XVII Escola Regional de Alto Desempenho do Estado do Rio Grande do Sul

# Curso A
# Fundamentos de Estatística para Análise de Desempenho

**Aishameriane Venes Schmidt**

**Francieli Zanon Boito**

**Laércio Lima Pilla**

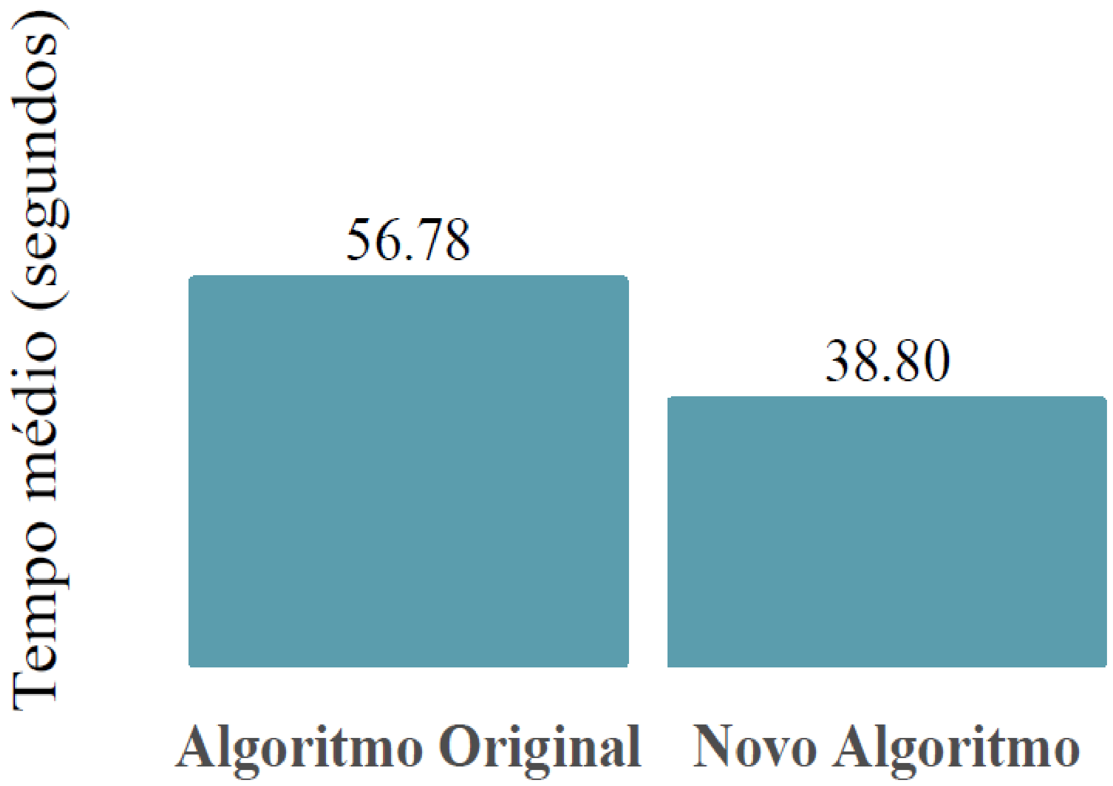
# MOTIVAÇÃO

# Motivação

- **Situação hipotética**
  - Temos um algoritmo de escalonamento original
  - Tentamos melhorar ele
  - Medimos o desempenho de ambas versões

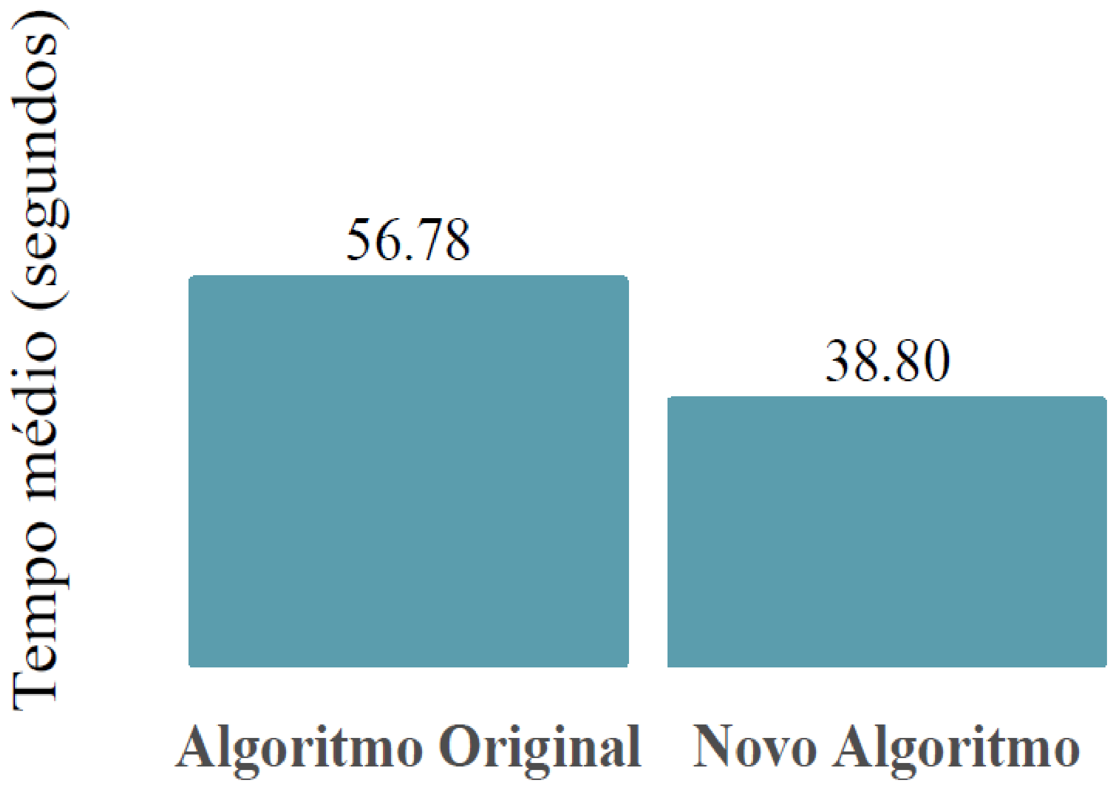
# Motivação

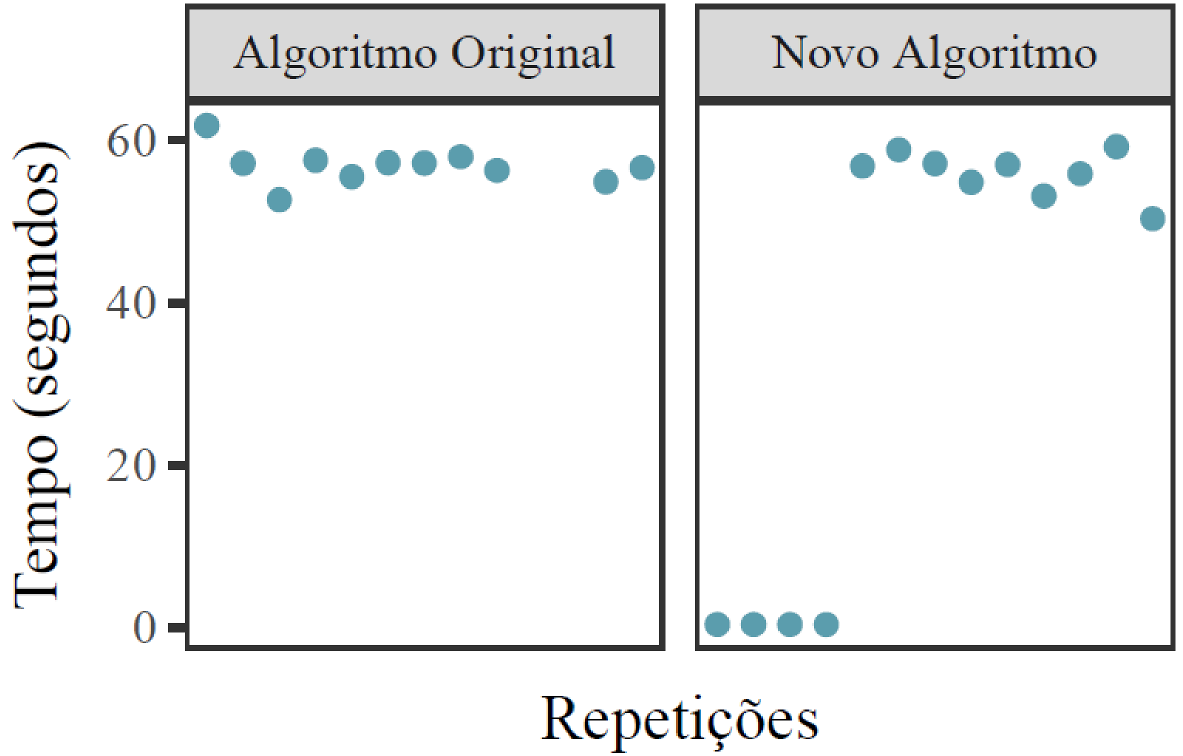
- **Resultados**



# Motivação

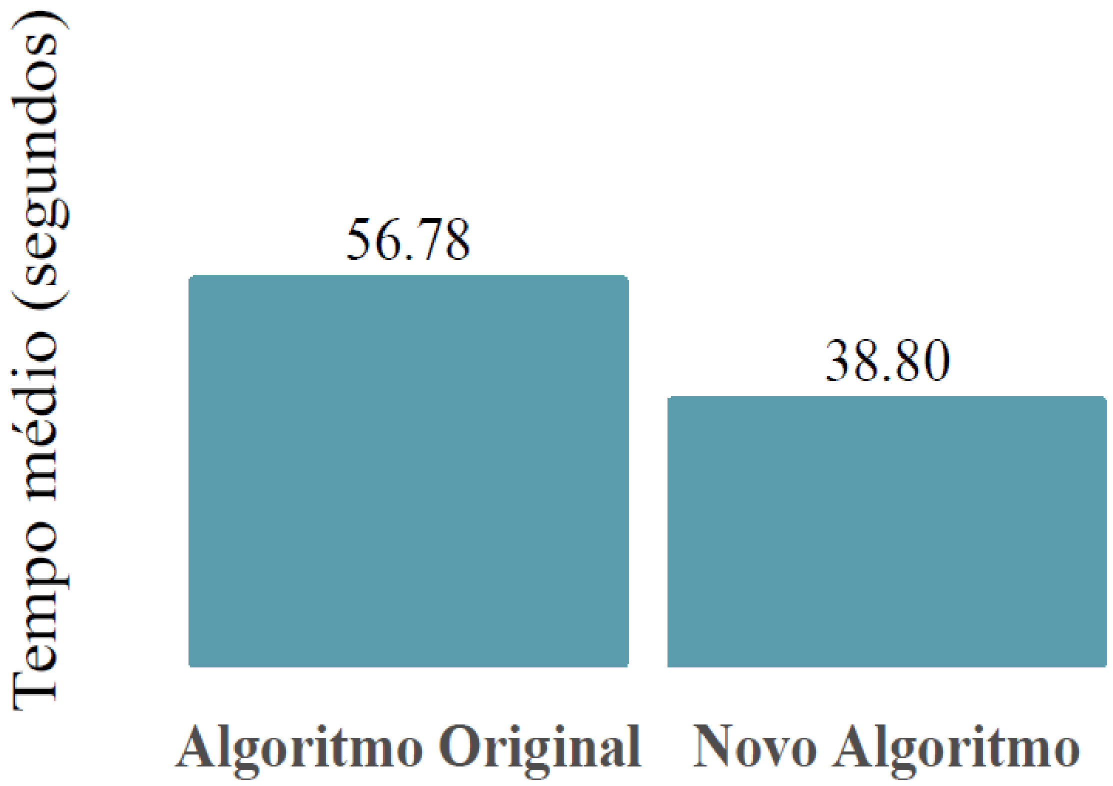
- **Resultados**

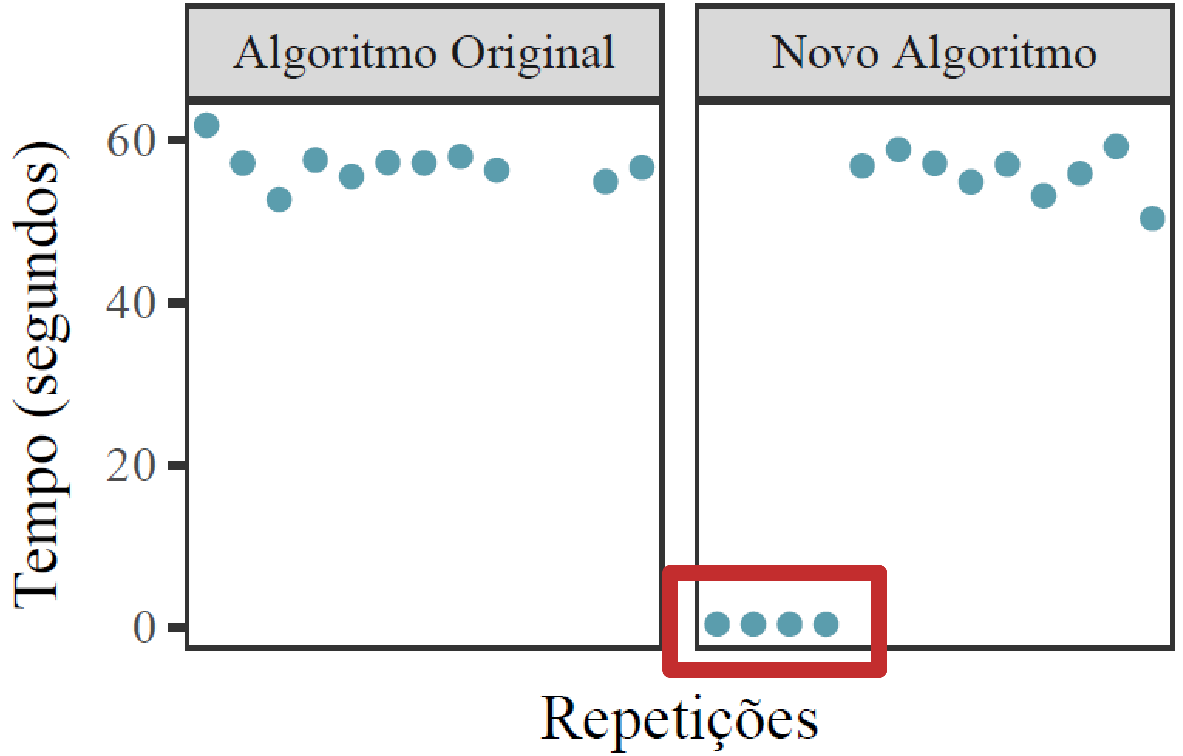


# Motivação

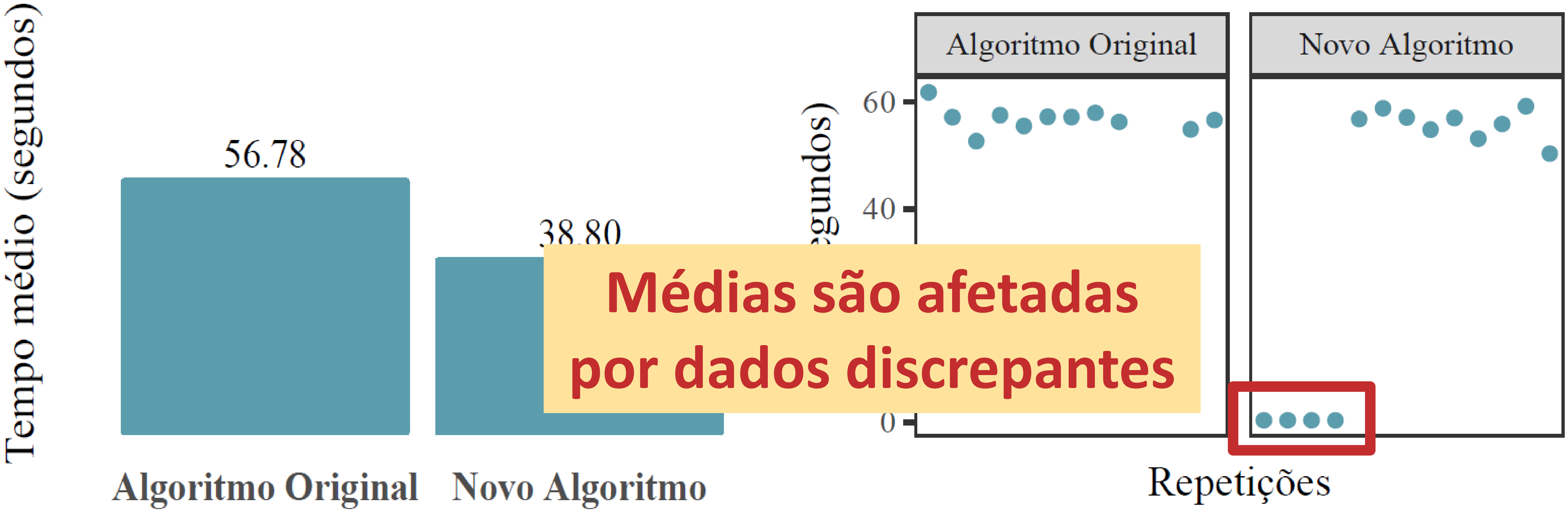
- **Resultados**



# Motivação

- **Resultados**




**Médias são afetadas por dados discrepantes**

# Motivação

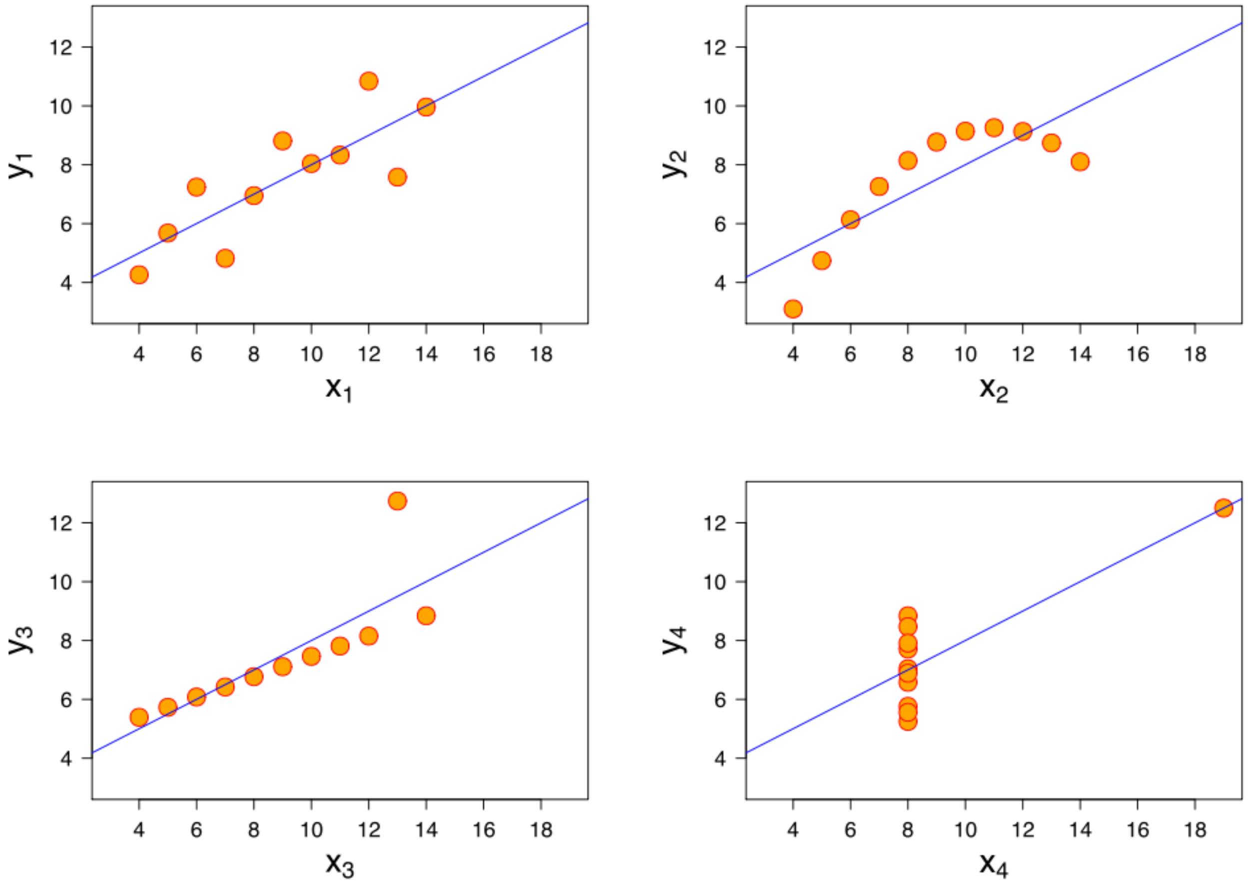
- **Comentário de [outro] avaliador de artigo**

  *“ [...] Outrossim, usar a média de 10 execuções é uma metodologia experimental bastante comum, e não parece óbvio para mim em qual sentido uma das metodologia mais sofisticadas sugeridas aumentaria a compreensão neste assunto em particular. ”*

