

## **Unità didattica N° 18**

### **Le medie**

- 1) Le medie**
- 02) Media aritmetica**
- 03) Media geometrica**
- 04) Media armonica**
- 05) Mediana**
- 06) Moda**
- 07) Media quadratica**

## Le medie

Una **media** è un numero che esprime una sintesi di una distribuzione statistica semplice . Indichiamo con  $x_i$  ( $i = 1, 2, \dots, n$ ) gli **n** valori assunti da una *variabile statistica* **X** , ciascuno di frequenza assoluta  $n_i$  ( $i = 1, 2, \dots, n$ ) . Il **valore medio statistico** è quel numero compreso tra il più piccolo e il più grande dei dati  $x_i$  , capace di riassumere significativamente l'insieme di tutti i dati . Nel caso di una variabile statistica divisa in classi converrà sostituire preliminarmente a ciascuna

classe il suo valore centrale :  $V_{ci} = \frac{x_{i-1} + x_i}{2}$  con  $i = 1, 2, \dots, n$

così da riportarsi al caso precedente . Per il calcolo della media si usano tutti i valori centrali con le rispettive frequenze .

Le MEDIE ( dette anche **parametri di posizione** ) vengono distinte in due grandi categorie :

1) **Medie algebriche** , o *analitiche* o ferme o di **calcolo** , se per la loro determinazione si tiene conto di tutti i valori della variabile statistica .

Appartengono a tale categoria la **media aritmetica** , la *media geometrica* , la media quadratica , la **media armonica** .

2) **medie di posizione** o lasche , se derivano dall'esame della posizione degli  $x_i$  all'interno della distribuzione . Appartengono a tale categoria la **moda** e la *mediana* .

*Medie algebriche*  $\left\{ \begin{array}{l} 1) \text{Media aritmetica} \\ 2) \text{Media geometrica} \\ 3) \text{Media armonica} \\ 4) \text{Media quadratica} \end{array} \right.$   
(si calcolano tenendo conto di tutti i valori della distribuzione)

*Medie di posizione*  $\left\{ \begin{array}{l} 1) \text{Moda} \\ 2) \text{Mediana} \end{array} \right.$   
(si calcolano tenendo conto solo di i particolari valori della distribuzione)

### Definizione di media secondo Chisini

Data una distribuzione  $x_1, x_2, \dots, x_n$  una sua media è la quantità  $\bar{x}$  che sostituita a ciascun termine della distribuzione lascia inalterato il risultato delle operazioni  $f$  eseguite sui termini della distribuzione .

**Media aritmetica ( A ,  $\bar{x}$  , m )**

Se un certo carattere **X** assume le **modalità**  $x_1, x_2, \dots, x_n$  , se la funzione delle **modalità**  $f(x_i)$  è del tipo :

$$f(x_1, x_2, \dots, x_n) = x_1 + x_2 + \dots + x_n \quad \text{definiamo } \textbf{media aritmetica}$$

**semplice** quel valore  $\bar{x} = m$  tale che  $f(x_i) = f(\bar{x}_i)$  , cioè tale che :

$$x_1 + x_2 + \dots + x_n = \bar{x} + \bar{x} + \dots + \bar{x} = n \cdot \bar{x}$$

$$\bar{x} = m = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Se le intensità del carattere **X** si presentano con una frequenza  $n_i$  , nel calcolo del valore medio dobbiamo tenere conto dei pesi  $n_i$  . Abbiamo la **media aritmetica ponderata** :

$$\bar{x} = m = \frac{n_1 x_1 + n_2 x_2 + \dots + n_n x_n}{n_1 + n_2 + \dots + n_n} = \frac{\sum_{i=1}^n n_i x_i}{\sum_{i=1}^n n_i} = \frac{\sum_{i=1}^n n_i x_i}{N}$$

Essa è denominata **media aritmetica ponderata** in quanto ogni modalità  $x_i$  interviene nel calcolo con un **peso** ( cioè un *coefficiente d ' importanza* ) pari alla sua frequenza . In questo caso la *funzione delle intensità* è del tipo :

$$f(x_i) = n_1 x_1 + n_2 x_2 + \dots + n_n x_n$$

$$f(x_i) = f(\bar{x}_i) \Rightarrow n_1 x_1 + n_2 x_2 + \dots + n_n x_n = n_1 \bar{x} + n_2 \bar{x} + \dots + n_n \bar{x} = (n_1 + n_2 + \dots + n_n) \bar{x}$$

$$\bar{x} = m = \frac{n_1 x_1 + n_2 x_2 + \dots + n_n x_n}{n_1 + n_2 + \dots + n_n} = \frac{\sum_{i=1}^n n_i x_i}{\sum_{i=1}^n n_i} = \frac{\sum_{i=1}^n n_i x_i}{N}$$

La funzione delle intensità :  $f(x_i) = x_1 + x_2 + \dots + x_n = n_1 x_1 + n_2 x_2 + \dots + n_n x_n$

è **invariante** se a ciascuna determinazione della variabile statistica si sostituisce la sua media aritmetica  $\bar{x}$  . Questo significa che la **media aritmetica** può e deve essere usata ogni volta che si desidera rappresentare sinteticamente la distribuzione con un unico parametro , senza alterare la somma delle intensità .

