

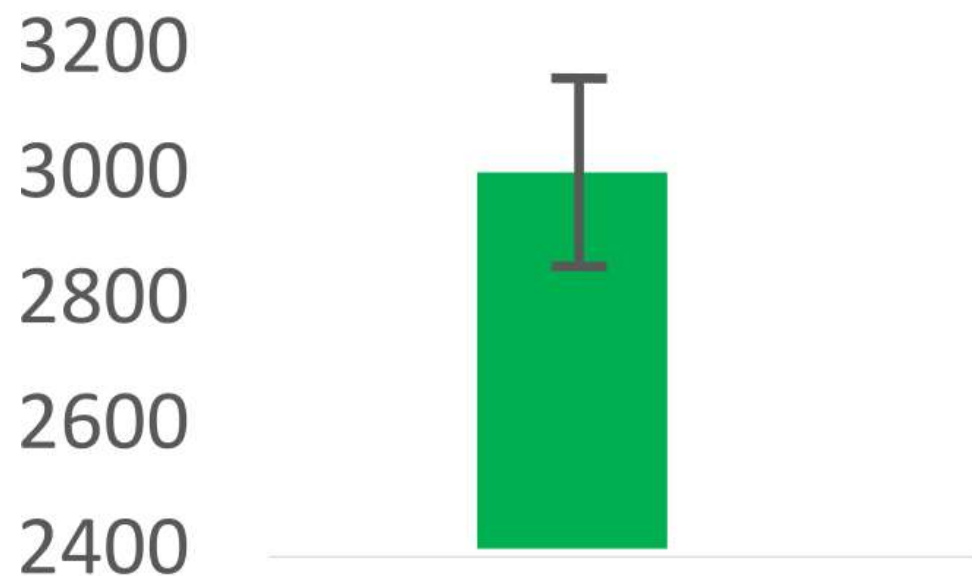
# **ESTATÍSTICA PARA SAÚDE COLETIVA**

## **Aula 6**

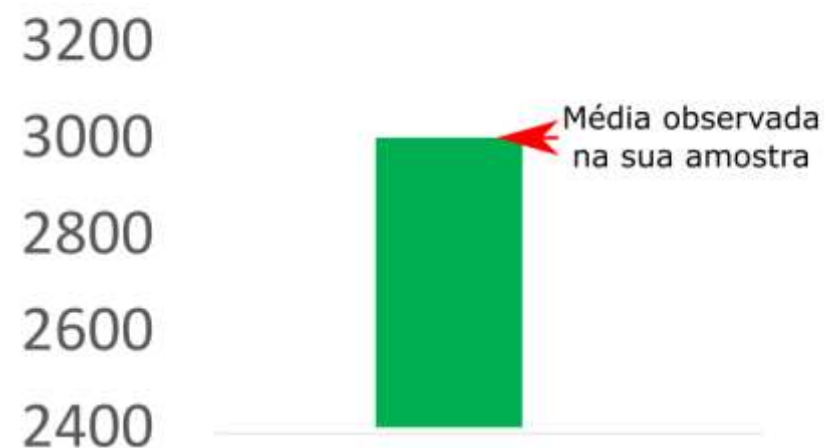
Lista de pessoas que não definiram artigo  
para apresentação

# Revisão ultima aula: Intervalo de confiança

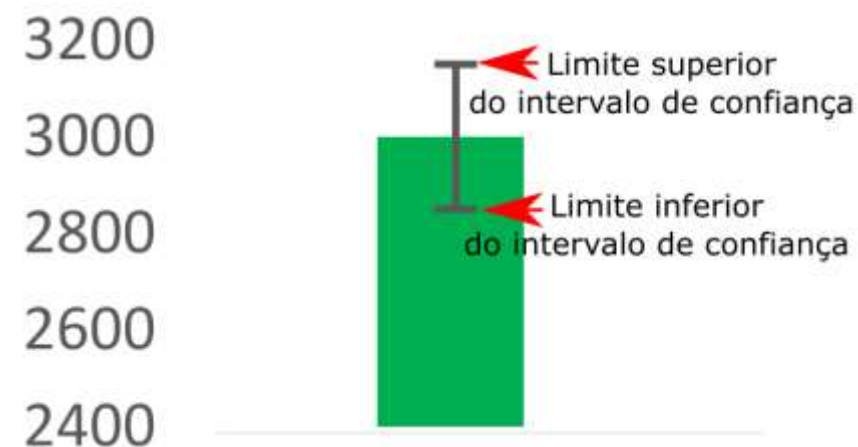
Peso dos bebês (g)



Peso dos bebês (g)

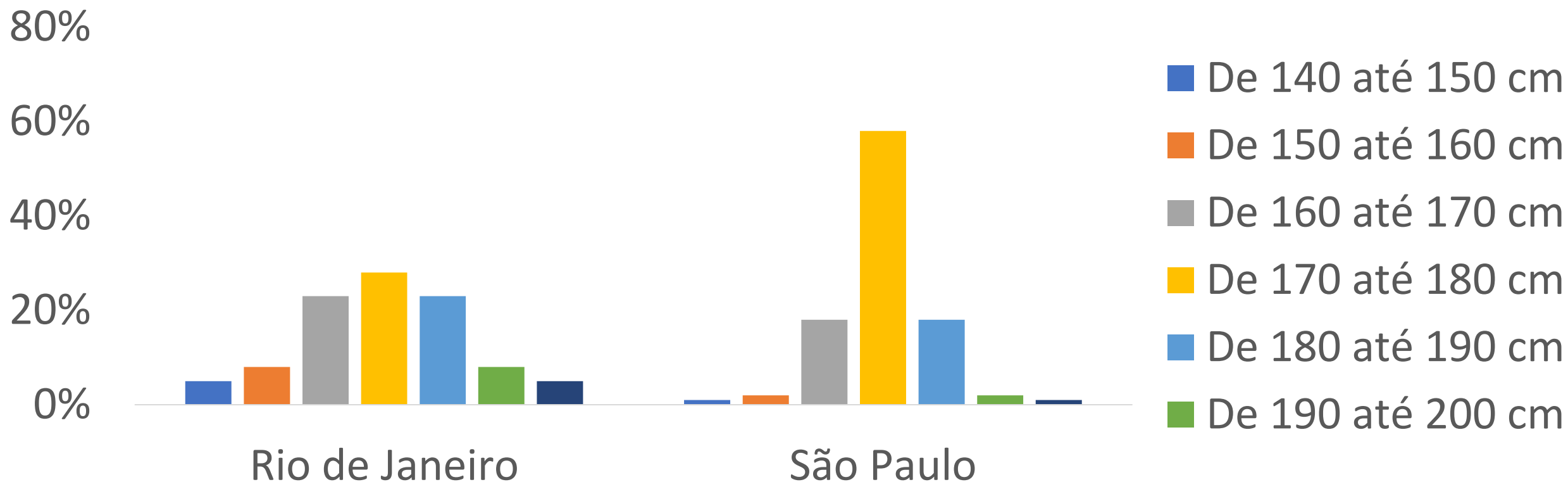


Peso dos bebês (g)

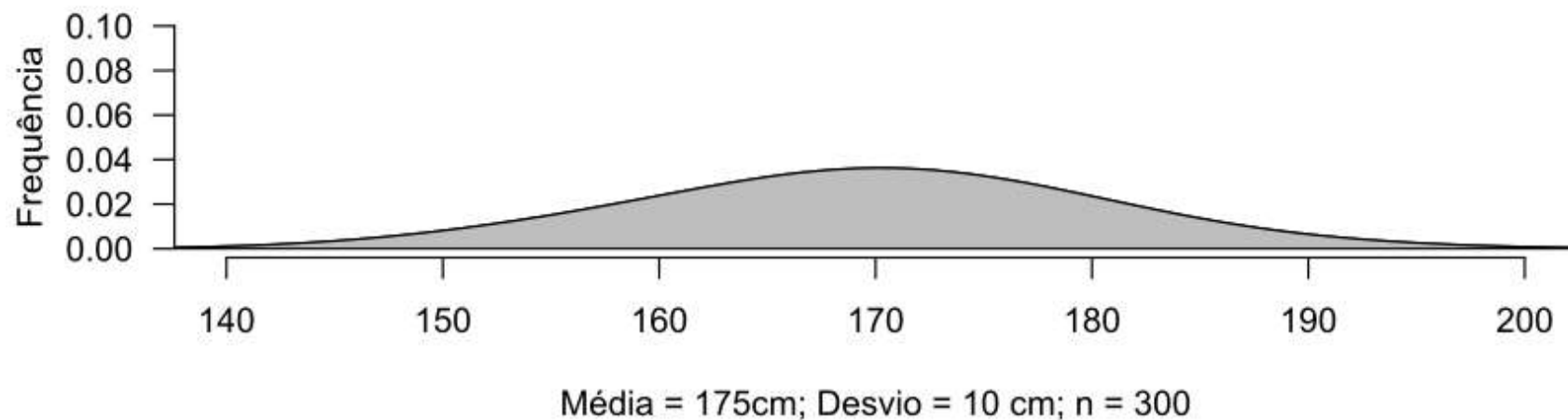
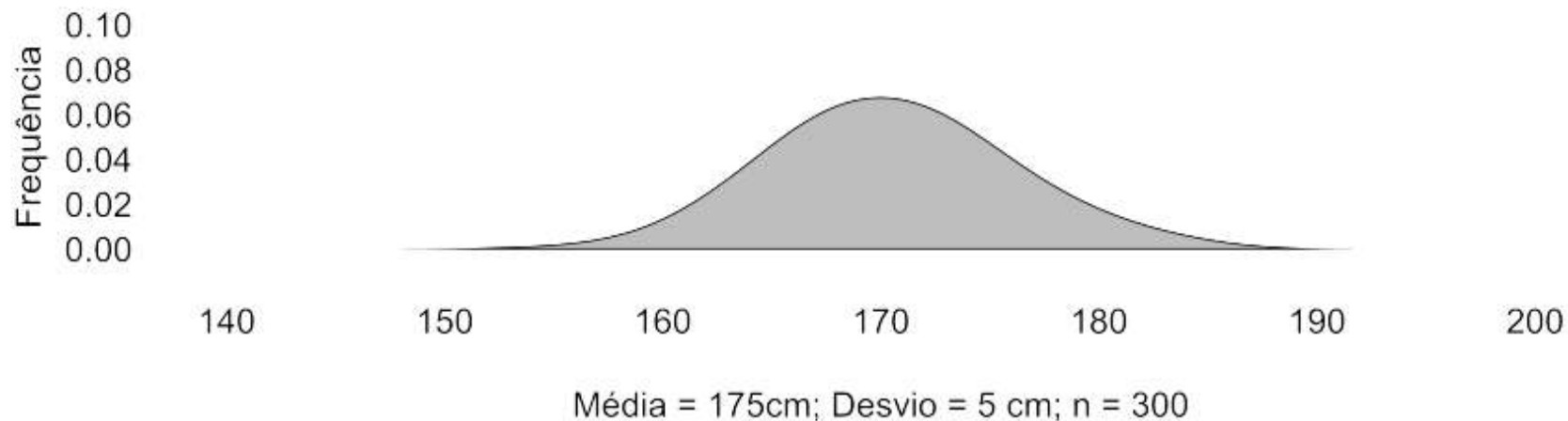


