



Prof<sup>a</sup>.

Paula Shinozaki

Monitora:

Luana Moisés

- Como instalar o R
- Normalidade dos dados
- Dados pareados e não pareados



# Como instalar o R



# Como instalar o R

<https://www.r-project.org/>

Google search results for "r project". The search bar shows "r project" and the results indicate approximately 3,470,000,000 results found in 0.74 seconds. The top result is "The R Project for Statistical Computing" with a green checkmark, described as a free software environment for statistical computing and graphics. Below this, there are search results from r-project.org, including links to "R-4.1.1 for Windows (32/64 bit)", "The Comprehensive R Archive...", "CRAN - Mirrors", "R-3.5.3 for Windows (32/64 bit)", and "Index of /src/base/R-4". A sidebar on the right shows a collage of R-related images and a description of R as a multi-paradigm programming language.

Google

r project

Todas Imagens Vídeos Notícias Livros Mais Ferramentas

Aproximadamente 3.470.000.000 resultados (0,74 segundos)

Exibindo resultados para **r project**  
Em vez disso, pesquisar por **r project**

<https://www.r-project.org> Traduzir esta página

**The R Project for Statistical Computing** ✓  
The **R Project** for Statistical Computing. Getting Started. R is a free software environment for statistical computing and graphics.

Resultados de r-project.org

**R-4.1.1 for Windows (32/64 bit)** ✓ Windows ✓  
If you want to double-check that the package you have ... base, Binaries for base distribution. This is what you want to install ...

**The Comprehensive R Archive...** **R-3.5.3 for Windows (32/64 bit)** ✓  
CRAN is a network of ftp and web servers around the world that ... If you want to double-check that the package you have ...

**CRAN - Mirrors** ✓  
If you want to host a new mirror at

**Index of /src/base/R-4** ✓  
Name · Last modified · Size ·

**R**  
Linguagem de programação

R é uma linguagem de programação multi-paradigma orientada a objetos, programação funcional, dinâmica, fracamente tipada, voltada à manipulação, análise e visualização de dados. Foi criado originalmente por Ross Ihaka e por Robert



# Como instalar o R

## □ Baixar o arquivo

### Download

CRAN



### R Project

About R

Logo

Contributors

What's New?

Reporting Bugs

Conferences

Search

Get Involved: Mailing Lists

Developer Pages

R Blog

### Brazil

<https://nbcgib.uesc.br/mirrors/cran/>

<https://cran-r.c3sl.ufpr.br/>

<https://cran.fiocruz.br/>

<https://vps.fmvz.usp.br/CRAN/>

<https://brieger.esalq.usp.br/CRAN/>

### Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#) ([Debian](#), [Fedora/Redhat](#), [Ubuntu](#))
- [Download R for macOS](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.



# Como instalar o R

## Subdirectories:

[base](#)

Binaries for base distribution. This is what you want to [install R for the first time](#).

[contrib](#)

Binaries of contributed CRAN packages (for R  $\geq$  2.13.x; managed by Uwe Ligges). There is also information on [third party software](#) available for CRAN Windows services and corresponding environment and make variables.

[old contrib](#)

Binaries of contributed CRAN packages for outdated versions of R (for R  $<$  2.13.x; managed by Uwe Ligges).

[Rtools](#)

Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.

[Download R 4.1.1 for Windows](#) (86 megabytes, 32/64 bit)

[Installation and other instructions](#)

[New features in this version](#)



# Como instalar o R

The screenshot displays the RGui (64-bit) application window. The menu bar includes Arquivo, Editar, Visualizar, Misc, Pacotes, Janelas, and Ajuda. The File menu is open, showing options like Interpretar código fonte R..., Novo script, Abrir script..., Exibir arquivo(s)..., Carregar área de trabalho..., Salvar área de trabalho... (Ctrl+S), Carregar Histórico..., Salvar Histórico..., Mudar dir..., Imprimir... (Ctrl+P), Salvar em arquivo..., and Sair. Below the menu, a snippet of R code is visible: `ou 'help.start()' para abrir o sistema`, `Digite 'q()' para sair do R.`, and `> |`.

