Caso studio 4

Un investitore deve decidere se investire il suo capitale di 10.000 euro in obbligazioni a tasso fisso o a tasso variabile. Il tasso fisso di interesse che gli viene proposto è del 4% annuo, per un investimento a 5 anni. Per le obbligazioni a tasso variabile gli viene invece proposto il seguente prospetto:

Anno	Tasso di interesse %
I	1,5
II	2,0
III	7,2
IV	7,4
V	9,0

Cosa gli conviene scegliere?

1

La media geometrica

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

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

- 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)

2

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:

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

3

- Perchè la media aritmetica 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:

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

4

Esercizio

Si considerino i dati del Caso Studio 4.

Si calcoli il tasso di incremento medio che, applicato a ciascun periodo, restituisca lo stesso valore di capitale al termine dell'investimento.

Si confronti il tasso di incremento medio con il tasso fisso e si stabilisca quali obbligazioni convenga acquistare.

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

$$\bar{x}_g = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \dots \cdot x_k^{n_k}} = \sqrt[n]{\prod_{i=1}^{k} x_i^{n_i}}$$

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

- Utilizzando le frequenze relative si ha:

$$\bar{x}_g = x_1^{f_1} \cdot x_2^{f_2} \cdot \dots \cdot x_k^{f_k} = \prod_{i=1}^{k} x_i^{f_i}$$

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

$$x_i = \frac{1}{2}(c_{i-1} + c_i)$$

al posto delle modalità.

Proprietà della media geometrica

- ***Proprietà 1*** *(consistenza)*: Se la distribuzione è costituita da *n* termine tutti pari ad *a*, la media geometrica 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 $x_i \leq y_i$ per ogni *i,* e almeno una volta $x_i < y_i$, 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

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

$$\bar{x}_g = \exp\left[\frac{1}{n}\sum_{i=1}^{k}n_i\log(x_i)\right]$$ (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

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

$$= \exp(0{,}121) = 1{,}129$$

Esercizio

Si considerino i dati del Caso Studio 4.

Si verifichi,

a) che la media geometrica gode della proprietà di internalità (proprietà 3),

b) che la media geometrica non è mai superiore alla media aritmetica (proprietà 5),

c) che il logaritmo della media geometrica è uguale alla media aritmetica dei logaritmi (proprietà 6).

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%

Caso studio 5

Famiglie e redditi (compresi fitti imputati) per classi di reddito (in migliaia) - Umbria 2004

Classi di reddito	% Famiglie	% Reddito	Reddito medio
Fino a 10	5,1	0,9	6.072
10-20	25,3	11,7	15.423
20-30	24,8	18,7	25.074
30-40	16,4	17,1	34.787
40-50	12,0	16,3	45.172
50-60	7,4	12,3	55.303
60-70	3,4	6,5	64.074
70 e oltre	5,6	16,5	97.338
Totale	**100**	**100**	**33.303**

Il reddito medio (33.303 euro) è un indicatore adatto a sintetizzare l'intera distribuzione? Osservando i dati si vede che in realtà più della metà della popolazione ha un reddito inferiore a 30.000 euro. Quale può essere un indicatore più appropriato? Da cosa dipende un valore così elevato del reddito medio?

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:

$$x_1 \leq x_2 \cdots \leq x_n$$

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

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

(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_i) o delle frequenze relative cumulate (F_i) come:

1. si individua la modalità x_i tale che:

$$N_{i-1} \leq n/2 < N_i \quad \text{oppure} \quad F_{i-1} \leq 1/2 < F_i$$

2. se $\begin{cases} N_{i-1} < n/2 \\ F_{i-1} < 1/2 \end{cases} \rightarrow M_e = x_i$

se $\begin{cases} N_{i-1} = n/2 \\ F_{i-1} = 1/2 \end{cases} \rightarrow M_e = (x_{i-1} + x_i)/2$

(solo per caratteri quantitativi)

Esempio

NUM. COMP.	Num. famiglie	Freq. ass. cum.	Freq. rel. cum.
1	27	27	0,27
2	30	57	0,57
3	22	79	0,79
4	17	96	0,96
5	3	99	0,99
6 o più	1	100	1,00
Totale	**100**		

Dato che $n/2$ = 100/2 = 50, con i = 2 si ha
$N_{i-1} < n/2 < N_i$ da cui $M_e = 2$.

Considerando invece le frequenze relative cumulate, con i = 2 si ha $F_{i-1} < 1/2 < F_i$ da cui $M_e = 2$.

Esempio

NUM. COMP.	Num. famiglie	Freq. ass. cum.	Freq. rel. cum.
1	24	24	0,24
2	26	50	0,50
3	29	79	0,79
4	17	96	0,96
5	3	99	0,99
6 o più	1	100	1,00
Totale	**100**		

Dato che $n/2$ = 100/2 = 50, con i = 3 si ha
$N_{i-1} = n/2 < N_i$ da cui $M_e = 2,5$.

