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 Duarte[1]

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 insights valiosos 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.

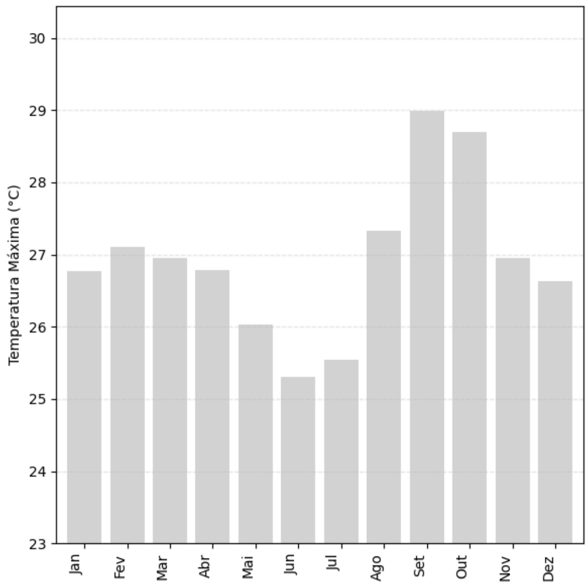


Figura 1 - Médias mensais de temperatura máxima em Brasília.

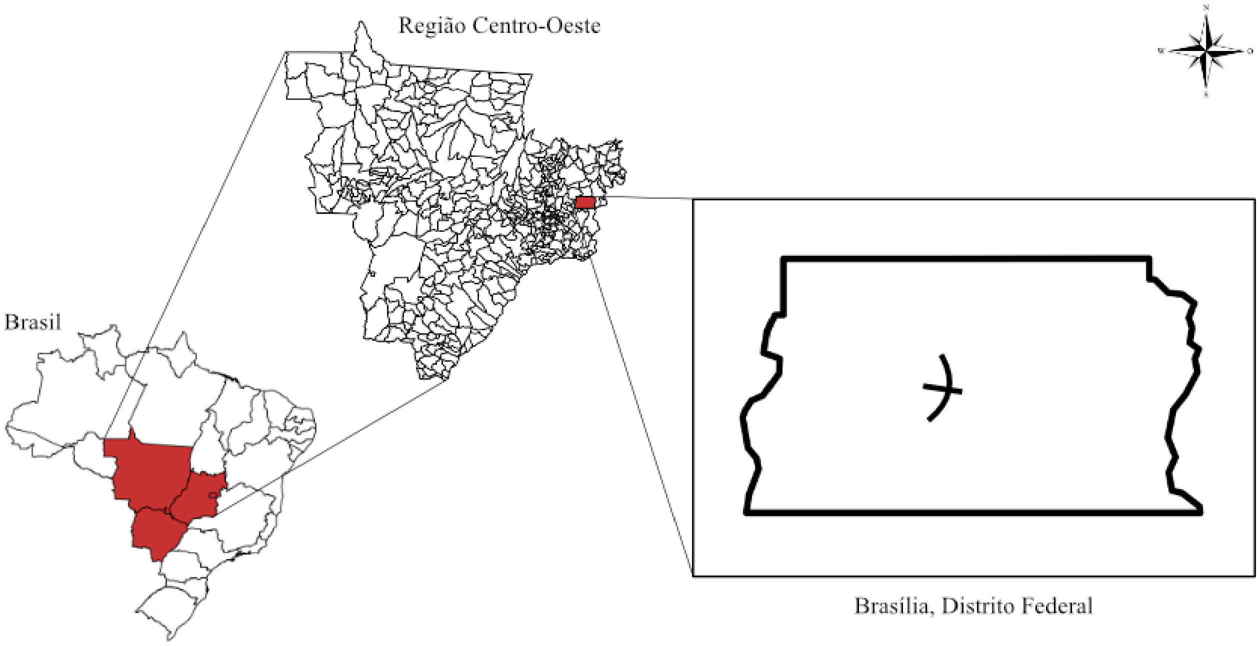


Figura 2 - Localização da cidade em estudo

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, levando em consideração 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 ~ N(μ, σ2) é comumente usada para denotar que a variável aleatória X segue uma distribuição normal com média μ e variância σ2. Para valores de −∞ < μ < ∞ e σ > 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 como a soma de várias variáveis aleatórias independentes. Na figura abaixo, é ilustrada a função de densidade de probabilidade para μ = 3 e dois diferentes valores de σ.