# Revisão ultima aula: variância/desvio-padrão

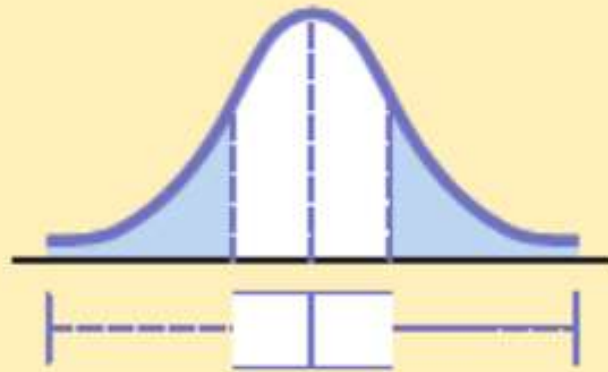
Altura de 1 milhão de pessoas amostradas nas cidades  
do Rio de Janeiro e São Paulo



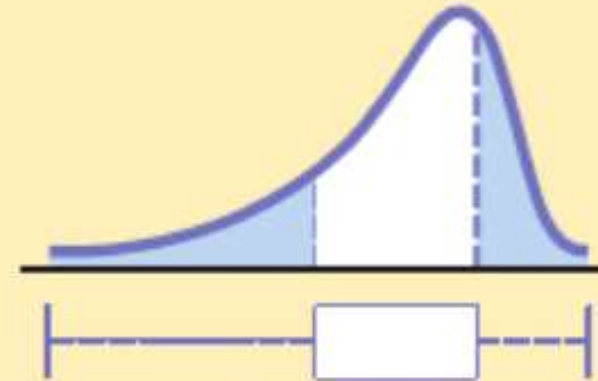
# Revisão ultima aula: variância/ desvio-padrão



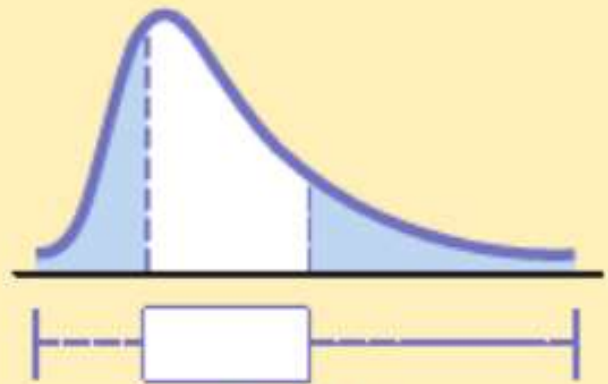
# Revisão ultima aula: Formato de curvas de distribuição



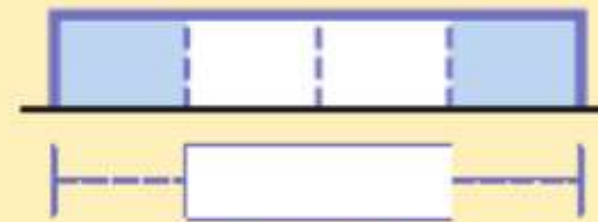
**Simétrico**



**Assimétrico à esquerda**

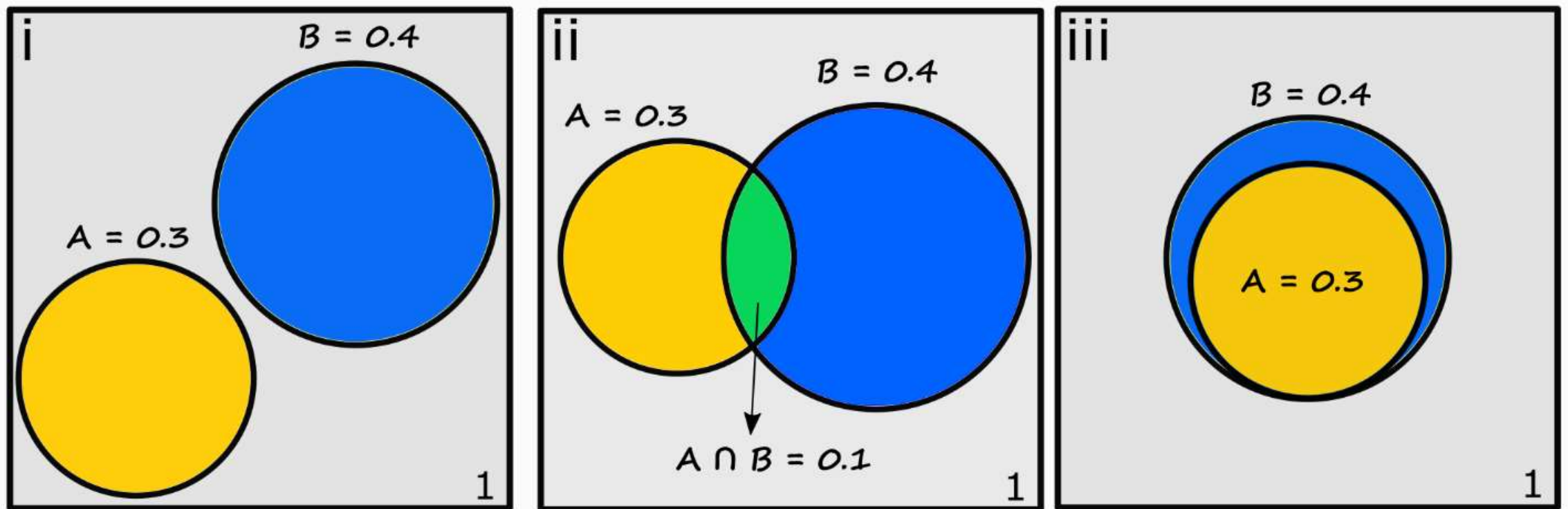


**Assimétrico à direita**



**Distribuição retangular**

# Revisão ultima aula: probabilidades



Representação de relações de probabilidades expressas por diagramas de Venn. Cada evento é representado por um círculo, com área proporcional a sua probabilidade de ocorrência 0.3 (A) e 0.4 (B). As relações podem ser: independência (i), sobreposição parcial (ii) ou total (iii).

# Exemplo

Revisão ultima aula: probabilidades

	Medicamento		Total
	Sim	Não	
Sobreviveu	8	7	15
Faleceu	50	35	85
Total	58	42	100

Probabilidade de morte =  $85/100 = 85\%$

Probabilidade de morte entre os que tomaram remédio =  $50/58 = 86\%$

Probabilidade de morte entre os que não tomaram remédio =  $35/42 = 83\%$



# Aula de hoje

- Fundamentos do teste de hipótese
- Inferência sobre normalidade dos dados
- Inferência sobre variâncias dos dados

# Fundamentos do teste de hipótese

# Nota

- Essa pergunta está intimamente ligada com algumas questões das atividades finais do curso:
- Avaliação do artigo: “Questão 2. Descreva o objetivo. Caso exista(m), a(s) hipótese(s) testada(s).”
- Proposta de pesquisa: hipóteses

O que é uma hipótese?

# O que é uma hipótese?

- Hipótese é a suposição de algo
- Não sabe se é verdade ou não

# Em estatística

- Afirmação feita sobre o parâmetro de uma população.
- Suposição que pode ser testada com base em dados observados.
- Suas hipóteses estatística obrigatoriamente vão envolver parâmetros populacionais!

# Em estatística

- Toda pergunta pode ser traduzida em :
  - Hipótese nula
  - Hipótese alternativa
- Hipótese nula
  - Geralmente associada que nada é diferente de nada, ou
  - Nada tem relação com nada
  - No dia-dia é a que corresponde ao *status quo*
  - Identificada pelo símbolo  $H_0$
- Hipótese alternativa
  - Oposto da hipótese nula

# Exemplos

- No Brasil, nascem mais meninos ou meninas?
  - A razão sexual de bebês nascidos no Brasil **é igual** à 1:1
  - A razão sexual de bebês nascidos no Brasil **é diferente** de 1:1
- A quantidade de álcool-gel escrita na embalagem de uma marca X está correta?
  - A quantidade média de álcool que entregue **é igual** à 500 ml
  - A quantidade média de álcool que entregue **é diferente** de 500 ml
- Homens são mais altos que mulheres?
  - A altura média de homens **é igual** a altura média de mulheres
  - A altura média de homens **é diferente** a altura média de mulheres



# Nunca esqueça

- Em estatística, quando formulamos hipóteses estamos interessados em inferir padrões ou processos que ocorrer a nível de população



# Pontos chave

- As hipóteses sempre são formuladas referindo-se a parâmetros da população e não da amostra
  - Você sempre sabe tudo que aconteceu na sua amostra (porque você tem os dados na mão). Não faria sentido criar hipóteses para **supor** parâmetros de alguma coisa você já conhece.
- A declaração correspondente à **hipótese nula** sempre contém um  **sinal de igualdade** com relação ao valor especificado para o parâmetro populacional
- A declaração correspondente à **hipótese alternativa** sempre contém um  **sinal de diferença** com relação ao valor especificado para o parâmetro populacional

# Pontos chave

- Se você está comparando coisas (ex altura entre pessoas de 2 grupos), a hipótese nula sempre pressupõe os parâmetros são iguais (ou seja, nada é diferente de nada)
- A hipótese alternativa é o oposto da hipótese nula
- **Só existem evidências de que a hipótese alternativa é verdadeira, quando você rejeita a hipótese nula**
- **Se você não rejeitar a hipótese nula, você não conseguiu comprovar a hipótese alternativa**
- Frequentemente a hipótese nula representa o *status quo* (aquilo que se acredita no momento)

# Exemplo relacionado ao *status quo*

- Você deseja fazer uma pesquisa para saber se o tempo de espera no atendimento em um hospital aumentou no ultimo mês. Você tem dados de tempo de espera de até 5 anos para trás. Esse histórico indica que o tempo médio de espera era de 10 min.
- Nesse caso, suas hipóteses seriam:
  - Hipótese nula ( $H_0$ ) = O tempo de espera **é igual** à 10 min
  - Hipótese alternativa ( $H_1$ ) = O tempo de espera **é diferente de** 10 min

# Hipótese nula vs. alternativa

- Na estatística frequentista , nós usamos técnicas para avaliar a possibilidade da hipótese nula ser falsa
  - Nota: existem outras abordagens (menos populares) como a estatística bayseana
- Hipótese alternativa representa a conclusão a que se chega se a hipótese nula for rejeitada
- Ou seja, a hipótese alternativa é aquela que você assume como verdade quando não existe evidência suficiente para acreditar que a hipótese nula seja verdadeira

# Como funciona?

1. Você aplica um teste
2. Observe o valor de  $p$
3. Interprete o resultado
  - **Se o valor de  $p$  calculado for maior que o valor que o nível de significância** (margem que você assumiu de erro): você assume que  $H_0$  é verdadeira.
  - **Se o valor de  $p$  calculado for menor que o valor que o nível de significância** (margem que você assumiu de erro): você rejeita  $H_0$  e assume que  $H_1$  é verdadeiro.

# Como funciona?

1. Você aplica um teste
2. Observe o valor de  $p$
3. Interprete o resultado
  - **Se o valor de  $p$  calculado for maior que o valor que o nível de significância** (margem que você assumiu de erro): você assume que  $H_0$  é verdadeira.
  - **Se o valor de  $p$  calculado for menor que o valor que o nível de significância** (margem que você assumiu de erro): você rejeita  $H_0$  e assume que  $H_1$  é verdadeiro.

