

TALLER PRÁCTICO E BÁSICO DE R

III Xornadas de Usuarios R en Galicia



M^a José Ginzo Villamayor
mariajose.ginzo@usc.es

Dpto. DE ANÁLISE MATEMÁTICA, ESTATÍSTICA E OPTIMIZACIÓN
SERVIZO DE CONSULTORÍA ESTATÍSTICA



DEPARTAMENTO DE ANÁLISE
MATEMÁTICA, ESTATÍSTICA
E OPTIMIZACIÓN





Contidos

A linguaxe de programación R

Representacións gráficas

Estatística básica

R Commander

Bibliografía

Curiosidades



Que é R?

- Forma parte do parte do proxecto GNU e se presenta como software libre (*liberdade dos usuarios para executar, copiar, distribuír, estudar, cambiar e mellorar o software*).
- **R** é unha linguaxe derivada de S (Bell Labs.) para o tratamento de datos, análise estatística e gráfica.
- Linguaxe estatístico e matemático.
- Entorno para o desenvolvemento de novos métodos de análise de datos.
 - **gratuíto** (\neq baixa calidade)
 - **dinámico** (\neq inestable)
- The **R** Project for Statistical Computing: <http://www.r-project.org>
- The Comprehensive **R** Archive Network: <http://cran.es.r-project.org/>



Que é R?

The screenshot shows the R GUI interface. The console window contains the following R code:

```
> ## volcano ## 87 x 61 matrix
> wireframe(volcano, shade = TRUE,
+          aspect = c(61/87, 0.4),
+          light.source = c(10,0,10))
>
> g <- expand.grid(x = 1:10, y = 5:15, gr = 1:2)
> g$z <- log((g$y*g$y + g$y^2) * g$gr)
> wireframe(z = x * y, data = g, groups = gr,
+          scales = list(arrows = FALSE),
+          draps = TRUE, colorkey = TRUE,
+          screen = list(z = 30, x = -60))
>
> cloud(Sepal.Length ~ Petal.Length + Petal.Width | Species, data = iris,
+       screen = list(x = -90, y = 70), distance = .4, zoom = .6)
> ## volcano ## 87 x 61 matrix
> wireframe(volcano, shade = TRUE,
+          aspect = c(61/87, 0.4),
+          light.source = c(10,0,10))
>
>
>
>
>
```

The 3D plot, titled "Device2 (ACTIVE)", displays a colorful surface of a volcano. The axes are labeled "column" and "row".

Below the console window, the documentation for the `mean` function is visible:

Description:
Generic function for the (trimmed) arithmetic mean.

Usage:
`mean(x, ...)`

Default S3 method:
`mean(x, trim = 0, na.rm = FALSE, ...)`

Arguments:
`x`: An R object. Currently there are methods for numeric/vectors and date, date-time and time interval objects, for data frames all of whose columns have a method. `C` vectors are allowed for 'trim = 0', only.

`trim`: the fraction (0 to 0.5) of observations to be trimmed from each end of 'x' before the mean is computed. Values of trim outside that range are taken as the nearest endpoint.



Por que R?

- **R** é unha plataforma independente e gratuíta.
- **R** pódese utilizar a distintos niveis.
- **R** é máis flexible que a maior parte do software utilizado en noutras ciencias.
- As análises non teñen que restrinxirse ás funcións implementadas na instalación básica.
- Continuamente pónense a disposición do usuario novas librarías coa implementación das metodoloxías máis modernas.



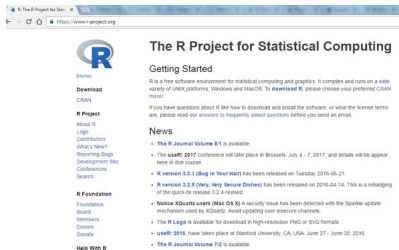
Instalación de R

- R pódese instalar en calquera sistema operativo con ou sen interfaz gráfica (Linux, MacOS, Windows, etc) sen importar a potencia do equipo.
- Desexable conexión a Internet para instalación, actualización e busca de novos procedementos.
- Instalación:
 - Arquivo instalación: 70 MB
 - Instalación básica: 172 MB
 - N° de paquetes totais: 9353
 - Instalación con todos os paquetes: \approx 2.5 GB



Instalación de R

O programa **R** está dispoñible para os principais sistemas operativos (Windows, Linux, MacOS, etc.), e pode descargarse en <http://www.r-project.org/>.



The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our [frequently asked questions](#) before you send an email.

News

- **The R Journal Volume 8(1)** is available.
- **The userR 2017** conference will take place in Brussels, July 4 - 7, 2017, and details will be appear here in due course.
- **R version 3.5.1 (Bug in Your Hair)** has been released on Tuesday 2016-06-21.
- **R version 3.3.8 (Very, Very Secure Dishes)** has been released on 2016-04-14. This is a rebadge of the quick-fix release 3.2.4-revised.
- **Notice XQuartz users (Mac OS X)** A security issue has been detected with the Sparkle update mechanism used by XQuartz. Avoid updating over insecure channels.
- **The R Logo** is available for download in high-resolution PNG or SVG formats.
- **userR 2016** have taken place at Stanford University, CA, USA, June 27 - June 30, 2016.
- **The R Journal Volume 7(2)** is available.

