Marco Minozzo
Laurea in Economia del Commercio Internazionale
Laurea in Economia e Amministrazione delle Imprese
Università degli Studi di Verona (sede di Vicenza)

Statistica Descrittiva
Soluzioni 4. Medie lasche

Esercizio A.
a) Per il calcolo delle medie richieste si consideri la tabella seguente dove con x_i si sono indicati i valori centrali delle classi e dove l'estremo inferiore e quello superiore sono stati posti pari a 12 e 50 rispettivamente,

Classi di età	< 15	15–19	20–24	25–29	30–34	35–39	> 39	
chiusura classi	$[12-15)$	$[15-20)$	$[20-25)$	$[25-30)$	$[30-35)$	$[35-40)$	$[40-50)$	
f_i	0	4	8	7	6	1	3	29
p_i	0	0,1379	0,2759	0,2414	0,2069	0,0345	0,1034	1
P_i	0	0,1379	0,4138	0,6552	0,8621	0,8966	1	–
d_i	3	5	5	5	5	5	10	
h_i	0	0,02758	0,05518	0,04828	0,04138	0,00690	0,01034	
x_i	13,5	17,5	22,5	27,5	32,5	37,5	45	
$\ln x_i$	2,603	2,862	3,114	3,314	3,481	3,624	3,807	

La classe modale è la classe con massima densità ed è quindi la classe $[20-25)$ a cui corrisponde una densità di 0,05518. Considerando che la classe che contiene la mediana è la classe $[25-30)$, sotto l'ipotesi di uniforme distribuzione all'interno delle classi, la mediana è data da $m_e = 25+(0,5-0,4138)/0,04828 = 26,79$. La media aritmetica è data da $m = \sum_{i=1}^{7} x_i p_i = 13,5 \cdot 0 + 17,5 \cdot 0,1379 + \cdots + 45 \cdot 0,1034 = 27,93$; la media geometrica risulta invece $m = \exp\left(\frac{1}{29}\sum_{i=1}^{7} f_i \cdot \ln x_i\right) = \exp(95,489/29) = 26,92$.
b) Tra le medie calcolate per sintetizzare la distribuzione in oggetto l'unica media non appropriata è la media geometrica.
c) Per i maschi della ULSS 22 si consideri la tabella:

Classi di età	< 15	15–19	20–24	25–29	30–34	35–39	> 39	
chiusura classi	$[12-15)$	$[15-20)$	$[20-25)$	$[25-30)$	$[30-35)$	$[35-40)$	$[40-50)$	
f_i	0	15	82	98	164	103	96	558
p_i	0	0,0269	0,1470	0,1756	0,2939	0,1846	0,1720	1
P_i	0	0,0269	0,1739	0,3495	0,6434	0,8280	1	–
x_i	13,5	17,5	22,5	27,5	32,5	37,5	45	

In questo caso la media aritmentica è pari a 32,82, mentre considerando che la classe mediana è $[30-35)$, la mediana è pari a $m_e = 30 + (0,5-0,3495) \cdot ((35-30)/0,2939) = 32,56$.

Esercizio B.
a) Ponendo l'estremo superiore dell'ultima classe pari a 100 anni non compiuti si ha
L'istogramma di frequenza si ottiene rappresentando per ogni classe un rettangolo di larghezza pari a d_i e altezza pari a h_i.
b) L'intervallo $[30,50)$ coinvolge due classi e quindi, sotto l'ipotesi di uniforme distribuzione, ad esso corrisponde una frequenza relativa teorica pari a $0,0143(45-30) + 0,0130(50-45) = 0,2795$ a cui corrisponde una frequenza assoluta di $0,2795 \cdot 830714 = 232184,6$. La mediana cade nella classe $[25,45)$

chiusura classi	$[0,1)$	$[1,5)$	$[5,10)$	$[10,15)$	$[15,25)$	$[25,45)$	$[45,65)$	$[65,100)$	Totale
f_i	6392	24696	34605	36251	94687	238052	216283	179748	830714
p_i	0,0077	0,0297	0,0416	0,0436	0,1140	0,2866	0,2604	0,2164	1
P_i	0,0077	0,0374	0,0790	0,1226	0,2366	0,5232	0,7836	1	—
d_i	1	4	5	5	10	20	20	35	—
h_i	0,0077	0,0074	0,0083	0,0087	0,0114	0,0143	0,0130	0,0061	—

e quindi è pari a $25 + 20(0,5 - 0,2366)/0,2866 = 43,38$. Analogamente, il primo quartile cade nella classe $[25,45)$ ed è pari a $Q_1 = 25 + 20(0,25 - 0,2366)/0,2866 = 25,94$ mentre il terzo quartile cade nella classe $[45,65)$ ed è pari a $Q_3 = 45 + 20(0,75 - 0,5232)/0,2604 = 62,42$.