The R Console window shows the following output:

```
R version 4.1.0 (2021-05-18) -- "Camp Pontanezen"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R é um software livre e vem sem GARANTIA ALGUMA.
Você pode redistribuí-lo sob certas circunstâncias.
Digite 'license()' ou 'licence()' para detalhes da licença.

R é um projeto colaborativo com muitos contribuintes.
Digite 'contributors()' para obter mais informações.
Digite 'citation()' para saber como citar o R ou para obter citações.

Digite 'demo()' para demonstrações, 'help()' para obter ajuda,
ou 'help.start()' para abrir o sistema de ajuda online.
Digite 'q()' para sair do R.
```

The script editor window, titled "Sem nome - Editor R", is currently blank.



# Como instalar o R

```
> ?shapiro
No documentation for 'shapiro' in specified packages and libraries:
you could try '??shapiro'
> ??shapiro
starting httpd help server ... done
> |
```

Search Results



Help pages:

[stats::shapiro.test](#)

Shapiro-Wilk Normality Test



# Como instalar o R

The screenshot shows the RGui (64-bit) interface. The 'Pacotes' menu is open, displaying options: 'Carregar pacote...', 'Escolher espelho CRAN...', 'Selecionar repositórios...', 'Instalar pacote(s)...', 'Atualizar pacotes...', and 'Install package(s) from local files...'. The 'R Console' window shows the following text:

```
R version 4.1.0 (2021-05-25)
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64

R é um software livre e vem sem GARANTIA ALGUMA.
Você pode redistribuí-lo sob certas circunstâncias.
Digite 'license()' ou 'licence()' para detalhes de distribuição.
Digite 'demo()' para demonstrações, 'help()' para o sistema on-line de ajuda
ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.
Digite 'q()' para sair do R.
```

The package selection dialog is open, showing a list of packages. The 'stats' package is selected. The list includes:

- base
- boot
- class
- cluster
- codetools
- compiler
- datasets
- foreign
- graphics
- grDevices
- grid
- KernSmooth
- lattice
- MASS
- Matrix
- methods
- mgcv
- nlme
- nnet
- parallel
- rpart
- spatial
- splines
- stats**
- stats4
- survival
- tcltk
- tools
- translations
- utils

The dialog has 'OK' and 'Cancelar' buttons at the bottom.





# Como instalar o R

## Com o pacote carregado: ?shapiro.test

shapiro.test {stats}

R Documentation

### Shapiro-Wilk Normality Test

#### Description

Performs the Shapiro-Wilk test of normality.

#### Usage

```
shapiro.test(x)
```

#### Arguments

**x**

a numeric vector of data values. Missing values are allowed, but the number of non-missing values must be between 3 and 5000.

#### Value

A list with class "htest" containing the following components:

**statistic**

the value of the Shapiro-Wilk statistic.

**p.value**

an approximate p-value for the test. This is said in Royston (1995) to be adequate for  $p.value < 0.1$ .

**method**

the character string "Shapiro-Wilk normality test".

**data.name**

a character string giving the name(s) of the data.