- Il numeratore delle media aritmetica è chiamato **ammontare del carattere** .

**Proprietà della media aritmetica**

**01 )** Ad ogni variabile statistica **X** si può associare una variabile statistica  $\hat{X}$  denominata **scarto** o **scostamento** dalla media aritmetica o più semplicemente *scarto o scostamento* . Essa è definita ,

per ciascuna unità osservata , dalla differenza  $\hat{x}_i = x_i - \bar{x}$  tra la modalità della stessa unità e la media aritmetica  $\bar{x}$  . **La somma algebrica degli scarti dalla media aritmetica è nulla** , cioè :

$$\sum_{i=1}^n \hat{x}_i = \sum_{i=1}^n (x_i - \bar{x}) = 0$$

Detta proprietà è del tutto intuitiva e deriva dalla circostanza che gli scarti sono in parte positivi e in parte negativi . Poiché la *media aritmetica* realizza l' ' equidistribuzione del carattere senza avanzi o disavanzi , la somma degli scarti positivi deve essere uguale alla somma in valore assoluto di quelli negativi .

Per le distribuzioni di frequenza ( dati raggruppati ) abbiamo :

$$\sum_{i=1}^n n_i \cdot \hat{x}_i = \sum_{i=1}^n n_i \cdot (x_i - \bar{x})$$

Una immediata conseguenza di questa proprietà è la seguente : se più numeri formano una progressione aritmetica , la loro media aritmetica semplice è uguale al termine centrale ( se il numero dei termini è dispari ) o alla media aritmetica dei due termini centrali ( se il numero dei termini è pari ) .

**02)** Non esiste un valore  $\alpha \neq \bar{x}$  per il quale risulti :  $\sum_{i=1}^n (x_i - \alpha) = 0$  cioè  $\sum_{i=1}^n (x_i - \alpha) \neq 0$

**03)** La somma dei quadrati di tutti gli scarti di un insieme di n valori  $x_1, x_2, \dots, x_n$  dalla loro media aritmetica  $\bar{x}$  è minima , cioè minore della somma dei quadrati degli scarti calcolati rispetto ad un altro qualsiasi valore  $\alpha \neq \bar{x}$  .

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \text{minimo}$$

$$\sum_{i=1}^n n_i (x_i - \bar{x})^2 = \text{minimo}$$

**04)** La media aritmetica è una **media interna** , nel senso che il suo valore risulta sempre non minore dell'intensità minima  $x_{\min}$  e non maggiore dell'intensità massima  $x_{\max}$  ; essa è , pertanto , una media nel senso di Cauchy . Questa proprietà è detta **proprietà di internalità**

**05)** La media aritmetica è quel valore  $\bar{x}$  che , sostituito ai dati , lascia invariata una loro determinata funzione , precisamente la **funzione somma totale**  $f(x_1, x_2, \dots, x_n) = x_1 + x_2 + \dots + x_n$  .

**06)** Quindi **la media aritmetica è fondata sull'invarianza dell'ammontare complessivo del carattere** . In altre parole la media aritmetica è quel valore costante  $\bar{x}$  che sostituito alle singole osservazioni ne lascia inalterata la somma .

**07 )** La media aritmetica gode della proprietà **traslativa** , nel senso che aggiungendo ( o sottraendo ) a tutte le osservazioni la stessa quantità **a**, anche la media aritmetica risulta aumentata ( diminuita ) di detta quantità :

$$\frac{\sum_{i=1}^n (x_i \pm a)}{n} = \frac{\sum_{i=1}^n x_i \pm \sum_{i=1}^n a}{n} = \frac{1}{n} \sum_{i=1}^n x_i \pm \frac{1}{n} \sum_{i=1}^n a = \bar{x} \pm \frac{n \cdot a}{n} = \bar{x} \pm a$$

**08) Proprietà associativa** La media della somma di due gruppi di n dati ciascuno è uguale alla somma delle due medie .

Indicando con  $x_i$  gli n dati di una variabile statistica **X** e con  $z_i$  gli n dati di un'altra variabile statistica **Z** , abbiamo :

$$\frac{\sum_{i=1}^n (x_i + z_i)}{n} = \frac{\sum_{i=1}^n x_i}{n} + \frac{\sum_{i=1}^n z_i}{n} \quad \text{cioè :} \quad \overline{x + z} = \bar{x} + \bar{z}$$

### Metodo abbreviato per il calcolo della media aritmetica

Sia **a** un qualsiasi valore diverso dalla media aritmetica  $\bar{x}$  , valore che consideriamo come **media**

**provvisoria** dei dati . Sia inoltre  $\frac{\sum_{i=1}^n (x_i - a)}{n}$  la **media degli scarti** dei dati  $x_i$  dalla media provvisoria **A** . Allora la media provvisoria **A** e la media aritmetica  $\bar{x}$  sono legate dalla relazione :

$$\bar{x} = a + \frac{\sum_{i=1}^n (x_i - a)}{n}$$

Una proprietà simile vale nel caso in cui vengono introdotte le frequenze  $n_i$  :

$$\bar{x} = a + \frac{\sum_{i=1}^n n_i (x_i - a)}{\sum_{i=1}^n n_i = N}$$

In questo caso conviene scegliere **A** uguale al valore  $x_i$  avente la frequenza  $x_i$  più elevata .

