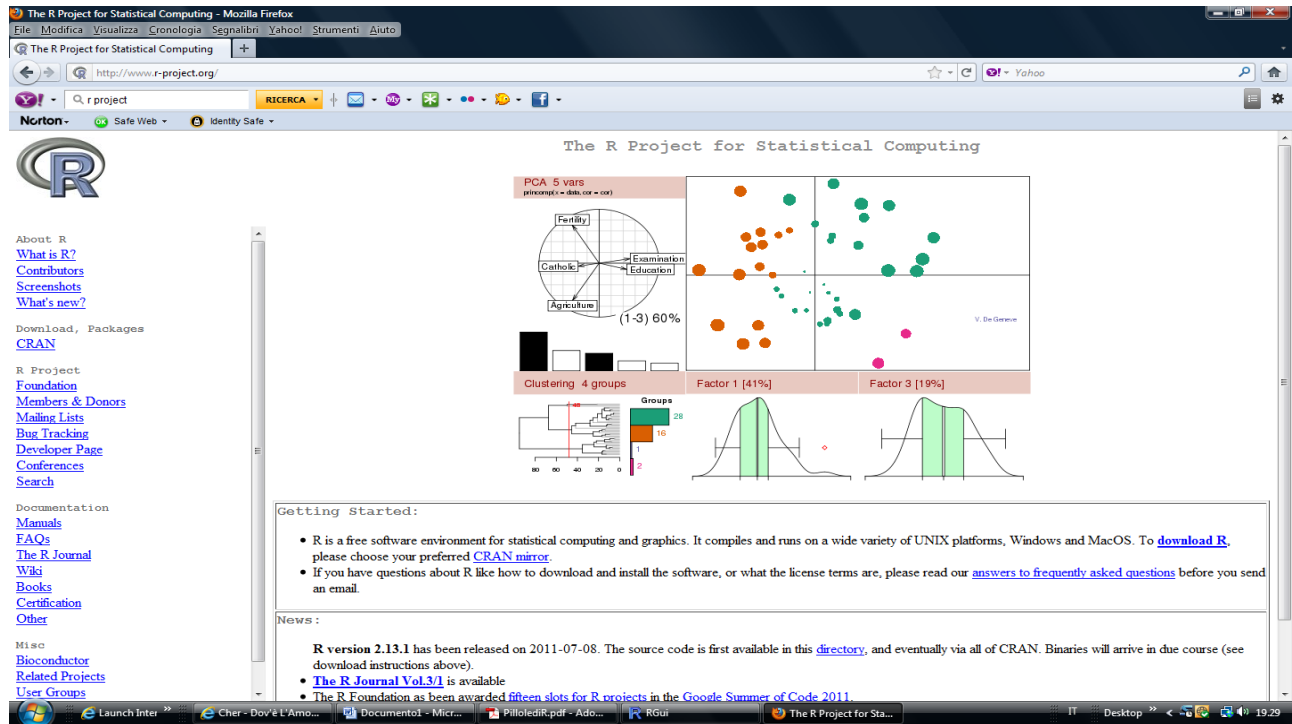


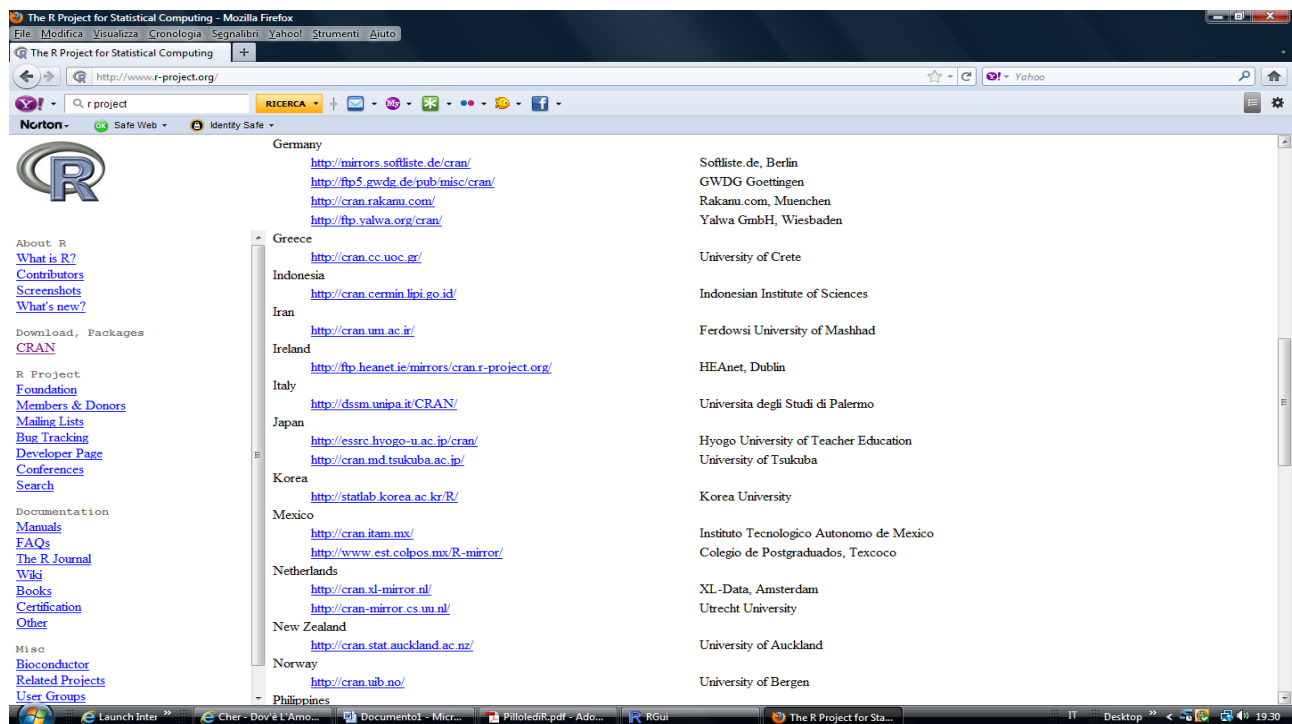
APPENDICE III. CONSIGLI PER L'USO DEL SOFTWARE R (a cura del Dott. Lorenzo Giolli)

