ℝ and R Commander (Rcmdr)[1] *by the example*
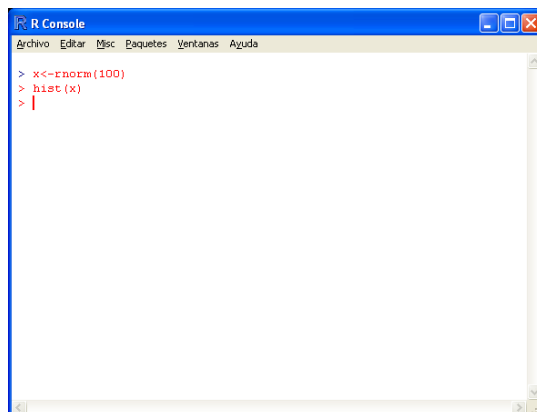
---

[1] Este documento ha sido adaptado a partir de otro (http://www.surgery.wisc.edu/admin/biostats.shtml), propuesto por Glen Leverson, Victoria Rajamanickam y Alejandro Muñoz del Río del departamento de "Surgery" de la Universidad de Wisconsin en Madison y que ya no se encuentra disponible en la red. El crédito es para ellos y los errores debo de haberlos introducido yo.

1. Introduction

The goal of this document is to be an introduction to using R "by-the-example". That is we assume the reader has already had some –probably short- exposition to R and R commander (Rcmdr from now on) and show how to use R to perform some typical biostatistical analyses.
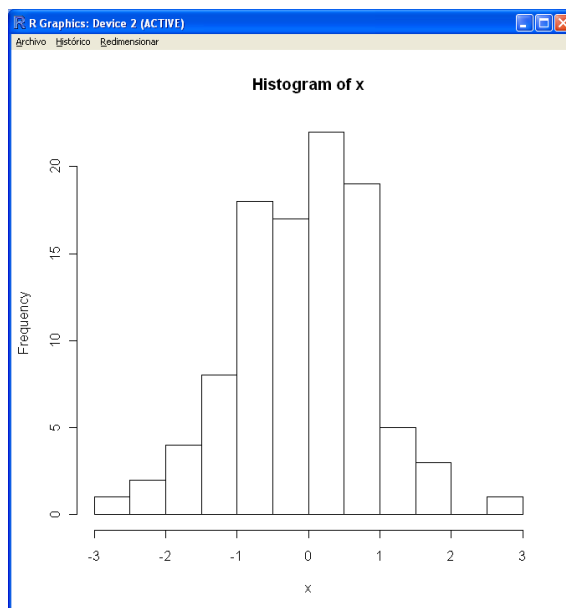
R is a console-based program, that is, you type the instructions you want to be executed

and the result –or an error message- appears in the screen.

This approach is very powerful but it has a slower learning curve than other typical statistical programs such as SPSS. To help new users to get acquainted with R capabilities, John Fox has developed a basic graphical user interface that allows using many R capabilities through a menu system. The name of this GUI is "R Commander" and it has been a successful project that can be followed at: (http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/). The project has indeed been so successful that many other developers have contributed their own extensions (called "plugins"). Altogether it has changed the R commander from a within-classroom application, to be used in Statistics courses, into a powerful

standalone system, that requires only having R available, allowing to use a great number of statistical techniques in an easy way. And for free, as R.

Conventions:
A sequence of words separated by arrows means the options being performed using the menus.
The "`OK`" button will not be mentioned, although it is usually the last thing that needs to be clicked on.

- `Courier:` Menu items, buttons, or text in dialog window, error or warning messages.
- *`Courier italic:`* Variables, R objects, user-supplied text; typed in or possibly double clicked
- ***`Large Courier bold italic:`*** First mention of a new variable.

Buttons will be shown as small boxes with the button text's within them **View data set**

The following pages assume we are working in windows and the operating system is in Spanish or Catalan, and so are the system menus of the R commander. Changes from this context are possible but should be easily manageable by any reader.

2. Starting R and Rcmdr

Starting R
  `Inicio -> Archivo de Programas -> R -> R.2.15`[2]

R basics:
  a. R is case sensitive: **`foo`**, **`Foo`**, and **`Foo`** would refer to three different objects
  b. R has variable/object types: numeric, character (factors), ordinal, dates, logical.
  c. R objects are persistent within an R session but not within sessions, unless the workspace is saved.
  d. The source code is real. The objects are realizations of ("the result of running") the source code: an object may be worthless unless we know how it was defined. Implication: **Save the code generated by your actions in your Rcmdr Script/Log window.**

Exiting R (5 ways):
`q( )`,
`q("no")`,
`q("yes")`,
from **`File -> Exit`** in RGui, or
from **`File -> Exit -> From Commander and R`** in Rcmdr

Warning: Exiting R from Rcmdr will not save the R workspace (the default memory zone where all objects that have been created during the session are still available).

