

**UNIVERSIDADE FEDERAL DE ALFENAS - UNIFAL-MG**

**JANAÍNA DE SOUZA FERREIRA**

**APLICAÇÃO DE TESTES NÃO PARAMÉTRICOS: UMA ANÁLISE DOS PREÇOS  
DE COMBUSTÍVEIS PRATICADOS NA CIDADE DE VARGINHA-MG**

**VARGINHA-MG**

**2025**

**JANAÍNA DE SOUZA FERREIRA**

**APLICAÇÃO DE TESTES NÃO PARAMÉTRICOS: UMA ANÁLISE DOS PREÇOS  
DE COMBUSTÍVEIS PRATICADOS NA CIDADE DE VARGINHA-MG**

Trabalho de Conclusão de PIEPEX apresentado como parte dos requisitos para obtenção do título de Bacharel Interdisciplinar em Ciência e Economia pela Universidade Federal de Alfenas.

Orientador: Prof. Dr. Deive Ciro de Oliveira

**VARGINHA-MG**

**2025**

Sistema de Bibliotecas da Universidade Federal de Alfenas  
Biblioteca Campus Varginha

Ferreira, Janaína de Souza.

Aplicação de testes não paramétricos : uma análise dos preços de combustíveis praticados na cidade de Varginha-MG / Janaína de Souza Ferreira. - Varginha, MG, 2025.

37 f. : il. -

Orientador(a): Deive Ciro de Oliveira.

Trabalho de Conclusão de Curso (Graduação em Bacharelado Interdisciplinar em Ciência e Economia) - Universidade Federal de Alfenas, Varginha, MG, 2025.

Bibliografia.

1. Estatística. 2. Preços. 3. Combustíveis. I. Oliveira, Deive Ciro de, orient. II. Título.

JANAÍNA DE SOUZA FERREIRA

**APLICAÇÃO DE TESTES NÃO PARAMÉTRICOS: UMA ANÁLISE DOS PREÇOS  
DE COMBUSTÍVEIS PRATICADOS NA CIDADE DE VARGINHA-MG**

O Presidente da banca examinadora abaixo assina a aprovação do Trabalho de Conclusão de Curso apresentado como parte dos requisitos para a obtenção do título de Bacharel em Ciência e Economia pela Universidade Federal de Alfenas.

Aprovado em: 19 de novembro de 2025.

Prof. Dr. Deive Ciro de Oliveira  
Presidente da Banca Examinadora  
Instituição: Universidade Federal de Alfenas

Prof. Dra. Gislene Araújo Pereira  
Instituição: Universidade Federal de Alfenas

Profa. Dra. Crysttian Arantes Paixão  
Instituição: Universidade Federal da Bahia



Documento assinado eletronicamente por **Deive Ciro de Oliveira, Professor do Magistério Superior**, em 17/12/2025, às 09:23, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [https://sei.unifal-mg.edu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.unifal-mg.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **1694911** e o código CRC **4EDAD3BE**.

---

## RESUMO

Este estudo analisou a aplicação de testes não paramétricos na comparação dos preços de etanol, gasolina e diesel praticados em Varginha-MG, com dados coletados entre os dias 20 e 24 dos meses de agosto e setembro de 2025 em 39 postos. Inicialmente, verificou-se a normalidade dos conjuntos por meio do Teste de Shapiro-Wilk, o qual indicou a rejeição do pressuposto de normalidade para todos os combustíveis. Diante desse resultado, aplicou-se o Teste U de Mann-Whitney para comparar as medianas entre os meses analisados. Os p-valores encontrados foram superiores ao nível de significância, indicando ausência de diferenças estatisticamente significativas entre os períodos. O estudo mostra a relevância do Teste U de Mann-Whitney como alternativa para análises inferenciais quando a distribuição dos dados não atende à normalidade.

Palavras-chave: estatística; preços; combustíveis.

## ABSTRACT

This study analyzed the application of nonparametric tests in comparing the prices of ethanol, gasoline, and diesel in Varginha, Brazil, using data collected from 39 gas stations between august and september 2025, specifically from the 20th to the 24th of each month. Initially, the normality of the data sets was assessed using the Shapiro–Wilk test, which indicated the rejection of the normality assumption for all fuels. Given this result, the Mann–Whitney U test was applied to compare the medians between the months analyzed. The p-values obtained were higher than the significance level, indicating no statistically significant differences between the periods. The study highlights the relevance of the Mann–Whitney U test as an appropriate alternative for inferential analysis when the data distribution does not meet the normality assumption.

Keywords: statistics; prices; fuels.

## LISTA DE FIGURAS

Figura 1 - Distribuição normal com média $\mu$ e desvio-padrão $\sigma$ .....	12
Figura 2 - Mudança de escala para unidades padronizadas.....	13
Figura 3 - Tabela Normal Padrão z dos valores 0,7 e 1,5.....	14
Figura 4 - Teste Shapiro-Wilk em duas distribuições distintas.....	22
Figura 5 - Histogramas de frequência dos preços entre 20 e 24 dos meses de agosto (esquerda) e setembro (direita).....	30

## LISTA DE GRÁFICOS

Gráfico 1 - Comparação entre uma curva normal e exponencial.....	15
Gráfico 2 - Teste Z para média .....	18
Gráfico 3 - Teste t de Student.....	20

## LISTA DE TABELAS

Tabela 1 - Resumo de Testes Shapiro-Wilk Simulados no R.....	21
Tabela 2 - Teste U de Mann-Whitney amostra pequena no R.....	26
Tabela 3 - Teste U de Mann-Whitney amostra grande no R.....	28
Tabela 4 - Resumo de Estatística Descritiva Agosto.....	31
Tabela 5 - Resumo de Estatística Descritiva Setembro.....	31
Tabela 6 - Resumo de Testes de Aderência à normalidade AGOSTO (Shapiro-Wilk).....	32
Tabela 7 - Resumo de Testes de Aderência à normalidade SETEMBRO (Shapiro-Wilk).....	32
Tabela 8 - Resumo dos Testes U de Mann-Whitney.....	32
Tabela 9 - Estimativa pontual das medianas dos preços.....	33

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>10</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO.....</b>	<b>11</b>
2.1	DISTRIBUIÇÃO NORMAL.....	11
<b>2.1.1</b>	<b>Não Normalidade.....</b>	<b>14</b>
<b>2.1.2</b>	<b>Testes de Hipóteses.....</b>	<b>15</b>
<b>2.1.3</b>	<b>Teste t de Student.....</b>	<b>18</b>
2.2	TESTES DE ADERÊNCIA.....	20
2.3	TESTES NÃO PARAMÉTRICOS.....	22
<b>2.3.1</b>	<b>Teste U de Mann-Whitney.....</b>	<b>23</b>
<b>3</b>	<b>O PROGRAMA DE EXTENSÃO: ESTATÍSTICA PARA TODOS.....</b>	<b>28</b>
4	APLICAÇÃO DO TESTE U: PREÇOS DE COMBUSTÍVEIS.....	28
4.1	RESULTADOS E DISCUSSÃO.....	33
<b>5</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>34</b>
	<b>REFERÊNCIAS.....</b>	<b>36</b>

## 1 INTRODUÇÃO

A Estatística tem sido um instrumento fundamental desde o surgimento das primeiras cidades. Através dela é possível enxergar que os dados vão além de “apenas” números. Seu papel é essencial na organização e no conjunto de métodos de análise dos dados. Uma vez que, seu uso orienta uma postura crítica acerca da interpretação e da formulação de conclusões sobre os dados (Callegari-Jacques, 2007).

Segundo Bussab e Morettin, “o uso de informações de uma amostra para concluir sobre o todo faz parte da atividade diária da maioria das pessoas” (2010, p. 261). Ao formular uma hipótese acerca de determinada população, pode-se verificá-la por meio da análise dos dados amostrais, se há evidências suficientes para rejeitá-la ou não.

Esses fenômenos ocorrem intuitivamente no cotidiano, por exemplo: ao escolher uma roupa nova que ajuste ao corpo; ao perceber que alguém conhecido está mentindo; ao perguntar se o dia está para chuva ou não; ao verificar se a quantidade de sal em uma refeição está adequada ou quando um feirante escolhe as frutas mais ‘bonitas’ para venda.

Muitas dos métodos aplicados em Estatística fundamentam-se no pressuposto de que as variáveis aleatórias analisadas seguem uma distribuição normal. Esse pressuposto é de suma importância, uma vez que assegura a validade dos procedimentos inferenciais para tal distribuição. No entanto, se este pressuposto for violado, torna-se imperativo a utilização de métodos mais “robustos”, chamados testes não paramétricos (Bussab; Morettin, 2010).

O presente Trabalho de Conclusão de PIEPEX (Programa Integrado de Ensino, Pesquisa e Extensão) tem como premissa apresentar metodologias estatísticas fundamentadas em uma revisão da literatura sobre testes não paramétricos. Neste contexto, considera-se a hipótese de que os dados observados sejam provenientes de uma população distinta da distribuição normal com aplicação do Teste U aos preços de combustíveis praticados em Varginha-MG entre os dias 20 e 24, de agosto e setembro de 2025. A escolha do tema se deu em razão da necessidade de compreender os métodos estatísticos empregados no Projeto: Inflação de Produtos e Serviços Relevantes na Comunidade desenvolvido no Programa de Extensão: Estatística para Todos, dessa forma, alinhando os objetivos deste trabalho aos eixos norteadores do PIEPEX. Assim como, na relevância dos métodos estatísticos para a pesquisa acadêmica, tendo em vista que a utilização de dados amostrais é prática comum em estudos científicos.

