Analisi dei dati quantitativi

Misure di posizione

L'analisi dei dati quantitativi

Consideriamo i seguenti dati:

dipendenti	salari settimanali
Rossi	250
Bianchi	320
Riccio	340
Porta	360
Esposti	400
Romano	350
Marzucca	300
Melito	260
Corsa	365
	totale

Guardando la tabella è subito chiaro che Rossi e Melito percepiscono lo stesso salario…così come Riccio e Corsa

È evidente che i salari vanno da un minimo di 250 ad un massimo di 400

Ma per fornire una rappresentazione più sintetica e completa di quanto descritto in tabella sarebbe più opportuno avere un unico numero che sintetizzi tutte le informazioni… ad esempio, il salario medio!!!

Media Aritmetica

dipendenti	salari settimanali
Rossi	250
Bianchi	320
Riccio	340
Porta	360
Esposti	400
Romano	350
Marzucca	300
Melito	260
Corsa	365
totale	2945

Per calcolare la media di un insieme di osservazioni è necessario dividere la somma dei valori osservati(modalità) per il numero totale di osservazioni

N.B. n è anche la somma delle frequenze

$$media = \frac{250+320+340+360+400+350+300+260+365}{9} = 327.22$$

È possibile svolgere i calcoli direttamente in tabella:

dipendenti	frequenze	salari settimanali
Rossi	1	250
Bianchi	1	320
Riccio	1	340
Porta	1	360
Esposti	1	400
Romano	1	350
Marzucca	1	300
Melito	1	260
Corsa	1	365
totale	9	2945
media		327.22

Il totale delle frequenze si indica con "n"

Indicando con X_1 il primo salario (250), con X_2 il secondo salario (260) e così via fino a X_9 = 400...la colonna generica del salario i-mo sarà intestata con il simbolo **Xi**

dipendenti	salari settimanali Xi	
Rossi	X_1=	250
Melito	X_2=	260
Marzucca	X_3=	300
Bianchi	X_4=	320
Riccio	X_5=	340
Romano	X_6=	350
Porta	X_7=	360
Corsa	X_8=	365
Esposti	X_9=	400

La media in simboli è,dunque, uguale a:

$$\boxed{\text{media} = \frac{X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9}{9}}$$

Una somma di più termini è anche definita SOMMATORIA e viene indicata con il simbolo Σ (sigma)

Pertanto:

$$\text{media} = \frac{X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9}{9} = \frac{\sum_{i=1}^{9} X_i}{9}$$

La media viene indicata con il seguente simbolo:

$$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$$

E SI LEGGE…

Sommatoria per i che va da 1 ad n di X_i diviso n

Ipotizzando, invece, i dati relativi a più di nove dipendenti e cioè a 12 dipendenti, quale dovrebbe essere il calcolo della media se la tabella fosse la seguente???

x_i=salari settimanali	n_i=dipendenti
250	3
320	2
340	1
360	2
400	4
Totale	12

$$\text{media} = \frac{250+250+250+320+320+340+360+360+400+400+400+400}{12} = 337.5$$

Che equivale alla seguente formulazione:

$$\boxed{\text{media} = \frac{(250\times 3)+(320\times 2)+(340\times 1)+(360\times 2)+(400\times 4)}{12}}$$

Quando si è in presenza di frequenze diverse da quelle unitarie, la media aritmetica si dice "ponderata" (ogni modalità viene pesata per il numero di volte con cui si presenta)

x_i=salari settimanali	n_i=dipendenti	x_i*n_i
250	3	750
320	2	640
340	1	340
360	2	720
400	4	1600
Totale	12	4050
media		337.5

In simboli, si ha:

$$\overline{X} = \frac{(X_1 \times n_1) + (X_2 \times n_2) + (X_3 \times n_4) + (X_4 \times n_4) + (X_5 \times n_5)}{N} = \frac{\sum_{i=1}^{s} X_i n_i}{N}$$

N.B.

