Introduzione ai metodi statistici per le applicazioni industriali parte 2

Antonio Panico

Department of Engineering for Industrial Systems and Technologies
University of Parma

18 giugno 2025

Media, Mediana e Moda: attenzione alla differenza!

UNIVERSITÀ DI PARMA

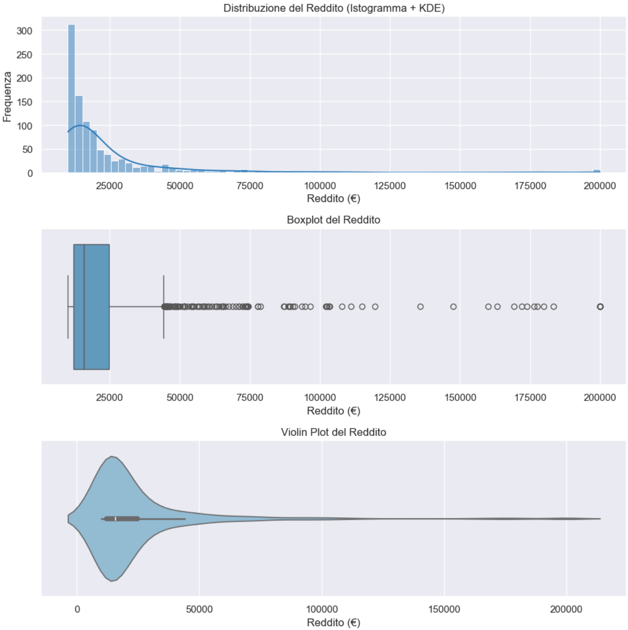
Cosa si intende per "media"?

Riassumere dei dati con un solo numero può sembrare semplice, ma esistono tre modi diversi:

- **Media (mean):** somma dei valori divisa per il numero di osservazioni
- **Mediana (median):** valore centrale dei dati ordinati
- **Moda (mode):** valore più frequente

Perché la media può essere fuorviante?

- La **media è sensibile agli outlier**: pochi valori estremi possono spostarla molto.
- Esempio: se una persona guadagna 1 milione in una classe, la media del reddito non rappresenta quasi nessuno.
- La **mediana** è più robusta e spesso più rappresentativa in distribuzioni sbilanciate.


Distribuzione del Reddito (Istogramma + KDE)
Frequenza
0
50
100
150
200
250
300
25000
50000
75000
100000
125000
150000
175000
200000
Reddito (€)
Boxplot del Reddito
25000
50000
75000
100000
125000
150000
175000
200000
Reddito (€)
Violin Plot del Reddito
0
50000
100000
150000
200000
Reddito (€)

Definizione: Sia $x_1, x_2, \ldots, x_n$ una **collezione di osservazioni** *(redditi individuali)*.

Formula della media

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Come si calcola la mediana

- Ordina i dati in ordine crescente: $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$
- Se n è dispari: la mediana è il valore centrale $x_{\left(\frac{n+1}{2}\right)}$
- Se n è pari: la mediana è la media tra i due valori centrali:

$$\text{Mediana} = \frac{x_{(n/2)} + x_{(n/2+1)}}{2}$$

la **moda** è il valore x che appare con la maggiore frequenza

$$\text{Moda} = \arg\max_x f(x)$$

dove $f(x)$ è la frequenza di x nel campione.

Esempio – Redditi simulati (distribuzione di Pareto)

- **Media (mean)**: €24.872.000
- **Mediana (median)**: €15.807.000
- **Moda (mode)**: €20.000

Nota: la moda può essere fortemente influenzata da valori estremi, soprattutto in distribuzioni asimmetriche.

Perché servono gli indici di dispersione?

UNIVERSITÀ DI PARMA

Non basta la media!

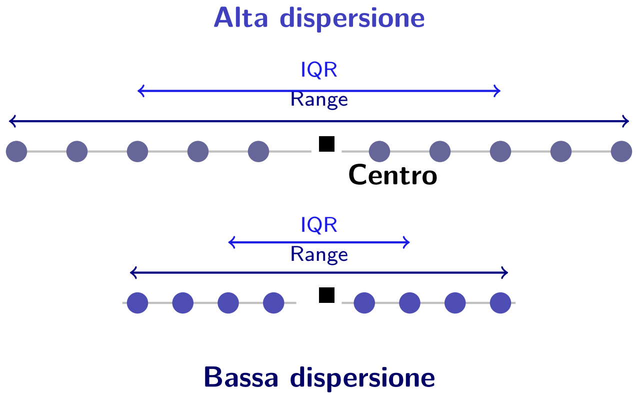
Conoscere solo la media non è sufficiente per descrivere un insieme di dati.
Esempio: la misura media delle scarpe maschili è utile, ma non dice nulla sulla **varietà di taglie** necessarie per produrre scarpe per tutti.