#### Source

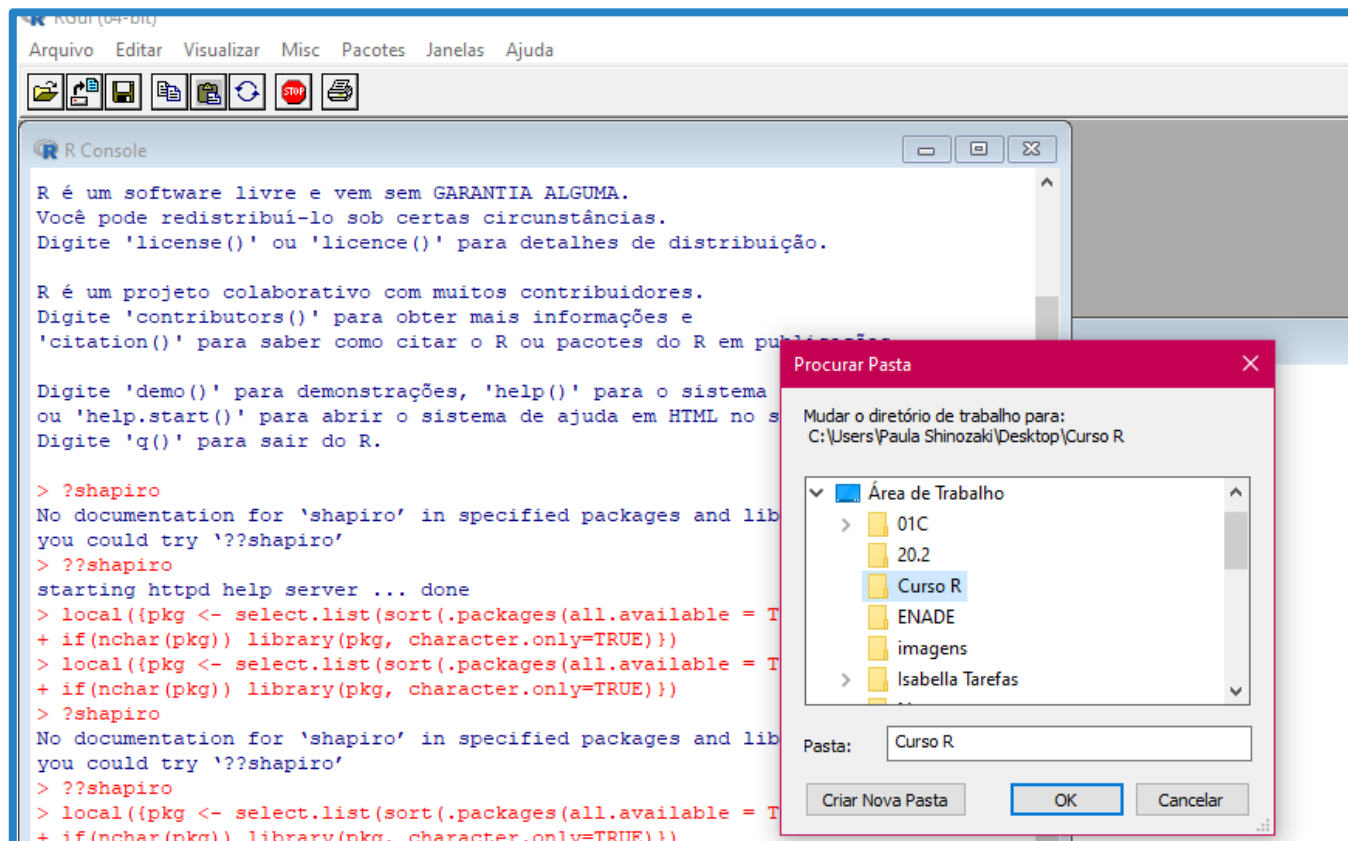
The algorithm used is a C translation of the Fortran code described in Royston (1995). The calculation of the p value is exact for  $n = 3$ , otherwise approximations are used, separately for  $4 \leq n \leq 11$  and  $n \geq 12$ .

#### References



# Como instalar o R

- Chamando o arquivo
  - ▣ Arquivo -> Mudar dir...





# Normalidade dos dados



# Normalidade dos dados

- Coleta dos dados
- Verificação de observações outliers
  - O que fazer?
- Análise descritiva dos dados
- Gráficos e tabelas
- Testes de hipóteses
  - Qual utilizar?
- Dados seguem uma distribuição normal?



# Normalidade dos dados

## □ Distribuição Normal

Utilizamos os testes de normalidade para verificar se um conjunto de dados seguem uma distribuição normal ou não.

## □ Característica de uma distribuição Normal

É uma distribuição de probabilidade absolutamente contínua parametrizada pela sua média ( $\mu$ ) e desvio padrão ( $\sigma$ )