O objetivo geral deste estudo consiste em aplicar os testes não paramétricos, com ênfase no Teste U, aos dados de preços coletados entre os dias 20 e 24, de agosto e setembro

de 2025, no âmbito do Projeto. Essa aplicação possibilitou o aprofundamento prático dos métodos estatísticos discutidos em sala de aula promovendo a integração entre teoria e prática.

A estrutura do TCP (Trabalho de Conclusão de PIEPEX) organiza-se da seguinte forma: na próxima seção, será discutida a questão da normalidade e não normalidade de distribuições, assim como, será apresentado o teste paramétrico t-Student, aplicável sob a hipótese de normalidade. Em seguida, ainda nesta seção, serão abordados os métodos utilizados para verificar a validade de normalidade, testes de aderência. A segunda seção, dedica-se à análise do Teste U. Na terceira seção será realizada a análise da estatística descritiva e aplicação dos testes estatísticos nos dados relativos aos preços dos combustíveis entre os dias 20 e 24, de agosto e setembro de 2025, por meio do software estatístico R versão 4.5.1, bem como a apresentação do Projeto: Inflação de Produtos e Serviços Relevantes na Comunidade, cujo responsável pela coleta dos dados. Por fim, serão apresentados os resultados obtidos por meio da aplicação dos métodos estatísticos estudados neste trabalho.

## 2 REFERENCIAL TEÓRICO

### 2.1 DISTRIBUIÇÃO NORMAL

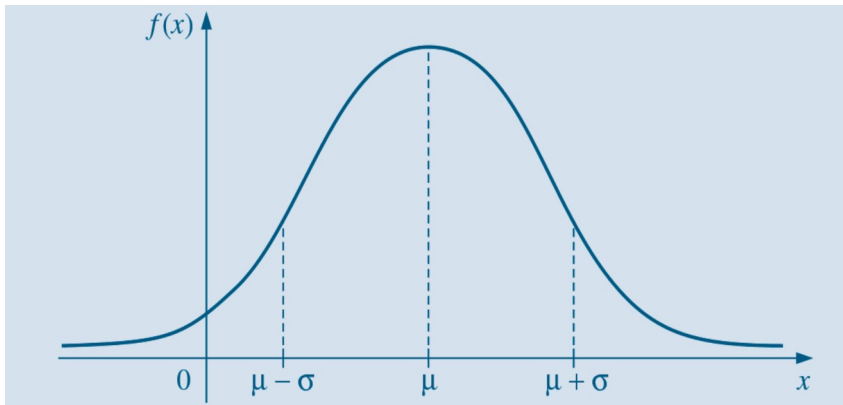
As variáveis aleatórias contínuas podem assumir diferentes distribuições de frequência, na maioria das vezes o histograma apresenta uma distribuição em forma de sino, denominada distribuição normal ou gaussiana. Essa distribuição possui particularidades que a torna uma das mais importantes no campo da Estatística, destacando-se a sua curva simétrica em torno da média ( $\mu$ ), localizada no centro da distribuição, ponto em que também coincidem a moda e a mediana (Vieira, 2016).

Para Larson (2015), a distribuição normal deve atender as seguintes propriedades: além da equivalência entre a média, mediana e moda; a área total sob a curva corresponde a 1, à medida que se afasta da média, a curva normal aproxima-se do eixo x, mas sem jamais tocá-lo. Entre  $\mu - \sigma$  e  $\mu + \sigma$ , o gráfico apresenta concavidade para baixo. Já à esquerda de  $\mu - \sigma$  e a direita de  $\mu + \sigma$ , a concavidade é voltada para cima. Nos pontos onde ocorrem as mudanças de concavidade são chamados pontos de inflexão,  $\mu - \sigma$  e  $\mu + \sigma$ , conforme representado na Figura 1.

Exemplo 1: Os preços de um determinado livro pedagógico, cotado para aquisição e distribuição nas escolas municipais é uma variável com distribuição normal de  $\mu = 75$  e desvio padrão ( $\sigma$ ) igual a 7. Isto é: 50% dos livros apresentam preço igual ou maior que R\$75,00;

50% apresentam preço igual ou menor que R\$75,00; não obteve valores muito abaixo desse valor e nem muito acima dessa média.

Figura 1 - Distribuição normal com média  $\mu$  e desvio-padrão  $\sigma$



Fonte: Morettin (2023, p. 196)

De acordo com as definições apresentadas por Bussab e Morettin, uma variável aleatória contínua  $X$  segue uma distribuição normal com parâmetros  $\mu$  e  $\sigma^2$ , cuja função densidade de probabilidade é expressa por:

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

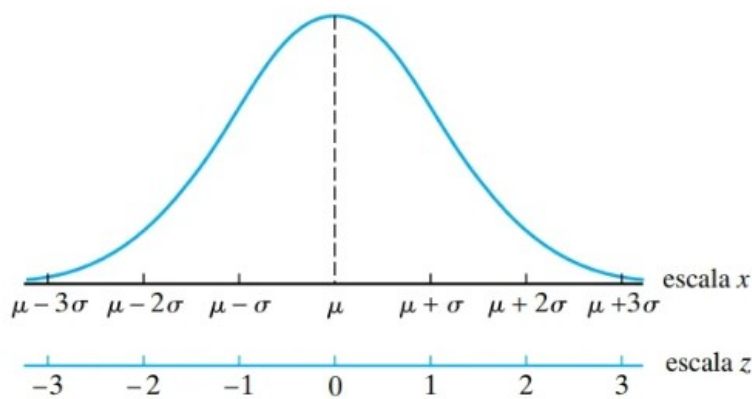
Para Anderson *et al.* (2020), uma variável aleatória que tem distribuição normal cuja média é igual a 0 e o desvio padrão é 1, a distribuição denomina-se normal padrão ou padronizada. O aspecto é como de uma distribuição normal, porém com os parâmetros definidos:  $\mu=0$  e  $\sigma=1$ . Segue a função densidade de probabilidade normal padronizada:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (2)$$

Para calcular as probabilidades, calcula-se as áreas sob o gráfico da função densidade de probabilidade. Assim, para encontrar a probabilidade de uma variável aleatória estar dentro de um intervalo deve-se calcular a área sob a curva normal desse intervalo. No entanto, para a distribuição padronizada as áreas foram calculadas e estão disponíveis em Tabelas<sup>1</sup> (Anderson *et al.*, 2020).

<sup>1</sup>ANDERSON, David R. *et al.* **Estatística aplicada a administração e economia**. 5. ed. Porto Alegre, 2020. Pag.669 e 670.

Figura 2 – Mudança de escala para unidades padronizadas



Fonte: Freund (2019, p. 219)

Dessa forma, ocorre uma mudança de escala original ( $X$ ) para unidades padronizadas de escala  $Z$  (Figura 2), ou seja, para calcular as probabilidades das distribuições, em primeiro lugar, transforma-se a variável original em uma distribuição normal padronizada por meio da equação (3). Então assim, para quaisquer  $\mu$  e  $\sigma$ , utiliza-se a distribuição normal padronizada para obter as probabilidades de uma distribuição normal (Freund, 2019).

$$Z = \frac{X - \mu}{\sigma} \quad (3)$$

Exemplo: Suponha uma distribuição normal com  $\mu=21$  e  $\sigma=4$ . Qual a probabilidade da variável aleatória  $x$  estar entre 24 e 27? Utilizamos a equação (3) para transformar a variável para padronizada:

$$Z_1 = \frac{24 - 21}{4} = 0,75$$

$$Z_2 = \frac{27 - 21}{4} = 1,5$$

Ou seja, após a transformação, a probabilidade mudou para 0,75 e 1,5, isto é,  $P(0,75 \leq x \leq 1,5) = P(z \leq 1,5) - P(z \leq 0,75)$ . Na Figura 3, demonstra os valores de  $z$  encontrados na Tabela (Anderson *et al.*, 2020, p. 670) para obter as probabilidades.

Calcula-se,  $P(z \leq 1,5) - P(z \leq 0,75) = 0,9332 - 0,7734 = 0,1598$ , ou seja,  $(24 \leq X \leq 27) = 0,1598$ . Portanto, a probabilidade de  $x$  estar entre 24 e 27 é 0,1598.

