La media geometrica

- Per una *distribuzione unitaria* di un carattere *quantitativo* di *n* termini, la **media geometrica** è definita come:

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \sqrt[n]{\prod_{1=1}^{n} x_i}$$

- Viene usata per sintetizzare dati che ha senso moltiplicare fra loro o per riassumere distribuzioni che hanno andamento geometrico

- Si applica per determinare un tasso di incremento / decremento medio (prezzi dei prodotti, andamento della popolazione, ecc)

Esempio

- In un determinato punto vendita si è osservato:

Anni	2005	2006	2007	2008
Vendite (milioni di euro)	207	189	246	298

Si vuole calcolare la variazione media nelle vendite

- Bisogna innanzitutto calcolare le variazioni annue:

Anni	2005	2006	2007	2008
Variazioni $x_i = \frac{V_i}{V_{i-1}}$	--	0,913	1,302	1,21

$$\bar{x}_g = \sqrt[3]{0{,}913 \cdot 1{,}302 \cdot 1{,}211} = 1{,}129$$

- La media aritmetica sarebbe stata invece:

$$\bar{x}_a = (0{,}913 + 1{,}302 + 1{,}211)/3 = 1{,}142$$

- Perchè la media geometrica non sarebbe stata appropriata?

Supponiamo che V_0 siano le vendite iniziali. Applicando le variazioni x_1, x_2, x_3 otteniamo:

$V_1 = V_0 \cdot x_1$
$V_2 = V_1 \cdot x_2 = V_0 \cdot x_1 \cdot x_2$
$V_3 = V_2 \cdot x_3 = V_0 \cdot x_1 \cdot x_2 \cdot x_3$

La variazione media è quella variazione costante, che applicata di anno in anno, deve restituire il valore corretto delle vendite a fine periodo, dato il valore iniziale.

$$V_3 = V_0 \cdot \bar{x} \cdot \bar{x} \cdot \bar{x} = V_0 \cdot \bar{x}^3$$

Sostituendo la media aritmetica e la media geometrica si ha:

$$V_0 \cdot \bar{x}_a^3 = 207 \cdot 1{,}142^3 = 308{,}03 \qquad V_0 \cdot \bar{x}_g^3 = 207 \cdot 1{,}129^3 = 298$$

- La media geometica può essere calcolata direttamente per una *distribuzione di frequenza* non in classi tramite la formula

$$\bar{x} = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \ldots \cdot x_k^{n_k}} = \sqrt[n]{\prod_{j=1}^{k} x_j^{n_j}}$$

che tiene conto del fatto che una modalità può ripetersi più volte

- Utilizzando le frequenze relative si ha:

$$\bar{x} = x_1^{f_1} \cdot x_2^{f_2} \cdot \ldots \cdot x_k^{f_k} = \prod_{j=1}^{k} x_j^{f_j}$$

- Se il carattere è in classi, si utilizzano i valori centrali

$$x_j = \frac{1}{2}(c_{j-1} + c_j)$$

al posto delle modalità.

Proprietà della media geometica

- ***Proprietà 1*** *(consistenza)*: Se la distribuzione è costituita da *n* termine tutti pari ad *a*, la media della distribuzione sarà anch'essa pari ad *a*:

$$\bar{x}_g = \sqrt[n]{\prod_{i=1}^{n} a} = a$$

- ***Proprietà 2*** *(monotonia)*: Date due distribuzioni unitarie con n termini, rispettivamente, $x_1, x_2, .., x_n$ e $y_1, y_2, .., y_n$, se vale la condizione $x_j \leq y_j$ per ogni *j,* e almeno una volta $x_j < y_j$, allora

$$\bar{x}_g < \bar{y}_g$$

- ***Proprietà 3*** *(di internalità)*: La media geometrica è sempre compresa tra il minimo e il massimo della distribuzione

$$x_1 \leq \bar{x}_g \leq x_k$$

- ***Proprietà 4*** (*invarianza rispetto a cambiamenti di scala*): se a ogni termine della distribuzione viene applicata la trasformazione *aX*, allora la media geometrica sarà pari a

$$a\bar{x}_g$$

- ***Proprietà 5***: La media geometrica non è mai superiore alla media aritmetica

$$\bar{x}_g \leq \bar{x}_a$$

per qualsiasi distribuzione

- ***Proprietà 6***: Il logaritmo della media geometrica è uguale alla media aritmetica dei logaritmi. Quindi, ad esempio, la media geometrica può essere calcolata come

$$\bar{x}_g = \exp\left[\frac{1}{n}\sum_{i=1}^{n}\log(x_i)\right] \qquad \text{(distribuzione unitaria)}$$

$$\bar{x}_g = \exp\left[\frac{1}{n}\sum_{j=1}^{k} n_j \log(x_j)\right] \qquad \text{(distribuzione di frequenze)}$$