AIII.1. Installazione del software R

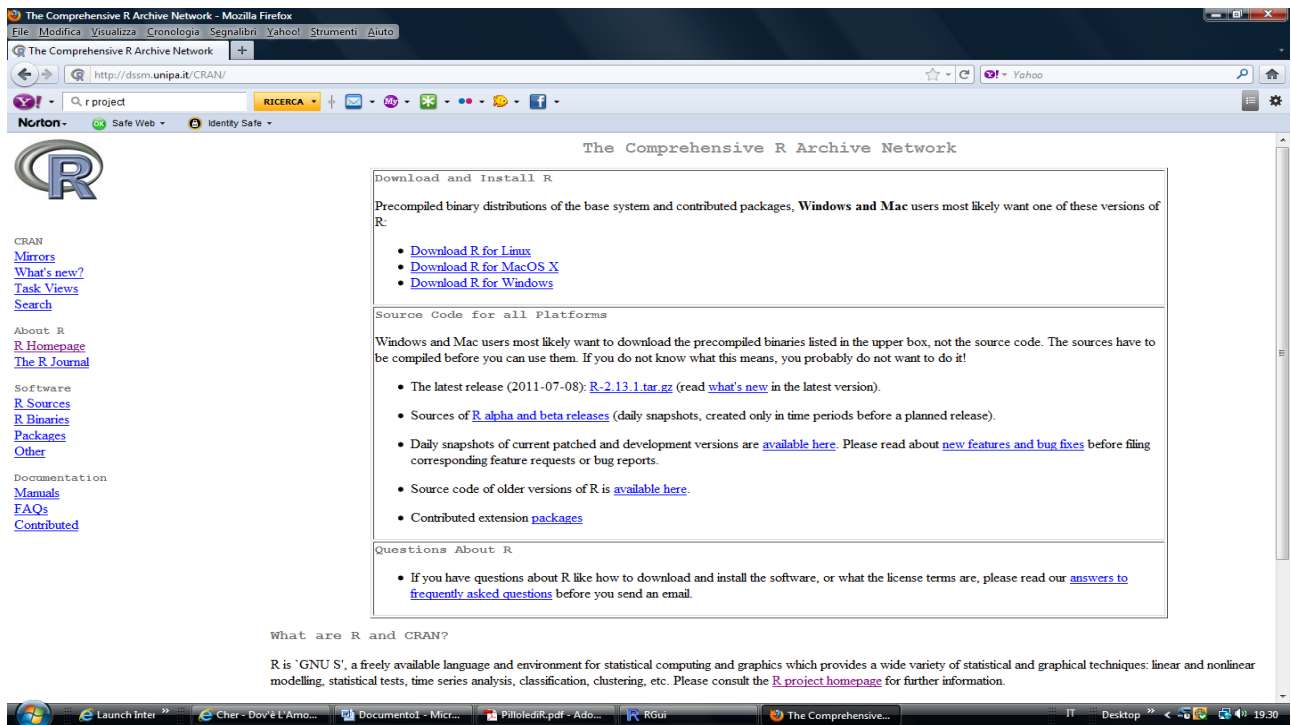
Aprire il browser e andare alla pagina www.r-project.org e cliccare CRAN (Download, Packages):



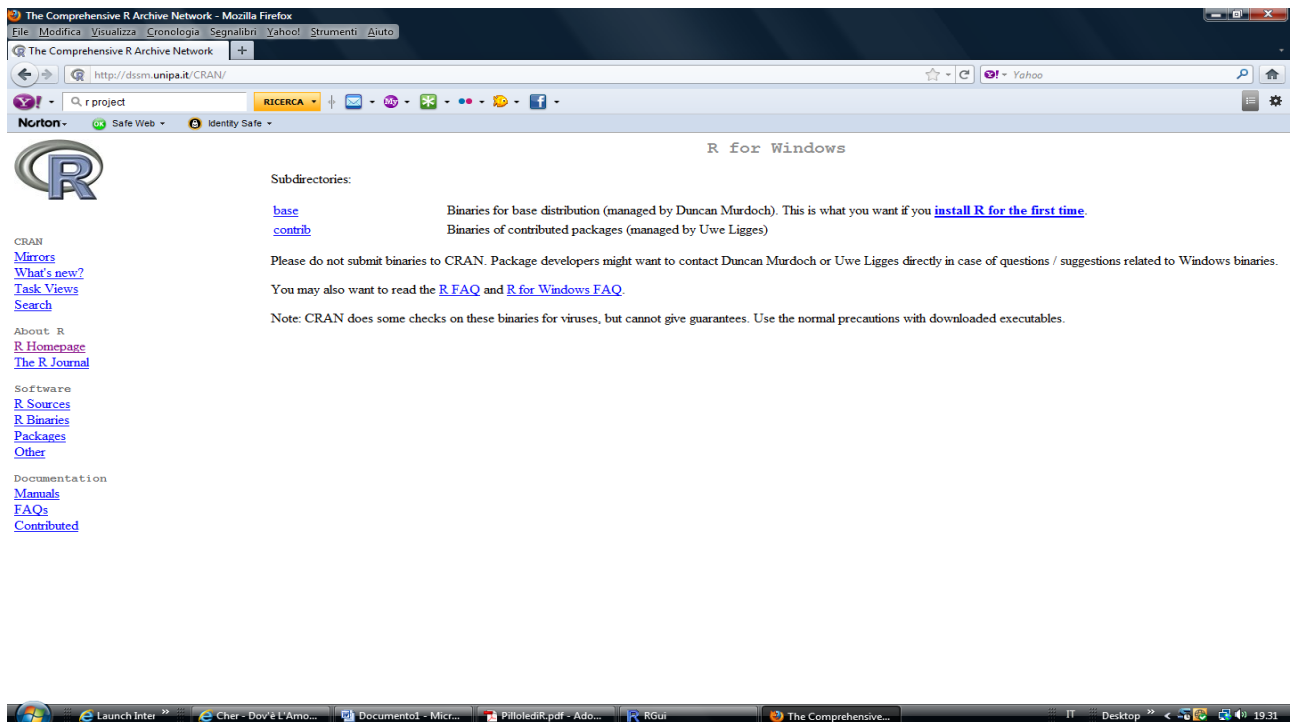
Cliccare <http://dssm.unipa.it/CRAN> (Italy, Università degli Studi di Palermo):