Figura 3 - Tabela Normal Padrão z dos valores 0,7 e 1,5

<i>z</i>	0,00	0,01	0,02	0,03	0,04	0,05
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394

Fonte: Anderson (2020, pag. 670)

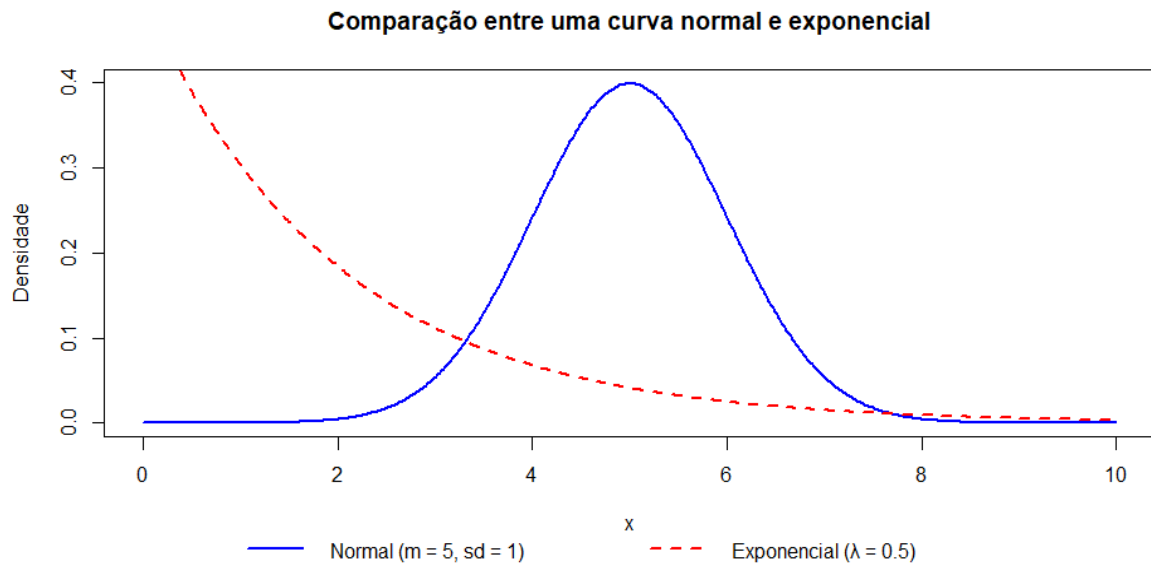
### 2.1.1 Não Normalidade

Como mencionado, é comum que as variáveis aleatórias sigam uma distribuição normal. A não normalidade, por sua vez, refere-se às situações em que determinada variável aleatória apresenta uma distribuição distinta da distribuição normal.

De acordo com Pino (2014), a não normalidade pode ser identificada de forma evidente em alguns casos. Um primeiro aspecto ocorre quando as distribuições apresentam assimetrias. Outra situação é observada quando as distribuições apresentam caudas pesadas, o que pode estar associado à ocorrência de observações extremas e discrepantes (*outliers*) em relação ao conjunto dos demais dados ou quando a variância é muito grande ou até infinita.

A distribuição exponencial-logarítmica é aplicável ao estudo do tempo de vida de objetos, organismos, dispositivos, entre outros (Garcia, 2013). Essa distribuição constitui um exemplo de distribuição não normal. No Gráfico 1, demonstra as diferenças do formato entre a curva de uma distribuição normal e a de uma distribuição exponencial, evidenciando como esta última viola os pressupostos de normalidade.

Gráfico 1- Comparação entre uma curva normal e exponencial



Fonte: Elaboração própria utilizando o software R

### 2.1.2 Testes de Hipóteses

Após o processo de amostragem, análise inicial das características amostrais, bem como a representação gráfica dos dados, por meio da estatística inferencial é possível estimar parâmetros ou testar hipóteses sobre o valor dos parâmetros sob um nível de significância conhecido.

De acordo com Morettin (2010), o teste de hipóteses é um método que permite realizar suposições acerca de uma ou mais populações a partir de uma ou mais amostras e, assim, testar as hipóteses e decidir entre duas alternativas. O teste é formado por duas hipóteses: a hipótese nula ou da existência ( $H_0$ ) e a hipótese alternativa ( $H_1$ ).

Assim, conforme Larson (2015), para construir um teste de hipóteses é necessário especificar a afirmação sobre o parâmetro a ser testado, tal que, a hipótese nula utiliza uma relação de igualdade ( $\leq$ ,  $=$  ou  $\geq$ ). Por outro lado, a hipótese alternativa será o complemento da hipótese nula e utiliza uma relação de desigualdade, como  $<$ ,  $\neq$  ou  $>$ . Por fim, a hipótese alternativa será aceita se a afirmação da hipótese nula for dada com falsa.

Quando realizado um teste de hipótese deve-se admitir a possibilidade de erros, visto que as informações se baseiam em dados amostrais, sendo assim há possibilidade de ocorrer dois tipos de erros. Primeiro, o erro de rejeitar  $H_0$  quando ela é verdadeira, conhecido como

erro do Tipo I. Segundo erro, pode ocorrer a situação inversa, em que a hipótese nula não é rejeitada, quando ela é falsa; trata-se do erro Tipo II. A probabilidade de cometer um erro do Tipo I é denominado nível de significância do teste e é representado por  $\alpha$  (Firmino, 2015).

Após a formulação das hipóteses é preciso calcular as estatísticas amostrais de interesse, chamada de estatística de teste. Em seguida, supondo que a hipótese nula seja verdadeira, transforma-se o valor da estatística de teste para uma estatística de teste padronizada, assim como:  $z$ ,  $t$  ou  $X^2$ . Com ela será tomada a decisão de rejeitar ou não a hipótese nula (Larson, 2015).

Os passos para realizar um teste de hipóteses conforme Firmino (2015) são:

1. Formular as hipóteses nula ( $H_0$ ) e alternativa ( $H_1$ );
2. Definir o nível de significância ( $\alpha$ ), isto é, a probabilidade de cometer o erro do tipo 1;
3. Escolher a estatística de teste a usar;
4. Determinar a região de rejeição (RR) ou região crítica (RC) e a região de não rejeição (RNR);
5. Obter o valor calculado e verificar se ele se encontra ou não na região de rejeição;
6. Se o valor calculado estiver na região de não rejeição, aceitar  $H_0$ ; se estiver na região de rejeição, rejeitar  $H_0$ ;
7. Concluir o problema.

Exemplo 02: Uma prefeitura implementou um projeto de profissionalização em determinada comunidade. A fim de avaliar seus impactos, um pesquisador buscou estudar se a renda média das famílias que participaram do projeto, na qual resultou em R\$1.650,00, difere da renda média mensal R\$1.500,00 com desvio padrão conhecido = R\$500,00, daquela comunidade.

Inicialmente, admite-se como verdadeira  $H_0$ , isto é,  $\mu = \mu_0 = 1500$ . Considerando que a renda média da comunidade é de R\$1.500,00, amostras aleatórias de 36 famílias que participaram do projeto apresentaram médias amostrais ( $\bar{x}$ = R\$1.650,00) distribuídas segundo uma curva normal. Calcula-se o desvio padrão amostral:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{500}{\sqrt{36}} = 83,33 \quad \text{reais} \quad (4)$$

Definido o nível de significância,  $\alpha = 0,05$ , estabelece-se um valor crítico ( $z_{\alpha}$ ) que

determina a região de rejeição ou região crítica de  $H_0$ , as caudas sombreadas de vermelho representada no Gráfico 2. O objetivo é verificar a quantos erros padrão corresponde o desvio entre  $\bar{x}$  e  $\mu_0$ . Caso o desvio não seja estatisticamente significativo, a  $H_0$  é aceita; caso contrário, é rejeitada. Determinando-se  $z_\alpha = 1,96$ . Logo, um desvio de até 1,96 erros padrão é admitido como não significativo.

Dado que a distribuição amostral de  $\bar{x}$  possui uma distribuição normal, a distribuição amostral (4) é uma distribuição normal padrão (Anderson *et al.*, 2020).

$$z_{calc} = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} \quad (5)$$

Ou seja, o valor do teste é:

$$z_{calc} = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{1650 - 1500}{83,33} = 1,80$$

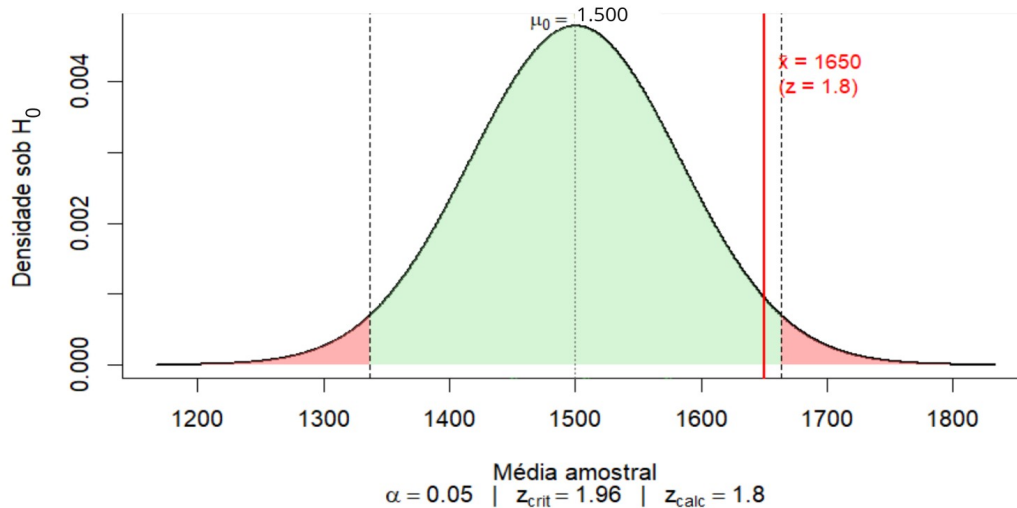
Esse resultado indica que o afastamento de  $\bar{x}$  em relação à  $\mu_0 = 1500$  é de 1,80 erro padrão. Como  $z_{calc}$  está na região de não rejeição (parte sombreada de verde) conforme mostra no Gráfico 2,  $|z_{calc}| = 1,80 < z_\alpha = 1,96$ , não se rejeita  $H_0$ . Dessa forma, conclui-se que não há evidências estatisticamente significativas, ao nível de 5%, para afirmar que a renda média das famílias que participaram do projeto difere da renda média da comunidade.