Allo scopo di operare con cifre più piccole , tenendo presente la proprietà **7)** , abbiamo :

$$\bar{x} = a + \frac{b}{n} \cdot \sum_{i=1}^n \left( \frac{x_i - a}{b} \right)$$

e nel caso in cui vengono introdotte le frequenze  $n_i$  abbiamo :

$$\bar{x} = a + \frac{b}{\sum_{i=1}^n n_i} \cdot \sum_{i=1}^n n_i \left( \frac{x_i - a}{b} \right)$$

Quindi per snellire ulteriormente il calcolo della media aritmetica si sottrae alle singole osservazioni  $x_i$  un valore opportuno **A** ( *media provvisoria* ) e si dividono gli scarti così ottenuti per un altro opportuno valore **b** . Si procede poi al calcolo della media aritmetica dei valori così trasformati , la si moltiplica per **b** e si aggiunge al risultato trovato la media provvisoria **A** .

A prima vista può sembrare che questo modo di calcolare la media aritmetica costituisca un'inutile complicazione rispetto al metodo originale , ma a volte , se non si dispone di una calcolatrice , l'uso del metodo abbreviato può risparmiare molto lavoro . Infatti il metodo abbreviato riduce notevolmente le dimensioni dei numeri che debbono essere sommati e tanto più vicina è la media provvisoria rispetto alla media reale tanto più piccoli sono i termini della somma . Questo metodo sarà particolarmente utile quando dovremo calcolare la media aritmetica partendo da dati raggruppati .

- La media aritmetica semplice applicata a dati statistici , trova per lo più applicazione nelle **serie**, sia di frequenza che di intensità , nelle **seriazioni di frequenza** trova generalmente applicazione la media ponderata applicata alle modalità quantitative , prendendo come pesi le frequenze .
- Nel caso di una variabile statistica divisa in classi il calcolo della *media aritmetica* deve essere preceduto dalla determinazione dei valori centrali delle classi . Il sostituire alle singole classi il loro valore centrale è giustificato dal fatto che si può quasi sempre ritenere che la frequenza sia distribuita uniformemente all' interno di ciascuna classe . Su questa stessa ipotesi si fonda , come abbiamo visto , anche la costruzione degli *istogrammi* .

### ESEMPI

■ Un alunno nelle prove scritte di matematica ha ottenuto i seguenti risultati : 4 , 6 , 8 , 7 , 5 . Calcolare la media

aritmetica :  $\bar{x} = \frac{4 + 6 + 8 + 7 + 5}{5} = 6$

| Voti in una classe |                 | colonna di servizio | 5 alunni hanno preso 4 , 2 alunni hanno preso 5 , 4 alunni hanno |
|--------------------|-----------------|---------------------|------------------------------------------------------------------|
| Voti $x_i$         | frequenze $n_i$ | $n_i x_i$           |                                                                  |
| 4                  | 5               | 20                  |                                                                  |

|               |               |           |                                                                                                                                     |
|---------------|---------------|-----------|-------------------------------------------------------------------------------------------------------------------------------------|
| 5             | 2             | 10        | <p>preso 6 , un alunno ha preso 7 , tre alunni hanno preso 8 .</p> $\bar{x} = \frac{\sum_{i=1}^5 n_i x_i}{N} = \frac{85}{15} = 5,6$ |
| 6             | 4             | 24        |                                                                                                                                     |
| 7             | 1             | 7         |                                                                                                                                     |
| 8             | 3             | 24        |                                                                                                                                     |
| <b>totale</b> | <b>N = 15</b> | <b>85</b> |                                                                                                                                     |

| Voti assegnati agli alunni della VC distribuiti per classi |                 |                          | Colonna di servizio | $\bar{x} = \frac{\sum_{i=1}^4 n_i \cdot V_c}{N} = \frac{61,5}{11} \approx 5,6$ <p>è la <b>media aritmetica</b> richiesta .</p> |
|------------------------------------------------------------|-----------------|--------------------------|---------------------|--------------------------------------------------------------------------------------------------------------------------------|
| Voti $x_i$                                                 | frequenze $n_i$ | Valore centrale $V_{ci}$ | $n_i \cdot V_{ci}$  |                                                                                                                                |
| 1 — 2                                                      | 3               | $1,5 = \frac{1+2}{2}$    | 4,5                 |                                                                                                                                |
| 2 — 4                                                      | 2               | 3                        | 6                   |                                                                                                                                |
| 4 — 8                                                      | 1               | 6                        | 6                   |                                                                                                                                |
| 8 — 10                                                     | 5               | 9                        | 45                  |                                                                                                                                |
| totale                                                     | N = 11          |                          | 61,5                |                                                                                                                                |