  *Avaliador anônimo, evento internacional em 2017*

# Motivação

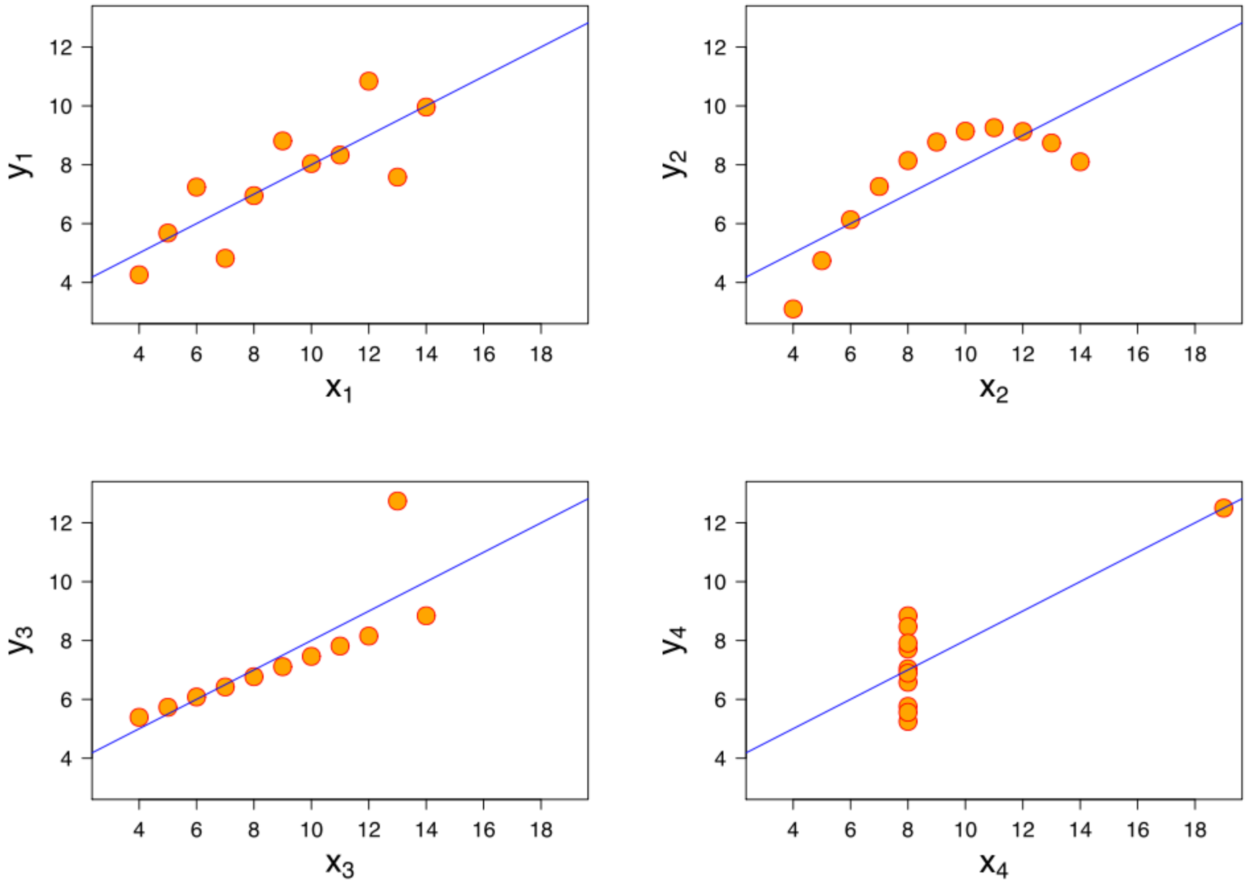
- **Quarteto de Anscombe**
  - –4 conjuntos de dados



# Motivação

- **Quarteto de Anscombe**
  - 4 conjuntos de dados

| Propriedade | Valor |
|---|---|
| Média de x | 9 |
| Variância de x | 11 |
| Média de y | 7,50 |
| Variância de y | 4,125 |
| Correlação | 0,816 |
| Regressão linear | y = 3,0 + 0,50x |




By Anscombe.svg: SchutzDerivative works of this file:(label using subscripts): Avenue - Anscombe.svg, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=9838454

# Motivação

- **Exemplos de "heurísticas" de avaliação de desempenho**
  - Descartar os *n* melhores e piores resultados
  - Selecionar apenas os *n* melhores resultados
  - Fazer *n* combinações de parâmetros e experimentos e comparar média a média

# Motivação

- **Exemplos de “heurísticas” de avaliação de desempenho**
  - Descartar os *n* melhores e piores resultados
  - Selecionar apenas os *n* melhores resultados
  - Fazer *n* combinações de parâmetros e experimentos e comparar média a média

Todas nós somos pessoas que já fizeram algo assim

# Motivação

- **Importância da análise [correta] de desempenho**
  - Resultados e conclusões mais provavelmente corretas
  - Reprodução dos resultados
  - Avanço da ciência e tecnologia

# ESTATÍSTICA E O MINICURSO

# Estatística e o Minicurso

- **Hipótese de pesquisa e experimentos**

Se a hipótese de pesquisa é

*“ o uso da estrutura de dados heap melhora o desempenho de operações sobre uma fila de prioridade ”*

então, com experimentos, vamos

*“ avaliar diferentes estruturas de dados para implementação de uma fila de prioridade ”*

e não

*“ provar que heap é a melhor estrutura para fila de prioridade ”*

# Estatística e o Minicurso

- **Cuidados com experimentos**
  - Coletar medidas que façam sentido
  - Documentar o ambiente experimental
  - Documentar a metodologia experimental
  - Lidar com variabilidade
    - DVFS, caches, interferência do SO
    - Ruídos e imprecisão nas medidas, interferência das medidas

# Estatística e o Minicurso

- **Fluxo de experimento (após planejamento)**

**Repetições do experimento**

**Maior número de resultados**

**Conclusões consistentes**

# Estatística e o Minicurso

- **Fluxo de experimento (após planejamento)**

**Repetições do experimento**

↓

**Maior número de resultados**

↓

**Maior probabilidade de obter conclusões consistentes usando técnicas apropriadas**

# Estatística e o Minicurso

- **Técnicas disponíveis**

**Descritivas**

Gráficos
Tabelas
Medidas Resumo
Análise de Cluster
Análise Fatorial
...

# Estatística e o Minicurso

- **Técnicas disponíveis**

**Descritivas**

Gráficos
Tabelas
Medidas Resumo
Análise de Cluster
Análise Fatorial
...

**Inferenciais**

Teste de hipóteses
Análise de Regressão
Séries Temporais
Métodos Causais
Análise de Experimentos
...

# Estatística e o Minicurso

- **Técnicas disponíveis**

**Descritivas**

Gráficos
Tabelas
Medidas Resumo
Análise de Cluster
Análise Fatorial
...

**Inferenciais**

Teste de hipóteses
Análise de Regressão
Séries Temporais
Métodos Causais
Análise de Experimentos
...

**Abordagem Bayesiana x Clássica**

**Métodos Paramétricos e Não-paramétricos**

# Estatística e o Minicurso

- **Técnicas disponíveis**

**Descritivas**

**Gráficos**
**Tabelas**
**Medidas Resumo**
Análise de Cluster
Análise Fatorial
...

**Inferenciais**

**Teste de hipóteses**
Análise de Regressão
Séries Temporais
Métodos Causais
Análise de Experimentos
...

**Abordagem Bayesiana x Clássica**

**Métodos Paramétricos e Não-paramétricos**

# Estatística e o Minicurso

- **O que o minicurso provê**
  - Mecanismos e ideias para avaliação de desempenho
  - Capítulo de livro com maior detalhamento
  - Repositório com exemplos de códigos em R
    - **https://github.com/llpilla/estatistica2017erad**

# MÉTODOS DESCRITIVOS

# Métodos Descritivos

Coletei os dados,
vamos testar as
hipóteses?

# Métodos Descritivos

- **Objetivos de métodos descritivos**
  - “Conhecer” a natureza dos dados
  - Detectar anomalias
  - Encontrar padrões

# Métodos Descritivos

- **Análise univariada**

**Representação**

Histograma
Boxplot
Tabela de Frequências
Gráfico de Pizza
Gráfico de Barras

**Medidas**

Média
Mediana
Desvio Padrão
Variância
Coeficiente de Variação
Moda
Quartis

# Métodos Descritivos

- **Análise univariada – Variáveis Quantitativas**

**Representação**

Histograma
Boxplot
Tabela de Frequências
Gráfico de Pizza
Gráfico de Barras

**Medidas**

Média
Mediana
Desvio Padrão
Variância
Coeficiente de Variação
Moda
Quartis

# Métodos Descritivos

- **Análise univariada – Variáveis Qualitativas**

**Representação**

Histograma

Boxplot

Tabela de Frequências

Gráfico de Pizza

Gráfico de Barras

**Medidas**

Média

Moda

Mediana

Quartis

Desvio Padrão

Variância

Coeficiente de Variação

# Métodos Descritivos

- **Análise univariada**
  - Exemplo de 5 medições para dois algoritmos A e B

| $t_A$ (u.t.) | $t_B$ (u.t.) |
|---|---|
| 10 | 30 |
| 12 | 30 |
| 13 | 30 |
| 15 | 30 |
| 100 | 30 |

# Métodos Descritivos

- **Análise univariada**
  - Exemplo de 5 medições para dois algoritmos A e B

| $t_A$ (u.t.) | $t_B$ (u.t.) |
|---|---|
| 10 | 30 |
| 12 | 30 |
| 13 | 30 |
| 15 | 30 |
| 100 | 30 |

$$\overline{t_A} = \overline{t_B} = 30$$

# Métodos Descritivos

- **Análise univariada**
  - Exemplo de 5 medições para dois algoritmos A e B

| $t_A$ (u.t.) | $t_B$ (u.t.) |
|---|---|
| 10 | 30 |
| 12 | 30 |
| 13 | 30 |
| 15 | 30 |
| 100 | 30 |

**50% dos tempos de execução observados do algoritmo A foram de até 13u.t.**

$$Med(t_A) = 13 < 30 = Med(t_B)$$

# Métodos Descritivos

- **Análise univariada**
  - Exemplo de 5 medições para dois algoritmos A e B

| $t_A$ (u.t.) | $t_B$ (u.t.) |
|---|---|
| 10 | 30 |
| 12 | 30 |
| 13 | 30 |
| 15 | 30 |
| 100 | 30 |

$$\boldsymbol{CV(A) = \frac{39,17}{30} \approx 1,31}$$

$$\boldsymbol{s_A = 39,17 \; e \; s_B = 0}$$

# Métodos Descritivos

**Média**

*"Centro de gravidade" dos dados, valor médio*

**Moda**

*Valor mais frequente*

**Mediana**

*Valor que divide os dados pela metade*

# Métodos Descritivos

**Média**

*"Centro de gravidade" dos dados, valor médio*

**Quartis**

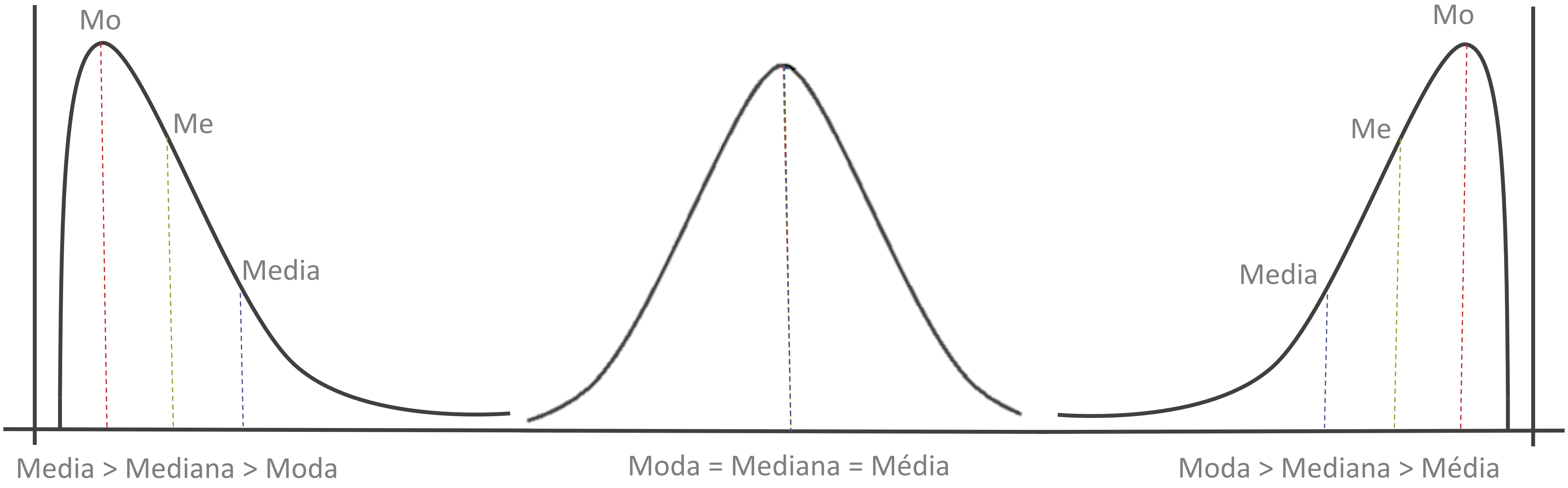
O 1º e o 3º quartil dividem a distribuição em 25% menores e 25% maiores

