

Introduzione al software R

Cristina Tortora¹

¹Università di Napoli
Federico II

cristina.tortora@unina.it

il software R

Si tratta di un software molto flessibile che permette di compiere praticamente qualsiasi tipo di analisi statistica sui dati.

- gratuito;
- open source;
- specializzato per uso statistico;
- può descrivere modelli anche molto complessi;
- adatto ad un uso interattivo;
- dispone di numerosi pacchetti aggiuntivi.



Come ottenere R

R è gratuito ed è possibile scaricarlo dal sito

<http://www.r-project.org/>

The R Project for Statistical Computing

PCA: 5 vari
 Clustering: 4 process
 Factor 1 (41%)
 Factor 3 (19%)

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 [answers to frequently asked questions](#) before you send an email.

News:

- R 2.11.0 pre-release version** will appear starting March 25. Final release is scheduled for April 22, 2010.



Come ottenere R

R è gratuito ed è possibile scaricarlo dal sito

<http://www.r-project.org/>

The screenshot shows the R Project website for CRAN mirrors. The page title is "CRAN Mirrors". Below the title, it states: "The Comprehensive R Archive Network is available at the following URLs, please choose a location close to you. Some statistics on the status of the mirrors can be found [here](#)." The page lists mirrors for various countries, including Argentina, Australia, Austria, Belgium, Brazil, Canada, Chile, and China. Each country entry includes a URL and the name of the hosting institution or organization.

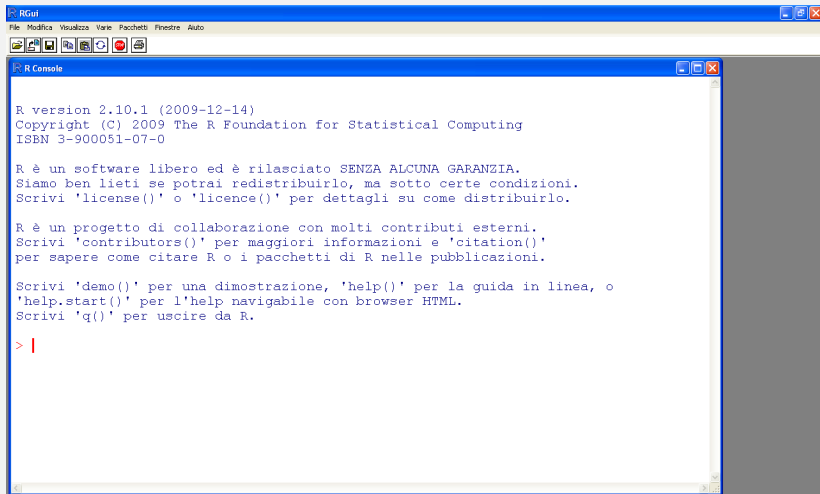
Country	URL	Host
Argentina	http://cran.patan.com.ar/	Patan.com.ar, Buenos Aires
Australia	http://cran.ms.unimelb.edu.au/	University of Melbourne
Austria	http://cran.at.r-project.org/	Wirtschaftsuniversität Wien
Belgium	http://www.froeststatistics.org/cran/	K.U.Leuven Association
Brazil	http://cran.br.r-project.org/	Universidade Federal do Paraná
	http://cran.fiocruz.br/	Oswaldo Cruz Foundation, Rio de Janeiro
	http://www.sps.fmvz.usp.br/CRAN/	University of Sao Paulo, Sao Paulo
	http://brieger.essli.usp.br/CRAN/	University of Sao Paulo, Piracicaba
Canada	http://cran.stat.sfu.ca/	Simon Fraser University, Burnaby
	http://mirror.in.dal.ca/cran/	Dalhousie University, Halifax
	http://probabilty.ca/cran/	University of Toronto
	http://cran.skakaforyou.com/	iWeb, Montreal
	http://cran.parentinginformed.com/	iWeb, Montreal
Chile	http://mirchlet.mat.puc.cl/	Pontificia Universidad Catolica de Chile, Santiago
China	http://ftp.ctex.org/mirrors/CRAN/	CTEX.ORG
	http://cran.csfb.cn/	Computer Network Information Center, CAS, Beijing

Su questo sito è possibile:

- scaricare R da diverse piattaforme;
- scaricare documentazione su R;
- scaricare *library* aggiuntive.



Primi passi su R



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

```
R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

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

> |
```



Primi passi su R

I comandi su R vanno inseriti dopo il prompt **>**

R è case sensitive Ciao \neq ciao

Per un aiuto è possibile usare:

- `help()`
- `help(comando)`
- `?comando`
- `example(comando)`



Primi passi su R

In R non è necessario inizializzare una variabile, è possibile assegnare direttamente un valore.

Per assegnare un valore ad una variabile è preferibile usare il comando `<-` al posto di `=`.

```
a<- 5   b<-'Statistica'  c<-TRUE
```

E' possibile vedere la classe di una variabile con il comando `class(nome variabile)`

Se voglio creare una variabile vuota o un valore mancante posso usare `NA` (not assigned) oppure `NaN` (not a number)



Operatori logici

TRUE FALSE NA

è possibile abbreviarli con T o F

operatori logici: `==`, `>`, `>=`, `<`, `<=`, `!=`.

Servono per verificare delle condizioni ad es. $(3*4) <= (2+7)$.

Per utilizzare più condizioni si possono usare i comandi `&` per indicare il verificarsi di entrambe le condizioni, `|` per una delle due condizioni.



