Statistica: principi e metodi

Capitolo 4

Indici di sintesi:
Medie

Medie

- le medie sono lo strumento con cui si **sintetizzano** i dati statistici.
- l'uso della media consente all'individuo di rappresentarsi mentalmente l'"**ordine di grandezza**" di un fenomeno, di effettuare **comparazioni** tra le manifestazioni di uno stesso fenomeno in tempi, luoghi o situazioni diverse, di comunicare ad altri tale informazione.

Operazioni che è possibile fare sulle modalità dei diversi tipi di carattere

		Operazioni		
		Uguaglianza/ Disuguaglianza	Ordinamento	Operazioni aritmetiche
Caratteri		**=, ≠**	**> , <**	**+, -, *, /**
Qualitativi	***sconnessi***	**si**	**no**	**no**
	ordinabili	**si**	**si**	**no**
Quantitativi		**si**	**si**	**si**

Medie che è possibile calcolare in relazione ai diversi tipi di carattere

		Operazioni		
		Moda	**Statistiche d'ordine** *(Mediana, Quartili, Decili, Percentili, Quantili)*	**Medie algebriche** (*media aritmetica, media armonica, media geometrica, media quadratica*)
Caratteri		**=, ≠**	**> , <**	**+, -, *, /**
Qualitativi	***sconnessi***	**si**	**no**	**no**
	ordinabili	**si**	**si**	**no**
Quantitativi		**si**	**si**	**si**

Moda

La **moda** di un collettivo, distribuito secondo un carattere di qualsiasi natura, è la modalità prevalente del carattere ossia quella **modalità a cui è associata la massima frequenza.**

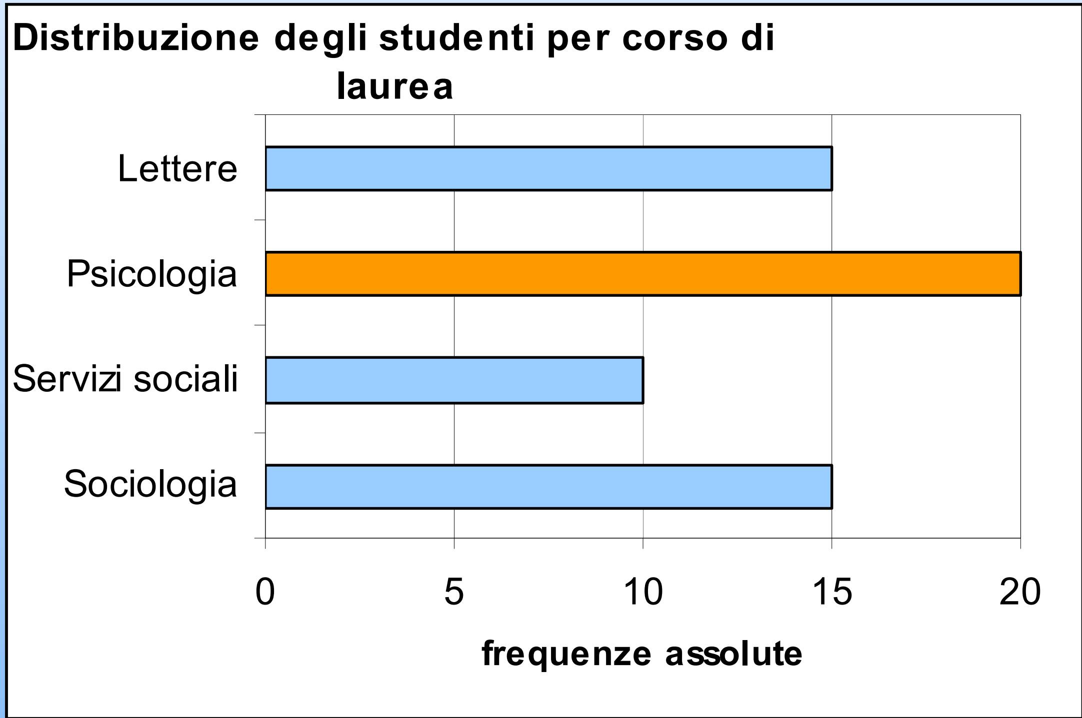
Quando il carattere è quantitativo e le modalità sono **raggruppate in classi**, si parla di **classe modale** con riferimento alla classe avente la **densità di frequenza più elevata**.