Outro ponto importante para esclarecer, seria as estatísticas amostrais para realizar os testes estatísticos conforme foi utilizado no exemplo anterior. A relação entre a distribuição amostral das médias e a população da qual provém é demonstrada no teorema do limite central. Se amostras de tamanho grande ( $n \geq 30$ ) são retiradas ao acaso de uma população, as médias amostrais se aproximam de uma distribuição normal. A medida que aumenta o tamanho da amostra, ela se aproxima cada vez mais da distribuição normal. Por fim, a variância das distribuições da média é igual  $1/n$  vezes  $\sigma^2$  e o desvio padrão amostral ( $\sigma_{\bar{x}}$ ) é igual a  $\frac{\sigma}{\sqrt{n}}$ , demonstrado na equação 4 (Larson, 2020).

Para Morettin (2023, p. 370):

Outra maneira de proceder consiste em apresentar a probabilidade de significância ou valor-p do teste. Os passos são muito parecidos aos já apresentados; a principal diferença está em não construir a região crítica. O que se faz é indicar a probabilidade de ocorrer valores da estatística mais extremos do que o observado, sob a hipótese de  $H_0$  ser verdadeira.

Gráfico 2 - Teste Z para média



Fonte: Elaboração própria utilizando o software R

### 2.1.3 Teste t de Student

Em 1908, W. S. Gosset, sob o pseudônimo “Student”, publicou um artigo no qual propôs a utilização da distribuição t de Student em substituição do valor crítico obtido pela distribuição normal nos casos em que a variância populacional ( $\sigma^2$ ) é desconhecida (Callegari-Jacques, 2007).

Se tomarmos várias amostras de mesmo tamanho de uma mesma população com distribuição normal e calcular as médias dessas amostras, essas médias seguem uma distribuição t de *student*. Assim o teste t, avalia se há diferenças significativas entre as médias. Nesse caso, utiliza-se a média amostral ( $\bar{x}$ ) e o estimador da variância populacional ( $\sigma^2$ ), chamado variância amostral ( $s^2$ ) no cálculo de t, posteriormente utiliza-se a Tabela t de Student<sup>2</sup>. Outro aspecto relevante, é o grau de liberdade (GL ou representado por  $\phi$ ) na primeira coluna da tabela que refere-se a quantidade de informações de cada uma das variáveis aleatórias escolhidas livremente (Castanheira, 2023).

De acordo com Morettin (2010), defina-se a variável com distribuição de t de Student com graus de liberdade ( $\phi$ ):

<sup>2</sup>CASTANHEIRA, Nelson Pereira. **Estatística aplicada a todos os níveis**. 3. ed. Curitiba, 2023. E-book. Pag. 256 e 257.

$$t_{\phi} = \frac{\bar{x} - \mu}{s_{\bar{x}}} \quad (6)$$

$$s_{\bar{x}} = \sqrt{\frac{s^2}{n}} = \frac{s}{\sqrt{n}} \quad (7)$$

Nesse contexto, os valores críticos de t, obtidos a partir da tabela da distribuição de t, dependem não apenas de  $\alpha$ , mas também dos graus de liberdade da amostra. Assim, quando  $\sigma$  é desconhecido, pode-se recorrer ao desvio padrão amostral. “Quando n é grande,  $s^2$  se aproxima bastante de  $\sigma^2$ , o que faz com que a variável t se aproxime da variável normal Z” (Morettin, 2010, p. 266).

Por outro lado, quando n é pelo menos 30 isso não se aplica, pois na equação 5,  $\sigma_{\bar{x}}$  é um denominador constante. Diferente de  $s_{\bar{x}}$  na equação 6, que é uma variável aleatória. Sendo assim, o uso da distribuição t é recomendada quando o tamanho da amostra é igual ou inferior a 30. Em amostras grandes, os valores de t aproximam-se dos de z (Morettin, 2010).

Considere-se uma amostra com  $n = 8$ , por exemplo, com  $\alpha = 0,05$ , seu valor crítico correspondente na tabela de distribuição normal padrão é  $z_{0,05} = 1,96$ . Entretanto, ao utilizar a tabela da distribuição t, o valor crítico para  $\alpha=0,05$  e  $\phi = n-1=7$  (GL) passa de 1,96 para  $t_{0,05;7} = 2,37$ . Dessa forma, para que a diferença entre as médias seja considerada estatisticamente significativa, o valor de t deve ser maior ou igual a 2,37.

Exemplo 03: Foi implementado certo projeto voltado a incentivar a permanência de estudantes nas escolas. Para avaliar seu impacto, coletou-se uma amostra de  $n=20$  estudantes, registrando-se uma média amostral ( $\bar{x}$ ) de 16 dias frequentados e  $s=3$  dias. A média de frequência antes da implementação era de 13 dias.

Solução:

$$H_0: \mu \leq 13$$

$$H_1: \mu > 13$$

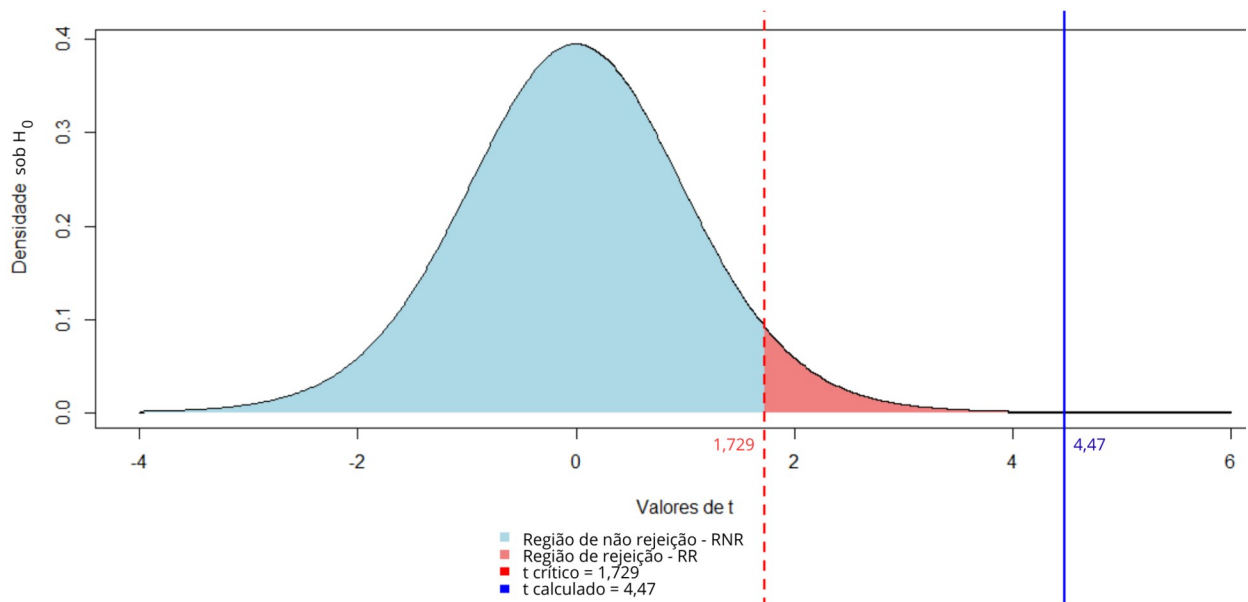
$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{3}{\sqrt{20}} \approx 0,671$$

$$\phi = n - 1 = 20 - 1 = 19$$

$$t_{calc} = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{16 - 13}{0,671} = 4,47$$

$$t_{crit} = 1,729$$

Gráfico 3 - Teste t de Student



Fonte: Elaboração própria utilizando o software R

Como  $t$  calculado  $\in$  RR representada pela parte sombreada em vermelho no Gráfico 3, rejeita-se  $H_0$ . Isto é, o programa aumentou significativamente a frequência média dos estudantes na escola ao nível de significância de 5%. Assim, a distribuição t de Student deve ser empregada quando o desvio padrão populacional é desconhecido, sendo este estimado pelo  $s_{\bar{x}}$ .