Starting R Commander (Rcmdr)
  > **`library( Rcmdr)`**   # ">" is the R prompt, "#" the R comment character

  The five sections of the Rcmdr Window, top to bottom:
  1. Menu bar: File, Edit, etc.
  2. Tool bar: Active data set, active model; both are clickable.
  3. Script/Log window: R commands. May ignore, but good way of learning R.
  4. Output window: R output.
  5. Message window: R warnings and errors.

  Script and output windows are resizable, via **`Tools -> Options`**

---

[2]   or whatever version you have installed.
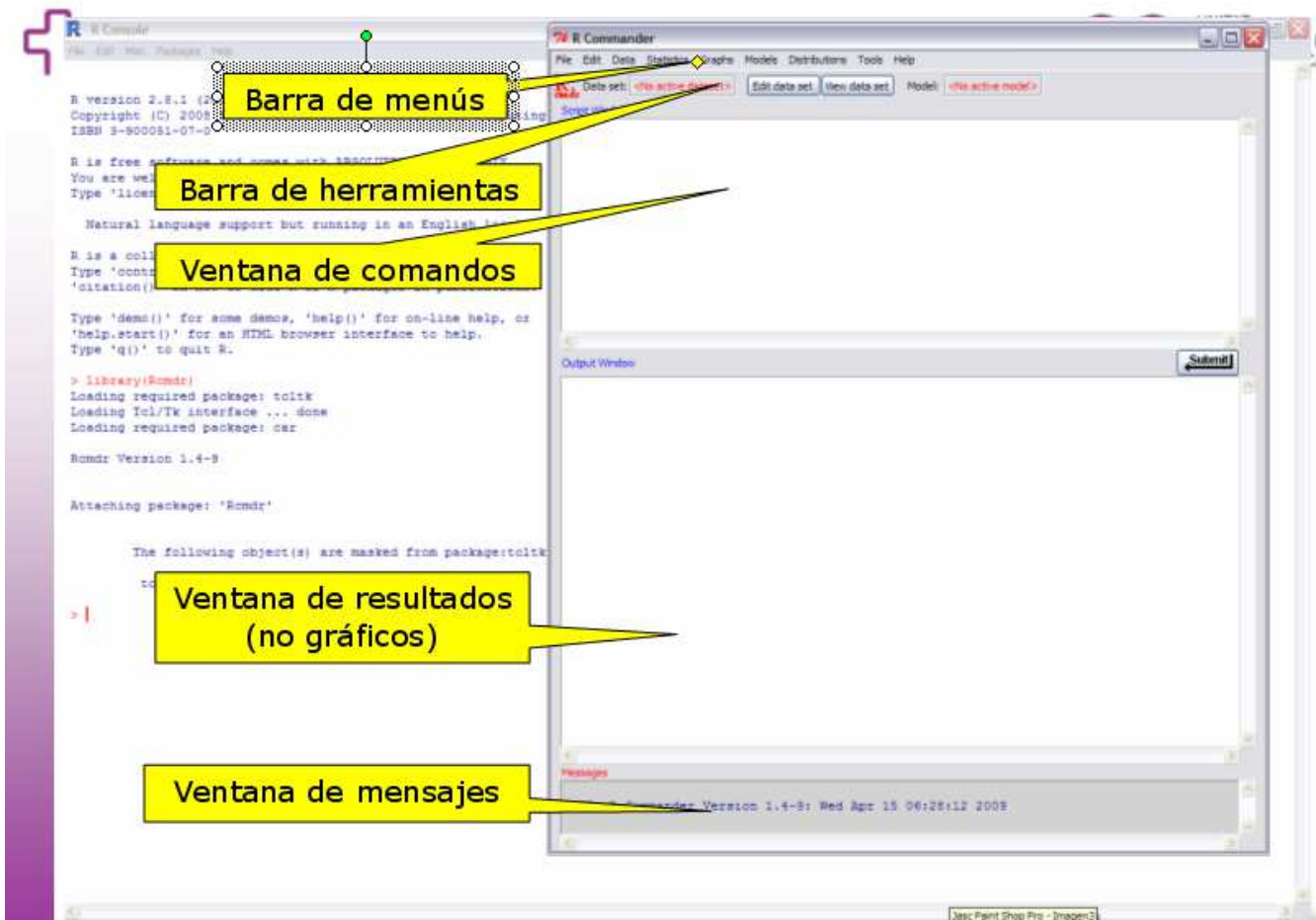
Exiting Rcmdr:
Fichero -> Salir -> De Commander y de R

(will be prompted to save script & output, **but not workspace!**)

3. Importing data (from Excel, SPSS, ...)

Although it is possible to enter and edit data using the R/Rcomander interface, most of the times the data will be available in other "popular" formats such as excel, access or SPSS.

