

nUniversità degli Studi di Napoli «Federico II»
Facoltà di Architettura

Upta

Corso di laurea in Urbanistica e Scienze della Pianificazione Territoriale e Ambientale
AA 2014-2015

Corso integrato di Matematica e statistica

Docente: Livia D'Apuzzo.

Statistica Descrittiva Univariata

Sintesi delle distribuzioni statistiche: indici di tendenza centrale

6. Le medie di posizione o medie lasche (*moda, mediana, quantili*)

Le medie lasche o medie di posizione

Le medie lasche sono gli indici di tendenza centrale nella cui determinazione intervengono solo alcuni valori o modalità del carattere. Esse sono:

- la **moda**, o **norma**, definibile per ogni tipo di carattere,
- la **mediana** e i **quantili** riferibili a caratteri misurabili in una scala ordinale, cioè caratteri quantitativi, per i quali mediana e quantili sono sempre determinabili, o caratteri qualitativi ordinabili, per i quali sono determinabili in alcuni casi: essi sono individuati dalla posizione che occupano in un allineamento crescente delle modalità del carattere.

1 MODA o Norma

La **moda** o **norma** (valore normale nel caso di una variabile statistica) è la **modalità** o **valore** del carattere **che ha la frequenza assoluta maggiore**, e quindi anche la frequenza relativa e la frequenza percentuale maggiore.

La moda non esiste se tutte le modalità hanno la stessa frequenza

- **Esempi di calcolo della moda nel caso di un carattere discreto**

Distribuzione di frequenza della variabile **NUMERO DI LIBRI LETTI**

	N. libri letti	n_i	
A)	1	4	Moda: 1
	2	3	
	3	2	
	4	1	

Distribuzione di frequenza della mutabile **CORPORATURA**

	Corporatura	n_i	n_i/n=pi	
B)	Gracile	4	0,2	Moda: Normale
	Normale	10	0,5	
	Robusta	6	0,3	
	Tot	20	1	

Distribuzione di frequenza della variabile **ORE DI SPORT**

	N.ore sport	n_i	
C)	1	4	Mode: 1 e 3 e 4
	2	3	
	3	4	
	4	4	
	5	2	

• Stima della moda nel caso di un carattere quantitativo con valori raggruppati in classi:

- si determina la classe modale che per definizione è:

“la classe alla quale corrisponde la densità maggiore”.

- individuata la classe modale, in mancanza di ulteriori informazioni, si prende, come stima della moda, il valore centrale della classe modale”.

Esempi di calcolo

1° caso: classi di ampiezze diverse

Distribuzione delle lunghezze in un gruppo di 200 tavoli

lunghezze	f_a	Densità	
1 –2	30	30	_30_
2 –5	70	23.333	_23.3_ _23.3_ _23.3_
5 –9	80	20	_20_ _20_ _20_ _20_
9 –15	20	3.333	
	200		

1|–2 classe modale

1.5 moda

2° caso: classi di uguale ampiezza-

In tal caso la classe modale coincide con la classe a cui corrisponde la frequenza più alta: infatti le densità si ottengono dalle frequenze assolute dividendole tutte per lo stesso numero (l'ampiezza comune a tutte le classi) e quindi sono proporzionali alle frequenze

Distribuzione degli addetti in 280 esercizi commerciali

n. addetti	f_a	δ
0–120	30	3/20
20–140	80	80/20
40–160	110	110/20
60–180	60	60/20
	280	

40–160 classe modale

50 moda

Tipologie di distribuzioni dipendenti dall'esistenza di uno o più mode

Per indicare l'esistenza di una o più mode le distribuzioni di frequenza, si aggettivano come segue:

distribuzione unimodale: distribuzione di frequenze con una sola moda;

distribuzione plurimodale: distribuzione di frequenze con più di una moda;
in particolare:

distribuzione bimodale: distribuzione di frequenze con due mode

Sono distribuzioni unimodali la distribuzione A), la distribuzione B) e la seguente

Censimento 1981 - Valori assoluti e relativi		
Ampiezza delle famiglie (n. di componenti)	FAMIGLIE	
	n.	%
1 componente	3.323.456	17,9
2 componenti	4.402.980	23,6
3 componente	4.177.217	22,1
4 componente	4.008.008	21,5
5 componente	1.773.621	9,5
6 componente	628.719	3,4
7 componente	224.115	1,2
8 e più	154.221	0,8

che ha **moda 2**

E' una distribuzione plurimodale la distribuzione C) che ha tre mode.

Per caratteri quantitativi si definisce anche il seguente tipo di distribuzione:

distribuzione zeromodale: distribuzione in cui la moda corrisponde al valore più piccolo del carattere.

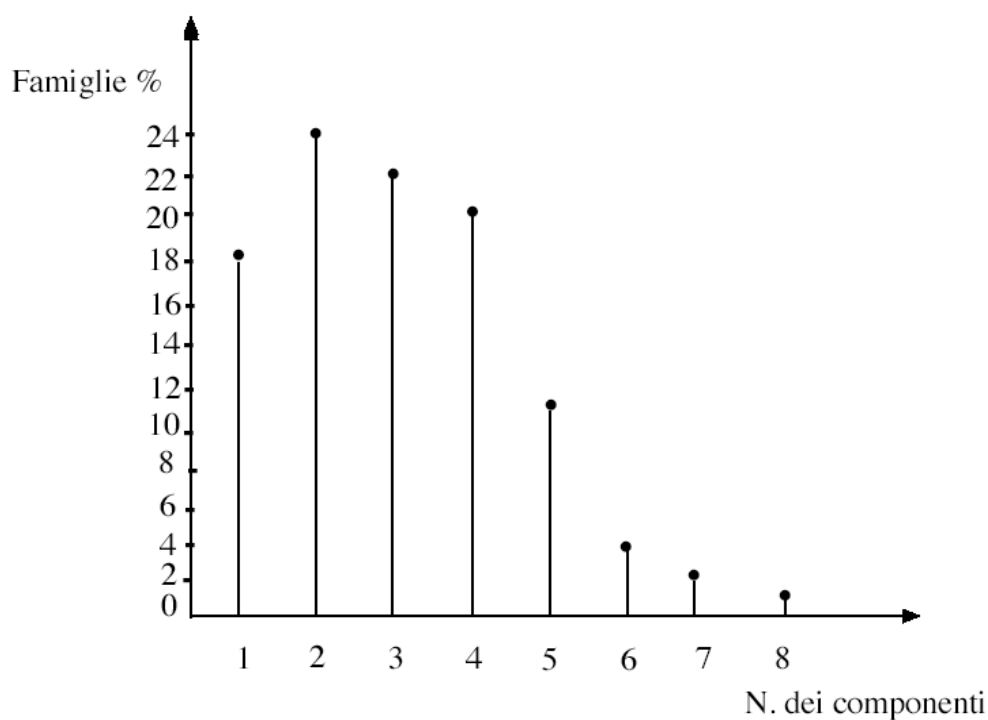
La tabella A) è un esempio di distribuzione sia unimodale sia zeromodale.

1.1 Individuazione della moda attraverso la rappresentazione grafica della distribuzione

- **carattere quantitativo discreto:** *la moda è quel valore cui corrisponde nel grafico cartesiano il punto con ordinata maggiore*

Censimento 1981 - Valori assoluti e relativi

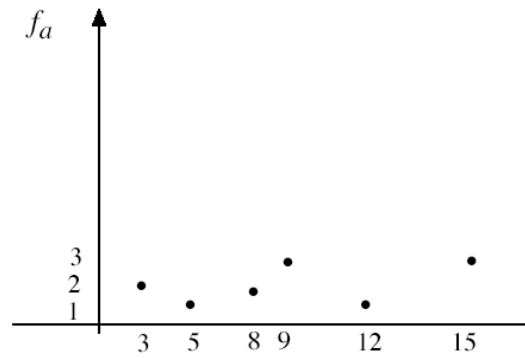
Ampiezza delle famiglie (n. di componenti)	FAMIGLIE	
	n.	%
1 componente	3.323.456	17,9
2 componente	4.402.980	23,6
3 componente	4.177.217	22,1
4 componente	4.008.008	21,5
5 componente	1.773.621	9,5
6 componente	628.719	3,4
7 componente	224.115	1,2
8 e più	154.221	0,8



Distribuzione unimodale.

