Esercizio 1

Un'azienda del settore abbigliamento operante nella provincia di Frosinone ha registrato i seguenti ricavi annui (in migliaia di Euro) derivanti dalla vendita di capi in pelle:

Anni	*1991*	*1992*	*1993*	*1994*	*1995*	*1996*	*1997*	*1998*	*1999*	*2000*	*2001*	*2002*
Ricavi	102,3	105,6	100,7	105,2	104,8	104,6	100,7	100,2	100,9	102,7	95,4	120,7

a) Calcolare la media aritmetica ed i quartili per i ricavi da vendite di capi in pelle concernenti il periodo 1991-1999.
b) Aggiungendo il dato relativo al solo anno 2000 ai dati considerati al punto a), discutere come variano la media aritmetica e la mediana considerando l'osservazione aggiuntiva relativa all'anno 2000.
c) Aggiungendo il dato relativo agli anni 2001 e 2002 ai dati considerati al punto b), discutere come variano la media aritmetica e la mediana considerando le osservazioni aggiuntive relative agli anni 2001 e 2002.
d) Relativamente ai dati dell'intero periodo 1991-2002, calcolare la media aritmetica dei ricavi in Lire Italiane.
e) Considerando unicamente l'intera successione delle intensità dei ricavi di vendita in Euro (seconda riga della tabella):
   1. suddividere tale successione in 4 classi equiampie e disegnare l'istogramma normalizzato di frequenza,
   2. suddividere tale successione in classi equiampie di ampiezza 5 e disegnare l'istogramma normalizzato di frequenza,
   3. suddividere tale successione in classi equifrequenti (frequenza 4) e disegnare l'istogramma normalizzato di frequenza.
f) A partire dalle distribuzioni in classi calcolare media e mediana e confrontarle con quelle ottenute dalla distribuzione originaria.
g) A partire dalla distribuzione in classi di intensità costruita al punto e3) determinare:
   1. Il quinto ed il 95-esimo percentile,
   2. La percentuale di ricavi compresa tra 100 e 104 milioni.
h) Determinare la moda della distribuzione originaria e delle tre distribuzioni in classi ottenute al punto e.

SVOLGIMENTO

Ai fini dello svolgimento dell'esercizio denotiamo i dati di input utilizzando differenti colori, come riportato nella seguente tabella:

Anni	*1991*	*1992*	*1993*	*1994*	*1995*	*1996*	*1997*	*1998*	*1999*	*2000*	*2001*	*2002*
Ricavi	102.3	105.6	100.7	105.2	104.8	104.6	100.7	100.2	100.9	102.7	95.4	120.7

In giallo si sono indicati i dati su cui effettuare inizialmente i calcoli di media e quartili (punto a) e con gli altri colori i dati relativi agli anni che via via si vengono ad aggiungere e su cui ripetere i calcoli di media e quartili effettuando il confronto con i risultati ottenuti in precedenza (punti b e c).

a)

Il calcolo della media aritmetica richiede il calcolo della somma dei ricavi annui.

$$\mu = \frac{\sum_{t=1991}^{1999} x_t}{n} = \frac{925}{9} = 102.78$$

Per il calcolo della mediana è invece necessario procedere prima ordinando i dati in ordine non decrescente, come di seguito:

Posiz.	1	2	3	4	5	6	7	8	9
Anni	*1998*	*1993*	*1997*	*1999*	*1991*	*1996*	*1995*	*1994*	*1992*
Ricavi	100.2	100.7	100.7	100.9	102.3	104.6	104.8	105.2	105.6

Essendo n dispari la mediana corrisponde all'intensità che occupa il posto (n+1)/2 ovvero (9+1)/2, cioè all'intensità del carattere che occupa la quinta posizione nella successione ordinata ed è quindi pari a 102.3.

Sulla successione ordinata è possibile calcolare il primo e il terzo quartile; tenendo conto che n è dispari si ha:

posiz. Q1 = (n+1)/4 = 10 / 4 = 2.5

posiz. Q3 = 3 (n+1)/4 = 30 / 4 = 7.5

A questo punto è possibile calcolare i quartili tenendo conto delle seguenti regole:

- Se il valore risultante dal calcolo (posizione del quartile) è un numero intero allora non si hanno problemi nella scelta
- Nel caso in cui il valore sia non intero esso sarà del tipo x.5: allora, per il calcolo del quartile, si effettua la media tra l'osservazione alla posizione x e quella alla posizione x+1

