

Università Ca' Foscari di Venezia

lrusnac@dsi.unive.it - Leonid Rusnac

14 ottobre 2014

ATTENZIONE: Si tratta di una mia soluzione ad un problema proposto nel corso di Probabilità e Statistica quindi soggetti ad errori e orrori, non consideratelo come la soluzione esatta in quanto si può arrivare in vari modi alla stessa soluzione e potrebbe non essere lo stesso metodo della professoressa. Usatelo a vostro rischio e pericolo.

Consegna: L'insieme di dati iris di R riguarda la misura di lunghezza e spessore dei petali e dei sepali di 3 specie di iris.

1. Costruire una tabella delle frequenze relative per Species.
2. Si può dire che la distribuzione di Species rappresenta un caso di massima mutabilità?
3. Rappresentare graficamente la variabile Species.
4. Rappresentare la variabile Sepal.Length mediante boxplot appaiati rispetto alle modalità di Species.
5. Quale specie di iris mostra una maggiore lunghezza dei sepali?

Soluzione: Innanzitutto è bene capire la struttura del dataset iris:

```
> head(iris)

  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
6          5.4         3.9          1.7          0.4  setosa
```

Possiamo vedere 4 variabili quantitative e una qualitativa. Guardiamo anche la numerosità campionaria:

```
> dim(iris)

[1] 150   5

> #salviamo la numerosita' in una variabile:
> n <- dim(iris)[1]
> n

[1] 150
```

per il primo punto possiamo creare prima la tabella delle frequenza assolute per poi dividerlo per la dimensione campionaria:

```
> #con l'operatore $ possiamo scegliere una variabile, potremmo usare
> #anche iris[,5]
> freq.ass <- table(iris$Species)
> freq.ass
```

```

      setosa versicolor virginica
      50         50         50

> #ora dividiamo per n
> freq.rel <- freq.ass/n
> freq.rel

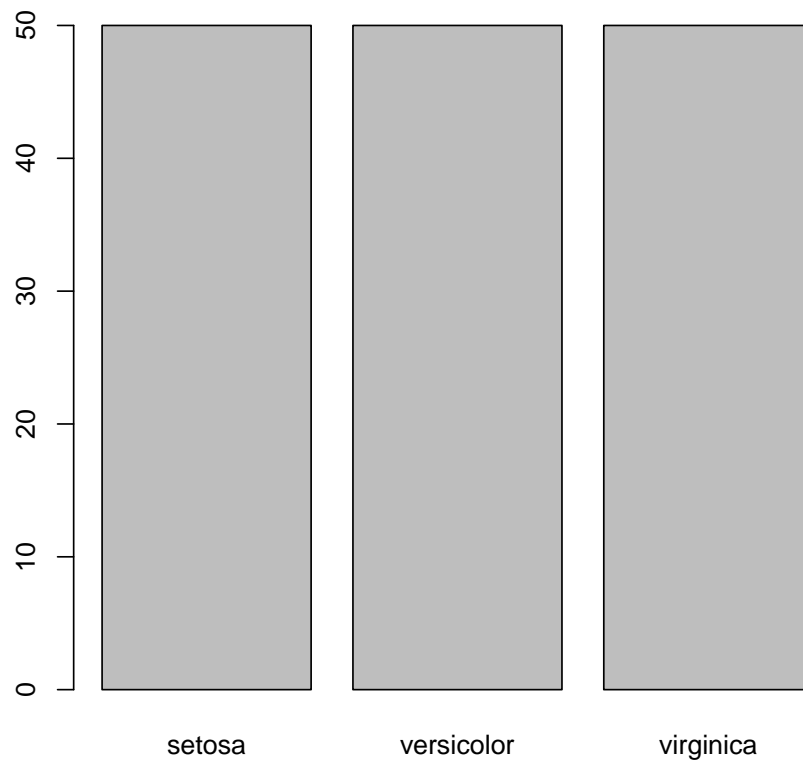
      setosa versicolor virginica
0.3333333  0.3333333  0.3333333

```

come si vede bene abbiamo lo stesso numero di esemplari di tutte le specie perciò possiamo dire che si tratta di un caso di massima mutabilità.

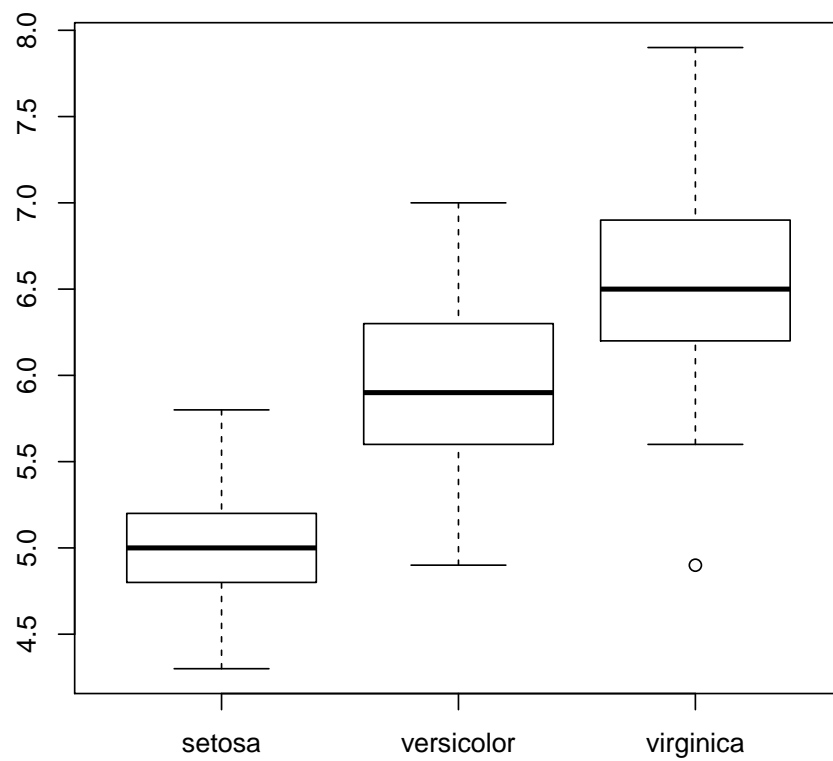
per rappresentare graficamente usiamo il comando plot:

```
> plot(iris$Species)
```



boxplot

```
> boxplot(iris$Sepal.Length ~ iris$Species)
```



dal boxplot si vede che virginica ha una maggiore lunghezza dei sepali