Si possono avere distribuzioni *unimodali, bimodali, multimodali e zeromodali*

Moda esempio (1)

Carattere qualitativo sconnesso

Corso di laurea x_i	Frequenze assolute n_i
Sociologia	15
Servizi sociali	10
Psicologia	**20**
Lettere	15
Totale	**60**



Distribuzione unimodale

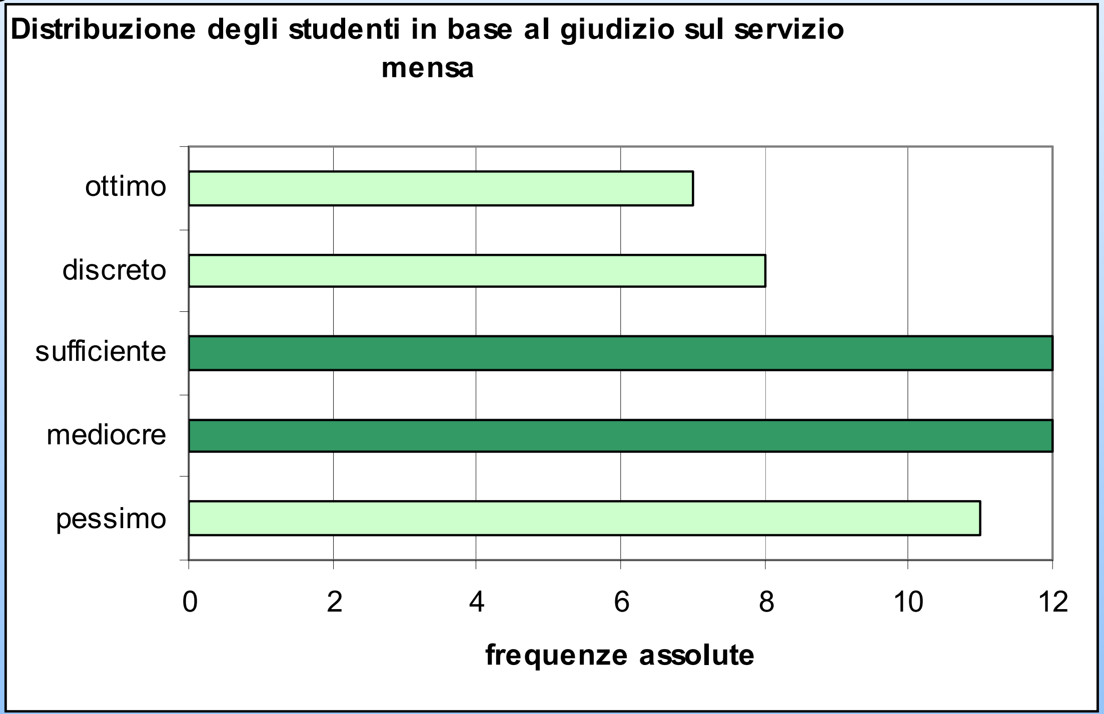
Moda: **Psicologia**

Frequenza massima: *20*

Moda esempio (2)

Carattere qualitativo ordinabile

Giudizio sul servizio mensa x_i	Frequenze assolute n_i
pessimo	11
mediocre	**12**
sufficiente	**12**
discreto	8
ottimo	7
Totale	**50**



