Evento:XXX Jornada de Pesquisa

TESTE DE NORMALIDADE PARA DADOS DE TEMPERATURA E UMIDADE DO AR[1]

Marcos Vicente Hart[2], Mateus D. Kuhn[3], Sandro Sawicki[4], Gerson Battisti[5]

[1] Artigo científico desenvolvido em disciplina do Curso de Mestrado do Programa de Pós-Graduação em Modelagem Matemática e Computacional, desenvolvido na UNIJUÍ.
[2] Bolsista CAPES; Mestrando em Modelagem Matemática e Computacional UNIJUÍ.
[3] Bolsista CNPq, Doutorado em Modelagem Matemática e Computacional (PPGMMC) da UNIJUÍ.
[4] Professor Doutor dos Cursos de Ciência da Computação, Engenharia de Software e do PPGMMC da UNIJUÍ; membro do Grupo de Pesquisa em Computação Aplicada (GCA) da UNIJUÍ.
[5] Professor Doutor dos Cursos de Ciência da Computação, Engenharia de Software e do PPGMMC da UNIJUÍ; membro do Grupo de Pesquisa em Computação Aplicada (GCA) da UNIJUÍ.

INTRODUÇÃO

Nas últimas décadas, os efeitos das mudanças climáticas têm apresentado eventos mais intensos e perceptíveis para a população em geral. Mudanças rápidas e frequentes de eventos extremos, como secas prolongadas ou grandes volumes de chuva em curto período, apresentam preocupação entre a população afetada. Diante disso, a compreensão dos fenômenos climáticos e a busca por modelos estatísticos que auxiliem na tomada de descrição de setores como a agricultura ou políticas públicas vêm se tornando importante para fortalecer a capacidade de adaptação da comunidade diretamente afetada.

No entanto, para o desenvolvimento de modelos estatísticos, a avaliação da normalidade é fundamental para garantir a validade do modelo. Um dos principais passos é o teste de normalidade. Neste trabalho, será apresentado o desempenho dos seguintes testes. O teste Anderson-Darling (ARSHAD et al. 2003), onde se baseia na distribuição empírica, Jarque-Bera (JARQUE e BERA, 1980), que leva em consideração diferenças entre simetria e curtose dos dados. Após isso, serão apresentados alguns métodos de transformação de dados visando melhorar a aderência à normalidade. Segundo De Assis et al., (2023), inúmeras técnicas de transformação de dados são aplicadas, entre as comumente utilizadas estão as transformações Logarítmica, Box-Cox, Raiz Quadrada e Yeo-Johnson.

METODOLOGIA

Para o desenvolvimento desta análise, foram utilizados dados de uma estação meteorológica instalada na cidade de Santa Rosa pelo Grupo de Computação Aplicada, na linha de pesquisa *Cidades Inteligentes e Sustentáveis Com Tecnologias Emergentes e*

Descentralizadas. Os dados utilizados são do ano de 2024 período em que foram registradas 28.875 leituras de temperatura e umidade.

A importância de se testar a normalidade tem sido amplamente reconhecida na literatura estatística, e diversos procedimentos têm sido propostos. A maioria baseia-se nos momentos amostrais, como assimetria (*skewness*) e curtose (*kurtosis*). Os mais conhecidos são o teste de D'Agostino-Pearson e o teste de Shapiro-Wilk (YAP, 2011). No entanto, esses testes são recomendados para amostras pequenas ou médias, sendo que o teste de Shapiro se limita à 5.000 leituras (MIOT, 2017).

No contexto da regressão, pode-se aplicar esses testes aos resíduos ou utilizar procedimentos mais gerais, baseados em funções de verossimilhança ou de pontuação (*score*). Dessa forma serão aplicados os testes de Anderson-Darling e Jarque-Bera.

