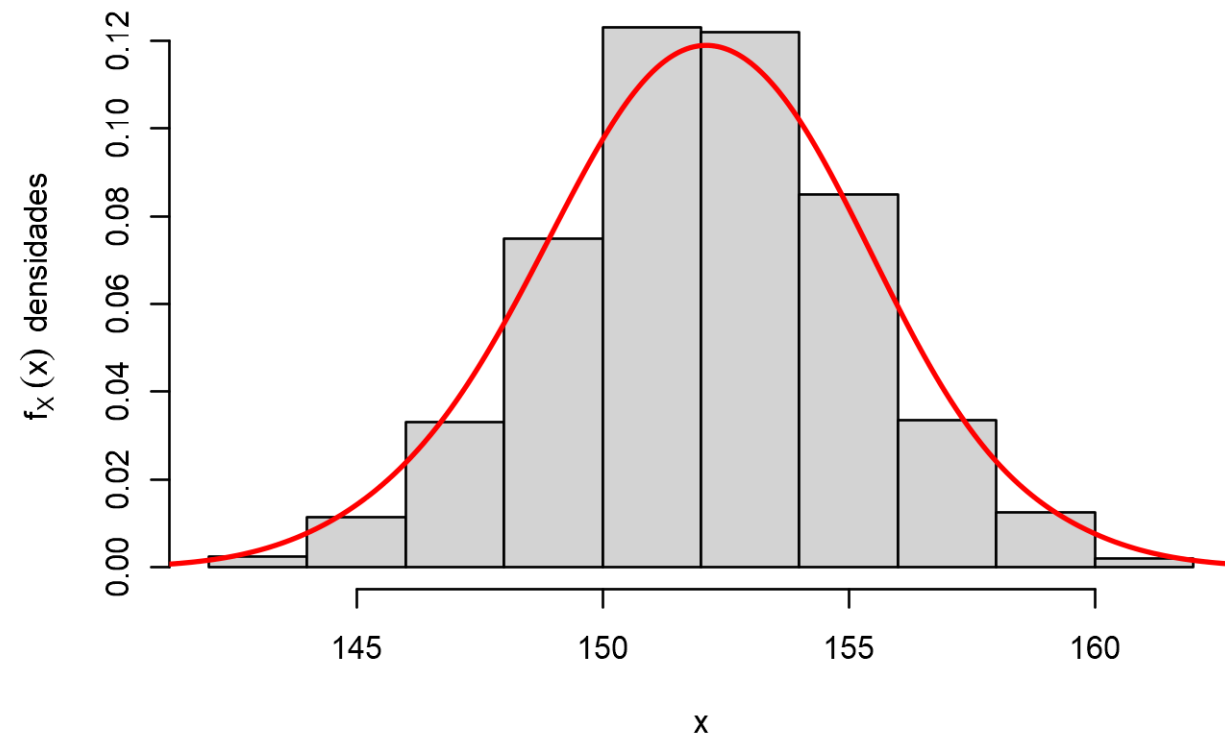




# Estatística Orientada à Ciência de Dados



# Distribuição normal



**Exemplo de distribuição normal**

**Teorema do limite central:** fundamental para o Teorema do Limite Central, que afirma que a soma de muitas variáveis aleatórias independentes e identicamente distribuídas tende a se aproximar de uma distribuição normal, independentemente da distribuição original das variáveis;

**“Regra”:** 68% dos dados estão dentro de **um desvio padrão da média**, 95% dentro de dois desvios padrão e 99,7% dentro de **três desvios padrão**.

# Testes de Hipótese

- **Processo estatístico para tomar decisões sobre uma afirmação com base em evidências e dados amostrais.**

Conceito	Definição
Hipótese nula ( $H_0$ )	A afirmação inicial que é testada
Hipótese alternativa ( $H_1$ )	A afirmação contrária à nula
Estatística de teste	Medida calculada a partir dos dados amostrais para avaliar a hipótese nula
Nível de significância	Probabilidade máxima que estamos dispostos a aceitar de cometer um erro de Tipo I (rejeitar a hipótese nula quando ela é verdadeira) em um teste de hipótese.
p-valor	A probabilidade de observar a estatística de teste ou algo mais extremo, assumindo que a hipótese nula seja verdadeira

Tipo de Teste	Hipótese Nula ( $H_0$ )	Hipótese Alternativa ( $H_1$ )	Avaliação da Evidência
Unicaudal à Direita	$H_0: \mu \leq \mu_0$	$H_a: \mu > \mu_0$	Evidência para um efeito maior
Unicaudal à Esquerda	$H_0: \mu \geq \mu_0$	$H_a: \mu < \mu_0$	Evidência para um efeito menor
Bicaudal	$H_0: \mu = \mu_0$	$H_a: \mu \neq \mu_0$	Evidência para uma diferença



# Hipóteses

• **Hipótese Nula (H0):** O medicamento não tem efeito significativo na redução da pressão arterial.