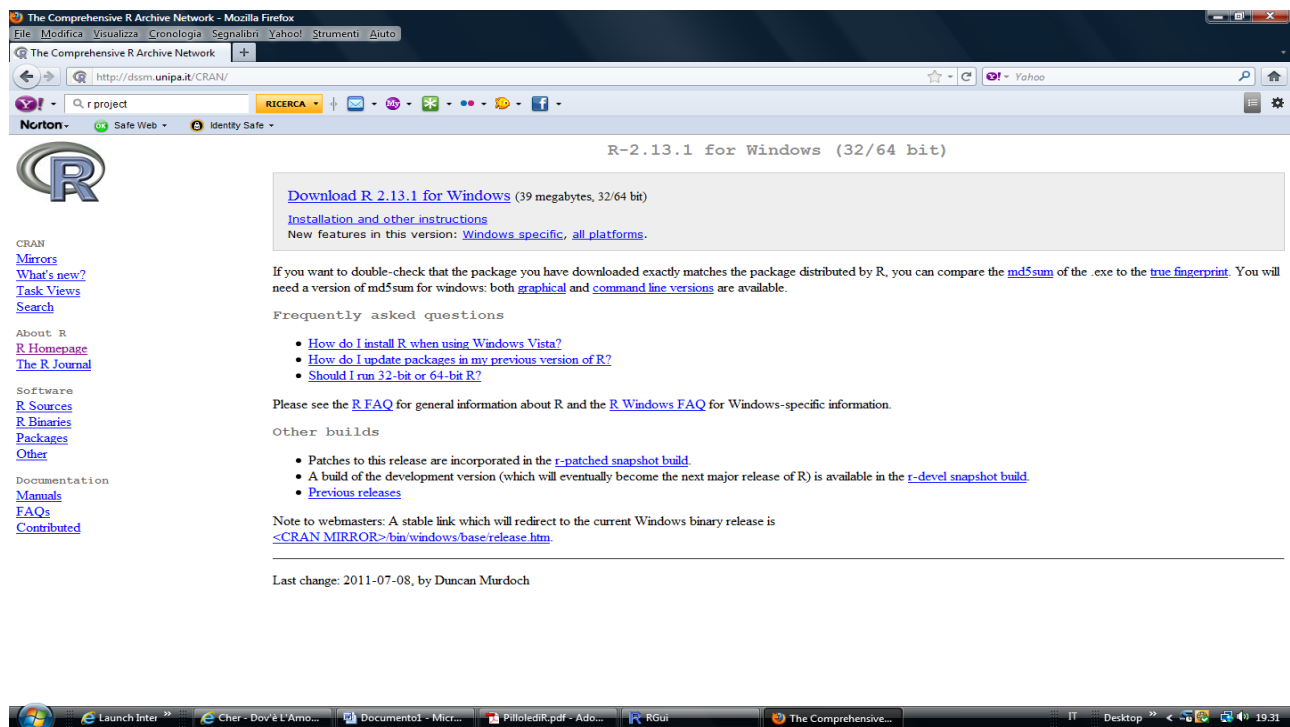
Selezionare [Download R for Windows](#) (Oppure Mac OS X o Linux se il sistema operativo del computer è diverso, in questo caso seguire le ulteriori istruzioni):



Cliccare [base](#) per una versione standard:



Selezionare Download R 2.13.1 for Windows (che in questo caso è l'ultima versione):

