

## Statistica Descrittiva

### Soluzioni 3. Medie potenziate

#### Introduzione

Per un carattere  $X$  con modalità  $x_i, i = 1, \dots, I$ , con frequenza assoluta  $f_i$ , la **media potenziata di ordine  $r$**  si calcola come segue

$$m_r = \left( \frac{\sum_{i=1}^I x_i^r \cdot f_i}{\sum_{i=1}^I f_i} \right)^{1/r}.$$

La media potenziata di ordine 1 corrisponde alla media aritmetica,  $m_1 = \left( \frac{\sum_{i=1}^I x_i \cdot f_i}{\sum_{i=1}^I f_i} \right)$ , quella di ordine -1 alla media armonica,  $m_{-1} = \left( \frac{\sum_{i=1}^I x_i^{-1} \cdot f_i}{\sum_{i=1}^I f_i} \right)^{-1}$ , mentre per  $r \rightarrow 0$  si ottiene la media geometrica,  $m_0 = \left( x_1^{f_1} * \dots * x_I^{f_I} \sum f_i \right)^{1/\sum f_i}$ .

#### Osservazioni

1) La media potenziata è una funzione non decrescente di  $r$ . Vale a dire che, al crescere di  $r$ , il valore della media potenziata  $m_r$  calcolata per valori positivi cresce o, al peggio, rimane costante:

$$\dots \leq m_{-1} \leq m_0 \leq m_1 \leq m_2 \leq \dots$$

2) Inoltre, indicando con  $x_{min}$  e con  $x_{max}$ , rispettivamente, il valore minimo ed il valore massimo delle  $N$  osservazioni positive del carattere  $X$  si ha che

$$\lim_{r \rightarrow -\infty} m_r = x_{min} \quad \text{e} \quad \lim_{r \rightarrow +\infty} m_r = x_{max}.$$

3) La media aritmetica è un operatore lineare. Si consideri la media  $m_1$  delle osservazioni del carattere  $X$ ,  $m_1 = (\sum_i x_i \cdot f_i) / (\sum_i f_i)$ . Allora, la trasformata lineare  $a + b \cdot X$ , dove  $a, b$  sono costanti reali qualunque, ha media aritmetica  $a + b \cdot m_1$ .

4) Sia data la distribuzione doppia di frequenza dei caratteri  $(X, Y)$ , con  $Y$  carattere con modalità  $y_j$  di frequenze assoluta  $f_j$ . Supponiamo che vengano forniti i valori della media aritmetica di  $X$  condizionatamente alle modalità di  $Y$ , cioè  $m_1(X|Y = y_1), \dots, m_1(X|Y = y_J)$ . Allora, la media dei valori medi  $m_1(X|Y = y_1), \dots, m_1(X|Y = y_J)$  data da

$$\frac{m_1(X|Y = y_1) * f_{y_1} + \dots + m_1(X|Y = y_J) * f_{y_J}}{N}$$

coincide con la media aritmetica marginale di  $X$ ,  $m_{1x} = \sum_{i=1}^I x_i f_i / N$ .

**Esercizio A.** Utilizzando le proprietà delle sommatorie risulta:

$$a) \sum_{i=1}^3 \sum_{j=1}^2 (a+i)b^j = \left[ \sum_{i=1}^3 (a+i) \right] \cdot \left[ \sum_{j=1}^2 b^j \right] = [(a+1) + (a+2) + (a+3)] \cdot (b+b^2) = (3a+6)(b^2+b);$$

$$b) \sum_{i=1}^2 \sum_{j=1}^2 (a^j - b^i) = \sum_{i=1}^2 \left[ \sum_{j=1}^2 a^j - \sum_{j=1}^2 b^i \right] = \sum_{i=1}^2 \sum_{j=1}^2 a^j - \sum_{j=1}^2 \sum_{i=1}^2 b^i = 2(a + a^2) - 2(b + b^2);$$

$$c) \sum_{i=1}^2 \sum_{j=1}^3 (a+b)^i j = \left[ \sum_{i=1}^2 (a+b)^i \right] \left[ \sum_{j=1}^3 j \right] = [(a+b) + (a+b)^2] \cdot [1+2+3] = 6(a+b+a^2+b^2+2ab);$$

$$d) \sum_{i=1}^2 \sum_{j=1}^2 a^{j-1} b^i = \left[ \sum_{i=1}^2 b^i \right] \cdot \left[ \sum_{j=1}^2 a^{j-1} \right] = (b + b^2)(1 + a).$$