Esempio

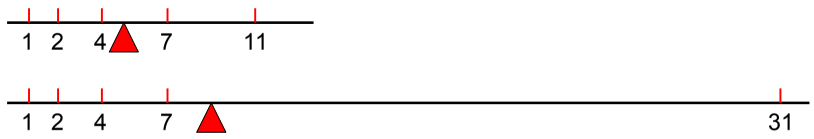
Anni	2005	2006	2007	2008
Variazioni $x_i = \frac{V_i}{V_{i-1}}$	--	0,913	1,302	1,21
Log(Variazioni)	--	-0,091	0,264	0,191

$$\bar{x}_g = \exp\left[\frac{1}{n}\sum_{i=1}^{n}\log(x_i)\right] = \exp\left[\frac{1}{3}(-0{,}091 + 0{,}264 + 0{,}191)\right] =$$
$$= \exp(0{,}121) = 1{,}129$$

La media troncata

- La media troncata al 50% è la media aritmetica calcolata sul 50% dei valori centrali della distribuzione.
- L'obiettivo della media troncata è eliminare l'effetto dei valori anomali sulla media aritmetica.

Media aritmetica



Media aritmetica troncata al 60%

La mediana

- Data una distribuzione secondo un carattere *qualitativo ordinato* o *quantitativo*, la **mediana** (M_e) è la modalità del carattere che divide il collettivo in due gruppi di uguale numerosità in modo tale che:
  1. le unità del primo gruppo hanno una modalità $\leq M_e$;
  2. le unità del secondo gruppo hanno una modalità $\geq M_e$
- Per calcolare la mediana di una *distribuzione unitaria* di un carattere quantitativo di n termini
  1. si ordinano le modalità in modo non decrescente:

$$x_1 \leq x_2 \leq \ldots \leq x_n$$

  2. se *n* è dispari $\rightarrow \quad M_e = x_{(n+1)/2}$

     se *n* è pari $\rightarrow \quad M_e = \left(x_{n/2} + x_{n/2+1}\right)/2$

(solo per caratteri quantitativi)

Esempio

- Si consideri la distribuzione dei voti di 5 studenti (27, 24, 30, 22, 28); la distribuzione ordinata è (22, 24, 27, 28, 30) da cui $M_e = x_{(5+1)/2} = x_3 = 27$.
- Se la distribuzione fosse stata (22, 24, 27, 28), avremmo avuto $M_e = (x_{4/2} + x_{4/2+1})/2 = (x_2 + x_3)/2 = (24+27)/2 = 25{,}5$.

- Per una *distribuzione di frequenza non in classi*, la mediana può essere calcolata sulla base delle frequenze cumulate (N_j) come:
  1. si individua la modalità x_j tale che: $N_{j-1} \leq n/2 < N_j$
  2. se $N_{j-1} < n/2 \quad \rightarrow \quad M_e = x_j$

     se $N_{j-1} = n/2 \quad \rightarrow \quad M_e = \left(x_{j-1} + x_j\right)/2$

(solo per caratteri quantitativi)

Esempio

NUM. COMP.	Num. famiglie	Freq. ass. cum.
1	2.883.250	2.883.250
2	3.048.249	5.931.499
3	2.352.645	8.284.144
4	1.667.391	9.951.535
5	391.376	10.342.911
6 o più	106.299	10.449.210
Totale	**10.449.210**	

Dato che $n/2$ = 10.449.210/2 = 5.224.605, con j = 2 si ha $N_{j-1} < n/2 < N_j$ da cui M_e = 2.

Esempio

NUM. COMP.	Num. famiglie	Freq. ass. cum.
1	1.803.250	1.803.250
2	2.714.299	4.517.549
3	2.352.640	6.870.189
4	1.667.271	8.537.460
5	391.348	8.928.808
6 o più	106.290	9.035.098
Totale	**9.035.098**	

Dato che $n/2$ = 9.035.098/2 = 4.517.549, con j = 3 si ha $N_{j-1} = n/2 < N_j$ da cui M_e = 2,5.