• Podes atopar:

- Distintas versións
- Paquetes
- Manuais de axuda
- Listas de correo ...

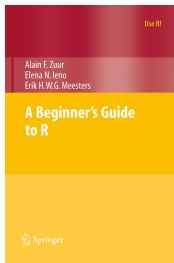
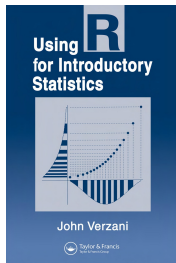


- <http://cran.es.r-project.org/>
 - Search-Menu CRAN
 - FAQ-Menu Documentation
 - Manuals-Menu Documentation
 - Contributed-Menu Documentation
 - Newsletter-Menu Documentation



Manuais, FAQs e publicacións

- A Beginner's Guide to R. Zuur, Ieno, Meesters. (2009)
- Using R for introductory Statistics. Verzani. (2005)

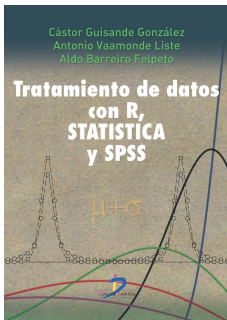


- Introducción a R. Descarga e instalación.
- Traballando con datos en R.
- Funcións simples.
- Ferramentas para gráficos.
- Programación básica en R.



Manuais, FAQs e publicacións

- Tratamiento de datos con R, STATISTICA y SPSS. Cástor Guisande González, Antonio Vaamonde Liste y Aldo Barreiro Felpeto (2011)



- Representación de datos, estadística descriptiva.
- Distribución, intervalos de confianza.
- Contrastes de hipótesis e contrastes de bondade de axuste.
- Contrastes de homoxeneidade en variables cualitativas, cuantitativas.
- Contrastes de independencia e asociación en variables cuantitativas.
- Regresión.
- Series de tempo.
- Análise multivariante: métodos factoriais, métodos de clasificación.
- Modelos de simulación.



Inicio das sesións en R

- Unha vez instalado, podemos xa executar o programa. Para iso, só fai falta facer click na icona ou desde o menú de programas seleccionar a icona correspondente.
- O primeiro que aparece na cabeceira é a versión que estamos a utilizar. Convén visitar habitualmente a páxina do proxecto para ver se hai unha nova versión. Se a hai e a queremos instalar, é recomendable que desinstalemos a anterior previamente.
- O comando `demo()` permite realizar demostracións do uso de R.
- Tal e como aparece no primeiro menú, R é un software libre e vén sen GARANTIA ALGUNHA, aínda que é posible redistribuílo baixo certas circunstancias. Máis info tecleando o comando `license()`.
- Mediante `citation()` podemos saber como citar R ou paquetes de R en publicacións e así dar o crédito á enorme cantidade de persoas que desenvolven este proxecto.



Entorno de traballo

Ao abrir o programa R, aparece a seguinte R-consola (fiestra de ordes) para traballo interactivo en modo comando:

```
RGui
Archivo  Editar  Visualizar  Misc  Paquetes  Ventanas  Ayuda
R Console
R version 2.14.1 (2011-12-22)
Copyright (C) 2011 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: i386-pc-mingw32/i386 (32-bit)

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribucion.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

> |
```

- Xusto despois da cabeceira, aparece unha liña en branco co símbolo > na marxe esquerda: *PROMPT*.
- A partir de aquí R espera que escribamos *COMANDOS* e instrucións para comezar a traballar.
- Para executar un comando escrito utilizamos a tecla de *intro*.
- O signo # indica a introdución dun comentario.
- As ordes elementais consisten en expresións ou en asignacións.



Entorno de traballo

- Na fiestra **R** Console cada liña na que o usuario pode introducir información iníciase co carácter `>` que pon o sistema **R**.
- As instrucións nunha mesma liña sepáranse por `;` para executar as instrucións que están nunha liña, púlsase a tecla *Retorno* ou a *Intro*.
- Pódense recuperar liñas de instrucións introducidas anteriormente premendo a tecla coa frecha ascendente do teclado, a fin de reexecutalas ou modificalas.

```
> 1+2*4  
[1] 9  
> (1+2)*4  
[1] 12  
> 3+5;3+6  
[1] 8  
[1] 9
```

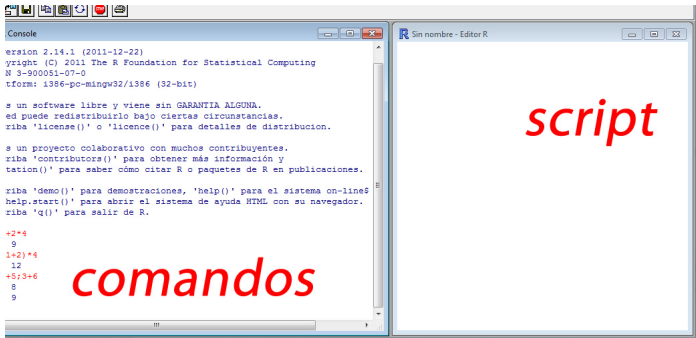


Fiestra de script

Un conxunto de instrucións poden integrarse nun único arquivo de texto para formar un *script* (este tipo de arquivos sóense gardar con extensión .R).