Distribuzione bimodale

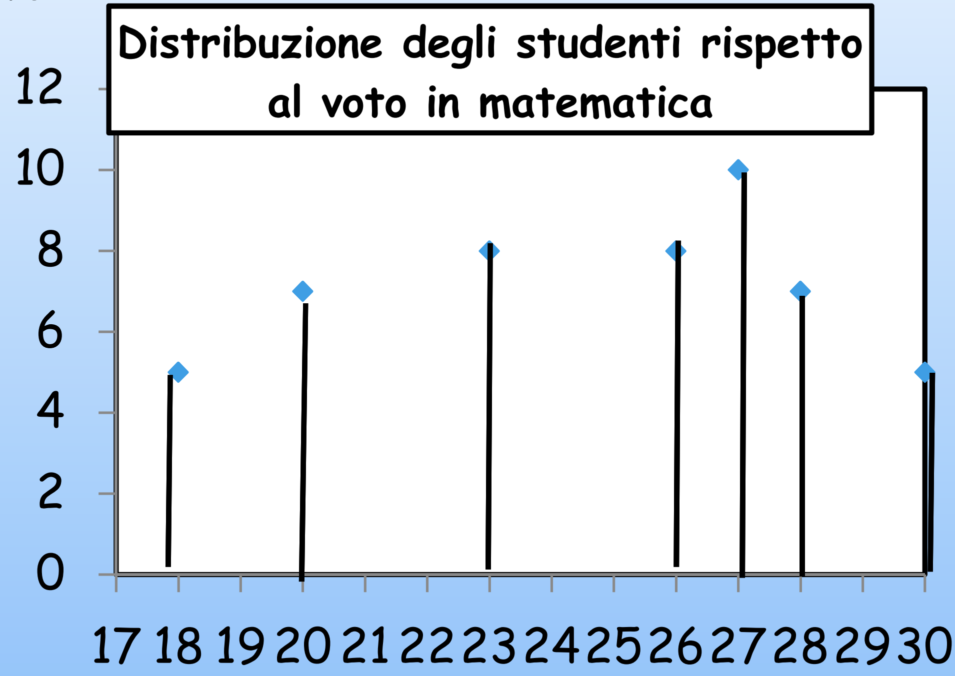
Mode: **mediocre** e **sufficiente**

Frequenza massima: ***12***

Moda esempio (3)

Carattere quantitativo discreto

Voto esame di matematica x_i	Frequenze assolute n_i
18	5
20	7
23	8
26	8
27	**10**
28	7
30	5
Totale	**50**



Distribuzione unimodale

Moda: **27**

Frequenza massima: ***10***

Moda per una distribuzione di frequenze a modalità singole

Distribuzione di frequenze della lunghezza dell'avambraccio (in cm) in 140 soggetti:

Modalità	*Frequenza*
41	3
42	2
43	6
44	11
45	8
46	17
47	**21**
48	14
49	17
50	15
51	10
52	10
53	5
54	1
Totale	140

- Moda: 47

Perché è la modalità con la frequenza più elevata (21)

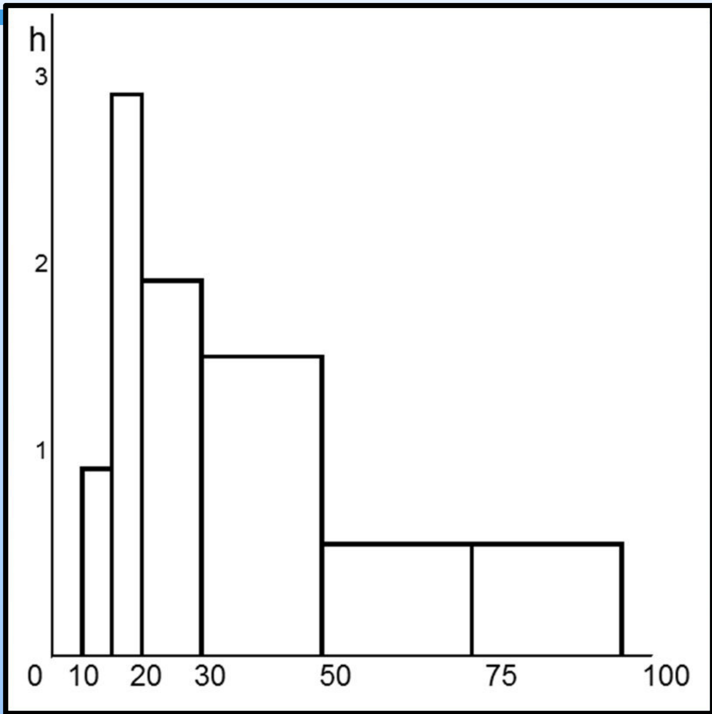
Carattere quantitativo continuo in classi: Classe modale