# Exemplo

- Suponha um estudo que avaliou a diferença de altura entre meninas e meninos.
- Por meio de um teste-t (método que vai ser apresentado da próxima aula), será calculada uma probabilidade de as alturas dos grupos comparados (meninos e meninas) ser diferente.
- Exemplos de possíveis resultados :



**p calculado = 0.80**

- Como 0.80 é maior que 0.05, conclusão:
- **Não existe evidência** de que existe diferença entre as médias comparadas. Portanto, a **altura média dos grupos é igual uma da outra.**

**p calculado = 0.03**

- Como 0.03 é menor que 0.05, conclusão:
- **Existe evidência** de que existe diferença entre as médias comparadas. Portanto, a **altura média dos grupos é diferente uma da outra.**



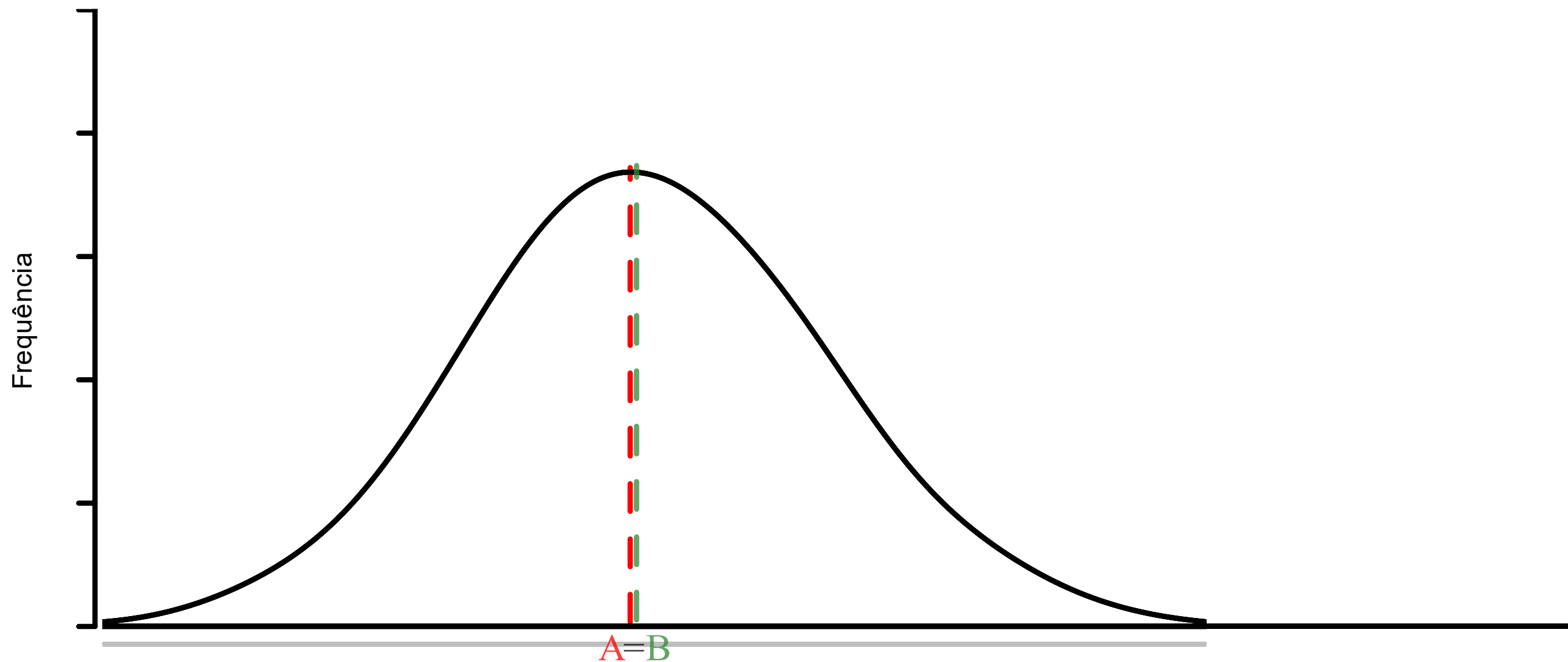
# Erro tipo I e Erro tipo II

<div>O que você assumiu como verdade</div> <div>O que de é fato</div>	Hipotese nula é verdadeira	Hipotese alternativa é verdadeira
Hipótese nula verdadeira	<b>Tudo certo</b>	<b>Erro tipo I</b>
Hipótese alternativa verdadeira	<b>Erro tipo II</b>	<b>Tudo certo</b>

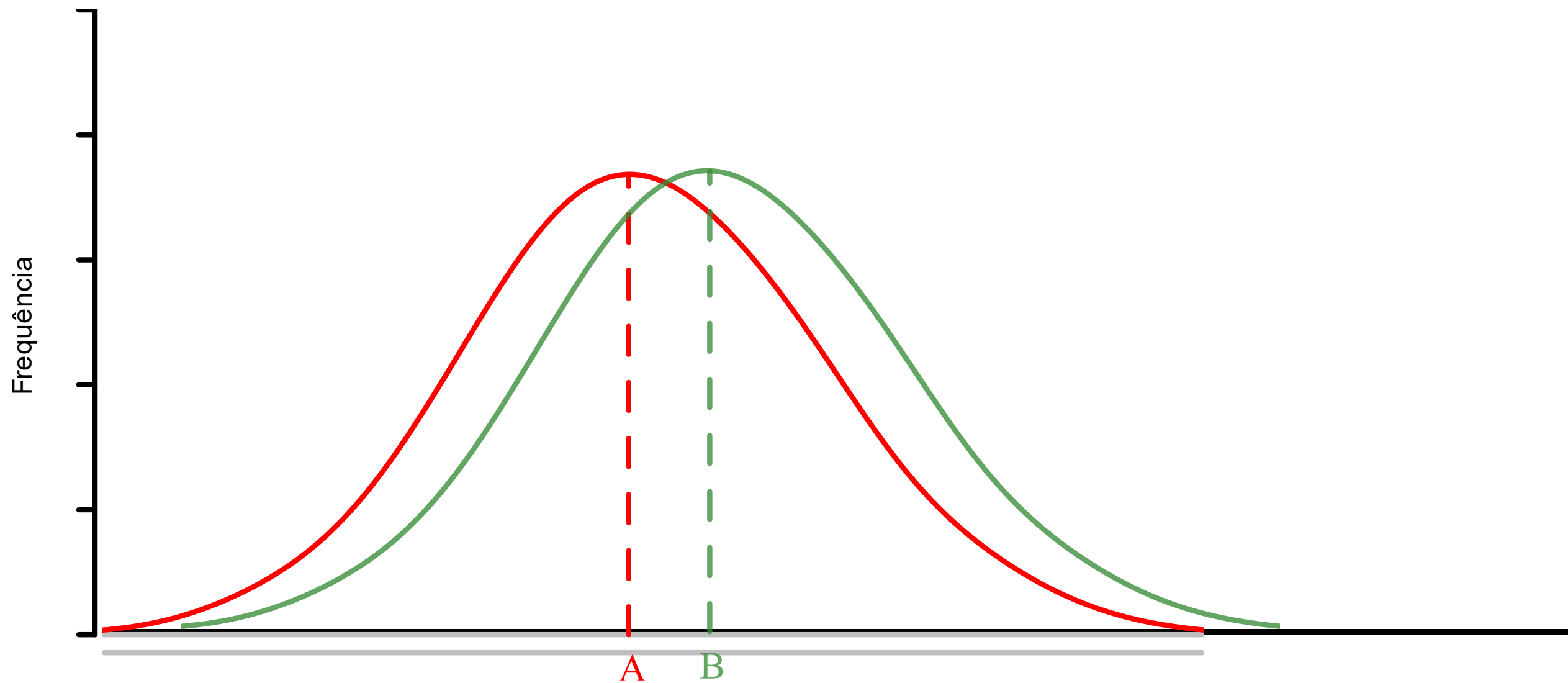
# Erro tipo I e Erro tipo II

O que você assumiu como verdade O que de é fato	Hipotese nula é verdadeira	Hipotese alternativa é verdadeira
Hipótese nula verdadeira	<b>Tudo certo</b>	Tradicionalmente utiliza-se 5%
Hipótese alternativa verdadeira	Depende de: <ul style="list-style-type: none"><li>• alfa</li><li>• tamanho amostral</li><li>• variância</li></ul>	<b>Tudo certo</b>

De onde vem o erro tipo 2?



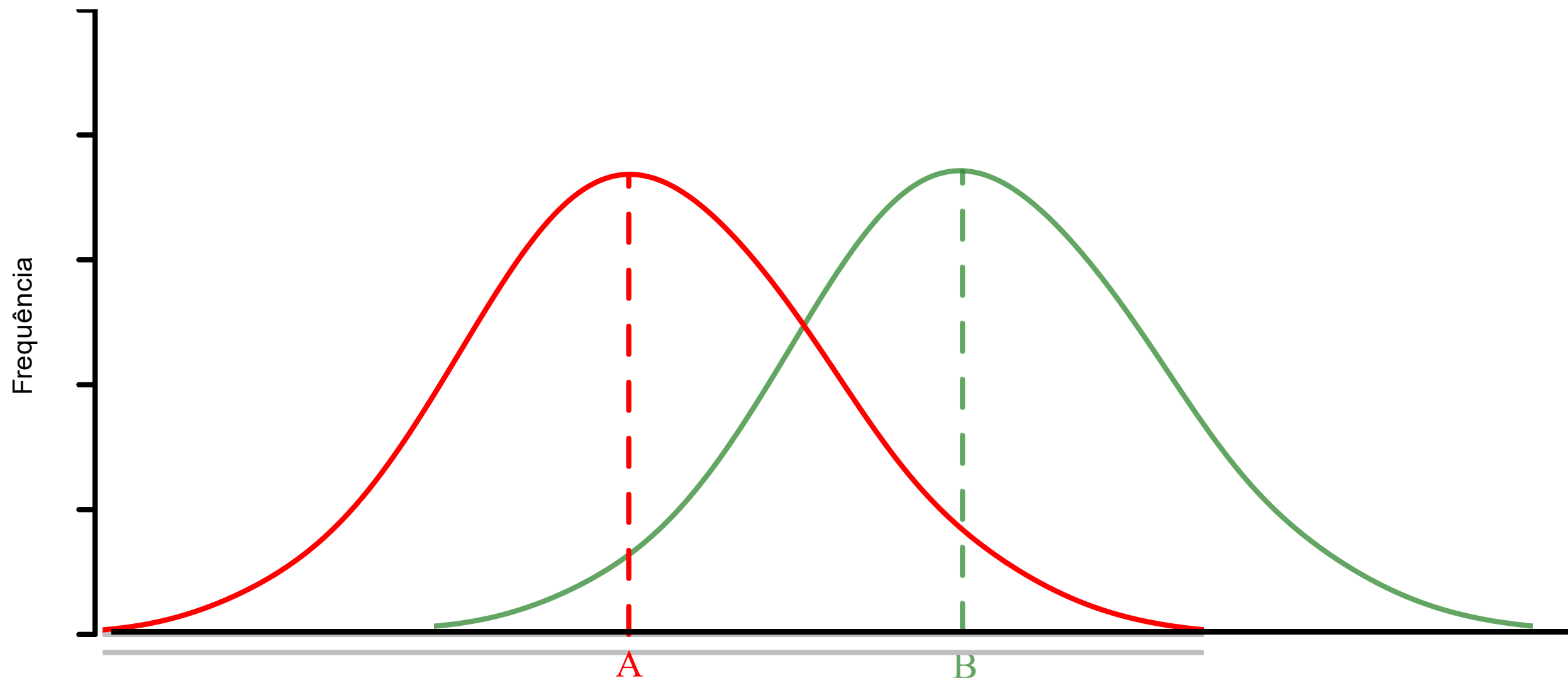
De onde vem o erro tipo 2?