Esercizio C. Si procede innanzitutto con il calcolo dei valori riportati in tabella dove con x_i si sono indicati i valori centrali di classe e con m_i le medie di classe ottenute come X_i/f_i:

Fatt.	[0;0,5)	[0,5;1)	[1;5)	[5;10)	[10;25)	[25;50)	[50;100)	[100;500)	[500;4000)	Tot.
f_i	8	136	2997	2954	3189	1446	802	646	131	12309
X_i	3	108	9585	21275	50427	50818	56120	126831	260018	575185
x_i	0,25	0,75	3	7,5	17,5	37,5	75	300	2250	—
m_i	0,375	0,794	3,198	7,202	15,813	35,144	69,975	196,33	1984,9	—
p_i	0,0006	0,0110	0,2435	0,2400	0,2591	0,1175	0,0652	0,0525	0,0106	1
P_i	0,0006	0,0116	0,2551	0,4951	0,7542	0,8717	0,9369	0,9894	1	—
d_i	0,5	0,5	4	5	15	25	50	400	3500	—

a) Sotto l'ipotesi di uniforme distribuzione la media aritmetica (utilizzando i valori centrali di classe) è pari a $m = (0,25 \cdot 8 + \cdots + 2250 \cdot 131)/(8 + \cdots + 131) = 689982,5/12309 = 56,06$. La mediana cade nella classe $[10;25)$ ed è pari a $m_e = 10 + 15(0,5 - 0,4951)/0,2591 = 10,284$. Il secondo e l'ottavo decile sono invece pari rispettivamente a $D_2 = 1 + 4(0,2 - 0,0116)/0,2435 = 4,095$ e a $D_8 = 25 + 25(0,8 - 0,7542)/0,1175 = 34,475$.
b) Utilizzando la distribuzione di quantità, la media aritmetica risulta pari a $575185/12309 = 46,729$. L'ipotesi di uniforme distribuzione appare più in conflitto con i dati per la classe [100;500) dove è massima, in valore assoluto, la differenza relativizzata $(x_i - m_i)/d_i$.

Esercizio D.
a) Fissando l'altezza minima e massima rispettivamente pari a 150 cm e a 199 cm, ed indicando con x_i i valori centrali di classe, per il Veneto si ha

chiusura classi	[149,5–159,5)	[159,5–164,5)	[164,5–169,5)	[169,5–179,5)	[179,5–184,5)	[184,5–189,5)	[189,5–199,5)
x_i	154,5	162	167	174,5	182	187	194,5
d_i	10	5	5	10	5	5	10
p_i	0,008	0,036	0,127	0,550	0,187	0,071	0,002
$h_i \cdot 100$	0,08	0,72	2,54	5,50	3,74	1,42	0,21
$P_i \cdot 100$	0,8	4,4	17,1	72,1	90,8	97,9	100,0
$\ln x_i$	5,040	5,088	5,118	5,162	5,204	5,231	5,270

da cui si vede che la classe modale è la classe 170–179, per cui la mediana è pari a $m_e(V) = 169,5 + 10(0,5 - 0,171)/0,55 = 175,48$. La media aritmetica è invece pari a $m_1(V) = \sum_{i=1}^{7} p_i x_i = 175,65$, mentre la media geometrica è data da $m_0(V) = \exp(\sum p_i \ \ln x_i) = \exp(5,0677) = 158,80$.
b) I quantili richiesti si possono ottenere con la formula $x(\alpha) = c_{h-1} + d_h(\alpha - P_{h-1})/p_h$, dove $P_{h-1} < \alpha < P_h$. Perciò si ottiene: per il primo quartile $Q_1 = 169,5 + 10(0,25 - 0,171)/0,55 = 170,94$; per il

terzo quartile $Q_3 = 179,5 + 5(0,75 - 0,721)/0,187 = 180,28$; per il primo decile $D_1 = 164,5 + 5(0,1 - 0,044)/0,127 = 166,71$; per il nono decile $D_9 = 179,5 + 5(0,9 - 0,721)/0,187 = 184,29$.
c) Per sintetizzare la distribuzione della statura degli iscritti con una media opportuna si possono utilizzare sia la mediana che la media aritmetica. Non è invece appropriato usare la media geometrica.
d) Procedendo come al punto a), per la regione Sicilia si ottiene una media aritmetica pari a $m_1(S) = 171,81$. Inoltre, essendo le frequenze percentuali cumulate pari a 3,0, 13,0, 37,2, 89, 97,6, 99,6, 100,0, la classe mediana è la classe $[169,5 - 179,5)$ e la mediana è pari a $m_e(S) = 169,5 + 10(0,5 - 0,372)/0,518 = 171,97$. Si può notare che mentre la distribuzione per il Veneto ha una leggera asimmetrica positiva e si ha $m_1(V) > m_e(V)$, la distribuzione per la Sicilia ha una leggera asimmetrica negativa e si ha $m_1(S) < m_e(S)$.

Esercizio E.
a) Per la funzione di ripartizione del numero di esami sostenuti si consideri la seguente tabella:

Numero esami	f_i	p_i	P_i
1	58	0,3867	0,3867
2	42	0,2800	0,6667
3	12	0,0800	0,7467
4	38	0,2533	1
	150	1	–

Considerando che il carattere è discreto, la funzione di ripartizione è una funzione a gradini ed è definita da

$$F(x) = \begin{cases} 0, & x < 1, \\ 0,3867, & 1 \le x < 2, \\ 0,6667, & 2 \le x < 3, \\ 0,7467, & 3 \le x < 4, \\ 1, & x \ge 4. \end{cases}$$

b) Il primo, il secondo (ovvero la mediana) ed il terzo quartile sono dati rispettivamente da $Q_1 = 1$, $Q_2 = m_e = 2$, $Q_3 = 4$.