

Laboratório de Estatística

Lupércio França Bessegato
Dep. de Estatística/UFJF

Roteiro Geral



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


Introdução ao Uso da Linguagem R




Introdução

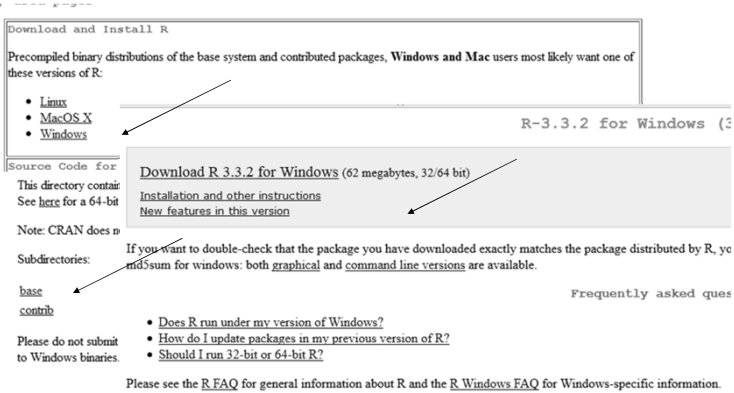


- Ferramenta de análise estatística de dados com muitos recursos e possibilidades;
- Constantemente atualizado e revisado incluindo novas técnicas e métodos;
- Programa de domínio público:
√ <http://cran.r-project.org>



Download do Programa





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.

[Download R 3.3.2 for Windows](#) (62 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 use 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.

Laboratório de Estatística - 2019

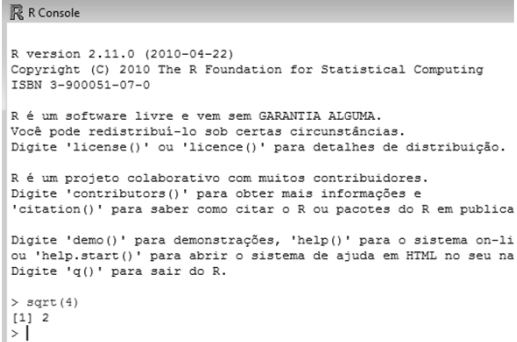
5



Comandos



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



```
R Console

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 publica

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

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

Laboratório de Estatística - 2019


6




- 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 virgula “,” quando há números com casas decimais.

Laboratório de Estatística - 2019

7



Alguns Comandos



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

Laboratório de Estatística - 2019

8



• 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	plot()	Produce a scatterplot
example()	View some examples on the use of a command	xyplot()	Lattice command for producing a scatterplot
c(), scan()	Enter data manually to a vector in R	lm()	Determine the least-squares regression line
seq()	Make arithmetic progression vector	aov()	Analysis of variance (can use results of lm())
rep()	Make vector of repeated values	predict()	Obtain predicted values from linear model
data()	Load (often into a dataframe) built-in datasets	summary()	estimate parameters of a nonlinear model
view()	View dataset in a spreadsheet-type format	residuals()	given (observed - predicted) for a model fit to data
str()	Display internal structure of an R object	sample()	take a sample from a vector of data
read.csv(), read.table()	Load into a dataframe an existing data file	replicate()	repeat some process a set number of times
library(), require()	Make available as R add-on package	cumsum()	produce running total of values for input vector
dim()	See dimensions (if of rows/columns) of dataframe	sd()	build empirical cumulative distribution function
length()	Give length of a vector	dnorm(), etc.	tools for normal distributions
ls()	List memory contents	pnorm(), etc.	tools for normal distributions
rm()	Remove an item from memory	qnorm(), etc.	tools for normal distributions
names()	List names of variables in a dataframe	pch(), etc.	tools for chi-square distributions
hist()	Command for producing a histogram	binom.test()	hypothesis test and confidence interval for 1 proportion
histogram()	Lattice command for producing a histogram	prop.test()	inference for 1 proportion using normal approx.
sum()	Make a sum plus	chisq.test()	carries out a chi-square test
table()	List all values of a variable with frequencies	fisher.test()	Fisher test for contingency table
xtabs()	Cross-tabulation tables using formulas	t.test()	student's t test for inference on population mean
mosaicplot()	Make a mosaic plot	qqnorm(), qqline()	tools for checking normality
cut()	Group values of a variable into larger bins	addmargins()	adds marginal sums to an existing table
mean(), median()	Identify "center" of distribution	prop.table()	compute proportions from a contingency table
by()	apply function to a column split by factors	par()	query and sets graphical settings
summary()	Display 5-number summary and mean	power.t.test()	power calculations for 1- and 2-sample t
var(), sd()	Find variance, sd of values in vector	anova()	compute analysis of variance table for fitted model
sum()	Add up all values in a vector		
quantile()	Find the position of a quantile in a dataset		
barplot()	Produce a bar graph		
barhplot()	Lattice command for producing bar graphs		
boxplot()	Produce a boxplot		
boxplot()	Lattice command for producing boxplots		



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`



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(datasestname) # para conjunto de dados datasestname
help(faithful) # para conjunto de dados faithful
```



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=-



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 “=”.



Tipos de Objetos



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



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
```



Operações com as Variáveis



- Calcular:

$$x^{x1} + \frac{z}{x2}$$

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



Escolha de Nomes de Objetos



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



Nomes Inválidos



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



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)).`



Limpeza do Console



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



Vetores



- Coleção unidimensional de dados pontuais de mesmo tipo (números, strings, lógico)

- Comando para armazenar dados em vetor
✓ `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
```



Vetores – Indexação



- Elementos particulares de uma estrutura de dados

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



Diretório de Trabalho



- Comando para verificar o diretório de trabalho que o R está usando:

✓ `getwd()`

- Ideal sempre deixar scripts e dados de trabalho no mesmo diretório!

- Comando para mudar o R para seu diretório

✓ `setwd("caminho_ate_diretorio")`

✓ Ex.: `"D:/Lupercio/Documents"`



Mudança Diretório – Barra de Ferramentas



- Sugestão:
 - ✓ Sempre mude para seu diretório de trabalho quando iniciar a sessão em R
 - ✓ Guarde nele seus dados, gráficos, scripts, etc





Outros Comandos Úteis



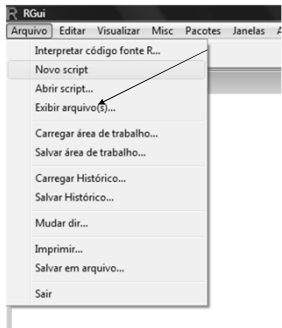
- Verifique se mudança funcionou com o comando `getwd()` :
- Verifique se seu diretório de trabalho está vazio:
√ `dir()`
- Verifique se o workspace está vazio:
√ `ls()`



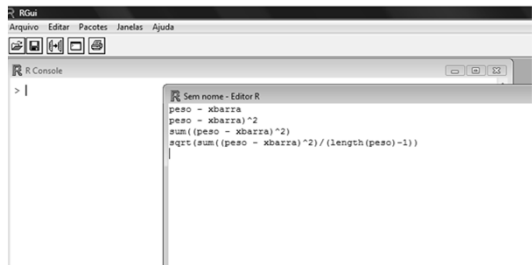
Uso de Script



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



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



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