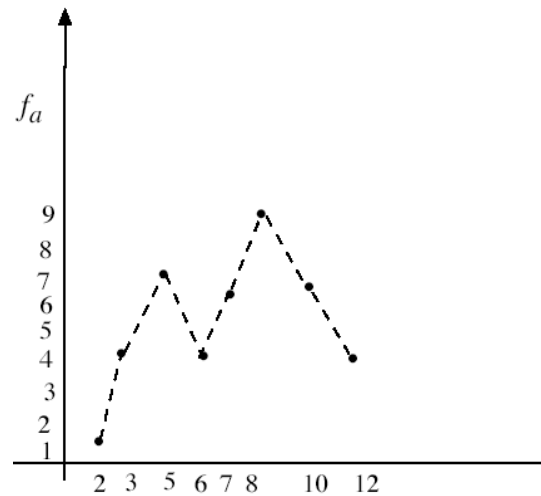
La moda è 2 ed è il valore (sull'asse delle ascisse) a cui nel grafico corrisponde il punto più alto: 2 è il punto di massimo della distribuzione di frequenze

Valori car.	Freq ass.
3	2
5	1
8	2
9	3
12	1
15	3



Distribuzione plurimodale: le moda sono 9 e 15

Valori car.	Freq ass.
2	1
3	4
5	7
6	4
7	6
8	9
10	6
12	4



moda: 8

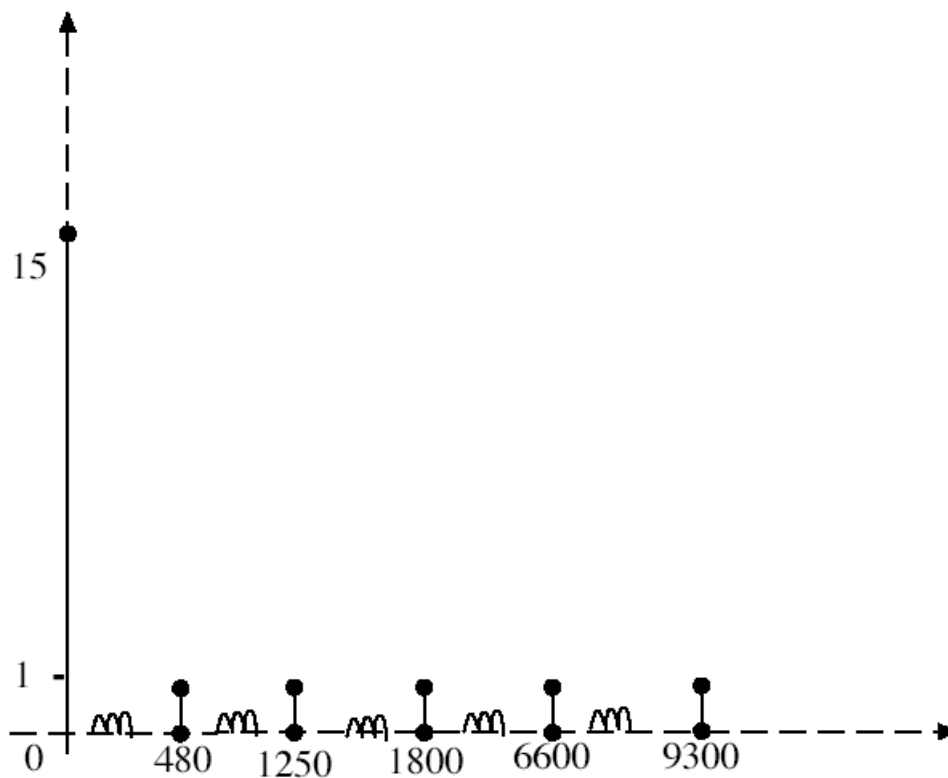
5?

5 non è, per la definizione data di moda, una moda. Tuttavia per la caratteristica che ha di essere un punto di massimo locale o relativo per la distribuzione di frequenze 5 è da alcuni statistici anch'esso chiamato moda.

Tabella sul chilometraggio effettuato in aereo in un anno da venti persone

Modalità= Chilom. ↓	Freq. ass. f_a	freq. rel. f_r	freq. perc. f_{per}
0	15	$15/20=3/4=0,75$	75%
480	1	$1/20 = 0,05$	5%
1250	1	$1/20$	5%
4800	1	$1/20$	5%
6600	1	$1/20$	5%
9300	1	$1/20$	5%

Distribuzione zero modale

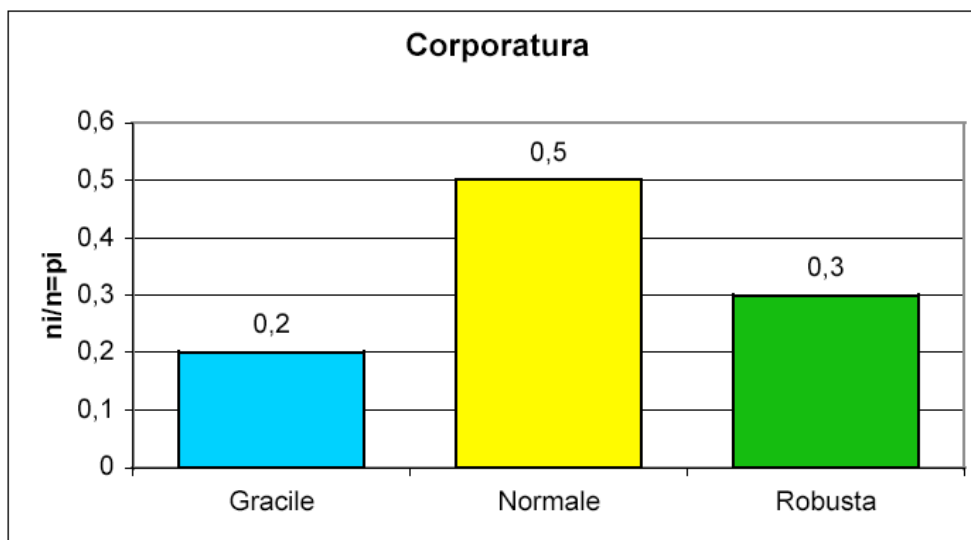


- **carattere qualitativo:** la moda è quella modalità a cui corrisponde nell'ortogramma la colonna più alta e nel diagramma circolare (delle frequenze relative o percentuali) il settore di area maggiore

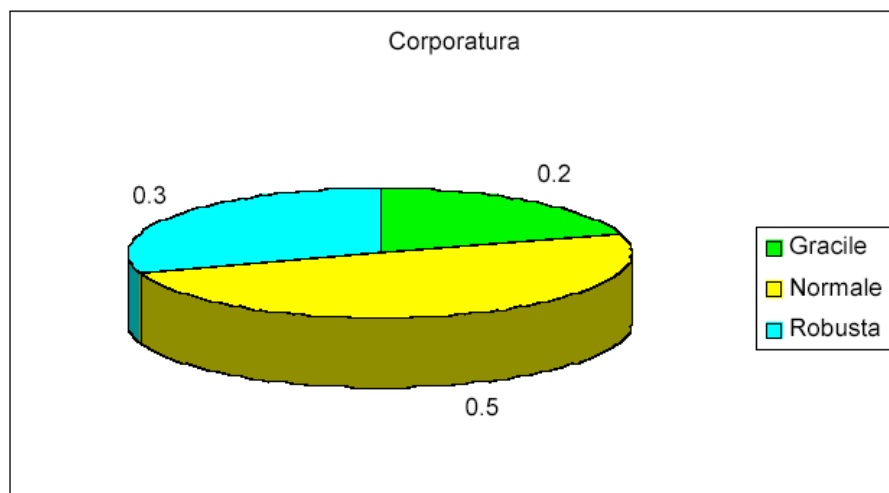
Distribuzione di frequenza della mutabile CORPORATURA (ordinale):

Corporatura	ni	ni/n=pi	Ni
Gracile	4	0,2	4
Normale	10	0,5	14
Robusta	6	0,3	20
Tot	20	1	

Rappresentazione grafica (1): ortogramma con frequenza relative



Rappresentazione grafica (2): torta (con frequenze relative)



