considering Normal and Student's t distributions. It addresses the importance of understanding climate variations and aims to determine the most suitable distribution to model the region's climate data.

Keywords - Frequency distribution. Time series. Goodness-of-fit test. T Student. Normal distribution.

¹ Aluno de graduação em Ciência de Dados e Inteligência Artificial no Instituto de Educação Superior de Brasília - IESB.

Introdução

A análise estatística das séries temporais de temperaturas máximas desempenha um papel fundamental na compreensão das variações climáticas em uma determinada região.

Localizada no coração do Planalto Central brasileiro, Brasília é conhecida por suas características climáticas distintas, marcadas por uma estação seca durante o inverno e uma estação chuvosa no verão. Sua localização geográfica privilegiada, com uma altitude média de aproximadamente 1.100 metros acima do nível do mar, contribui para a formação de um clima tropical de altitude, influenciado por fatores como a topografia da região, os ventos predominantes e a sazonalidade das chuvas.

A compreensão das distribuições de probabilidade é essencial para modelar e interpretar adequadamente os padrões climáticos observados em Brasília. A distribuição normal, também conhecida como gaussiana, é amplamente utilizada na análise estatística devido à sua simetria e à forma de sino característica, sendo aplicável em muitos contextos climáticos. Por outro lado, a distribuição t de Student, desenvolvida por William Sealy Gosset no início do século XX, é particularmente útil quando se lida com amostras pequenas ou quando a variabilidade dos dados não pode ser adequadamente representada pela distribuição normal.

Ao aplicar essas distribuições de probabilidade às séries de temperaturas máximas em Brasília, é possível investigar uma variedade de questões climáticas relevantes, incluindo a variabilidade sazonal das temperaturas, a identificação de tendências de longo prazo e a avaliação de eventos extremos, como ondas de calor ou períodos de frio intenso. Além disso, a análise probabilística das temperaturas máximas pode oferecer visões valiosas para setores como agricultura, saúde pública, planejamento urbano e gestão de recursos hídricos, ajudando a mitigar os impactos adversos das mudanças climáticas na região.

O objetivo deste estudo foi determinar, entre as distribuições de probabilidade Normal e T de Student, a mais adequada para modelar as séries diárias de temperatura máxima em Brasília, Distrito Federal, ao longo do período entre 1980 a 2023, considerando o seu clima tropical de altitude.

Material e métodos

Para analisar as distribuições normal e t dos dados de temperatura máxima diária na cidade de Brasília, foi adotada uma metodologia consistente, considerando as particularidades climáticas e geográficas da região. Os dados utilizados foram obtidos junto ao órgão competente (INMET), abrangendo um período de 43 anos, de 1980 a 2023.

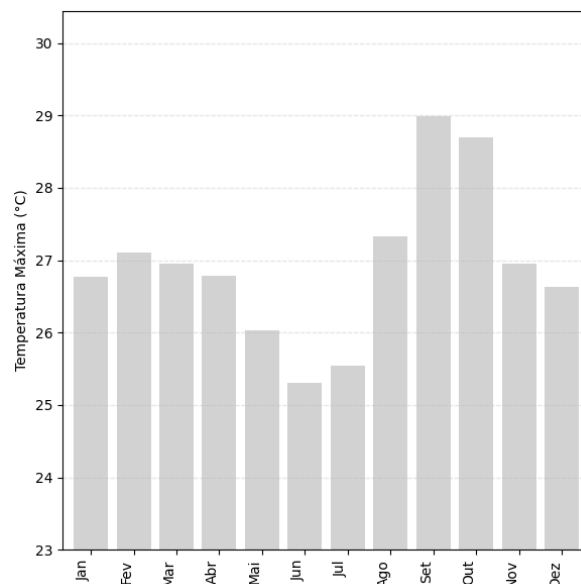


Figura 1 - Médias mensais de temperatura máxima em Brasília.

