Caso studio 2

Un investitore sta valutando i rendimenti di due titoli del settore Petrolio e Gas naturale. Sulla base dei rendimenti giornalieri della settimana passata vuole cercare di prevedere il rendimento per la prossima settimana e quale dei due titoli avrà il rendimento più alto. I dati sono riportati di seguito:

Data	ERG	Gas Plus
01/02/2015	-0,17	2,22
02/02/2015	1,22	-2,48
03/02/2015	-2,92	0,63
04/02/2015	0,53	-4,10
05/02/2015	-0,88	1,64

Le medie

- **Medie**: permettono di sintetizzare una distribuzione sulla base di un solo valore. Possono essere classificate in:
  - ***Medie analitiche***: calcolate tramite operazioni algebriche sui valori del carattere ➡ solo per *caratteri quantitativi*
  - ***Medie di posizione***: non richiedono operazioni algebriche ➡ anche per *caratteri qualitativi*

Medie analitiche	*Medie di posizione*
Media aritmetica	Mediana
Media geometrica	Moda
Media troncata	Percentili

La media aritmetica

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

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Esempio

- Distribuzione del voto in Statistica per un gruppo di 6 studenti:

Unità statistica (*i*)	Voto (x_i)
1	27
2	21
3	28
4	30
5	21
6	27
Totale	**154**

$$\bar{x} = \frac{1}{6} \cdot 154 = 25{,}667$$

Esercizio

Si considerino i dati del Caso Studio 2.

Si calcoli il rendimento medio dei due titoli (come previsione del rendimento della prossima settimana e per verificare quale titolo ha avuto in media il rendimento migliore).

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

$$\bar{x} = \frac{1}{n}\left(x_1 n_1 + x_2 n_2 + \ldots + x_k n_k\right) = \frac{1}{n}\sum_{i=1}^{k} x_i n_i$$

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

- Per il calcolo si imposta una tabella del tipo:

Modalità (x_i)	**Frequenze (n_i)**	**$x_i n_i$**
x_1	n_1	$x_1 n_1$
x_2	n_2	$x_2 n_2$
…	…	…
x_k	n_k	$x_k n_k$
Totale	n	$\Sigma_i x_i n_i$

Esempio

Voto (x_j)	**Frequenza (n_j)**	**$x_j n_j$**
21	2	42
27	2	54
28	1	28
30	1	30
Totale	**6**	**154**

$$\bar{x} = \frac{1}{6} \cdot 154 = 25{,}667$$

- Alternativamente si può utilizzare la formula

$$\bar{x} = x_1 f_1 + x_2 f_2 + \ldots + x_k f_k = \sum_{i=1}^{k} x_i f_i$$

basata sulle frequenze relative.

- La tabella seguente fornisce direttamente il valore della media aritmetica.

Modalità (x_i)	**Frequenze (n_i)**	**Freq. relative (f_i)**	**$x_i f_i$**
x_1	n_1	f_1	$x_1 f_1$
x_2	n_2	f_2	$x_2 f_2$
…	…	…	…
x_k	n_k	f_k	$x_k f_k$
Totale	n	1	$\Sigma_i x_i f_i$

- Il risultato può risentire di approssimazioni nel calcolo delle frequenze relative

Esempio

Voto (x_i)	**Frequenza (n_i)**	**Freq. relativa (f_i)**	**$x_i f_i$**
21	2	0,333	6,993
27	2	0,333	8,991
28	1	0,167	4,676
30	1	0,167	5,010
Totale	**6**	**1**	**25,670**

$$\bar{x} = 25{,}670 \approx 25{,}667$$

- Quando il carattere è in classi, si ha una perdita di informazione in quanto non si conosce più con esattezza la modalità di ogni unità statistica. La media sarebbe calcolabile esattamente conoscendo le medie di classe ($\bar{x}_i$) come:

$$\bar{x} = \frac{1}{n}\left(\bar{x}_1 n_1 + \bar{x}_2 n_2 + \ldots + \bar{x}_k n_k\right) = \frac{1}{n}\sum_{i=1}^{k} \bar{x}_i n_i$$

- Non conoscendo le medie di classe, si usano invece i **valori centrali**

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

che corrispondono alle medie di classe sotto l'ipotesi di uniforme distribuzione. La media aritmetica viene quindi calcolata come

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i \qquad \text{oppure} \qquad \bar{x} = \sum_{i=1}^{k} x_i f_i$$

- Il risultato è solo approssimato, a causa della perdita di informazione.

