LA STATISTICA

Daniela Tondini

dtondini@unite.it

Facoltà di Bioscienze e Tecnologie agro-alimentari e ambientali e Facoltà di Medicina Veterinaria

C.L. in Biotecnologie

Università degli Studi di Teramo

INDICI STATISTICI

Nella ricerca scientifica e tecnologica è importante misurare la reale efficacia di interventi sul sistema oggetto di studio, ovvero valutare gli effetti complessivi indotti da una causa nota, pur nella mutevolezza ed instabilità dei risultati individuali. A tal riguardo, la Statistica ha proposto numerosi *indici statistici*, aventi quale obiettivo proprio la misurazione di due componenti del fenomeno oggetto di studio e di interesse scientifico: la *consistenza della sistematicità*, cioè la *centralità*, ovvero l'attitudine che hanno i fenomeni ad assumere tendenzialmente una certa dimensione all'osservazione, e la *variabilità* o *mutabilità*, cioè la *dispersione*, ovvero l'attitudine che hanno i fenomeni ad assumere dimensioni e tendenze diverse all'osservazione, nel tempo e nello spazio.

In particolare, la *centralità* è misurata dai cosiddetti *indici di posizione* (o *indici di tendenza centrale* o *indicatori di posizione* o *misure di tendenza centrale*) o *medie statistiche* o ancora più semplicemente *medie*, in grado di esprimere e sintetizzare la posizione di una distribuzione di frequenza mediante un valore reale rappresentativo della globalità del fenomeno, riassumendone gli aspetti ritenuti più importanti.

INDICI STATISTICI

Tali indici si possono ricavare effettuando operazioni che coinvolgono:

- tutti i termini della serie; in tal caso gli indici di posizione maggiormente usati, denominati *medie analitiche* o *di calcolo*, sono la media aritmetica M_a, la media geometrica M_g, la media armonica M_h e la media quadratica M_p tra le quali sussiste la seguente relazione:

$$\boxed{M_h \leq M_g \leq M_a \leq M_p}$$

- solo alcuni termini della serie, che si differenziano dagli altri per particolari caratteristiche; in tal caso gli indici di posizione maggiormente usati, denominati *medie posizionali* o *di posizione* o *lasche*, sono la mediana, la moda, i quartili.

INDICI STATISTICI

La ***media aritmetica semplice***, denominata semplicemente *media* ed indicata con M_a, usata per riassumere con un solo numero un insieme di n dati relativi ad un fenomeno misurabile, ovvero in presenza di <u>variabili quantitative</u> qualora la differenza tra un dato ed il precedente risulti costante, è ottenuta dividendo la somma di tutti gli n valori per il numero n di osservazioni; in formule è data da:

$$\boxed{M_a = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \ldots + x_n}{n}}$$

avendo indicato con n_i le frequenze delle x_i.

La media aritmetica di n numeri, dunque, è quel numero che, sostituito a ciascuno di essi, lascia invariata la somma totale e non può essere maggiore del valore più grande né minore del valore più piccolo.

INDICI STATISTICI

Esempio

La media aritmetica dei seguenti 5 = n numeri:

$$x_1 = 10; \quad x_2 = 13; \quad x_3 = 9; \quad x_4 = 7; \quad x_5 = 12$$

è data da:

$$M_a = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{1}{5}(10+13+9+7+12) = \frac{1}{5}(51) = \frac{51}{5} = 10,2$$

Si osservi che, sostituendo a ciascun x_i (i = 1, …, 5) il valore della media M_a e sommando i risultati, si ottiene;

$$10,2+10,2+10,2+10,2+10,2 = 5 \cdot M_a = 5 \cdot 10,2 = 51$$

che è proprio la somma degli x_i, 10 + 13 + 9 + 7 + 12 = 51.