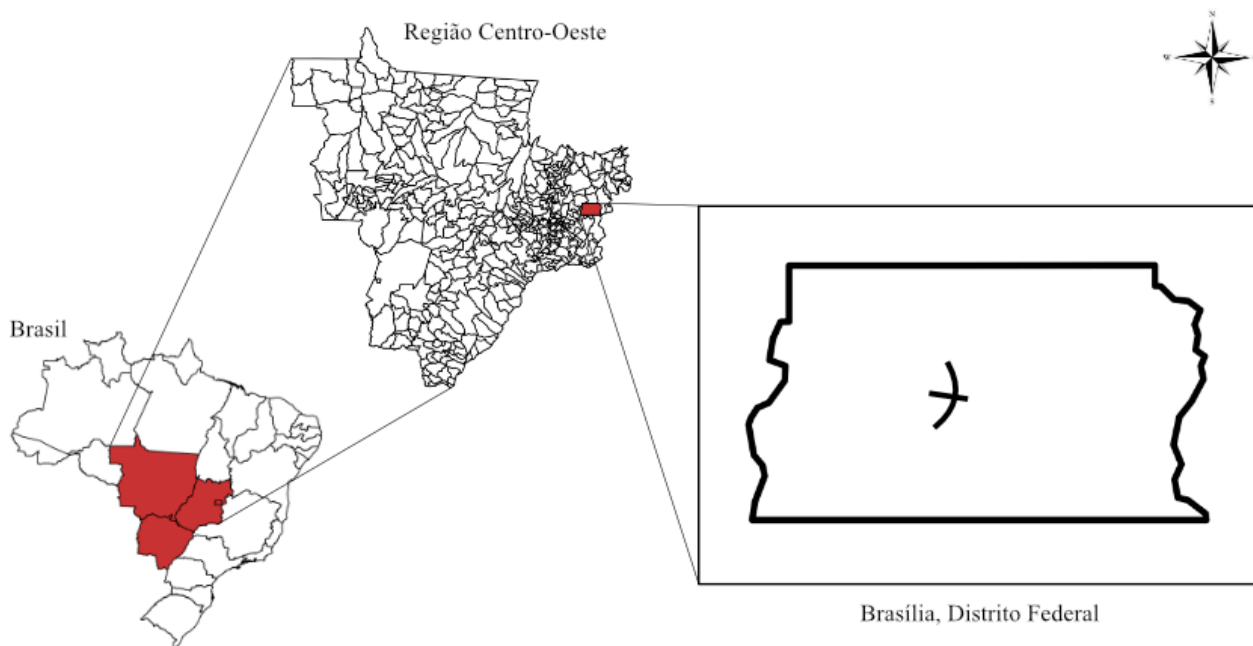


Figura 2 - Localização da cidade em estudo

No contexto climático, Brasília é caracterizada por um clima tropical de altitude, com duas estações bem definidas: uma estação seca e outra chuvosa. A temperatura máxima diária é influenciada por essas variações sazonais, bem como pela topografia da região e sua altitude média de 1.158 metros acima do nível do mar. Esse contexto proporciona condições ideais para a realização de estudos climáticos.

Foram examinadas as distribuições normal e T de Student. A distribuição normal é frequentemente utilizada para modelar fenômenos naturais, enquanto a distribuição t é empregada em situações em que o tamanho da amostra é pequeno ou a variabilidade da população é desconhecida. Nesse contexto, foram realizadas análises estatísticas para avaliar a adequação dessas distribuições aos dados de temperatura máxima diária em Brasília.

A abordagem metodológica adotada neste estudo visa compreender os padrões climáticos da cidade e modelar a distribuição da temperatura máxima diária, considerando as características específicas do clima local. Isso proporcionará insights importantes para o entendimento do clima de Brasília e contribuirá para estudos futuros relacionados ao clima e suas variações.

Na distribuição normal, a notação $X \sim N(\mu, \sigma^2)$ é comumente usada para denotar que a variável

aleatória X segue uma distribuição normal com média μ e variância σ^2 . Para valores de $-\infty < \mu < \infty$ e $\sigma > 0$, a função de densidade de probabilidade é definida como:

$$f(x) = \frac{e^{\frac{-(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma}}$$

Figura 3 - Função Distribuição Normal

Acima, a função de densidade de probabilidade da distribuição normal é ilustrada para diferentes valores de μ e σ .

μ (mu): Representa a média da distribuição normal. Indica o ponto central da distribuição, onde a maioria dos dados tende a se concentrar.