### Media geometrica ( G , $M_o$ )

Se un certo carattere **X** assume le modalità  $x_1, x_2, \dots, x_n$  , se la funzione delle modalità  $f(x_1, x_2, \dots, x_n) = x_1 \cdot x_2 \cdots x_n$  **[A]** è **invariante** rispetto al prodotto delle modalità del carattere quantitativo **X** definiamo **media geometrica semplice** il numero  $M_o = M_g = G$  per il quale risulta :

$$x_1 \cdot x_2 \cdots x_n = M_o \cdot M_o \cdot M_o \cdots M_o = M_o^n = G^n$$

$$G = M_o = \sqrt[n]{x_1 \cdot x_2 \cdots x_n} = \sqrt[n]{\prod_{i=1}^n x_i} \quad [B]$$

Quindi nel caso della media geometrica semplice si assume come invariante il prodotto delle intensità  $x_i$  , cioè << la **media geometrica semplice** è quel valore  $M_o$  che sostituito alle singole osservazioni  $x_i$  ne lascia inalterato il prodotto >> . La media geometrica semplice  $M_o$  rappresenta e sostituisce la variabile statistica considerata , e quindi l'intera distribuzione , ai fini della valutazione sintetica dell'espressione [A] , assunta come invariante . <sup>(1)</sup>

Nel caso di una variabile statistica con frequenza l'invariante [A] diventa :

<sup>(1)</sup> Dicesi media geometrica di n termini positivi , quel valore che sostituito a ciascun termine lascia immutato il loro prodotto .

$$x_1^{n_1} \cdot x_2^{n_2} \cdot \dots \cdot x_n^{n_n} = M_o^{n_1} \cdot M_o^{n_2} \cdot \dots \cdot M_o^{n_n} = M_o^N \quad \text{con} \quad N = n_1 + \dots + n_n = \sum_{i=1}^n n_i$$

$$G = M_o = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \dots \cdot x_n^{n_n}} = x_1^{\frac{n_1}{N}} \cdot x_2^{\frac{n_2}{N}} \cdot \dots \cdot x_n^{\frac{n_n}{N}} = x_1^{r_1} \cdot x_2^{r_2} \cdot \dots \cdot x_n^{r_n} = \sqrt[n]{\prod_{i=1}^n x_i^{n_i}}$$

$M_o$  prende il nome di **media geometrica ponderata**. Questo tipo di media analitica è detta geometrica perché coincide col termine centrale di una successione geometrica con un numero  $n$  dispari di termini; mentre se  $n$  è pari coincide con la radice quadrata del prodotto dei due termini centrali. Anche per la media geometrica, come si è osservato per la media aritmetica, valgono le medesime considerazioni, quando i valori del carattere quantitativo siano ripartiti per classi di intensità. La media geometrica, semplice o ponderata, è di uso meno frequente della media aritmetica. In ogni caso, viene usata quando il carattere è moltiplicativo e si debba determinare una distribuzione uniforme dei suoi valori, senza alterarne il prodotto.

### Media armonica ( $H$ , $M_{-1}$ )

Se un certo carattere  $X$  assume le intensità  $x_1, x_2, \dots, x_n$ , se la funzione delle intensità è del tipo

$$f(x_i) \text{ è del tipo : } f(x_1, x_2, \dots, x_n) = \frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} \quad [*]$$

definiamo **media armonica semplice** quel valore  $H$  ( $M_{-1}$ ) tale che :  $f(x_i) = f(H)$  cioè tale che

$$\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} = \frac{1}{H} + \frac{1}{H} + \dots + \frac{1}{H}$$

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

Quindi nel caso della *media armonica* si assume come **invariante** la somma dei reciproci delle intensità, cioè << la media armonica è il numero  $H$  che sostituito alle singole osservazioni ne lascia invariata la somma dei reciproci .>> La **media armonica semplice** rappresenta e sostituisce la variabile statistica considerata, e quindi l'intera distribuzione, ai fini della valutazione sintetica dell'espressione [\*] assunta come invariante. Nel caso di distribuzioni di

frequenze l'invariante [\*] diventa :

$$\frac{n_1}{x_1} + \frac{n_2}{x_2} + \dots + \frac{n_n}{x_n} = \frac{n_1}{H} + \frac{n_2}{H} + \dots + \frac{n_n}{H} =$$

$$\frac{n_1 + n_2 + \dots + n_n}{H} = \frac{N}{H}$$



$$H = \frac{\frac{n_1}{x_1} + \frac{n_2}{x_2} + \dots + \frac{n_n}{x_n}}{\frac{n_1}{x_1} + \frac{n_2}{x_2} + \dots + \frac{n_n}{x_n}} = \frac{\sum_{i=1}^n \frac{n_i}{x_i}}{\sum_{i=1}^n \frac{n_i}{x_i}} = \frac{N}{\sum_{i=1}^n \frac{n_i}{x_i}}$$

La **media armonica** non è molto utilizzata nelle applicazioni statistiche . Tuttavia , se due fenomeni sono connessi fra loro in maniera inversa ( come prezzi e potere d'acquisto della moneta , tempi di percorrenza e velocità ,...) e se uno di essi viene sintetizzato come la media aritmetica l'altro deve essere sintetizzato come la media armonica .