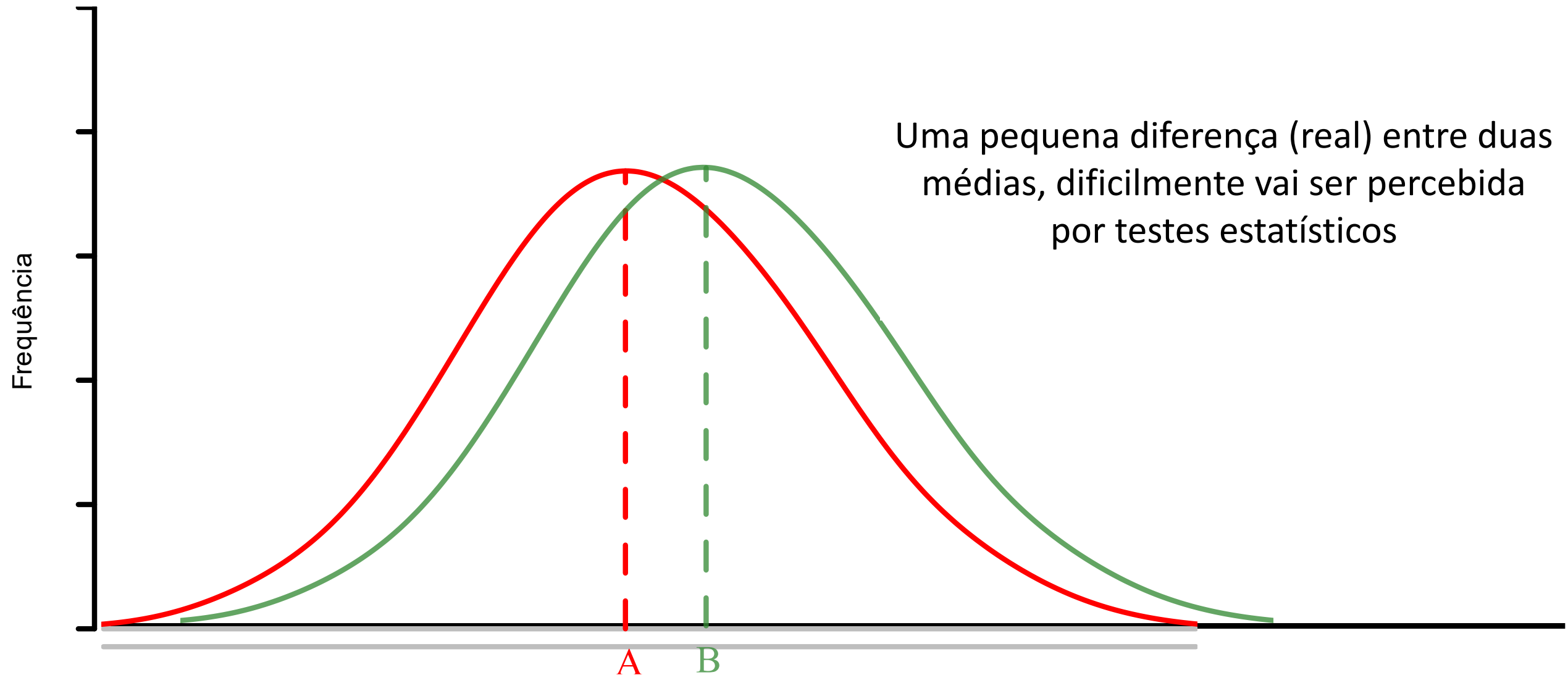
De onde vem o erro tipo 2?



De onde vem o erro tipo 2?



# De onde vem o erro tipo 2?



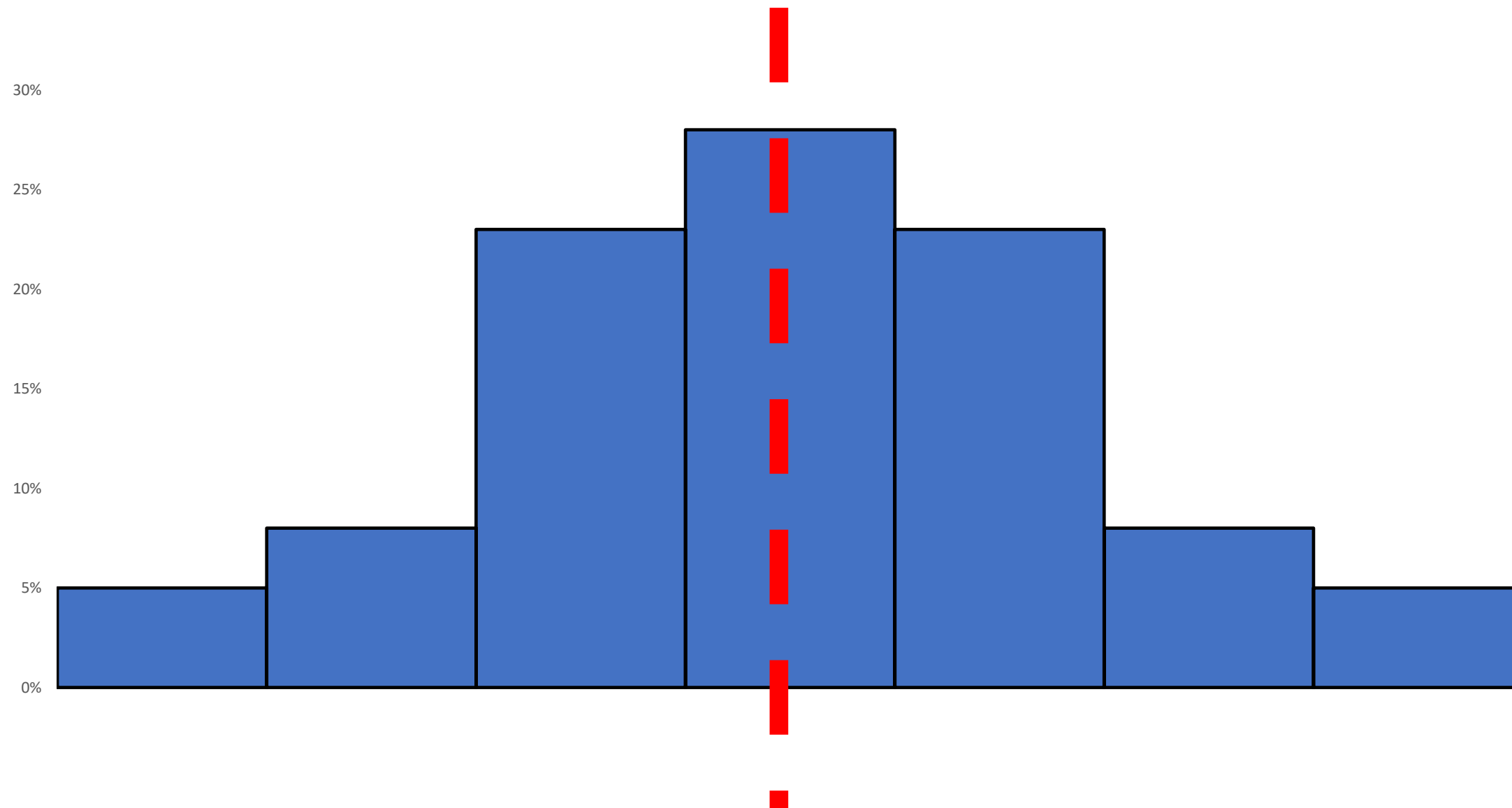
# Como testar hipóteses? Métodos que serão vistos a partir da semana que vem.

- Você está querendo estimar um parâmetro?
  - Ex. A razão sexual de bebês é **igual** à 1:1 (0.5)
  - Opção 1: Estima qual é o parâmetro, e avalia o seu parâmetro esperado está dentro do intervalo de confiança (assunto da ultima aula)
  - Opção 2: teste-t para 1 amostra (próxima aula)
- Você quer comparar médias?
  - Estima o parâmetros nos grupos e compara se o valor de um é diferente do outro (próximas duas aulas)
- Correlações, regressões, etc:
  - Assuntos para o mês de outubro

