

Laboratório de Estatística

Lupércio França Bessegato
Dep. de Estatística/UFJF

Introdução ao Uso da Linguagem R



Roteiro Geral

1. Fundamentos da linguagem R
2. Probabilidade em R
3. Gráficos em R
4. Tópicos especiais
5. Referências



2

Laboratório de Estatística - 2019



Introdução

- Ferramenta de análise estatística de dados com muitos recursos e possibilidades;
- Constantemente atualizado e revisado incluindo novas técnicas e métodos;
- Programa de domínio público:
 - ✓ <http://cran.r-project.org>



4

Laboratório de Estatística - 2019

Download do Programa

The screenshot shows the 'Download and Install R' page. It lists precompiled binary distributions for Windows and Mac users. A red arrow points from the 'Windows' option in the dropdown menu to the 'Download R 3.3.2 for Windows' link. Another red arrow points from the 'base' link in the 'Subdirectories:' section to the 'base' link in the 'Frequently asked questions' section.

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:

- Linux
- MacOS X
- Windows

Source Code for
This directory contains
See [here](#) for a 64-bit
Note: CRAN does not
Subdirectories:
[base](#) [contrib](#)
Please do not submit
to Windows binaries.

- 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.

5

Laboratório de Estatística - 2019

Comandos

- Para solicitar uma tarefa do R podemos digitar uma linha de comando no console

The screenshot shows the R Console window. It displays the R version information and a help message about redistribution. A red arrow points from the 'sqrt(4)' command in the console to its output '[1] 2'. The console also shows the prompt '>'.

```
R version 2.11.0 (2010-04-22)
Copyright (C) 2010 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

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.

R é um projeto colaborativo com muitos contribuidores.
Digite 'contributors()' para obter mais informações e
'citation()' para saber como citar o R ou pacotes do R em publicação.

Digite 'demo()' para demonstrações, 'help()' para o sistema on-line
ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.
Digite 'q()' para sair do R.

> sqrt(4)
[1] 2
> |
```

6

Laboratório de Estatística - 2019

Características do R

- Todas as funções do R devem ser digitadas em letras **minúsculas**
 - ✓ O R é sensível a letras maiúsculas e minúsculas.
- Todas as palavras-chaves do R estão em letras minúsculas
- R usa um ponto “.” em vez de vírgula “,” quando há números com casas decimais.

7

Laboratório de Estatística - 2019

Alguns Comandos

- Soma → +
- Subtração → -
- Divisão → /
- Multiplicação → *
- Potenciação → ** ou ^
- Raiz quadrada de n → sqrt(x)
- Logaritmo natural → log(x)
- Logaritmo decimal → log10(x)

8

Laboratório de Estatística - 2019

 • Consulte Table of useful R commands
 ✓ http://www.ufjf.br/lupercio_bessegato

Table of Useful R commands

Command	Purpose	Command	Purpose
help()	Obtain documentation for a given R command	hist()	Produces a histogram
example()	Enter data manually to a vector in R	lattice()	Lattice command for producing a scatterplot
c(), scan()	Make arithmetic progress vector	rsq()	Determine the least-squares regression line
seq()	Generate of regular values	lm()	Analysis of variance (can use on factor)
rep()	Load (observed) data frame built in dataset	anova()	Other analysis of variance tools
data()	View dataset in a spreadsheet type format	gnls()	estimate parameters of a nonlinear model
View()	Display internal structure of an R object	nls()	gives (observed - predicted) for a model fit to data
str()	Load into a function an existing data file	residuals()	take a sample from a vector of data
read.table(), read.csv()	Read data from a file	sample()	represents a random sample of times
library(), require()	Load a library	sum()	produce running total of values for input vector
dim()	Give dimensions (# of rows/columns) of dataframes	ecdf()	build empirical cumulative distribution function
length()	Give length of a vector	dhists()	tools for binomial distributions
length()	List memory contents	dt()	tools for discrete distributions
ls()	Print names of objects from memory	dnorm()	tools for normal distributions
names()	List names of variables in a data.frame	pnorm()	tools for student t distributions
hist()	Command for producing a histograms	qt()	tools for chi-square distributions
histogram()	Lattice command for producing a histogram	pcpnorm()	hypergeometric distribution
image()	Make a 2D plot	prop.test()	inference for 1 proportion using normal approx.
table()	List all values of a variable with frequencies	chisq.test()	carries out a chi-square test
xtabs()	Cross-tabulation tables using formulas	fisher.test()	Fisher test for contingency table
mosaicplot()	Produce a mosaic plot	t.test()	standard two-sample test on population mean
cut()	Breaks a variable into larger bins	apply()	apply function to a column/split by factors
mean(), median()	Identify "center" of distribution	summary()	Display Summary and mean
by()	apply function to a column/split by factors	var()	standard deviation of vector
summary()	Display Summary and mean	sadef()	standard deviation of vector
var(), sd()	standard deviation of vector	sum()	Add up all values in a vector
sum()	Find the position of a quantile in a dataset	prop.table()	compute proportions from a contingency table
quantile()	Find the position of a quantile in a dataset	pt()	probabilities for 1-tail t-test
barplot()	Produce a bar graph	power.t.test()	power calculations for 1- and 2-sample t
boxplot()	Lattice command for producing boxplots	stovesc()	compute analysis of variance table for fitted model
bpplot()	Produce a bpplot		