Classi di peso (in Kg) c_{i-1}|-c_i	Frequenza assoluta n_i	Ampiezza di classe d_i	Densità di frequenza h_i
10 |-- 15	5	5	1
15 |-- 20	15	5	**3**
20 |-- 30	20	10	2
30 |-- 50	30	20	1,5
50 |-- 75	15	25	0,6
75 |-- 100	15	25	0,6
Totale	100		

$$h_i = \frac{n_i}{d_i}$$

Densità Massima

$$d_i = c_i - c_{i-1}$$



Per determinare la moda è necessario calcolare la densità di frequenza h_i, data dal rapporto fra frequenza assoluta n_i e ampiezza di classe d_i.

<u>La classe modale è la classe con maggiore densità di frequenza</u>

<u>Distribuzione unimodale</u>

Classe modale: 15|--20

Densità di frequenza massima: ***3***

Classe modale: esempio

Distribuzione di frequenze degli studenti di un corso di Statistica secondo l'età:

Classe di età	Studenti n_i	$d_i = c_i - c_{i-1}$	$hi = n_i / d_i$
19-21	31	2	15.5
21-24	45	3	15
24-27	5	3	1,7
27-30	1	3	0,3
Totale	82		

- La classe modale è 19-21

Perché ha la densità di frequenza più elevata

Mediana

La **mediana** è quella modalità che suddivide ogni **distribuzione ordinata** in due distribuzioni aventi ciascuna una numerosità (o una quantità) che è il 50% della numerosità (o della quantità) della distribuzione totale.

Si può calcolare per i caratteri qualitativi ordinabili e quantitativi

Sia $x_1, x_2, \ldots, x_N$, una distribuzione statistica disaggregata.

Sia $y_1, y_2, \ldots, y_N$, con $y_1 \le y_2 \le \ldots, \le y_N$, la corrispondente **distribuzione dei termini ordinati.**

Mediana per N dispari

➢ **<u>N dispari</u>** : la mediana è la modalità che nella distribuzione ordinata occupa il posto $\frac{N+1}{2}$

Distribuzione disaggregata dei voti riportati a N=9 esami

$x_1=21 \quad x_2=18 \quad x_3=28 \quad x_4=27 \quad x_5=30 \quad x_6=28 \quad x_7=30 \quad x_8=25 \quad x_9=22$

Distribuzione ordinata

$y_1=18 \quad y_2=21 \quad y_3=22 \quad y_4=25 \quad y_5=27 \quad y_6=28 \quad y_7=28 \quad y_8=30 \quad y_9=30$

Posizione occupata dalla mediana: $$\frac{\mathbf{N+1}}{\mathbf{2}} = \frac{\mathbf{9+1}}{\mathbf{2}} = \mathbf{5}$$

Mediana: $$\mathbf{y}_{\frac{\mathbf{N+1}}{\mathbf{2}}} = \mathbf{y}_5 = 27$$

Mediana per N pari

N pari : si hanno due modalità mediane che nella **distribuzione ordinata** occupano rispettivamente i posti $\frac{N}{2}$ e $\frac{N}{2}+1$