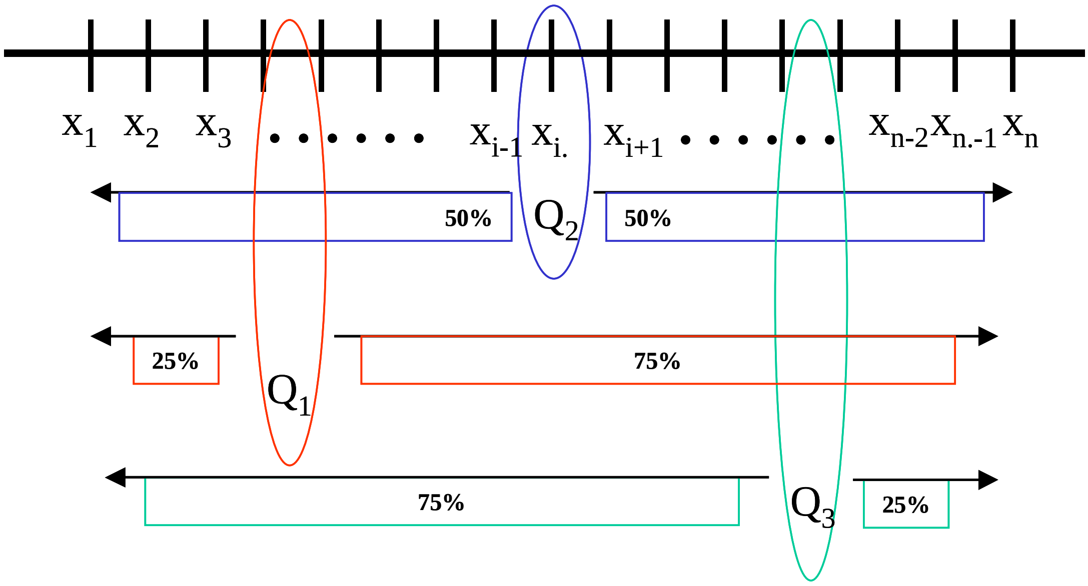
Seguendo tali regole, il primo quartile, nell'esempio, corrisponde alla media tra il secondo e il terzo ricavo nella successione ordinata, ovvero:

Q1 = (x2 + x3) / 2 = 100.7

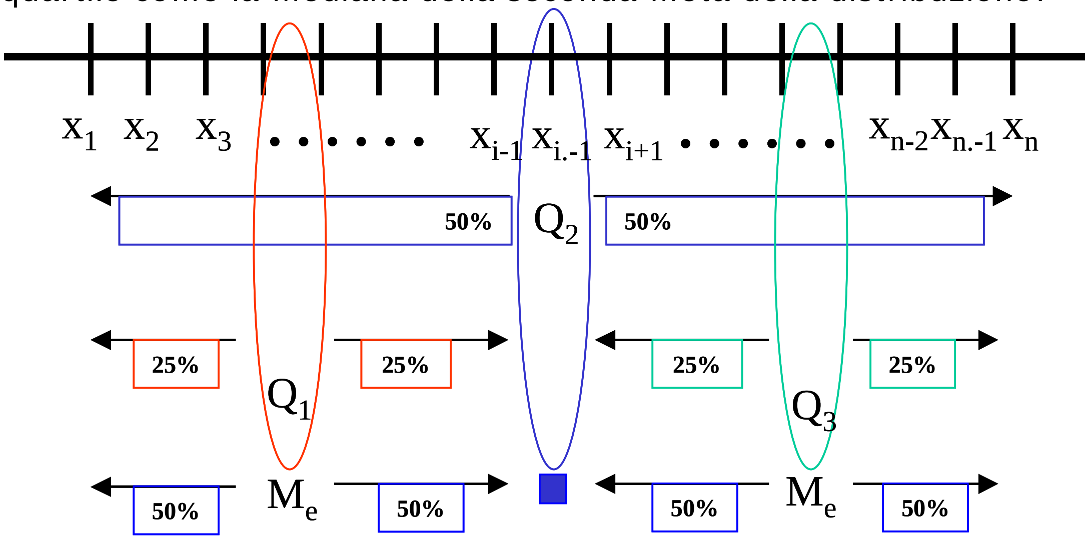
essendo pari a 100.7 entrambe le osservazioni in questione. Il terzo quartile corrisponde invece alla media tra la settima e l'ottava osservazione:

Q3 = (x7 + x8) / 2 = (104.8 + 105.2) / 2 = 105

E' possibile giungere agli stessi risultati per i quartili applicando le regole utilizzate per la mediana considerando separatamente le due metà della successione ordinata. Il primo quartile, infatti, è per definizione quel valore della successione ordinata che lascia a sinistra il 25% e a destra il 75% mentre il terzo quartile è quel valore che lascia a sinistra il 75% e a destra il 25%.



Dalla figura sopra risulta evidente come il primo quartile possa essere interpretato come la mediana della prima metà della distribuzione e il terzo quartile come la mediana della seconda metà della distribuzione:



Nell'applicare questa seconda modalità di calcolo bisogna tenere conto dei due casi seguenti:

- Se il numero delle osservazioni è pari è possibile dividere perfettamente la successione ordinata in due distribuzioni di pari numerosità su cui applicare (separatamente) le regole per il calcolo della mediana in modo da ottenere il primo ed il terzo quartile,
- Se il numero delle osservazioni è dispari, nel dividere in due la successione ordinata, al fine di ottenere due successioni di pari numerosità non si considera l'osservazione mediana.