PERTANTO, LA MEDIA ARITMETICA PONDERATA VIENE ESPRESSA NEL MODO SEGUENTE:

$$\boxed{\bar{X} = \frac{\sum_{i=1}^{s} X_i n_i}{N}}$$

E SI LEGGE…

Sommatoria per i che va da 1 ad n del prodotto di X_i per n_i diviso n

Limite nell'utilizzo della media aritmetica

La media aritmetica soffre però di un limite: è influenzata dai <u>valori estremi (troppo piccoli o troppo grandi!!!)</u>

dipendenti	frequenze	salari settimanali
Rossi	1	250
Bianchi	1	320
Riccio	1	340
Porta	1	360
Esposti	1	400
Romano	1	350
Marzucca	1	1000
Melito	1	260
Corsa	1	365
totale	9	3645
media		405.00

Supponiamo che Marzucca guadagni 1000 anziché 300

Cosa succede al valore della media aritmetica?

Aumenta considerevolmente da 327 a 405

Solo Esposti presentava in origine un salario così elevato (pari a 400)

Ora, invece, questo valore rappresenta il salario medio

la media aritmetica può essere influenzata dalla presenza di valori molto grandi o molto

Mediana

La mediana rappresenta il valore centrale di una distribuzione ordinata di dati e non risulta influenzata da valori estremi. Quando la distribuzione presenta valori anomali (troppo piccoli o troppo grandi!), è, pertanto, preferibile utilizzare la mediana in alternativa alla media aritmetica.

dipendenti	frequenze	salari settimanali
Rossi	1	250
Bianchi	1	320
Riccio	1	340
Porta	1	360
Esposti	1	400
Romano	1	350
Marzucca	1	1000
Melito	1	260
Corsa	1	365

È necessario pertanto ordinare i dati

dipendenti	salari settimanali
Rossi	250
Melito	260
Bianchi	320
Riccio	340
Romano	350
Porta	360
Corsa	365
Esposti	400
Marzucca	1000

4 prima

4 dopo

La mediana è 350 euro e ricade nella 5° posizione

La mediana

- La mediana è il valore centrale di un insieme di dati ordinato (in senso crescente o decrescente). Poiché le modalità devono essere preventivamente ordinate, il calcolo della mediana è possibile solo su caratteri quantitativi e qualitativi ordinali

La mediana, indicata con *Me*, precede il 50% dei dati più piccoli ed è seguita dal 50% dei dati più grandi.

- Non è influenzata da dati anomali (se si sostituisse il valore 1000 a 300 la mediana non ne risulterebbe influenzata, il valore sarebbe sempre uguale a 350)

- Per calcolarla è opportuno distinguere tra numero di osservazioni, *"N"* , :

  1. Dispari

  2. Pari

Mediana per n pari

Supponiamo di aver intervistato un altro dipendente (ROCCA, 410€), il numero totale "N" di frequenze è 10 (intervistati):

dipendenti	salari settimanali
Rossi	250
Melito	260
Bianchi	320
Riccio	340
Romano	350
Porta	360
Corsa	365
Esposti	400
Rocca	410
Marzucca	1000

4 prima

5 dopo

La mediana non è più 350 euro

…ma …

Mediana per n pari

Supponiamo di avere un dipendente in più intervistato, in totale ne sono 10:

dipendenti	salari settimanali
Rossi	250
Melito	260
Bianchi	320
Riccio	340
Romano	350
Porta	360
Corsa	365
Esposti	400
Rocca	410
Marzucca	1000

4 prima

4 dopo

Ma sta al centro tra 350 e 360!!! Sta pertanto tra la 5° e la 6° posizione

Esiste una regola allora!!!

Se n è pari:

la mediana si colloca al centro TRA le due posizioni (non è mai un'unica posizione!!!) corrispondenti a

$$\boxed{\frac{N}{2} \quad \text{e} \quad \frac{\mathrm{N}}{2}+1}$$

Infatti: $(10)/2 = 5$ e $(10/2)+1 = 5+1 = 6$

LA MEDIANA SI TROVA INFATTI tra la 5° e la 6° POSIZIONE e corrisponde alla media dei valori che occupano tali posizioni:

Me= (350+360)/2 = 355

Se n è dispari:

Nel primo caso (N=9), al contrario, la mediana si colloca nella posizione corrispondente a:

$$\boxed{\frac{N+1}{2}}$$

INFATTI:

$$(9+1)/2 = 10/2 = 5$$

LA MEDIANA SI TROVA INFATTI NELLA 5° POSIZIONE e corrispondeva a:

Me=350

Esempio 1

4 amici universitari registrano la distanza percorsa per raggiungere il proprio posto di lavoro.

Sintetizzare questa informazione calcolando il valore mediano.