### Proprietà della media armonica

1) La somma degli scarti di n termini dalla loro media armonica rapportati ai termini stessi è **nulla** ,

cioè :

$$\sum_{i=1}^n \frac{x_i - H}{x_i} = 0$$

2) La somma dei quadrati degli scarti dei reciproci dei termini dal reciproco della media armonica è

minima , cioè :

$$\sum_{i=1}^n \left( \frac{1}{x_i} - \frac{1}{H} \right)^2 = \min$$

### Media quadratica ( Q , M<sub>2</sub> )

Se come **invariante** assumiamo la somma dei quadrati delle intensità x<sub>i</sub> , espressa dalla funzione

$$f(x_1, x_2, \dots, x_n) = x_1^2 + x_2^2 + \dots + x_n^2 \quad [D]$$

la quantità M<sub>Q</sub> = M<sub>2</sub> tale che f(x<sub>i</sub>) = f(M<sub>Q</sub>) prende il nome di **media quadratica semplice** .

$$f(x_i) = f(M_Q) \Rightarrow x_1^2 + x_2^2 + \dots + x_n^2 = (M_Q)^2 + (M_Q)^2 + \dots + (M_Q)^2 \Rightarrow$$

$$M_Q = M_2 = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}$$

Quindi la **media quadratica semplice** è quel valore M<sub>Q</sub> = M<sub>2</sub> che sostituito alle singole osservazioni ne lascia inalterata la somma dei quadrati .

La **media quadratica semplice** rappresenta e sostituisce la variabile statistica considerata ai fini della valutazione sintetica della [D] , assunta come **invariante distributivo** .

Nel caso di **dati raggruppati** ( distribuzioni di frequenza diverse dall'unità ) l'invariante [D] si scrive :

$$f(x_1, x_2, \dots, x_n) = n_1 x_1^2 + n_2 x_2^2 + \dots + n_n x_n^2$$

In questo caso la *condizione di invarianza* diventa :

$$n_1x_1^2 + n_2x_2^2 + \dots + n_nx_n^2 = n_1(M_Q)^2 + n_2(M_Q)^2 + \dots + n_n(M_Q)^2$$

da cui ricaviamo il valore della **media quadratica ponderata** :

$$M_Q = M_2 = \sqrt{\frac{n_1x_1^2 + n_2x_2^2 + \dots + n_nx_n^2}{n_1 + n_2 + \dots + n_n}} = \sqrt{\frac{\sum_{i=1}^n n_i x_i^2}{\sum_{i=1}^n n_i}} = \sqrt{\frac{\sum_{i=1}^n n_i x_i^2}{N}} \quad [T]$$

**Disposizione di un collettivo di famiglie secondo il numero dei componenti**

| Componenti $x_i$ | Famiglie $n_i$ | $x_i^2$ | $n_i \cdot x_i^2$ |
|------------------|----------------|---------|-------------------|
| 1                | 890            | 1       | 890               |
| 2                | 1.512          | 4       | 6.048             |
| 3                | 1.830          | 9       | 16.470            |
| 4                | 1415           | 16      | 22.640            |
| 5                | 802            | 25      | 20.050            |
| 6                | 304            | 36      | 10.944            |
| totale N         | 6.753          |         | 77.042            |

$$M_Q = \sqrt{\frac{77042}{6753}} = \sqrt{11,4086} = 3,38$$

$$\sum_{i=1}^6 n_i x_i^2 = 77.042$$

### Mediana ( $M_e$ )

La statistica si avvale anche di valori medi che non dipendono da tutti i valori osservati , ma si basano soltanto su alcuni valori particolari . Tra questi valori , che prendono il nome di **medie di posizione** o medie lasche , i più usati sono la **mediana** e la moda .

In questo paragrafo sarà presa in considerazione la **mediana** . Si tratta di una *media di posizione* che riveste una particolare importanza in diversi fenomeni statistici .

Dati  $n$  grandezze  $x_1, x_2, \dots, x_n$  disposte in ordine non decrescente , cioè tali che  $x_1 \leq x_2 \leq \dots \leq x_n$  definiamo **mediana** quella grandezza che divide la distribuzione in due parti tali che il numero dei termini che la precedono sia uguale al numero dei termini che la seguono . Se il numero dei dati disponibili è **dispari** la mediana  $M_e$  è rappresentata dal termine centrale , cioè quello che occupa il

posto  $\frac{n+1}{2}$  della successione esaminata . Se  $n$  è **pari** non esiste un termine mediano , bensì una

coppia di valori mediani . In questo caso si ha un **intervallo mediano** o zona mediana , mentre la mediana risulta indeterminata , potendosi assumere come tale un qualsiasi valore dell'intervallo mediano . Nella pratica , tuttavia , è consuetudine adottare come mediana la semisomma dei due termini centrali che occupano rispettivamente i posti  $\frac{n}{2}$  ed  $\frac{n}{2} + 1$  .

