

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*

<i>Medie analitiche</i>	<i>Medie di posizione</i>
Media aritmetica	Mediana
Media geometrica	Moda
Media troncata	Percentili

- La media aritmetica può essere calcolata direttamente per una *distribuzione di frequenza* non in classi tramite la formula
$$\bar{x} = \frac{1}{n} (x_1 n_1 + x_2 n_2 + \dots + x_k n_k) = \frac{1}{n} \sum_{j=1}^k x_j n_j$$
che tiene conto del fatto che una modalità può ripetersi più volte
- Per il calcolo si imposta una tabella del tipo:

Modalità (x _j)	Frequenze (n _j)	x _j n _j
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	Σ _j x _j n _j

La media aritmetica

- Per una *distribuzione unitaria* di un carattere *quantitativo* di *n* termini, la **media aritmetica** è definita come:
$$\bar{x} = \frac{1}{n} (x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i$$
- Distribuzione del voto in Statistica per un gruppo di 6 studenti:

Esempio

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

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

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

$$\bar{x} = x_1f_1 + x_2f_2 + \dots + x_kf_k = \sum_{j=1}^k x_jf_j$$

basata sulle frequenze relative.

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

Modalità (x _j)	Frequenze (n _j)	Freq. relative (f _j)	x _j f _j
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	Σ _j x _j f _j

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

- 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 ($\bar{x}_j$) come:

$$\bar{x} = \frac{1}{n}(\bar{x}_1n_1 + \bar{x}_2n_2 + \dots + \bar{x}_kn_k) = \frac{1}{n}\sum_{j=1}^k \bar{x}_jn_j$$

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

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

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

$$\bar{x} = \frac{1}{n}\sum_{j=1}^k x_jn_j \qquad \text{oppure} \qquad \bar{x} = \sum_{j=1}^k x_jf_j$$

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

Esempio

Voto (x _j)	Frequenza (n _j)	Freq. relativa (f _j)	x _j f _j
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

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

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

Classi (c _{j-1} - c _j)	Frequenze (n _j)	Valori centrali (x _j)	x _j n _j
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	-	Σ _j x _j n _j

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

Classi (c _{j-1} - c _j)	Frequenze (n _j)	Freq. relative (f _j)	Valori centrali (x _j)	x _j f _j
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	-	Σ _j x _j f _j

Esempio

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

Classi ($c_{j-1} - c_j$)	Frequenze (n_j)	Valori centrali (x_j)	$x_j n_j$	Freq. relative (f_j)	$x_j f_j$
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

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

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

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

- 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

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

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, $x_1, x_2, \dots, x_n$ e $y_1, y_2, \dots, y_n$, se vale la condizione $x_j \leq y_j$ per ogni j , e almeno una volta $x_j < y_j$, allora

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

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

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

Esempio

Unità statistica (i)	Voto in stat. (x _i)	Scarto
1	27	1,333
2	21	-4,667
3	28	2,333
4	30	4,333
5	21	-4,667
6	27	1,333
Totale	154	0

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

- **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

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

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

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

Esempio

Unità statistica (i)	Voto in stat. (x _i)	Scarto x _i - $\bar{x}$	Scarto (x _i - $\bar{x}$) ²
1	27	1,333	1,778
2	21	-4,667	21,778
3	28	2,333	5,444
4	30	4,333	18,778
5	21	-4,667	21,778
6	27	1,333	1,778
Totale	154	0	71,333

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

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

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

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

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 $y = 16 + 1,2 \cdot x$, dove x indica il numero di componenti e y la spesa giornaliera.

Numero di componenti (x_j)	Numero di famiglie (n_j)	$x_j n_j$	Spesa giornaliera $y_j = 16 + 1,2 \cdot x_j$	$y_j n_j$
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

$$\bar{x} = \frac{1457088}{549474} = 2,65 \qquad \bar{y} = \frac{10540090}{549474} = 19,182$$
$$\bar{y} = 16 + 1,2 \cdot \bar{x} = 16 + 1,2 \cdot 2,65 = 19,182$$

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

Esempio

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

Classi ($c_{j-1} - c_j$)	Frequenze (n_j)	Valori centrali (x_j)	$x_j n_j$	Medie di classe $\bar{x}_{(j)}$	$\bar{x}_{(j)} n_j$
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 ($c_{j-1} - c_j$)	Frequenze (n_j)	Ammontare di fatturato (A_j)
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

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