

Trovare il Centro

Una guida professionale alla scelta tra Media, Moda e Mediana.



La cassetta degli attrezzi del data analyst.



La sfida: come si riassume un insieme di dati con un solo valore?

Di fronte a una serie di numeri, che si tratti di voti, prezzi o punteggi, sorge una domanda fondamentale: qual è il valore “tipico” o “centrale”?

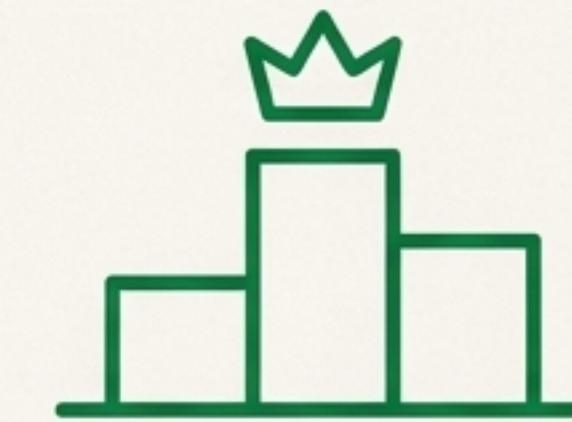
La risposta non è sempre ovvia. A seconda dello strumento che usiamo, possiamo raccontare storie molto diverse. Questa guida ti insegnerà a scegliere lo strumento giusto per il lavoro.

I tre indicatori di tendenza centrale.



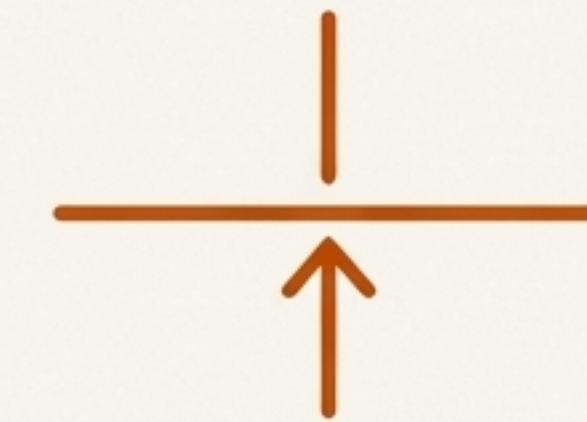
LA MEDIA

Il punto di equilibrio
del dataset.



LA MODA

Il valore più popolare
o frequente.



LA MEDIANA

Il valore che si trova
esattamente nel mezzo.

Strumento 1: La Media Aritmetica



La media aritmetica è il valore che si ottiene sommando tutti i dati e dividendo il risultato per il numero totale di dati. È il baricentro matematico dell'insieme.

Formula:

$$\text{Media} = \frac{\text{Somma dei valori}}{\text{Numero dei valori}}$$

Come si calcola la Media: un esempio pratico

Analizziamo questa serie di voti:

Dati:



Passo 1: Somma

$$7 + 8 + 6 + 7 + 8 + 6 + 9 + 5 = \mathbf{56}$$

Passo 2: Divisione

$$56 / \mathbf{8} \text{ (numero di voti)}$$

Risultato:

Media = 7

Nota: Ogni voto contribuisce equamente al risultato finale.

Strumento 2: La Moda



La moda è il valore che compare **con maggiore frequenza** all'interno di un insieme di dati.

Caratteristiche chiave:

- È l'unico indicatore utilizzabile anche per dati non numerici (es. colori, categorie).
- Un dataset può non avere una moda, oppure averne più di una (bimodale, trimodale, ecc.).

Come si individua la Moda: l'arte di contare.

Esempio 1 (Numerico)

Dati: Numeri di scarpe:

37, **38**, **38**, **38**, **38**, 39, 40, 37

Risultato:

Moda = **38** (compare 4 volte)

Esempio 2 (Non Numerico)

Dati: Colori preferiti:

rosso, **blu**, **blu**, **blu**, **blu**, verde, rosso,
giallo

Risultato:

Moda = **blu** (compare 4 volte)

Strumento 3: La Mediana



La mediana è il **valore centrale** di un **insieme di dati** che sono stati **preventivamente ordinati** dal più piccolo al più grande.

Regola Fondamentale:

L'ordinamento dei dati è il primo passo, non negoziabile. La Mediana non si **preoccupa dei valori, ma solo della loro posizione.**

Come si calcola la Mediana: un processo in due fasi.

Passo 1 (Universale): Ordinare sempre i dati in ordine crescente.

Scenario A: Numero di dati DISPARI

Esempio (Altezze in cm):

Dati iniziali: 142, 155, 138, 160, 147



Dati ordinati: 138, 142, **147**, 155, 160

Risultato:

La mediana è il valore che si trova esattamente al centro.