Moda: corporatura normale

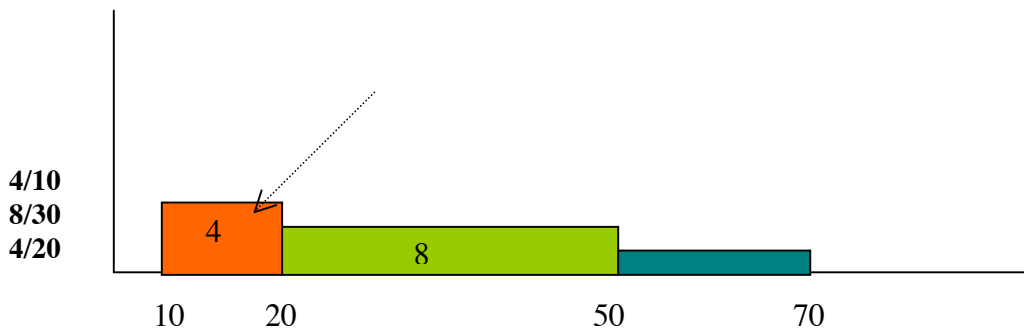
• **carattere quantitativo con valori raggruppati in classi:** *la classe modale è quella cui corrisponde nell'istogramma (costruito attraverso le densità di frequenza) il rettangolo di altezza maggiore.*

Classi di valori	f_a	dens
10 — 20	4	4/10
20 — 50	8	8/30
50 — 70	4	4/20

classe modale: 10 — | 20

moda: 15

16

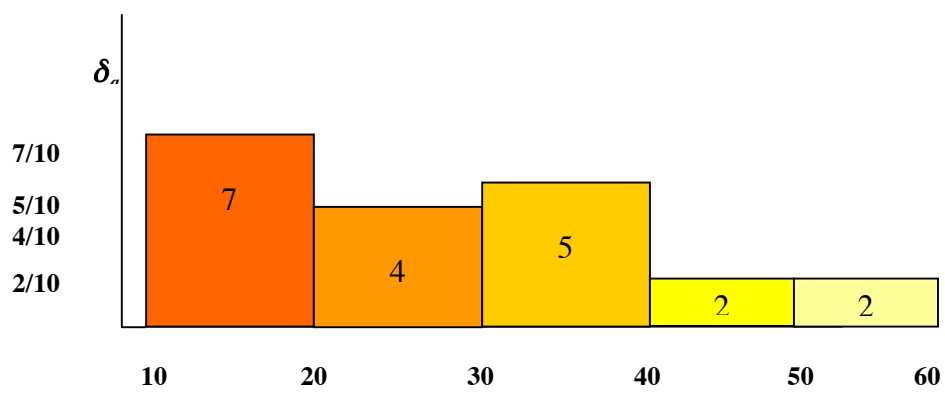


Se le classi sono di uguale ampiezza la classe modale è rappresentata anche dal rettangolo di area maggiore perché questo coincide con quello di altezza massima

Classi di valori	f_a	$\delta = \text{dens. di frequenza}$	f_{cum}
10 — 20	7	7/10=0,7	7
20 — 30	4	4/10=0,4	11
30 — 40	5	5/10=0,5	16
40 — 50	2	2/10=0,2	18
50 — 60	2	2/10=0,2	20
	20		

10 — | 20 classe modale

15 moda



2. MEDIANA

Il concetto di mediana è riferito ai valori di un carattere quantitativo o alle modalità di un carattere qualitativo ordinabile: essa è, in prima approssimazione, il valore o la modalità che occupa il posto centrale nell'allineamento in ordine crescente (o decescente) di tutti i valori o tutte le modalità rilevate (ognuna riportata tante volte quante volte è stata rilevata). Una definizione precisa di mediana deve tener conto della *tipologia* del carattere e *del numero n dei casi osservati*.

Se il carattere è qualitativo:

- *nel caso n dispari*, la mediana è la modalità che occupa il posto centrale nella gerarchia delle modalità rilevate;
- *nel caso n pari*, la mediana non è definita se nei due posti centrali ci sono modalità diverse, altrimenti è la modalità che occupa i due posti centrali.

Se il carattere è quantitativo:

- *nel caso n dispari*, la mediana è il valore che occupa il posto centrale nell'ordinamento crescente o decrescente dei valori rilevati;
- *nel caso n pari*, la mediana è il valore dato dalla media aritmetica dei valori che occupano i due posti centrali nell'ordinamento.

-
Nel caso di caratteri quantitativi raggruppati in classi la mediana non è facilmente calcolabile: di essa si dà allora una stima che rappresenta un'approssimazione del valore effettivo (vedi paragrafo 2.3)

Illustriamo di seguito più in dettaglio la definizione di mediana nei vari casi considerati attraverso l'indicazione del posto o dei posti centrali dell'ordinamento.

2.1 La mediana nel caso di caratteri qualitativi ordinabili

$n =$ il numero delle unità statistiche (o dei casi osservati)

• **Determinazione della mediana a partire dai dati grezzi (o tabella unità /modalità)**

Si ordinano le modalità secondo la naturale gerarchia ripetendo ogni modalità tante volte quante volte è stata rilevata; quindi si individua il posto centrale o i posti centrali e le modalità che occupano tali posti:

Caso n dispari $n = 2k - 1$

posti 1° 2° ... $(k-1)^\circ$ k° $(k+1)^\circ$... $n^\circ = (2k-1)^\circ$

$m_1, m_2, \dots, m_{k-1}, m_k, m_{k+1}, \dots, m_{2k-1}$

$k-1$ mod. prima di m_k

$k-1$ mod. dopo di m_k

allineamento
delle
modalità

m_1

$Me = m_k$

$m_n = m_{2k-1}$

il posto centrale è determinato dalla formula $k = (n+1)/2$, e la mediana è la modalità m_k che occupa il posto centrale

Caso n pari $n = 2k$

posti 1° 2° ... k° $(k+1)^\circ$... $n^\circ = 2k^\circ$

$m_1, m_2, \dots, m_k, m_{k+1}, \dots, m_{2k}$

$Me = m_k$ se $m_k = m_{k+1}$.

i due posti centrali sono determinati dalle formule

$k = n/2$ e $(k+1) = (n/2) + 1 = (n+2)/2$,

e, indicati con m_k e m_{k+1} le modalità che occupano tali posti la mediana:

“è la modalità m_k se $m_k = m_{k+1}$ ”

“non esiste se nei due posti centrali ci sono modalità diverse”.

Esempio M1 (per il caso n dispari).

E' stato chiesto a nove persone di esprimere un giudizio sui trasporti pubblici di Napoli. Le risposte sono state:

Buono, buono, mediocre, ottimo, mediocre, insufficiente, buono, mediocre, insufficiente.

Per ottenere la mediana degli n (=9) intervistati disponiamo i giudizi dal peggiore al migliore

1° 2° 3° 4° **5°** 6° 7° 8° 9°
i. i. m. m. **m.** b. b. b. o. (A)

Unità	Modalità(giudizi) ordinate
1°	insufficiente
2°	insufficiente
3°	mediocre
4°	mediocre
5°	mediocre
6°	buono
7°	buono
8°	buono
9°	Ottimo

**Al posto centrale 5 (=9+1)/2
compare la modalità mediocre
quindi
Mediana = mediocre**

NOTA: ordinando in senso inverso, senso decrescente, le modalità del carattere si giunge alle stesse conclusioni

Esempi M2. (per il caso n pari).

Supponiamo che ai nove intervistati di cui all'esempio M1 se ne aggiunga un altro. Allora $n = 10 = 2 \cdot 5$

1° Cosa accade se il giudizio del nuovo arrivato è insufficiente?

L'allineamento delle modalità ordinate è ora il seguente

1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	
i.	i.	i.	m.	m.	m.	b.	b.	b.	o.	(A')

I due posti centrali sono 5(= n/2) e 6 e sono occupati dalla stessa modalità (mediocre). Quindi la mediana esiste ed è “mediocre”

2° Cosa accade se il giudizio del nuovo arrivato è buono?