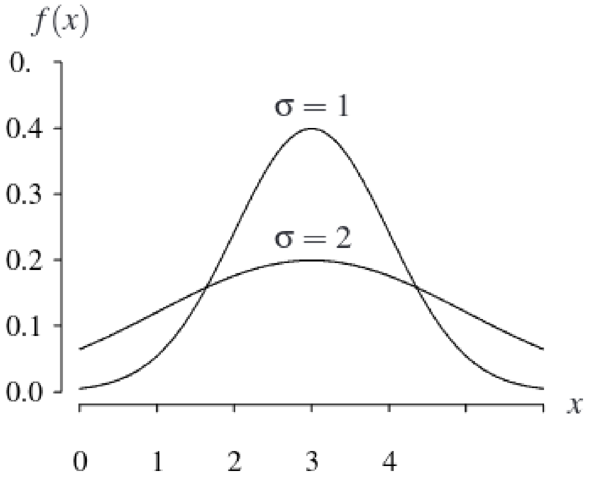


Figura 4 - Normal para μ = 3 e para σ=1 e σ=2.

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 em que 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 ~ 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:

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

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 nn, 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 que há 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.

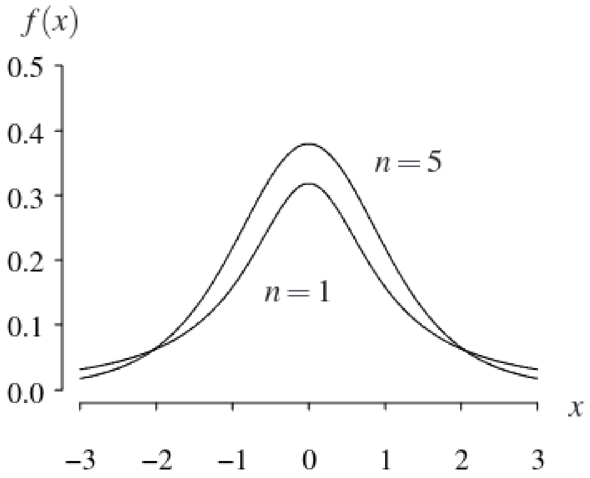


Figura 6 - Distribuição de Cauchy e T quando n=5

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

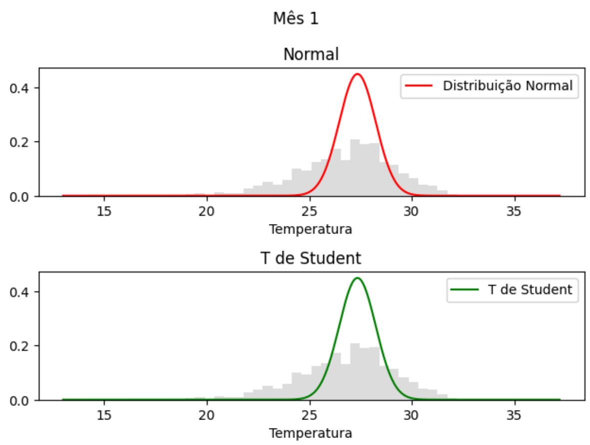
$$E[X] = 0, \; para\, n > 1$$

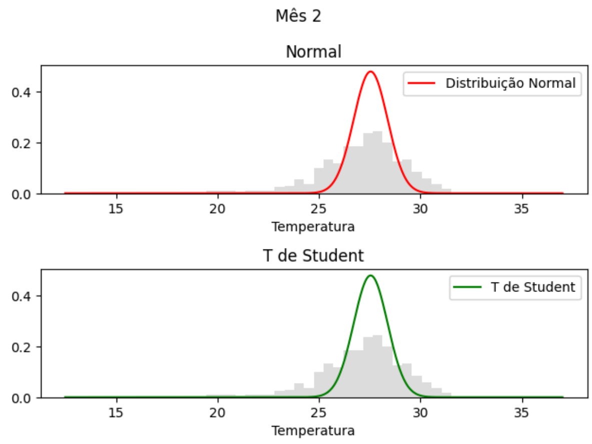
$$V[X] = n/(n-2), \; para\, n > 2$$

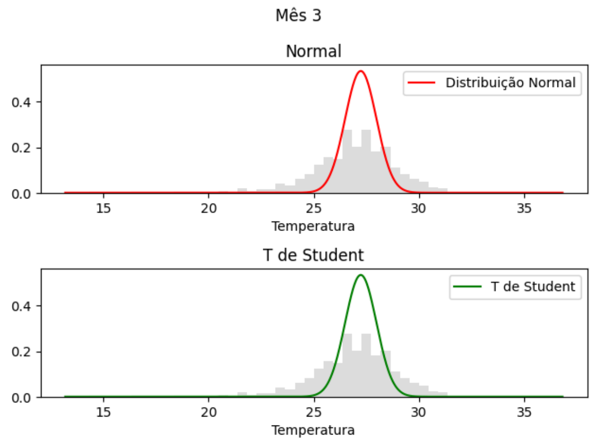
$$E[(X - \mu)/\sigma)^3] = 0, \; para\, n > 3$$

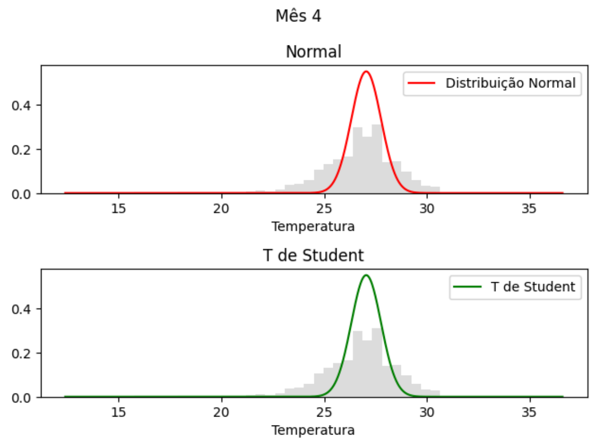
$$E[(X - \mu)/\sigma)^4] = [3(n-2)]/(n-4), \; para\, n > 4$$