# Normalidade dos dados

- Notação

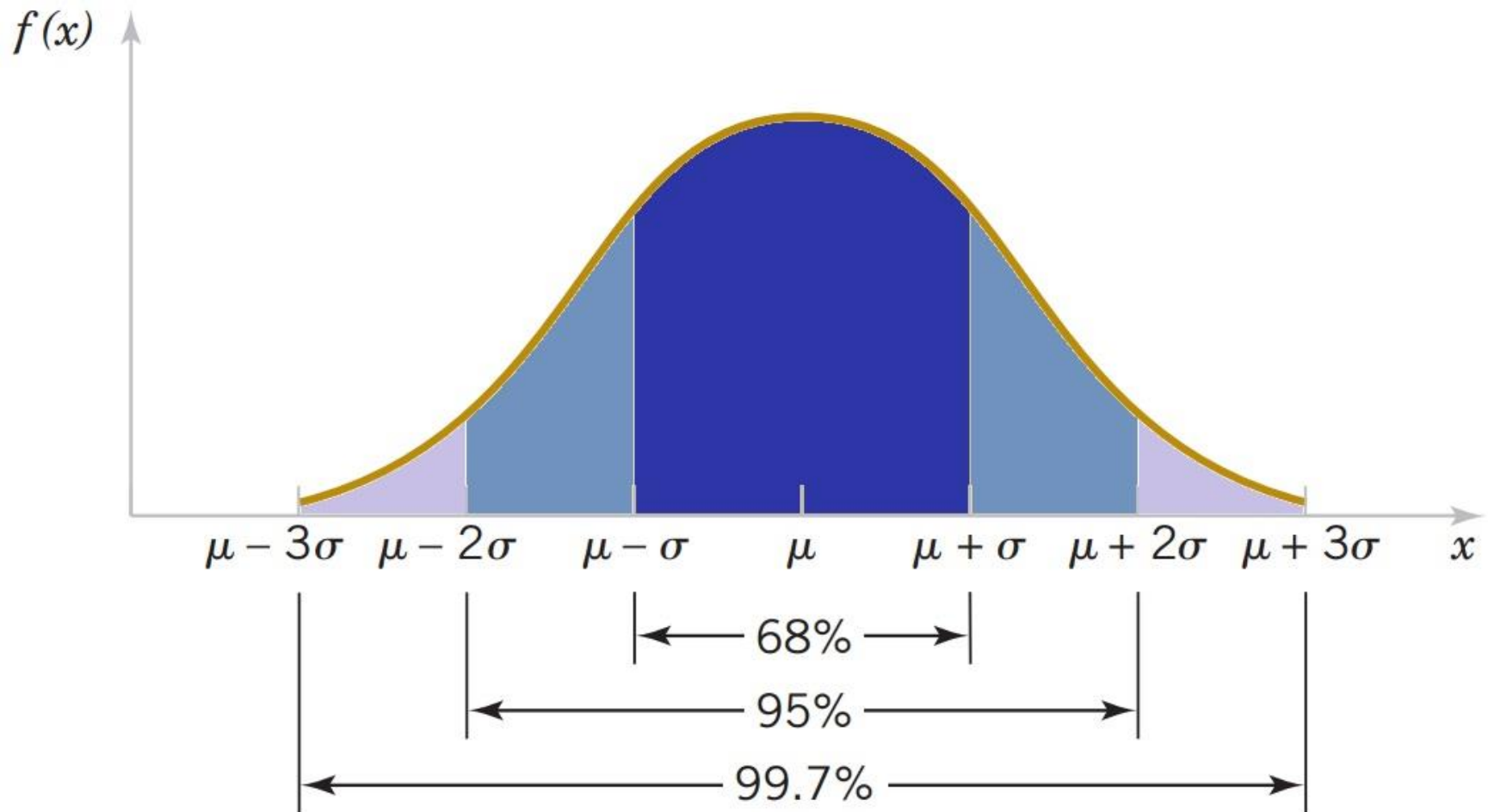
$$X \sim N(\mu; \sigma^2)$$

- Densidade de probabilidade

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



# Normalidade dos dados





# Normalidade dos dados

- Teste de hipóteses

- ▣ Erro tipo I e erro tipo II

Uma indústria usa, em um dos seus componentes, um parafuso importado, que deve satisfazer a algumas exigências. Esses parafusos são fabricados em alguns países e suas especificações mudam de acordo com cada país. O catálogo do país **A** afirma que a resistência média da tração é de 145 kg, com desvio de 12 kg. Já o país **B**, a média é de 155 kg e desvio de 20 kg.

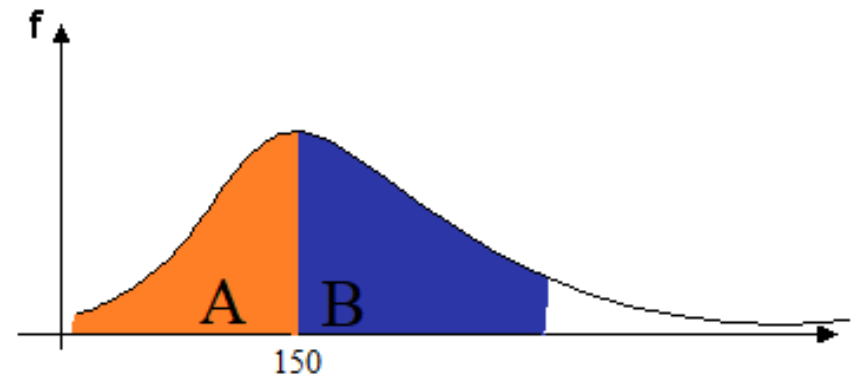




# Normalidade dos dados

Um lote será leiloadado por um preço convidativo e para que a indústria faça ou não uma oferta, ela necessita saber qual país produziu tais parafusos. Pouco antes do leilão será divulgada a resistência média de uma amostra de 25 parafusos. Qual a regra a ser usada para dizer se os parafusos são do país **A** ou **B**?