**Esercizio B.**

a) Per ricavare la funzione di ripartizione, si calcolano prima le frequenze relative cumulate  $P_i$ :

$x_i$	0	1	2	3	4	5	6	7	8	9	10	11
$f_i$	5	6	8	19	25	56	23	11	4	2	0	2
$F_i$	5	11	19	38	63	119	142	153	157	159	159	161
$P_i$	0,031	0,068	0,118	0,236	0,391	0,739	0,882	0,950	0,975	0,988	0,988	1

Quindi, considerando che stiamo trattando la distribuzione di frequenza di un carattere quantitativo discreto, la funzione di ripartizione è una funzione a gradini definita da

$$F(x) = \begin{cases} 0, & x < 0, \\ 0,031, & 0 \leq x < 1, \\ 0,068, & 1 \leq x < 2, \\ 0,118, & 2 \leq x < 3, \\ 0,236, & 3 \leq x < 4, \\ 0,391, & 4 \leq x < 5, \\ 0,739, & 5 \leq x < 6, \\ 0,882, & 6 \leq x < 7, \\ 0,950, & 7 \leq x < 8, \\ 0,975, & 8 \leq x < 9, \\ 0,988, & 9 \leq x < 11, \\ 1, & x \geq 11. \end{cases}$$

**Esercizio C.**

a) Le due funzioni di ripartizione si ricavano da:

Classi di abitanti	chiusura delle classi	Belluno		Rovigo	
		$f_i$	$P_i$	$f_i$	$P_i$
$\leq 3000$	$[-0,5 - 3000,5]$	50	0,725	29	0,58
3001 – 5000	$[3000,5 - 5000,5]$	12	0,899	11	0,80
5001 – 10000	$[5000,5 - 10000,5]$	5	0,971	3	0,86
10001 – 20000	$[10000,5 - 20000,5]$	1	0,986	5	0,96
20001 – 30000	$[20000,5 - 30000,5]$	1	1	1	0,98
30001 – 100000	$[30000,5 - 100000,5]$	0	1	1	1
		69	–	50	–

Per rappresentare graficamente la funzione di ripartizione dei comuni della provincia di Belluno (si procede analogamente per la provincia di Rovigo) basta congiungere i seguenti punti con dei segmenti di retta:  $(-0,5; 0)$ ,  $(3000,5; 0,725)$ ,  $(5000,5; 0,899)$ ,  $(10000,5; 0,971)$ ,  $(20000,5; 0,986)$ ,  $(30000,5; 0,988)$ ,  $(100000,5; 1)$ .

(30000, 5; 1), (100000, 5; 1). Ovviamente, la funzione di ripartizione è nulla per valori di  $x$  minori di  $-0,5$ , ed uguale a 1 per valori di  $x$  maggiori di 100000,5.

b) Il confronto fra le due funzioni di ripartizione mostra che la funzione di ripartizione relativa a Belluno è, per ogni  $x$ , sempre maggiore od uguale alla funzione di ripartizione relativa a Rovigo. In questo caso si dice che la distribuzione per Rovigo è statisticamente superiore alla distribuzione per Belluno. In altre parole, la distribuzione relativa a Rovigo è spostata a destra, verso valori più alti, rispetto alla distribuzione relativa a Belluno. Possiamo quindi dire che i comuni di Belluno hanno in genere dimensione più piccola rispetto a quelli di Rovigo.

#### Esercizio D.

1) La media potenziata di ordine  $r = 1$ , o media aritmetica, è

$$(1355 + 763 + 953 + 973 + 1113 + 1030 + 1244)/7 = 1061,571,$$

dove si consideri che le frequenze assolute sono tutte pari a 1.

La media potenziata di ordine  $r = 2$  è

$$\{(1355^2 + 763^2 + 953^2 + 973^2 + 1113^2 + 1030^2 + 1244^2)/7\}^{1/2} = 1077,055$$

e quella di ordine 3 è

$$(1355^3 + 763^3 + 953^3 + 973^3 + 1113^3 + 1030^3 + 1244^3)/7 = 1092,005$$

Si noti che il valore della media potenziata cresce al crescere dell'ordine  $r$ .