**Mediana**

*Valor que divide os dados pela metade*

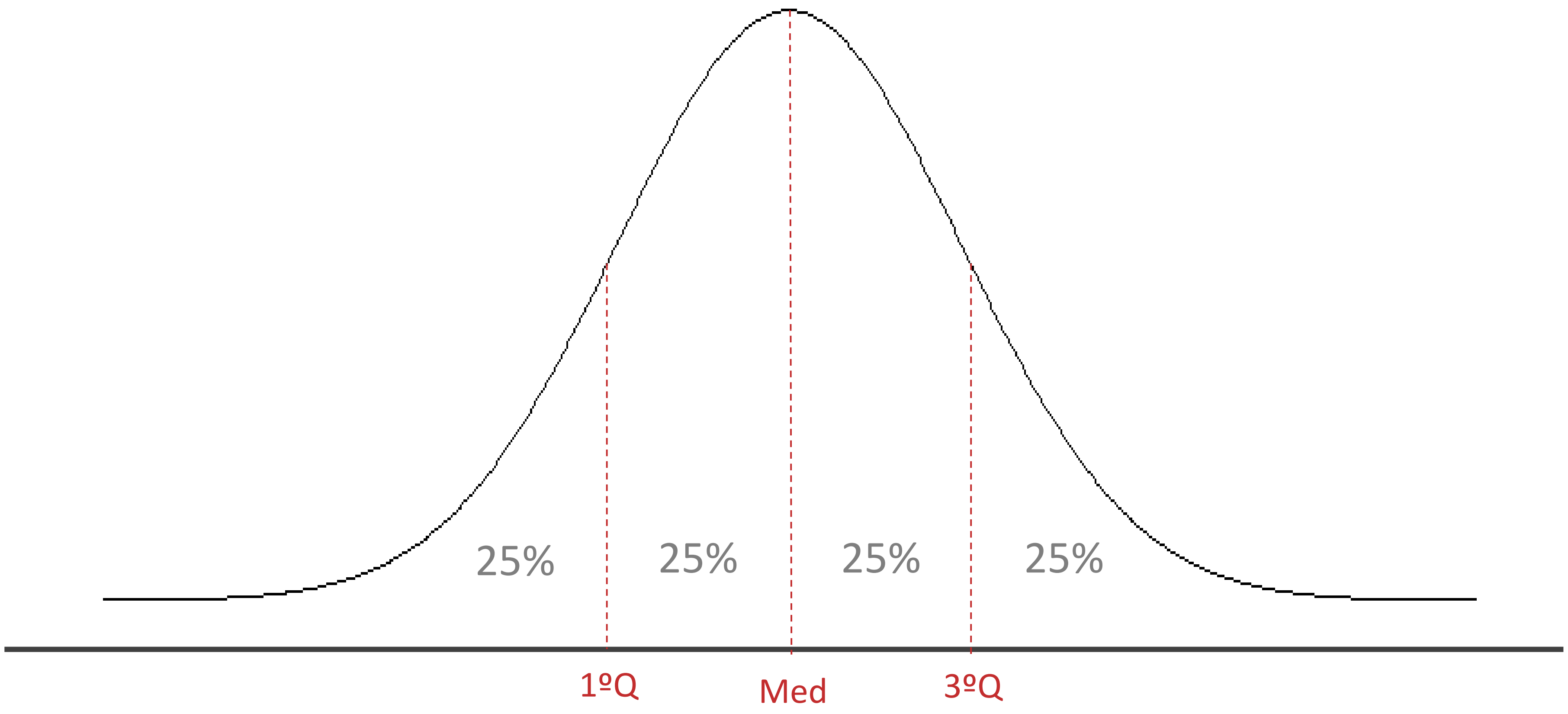
# Métodos Descritivos

- **Média, moda e mediana**



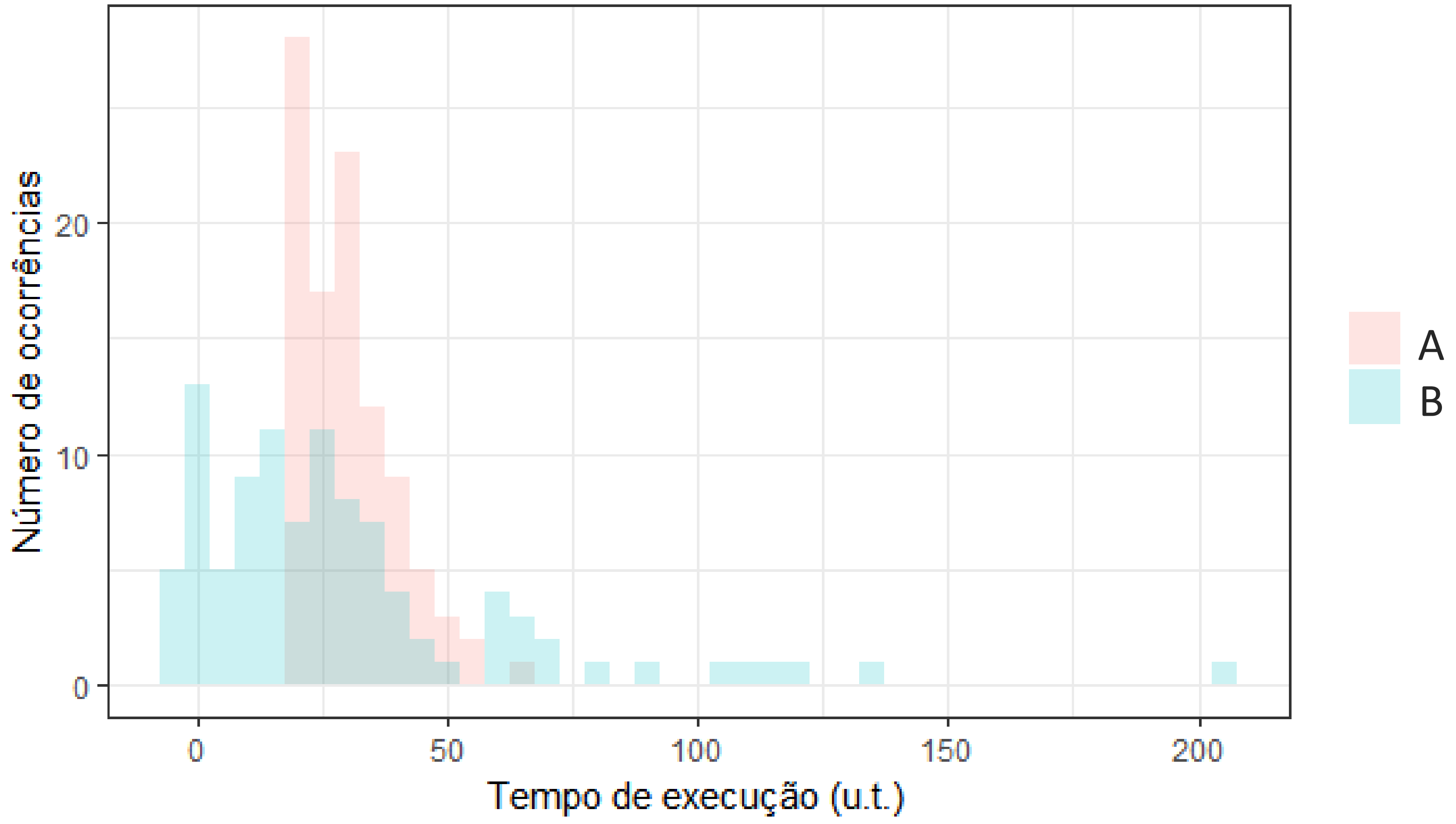
# Métodos Descritivos

- **Mediana e quartis**



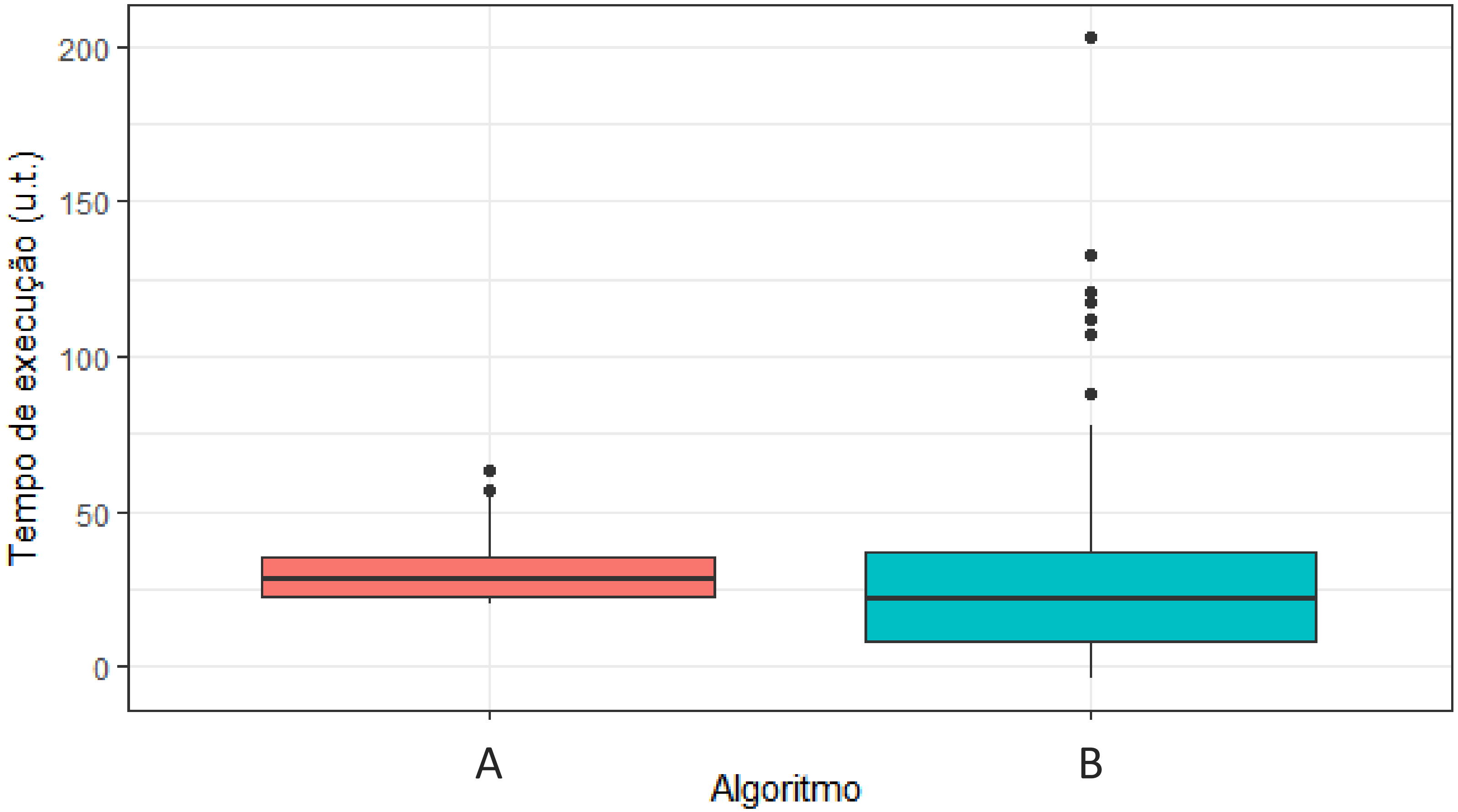
# Métodos Descritivos

- **Exemplo: Histograma**



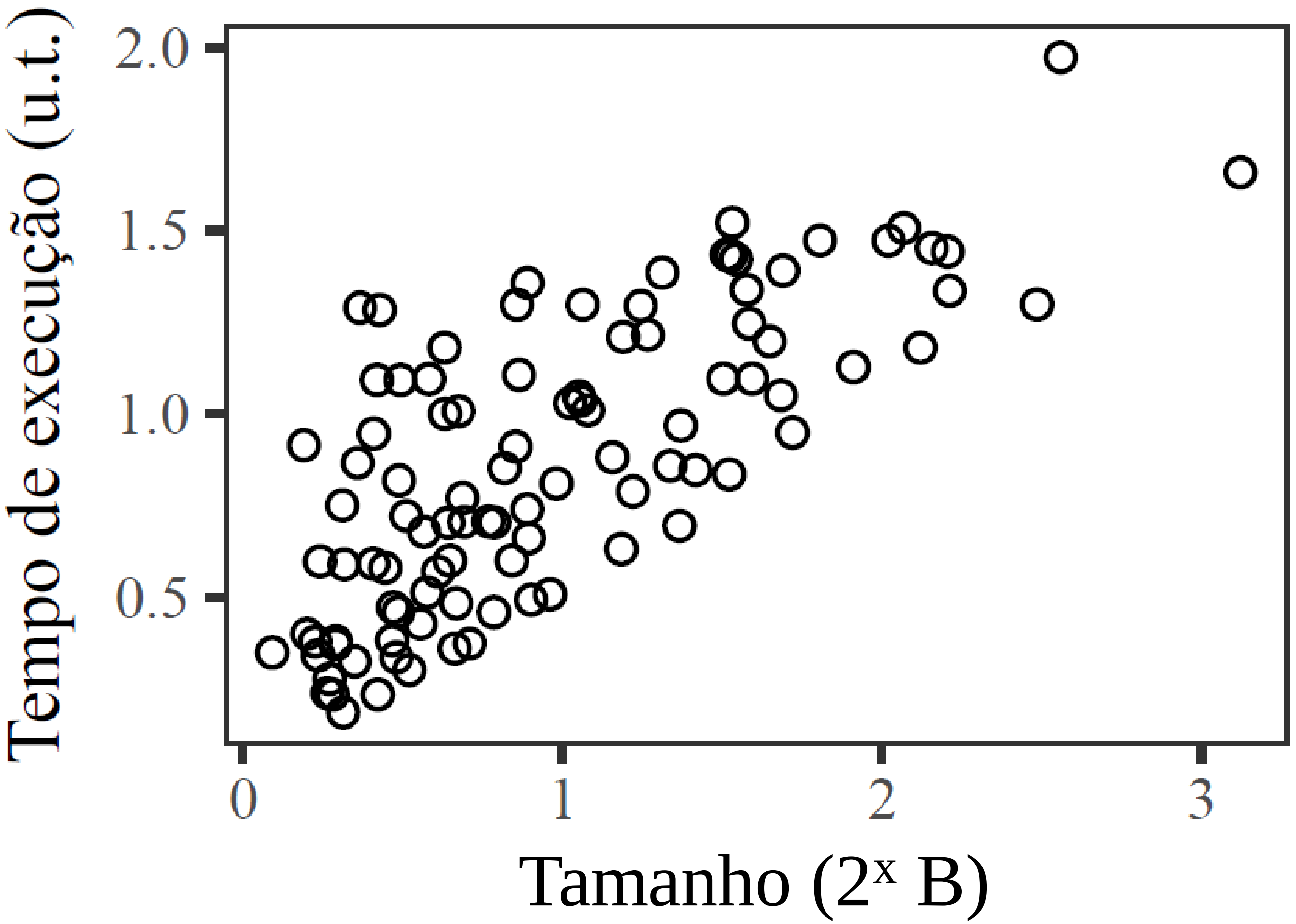
# Métodos Descritivos

- **Exemplo: Boxplot**



# Métodos Descritivos

- **Gráfico de dispersão**



# Métodos Descritivos

*Para dados qualitativos:*

- *Tabelas de frequências*
- *Tabelas cruzadas*
- *Gráficos de Barras*
- *Gráficos de Pizzas*

# Métodos Descritivos

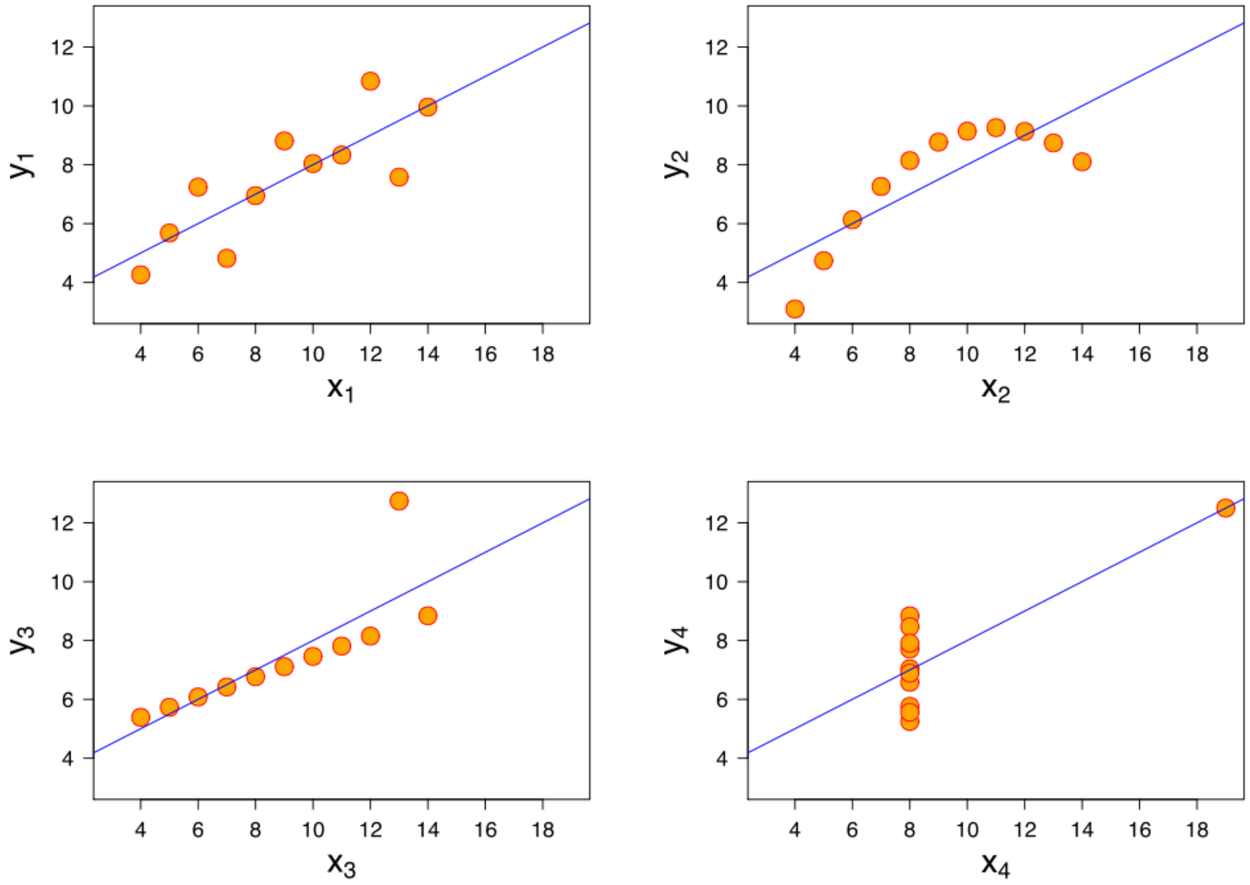
- **Correlação**

# Métodos Descritivos

- **Correlação**
  - Correlação de Pearson
    - Grau de associação linear entre duas variáveis continuas
  - Correlação de Spearman
    - Grau de associação entre duas variáveis continuas

# Métodos Descritivos

| Correlação (Pearson) | 0,816 |
|---|---|




By Anscombe.svg: SchutzDerivative works of this file:(label using subscripts): Avenue - Anscombe.svg, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=9838454

# INFERÊNCIA ESTATÍSTICA

# Inferência Estatística

**O problema fundamental da inferência estatística é inferir, através das estimativas obtidas com base em uma amostra, características dos parâmetros populacionais.**

# Inferência Estatística

**População-alvo**

**Parâmetros:** $\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{p}, \boldsymbol{\beta}$

# Inferência Estatística

**População-alvo**

A única forma de ter as informações exatas dos parâmetros é através de um CENSO.

Processo caro, trabalhoso, difícil.

**Parâmetros:** $\boldsymbol{\mu, \sigma^2, p, \beta}$

# Inferência Estatística

## População-alvo

**Amostra**

**Estimadores:**

$$\bar{X} = \frac{\sum X_i}{n}, s^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}, \hat{p}, \hat{\beta}$$

**Parâmetros:** $\mu, \sigma^2, p, \beta$

# Inferência Estatística

## População-alvo

### Amostra

**Quando "acoplamos" os dados coletados nas equações dos estimadores, obtemos as estimativas.**

### Estimadores:

$$\bar{X} = \frac{\sum X_i}{n}, s^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}, \hat{p}, \hat{\beta}$$

**Parâmetros:** $\mu, \sigma^2, p, \beta$

## Na inferência clássica

**Parâmetro:** constante fixa e desconhecida

# Inferência Estatística

## Na inferência clássica

**Parâmetro:** constante fixa e desconhecida

**Amostra:** os dados vem de uma distribuição populacional com parâmetros desconhecidos

# Inferência Estatística

## Na inferência clássica

**Parâmetro:** constante fixa e desconhecida

**Amostra:** os dados vem de uma distribuição populacional com parâmetros desconhecidos

**Estimadores:** estatísticas

(funções da amostra que não dependem de parâmetros)

## Na inferência clássica

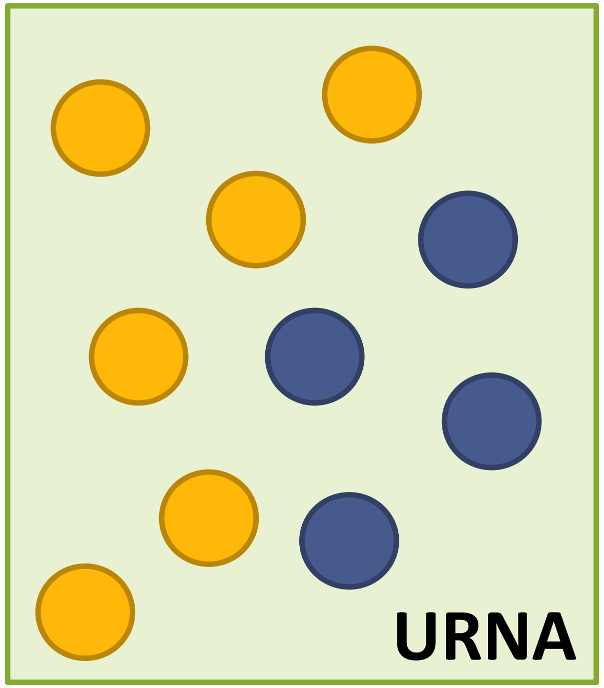
Calculamos as probabilidades de observamos os dados supondo uma distribuição populacional.

## Na inferência clássica

Calculamos as probabilidades de observamos os dados supondo uma distribuição populacional.

Buscamos valores de parâmetros que maximizem a chance dos dados terem vindo de determinada distribuição.

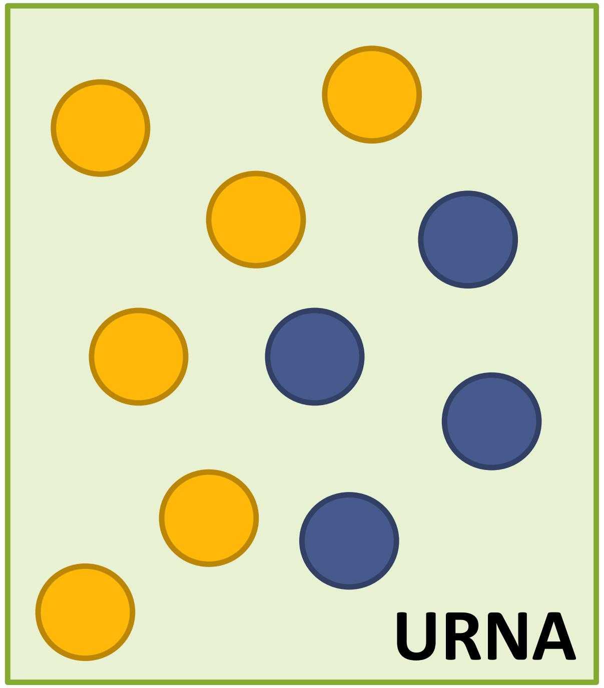
- **Independência**




1. Qual a probabilidade de sortear uma bola amarela?

# Inferência Estatística

- **Independência**

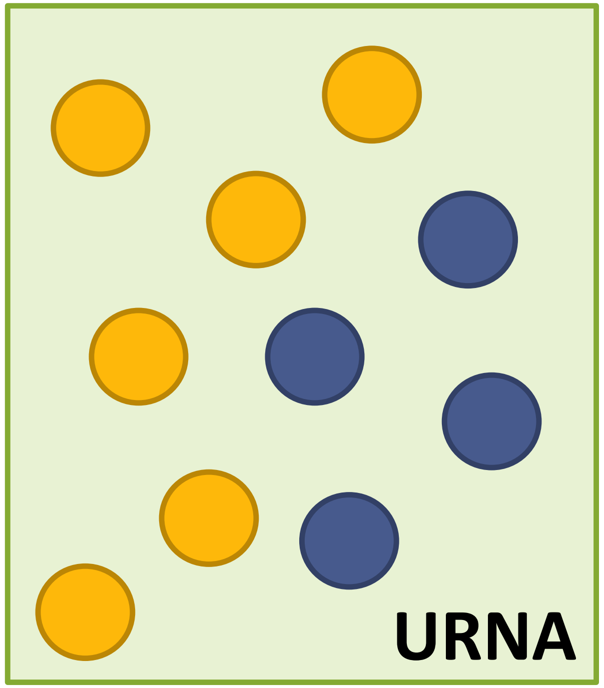



1. Qual a probabilidade de sortear uma bola amarela?

**R: 6/10**

# Inferência Estatística

- **Independência**

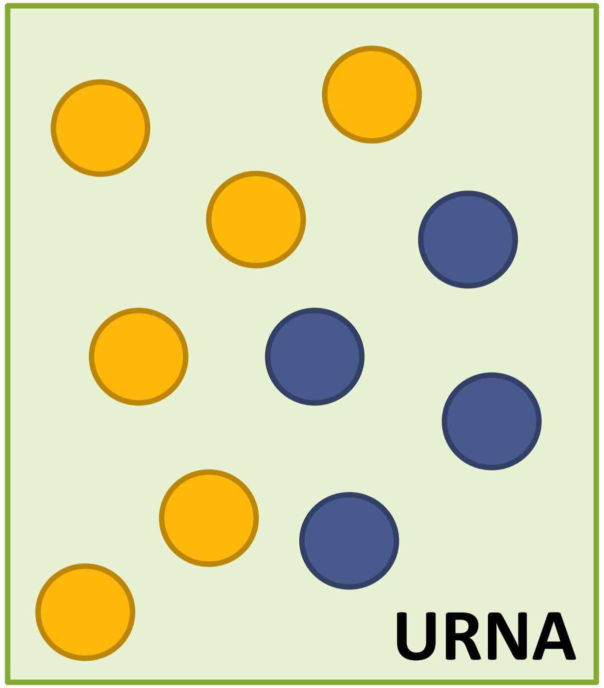



1. Qual a probabilidade de sortear uma bola amarela?

   **R: 6/10**

2. Qual a probabilidade de sortear uma bola amarela e uma azul?

- **Independência**



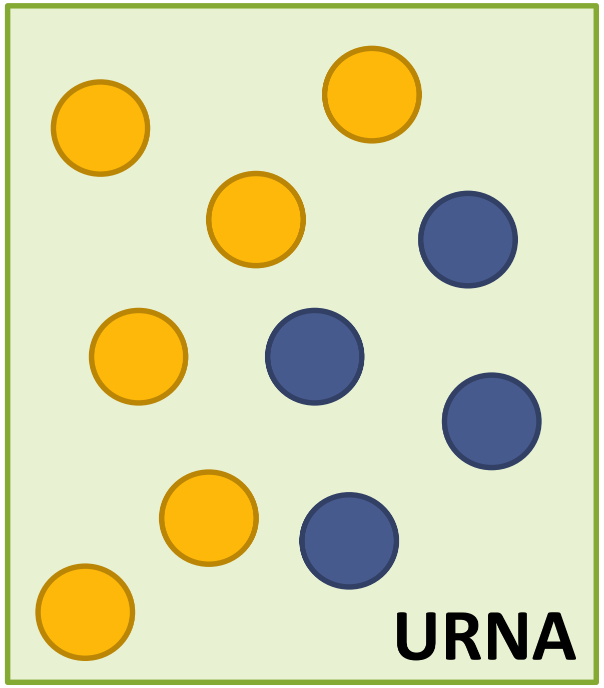

1. Qual a probabilidade de sortear uma bola amarela?

   **R: 6/10**

2. Qual a probabilidade de sortear uma bola amarela e uma azul?

# Inferência Estatística

- **Independência**




1. Qual a probabilidade de sortear uma bola amarela?

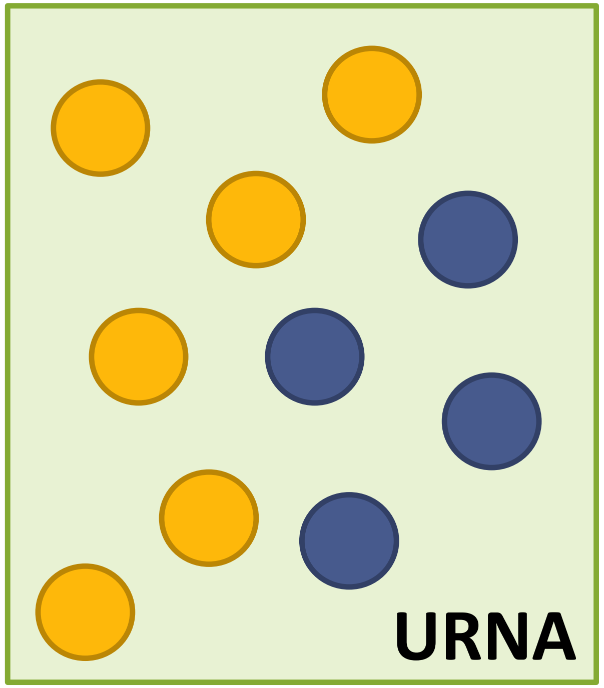
   **R: 6/10**

2. Qual a probabilidade de sortear uma bola amarela e uma azul?
3. Qual a probabilidade de sortear uma bola amarela e uma azul com reposição?

   **R: 6/10*4/10**

# Inferência Estatística

- **Independência**




1. Qual a probabilidade de sortear uma bola amarela?

   **R: 6/10**

2. Qual a probabilidade de sortear uma bola amarela e uma azul?
3. Qual a probabilidade de sortear uma bola amarela e uma azul com reposição?

   **R: 6/10*4/10**

4. Qual a probabilidade de sortear uma bola amarela e uma azul sem reposição?

   **R: 6/10*4/9**

**Os métodos que serão apresentados pressupõe que as observações são independentes entre si.**
**Sem ela, não há nenhuma garantia sobre os resultados.**

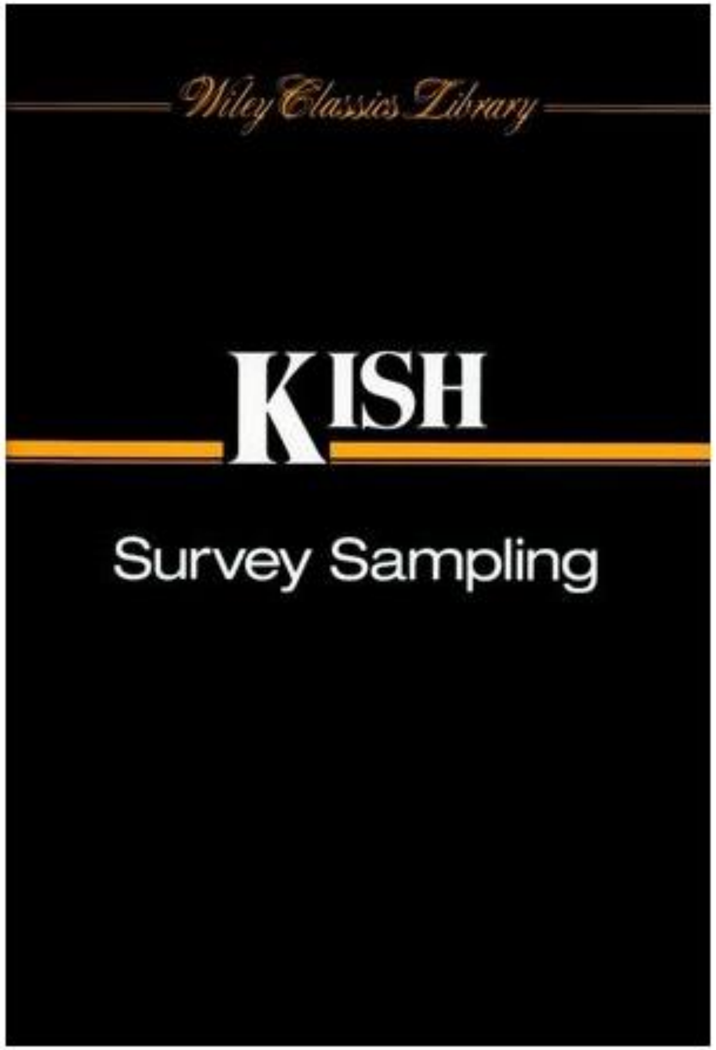
# Inferência Estatística

“Todas as generalizações são perigosas, incluindo esta.”