Nel caso dell'esempio in questione la numerosità è dispari: si può quindi dividere la successione ordinata in due sottodistribuzioni di cui la prima contiene le prime quattro osservazioni e la seconda le ultime quattro osservazioni tralasciando quindi la quinta osservazione (osservazione mediana). Calcolando la mediana delle due successioni ordinate così ottenute si ottengono gli stessi risultati ottenuti in precedenza applicando direttamente le formule per il calcolo del primo e del terzo quartile.

b)

Il calcolo della media con l'inserimento del dato relativo all'anno 2000 richiede nuovamente di calcolare il totale dei ricavi annui:

$$\mu = \frac{\sum_{t=1991}^{2000} x_t}{n} = \frac{925+102.7}{10} = \frac{1027.7}{10} = 102.77$$

E' naturalmente possibile calcolare la nuova media partendo dal valore della media precedente (e dalla numerosità su cui essa è calcolata) e sfruttando una delle proprietà della media aritmetica:

$$\mu = \frac{102.78 \times 9 + 102.7}{10} = \frac{925+102.7}{10} = \frac{1027.7}{10} = 102.77$$

Il calcolo della mediana richiede di verificare la posizione assunta dal dato relativo all'anno 2000 nella successione ordinata delle intensità del carattere:

Posiz.	1	2	3	4	5	6	7	8	9	10
Anni	*1998*	*1993*	*1997*	*1999*	*1991*	*2000*	*1996*	*1995*	*1994*	*1992*
Ricavi	100.2	100.7	100.7	100.9	102.3	102.7	104.6	104.8	105.2	105.6

L'aggiunta del nuovo dato comporta naturalmente la variazione della numerosità di riferimento: in questo caso n diventa pari ed è quindi possibile ripartire la distribuzione esattamente in due parti (i primi 5 valori e i restanti 5 della successione ordinata). Vengono quindi individuate due modalità mediane corrispondenti, rispettivamente, all'estremo superiore della prima distribuzione (il valore che occupa il posto 5) e all'estremo inferiore della seconda distribuzione (il valore che occupa il posto 6). Poiché il carattere in questione è quantitativo è possibile calcolare la mediana come media di tali due intensità, ovvero come:

$$\frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = \frac{102.3+102.7}{2} = 102.5$$

Sulla successione ordinata è possibile calcolare il primo e il terzo quartile utilizzando uno dei due procedimenti sopra illustrati.

Partendo dalle formule per il calcolo diretto dei quartili per n pari si ha:

posiz. Q1 (prima modalità)= n/4 = 10 / 4 = 2.5

posiz. Q1 (seconda modalità)= n/4 + 1 = 10 / 4 + 1 = 3.5

posiz. Q3 (prima modalità) = 3 n / 4 = 30 / 4 = 7.5

posiz. Q3 (seconda modalità) = 3 n / 4 + 1 = 30 / 4 + 1= 8.5

da cui risulta che il primo quartile corrisponde al valore che occupa la terza posizione nella successione ordinata, ovvero a 100.7, mentre il terzo quartile corrisponde al valore che occupa l'ottava posizione nella successione ordinata, ovvero a 104.8. A differenza del caso di n dispari, per n pari si fa riferimento a due osservazioni per ciascun quartile: in realtà, come risulta dell'esempio, le due osservazioni non sempre sono osservazioni reali. Non esiste infatti nessuna osservazione di posizione 2.5 e 3.5 (nel caso del primo quartile) e analogamente nessuna osservazione di posizione 7.5 e 8.5 (nel caso del terzo quartile). Ci si imbatterà in situazioni del genere ogni volta che il risultato della divisione per due della numerosità della successione sia un numero dispari, come risulta evidente dall'esame della seguente tabella che riporta il calcolo delle due modalità del primo e del terzo quartile per differenti valori di n:

Numerosità (pari)	Q1 (prima modalità)	Q1 (seconda modalità)	Q3 (prima modalità)	Q3 (seconda modalità)
4	1	2	3	4
6	1.5	2.5	4.5	5.5
8	2	3	6	7
10	2.5	3.5	7.5	8.5
12	3	4	9	10
14	3.5	4.5	10.5	11.5
16	4	5	12	13
18	4.5	5.5	13.5	14.5
20	5	6	15	16
22	5.5	6.5	16.5	17.5
24	6	7	18	19
26	6.5	7.5	19.5	20.5
28	7	8	21	22
30	7.5	8.5	22.5	23.5
32	8	9	24	25
34	8.5	9.5	25.5	26.5
36	9	10	27	28
38	9.5	10.5	28.5	29.5
40	10	11	30	31
...	...	...	...	...

Nel caso di n pari è quindi possibile elencare le due seguenti regole:

- se le due osservazioni da considerare corrispondono a due posizioni reali allora si considera la media di tali osservazioni
- se le posizioni risultanti dal calcolo sono del tipo x.5 e (x+1).5 allora si considera la posizione centrale tra le due posizioni ottenute, ovvero x+1, ottenendo così direttamente una singola posizione reale che corrisponde al quartile di interesse.

