

AULA PRÁTICA - 

| | | |
|---------------|-----------------------------|--------|
| DEST-IME-UFBA | MATB59-Estatística Básica A | 2019.1 |
|---------------|-----------------------------|--------|

Materiais utilizados para elaboração dos Slides

Tutorial Introdução ao

Orientadora: Profª Edleide de Brito
Bolsista Permanecer 2015: Antonio de Jesus Nascimento

Slides de aula

Profª Giovana Oliveira Silva

3

Introdução ao R

R é um ambiente de *software* livre para computação estatística e gráficos. Ele compila e roda em uma ampla variedade de plataformas Linux, Windows e Mac.

www.r-project.org

DEST-IME-UFBA Profº Andrea A. Prudente MATB59 - 2019.1

Sobre o R

4

- O R é uma linguagem para manipular objetos.
- Os objetos podem ser conjunto de dados, vetores, funções, matrizes, etc.
- As manipulações podem ser cálculos, entrada e saída de dados, análises estatísticas e gráficos.

DEST-IME-UFBA Profº Andrea A. Prudente MATB59 - 2019.1

Características

5

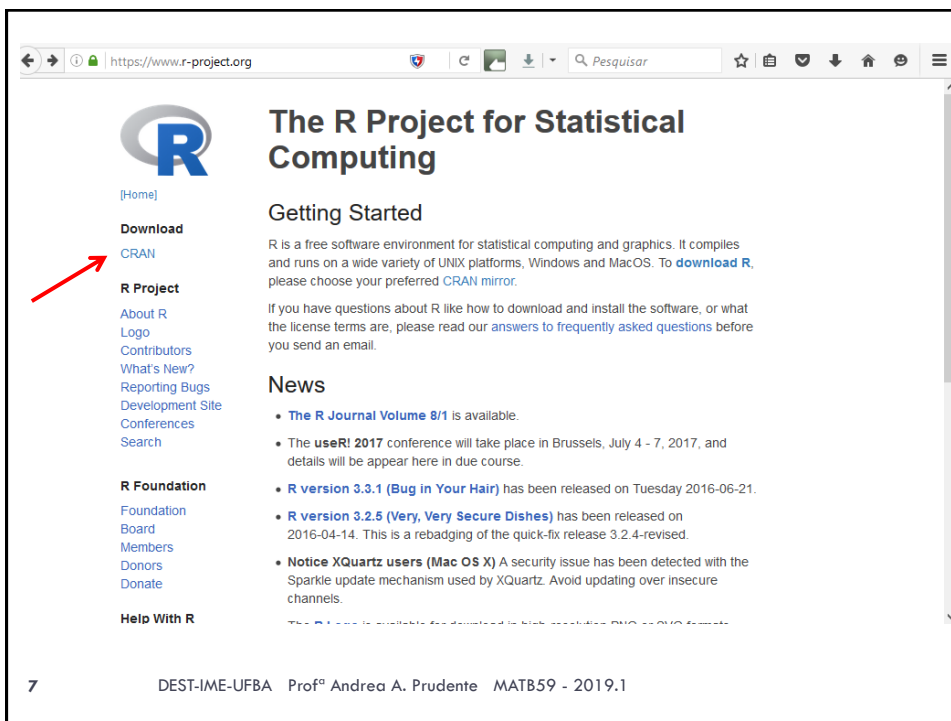
- É gratuito, diversas funções estatísticas disponíveis na versão básica.
- Grande variedade de pacotes/bibliotecas com funções específicas disponíveis que podem ser instalados pela Internet, através do próprio programa.
- Conta com inúmeros colaboradores no mundo inteiro que criam, testam e corrigem as funções que podem ser usadas por qualquer pessoa.

