

Boxplot

o Grafici a scatola e baffi

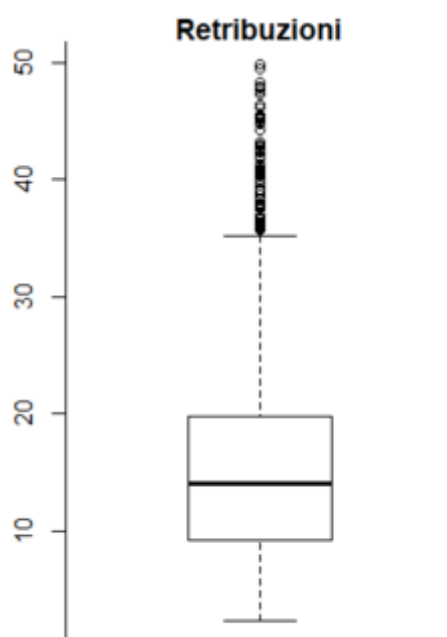
In una distribuzione normale, media e mediana coincidono, e i quantili sono simmetrici rispetto al valore centrale. È dunque pratica comune studiare la forma di una distribuzione con riferimento a tali misure.

Un tipo di rappresentazione grafica adatta a tale scopo è il diagramma “a scatole e baffi”, o **boxplot**, o **box and whiskers plot**.

Boxplot con una variabile

```
# dati
data(SLID, package = "carData")

par(bty = "l") # box del grafico "a elle"
boxplot(SLID$wages, main="Retribuzioni")
```



Il grafico rappresenta la variabile mediante una “scatola” delimitata in alto e in basso dal primo e dal terzo quartile, e divisa da una linea che corrisponde alla mediana. I “baffi” — i due segmenti che escono dalla “scatola” — indicano il valore minimo e il valore massimo, oltre i quali sono rappresentati i punti degli **outliers**, casi con valori estremi.

**Se i due baffi hanno la stessa lunghezza, e la scatola è divisa a metà dalla riga, la distribuzione è simmetrica*.*

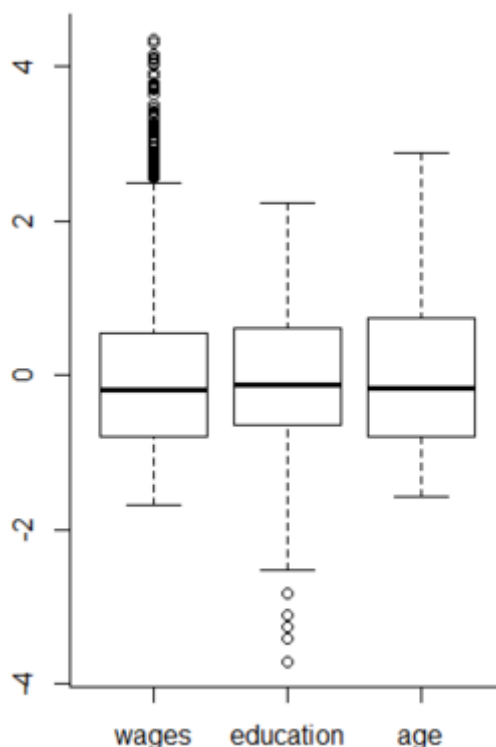
Nella figura sopra, le retribuzioni si concentrano sui valori bassi della scala: il baffo inferiore è più corto, e la linea della mediana, all'interno della “scatola”, è leggermente spostata verso il

basso.

Confronto fra più variabili

Possiamo confrontare più variabili con unità di misura diverse standardizzandole, ovvero trasformandole in valori z, con la funzione ``scale()``. Da riga di comando, scriveremo:

```
par(bty = "l")
boxplot(scale(SLID[1:3]))
```

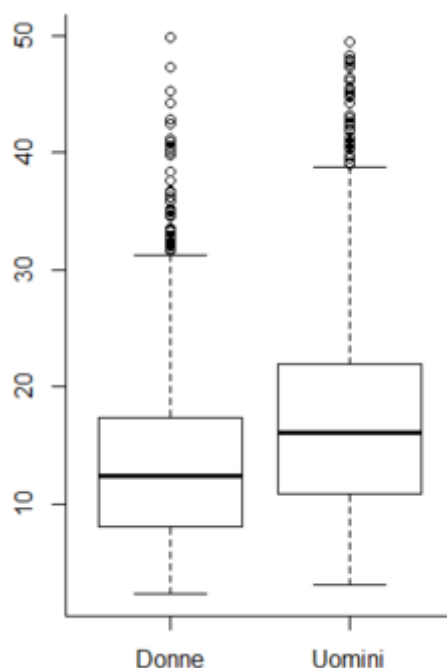


La forma delle variabili e il loro maggiore o minore discostamento dalla distribuzione normale è qui immediatamente confrontabile.

Boxplot bivariato

Per confrontare la forma delle distribuzioni delle variabili numeriche fra donne e uomini, si deve indicare una **formula** per rappresentare la relazione fra le due: ``wages ~ sex``, ovvero $\$wages = f(\$sex)$.

```
par(bty = "l")
boxplot(wages ~ sex, data=SLID,
        names = c("Donne", "Uomini"))
```



La distribuzione delle retribuzioni orarie delle donne è interamente spostata sui valori bassi.

Altro

Vedi anche:

- [Gestione dei colori](#)
- [Titoli](#)
- [Etichette e annotazioni](#)
- [Parametri grafici](#)

Script di esempio

E' possibile scaricare ed eseguire lo script dell'esempio:

[es_boxplot.R](#)

```
# dati
data(SLID, package = "carData")

# bordo del grafico
par(bty = "l")

# boxplot semplice
boxplot(SLID$wages, main="Retribuzioni")

# variabili standardizzate
boxplot(scale(SLID[1:3]))
```

```
# biviato  
boxplot(wages ~ sex, data=SLID,  
        names = c("Donne", "Uomini"))
```

[Grafici](#), [Analisi monovariata](#), [Analisi bivariata](#)

From:

<https://www.agnesevardanega.eu/wiki/> - **Ricerca Sociale con R**

Permanent link:

<https://www.agnesevardanega.eu/wiki/r/grafici/boxplot?rev=1539152522>

Last update: **10/10/2018 06:22**