## 2.2 TESTES DE ADERÊNCIA

Surgem as questões: a amostra deriva de uma população com distribuição normal ou não? Como validar sua normalidade? O que fazer quando a população não tem distribuição normal? Em resposta à primeira pergunta, é possível validar a normalidade de uma distribuição por meio de gráficos de frequência, no qual permite-se identificar assimetrias nos dados obtidos. Caso haja suspeita de não normalidade na distribuição, recorre-se aos testes de hipóteses ( $H_0: P=P_0$ ) em que a hipótese nula determina a validação de normalidade, enquanto que a hipótese alternativa viola os pressupostos de normalidade (Bussab; Morettin, 2012). Isto é:

$H_0$ : a distribuição segue uma distribuição normal;

$H_1$ : a distribuição não segue uma normalidade.

De acordo com Pino (2014), um teste de aderência clássico é o chamado Teste Qui-Quadrado, no qual verifica as diferenças entre as frequências observadas e esperadas. Para isso, utiliza-se a expressão:

$$X^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i} = \sum_{i=1}^k \frac{o_i^2}{e_i} - n \quad (8)$$

Onde:

$o_i$  representa a frequência observada na amostra

$e_i$  a frequência esperada sob distribuição de  $H_0$  (Normal)

Destaca-se o teste de Shapiro-Wilk, no qual fornece um valor de prova (p-valor, *p-value* ou nível de significância), este indica o grau de concordância entre os dados analisados e a hipótese nula  $H_0$ . Dessa forma, a decisão de rejeitar ou não se a distribuição observada segue uma distribuição normal segue a regra: (i) se  $p\text{-valor} \leq \alpha$ , rejeita-se  $H_0$ , ou seja, conclui-se que os dados não seguem uma distribuição normal; (ii) se  $p\text{-valor} > \alpha$ , não se rejeita  $H_0$  (Lopes; Castelo Branco; Soares, 2013).

Exemplo: Foram gerados dois conjuntos de dados com 50 observações um com distribuição aproximadamente normal ( $\mu=10$  e  $\sigma=2$ ) e o outro seguindo uma distribuição exponencial no R, a fim de aplicar o teste a cada conjunto numérico e interpretar os resultados.

$H_0$ : a distribuição segue uma distribuição normal;

$H_1$ : a distribuição não segue uma normalidade.

Decisão:

Se  $p\text{-valor} \leq 0,05$ , rejeita-se  $H_0$ , ou então, se  $p\text{-valor} > 0,05$ , não se rejeita  $H_0$ .

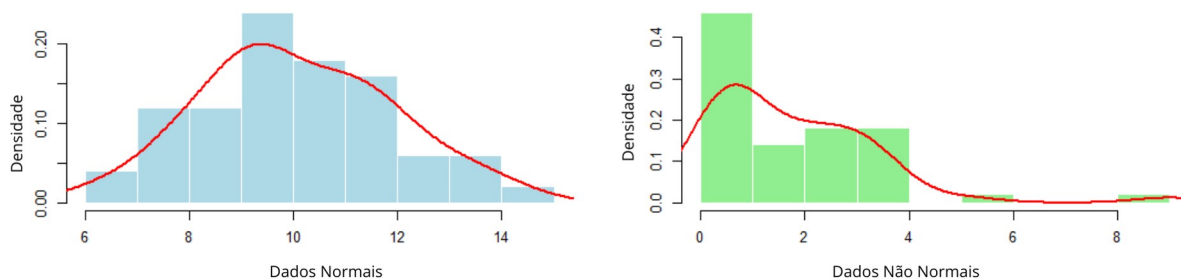
Tabela 1-Resumo de Testes Shapiro-Wilk Simulados no R

$H_0$ : A distribuição segue uma distribuição Normal ( $\alpha=0,05$ )

Conjunto	N	Estatística W	P-valor	Conclusão
Dados Normais	50	0,9893	0,9278	Não Rejeita $H_0$ (Normal)
Dados não Normais	50	0,8249	0,0000	Rejeita $H_0$ (NÃO Normal)

Fonte: Elaboração própria utilizando o software R

Figura 4 - Teste Shapiro-Wilk em duas distribuições distintas



Fonte: Elaboração própria utilizando o software R

Além do teste Qui-Quadrado e Shapiro-Wilk, atualmente, encontram-se mais de 50 métodos para verificar se os valores observados provém de uma distribuição normal (Henze; Wagner, 1997 apud Pinho, 2014).

Na ausência da condição de normalidade, isto é, quando a suposição de que a distribuição segue uma distribuição normal é rejeitada, recomenda-se os procedimentos de transformação dos dados, de modo que estes se aproximem de uma distribuição normal ou aproximada. Outra forma consiste na aplicação de métodos estatísticos mais “robustos”, denominados testes não paramétricos, os quais apresentam menor sensibilidade ao afastamento da normalidade e serão aprofundados na seção seguinte (Pino, 2014).

### 2.3 TESTES NÃO PARAMÉTRICOS

Como a própria denominação sugere, os testes paramétricos constituem procedimentos estatísticos baseados em amostras de populações que apresentam parâmetros específicos. Tais métodos pressupõem a condição de normalidade da distribuição. Em contraposição, os métodos não paramétricos são utilizados em situações que violam os pressupostos de normalidade, embora alguns testes dependam de um parâmetro, como por exemplo, a mediana (Triola, 1999).

Ainda conforme Triola (1999), a utilização de métodos não paramétricos apresenta algumas vantagens relevantes. Dentre elas, destacam-se a facilidade de aplicação, não exigem os pressupostos de normalidade e a possibilidade de serem aplicados em dados não numéricos. Ademais, esses métodos tendem a ser mais eficientes quando o tamanho da amostra é grande. Por outro lado, apresentam algumas limitações, como a perda de

informações, amostras grandes podem dificultar o cálculo manual, além de, as tabelas não serem amplamente disponíveis.

Vieira (2023, pag. 87) diz que:

É importante lembrar, de começo, que muitos pesquisadores iniciam as análises de dados – quando a variável é numérica – calculando médias e desvios padrão. No entanto, os testes não paramétricos não trabalham com os dados coletados, mas com seus postos. A inferência é feita com base nos postos dos dados – e são, portanto, as médias dos postos que devem ser comparadas [...]. Para obter os postos, primeiro colocam-se os  $n$  dados observados em ordem crescente. Depois, atribui-se um número de ordem a cada dado observado. Esse número é o posto (em inglês, *rank*) [...]. O menor posto é 1 e o maior posto é  $n$ .

Há uma variedade de testes não paramétricos, dentre os quais se destacam os Teste dos Sinais e o Teste de Wilcoxon, que empregam a mediana como parâmetro nos testes de hipóteses. Ademais, o Teste Exato de Fisher é utilizado na análise de variáveis discretas, assim como, o Teste de Kruskal-Wallis que indica diferenças estatisticamente significativas entre pelo menos duas amostras (Firmino, 2015). O presente estudo concentra-se especificamente na fundamentação teórica e na aplicação do Teste U de Mann-Whitney.

### 2.3.1 Teste U de Mann-Whitney

Em 1945, F. Wilcoxon desenvolveu o Teste U com o objetivo de comparar os parâmetros de duas amostras com tamanhos equivalentes. Posteriormente, em 1947, H. B. Mann e D. R. Whitney aprimoraram a técnica, estendendo a sua aplicação às amostras de tamanhos diferentes (Callegari-Jacques, 2007).

Assim, consolidou o teste denominado teste U ou teste da soma de postos de Wilcoxon, ou ainda, teste de Mann-Whitney, como uma alternativa não paramétrica ao teste  $t$  para duas amostras independentes (não emparelhadas), isto é, não há necessidade de ter qualquer relação entre as amostras estudadas. Tal que, por meio do teste U seja possível identificar a equivalência dos parâmetros de duas populações (Freund, 2006).

Um aspecto de relevância na aplicação do Teste U refere-se ao tamanho das amostras. Quando a amostra apresenta o tamanho menor que 16 unidades ( $n < 16$ ) recomenda-se a utilização da Tabela U para determinação dos valores críticos. Contudo, quando quaisquer amostras apresentam  $n > 16$ , a estatística do teste passa a seguir, de forma aproximada, uma distribuição normal com média  $\mu_U = \frac{n_1 n_2}{2}$ .

- **Exemplo de aplicação para amostras pequenas:**

Suponha que se deseje comparar o desempenho dos estudantes de duas escolas em um teste de conhecimento aplicado. As escolas apresentaram os seguintes notas:

Escola A: 12,15, 14

Escola B: 10,13,12,11

O objetivo é verificar, por meio do Teste U, se há diferenças significativas entre as distribuições das notas. A amostra com tamanho menor é designada  $n_1$  e a maior  $n_2$ . Em seguida, devem-se ordenar os dados das duas amostras conjuntamente, como se tivesse obtido sido obtida uma única amostra. No caso de empate, atribui-se a média dos postos (ou *ranks*) a cada observação empatada. Estes postos podem ser verificados nas colunas abaixo. Se houver diferenças significativas entre as medianas das duas amostras, a maioria dos postos mais baixos tende a pertencer aos valores de uma das amostras, enquanto a outra amostra tende a possuir os postos mais altos.

Após o ordenamento, a identificação de empates e a determinação dos postos, calculam-se  $W_1$  e  $W_2$ , que são as somas dos postos da Escola A e Escola B, respectivamente.  $N$  representa a soma de  $n_1+n_2$ , e a condição de verificação é que,  $W_1+W_2$  deve ser igual a  $\frac{N(N+1)}{2}$ .