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

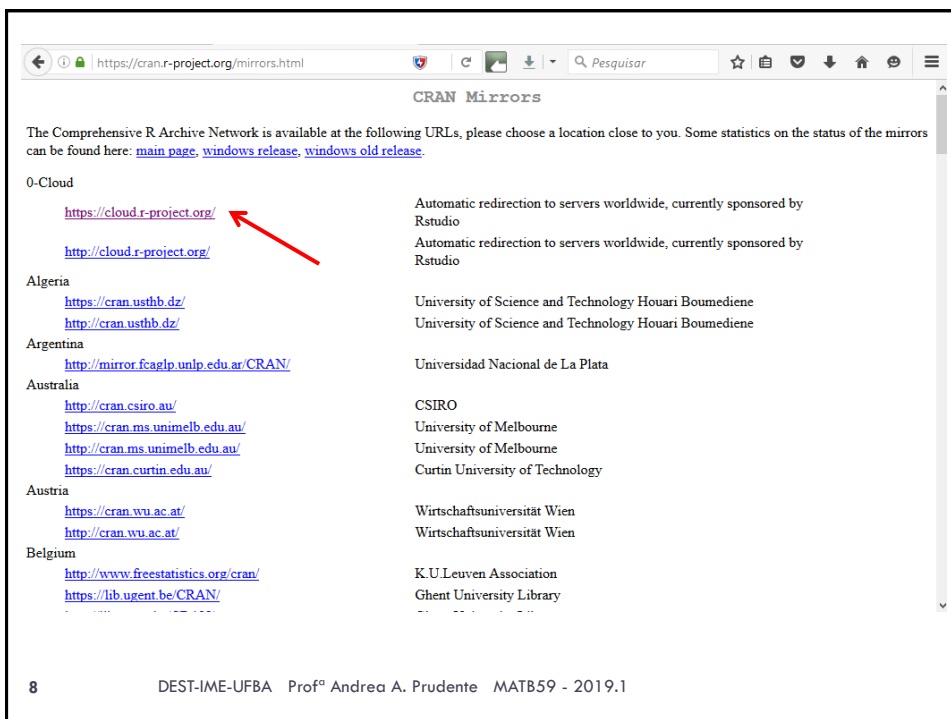
6

Download do R

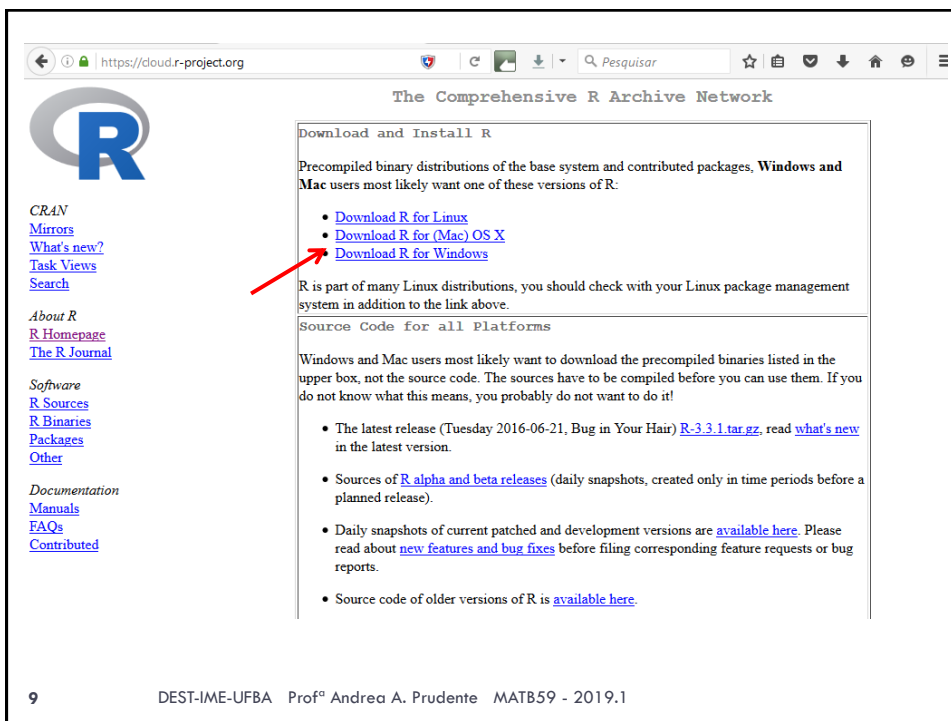
DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1



7 DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1



8 DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1



The Comprehensive R Archive Network

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](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

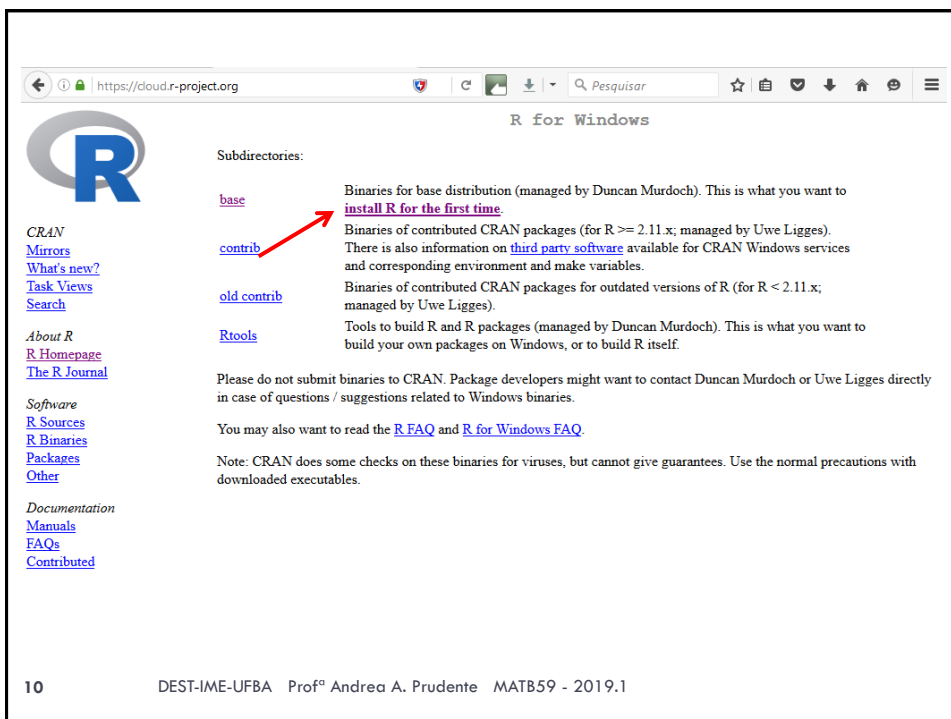
R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (Tuesday 2016-06-21, Bug in Your Hair) [R-3.3.1.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).

9 DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1



R for Windows

Subdirectories:

- [base](#)
- [contrib](#)
- [old contrib](#)
- [Rtools](#)

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

Binaries of contributed CRAN packages (for R >= 2.11.x; managed by Uwe Ligges). There is also information on [third party software](#) available for CRAN Windows services and corresponding environment and make variables.

Binaries of contributed CRAN packages for outdated versions of R (for R < 2.11.x; managed by Uwe Ligges).