Alexandre Dumas (1824-1895)

# Inferência Estatística

**Generalizar resultados requer uma amostra representativa da população.**



# TESTES DE HIPÓTESES

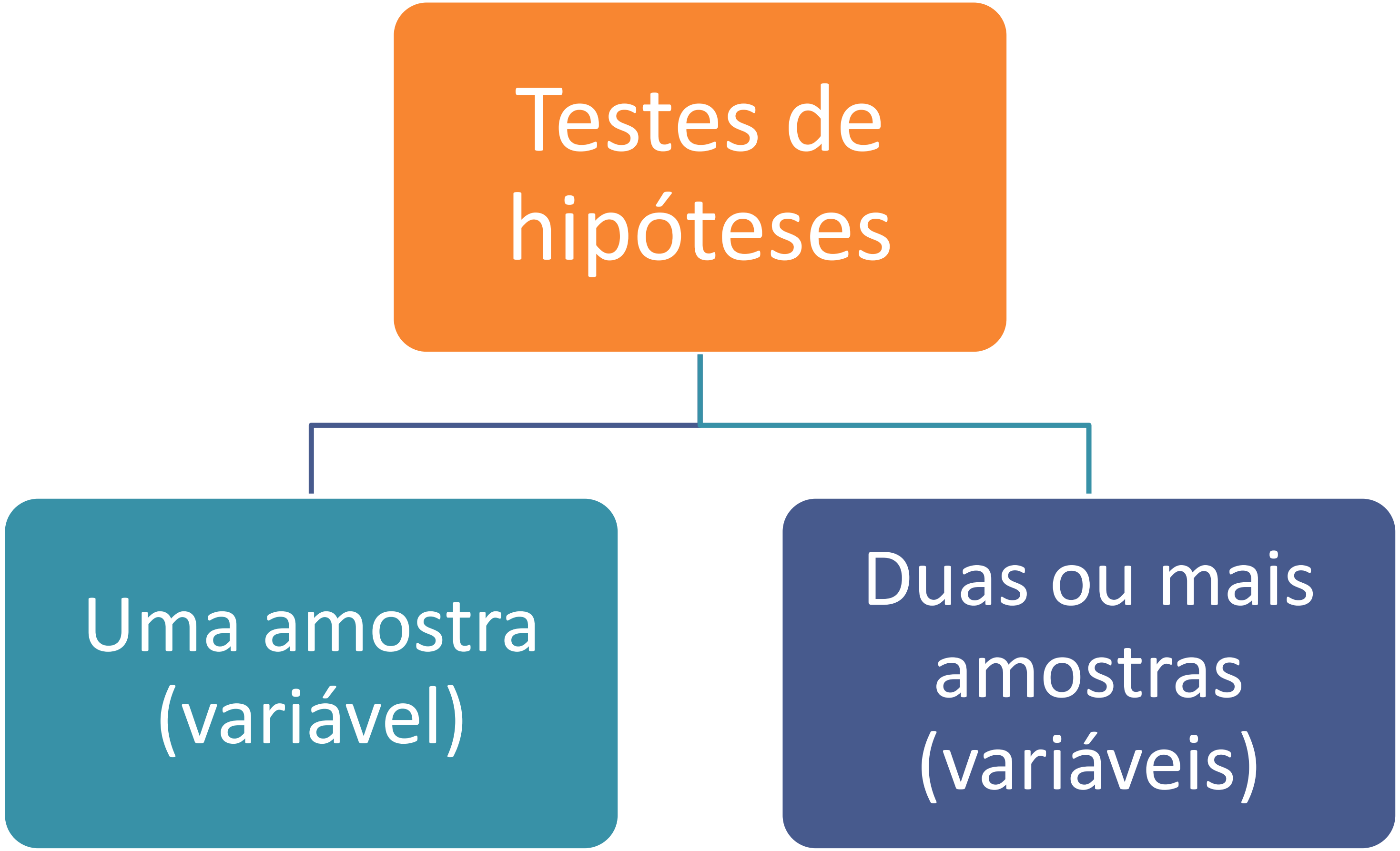
# Testes de Hipóteses

- Inferência estatística
  - Testes de hipóteses
  - Estimação de parâmetros

# Testes de Hipóteses

- Inferência estatística
  - Testes de hipóteses
  - Estimação de parâmetros

# Testes de Hipóteses



# Testes de Hipóteses

- **Testes com uma amostra**
  - Exemplos
    - Estes dados parecem ter sido gerados por um processo com distribuição $Normal(\mu,\sigma^2)$?

# Testes de Hipóteses

- **Testes com uma amostra**
  - Exemplos
    - Estes dados parecem ter sido gerados por um processo com distribuição $Normal(\mu,\sigma^2)$?
    - O algoritmo tem uma média de 5 erros por execução?

# Testes de Hipóteses

- **Testes com uma amostra**
  - Exemplos
    - Estes dados parecem ter sido gerados por um processo com distribuição $Normal(\mu,\sigma^2)$?
    - O algoritmo tem uma média de 5 erros por execução?
    - A proporção de erros de execução por pacote é inferior a 20%?

# Testes de Hipóteses

- **Testes com uma amostra**
  - Pontos importantes
    - Verificar distribuições

# Testes de Hipóteses

- **Testes com uma amostra**
  - Pontos importantes
    - Verificar distribuições
    - Verificar se os dados parecem ter vindo de uma distribuição com uma média, proporção ou variância específica, etc.

# Testes de Hipóteses

- **Testes com duas ou mais amostras**
  - Exemplos
    - O algoritmo $A$ tem tempo de execução menor que o algoritmo $B$?

# Testes de Hipóteses

- **Testes com duas ou mais amostras**
  - Exemplos
    - O algoritmo $A$ tem tempo de execução menor que o algoritmo $B$?
    - A taxa de falhas de cache exibida nos procedimentos do tipo $M$ é maior do que nos procedimentos do tipo $N$?

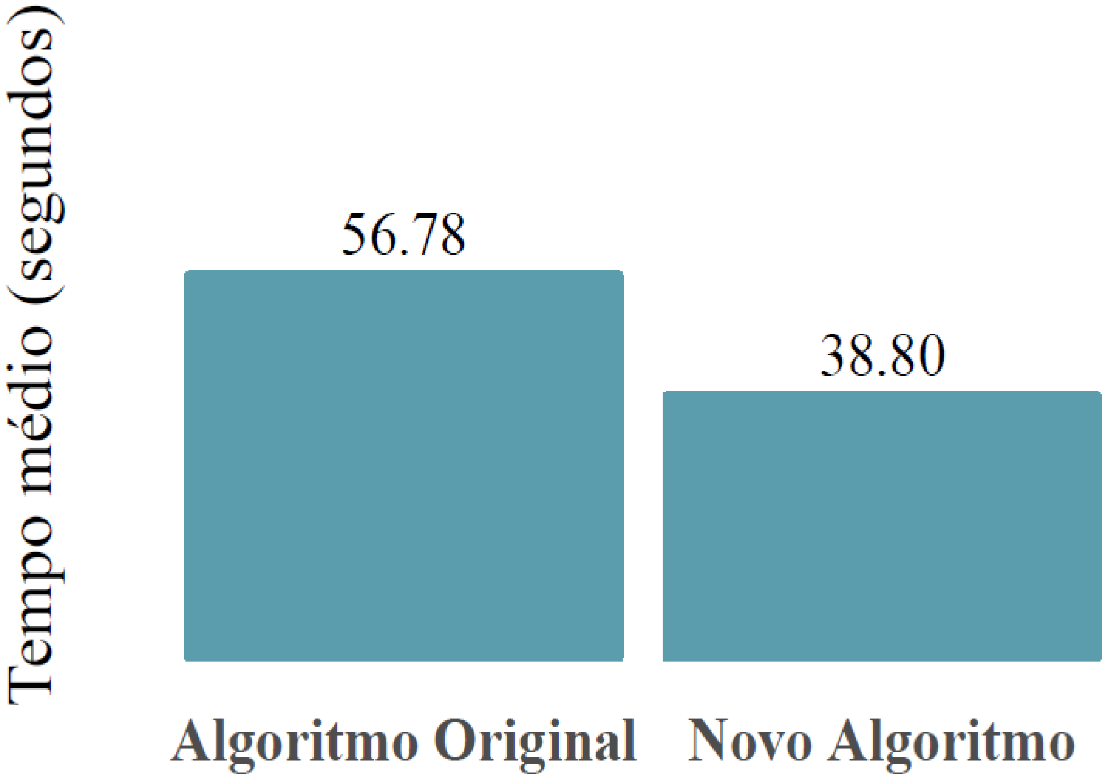
# Testes de Hipóteses

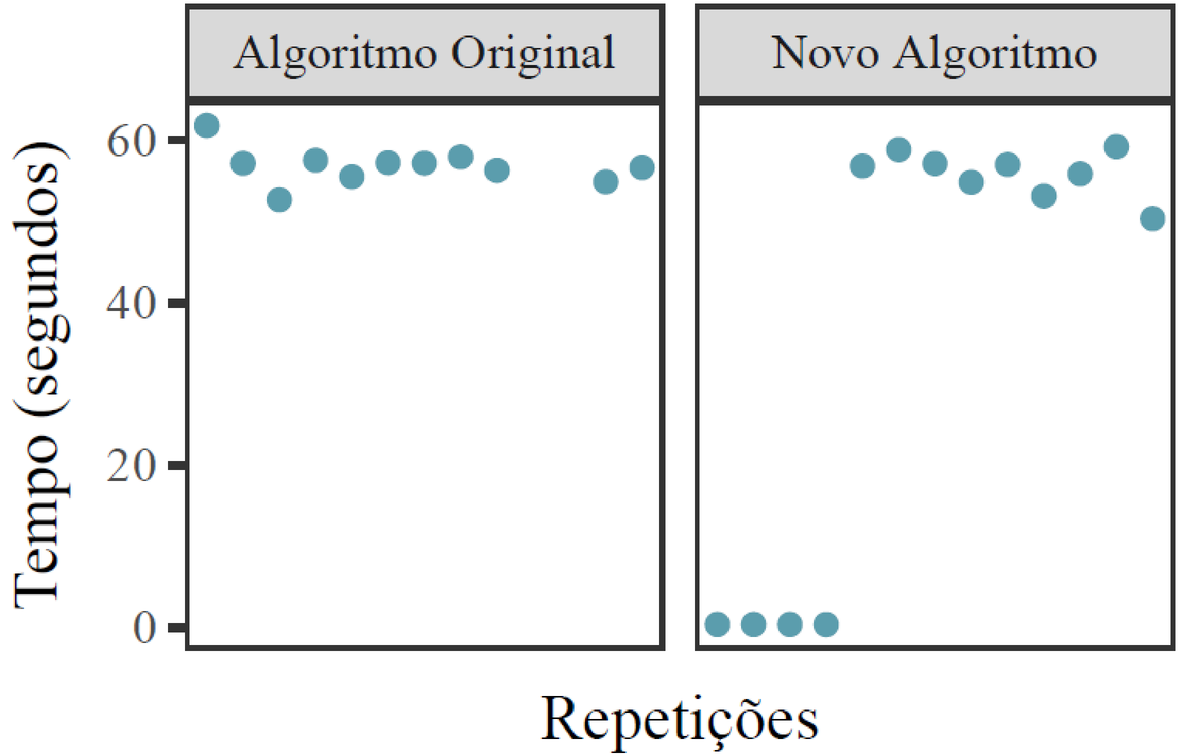
- **Testes com duas ou mais amostras**
  - Pontos importantes
    - Comparações dos dados para verificar se os dados coletados parecem ser provenientes de uma mesma distribuição (com mesmos parâmetros)

**Por que não comparamos diretamente as médias?**

# Testes de Hipóteses

- **Resultados apresentados na Motivação**



# Testes de Hipóteses

- **Questões que queremos responder**
  - Estes dados parecem ter sido gerados por um processo com distribuição $Normal(\mu, \sigma^2)$?

# Testes de Hipóteses

- **Questões que queremos responder**
  - Estes dados parecem ter sido gerados por um processo com distribuição $Normal(\mu,\sigma^2)$?
  - O algoritmo tem uma média de 5 erros por execução?

# Testes de Hipóteses

- **Questões que queremos responder**
  - Estes dados parecem ter sido gerados por um processo com distribuição $Normal(\mu,\sigma^2)$?
  - O algoritmo tem uma média de 5 erros por execução?
  - A proporção de erros de execução por pacote é inferior a 20%?

# Testes de Hipóteses

- **Questões que queremos responder**
  - Estes dados parecem ter sido gerados por um processo com distribuição $Normal(\mu,\sigma^2)$?
  - O algoritmo tem uma média de 5 erros por execução?
  - A proporção de erros de execução por pacote é inferior a 20%?
  - O algoritmo $A$ tem tempo de execução menor que o algoritmo $B$?

# Hipóteses estatísticas

## $H_0$: Hipótese nula

## $H_1$: Hipótese alternativa

# Testes de Hipóteses

- **Questões que queremos responder**
  - Estes dados parecem ter sido gerados por um processo com distribuição $Normal(\mu,\sigma^2)$?

# Testes de Hipóteses

- **Questões que queremos responder**
  - Estes dados parecem ter sido gerados por um processo com distribuição $Normal(\mu,\sigma^2)$?
    - $H_0$: Os dados aderem a uma $Normal(\mu, \sigma^2)$
    - $H_1$: Os dados não aderem a uma distribuição $Normal(\mu, \sigma^2)$

# Testes de Hipóteses

- **Questões que queremos responder**
  - O algoritmo tem uma média de 5 erros por execução?

# Testes de Hipóteses

- **Questões que queremos responder**
  - O algoritmo tem uma média de 5 erros por execução?
    - $H_0$: Os dados são provenientes de uma distribuição com média igual a 5 ($H_0$: $\mu = 5$ )
    - $H_1$: Os dados são provenientes de uma distribuição com média diferente de 5 ($H_1$: $\mu \neq 5$)

# Testes de Hipóteses

- **Questões que queremos responder**
  - A proporção de erros de execução por pacote é inferior a 20%?

# Testes de Hipóteses

- **Questões que queremos responder**
  - A proporção de erros de execução por pacote é inferior a 20%?
    - $H_0$: Os dados são provenientes de uma distribuição com proporção maior ou igual a 20% ($H_0: p \geq 0.20$)
    - $H_1$: Os dados são provenientes de uma distribuição com proporção inferior a 20% ($H_1: p < 0.20$)

# Testes de Hipóteses

- **Questões que queremos responder**
  - O algoritmo $A$ tem tempo de execução menor que o algoritmo $B$?

# Testes de Hipóteses

- **Questões que queremos responder**
  - O algoritmo $A$ tem tempo de execução menor que o algoritmo $B$?
    - $H_0$: A diferença nos tempos de execução entre algoritmo A e o algoritmo B é maior ou igual a 0 ($H_0$: $\mu_A \geq \mu_B \Rightarrow \mu_A - \mu_B \geq 0$ )
    - $H_1$: A diferença nos tempos de execução entre algoritmo A e o algoritmo B é menor que 0 ($H_1$: $\mu_A - \mu_B < 0$)

# Testes de Hipóteses

- **Hipóteses estatísticas e erros**

| Decisão do teste/ Realidade | Aceitar H0 | Rejeitar H0 |
|---|---|---|
| | | |
| | | |

# Testes de Hipóteses

- **Hipóteses estatísticas e erros**

| Decisão do teste/ Realidade | | |
|---|---|---|
| H0 é verdadeira | | |
| H0 é falsa | | |

# Testes de Hipóteses

- **Hipóteses estatísticas e erros**

| Decisão do teste/ Realidade | Aceitar H0 | Rejeitar H0 |
|---|---|---|
| H0 é verdadeira | Decisão correta | |
| H0 é falsa | | Decisão correta |

# Testes de Hipóteses

- **Hipóteses estatísticas e erros**

| Decisão do teste/ Realidade | Aceitar H0 | Rejeitar H0 |
|---|---|---|
| H0 é verdadeira | Decisão correta | **Erro tipo I** $\mathbb{P}(\textbf{erro tipo I}) = \boldsymbol{\alpha}$ |
| H0 é falsa | **Erro tipo II** $\mathbb{P}(\textbf{erro tipo II}) = \boldsymbol{\beta}$ | Decisão correta |

# Testes de Hipóteses

- **Mecanismos dos testes paramétricos**

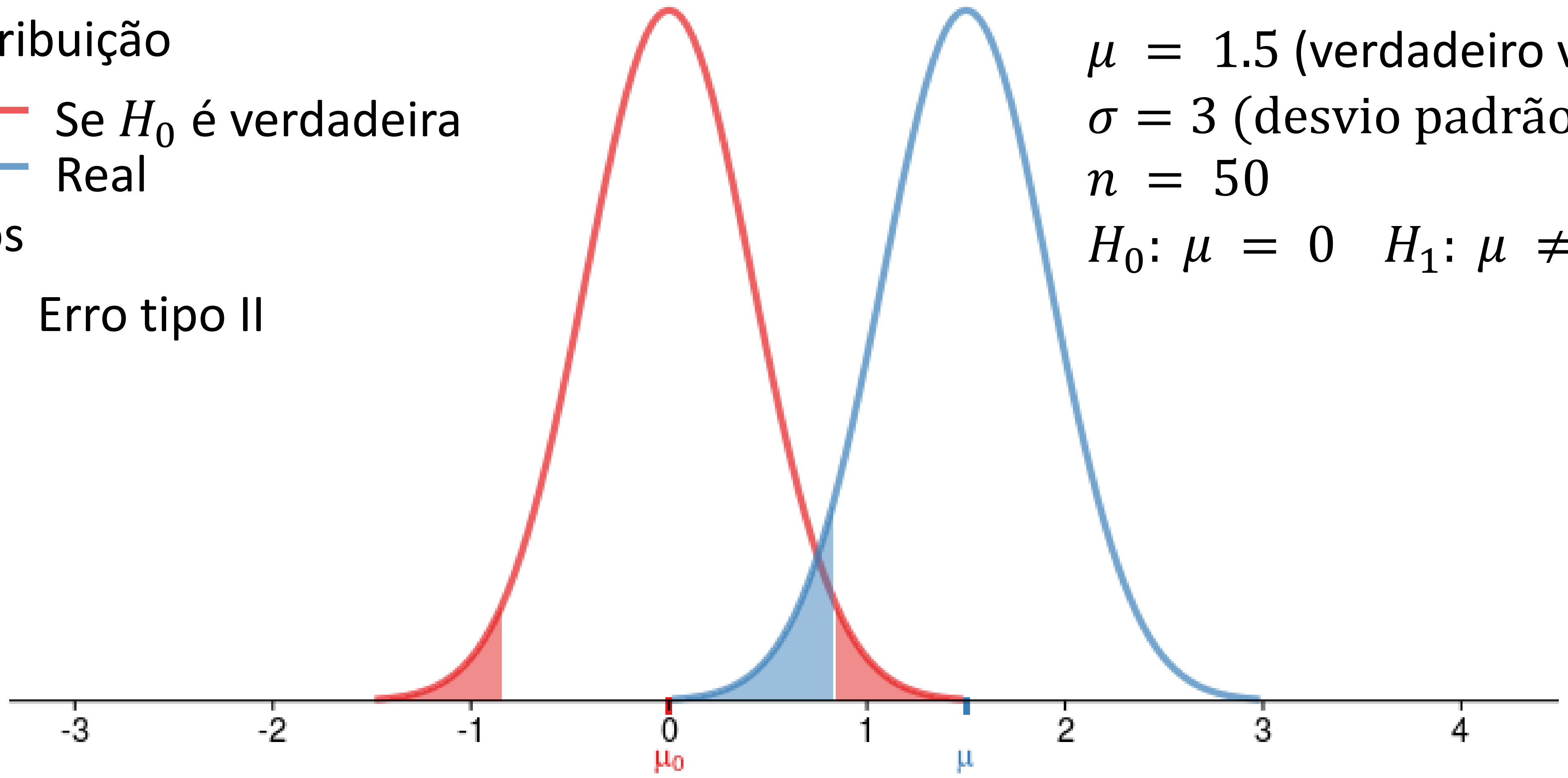
Distribuição

Se $H_0$ é verdadeira

Real

Erros

Erro tipo II



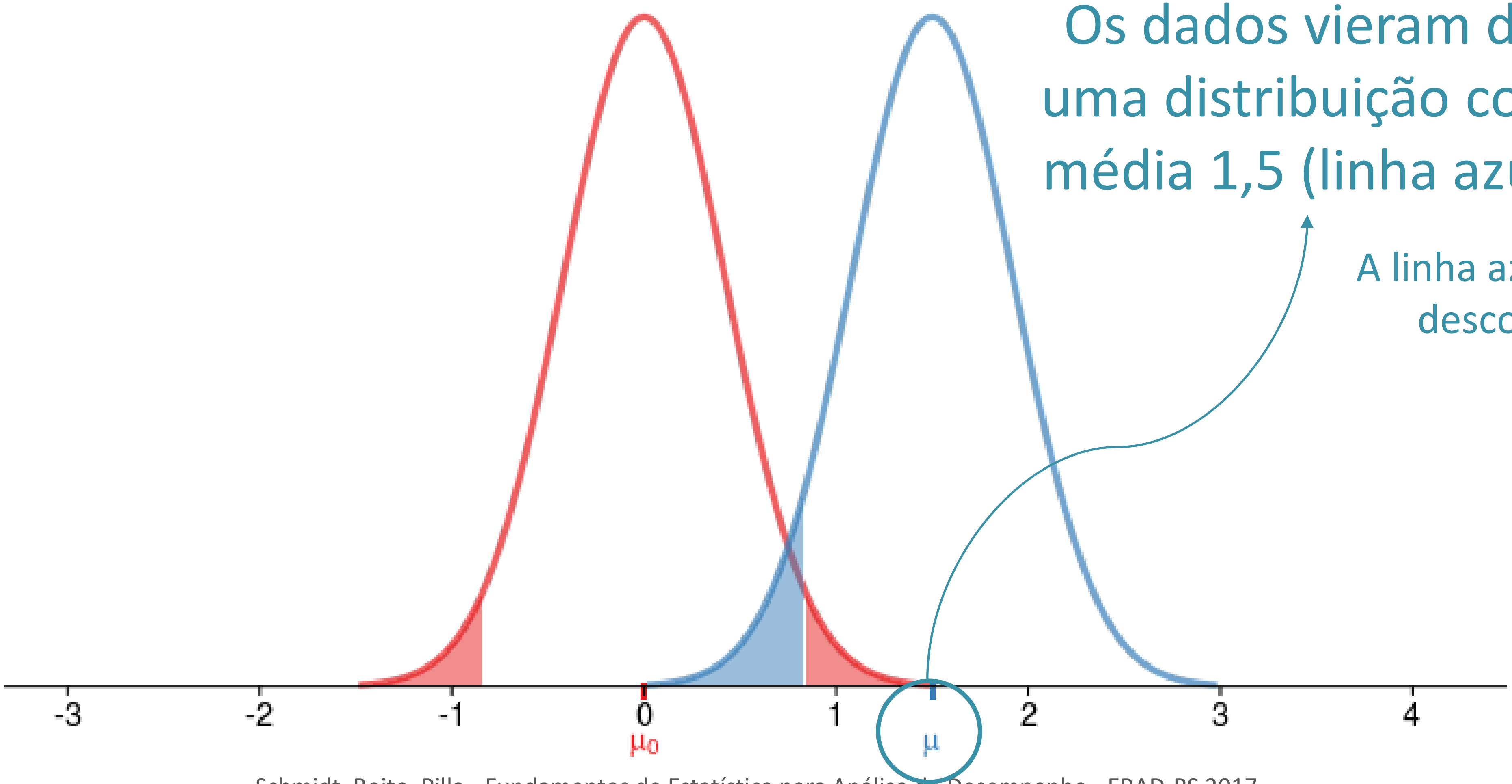

$\mu = 1.5$ (verdadeiro valor)

$\sigma = 3$ (desvio padrão)