È possibile utilizzare il procedimento della mediana per distribuzioni di frequenza considerando le frequenza unitarie???

Tab.: Distanze giornaliere	
Soggetti	Dist.Giorn.
Fabio	250
Laura	90
Mara	12
Luca	18

1 step: ordino	
Soggetti	Dist. Giorn.
Mara	12
Luca	18
Laura	90
Fabio	250

2 step: frequenze cumulate			
Soggetti	Dist. Giorn.	n_i	N_i
Mara	12	1	1
Luca	18	1	2
Laura	90	1	3
Fabio	250	1	4

3 step: posizione mediana

Numerosità pari

N=4

$$posizione_{[Me]} = \frac{N}{2} \; e \; \frac{N}{2}+1$$

Posizione 2 : trovo il valore 18

Posizione 3 : trovo il valore 90

4 step: valore mediano

$$Me = \frac{18+90}{2} = 54$$

Mediana per distribuzioni di frequenza

Se le modalità si ripetono con frequenze diverse (dall'unità)…

…la posizione della mediana va ricercata sulle frequenze cumulate

Xi=costo biglietto dei voli alitalia	ni= n. vendite
100	50
200	20
300	25
400	33
totale	128

N in questo caso è pari (128) per cui la posizione della mediana ricade tra la 64°e 65° posizione…

$$\boxed{\frac{N}{2} \quad e \quad \frac{\mathrm{N}}{2}+1}$$

$$\boxed{\frac{128}{2}=64 \quad e \quad \frac{128}{2}+1=65}$$

…e corrisponde alla media dei valori che occupano tali posizioni…

Mediana per distribuzioni di frequenza

costo biglietto dei voli alitalia	n. vendite	Ni
100	50	50
200	20	70
300	25	95
400	33	128
totale	128	

La mediana è 200

perché tale valore (modalità) occupa sia la 64° che la 65° posizione

Vediamo come…

Mediana per distribuzioni di frequenza

costo biglietto dei voli alitalia	n. vendite	Ni
100	50	50
200	20	70
300	25	95
400	33	128
totale	128	

Si parte dall'osservazione che le prime 50 posizioni sono occupate dalla modalità 100

mentre le posizioni che vanno dalla 51-esima alla 70-esima sono occupate dalla modalità 200

La distribuzione espressa sotto forma di successione è la seguente:

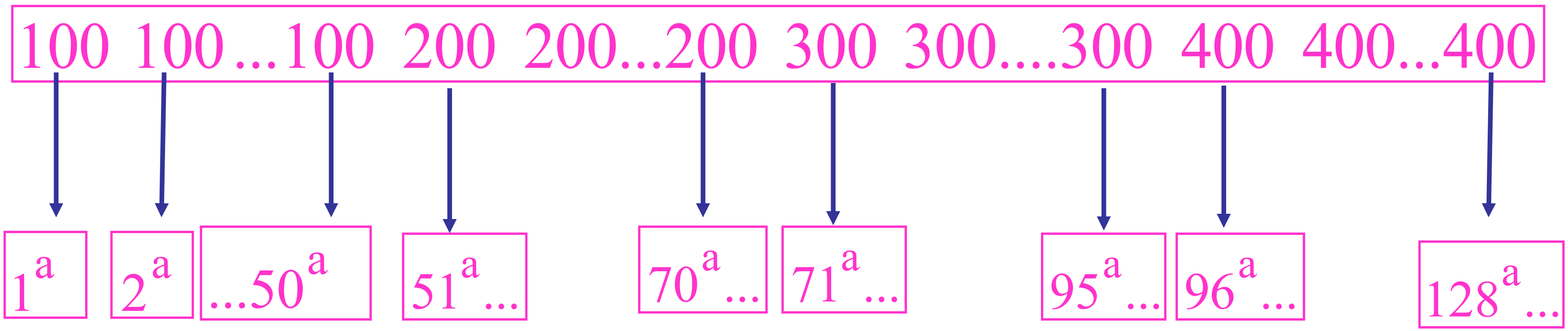
100 100 100 ...(50 volte) 200 200 200 (20 volte) ...300 300

100 occupa pertanto le prime 50 posizioni

200 occupa le posizioni dalla 51-ma alla 70-ma

300 dalla 71-ma alla 95ma

Mediana per distribuzioni di frequenza



Le frequenze cumulate indicano proprio le posizioni delle modalità:

costo biglietto dei voli alitalia	n_i= n. vendite	Ni
100	50	50
200	20	70
300	25	95
400	33	128
totale	128	