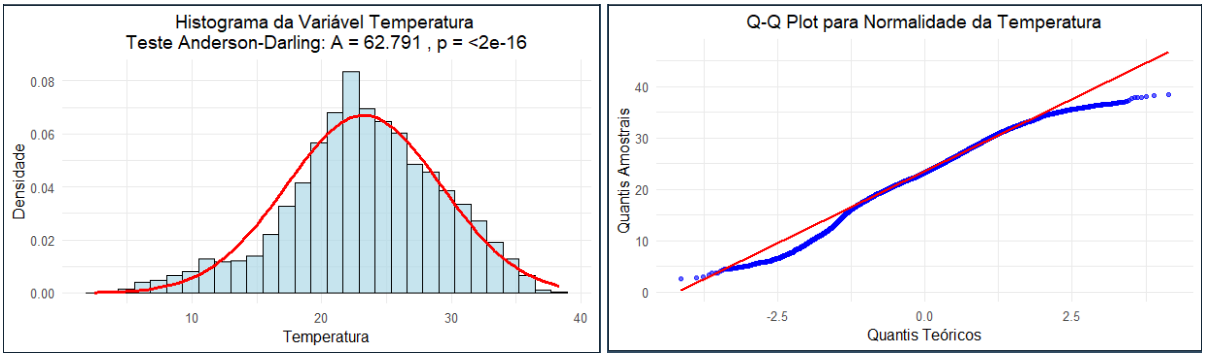
Para a apresentação e identificação de uma distribuição com normalidade, existem opções gráficas comumente utilizadas, dentre as principais o histograma, gráfico de dispersão normal Q-Q *plot* e *box plot* (FIELD, 2009). Contudo, apenas apresentar os dados não permite identificar quantos deles estão dentro da normalidade. Sendo assim, são necessários métodos objetivos que determinam se existe uma normalidade nos dados (BARROS et al., 2012).

RESULTADOS E DISCUSSÃO

Inicialmente, foram realizados, no *software R* os testes de normalidade, esse ambiente é voltado para o desenvolvimento de modelos estatísticos, abrangendo desde a geração de modelos lineares e não lineares, séries temporais e testes estatísticos clássicos (BATTISTI, 2019).

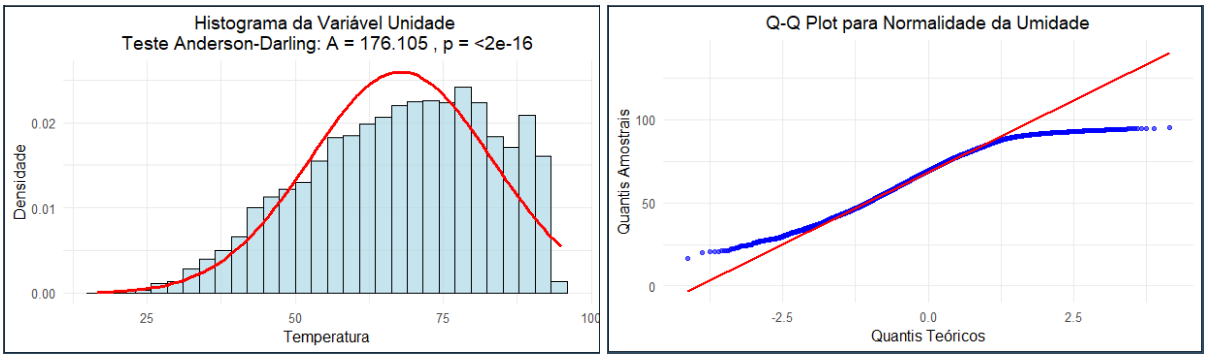
O primeiro teste realizado foi o de Anderson-Darling, que se baseia na distribuição empírica, com parâmetros como os percentis ordenados (FERREIRA, 2022).

Imagem 01 – Teste de normalidade temperatura Anderson-Darling.



Fonte: Autor 2025

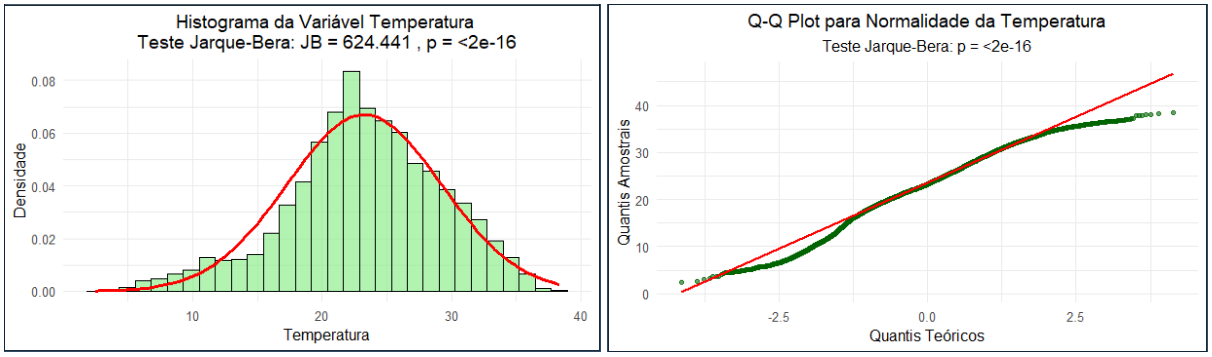
Imagem 02 – Teste de normalidade umidade Anderson-Darling.



