Lezione 4: Indici di posizione

Corso di Statistica

Facoltà di Economia

Università della Basilicata

Prof. Massimo Aria

- aria@unina.it

Indice di posizione

Obiettivo di **una misura di posizione** è quello di sintetizzare in un singolo valore numerico l'intera distribuzione di frequenza così da consentire confronti nel tempo, nello spazio o tra situazioni differenti.

In altre parole, si intende individuare un valore che rappresenti la direzione, la tendenza centrale di una variabile statistica.

Tipologia di indici di posizione

Nel contesto della statistica descrittiva sono stati proposti numerosi indici di posizione.

Tra questi il concetto di media ha assunto un ruolo preponderante nella rappresentazione del concetto di posizione (o *centro*) di una distribuzione.

Tra gli indici di posizione maggiormente utilizzati si fa riferimento a:

- **Media**
- **Moda**
- **Mediana**

Il concetto di media

La **media** è un *concetto primitivo* per l'essere umano.
Pur non avendo nozioni di statistica, la moltitudine delle persone fa quotidianamente riferimento alla media per esprimere una sintesi.

La media è un indice di posizione per variabili quantitative.

Criterio della trasferibilità
Scelta una funzione "f", la media si definisce, secondo Chisini, come quel valore che se sostituito alle osservazioni di una distribuzione non ne muta il valore della funzione.

A seconda del criterio della trasferibilità considerato si distinguono differenti **tipologie di media**:

- *Media aritmetica ("f"=somma)*
- *Media geometrica ("f"=prodotto)*
- *Media armonica ("f"=somma degli inversi)*

Medie secondo Chisini

$$f(x_1, x_2, \cdots, x_n) = f(M, M, \cdots, M)$$

con

$$x_1 \le x_2 \le \cdots \le M \le \cdots \le x_{n-1} \le x_n$$

Criterio della trasferibilità di Chisini.

Media aritmetica

È il valore di posizione per eccellenza, a cui spesso si fa riferimento con il semplice termine "media".

La **media aritmetica** si definisce come "quel valore che, sostituito alle N osservazioni della distribuzione, non ne muta la somma degli elementi" (Chisini).

A seconda della distribuzione:

• Per distribuzioni unitarie della variabile X si parlerà di ***media aritmetica semplice***.

Ottenuta come somma dei valori assunti dalla distribuzione rapportata al numero di osservazioni.

• Per distribuzioni di frequenza della variabile X si parlerà di ***media aritmetica ponderata***.

Ottenuta come somma dei valori della X ponderati per le rispettive frequenze e rapportata al numero delle osservazioni.

Media *aritmetica* semplice

$$f(x_1, x_2, \cdots, x_N) = \sum_{l=1}^{N} x_l \Rightarrow$$

$$f(x_1, x_2, \cdots, x_N) = f(\mu, \mu, \cdots, \mu) \Leftrightarrow$$

$$\sum_{l=1}^{N} x_l = \sum_{l=1}^{N} \mu \Leftrightarrow \sum_{l=1}^{N} x_l = N\mu \Rightarrow$$

$$\boxed{\frac{\sum_{l=1}^{N} x_l}{N} = \mu}$$

Consideriamo come criterio f la somma degli elementi

Definizione della media aritmetica semplice

Media aritmetica ponderata

Consideriamo una variabile X nella forma:

$$(x_i, n_i) \qquad i = 1, 2, \cdots, k$$

…la media aritmetica sarà:

$$\mu = \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i} \qquad \mu = \sum_{i=1}^{k} x_i f_i$$

Utilizzando le frequenze assolute — Utilizzando le frequenze relative

Definizione della media aritmetica ponderata

Proprietà della media aritmetica

La media aritmetica gode delle seguenti proprietà:

1) **Internalità**. *La media è sempre compresa tra il valore minimo e il valore massimo della distribuzione*

2) **Baricentro**. *La media è pari a quel valore che rende nulla la somma degli scarti* (si definisce scarto la differenza tra un'osservazione e la media)

3) **Linearità**.
   a) *Se si aggiunge (o si sottrae) una costante a tutti i valori della distribuzione, la nuova media sarà pari alla vecchia media più (o meno)la costante*
   b) *Se si moltiplica tutti i valori della distribuzione per una costante, la nuova media sarà pari alla vecchia media per la costante*

4) **Associatività**. *Se si divide la distribuzione in g gruppi e se ne calcolano le medie, la media generale sarà pari alla media ponderata delle g medie dei gruppi*

5) **Minimo della somma degli scarti al quadrato**. *La media è pari a quel valore che rende minima la somma degli scarti al quadrato.*