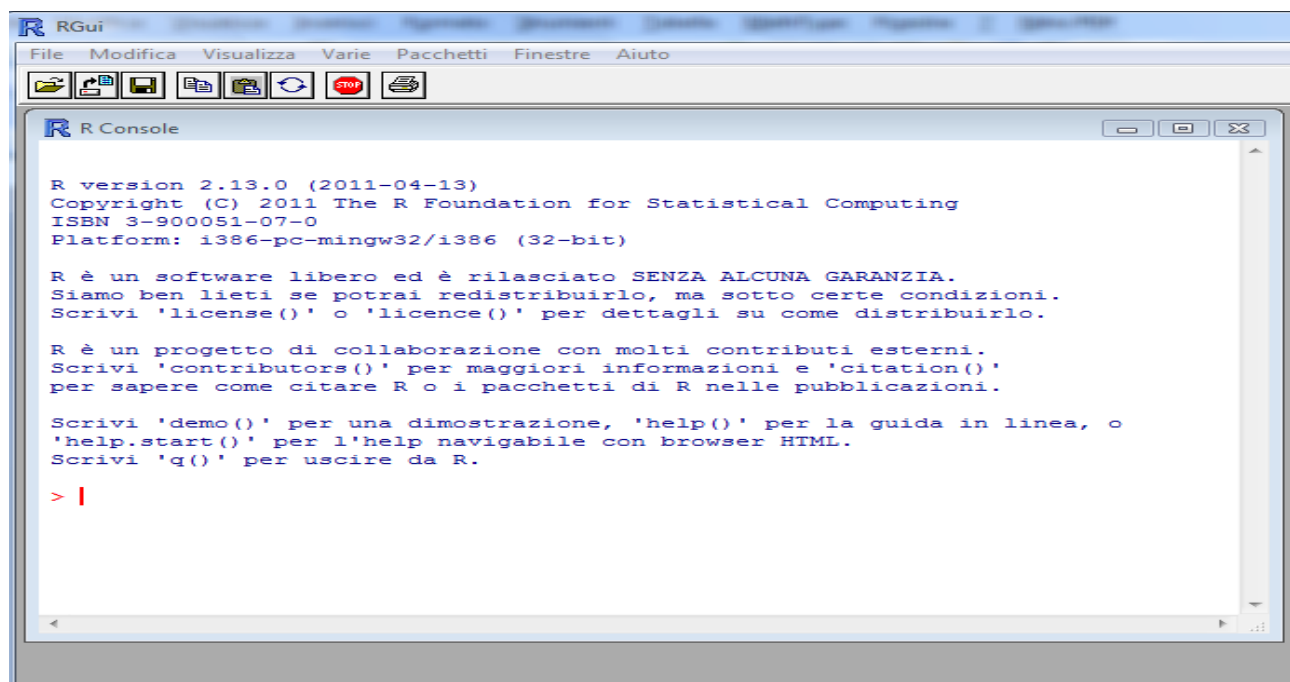


The screenshot shows a web browser window displaying the CRAN website. The main heading is "R-2.13.1 for Windows (32/64 bit)". Below this, there is a prominent link: "Download R 2.13.1 for Windows (39 megabytes, 32/64 bit)". Underneath, there are links for "Installation and other instructions" and "New features in this version: Windows specific, all platforms." A paragraph of text explains how to verify the download using md5sum. A section titled "Frequently asked questions" lists three items: "How do I install R when using Windows Vista?", "How do I update packages in my previous version of R?", and "Should I run 32-bit or 64-bit R?". There is also a link to the "R FAQ" for general information and a "Windows FAQ" for Windows-specific information. A section for "Other builds" lists "r-patched snapshot build" and "r-devel snapshot build". A note at the bottom states: "Note to webmasters: A stable link which will redirect to the current Windows binary release is <CRAN MIRROR>/bin/windows/base/release.htm". The last change is noted as "2011-07-08, by Duncan Murdoch".

Scaricare il file exe ed eseguirlo per installare, seguendo le istruzioni sul video, l'ultima versione di R (durante l'installazione si consiglia di creare una icona sul desktop e di selezionare i manuali in PDF).

AIII.2. Introduzione all'ambiente R

Dopo avere installato il software R cliccando sull'icona di R, opportunamente creata sul desktop durante l'installazione stessa, si entra nell'ambiente di lavoro R. La console in attesa di ricevere comandi si presenta nel seguente modo:



The screenshot shows the RGui application window. The title bar reads "RGui". The menu bar includes "File", "Modifica", "Visualizza", "Varie", "Pacchetti", "Finestre", and "Aiuto". Below the menu bar is a toolbar with icons for file operations and execution. The main area is the "R Console", which displays the following text:

```
R version 2.13.0 (2011-04-13)
Copyright (C) 2011 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: i386-pc-mingw32/i386 (32-bit)

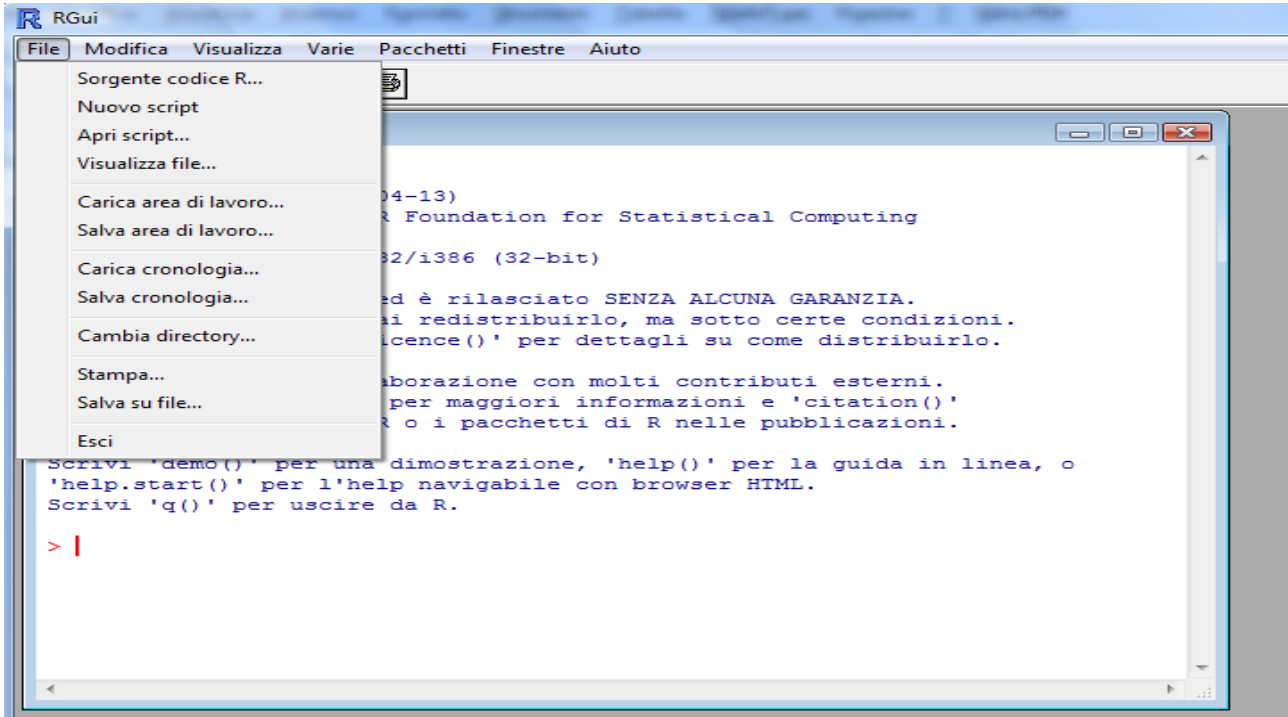
R è un software libero ed è rilasciato SENZA ALCUNA GARANZIA.
Siamo ben lieti se potrai redistribuirlo, ma sotto certe condizioni.
Scrivi 'license()' o 'licence()' per dettagli su come distribuirlo.

R è un progetto di collaborazione con molti contributi esterni.
Scrivi 'contributors()' per maggiori informazioni e 'citation()'
per sapere come citare R o i pacchetti di R nelle pubblicazioni.

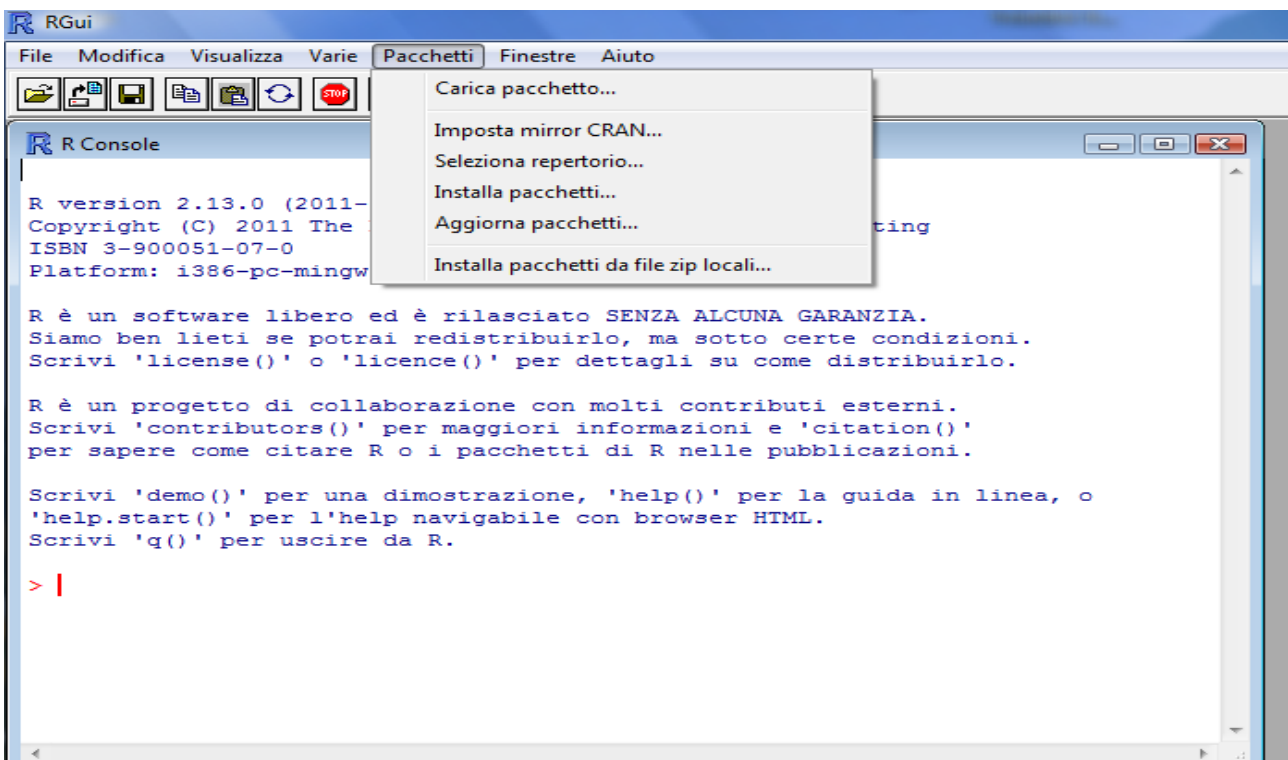
Scrivi 'demo()' per una dimostrazione, 'help()' per la guida in linea, o
'help.start()' per l'help navigabile con browser HTML.
Scrivi 'q()' per uscire da R.

> |
```