>Archivo<Nuevo script (Ctrl+N)

As instrucións do script cópanse e péganse na fiestra de comandos para ser executadas.





Fiestra de script

- Os arquivos script pódense gardar utilizando
`>Archivo>Guardar como ...`
e elixindo a continuación a ubicación que interesa.
- Por defecto **R** utiliza un cartafol de traballo onde gardará a información.
`>getwd() #devolve a carpeta de traballo`
- Para cambiar a carpeta de traballo pódese utilizar
`>Archivo>Cambiar dir....`
ou directamente por código utilizando `setwd` (carpeta). Por exemplo:
`> setwd("c:/")`



Editores de script

- Existen diferentes editores que poden facilitar o traballo con **R**: cores de sintaxes, completa parénteses, etc.
 - RStudio, TinnR
 - Notepad, (X)Emacs
 - RWinEdt, word
- **RWinEdt** é un editor para scripts dispoñible para usuarios de **R** no sistema de operativo Windows.
- Require que o editor **WinEdt** (programa de pago), <http://www.winedt.com>, estea instalado no computador.



- Outra alternativa é utilizar o editor **TinnR** (acrónimo de Tinn is not notepad) que é similar ao anterior pero de libre distribución.



-
-
-



Editores de script: RStudio

- RStudio é outro editor para scripts dispoñible para usuarios de R en todos os sistemas operativos.
- Os autores considéranlo un entorno de desenvolvemento integrado que combina unha interface moi intuitiva con ferramentas de código moi potentes que permiten sacar o máximo proveito a R.
- Pódese descargar de <http://www.rstudio.org/>
- Como os anteriores a vantaxe deste editor é que ofrece unha serie de opcións non existentes en R, entre outras, por exemplo, comprobar rapidamente que ningunha paréntese queda sen pecharse ou marcar, copiar e pegar columnas.
- Pero ademais danos unha listaxe das variables e danos unha descrición dos bancos de datos que introducimos. Tamén ten unha lista dos paquetes instalados e os gráficos realizados





Axuda en R

- `help.search("quero esto")`
- `help(function)`
- `?function`
- `help.start()` Abre a axuda en HTML



R como unha calculadora (sofisticada)

- R permite realizar cálculos aritméticos e, nese sentido, poderíase utilizar como unha potente calculadora:

```
x<-3+1
```

Asignación con: `<-`, `<<-`, `=`, `->`, `->>`.

- Podemos utilizar funcións matemáticas como:

```
y<-log(5)
```

```
z<-exp(x)
```

- Calquera función matemática que se nos ocorra está en **R**:
`log(x)`, `exp(x)`, `log(x,n)`, `log10(x)`, `sqrt(x)`, `factorial(x)`, `choose(n,x)`, `gamma(x)`, `lgamma(x)`, `floor(x)`, `ceiling(x)`, `trunc(x)`, `round(x,digits=0)`, `signif(x,digits=6)`, `cos(x)`, `sin(x)`, `tan(x)`, `acos(x)`, `asin(x)`, `atan(x)`, `acosh(x)`, `asinh(x)`, `atanh(x)`, `abs(x)` e podemos aplicala sobre calquer número.

- É "case sensitive":

```
var<-2*5
```

```
VAR<-2^3
```



Sobre o directorio de traballo

- En todo o manexo de arquivos, debemos ter en conta cal é o directorio de traballo (ou working directory).
- O máis importante é situar a **R** no directorio de traballo do noso ordenador, é dicir, o sitio onde localizaremos todos os datos, os resultados, os gráficos, ... da nosa análise.

```
getwd() # Directorio de traballo actual
```

- Podemos situarnos nun directorio concreto de dúas formas:
 - 1 Mediante a función `setwd()`. Como argumento desta función debemos escribir a ruta que conduce no noso ordenador ao directorio de traballo, entre comillas. Por exemplo,

```
setwd("D:/TallerR")
```
 - 2 Utilizando a opción **Archivo ► Cambiar dir...** do menú de **R**



As librarías en R

- Unha librería ou paquete é un conxunto de funcións sobre unha temática común. A instalación por defecto de **R** inclúe as librarías básicas implementadas polo grupo central de deseñadores.
- Cando se inicia **R**, só as funcións contidas en determinadas librarías do núcleo básico están dispoñibles para ser utilizadas.
- Sen embargo, moitas funcións interesantes non se instalan co paquete básico senón que están contidas en paquetes ou librarías externas que hai que instalar por separado.
- En moitas ocasións é necesario cargar librarías específicas que teñan implementadas algunhas funcións concretas.
- Para cargar unha librería hai que utilizar a función `library()`.
- Por exemplo para cargar a librería 'survival' (para a análise de supervivencia), que se instala con paquete básico pero non se carga ao inicio, habería que escribir `library(survival)`.