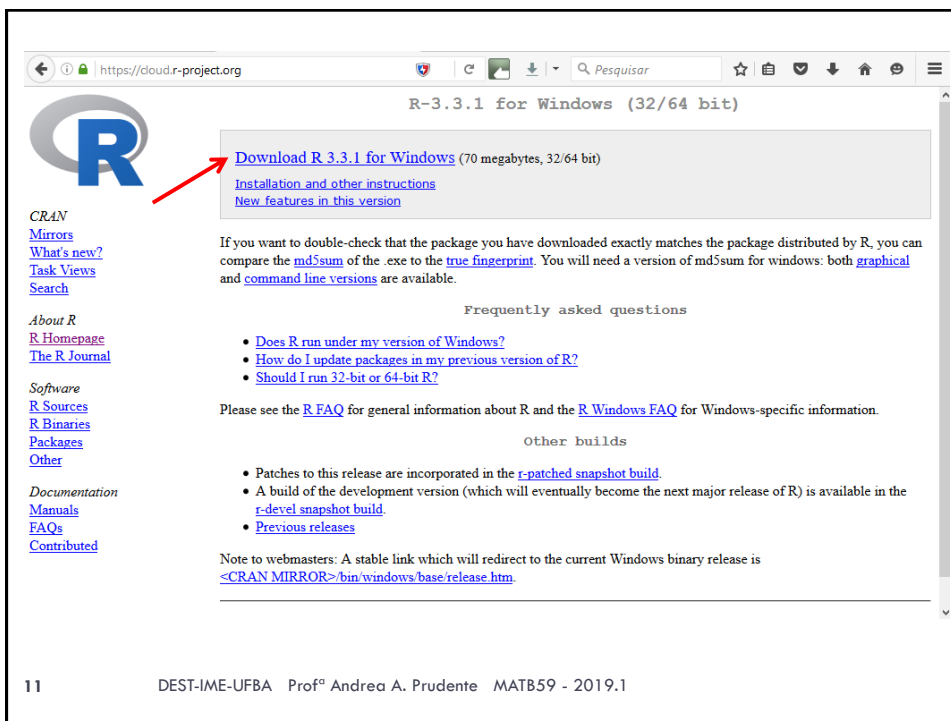
Tools to build R and R packages (managed by Duncan Murdoch). This is what you want to build your own packages on Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Duncan Murdoch or Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

10 DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1



The screenshot shows the R Project website for Windows. The browser address bar displays <https://cloud.r-project.org>. The page title is "R-3.3.1 for Windows (32/64 bit)". A red arrow points to the "Download R 3.3.1 for Windows (70 megabytes, 32/64 bit)" link. Below this link are links for "Installation and other instructions" and "New features in this version". The page includes a sidebar with links for CRAN, Mirrors, What's new?, Task Views, Search, About R, R Homepage, The R Journal, Software, R Sources, R Binaries, Packages, Other, Documentation, Manuals, FAQs, and Contributed. The main content area contains a paragraph about double-checking the package, a section for frequently asked questions, and a note about CRAN mirrors.

CRAN
[Mirrors](#)
[What's new?](#)
[Task Views](#)
[Search](#)

About R
[R Homepage](#)
[The R Journal](#)

Software
[R Sources](#)
[R Binaries](#)
[Packages](#)
[Other](#)

Documentation
[Manuals](#)
[FAQs](#)
[Contributed](#)

R-3.3.1 for Windows (32/64 bit)

[Download R 3.3.1 for Windows](#) (70 megabytes, 32/64 bit)
[Installation and other instructions](#)
[New features in this version](#)

If you want to double-check that the package you have downloaded exactly matches the package distributed by R, you can compare the [md5sum](#) of the .exe to the [true fingerprint](#). You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

Frequently asked questions

- [Does R run under my version of Windows?](#)
- [How do I update packages in my previous version of R?](#)
- [Should I run 32-bit or 64-bit R?](#)

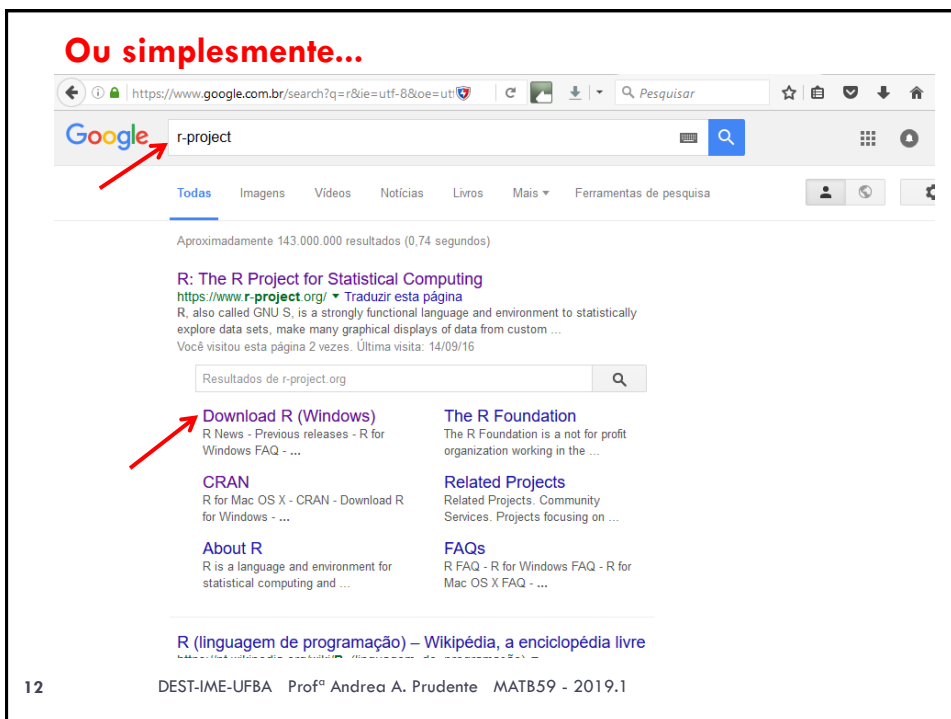
Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.

Other builds

- Patches to this release are incorporated in the [r-patched snapshot build](#).
- A build of the development version (which will eventually become the next major release of R) is available in the [r-devel snapshot build](#).
- [Previous releases](#)

Note to webmasters: A stable link which will redirect to the current Windows binary release is
CRAN.MIRROR>/bin/windows/base/release.htm

11 DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1



The screenshot shows a Google search for "r-project". The browser address bar displays <https://www.google.com.br/search?q=r&ie=utf-8&oe=utf-8>. The search results show "R: The R Project for Statistical Computing" with the URL <https://www.r-project.org/>. A red arrow points to the "Download R (Windows)" link in the search results. The page includes a sidebar with links for Todas, Imagens, Vídeos, Notícias, Livros, Mais, and Ferramentas de pesquisa. The main content area contains a paragraph about R, a section for frequently asked questions, and a note about CRAN mirrors.

Ou simplesmente...

