

Indici di tendenza centrale

Media, mediana e moda.



SAPIENZA
UNIVERSITÀ DI ROMA

Indici di tendenza centrale

- Gli indici di tendenza centrale individuano gli aspetti “tipici”, ovvero i valori più rappresentativi della distribuzione
- Questi indici consentono di riassumere un’intera distribuzione di frequenza in un unico numero
- Gli indici di tendenza centrale più utilizzati sono:
Moda → (scale nominali, ordinali, a intervalli e a rapporti)
- Mediana → (scale ordinali, a intervalli e a rapporti)
- Media → (solo scale a intervalli e a rapporti)

Esempio

- Dati 35 voti in una classe di scuola media superiore:
- 5, 8, 2, 5, 7, 3, 7, 4, 8, 4, 5, 7, 5, 3, 5, 6, 5, 6, 6, 7, 6, 5, 6, 5, 4, 6, 6, 6, 7, 8, 4, 9, 4, 6, 6
- Quale potrebbe considerarsi un valore rappresentativo del campione?

La moda

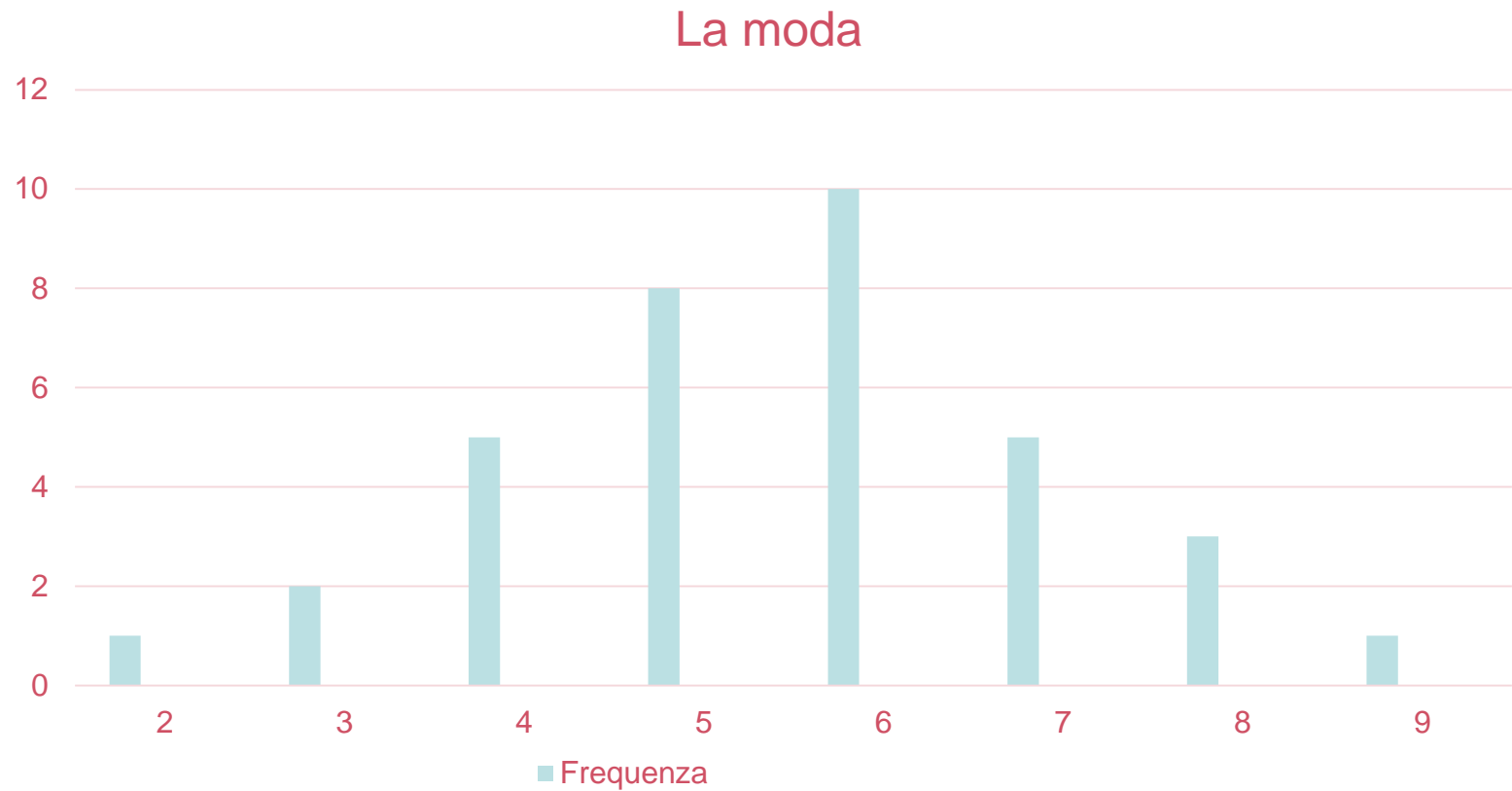
- La moda è l'osservazione che si presenta con maggiore frequenza nella distribuzione dei dati
- Può essere calcolata con qualunque tipo di scala di misura (nominale, ordinale, a intervalli o a rapporti)
- E' l'unica misura di tendenza centrale per dati su **scala nominale**, poiché dipende esclusivamente dal calcolo delle **frequenze**