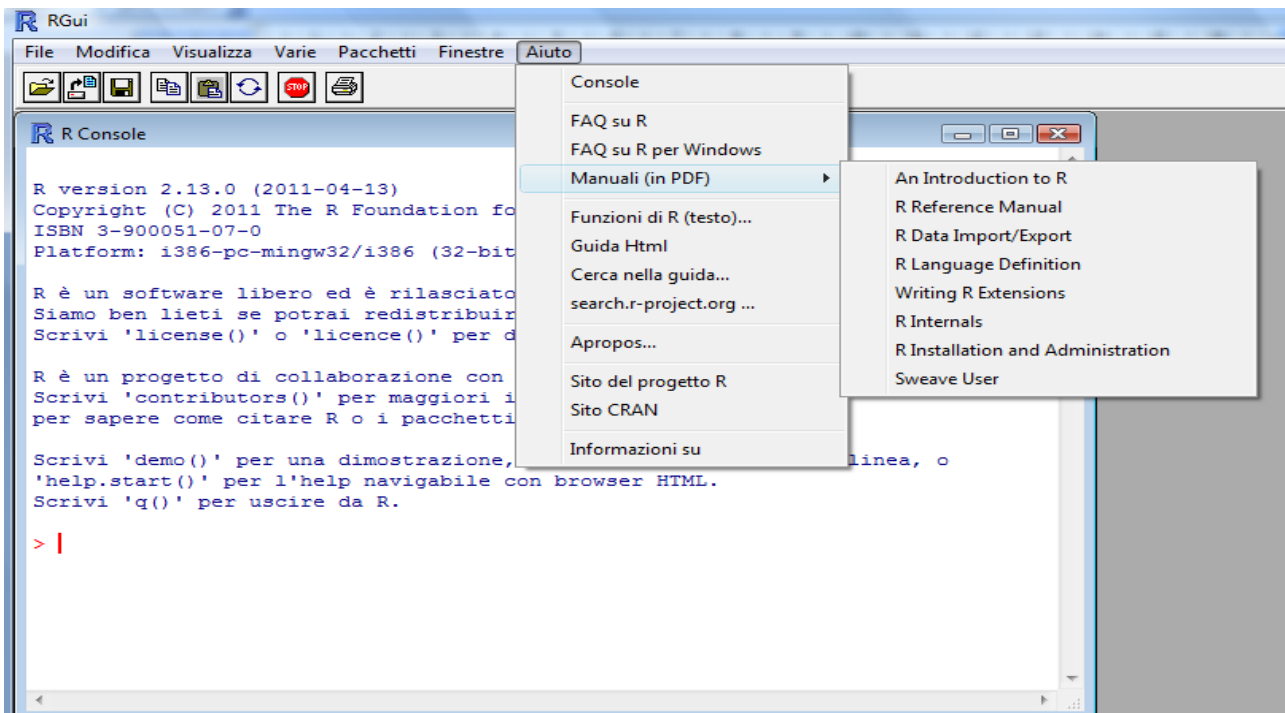
Si descrive di seguito la barra dei menu. Il menu **File** consente di salvare sia l'area di lavoro con la voce "Salva area di lavoro..." (Per poi caricarla in una sessione successiva) che la cronologia dei comandi con la voce "Salva cronologia...". Per caricare un file contenente comandi R si usa la voce "Apri script...", invece per creare un nuovo file contenente comandi R si usa la voce "Nuovo script..." (Altrimenti è possibile inserire direttamente i comandi tramite la console).



Il menu **Pacchetti** consente di installare pacchetti (chiamati anche librerie) sia direttamente da CRAN con la voce "Installa pacchetti..." che da file zip locali scaricati precedentemente da CRAN con la voce "Installa pacchetti da file zip locali..."



Il menu **Aiuto** contiene la voce "Manuali (in PDF)" che comprende una serie di manuali in PDF tra i quali si segnalano "An Introduction to R" (in inglese) dove in Appendice A è riportata una sessione di esempio che si invita ad eseguire.



AIII.3. Lettura di file ASCII e CSV

In questo paragrafo vengono descritti i comandi per la lettura di files di tipo ASCII (di solito indicati con estensione txt) e di tipo CSV (salvati in questo formato tramite il software Excel). Se il file che si vuole leggere si chiama prova.txt (contenente un vettore colonna di dati senza il nome della colonna nella prima riga) che si trova nella directory mydat del disco C: il comando da utilizzare sarà:

```
prova <- scan("c:/mydat/prova.txt")
```

Se il file prova.txt contiene due e più colonne separate da spazi vuoti con nome delle colonne nella prima riga il comando da utilizzare è il seguente:

```
prova <- read.table("c:/mydat/prova.txt", header=TRUE)
```

Se nella prima colonna ci sono i nomi di riga il comando è:

```
prova <- read.table("c:/mydat/prova.txt", header=TRUE, row.names=1)
```

Se il file prova.csv (salvato con una versione italiana di Excel, vale a dire come separatore decimale la virgola) contiene due e più colonne con nome delle colonne nella prima riga il comando da utilizzare è il seguente:

```
prova <- read.csv2("c:/mydat/prova.csv", header=TRUE)
```

Se nella prima colonna ci sono i nomi di riga il comando è:

```
prova <- read.csv2("c:/mydat/prova.csv", header=TRUE, row.names=1)
```

Se il file prova.csv è stato salvato con la versione inglese di Excel, vale a dire come separatore decimale il punto, il comando è:

```
prova <- read.csv("c:/mydat/prova.csv", header=TRUE)
```

Se nella prima colonna ci sono i nomi di riga il comando è:

```
prova <- read.csv("c:/mydat/prova.csv", header=TRUE, row.names=1)
```

AIII.4. I comandi attach() e detach()

