

Inferenza Statistica

Breve introduzione ad R

Dipartimento di Scienze Economiche, Aziendali,
Matematiche e Statistiche “B. de Finetti”

a.a. 2021/2022

Che cos'è R ?

 è un programma per la grafica e l'analisi statistica. È gratuito e disponibile per tutti i principali sistemi operativi (Unix, Windows, Mac OS) alla pagina web

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

- R fornisce un insieme piuttosto ampio di funzioni statistiche per l'analisi dei dati e le elaborazioni grafiche;
- è basato sul linguaggio S, linguaggio e ambiente di sviluppo di alto livello per l'analisi dei dati e la grafica;
- dispone di applicazioni per la grafica estremamente potenti che possono essere facilmente adattate ai dati a disposizione;
- è un potente linguaggio ad oggetti che può essere facilmente esteso dall'utente (si rendono continuamente disponibili nuovi pacchetti, al momento sono 12751).

R Studio

R Studio è un ambiente di sviluppo integrato per R disponibile alla pagina <http://www.rstudio.com>. Si presenta suddiviso in

- finestra del codice: in quest'area è possibile aprire, creare e scrivere i vostri script;
- finestra della console: questa zona è la console R effettiva in cui vengono eseguiti i comandi e visualizzato l'output;
- finestra degli oggetti: in quest'area è possibile trovare una lista di tutti gli oggetti creati nello spazio di lavoro in cui si sta lavorando;
- la finestra dei pacchetti-dei grafici-dell'aiuto in linea: in quest'area è possibile caricare facilmente i pacchetti, aprire file di aiuto R e visualizzare i grafici.

R studio

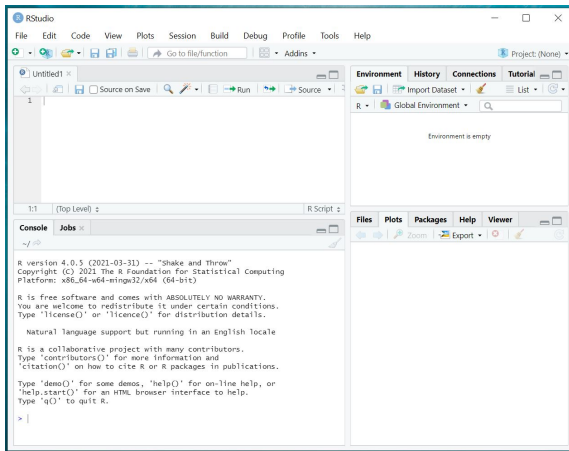


Figure: L'interfaccia grafica di R Studio.

Primo avvio di R

Avviando una sessione di R appare una “console” con un cursore, rappresentato dal simbolo ‘>’, dove digitare i comandi e le istruzioni, che vengono eseguite dando l’invio.

Possiamo memorizzare il codice creato in un file di estensione .R (*script*), per poterle riutilizzare in seguito.

Una volta installato, R contiene un set standard di **pacchetti** (librerie di funzioni), che comprendono funzioni R, dati, e codice compilato in un formato ben definito. Esistono molti altri pacchetti per analisi più specifiche disponibili per il download e l’installazione.

Oggetti in R

R è orientato agli **oggetti**, nel senso che possiamo creare entità denominate oggetti, a cui attribuiamo un nome. Per costruire un oggetto viene utilizzato l'operatore di assegnazione `<-`. Gli oggetti possono essere

scalari	contengono un singolo valore
vettori	insieme lineare di elementi omogenei per tipologia, tutti numeri, tutte stringhe, ecc.
fattori	vettore utilizzato per classificare o suddividere in "livelli" gli elementi di un altro vettore
array e matrici	sono vettori multidimensionali, le matrici hanno due dimensioni: righe e colonne
liste	insieme di elementi disomogenei tra loro
data frame	sono matrici bidimensionali dove ogni colonna può avere un tipo di dato diverso dalle altre

Semplici regole

È importante ricordare che:

- R è “case sensitive” cioè considera come caratteri differenti le lettere maiuscole e minuscole
- tutte le lettere ed i numeri possono essere utilizzati per dare nomi a variabili, funzioni e procedure (esclusi gli spazi e tutti i segni di punteggiatura ad eccezione del punto)
- testi di commento possono essere aggiunti dopo il simbolo cancelletto ('#')
- se si va a capo prima della conclusione di un comando R mostrerà all'inizio della riga il simbolo più ('+');
- è possibile richiamare i comandi impartiti in precedenza utilizzando la freccia su' della tastiera.