Google

Todas Imagens Vídeos Notícias Livros Mais Ferramentas de pesquisa

Aproximadamente 143.000.000 resultados (0,74 segundos)

R: The R Project for Statistical Computing
<https://www.r-project.org/> Traduzir esta página
 R, also called GNU S, is a strongly functional language and environment to statistically explore data sets, make many graphical displays of data from custom ...
 Você visitou esta página 2 vezes. Última visita: 14/09/16

Resultados de r-project.org

[Download R \(Windows\)](#)
 R News - Previous releases - R for Windows FAQ - ...

[The R Foundation](#)
 The R Foundation is a not for profit organization working in the ...

[CRAN](#)
 R for Mac OS X - CRAN - Download R for Windows - ...

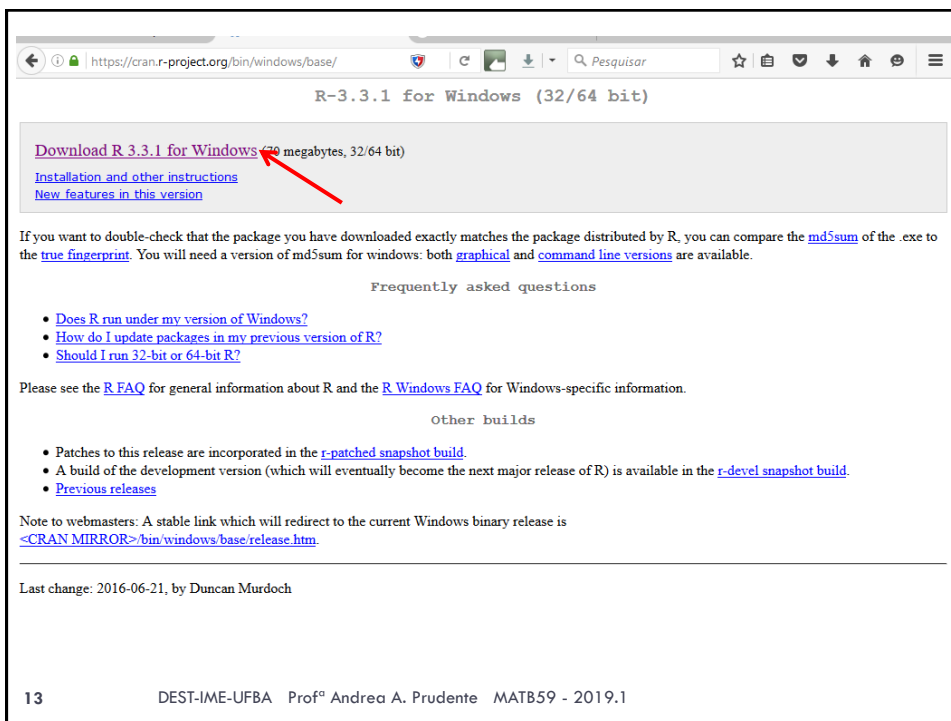
[Related Projects](#)
 Related Projects. Community Services. Projects focusing on ...

[About R](#)
 R is a language and environment for statistical computing and ...

[FAQs](#)
 R FAQ - R for Windows FAQ - R for Mac OS X FAQ - ...

R (linguagem de programação) – Wikipédia, a enciclopédia livre

12 DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1



R-3.3.1 for Windows (32/64 bit)

[Download R 3.3.1 for Windows](#) (70 megabytes, 32/64 bit)

[Installation and other instructions](#)

[New features in this version](#)

If you want to double-check that the package you have downloaded exactly matches the package distributed by R, you can compare the [md5sum](#) of the .exe to the [true fingerprint](#). You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

Frequently asked questions

- [Does R run under my version of Windows?](#)
- [How do I update packages in my previous version of R?](#)
- [Should I run 32-bit or 64-bit R?](#)

Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.

Other builds

- Patches to this release are incorporated in the [r-patched snapshot build](#).
- A build of the development version (which will eventually become the next major release of R) is available in the [r-devel snapshot build](#).
- [Previous releases](#)

Note to webmasters: A stable link which will redirect to the current Windows binary release is [<CRAN MIRROR>/bin/windows/base/release.htm](#).