Nel caso dell'esempio la regola da applicare è la seconda: il primo quartile corrisponde alla 3^ osservazione (100.7) mentre il terzo quartile corrisponde alla 8^ osservazione (104.8).
E' immediato verificare che si ottengono gli stessi risultati dividendo la successione ordinata in due sottodistribuzioni (la prima contenente le prime 5 osservazioni e la seconda le restanti 5) ed andando a calcolare la mediana di ciascuna di esse.

c)
L'inserimento dei valori relativi agli anni 2001 e 2002 viene di seguito considerato aggiungendo un dato per volta, in modo da valutare l'incidenza dello stesso sui valori di media e quartili.
In particolare, inserendo il valore relativo all'anno 2001 e seguendo il procedimento riportato al punto precedente, la media risulta pari a **102.10**.
L'aggiunta del dato relativo al 2002 comporta un'ulteriore variazione nella media, che assume il valore di **103.65**
Per quanto riguarda i quartili, invece, l'inserimento del valore del ricavo del 2001 comporta la seguente successione ordinata:

Posiz.	1	2	3	4	5	6	7	8	9	10	11
Anni	2001	1998	1993	1997	1999	1991	2000	1996	1995	1994	1992
Ricavi	95.4	100.2	100.7	100.7	100.9	102.3	102.7	104.6	104.8	105.2	105.6

Essendo n dispari la mediana è facilmente individuabile come valore centrale della successione ordinata, ovvero come il valore 102.3, che occupa la posizione 6 in tale successione.
Sulla successione ordinata è possibile calcolare il primo e il terzo quartile; tenendo conto che la numerosità è dispari si ha:
posiz. Q1 = (n+1)/4 = 12 / 4 = 3 → Q1 = 100.7
posiz. Q3 = 3 (n+1)/4 = 36 / 4 = 9 → Q3 = 104.8
Gli stessi risultati si ottengono dividendo in due parti la successione ordinata tralasciando l'osservazione mediana (essendo n dispari) e calcolando la mediana delle prime 5 osservazioni e delle ultime cinque osservazioni.

Se si considera anche il dato relativo all'anno 2002 si ha la seguente successione ordinata:

Posiz.	1	2	3	4	5	6	7	8	9	10	11	12
Anni	2001	1998	1993	1997	1999	1991	2000	1996	1995	1994	1992	2002
Ricavi	95.4	100.2	100.7	100.7	100.9	102.3	102.7	104.6	104.8	105.2	105.6	120.7

da cui si ricava che la mediana può essere calcolata come media della 6^ (102.3) e della 7^ intensità (102.7), ovvero risulta pari a 102.5.
Sulla successione ordinata è possibile calcolare il primo e il terzo quartile; tenendo conto che la numerosità è pari si ha:
posiz. Q1 (prima modalità)= n/4 = 12 / 4 = 3
posiz. Q1 (seconda modalità)= n/4 + 1 = 12 / 4 + 1 = 4
Q1 = (x3 + x4) / 2 = 100.7

posiz. Q3 (prima modalità) = 3 n / 4 = 36 / 4 = 9
posiz. Q3 (seconda modalità) = 3 n / 4 + 1 = 36 / 4 + 1= 10
Q3 = (x9 + x10) / 2 = 104.8 + 105.2) / 2 = 105
Anche qui è facile verificare come si ottengono gli stessi risultati dividendo la successione ordinata in due distribuzioni, contenenti, rispettivamente, le prime 6 osservazioni e le restanti sei osservazioni, ed andando a calcolare la mediana.

b e c)
I risultati ottenuti in precedenza sono riassunti nella seguente tabella:

				I	II	III
Media (1991-1999)	102.78		Quartile (1991-1999)	100.7	102.3	105
Media (1991-2000)	102.77		Quartile (1991-2000)	100.7	102.5	104.8
Media (1991-2001)	102.10		Quartile (1991-2001)	100.7	102.3	104.8
Media (1991-2002)	103.65		Quartile (1991-2002)	100.7	102.5	105

Il confronto tra le variazioni subite dalla media e dalla mediana in seguito all'inserimento di nuovi dati può essere interpretato alla luce delle variazioni nella successione ordinata delle intensità del carattere, riportate nuovamente di seguito:

Dati 1991- 1999

Posiz.	1	2	3	4	5	6	7	8	9
Anni	1998	1993	1997	1999	1991	1996	1995	1994	1992
Ricavi	100.2	100.7	100.7	100.9	102.3	104.6	104.8	105.2	105.6

Dati 1991- 2000

Posiz.	1	2	3	4	5	6	7	8	9	10
Anni	1998	1993	1997	1999	1991	2000	1996	1995	1994	1992
Ricavi	100.2	100.7	100.7	100.9	102.3	102.7	104.6	104.8	105.2	105.6

Dati 1991- 2001

Posiz.	1	2	3	4	5	6	7	8	9	10	11
Anni	2001	1998	1993	1997	1999	1991	2000	1996	1995	1994	1992
Ricavi	95.4	100.2	100.7	100.7	100.9	102.3	102.7	104.6	104.8	105.2	105.6

Dati 1991- 2002

Posiz.	1	2	3	4	5	6	7	8	9	10	11	12
Anni	2001	1998	1993	1997	1999	1991	2000	1996	1995	1994	1992	2002
Ricavi	95.4	100.2	100.7	100.7	100.9	102.3	102.7	104.6	104.8	105.2	105.6	120.7

Come si nota il dato relativo al 2000 si inserisce nella parte centrale della distribuzione ed è praticamente uguale alla media: questo comporta una variazione praticamente nulla nella media e una variazione maggiore nella mediana.
I valori relativi agli anni 2001 e 2002, invece, risultano valori estremi nella successione ordinata: questo causa variazioni maggiori nella media rispetto alla mediana, confermando quanto noto riguardo alla robustezza dei due indici di posizione.