$n = 50$

$H_0$: $\mu = 0$ $H_1$: $\mu \neq 0$

- **Mecanismos dos testes paramétricos**

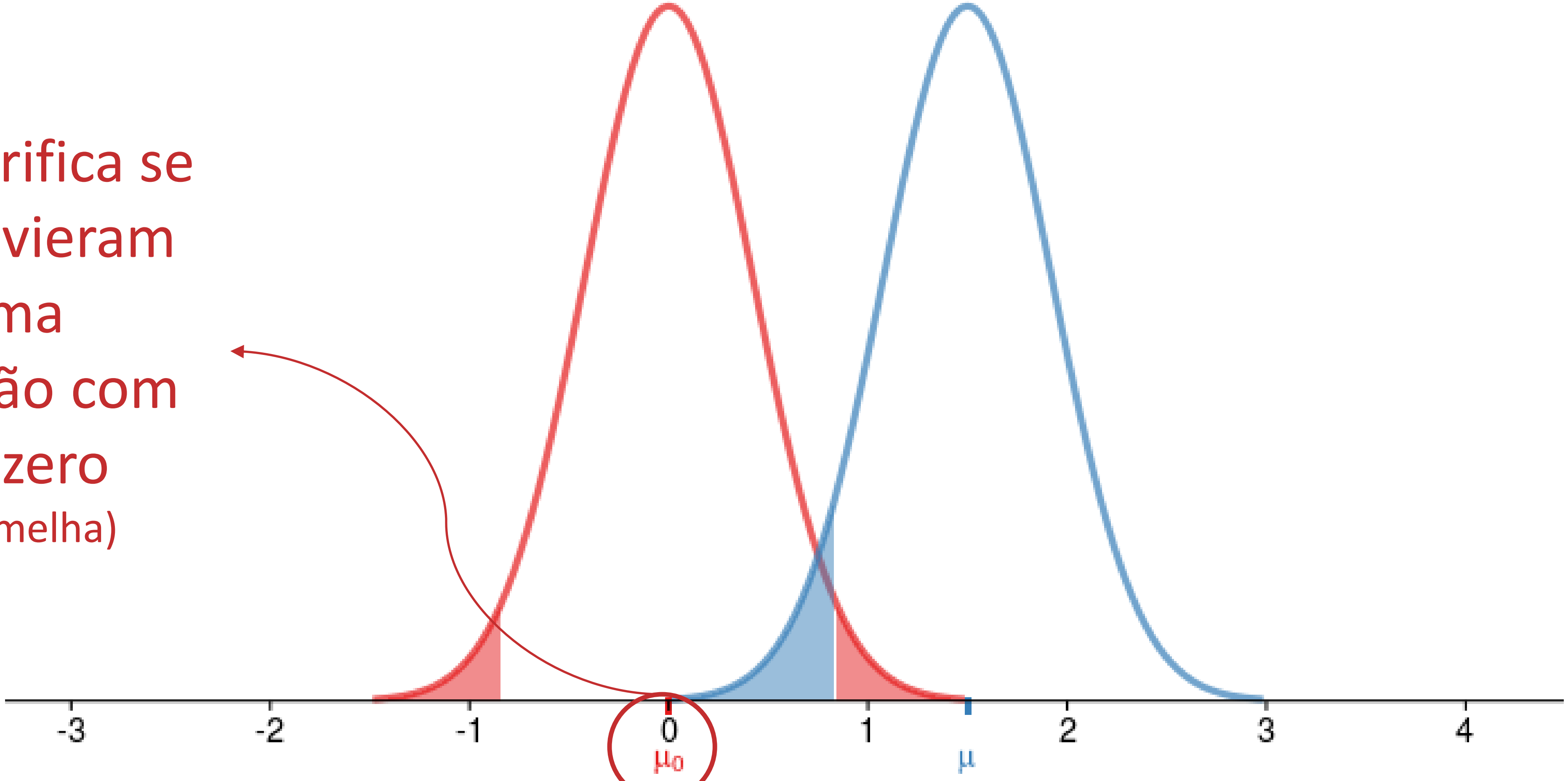



Os dados vieram de uma distribuição com média 1,5 (linha azul)

A linha azul é sempre desconhecida.

- **Mecanismos dos testes paramétricos**

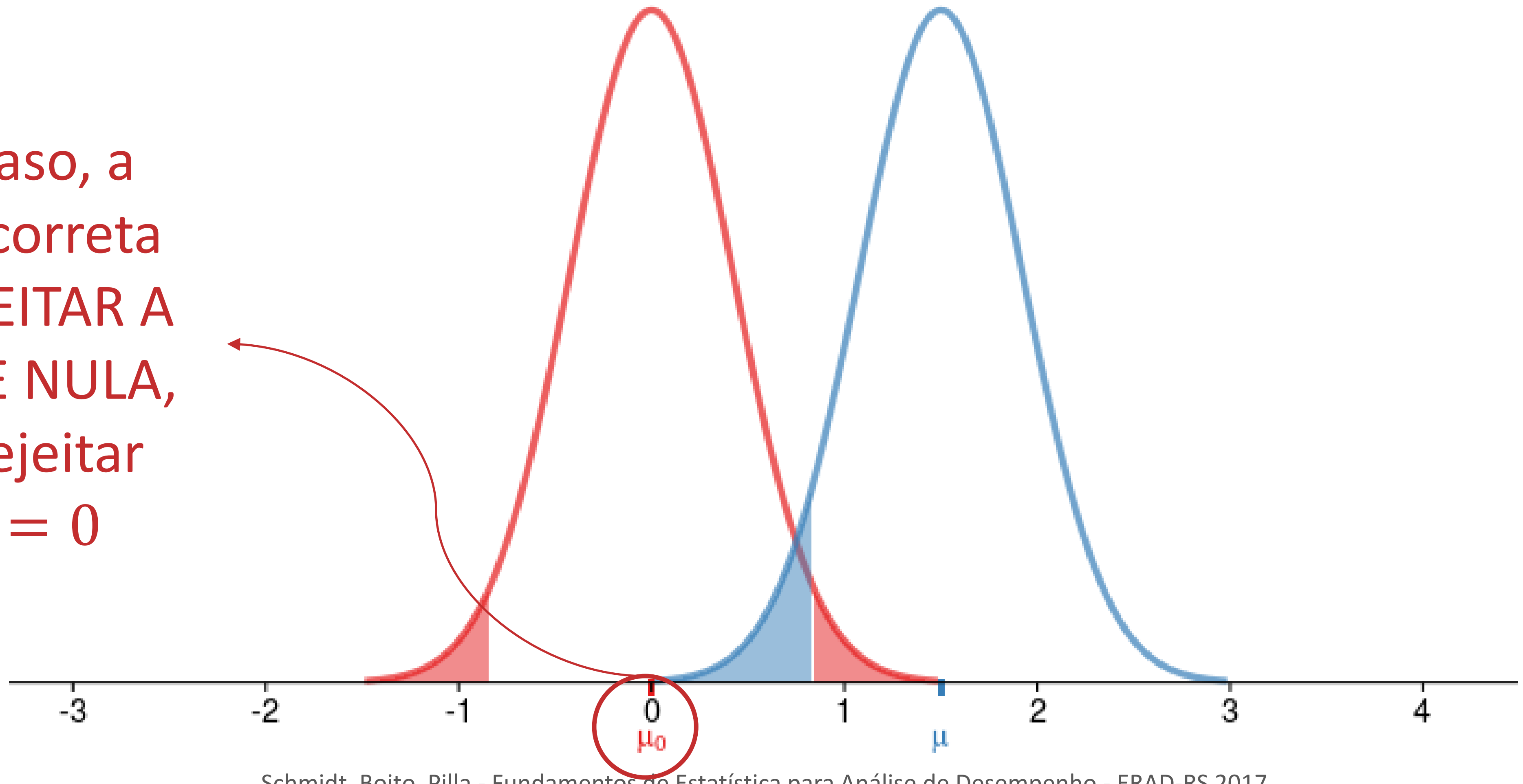
O teste verifica se os dados vieram de uma distribuição com média zero
(linha vermelha)



# Testes de Hipóteses

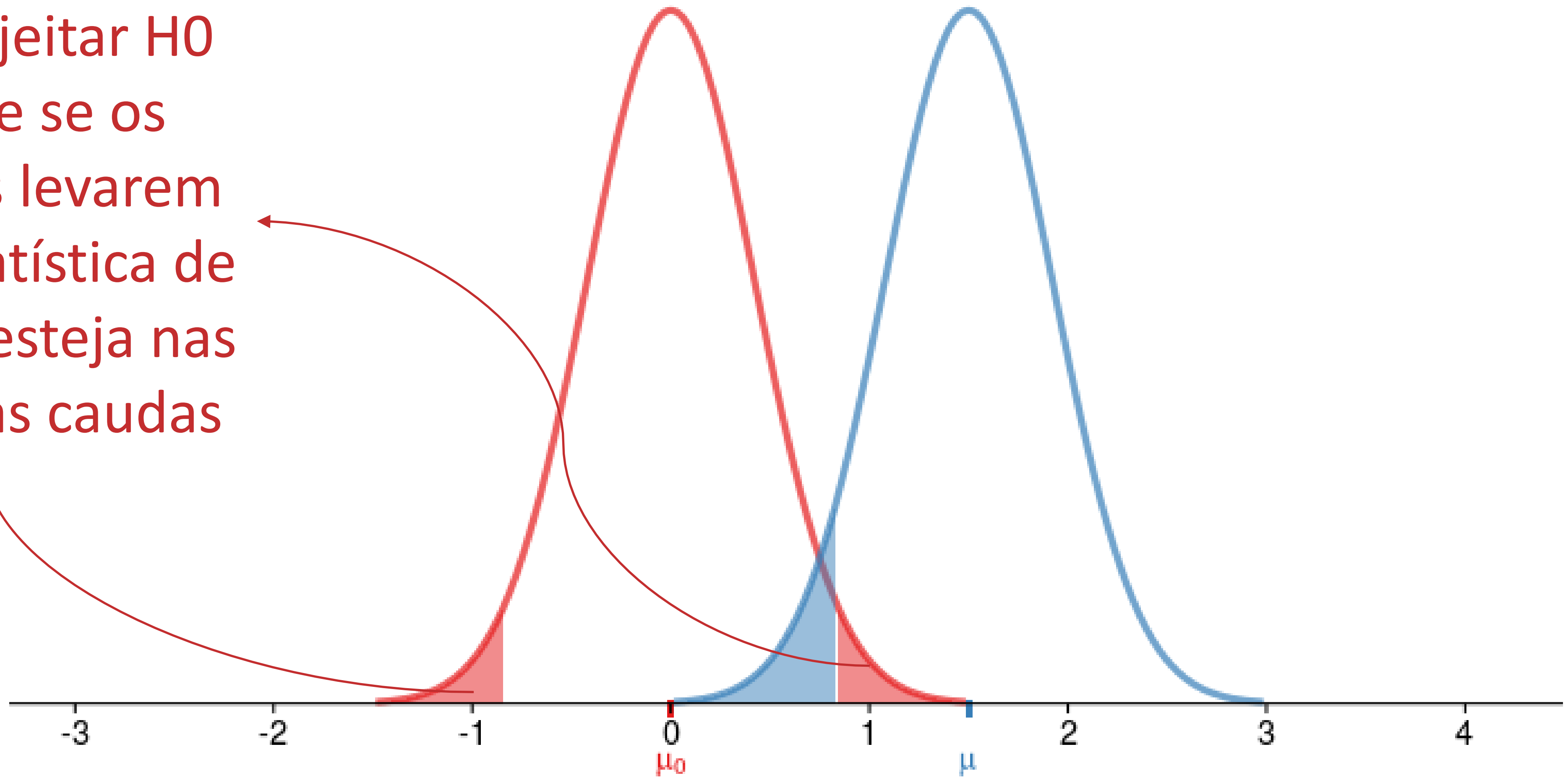
- **Mecanismos dos testes paramétricos**

Neste caso, a decisão correta seria REJEITAR A HIPÓTESE NULA, isto é, rejeitar $H_0$: $\mu = 0$



- **Mecanismos dos testes paramétricos**

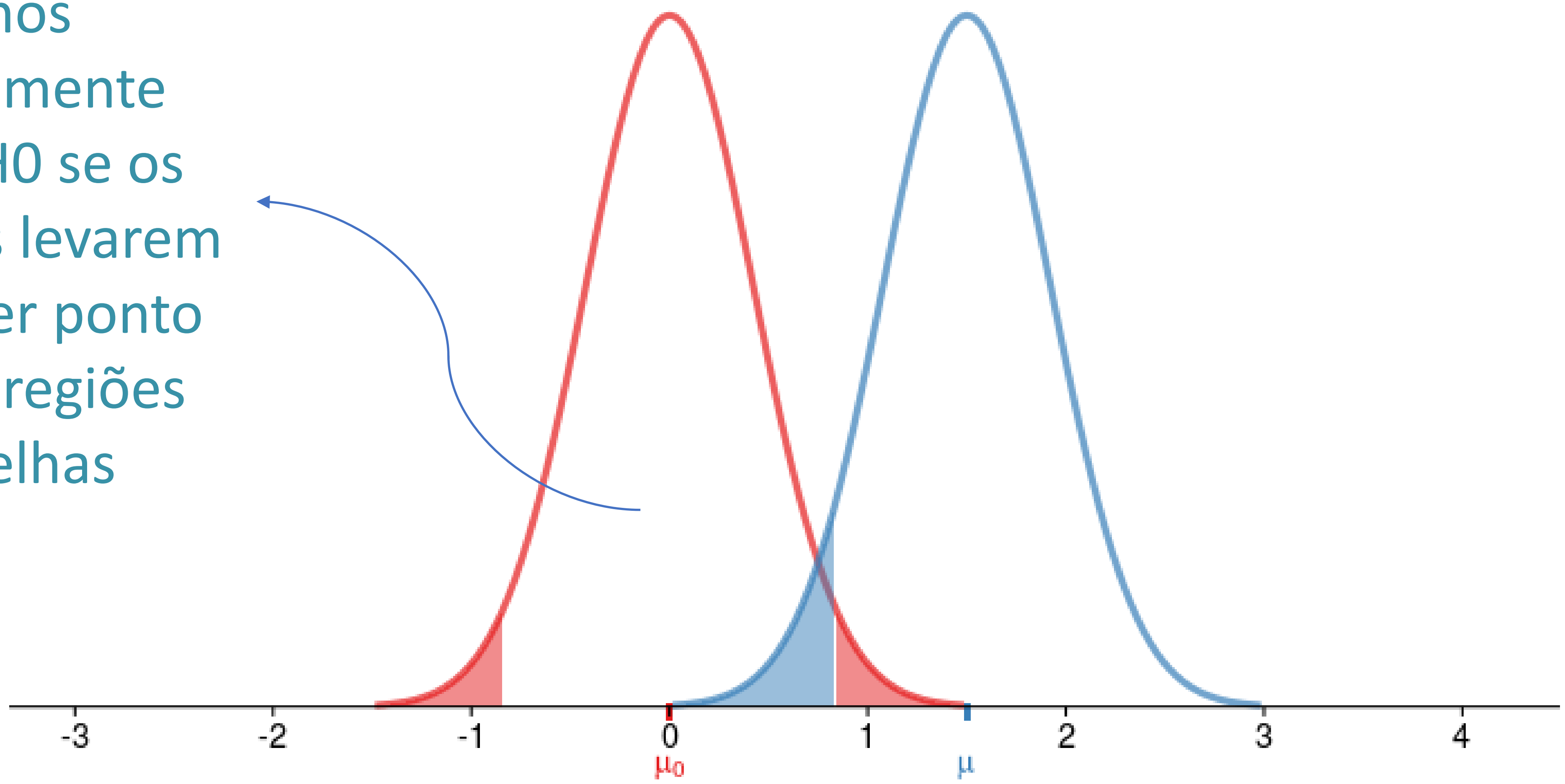
Iremos rejeitar H0 somente se os dados nos levarem a uma estatística de teste que esteja nas regiões das caudas



# Testes de Hipóteses

- **Mecanismos dos testes paramétricos**

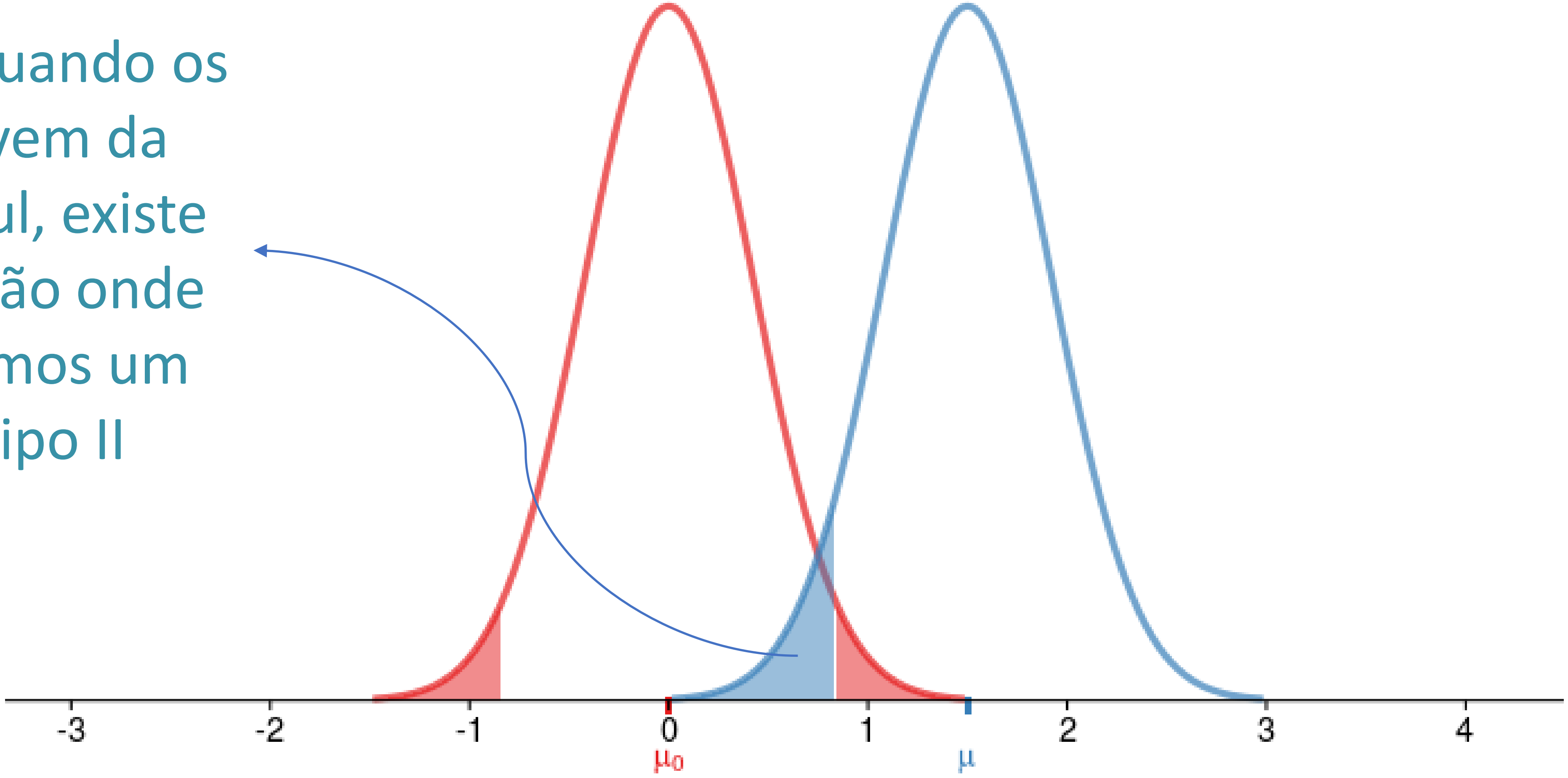
Iremos erroneamente aceitar H0 se os dados nos levarem a qualquer ponto entre as regiões vermelhas



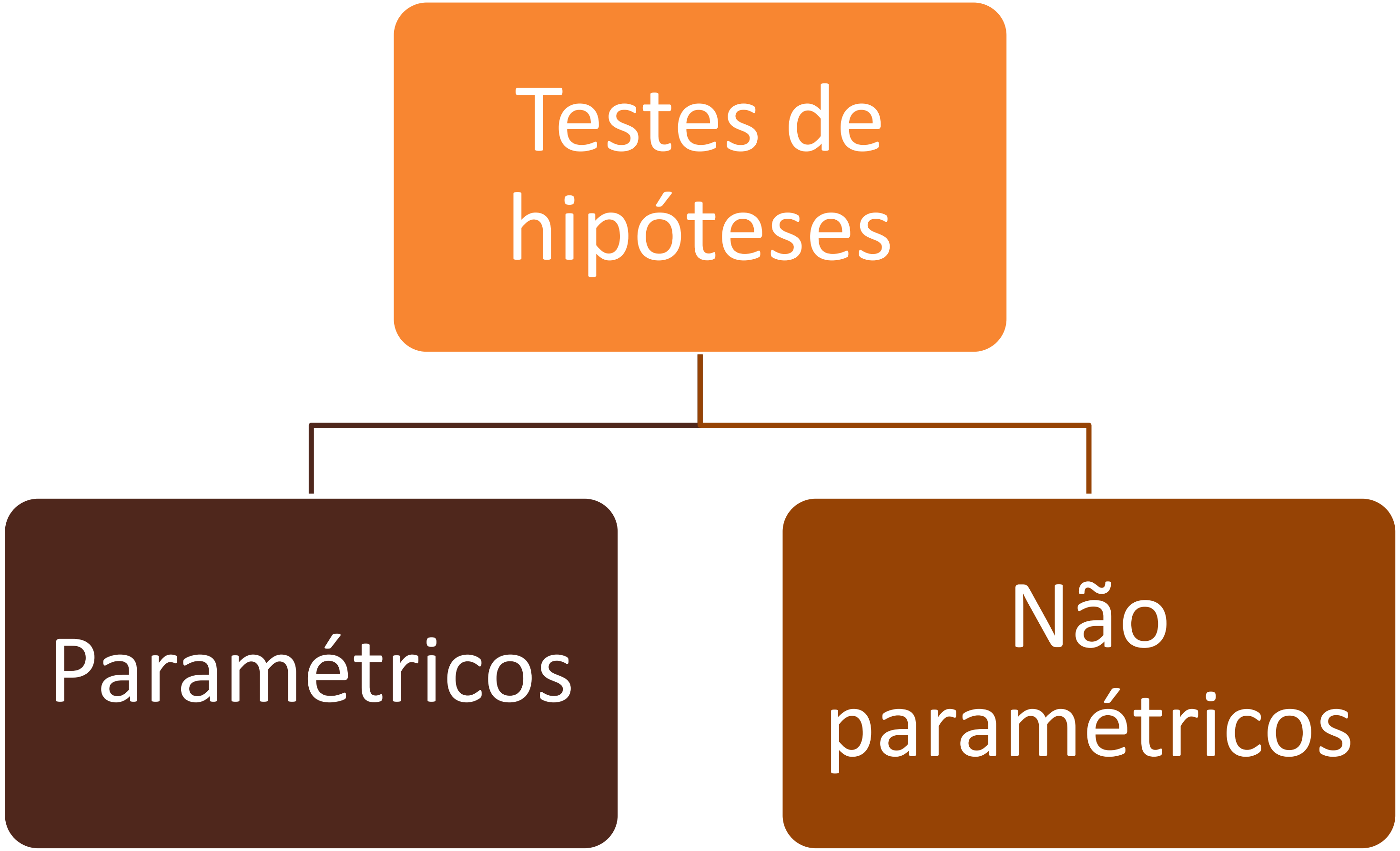
# Testes de Hipóteses

- **Mecanismos dos testes paramétricos**

Mesmo quando os dados vem da curva azul, existe essa região onde cometemos um erro tipo II



# Testes de Hipóteses



# Testes de Hipóteses

- **Roteiro básico**
  1. Estabeleça as hipóteses com base nas suas perguntas de pesquisa.