INDICI STATISTICI

La ***media aritmetica ponderata***, invece, è ottenuta dividendo la somma di tutti gli *n* valori, moltiplicati per le rispettive frequenze, per il numero *n* di osservazioni; in formule è data da:

$$\boxed{M_a = \frac{1}{n}\sum_{i=1}^{s} x_i n_i = \frac{x_1 n_1 + x_2 n_2 + \ldots + x_n n_s}{n}}$$

avendo indicato con n_i le frequenze delle x_i e con *n* la somma delle n_i.
Tale denominazione deriva dal fatto che, a volte, le n_i non esprimono le frequenze ma opportuni *pesi di ponderazione* che tengono conto di altri aspetti rilevanti: -basti pensare, ad esempio, ai prezzi delle merci che vengono ponderati con cifre che esprimono le quantità vendute di ciascuna merce, allo scopo proprio di tener conto del valore globale (prezzo per quantità) degli scambi effettuati sul mercato considerato.

INDICI STATISTICI

Esempio

Se i voti riportati in matematica da n = 20 alunni di una scuola media di secondo grado sono riassunti nella seguente tabella:

Voti x_i	Alunni n_i
3	1
4	2
5	5
6	7
7	4
8	1
Tot.	20

allora la media aritmetica è data da:

$$M_a = \frac{1}{n}\sum_{i=1}^{s} x_i n_i = \frac{3\cdot 1 + 4\cdot 2 + 5\cdot 5 + 6\cdot 7 + 7\cdot 4 + 8\cdot 1}{20} = \frac{114}{20} = 5{,}7$$

INDICI STATISTICI

Se poi la v.s. X è divisa in intervalli, si può fare l'ipotesi che le intensità di X di ogni intervallo siano concentrate nel valore centrale della classe, in modo da riportarsi al caso discreto.

Esempio

Calcolare la statura media (aritmetica) dei coscritti italiani nati nel 1955.

Classi di statura (in cm)	**Valori centrali delle classi x_i**	**Frequenze n_i**	**Prodotti $x_i * n_i$**
meno d1 150	145	300	43500
150 ---160	155	12200	1891000
160 ---170	165	120800	19932000
170 ---180	175	160400	28070000
180 e oltre	185	36300	6715500
Tot.		**330000**	**56652000**

$$M_a = \frac{1}{n}\sum_{i=1}^{s} x_i n_i = \frac{56652000}{330000} = 171,67\,cm$$

La sostituzione delle singole classi con il valore centrale introduce un errore di approssimazione poco rilevante, anche se, tuttavia, si perde informazione.

INDICI STATISTICI

La media aritmetica, quindi, rappresenta quel valore che si può attribuire singolarmente a ciascuna unità statistica del collettivo lasciando invariato l'ammontare complessivo del carattere.

La media aritmetica di *n* numeri, dunque, rappresenta il *baricentro* dei dati e, quindi, propone un valore che equi-ripartisce il fenomeno tra le unità statistiche, pervenendo così a decisioni nelle quali contano, a parità numerica, gli estremi molto più dei valori centrali: la media aritmetica, infatti, costituisce un indice di equilibrio generale. Essendo, inoltre, la media statistica per eccellenza, consente un'ottima correzione degli errori accidentali commessi in una rilevazione statistica, risultando così utile, nonostante la sua scarsissima resistenza ai valori eccezionali, in tutti i campi della scienza e della tecnica in cui vengono effettuate misurazioni di qualunque genere.

Se la media coincide con una delle modalità viene detta *media effettiva* o *reale*; se, invece, non coincide con una delle modalità è detta *media di conto*.

INDICI STATISTICI

La ***media geometrica semplice***, usata quando le variabili quantitative risultano non lineari ma ottenute da un prodotto o da un rapporto di valori lineari non negativi e diversi da zero, si ottiene estraendo la radice *n*-esima del prodotto degli *n* termini; in formule è data da:

$$\boxed{M_g = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}}$$

dove Π è il simbolo di prodotto.

La media geometrica, considerata come quel valore che sostituito a ciascuno degli *n* dati ne lascia inalterato il prodotto, è usata soprattutto quando i dati non sono numerosi, i termini della distribuzione presentano valori molto differenti tra loro ed il rapporto tra un dato ed il precedente risulta costante (ad esempio, la determinazione del tasso di interesse medio equivalente alla sequenza dei tassi variabili, nel regime di capitalizzazione composta).