σ (sigma): Representa o desvio padrão da distribuição normal. Indica a dispersão dos dados em torno da média. Quanto maior o valor de σ , maior é a dispersão dos dados.

A função de densidade de probabilidade mostra como a probabilidade está distribuída ao longo do eixo x para diferentes valores de μ e σ . Em uma distribuição

normal, a forma da curva é simétrica em relação à média μ , e o desvio padrão σ influencia a largura da curva. Quanto maior o desvio padrão, mais achatada é a curva.

A distribuição normal encontra aplicações em diversos campos, como modelagem de alturas de adultos, pesos de recém-nascidos e temperaturas de cidade. Além disso, ela pode ser uma boa aproximação para a distribuição binomial em casos onde n é grande e p está próximo de $1/2$, assim como para a distribuição de Poisson quando n é grande e p é pequeno. O teorema do limite central destaca a utilidade da distribuição normal na modelagem de variáveis aleatórias que podem ser consideradas a soma de várias variáveis aleatórias independentes. Na figura abaixo, é ilustrada a função de densidade de probabilidade para $\mu = 3$ e dois diferentes valores de σ .

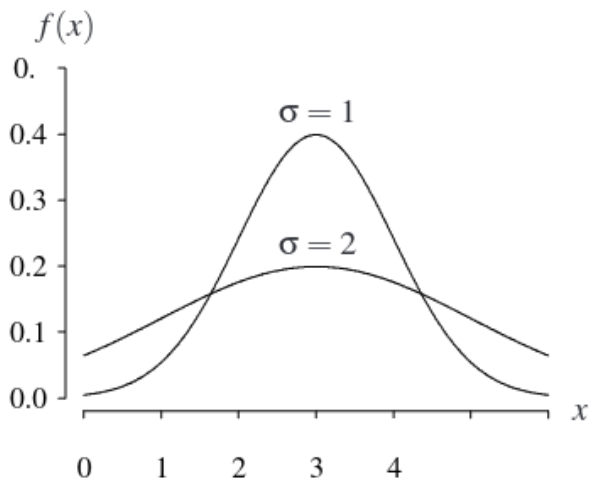


Figura 4 - Normal para $\mu = 3$ e para $\sigma=1$ e $\sigma=2$.

A distribuição normal, também conhecida como distribuição gaussiana, é uma das distribuições de probabilidade mais importantes. É usada para modelar uma ampla variedade de fenômenos naturais e humanos, devido à sua propriedade de simetria e ao Teorema do Limite Central. Ela descreve muitos conjuntos de dados na qual a maioria das observações se agrupa em torno da média, com poucas observações mais distantes. Além disso, a função densidade de probabilidade normal é usada em testes de hipóteses, intervalos de confiança e muitas outras análises estatísticas.

Os parâmetros populacionais de média, variância, assimetria e curtose de X são, respectivamente:

$$E[X] = \mu$$

$$V[X] = \sigma^2$$

$$E[(X - \mu)/\sigma^3] = 0$$

$$E[(X - \mu)/\sigma^4] = 3$$

Na distribuição T, a abreviação $X \sim t(n)$ é utilizada para indicar que a variável aleatória X segue com parâmetro n positivo e real, conhecido como graus de liberdade. Na maioria das aplicações, n é um número inteiro positivo. Uma variável aleatória t X com n graus de liberdade possui função de densidade de probabilidade:

$$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\sqrt{n\pi}} \frac{1}{\left(1 + \frac{x^2}{n}\right)^{\frac{(n+1)}{2}}}; x \in \mathbb{R} \text{ e } n > 0$$

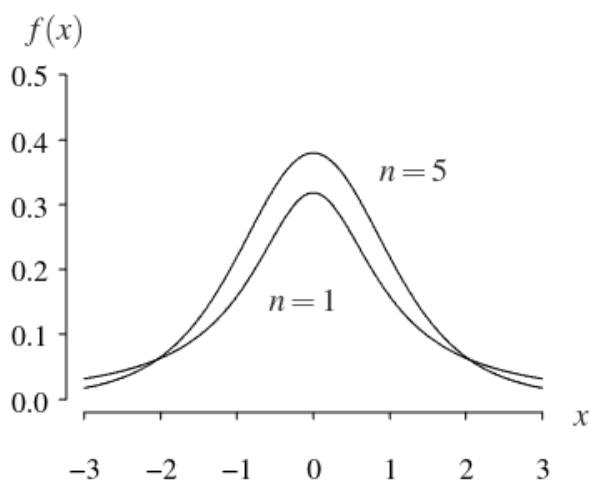
Figura 5 - Função Distribuição T de Student

Na figura acima, a função de densidade de probabilidade da distribuição t de Student é ilustrada para diferentes valores de graus de liberdade, representados por n .

n : Representa os graus de liberdade da distribuição t. Indica o número de observações independentes na amostra. Quanto maior o valor de n , mais a distribuição t se aproxima de uma distribuição normal.