- Em termos estatísticos, isso pode ser representado como  $H0: \mu_{\text{medicamento}} - \mu_{\text{placebo}} = 0$ , onde  $\mu_{\text{medicamento}}$  é a média da pressão arterial no grupo que recebeu o medicamento, e  $\mu_{\text{placebo}}$  é a média da pressão arterial no grupo do placebo.

• **Hipótese Alternativa (H1):** O medicamento tem um efeito significativo na redução da pressão arterial.

- Em termos estatísticos, isso pode ser representado como  $H1: \mu_{\text{medicamento}} - \mu_{\text{placebo}} < 0$ , indicando que a média da pressão arterial no grupo que recebeu o medicamento é menor do que a média no grupo do placebo.

# Estatística de teste

A estatística de teste é uma **medida numérica calculada a partir dos dados amostrais durante a realização de um teste estatístico;**

Essa estatística é **usada para avaliar a evidência contra a hipótese nula e determinar se os resultados da amostra são consistentes com a hipótese nula ou se indicam que a hipótese nula deve ser rejeitada em favor da hipótese alternativa.**

# Nível de significância

É utilizado para **tomar decisões sobre as hipóteses nula e alternativa em um teste estatístico;**

O nível de significância, frequentemente representado pela letra grega  $\alpha$  (alfa), **determina o limite para considerar um resultado como estatisticamente significativo.**

O nível de significância é a probabilidade máxima que estamos dispostos a aceitar de cometer um erro de Tipo I (rejeitar a hipótese nula quando ela é verdadeira) em um teste de hipótese. Em outras palavras, é a probabilidade de obter um resultado tão extremo ou mais extremo do que o observado, sob a hipótese nula.

Quanto menor for o nível de significância escolhido (por exemplo,  $\alpha = 0,01$ ), mais rigorosa será a evidência necessária para rejeitar a hipótese nula. Por outro lado, quanto maior for o nível de significância escolhido (por exemplo,  $\alpha = 0,10$ ), mais permissivos seremos ao rejeitar a hipótese nula.

Suponha que você esteja conduzindo um teste de hipótese para determinar se um novo tratamento é eficaz na redução do tempo de cicatrização de feridas. Se você escolher um nível de significância de  $\alpha = 0,05$ , isso significa que você está disposto a aceitar uma probabilidade de 5% de cometer um erro de Tipo I. Se o valor de p (p-valor) calculado for menor que 0,05, você rejeitaria a hipótese nula em favor da hipótese alternativa.



p-valor

O p-valor é a probabilidade de obter uma estatística de teste tão extrema (ou mais extrema) quanto a observada, sob a hipótese nula;

É a probabilidade de observar os dados que temos ou resultados mais extremos, assumindo que a hipótese nula seja verdadeira.

Não é a probabilidade de a hipótese nula ser verdadeira ou falsa;

Em vez disso, ele quantifica a probabilidade de observar os dados da amostra ou resultados mais extremos sob a hipótese nula;

A interpretação do p-valor deve ser feita em conjunto com o contexto do problema, as premissas do teste e a área de pesquisa.

- **p-valor baixo (geralmente menor que o nível de significância escolhido):**
  - sugere que os dados são improváveis de ocorrer sob a hipótese nula, o que nos leva a questionar a validade da hipótese nula.
- **p-valor alto:**
  - sugere que os dados são compatíveis com a hipótese nula, o que significa que não temos evidências convincentes para rejeitá-la.

- Se o p-valor for menor do que o nível de significância (geralmente denotado por  $\alpha$ ), geralmente 0,05, rejeitamos a hipótese nula em favor da hipótese alternativa;
- Se o p-valor for maior do que  $\alpha$ , não temos evidências suficientes para rejeitar a hipótese nula.

Suponha que um pesquisador esteja conduzindo um estudo para determinar se a média de altura das pessoas em uma determinada cidade é diferente da média de altura nacional.