# Testes de Hipóteses

- **Roteiro básico**

1. Estabeleça as hipóteses com base nas suas perguntas de pesquisa.
2. Verifique quais testes são adequados para suas hipóteses e tipo de dado. Se as suposições forem atendidas, um teste paramétrico é sempre melhor que o não paramétrico.

# Testes de Hipóteses

- **Roteiro básico**

1. Estabeleça as hipóteses com base nas suas perguntas de pesquisa.
2. Verifique quais testes são adequados para suas hipóteses e tipo de dado. Se as suposições forem atendidas, um teste paramétrico é sempre melhor que o não paramétrico.
3. Estabeleça um nível de significância $\alpha$.

# Testes de Hipóteses

- **Roteiro básico**

1. Estabeleça as hipóteses com base nas suas perguntas de pesquisa.
2. Verifique quais testes são adequados para suas hipóteses e tipo de dado. Se as suposições forem atendidas, um teste paramétrico é sempre melhor que o não paramétrico.
3. Estabeleça um nível de significância $\alpha$.
4. Calcule a estatística de teste e o p-valor associado.

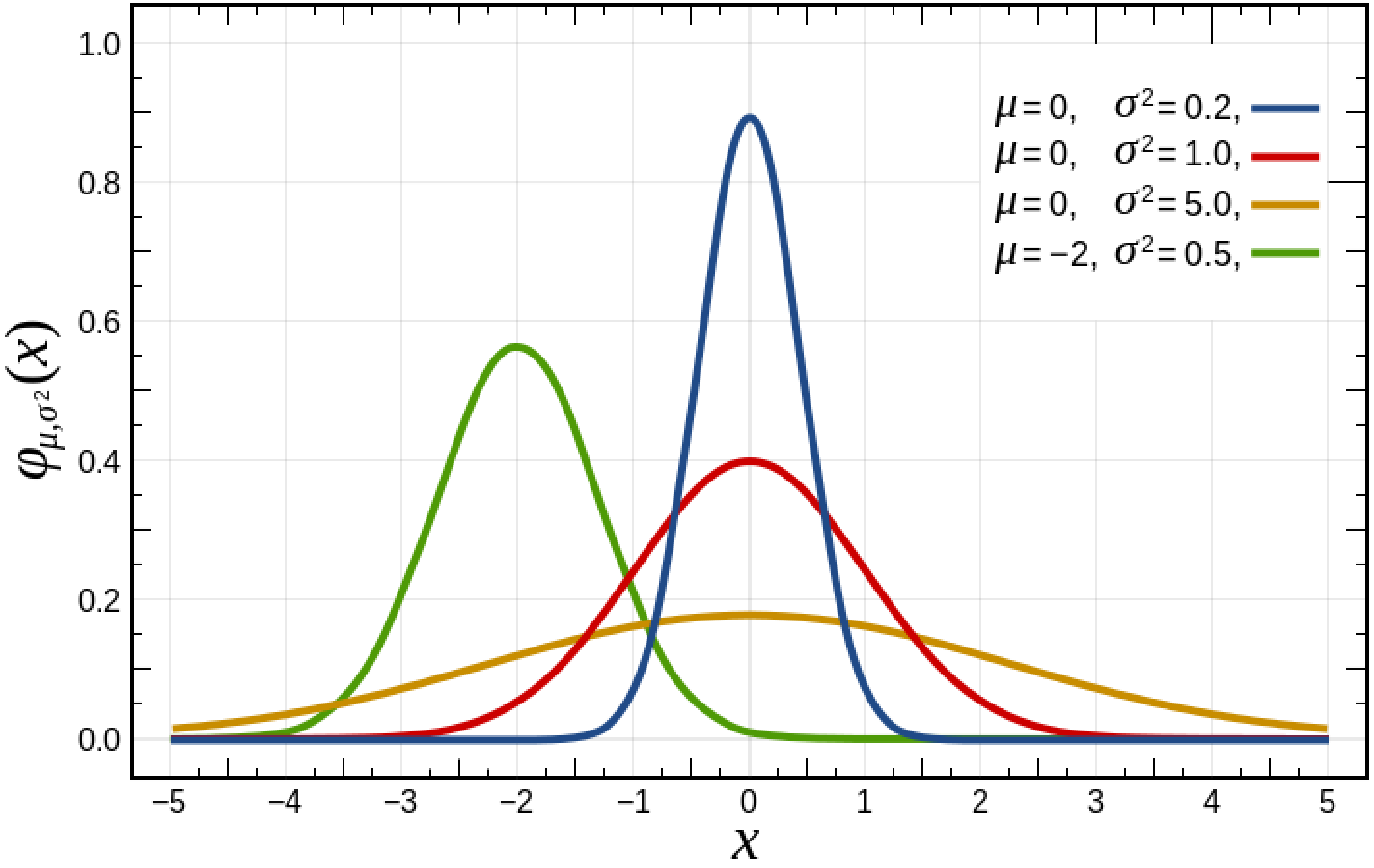
# Testes de Hipóteses

- **Roteiro básico**

1. Estabeleça as hipóteses com base nas suas perguntas de pesquisa.
2. Verifique quais testes são adequados para suas hipóteses e tipo de dado. Se as suposições forem atendidas, um teste paramétrico é sempre melhor que o não paramétrico.
3. Estabeleça um nível de significância $\alpha$.
4. Calcule a estatística de teste e o p-valor associado.
5. Interprete os resultados.

# TESTE DE NORMALIDADE

# Distribuição Normal




By Inductiveload - self-made, Mathematica, Inkscape, Public Domain,
https://commons.wikimedia.org/w/index.php?curid=3817954

# Teste de Normalidade

- **Exemplo**
  - Um algoritmo foi executado *n* = 100 vezes e foram anotados os tempos de cada execução.
  - Deseja-se testar, com significância $\alpha$ = 5% as hipóteses:
    - $H_0$: os tempos do algoritmo seguem uma distribuição normal com média $\bar{X}$ e desvio padrão $s\,/\sqrt{n}$
    - $H_1$: os tempos do algoritmo não seguem uma distribuição normal com média $\bar{X}$ e desvio padrão $s\,/\sqrt{n}$

# Teste de Normalidade

- **Teste de Kolmogorov-Smirnov (KS)**
  - Verifica se os dados “aderem” a uma distribuição contínua específica

# Teste de Normalidade

- **Teste de Kolmogorov-Smirnov (KS)**
  - Verifica se os dados "aderem" a uma distribuição contínua específica
  - Necessita que os dados estejam pelo menos em escala ordinal

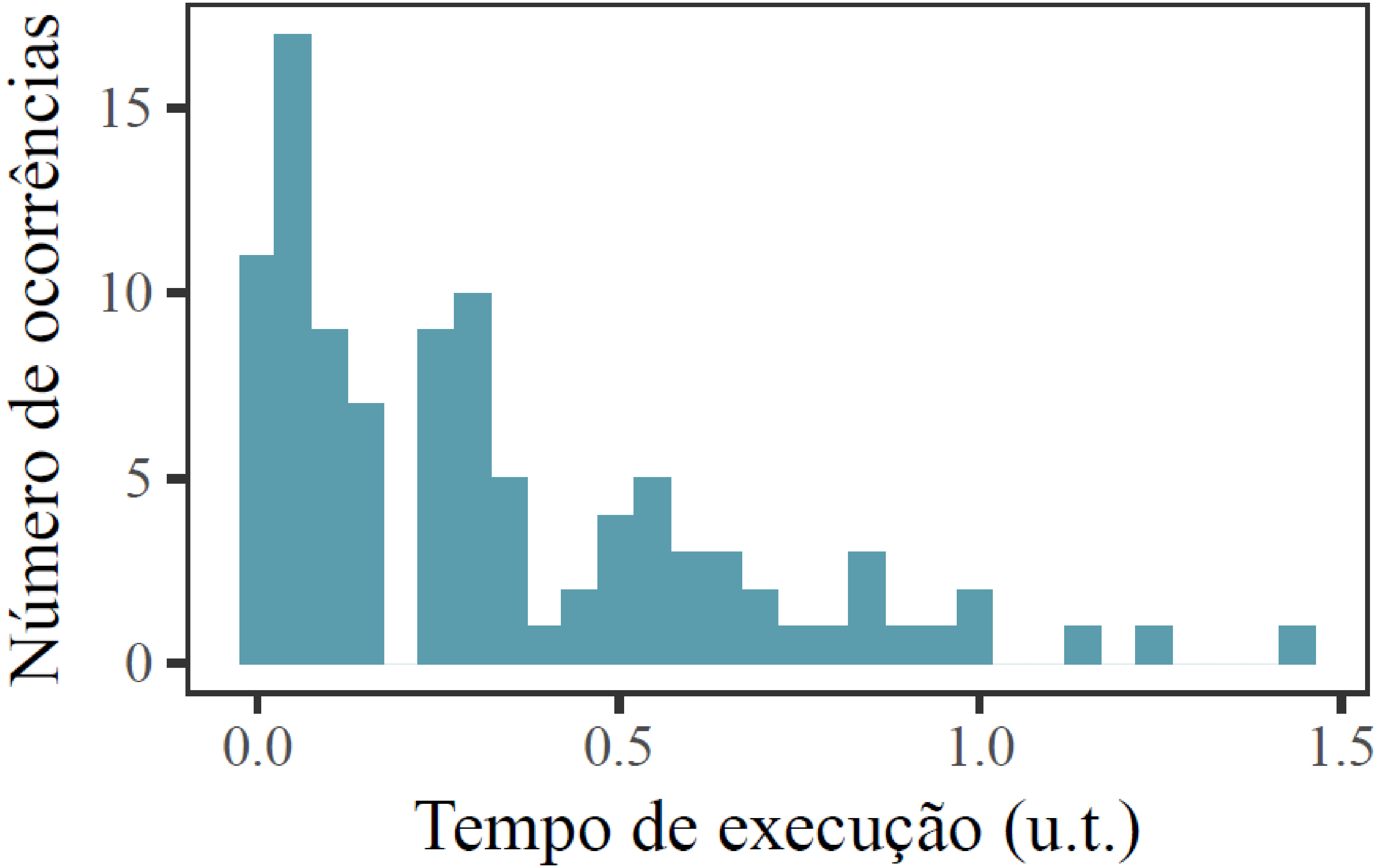
# Teste de Normalidade

- **Teste de Kolmogorov-Smirnov (KS)**
  - Verifica se os dados "aderem" a uma distribuição contínua específica
  - Necessita que os dados estejam pelo menos em escala ordinal
  - Verifica se a maior distância entre a distribuição de frequência acumulada dos dados e a frequência acumulada teórica pode ser considerada "ao acaso"

# Teste de Normalidade

- **Teste de Kolmogorov-Smirnov (KS)**
  - Verifica se os dados "aderem" a uma distribuição contínua específica
  - Necessita que os dados estejam pelo menos em escala ordinal
  - Verifica se a maior distância entre a distribuição de frequência acumulada dos dados e a frequência acumulada teórica pode ser considerada "ao acaso"
  - Apresenta limitações para lidar com observações repetidas

# Teste de Normalidade



# Teste de Normalidade

```r
# Exemplos do texto do curso

## Exemplo 1.7 (p. 15) - teste de normalidade

### Preparação de dados sintéticos padronizados
set.seed(1234)                      # fixa a semente aleatória
lambda <- 3                         # parâmetro lambda a ser usado para a geração de pontos em uma distribuição exponencial
n <- 100                            # número de repetições
tempos <- rexp(n = n, rate = lambda)  # geração de tempos segundo uma distribuição exponencial
```

```r
### Aplicação do teste de Kolmogorov-Smirnov para verificarmos normalidade
media <- mean(tempos)          # media a ser usada para a geração da distribuição normal
desvio <- sd(tempos)           # desvio padrão para ser usado na geração da distribuição normal
ks.test(tempos, 'pnorm', media, desvio/sqrt(n)) # 'desvio/sqrt(n)' é uma aplicação errada do Teorema do Limite Central (proposital aqui)
# O resultado acima é um pouco diferente do valor no texto (0.49403 vs 0.50073)
# porque os valores originais usavam as medias e desvio padrão
# populacionais ao invés dos amostrais. No caso abaixo usamos os parâmetros populacionais
# que nós conhecemos da distribuição exponencial
ks.test(tempos, 'pnorm', 1/lambda, (1/lambda)/sqrt(n))
```

# Teste de Normalidade

One-Sample Kolmogorov-Smirnov test
D = 0.49403, p-value < 2.2e-16

A probabilidade de que os dados que produziram a estatística de teste calculada tenham saído de uma distribuição normal com média aproximadamente igual a 1/3 e desvio padrão aproximadamente igual a 1/30 é menor do que $2.2*10^{-16}$

# Teste de Normalidade

One-Sample Kolmogorov-Smirnov test
D = 0.49403, p-value < 2.2e-16

A probabilidade de que os dados que produziram a estatística de teste calculada tenham saído de uma distribuição normal com média aproximadamente igual a 1/3 e desvio padrão aproximadamente igual a 1/30 é menor do que $2.2*10^{-16}$

Mas a amostra não era grande?

# Teste de Normalidade

"Sejam $X_1, X_2, \dots, X_n, \dots$, uma sequência de variáveis aleatórias i.i.d., cujas funções geradoras de momentos existem na proximidade de $0$. Sejam $\mathbb{E}[X_i] = \mu$ e $Var[X_i] = \sigma^2 > 0$.

Adaptado de: CASELLA, G., BERGER, R.. *Inferência Estatística.* São Paulo, 2010

# Teste de Normalidade

"Sejam $X_1, X_2, \ldots, X_n, \ldots$, uma sequência de variáveis aleatórias i.i.d., cujas funções geradoras de momentos existem na proximidade de 0. Sejam $\mathbb{E}[X_i] = \mu$ e $Var[X_i] = \sigma^2 > 0$.

Defina:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Então, a quantidade:

$$Z = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$$

Tem uma distribuição normal padrão limite".

Adaptado de: CASELLA, G., BERGER, R.. *Inferência Estatística.* São Paulo, 2010

# Teste de Normalidade

“Sejam $X_1, X_2, \ldots, X_n, \ldots$, uma sequência de variáveis aleatórias i.i… …as de momentos existem na pr… $[X_i] = \mu$ e $Var[X_i] = \sigma^2 > 0$. Defina:

$$\ldots_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Podem ter QUALQUER distribuição, atendidas as condições de variância e média finitas

Então, a quantidade:

$$Z = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$$

Distribuição normal padrão assintótica (significa n grande)

Tem uma distribuição normal padrão limite”.

Adaptado de: CASELLA, G., BERGER, R.. *Inferência Estatística.* São Paulo, 2010

# Teste de Normalidade

É possível mostrar que a média e a variância populacionais de $\bar{X}_n$ serão iguais a:

$$\mathbb{E}[\bar{X}_n] = \mu$$
$$Var[\bar{X}_n] = \frac{\sigma}{\sqrt{n}}$$

Utilizaremos o fato de que a média amostral e variância amostral de $\bar{X}_n$ irão convergir para os valores populacionais para valores grandes de *n*.

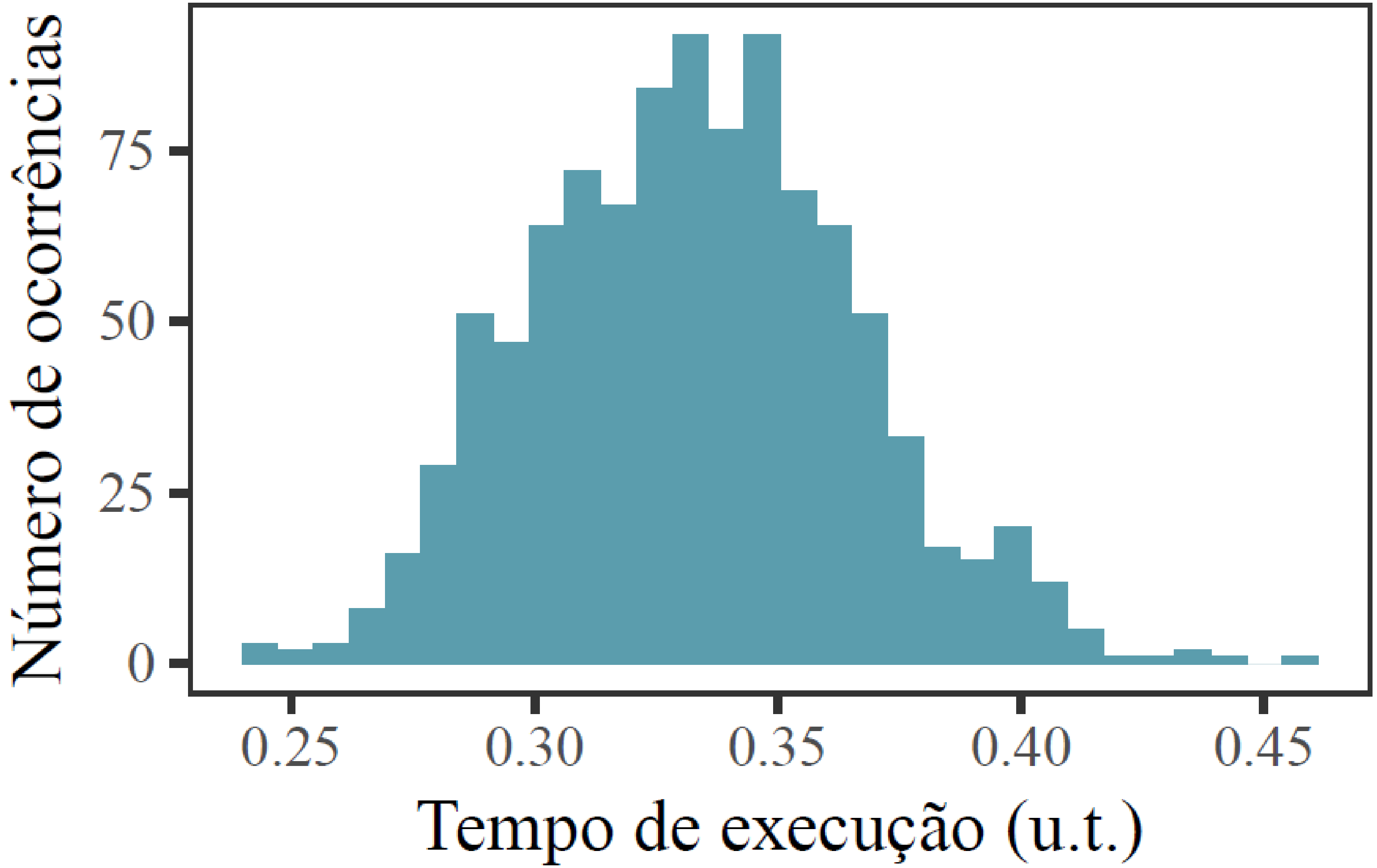
# Teste de Normalidade

- **Retorno ao exemplo**
  - Um algoritmo foi executado n = 100 vezes e foi calculada a **média** dos tempos de execução. Esse procedimento foi repetido 1000 vezes.

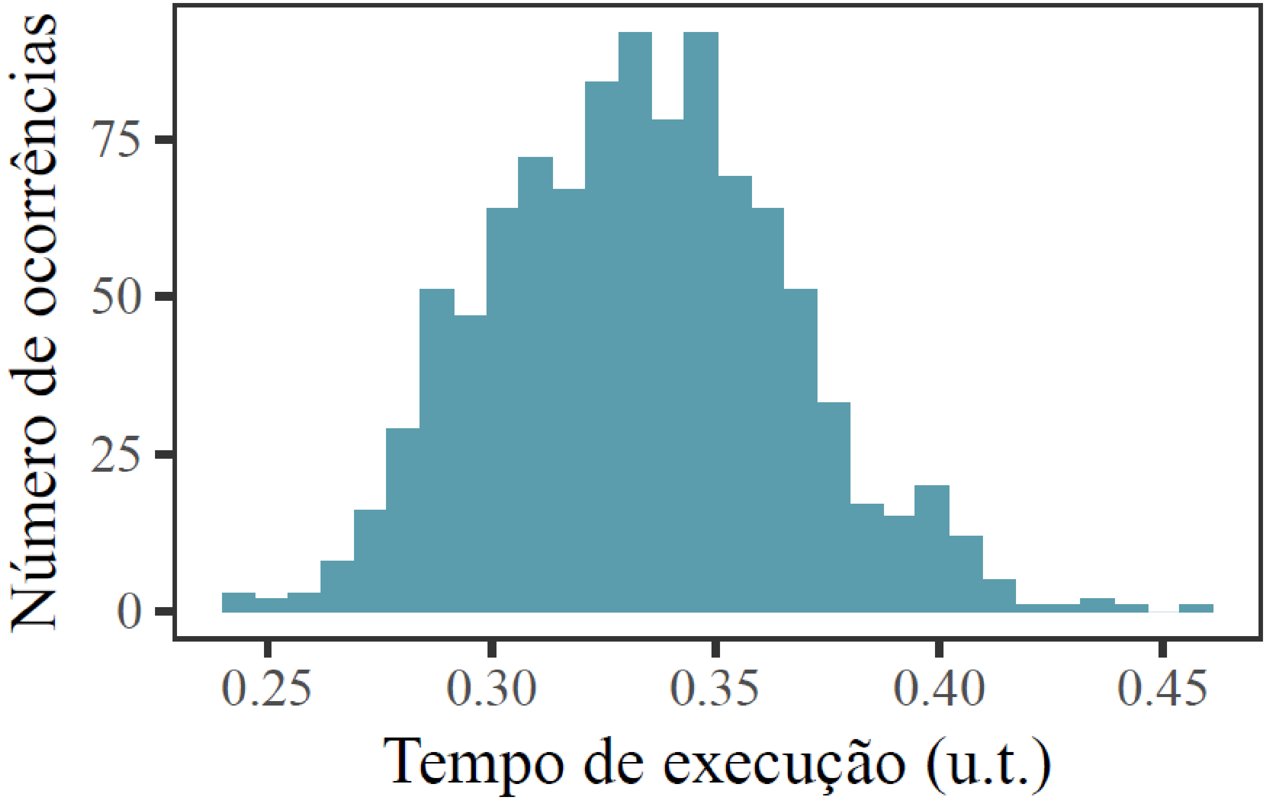
# Teste de Normalidade

- **Retorno ao exemplo**
  - Um algoritmo foi executado n = 100 vezes e foi calculada a **média** dos tempos de execução. Esse procedimento foi repetido 1000 vezes.
  - Deseja-se testar, com significância $\alpha$ = 5% as hipóteses:
    - $H_0$: a média dos tempos seguem uma distribuição normal com média $\bar{X}$ e desvio padrão $s\ /\sqrt{n}$
    - $H_1$: a média dos tempos não seguem uma distribuição normal com média $\bar{X}$ e desvio padrão $s\ /\sqrt{n}$

# Teste de Normalidade



# Teste de Normalidade




One-Sample Kolmogorov-Smirnov test
D = 0.01956, p-value =0.8388

A probabilidade de que os dados que produziram a estatística de teste calculada tenham saído de uma distribuição normal com média aproximadamente igual a 1/3 e desvio padrão aproximadamente igual a 1/3 é igual a 0.8388.