A função de densidade de probabilidade mostra como a probabilidade está distribuída ao longo do eixo x para diferentes valores de n . A distribuição t é semelhante à distribuição normal, mas possui caudas mais pesadas, especialmente quando o número de graus de liberdade é baixo. Isso significa haver uma maior probabilidade de observar valores extremos em comparação com a distribuição normal, especialmente em amostras pequenas.

A distribuição t surge em testes de hipóteses relacionados à comparação de (a) uma média amostral com um padrão, ou (b) a diferença entre duas médias. A função de densidade de probabilidade para $n = 1$ (um caso especial da distribuição t conhecido como distribuição de Cauchy) e $n = 5$ está ilustrada abaixo.



$$E[X] = 0, \text{ para } n > 1$$

$$V[X] = n/(n - 2), \text{ para } n > 2$$

$$E[(X - \mu)/\sigma]^3 = 0, \text{ para } n > 3$$

$$E[(X - \mu)/\sigma]^4 = [3(n - 2)]/(n - 4), \text{ para } n > 4$$

Distribuições Mensais das Temperaturas Máximas

A seguir, são apresentadas as distribuições Normal e T de Student das temperaturas máximas para cada mês do período em estudo:

Figura 6 - Distribuição de Cauchy e T quando n=5

Os parâmetros populacionais de média, variância, assimetria e curtose de X são, respectivamente:

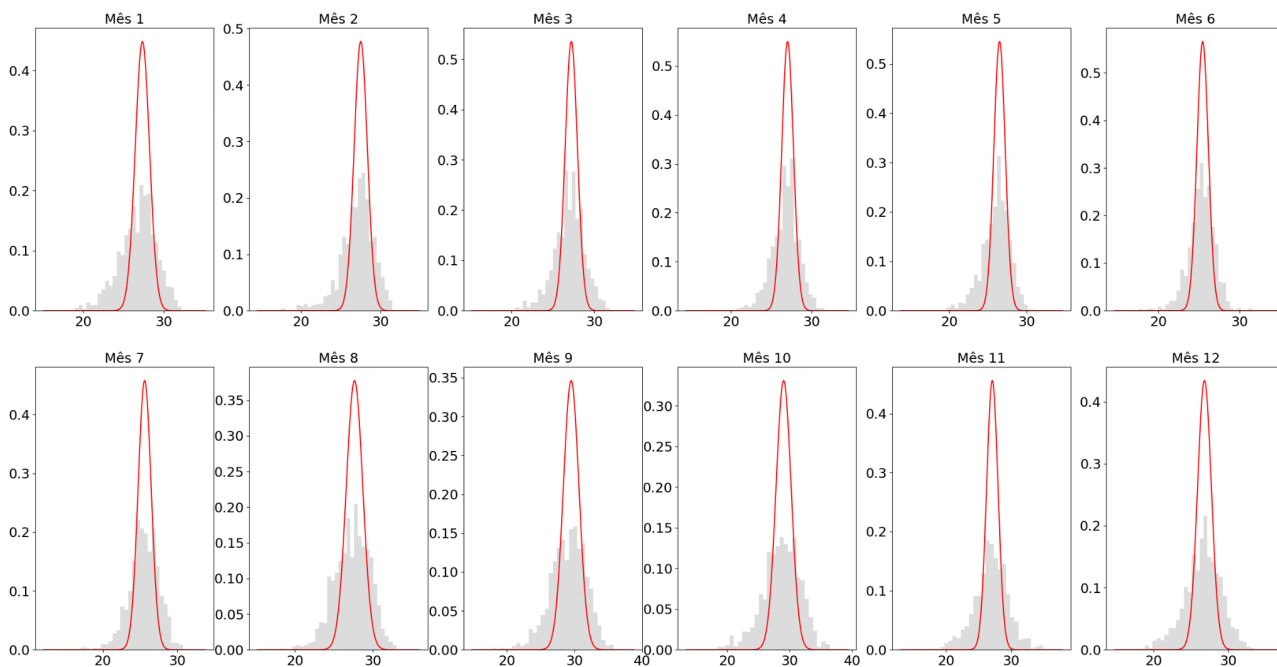


Figura 7: Distribuição Normal de cada mês, sendo o eixo x a temperatura e o eixo y a densidade de probabilidade.

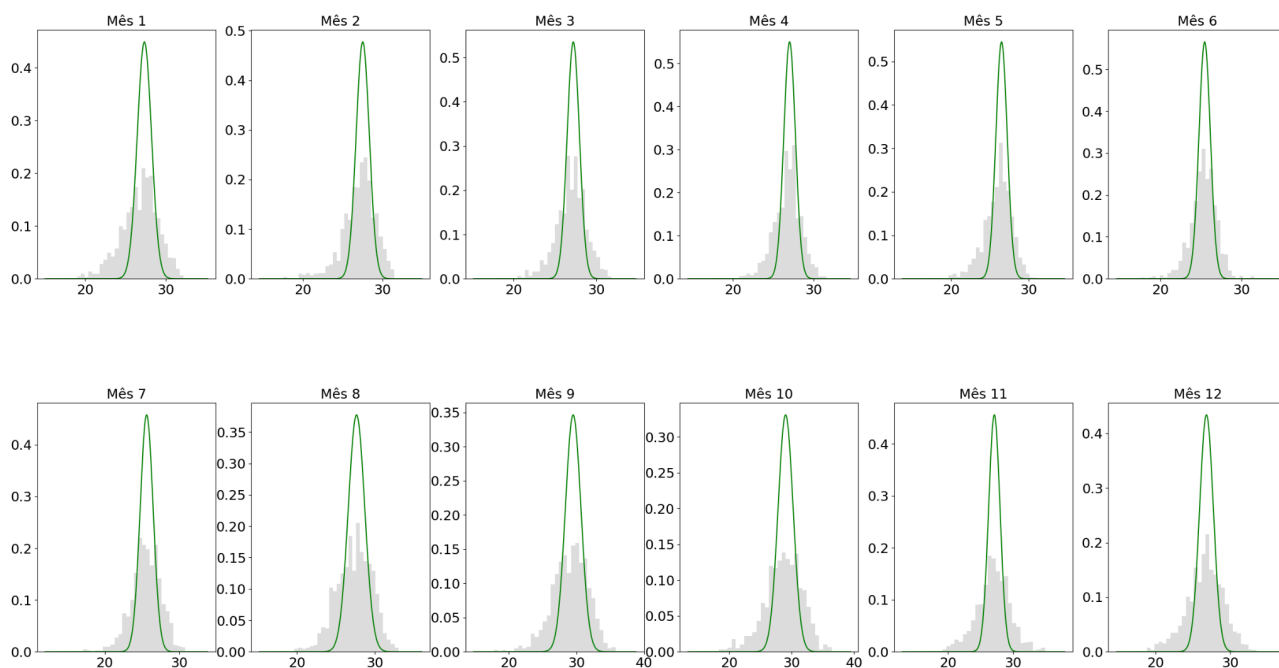


Figura 8: Distribuição T de cada mês, sendo o eixo x a temperatura e o eixo y a densidade de probabilidade.

Resultados e discussão

Análise dos Testes de Aderência

Para determinar a adequação das distribuições Normal e T de Student aos dados de temperatura máxima diária em Brasília, foram realizados vários testes de aderência, incluindo os testes de Cramer, Kolmogorov-Smirnov, Qui-Quadrado, Shapiro, Lilliefors e Anderson. A seguir, serão apresentados os resultados desses testes.

Tabela 1 - Aderência das funções distribuição de probabilidade para o teste Cramer. Onde (*) - significativo e (n) - não significativo, ao nível de 5% de probabilidade.