Laboratório de Estatística - 2019

9

 Help

- Help no R é bastante completo e autosuficiente.
- Se não conhecemos o comando de alguma técnica podemos chamar a função *help*:
 ✓ `help(palavra-chave)` ou `?palavra-chave`

Laboratório de Estatística - 2019

10

 Alguns Comandos para Ajuda

- Sempre procure ajuda no Google!

```

help.start()          # abre página de ajuda geral
help(hist)           # ajuda sobre a função hist
?hist                # idem
help.search("variance") # ajuda para localizar string
??variance           # idem
apropos("test")      # lista funções contendo a string test
example(hist)        # mostra um exemplo da função hist

# procura por hist em manuais de ajuda e mailing lists
RSiteSearch("hist")

# vignettes de pacotes instalados e em uso
vignette()           # mostra vignettes disponíveis
vignette("foo")       # mostra vignette específica

# ajuda sobre conjunto de dados de pacotes instalados e em uso
help(datasetname)    # para conjunto de dados datasetname
help(faithful)        # para conjunto de dados faithful
  
```

Laboratório de Estatística - 2019

11

 Help na Web

- R Project
 ✓ <http://www.r-project.org>
- Tutorial de Introdução ao R
 ✓ <http://ecologia.ib.usp.br/bie5782/doku.php?id=start>
- R Graphical Manual
 ✓ http://www.imsbio.co.jp/RGM/R_image_list?page=1&sort=-

Laboratório de Estatística - 2019

12

Armazenando dados

- Tipos de dados:
 - ✓ Numéricos
 - ✓ Caracteres: compostos por letras ou palavras.
 - ✓ Lógicos:
- Quando os dados são armazenados, eles são chamados de objetos.
- Para armazenar um objeto basta utilizar o símbolo “`<-`” ou “`=`”.

Laboratório de Estatística - 2019

13

Tipos de Objetos

- Variável
- Vetor
- Matriz
- Data Frame
- Array
- Lista
- Fatores

Laboratório de Estatística - 2019

14

Comando de Atribuição – Variável

- Atribuição de valor à variável `x` (ou qualquer outro ‘nome’)


```
x <- 4      # atribui o valor 4 à variável x
x1 <- -4     # atribui o valor -4 à variável x1
x2 <- 6      # atribui o valor 6 à variável x2
x; x1; x2    # print no console das variáveis x, x1 e x2
```

 - ✓ Evite usar acentos ou ç na denominação das variáveis
- Pode-se aplicar funções e operações à variável


```
z <- x1 + x2 # atribua à variável z a soma de x1 e x2
```

Laboratório de Estatística - 2019

15

Operações com as Variáveis

- Calcular:
$$x^{x1} + \frac{z}{x2}$$

```
x^x1 + z/x2  # calcula a expressão
[1] 0.3372396 # resultado da expressão
```

Laboratório de Estatística - 2019

16



Escolha de Nomes de Objetos



- Três princípios para escolha do nome:
 - Deve significar alguma coisa
 - Deve ser o mais curto possível
 - Equilíbrio (1) e (2)
 - Use seu discernimento