d)
Per calcolare la media dei valori in £ è possibile procedere trasformando prima i valori in € nei corrispondenti valori in £ (moltiplicando per il fattore fisso di cambio 1936.27) ottenendo i valori riportati nella seguente tabella:

Anni	Ricavi €	Ricavi £
1991	102.3	198'080.42
1992	105.6	204'470.11
1993	100.7	194'982.39
1994	105.2	203'695.60
1995	104.8	202'921.10
1996	104.6	202'533.84
1997	100.7	194'982.39
1998	100.2	194'014.25
1999	100.9	195'369.64
2000	102.7	198'854.93
2001	95.4	184'720.16
2002	120.7	233'707.79

A questo punto la media dei valori in £ è facilmente calcolabile:
media ricavi in £ = somma ricavi in £ / nr. Osservazioni =
= 2'408'332.63 / 12 = 200'694.39
E' possibile però sfruttare la proprietà dell'invarianza di cui gode la media per trasformazioni lineari nei dati. In particolare, essendo:
valori in £ = valori in € * 1936.27

Si ha che:
media in £ = media in € * 1936.27 = 103.65 * 1936.27 = 200'694.39

e1)
La costruzione della distribuzione in 4 classi equiampie richiede innanzitutto la determinazione dell'ampiezza delle classi da costruire. A tal fine è sufficiente determinare il valore minimo della distribuzione (pari a 95.4) e il valore massimo (pari a 120.7) per poi calcolare come segue l'ampiezza da attribuire alle classi:
Ampiezza Classi = (Valore max – Valore min) / nr. Classi =
= (120.7 – 95.4) / 4 = 6.325
classe 1 → da 95.4 a 101.725 (95.4 + 6.325)
classe 2 → da 101.725 a 108.05 (101.725 + 6.325)
classe 3 → da 108.05 a 114.375 (108.05 + 6.325)
classe 4 → da 114.375 a 120.7 (114.375 + 6.325)
E' quindi possibile procedere alla costruzione della distribuzione di frequenza in classi secondo tale suddivisione dei valori del carattere. Tale distribuzione è riportata nella seguente tabella, dove sono calcolate le frequenze assolute e relative, sia semplici che cumulate.

		Frequenza		Frequenza Cumulata	
	Classi Reddito	Assoluta	Relativa	Assoluta	Relativa
1	[95.4, 101.725 [	5	0.417	5	0.417
2	[101.725, 108.05 [	6	0.5	11	0.917
3	[108.05, 114.375 [	0	0	11	0.917
4	[114.375, 120.7]	1	0.083	12	1.000
		12	1.000		

e2)
Nel caso in cui si richiede la distribuzione in classi equiampie a partire da un valore fissato dell'ampiezza delle classi la procedura è ancora più agevole. In tal caso è sufficiente considerare il valore minimo e il valore massimo (quest'ultimo serve in particolare per capire quando si è raggiunta l'ultima classe).
Gli estremi delle classi per un'ampiezza pari a 5 sono:
classe 1 → da 95.4 a 100.4 (95.4 + 5)
classe 2 → da 100.4 a 105.4 (100.4 + 5)
classe 3 → da 105.4 a 110.4 (105.054 + 5)
classe 4 → da 110.4 a 115.4 (110.4 + 5)
classe 5 → da 115.4 a 120.7
In particolare si è preferito considerare la classe 5 leggermente più ampia delle precedenti (arriva a 120.7 anziché a 120.4) in maniera da evitare una sesta classe (da 120.4 a 125.4) praticamente inutile considerando il valore massimo effettivo dei dati. Le differenti distribuzioni di frequenze sono riportate nella seguente tabella:

		Frequenza		Frequenza Cumulata	
	Classi Reddito	Assoluta	Relativa	Assoluta	Relativa
1	[95.4, 100.4 [	2	0.167	2	0.167
2	[100.4, 105.4 [	8	0.667	10	0.833
3	[105.4, 110.4 [	1	0.083	11	0.917
4	[110.4, 115.4 [	0	0	11	0.917
5	[115.4, 120.7]	1	0.083	12	1.000
		12	1.000		

Per la costruzione dell'istogramma è necessario calcolare le densità di frequenza in quanto l'ultima classe è leggermente più estesa delle prime quattro (anche se la differenza risulta poco sensibile come si vede dal confronto con la densità di frequenza relativa alla classe 3 che ha stessa ampiezza e stessa frequenza):

	Classi Reddito	Ampiezza	Frequenza Assoluta	Densità Frequenza
1	[95.4, 100.4 [	5	2	0.400
2	[100.4, 105.4 [	5	8	1.600
3	[105.4, 110.4 [	5	1	0.200
4	[110.4, 115.4 [	5	0	0
5	[115.4, 120.7]	5.3	1	0.189
			12	

e3)

La costruzione in classi equifrequenti di frequenza 4 richiede che si parta dai dati ordinati:

Anni	Ricavi €	
2001	95.4	Classe 1
1998	100.2	
1993	100.7	
1997	100.7	
1999	100.9	Classe 2
1991	102.3	
2000	102.7	
1996	104.6	
1995	104.8	Classe 3
1994	105.2	
1992	105.6	
2002	120.7	