Le prime **70** sono occupate dal 100 e dal 200

Esempio 2

La ricerca "Donne più *giovani* domani" ha rilevato la distribuzione del n° di prodotti di bellezza acquistati in 1 mese da 13 donne

n° prodotti: 15, 12, 13, 5, 10, 11, 8, 9, 14, 3, 16, 7, 6

Individuare il n° di prodotti mediano acquistato dal gruppo di donne intervistate

n° prodotti		
15		
12		
13		
5		
10		
11		
8		
9		
14		
3		
16		
7		
6		

	1°step:ordino	
n° prodotti	n° prodotti	
15	3	
12	5	
13	6	
5	7	
10	8	
11	9	
8	10	
9	11	
14	12	
3	13	
16	14	
7	15	
6	16	

	1 step: ordino	2° step: frequenze cumulate	
n° prodotti	n° prodotti	n_i	N_i
15	3	1	1
12	5	1	2
13	6	1	3
5	7	1	4
10	8	1	5
11	9	1	6
8	10	1	7
9	11	1	8
14	12	1	9
3	13	1	10
16	14	1	11
7	15	1	12
6	16	1	13

3° step: posizione mediana

Numerosità dispari
N=13

$$posizione_{[Me]} = \frac{N+1}{2} = \frac{13+1}{2} = 7$$

4° step: valore mediano

$$\mathrm{Me} = 10$$

Esempio 3

La ricerca "Donne più giovani domani" ha rilevato la distribuzione del n° di prodotti di bellezza acquistati in 1 mese da 16 donne

n° prodotti: 3, 3, 5, 5, 5, 5, 6, 6, 6, 6, 10, 10, 10, 10, 10, 10,

Individuare il n° di prodotti mediano acquistato dal gruppo di donne intervistate

Distribuzione di frequenze		1 step: calcolo frequenze cumulate
x_i	n_i	N_i
3	2	2
5	4	6
6	4	10
10	6	16
Totale	16	

Distribuzione di frequenze		**1 step: calcolo frequenze cumulate**	
x_i	n_i	N_i	
3	2	2	da 1 a 2
5	4	6	da 3 a 6
6	4	10	da 7 a 10
10	6	16	da 11 a 16
Totale	16		

2 step: posizione mediana

Numerosità pari

N=16

$$posizione_{[Me]} = \frac{N}{2} \quad e \quad \frac{N}{2}+1$$

Posizione 8 : trovo il valore 6

Posizione 9 : trovo il valore 6

3 step: valore mediano

$$\mathrm{Me} = 6$$

Se la Me si colloca tra due modalità???

Accade solo quando N è pari.

Se la posizione della mediana ricade tra due frequenze cumulate consecutive (è, in sintesi a metà tra due posizioni), si sceglie come valore mediano la media delle due modalità consecutive corrispondenti alle posizioni individuate

Esempio 4

La ricerca "Donne più giovani domani" ha rilevato la distribuzione del n° di prodotti di bellezza acquistati in 1 mese da 16 donne

n° prodotti: 3, 3, 5, 5, 5, 5, 5, 5, 6, 10, 10, 10, 10, 10, 10, 10

Individuare il n° di prodotti mediano del gruppo di donne

Distribuzione di frequenze		1 step: calcolo frequenze cumulate	
x_i	n_i	N_i	
3	2	2	da 1 a 2
5	6	8	da 3 a 8
6	1	9	da 9 a 9
10	7	16	da 10 a 16
Totale	16		

2 step: posizione mediana

Numerosità pari

N=16

$$posizione_{[Me]} = \frac{N}{2} \quad e \quad \frac{N}{2}+1$$

Posizione 8 : trovo il valore 5

Posizione 9 : trovo il valore 6

Distribuzione di frequenze		1 step: calcolo frequenze cumulate	
x_i	n_i	N_i	
3	2	2	da 1 a 2
5	6	8	da 3 a 8
6	1	9	da 9 a 9
10	7	16	da 10 a 16
Totale	16		