17

Laboratório de Estatística - 2019



Nomes Inválidos



```

24e<-4 #O nome do objeto começa com números
12<-2 #O nome do objeto é um número
e*2<-
x->5 #A seta está para o lado errado
  
```

18

Laboratório de Estatística - 2019



Apagar Objetos



- Comando `rm(objeto)`
 - ✓ Cuidado! Não tem undo
 - ✓ `rm(x, x1, x2, z)` ou `rm(x); rm(x1); rm(x2); rm(z)`
- Apagar tudo!
 - ✓ `rm(list=ls(all=TRUE))`.

19

Laboratório de Estatística - 2019



Limpeza do Console



- Para limpar o console usa-se **CTRL + L**.
-

20

Laboratório de Estatística - 2019

Vetores

- Coleção unidimensional de dados pontuais de mesmo tipo (números, strings, lógico)
- Comando para armazenar dados em vetor
 $\sqrt{c(dados)}$ # concatenar

```
xNum <- c(1, 2, 5.3, 6, -2, 4)          # vetor numérico
xCar <- c("um", "dois", "três")        # vetor de caracteres
xLog <- c(TRUE, TRUE, TRUE, FALSE, TRUE, FALSE) # vetor lógico
```

Laboratório de Estatística - 2019

21

Vetores – Indexação

- Elementos particulares de uma estrutura de dados

<code>xNum[2]</code>	# retorna o 2º elemento do vetor
<code>xCar[c(1, 3)]</code>	# retorno o 1º e o 3º elementos do vetor

Laboratório de Estatística - 2019

22

Diretório de Trabalho

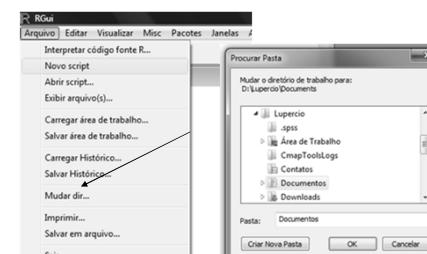
- Comando para verificar o diretório de trabalho que o R está usando:
 $\sqrt{getwd()}$
- Ideal sempre deixar scripts e dados de trabalho no mesmo diretório!
- Comando para mudar o R para seu diretório
 $\sqrt{setwd("caminho_ate_diretorio")}$
 $\sqrt{\text{Ex.: "D:/Lupercio/Documents"}}$

Laboratório de Estatística - 2019

23

Mudança Diretório – Barra de Ferramentas

- Sugestão:
 $\sqrt{\text{Sempre mude para seu diretório de trabalho quando iniciar a sessão em R}}$
 $\sqrt{\text{Guarde nele seus dados, gráficos, scripts, etc}}$



Laboratório de Estatística - 2019

24

Outros Comandos Úteis

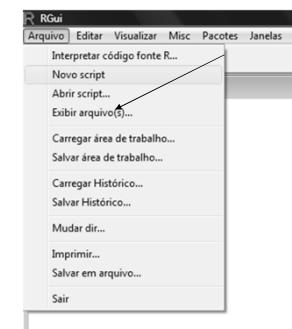
- Verifique se mudança funcionou com o comando `getwd()`:
- Verifique se seu diretório de trabalho está vazio:
✓ `dir()`
- Verifique se o workspace está vazio:
✓ `ls()`

Laboratório de Estatística - 2019

25

Uso de Script

- Facilita para:
 - ✓ Correção ou expansão de comandos
 - ✓ Repetição de comandos
 - ✓ Armazenamento de resultados



Laboratório de Estatística - 2019

26

Compilação do Script



- Usa-se a tecla F5 para compilar:
 - ✓ A linha em que se encontra o cursor (no script)
 - ✓ As linhas selecionadas (no script)
- Resultado compilação no console

Laboratório de Estatística - 2019

27

Vantagens

- Facilidade para corrigir os comandos ou valores
- Possibilidade de armazenar todos os resultados
- Repetição dos passos corretos de toda a sessão de trabalho
- Trabalho ‘limpo’
 - ✓ Para limpar o console usa-se `CTRL + L`

Laboratório de Estatística - 2019

28