Last change: 2016-06-21, by Duncan Murdoch

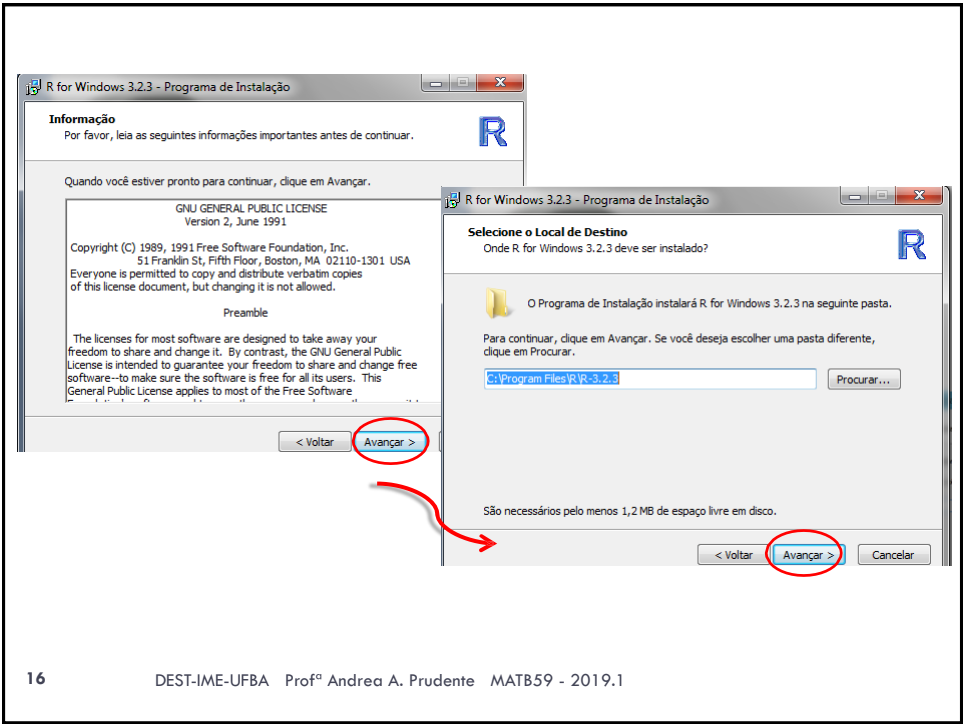
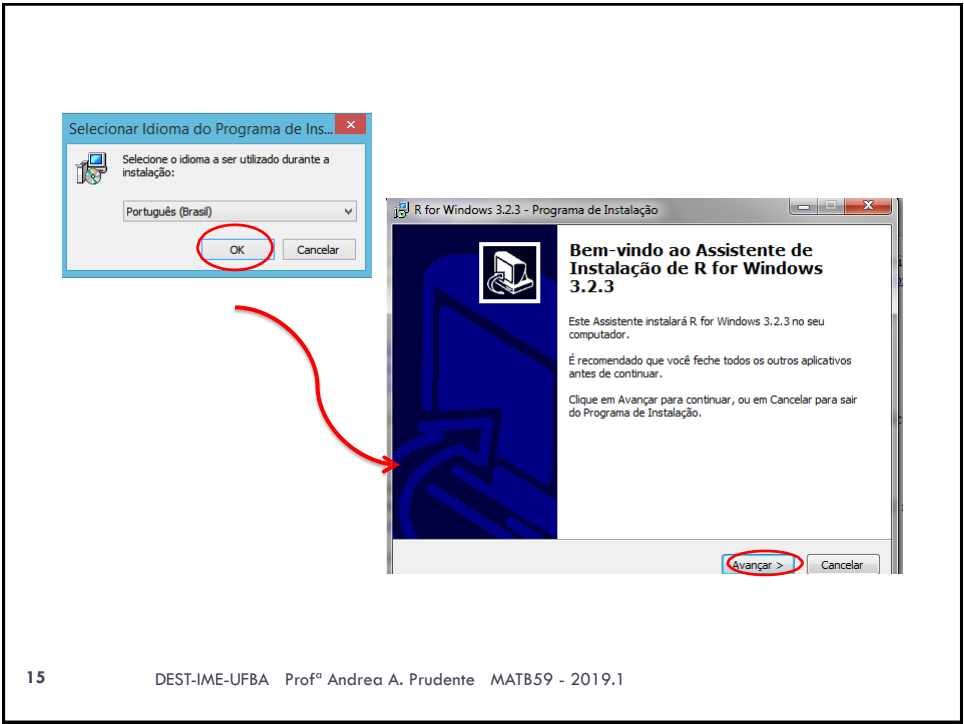
13 DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

14

Instalando o R

Após selecionar o idioma **“Português”** clique em **“Avançar”** até a última janela denominada Selecionar Tarefas Adicionais. Com o término da instalação, aparecerá uma janela de finalização do instalador, clique em **“Concluir”**. A partir daí, o R já pode ser usado.

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1



17

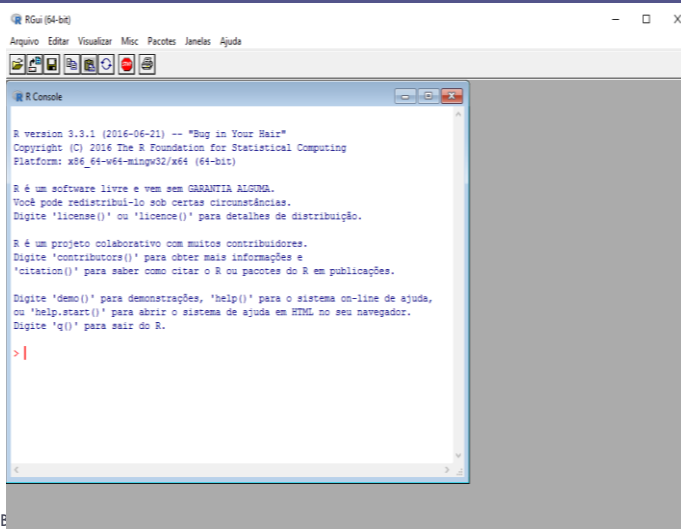
Iniciando o R

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Interface do R

18

Com o R iniciado você verá o símbolo “>|” em vermelho, que é o prompt do R (conhecido também como R Console) indicando que o sistema está pronto para receber seus comandos.



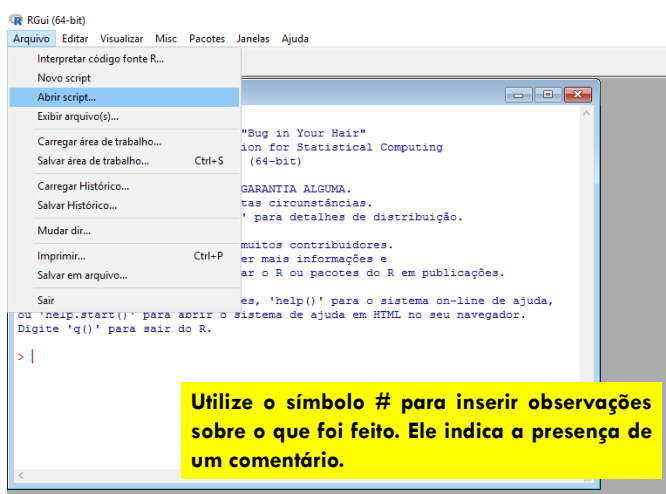
DEST-IME-UFBA

Script do R (Editor do R)

19

Uma maneira que otimiza o uso do R e que poupa muito tempo é usar um editor de texto para digitar seus comandos (script).

Para que o comando digitado no script seja enviado para a linha de comando do R basta apertar **Ctrl+r**.



DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Ajuda do R

20

- Para obter ajuda sobre algum comando basta digitar `help(nome da função)` ou `?nome da função`.

