

Standardizzazione delle variabili

La standardizzazione delle variabili consente il confronto fra dati che appartengono a variabili o a campioni diversi.

La formula della standardizzazione è: $Z_x = \frac{x_i - X}{\sigma_x}$ dove:

- X = media aritmetica ([vedi](#));
- σ_x = deviazione standard, o scarto quadratico medio, o scarto tipo.

In R, è possibile standardizzare una o più variabili di un dataset con la formula:

```
scale(x, center = TRUE, scale = TRUE)
```

Ad esempio, il comando

```
scale(ChickWeight$weight)
```

restituisce il seguente output (una [matrice](#)):

```
      [,1]
[1,] -1.123063717
[2,] -0.996431497
...
[577,]  2.000531035
[578,]  2.000531035
attr(,"scaled:center")
[1] 121.8183
attr(,"scaled:scale")
[1] 71.07196
```

Gli ultimi due valori (attributi), `center` e `scale` sono, rispettivamente, la media aritmetica della variabile originaria, e la sua deviazione standard.

La funzione può essere quindi utilizzata per trasformare le variabili in base a valori diversi.

È anche possibile trasformare un dataset, purché composto interamente da vettori numerici:

```
Zcars <- scale(cars)
```

Se un dataset è composto da vettori e fattori, si potranno aggiungere le variabili standardizzate al dataset originario. Es.:

```
# faccio una copia del dataset originario (per sicurezza)
mydata<-ChickWeight
```

```
# aggiungo la variabile standardizzata  
mydata$z_weight <- scale(mydata$weight)
```

Script di esempio

Es Standardizz.R

```
# UNA VARIABILE  
scale(ChickWeight$weight)  
  
# UN DATASET  
Zcars <- scale(cars)  
  
# UNA VARIABILE IN UN DATASET  
# faccio una copia del dataset originario (per sicurezza)  
mydata<-ChickWeight  
  
# aggiungo la variabile standardizzata  
mydata$z_weight <- scale(mydata$weight)
```

Gestione dei dati

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

Permanent link:
https://www.agnesevardanega.eu/wiki/r/gestione_dei_dati/standardizzazione?rev=1632642375

Last update: 26/09/2021 07:46