La moda è il valore con la frequenza maggiore, ovvero più elevata

Voti	Frequenze
2	1
3	2
4	5
5	8
6	10
7	5
8	3
9	1
TOTALE	35

La moda è **6**,
attenzione da non
confondere con la
frequenza (che invece
è **10**)

La moda in istogramma



Distribuzioni bimodali

In senso stretto



In senso lato



La mediana

- La mediana è il valore **che divide la distribuzione in due parti uguali**
- È il valore della distribuzione al di sopra o al di sotto del quale cade un ugual numero di osservazioni
- È il valore che occupa la posizione centrale in una serie ordinata di dati: metà osservazioni ottengono punteggi inferiori alla mediana, l'altra metà ottiene punteggi superiori
- Può essere calcolata su scale ordinali, a intervalli e a rapporti ma non su scale nominali

Calcolare la mediana

- Ordinare le osservazioni in ordine crescente
- Se dispari utilizzare l'operazione $(n + 1) / 2$ per individuare la posizione cui corrisponderà il valore (mediana)
- Se pari utilizzare l'operazione $n/2$ e $n/2+1$ identificando così due posizioni e due valori associati (coppia mediana)

Esempio di calcolo

- Supponiamo di misurare l'attenzione di 11 bambini durante un compito in classe misurando quante volte mettono in atto comportamenti di distrazione:
 - $N=11$
 - Numero di comportamenti per ciascun alunno: 7 5 1 3 4 10
8 4 0 8 9
- Per calcolare la mediana disponiamo i tutti i valori in ordine crescente
 - 0 1 3 4 4 5 7 8 8 9 10

Esempio di calcolo della mediana

- Dato che $n=11$ l'operazione da eseguire è $(n+1)/2$, quindi $(11+1)/2$
- $12/2= 6$
- 6 indica la POSIZIONE della mediana sulla sequenza ordinata dei valori.
- Quindi...
 - 0 1 3 4 4 **5** 7 8 8 9 10
 - 5 è il VALORE associato alla posizione indicata dalla formula

Esempio di calcolo coppia mediana

- Consideriamo la stessa serie di valori tranne l'ultimo:

1a 2a 3a 4a 5a 6a 7a 8a 9a 10a

0 1 3 4 4 5 7 8 8 9

- In questo caso il numero di osservazioni è pari ($n=10$), quindi la mediana è una coppia di valori (4 e 5) che si trova nelle seguenti posizioni:

$$n/2 = 10/2 = 5$$

$$n/2+1 = (10/2)+1 = 5+1 = 6$$

- Se si vuole esprimere la tendenza centrale con un solo valore (anche quando n è pari), si può utilizzare la media delle due mediane
- Nell'esempio: $(4 + 5) / 2 = 4.5$

La mediana e la tabella di frequenze

- Quando i dati sono presentati in una tabella di frequenza, è possibile calcolare la mediana attraverso l'uso delle **frequenze cumulate**
- Le frequenze cumulate corrispondono al valore della frequenza assoluta più la somma di tutte le frequenze di ordine inferiore. L'ultima frequenza cumulata corrisponde al totale delle osservazioni.

Frequenze cumulate e mediana

Voti	Frequenze	Frequenze cumulate
2	1	1
3	2	3
4	5	8
5	8	16
6	10	26
7	5	31
8	3	34
9	1	35
TOTALE	35	

Calcolo della mediana

- 1 – si calcolano le frequenze cumulate
- 2 – si identifica la posizione in cui si colloca la mediana (POS Me), tramite le solite formule:
 - $\text{POS Me} = (n+1)/2 \rightarrow$ se n è dispari
 - $\text{POS Me} = n/2$ e $(n/2+1) \rightarrow$ se n è pari
- 3 – si individua la mediana: la mediana è il **valore** che si trova in **corrispondenza della frequenza cumulata** uguale a **POS Me**
- se POS Me non corrisponde a nessuna frequenza cumulata, la mediana si trova nella prima frequenza cumulata superiore a POS Me

Esempio di calcolo della mediana data la tabella di frequenze

- $POS Me = (n+1)/2$
- $36/2 = 18$

La frequenza cumulata **18 NON** è presente nella tabella.

26 è la prima frequenza cumulata maggiore a POS Me (quella prima è 16).

In pratica tutte le posizioni tra 17 e 26 fanno parte del valore 6. Dato che la POS Me calcolata è la 18esima posizione il valore associato a quella posizione è 6.