### **Determinazione della mediana $M_e$ per dati non raggruppati**

I valori osservati dalla variabile statistica siano  $x_1, x_2, \dots, x_n$  , ciascuno con frequenza unitaria , disposti in ordine crescente . Per i posti centrali si ha :

**a) n dispari** : unico posto centrale  $P = \frac{n + 1}{2}$

**b) n pari** : si hanno due posti centrali ;  $P_1 = \frac{n}{2}$  ( primo posto centrale)  $P_2 = \frac{n}{2} + 1$  ( secondo posto centrale )

La **mediana** si determina immediatamente non appena si siano individuate , nella successione ordinata , le intensità corrispondenti ai suddetti posti centrali .

#### **ESEMPIO : calcolo della mediana ; dati non raggruppati ; numero dispari di osservazioni**

Le **modalità** della variabile statistica ( ad esempio i voti riportati all'esame di statistica ) siano 21 , 25 , 27 , 28 , 30 . Poiché il numero delle unità osservate ( studenti ) è  $n = 5$  , numero dispari , l'unico posto centrale è dato da  $P = \frac{5 + 1}{2} = 3$  . Si tratta del terzo posto centrale della sequenza crescente delle intensità .

Ad esso corrisponde l'intensità  $M_e = 27$  che è la **mediana** della successione . In effetti , l'intensità mediana  $M_e = 27$  separa due gruppi ugualmente numerosi , ciascuno costituito da due osservazioni .

#### **ESEMPIO : calcolo della mediana ; dati non raggruppati ; numero pari di osservazioni**

Le modalità della variabile statistica , ad esempio i voti riportati all'esame di Economia Politica , siano **18 , 19 , 20 , 22 , 23 , 26 , 28 , 30** . Poiché il numero delle unità osservate ( studenti ) è  $n = 8$  , numero pari , i due posti centrali sono  $P_1 = \frac{8}{2} = 4$  ( quarto posto ) e  $P_2 = \frac{8}{2} + 1 = 4 + 1 = 5$  ( quinto posto ) . Le **modalità** ad essi corrispondenti sono , rispettivamente , **22 e 23** . Può quindi

assumersi come **mediana** un qualsiasi valore dell'intervallo [22,23] e in particolare il suo valore

centrale :

$$M_e = \frac{22 + 23}{2} = 22,5 .$$

Detto valore isola a destra e a sinistra due insiemi ugualmente numerosi , ciascuno costituito da quattro elementi ( studenti ) .

### Determinazione della mediana $M_e$ per dati raggruppati e caratteri discontinui

In questo caso , nota la tabella dei dati raggruppati si costruisce la tabella delle frequenze cumulate .

Se  $\sum n_i$  è **dispari** si ha un solo posto centrale , se N è **pari** si hanno due posti centrali i quali molto spesso coincidono , in tal caso la mediana è il loro valore comune .

Se non coincidono , si può assumere come mediana una qualunque intensità dell'intervallo mediano da essi definito o , più semplicemente , la semisomma dei due valori centrali . Sia data la seguente distribuzione di frequenza di una variabile statistica discreta :

|       |    |    |    |    |    |    |
|-------|----|----|----|----|----|----|
| $x_i$ | 3  | 4  | 5  | 6  | 7  | 8  |
| $n_i$ | 11 | 17 | 38 | 30 | 45 | 50 |

Per potere stabilire qual è il dato centrale di questa distribuzione si procede come segue :

- Si controlla che le modalità  $x_i$  siano tutti disposti in ordine crescente
- Si sommano tutte le frequenze assolute  $n_i$  e si divide il risultato per due

$$\frac{11 + 17 + 38 + 30 + 45 + 50}{2} = 95,5$$

individuando la posizione centrale ( o le posizioni centrali se i dati sono in numero pari ) . Nel nostro caso la **posizione centrale** è la 96-esima .

c) per trovare quale dato corrisponde alla posizione centrale si calcolano le *frequenze cumulate* finché non si arriva ad una frequenza cumulata uguale o maggiore della posizione centrale .

| $x_i$    | $n_i$ | frequenze assolute cumulate $C_i$ |
|----------|-------|-----------------------------------|
| 3        | 11    | 11                                |
| 4        | 17    | 28                                |
| 5        | 38    | 66                                |
| <b>6</b> | 30    | <b>96</b>                         |
| 7        | 45    | 141                               |
| 8        | 50    | 191                               |

|                |
|----------------|
| Totale 191 = N |
|----------------|

La **mediana** è **6** perché corrisponde alla frequenza cumulata 96 . Infatti , se elencassimo in ordine crescente i dati uno ad uno , troveremmo il valore **6** al 67-esimo posto , e quindi , per quanto ci interessa nella 96-esima posizione .