2) Si consideri che la media aritmetica è un operatore lineare, di modo che  $\sum_i(100 + 0,8 \cdot x_i)/N = 100 + 0,8 \cdot \sum_i x_i/N = 100 + 0,8 \cdot 1061,571 = 949,2568$ .

#### Esercizio E.

a) Una media opportuna dei quozienti di natalità dei quattro Paesi in questione è data dalla media aritmetica degli stessi ponderata con le popolazioni dei rispettivi Paesi:

$$m = \frac{\sum_{i=1}^4 (\text{quoz.})_i (\text{pop.})_i}{\sum_{i=1}^4 (\text{pop.})_i} = \frac{387767,45}{31177} = 12,4.$$

Infatti

$$m = \frac{\sum_{i=1}^4 (\text{quoz.})_i (\text{pop.})_i}{\sum_{i=1}^4 (\text{pop.})_i} = \frac{\sum_{i=1}^4 \frac{(\text{nati})_i 1000}{(\text{pop.})_i} (\text{pop.})_i}{\sum_{i=1}^4 (\text{pop.})_i} = \frac{\sum_{i=1}^4 (\text{nati})_i}{\sum_{i=1}^4 (\text{pop.})_i} \cdot 1000,$$

non è altro che il quoziente di natalità complessivo dell'area geografica risultante dall'unione dei quattro Paesi.

#### Esercizio F.

1) La distribuzione di frequenza doppia che si ricava dalla Tavola 1.2 è la seguente

	Maschi	Femmine
0 – 20	439271	414166
20 – 40	700362	657834
40 – 70	923247	932415
> 70	238648	394007
Totale	2301528	2398422

2) Poniamo 101 come limite superiore alla quarta classe di età e consideriamo il valore centrale di ogni classe di età (10; 30; 55; 85,5), assumendo che vi sia un distribuzione uniforme delle unità statistiche all'interno di ogni classe. Allora la media per i Maschi è data da

$$(10 \cdot 439271 + 30 \cdot 700362 + 55 \cdot 923247 + 85,5 \cdot 238648) / 2301528 = 41,96628$$

e quella per le femmine è

$$(10 \cdot 414166 + 30 \cdot 657834 + 55 \cdot 932415 + 85,5 \cdot 394007) / 2398422 = 45,3828$$

3) La media totale, senza distinzione per sesso, è data da

$$(10 \cdot (439271 + 414166) + 30 \cdot (700362 + 657834) + 55 \cdot (923247 + 932415) + 85,5 \cdot (238648 + 394007)) / (2301528 + 2398422) = 43,70975$$

Si noti che calcolare la media delle medie condizionate al sesso, vale a dire

$$(41,96628 \cdot 2301528 + 45,3828 \cdot 2398422) / (2301528 + 2398422) = 43,70975$$

fornisce lo stesso valore della media marginale, calcolata senza considerare il condizionamento al sesso.

### Esercizio G.

a) Assumendo che la distanza tra le 4 stazioni sia rispettivamente di 11, 12 e 15 Km, per calcolare la velocità media tra la stazione di partenza e quella di arrivo si può usare la media armonica pesata

$$m = \frac{11 + 12 + 15}{\frac{1}{100} \cdot 11 + \frac{1}{80} \cdot 12 + \frac{1}{70} \cdot 15} = \frac{38}{0,474} = 80,12.$$

Questa corrisponde alla velocità media cercata in quanto il numeratore non è altro che la distanza complessiva, mentre il denominatore è il tempo totale impiegato per percorrere l'intera tratta.

### Esercizio H.

a) Partendo con un capitale iniziale pari a  $C_0$ , l'ammontare totale in possesso del giocatore dopo la prima giocata è pari a  $C_0 \cdot 3$ , dopo la seconda giocata è pari a  $C_0 \cdot 3 \cdot 8$  e dopo la terza giocata è pari a  $C_0 \cdot 3 \cdot 8 \cdot 9$ . Se l'incremento relativo fosse rimasto costante, diciamo pari a  $m$ , per tutte le giocate, allora il giocatore si sarebbe trovato con  $C_0 \cdot m$  dopo la prima giocata, con  $C_0 \cdot m \cdot m$  dopo la seconda giocata, e con  $C_0 \cdot m \cdot m \cdot m$  dopo la terza giocata. Imponendo che il capitale finale sia lo stesso nei due casi, abbiamo

$$C_0 \cdot 3 \cdot 8 \cdot 9 = C_0 \cdot m \cdot m \cdot m,$$

da cui si ricava  $m = \sqrt[3]{3 \cdot 8 \cdot 9}$ , cioè  $m$  è la media geometrica.