2 step: posizione mediana

Numerosità pari

N=16

$$posizione_{[Me]} = \frac{N}{2} \; e \; \frac{N}{2}+1$$

Posizione 8 : trovo il valore 5

Posizione 9 : trovo il valore 6

3 step: valore mediano

$$Me = \frac{5+6}{2} = 5.5$$

Mediana di una distribuzione di frequenza con classi di valori

Classi di superficie (in ettari)	Numero aziende (n_j)	Freq. cumulate (N_j)
0-1	120	120
1-2	160	280
2-3	220	500
3-5	212	712
5-10	205	917
10-20	110	1027
20-40	65	1092
Oltre 40	21	1113

$$posizione\ \ mediana = \frac{n+1}{2} =$$

$$= \frac{1113+1}{2} = 557$$

L'elemento che occupa la posizione 557 è uno dei 212 valori della classe 3-5

La mediana è contenuta nella classe 3-5

Mediana

Ricapitolando…

STEP:

1. Ordinare i dati in senso crescente
2. Calcolare le frequenze cumulate
3. Determinare la posizione della mediana
4. Individuare il valore mediano (o mediana)

La scelta tra media e mediana

FIGURA 30

«Prendiamo la media aritmetica dell'altezza e spaventiamo gli avversari o li vogliamo illudere prendendo il mediano?»

Fonte: Walter Kramer (2009), Le bugie della statistica, Nimesis

Moda

- È la modalità più frequente
- In un insieme di valori: quel termine che si ripete più volte
- In una distribuzione di frequenza: quella modalità che ha la frequenza più alta
- In una distribuzione di frequenza con classi di valori: ogni valore della classe con la più alta densità di frequenza

Moda di un insieme di valori

Punti vendita	Genere respons.
1	maschio
2	maschio
3	femmina
4	femmina
5	maschio
6	maschio
7	maschio
8	femmina
9	femmina

La modalità del carattere "Genere del responsabile" che si ripete più volte (5 volte) è "maschio"

Moda="maschio"

La maggioranza dei punti vendita ha come responsabile un uomo

Moda di una distribuzione di frequenza

Addetti (valori distinti)	Numero punti vendita (frequenze)
3	2
4	1
6	$\boxed{3}$
7	1
10	2

La frequenza maggiore è 3

La modalità del carattere "Numero di addetti" cui è associata la frequenza maggiore è 6

Moda=6

La maggioranza dei punti vendita ha un numero di addetti pari a 6

Moda di una distribuzione di frequenza con classi di valori

Classi di superficie (in ettari)	Numero aziende (n_j)	Ampiezza classe (a_j)	Densità di freq (h_j)
0-1	120	1	120
1-2	160	1	160
2-3	220	1	$\boxed{220}$
3-5	212	2	106
5-10	205	5	41
10-20	110	10	11
20-40	65	20	3,25
40-80	21	40	0,525

In presenza di classi di ampiezza diversa,
la classe modale è quella che ha la densità di frequenza maggiore

La classe modale è 2-3

Moda

- Può non esistere
- Può non essere unica
- Può essere una modalità “poco rappresentativa” del fenomeno

- Per chi vende abbigliamento, la moda rappresenta un parametro utile per decidere in merito a come rifornire il negozio: saranno ordinati più capi delle taglie più diffuse

La scelta tra media, moda e mediana

Consideriamo ora le tre misure della tendenza centrale per calcolare lo stipendio medio di un gruppo di 41 dipendenti di una compagnia.

X = una persona

Numero dei dipendenti	Stipendio	
XX	€ 5000	
XXXXXX	€ 8000	
XXXXXXXX	€ 12 000	— moda: quello riscontrato più frequentemente
XXXX	€ 20 000	
X	€ 25 000	— mediana: quello a metà, con venti dipendenti sopra e venti sotto
XXXX	€ 32 000	
XXX	€ 37 000	
XXXXX	€ 40 000	
XX	€ 47 000	
XXXX	€ 55 000	
X	€ 260 000	
X	€ 450 000	

Il valore medio = **€ 42 000**
Il valore modale (con otto dipendenti) = **€ 12 000**
Il valore mediano = **€ 25 000**

Fonte: Magnello e Van Loon (2011), La statistica a fumetti, Raffaello Cortina Editore

Calcolo dei valori medi in base al tipo di carattere

		Caratteri	
	Quantitativi	Qualitativi ordinati	Qualitativi sconnessi
Media	✓		
Mediana	✓	✓	
Moda	✓	✓	✓