Distribuzione disaggregata dei voti riportati a N=10 esami

x_1=21 x_2=18 x_3=28 x_4=27 x_5=30 x_6=28 x_7=30 x_8=25 x_9=22 x_{10}=28

Distribuzione ordinata

y_1=18 y_2=21 y_3=22 y_4=25 y_5=27 y_6=28 y_7=28 y_8=28 y_9=30 x_{10}=30

Posizioni occupate dalle mediane

$$\mathbf{\frac{N}{2}=\frac{10}{2}=5;\ \frac{N}{2}+1=5+1=6}$$

Mediane

$$\mathbf{y}_{\frac{\mathbf{N}}{\mathbf{2}}}=\mathbf{y}_{\mathbf{5}}=27;\ \mathbf{y}_{\frac{\mathbf{N}}{\mathbf{2}}+1}=\mathbf{y}_{\mathbf{6}}=28$$

Mediana

$$\frac{\mathbf{y}_{\frac{\mathbf{N}}{\mathbf{2}}}+\mathbf{y}_{\frac{\mathbf{N}}{\mathbf{2}}+1}}{2}=\frac{27+28}{2}=27.5$$

Calcolo della Mediana

Se ***N* è dispari**, si chiama mediana della distribuzione la quantità, *m*, che occupa il posto centrale, cioè il posto $(N+1)/2$, della **graduatoria dei termini ordinati**.

$$m = y_{\frac{N+1}{2}}$$

Se ***N* è pari**, si chiamano mediane della distribuzione la quantità, m_1 e m_2 , che occupano le due posizioni centrali della graduatoria dei termini ordinati, ossia le posizioni $N/2$ e $N/2 + 1$. della **graduatoria dei termini ordinati**.

$$m_1 = y_{\frac{N}{2}}; \quad m_2 = y_{\frac{N}{2}+1}$$

Calcolo della Mediana

Se *N* è **pari**, **<u>e il carattere è quantitativo</u>** si può assumere come mediana la media aritmetica dei termini che occupano le due posizioni centrali della graduatoria dei termini ordinati, ossia le posizioni *N*/2 e *N*/2 + 1.

$$m = \frac{y_{\frac{N}{2}} + y_{\frac{N}{2}+1}}{2}$$

Esempio: Mediana per caratteri qualitativi ordinabili - N dispari

Distribuzione disaggregata:

Sufficiente, Pessimo, Insufficiente, Ottimo, Sufficiente, Insufficiente, Ottimo

Distribuzione ordinata:

Pessimo, Insufficiente, Insufficiente, Sufficiente, Sufficiente, Ottimo, Ottimo

Posizione occupata dalla mediana

$$\frac{N+1}{2} = \frac{7+1}{2} = 4$$

Mediana:

$$y_{\frac{N+1}{2}} = y_4 = \text{Sufficiente}$$

Esempio: Mediana per caratteri qualitativi ordinabili - N pari

Distribuzione disaggregata:

Licenza media, Diploma, Diploma, Laurea, Licenza media, Licenza elementare

Distribuzione ordinata:

Licenza elementare, Licenza media, Licenza media, Diploma, Diploma, Laurea,

Posizioni occupate dalle mediana

$$\frac{\mathbf{N}}{\mathbf{2}} = \frac{6}{\mathbf{2}} = \mathbf{3}; \quad \frac{\mathbf{N}}{\mathbf{2}} + 1 = \frac{6}{\mathbf{2}} + 1 = \mathbf{4};$$

Mediane

$$\mathbf{y}_{\frac{\mathbf{N+1}}{\mathbf{2}}} = \mathbf{y}_{\mathbf{3}} = \text{Licenza media;}$$

$$\mathbf{y}_{\frac{\mathbf{N}}{\mathbf{2}}+\mathbf{1}} = \mathbf{y}_{\mathbf{4}} = \text{Diploma}$$

Esempio: Mediana per caratteri quantitativi discreti - N dispari

Ritardi (in minuti) di un treno a lunga percorrenza alla stazione di Roma Termini, registrati in un campione di 7 osservazioni (i valori vengono ordinati in senso crescente

0, 9, 5, 6, 8, 10, 12

- Mediana

  - Fase 1: ordinamento dei termini

0, 5, 6, 8, 9, 10, 12

8 è il termine che occupa il quarto posto

  - Fase 2: individuazione del posto centrale della graduatoria

$$\frac{N+1}{2} = \frac{7+1}{2} = 4$$

  - Fase 3: individuazione della mediana

$$m = 8$$

Esempio: Mediana per caratteri quantitativi discreti - N pari

Quotazioni di borsa di un titolo azionario in 8 sedute successive:

12.8, 13.0, 13.4, 13.4 , 13.6, 13.5, 13.6, 13.7

- Mediana

  - Fase 1: ordinamento dei termini

12.8, 13.0, 13.4, 13.4 , 13.5, 13.6, 13.6, 13.7

13.4 e 13.5 sono i termini che occupano i posti quarto e quinto

  - Fase 2: individuazione dei posti centrali della graduatoria:

$$\frac{N}{2} = \frac{8}{2} = 4, \quad \frac{N}{2} + 1 = \frac{8}{2} + 1 = 5$$

  - Fase 3: individuazione della mediana

$$m = \frac{13.4 + 13.5}{2} = 13.45$$

Proprietà della mediana

- **INTERNALITA'**: se a e b sono il minimo e il massimo dell'insieme dei numeri $x_1, x_2, \ldots, x_N$, la mediana è compresa tra queste due quantità: $a \leq m \leq b$
- La somma degli scarti in valore assoluto dei valori $x_1, x_2, \ldots, x_N$ da una costante c è minima quando c è uguale alla mediana

Quartili

Sia $x_1, x_2, \ldots, x_N$ una distribuzione disaggregata.

Sia $y_1, y_2, \ldots, y_N$ la corrispondente distribuzione di termini ordinati, con $y_1 \leq y_2 \leq \ldots \leq y_N$.

- Il **primo quartile**, q_1, è la quantità che non è superata da un quarto (25%) dei termini ordinati della distribuzione
- Il **secondo quartile**, q_2, è la quantità che non è superata dalla metà (50%) dei termini ordinati.
- Il **terzo quartile**, q_3, è la quantità che non è superata dai tre quarti (75%) dei termini ordinati della distribuzione.

N.B.: Il secondo quartile coincide con la mediana

Quartili: calcolo quando *N* è multiplo di 4

Quotazioni di borsa di un titolo azionario in 8 sedute successive:

12.8, 13.0, 13.4, 13.4, 13.6, 13.5, 13.6, 13.7

- Fase 1: ordinamento dei termini

12.8, 13.0, 13.4, 13.4, 13.5, 13.6, 13.6, 13.7

Quartili: calcolo quando N è multiplo di 4

- Fase 2: suddivisione dei termini ordinati in quattro gruppi della stessa numerosità

12.8, 13.0 | 13.4, 13.4 | 13.5, 13.6 | 13.6, 13.7

- Fase 3: individuazione dei quartili:

  il primo quartile è la media aritmetica dei termini 13.0 e 13.4, cioè 13.20,

  il secondo quartile è la media aritmetica tra 13.4 e 13.5, cioè 13.45,

  il terzo quartile è la media aritmetica tra 13.6 e 13.6, cioè 13.6.

Quartili: definizione operativa

Sia $x_1, x_2, \ldots, x_N$ una distribuzione disaggregata.

Sia $y_1, y_2, \ldots, y_N$ la corrispondente distribuzione di termini ordinati, con $y_1 \leq y_2 \leq \ldots \leq y_N$.

Primo quartile : Posto

$$h = N \cdot \frac{1}{4}$$

q_1 è dato da

$$q_1 = \begin{cases} y_h \;\; e \;\; y_{h+1} & \text{se } h \text{ è un numero intero, se y è quantitativo } q_1 = \dfrac{y_h + y_{h+1}}{2} \\ y_{[h]+1}, & \text{altrimenti} \end{cases}$$

dove $[h]$ è la parte intera di h.