L'allineamento delle modalità ordinate è il seguente con posti centrali : 5 e 6

1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	
i.	i.	m.	m.	m.	b.	b.	b.	b.	o.	(A'')

I due posti centrali 5(= n/2) e 6 sono occupati da modalità diverse. In tal caso la mediana non esiste

• **Determinazione della mediana a partire da una tabella di dati elaborati (tabella modalità/ frequenza).**

Nella tabella modalità/frequenze le modalità devono essere indicate rispettando il livelli della naturale gerarchia; in tal modo si legge l'allineamento crescente e, contando le frequenze dalla prima all'ultima, si contano i posti che le corrispondenti modalità occupano nell'allineamento; si possono inoltre calcolare le frequenze cumulate, utili anche esse per la determinazione della mediana.

Esempio M1'.

Ripartiamo dall'esempio M1. Le 9 modalità rilevate sono:

Buono, buono, mediocre, ottimo, mediocre, insufficiente, buono, mediocre, insufficiente.

Costruiamo la tabella *modalità/frequenze* indicando le modalità distinte nell'ordine della naturale gerarchia e determiniamo il posto centrale

Giudizi	f_a (posti)	f_a cum	$F=f_r$ cum	$n=9$
Insufficiente	2 (1-2)	2	$2/9=0,222..$	$k = (n+1) / 2 = 5$ posto centr.
mediocre	3 (3-4-5)	5	$5/9=0,5555..$	
Buono	3 (6-7-8)	8	$8/9=0,888.$	
Ottimo	1 (9)	9	$9/9=1$	
	9			

La prima frequenza è 2 e indica che nei primi due posti nell'allineamento (A) c'è "insufficiente", la seconda frequenza è 3 e indica che nei successivi tre posti (3°, 4°, 5°) c'è "**mediocre**" che è la mediana perché occupa il posto centrale; così continuando si ottiene tutto l'allineamento:

1°	2°	3°	4°	5°	6°	7°	8°	9°	
i.	i.	m.	m.	m.	b.	b.	b.	o.	(A)

Alla mediana "**mediocre**" corrisponde una frequenza cumulata pari a 5 posto centrale e la prima frequenza relativa cumulata maggiore di 0,5.

Si può determinare la mediana considerando le frequenze cumulate e leggendo ogni frequenza cumulata t come indicatrice dei primi t posti.

- **Regola per la determinazione della mediana dalla tabella modalità/frequenze attraverso le frequenze cumulate e la funzione di ripartizione F , nel caso n dispari.**

La mediana è la modalità a cui corrisponde la prima frequenza cumulata che comprende il posto centrale k ed è cioè maggiore o uguale a k (5 nell'esempio M1'); è anche la modalità cui corrisponde la prima frequenza relativa cumulata che risulta $> 1/2 = 0.5$.

Esempio M2'(per il caso n pari).

Ripartiamo dall'esempio **M2**. Le modalità rilevate sono:

Buono, buono, mediocre, ottimo, mediocre, insufficiente, buono, mediocre, insufficiente, insufficiente.

Costruiamo la tabella *modalità/ frequenze* riportando le modalità secondo la naturale gerarchia

giudizi	f_a	$f_{a\ cum}$	$F=f_{r\ cum}$	
<i>Insufficiente</i>	3	3	$3/10=0,3$	$n=10$
<i>mediocre</i>	3	6	$6/10=0,6$	$k = (n) / 2 = 5$ 1° posto cent.
<i>buono</i>	3	9	$9/10=0,9$	$k+1 = 6$ 2° posto cent.
<i>ottimo</i>	1	10	$10/10=1$	
	10			

La prima modalità nella gerarchia, “sufficiente”, ha frequenza 3 e quindi occupa i primi 3 posti; segue “mediocre” che occupa i successivi 3 posti, cioè i posti 4°, 5°, 6°: questi includono i posti centrali e quindi la mediana è “mediocre”; “buono” occupa i 3 posti successivi, cioè i posti 7°, 8°, 9°, e “ottimo” occupa l’ottavo e ultimo posto.

1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	(A')
i.	i.	i.	m.	m.	m.	b.	b.	b.	o.	

Nota che alla mediana “mediocre” corrisponde una frequenza cumulata pari a 6, che comprende quindi i due posti centrali, e la prima frequenza relativa cumulata maggiore di 0,5.

Supponiamo che le modalità rilevate siano

Buono, buono, mediocre, ottimo, mediocre, insufficiente, buono, mediocre, insufficiente, buono.

La tabella *modalità/frequenze* costruita indicando le modalità distinte in modo da rispettare la gerarchia è la seguente

Giudizi	f_a	$f_{a\ cum}$	F
<i>Insufficiente</i>	2	2	0,2
<i>Mediocre</i>	3	5	0,5
<i>Buono</i>	4	9	0,9
<i>Ottimo</i>	1	10	1
	10		

$n=10$
 $k = (n)/2 = 5$ 1° p. cent.
 $k+1 = 6$ 2° p. cent.

e da essa si risale al corrisponde allineamento (A'') dell'esempio M2

1°	2°	3°	4°	5°	6°	7°	8°	9°	10°
i.	i.	m.	m.	m.	b.	b.	b.	b.	o.

(A'')

da cui deduciamo la non esistenza della mediana

Leggiamo la non esistenza della mediana direttamente dalla tabella:

la prima frequenza cumulata che include il primo posto centrale è 5; essa non include il secondo posto centrale 6, che è invece incluso nella frequenza cumulata successiva: ciò vuol dire che ai posti centrali corrispondono modalità diverse e di conseguenza la mediana non esiste.

Nota: *in questo caso particolare c'è una frequenza relativa cumulata pari a 0.5 = 1/2 e corrisponde alla modalità che occupa il primo posto centrale.*

- **Regola per la determinazione della mediana dalla tabella *modalità/frequenze* attraverso le frequenze cumulate e la funzione di ripartizione F , nel caso n pari.**

La mediana esiste se e solo nella tabella non compare una frequenza assoluta pari al primo posto centrale e una frequenza relativa cumulata pari a 0.5: essa è allora la modalità corrispondente alla prima frequenza relativa cumulata >0.5

Esempio M3. è data la seguente tabella *modalità / frequenze* relativa al carattere titolo di studio. Determinare la mediana

<i>Titolo di studio</i>	<i>popolazione = f_a</i>	<i>f_{cum}</i>
<i>Nessun titolo</i>	<i>10.994.700</i>	<i>10.994.700</i>
<i>Licenza elementare</i>	<i>21.153.950</i>	<i>32.148.650</i>
<i>Licenza media</i>	<i>12.375.350</i>	<i>44.524.000</i>
<i>Diploma</i>	<i>5.881.900</i>	<i>50.405.900</i>
<i>Laurea</i>	<i>1.423.550</i>	<i>51.829.450</i>
<i>Tot.</i>	<i>51.829.450</i>	

La frequenza totale è **51.829.450** e ciò vuol dire che sono stati rilevati i titoli di studio in **51.829.450** casi; quindi le modalità non distinte (cioè le risposte all'indagine) vanno a disporsi in **51.829.450** posti. I posti centrali sono 2, il k-esimo e il (k+1) esimo, con

$$k = \frac{n}{2} = \frac{51.829.450}{2} = 25.914.750 \quad e \quad k + 1 = 25.914.751$$

I posti centrali sono inclusi nella seconda frequenza cumulata e sono occupati entrambi dalla modalità licenza elementare, che risulta la mediana della distribuzione.