# TESTE T PARA MÉDIAS

# Teste t para médias

- **Teste t para uma média**
  - **Exemplo**
    - Suponha que você está testando a eficiência de um novo algoritmo e deseja verificar se ela é maior que 80%.
    - Deseja-se testar, com significância a = 5% as hipóteses:
      - $H_0$: eficiência menor ou igual que 80%
      - $H_1$: eficiência maior que 80%

# Teste t para médias

- **Teste t para uma média**
  - **Suposições**
    - Os dados amostrais vêm de uma mesma população com distribuição normal
    - As observações da amostra são independentes entre si

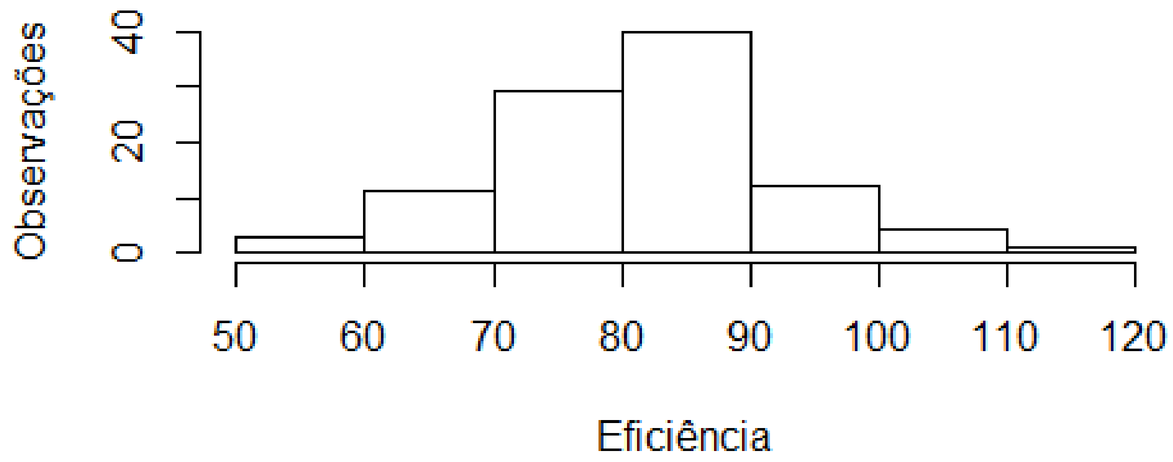
# Teste t para médias

- **Teste t para uma média**
  - Suposições
    - Os dados amostrais vem de uma mesma população com distribuição normal
    - As observações da amostra são independentes entre si
  - **Estatística de teste**

$$t_{teste} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

# Teste t para médias

- **Teste t para uma média**
  - **Exemplo**

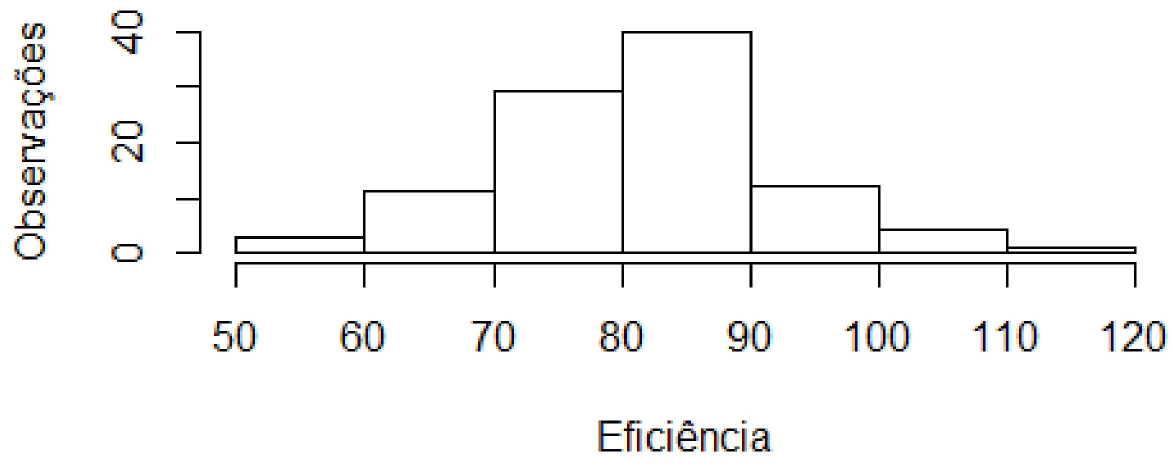



One-Sample Kolmogorov-Smirnov test
D = 0.076876, p-value =0.5957

One-Sample t-test
t = 0.7431, p-value =0.2296
$H_1\colon\ \mu > 80$
95% IC: $(79.03; \infty)$
Sample mean: 80.78

# Teste t para médias

- **Teste t para uma média**
  - **Exemplo**




One-Sample Ko

D = 0.076876,

One-Sample t-t

t = 0.7431, p-value =0.2296

$H_1:\ \mu > 80$

95% IC: $(79.03; \infty)$

Sample mean: 80.78

**Não é possível rejeitar a hipótese nula de que a eficiência é menor ou igual a 80%.**

# Teste t para médias

- **Teste t para duas médias e variâncias iguais**
  - **Exemplo**
    - Suponha que você está testando a eficiência de um algoritmo $\theta$ com um algoritmo $\psi$.

# Teste t para médias

- **Teste t para duas médias e variâncias iguais**
  - **Exemplo**
    - Suponha que você está testando a eficiência de um algoritmo $\theta$ com um algoritmo $\psi$.
    - Para executar um teste t, precisamos primeiro testar as variâncias utilizando um teste F.
      - $H_0$: $\sigma_\psi^2 = \sigma_\theta^2$
      - $H_1$: $\sigma_\psi^2 \neq \sigma_\theta^2$

# Teste t para médias

- **Teste F de homogeneidade de variâncias**
  - Estatística de teste:
    - $F_{teste} = \frac{S_{\psi}^2}{S_{\theta}^2} \sim F_{n-1,n-2}$

# Teste t para médias

- **Teste F de homogeneidade de variâncias**
  - Estatística de teste:
    - $F_{teste} = \frac{S_{\theta}^2}{S_{\psi}^2} \sim F_{n-1,n-2}$
  - Suposições:
    - Os dados de cada amostra vêm de uma mesma população com distribuição normal

# Teste t para médias

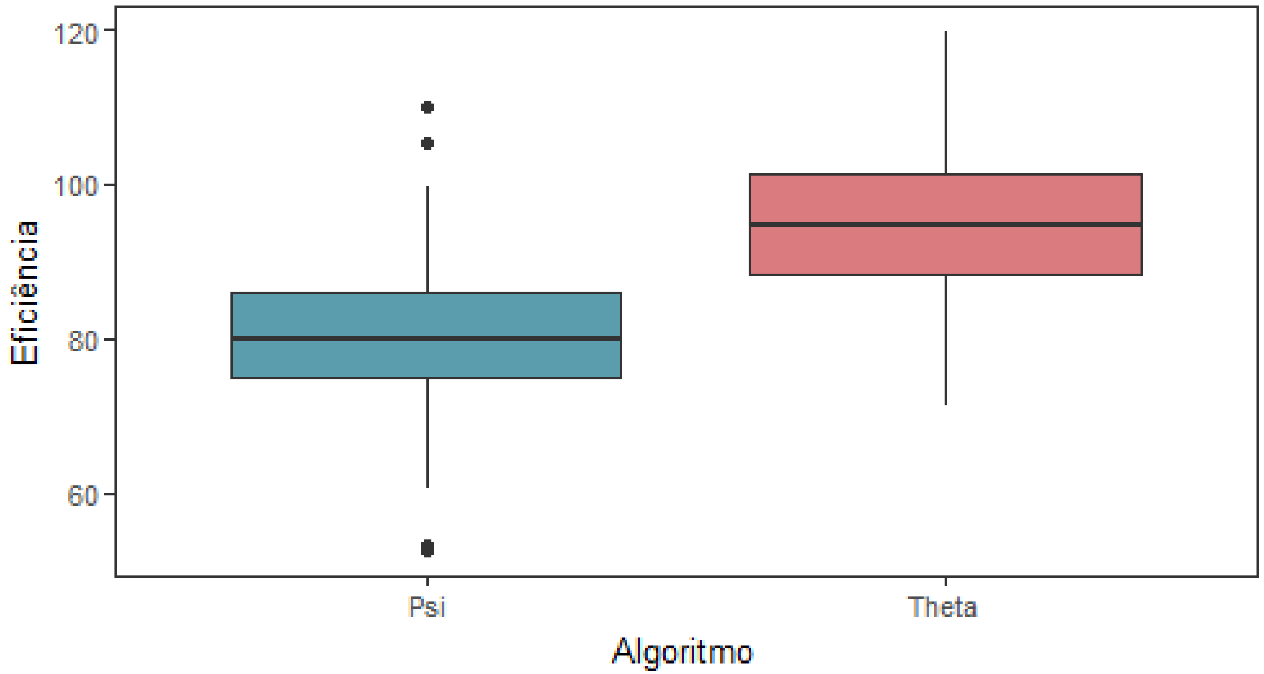
- **Teste F de homogeneidade de variâncias**
  - Estatística de teste:
    - $F_{teste} = \frac{S_\theta^2}{S_\psi^2} \sim F_{n-1,n-2}$
  - Suposições:
    - Os dados de cada amostra vêm de uma mesma população com distribuição normal
    - As observações dentro amostra são independentes entre si

# Teste t para médias

- **Teste F de homogeneidade de variâncias**
  - Estatística de teste:
    - $F_{teste} = \frac{S_{\theta}^2}{S_{\psi}^2} \sim F_{n-1,n-2}$
  - Suposições:
    - Os dados de cada amostra vêm de uma mesma população com distribuição normal
    - As observações dentro amostra são independentes entre si
    - As amostras são independentes entre si

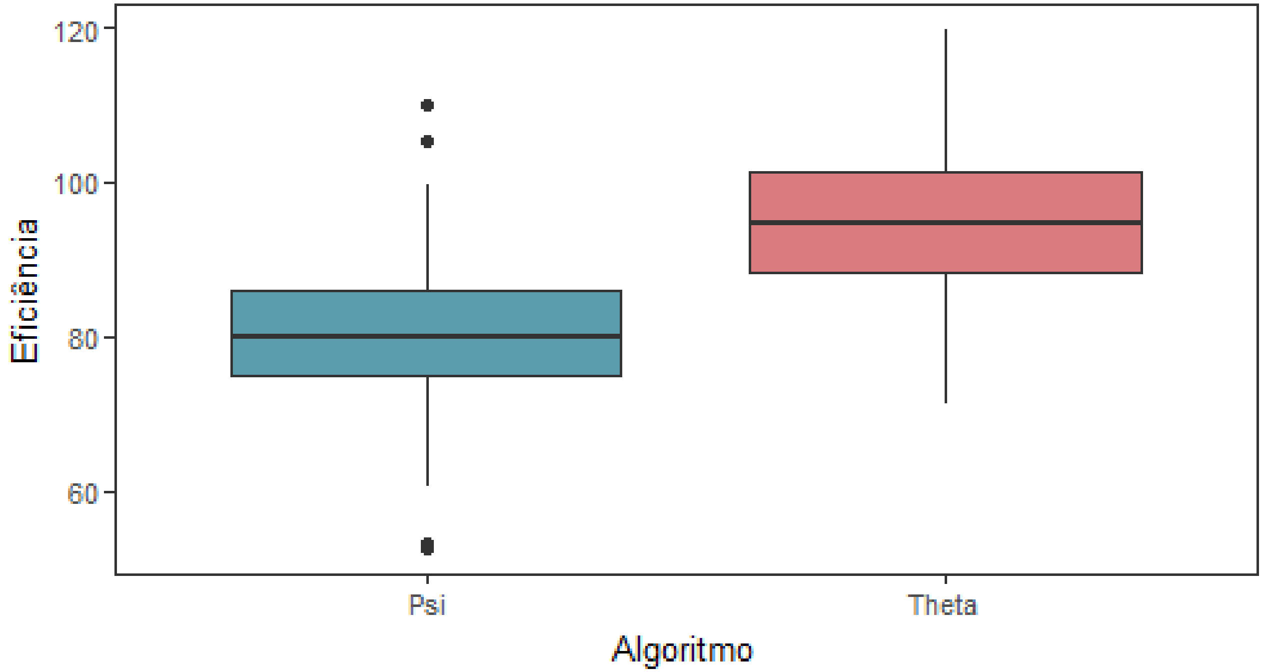
# Teste t para médias

- **Teste de normalidade e teste F**
  - Exemplo



# Teste t para médias

- **Teste de normalidade e teste F**
  - **Exemplo**

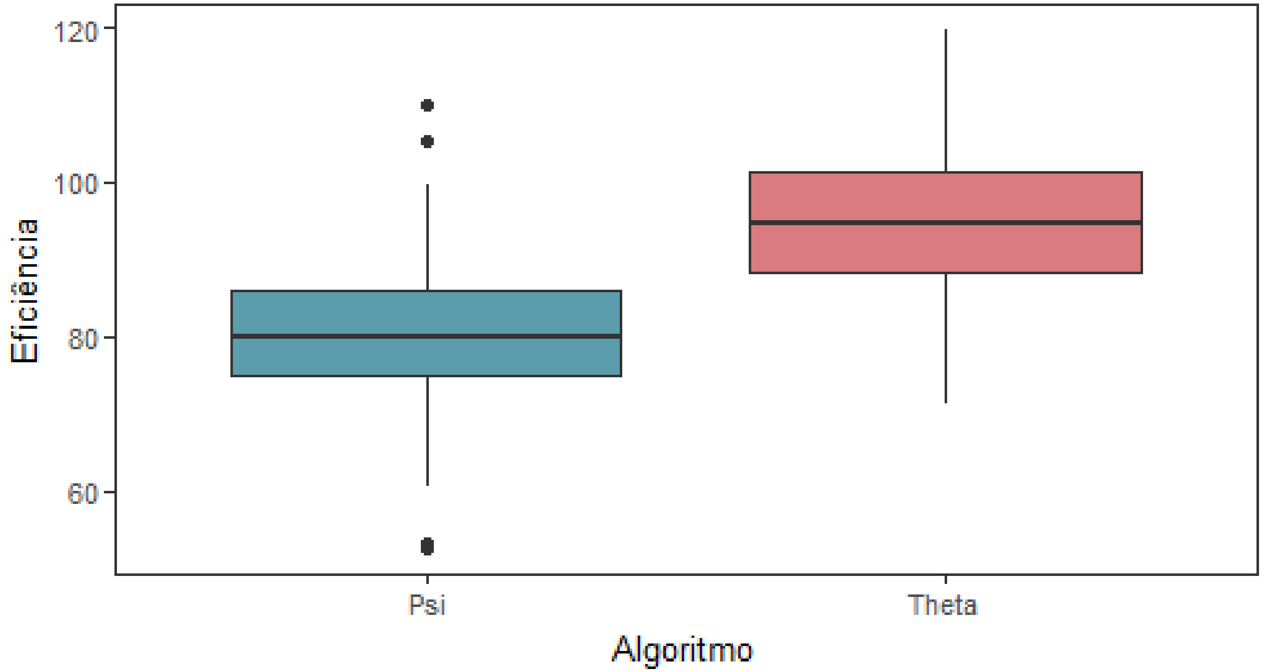



Theta:
One-Sample Kolmogorov-Smirnov test
D = 0.03, p-value =0.9922

Psi:
One-Sample Kolmogorov-Smirnov test
D = 0.046001, p-value =0.7912

# Teste t para médias

- **Teste de normalidade e teste F**
  - **Exemplo**




Theta:
One-Sample Kolmogorov-Smirnov test
D = 0.03, p-value =0.9922

Psi:
One-Sample Kolmogorov-Smirnov test
D = 0.046001, p-value =0.7912

F test to compare two variances
F = 0.92, p-value =0.5582
$H_1: \sigma_\psi^2/\sigma_\theta^2 \neq 1$
95% IC: $(0.6964; 1.2160)$
Ratio of variances: 0.92

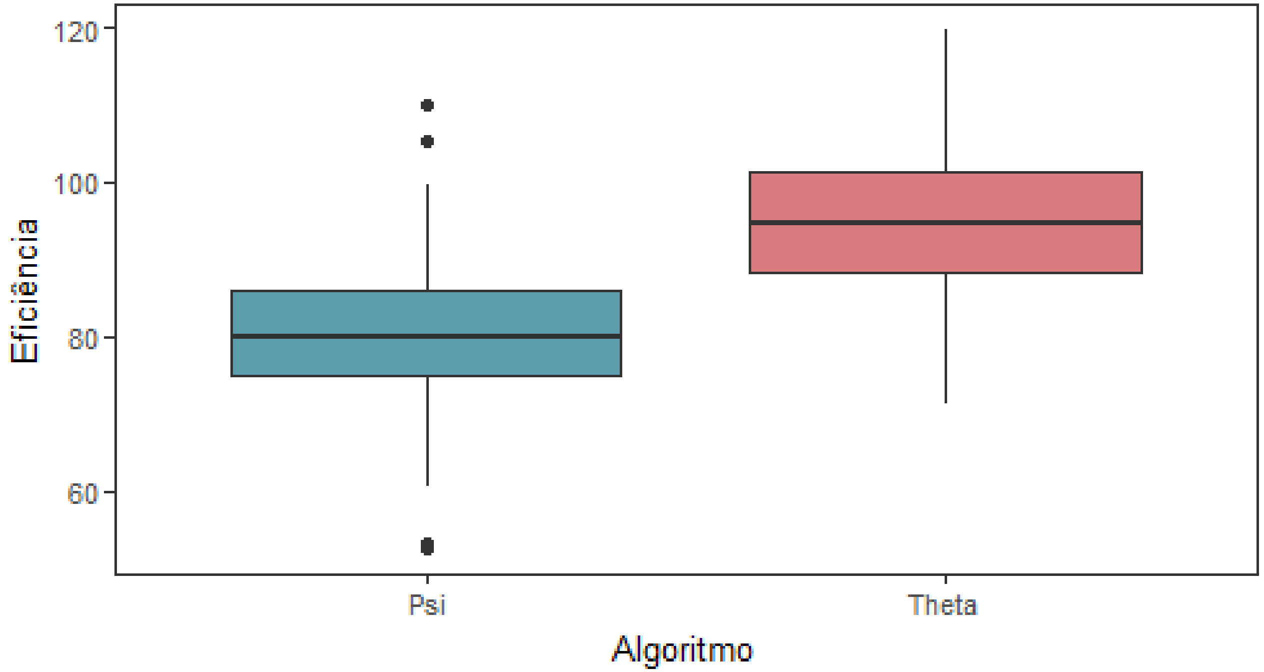
# Teste t para médias

- **Teste t para duas médias e variâncias iguais**
  - **Exemplo**
    - Agora testamos as médias:
      - $H_0$: $\mu_\theta = \mu_\psi$
      - $H_1$: $\mu_\theta \neq \mu_\psi$

    (valem as mesmas suposições do teste F)
    - Estatística de teste:
      - $t_{teste} = \frac{\overline{X_1} - \overline{X_2}}{S_{\overline{X_1} - \overline{X_2}}} \sim t_{n_1+n_2-2}$

# Teste t para médias

- **Teste t para duas médias e variâncias iguais**
  - **Exemplo**




Two Sample t-test
t = -14.943, p-value < 0.001
$H_1$: $\mu_\theta - \mu_\psi \neq 0$
95% IC: $(-16.37; -12.56)$
Sample estimates:
Theta: 94.89
Psi: 80.43

# TESTE NÃO PARAMÉTRICO PARA DUAS AMOSTRAS

# Teste não paramétrico para duas amostras

- **Teste de Wilcoxon-Mann-Whitney**
  - **Suposições**
    - Duas amostras independentes*
    - Mensuração pelo menos ordinal

# Teste não paramétrico para duas amostras

- **Teste de Wilcoxon-Mann-Whitney**
  - **Suposições**
    - Duas amostras independentes*
    - Mensuração pelo menos ordinal
  - **Mecanismo**
    - Similar ao procedimento do KS, este é um teste não paramétrico

      -> Utiliza a distribuição empírica dos dados

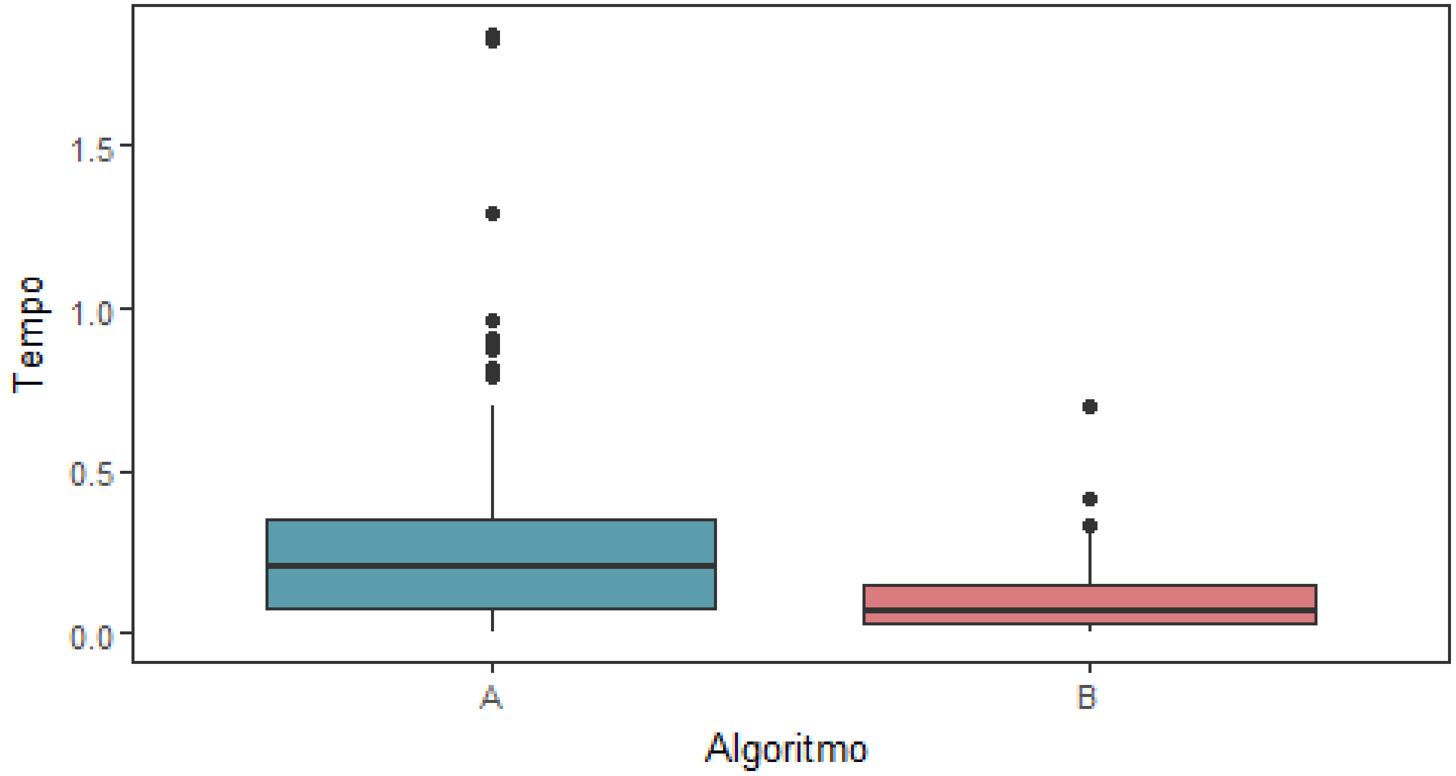
      -> Calcula ranks (postos)

# Teste não paramétrico para duas amostras

- **Teste de Wilcoxon-Mann-Whitney**
  - **Exemplo**
  - Temos dois algoritmos, A e B, e resultados de 100 execuções independentes de cada um.
  - Desejamos testar, com nível de significância $\alpha$ = 5%, as hipóteses
    - $H_0$: os tempos de execução dos algoritmos são provenientes de uma mesma distribuição
    - $H_1$: os tempos são provenientes de distribuições diferentes

# Teste não paramétrico para duas amostras

- **Teste de Wilcoxon-Mann-Whitney**
  - **Exemplo**




Wilcoxon rank sum test

W = 7460, p-value < 0.001

$H_1: Os\ dados\ não\ vêm\ de\ uma\ mesma\ distribuição$

# E se eu tiver mais do que duas médias? Faço testes duas a duas?

# O problema dos testes sequenciais

- Suponha que você tem uma moeda honesta.