The following examples assume the spreadsheet "VASopenlap.xls" is available in the *working directory,* that is the directory ("folder") where actions are applied if no other directory is specified.

The working directory can be set by doing:
        Fichero -> Cambiar directorio de trabajo … -> *<PATH TO MY DIRECTORY>*

Notice that, when this action is applied using the manus a new line of code appears in the Script window first, and in the Results Window after pressing the OK button.

```
> setwd("/home/alex/Documents/Classes/einesbioestbioinf/intro-r")
```

To import a data file we must work with the "Datos" menu. To import an Excel file[3] we proceed as follows:

> **Datos -> Importar Datos -> desde conjunto de datos Excel, Access o dBase...**

1. We will be prompted for a dataset name, Enter **vas** for name of data set
2. Choose Excel file "VASopenlap.xls"
3. Select the shhet to import (Sheet1)
   [Notice how, now, **vas** is active data set in Rcmdr Tool bar]

Exercise: Proceed similarly to import another excel file or a file with data from other statistical packages such as SPSS or STATA.

4. Viewing and Editing Data

Clicking on Visualizar Conjunto de Datos button in Tool bar shows data in the data set. This dataset cannot be edited because it contains non-editable (from within Rcmdr) data types such as dates. See next point to handle data.

5. Working with Variables *(may prefer to do this in Excel before importing)*

Create new variable:

> **Datos ->**
> **Modificar variables del conjunto de datos activo ->**
> **Calcular nueva variable**

bmi for name of new variable
double click on current variables to get **weight / height^2**
Similarly, compute **survyrs** as **(fudate - surgdate)/365.25**
Similarly, compute **VASdelta** as **preVAS - postVAS**

Group BMI into three equal-width groups:
Datos ->
Modificar variables del conjunto de datos activo ->
Segmentar variable numérica
Select **bmi**, name new variable **bmigroup**, select 3 bins, labeled by ranges

*Convert complication (**complic30d**) from numeric to factor*
Datos
-> Modificar variables del conjunto de datos activo ->
-> Convertir variables numéricas en factores
Under **Niveles del factor**, leave **Asignar nombres a los niveles** (default) marked;
0 = none, 1 = yes

---

[3]   The option for importing from "Excel" is not availabe in linux versions of R/Rcmdr. In this cas you must first export your spreadsheet as a tab-delimited or comma-separated file, and use this as the import source.

Convert surgery type (`surgtype`) from factor to numeric

```
Datos
    -> Modificar variables del conjunto de datos activo ->
        -> Recodificar variables
```

Select **`surgtype`**, type **`surgopen`** as new variable name, uncheck factor box, and type
`"open" = 1 and "lap" = 0` in the `Recode directives` part of the box.

6. Summary and Descriptive Statistics

Compute easy summaries (a different type of summary for different variable types).
```
Estadísticos -> Resúmenes -> Conjunto Activo de Datos.
```

Compute summary statistics at each level of a factor variable
```
Estadísticos -> Resúmenes -> Tabla de Estadísticas
```
Choose **`surgtype`** as factor, **`bmi`** as response

```
FACT     surgopen
 none:13   0:12
 yes : 9   1:10
```

Plot from simple to complicated figures

```
Gráficas -> Gráfico de tallo y hojas
```
Select **`bmi`** and accept all default options.

```
Gráficas -> Diagrama de caja
```
Select **`weight`**, as variable, **`bmigroup`** as grouping variable ("Graficas por grupo") and press OK

```
Gráficas -> Matriz de diagramas de dispersión...
```
Select >2 desired variables (at least three: **`weight, height, bmi`**,…)
Uncheck **`línea de mínimos cuadrados`**

7. Basic tests

7.1 Is surgery beneficial?

`Graficos -> Diagrama de dispersión... -> ` ***preVAS* (x)** vs. ***postVAS* (y)**
Check boxes to jitter both x & y
Click on "`Plot by groups...`" and select ***surgtype***

Paired Student's t-test of Pre- vs. Post-VAS scores
`Statistics -> Means -> Paired t-test, ` ***preVAS*** as 1st variable, ***postVAS*** as 2nd one.
P = 0.001116 (two-sided)

Non-parametric equivalent: **Wilcoxon signed rank test**
`Statistics -> Nonparametric tests -> Paired-samples Wilcoxon test`
Select ***preVAS*** as 1st variable, ***postVAS*** as 2nd one.
P-value = 0.003297 (default test)

Equivalent to paired t, *one sample Student's t-test* of VAS improvement
`Statistics -> Means -> Single-sample t-test`, select ***VASdelta*** as the response.
P = 0.001116 (two-sided)