Função	Aderência para Cramer											
	J	F	M	A	M	J	J	A	S	O	N	D
Normal	n	n	n	n	n	n	*	*	*	*	n	*
T	n	n	n	n	n	n	n	n	n	n	n	n

No teste de Cramer, a distribuição Normal apresentou rejeição da hipótese nula (H_0) na maioria dos meses, exceto nos meses julho, agosto, setembro, outubro e dezembro. Em contraste, a distribuição T de Student teve a H_0 rejeitada para todos os meses. Esses resultados sugerem que, enquanto ambas as distribuições enfrentam dificuldades em ajustar-se perfeitamente aos dados, a distribuição Normal teve um desempenho ligeiramente melhor.

Tabela 2 - Aderência das funções distribuição de probabilidade para o teste Kolmogorov-Smirnov. Onde (*) - significativo e (n) - não significativo, ao nível de 5% de probabilidade.

Função	Aderência para Kolmogorov											
	J	F	M	A	M	J	J	A	S	O	N	D
Normal	n	n	n	n	n	n	n	n	n	*	n	n
T	n	n	n	n	n	n	n	n	n	*	n	n

O teste de Kolmogorov-Smirnov revelou que a distribuição Normal foi rejeitada na maioria dos meses, exceto no mês de outubro, assim como a distribuição T de Student.

Tabela 3 - Aderência das funções distribuição de probabilidade para o teste Qui-quadrado. Onde (*) - significativo e (n) - não significativo, ao nível de 5% de probabilidade.

Função	Aderência para Qui-Quadrado											
	J	F	M	A	M	J	J	A	S	O	N	D
Normal	*	n	n	n	n	n	*	*	*	*	*	*
T	*	n	n	n	n	n	n	n	*	n	n	*

O teste Qui-Quadrado mostrou que a distribuição Normal não rejeitou a H_0 em alguns meses (janeiro, julho, agosto, setembro, outubro, novembro e dezembro), enquanto a distribuição T de Student foi

rejeitada em todos os meses, exceto nos meses de janeiro, setembro e dezembro. Este resultado destaca a distribuição Normal como um modelo potencialmente mais adequado em alguns períodos específicos do ano.

Tabela 4 - Aderência das funções distribuição de probabilidade para o teste Shapiro. Onde (*) - significativo e (n) - não significativo, ao nível de 5% de probabilidade.

Função	Aderência para Shapiro											
	J	F	M	A	M	J	J	A	S	O	N	D
Normal	n	n	n	n	n	n	n	n	n	n	n	n

Tabela 5 - Aderência das funções distribuição de probabilidade para o teste Lilliefors. Onde (*) - significativo e (n) - não significativo, ao nível de 5% de probabilidade.

Função	Aderência para Lilliefors											
	J	F	M	A	M	J	J	A	S	O	N	D
Normal	n	n	n	n	n	n	n	n	n	n	n	n

Nos testes de Shapiro e Lilliefors, a distribuição Normal foi consistentemente rejeitada em todos os meses. Isso indica que, de acordo com esse teste, nenhuma das distribuições analisadas proporciona um ajuste perfeito aos dados de temperatura máxima diária.

Tabela 6 - Aderência das funções distribuição de probabilidade para o teste Anderson. Onde (*) - significativo e (n) - não significativo, ao nível de 5% de probabilidade.

Função	Aderência para Anderson											
	J	F	M	A	M	J	J	A	S	O	N	D
Normal	n	n	n	n	n	n	n	n	n	n	n	n

Finalmente, o teste de Anderson também rejeitou a distribuição para todos os meses, confirmando a dificuldade de ajustar os dados de temperatura máxima diária a qualquer das distribuições de forma satisfatória.

Em resumo, os resultados dos testes de aderência indicam que ambas as distribuições, Normal e T de Student, foram rejeitadas para a maioria dos dados mensais e para o conjunto total de dados. No entanto, a distribuição Normal apresentou um desempenho ligeiramente melhor, especialmente em testes como o Qui-Quadrado e Cramer. Esses achados

sugerem que, embora nenhuma das distribuições seja perfeita, a distribuição Normal pode ser considerada a melhor aproximação entre as duas distribuições testadas para modelar as temperaturas máximas diárias em Brasília.