Se si vogliono utilizzare singolarmente le colonne che compongono un data frame si deve svincolare le singole suddette colonne tramite il comando attach(). Il comando da usare per il data frame prova è il seguente:

```
attach(prova)
```

Il comando detach() è il comando inverso, vale a dire attacca nuovamente le variabili colonna dentro il data frame in modo che non è possibile effettuare operazioni sulle singole colonne, per il data frame prova il comando è:

```
detach(prova)
```

AIII.5. Utilizzo dei Data Sets di R

L'ambiente R carica in modo automatico un certo numero di pacchetti fra i quali il pacchetto datasets. Il suddetto pacchetto contiene molti data sets. Per avere una lista dei data sets disponibili si usa il comando:

```
library(help="datasets")
```

Per disporre dei data sets occorre richiamarli nell'ambiente R, ad esempio caricare il data set Orange e renderlo disponibile si usa il comando:

```
data(Orange)
```

AIII.6. Introduzione alla statistica con R

Nella Tabella AIII.6.1 sono riportate alcune funzioni statistiche elementari:

Tab. AIII.6.1. Funzioni statistiche elementari

mean(x)	Media del vettore x
median(x)	Mediana del vettore x
quantile(x,prob)	Quantile del vettore x
range(x)	Range del vettore x
var(x)	Varianza campionaria del vettore x
var(x,y)	Covarianza tra il vettore x e il vettore y
cor(x,y)	Correlazione tra il vettore x e il vettore y

Per ulteriori approfondimenti sulla statistica con R si rimanda alla bibliografia ed in particolare al "Formulario di Statistica con R" e al documento introduttivo "An Introduction to R". Nel "Laboratorio di

Simulazione”, contenuto in ogni capitolo, tranne il 12, del presente manuale ,vengono presentati esempi approfonditi per l'uso della statistica con R con allegati i comandi di R necessari per svolgere tali esempi.

Di seguito si vuole introdurre il lettore alla comprensione e utilizzo delle capacità grafiche di base di R. Il comando per costruire un diagramma di dispersione è:

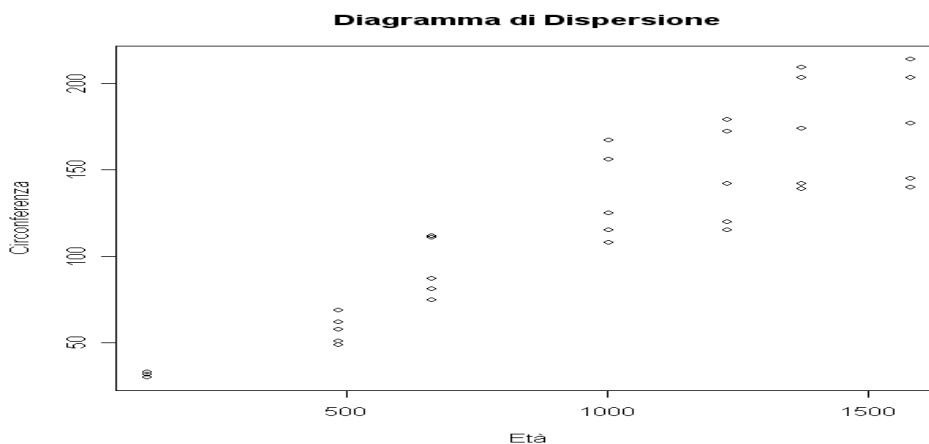
```
plot(x,y)
```

dove x e y sono vettori della stessa lunghezza. Il primo vettore viene rappresentato sull'asse delle ascisse mentre il secondo sull'asse delle ordinate. Ad esempio il data set Orange contiene le variabili età (age) e circonferenza (circumference) con i comandi:

```
data(Orange)
attach(Orange)
plot(age, circumference, main ='Diagramma di Dispersione', xlab='Età', ylab='Circonferenza')
```

si ottiene il seguente Grafico AIII.6.1.

Grafico AIII.6.1. Grafico a dispersione



L'istogramma e il box-plot, rappresentati nei Grafici AIII.6.2 e AIII.6.3, sono altrettanto importanti e si ottengono con i comandi hist() e boxplot() :

```
data(Orange)
attach(Orange)
hist(circumference, main ='Istogramma di Circonferenza', xlab='Circonferenza', ylab='Frequenza')
boxplot(circumference, main ='Box-Plot di Circonferenza')
```

Grafico AIII.6.2. Grafico a istogrammi

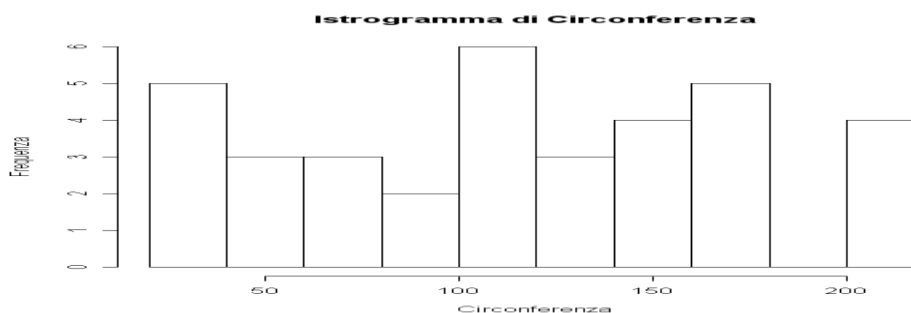
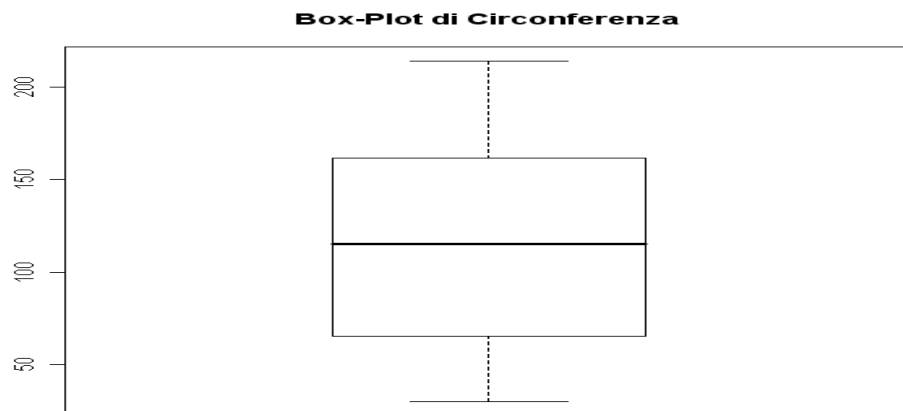