La costruzione della distribuzione in classi risulta agevole ed è mostrata nella seguente tabella. E' da notare che in questo caso è necessario, per evitare errate interpretazione dei dati, procedere al calcolo della densità di frequenza poiché le classi hanno differente ampiezza:

	Classi Reddito	Frequenza	Ampiezza Classe	Densità frequenza
1	[95.4, 100.7]	4	5.3	0.755
2	] 100.7, 104.6]	4	3.7	1.081
3	] 104.6, 120.7]	4	15.9	0.252

f)

Il calcolo della media a partire dalla distribuzione in classi richiede di attribuire un valore singolo a ciascuna classe, valore da utilizzare poi nel calcolo. Solitamente si sceglie come valore centrale della classe la semisomma dei due estremi di classe. Alcuni problemi potrebbero riscontrarsi in relazione alle classi estremi per cui potrebbe non essere agevole calcolare tale valore (si pensi ad esempio alle classi "minore di..." e "maggiore di..." per le quali è necessario introdurre qualche ipotesi ai fini della scelta del valore centrale).

La media aritmetica può essere calcolata indifferentemente a partire dalla distribuzione delle frequenze assolute o relative, come segue:

$$\mu = \frac{\sum_{i=1}^{n} x_i n_i}{n} = \sum_{i=1}^{n} x_i \frac{n_i}{n} = \sum_{i=1}^{n} x_i f_i$$

I calcoli sono riportati di seguito per le tre distribuzione in classi considerate.

	Classi Reddito	Valore Centrale	Frequenza Assoluta	Frequenza Relativa	x_i x n_i	x_i x f_i
1	[95.4, 101.725 [	98.563	5	0.417	492.813	41.068
2	[101.725, 108.05 [	104.888	6	0.5	629.325	52.444
3	[108.05, 114.375 [	111.213	0	0	0	0
4	[114.375, 120.7]	117.538	1	0.083	117.538	9.795
			12	1.000	1239.675	103.306

	Classi Reddito	Valore Centrale	Frequenza Assoluta	Frequenza Relativa	x_i x n_i	x_i x f_i
1	[95.4, 100.4 [	97.9	2	0.167	195.800	16.317
2	[100.4, 105.4 [	102.9	8	0.667	823.200	68.600
3	[105.4, 110.4 [	107.9	1	0.083	107.900	8.992
4	[110.4, 115.4 [	112.9	0	0	0.000	0.000
5	[115.4, 120.7]	118.05	1	0.083	118.050	9.838
			12	1.000	1244.950	103.746

			Frequenza			
	Classi Reddito	Valore Centrale	Assoluta	Relativa	x_i x n_i	x_i x f_i
1	[95.4, 100.7]	98.05	4	0.333	392.200	32.683
2	] 100.7, 104.6]	102.65	4	0.333	410.600	34.217
3	] 104.6, 120.7]	112.65	4	0.333	450.600	37.550
			12	1.000	1253.400	104.450

Per il calcolo della mediana è necessario ricorrere alle frequenze cumulate. Per la prima categorizzazione usata si ha:

		Frequenza		Frequenza Cumulata	
	Classi Reddito	Assoluta	Relativa	Assoluta	Relativa
1	[95.4, 101.725 [	5	0.417	5	0.417
2	[101.725, 108.05 [	6	0.5	11	0.917
3	[108.05, 114.375 [	0	0	11	0.917
4	[114.375, 120.7]	1	0.083	12	1.000
		12	1.000		

Utilizzando la frequenza relativa cumulata (funzione di ripartizione empirica), la mediana corrisponde a quel valore x per cui la funzione di ripartizione empirica F(x)=0.5. La classe mediana è quindi facilmente individuabile nella seconda classe in quanto il valore della funzione di ripartizione con riferimento alla prima classe è 0.417 mentre con riferimento alla seconda classe questa è pari a 0.917.

Al fine di determinare un valore approssimato per la mediana è possibile ricorrere al seguente metodo di interpolazione lineare.

Sia data la seguente notazione:

X_{me-1} - X_{me} → classe mediana

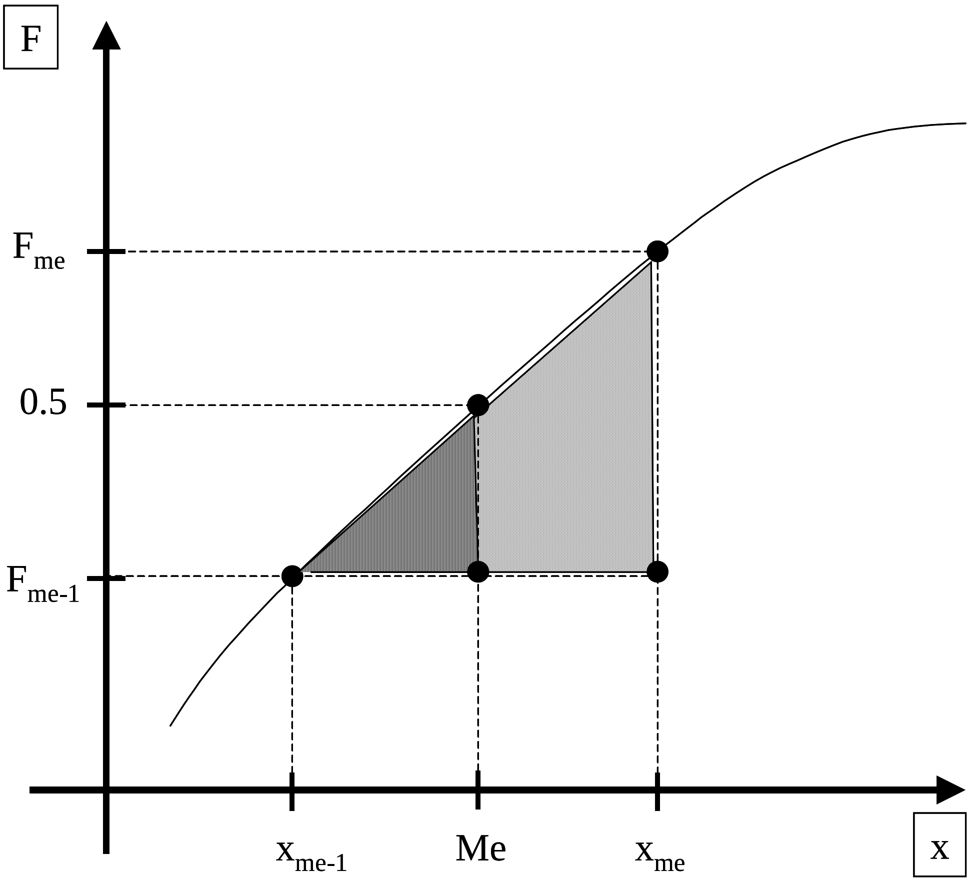
X_{me-1} → estremo inferiore della classe mediana

X_{me} → estremo superiore della classe mediana

F_{me} → valore della funzione di ripartizione empirica in corrispondenza della classe mediana

F_{me-1} → valore della funzione di ripartizione empirica in corrispondenza della classe che precede la classe mediana

Il seguente grafico riporta tali grandezze nell'ipotesi di equispaziatura dei dati all'interno della classe.



Poiché i due triangoli evidenziati in figura sono simili si ha:

$$\frac{(M_e - x_{me-1})}{(0.5 - F_{me-1})} = \frac{(x_{me} - x_{me-1})}{(F_{me} - F_{me-1})} \quad \Rightarrow \quad M_e = x_{me-1} + \frac{(x_{me} - x_{me-1})(0.5 - F_{me-1})}{(F_{me} - F_{me-1})}$$

Nel caso dell'esempio in questione si ha:

$$M_e = x_{me-1} + \frac{(x_{me} - x_{me-1})(0.5 - F_{me-1})}{(F_{me} - F_{me-1})} = 101.725 + \frac{(108.05 - 101.725)(0.5 - 0.417)}{(0.917 - 0.417)} = 102.9$$

Per la seconda categorizzazione si ha:

		Frequenza		**Frequenza Cumulata**	
	Classi Reddito	**Assoluta**	**Relativa**	**Assoluta**	**Relativa**
1	[95.4, 100.4 [	2	0.167	2	0.167
2	[100.4, 105.4 [	8	0.667	10	0.833
3	[105.4, 110.4 [	1	0.083	11	0.917
4	[110.4, 115.4 [	0	0	11	0.917
5	[115.4, 120.7 [	1	0.083	12	1.000
		12	1.000		

La classe mediana, anche in questo caso, corrisponde alla seconda classe, in quanto il valore della funzione di ripartizione con riferimento alla prima classe è 0.167 mentre con riferimento alla seconda classe è pari a 0.833.

$$M_e = x_{me-1} + \frac{(x_{me} - x_{me-1})(0.5 - F_{me-1})}{(F_{me} - F_{me-1})} = 100.4 + \frac{(105.4 - 100.4)(0.5 - 0.167)}{(0.833 - 0.167)} = 102.8$$

Gli stessi calcoli, infine, sono riportati per la terza categorizzazione utilizzata:

		Frequenza		Frequenza Cumulata	
	Classi Reddito	Assoluta	Relativa	Assoluta	Relativa
1	[95.4, 100.7]	4	0.333	4	0.333
2	] 100.7, 104.6]	4	0.333	8	0.667
3	] 104.6, 120.7]	4	0.333	12	1.000
		12			

La classe mediana è nuovamente la seconda classe (il valore della funzione di ripartizione con riferimento alla prima classe è 0.333 mentre con riferimento alla seconda classe è pari a 0.667).

$$M_e = x_{me-1} + \frac{(x_{me} - x_{me-1})(0.5 - F_{me-1})}{(F_{me} - F_{me-1})} = 100.7 + \frac{(104.6 - 100.7)(0.5 - 0.333)}{(0.667 - 0.333)} = 102.7$$

Si può notare come i valori di media e mediana siano differenti a seconda della categorizzazione utilizzata per la variabile e come tali valori rappresentino comunque un'approssimazione rispetto ai valori assunti da tali indici qualora questi vengano calcolati a partire dai dati originari.

g1)

