# Dispense sull'analisi delle componenti principali

Giorgio Marrubini e Camillo Melzi

# Indice

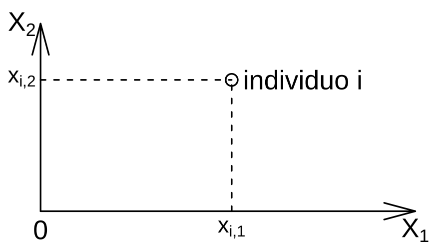
# Capitolo 1

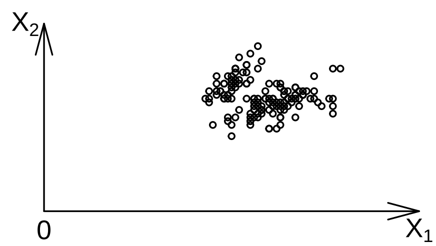
# Dati multidimensionali

## 1.1 Rappresentazione matriciale e geometrica

Tabella 1.1: Rappresentazione matriciale

| Indiv. | $X_1$ | $X_2$ | ... | $X_p$ |
|---|---|---|---|---|
| 1 | $x_{11}$ | $x_{12}$ | ... | $x_{1p}$ |
| 2 | $x_{21}$ | $x_{22}$ | ... | $x_{2p}$ |
| ... | | | | |
| m | $x_{m1}$ | $x_{m2}$ | ... | $x_{mp}$ |



## 1.2 Trasformazione delle variabili: centratura e standardizzazione

Indichiamo con $\bar{x_1}, \dots, \bar{x_p}$ le medie delle variabili $X_1, \dots, X_p$, cioè le $p$ medie delle $p$ colonne della Tabella 1.1, e con $\sigma_1^2, \dots, \sigma_p^2$ le rispettive varianze.
Il vettore $\bar{x} = (\bar{x_1}, \dots, \bar{x_p})$ viene chiamato **baricentro**.

**Centratura**: semplice traslazione del baricentro nell'origine

$$x_{ij}^{'} = x_{ij} - \bar{x_j} \tag{1.1}$$

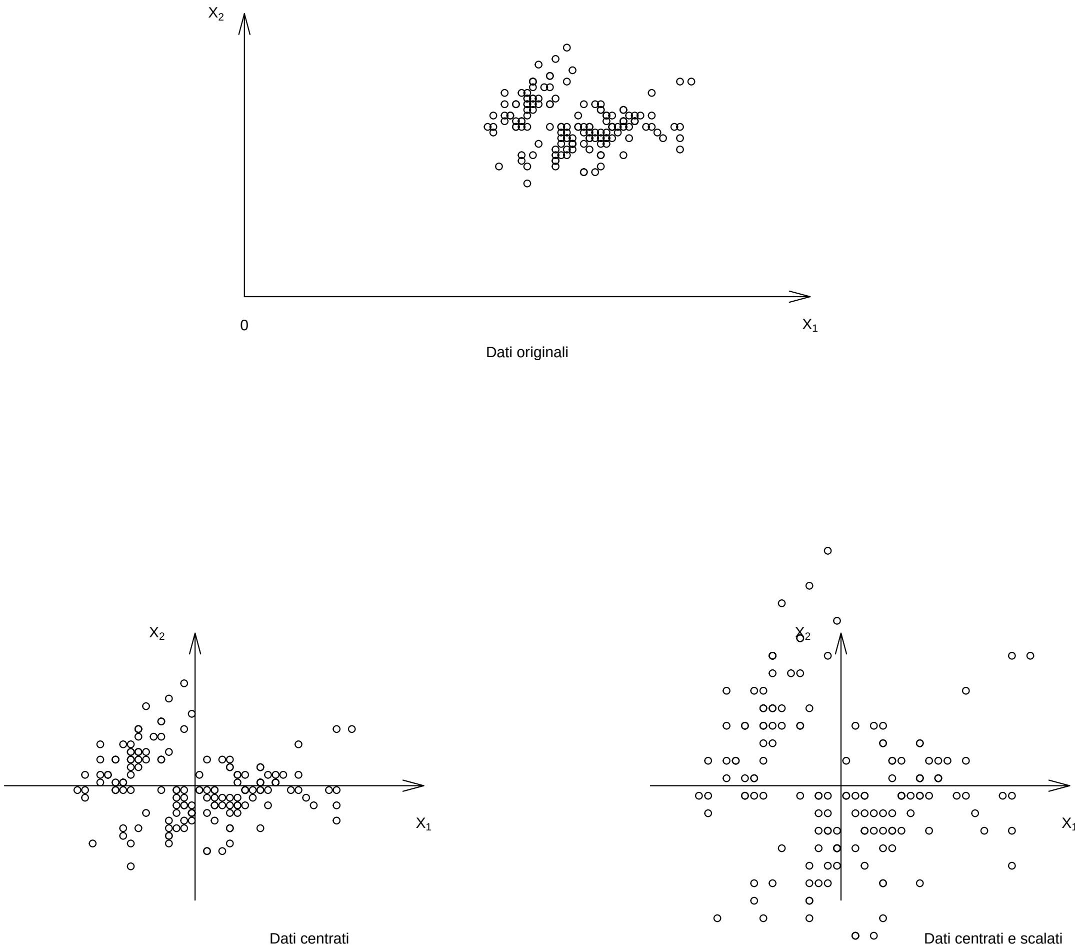
- non perdo informazione sulla distanza tra i punti (la geometria della nuvola di punti rimane invariata)
- perdo solo informazione sul baricentro
- semplifica formule e conti (da ora in poi useremo sempre dati centrati)

**Standardizzazione**: questa trasformazione porta ogni variabile ad avere varianza 1 (in generale questa trasformazione viene fatta insieme alla centratura)

$$x_{ij}^{'} = \frac{x_{ij} - \bar{x_j}}{\sigma_j} \tag{1.2}$$

- questa trasformazione rende le variabili degli scalari (numeri puri)
- questa trasformazione è necessaria quando si vogliono confrontare variabili con differenti unità di misura (le variabili devono essere omogenee per essere confrontabili)
- tutte le variabili hanno lo stesso "peso"
- cambia la distanza (la geometria) tra i punti. E' una dilatazione o contrazione.

Si veda la seguente figura per una rappresentazione grafica di dati centarti e scalati per una matrice di dati di 2 variabili


X2
0
X1
Dati originali
X2
X1
Dati centrati
X2
X1
Dati centrati e scalati