As librarías en R

- Existen numerosas librarías de funcións de **R** que poden sernos de utilidade e que hai que instalar manualmente.
- Para instalar paquetes de **R** (precísase conexión a internet):
 - Accedemos ao menú **Paquetes ► Instalar paquete(s)**
 - Seleccionar o CRAN-mirror, por comodidade podemos elixir o de España.
 - Logo seleccionamos o paquete que queremos instalar. Se o paquete necesita outros paquetes, instálalos automaticamente.
- A instalación non implica que os paquetes xa se poidan utilizar. É necesario cargar as librarías antes de empezar a usalas mediante a función **library()**.



Algunhas librarías

- Librería **stats**: Funcións estatísticas básicas
- Librería **graphics**: Funcións gráficas básicas
- Librería **languageR**: Conxuntos de datos lingüísticos e métodos estadísticos básicos aplicados a estes datos
- Librería **sp**: Funcións para traballar con datos espaciais
- Librería **maptools**: Funcións para traballar con datos espaciais e representación de mapas
- ...



Conxuntos de datos dispoñibles en R

- R trae coa instalación múltiples conxuntos de datos.
- Para ver un listado dos mesmos basta teclear `data()`
- Para cargar un conxunto en concreto emprégase o comando `data(name)`, por exemplo:

```
data(EuStockMarkets)
```

```
help(EuStockMarkets)
```




Importando datos en R

- Introducir datos a man pode converterse nunha tarefa moi pesada a pouco que o número de casos ou de variables sexa medianamente alto.
- Por outra parte, é bastante común ter os datos almacenados en arquivos non propios de R, senón doutros programas (Excel, bases de datos, etc.).
- Como parece natural, podremos facer que R lea estes datos.



Importando datos en R

- Os formatos de arquivos máis habituais nos que podemos atopar os datos son:
 - Arquivos de texto **.txt**.
 - Arquivos **.csv**.
 - Arquivos de Microsoft Excel **.xls**.
 - Arquivos de SPSS **.sav**.
 - Arquivos de BBDD como **.dbf**.
- Existen outros moitos formatos, pero casi sempre son convertibles a estes tipos. De feito, o propio Excel ou o seu análogo en OpenOffice, permiten transformar arquivos de texto en arquivos **.xls** e viceversa.



Importando datos en R. Ficheiros de texto

- Os arquivos de tipo texto que conteñen datos adoitan ter unha estrutura na que os individuos están nas distintas filas e as variables están separadas por algún tipo de carácter, tales como comas, tabulacións, espazos ou outros (**sep**)

```
e1<-read.table("EuStockMarkets_1.txt",sep=",")
```

```
e2<-read.table("EuStockMarkets_2.txt",sep="\t")
```

- É posible que a primeira fila conteña os nomes das variables (**header**)

```
e3<-read.table("EuStockMarkets_3.txt",sep="\t",header=TRUE)
```

- Tamén é necesario fixarse en como están especificados os decimais (**dec**)

```
e4<-read.table("EuStockMarkets_4.txt",sep="\t",header=TRUE,dec=",")
```

- Exercicio:** Os arquivos `houses.txt`, `twosample.txt`, `twoseries.txt`, `places.txt`, `mmreg.txt` conteñen diferentes conxuntos de datos. Ábreos cun editor de texto para comprobar a súa estrutura e impórtaos en R.



Principais obxectos en R

Principais tipos de obxectos en R

- numeric
- character
- matrix
- array
- logical
- factor
- data.frame
- list
- ...

Principais funcións relacionados co manexo de obxectos en R

- > `ls()` Lista os obxectos no directorio de traballo
- > `objects()`
- > `class(x)` Determina a clase do obxecto x
- > `apropos()` Búsqueda de obxectos
- > `rm(x)` Elimina o obxecto x



Operacións básicas en R

- Vemos o que creamos:

```
> x=3
```

```
> ls()
```

```
> objects()
```

```
> rm(x)
```

```
> rm()
```

```
> rm(list=ls())
```



Vectores en R

- Vector numérico
 - > `x<-c(1,2,5,6,9)`
- Vector carácter
 - > `nomes<-c("Luis","Mar")`
- Vector lóxico
 - > `lox<-c(F,F,T,F)`
- Vector enteiro
 - > `y<-2:6`
- Outras funcións para definir vectores
 - > `z1<-seq(0,8,by=4)`
 - > `z2<-seq(0,8,length=4)`
 - > `z3<-rep(5,4)`

Os argumentos das funcións en R

- **R** substitúe os argumentos das funcións na mesma orde na que están definidos a non ser que lle indiquemos explícitamente o nome do argumento.