- Identifique os empates:

Empate dos Posto 3 e Posto 4 =  $(3+4) / 2 = 3,5$

Nota	Postos		Nota	Postos
10	1		10	1
11	2		11	2
12	3	} (3 + 4)/2 = 3,5 }	12	3,5
12	4		12	3,5
13	5		13	5
14	6		14	6

Escola A		Escola B	
Valores	Postos	Valores	Postos
12	3,5	10	1
15	7	13	5
14	6	12	3,5
$n_1=3$	$W_1=16,5$	11	2
		$n_2=4$	$W_2=11,5$

- Posteriormente, calculam-se  $U_1$  e  $U_2$  com a seguinte expressão:

$$U_1 = W_1 - \frac{n_1(n_1+1)}{2} = 16,5 - \frac{3(3+1)}{2} = 10,5 \quad (9)$$

$$U_2 = W_2 - \frac{n_2(n_2+1)}{2} = 11,5 - \frac{4(4+1)}{2} = 1,5 \quad (10)$$

- Determine  $U_{calc}$ , o menor valor entre  $U_1$  e  $U_2$ :

$$U_{calc} = \min(U_1; U_2)$$

- Em seguida, defina-se o Teste de Hipóteses com  $\alpha=0,05$ :

$H_0$ : as medianas das notas não diferem

$H_1$ : as medianas das notas diferem

O valor crítico  $U_{\alpha;n_1;n_2}$  encontra-se na Tabela VII (Freund, 2006, p. 507) isto é,  $U_{\alpha;n_1;n_2} = U_{0,05;3;4} = 0$ .

Considerando tais valores,  $U_{calc} = 1,5 > U_{0,05;3;4} = 0$ . Como o valor calculado é maior do que o valor crítico (critério para rejeição do Teste U para amostras pequenas), não se rejeita  $H_0$ . Em outras palavras, não há evidências estatísticas para concluir que exista uma diferença significativa nas distribuições das notas.

Nesse exemplo foi utilizado o R para realização dos cálculos (Tabela 2). Tal que, a estatística  $W = 10,5$  coincidiu com  $U_1$  e o p-valor=0,1536. Como  $0,1536 > 0,05$ , concluiu-se novamente, que a hipótese nula não pode ser rejeitada.

Tabela 2 – Teste U de Mann-Whitney amostra pequena no R

Teste U de Mann-Whitney exemplo em amostras pequenas

Conjunto	N	Mediana	$\alpha=0,05$
Escola A	3	14	$W=10,5$
Escola B	4	11,5	P-valor =0,1536

Fonte: Elaboração própria utilizando o software R

- **Exemplo para amostras grandes:**

Escola A: 12, 15, 14, 10, 9, 16,17, 18, 13, 12, 15, 13, 15, 16, 18, 19, 20, 15, 12, 17, 18  
( $n_1=21$ )

Escola B: 10, 9, 15, 16, 13, 12, 11, 13, 14, 15, 19, 15, 17, 18, 13, 17, 18, 19, 11, 13  
( $n_2=20$ )

De acordo com Freund (2006), quando os tamanhos das amostras são grandes, a estatística U pode ser aproximada por uma distribuição normal padronizada (equação 13) e, tanto  $U_1$  quanto  $U_2$  podem ser utilizados. A suposição é que as amostras tenham parâmetros equivalentes. Calculam-se a média ( $\mu_U$ ) e o desvio-padrão da distribuição de U ( $\sigma_U$ ):

$$\mu_U = \frac{n_1 n_2}{2} \quad (11)$$

$$\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} \quad (12)$$

A distribuição amostral de  $U_1$ (ou  $U_2$ ) pode ser padronizada, sendo  $z_{calc}$  obtido por:

$$z_{calc} = \frac{U_1 - \mu_{U_1}}{\sigma_{U_1}} \quad (13)$$

Solução:

- $H_0: \mu_1 = \mu_2$  (As medianas são iguais)
- $H_1: \mu_1 \neq \mu_2$  (As medianas se diferem)
- $\alpha = 0,05$

1. Cálculo de  $U_1$ :

$W_1 = 463$  (Soma dos postos de  $W_1$ )

$$U_1 = W_1 - \frac{n_1(n_1+1)}{2} = 463 - \frac{21(21+1)}{2} = 463 - 231 = 232$$

2. Cálculo da média ( $\mu_{U_1}$ ) e o desvio-padrão ( $\sigma_{U_1}$ ):

$$\mu_{U_1} = \frac{n_1 n_2}{2} = \frac{21 \cdot 20}{2} = 210 \quad (14)$$

$$\sigma_{U_1} = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{21 \cdot 20 (21 + 20 + 1)}{12}} \approx 38,34 \quad (15)$$

3. Cálculo da Estatística Z:

$$z_{calc} = \frac{U_1 - \mu_{U_1}}{\sigma_{U_1}} = \frac{232 - 210}{38,34} \approx 0,5738 \quad (16)$$

4. Decisão:

Para um teste com  $\alpha = 0,05$ , o valor crítico de  $z_{0,05} = 1,96$ . Como  $|z_{calc}| = 0,538 < z_{0,05} = 1,96$ , não se rejeita a hipótese nula. Portanto, não há evidências estatísticas para concluir que haja diferença significativa entre as notas dos alunos da Escola A e a Escola B.

O cálculo no R para amostras grandes é demonstrado na Tabela 3, resultou a estatística  $W = 232$  igual a  $U_1$ , e o p-valor = 0,5726. Como  $0,5726 > 0,05$ , conclui-se que não se rejeita a hipótese nula, obtendo o mesmo resultado dos cálculos anteriores.

Tabela 3 - Teste U de Mann-Whitney amostra grande no R

Teste U de Mann-Whitney exemplo em amostras grandes			
Conjunto	N	Mediana	$\alpha = 0,05$
Escola A	21	15	W=232
Escola B	20	14,5	P-valor = 0,5726

Fonte: Elaboração própria utilizando o software R

### 3 O PROGRAMA DE EXTENSÃO: ESTATÍSTICA PARA TODOS

O Programa de Extensão Estatística para Todos, vinculado à Universidade Federal de Alfenas (UNIFAL-MG), tem como finalidade difundir o conhecimento estatístico e fortalecer a interação entre a universidade e a comunidade externa. As ações do programa concentram-se na produção e divulgação de conteúdos digitais acessíveis, veiculados na plataforma Instagram. O projeto é formado por docentes coordenadores, discentes bolsistas e voluntários, que atuam de forma integrada nas atividades de ensino, pesquisa e extensão (Unifal, 2025).

No âmbito deste estudo, a participação do programa concentrou-se na análise da inflação dos preços de combustíveis praticados na cidade de Varginha-MG. A coleta de dados foi realizada entre abril e outubro de 2025 por meio do aplicativo Nota Fiscal Mineira, abrangendo 39 postos de combustíveis, com registros sistemáticos dos preços de etanol, gasolina e diesel. Os dados foram submetidos às estatísticas descritivas e inferenciais, seguidas da estimação pontual, visando mensurar as variações de preços no período analisado. A divulgação dos resultados ocorreu por meio de publicações na rede social Instagram, em consonância com o objetivo do projeto de promover o acesso ao conhecimento estatístico.

### 4 APLICAÇÃO DO TESTE U: PREÇOS DE COMBUSTÍVEIS

Nesta seção, são abordadas as estatísticas descritivas e inferenciais empregadas no Programa de Extensão Estatística para Todos, com o propósito de demonstrar a aplicabilidade do teste U. A análise fundamenta-se nos dados coletados entre os dias 20 e 24, no mês de agosto e setembro de 2025, referentes aos preços dos combustíveis etanol, gasolina e diesel, obtidos junto ao aplicativo Nota Fiscal Mineira. Todos os procedimentos estatísticos foram realizados no software R versão 4.5.1.

Os dados utilizados neste estudo foram coletados em 39 postos de combustíveis.

Entretanto, o tamanho da amostra varia entre os períodos analisados, devido a fatores como fechamento de alguns postos, bem como ausência ou indisponibilidade de informações em determinados meses. Dessa forma, o n pode apresentar oscilações conforme a disponibilidade dos dados.