Utilizzando le frequenze relative cumulate, con i = 3 si ha
$F_{i-1} = 1/2 < F_i$ da cui $M_e = 2,5$.

Esempio

NUM. COMP.	Num. famiglie	Freq. ass. cum.	Freq. rel. cum.
1	27	27	0,267
2	30	57	0,564
3	22	79	0,782
4	17	96	0,950
5	3	99	0,980
6 o più	2	101	1,000
Totale	**101**		

Dato che $n/2$ = 101/2 = 50,5, con i = 2 si ha
$N_{i-1} < n/2 < N_i$ da cui $M_e = 2$.

Considerando invece le frequenze relative cumulate, con i = 2 si ha $F_{i-1} < 1/2 < F_i$ da cui $M_e = 2$.

Esempio

NUM. COMP.	Num. famiglie	Freq. ass. cum.	Freq. rel. cum.
1	24	24	0,238
2	26	50	0,495
3	29	79	0,782
4	17	96	0,950
5	3	99	0,980
6 o più	2	101	1,000
Totale	**101**		

Dato che $n/2$ = 100/2 = 50,5, con i = 3 si ha
$N_{i-1} < n/2 < N_i$ da cui $M_e = 3$.

Utilizzando le frequenze relative cumulate, con i = 3 si ha
$F_{i-1} < 1/2 < F_i$ da cui $M_e = 3$.

Esempio

NUM. COMP.	Num. famiglie	Freq. ass. cum.	Freq. rel. cum.
1	2.883.250	2.883.250	0,276
2	3.048.249	5.931.499	0,568
3	2.352.645	8.284.144	0,793
4	1.667.391	9.951.535	0,952
5	391.376	10.342.911	0,990
6 o più	106.299	10.449.210	1,000
Totale	**10.449.210**		

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

Considerando invece le frequenze relative cumulate, con i = 2 si ha $F_{i-1} < 1/2 < F_i$ da cui M_e = 2.

21

Esempio

NUM. COMP.	Num. famiglie	Freq. ass. cum.	Freq. rel. cum.
1	1.803.250	1.803.250	0,200
2	2.714.299	4.517.549	0,500
3	2.352.640	6.870.189	0,760
4	1.667.271	8.537.460	0,945
5	391.348	8.928.808	0,988
6 o più	106.290	9.035.098	1,000
Totale	**9.035.098**		

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

Utilizzando le frequenze relative cumulate, con i = 3 si ha $F_{i-1} = 1/2 < F_i$ da cui M_e = 2,5.

22

- Se il carattere è in *classi*:
    1. si individua la classe mediana, $c_{i-1} – c_i$, tale che:

$$N_{i-1} \le n/2 < N_i \quad \text{oppure} \quad F_{i-1} \le 1/2 < F_i$$

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

$$M_e = c_{i-1} + \frac{n/2 - N_{i-1}}{n_i} \cdot a_i$$

$$M_e = c_{i-1} + \frac{1/2 - F_{i-1}}{f_i} \cdot a_i$$

23

Esempio

Classi	n_i	N_i	f_i	F_i
0--250	2.156	2.156	0,482	0,482
250--500	666	2.822	0,149	0,631
500--1.000	517	3.339	0,116	0,747
1.000--2.500	539	3.878	0,121	0,867
2.500--5.000	260	4.138	0,058	0,926
5.000--10.000	171	4.309	0,038	0,964
10.000--25.000	95	4.404	0,021	0,985
25.000--50.000	32	4.436	0,007	0,992
50.000--100.000	35	4.471	0,008	1,000
Totale	**4.471**		**1,000**	

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

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

$$M_e = c_{i-1} + \frac{1/2 - F_{i-1}}{f_i} \cdot a_i = 250 + \frac{0,5 - 0,482}{0,149} \cdot 250 = 279,84$$

24

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 $x_i \leq y_i$ per ogni *i*, e almeno una volta $x_i < y_i$, 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, $c_{i-1} - c_i$, tale che:

  $$N_{i-1} \leq n/4 < N_i \quad \text{oppure} \quad F_{i-1} \leq 1/4 < F_i$$

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

$$Q_1 = c_{i-1} + \frac{n/4 - N_{i-1}}{n_i} \cdot a_i \quad \text{oppure} \quad Q_1 = c_{i-1} + \frac{1/4 - F_{i-1}}{f_i} \cdot a_i$$

Esempio

Classi di fatturato	***n**$_i$*	***N**$_i$*	***f**$_i$*	***F**$_i$*
0--250	2156	2156	0,482	0,482
250--500	666	2822	0,149	0,631
500--1.000	517	3339	0,116	0,747
1.000--2.500	539	3878	0,121	0,867
2.500--5.000	260	4138	0,058	0,926
5.000--10.000	171	4309	0,038	0,964
10.000--25.000	95	4404	0,021	0,985
25.000--50.000	32	4436	0,007	0,992
50.000--100.000	35	4471	0,008	1,000
Totale	**4471**		1,000	

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

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

$$Q_1 = c_{i-1} + \frac{1/4 - F_{i-1}}{f_i} \cdot a_i = 0 + \frac{0{,}25 - 0}{0{,}482} \cdot 250 = 129{,}61$$

Esercizio

Si considerino i dati del Caso Studio 5.