Definición da función `seq`: `seq(from,to,by,length.out,...)`

> `seq(1,10,2)`

> `seq(1,2,length=2)`



Vectores en R

- Como escollemos as compoñentes dun vector

```
> x[2]
> x[-3]
> x[1:3]
> x[c(1,5)]
> x[x>4]
```

Algunhas funcións útiles

```
> length(x)
> min(x)
> max(x)
> sort(x)
> order(xx)
```

- Consulta a axuda para saber como utilízalas



Operacións con vectores R

- Operadores matemáticos: $+ - / \times$
- Para vectores da mesma dimensión, **R** realiza a operación desexada entre cada compoñente i dun vector coa correspondente compoñente i do outro vector.

```
> x+y
```

```
> x/y
```

```
> 6*x
```

- Que pasa cando operamos con vectores de distinta dimensión?

```
> x+z1
```




Matrices en R

- Como definir matrices
- Creando matrices a partir dun vector:

```
> m<-matrix(0,nrow=2,ncol=3)
> x<-1:6
> m1<-matrix(x,nrow=2)
> m2<-matrix(x,5,2)
```

(Replica o vector x ata completar a dimensión da matriz)

A función *matrix*

```
matrix(data = NA, nrow = 1, ncol = 1,
byrow = FALSE, dimnames = NULL)
```

- Exemplo:

```
Matriz <- matrix(c(1,2,3,11,12,13),nrow=2,ncol=3,
byrow=TRUE,dimnames=list(c("f1","f2"),
c("c1","c2","c3")))
```



Matrices en R

- Como escollemos as compoñentes dunha matrix

```
> A<-matrix(c(2,1,3,4),nrow=2,ncol=2)
```

```
> B<-matrix(c(3,8),nrow=2,ncol=1)
```

```
> C<-matrix(c(2,4,2,1),nr=2,nc=2)
```

```
> A[1,2]
```

```
> A[,1]
```

```
> A[2,]
```

$$A = \begin{pmatrix} 2 & 3 \\ 1 & 4 \end{pmatrix}$$

$$B = \begin{pmatrix} 3 \\ 8 \end{pmatrix}$$

$$C = \begin{pmatrix} 2 & 2 \\ 4 & 1 \end{pmatrix}$$

Algunhas funcións útiles

```
> dim(A)
```

```
> nrow(A)
```

```
> ncol(A)
```

```
> t(B)
```

```
> diag(nr=3)
```

- Consulta a axuda para saber como utilízalas



Operacións con matrices en R

- Operadores matemáticos: + - / * %*%
- Para matrices da mesma dimensión, **R** realiza a operación desexada entre cada compoñente (i,j) dunha matriz coa correspondente compoñente (i,j) da outra matriz

> A+C

> A/C

> A*C

- Que pasa cando operamos con matrices de distinta dimensión?
- Produto matricial (%*%) e produto compoñente a compoñente (*)

> A*B

> A%*%B

- Función outer

> outer(A,B,"+")

$$A = \begin{pmatrix} 2 & 3 \\ 1 & 4 \end{pmatrix}; B = \begin{pmatrix} 3 \\ 8 \end{pmatrix}; C = \begin{pmatrix} 2 & 2 \\ 4 & 1 \end{pmatrix}$$



Tipos de obxetos en R: data.frame

- Un **data.frame** é o tipo de obxeto de **R** que utiliza para almacenar conxuntos de datos.
- Un data.frame é unha táboa onde tipicamente as columnas úsanse para representar variables e as filas para representar individuos ou observacións.
- Creamos un data.frame cos datos Germany DAX (Ibis), Switzerland SMI do conxunto de datos d4,

```
datos<-data.frame(d4$DAX,d4$SMI)
```

```
class(datos)
```

```
names(datos)
```

```
names(datos)[1]<- "DAX"
```

```
names(datos)[2]<- "SMI"
```

```
names(datos)
```

```
fix(datos)
```

- Seleccionamos datos dun data.frame de distintas formas:

```
datos[1,] # Eleximos os datos da primeira fila
```

```
datos[,2] # Eleximos os datos da segunda columna
```

```
datos[,"DAX"] # Eleximos os datos da primeira columna
```

```
datos$DAX # Eleximos os datos da primeira columna
```



As funcións outer e apply

outer

```
> mifun<-function(x,y){cos(y)/(1+x^2)}  
> z<-outer(x,y,mifun)  
> class(z)
```

apply

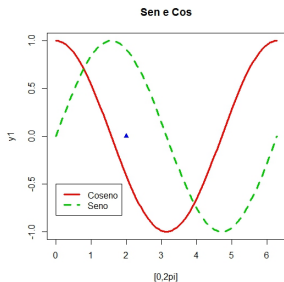
```
> m1  
> apply(m1,1,sum) #apply(objeto, dim, func)  
> clase<-c("a","b","a","a","b","a","b")  
> x<-c(1,2,1,1,2,1,2)  
> tapply(x,clase,mean)
```



Un exemplo de gráfico en R

```
> x<-seq(0,2*pi,length=100)
> y1<-cos(x)
> y2<-sin(x)

> plot(x,y1,col=2,lwd=3,xlab="[0,2pi]",type="l",main="Sen e Cos")
> lines(x,y2,col=3,lwd=3,lty=2)
> points(2,0,pch=17,col=4)
> legend(0,-0.5,c("Coseno","Seno"),col=2:3,lty=1:2,lwd=3)
```



Alguns argumentos útiles

- col** Color
- lwd** Anchura de liña
- lty** Tipo de liña
(continua,discontinua, ...)
- pch** Tipo de punto
- main** Título principal
- xlab** Título eixo X
- ylab** Título eixo Y