Estatísticas Descritivas

No estudo das temperaturas de Brasília ao longo do período estudado, o mês de outubro destacou-se como um mês com características climáticas únicas. Utilizando métodos estatísticos, foram calculadas diversas métricas descritivas para compreender melhor o comportamento das temperaturas nesse mês específico. A seguir, serão apresentados os resultados da análise das temperaturas de Brasília do mês de outubro do período estudado.

A média das temperaturas registradas em Brasília durante o mês de outubro foi de 28,69 graus Celsius, sugerindo que outubro é um mês geralmente quente na capital brasileira. A variância das temperaturas em outubro foi de 8,97 graus Celsius, indicando uma moderada variabilidade nas temperaturas diárias. Isso implica que, embora as temperaturas sejam elevadas em média, há flutuações que podem trazer dias significativamente mais quentes ou mais frescos.

A curtose das temperaturas em outubro foi 0,41, um valor ligeiramente superior a zero, indicando uma distribuição levemente leptocúrtica. Isso significa que a distribuição das temperaturas tem caudas um pouco mais pesadas do que uma distribuição normal, sugerindo a presença de alguns valores extremos (dias muito quentes ou, menos frequentemente, dias mais frescos) que, embora raros, podem ocorrer.

A simetria (ou skewness) das temperaturas foi medida em -0,45, indicando uma leve assimetria à esquerda. Isto implica que a distribuição das temperaturas é ligeiramente inclinada para a esquerda, com uma prevalência de temperaturas mais altas. Em outras palavras, enquanto a maioria das temperaturas está próxima ou acima da média, existem alguns dias com temperaturas significativamente mais baixas que puxam a cauda da distribuição para a esquerda.

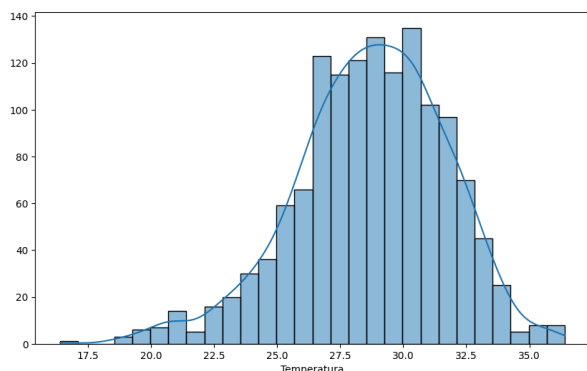


Figura 9: Distribuição de temperaturas do mês de outubro

Para complementar a análise estatística, foi gerado um histograma com a densidade KDE (Kernel Density Estimation) das temperaturas de outubro. A visualização gráfica reforça os resultados numéricos, mostrando uma distribuição com uma leve concentração de valores próximos à média, mas com algumas temperaturas extremas presentes. A leve assimetria à esquerda também é perceptível, confirmando a presença de dias com temperaturas abaixo da média, embora menos frequentes.

Conclusão

Com base nos resultados obtidos na análise da temperatura máxima em Brasília ao longo de 43 anos (1980-2023) utilizando as distribuições Normal e T de Student, concluiu-se que:

1. A distribuição Normal se mostrou mais adequada para modelar os dados de temperatura máxima diária, conforme indicado pelos testes de aderência, especialmente o teste de Cramer e o teste de Qui-Quadrado.
2. Embora nenhuma das distribuições tenha sido aceita em todos os meses, a distribuição Normal apresentou um desempenho ligeiramente melhor em comparação com a distribuição T de Student.
3. A análise das temperaturas em Brasília no mês de outubro revelou um panorama climático de um mês majoritariamente quente, com moderada variabilidade e a presença ocasional de temperaturas extremas. A distribuição Normal destacou-se pela consistência das temperaturas elevadas, enquanto indica uma leve tendência para temperaturas mais altas e a eventualidade de dias excepcionalmente quentes ou frescos.

Referências

BENISTON, M. (2012). Exploring the behaviour of atmospheric temperatures under dry conditions in Europe: evolution since the mid-20th century and projections for the end of the 21st century. **International Journal of Climatology**, 33(2), 457-462. <https://doi.org/10.1002/joc.3436>

T distribution , Normal distribution. Univariate Distribution Relationships. <https://www.math.wm.edu/~leemis/chart/UDR/UDR.html>