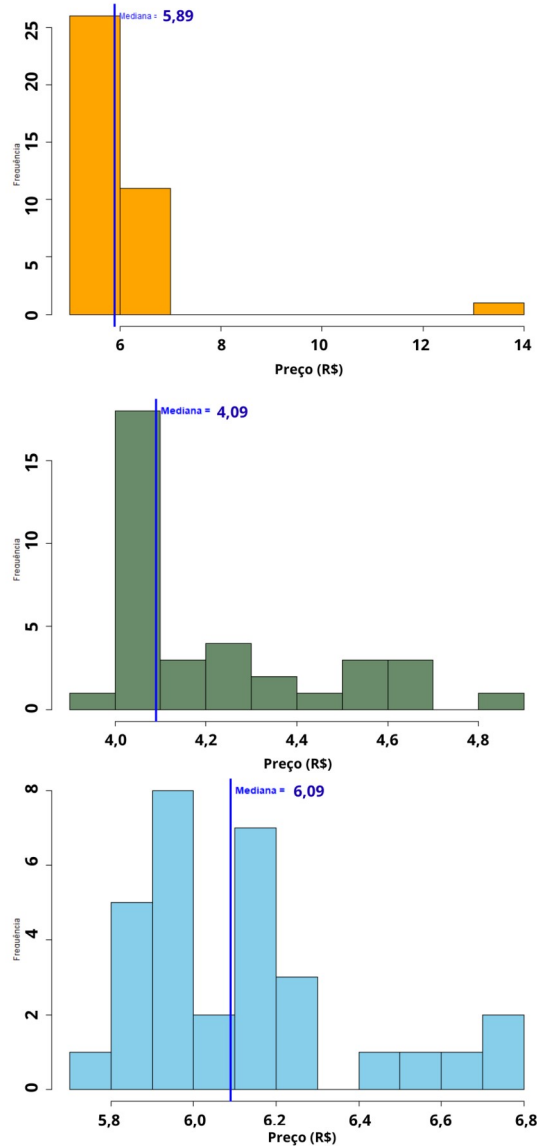
A priori, foram construídos os histogramas das distribuições de frequência para cada tipo de combustível, com o objetivo de observar o comportamento dos dados e identificar possíveis desvios de normalidade. Em seguida, realizou-se a análise descritiva, com as medidas de tendências centrais e de dispersão. Posteriormente, aplicou-se o teste de aderência Shapiro-Wilk, a fim de verificar a compatibilidade das distribuições à normalidade. Na sequência procedeu-se à aplicação do teste U de Mann-Whitney para amostras independentes (não pareadas) e, por fim, à estimativa percentual das medianas.

É importante esclarecer que, optou-se pela aplicação de testes para amostras independentes (não pareadas), uma vez que as variações no tamanho da amostra entre os períodos é influenciada por fatores de mercado. Como houve fechamento de estabelecimentos, interrupções operacionais e alguns preços não encontrados em determinados meses, não foi possível manter o tamanho de amostra.

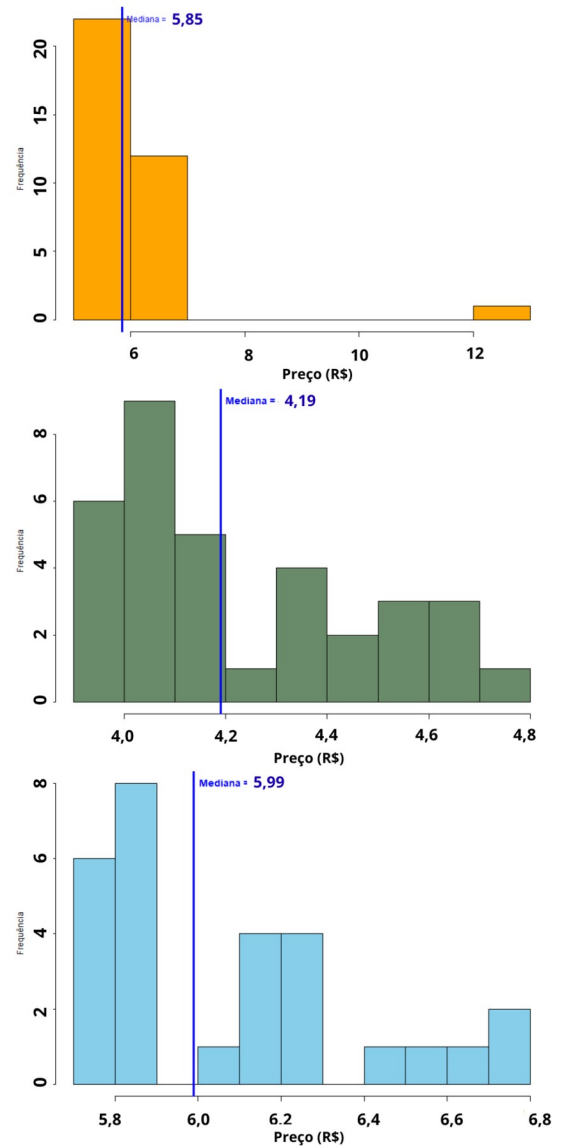
A primeira etapa consistiu na elaboração dos histogramas de frequência para as variáveis analisadas. Conforme ilustrado na Figura 5, as distribuições apresentam assimetrias e presenças de valores discrepantes nos preços da gasolina indicando a violação do pressuposto de normalidade. É possível perceber por meio dos histogramas que menos de cinco postos praticam valores acima de R\$12,00.

Figura 5 – Histogramas de frequência dos preços entre 20 e 24 dos meses de agosto (esquerda) e setembro (direita)

Histogramas de frequência dos preços de combustíveis no período entre 20 a 24 de agosto de 2025



Histogramas de frequência dos preços de combustíveis no período entre 20 a 24 de setembro de 2025



■ Gasolina  
■ Etanol  
■ Diesel  
— Mediana

Fonte: Elaboração própria utilizando o software R

Nas Tabelas 4 e 5 é apresentado o resumo da estatística descritiva. Observa-se que a gasolina apresentou maior dispersão, com desvio-padrão de 1,3058 no mês de agosto e 1,1983 no mês de setembro, indicando maior variabilidade dos preços entre os postos analisados. Além disso, a gasolina indicou, respectivamente dos meses de agosto e setembro, o preço máximo de R\$13,90 e R\$12,90, com médias R\$6,19 e R\$6,17. Em contrapartida, o etanol indicou menores variações (Desvio Padrão = 0,2369 e 0,2492), sugerindo maior estabilidade nos valores praticados.

Tabela 4 - Resumo de Estatística Descritiva Agosto

<b>Conjunto</b>	<b>N</b>	<b>Média</b>	<b>Mediana</b>	<b>Mínimo</b>	<b>Máximo</b>	<b>Desvio Padrão</b>
Etanol	36	4,2577	4,09	3,99	4,90	0,2369
Gasolina	38	6,1908	5,89	5,75	13,90	1,3058
Diesel	31	6,1470	6,09	5,79	6,79	0,2697

Fonte: Elaboração própria utilizando o software R

Tabela 5 - Resumo de Estatística Descritiva Setembro

<b>Conjunto</b>	<b>N</b>	<b>Média</b>	<b>Mediana</b>	<b>Mínimo</b>	<b>Máximo</b>	<b>Desvio Padrão</b>
Etanol	34	4,2691	4,19	3,99	4,79	0,2492
Gasolina	35	6,1797	5,85	5,75	12,90	1,1983
Diesel	28	6,1071	5,99	5,79	6,79	0,2702

Fonte: Elaboração própria utilizando o software R

Assim, como observado nos histogramas, a análise descritiva revela a existência de outliers nos preços da gasolina, o que reforça a hipótese de não normalidade das distribuições.

Em seguida, foi aplicado o teste de aderência Shapiro-Wilk para verificar a conformidade com a distribuição normal. Os resultados estão apresentados na Tabela 6 e 7. Conforme apresentado, todas as variáveis analisadas rejeitaram a hipótese nula de normalidade, ao nível de significância de 5%, ou seja, com todos p-valores <0,05. Indicando que os preços não seguem uma distribuição normal.

Tabela 6 - Resumo de Testes de Aderência à normalidade AGOSTO (Shapiro-Wilk)

---

$H_0$ : A distribuição dos preços segue a curva Normal ( $\alpha=0,05$ )

---

Conjunto	N	Estatística W	P-valor		Conclusão
Etanol	31	0,8697	0,0014	Rejeita $H_0$	(NÃO Normal)
Gasolina	36	0,7888	0,0000	Rejeita $H_0$	(NÃO Normal)
Diesel	38	0,2823	0,0000	Rejeita $H_0$	(NÃO Normal)

---

Fonte: Elaboração própria utilizando o software R

Tabela 7 - Resumo de Testes de Aderência à normalidade SETEMBRO (Shapiro-Wilk)

---

$H_0$ : A distribuição dos preços segue a curva Normal ( $\alpha=0,05$ )

---

Conjunto	N	Estatística W	P-valor		Conclusão
Etanol	35	0,3240	0,0000	Rejeita $H_0$	(NÃO Normal)
Gasolina	28	0,8547	0,0012	Rejeita $H_0$	(NÃO Normal)
Diesel	34	0,8800	0,0014	Rejeita $H_0$	(NÃO Normal)

---

Fonte: Elaboração própria utilizando o software R

Diante da rejeição da hipótese de normalidade, procedeu-se à aplicação do Teste U de Mann-Whitney para amostras independentes. Com as seguintes hipóteses:

$H_0$ : as medianas dos preços não diferem

$H_1$ : as medianas dos preços diferem

Em todos os casos observados na Tabela 8, os p-valores foram superiores ao nível de significância de 5%. Assim, não há evidências significativas para rejeitar a hipótese nula. Isto é, as medianas dos preços entre setembro e agosto podem ser considerados estatisticamente equivalentes para os três tipos de combustíveis.