- En **R** pódense realizar multitude de gráficos.
- Algúns comandos para realizar gráficos:

| Función | Tipo de gráfico |
|---------------------------|----------------------------|
| <code>plot()</code> | Barras, dispersión, caixas |
| <code>hist()</code> | Histograma |
| <code>barplot()</code> | Diagrama de barras |
| <code>boxplot()</code> | Gráfico de caixas |
| <code>stripchart()</code> | Diagrama de dispersión |
| <code>pie()</code> | Gráfico de sectores |



Os gráficos en R

- A maioría das funcións gráficas dispoñen de argumentos para editar o gráfico:

| Argumentos gráficos | |
|--|---|
| <code>type = " "</code> | Especifica o tipo de gráfico: "p" puntos "l" liñas "n" non gráfico "o" puntos e liñas |
| <code>xlab = "texto"</code> <code>ylab = "texto"</code> | Engade texto ao eixe |
| <code>main = "texto"</code> | Título do gráfico |
| <code>xlim = c(mínimo, máximo)</code> <code>ylim = c(mínimo, máximo)</code> | Escalas dos eixos |

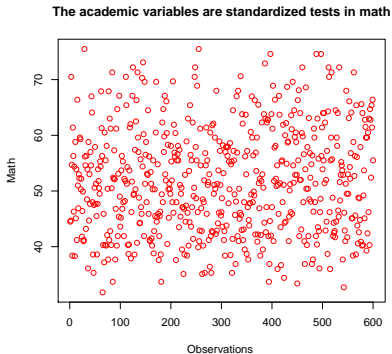
- A opción **Archivo** permite gardar (**Metafile, Postscript, PDF, Png, BMP y JPEG**), copiar ou imprimir a imaxe.



Gráficos de dispersión

- Comando “plot”

```
> plot(mmreg$math,col="red",xlab="Observations ",  
+ ylab="Math",main="The academic variables are  
+ standardized tests in math")
```

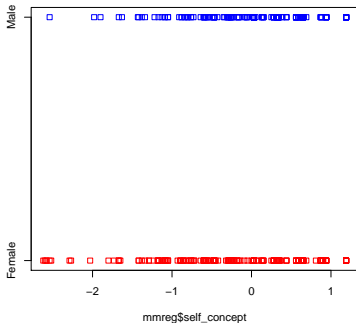




Gráficos de dispersión

- Comando “stripchart”

```
> stripchart(mmreg$self_concept~mmreg$sex,col=c("red","blue"))
```



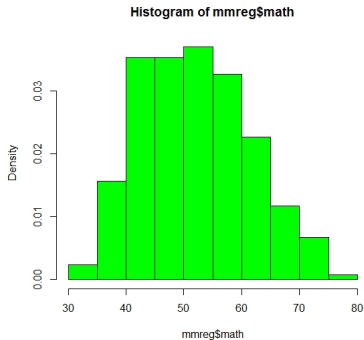
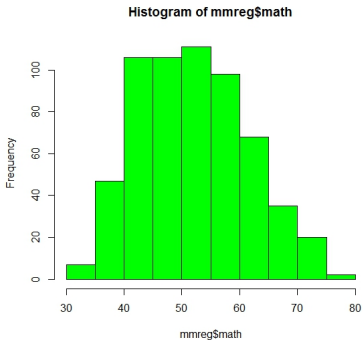


Histograma

• Comando "hist"

```
> hist(mmreg$math, col="green")
```

```
> hist(mmreg$math, freq=FALSE, col="green")
```





Diagramas de caixa

Os diagramas de caixa (boxplots) dan información visual sobre como están distribuidos os datos. Consta de:

- unha caixa central delimitada por los cuartiles Q_1 y Q_3 .
 - A liña do interior da caixa que representa a mediana (cuartil Q_2).
 - Os datos que caen fora dos bigotes represéntanse individualmente mediante "o" (datos atípicos).
- **Comando "boxplot"**

```
> boxplot(mmreg$write~mmreg$sex, col=c("red","blue"))
```

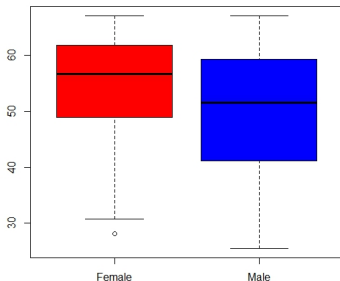




Diagrama de barras

- Comando “barplot”

```
> tab1<-xtabs(~mmreg$female+mmreg$motivationb)
```

```
> tab1
```

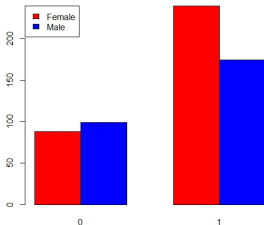
```
      no yes
```

```
Female 88 239
```

```
Male   99 174
```