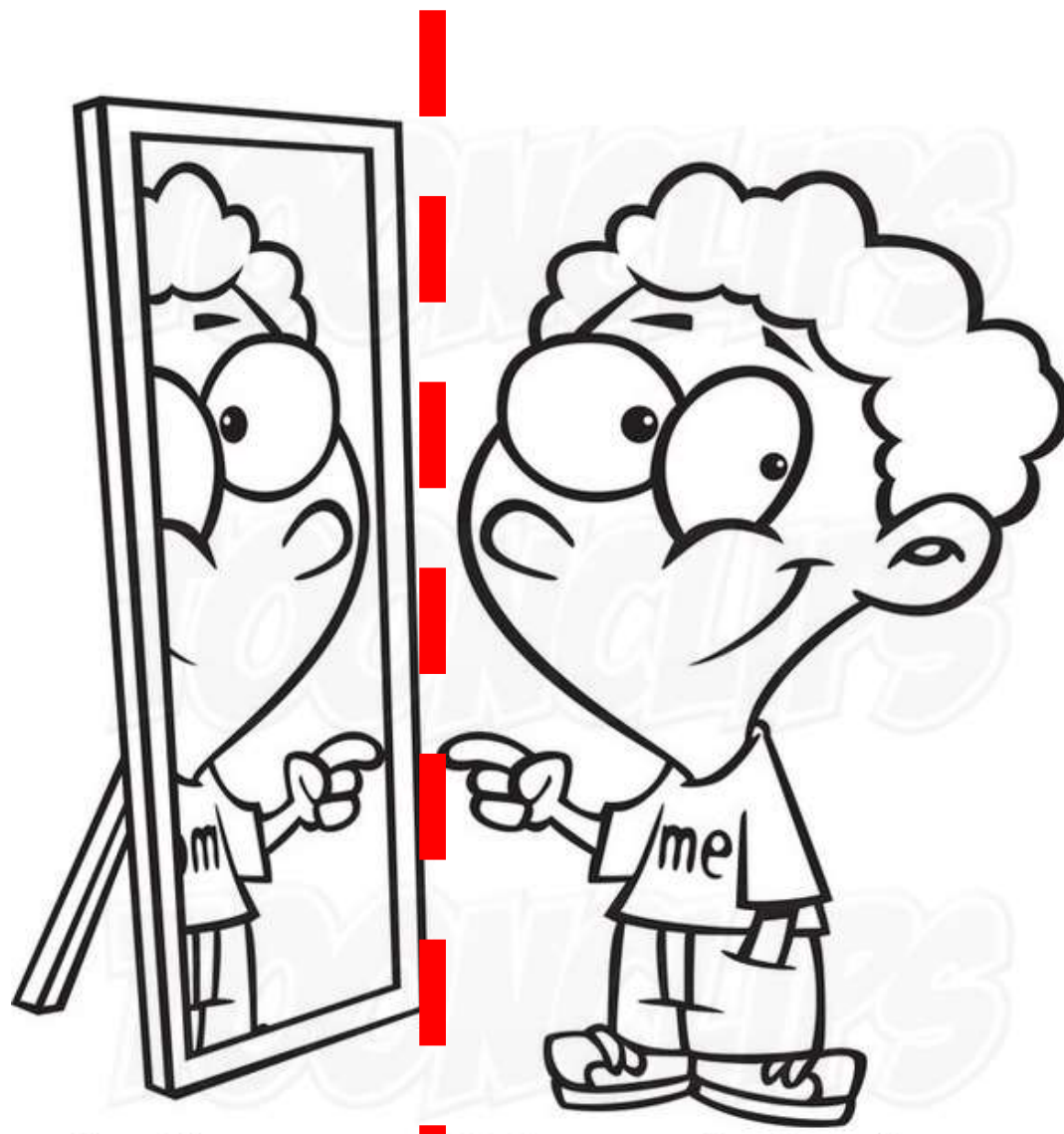


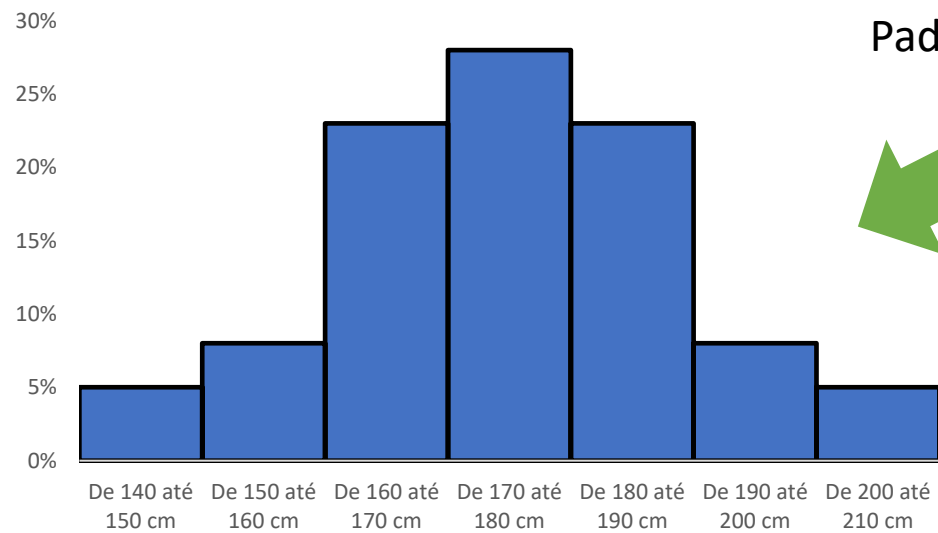
Um determinado conjunto de dados tem distribuição normal?

# Simetria dos dados



# Simetria dos dados simétricos

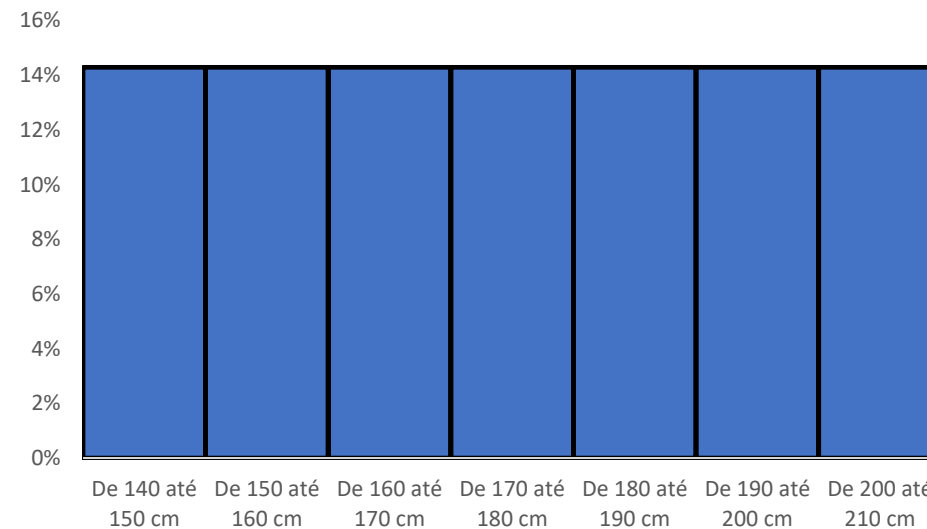
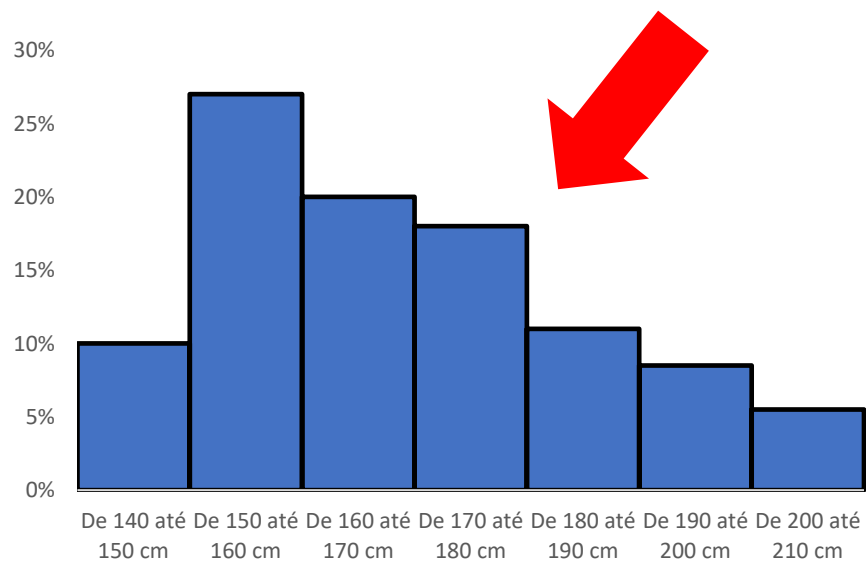




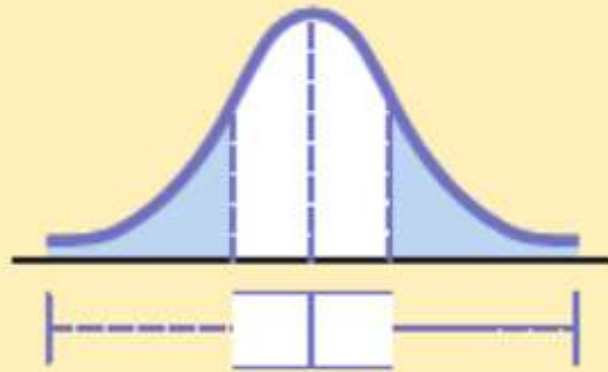
Padrão que sugere que os dados apresentam distribuição normal



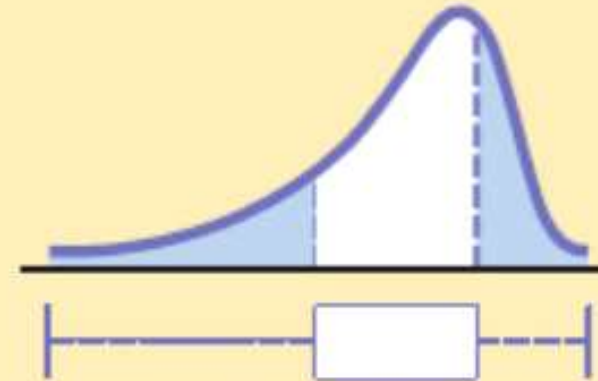
Padrão que sugere que os dados apresentam distribuição não-normal



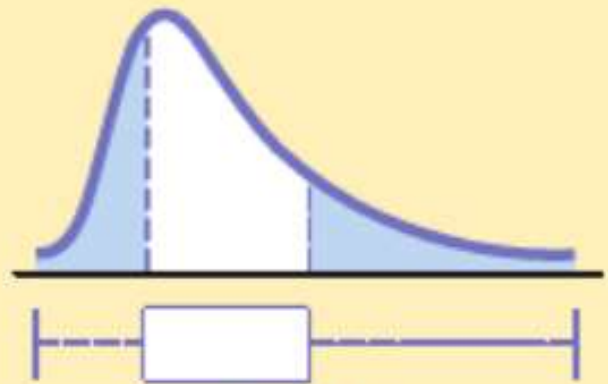
# Um determinado conjunto de dados tem distribuição normal?