Numeri, vettori e matrici

Sulle variabili numeriche è possibile effettuare le operazioni aritmetiche con i comandi `+` `-` `*` `/`

Per creare un vettore si può usare l'operatore `c` che permette di concatenare

```
a<-c(1,2,4)
```

In questo modo si otterrà un vettore colonna, per ottenere un vettore riga si usa il comando `t()` che traspone la matrice



Numeri, vettori e matrici

Per generare una sequenza di numeri è possibile usare i comandi:

```
x <- 1:5; seq(1,5);
```

è possibile specificare anche il passo ad es. `seq(8,2,-2)`

Per creare una matrice si usa il comando

```
matrix(valori, nrighe, ncolonne, byrow)
```

```
c<-matrix(c(1,2,3,4),2,2,1)
```



Esercizio 1

Generare

$$A = (5, 2, 6)$$

$$B = \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix}$$

$$C = \begin{pmatrix} 2 & 3 & 0 \\ 3 & 6 & 2 \\ 8 & 1 & 1 \\ 4 & 1 & 1 \end{pmatrix}$$



Operazioni sulle matrici

Si può accedere agli elementi di una matrice:

- `c[1,2]` accede ad un singolo elemento della matrice;
- `c[1,]` accede alla prima riga della matrice;
- `c[,2]` accede alla seconda colonna della matrice.

Operazioni sulle matrici

- `%*%` prodotto matriciale;
- `det(c)` determinante della matrice;
- `solve(c)` inversa della matrice;
- `diag(c)` ottengo la diagonale di `c`;
- `diag(n)` ottengo una matrice identità di dimensione `n`



Esercizio 2

Generare

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

Generare

$$B = \begin{pmatrix} 2 & 3 & 0 \\ 0 & 0 & 2 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

Calcolare $C = A * B$ sostituire la seconda colonna di C con i valori -2,-1,1 e calcolare il determinante di C.



Liste

E' possibile creare una variabile che contiene matrici di classi e lunghezza diverse con il comando `list(a,b,c)`

```
ad es.  N<-c('Chiara','Fabio','Cristina')
        Età<-c(20,31,15,12)
        l<-list(N,Età)
```

Si può accedere ad un elemento della lista digitando `l[[1]]`



Data Frame

Il comando più usato per raggruppare dati di tipo diverso è il data frame, i vettori possono essere di classi diverse ma devono avere la stessa lunghezza. `data.frame(a,b,c)`

```
ad es.  N<-c('Chiara','Fabio','Cristina')
        E<-c(20,31,15)
        df<-data.frame(Nome=N,Età=E)
```

- Per accedere ad un elemento del data frame si usa la sintassi utilizzata per le matrici;
- Per accedere ad un oggetto digitiamo ad es. `df$Nome`;
- Se digitiamo `attach(df)` non dobbiamo scrivere `df$` per accedere agli elementi del data frame, digitando `detach(df)` annulliamo il comando.



Installare e caricare un package

Sul sito del CRAN-R sono disponibili numerosi package aggiuntivi.

- pacchetti, installa pacchetti
- scegliere una piattaforma
- scegliere il package da installare

Per caricare un pacchetto si può usare il comando `library(nome package)`



Caricare un dataset

In molti package ci sono dei dataset,
per conoscere i dataset contenuti in un package si usa
`data(package="nomePackage")`

per caricare un dataset usiamo `data(nomedataset)`

Si può accedere ad una singola variabile del dataset digitando
`nomedataset$nomevariabile`

Il dataset è una matrice pertanto è possibile operare sul dataset
usando i comandi per le matrici.



Funzioni statistiche

Sulle variabili quantitative è possibile applicare le seguenti funzioni:

- `table(nomedataset[,2])` distribuzione di frequenza della seconda variabile
- `cut(nomedataset[,1], breaks=c(v1,v2,v3))` divide la prima variabile in intervalli
- `range(nomedataset[,3])` da il campo di variazione della terza variabile

Esercizio 3

Calcolare il dataset `iris`, calcolare la distribuzione di frequenza della variabile `Species`, dividere in 2 classi la variabile `Sepal.Length`, prima classe da 4.3 a 6, seconda classe da 6 a 7.9.



Funzioni statistiche

<code>rnorm, dnorm, pnorm, qnorm</code>	Normal distribution random sample, density, cdf and quantiles
<code>lm, glm, anova</code>	Model fitting
<code>loess, lowess</code>	Smooth curve fitting
<code>sample</code>	Resampling (bootstrap, permutation)
<code>.Random.seed</code>	Random number generation
<code>mean, median</code>	Location statistics
<code>var, cor, cov, mad, range</code>	Scale statistics
<code>svd, qr, chol, eigen</code>	Linear algebra



Grafici

In R esistono funzioni per produrre grafici

- `plot` scatter plot;
- `pie` grafico a torta;
- `hist` istogramma;
- `boxplot` calcola direttamente i quantili.

Esercizio 4

Fare un grafico a torta della variabile `Species`, un boxplot e un istogramma della variabile `Sepal.Width` ed uno scatter plot con ascissa la variabile `Sepal.Length` e ordinata `Sepal.Width`.

N.B. i grafici si sovrappongono, per aprire una nuova finestra per il grafico digitare `windows()`.



Come salvare

Per salvare il contenuto del workspace è necessario scegliere una cartella dove conservare i file nel menu *file, cambia directory*.

- `save.image()` salva tutto il contenuto del workspace;
- `save(x, file='prova.r')` salva la variabile `x` nel file `prova.r`;
- `save(list=ls(), file='prova.R')` salva tutte le variabili nel file `prova.r`;
- `load('prova.r')` carica il file salvato.