Spazio di lavoro

Dopo aver avviato R è necessario impostare la *directory*, ovvero la cartella di lavoro, che contiene i dati che si intende utilizzare

- `setwd()` per scegliere la directory di lavoro (oppure si può utilizzare la finestra di dialogo **Change directory** dal menù **File**)
- `getwd()` restituisce l'indirizzo relativo alla *directory* di lavoro corrente
- `ls()` per controllare gli oggetti presenti nella *directory* di lavoro corrente
- R salva gli oggetti creati (*workspace*) in un file con estensione `.RData`, che può essere ripristinato in una sessione successiva (**Load Workspace**)
- `q()` consente di chiudere R .

Uso dell'Help

R dispone di un sistema di *help* interattivo versatile ed articolato, attraverso il quale si possono consultare manuali, ricercare e/o identificare librerie, funzioni ed esempi:

- `help.start()` attiva il browser di sistema utilizzato per navigare nella documentazione disponibile
- `help("nomeComando")` (o `?nomeComando`) digitato dal prompt fornisce informazioni dettagliate su un comando specifico: gli argomenti richiesti dalla funzione, il loro nome, il loro ordine, quelli necessari e quelli opzionali, ovvero quelli per cui un valore di default è già impostato.

Operatori e alcune semplici funzioni

- operatori aritmetici: $+$, $-$, $*$, $/$, $^$
- operatori logici: $\&$ (AND), $|$ (OR), $!$ (NOT)
- operatori di confronto: $<$ (minore), $>$ (maggiore), $<=$ (minore o uguale), $>=$ (maggiore o uguale), $==$ (uguale), $!=$ (diverso)
- La valutazione di espressioni che coinvolgono operatori di confronto e operatori logici fornisce come risultato TRUE o FALSE
- funzioni matematiche: `sqrt()`, `abs()`, `exp()`, `log()`, `log2()`, `sin()`, `cos()`, `sum()`, `max()`, `min()`, **etc.**

Operatori e alcune semplici funzioni

Esempio: Vettori

```
> v<-c(100,18,41,21,53)
> v
[1] 100 18 41 21 53
> v[2]
[1] 18
> v[c(1:3)]
[1] 100 18 41
> v>50
[1] TRUE FALSE FALSE FALSE TRUE
> v[v>20 & v<80]
[1] 41 21 53
> v*3
[1] 300 54 123 63 159
> length(v)
[1] 5
```

Operatori e alcune semplici funzioni

Esempio

```
> v<-c(100,18,41,21,53)
```

```
> sum(v)
```

```
[1] 233
```

```
> cumsum(v)
```

```
[1] 100 118 159 180 233
```

```
> mean(v)
```

```
[1] 46.6
```

```
> sortvett<-sort(v)
```

```
> sortvett
```

```
[1] 18 21 41 53 100
```

```
> median(v)
```

```
[1] 41
```

I pacchetti

La directory in cui sono presenti i pacchetti è detta libreria. R fornisce una cospicua varietà di procedure e funzioni per analisi statistiche anche molto sofisticate raccolte in **pacchetti aggiuntivi**, che possono essere installati e caricati facilmente dalla console

```
> install.packages("ggplot2")  
> library(ggplot2)
```

o considerando le opzioni del menù **Packages**.

Il comando `??ggplot2` permetterà di accedere all'*Help* e alla documentazione relativa alle funzioni presenti nel pacchetto.

Definizione di una nuova funzione

In R è possibile definire nuove funzioni. La definizione di una funzione deve rispettare la seguente sintassi

```
nomefunzione <- function(arg1,arg2,...){espressione}
```

- `espressione` è un'espressione, o un insieme di espressioni e `arg1, arg2, ...` sono gli input passati alla funzione.
- Il comando che attiva la funzione è

```
> nomefunzione(espressione1, espressione2, ...)
```
- Il risultato coincide con il valore determinato dall'ultima espressione del corpo della funzione o con quella individuata dal comando `return()`.
- Le funzioni in R possono essere anche ricorsive!

Condizione *if*, cicli *for* e *while*

Costrutti condizionali

```
if (condizione) {istruzione}
```

```
if (condizione) {istruzione1}  
  else {istruzione2}
```

Costrutti iterativi

```
for (nome.var in sequenza)  
  {corpo del ciclo}
```

```
while (condizione)  
  {corpo del ciclo}
```

Laboratori del corso

1. Distribuzioni di probabilità discrete e continue
2. Modello Gaussiano bivariato, somme di variabili aleatorie, applicazioni del Teorema limite centrale
3. Distribuzioni campionarie e introduzione ai metodi Monte Carlo
4. Stimatori e loro proprietà, funzione di verosimiglianza
5. Intervalli di confidenza e verifica di ipotesi