

R per l'analisi statistica multivariata

Statistica descrittiva con R

Tommaso Rigon

Università Milano-Bicocca

- Un dataframe è essenzialmente una matrice dei dati. Ciascuna riga rappresenta un'**unità statistica**, ciascuna colonna rappresenta una **variabile**.
- Un dataframe assomiglia ad una matrice (`matrix`), ma può contenere anche valori non numerici.
- Come vedremo nelle unità successive, il dataframe viene generalmente caricato all'interno di **R** utilizzando ad esempio la funzione `read.table`.
- Per cominciare, consideriamo un dataframe già presente in **R**, a scopo illustrativo.

```
# Dataset di esempio già presente in R  
data(iris) # Caricamento in memoria dell'oggetto iris
```

- Il dataset `iris` contiene 150 osservazioni e 5 variabili, come si può vedere dal comando `dim(iris)`.
- Conviene quindi considerare solamente un piccolo numero di osservazioni, ad esempio le prime 5.

```
iris[1:5, ] # Estrazione delle prime 5 osservazioni
```

##	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
## 1	5.1	3.5	1.4	0.2	setosa
## 2	4.9	3.0	1.4	0.2	setosa
## 3	4.7	3.2	1.3	0.2	setosa
## 4	4.6	3.1	1.5	0.2	setosa
## 5	5.0	3.6	1.4	0.2	setosa