La seguente tabella ci dice quali sono i voti riportati da un certo gruppo di studenti.

| Voti<br>$x_i$ | frequenze<br>assolute $n_i$ | frequenze<br>cumulate $C_i$ |
|---------------|-----------------------------|-----------------------------|
| 3             | 1                           | 1                           |
| 4             | 2                           | 3                           |
| 5             | 5                           | 8                           |
| 6             | 7                           | 15 → 10,5                   |
| 7             | 4                           | 19                          |
| 8             | 1                           | 20                          |
| totale        | 20                          |                             |

$$P_1 = \frac{N}{2} = \frac{20}{2} = 10 \text{ 1° posto centrale}$$

$$P_2 = \frac{N}{2} + 1 = \frac{20}{2} + 1 = 11 \text{ 2° posto centrale}$$

$$P = \frac{P_1 + P_2}{2} = \frac{10 + 11}{2} = 10,5 \text{ Posto centrale}$$

Come **posizione centrale teorica** possiamo assumere il valore 10,5 . Fra le frequenze cumulate il primo valore uguale o maggiore di 10,5 è **15** a cui corrisponde l'intensità **6** che è il **valore mediano** cercato . Concludendo possiamo affermare che la mediana vale :  $M_e = 6$  .

#### Osservazione

Geometricamente la **mediana** è il valore  $x_e$  di **X** ( ascissa ) corrispondente alla retta verticale ( di equazione  $x = x_e$  ) che divide un **istogramma** in due parti equivalenti ( cioè aventi la stessa area )

**Calcolo della mediana di una variabile statistica continua i cui dati sono distribuiti in classi .**

In questo caso non è possibile ricostruire la graduatoria delle osservazioni in quanto queste non si conoscono singolarmente , ma si sa solo quante di esse cadono in ciascuna classe .

Nella pratica la **serie delle frequenze cumulate** , analogamente a quanto si è visto nel caso di una variabile statistica per singoli valori , ci consente di accertare che la mediana capita all'interno della prima classe per la quale la frequenza cumulata uguaglia o supera il numero

$$\frac{N}{2} = \frac{n_1 + n_2 + \dots + n_n}{2} , \text{ ossia metà delle osservazioni } N .$$

Indichiamo con  $x_{r-1} - x_r$  tale **classe mediana** ( $x_{r-1}$  limite inferiore della classe mediana ,  $x_r$  limite superiore della classe mediana ) , con  $c_r$  ( $c_{r-1}$  ) la frequenza cumulata corrispondente

all'intervallo mediano ( intervallo che precede quello mediano ) , e con  $\bar{c}$  l'elemento della distribuzione cui corrisponde il valore mediano  $M_e$  cercato , avremo :

$$M_e = x_{r-1} + \frac{x_r - x_{r-1}}{C_r - C_{r-1}} (\bar{C} - C_{r-1}) \quad [A]$$

dove  $\bar{C}=P$  con :  $P = \frac{N+1}{2}$  se N è dispari ,  $P = \frac{P_1 + P_2}{2}$  se N è pari

[ con  $P_1 = \frac{N}{2}$  = primo posto centrale ,  $P_2 = \frac{N}{2} + 1$  = secondo posto centrale ]

Di solito per N pari si prende  $P = \frac{N}{2}$  e non la semisomma dei due valori centrali . Qualche autore

prende  $P = \frac{N}{2}$  per n pari o dispari .

$x_r - x_{r-1}$  = **ampiezza della classe mediana**

$\bar{C}-C_{r-1}$  = **frequenza assoluta della classe mediana**

<< **Pellicole cinematografiche programmate in Italia nel 1976 per classi d'incasso**  
( in milioni di lire ) >>

| classi d'incasso $x_i$ | Pellicole $n_i$ | frequenze assolute cumulate<br>$C_i$ |
|------------------------|-----------------|--------------------------------------|
| 0 — 2                  | 22              | 22                                   |
| 2 — 5                  | 20              | 42                                   |

|                        |     |               |
|------------------------|-----|---------------|
| 5—10                   | 18  | 60            |
| 10—20                  | 35  | 95            |
| 20—50 $x_{r-1}$        | 47  | 142 $C_{r-1}$ |
| $x_{r-1}$ 50—100 $x_r$ | 52  | 194 $C_r$     |
| 100—200                | 48  | 242           |
| 200— $\omega$          | 81  | 323           |
| Totale N               | 323 |               |

$$162 = \bar{C}$$

La **frequenza cumulata** del posto centrale è  $\bar{C} = P = \frac{323+1}{2} = 162$ . La pellicola che fa registrare l' **incasso mediano** è al 162<sup>a</sup>. Essa si trova nella sesta frequenza cumulata ( $C_r = C_6 = 194$ ), per cui la **classe mediana** è quella che va da 50 a 100 milioni (50—100).

Risulta:  $x_{r-1} = 50$ ,  $x_r = 100$ ,  $x_r - x_{r-1} = 100 - 50 = 50$ ,  $C_r = 194$ ,  $C_{r-1} = 142$ ,  $\bar{C} = 162$

$$M_e = x_{r-1} + \frac{x_r - x_{r-1}}{C_r - C_{r-1}} (\bar{C} - C_{r-1}) = 50 + \frac{100 - 50}{194 - 142} (162 - 142) = 50 + \frac{50}{52} \cdot 20 =$$

$$= 50 + 19,2 = 69,2 \text{ milioni.}$$

### Moda ( $M_d$ )

La **moda** ( detta anche *valore modale*) è un altro parametro molto usato per la descrizione sintetica della distribuzioni di frequenze.