Per la determinazione del quinto e del 95-esimo percentile si procede in maniera analoga a quanto visto per il calcolo della mediana. La distribuzione in classi su cui effettuare il calcolo è quella in classi equifrequenti ed è riportata nuovamente di seguito:

		Frequenza		Frequenza Cumulata	
	Classi Reddito	Assoluta	Relativa	Assoluta	Relativa
1	[95.4, 100.7]	4	0.333	4	0.333
2	] 100.7, 104.6]	4	0.333	8	0.667
3	] 104.6, 120.7]	4	0.333	12	1.000
		12			

In particolare il quinto percentile corrisponde al valore per cui la funzione di ripartizione empirica è pari a 0.05 mentre il 95-esimo percentile corrisponde al valore per cui la funzione di ripartizione empirica è pari a 0.95. E' possibile applicare nuovamente il metodo di interpolazione illustrato in precedenza per la mediana cambiando opportunamente il valore di interesse per la funzione di ripartizione empirica, come di seguito riportato:

$$P_{0.05} = x_{P_{0.05-1}} + \frac{(x_{P_{0.05}} - x_{P_{0.05}-1})(0.05 - F_{P_{0.05}-1})}{(F_{P_{0.05}} - F_{P_{0.05}-1})} = 95.4 + \frac{(100.7 - 95.4)(0.05 - 0)}{(0.333 - 0)} = 96.2$$

$$P_{0.95} = x_{P_{0.95-1}} + \frac{(x_{P_{0.95}} - x_{P_{0.95}-1})(0.95 - F_{P_{0.95}-1})}{(F_{P_{0.95}} - F_{P_{0.95}-1})} = 104.6 + \frac{(120.7 - 104.6)(0.95 - 0.667)}{(1 - 0.667)} = 118.3$$

g2)

Il calcolo della percentuale di ricavi tra 100 e 104 milioni richiede i seguenti passi:

1) Calcolo delle densità di frequenze.

La tabella, calcolata già in precedenza, viene di seguito riportata:

	Classi Reddito	Frequenza	Ampiezza Classe	Densità frequenza
1	[95.4, 100.7]	4	5.3	0.755
2	] 100.7, 104.6]	4	3.7	1.081
3	] 104.6, 120.7]	4	15.9	0.252

2) Calcolo della frequenza relativa delle aziende con fatturato tra 100 e 104 milioni.

Tali aziende sono in parte appartenenti alla prima classe (aziende con fatturato tra 100 e 100.7) ed in parte alla seconda classe (aziende con fatturato tra 100.7 e 104).
Con riferimento alle aziende con fatturato tra 100 e 100.7 è possibile calcolare la densità di frequenza corrispondente a tale aziende secondo la seguente proporzione:
0.755 : 5.3 = x : (100.7 – 100) → x = ((100.7 – 100) * 0.755)/5.3=0.100
Per calcolare la frequenza relativa delle aziende comprese tra 100 e 100.7 basta moltiplicare il valore appena ottenuto per l'ampiezza del reddito relativo a tali aziende (vale a dire 100.7-100):
0.100*(100.7-100)=0.07

Per le aziende comprese tra 100.7 e 104 si procede invece come di seguito illustrato:
1.081 : 3.7 = x : (104 – 100.7) → x = ((104 – 100.7) * 1.081)/3.7=0.964
Per calcolare la frequenza relativa delle aziende comprese tra 100.7 e 104 basta moltiplicare il valore appena ottenuto per l'ampiezza del reddito relativo a tali aziende (vale a dire 104-100.7):
0.964*(104-100.7)=3.1812

3) Calcolo del numero complessivo di aziende nell'intervallo di interesse.

Sommando adesso le due frequenze relative si ottiene:
0.07 + 3.1812 = 3.2512
che arrotondato per difetto alla prima cifra intera indica che il numero di aziende tra 100 e 104 è pari a 3

h)

Per quanto riguarda il calcolo della moda a partire dalla distribuzione originaria, essa assume scarso significato. Essendo il carattere quantitativo continuo, infatti, è facile imbattersi in una serie costituita da valori tutti differenti. Nel caso dell'esempio fa eccezione il valore 100.7 che si ripete per gli anni 1993 e 1997, e che potrebbe, stando alla definizione, essere considerato come valore modale o moda.
Tale indice è più informativo a partire dalla distribuzione di frequenza in classi. Una volta costruita tale distribuzione basta semplicemente controllare quale/quali classe presenti il valore più elevato per la frequenza.
Nel caso della distribuzione in classi equiampie a partire dal numero delle classi (punto e1) la classe modale è la seconda classe (da 101.725 a 108.05) che presenta frequenza pari a 6.

Se invece si considera la distribuzione in classi equiampie a partire da un'ampiezza data da attribuire alle classi (punto e2), la classe modale deve essere stabilita tenendo conto della densità di frequenza piuttosto che della frequenza assoluta. In tale caso la classe modale è la seconda classe (va da 100.4 a 105.4) che presenta la densità di frequenza più alta e pari a 1.6.
Un analogo discorso circa la necessità di considerare le densità di frequenza va fatta per il caso di classi equifrequenti (punto e3): la classe modale, per tale distribuzione è ancora la seconda (che va però da 100.9 a 104.6) la cui densità di frequenza è pari a 1.081.