QUARTILI

I QUARTILI SONO COSTITUITI DALLE MODALITA' CHE SUDDIVIDONO LA DISTRIBUZIONE IN 4 PARTI UGUALI

NATURALMENTE, I VALORI CHE SUDDIVIDONO L'INTERA DISTRIBUZIONE IN QUATTRO PARTI SONO 3 :

I QUARTILE Q1

II QUARTILE Q2 pari anche alla Mediana

III QUARTILE Q3

Quartili

Primo Quartile (Q_1)

Suddivide la distribuzione, ordinata in senso crescente, in due parti: la prima alla propria sinistra costituisce 1/4 (25%) dei dati più piccoli e la seconda alla sua destra i restanti 3/4 (75%) dei dati più grandi (*individua il VALORE che separa il primo 25% del collettivo statistico dal restante 75%*)

Secondo Quartile (Q_2) o Me

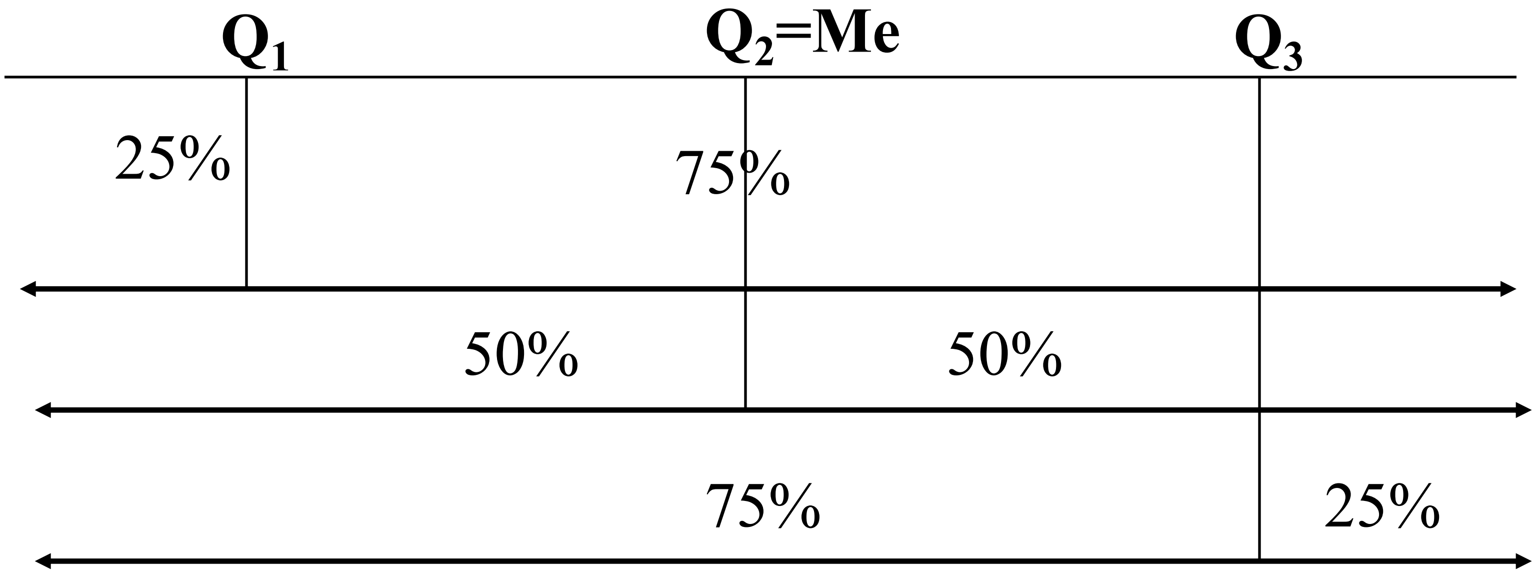
Suddivide le modalità, ordinate in senso crescente, in due parti uguali: si lascia alla sua sinistra il 50% (1/2) dei dati più piccoli e alla sua destra il 50% (1/2) dei dati più grandi (*individua il valore che separa il primo 50% del collettivo statistico dal restante 50%*)

Terzo Quartile (Q_3)

Suddivide le modalità, ordinate in senso crescente, lasciando alla sua sinistra 3/4 (75%) dei termini più piccoli e alla sua destra 1/4 (25%) dei dati più grandi (*individua il valore che separa il primo 75% del collettivo statistico dal restante 25%*)