Per *variabili statistiche discontinue*, essa può definirsi come **quella modalità del carattere che si osserva nel maggior numero dei casi**: in corrispondenza di essa la **frequenza assume un valore massimo, per cui la moda è una sede di massimo delle frequenze**.

In molte applicazioni pratiche, la **moda** è preferita alle medie analitiche ed alla stessa media aritmetica. Questo si spiega tenendo presente che essa, al contrario delle medie analitiche, non rappresenta il risultato di una manipolazione dei dati e quindi ha un carattere meno astratto.

La **moda** può non esistere ( quando tutti i valori hanno la stessa frequenza ) e se esiste può non essere unica. Se esiste ed è unica si parla di **distribuzione unimodale**, se invece non è unica la distribuzione è detta **plurimodale**. E' possibile, quindi, imbattersi in **distribuzioni zeromodali** (cioè prive di moda), **unimodali** ( con una sola moda ) e **plurimodali**.

Se un collettivo è distribuito secondo modalità ( non raggruppate in classi ) di un carattere discreto , l'identificazione della moda è immediata : basta scorrere la colonna delle frequenze e individuare il dato che corrisponde alla **massima frequenza** .

Se il carattere è presentato in classi , occorre distinguere due situazioni :

1°) le classi hanno tutte la stessa ampiezza : la moda cade nella classe di maggiore frequenza

2°) **le classi hanno ampiezze diverse** : occorre allora eliminare l'influenza perturbatrice di tale diversità .

Assumendo il **modello di equiripartizione** , occorre dividere le frequenze delle classi per il numero dei valori del carattere in esse contenuto . La *moda cade nella classe per la quale si riscontra il più elevato quoziente* . Quindi nel caso di distribuzioni di frequenza con classi aventi ampiezze diverse , la **classe modale** coincide con la classe avente maggiore **densità di frequenza** , intendendo per densità di frequenza di una classe il rapporto fra la frequenza della classe e l'ampiezza della classe stessa .

In prima approssimazione possiamo identificare la **moda** con la posizione centrale ( semisomma degli estremi ) della classe modale .

### Esempi

| Addetti  | Unità commerciali |
|----------|-------------------|
| 1—2      | 40                |
| 3—4      | 90                |
| 5—6      | 120               |
| 7—8      | 70                |
| 9—10     | 15                |
| Totale N | 335               |

In questa tabella , le classi sono tutte di uguale ampiezza ;  
poiché la frequenza massima si riscontra  
in corrispondenza della classe 5—6 , quest'ultima è  
la **classe modale**

<< Unità commerciali di un comune secondo il numero degli addetti >>



| Addetti  | Unità commerciali | densità di frequenza |
|----------|-------------------|----------------------|
| 1—2      | 20                | 20                   |
| 3—5      | 90                | 45                   |
| 6—10     | 125               | 31,25                |
| 11—20    | 100               | 11,11                |
| 21—35    | 75                | 5,36                 |
| Totale N | 410               |                      |

In questa tabella le classi hanno diversa ampiezza .  
Per ciascuna classe bisogna calcolare la **densità di frequenza** . La massima densità di frequenza è 45 che si riscontra per la seconda classe 3—5 che è la **classe modale** richiesta .

<< Consideriamo gli occupati in agricoltura secondo il numero di ore settimanali >>

| Classi | occupati | Densità di frequenza |
|--------|----------|----------------------|
| 1—16   | 5.600    | 373,33               |
| 16—21  | 7.600    | 1250                 |
| 21—26  | 4.400    | 880                  |
| 26—31  | 6.500    | 1300                 |
| 31—36  | 1.700    | 340 $d_{-1}$         |
| 36—41  | 18.600   | 3720 $d_o$           |
| 41—46  | 8.700    | 1740 $d_1$           |

La massima **densità di frequenza** è **3720** a cui corrisponde la classe 36—41 che è , pertanto , la **classe modale** .

Potremmo assumere come **moda** il seguente valore

$$\text{centrale : } M_d = \frac{36 + 41}{2} = 38,5$$

Un valore più preciso della **moda** si ottiene applicando la seguente formula :

$$M_d = x_o + \frac{\alpha}{2} \cdot \frac{d_1 - d_{-1}}{2d_o - d_1 - d_{-1}} \quad [\text{B}] \quad \text{dove :}$$

$d_o$  è la **densità di frequenza** della classe modale ,

$d_{-1}$  è la densità di frequenza della classe che precede la classe modale

$d_1$  è la *densità di frequenza* della classe che segue la classe modale

$x_o$  è il punto ( o valore ) centrale ( semisomma degli estremi ) della classe modale

$\alpha$  è l' **ampiezza** della classe modale

$\alpha$

$$M_d = 38,5 + \frac{5}{2} \cdot \frac{1740 - 340}{2 \cdot 3720 - 1740 - 340} = 39,2$$