### Esercizio I.

a) Scegliendo come estremo superiore 19999 addetti e congiungendo le classi aggiungendo e togliendo 0,5 agli estremi di ogni classe si ottiene:

Classi di addetti	chiusura delle classi	$p_i$	$d_i$	$h_i = p_i/d_i$
Fino a 49	$[-0,5 - 49,5)$	0,0938	50	0,001877
da 50 a 99	$[49,5 - 99,5)$	0,0874	50	0,001747
da 100 a 199	$[99,5 - 199,5)$	0,1319	100	0,001319
da 200 a 499	$[199,5 - 499,5)$	0,1669	300	0,000556
da 500 a 999	$[499,5 - 999,5)$	0,1587	500	0,000317
da 1000 a 1999	$[999,5 - 1999,5)$	0,1112	1000	0,000111
da 2000 a 4999	$[1999,5 - 4999,5)$	0,0714	3000	0,000024
da 5000 a 9999	$[4999,5 - 9999,5)$	0,1787	5000	0,000018
oltre 10000	$[9999,5 - 19999,5)$	0	10000	0

L'istogramma si ottiene rappresentando graficamente 9 rettangoli adiacenti di base  $d_i$  e altezza  $h_i$  (l'ultimo rettangolo ha altezza nulla).

b) La frequenza che spetterebbe all'intervallo  $[10 - 150)$  sotto l'ipotesi semplificatrice di uniforme distribuzione all'interno delle classi è pari alla differenza  $F(150) - F(10)$ . Utilizzando le densità relative si ottiene:

$$F(10) = 0,001877 \cdot 10,5 = 0,01971;$$

$$F(150) = 0,0938 + 0,0874 + 0,001319(150 - 99,5) = 0,2478;$$

$$F(150) - F(10) = 0,2281.$$

Quindi, considerando che tutte le imprese sono 7544, la frequenza assoluta cercata è pari a  $0,2281 \cdot 7544 = 1720,8$  imprese.

### Esercizio L.

a) Per disegnare la funzione di ripartizione dobbiamo calcolare le frequenze relative cumulate e decidere delle opportune chiusure per le classi, tenendo in considerazione che il carattere rilevato è discreto:

Classi di Voto	$[18 - 22)$	$[22 - 25)$	$[25 - 28)$	$[28 - 30]$	
chiusura delle classi	$[17,5 - 21,5)$	$[21,5 - 24,5)$	$[24,5 - 27,5)$	$[27,5 - 30,5]$	
$f_i$	12	20	15	16	63
$p_i$	0,1905	0,3175	0,2381	0,2540	1
$P_i$	0,1905	0,5079	0,7460	1	-

Per disegnare la funzione di ripartizione  $F(x)$ , basta congiungere con dei segmenti di retta i punti:  $(17,5; 0)$ ,  $(21,5; 0,1905)$ ,  $(24,5; 0,5079)$ ,  $(27,5; 0,7460)$ ,  $(30,5; 1)$ . Ovviamente, la funzione di ripartizione è nulla per valori di  $x$  minori di 17,5, ed uguale a 1 per valori di  $x$  maggiori di 30,5.

b) La proporzione di studenti che hanno preso un voto di almeno 25 è data da  $0,2381 + 0,2540 = 1 - 0,5079 = 0,4921$ .

c) La proporzione di studenti che hanno preso un voto inferiore a 28 è pari a  $1 - 0,2540 = 0,7460$ . (Si noti che per la rappresentazione grafica della  $F(x)$  si è fatto ricorso alla correzione per continuità, mentre le proporzioni sono state calcolate utilizzando le frequenze cumulate indicate nella tabella. Le proporzioni appena calcolate, rispetto alla  $F(x)$  ottenuta, sono date da  $F(24,5) = 0,4921$  e  $F(27,5) = 0,7460$ .)

**N.B.** Si assuma che l'esatta distribuzione di frequenza del voto sia la seguente:

Voto	18	19	20	21	22	23	24	25	26	27	28	29	30	
$f_i$	3	4	3	2	0	20	0	5	5	5	8	8	0	63

Si ricavi la funzione di ripartizione del voto sulla base di questa nuova distribuzione e la si confronti con la funzione di ripartizione ottenuta in precedenza. In particolare si noti che mentre nella prima e nella terza classe può ritenersi valida l'ipotesi di uniforme distribuzione all'interno delle classi, nella

seconda e nella quarta classe tale ipotesi non è valida e la funzione di ripartizione precedentemente calcolata non approssima, in queste classi, la nuova funzione di ripartizione a gradini.