Mediana = **147 cm**

Scenario B: Numero di dati PARI

Esempio (Pesi in kg):

Dati iniziali: 45, 52, 48, 55, 50, 47



Dati ordinati: 45, 47, **48, 50**, 52, 55

Risultato:

La mediana è la media dei due valori centrali.

Mediana = $(48 + 50) / 2 = \mathbf{49 \text{ kg}}$

Campi di applicazione: dove usare ogni strumento.

Media

Applicazioni: Media dei voti, temperatura media, altezza media di una classe, punteggio medio in un test.



Moda

Applicazioni: Taglia più venduta, gusto di gelato più richiesto, canzone più ascoltata, colore di auto più popolare.



Mediana

Applicazioni: Reddito mediano di una popolazione, prezzo mediano delle case in una città, tempo mediano di percorrenza.



Scegliere lo strumento giusto: una guida alla decisione.



IDEALE QUANDO...

- ✓ I dati sono distribuiti in modo omogeneo e non ci sono valori anomali.

ATTENZIONE A...

- ⚠ Valori estremi (molto alti o molto bassi) possono distorcerla significativamente.



IDEALE QUANDO...

- ✓ Si cerca il valore più frequente o si lavora con dati categorici (non numerici).

ATTENZIONE A...

- ⚠ Può non esistere, oppure possono essercene più di una, rendendo l'interpretazione ambigua.



IDEALE QUANDO...

- ✓ Nel dataset sono presenti valori estremi (outlier) che falserebbero la media.

ATTENZIONE A...

- ⚠ Richiede l'ordinamento preliminare dei dati, che può essere computazionalmente oneroso per dataset molto grandi.

Caso di studio: Punti segnati in 7 partite di basket.

Dati:

12

8

15

8

10

8

22

 Media

$$(12 + 8 + 15 + 8 + 10 + 8 + 22) / 7 = 83 / 7 \approx \mathbf{11.9 \text{ punti}}$$

 Moda

8 punti (è il valore che compare più frequentemente, 3 volte)

 Mediana

Dati ordinati: 8, 8, 8, **10**, 12, 15, 22

Valore centrale: **10 punti**

Interpretazione dei risultati: tre numeri, tre storie

Risultati a confronto:



Media: **11.9**



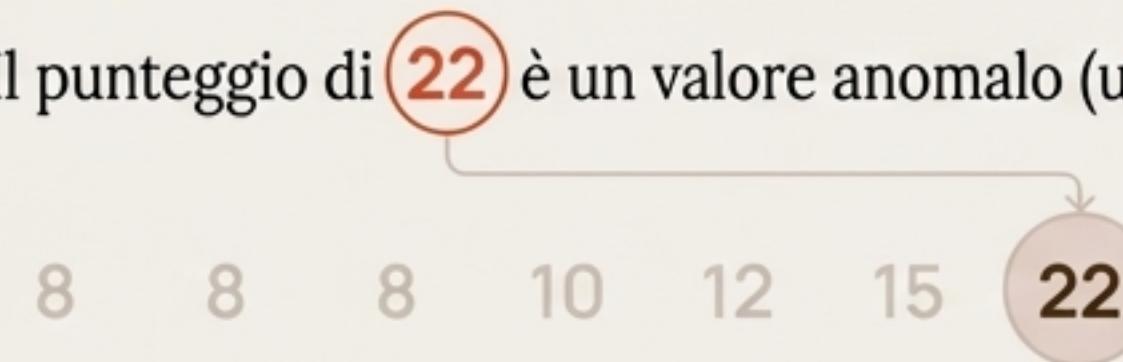
Mediana: **10**



Moda: **8**

Analisi Chiave:

Il punteggio di **22** è un valore anomalo (un ‘outlier’).



Spiegazione:

Questo valore estremo ha ‘gonfiato’ la **Media**, portandola a quasi 12, un valore superiore a 5 delle 7 prestazioni. La **Mediana** e la **Moda**, invece, non sono influenzate da questo valore estremo, e forniscono una rappresentazione forse più ‘onesta’ della prestazione tipica del giocatore.

La scelta non è matematica, è interpretazione.

Media, Moda e Mediana non sono in competizione.
Sono strumenti diversi che illuminano aspetti diversi dei tuoi dati.

 La **Media** descrive il baricentro.

 La **Moda** rivela la tendenza più comune.

 La **Mediana** indica il vero punto centrale, immune agli estremi.

Il tuo compito da analista non è solo calcolare un numero, ma scegliere lo strumento che racconta la storia più accurata e rappresentativa del fenomeno che stai osservando.