2.2. La mediana nel caso di caratteri quantitativi discreti (con valori non raggruppati in classi)

$n =$ il numero delle unità statistiche (o dei casi osservati)

• **Determinazione della mediana a partire dai dati grezzi (o tabella unità /modalità)**

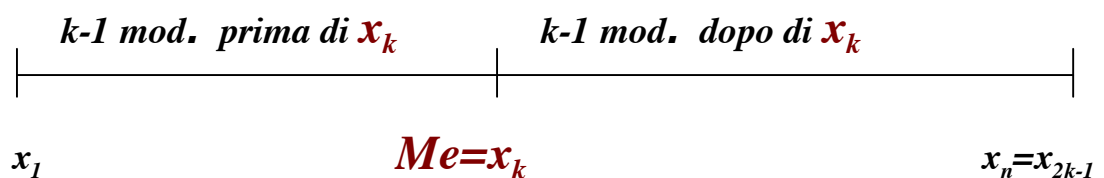
Si allineano i valori del carattere secondo l'ordine crescente (o decrescente), riportando ogni valore tante volte quante volte è stato rilevato; quindi

nel Caso n dispari $n = 2k-1$

$$x_1, x_2, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_n$$

$$\text{mediana} = x_k$$

allineamento
crescente o
decr. dei valori



Il posto centrale è dato dalla formula $k = (n+1)/2$, e la mediana è il valore x_k che occupa il posto centrale

nel Caso $n = 2k$

$$x_1, x_2, \dots, x_k, x_{k+1}, \dots, x_{2k}$$

allineamento
crescente o decr.
dei valori

I due posti centrali sono dati dalle formule: $k = n/2$ e $k+1 = (n/2)+1 = (n+2)/2$, e

La mediana è la media aritmetica dei valori x_k e x_{k+1} che occupano i due posti centrali

$$\text{mediana} = \frac{x_k + x_{k+1}}{2}$$

Esempi M4 (per il caso n dispari)

- *E' stato rilevato il numero delle monete da 10 cents immesse nella cassetta dell'ascensore in 11 giorni*

Giorni	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	11°
	270	223	313	277	223	289	270	292	313	259	313

1° metodo: allineiamo i valori del carattere in ordine crescente

	223	223	259	270	270	277	289	292	313	313	313	(*)
posti	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	11°	

n=11 posto centrale $k=(n+1)/2 =12/2 =6$

mediana 277

2° metodo: determiniamo la mediana a partire da una tabella di dati elaborati (tabella modalità/ frequenza) in cui i valori distinti sono indicati in ordine crescente; in tal modo si legge l'allineamento crescente e contando le frequenze dalla prima all'ultima si contano i posti che le corrispondenti modalità occupano nell'allineamento; si possono inoltre calcolare le frequenze cumulate, utili anche esse per la determinazione della mediana.

Valori	$f_i = \text{fr.ass.}$	$f_c = \text{fr.ass.cum.}$	F
223	2	2 (1-2)	$2/11 = 0.1818...$
259	1	3 (3)	$3/11 = 0.2727...$
270	2	5 (4-5)	$5/11 = 0.4545...$
277	1	6 (6)	$6/11$
289	1	7	$=0.5454 > 1/2$
292	1	8	$7/11 = 0.6363...$
313	3	11	$8/11 = 0.7272...$
	11		$11/11 = 1$

$$n=11 = \sum_i f_i$$

$$k = (11+1)/2 = 6 \text{ posto centrale}$$

$$k = \frac{n+1}{2} = \frac{\sum_i f_i + 1}{2}$$

Da tale tabella risaliamo all'allineamento () dove leggiamo che il posto centrale 6 è occupato da 277 che è la mediana*

Leggiamo direttamente la mediana dalla tabella:

La prima frequenza cumulata che include il posto centrale 6 è 6 corrisponde al valore 277 del carattere: allora la mediana è 277.

Notiamo che alla mediana 277 corrisponde la prima frequenza cumulata relativa maggiore di 1/2

- Consideriamo una diversa distribuzione di valori

	223	223	259	270	270	<u>277</u>	277	289	292	313	313
posti	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	11°

n=11 **posto centrale 6**

Costruiamo una tabella di frequenze in cui immettiamo i valori del carattere in ordine crescente

Valori car.	$f_i = \text{fr.ass.}$	$f_c = \text{fr.ass.cum.}$	F
223	2	2 (1-2)	$2/11 = 0.1818\dots$
259	1	3 (3)	$3/11 = 0.2727\dots$
270	2	5 (4-5)	$5/11 = 0.4545\dots$
<u>277</u>	2	<u>7</u> (6-7)	<u>$7/11 = 0.63636 > 1/2$</u>
289	1	8	$8/11 = 0.727\dots$
292	1	9	$9/11 = 0.8181\dots$
313	2	11	$11/11 = 1$
	11		

Nelle prime due unità la determinazione del carattere è 223, nella terza è 259, nella quarta e quinta è 270 nella sesta e settima è 277; quindi nell'unità al sesto posto il valore del carattere è 277 e questo è il valore della mediana

La prima frequenza cumulata che include il posto centrale 6 è 7 e corrisponde al valore 277: allora la mediana è 277.

Nota alla mediana corrisponde la prima frequenza cumulata maggiore di 6 e la prima frequenza relativa cumulata maggiore di 1/2

Esempio M5(per il caso n pari).

Monetine immerse nella cassetta dell'ascensore in 10 giorni

Giorni	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°
	270	223	313	277	223	289	270	292	313	259

1° metodo: allineiamo i valori del carattere in ordine crescente

	223	223	259	270	270	277	289	292	313	313	(°)
posti	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	

n=10 posti centrali $k = 10/2 = 5$ e $k+1= 6$

mediana $(270+277)/2= 273,5$

2° metodo: determiniamo la mediana a partire da una tabella di dati elaborati (tabella modalità/ frequenza) in cui i valori distinti sono indicati in ordine crescente; in tal modo si legge l'allineamento crescente e contando le

Valori carat.	$f_i = \text{fr. ass.}$	$f_c = \text{fr. ass. cum}$	F
223	2	2	$2/10=0,2$
259	1	3	$3/10=0,3$
270	2	5	$5/10=0,5$
277	1	6	$6/10=0,6$
289	1	7	$7/10=0,7$
292	1	8	$8/10=0,8$
313	2	10	1

10

da tale tabella si può risalire all'allineamento (°) e determinare la mediana

Leggiamo la mediana direttamente dalla tabella:

I due posti centrali corrispondono proprio alle frequenze cumulate 5 e 6 e sono occupati rispettivamente dai valori 270 e 277 allora

mediana $(270+277)/2= 273,5$

Nota: in questo caso particolare il primo dei due posti centrali corrisponde alla frequenza cumulata relativa pari a $0.5 = 1/2$.

- consideriamo una diversa distribuzione di dati e di avere già la tabella valori/frequenze

Valori carat.	$f_i = \text{fr. ass.}$	$f_c = \text{fr. ass. cum}$	F
223	2	2	2/10=0,2
259	2	4	4/10=0,4
270	2	6	6/10= 0,6
277	1	7	7/10 =0,7
289	1	8	8/10 =0,8
292	1	9	9/10=0,9
313	1	10	1

10

Da tale tabella si risale all'allineamento

223 223 259 259 270 270 277 289 292 313 **n=10**

posti centrali $k = n/2 = 10/2 = 5$ e $k+1 = (n/2)+1 = 5+1=6$

Mediana = $(270+270)/2 = 270$

Leggiamo la mediana direttamente dalla tabella:

*I due posti centrali sono compresi nella prima frequenza cumulata maggiore a $k=5$ e sono occupati dallo stesso valore del carattere **270** che costituisce la mediana*

Nota: *i due posti centrali corrispondono alla prima frequenza cumulata relativa maggiore di **0.5 = 1/2***

••Regola per la determinazione della mediana utilizzando le frequenze cumulate o la funzione di ripartizione F

Nel caso n dispari in cui il posto centrale è

$$k = \frac{n+1}{2} = \frac{\sum_i f_i + 1}{2} \quad \text{posto centrale}$$

la mediana è quel valore x_i a cui corrisponde la prima frequenza cumulata maggiore o uguale a k e la prima frequenza relativa cumulata maggiore di $\frac{1}{2} = 0.5$; risulta cioè:

- $F(x_{i-1}) < \frac{1}{2} = 0.5$
- $F(x_i) > \frac{1}{2} = 0.5$

Nel caso n pari i due posti centrali sono

$$k = \frac{n}{2} = \frac{\sum_i f_i}{2}, \quad k+1 = \frac{n}{2} + 1 = \frac{n+2}{2} = \frac{\sum_i f_i}{2} + 1 = \frac{(\sum_i f_i) + 2}{2}$$

posti centrali

- se c'è una frequenza cumulata uguale a $k = n/2$ (e quindi una frequenza cumulata relativa pari a $1/2 = 0.5$) la mediana è la media aritmetica del valore a cui corrisponde la frequenza cumulata k e del valore a cui corrisponde la frequenza cumulata successiva;
- se non c'è una frequenza cumulata uguale a $k = n/2$ la mediana è il valore cui corrisponde la prima frequenza cumulata maggiore di k e quindi la prima frequenza cumulata relativa maggiore di 0.5.