- Se  $\bar{x} \leq 150$ , são do A
- Se  $\bar{x} > 150$ , são do B

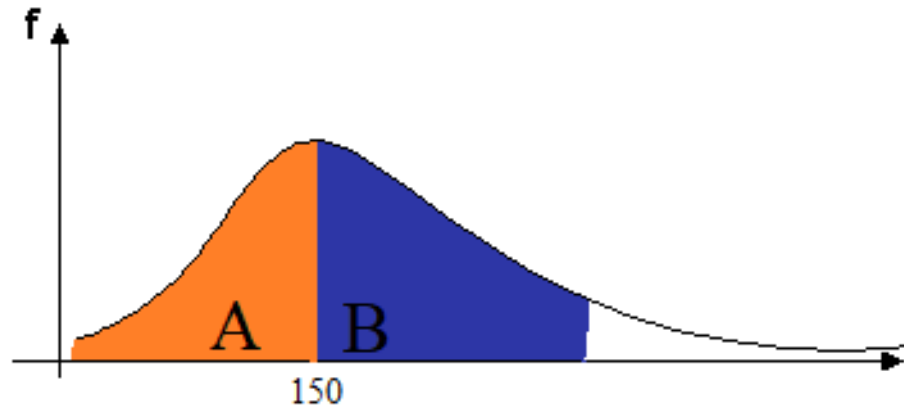




# Normalidade dos dados

No dia do leilão fomos informados que  $\bar{x} = 148$  então pela nossa regra seria do país **A**.

- Podemos estar errados?
- É possível que no país B tenha  $\bar{x} = 148$ ?





# Normalidade dos dados

## □ ERRO TIPO I

Dizer que os parafusos são de **A** quando na verdade são de **B**. Ou melhor, quando uma amostra de 25 parafusos de B apresentam  $\bar{x} \leq 150$

## □ ERRO TIPO II

Dizer que os parafusos são de **B** quando na verdade são de **A**. Isso ocorre quando uma amostra de 25 parafusos de **A** apresentam  $\bar{x} > 150$



# Normalidade dos dados

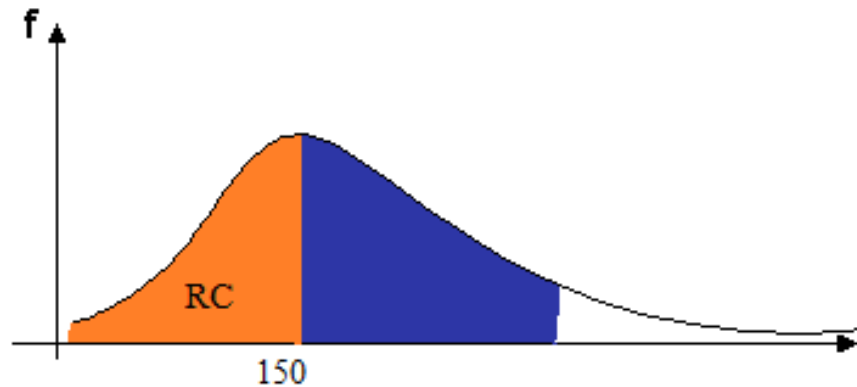
## Hipóteses enumeradas

- $H_0$ : *Os parafusos são de B*  
 $\approx X \sim (155; 20)$
- $H_1$ : *Os parafusos são de A*  
 $\approx X \sim (145; 12)$



# Normalidade dos dados

Considere RC por região crítica, isto é, região correspondente aos valores menores que 150



$$P(\text{erro I}) = P(\bar{x} \in RC | H_0 \text{ é verdadeira}) = \alpha$$
$$P(\text{erro II}) = P(\bar{x} \notin RC | H_1 \text{ é verdadeira}) = \beta$$