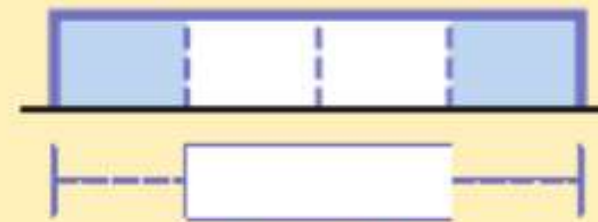
**Simétrico**



**Assimétrico à esquerda**



**Assimétrico à direita**



**Distribuição retangular**

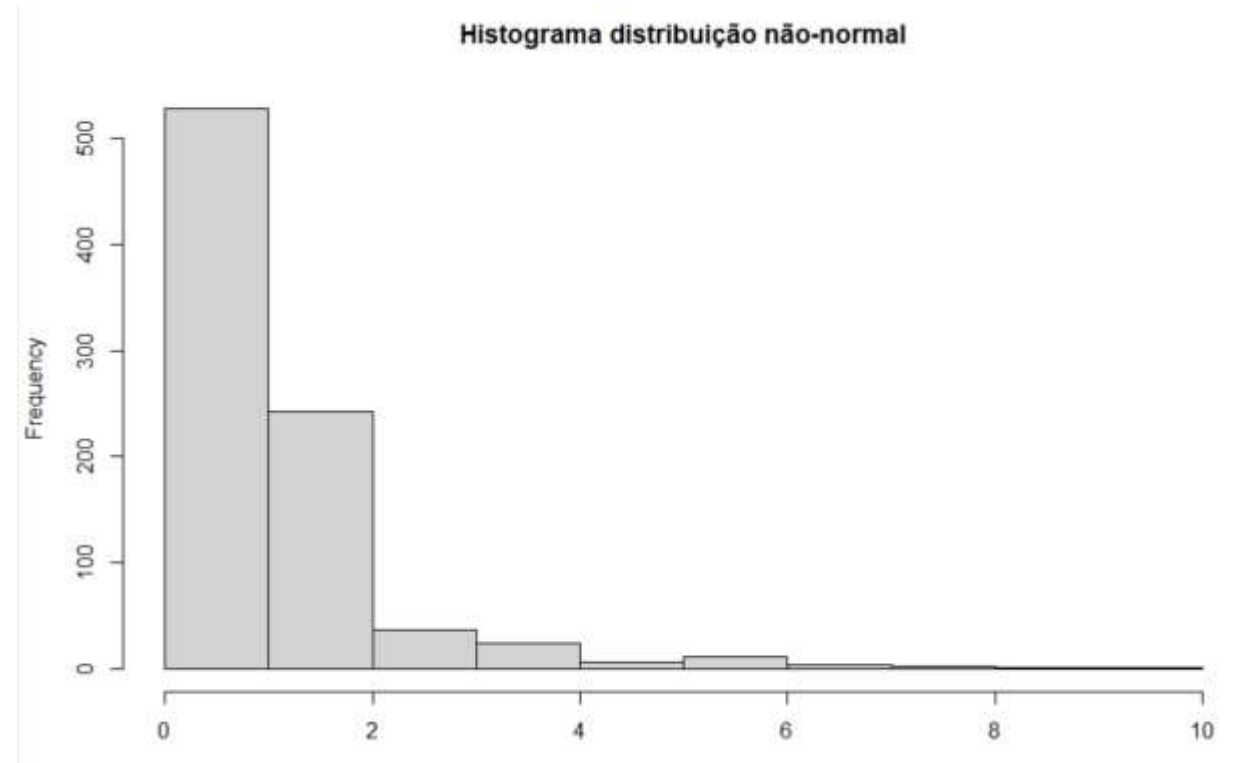
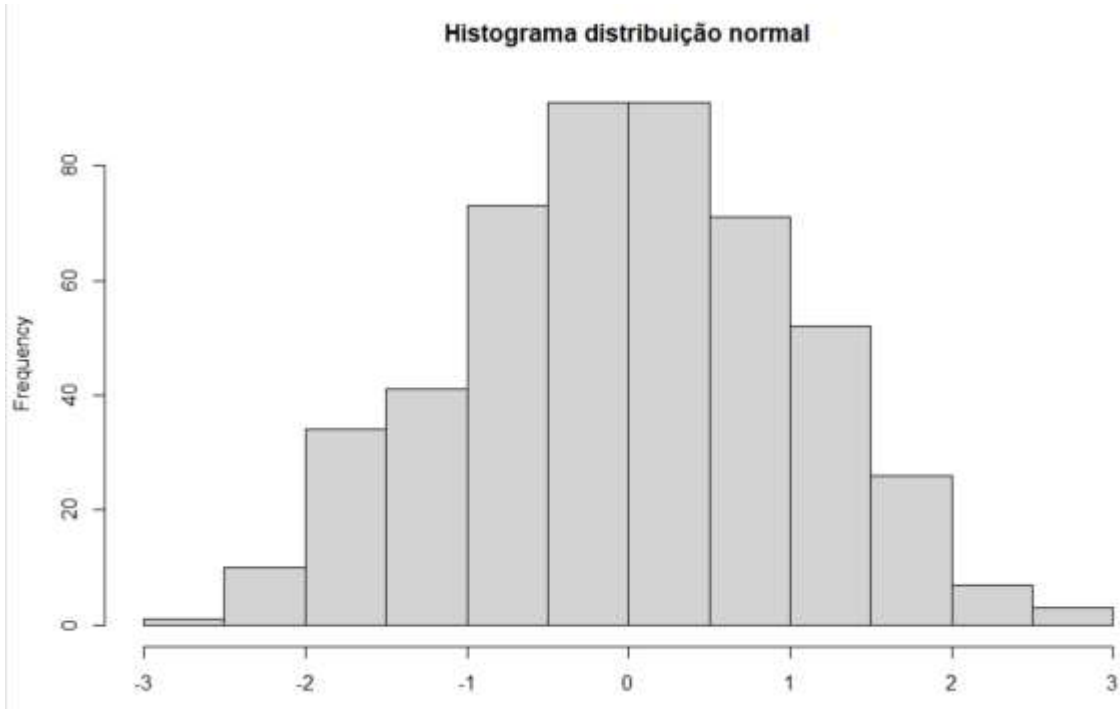
# Teste de normalidade

- Existem varias formas de avaliar normalidade nos dados
- Opção 1: inspeção visual (histogramas)
- Opção 2: testes estatístico (nosso primeiro exemplo de teste de hipótese)
  - Shapiro-Wilk
  - Kolmogorov-Smirnov

# Informações importantes sobre normalidade

- Muitos autores argumentam que a normalidade não é um problema serio (Sokal & Rohlf 1995; Zar 1999).
- Alguns autores argumentam que se sua amostra for grande, você não precisa nem se preocupar com normalidade (Fitzmaurice et al 2004).
- A normalidade da variável estudada deve pode ser conferida fazendo um histograma.
- Se por acaso você tem poucas observações é melhor fazer o histograma dos resíduos (assunto de outra aula). Isso não implica em dizer que você testou a normalidade da população, mas é o melhor que você poderia fazer.
- Você deve se preocupar com a normalidade da sua variável preditora, e não da variável resposta (Pag 20, Zuur et al 2009 Mixed Effects Models and Extensions in Ecology with R)

# Opção 1: inspeção visual (histogramas) – minha recomendação





# Teste de Shapiro-Wilk

- O teste de Shapiro-wilk testa se os dados diferem de uma distribuição normal
  - $H_0$ : dados não diferem de uma distribuição normal
  - $H_1$ : dados diferem de uma distribuição normal
- Valores de  $p < 0,05$  indicam que os dados não são normais

Shapiro-wilk normality test

```
data: SeusDados  
W = 0.99309, p-value = 0.4718
```

Shapiro-wilk normality test

```
data: tabela1$EDV  
W = 0.9201, p-value = 2.709e-08
```

# Teste de Shapiro-Wilk

- O teste de Shapiro-wilk testa se os dados diferem de uma distribuição normal
  - $H_0$ : dados não diferem de uma distribuição normal
  - $H_1$ : dados diferem de uma distribuição normal
- Valores de  $p < 0,05$  indicam que os dados não são normais

Shapiro-wilk normality test

```
data: SeusDados  
W = 0.99309, p-value = 0.4718
```

Shapiro-wilk normality test

```
data: tabela1$Env  
W = 0.9201, p-value = 2.709e-08
```

Hipóteses:

$H_0$ : dados não diferem de uma distribuição normal

$H_1$ : dados diferem de uma distribuição normal

Interpretação do resultado: 0.47 é menor que 0.05?

- Não, portanto  $H_0$  é assumida como verdadeira

Shapiro-wilk normality test

data: SeusDados

$w = 0.99309$ ,  $p\text{-value} = 0.4718$

Hipóteses:

$H_0$ : dados não diferem de uma distribuição normal

$H_1$ : dados diferem de uma distribuição normal

OBS:  $2.709e-08 = 2.709 \times 10^{-8} = 0.00000002709$

Interpretação do resultado: 0.00000002709 é menor que 0.05?

- Sim, portanto rejeitamos  $H_0$ , e assumimos que  $H_1$  é verdadeira

Shapiro-Wilk normality test

data: tabela1\$Env  
W = 0.9201, p-value = 2.709e-08

# Smirnov-kolomogorov

- O teste de Smirnov-kolomogorov é usado para avaliar se dois conjuntos de dados apresentam uma distribuição semelhante
  - $H_0$ : os dois conjunto de dados apresentam distribuição semelhante
  - $H_1$ : os dois conjunto de dados não apresentam distribuição semelhante
- $p < 0,05$  indicam que os dados apresentam distribuições distintas
- Você pode comparar a distribuição dos seus dados com umas distribuição conhecida (ex. normal) e avaliar evidencias de que os dados não apresentam distribuição normal
- Teste apresentado no Script suplementar

Um determinado dois conjunto  
de dados tem variâncias  
semelhantes?

Isso é = homogeneidade de variâncias

# Comparar variâncias

- Existem varias formas de avaliar normalidade nos dados
- Opção 1: inspeção visual (boxplot)
- Opção 2: testes estatístico
  - F test

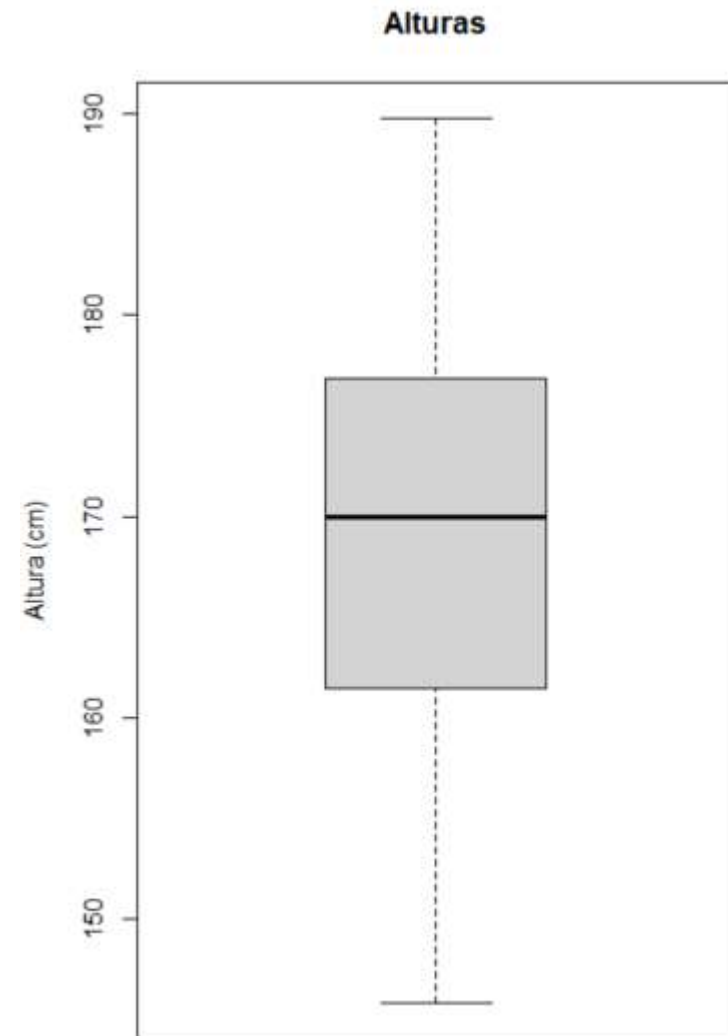
# Opção 1: inspeção visual (boxplot) – minha recomendação

- Você viu na aula 4 que um conjunto de dados pode ser



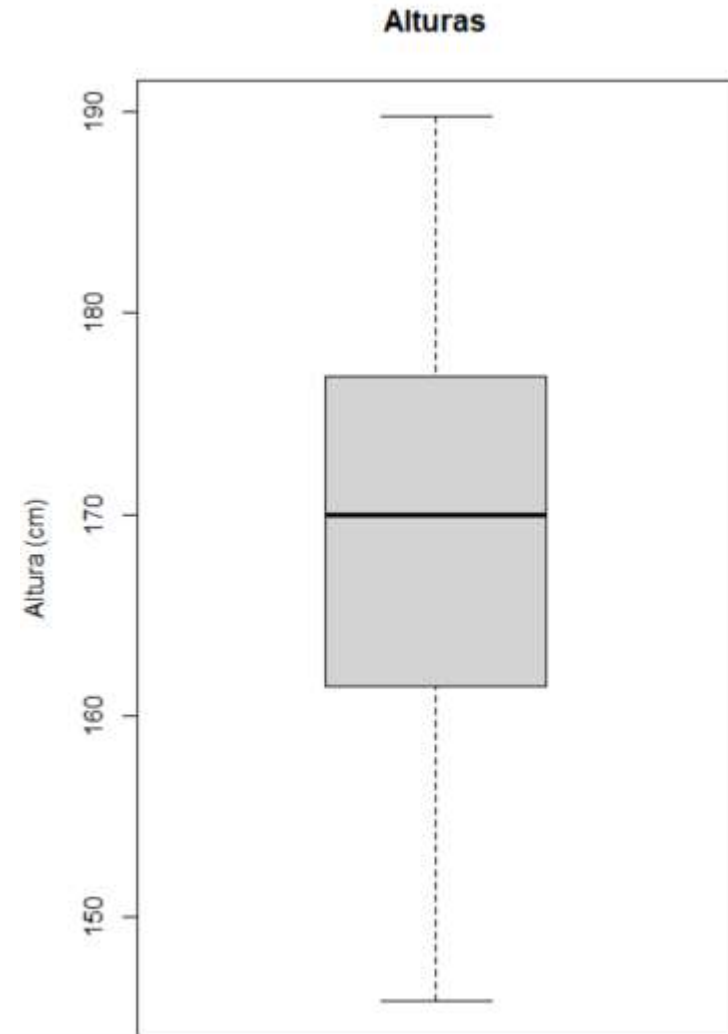
# Opção 1: inspeção visual (boxplot) – minha recomendação