NOTA: le regole per la determinazione della mediana per un **carattere quantitativo discreto** e per un **carattere qualitativo**, che è anch'esso considerato discreto, possono essere riassunte nella seguente regola

Regola per la determinazione della mediana di un carattere discreto utilizzando le frequenze cumulate o la funzione di ripartizione F

- *Nel caso n dispari in cui il posto centrale è*

$$k = \frac{n+1}{2} = \frac{\sum_i f_i + 1}{2} \quad \text{posto centrale}$$

la mediana è quel valore x_i o quella modalità m_i a cui corrisponde la prima frequenza cumulata maggiore o uguale a k e la prima frequenza relativa cumulata maggiore di $\frac{1}{2} = 0.5$

- *Nel caso n pari i due posti centrali sono*

$$k = \frac{n}{2} = \frac{\sum_i f_i}{2}, \quad k+1 = \frac{n}{2} + 1 = \frac{n+2}{2} = \frac{\sum_i f_i}{2} + 1 = \frac{(\sum_i f_i) + 2}{2}$$

posti centrali

e si possono presentare due sotto casi:

1° - c'è una frequenza cumulata uguale a $k = n/2$ (e quindi una frequenza cumulata relativa pari a $1/2 = 0.5$), allora:

- se il carattere è quantitativo la mediana è la media aritmetica del valore a cui corrisponde la frequenza cumulata k e del valore successivo;
- se il carattere è qualitativo la mediana non esiste

2° - non c'è una frequenza cumulata uguale a $k = n/2$; allora la mediana è il valore o la modalità cui corrisponde la prima frequenza cumulata maggiore di k e quindi la prima frequenza cumulata relativa maggiore di 0.5.

2.3 Stima della mediana nel caso di caratteri quantitativi con valori raggruppati in classi

La mediana nel caso considerato è teoricamente definita come nel caso discreto. Tuttavia, cadendo la mediana in una delle classi e non conoscendo come i valori sono distribuiti in tale classe, non si può ottenere un valore esatto di essa; si può solo ottenere una stima della stessa ipotizzando una particolare distribuzione delle frequenze cumulate all'interno delle classi.

Per ottenere una stima (approssimata) della mediana è importante in ogni caso prima determinare la classe che contiene la mediana e che è detta **classe mediana; essa**

- è la classe contenente il valore del carattere di posto

$k = (n+1)/2$ nel caso **n dispari**

- è la classe contenente i valori del carattere di posto

$k = n/2$ e $k+1 = (n/2)+1$ nel caso **n pari**

Nel caso particolare in cui n è pari ma i valori nei posti k e $k+1$ sono in classi diverse si considerano due classi mediane contigue. In tal caso si conviene di scegliere come mediana l'estremo comune alle due classi

Esempi:

ESERCIZI ALBERGHIERI CON OLTRE 20 ADDETTI NEL 1977

Classi di fatturato in milioni di lire	$f_i =$ n. di esercizi	$f_c =$ n. di esercizi cumulado
≤ 50	$f_1 = 5$	$f_1 = 5 = f_c (50)$
50 -- 100	$f_2 = 16$	$f_1 + f_2 = 21 = f_c (100)$
100 -- 500	$f_3 = 442$	$f_1 + f_2 + f_3 = 463 = f_c (500)$
500 -- 1000	$f_4 = 128$	591
1000 -- 5000	$f_5 = 87$	678
> 5000	$f_6 = 15$	693

$$n = 693$$

$$\text{Posto centrale } k = \frac{694}{2} = 347 \Rightarrow \text{Classe mediana } 100 \text{ ---| } 500$$

caso particolare

Classi	f_a	f_c
≤ 50	$f_1 = 5$	$f_1 = 5 = f_c(50)$
50 -- 100	$f_2 = 16$	$f_1 + f_2 = 21 = f_c(100)$
<u>100 -- 500</u>	$f_3 = 421$	$f_1 + f_2 + f_3 = 442 = f_c(500)$
<u>500 -- 1000</u>	$f_4 = 128$	570
1000 -- 500	$f_5 = 87$	657
> 5000	$f_6 = 227$	884

$n = 884$

Posti centrali: $k = \frac{884}{2} = 442$ e $k+1 = 443$

in tal caso *accade che i valori che occupano i due posti centrali appartengono a due classi distinte ma contigue*: avremo allora due classi mediane

Classi mediane: 100 --| 500 500 --| 1000

Si conviene allora di scegliere come mediana l'estremo comune alle due classi

Mediana = 500

DALLA CLASSE MEDIANA ALLA STIMA DELLA MEDIANA nel caso in cui comporta la distribuzione di frequenza cumulata risulti lineare all'interno delle classi (*cioè* nel caso di distribuzione uniforme *delle frequenze assolute*)

Individuata la classe mediana e supposto che $x_{(r)}$ e $x_{(r+1)}$ siano i suoi estremi, si procede come segue:

- nel caso **n dispari** si pone

$$Mediana = x_{(r)} + \frac{x_{(r+1)} - x_{(r)}}{f_{cum}(x_{(r+1)}) - f_{cum}(x_{(r)})} \left(\frac{n+1}{2} - f_{cum}(x_{(r)}) \right) \quad (1)$$

- nel caso **n pari** si pone convenzionalmente

$$Mediana = x_{(r)} + \frac{x_{(r+1)} - x_{(r)}}{f_{cum}(x_{(r+1)}) - f_{cum}(x_{(r)})} \left(\frac{n}{2} - f_{cum}(x_{(r)}) \right) \quad (2)$$

anche se la definizione di mediana suggerirebbe, nel caso n pari, di porre la mediana uguale alla media aritmetica del valore (2) e del seguente altro

$$x_{(r)} + \frac{x_{(r+1)} - x_{(r)}}{f_{cum}(x_{(r+1)}) - f_{cum}(x_{(r)})} \left(\frac{n}{2} + 1 - f_{cum}(x_{(r)}) \right) \quad (3)$$

Legenda

$x_{(r+1)} - x_{(r)} =$ *ampiezza della classe mediana*

$f_{cum}(x_{(r+1)}) - f_{cum}(x_{(r)}) =$ *frequenza corrispondente alla classe mediana*

$\frac{n+1}{2} - f_{cum}(x_{(r)}) =$ *numero dei casi da aggiungere ai casi con determinazione del carattere $\leq x_{(r)}$ per giungere al posto centrale nel caso n dispari*

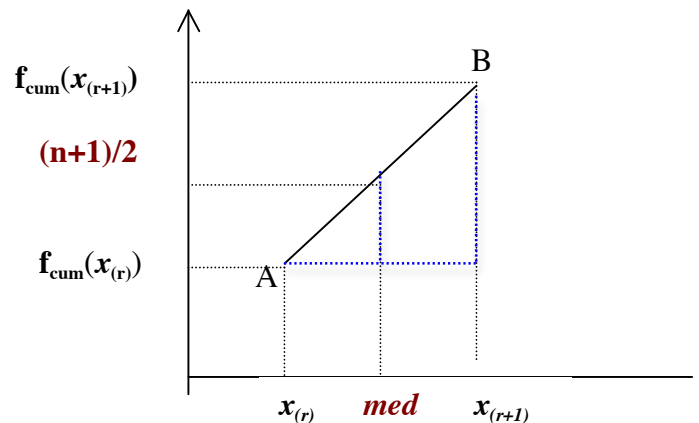
• *Giustificazione della formula (1) per il caso n dispari*

classi	$f_{(i)} = \text{fr. ass}$	f_{cum}
$x_{(0)} - - x_{(1)}$	$f_{(1)}$	$f_{(1)} = f_{\text{cum}}(x_{(1)})$
•	•	
•	•	
•	•	
$x_{(r-1)} - x_{(r)}$	$f_{(r)}$	$f_{(1)} + \dots + f_{(r)} = f_{\text{cum}}(x_{(r)})$
$x_{(r)} - x_{(r+1)}$	$f(x_{(r+1)})$	$f_{(1)} + \dots + f_{(r)} + f(x_{(r+1)}) = f_{\text{cum}}(x_{(r+1)})$
•	•	
•	•	
•	•	
$x_{(k)} - x_{(k+1)}$	$f_{(k)}$	$f_{(1)} + \dots + f_{(k)} = n$
	$n=2k-1$	

$$f_{\text{cum}}(x_{(r)}) < (n+1)/2 \leq f_{\text{cum}}(x_{(r+1)}) \Leftrightarrow x_{(r)} < \text{med.} \leq x_{(r+1)}$$