- Se il carattere è in *classi*:

  1. si individua la classe mediana, $c_{j-1} - c_j$, tale che:

$$N_{j-1} \leq n/2 < N_j$$

  2. sulla base dell'*ipotesi di uniforme distribuzione*

$$M_e = c_{j-1} + \frac{n/2 - N_{j-1}}{n_j} \cdot a_j$$

Esempio

Classi di fatturato ($c_{j-1} - c_j$)	**Frequenze assolute (n_j)**	**Frequenze assolute cumulate (N_j)**
0--250	2.156	2.156
250--500	666	2.822
500--1.000	517	3.339
1.000--2.500	539	3.878
2.500--5.000	260	4.138
5.000--10.000	171	4.309
10.000--25.000	95	4.404
25.000--50.000	32	4.436
50.000--100.000	35	4.471
Totale	**4.471**	

Dato che n/2 = 2.235,5, la classe mediana è 250—500 da cui

$$M_e = c_{j-1} + \frac{n/2 - N_{j-1}}{n_j} \cdot a_j = 250 + \frac{2.235{,}5 - 2.156}{666} \cdot 250 = 279{,}84$$

Proprietà della mediana

- ***Proprietà 1*** *(consistenza)*: Se la distribuzione è costituita da *n* termine tutti pari ad *a*, la mediana della distribuzione sarà anch'essa pari ad *a*.

- ***Proprietà 2*** *(monotonia in senso debole)*: Date due distribuzioni unitarie con n termini, rispettivamente, $x_1, x_2, .., x_n$ e $y_1, y_2, .., y_n$, se vale la condizione $x_j \leq y_j$ per ogni *j,* e almeno una volta $x_j < y_j$, allora

$$Me_x \leq Me_y$$

- ***Proprietà 3***: La mediana minimizza la distanza di ogni modalità da una costante

$$\sum_{i=1}^{n} |x_i - c| \quad \text{è minimo per} \quad c = M_e$$

- ***Proprietà 4*** *(di internalità)*: La mediana è sempre compresa tra il minimo e il massimo della distribuzione

$$x_1 \leq M_e \leq x_k$$

- ***Proprietà 5*** (*invarianza rispetto a trasformazioni lineari*): se a ogni termine della distribuzione viene applicata la trasformazione *aX + b*, allora la mediana sarà pari a

$$aM_e + b$$

I percentili

- I percentili sono quei valori che dividono la distribuzione in 100 parti uguali.

- I percentili più usati sono il 25° (Q_1), il 50° (M_e) e il 75° (Q_3).

- Ad esempio se il carattere è in *classi,* per calcolare Q_1:

  1. si individua la classe che contiene Q_1, $c_{j-1} - c_j$, tale che:

$$N_{j-1} \leq n/4 < N_j$$

  2. sulla base dell'*ipotesi di uniforme distribuzione*

$$Q_1 = c_{j-1} + \frac{n/4 - N_{j-1}}{n_j} \cdot a_j$$

Esempio

Classi di fatturato ($c_{j-1} - c_j$)	**Frequenze assolute (n_j)**	**Frequenze assolute cumulate (N_j)**
0--250	2156	2156
250--500	666	2822
500--1.000	517	3339
1.000--2.500	539	3878
2.500--5.000	260	4138
5.000--10.000	171	4309
10.000--25.000	95	4404
25.000--50.000	32	4436
50.000--100.000	35	4471
Totale	**4471**	

Dato che n/4 = 1117,75, la classe contenente Q_1 è 0—250 da cui

$$Q_1 = c_{j-1} + \frac{n/4 - N_{j-1}}{n_j} \cdot a_j = 0 + \frac{1117{,}75 - 0}{2156} \cdot 250 = 129{,}61$$

Moda e classe modale

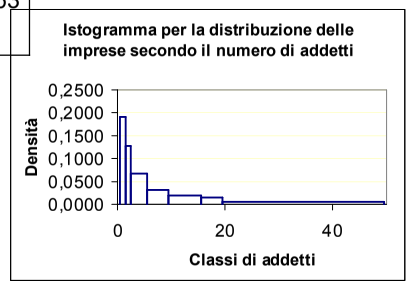
- Nel caso di una distribuzione di un carattere non in classi, la **moda** è definita come modalità del carattere che si presenta con la maggiore frequenza.
- Se il carattere è in classi, la classe modale è la classe che presenta la maggiore densità di frequenza.
- In alcuni casi la moda (o classe modale) può non essere unica; si ha quindi una **distribuzione plurimodale**

Esempio

NUM. COMP.	Num. famiglie
1	2.883.250
2	3.048.249
3	2.352.645
4	1.667.391
5	391.376
6 o più	106.299
Totale	**10.449.210**

La moda è pari a 2.

Esempio

Classi di addetti	**Frequenze relative (f_j)**	**Ampiezza classi (a_j)**	**Densità di frequenza (h_j)**
0,5--1,5	0,1915	1	0,19155
1,5--2,5	0,1280	1	0,12799
2,5--5,5	0,2037	3	0,06792
5,5--9,5	0,1288	4	0,03221
9,5--15,5	0,1180	6	0,01966
15,5--19,5	0,0610	4	0,01525
19,5--49,5	0,1689	30	0,00563
Totale:	**1**		

La moda è pari a 1 addetto.