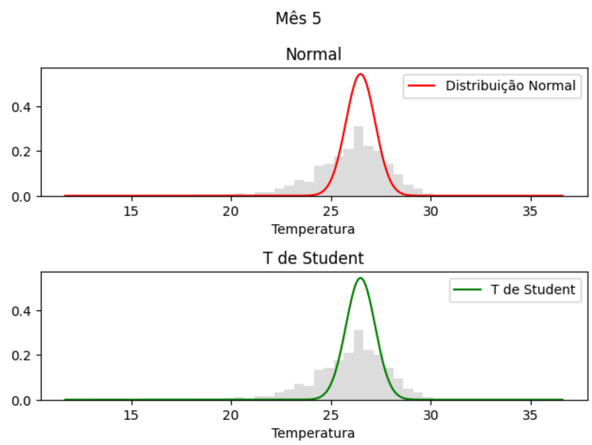
Resultados e discussão

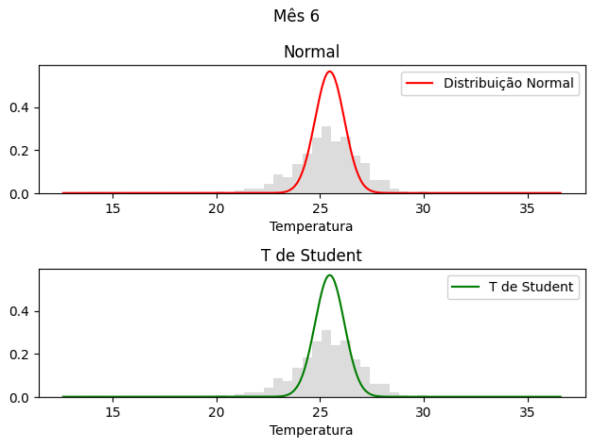


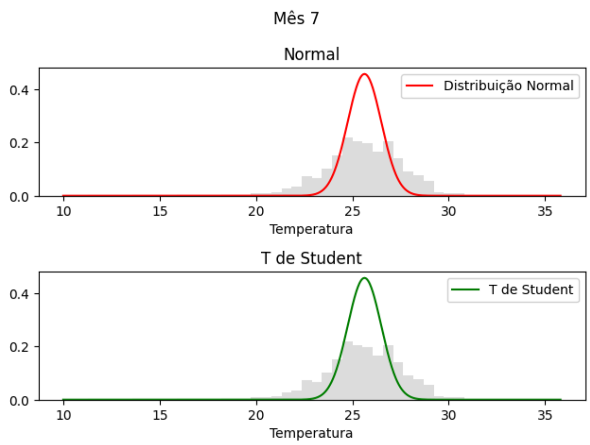


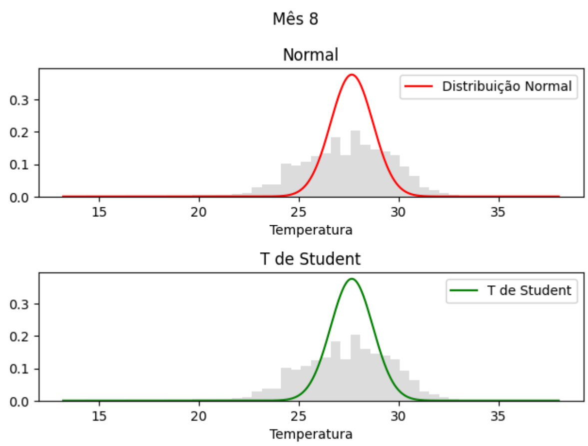


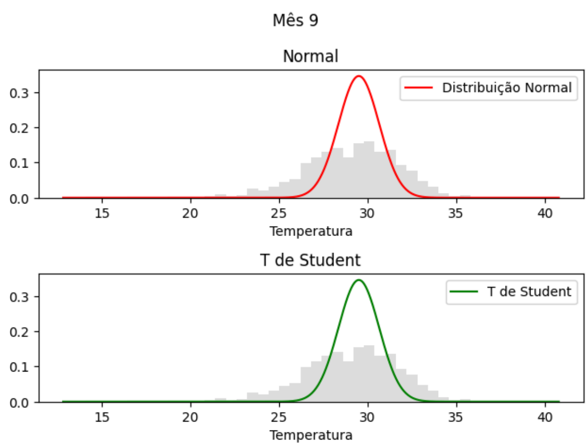