INDICI STATISTICI

Esempio

Uno studente ha sostenuto 6 = *n* esami riportando i seguenti voti:

$$x_1 = 21; \quad x_2 = 20; \quad x_3 = 24; \quad x_4 = 30; \quad x_5 = 28; \; x_6 = 25$$

La media geometrica dei voti è data da:

$$M_g = \sqrt[6]{\prod_{i=1}^{6} x_i} = \sqrt[6]{21 \cdot 20 \cdot 24 \cdot 30 \cdot 28 \cdot 25} = \sqrt[6]{211680000} \approx 24,41$$

INDICI STATISTICI

La ***media geometrica ponderata*** è usata, invece, qualora ci si trovi in presenza di una distribuzione costituita da *n* osservazioni e dalle relative frequenze; in formule, è data da:

$$\boxed{M_g = \sqrt[n]{\prod_{i=1}^{s} x_i^{n_i}} = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \ldots \cdot x_s^{n_s}}}$$

dove Π è il simbolo di prodotto ed $n = n_1 + n_2 + \ldots + n_s$.
Ogni termine, dunque, viene ponderato, ad esponente, con la relativa frequenza.
Analogamente, si può utilizzare la seguente formula:

$$\boxed{M_g = 10^{\left(\frac{\sum_{i=1}^{s} n_i \log x_i}{n}\right)}}$$

INDICI STATISTICI

Esempio

La seguente tabella riporta i voti ottenuti da un gruppo di studenti all'esame di Matematica:

Voti x_i	**Numeri di studenti** n_i
21	**5**
24	**6**
26	**10**
30	**4**
Tot.	**25**

La media geometrica ponderata è data da:

$$M_g = \sqrt[25]{21^5 \cdot 24^6 \cdot 26^{10} \cdot 30^4} =$$

$$= \sqrt[25]{4084101 \cdot 191102976 \cdot 141167095653376 \cdot 810000} \approx 25{,}00479$$

INDICI STATISTICI

Analogamente, utilizzando i logaritmi, si può impostare la seguente tabella:

Voti x_i	Numeri di studenti n_i	Logaritmi dei voti $\log x_i$	Prodotti $n_i \bullet \log x_i$
21	5	1,322219	6,611096
24	6	1,380211	8,281267
26	10	1,414973	14,149733
30	4	1,4771121	5,908485
Tot.	25		34,950582

Essendo, poi,

$$\frac{\sum_{i=1}^{4} n_i \log x_i}{n} = \frac{34,950582}{25} = 1,398023297$$

si ha la seguente media geometrica ponderata:

$$M_g = 10^{1,398023297} = 25,00479$$

INDICI STATISTICI

La ***media armonica semplice***, usata nello studio di variabili quantitative tra loro inversamente proporzionali, ovvero quando si deve trovare il valore medio, non del fenomeno considerato, ma di un fenomeno che è l'inverso del primo (ad esempio, prezzo di un bene e potere di acquisto della moneta, interesse effettivo che cresce al decrescere del costo del titolo, …), è pari al reciproco della media aritmetica dei reciproci dei termini; in formule è data da:

$$\boxed{M_h = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}}}$$

La media armonica, dunque, è quel valore tale che il suo reciproco, sostituito ai dati, che devono essere tutti positivi, fa rimanere invariata la somma dei reciproci dei dati stessi: viene usata, infatti, per mediare rapporti di tempo.

INDICI STATISTICI

Esempio

La media armonica dei seguenti 5 = n numeri:

$$x_1 = 10; \quad x_2 = 13; \quad x_3 = 9; \quad x_4 = 7; \quad x_5 = 12$$

è data da:

$$M_h = \frac{5}{\sum_{i=1}^{5} \frac{1}{x_i}} = \frac{5}{\frac{1}{10} + \frac{1}{13} + \frac{1}{9} + \frac{1}{7} + \frac{1}{12}} =$$

$$= \frac{5}{\frac{1638 + 1260 + 1820 + 2340 + 1365}{16380}} = \frac{5}{\frac{8423}{16380}} = 5 \cdot \frac{16380}{8423} \approx 9,72$$

INDICI STATISTICI

La ***media armonica ponderata***, invece, è data da:

$$\boxed{M_h = \frac{n}{\sum_{i=1}^{s} \frac{n_i}{x_i}} = \frac{n}{\frac{n_1}{x_1} + \frac{n_2}{x_2} + \ldots + \frac{n_s}{x_s}}}$$

dove $n = n_1 + n_2 + \ldots + n_s$.
La media armonica, dunque, è pari al valore reciproco della media aritmetica dei reciproci dei termini.