Exemplo

`help(mean)` ou `?mean`

- Para iniciar ajuda no browser padrão instalado basta digitar `help.start()`.

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

21

Base de dados

O R pode ler arquivos de texto (ASCII) além de outros formatos (Excel, SAS, SPSS etc.), porém, as funções necessárias à realização de algumas dessas operações não se encontram na biblioteca BASE.

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

22

Entrando com os dados

Pode-se entrar com dados no R de diferentes formas. O formato mais adequado vai depender do tamanho do conjunto de dados e se os dados já existem em outro formato para serem importados ou se serão digitados diretamente no R.

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Definindo vetores

23

- Esta forma de entrada de dados é conveniente quando:
 - ▣ se tem um pequeno número de dados; ou
 - ▣ quando se tem algum “padrão” tal como elementos repetidos ou números sequenciais (podem-se usar os mecanismos do R para facilitar a entrada dos dados como vetores).

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

- Podemos entrar com dados definindo vetores com o comando “c()” (c corresponde a “concatenate”), ou usando funções que criam vetores.

Exemplos:

```
a1 <- c(2,5,8) #cria vetor a1 com os dados 2, 5 e 8
a3 <- 1:10 #cria vetor com números sequenciais de 1 a 10
a5 <- rep(3,5) #cria vetor com elemento 3 repetido 5 vezes
a6 <- rep(c(5,8),3) #cria vetor repetindo 3 vezes 5 e 8
a7 <- rep(c(5,8),each = 3) #cria o mesmo vetor, mas de forma ordenada
```

24

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Usando a função `scan()`

25

- Esta função coloca o R em modo *prompt* onde o utilizador deve digitar cada dado seguido da tecla Enter. Para encerrar a entrada de dados basta digitar Enter novamente.
- Este formato pode ser mais ágil que o anterior e é conveniente para digitar vetores longos.

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Exemplo: utilizando a base de dados, `Empresas.xlsx`, copiar os dados da variável `tempo` (tempo de existência da empresa) e em seguida digitar no R

```
tempo <- scan()
```

e teclar o botão Enter, depois colar os dados selecionados e, novamente, teclar o Enter.

- Com esse procedimento podemos montar uma matriz com os dados de uma planilha eletrônica, basta importar quantas variáveis desejarmos.

26

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Exemplo: Utilizaremos mais uma variável empregado (número total de empregados - 1999). Copiar os dados da variável e em seguida digitar no R

```
empregado <- scan()
```

e teclar o botão Enter, depois colar os dados selecionados e, novamente, teclar o Enter.

Uma matriz com os dados poderá ser obtida fazendo-se:

```
dados <- cbind(tempo, empregado)
```

27

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

- As funções *rbind* e *cbind* são muito importantes no processo de criação de matrizes, pois possibilita a adição de novas linhas e colunas, respectivamente.
- É recomendado armazenar os dados em formato de *data.frame*, que é muito parecido com matrizes. Entretanto, diferentemente das matrizes, cada coluna pode armazenar elementos de diferentes tipos. Por exemplo, a primeira coluna pode ser numérica, enquanto a segunda, constituída de caracteres.

```
dados <- data.frame(dados)
```

28

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Usando a função readLines()

29

- Esta função é particularmente útil para ler entradas na forma de texto (*strings*).

Exemplo: Utilizaremos mais uma variável porte (porte da empresa). Copiar os dados da variável e em seguida digitar no R

```
porte <- readLines(n = 106)
```

e teclar o botão Enter, depois colar os dados selecionados e, novamente, teclar o Enter.

```
dados <- cbind(dados, porte)
```

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Usando a função edit()

30

- O comando `edit(data.frame())` abre uma janela para digitação de dados que são armazenados como `data.frame`, observe que nada mais é do que uma planilha eletrônica.

Exemplo:

```
a8 <- edit(data.frame())
```

Se for necessário abrir novamente a janela com os dados, para fazer correções e/ou inserir mais dados use o comando `fix()`.

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Lendo dados de um arquivo externo

31

- Uma tarefa muito utilizada no R é a leitura de arquivos externos (planilha do *Excel*, por exemplo). Se os dados já estão disponíveis em formato eletrônico, isto é, se já foram digitados em outro programa, podemos importar para o R sem a necessidade de digitá-los novamente.

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

- Primeiro faz-se necessário informar ao R em qual pasta (diretório) está seu arquivo.
 - Com o R aberto, clique no menu “Arquivo” e escolha a opção “Mudar dir...”, clicando sobre ela.
 - Aparecerá uma janela onde você deve buscar a pasta em que salvou o arquivo a ser lido, quando a encontrar, clique sobre ela e dê “Ok”.
 - Esse procedimento indica pasta/diretório onde o arquivo está localizado. Que nada mais é do que uma pasta que o R usa para gravar, ler, importar e exportar arquivos quando nenhum outro caminho é explicitado nas funções usadas para esses procedimentos.

32

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Lendo arquivo com extensão csv

33

- Para que os dados apareçam no *Console* do R devemos localizá-lo, atribuindo um nome ao objeto. No nosso exemplo, o arquivo foi salvo como 'Empresas'. Então, para ler um arquivo com extensão csv devemos fazer:

```
Dados <- read.csv("Empresas.csv", header=TRUE, sep=";")
mean(tempo)
mean(Dados[,7]) ou mean(Dados$tempo)
attach(Dados) #os objetos no banco de dados podem ser
               acessados por seus nomes (detach(Dados)).
DadosPeq <- subset(Dados, porte=="Pequena")
```

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

- Uma forma alternativa é copiar o endereço da pasta em que o arquivo foi salvo e colá-lo dentro da função, servirá como um 'caminho' até o arquivo.
- Para que o procedimento seja realizado com sucesso, devemos acrescentar uma barra "\" ao lado da existente ou invertê-la ("/").