Indici di dispersione principali:

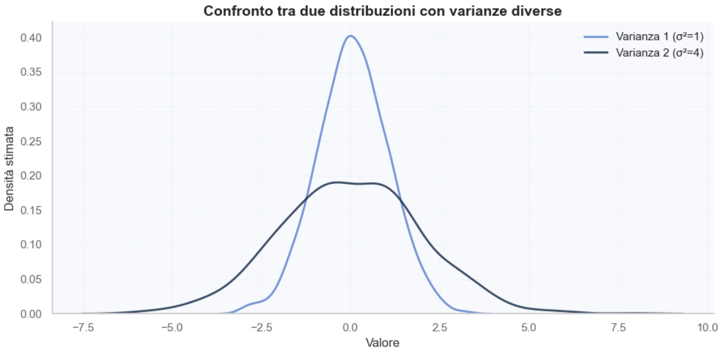
- **Range**: differenza tra valore massimo e minimo. Molto sensibile agli estremi.
- **Intervallo interquartile (IQR)**: differenza tra il 75º e il 25º percentile. Robusto contro outlier.
- **Deviazione standard**: misura quanto i valori si discostano dalla media. Adatta a distribuzioni simmetriche.

Dispersione, Range e IQR

UNIVERSITÀ DI PARMA



La deviazione standard (σ) è una misura della **dispersione** dei dati rispetto alla media (μ).



- Se σ è **piccola**, i dati sono **concentrati** intorno alla media.
- Se σ è **grande**, i dati sono **più dispersi**.

Definizione Formale

Sia data una popolazione di N elementi:

$$x_1, x_2, \ldots, x_N$$

La **varianza della popolazione** è:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

La **deviazione standard della popolazione** è:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

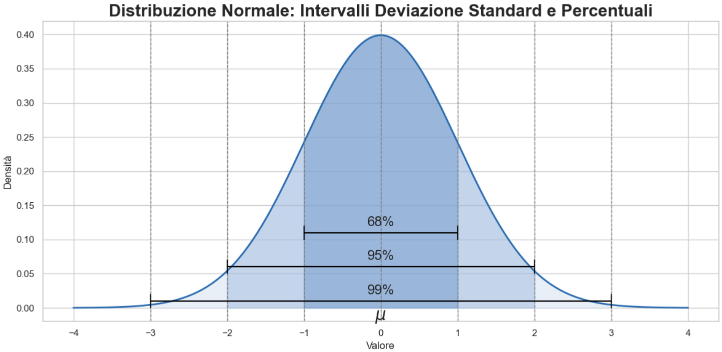
Interpretazione della Deviazione Standard

UNIVERSITÀ DI PARMA

Se i dati seguono (approssimativamente) una **distribuzione normale**, possiamo interpretare σ come intervallo di confidenza empirico:

- Circa **68%** dei dati si trova in $[\mu - \sigma, \mu + \sigma]$
- Circa **95%** dei dati si trova in $[\mu - 2\sigma, \mu + 2\sigma]$
- Circa **99.7%** dei dati si trova in $[\mu - 3\sigma, \mu + 3\sigma]$

Questo è noto come la **regola empirica** o **regola dei tre sigma**.


Distribuzione Normale: Intervalli Deviazione Standard e Percentuali
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Densità
68%
95%
99%
μ
−4
−3
−2
−1
0
1
2
3
4
Valore

Coefficiente di Variazione (CV)

UNIVERSITÀ DI PARMA

Il **coefficiente di variazione** è una misura di dispersione relativa, utilizzata per confrontare la variabilità di dataset con unità di misura o scale diverse.

Definizione:

$$\mathrm{CV} = \frac{\sigma}{\mu} \times 100\%$$

Dove:

- $\sigma =$ deviazione standard del campione
- $\mu =$ media del campione

Interpretazione:

- Valori più alti indicano maggiore variabilità relativa.
- Utile per confrontare variabilità tra gruppi anche con medie diverse.

Confronto della variabilità della domanda

UNIVERSITÀ DI PARMA

Obiettivo

Capire quale prodotto ha una *domanda più variabile* nel tempo.

Dati mensili di vendita

Prodotto	Media vendite	Dev. standard	Coef. Variazione
A	1000 unità	200	0.20
B	100 unità	40	0.40

Attenzione

Non possiamo confrontare direttamente le **deviazioni standard** di due serie con scale diverse: il prodotto B ha una variabilità relativa maggiore nonostante la dev. standard sia inferiore.

Cos'è la Covarianza?

UNIVERSITÀ DI PARMA

Definizione

La **covarianza** misura il modo in cui due variabili quantitative variano insieme. Indica se, e in che direzione, due variabili tendono a muoversi in relazione l'una all'altra.

Formula:

$$\mathrm{Cov}(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

- Se > 0: le variabili crescono insieme
- Se < 0: una cresce mentre l'altra decresce
- Se $= 0$: assenza di relazione lineare

Esempio: Calcolo della Covarianza

UNIVERSITÀ DI PARMA

Dati:

- X = [1, 2, 3], Y = [2, 4, 6]
- $\bar{x} = 2, \quad \bar{y} = 4$

Calcolo:

$$\mathrm{Cov}(X, Y) = \frac{(1-2)(2-4) + (2-2)(4-4) + (3-2)(6-4)}{3} = \mathbf{1.33}$$

Interpretazione

Cov(X, Y) = 1.33 → esiste una relazione positiva tra X e Y: quando X aumenta, anche Y tende ad aumentare.

Limiti della Covarianza: un esempio

UNIVERSITÀ DI PARMA

Due coppie di variabili

X	1	2	3	4
Y	2	4	6	8
W	10	20	30	40
Z	20	40	60	80

Covarianze:

$$\mathrm{Cov}(X,Y) = 3.33 \qquad \mathrm{Cov}(W,Z) = 333.33$$

Problema

La relazione tra le variabili è identica in entrambi i casi, ma la covarianza cambia a causa della scala. **Non possiamo confrontarle direttamente.**