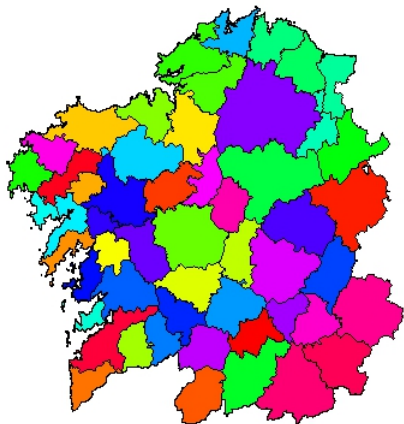
```
> barplot(tab1,beside=TRUE,col=c("red","blue"),
```

```
+ legend.text=TRUE,args.legend=list(x = "topleft"))
```



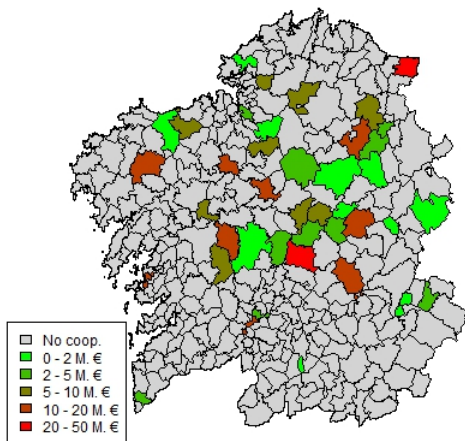


Comarcas de Galicia





Facturación cooperativas 2007





- Generación de variables (e vectores) aleatorias:

```
> x<-rnorm(100,0,1)
> mean(x)
> var(x) #É a cuasivarianza
> var(x)*(length(x)-1)/length(x)
> summary(x)
```

- Más estadística descriptiva:

```
> hist(x,freq=F)
> lines(density(x),col=3,lwd=2) #estima densidade
> rug(x)
> stem(x)
> boxplot(x)
> plot(ecdf(x),do.points=F,verticals=T)
```




- Más estadística descriptiva:

```
> shapiro.test(x)
> ks.test(x, "pnorm", mean=mean(x), sd=sqrt(x))
> y<-rexp(1000,rate=5)
> ks.test(y, "pexp", rate=1/mean(y))
> z<-rnorm(100,mean=5,sd=3)
> boxplot(x,z)
> t.test(x,z)
> var.test(x,z)
```



- Densidade, distribución, cuantís e xeración de variables aleatorias: d^* , p^* , q^* e r^* .

```
> help("rnorm")  
> x<-seq(-4,4,by=0.25)  
> plot(x,pnorm(x),type="l",col=2)  
> lines(x,dnorm(x))  
> y<-seq(0,1,length=100)  
> plot(y,qnorm(y))
```



- Regresión lineal:

```
> x<-rnorm(100,0,1)
> y<-2+3*x+rnorm(100,0,2*abs(x))
> plot(x,y)
> modelo<-lm(y~x)
> names(modelo)
> plot(modelo)
```



R Commander. Unha pequena (ou gran) axuda

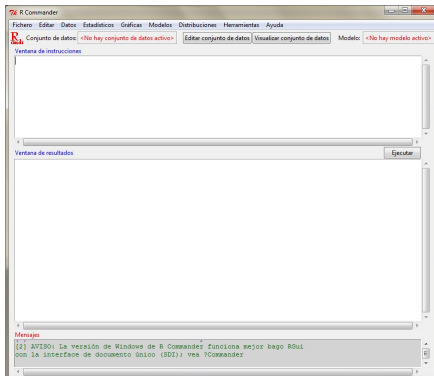
- R Commander é unha interfaz que cubre a maior parte das análises estatísticas máis habituais mediante menús despregables como os da maioría dos programas habituais.
- R Commander permite acceder a moitas capacidades da contorna estatística R sen que o usuario teña que coñecer a linguaxe de programación.
- Para utilizar R Commander, simplemente hai que cargar o paquete mediante a sentenza `library(Rcmdr)`.
- Tras cargar o paquete Rcmdr, iníciase a interfaz gráfica de R Commander.



R Commander. Unha pequena (ou gran) axuda

Una vez cargado R Commander vemos unha fiestra na que distinguimos:

- 1 O menú de fiestras desplegables.
- 2 A fiestra de instrucións: Cada vez que executemos algunha acción do menú, R Commander traducirá dita acción a código de R e escribirá na fiestra.
- 3 A fiestra de resultados: En xeral, calquer resultado de R Commander mostrarase aquí.
- 4 A fiestra de mensaxes: Serve para que R Commander nos informe de calquera aspecto, especialmente dos erros cometidos.





R Commander. Unha pequena (ou gran) axuda

- **Fichero:** Menú para abrir e gardar arquivos, abrir e gardar contornas de traballo de R, salir de **R** Commander,...
- **Editar:** Menú para editar contidos dun arquivo (copiar, pegar, cortar,...).
- **Datos:** Menú para ler e manipular datos.
- **Estatísticos:** Menú para análises estatísticas básicas.
- **Gráficos:** Menú para crear gráficos estatísticos básicos.
- **Modelos:** Menú para resumos estadísticos, intervalos de confianza, contrastes de hipóteses,...
- **Distribucións:** Probabilidades, cuantiles e gráficas de distribucións estatísticas habituais.
- **Ferramentas:** Menú para cargar outros paquetes de R ou establecer diferentes opcións de visualización de **R** Commander.
- **Axuda:** Información sobre **R** Commander e manual.