Se la funzione di distribuzione delle frequenze cumulate è supposta lineare in ogni classe si ha la seguente uguaglianza di rapporti

$$\frac{\text{med.} - x_{(r)}}{\frac{n+1}{2} - f_{\text{cum}}(x_{(r)})} = \frac{x_{(r+1)} - x_{(r)}}{f_{\text{cum}}(x_{(r+1)}) - f_{\text{cum}}(x_{(r)})}$$



$$\text{med.} - x_{(r)} = \frac{x_{(r+1)} - x_{(r)}}{f_{\text{cum}}(x_{(r+1)}) - f_{\text{cum}}(x_{(r)})} \left(\frac{n+1}{2} - f_{\text{cum}}(x_{(r)}) \right)$$

$$\text{med.} = x_{(r)} + \frac{x_{(r+1)} - x_{(r)}}{f_{\text{cum}}(x_{(r+1)}) - f_{\text{cum}}(x_{(r)})} \left(\frac{n+1}{2} - f_{\text{cum}}(x_{(r)}) \right)$$

Allo stesso risultato si giunge applicando l'equazione della retta per A ($x_{(r)}, f_{\text{cum}}(x_{(r)})$) e B ($x_{(r+1)}, f_{\text{cum}}(x_{(r+1)})$) a cui deve appartenere il punto M($\text{med.}, (n+1)/2$)

- Giustificazione della formula (2) per il caso n pari

Nel caso n pari la formula (2) per il calcolo della mediana si determina allo stesso modo in cui si determina la formula (1), convenendo però che la mediana sia il valore che occupa il primo posto centrale

Esempi

ESERCIZI ALBERGHIERI CON OLTRE 20 ADDETTI NEL 1977

Classi di fatturato in migliaia di euro	f_i = n. di esercizi	f_c = n. di esercizi cumulato
≤ 50	$f_1 = 5$	$f_1 = 5 = f_c(50)$
50 -- 100	$f_2 = 16$	$f_1 + f_2 = 21 = f_c(100)$
100 -- 500	$f_3 = 442$	$f_1 + f_2 + f_3 = 463 = f_c(500)$
500 -- 1000	$f_4 = 128$	591
1000 -- 5000	$f_5 = 87$	678
> 5000	$f_6 = 15$	693

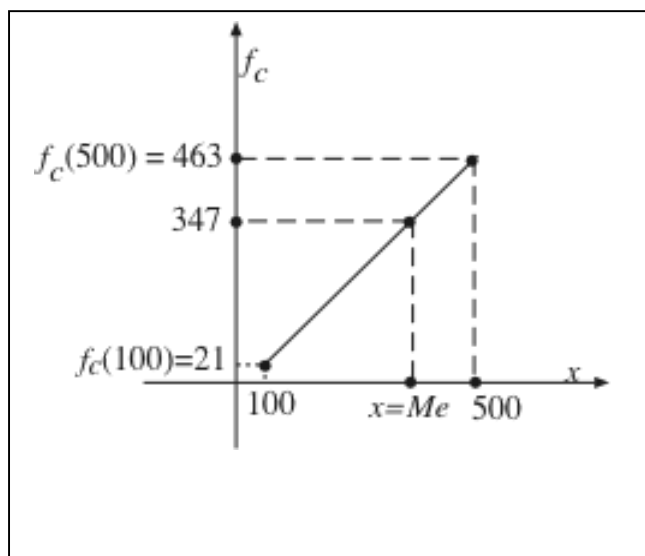
$$n = 693$$

Posto centrale $k = \frac{694}{2} = 347 \Rightarrow$ Classe mediana 100 --| 500

$$x - 100 = \frac{500 - 100}{463 - 21} (347 - 21)$$

$$x = 100 + \frac{500 - 100}{442} (347 - 21)$$

100 = lim. inf classe med.
500 = lim. sup. classe med.
21 = $f_c(100) = f_1 + f_2$
463 = $f_c(500) = f_1 + f_2 + f_3$
442 = freq. class med.
347 = $\frac{n+1}{2}$



quindi è $x = 100 + \frac{400}{442} 326 = 395,022626$

Un caso particolare per n dispari (per classi di valori chiuse a destra):

Se n è dispari ed una frequenza cumulata f_c coincide con k essendo k il numero che indica il posto centrale: allora la mediana coincide con l'estremo superiore della classe mediana (estremo che coincide con il valore a cui è associata la frequenza cumulata uguale a k)

Classi	f_a	f_c
≤ 50	$f_1 = 5$	$f_1 = 5 = f_c (50)$
50 -- 100	$f_2 = 16$	$f_1 + f_2 = 21 = f_c (100)$
100 -- 500	$f_3 = 421$	$f_1 + f_2 + f_3 = 442 = f_c (500)$
500 -- 1000	$f_4 = 128$	570
1000 -- 500	$f_5 = 87$	657
> 5000	$f_6 = 226$	884

$$n = 883$$

Posto centrale: $k = \frac{n+1}{2} = \frac{884}{2} = 442$ classe mediana **100 --| 500**

se 500 è un valore preso effettivamente allora 500 è il valore che occupa il posto centrale e ha senso allora porre:

$$\text{Mediana} = 500$$

D'altra parte applicando la formula (1) ritroviamo lo stesso risultato

caso particolare per n

Classi	f_a	f_c
≤ 50	$f_1 = 5$	$f_1 = 5 = f_c (50)$
50 -- 100	$f_2 = 16$	$f_1 + f_2 = 21 = f_c (100)$
100 -- 500	$f_3 = 421$	$f_1 + f_2 + f_3 = 442 = f_c (500)$
500 -- 1000	$f_4 = 128$	570
1000 -- 500	$f_5 = 87$	657
> 5000	$f_6 = 227$	884

$$n = 884$$

Posti centrali: $k = \frac{884}{2} = 442$ e $k+1 = 443$

Classi mediane: **100 --| 500** **500 --| 1000**

Mediana = 500 come convenuto a pag. 26 e come si ricava applicando la formula (2)

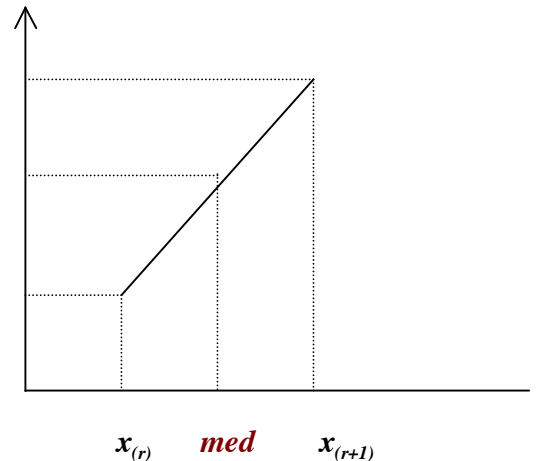
Formule equivalenti alla (1) per il calcolo della mediana

La formula che dà la mediana non cambia se si utilizzano le frequenze cumulate relative e il posto centrale viene diviso per il totale delle unità statistiche

$$\begin{aligned}
 med. &= x_{(r)} + \frac{x_{(r+1)} - x_{(r)}}{f_{cum}(x_{(r+1)}) - f_{cum}(x_{(r)})} \left(\frac{n+1}{2} - f_{cum}(x_{(r)}) \right) = \\
 &\quad \Downarrow \\
 med. &= x_{(r)} + \frac{x_{(r+1)} - x_{(r)}}{\frac{1}{n} (f_{cum}(x_{(r+1)}) - f_{cum}(x_{(r)}))} \frac{1}{n} \left(\frac{n+1}{2} - f_{cum}(x_{(r)}) \right) \\
 &\quad \Downarrow \\
 x &= x_{(r)} + \frac{x_{(r+1)} - x_{(r)}}{\frac{f_{cum}(x_{(r+1)})}{n} - \frac{f_{cum}(x_{(r)})}{n}} \left(\frac{n+1}{2n} - \frac{f_{cum}(x_{(r)})}{n} \right) \\
 &\quad \Downarrow \\
 med. &= x_{(r)} + \frac{x_{(r+1)} - x_{(r)}}{F(x_{(r+1)}) - F(x_{(r)})} \left(\frac{n+1}{2n} - F(x_{(r)}) \right) \quad (4)
 \end{aligned}$$