Quartili: definizione operativa

Secondo quartile: q_2 si determina con lo stesso procedimento, ponendo

$$h = N \cdot \frac{2}{4}$$

Terzo quartile: q_3 si determina con lo stesso procedimento, ponendo

$$h = N \cdot \frac{3}{4}$$

Quartili: calcolo per un *N* qualsiasi

Previsione da parte di 14 economisti della variazione media percentuale dei prezzi al consumo per il prossimo anno:

2.1, 2.2, 1.8, 2.4, 2.5, 2.8, 2.1, 2.2, 2.1, 1.9, 1.8, 2.4, 2.9, 2.4

- **Fase 1: ordinamento dei termini**

1.8, 1.8, 1.9, 2.1, 2.1, 2.1, 2.2, 2.2, 2.4, 2.4, 2.4, 2.5, 2.8, 2.9

Quartili: calcolo per un N qualsiasi
Carattere quantitativo

$$q = \begin{cases} \dfrac{y_h + y_{h+1}}{2}, & \text{se } h \text{ è un numero intero} \\ y_{[h]+1}, & \text{altrimenti} \end{cases}$$

1.8, 1.8, 1.9, 2.1, 2.1, 2.1, 2.2, 2.2, 2.4, 2.4, 2.4, 2.5, 2.8, 2.9

- **Fase 2: individuazione dei quartili**

Primo quartile:

$$h = N \cdot \frac{1}{4} = 14/4 = 3.5 \Rightarrow [h] = 3 \Rightarrow q_1 = y_4 = 2.1$$

Secondo quartile:

$$h = N \cdot \frac{2}{4} = 14 \cdot 2/4 = 7 \Rightarrow q_2 = \frac{y_7 + y_8}{2} = \frac{2.2 + 2.2}{2} = 2.2$$

Terzo quartile:

$$h = N \cdot \frac{3}{4} = 14 \cdot 3/4 = 10.5 \Rightarrow [h] = 10, \Rightarrow q_3 = y_{11} = 2.4$$

Decili: definizione

In termini discorsivi, i decili si possono definire come medie di posizione tali che:

Il primo decile: è la quantità che non è superata da un decimo (10%) dei termini ordinati

Il secondo decile: è la quantità che non è superata da due decimi (20%) dei termini ordinati

...

N.B.: i decili sono 9.

N.B.: Il quinto decile coincide con la mediana

Il calcolo dei decili si effettua con lo stesso procedimento descritto per i quartili

Decili: definizione operativa

Sia $x_1, x_2, \ldots, x_N$ una distribuzione disaggregata.

Sia $y_1, y_2, \ldots, y_N$ la corrispondente distribuzione di termini ordinati, con $y_1 \leq y_2 \leq \ldots \leq y_N$.

Primo decile: Posto

$$h = N \cdot \frac{1}{10}$$

il primo decile, d_1, è dato da

$$d_1 = \begin{cases} y_h \text{ e } y_{h+1} & \text{se } h \text{ è un numero intero, se } y \text{ è quantitativo } d_1 = \dfrac{y_h + y_{h+1}}{2} \\ y_{[h]+1}, & \text{altrimenti} \end{cases}$$

dove $[h]$ è la parte intera di h.

Decili: definizione operativa

Secondo decile: Posto

$$h = N \cdot \frac{2}{10}$$

il primo decile, d_2, è dato da

$$d_2 = \begin{cases} y_h \ \text{ e } \ y_{h+1} & \text{se h è un numero intero, se y è quantitativo } d_2 = \dfrac{y_h + y_{h+1}}{2} \\ y_{[h]+1}, & \text{altrimenti} \end{cases}$$

dove $[h]$ è la parte intera di h.

...

Mediana e quartili nel caso delle distribuzioni di frequenze a modalità singole

Per il calcolo della mediana e dei quartili in caso di distribuzione di frequenze sono immediatamente applicabili le formule viste in precedenza.