# Normalidade dos dados

Admitindo que  $\bar{x} \sim N(155; 16)$

$$\begin{aligned} P(\text{erro tipo I}) &= P\left(z \leq \frac{150 - 155}{4}\right) = \\ &= P(z \leq -1,25) = 0,1056 = \\ &= 10,56\% = \alpha \end{aligned}$$



# Normalidade dos dados

Admitindo que  $\bar{x} \sim N(145; 5,76)$

$$\begin{aligned} P(\text{erro tipo II}) &= P\left(z > \frac{150 - 145}{2,4}\right) = \\ &= P(z > 2,08) = 0,0188 = \\ &= 1,88\% = \beta \end{aligned}$$



# Normalidade dos dados

Origem real	Decisão	
	A	B
A	Sem erro	Erro tipo II ( $\beta = 1,88\%$ )
B	Erro tipo I ( $\alpha = 10,56\%$ )	Sem erro





# Normalidade dos dados

- Para qualquer  $\alpha$  fixo, um aumento no tamanho de  $n$  descrece  $\beta$ ;
- Para qualquer  $n$  fixo, um decréscimo em  $\alpha$  provoca um aumento em  $\beta$ ;
- Para diminuir ambos,  $n$  deve ser aumentado.



# Normalidade dos dados

## PROCEDIMENTO GERAL DO TESTE DE HIPÓTESES

$$H_0: \theta = \theta_0$$

em que

$\theta \rightarrow$  parâmetro desconhecido da população

$\theta_0 \rightarrow$  estimativa de  $\theta$  baseado na amostra

$$H_1: \theta \neq \theta_0 \text{ ou } H_1: \theta > \theta_0 \text{ ou } H_1: \theta < \theta_0$$



# Normalidade dos dados

- Esta decisão é formada através da consideração da RC. Caso o valor observado da *estatística* pertença a essa região, **rejeitamos  $H_0$** ; caso contrário não rejeitamos  $H_0$
- A região crítica é sempre construída sob a hipótese nula ser verdadeira. A determinação de  $\beta$  é mais difícil uma vez que não fixamos os valores para os parâmetros.