A hipótese nula ( $H_0$ ) afirma que a média de altura na cidade é igual à média nacional, enquanto a hipótese alternativa ( $H_1$ ) afirma que há diferença.

O pesquisador, então, coleta uma amostra de 50 pessoas na cidade e mede suas alturas. Em seguida, calcula a média da amostra e supõe que é um pouco mais alta do que a média nacional. Todavia, é necessário testar se essa diferença é estatisticamente significativa.

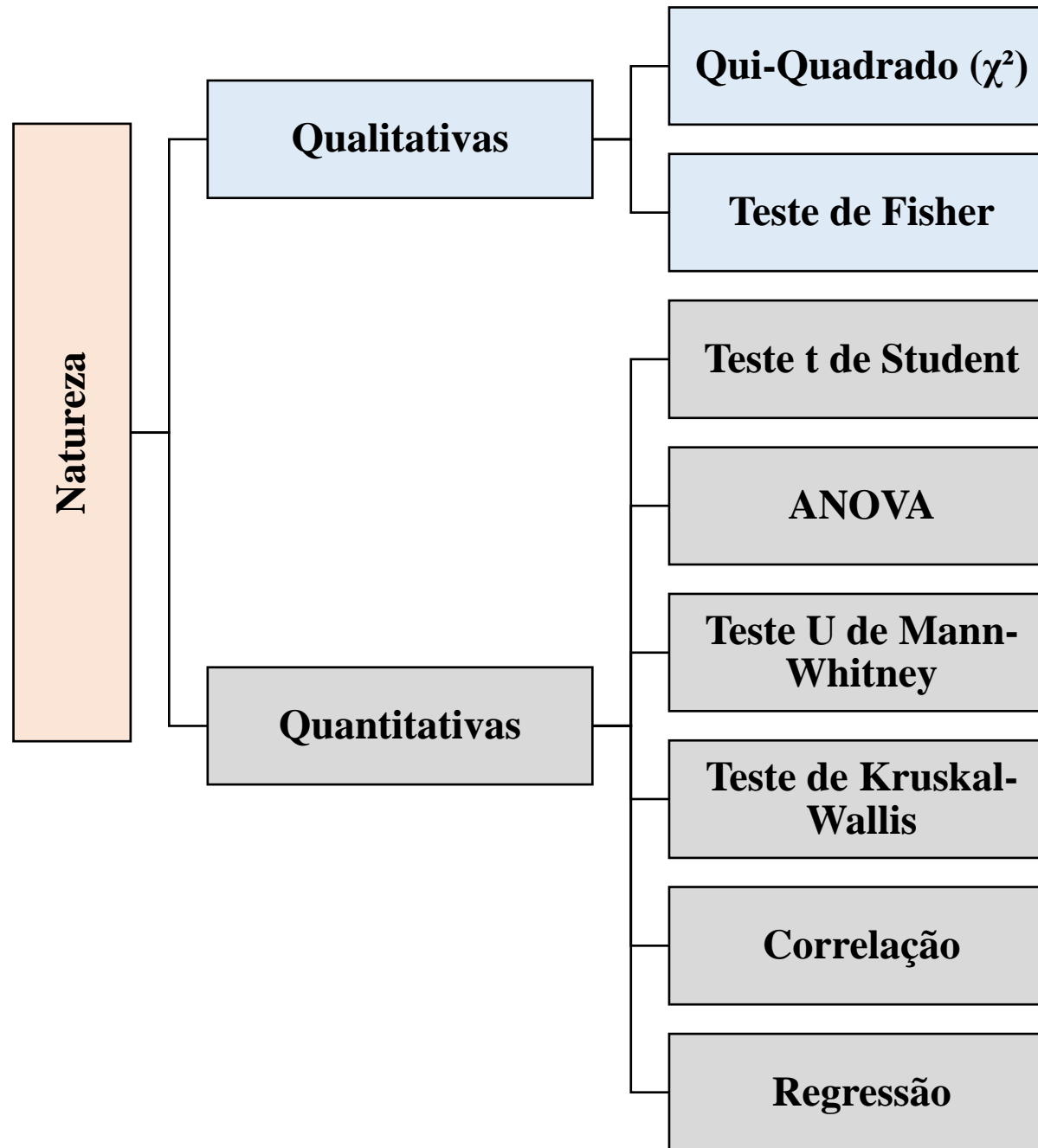
Para tanto, o pesquisador realiza um teste de hipótese, onde a estatística de teste é baseada nas alturas da amostra e segue uma distribuição t de Student. O resultado é uma estatística de teste de 2,5.