- Você viu na aula 4 que:
  - boxplot = diagrama de caixa, diagrama de extremos e quartis
  - Apresenta a variação de dados observados numérica por meio de quartis

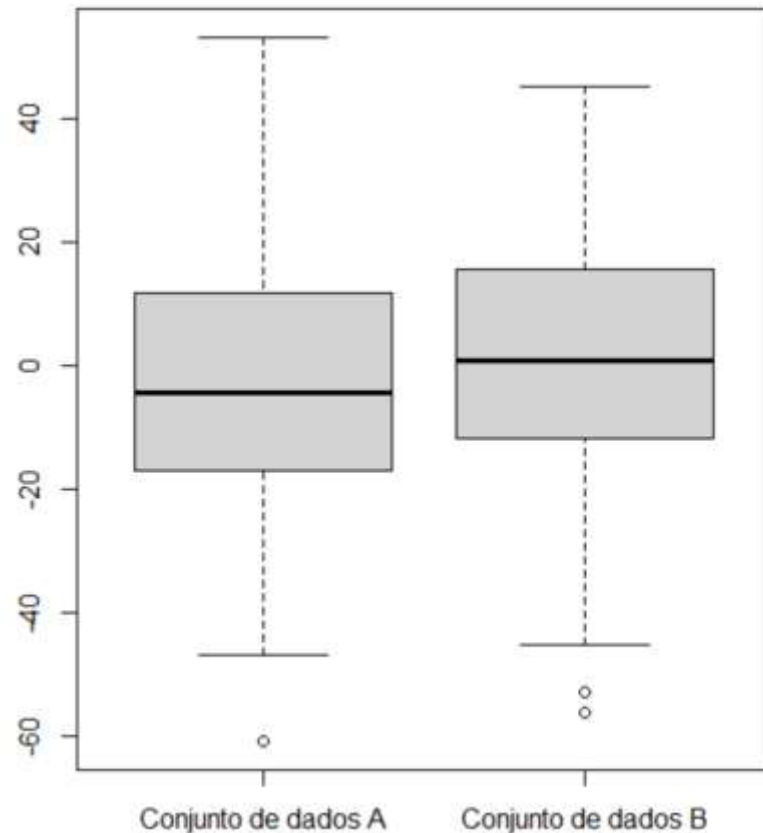


# Opção 1: inspeção visual (boxplot) – minha recomendação

- Portanto, você pode utilizar um boxplot para ter uma ideia geral da dispersão dos dados
- Construindo 2 boxplot, lado-a-lado é possível detectar evidências de variâncias desiguais



# Opção 1: inspeção visual (boxplot) – minha recomendação

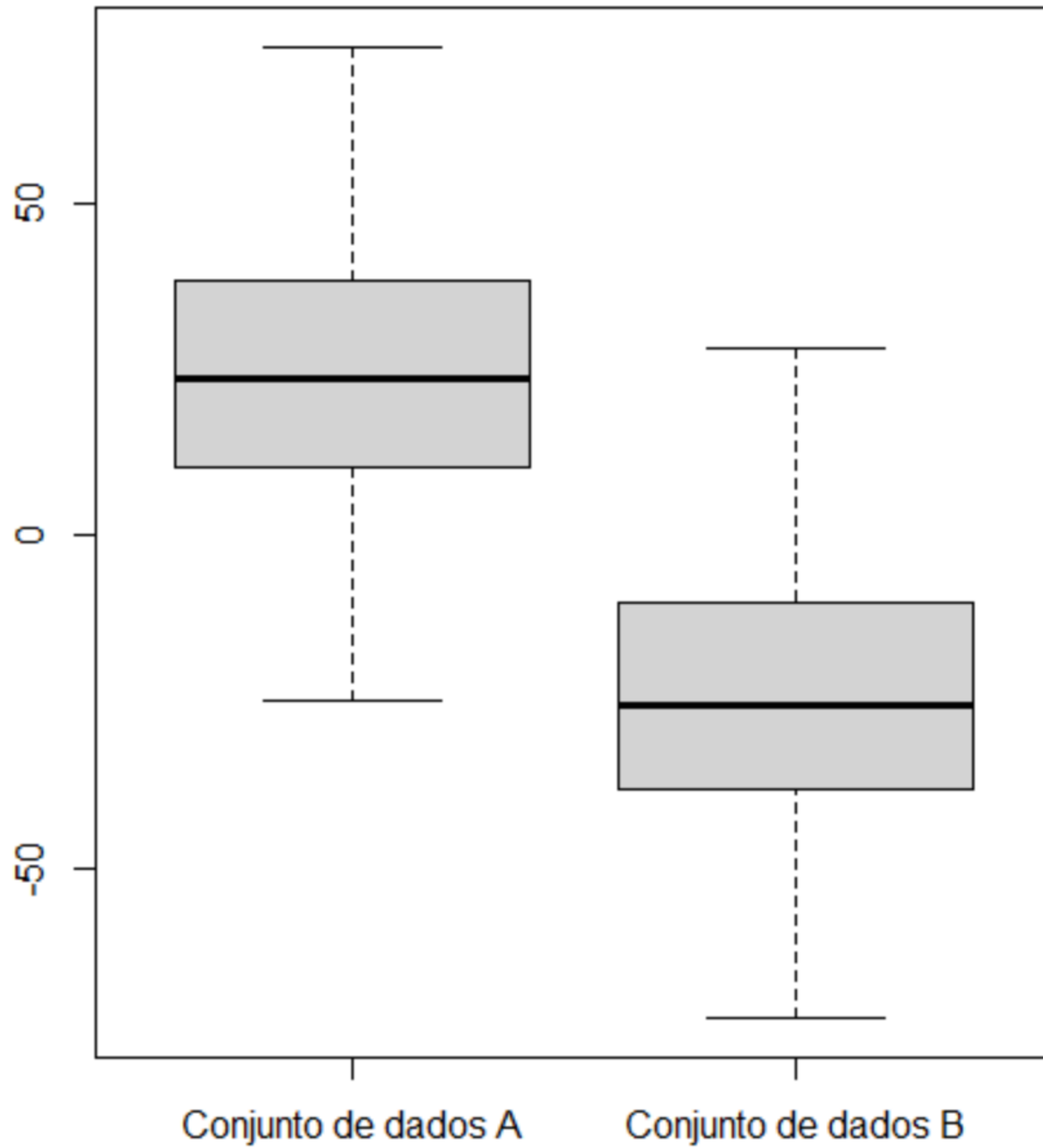


Aqui você deve comparar visualmente o tamanho das caixinhas cinza e suas caudas. Se as variâncias forem iguais o padrão geral dos dois boxplot deve ser semelhante.

OBS: **não se preocupe** com onde passa a linha preta central (mediana).

Dois conjuntos de dados podem ter variâncias iguais, e medianas muito distintas (isso é esperado quando as médias também são distintas)

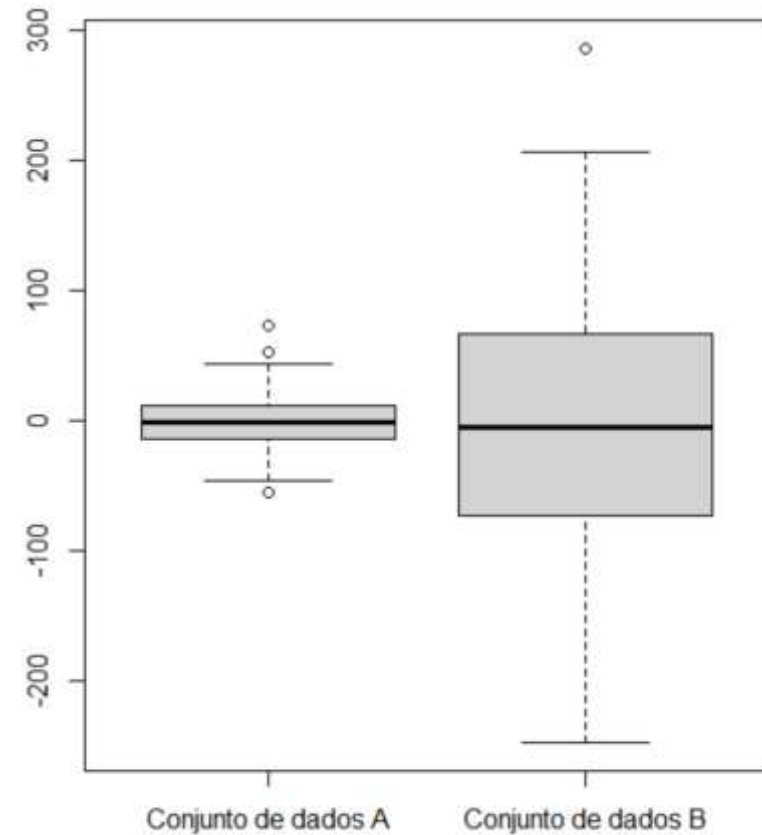
## Boxplot (boxplot) – minha



Exemplo de variâncias iguais e medianas distintas

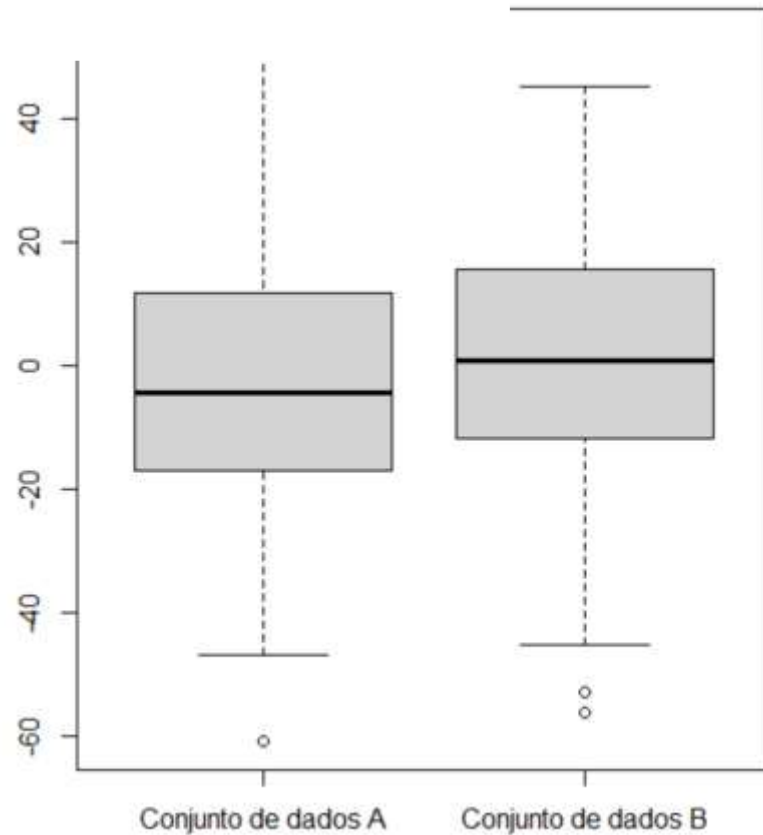
# Opção 1: inspeção visual (boxplot) – minha recomendação