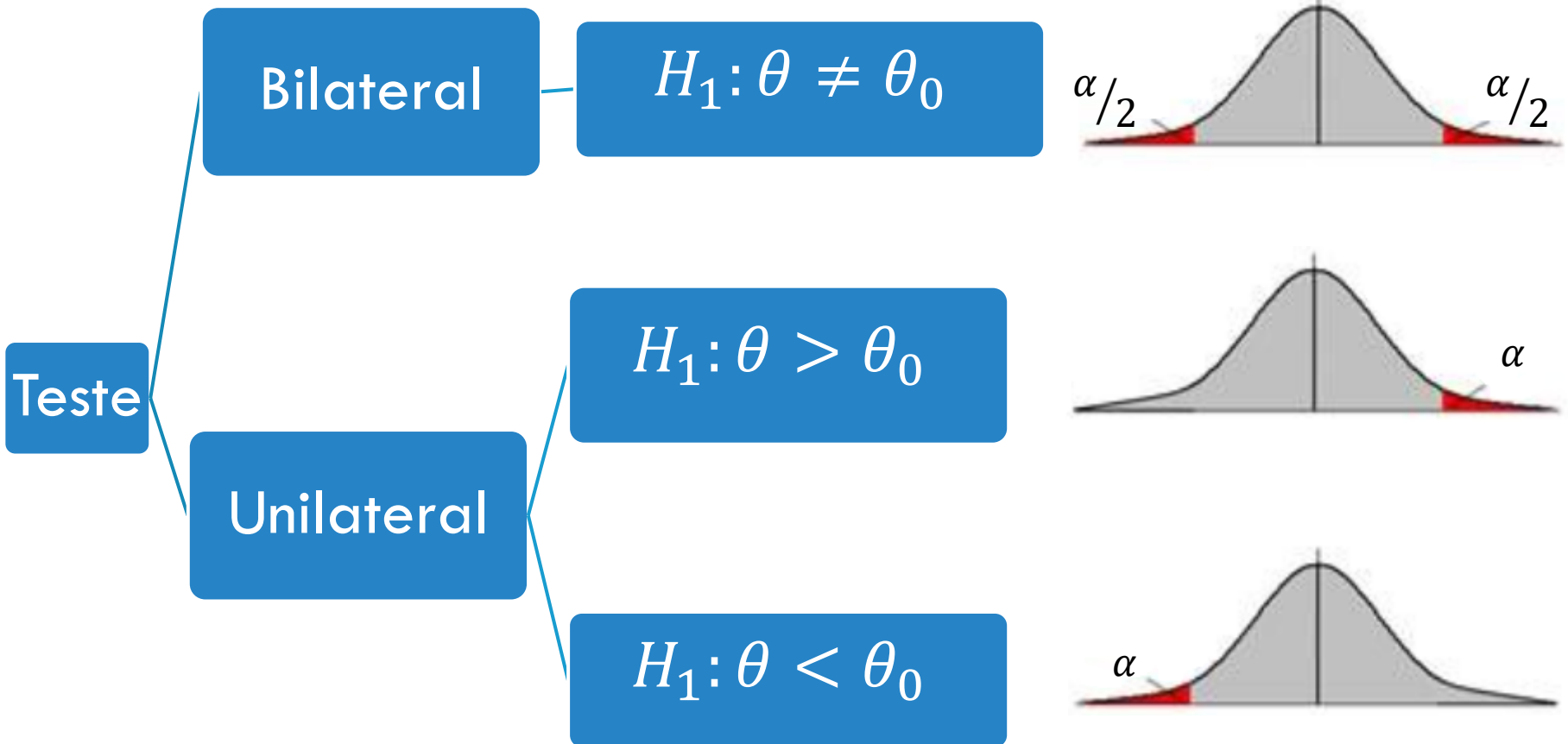


# Normalidade dos dados

A probabilidade de  $\alpha$  de se cometer o erro tipo I recebe o nome de **nível de significância** do teste e é um valor arbitrário.



# Normalidade dos dados





# Normalidade dos dados

## Testes de normalidade

```
graph TD; A[Testes de normalidade] --> B[Shapiro-Wilk]; A --> C[Kolmogorov-Smirnov]
```

A hierarchical diagram showing 'Testes de normalidade' at the top, branching down to 'Shapiro-Wilk' and 'Kolmogorov-Smirnov'.

Shapiro-Wilk

Kolmogorov-Smirnov



# Normalidade dos dados

**Shapiro-Wilk**

**Kolmogorov-Smirnov**

- Quando utilizar cada um?

Para amostras de tamanho superior ou igual a 30 aconselha-se o teste de Kolmogorov-Smirnov e para amostras menores é mais indicado o teste de Shapiro-Wilk.



# Normalidade dos dados

- Lendo os dados

```
notas<-read.table("notas.txt", head=T)
```

```
attach(notas)
```

```
Notas
```

```
shapiro.test(Notas)
```





# Normalidade dos dados

## □ Shapiro-Wilk

$\left\{ \begin{array}{l} H_0: \text{Os dados seguem uma distribuição normal} \\ H_1: \text{Os dados não seguem uma distribuição normal} \end{array} \right.$

```
> shapiro.test(Notas)
```

Shapiro-Wilk normality test

data: Notas

W = 0.89768, p-value = 0.0004031



# Normalidade dos dados

## □ Shapiro-Wilk

O valor  $p$  indica a probabilidade de irmos a favor da hipótese nula, que no teste em questão essa hipótese afirma que os dados seguem uma distribuição normal.

Dado o nível de significância  $\alpha$ , que geralmente utiliza-se 5%, ou seja, 0,05, quando o valor  $p$  for menor que  $\alpha$  dizemos que existem evidências para irmos contra  $H_0$ .



# Normalidade dos dados

- Shapiro-Wilk

Valor  $p = 0,0004031 \rightarrow$  Rejeita-se  $H_0$

Logo, existem evidências que os dados não seguem uma distribuição normal



# Normalidade dos dados

## □ Kolmogorov-Smirnov

Esse teste compara distribuições. No nosso caso, como estamos utilizando para testar se os dados seguem uma distribuição **Normal**, precisamos gerar um vetor de uma normal para usar como comparação com os nossos dados.

```
x ← rnorm(50)
```

$\left\{ \begin{array}{l} H_0: \text{Os dados seguem uma distribuição normal} \\ H_1: \text{Os dados não seguem uma distribuição normal} \end{array} \right.$



# Normalidade dos dados

## □ Kolmogorov-Smirnov

```
> ks.test(Notas,x)
```

Two-sample Kolmogorov-Smirnov test

data: Notas and x

D = 1, p-value < 2.2e-16

alternative hypothesis: two-sided

Warning message:

In ks.test(Notas, x) : cannot compute exact p-value with ties



# Normalidade dos dados

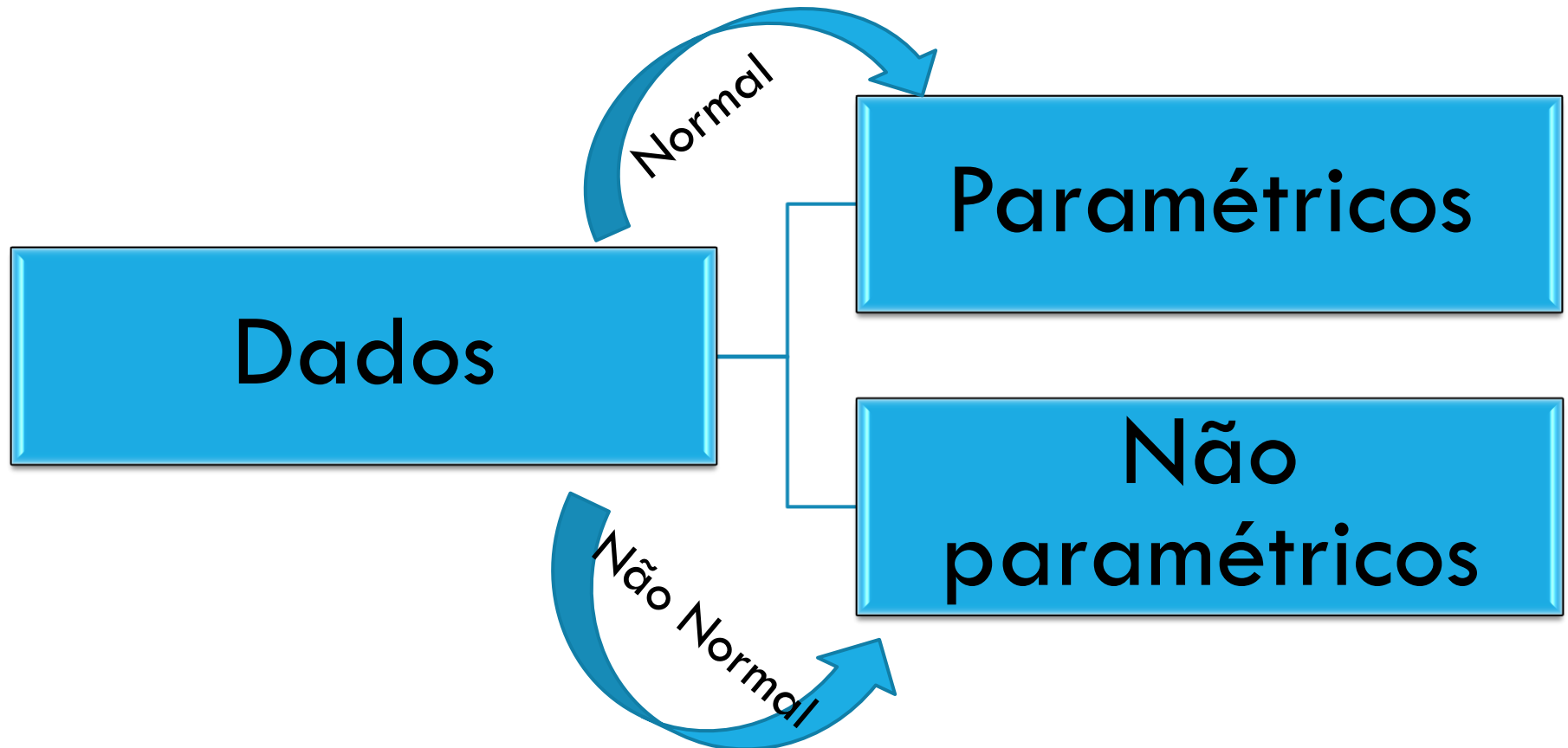
## □ Kolmogorov-Smirnov

Valor  $p < 2,2e-16 \rightarrow$  Rejeita-se  $H_0$

Logo, existem evidências que os dados não seguem uma distribuição normal.



# Normalidade dos dados





# Normalidade dos dados

## EXERCÍCIOS DE FIXAÇÃO

- No arquivo tarefa.txt efetue os testes de normalidades na variável Tarefa com os métodos:
  - 1) Shapiro-Wilk
  - 2) Kolmogorov-Smirnov

Bom trabalho!