Si calcoli il reddito mediano, ossia quel valore del reddito al di sotto del quale troviamo il 50% delle famiglie umbre.

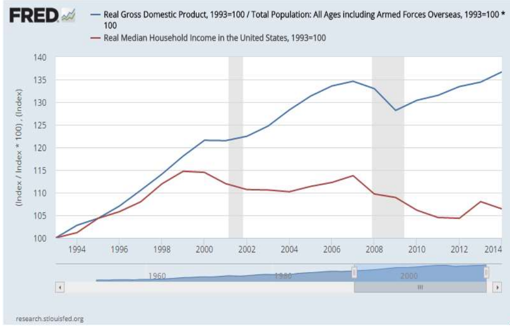
Si confronti tale valore con quello del reddito medio e si spieghino le eventuali differenze.

Si determini il valore del reddito al di sotto del quale troviamo il quarto più povero delle famiglie.

Si determini il valore del reddito al di sopra del quale troviamo il quarto più ricco delle famiglie.

Esempio

Il grafico mostra l'andamento del reddito medio e del reddito mediano negli Stati Uniti dal 1994 ad oggi. Negli ultimi anni le due curve tendono a divergere in maniera significativa. Perché?



Facciamo un esempio semplice: 10 cassieri di banca stanno bevendo una birra al bar. Ognuno di loro guadagna 30.000 dollari l'anno, pertanto il reddito medio e il reddito mediano del gruppo è di 30.000 dollari. Entra l'amministratore delegato e ordina una birra. Adesso il reddito medio di questo gruppo schizza alle stelle ma quello mediano non cambia minimamente.

In genere, più sono disuguali i redditi, più la media tende a divergere dalla mediana. E' quanto è successo non solo in questo bar ipotetico ma anche in tutti gli Stati Uniti.

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

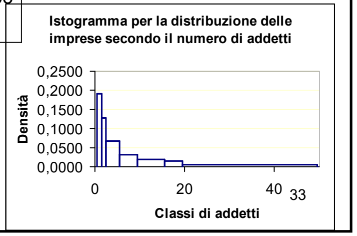
NUM. COMP.	Num. famiglie
1	2.883.250
2	3.048.249
3	2.352.645
4	1.667.391
5	391.376
6 o più	106.299
Totale	**10.449.210**

La moda è pari a 2 componenti.

Esempio

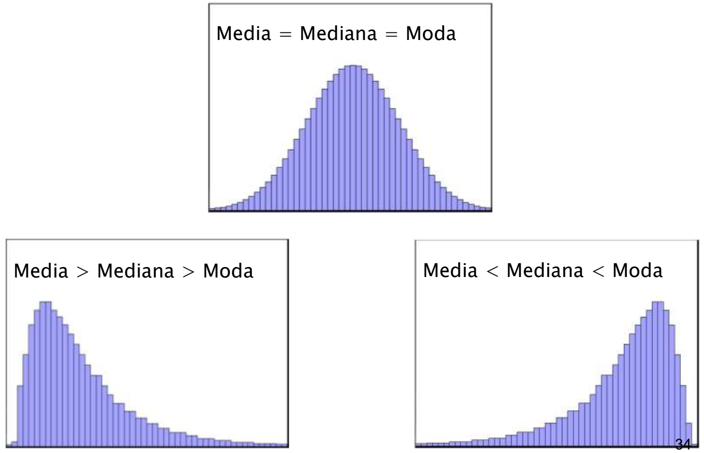
Classi di addetti	Frequenze relative (f_i)	Ampiezza classi (a_i)	Densità di frequenza (h_i)
0,5--1,5	0,1915	1	0,19155
1,5--2,5	0,1280	1	0,12799
2,5--5,5	0,2037	3	0,06792
5,5--9,5	0,1288	4	0,03221
9,5--15,5	0,1180	6	0,01966
15,5--19,5	0,0610	4	0,01525
19,5--49,5	0,1689	30	0,00563
Totale:	**1**		

La moda è pari a 1 addetto.



Esempio

Media, moda e mediana per distribuzioni con varie forme



Esercizio

Si considerino i dati del Caso Studio 5.

Si disegni l'istogramma e si evidenzino sull'istogramma la media, la mediana e la moda della distribuzione del reddito.

Esercizio

Per un insieme di 5 numeri interi, si ha che la media è 4, la moda è 1, la mediana è 5. Quali sono questi 5 numeri?

Dove e come studiare

- Libro di testo: S. Borra, A. Di Ciaccio (2014), Cap. 3
- Svolgere 'Esercitazione 2'
- Svolgere i punti non precedentemente svolti degli esercizi nel file 'Esercizi su medie.xls'. Escludere il calcolo della deviazione standard (Foglio 2, seconda parte del punto b, Foglio 4, seconda parte del punto d)