Proprietà della media aritmetica

1) Internalità $\quad x_{\min} \le \mu \le x_{\max}$

2) Baricentro $\quad \sum (x_l - \mu) = 0$

3) Linearità $\quad \frac{1}{N} \sum (\alpha + \beta x_l) = \alpha + \beta\mu$

4) Associativa $\quad \mu = \frac{1}{N} \sum_{i=1}^{g} \mu_i \cdot n_i$

5) Minimo della somma degli scarti al quadrato

$$\sum (x_l - \mu)^2 = \min$$

Media aritmetica per distribuzioni in classi

In presenza di una distribuzione di frequenza in classi (nel caso di variabili continue), nel calcolo della media è necessario individuare, per ogni classe un valore rappresentativo della stessa.

Si introduce il concetto di **valore centrale della classe**.

Esso è ottenuto come semi-somma degli estremi di ogni classe.

La media è calcolata come media ponderata dei valori centrali per le rispettive frequenze delle classi.

media aritmetica per distribuzioni in classi

Si considerano i valori centrali di ogni classe

$$c_i = \frac{x_{i-1} + x_i}{2}$$

…la media aritmetica sarà:

$$\mu = \frac{\sum_{i=1}^{k} c_i n_i}{N} \qquad \mu = \sum_{i=1}^{k} c_i f_i$$

Utilizzando le frequenze assolute — Utilizzando le frequenze relative

Media geometrica

Se si definisce come criterio della trasferibilità il prodotto, la media secondo Chisini prende il nome di **media geometrica**.

Essa si definisce come il *la radice (di ragione N) del prodotto dei valori della distribuzione della variabile X.*

La media geometrica, coinvolgendo il prodotto delle modalità, va utilizzata in quei casi in cui la variabile varia in ragione proporzionale.

Esempio:

Si pensi alla necessità di calcolare un tasso di interesse medio di un investimento bancario di durata pluriennale.

Essendo il tasso variabile nel tempo allora la media geometrica può consentire di individuare il tasso medio per un dato montante dopo k anni.

Media geometrica semplice (per distribuzioni unitarie)

$$\mu_g = \sqrt[N]{\prod_{l=1}^{N} x_l} = \left(\prod_{l=1}^{N} x_l\right)^{\frac{1}{N}}$$

Media geometrica ponderata (per distribuzioni di frequenza)

$$\mu_g = \sqrt[N]{\prod_{i=1}^{k} (x_i)^{n_i}} = \left(\prod_{i=1}^{k} (x_i)^{n_i}\right)^{\frac{1}{N}}$$

Formulazione della media geometrica

Media armonica

Se si definisce come criterio della trasferibilità la somma degli inversi, la media secondo Chisini prende il nome di **media armonica**.

Essa si definisce come il *la somma degli inversi dei valori della distribuzione della variabile X rapportata al numero di osservazioni.*

La media armonica è utile quando occorre sintetizzare un rapporto tra variabili quando la somma dei termini a denominatore è pari ad una costante.

Esempio:

Si pensi al calcolo della velocità media di un automobile che ha percorso spazi diversi in tempi diversi.

La velocità media è la media armonica delle singole velocità ponderate per gli spazi percorsi nei singoli tratti.

Media armonica semplice (per distribuzioni unitarie)

$$\mu_a = \frac{N}{\sum_{l=1}^{N} \frac{1}{x_l}}$$

Media armonica ponderata (per distribuzioni di frequenza)

$$\mu_a = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} \frac{n_i}{x_i}}$$

Formulazione della media armonica

Medie lasche

Si definiscono medie lasche (o medie di posizione) quelle che utilizzano alcuni valori particolari della distribuzione di frequenza, per individuare una particolare modalità che ha una collocazione centrale rispetto a tutte le altre.

Tra le medie lasche, sono riconducibili:

• **Moda**
• **Mediana**

Moda

La **Moda** M_o (o *norma*) di una distribuzione di frequenza è la modalità cui corrisponde la massima frequenza, assoluta o relativa.

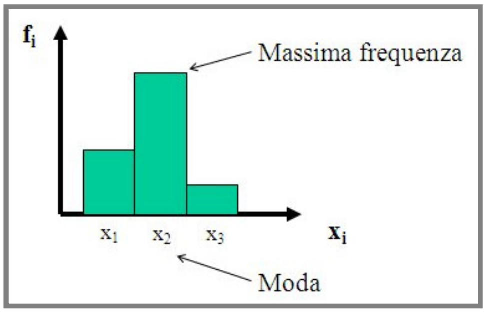
Per distribuzioni che presentano due o più modalità con massima frequenza si parla di *distribuzioni bimodali* o *plurimodali*.