Exemplo:

```
Dados<-read.csv("H:\\_UFBA\\_UFBA 2017.1\\MATB59\\Aulas
práticas\\R\\Empresas.csv", h=T, sep=";")
ou
Dados<-read.csv("H:/_UFBA/_UFBA 2017.1/MATB59/Aulas
práticas/R/Empresas.csv", h=T, sep=";")
```

34

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Algumas funções

35

- Somatório: `sum()`
- Produtório: `prod()`
- Elemento máximo: `max()`
- Elemento mínimo: `min()`
- Elemento máximo e mínimo: `range()`
- Número total de elementos: `length()`
- Raiz quadrada: `sqrt()`
- Valor absoluto: `abs()`
- Logaritmo na base 10: `log10()`

DEST-IME-UFBA Profª. Andrea A. Prudente MATB59 - 2017.1

Lendo arquivo com extensão *xls* ou *xlsx*

Comando `read_excel`

36

- O comando `read_excel()`, do pacote `readxl`, lê o conteúdo de uma planilha eletrônica para o R com a estrutura de dados de um `data.frame`. Porém, ele não consta na instalação base do R e, por isso, deve ser instalado antes de usar o comando `read_excel()`.

```
install.packages("readxl")
```

DEST-IME-UFBA Profª. Andrea A. Prudente MATB59 - 2019.1

- Após instalação do pacote, este precisa ser carregado.

```
library(readxl)
```

- O comando que lê a planilha é

```
dados<-read_excel("Empresas.xlsx", sheet = 1)
```

```
attach(dados) #os objetos no banco de dados podem
               ser acessados por seus nomes
```

```
detach(dados) #apaga da memória do programa,
               evitando possíveis erros quando
               for trabalhar com outro banco de
               dados
```

37

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Lendo arquivo com extensão *txt*

38

- A forma mais fácil de ler um conjunto de dados no R é trabalhar com arquivos *txt*, pois a leitura desse tipo de arquivo é nativa no R.

```
dados<-read.table("notas_1AV.txt", h=T, dec=",")
```

- Ainda é possível utilizar a função `read.csv` para ler arquivos no formato *txt*, mas faz-se necessário acrescentar o argumento `sep="\t"`

```
dados1<-read.csv("Nome do Arquivo.txt", h=T, sep="\t")
```

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

39

Estatística Descritiva

Aqui é uma orientação simples e direta de procedimentos utilizados na análise descritiva.

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Função table()

40

- Função utilizada para construir tabelas simples e de múltiplas entradas utilizando frequências simples absoluta.

```
empresas <- read.csv("Empresas.csv",header=TRUE,sep=";")
```

```
attach(empresas)
```

```
tabela1<-table(constituicao) #tabela simples
```

```
tabela2<-table(constituicao,porte) #tabela de dupla entrada
```

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Função prop.table()

41

- Função utilizada para construir tabela de frequências simples relativa.

```
prop.table(tabela1)
```

```
prop.table(tabela2)
```

- Para arredondar os valores das frequências, basta utilizar a função

```
round(prop.table(tabela1),4)*100
```

```
round(prop.table(tabela2),4)*100
```

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Função summary()

42

- Função utilizada para calcular as medidas descritivas: média, mediana, quartis, mínimo e máximo da variável.

```
summary(empregado)
```

- Outras medidas podem ser calculadas, como:

```
range(empregado) #amplitude total
```

```
var(empregado) #variância da variável empregado
```

```
sd(empregado) #desvio padrão da variável empregado
```

```
quantile(empregado,0.8) # Percentil 80
```

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

- É muito comum quando estamos coletando dados para montar a base de dados existir observações faltantes em uma determinada variável, isso pode causar alguns problemas, principalmente se forem variáveis quantitativas. Pois, o programa pode não apresentar as medidas descritivas quando solicitarmos, interpretando com NA (ou seja, valor ausente).

```
summary(empregado) #calcula as medidas e ao final indica a
                    presença de NA
```

```
mean(empregado, na.rm=T) #o programa entende que é para
                          remover o valor ausente (NA)
```

43

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Distribuição de frequências

44

```
notas<-read.table("notas_1AV.txt", h=T, dec=",")
attach(notas)
notasT1 <- subset(notas, turma==1)
notasT1<-notasT1$nota
```

```
summary(notasT1)
n<-length(notasT1);n
k<-1+3.3*log10(n);k
AT<-max(notasT1)-min(notasT1);AT
h<-AT/k;h
```

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Distribuição de frequências

45

```
install.packages("fdth")
library(fdth)

tab<-fdt(notas,start=3,h=1,end=9);tab
plot(tab) #histograma
plot(tab,type='cfp') #ogiva
plot(tab,type='rfp') #polígono de frequência
```

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Gráficos - Histograma

46

- No R, usamos a função *hist()*, para gerar um histograma.

```
hist(notasT1)
```

- É possível alterar o título, os títulos dos eixos, alterar a cor e etc.

```
hist(notasT1,xlab="Notas", ylab="Frequência",
col="pink", main="Histograma da nota da primeira
avaliação da turma 02 de MATB59")
```

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

- Percebemos que novos gráficos sobrescrevem o gráfico anterior, pois são plotados na mesma janela. Portanto, é necessário salvar o gráfico anterior antes de plotar um novo gráfico.
- É possível salvar os gráficos em PDF, JPEG, BMP, PNG ou outros formatos. Para salvar clique sobre o gráfico e depois clique em “Arquivo”, em seguida clique “Salvar como” e escolha o formato que deseja salvar.

47

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

- Uma alternativa é utilizarmos o comando `dev.new()` que cria uma nova janela gráfica e permite o surgimento de um novo gráfico sem sobrescrever o antigo. Sendo útil, também, na comparação entre gráficos.
- Este comando deve ser digitado antes da solicitação do novo gráfico, isso porque, é na nova janela que surge que o novo gráfico aparecerá, sobrepondo ao anterior.