R Commander. Unha pequena (ou gran) axuda

- A maioría dos procedementos en **R** Commander requiren que haxa un conxunto de datos activo. Se existen varios conxuntos de datos cargados en memoria podremos eleixir entre eles, aínda que só un poderá estar activo.
- Formas de introducir datos en **R**:
 - Directamente a través do menú **Datos ► Novo conxunto de datos**.
 - Importando os datos dun arquivo de texto ou de outro programa.
 - Importando un conxunto de datos de R.



R Commander. Resúmenes numéricos e gráficos

- Una vez que hai un conxunto de datos activo, podemos usar os menús de **R Commander** para xerar varios resumos numéricos e gráficos.
- Pódese cambiar o conxunto de datos activo premendo no botón horizontal co nome co conxunto de datos activo.
- Seleccionando **Estadísticos ► Resúmenes ► Conxunto de datos activo**, obtense, para cada variable numérica no conxunto de datos, os valores máximos e mínimos, o primeiro e o terceiro cuartil, a mediana, a media, así como o número de valores que faltan. Para as variables categóricas, obtemos o número de observacións en cada nivel do factor.
- Observamos que o resultado obtense do comando **summary**.



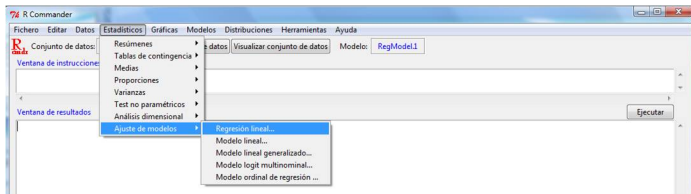
R Commander. Resúmenes numéricos e gráficos

- A realización de gráficos en **R** Commander é bastante intuitiva e lévase a cabo moi facilmente a través dos menús
 - Gráficas ▶ Histograma
 - Gráficas ▶ Diagrama de caixas
 - Gráficas ▶ Diagrama de dispersión
 - Gráficas ▶ Gráfica de barras
 - Gráficas ▶ Gráfica de sectores



R Commander. Modelos estadísticos

- R Commander permite axustar varios tipos de modelos estadísticos usando as opciones do menú **Estadísticos** ► **Axuste de modelos**





Máis sobre R Commander

- Baixo a fiestra de instrucións hai unha fiestra grande para os resultados que permite editar o texto e desprazarse arriba e abaixo sobre el.
- As instrucións escritas nesta fiestra aparecen en **vermello** e os resultados en **azul escuro** (como na consola de **R**).
- Ao final hai unha pequena fiestra de texto **gris** para mensaxes. As mensaxes de erro móstranse en vermello, os avisos en verde e outras mensaxes en azul escuro.
- Podemos reaxustar ou maximizar a fiestra **R Commander** da maneira usual. Se reaxustamos **R Commander**, o ancho dos resultados **R** subseguintes automaticamente axustarase para encaixar na fiestra de resultados.
- É posible cambiar a configuración de **R Commander** a través do menú **Herramientas**.



- A bibliografía de R é extensísima! Imposible dala toda enteira.
- Na páxina web do proxecto existe dispoñible moita *Contributed Documentation* que ademais é de libre distribución.
- Entre a enorme cantidade de libros (posiblemente na actualidade haxa máis de 100!) enteiramente dedicados a **R** destacamos unicamente dous:
 - John M. Chambers (2008), *Software for Data Analysis: Programming with R*. Springer, New York.
 - Peter Dalgaard (2008), *Introductory Statistics with R*, 2nd edition. Springer.
- Unha moi boa referencia é: Michael J. Crawley (2007), *The R book*. John Wiley and Sons, Ltd.



Curiosidades de R

- Existen dous congresos: “Use R” (anual e para usuarios) e “DSC, Directions in Statistical Computing” (bianual e para programadores) enteiramente adicados ao seu uso.
- Ata no New York Times escribiuse sobre as excelencias de R:
http://www.nytimes.com/2009/01/07/technology/businesscomputing/07program.html?_r=2
- Desde 2009 existe unha publicación [The R Journal](#) enteiramente dedicada a artigos sobre o desenvolvemento e a aplicación de R. O seu predecesor era [R News](#).
- A nivel nacional: Celebración das [VIII Xornadas de usuarios de R](#) anuais, <http://res.org/8jornadasR/>, (Albacete, 17-18 de novembro 2016).
- A nivel galego: Celebración das [III Xornadas de Usuarios R en Galicia](#), <https://www.r-users.gal/>, (Santiago de Compostela, 20 de outubro 2016).



TALLER PRÁCTICO E BÁSICO DE R

III Xornadas de Usuarios R en Galicia

M^a José Ginzo Villamayor
mariajose.ginzo@usc.es

Dpto. DE ANÁLISE MATEMÁTICA, ESTATÍSTICA E OPTIMIZACIÓN
SERVIZO DE CONSULTORÍA ESTATÍSTICA



DEPARTAMENTO DE ANÁLISE
MATEMÁTICA, ESTATÍSTICA
E OPTIMIZACIÓN