7.2 Is there a difference in pre to post VAS changes by surgery type?

Scatterplot of VAS improvement vs. surgery type Note: x and y should both be numeric variables.
`Graphs -> Scatterplot... -> ` ***surgopen* (x)** vs. ***VASdelta* (y)**
Only checked box should be the one to jitter y (least-squares line will join means)

Boxplot of VAS improvement vs. surgery type Note: y is numeric, grouping must be factor.
`Graphs -> Boxplot...,` select ***VASdelta***,
Click on "`Plot by groups`" and select ***surgtype***

Two sample Student's t-test of VAS improvement by surgery type
`Statistics -> Means -> Independent samples t-test`,
select ***surgtype*** as the grouping variable, ***VASdelta*** as the response.
P = 0.7387 (Welch test = default, allows for unequal variances)
P = 0.7396 (variances assumed equal)

Non-parametric equivalent: **Wilcoxon rank sum test**
`Statistics -> Nonparametric tests -> Two-sample Wilcoxon test`
Select ***surgtype*** as grouping variable, ***VASdelta*** as response.
P = 0.5708

7.3 Is there a difference in VAS improvement due to BMI groups?

Boxplot of VAS improvement vs. BMI group
`Graphs -> Boxplot...,` select ***VASdelta***,
Click on "`Plot by groups`" and select ***bmigroup***

ANOVA (one-way)
`Statistics -> Means -> One-way ANOVA...,` check "`Pairwise comparison of means`"
select ***bmigroup*** as the grouping variable, ***VASdelta*** as the response.
P = 0.8209 (variances assumed equal; produces means, SDs, counts per group.)

Non-parametric equivalent: **Kruskal-Wallis test**
`Statistics -> Nonparametric tests -> Kruskal-Wallis test`
Select ***bmigroup*** as grouping variable, ***VASdelta*** as response.
P = 0.4892

Scatterplot of VAS improvement by BMI, to prevent sacrificing information due to binning
`Graphs -> Scatterplot,` select ***bmi*** (x) and ***VASdelta*** (y)
Uncheck "`Smooth line`"
Click on "`Plot by groups`", select ***surgtype*** and check "`Plot lines by group`"

7.4 Is the 30-day complication rate the same for both surgery types?

Test of two-proportions (Normal approximation)
`Statistics -> Proportions -> Two-sample proportions test`
Select ***surgtype*** as the grouping variable and ***complic30d*** as the response variable
P = 1, so no difference between observed complication rates (41.7% lap vs. 40% open)
Note warning: `"Chi square approximation may be incorrect."`

A better way: *Fisher's exact test*
`Statistics -> Contingency tables -> Two-way table`
Select ***surgtype*** as the row variable and ***complic30d*** as the column variable;
Select `row percentages`
Check "`Fisher's exact test`" and "`Print expected frequencies`"
P = 0.9362, and note two cells have expected totals < 5.

7.5 Can complications be predicted as a function of demographic and operative variables?

Logistic regression
`Statistics -> Fit Models -> Generalized linear models...`
Click on variables so that the resulting model is:
complic30d ~ bmi + height + sex + surgtype + weight + preVAS
Leave "`binomial`" and "`logit`" (defaults) under family and link function untouched.
No variable is significant. We confirm that surgery type does not affect complication rates (P=0.5030).

8. Exiting R and Rcmdr

Script: `File -> Save script…`
Output: `File -> Save output…`
Workspace: `File -> Save R workspace…`

9. Getting help, learning more (in suggested order for consulting)

Context-specific help in dialog boxes via `Help` button
R Commander Help (`Help` menu item)
Getting Started with Rcmdr PDF (`Help -> Introduction to the R Commander`)
R PDF manuals in `c:/path/to/R/version/doc/manual`; supply correct path as appropriate
Dalgaard, Peter (2002) *Introductory Statistics with R*. Springer. ISBN 0-387-95475-9

10. Installation

R installation:
  1. Download from a nearby CRAN (Comprehensive R Archive Network) mirror http://cran.es.r-project.org/
  2. Select OS (Linux, Mac, Windows)
  3. Select base
  4. Download R-2.15.0 for Windows (2.15.0 is most recent as of May, 2012)

Rcmdr installation
See http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/installation-notes.html for OS-specific instructions.
  1. Start R
  2. From the prompt type: `> install.packages("Rcmdr", dependencies=TRUE)` #
  3. and choose repository/mirror,
  4. If prompted to install missing packages, accept.

12. Giving credit:

```
> citation()          # R
> citation("Rcmdr")   # Rcmdr
```

Acknowledgments:
  Andrew Gay, Ebling IT
  R: Ross Ihaka, Robert Gentleman, University of Auckland
  R Commander: John Fox