48

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

- Para construir um histograma definindo os intervalos de classes basta acrescentar o argumento `breaks = c()` à função `hist()`.

```
hist2<-hist(notasT1, breaks = c(3,5,7,10), xlab="Notas",
ylab="Densidade", main="")
```

```
hist2$counts #retorna as frequências de cada classe
```

- É possível alterar os limites dos eixos.

```
hist(notasT1,breaks = c(3,5,7,10), xlim=c(0,10),
ylim=c(0,0.30), xlab="Notas", ylab="Densidade", main="")
```

49

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Gráficos – Box plot

50

- No R, usamos a função `boxplot()`, para gerar um *box plot*.

```
library(readxl)
dados<-read_excel("Empresas.xlsx")
attach(dados)
boxplot(empregado, ylab="Número de funcionários")
boxplot(tempo, xlab="Idade da empresa", ylab="Anos")
```

- Para plotar dois gráficos numa mesma janela.

```
par(mfrow = c(1, 2))
```

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

- Construindo um *box plot* envolvendo uma variável quantitativa e outra qualitativa.

```
boxplot(empregado~porte, xlab="Porte da empresa",
ylab="Número de funcionários")
```

```
boxplot(empregado~porte, xlab="Porte da empresa",
ylab="Número de funcionários", names=c("grande", "média",
"pequena"))
```

- É possível construir o *box plot* com apenas uma categoria da variável qualitativa.

```
boxplot(empregado~porte, subset = porte == "Pequena",
xlab="Empresas de pequeno porte", ylab="Número de
funcionários")
```

51

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Gráficos – Colunas e Barras

52

- No R, usamos a função *barplot()*, para gerar gráficos em colunas ou barras.

```
tabela1<-table(constituicao)
```

```
barplot(tabela1, xlab="Constituição jurídica da empresa",
ylab="Frequência")
```

```
barplot(tabela1,col=c("steelblue","skyblue1","greenyellow"),
xlab="Constituição jurídica da empresa", ylab="Frequência")
```

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Eixo y - Percentual

```
tabela1.prop<-round(prop.table(tabela1),4)*100
barplot(tabela1.prop, xlab="Constituição jurídica da
empresa", ylab="Percentual")
```

Gráfico em colunas agrupadas

```
tabela2<-table(constituicao,porte)
barplot(tabela2,legend.text = c("Individual", "Ltda.",
"S.A."), beside = TRUE)
```

Gráfico em barras

```
barplot(tabela1, ylab="Constituição jurídica da empresa",
xlab="Frequência", horiz = TRUE)
```

53

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Gráficos – Setores (Pizza)

54

- No R, usamos a função *pie()*, para gerar gráficos em setores.

```
pie(tabela1)
```

Acrescentando o percentual aos rótulos dos setores

```
rotulos<-paste(names(tabela1.prop), "(", tabela1.prop, "%)",
sep="")
pie(tabela1, labels=rotulos)
```

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Gráficos – *Dispersão*

55

- No R, usamos a função `plot()`, para gerar gráficos de dispersão.

```
anos_servico<-c(2,3,4,5,4,6,7,8,8,10)
```

```
clientes<-c(48,50,56,52,43,60,62,58,64,72)
```

```
plot(anos_servico, clientes)
```

```
plot(anos_servico, clientes, cex=1.5, pch=3)
```

```
?par #mais argumentos para a função plot()
```

```
cor(anos_servico, clientes) #calcula o coeficiente de correlação de Pearson
```

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

56

Distribuição de Probabilidade

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Distribuição Binomial

57

- Quando o número de tentativas (*size*) e a probabilidade de sucesso são conhecidos para cada evento (*prob*) é possível utilizar o comando abaixo para descobrir a probabilidade para qualquer valor da variável *x*.

`dbinom(x, size, prob)`

`pbinom(q, size, prob, lower.tail = TRUE)` # função de distribuição acumulada

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Distribuição Binomial

58

EXEMPLOS:

1. Seja 0,2 a probabilidade do nível de poluição do ar em certa região ultrapassar o limite de segurança. Qual a probabilidade de tal ocorrência em 7 dias num mês de 30 dias?

`dbinom(7, 30, 0.2)`

2. Três em cada quatro alunos de uma universidade fizeram cursinho antes de prestar vestibular. Se 16 alunos são selecionados ao acaso, qual é a probabilidade de que pelo menos 12 tenha feito cursinho?

`1 - (pbinom(11, 16, 3/4))`

ou `pbinom(11, 16, 3/4, lower.tail = FALSE)`

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Distribuição Binomial

59

EXEMPLOS:

3. Calcular a $P(45 < X < 55)$, sendo X uma variável aleatória com distribuição Binomial(100, 0.5)

`sum(dbinom(46:54, 100, 0.5))`

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Distribuição Normal

60

Seja X uma variável aleatória contínua. Dizemos que X tem distribuição normal com média μ e variância σ^2 , ou $X \sim N(\mu; \sigma^2)$.

Suponha que um pesquisador coletou dados de estatura de jovens em idade de alistamento militar. Sabe-se que a estatura de certa população segue a distribuição normal, com média 170 cm e variância 36 cm². O pesquisador pode assim escrever: $X \sim N(170; 36)$, em que X é a variável aleatória altura, com unidade em centímetros.

Qual a probabilidade de encontrarmos um jovem com mais de 1,79 metro de altura?

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Distribuição Normal

61

Quando se tem a média e o desvio padrão da população você pode utilizar o comando abaixo para descobrir a probabilidade para qualquer intervalo.

```
pnorm(x, mean, sd, lower.tail = TRUE)
```

No caso do exemplo, para descobrirmos qual a probabilidade de encontrar um jovem com mais de 1,79 metro de altura, precisamos digitar o seguinte comando:

```
pnorm (179, mean = 170, sd = 6, lower.tail = FALSE)
```

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1

Distribuição Normal

62

Imagine que se tenha uma população com média 150 cm e um desvio padrão de 10 cm, para descobrir o valor da distribuição que acima deste valor temos 5% de probabilidade devemos usar o seguinte comando:

```
qnorm (0.95, mean = 150, sd = 10)
```

DEST-IME-UFBA Profª Andrea A. Prudente MATB59 - 2019.1