Tabela 8 - Resumo dos Testes U de Mann-Whitney

---

$H_0$ : As medianas dos preços de Setembro e Agosto são iguais

---

Conjunto	N_Set	N_Ago	Estatística W	P-valor		Conclusão
Etanol	35	38	607,0	0,5195	Não Rejeita $H_0$	Medianas podem ser iguais
Gasolina	28	31	369,5	0,3292	Não Rejeita $H_0$	Medianas podem ser iguais
Diesel	34	36	617,0	0,9566	Não Rejeita $H_0$	Medianas podem ser iguais

---

Fonte: Elaboração própria utilizando o software R

Por fim, foi realizada a estimativa percentual das medianas entre os meses analisados. A Figura 5 apresenta as medianas de cada combustível no mês estudado. A Tabela 9 apresenta os resultados das variações: apenas o etanol teve um aumento de 2,44%, enquanto que a gasolina e o diesel tiveram uma redução, respectivamente de, -0,68% e -1,64%.

Tabela 9 - Estimativa pontual das medianas dos preços

<b>Conjunto</b>	<b>Ago</b>	<b>Set</b>	<b>Variação</b>
Etanol	4,09	4,19	2,444%
Gasolina	5,89	5,85	-0,679%
Diesel	6,09	5,99	-1,642 %

Fonte: Elaboração própria utilizando o software R

#### 4.1 RESULTADOS E DISCUSSÃO

Os resultados obtidos a partir das análises realizadas no software R possibilitaram a avaliação do comportamento dos preços de etanol, gasolina e diesel praticados no município de Varginha-MG.

Inicialmente, a análise exploratória de dados, por meio dos histogramas de frequência e os resultados das estatísticas descritivas, evidenciou a presença de assimetria e a ocorrência de valores discrepantes (*outliers*). Tais aspectos implicam a violação do pressuposto de normalidade, o que foi subsequentemente comprovado pelo Teste de Aderência de Shapiro Wilk, cujos resultados (Tabela 6 e 7) conduziram à rejeição da hipótese nula de normalidade.

Diante da constatação de não normalidade dos dados, aplicou-se o Teste U de Mann-Whitney para amostras não pareadas (independentes), com o objetivo de comparar as medianas dos preços entre os dias 20 e 24, de setembro e agosto de 2025. Conforme resumo da Tabela 8, verificou-se que os p-valores apurados em todos os conjuntos analisados foram superiores ao nível de significância.

Por conseguinte, não foram identificadas evidências estatísticas suficientes para rejeitar a hipótese nula de equivalência das medianas de preços entre os períodos de agosto e setembro. Sendo assim, os resultados sugerem que os valores das medianas dos preços dos combustíveis permaneceram estáveis no intervalo investigado, sem a ocorrência de variações estatisticamente significantes. A Tabela 9, que apresenta a estimativa pontual das medianas,

reforça esta conclusão. Embora as oscilações específicas possam ocorrer, originadas por outros fatores, tais como custos logísticos ou as dinâmicas de mercado, tais variações não foram estatisticamente significativas para promover alterações relevantes nas medianas amostrais.

Os resultados reforçam a eficácia e validade do Teste U de Mann-Whitney enquanto alternativa metodológica robusta aos testes paramétricos, principalmente em contextos de não normalidade. Ademais, destaca-se a relevância desta aplicação no âmbito do Programa de Extensão: Estatística para Todos, alcançando assim, os objetivos de promover a análise de variações de preços em contextos reais.

## **5 CONSIDERAÇÕES FINAIS**

A análise desenvolvida demonstrou que a aplicação de testes não paramétricos em cenários de dados reais, vai além do escopo da variação de preços dos combustíveis, desdobrando-se com pertinência em uma ampla gama de contextos e áreas do conhecimento, tais como as ciências sociais, humanas e biológicas.

Constatou-se que a análise de dados que seguem uma distribuição normal apresenta maior flexibilidade metodológica, isto é, admitindo a utilização tanto de testes paramétricos quanto de testes não paramétricos. Em contrapartida, esta versatilidade não se verifica em situações de não normalidade que exige métodos estatísticos mais robustos para garantir sua validade e eficiência das inferências.

A escolha pelo teste não paramétrico para amostras não emparelhadas (ou independentes) foi determinada em função das dinâmicas voláteis do mercado. Fatores como abertura e fechamento de postos, a circulação de notícias sobre adulteração, as estratégias de precificação e a concorrência local justificam o aspecto dos dados como não dependentes.

Adicionalmente, este trabalho atingiu o objetivo de integrar ensino, pesquisa e extensão, conforme os eixos norteadores do PIEPEX. Os resultados do presente estudo se consolidam como um recurso fundamental para os estudos futuros do Programa, servindo como uma ferramenta de pesquisa.

Por fim, reconhecendo que o presente estudo foi restringido ao âmbito do Programa de Extensão: Estatística para Todos, sugere-se, como perspectiva para futuras investigações, a expansão do estudo para o aprofundamento da análise dos cálculos da inflação para uma cesta de bens mais abrangente, tal que permita a utilização de métodos estatísticos mais complexos

e complementarão a análise do pontual de preços de combustíveis realizada nesta etapa.

## REFERÊNCIAS

ANDERSON, David R. *et al.* **Estatística aplicada a administração e economia**. 5. ed. São Paulo, 2020. E-book. Disponível em: <https://integrada.minhabiblioteca.com.br/books/9786555583991>. Acesso em: 1 dez. 2025.

BUSSAB, Wilton O.; MORETTIN, Pedro Alberto. **Estatística básica**. 6. ed. rev. e atual. São Paulo: Saraiva, 2010.

CALLEGARI-JACQUES, Sidia M. **Bioestatística: princípios e aplicações**. Porto Alegre: Artmed, 2007.

CASTANHEIRA, Nelson Pereira. **Estatística aplicada a todos os níveis**. 3. ed. Curitiba, 2023. E-book. Disponível em: <https://plataforma.bvirtual.com.br>. Acesso em: 03 dez 2025.

FIRMINO, Maria José de Almeida Caetano de Sousa. **Testes de hipóteses: uma abordagem não paramétrica**. 2015. Dissertação (Mestrado em Matemática para Professores) – Universidade de Lisboa, Faculdade de Ciências, Lisboa, 2015.

FREUND, John E. **Estatística aplicada: economia, administração e contabilidade**. 11. ed. Porto Alegre: Bookman, 2006.

GARCIA, Livia Matos. **Análise Bayesiana para a distribuição Exponencial-Logarítmica**. 2013. 65 f. Dissertação (Mestrado em Matemática Aplicada e Computacional) – Faculdade de Ciências e Tecnologia, Universidade Estadual Paulista "Júlio de Mesquita Filho", Presidente Prudente, 2013. Disponível em: <http://hdl.handle.net/11449/94322>. Acesso em: 30 set. 2025.

HENZE, N.; WAGNER, T. A new approach to the BHEP tests for multivariate normality. **Journal of Multivariate Analysis**, v. 62, p. 1-23, 1997.

LARSON, Roland Edwin. **Estatística aplicada**. Tradução de José Fernando Pereira Gonçalves. Revisão de Manoel Henrique Salgado. 6. ed. São Paulo, SP: Pearson, 2015. E-book. Disponível em: <https://middleware-bv.am4.com.br/SSO/unifalmg/9788543004778>. Acesso em: 1 dez. 2025.

LOPES, Manuela de Mesquita; CASTELO BRANCO, Verônica Teixeira Franco; SOARES, Jorge Barbosa. **Utilização dos testes estatísticos de Kolmogorov-Smirnov e Shapiro-Wilk para verificação da normalidade para materiais de pavimentação**. Transportes, São Paulo-SP, v. 21, n. 1 p. 59–66, 2013.

MORETTIN, Luiz Gonzaga. **Estatística básica: probabilidade e inferência**. São Paulo: Pearson; Makron Books, 2010.

MORETTIN, Pedro A. **Estatística básica**. 10. ed. São Paulo: Saraiva Uni, 2023. E-book. Disponível em: <https://integrada.minhabiblioteca.com.br/books/9788571441484>. Acesso em: 30 set. 2025.

PINO, Francisco Alberto. A questão da não normalidade: uma revisão [The Question of Non-Normality: a review]. **Revista de Economia Agrícola**, São Paulo, v. 61, n. 2, p. 1–92, jul./dez. 2014.

TRIOLA, Mario F. **Introdução à estatística**. 7. ed. Rio de Janeiro: LTC, 1999.

UNIVERSIDADE FEDERAL DE ALFENAS. **Ação de Extensão Vigentes**. Disponível em: [https://sistemas.unifal-mg.edu.br/app/caex/comum/paginas/acoesNoPortal.php?proposta\\_id=7980&tipo\\_saiba\\_mais=Programa](https://sistemas.unifal-mg.edu.br/app/caex/comum/paginas/acoesNoPortal.php?proposta_id=7980&tipo_saiba_mais=Programa). Acesso em: 22 out. 2025.

VIEIRA, Sonia. **Bioestatística: tópicos avançados**. 5. ed. Rio de Janeiro: GEN Guanabara Koogan, 2023. E-book. ISBN 9788595159594. Disponível em: <https://integrada.minhabiblioteca.com.br/books/9788595159594>. Acesso em: 7 nov. 2025.

VIEIRA, Sonia. **Introdução à Bioestatística**. 5. ed. Rio de Janeiro: Elsevier, 2016.