  A probabilidade de obter pelo menos uma cara em 2 lançamentos é 1 menos a probabilidade de obter exatamente duas coroas:

# O problema dos testes sequenciais

- Suponha que você tem uma moeda honesta.
  - A probabilidade de obter pelo menos uma cara em 2 lançamentos é:

$$1 - \underbrace{(0.5 * 0.5)}_{\text{Probabilidade de 2 coroas}} = 1 - 0.25 = 0.75$$

# O problema dos testes sequenciais

E a probabilidade de obter pelo menos uma cara em 4 lançamentos é:

$$1 - (0.5)^4 = 1 - 0.0625 = 0.9375$$

# O problema dos testes sequenciais

- De maneira similar, quando efetuamos um teste de hipóteses, a probabilidade de **não** cometer um erro tipo I é dada por $1 - \alpha$ (usualmente 95%).

# O problema dos testes sequenciais

- De maneira similar, quando efetuamos um teste de hipóteses, a probabilidade de **não** cometer um erro tipo I é dada por $1 - \alpha$ (usualmente 95%).
- Para dois testes sequenciais, a probabilidade de não cometer um erro tipo I será $(1 - \alpha)^2$ e, portanto, a probabilidade de cometer o erro tipo I será $1 - (1 - \alpha)^2$.

# O problema dos testes sequenciais

- De maneira similar, quando efetuamos um teste de hipóteses, a probabilidade de **não** cometer um erro tipo I é dada por $1-\alpha$ (usualmente 95%).
- Para dois testes sequenciais, a probabilidade de não cometer um erro tipo I será $(1-\alpha)^2$ e, portanto, a probabilidade de cometer o erro tipo I será $1-(1-\alpha)^2$.
  - Por exemplo, se tivermos dois testes com nível de significância de 95%:

$$\mathbb{P}(erro\ tipo\ I) = 1-(1-0.05)^2 = 0.0975$$

# ONE-WAY ANOVA

# One-Way ANOVA

- **Suposições**
  - Deve haver uma variável dependente e três ou mais grupos;

# One-Way ANOVA

- **Suposições**
  - Deve haver uma variável dependente e três ou mais grupos;
  - A variável dependente deve ter distribuição normal para cada um dos grupos;

# One-Way ANOVA

- **Suposições**
  - Deve haver uma variável dependente e três ou mais grupos;
  - A variável dependente deve ter distribuição normal para cada um dos grupos;
  - Cada um dos grupos deve ser independente dos demais;

# One-Way ANOVA

- **Suposições**
  - Deve haver uma variável dependente e três ou mais grupos;
  - A variável dependente deve ter distribuição normal para cada um dos grupos;
  - Cada um dos grupos deve ser independente dos demais;
  - A observações dentro de cada um dos grupos devem ser independentes.

# One-Way ANOVA

- **Hipóteses**
  - $H_0$: $\mu_i = \mu_j \; \forall \; i, j$
  - $H_1$: Pelo menos uma média é diferente das demais

# One-Way ANOVA

- **Hipóteses**
  - $H_0$: $\mu_i = \mu_j \ \forall \ i, j$
  - $H_1$: Pelo menos uma média é diferente das demais
    - **Qual ou quais delas?**

# One-Way ANOVA

- **Hipóteses**
  - $H_0$: $\mu_i = \mu_j \; \forall \; i, j$
  - $H_1$: Pelo menos uma média é diferente das demais

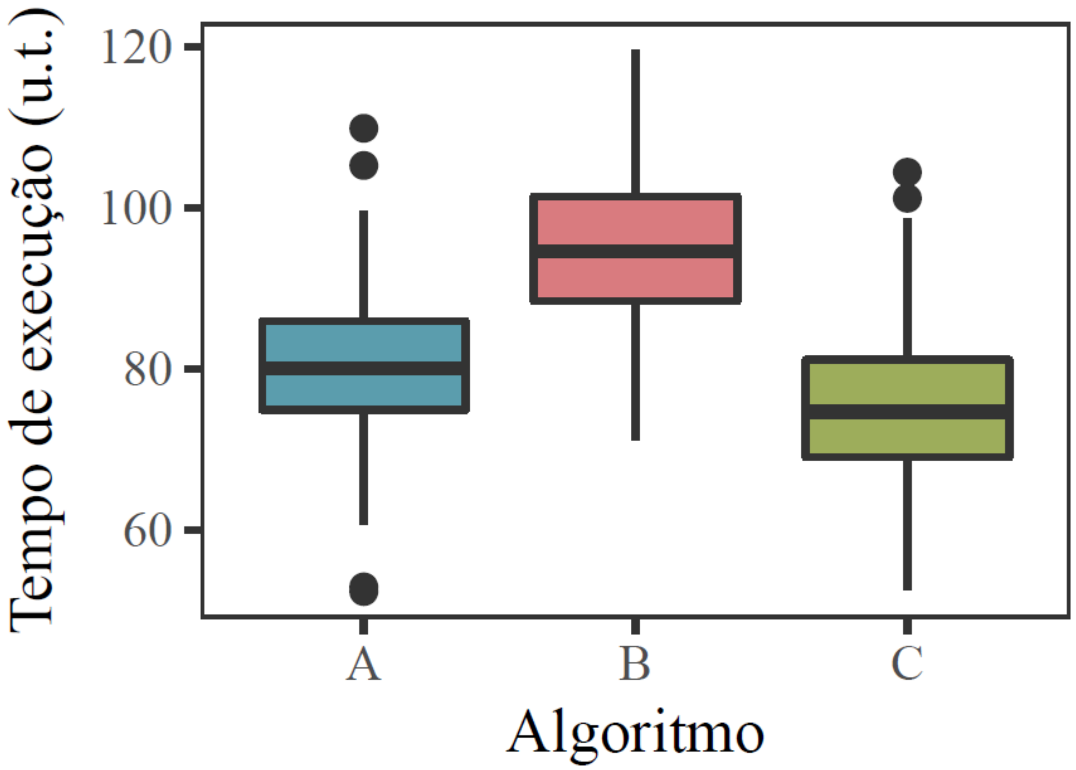
A ANOVA é complementada pelo teste de Tukey, que verifica quais grupos apresentam médias iguais.

As suposições são as mesmas da ANOVA e ele somente deve ser aplicado quando rejeitamos a hipótese nula da ANOVA.

# One-Way ANOVA

- **Exemplo**
  - Temos tempos de 200 execuções de três algoritmos, A, B e C.

  Após utilizar o teste de normalidade, queremos verificar se algum dos algoritmos é mais rápido que os demais.



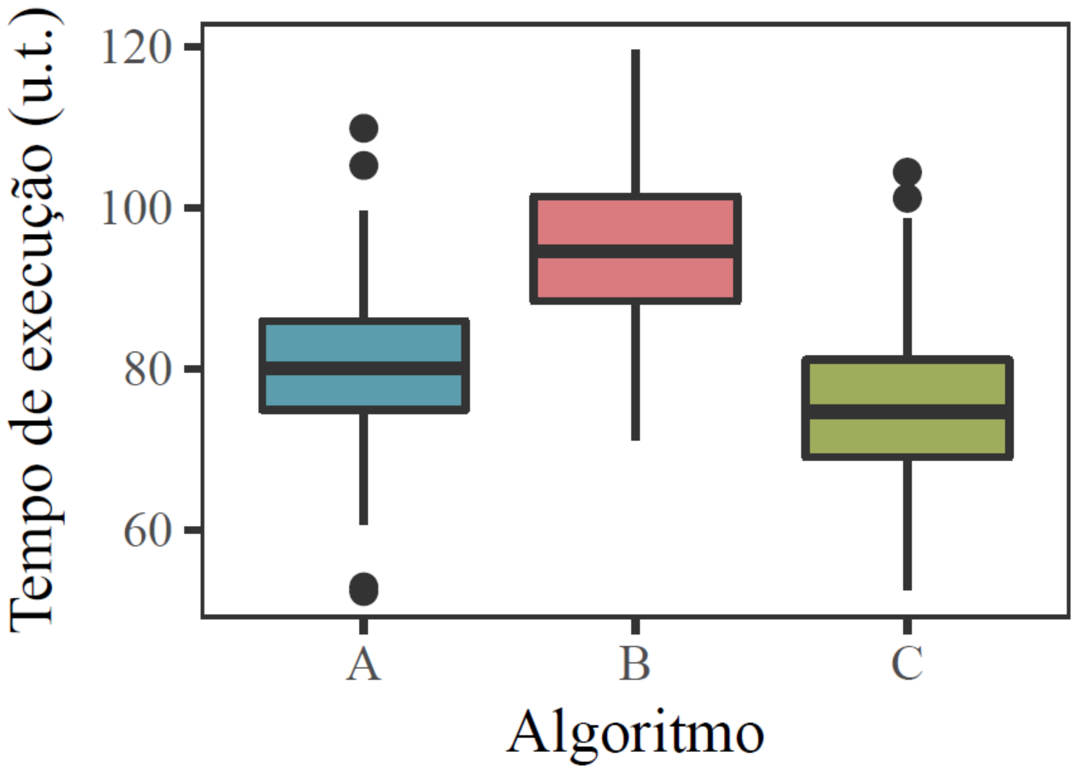
# One-Way ANOVA

- **Exemplo**

A:
One-Sample Kolmogorov-Smirnov test
$D \cong 0.05$, p-value = 0.7912

B:
One-Sample Kolmogorov-Smirnov test
$D \cong 0.03$, p-value = 0.9922

C:
One-Sample Kolmogorov-Smirnov test
$D \cong 0.04$, p-value = 0.7489



# One-Way ANOVA

- **Exemplo**
  - Fazemos um teste para verificar
    - $H_0$: $\mu_A = \mu_B = \mu_C$
    - $H_1$: Pelo menos uma média é diferente das demais

# One-Way ANOVA

- **Exemplo**
  - Fazemos um teste para verificar
    - $H_0$: $\mu_A = \mu_B = \mu_C$
    - $H_1$: Pelo menos uma média é diferente das demais
  - Antes, verificamos a homogeneidade das variâncias
  - E após, complementamos a ANOVA, se for o caso.

# One-Way ANOVA

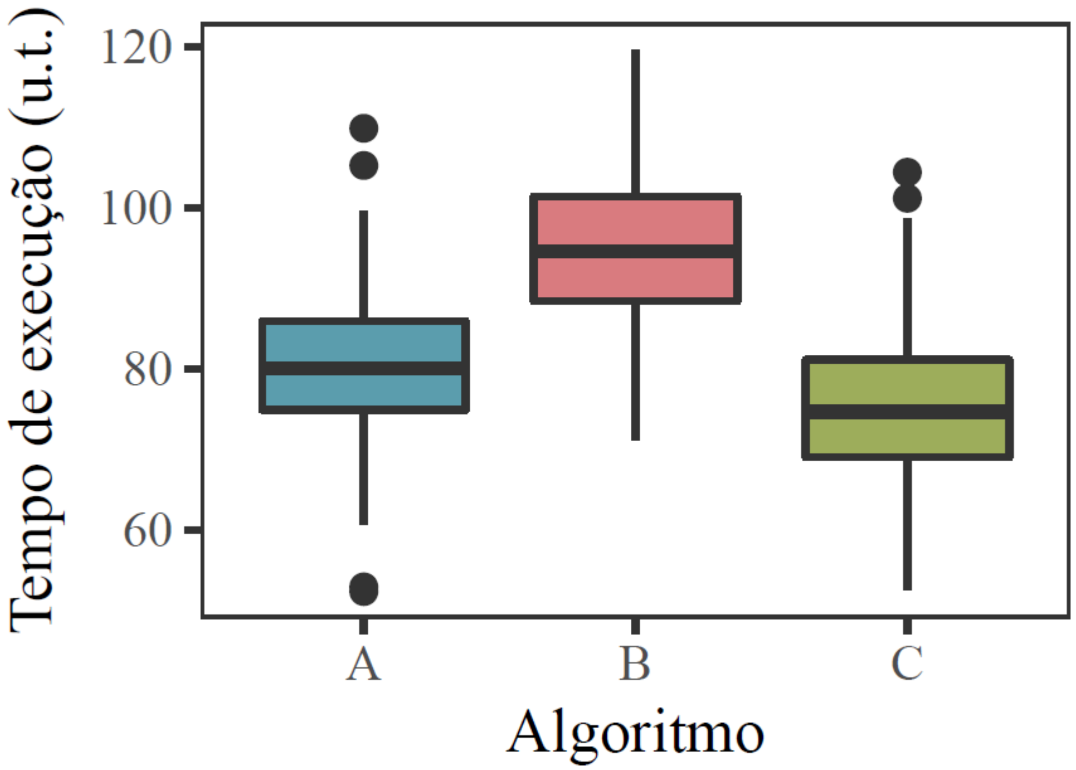
- **Exemplo**

Levene’s test for Homogeneity of Variance
F = 0.4316, p-value = 0.6497

Summary(ANOVA)
F = 214.3 p-value < 0.001

HSD Test

| Algoritmo | Média | Grupo |
|---|---|---|
| B | 94.89 | a |
| A | 80.43 | b |
| C | 75.27 | c |



# INTERVALOS DE CONFIANÇA

# Intervalos de Confiança

- **Ideias**
  - Os testes de hipóteses apenas rejeitam ou aceitam as hipóteses com base em $H_0$

# Intervalos de Confiança

- **Ideias**
  - Os testes de hipóteses apenas rejeitam ou aceitam as hipóteses com base em $H_0$
  - Algumas vezes pode ser do nosso interesse ter uma medida de magnitude:
    - 80 é diferente de 85; e
    - 80 é diferente de 160

# Intervalos de Confiança

- **Ideias**
  - Sob o enfoque frequentista, as estatísticas, enquanto funções da amostra, são variáveis aleatórias.

# Intervalos de Confiança

- **Ideias**
  - Por exemplo, sob algumas hipóteses sobre os dados, temos que:

$$\mathbb{P}\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

# Intervalos de Confiança

- **Ideia**
  - Após observar a amostra, o intervalo deixa de ser uma quantia aleatória e por isso dizemos que ele é um intervalo de $(1-\alpha)\%$ de confiança para $\mu$

# Intervalos de Confiança

- **Exemplo – diferença de médias**
  – No exemplo para comparação de duas médias, o próprio R fornece o intervalo de confiança para aquelas amostras

# Intervalos de Confiança

- **Exemplo – diferença de médias**
  - No exemplo para comparação de duas médias, o próprio R fornece o intervalo de confiança para aquelas amostras

Two Sample t-test

t = -14.943, p-value < 0.001

$H_1\colon \mu_\theta - \mu_\psi \neq 0$

**95% IC: $(-16.37; -12.56)$**

Sample estimates:

Theta: 94.89

Psi: 80.43

# Intervalos de Confiança

- **Exemplo – homogeneidade de variâncias**
  - O teste F também nos dá um intervalo de confiança.
    - Agora nosso interesse é saber se o valor 1 pertence ao intervalo

# Intervalos de Confiança

- **Exemplo – homogeneidade de variâncias**
  - O teste F também nos dá um intervalo de confiança.
    - Agora nosso interesse é saber se o valor 1 pertence ao intervalo

F test to compare two variances

F = 0.92, p-value =0.5582

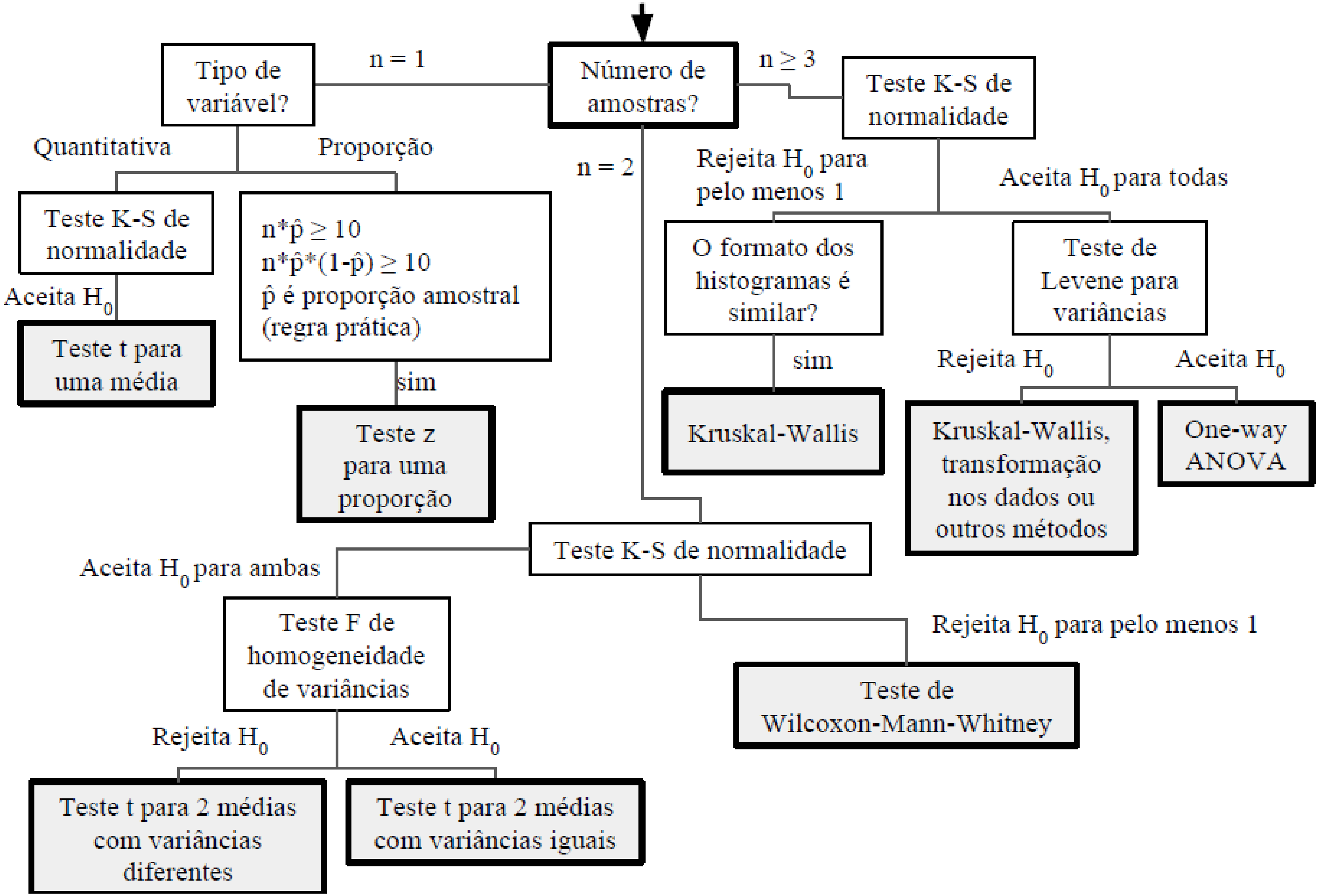
$H_1: \sigma_{\psi}^2/\sigma_{\theta}^2 \neq 1$

95% IC: $(0.6964; 1.2160)$

Ratio of variances: 0.92

# FECHAMENTO

# Fechamento



Número de amostras?
n = 1
Tipo de variável?
Quantitativa
Proporção
Teste K-S de normalidade
Aceita $H_0$
Teste t para uma média
$n*\hat{p} \geq 10$
$n*\hat{p}*(1-\hat{p}) \geq 10$
$\hat{p}$ é proporção amostral
(regra prática)
sim
Teste z para uma proporção
n ≥ 3
Teste K-S de normalidade
Rejeita $H_0$ para pelo menos 1
Aceita $H_0$ para todas
O formato dos histogramas é similar?
sim
Kruskal-Wallis
Teste de Levene para variâncias
Rejeita $H_0$
Aceita $H_0$
Kruskal-Wallis, transformação nos dados ou outros métodos
One-way ANOVA
n = 2
Teste K-S de normalidade
Aceita $H_0$ para ambas
Teste F de homogeneidade de variâncias
Rejeita $H_0$
Aceita $H_0$
Teste t para 2 médias com variâncias diferentes
Teste t para 2 médias com variâncias iguais
Rejeita $H_0$ para pelo menos 1
Teste de Wilcoxon-Mann-Whitney

# XVII Escola Regional de Alto Desempenho do Estado do Rio Grande do Sul

**Muito obrigado.**

Aishameriane Venes Schmidt

Francieli Zanon Boito

Laércio Lima Pilla