- Per il calcolo si imposta una tabella del tipo (quando si applica la formula basata sulle frequenze assolute)

Classi (c_{i-1} - c_i)	**Frequenze (n_i)**	**Valori centrali (x_i)**	**$x_i n_i$**
c_0 - c_1	n_1	x_1	$x_1 n_1$
$c_1 - c_2$	n_2	x_2	$x_2 n_2$
…	…	…	…
c_{k-1} - c_k	n_k	x_k	$x_k n_k$
Totale	n	-	$\Sigma_i x_i n_i$

- Oppure una tabella del tipo (quando si applica la formula basata sulle frequenze relative)

Classi (c_{i-1} - c_i)	**Frequenze (n_i)**	**Freq. relative (f_i)**	**Valori centrali (x_i)**	**$x_i f_i$**
c_0 - c_1	n_1	f_1	x_1	$x_1 f_1$
$c_1 - c_2$	n_2	f_2	x_2	$x_2 f_2$
…	…	…	…	…
c_{k-1} - c_k	n_k	f_k	x_k	$x_k f_k$
Totale	n	1	-	$\Sigma_i x_i f_i$

Esempio

- Per la distribuzione dell'altezza per un collettivo di 50 persone:

Classi (c_{i-1} - c_i)	**Frequenze (n_i)**	**Valori centrali (x_i)**	**$x_i n_i$**	**Freq. relative (f_i)**	**$x_i f_i$**
150 – 160	1	155	155	0,02	3,1
160 – 170	10	165	1650	0,20	33,0
170 – 180	35	175	6125	0,70	122,5
180 - 200	4	190	760	0,08	15,2
Totale	50	-	8690	1	173,8

$$\bar{x} = 8690 / 50 = 173{,}8$$

Caso studio 3

Un investitore vuole confrontare il rendimento di due portafogli di titoli diversi. Avendo a disposizione, per ogni portafoglio, i titoli contenuti, il peso con cui sono presenti nel portafoglio e il rendimento annuo di ciascun titolo, l'investitore vuole calcolare qual è stato il rendimento annuo di ciascun portafoglio. I dati sono riportati di seguito:

Titoli	Peso	Rendimento
A	0,20	2,10
B	0,30	1,50
C	0,10	0,50
D	0,40	0,10

Titoli	Peso	Rendimento
E	0,20	1,10
F	0,30	1,00
G	0,30	1,50
H	0,20	2,00

- In alcune situazioni si ha una ***distribuzione ponderata*** in cui a ogni modalità viene associato un peso che ne quantifica l'importanza

Unità (*i*)	**Modalità (x_i)**	**Peso (w_i)**
1	x_1	w_1
2	x_2	w_2
…	…	…
n	x_n	w_n

- Per calcolare la media aritmetica, il peso viene trattato come una frequenza

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}$$

Esempio

- Voti di uno studente del primo anno di Economia:

Esame (*i*)	**Voto (x_i)**	**CFU (w_i)**	**$x_i w_i$**
Microeconomia	28	12	336
Matematica	25	12	300
Storia Economica	30	6	180
Informatica	30	6	180
Inglese	28	6	168
Ec. aziendale	27	9	243
Ist. Dir. Pub.	30	9	270
Totale	-	60	1677

$$\bar{x} = 1677 / 60 = 27{,}95$$

Esercizio

Si considerino i dati del Caso Studio 3.

Si calcoli il rendimento medio dei due portafogli di titoli.

Proprietà della media aritmetica

- ***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*:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} a = a$$

- ***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

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

- ***Proprietà 3***: La somma algebrica degli scarti dalla media è nulla:

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

Esempio

Unità statistica (i)	Voto in stat. (x_i)	Scarto
1	27	1
2	21	-5
3	28	2
4	30	4
5	23	-3
6	27	1
Totale	**156**	**0**

$$\bar{x} = \frac{1}{6} \cdot 156 = 26$$

17

- ***Proprietà 4***: La media minimizza la distanza al quadrato di ogni modalità da una costante

$$\sum_{i=1}^{n}\left(x_i - c\right)^2 \quad \text{è minimo per} \quad c = \bar{x}$$

Esempio

Unità statistica (i)	Voto in stat. (x_i)	Scarto $x_i - \bar{x}$	Scarto $\left(x_i - \bar{x}\right)^2$
1	27	1	1
2	21	-5	25
3	28	2	4
4	30	4	16
5	23	-3	9
6	27	1	1
Totale	**154**	**0**	**56**

$$\bar{x} = \frac{1}{6} \cdot 156 = 26$$

18

- ***Proprietà 5*** *(di internalità)*: La media è sempre compresa tra il minimo e il massimo della distribuzione

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

- ***Proprietà 6*** (*invarianza rispetto a trasformazioni lineari*): se a ogni termine della distribuzione viene applicata la trasformazione *aX* + *b*, allora la media sarà pari a

$$a\bar{x} + b$$

19

Esempio

- Da un sito web americano vengono acquistati 5 libri. Il prezzo è espresso in dollari e si vuole conoscere il prezzo medio in euro

Unità statistica (i)	Prezzo in \$ (x_i)	Prezzo in € (y_i =0,7431•x_i)
1	20	14,862
2	16	11,890
3	30	22,293
4	15,5	11,518
5	21	15,605
Totale	**102,5**	**76,168**

$$\bar{x} = \frac{1}{5} \cdot 102{,}5 = 20{,}5 \qquad \bar{y} = \frac{1}{5} \cdot 76{,}168 = 15{,}234$$

$$\bar{y} = 0{,}7431 \cdot \bar{x} = 0{,}7431 \cdot 20{,}5 = 15{,}234$$

20

Esempio

- Si conosce il numero di componenti delle famiglie residenti nelle Marche. Si vuole calcolare la spesa media giornaliera per consumi alimentari sapendo che vale approssimativamente la relazione y = 16+1,2•x, dove x indica il numero di componenti e y la spesa giornaliera.