Alla (4) si giunge
Utilizzando
l'ogiva delle
frequenze
cumulate relative e
seguendo il
ragionamento
analogo a quello
svolto per
dimostrare la (1)

$$\begin{aligned}
 F(x_{(r+1)}) &= f_{cum}(x_{(r+1)})/n \\
 & \quad (n+1)/2n \\
 F(x_{(r)}) &= f_{cum}(x_{(r)})/n
 \end{aligned}$$



Moltiplicando per 100 numeratore e denominatore del rapporto che segue $x_{(r)}$ + nella formula (4) si giunge alla seguente formula che utilizza le frequenze cumulate percentuali $F_{\%}(x) = F(x) \times 100$

$$x_{(r)} + \frac{x_{(r+1)} - x_{(r)}}{F_{\%}(x_{(r+1)}) - F_{\%}(x_{(r)})} \left(\frac{n+1}{n} 100 - F_{\%}(x_{(r)}) \right)$$

Formule equivalenti alla (2) per il calcolo della mediana

Anche in questa formula le frequenze cumulate assolute possono essere sostituite da quelle cumulate relative o percentuali. osserviamo che :

- se si usano le frequenze cumulate relative la (2) si trasforma in

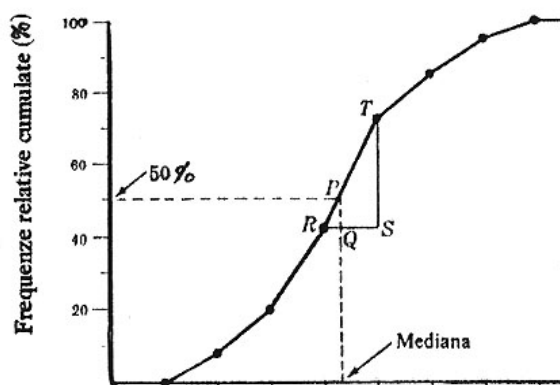
$$x_{(r)} + \frac{x_{(r+1)} - x_{(r)}}{\frac{f_{cum}(x_{(r+1)})}{n} - \frac{f_{cum}(x_{(r)})}{n}} \left(\frac{1}{2} - \frac{f_{cum}(x_{(r)})}{n} \right) \quad (5)$$

cioè

$$x_{(r)} + \frac{x_{(r+1)} - x_{(r)}}{F(x_{(r+1)}) - F(x_{(r)})} (0,5 - F(x_{(r)})) \quad (6)$$

- se si usano le frequenze cumulate relative percentuali $F_{\%}(x) = F(x) \times 100$, la (2) si trasforma in

$$x_{(r)} + \frac{x_{(r+1)} - x_{(r)}}{F_{\%}(x_{(r+1)}) - F_{\%}(x_{(r)})} (50 - F_{\%}(x_{(r)})) \quad (7)$$



Nota. Se sono note solo le frequenze cumulate relative e non sappiamo se n è pari o dispari si utilizza per il calcolo della mediana la formula (5) o (6)

Si può utilizzare equivalentemente la formula (7) con le frequenze cumulate percentuali

$$x_{(r)} + \frac{x_{(r+1)} - x_{(r)}}{\frac{f_{cum}(x_{(r+1)})}{n} \times 100 - \frac{f_{cum}(x_{(r)})}{n} \times 100} \left(50 - \frac{f_{cum}(x_{(r)})}{n} \times 100 \right)$$

Esempio:

Classi di statura

Classi	Val. centrale	$f_i = \text{freq. rel}$	F_i
[155-164)	159.5	0.093	0.093
[164-169)	166.5	0.194	0.287
[169-174)	171.5	0.290	0.577
[174-179)	176.5	0.248	0.825
[179-184)	181.5	0.126	0.951
[184-194)	189	0.049	1.000
Totali		1.000	

Le frequenze assegnate sono frequenze relative. La classe mediana è la classe [169-174) a cui corrisponde la prima frequenza relativa cumulata $F(x_{r+1}) = F(174) \geq 0.5$. Utilizzando la formula (6), otteniamo:

$$Me = 169 + \frac{5}{0.577 - 0.287} (0.5 - 0.287) = 172.67$$

2.4. Una proprietà della mediana per un carattere quantitativo discreto

Se ci riferiamo a un carattere quantitativo discreto, la mediana è quel numero che minimizza la somma degli scarti presi in valore assoluto

$$\sum_i |a_i - \text{Med}(a_i)| < \sum_i |a_i - y| \quad \forall y \neq \text{Med}(a_i)$$

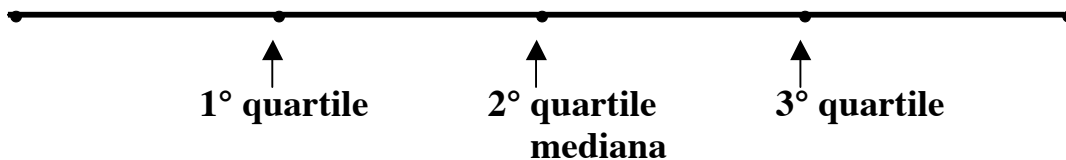
Problema

La ditta Delta intende costruire tre nuove unità locali alla periferia di Bologna.

Il progetto prevede che esse siano disposte lungo un rettilineo alle seguenti distanze dall'entrata in città: 800m, 2km, 5km; è previsto anche che esse facciano capo a un unico stabilimento fornitore delle materie prime.

A quale distanza dall'entrata in città bisogna collocare questo stabilimento se si vuole minimizzare la spesa per il trasporto delle materie prime verso le tre unità locali?

I quartili sono 4-quantili e sono in numero di tre: il primo quartile è preceduto dal primo quarto (25%) dei valori allineati e seguito dai restanti tre quarti (75%); il secondo è preceduto dai primi due quarti (50%) dei valori allineati e seguito dai restanti due quarti (50%) e coincide quindi con la mediana; il terzo quartile è preceduto dai primi tre quarti (75%) dei valori allineati e seguito dall'ultimo quarto (25%).

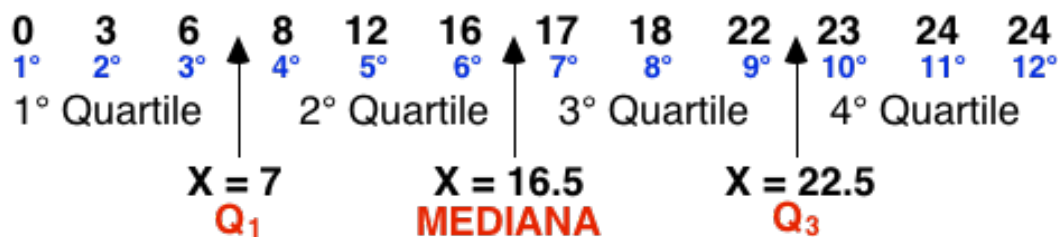


Esempio

VARIABILE: Numero di errori nella lettura di un brano (N=12)

Distribuzione ordinata dei punteggi (e posizioni):

Dividiamo la distribuzione dei valori in quattro parti di uguale numerosità: allora ogni parte è costituita da tre valori.



Il primo quartile si troverà tra il 3° e 4° posto e assume il valore dato dalla media aritmetica dei valori che occupano tali posti. Il secondo quartile si troverà tra il 6° e 7° posto e assume il valore dato dalla media aritmetica dei valori che occupano tali posti. Il terzo quartile si troverà tra il 9° e 10° posto e assume il valore dato dalla media aritmetica dei valori che occupano tali posti

$$\text{Posti centrali } n/2 = 12/2 = 6 \text{ e } (n/2) + 1 = 7$$

$$\downarrow$$

$$\mathbf{Q_2 = Mediana = (16+17)/2 = 16.5}$$

$$n/4 = 12/4 = 3 \text{ e } (n/4) + 1 = 4$$

$$\downarrow$$

$$\mathbf{Q_1 = (6+8)/2 = 7}$$

$$3n/4 = 36/4 = 9 \text{ e } (3n/4) + 1 = 10$$

$$\downarrow$$

$$\mathbf{Q_3 = (22+23)/2 = 22,5}$$