Fonte: Autor 2025

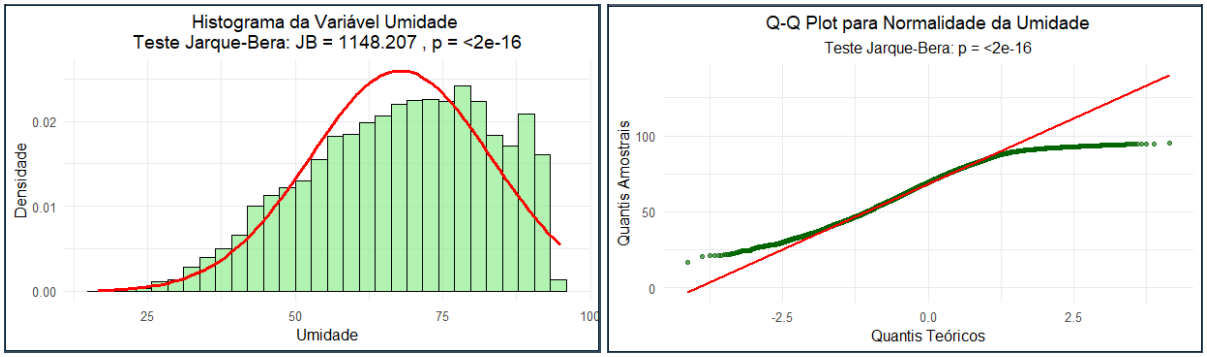
Em seguida foram realizados os mesmos testes utilizando o método Jaque-Bera.

Imagem 03 – Teste de normalidade temperatura Jarque-Bera.



Fonte: Autor 2025

Imagem 04 – Teste de normalidade umidade Jarque-Bera.

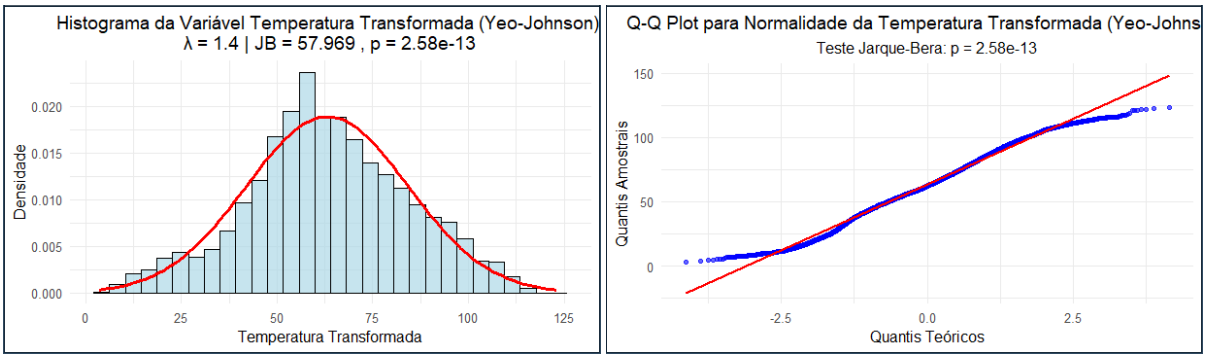


Fonte: Autor 2025

Realizando uma análise visual dos gráficos gerados, observa-se a mesma distribuição dos dados, o que indica o funcionamento e valida os códigos escritos.

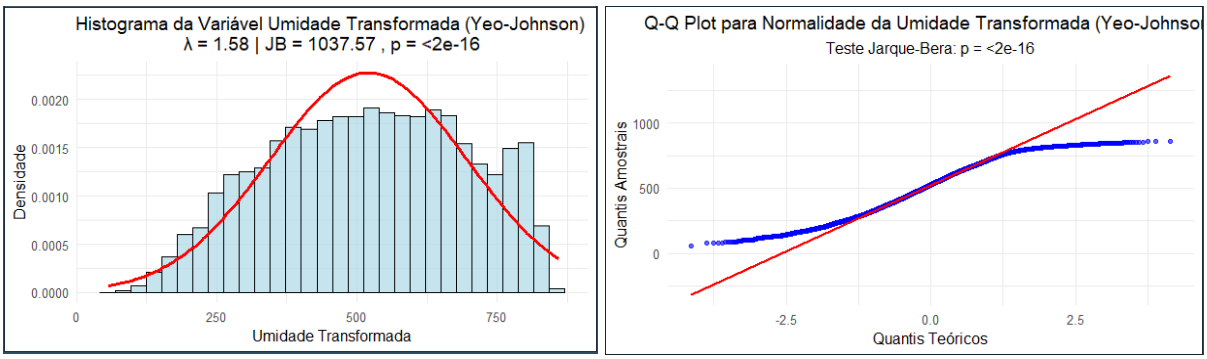
O segundo momento consiste na transformação desses dados por meio dos métodos Logarítmico, Box-Cox, Yeo-Johnson e Raiz Quadrada. Dentre esses métodos foi escolhido Yeo-Johnson por ser o que melhor ajustou os dados.