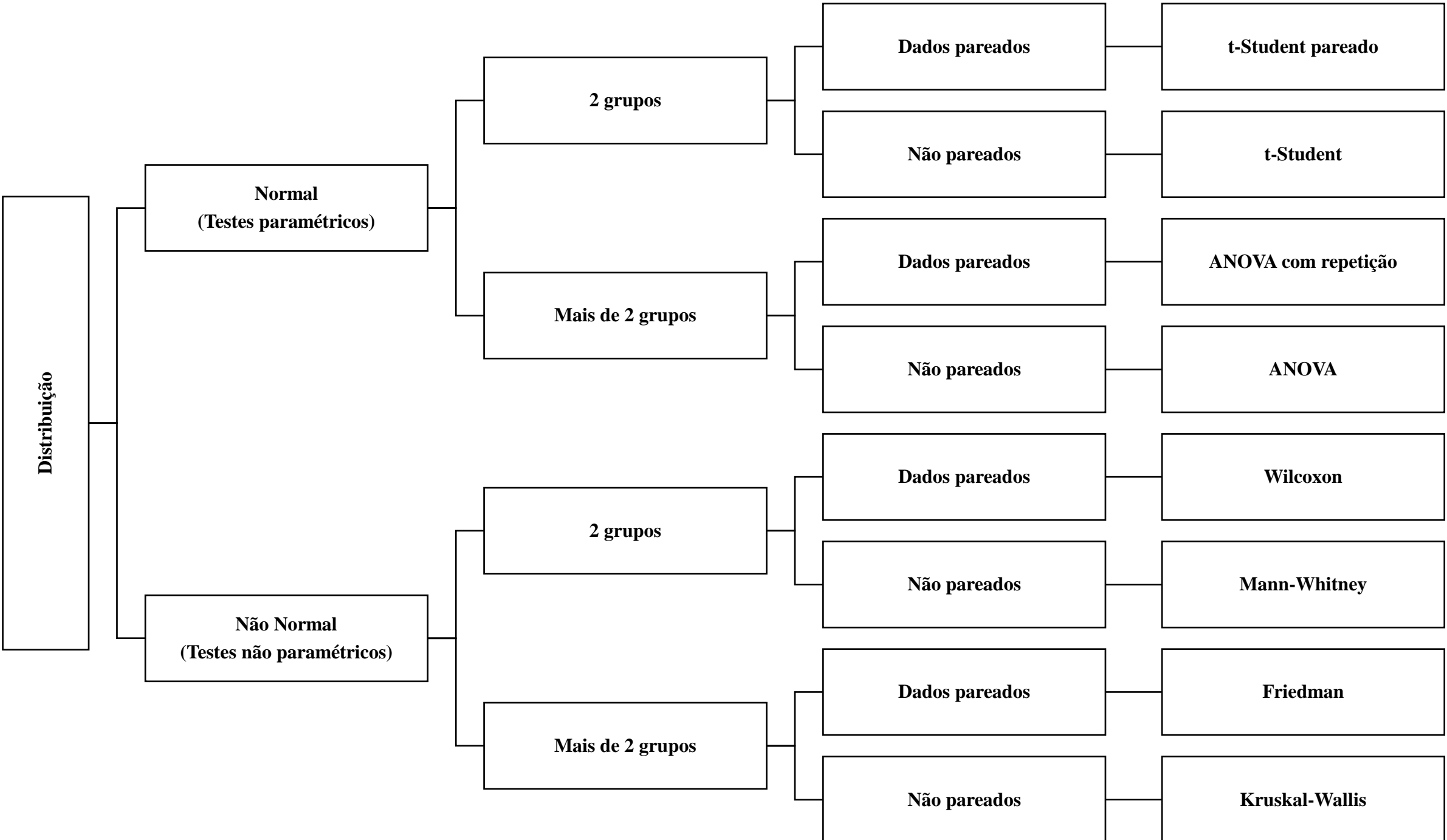
O pesquisador calcula o p-valor associado a essa estatística de teste. Neste teste, o p-valor foi igual 0,018. Isso significa que, se a hipótese nula ( $H_0$ ) de igualdade de médias fosse verdadeira, a probabilidade de obter uma estatística de teste tão extrema quanto 2,5, ou ainda mais extrema, seria de apenas 0,018.