Numero di componenti (x_i)	**Numero di famiglie (n_i)**	**$x_i n_i$**	**Spesa giornaliera $y_i = 16+1,2 \cdot x_i$**	**$y_i n_i$**
1 persona	124143	124143	17,2	2135260
2 persone	149531	299062	18,4	2751370
3 persone	124394	373182	19,6	2438122
4 persone	107992	431968	20,8	2246234
5 persone	31751	158755	22	698522
6 persone	11663	69978	23,2	270581,6
Totale	**549474**	**1457088**	**-**	**10540090**

$$\bar{x} = \frac{1457088}{549474} = 2,65 \qquad \bar{y} = \frac{10540090}{549474} = 19,182$$

$$\bar{y} = 16 + 1,2 \cdot \bar{x} = 16 + 1,2 \cdot 2,65 = 19,182$$

Esercizio

Si considerino i dati del Caso Studio 2.

Si verifichi, separatamente per i due titoli,

a) che la media aritmetica annulla la somma degli scarti da se stessa (proprietà 3),

b) che la media aritmetica gode della proprietà di internalità (proprietà 5).

Esercizio

Si considerino i dati del Caso Studio 3.

Si verifichi, separatamente per i due portafogli,

a) che la media aritmetica annulla la somma degli scarti da se stessa (proprietà 3),

b) che la media aritmetica gode della proprietà di internalità (proprietà 5).

- ***Proprietà 7*** *(associativa)*: se un collettivo di n unità è composto da L sottocolletivi con numerosità $n_1, n_2, \ldots, n_L$ e medie $\bar{x}_{(1)}, \bar{x}_{(2)}, \ldots, \bar{x}_{(L)}$, la media complessiva si può ottenere come

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{L} \bar{x}_{(i)} n_i$$

Esempio

- Per la distribuzione delle industrie per classi di fatturato (milioni):

Classi (c_{i-1} - c_i)	**Frequenze (n_i)**	**Valori centrali (x_i)**	**$x_i n_i$**	**Medie di classe $\bar{x}_{(i)}$**	**$\bar{x}_{(i)} n_i$**
0 – 1	144	0,5	72	0,604	87,0
1 – 5	457	3,0	1371	2,556	1168,1
5 – 10	171	7,5	1282,5	7,018	1200,1
10 – 25	112	17,5	1960	16,045	1797,0
25 – 100	27	62,5	1687,5	47,333	1278,0
Totale	911	-	6373	-	5530,2

$$\bar{x} = 6373/911 = 6,996 \qquad \bar{x} = 5530,2/911 = 6,070$$

Esempio

Se si conosce l'ammontare totale di carattere posseduto dalla classe (ossia se è disponibile la *distribuzione di quantità* associata alla distribuzione in classi) è possibile calcolare la media in maniera esatta:

Classi di fatturato (c_{i-1} - c_i)	**Frequenze (n_i)**	**Ammontare di fatturato (A_i)**
0 – 1	144	87,0
1 – 5	457	1168,1
5 – 10	171	1200,1
10 – 25	112	1797,0
25 – 100	27	1278,0
Totale	**911**	**5530,2**

$$\bar{x} = 5530{,}2 / 911 = 6{,}070$$

25

Esercizio

La tabella sottostante riporta le unità locali attive nel settore dei servizi di ristorazione a Macerata nel 2011, per classi di addetti (dati ISTAT). L'ultima colonna presenta il numero di addetti in ogni classe. Calcolare il numero medio di addetti, sulla base della distribuzione di frequenza e sulla base della distribuzione di quantità e confrontare i risultati.

Classe di addetti	Numero unità locali	Numero addetti
0	6	0
1	45	45
2	52	104
3-5	68	252
6-9	19	132
10-15	7	83
16-19	3	54
20-49	3	81
50-99	0	0
100-199	0	0
totale	203	751

26

Esercizio

Si consideri la seguente distribuzione riguardante le aziende zootecniche con suini nel Centro Italia, secondo il numero di capi allevati (Fonte: Ministero dell'Agricoltura, Anno 2004):

Numero capi	Aziende	Suini in migliaia
Fino a 5	43352	71
6-9	1215	8
10-19	904	12
20-49	572	17
50-99	221	15
100 e oltre	570	536
Totale	46834	659

Si richiede di: (a) valutare la consistenza dei dati sulla base della distribuzione di quantità; (b) calcolare la media aritmetica dei capi allevati, nell'ipotesi di uniforme distribuzione all'interno delle classi (scegliendo opportunamente l'estremo superiore dell'ultima classe); (c) calcolare la media aritmetica dei capi allevati facendo uso della distribuzione di quantità; (d) commentare i due valori medi ottenuti, valutando anche la plausibilità dell'ipotesi di uniforme distribuzione per le diverse classi.

27