Quando as variâncias são diferentes, o padrão é como ao lado: Uma caixa tem proporções diferentes da outra

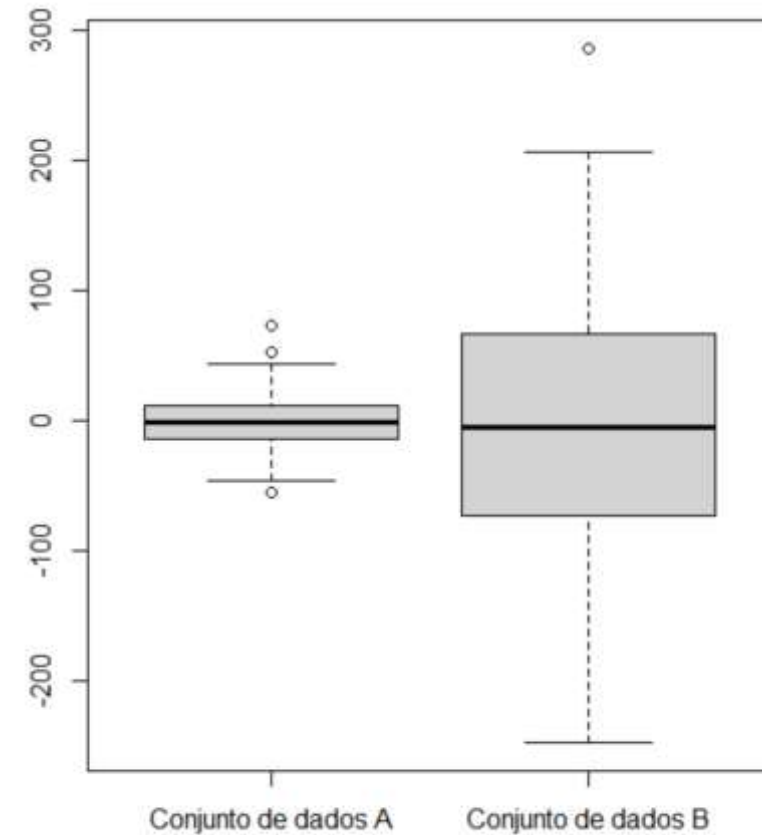


# Opção 1: inspeção visual (boxplot) – minha recomendação

Padrão que sugere variâncias iguais



Padrão que sugere variâncias diferentes



# Teste F para comparar variâncias

- O teste F para comparar variâncias, avalia se os dados apresentam evidencias divergirem quanto suas variâncias
  - $H_0$ : os dois conjuntos de dados **não apresentam diferenças** nas variâncias
  - $H_1$ : os dois conjuntos de dados **apresentam diferenças** nas variâncias
- Valores de  $p < 0,05$  indicam que as variâncias são desiguais

# Teste F para comparar variâncias

F test to compare two variances

```
data: SeusDados1 and SeusDados2
F = 0.86464, num df = 199, denom df
= 199, p-value = 0.3058
alternative hypothesis: true ratio of varia
nces is not equal to 1
95 percent confidence interval:
 0.6543445 1.1425089
sample estimates:
ratio of variances
      0.8646354
```

F test to compare two variances

```
data: SeusDados1 and SeusDados2
F = 0.69848, num df = 199, denom df
= 199, p-value = 0.0117
alternative hypothesis: true ratio of varia
nces is not equal to 1
95 percent confidence interval:
 0.5286030 0.9229597
sample estimates:
ratio of variances
      0.6984835
```



# Teste F para comparar variâncias

F test to compare two variances

```
data: SeusDados1 and SeusDados2
F = 0.86464, num df = 199, denom df
= 199, p-value = 0.3058
alternative hypothesis: true ratio of varia
nces is not equal to 1
95 percent confidence interval:
 0.6543445 1.1425089
sample estimates:
ratio of variances
 0.8646354
```

$p > 0,05$  = variâncias iguais

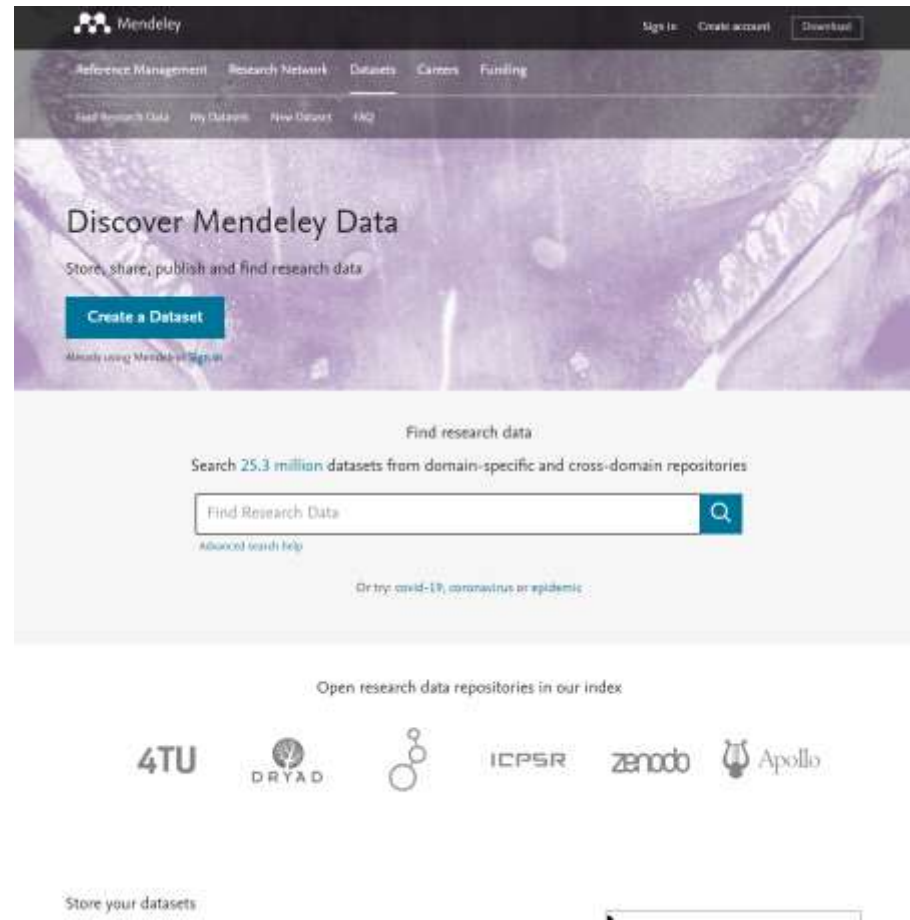
F test to compare two variances

```
data: SeusDados1 and SeusDados2
F = 0.69848, num df = 199, denom df
= 199, p-value = 0.0117
alternative hypothesis: true ratio of varia
nces is not equal to 1
95 percent confidence interval:
 0.5286030 0.9229597
sample estimates:
ratio of variances
 0.6984835
```

$P < 0,05$  = variâncias desiguais

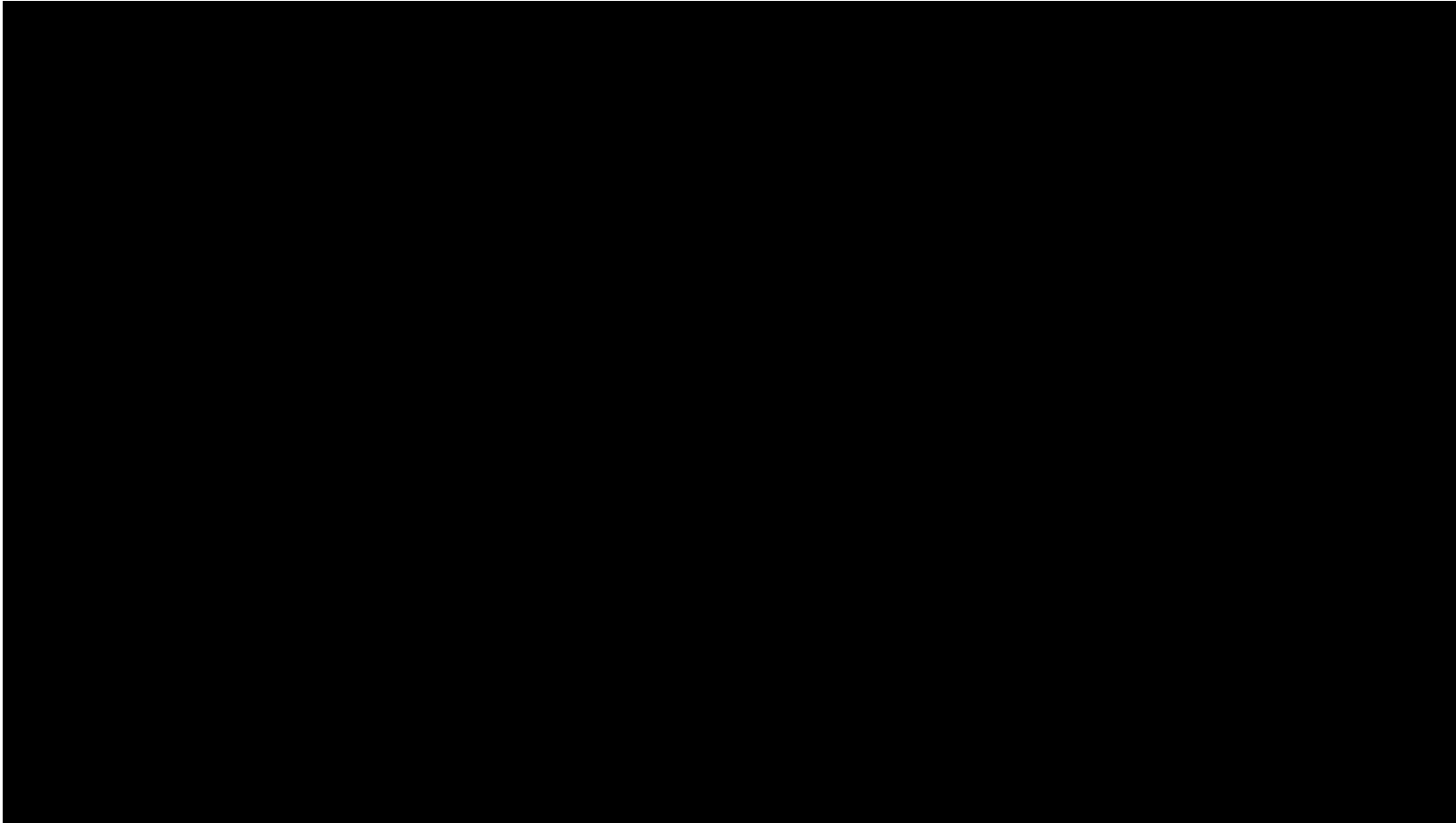
Nota: de onde vem os dados que estamos analisando nos exemplos?

# Origem dos dados a serem analisados



<https://data.mendeley.com/>

# Origem dos dados a serem analisados



<https://repositoriodatasharingfapesp.uspdigital.usp.br/handle/item/1>