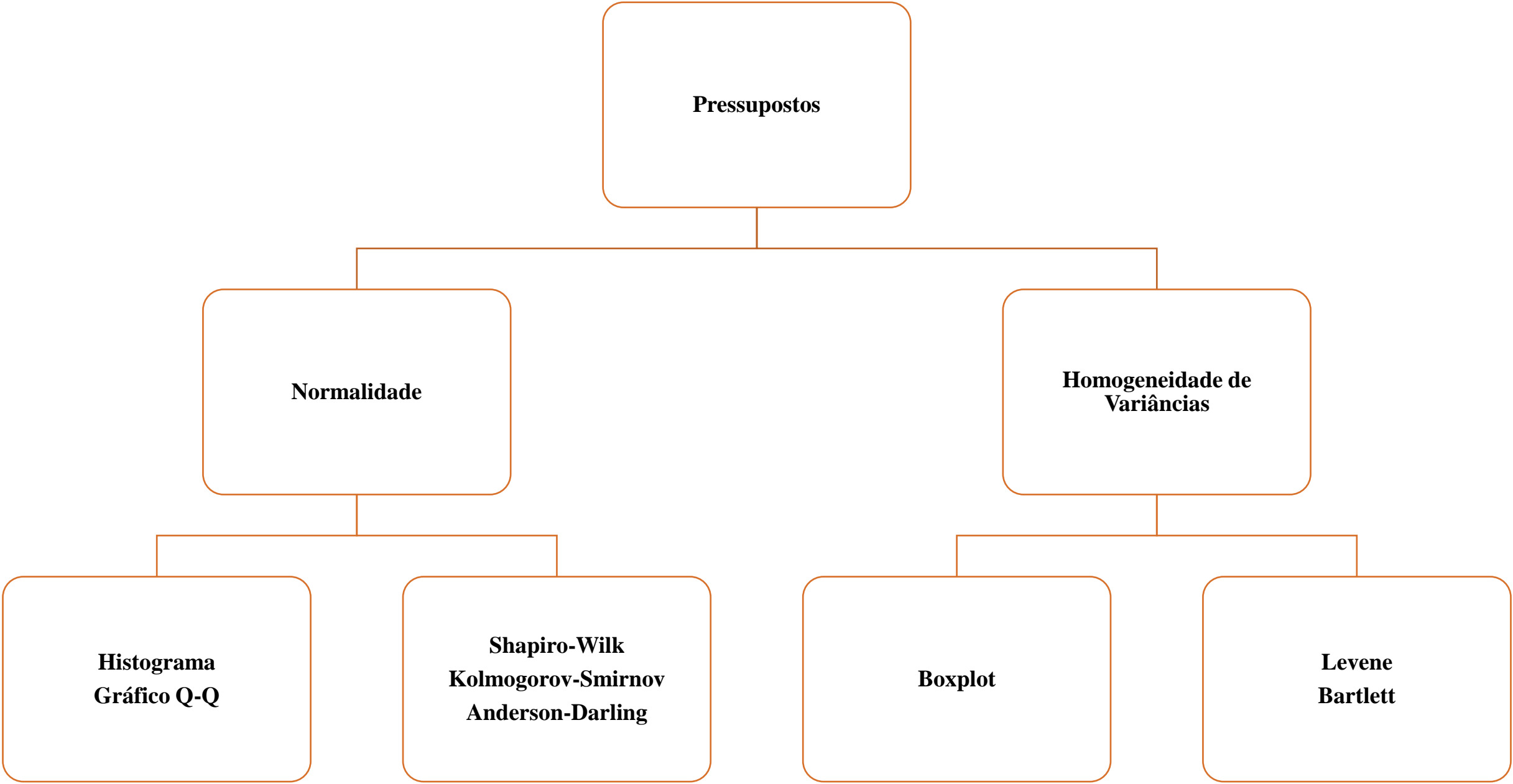
- $p\text{-valor} = 0,018$ : valor menor do que um nível de significância comum de, digamos, 0,05. Isso significa que há evidências estatísticas suficientes para rejeitar a hipótese nula;
- Conclusão: Com base no  $p$ -valor baixo, o pesquisador rejeita a hipótese nula e conclui que há uma diferença estatisticamente significativa entre a média de altura na cidade e a média nacional.

# Testes





# Pressupostos





# Prática