INDICI STATISTICI

Esempio

Si consideri la seguente tabella la seguente tabella:

Voti x_i	Numeri di studenti n_i
20	2
21	3
22	6
23	2
24	1
Tot.	14

INDICI STATISTICI

Ne segue, allora, che la media armonica ponderata è data da:

$$M_h = \frac{n}{\sum_{i=1}^{s} \frac{n_i}{x_i}} = \frac{14}{\frac{2}{20} + \frac{3}{21} + \frac{6}{22} + \frac{2}{23} + \frac{1}{24}} = 22$$

INDICI STATISTICI

La ***media quadratica semplice*** si ottiene estraendo la radice quadrata della media aritmetica dei quadrati degli *n* termini; in formule è data da:

$$\boxed{M_2\left(x_1, x_2, \ldots, x_n\right) = \sqrt[2]{\frac{1}{n}\sum_{i=1}^{n} x_i^2} = \sqrt[2]{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n}}}$$

Tale media, denominata anche *media di precisione*, usata tutte le volte che alle differenze tra i termini ed il valore medio si dà il significato di *deviazione* o *errore del valore esatto*, ovvero nei casi in cui alcuni termini considerati risultano negativi e si desidera quindi eliminare la loro influenza, trova applicazione soprattutto nell'ambito della teoria degli errori.
Generalizzando ora il concetto di media quadratica, si può definire la cosiddetta ***media di potenza di indice t*** data da:

$$\boxed{M_t\left(x_1, x_2, \ldots, x_n\right) = \sqrt[t]{\frac{1}{n}\sum_{i=1}^{n} x_i^t} = \sqrt[t]{\frac{x_1^t + x_2^t + \ldots + x_n^t}{n}}}$$

INDICI STATISTICI

Esempio

La media quadratica dei seguenti 10 = *n* numeri:

$$x_1 = 1;\ x_2 = 1;\ x_3 = 2;\ x_4 = 2;\ x_5 = 3;\ x_6 = 3;\ x_7 = 4;\ x_8 = 4;\ x_9 = 5;\ x_{10} = 5$$

è data da:

$$M_2\left(x_1, x_2, \ldots, x_{10}\right) = \sqrt{\frac{1}{10}\sum_{i=1}^{10} x_i^2} =$$

$$= \sqrt{\frac{1}{10}\left(1^2 + 1^2 + 2^2 + 2^2 + 3^2 + 3^2 + 4^2 + 4^2 + 5^2 + 5^2\right)} =$$

$$= \sqrt{\frac{1}{10}\left(1 + 1 + 4 + 4 + 9 + 9 + 16 + 16 + 25 + 25\right)} =$$

$$= \sqrt{\frac{1}{10}\left(1 + 1 + 4 + 4 + 9 + 9 + 16 + 16 + 25 + 25\right)} =$$

$$= \sqrt{\frac{110}{10}} = \sqrt{11} \approx 3{,}31$$

INDICI STATISTICI

La ***media quadratica ponderata***, invece, è data da:

$$M_2(x_1, x_2, ..., x_n) = \sqrt[2]{\frac{1}{n}\sum_{i=1}^{s} x_i^2 n_i} = \sqrt[2]{\frac{x_1^2 \cdot n_1 + x_2^2 \cdot n_2 + ... + x_n^2 \cdot n_s}{n}}$$

dove *n* è sempre la somma delle n_i. La precedente espressione, generalizzata alle potenze di indice *t*, diventa:

$$M_t(x_1, x_2, ..., x_n) = \sqrt[t]{\frac{1}{n}\sum_{i=1}^{s} x_i^t n_i} = \sqrt[t]{\frac{x_1^t \cdot n_1 + x_2^t \cdot n_2 + ... + x_n^t \cdot n_s}{n}}$$

dove *n* è sempre la somma delle n_i.