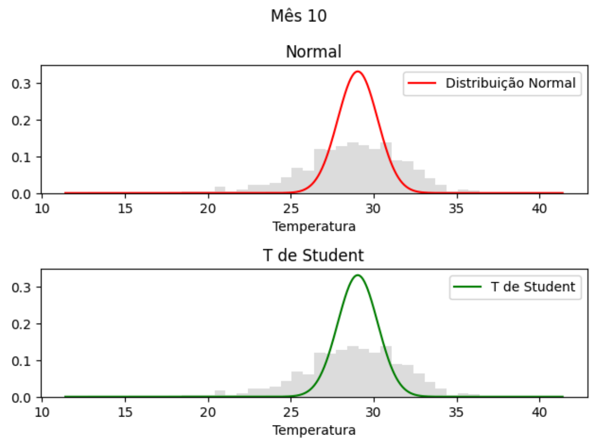


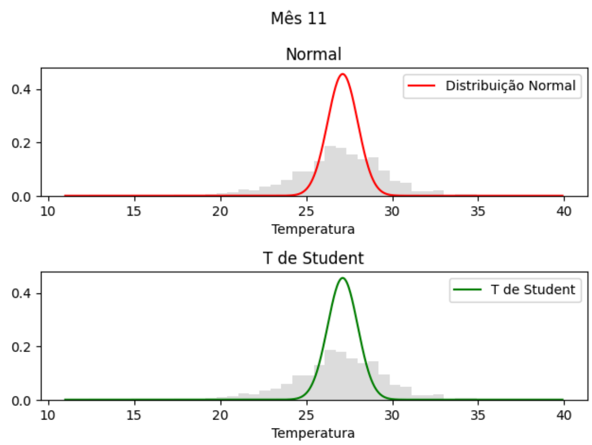


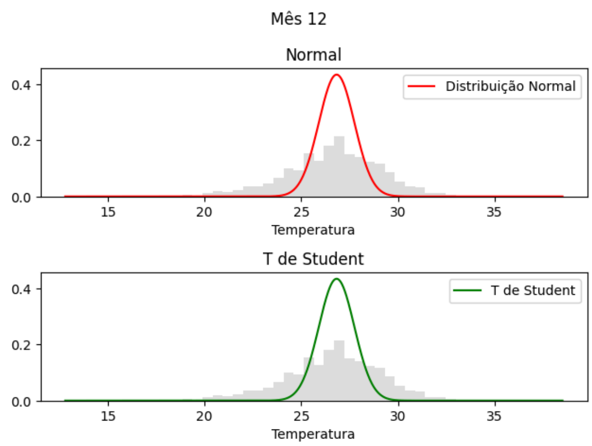












Referências

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