La moda può essere calcolata per qualunque carattere statistico, sia esso qualitativo che quantitativo.

Nella realtà, essa trova però scarsa applicazione nel caso di variabili continue (distribuzione in classi) in quanto:

• appare irreale immaginare che vi sia un unico valore moda della distribuzione

• affermare che per una classe modale (classe con maggior frequenza) il valore centrale sia rappresentativo della stessa è una semplificazione non facilmente accettabile



Esempio della moda di una distribuzione

Mediana per variabili discrete

La **Mediana** M_e è il valore che bipartisce la distribuzione ordinata delle modalità di una variabile X, in modo tale che la metà delle osservazioni sia inferiore alla mediana e l'altra metà sia invece superiore.

La mediana è quindi quel valore per cui la funzione di ripartizione empirica vale ½: $F(M_e)=0,5$.

Il calcolo della mediana si differenzia a seconda che esso riguardi una distribuzione unitaria o una distribuzione di frequenza.

NB. Prima di calcolare la mediana bisogna sempre ordinare la distribuzione in ordine crescente sulla base delle modalità della variabile X.

La mediana per variabili discrete

(data una distribuzione ordinata della X)

Distribuzione unitaria

$$Me = \begin{cases} \dfrac{X_{(N/2)} + X_{(N/2)+1}}{2} & \text{N = pari} \\ X_{\left(\frac{N+1}{2}\right)} & \text{N = dispari} \end{cases}$$

Distribuzione di frequenza

$$Me = x_i \quad \text{tale che} \quad F(Me) = 0.5$$

Formulazione della mediana

Esempi

Distrib. unitaria

$\boxed{1, 2, 3, 4, 5, 5}$ $\qquad Me = \dfrac{x_3 + x_4}{2} = \dfrac{3+4}{2} = 3.5$

Distrib. di freq.

X	Fr. relative	Fr. cumulate
1	0.2	0.2
3	0.1	0.3
5	0.4	0.7
6	0.3	1
	1	

$\boxed{Me = x_i \text{ tale che } F(Me) = 0.5}$

$\boxed{Me = 5}$

Esempi del calcolo della mediana

Mediana per variabili continue

Per distribuzioni in classi, la mediana si calcola in due passi.

Innanzi tutto si perviene alla identificazione della classe mediana come quella classe la cui funzione di ripartizione F è pari a ½.

Successivamente si identifica, attraverso un calcolo proporzionale, il valore mediano all'interno della classe mediana.

La mediana per variabili continue

X	Fr. relative	Fr. Cumul.
$x_0 — x_1$	f_1	F_1
$x_1 — x_2$	f_2	F_2
•	•	•
$x_{i-1} — x_i$	f_i	F_i
•	•	•
$x_{k-1} — x_k$	f_k	1

Proporzione

$$(Me - x_{i-1}) : (x_i - x_{i-1}) = (0.5 - F_{i-1}) : (F_i - F_{i-1})$$

$$Me \cong x_{i-1} + (x_i - x_{i-1}) \frac{0.5 - F_{i-1}}{F_i - F_{i-1}}$$

Formulazione della mediana per variabili continue

Quartili

Come già affermato, la mediana è un indice che bipartisce egualmente la distribuzione ordinata.

Estendendo questo concetto a più ripartizioni è possibile definire i quartili.

Dividendo egualmente la distribuzione in quattro parti, si identificano:

• il **primo quartile** Q1. Rappresenta quella modalità tale che il 25% delle osservazioni assumono valori inferiori ad essa mentre il restante 75% hanno valori superiori.

• il **secondo quartile** Q2 che equivale alla mediana.

• il **terzo quartile** Q3. Rappresenta quella modalità tale che il 75% delle osservazioni assumono valori inferiori ad essa mentre il restante 25% hanno valori superiori.

Analogamente, ripartendo la distribuzione in dieci o cento parti, si possono definire i *decili* così come i *percentili*.

La mediana corrisponderà al quinto decile e al cinquantesimo percentile.

Quartili

(data una distribuzione ordinata della X)

$$Q_1 = x_i \quad \text{tale che} \quad F(Q_1) = 0.25$$

$$Q_2 = x_i \quad \text{tale che} \quad F(Q_2) = F(Me) = 0.50$$

$$Q_3 = x_i \quad \text{tale che} \quad F(Q_3) = 0.75$$

Formulazione dei quartili

Nella prossima lezione

Nella prossima lezione si affronteranno i seguenti argomenti:

- concetto di variabilità di una distribuzione
- indici statistici di variabilità assoluta e relativa