Imagem 05 – Yeo-Johnson temperatura transformada.



Fonte: Autor 2025

Imagem 06 – Yeo-Johnson umidade transformada.



Fonte: Autor 2025

Mesmo realizando a transformação de dados, os resultados não apresentaram uma distribuição normal. Dessa forma, apenas a transformação não é suficiente, sendo necessário a abordagem de diferentes alternativas ou até mesmo à associação de diferentes métodos.

Os testes foram realizados no *software R*, avaliando a distribuição dos dados sem nenhum tipo de tratamento. Os dados foram testados pelo método de Anderson-Darling e Jarque-Bera, apresentando a mesma distribuição. Dessa forma, é possível validar os testes, indicando que eles foram realizados corretamente.

Os resultados obtidos por meio de testes indicam que, mesmo após transformação dos dados, não foi possível alcançar a normalidade. Estes valores quantificam o grau de afastamento em relação à distribuição normal perfeita. No caso da temperatura, a transformação reduziu os valores estatísticos de 624,44 para 57,97 apresentado uma melhoria na distribuição, reduzindo substancialmente tanto a assimetria quanto a curtose excessiva presente nos dados originais. No entanto, esse valor ainda se mantem acima do limite definido pelo método.

Da mesma forma, a transformação dos dados de umidade não foi suficiente para normalizar a distribuição, apresentando pouca eficácia tanto na redução da assimetria quanto no excesso de curtose em relação à distribuição normal padrão. Os valores estatísticos foram reduzidos de 1148,20 para 1037,56, indicando um desempenho inferior ao observado para a variável temperatura.

CONSIDERAÇÕES FINAIS

Os resultados ilustram que as transformações convencionais não são suficientes para corrigir a não normalidade do conjunto de dados, que apresenta desvios distribucionais. Dessa forma, torna-se necessária uma abordagem mais profunda, com a aplicação de técnicas adicionais aos tratamentos. Como alternativa, pode-se considerar a análise da presença de outliers ou, ainda, métodos de normalização que sejam adequados às características específicas dos dados analisados.

Palavras-chave: Temperatura. Umidade. Anderson-Darling. Jarque-Bera. Yeo-Johnson.

REFERÊNCIAS BIBLIOGRÁFICAS

ARSHAD, M.; RASOOL, M. T.; AHMAD, M. I. Anderson darling and modified anderson darling tests for. Pakistan Journal of Applied Sciences, v. 3, n. 2, p. 85-88, 2003.

BARROS, Mauro Virgílio Gomes de et al. Análise de dados em saúde. 2012.

BATTISTI, Iara Denise Endruweit; DA SILVA SMOLSKI, M. Software R: Análise estatística de dados utilizando um programa livre. Bagé: Faith, 2019.

DE ASSIS, JANILSON PINHEIRO et al. Transformação de dados aplicada à estatística. 2023.

FERREIRA, Amanda Ramalho. Testes de normalidade: Um estudo comparativo. Trabalho de Conclusão de Curso (Bacharel em Estatística). Universidade Federal Fluminense, Niterói, Rio de Janeiro. 37p, 2022.

FIELD, Andy. Descobrindo a estatística usando o SPSS-5. Penso Editora, 2020.

JARQUE, Carlos M.; BERA, Anil K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics letters, v. 6, n. 3, p. 255-259, 1980.

LEOTTI, Vanessa Bielefeldt; BIRCK, Alan Rodrigues; RIBOLDI, João. Comparação dos Testes de Aderência à Normalidade Kolmogorov-smirnov, Anderson-Darling, Cramer–Von Mises e Shapiro-Wilk por Simulação. Anais do 11º Simpósio de Estatística Aplicada à Experimentação Agronômica, 2005.

MIOT, Hélio Amante. Avaliação da normalidade dos dados em estudos clínicos e experimentais. Jornal vascular brasileiro, v. 16, n. 2, p. 88-91, 2017.

YAP, Bee Wah; SIM, Chiaw Hock. Comparisons of various types of normality tests. Journal of Statistical Computation and Simulation, v. 81, n. 12, p. 2141-2155, 2011.