La nostra mediana è 6

Voti	Frequenze	Frequenze cumulate
2	1	1
3	2	3
4	5	8
5	8	16
6	10	26
7	5	31
8	3	34
9	1	35
TOTALE	35	

Esempio di calcolo della mediana data la tabella di frequenze

- Infatti, se disponiamo le osservazioni in una sequenza ordinata:

2 3 3 4 4 4 4 4 5 5 5 5 5 5 5 5 6 6 6 6 6 6 6 6
6 7 7 7 7 7 8 8 8 9

6 è il valore che si trova nella 18° posizione

Circa metà soggetti ha ottenuto un voto inferiore a 6

L'altra metà ha ottenuto un voto superiore a 6

In poche parole, 6 è il valore che divide la distribuzione in due parti uguali

Indici di posizione

- Oltre alla mediana, che rappresenta un indicatore di tendenza centrale della distribuzione, vi sono altri indici di posizione, che vengono calcolati in maniera del tutto analoga alla mediana:
- Quartili → dividono la distribuzione in quattro parti uguali
- Decili → dividono la distribuzione in dieci parti uguali
- Percentili → dividono la distribuzione in cento parti uguali

I Quartili

- Quartili: sono quei tre valori che dividono la distribuzione in quattro parti uguali:
- il primo quartile (Q1) è il valore al di sotto del quale cade il 25% dei casi
- il secondo quartile (Q2) è il valore al di sotto del quale cade il 50% dei casi (corrisponde alla mediana)
- il terzo quartile (Q3) è il valore al di sotto del quale cade il 75% dei casi

Calcolo dei Quartili

- Per calcolare un quartile il procedimento è simile a quello del calcolo della mediana.
- Si calcolano le frequenze cumulate
- Si identifica la POSIZIONE occupata dal quartile con le seguenti operazioni

Primo quartile= $(n+1)/4$

Secondo quartile= $((n+1)/4) * 2$

Primo quartile= $((n+1)/4) * 3$

Primo quartile= $((n+1)/4) * 4$

- Si trova, nella distribuzione delle frequenze cumulate, la posizione (POS) del quartile che si intende calcolare e si legge il valore corrispondente

Esempio precedente

- Calcolate i quartili dell'esempio precedente ($n=35$)

- Pos Q1=9
- Pos Q2=18
- Pos Q3=27

- Q1=5
- Q2=6
- Q3=7

Voti	Frequenze	Frequenze cumulate
2	1	1
3	2	3
4	5	8
5	8	16 Pos Q1
6	10	26 Pos Q2
7	5	31 Pos Q3
8	3	34
9	1	35
TOTALE	35	

Decili

- sono nove valori che dividono la distribuzione in dieci parti uguali:
- il primo decile (D1) è il valore al di sotto del quale cade il 10% dei casi
- il secondo decile (D2) è il valore al di sotto del quale cade il 20% dei casi
- il quinto decile (D5), è il valore al di sotto del quale cade il 50% casi (**corrisponde alla mediana**)
- il nono decile (D9), è il valore al di sotto del quale cade il 90% casi
- La formula è $(n+1)/10$ * il decile che ci interessa indentificare

I percentili

- Percentili: sono 99 valori che dividono la distribuzione in cento parti uguali:
- il primo percentile (C1) è il valore al di sotto del quale cade l'1% dei casi
- il secondo percentile (C2) è il valore al di sotto del quale cade il 2% dei casi
- il 50-esimo percentile (C50), è il valore al di sotto del quale cade il 50% casi (corrisponde alla mediana)
- il 99-esimo percentile (C99), è il valore al di sotto del quale cade il 99% casi
- La formula è $(n+1)/100$ * il percentile che ci interessa indentificare

Media

- Volendo riassumere in un solo numero un insieme di dati, possiamo calcolarne il valore medio (es. la media dei vostri esami)
- La formula per calcolare la media è

- $$\frac{\bar{x} = x_1 + x_2 + x_3 + \dots + x_n}{n}$$

- $$M = \sum_1^n i x_i$$

- Ovvero la somma di tutti i punteggi diviso il numero di punteggi osservati.

Calcolo della media

- Nell'esempio precedente
- $\bar{x} = \frac{2+3+\dots}{35} = 5,6$

Calcolo della media attraverso le frequenze

- Quando si lavora direttamente sulle tabelle i frequenza, una formula alternativa per il calcolo della media prevede di:
- Sommare i VALORI della variabile moltiplicati per la rispettiva FREQUENZA
- $(2*1)+(3*2)+(4*5)+(5*8)+(6*10)+(7*5)+(8*3)+(9*1)$
- Si divide per n (n=35)
- $M=5,6$

Moda, mediana e media

- La moda è l'indice meno informativo, viene infatti calcolata sulla sole frequenze.
- La mediana è più informativa della moda: anch'essa si basa sulle frequenze ma considera anche l'ordine tra le osservazioni
- In genere la media è l'indice di tendenza centrale più informativo, in quanto considera i valori numerici di tutte le osservazioni
- Non sempre, però, la media è l'indice più adatto ad individuare il valore più rappresentativo di una distribuzione
- Un limite della media è la sua sensibilità ai valori estremi, o mancanza di "robustezza" (soprattutto quando la distribuzione è formata da poche unità)