### Esercizio M.

a) La media aritmetica del fatturato delle imprese si può ottenere come

$$m = \frac{260 + 1031 + 3032 + 2833 + 2551 + 11458}{71 + 137 + 189 + 80 + 36 + 51} = \frac{21165}{564} = 37,53.$$

b) Per valutare l'ipotesi di uniforme distribuzione all'interno delle classi possiamo confrontare i valori centrali di classe  $x_i$  con le medie di classe  $m_i$ :

Fatturato	[0 – 5)	[5 – 10)	[10 – 25)	[25 – 50)	[50 – 100)	[100 – 500)
$m_i$	3,66	7,53	16,04	35,4	70,86	224,67
$x_i$	2,5	7,5	17,5	37,5	75	300

Nel caso di uniforme distribuzione all'interno delle classi, le medie di classe  $m_i$  dovrebbero essere approssimativamente uguali ai valori centrali di classe  $x_i$ . Infatti, sotto l'ipotesi di uniforme distribuzione le osservazioni all'interno delle classi formerebbero delle progressioni aritmetiche e le medie di classe sarebbero uguali ai valori centrali (o alle semisomme dei valori centrali) delle progressioni, e quindi sarebbero vicine ai valori centrali delle classi. Nel nostro caso possiamo ritenere valida tale ipotesi, a parte forse per la prima e per l'ultima classe.

### Esercizio N.

a) Indicando con  $x_j$  il tempo impiegato da una generica magliaia  $j$  per produrre il maglione, si ha che la quantità di maglioni prodotti nell'unità di tempo è pari a  $1/x_j$ . Pertanto, la relazione di eguaglianza da cui ricavare il valore medio è

$$\sum_{j=1}^3 \frac{1}{x_j} = \sum_{j=1}^3 \frac{1}{m},$$

e quindi il tempo medio (in minuti) di produzione richiesto è dato dalla media armonica

$$m = \frac{3}{\frac{1}{125} + \frac{1}{85} + \frac{1}{107}} = 103,06.$$

### Esercizio O.

a) Sfruttando la proprietà associativa valida nel caso di un miscuglio, la media aritmetica senza distinzione di qualifica può ottenersi come media aritmetica delle medie aritmetiche distinte per qualifica, ponderate con le rispettive frequenze. Per gli insegnanti di *Scuola elementare* risulta che

$$m = (1205 \times 18 + 950 \times 10)/28 = 1113,93.$$

Analogamente, per gli insegnanti di *Scuola media* si ha che

$$m = (1310 \times 25 + 1020 \times 9)/34 = 1233,24.$$

b) Lo stipendio medio di tutti gli insegnanti considerati è pari a

$$m = (1113,93 \times 28 + 1233,24 \times 34)/62 = 1179,36.$$

**Esercizio P.**

a) La media aritmetica (esatta) del fatturato delle imprese si può calcolare come media aritmetica ponderata delle medie aritmetiche di classe  $m_i = X_i/f_i$ , dove con  $X_i$  si è indicato il fatturato della classe  $i$ -esima e con  $f_i$  la frequenza assoluta della classe. Oppure, considerando il totale del fatturato, come  $m = 3015211/111 = 27164,06$ .

b) Considerando 0 come estremo inferiore e 300 come estremo superiore, si consideri la tabella

Fatturato	[0–5)	[5–10)	[10–25)	[25–50)	[50–100)	[100–300)	Totale
$f_i$	26	35	22	14	10	4	111
$X_i$	88690	249448	370671	454936	724037	1127429	3015211
$x_i$	2,5	7,5	17,5	37,5	75	200	
$m_i$	3411,2	7127,1	16848,7	32495,4	72403,7	281857,3	

L'ipotesi di uniforme distribuzione all'interno delle classi può essere discussa confrontando i valori centrali di classe  $x_i$  con le medie di classe  $m_i$ . Questa ipotesi è chiaramente non valida per la prima e l'ultima classe, mentre si può ritenere sostanzialmente valida per la seconda, la terza e la quinta classe.