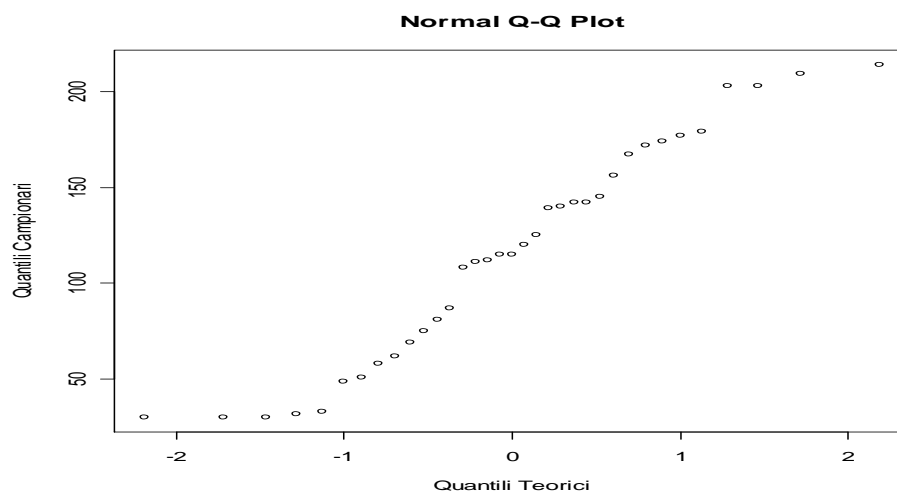
Grafico AIII.6.3. Box-plot



Il Grafico AIII.6.4 Normal Q-Q Plot è utile per verificare la normalità di una serie di dati e si ottiene con il comando `qqnorm()`:

```
data(Orange)
attach(Orange)
qqnorm(circumference, xlab='Quantili Teorici', ylab='Quantili Campionari')
```

Grafico AIII.6.4. Grafico Normal Q-Q Plot



Per informazioni aggiuntive su come creare grafici con R si rimanda al testo "Laboratorio di statistica con R" e ai manuali scaricabili dal sito www.r-project.org sotto la voce "Manuals" disponibili sia in Italiano che in Inglese.

AIII.7. Un esempio di sessione di lavoro

Qui di seguito sono riportati i comandi di una breve sessione di lavoro di R.

```
#
# CHIEDERE AIUTO
#
?hist
help("hist") # evidenzia la pagina di help sulla funzione hist
help.search("hist") # evidenzia tutti i comandi di hist
#
# LIBRERIE
#
library() #mostra le librerie dell'area di lavoro
library(MASS) # carica le libreria MASS
#
# OPERATORI MATEMATICI E LOGICI
#
7 + 3
3 + 4*2
sqrt(81)
log(10)
log10(10)
sum(2, 3, 7, 5, 8)
#
# MATRICI
#
x <- seq(18)
#
matrice_1 <- matrix(x, 3, 6,byrow=FALSE)
matrice_1
matrice_2 <- matrix(x, 3, 6,byrow=TRUE)
matrice_2
ncol(matrice_1)
nrow(matrice_1)
#
# DATAFRAME
#
tabella_1 <- matrix(rnorm(12),4, 3)
tabella_1
tabella_2 <- as.data.frame(tabella_1)
tabella_2
tabella_2[2,1]
row.names(tabella_2) <- c("A", "B", "C", "D")
names(tabella_2) <- c("X", "Y", "Z")
tabella_2
DATI <- data.frame(a = c(10, 20, 30, 40), sesso = c("M", "F", "F", "M"))
DATI
#
# FINE SESSIONE
```

ATTENZIONE!

Si fa presente al lettore che, dopo aver scaricato la versione di R, può facilmente gestire le sessioni di lavoro riferite ai casi riepilogativi e personalizzati riportati nel Laboratorio di Simulazione in tutti i Capitoli del Manuale, tranne il 12 e 24, utilizzandone i dati. Si ricorda, inoltre, che può disporre di una guida completa all'uso di R sul sito della Maggioli all'indirizzo:

www.approfondimenti.maggioli.it/statistica

**Bibliografia**

- **Formulario di Statistica con R** di F. Frascati.
- **Una guida all'utilizzo dell'ambiente statistico R** di Angelo M. Mineo
- **Pillole di R** di Laura Grassini
- **Alcune note di introduzione all'uso di R** di Lucia Buzzigoli
- **An introduction to R** di W.N. Venables, D.M. Smith e the R Development Core Team
- **Laboratorio di statistica con R** di Stefano Iacus e Guido Masarotto