Quartili

Dati (modalità)ordinati in senso crescente



Primo Quartile(Q_1) distribuzioni di unità e di frequenze

1. Ordinare i dati
2. Calcolare le frequenze cumulate
3. Individuare la posizione di Q_1:

$$posizione_{[Q_1]} = \left(N+1\right) \times 0.25 = \frac{(N+1)}{4}$$

4. Associare il valore della modalità corrispondente

N.B. Q_1 è calcolato anche come il primo valore x_i per il quale $F_j \geq 0{,}25$

Terzo Quartile (Q_3) distribuzioni di unità e di frequenza

1. Ordinare i dati
2. Calcolare le frequenze cumulate
3. Individuare la posizione di Q_3:

$$posizione_{[Q_3]} = (N+1) \times 0.75 = \frac{3}{4} \times (N+1)$$

4. Associare il valore della modalità corrispondente

N.B. Q_3 è calcolato anche come il primo valore x_i in corrispondenza del quale $F_j \geq 0,75$

Esempio 1

	1 step: ordino	2 step: frequenze cumulate	
n° prodotti	n° prodotti	n_i	N_i
15	3	1	1
12	5	1	2
13	6	1	3
5	7	1	4
10	8	1	5
11	9	1	6
8	10	1	7
9	11	1	8
14	12	1	9
3	13	1	10
16	14	1	11
7	15	1	12
6	16	1	13

4 step: valore Q1

$$Q_1 = \frac{6+7}{2} = 6.5$$

3 step: posizione Q1

$$posizione_{[Q1]} = \frac{(N+1)}{4} = \frac{(13+1)}{4} = 3.5$$

Esempio 2

	1 step: ordino	2 step: frequenze cumulate	
n° prodotti	n° prodotti	n_i	N_i
15	3	1	1
12	5	1	2
13	6	1	3
5	7	1	4
10	8	1	5
11	9	1	6
8	10	1	7
9	11	1	8
14	12	1	9
3	13	1	10
16	14	1	11
7	15	1	12
6	16	1	13

4 step: valore Q3

$$Q_3 = \frac{13+14}{2} = 13.5$$

3 step: posizione Q3

$$posizione_{[Q3]} = \frac{3}{4}(N+1) = \frac{3}{4}(13+1) = 10.5$$

Calcolo dei quartili

Ricavi
350
200
600
500
270
180
205
340
280

Ricavi (valori ordinati)	Freq. cum. rel.
$X_{(1)}=180$	1/9=0,11
$X_{(2)}=200$	2/9=0,22
$X_{(3)}=205$	3/9=0,33
$X_{(4)}=270$	4/9=0,44
$X_{(5)}=280$	5/9=0,56
$X_{(6)}=340$	6/9=0,67
$X_{(7)}=350$	7/9=0,78
$X_{(8)}=500$	8/9=0,89
$X_{(9)}=600$	9/9=1

La prima F_i ad essere maggiore o uguale a 0,25 è la terza

$$Q_1 = x_{(3)} = 205$$

Il 25% dei punti vendita con i ricavi più bassi registrano ricavi che non superano 205 mila euro

La prima F_i ad essere maggiore o uguale a 0,75 è la settima

$$Q_3 = x_{(7)} = 350$$

Per essere nel 25% dei punti vendita con i ricavi più alti si devono superare 350 mila euro di ricavi

Se le posizioni di Q_1 e Q_3 sono numeri decimali???

1. Se la posizione calcolata presenta un decimale pari a 5 (esempio: 12.5) ed il quartile si colloca tra due modalità intere consecutive, si sceglie come quartile la loro semisomma (o media)
2. Se la posizione calcolata non corrisponde né a un numero intero (12.0) né ad un numero con decimale pari a 5, vige la regola dell'approssimazione del numero per difetto (<=4) e/o per eccesso (>=5), che fornisce un intero a cui si fa corrispondere la posizione dei quartili (Q1 o Q3)

Percentili

Sono quei valori che dividono la distribuzione in cento parti di uguale numerosità

Mediana=50-esimo percentile

Q3= 75-esimo percentile

P10 = decimo percentile: lascia alla sua sinistra il 10% dei valori

P90 = novantesimo percentile: lascia alla sua destra il 10% dei valori