<u>Occorre tenere presente che i posti in graduatoria delle diverse modalità si si deducono dalle frequenze cumulate.</u>

Quartili per una distribuzione di frequenze a modalità quantitative singole: calcolo

$$q = \begin{cases} \dfrac{y_h + y_{h+1}}{2}, & \text{se } h \text{ è un numero intero} \\ y_{[h]+1}, & \text{altrimenti} \end{cases}$$

Distribuzione di frequenze della lunghezza dell'avambraccio (in cm)

x_i	n_i	N_i
41	3	3
42	2	5
43	6	11
44	11	22
45	8	30
46	17	47
47	21	68
48	14	82
49	17	99
50	15	114
51	10	124
52	10	134
53	5	139
54	1	140
Totale	140	

35
36

- **Primo quartile:**

$$h = N \cdot \frac{1}{4}$$

$$h = 140 \cdot \frac{1}{4} = 35 \Rightarrow q_1 = \frac{y_{35} + y_{36}}{2}$$

$$= \frac{46 + 46}{2} = 46$$

perché dalle frequenze cumulate si evince che il 35-esimo e il 36-esimo posto in graduatoria sono occupati dal termine 46.

Quartili per una distribuzione di frequenze a modalità quantitative singole: calcolo

$$q = \begin{cases} \dfrac{y_h + y_{h+1}}{2}, & \text{se } h \text{ è un numero intero} \\ y_{[h]+1}, & \text{altrimenti} \end{cases}$$

Distribuzione di frequenze della lunghezza dell'avambraccio (in cm)

x_i	n_i	N_i
41	3	3
42	2	5
43	6	11
44	11	22
45	8	30
46	17	47
47	21	68
48	14	82
49	17	99
50	15	114
51	10	124
52	10	134
53	5	139
54	1	140
Totale	140	

70
71

- **La mediana:** $h = N \cdot \dfrac{2}{4}$

$$h = 140 \cdot \frac{2}{4} = 70 \Rightarrow q_2 = m = \frac{y_{70} + y_{71}}{2} = \frac{48+48}{2} = 48$$

perché dalle frequenze cumulate si evince che il 70-esimo e il 71-esimo posto nella graduatoria dei termini della distribuzione sono occupati dal termine 48.

Quartili per una distribuzione di frequenze a modalità quantitative singole: calcolo

$$q = \begin{cases} \dfrac{y_h + y_{h+1}}{2}, & \text{se } h \text{ è un numero intero} \\ y_{[h]+1}, & \text{altrimenti} \end{cases}$$

Distribuzione di frequenze della lunghezza dell'avambraccio (in cm)

x_i	n_i	N_i
41	3	3
42	2	5
43	6	11
44	11	22
45	8	30
46	17	47
47	21	68
48	14	82
49	17	99
50	15	114
51	10	124
52	10	134
53	5	139
54	1	140
Totale	140	

105
106

❑ **Terzo quartile:** $h = N \cdot \dfrac{3}{4}$

$$h = 140 \cdot \frac{3}{4} = 105 \Rightarrow q_3 = \frac{y_{105} + y_{106}}{2}$$

$$= \frac{50 + 50}{2} = 50$$

perché dalle frequenze cumulate si evince che il 105-esimo e il 106-esimo posto nella graduatoria dei termini della distribuzione sono occupati dal termine 50.

Decili per una distribuzione di frequenze a modalità quantitative singole: calcolo

$$d = \begin{cases} \dfrac{y_h + y_{h+1}}{2}, & \text{se } h \text{ è un numero intero} \\ y_{[h]+1}, & \text{altrimenti} \end{cases}$$

Distribuzione di frequenze della lunghezza dell'avambraccio (in cm)

x_i	n_i	N_i
4 1	3	3
42	2	5
43	6	11
44	11	22
45	8	30
46	17	47
47	21	68
48	14	82
49	17	99
50	15	114
5 1	10	124
52	10	134
53	5	139
54	1	140
Totale	140	

42
43

$$h = N \cdot \frac{3}{10}$$

- **Terzo decile:**

$$h = 140 \cdot \frac{3}{10} = 42 \Rightarrow d_3 = \frac{y_{42} + y_{43}}{2}$$

$$= \frac{46 + 46}{2} = 46$$

In modo analogo si determinano gli altri decili.

Mediana e quartili: definizione operativa basata sulle frequenze percentuali cumulate

x_i	n_i	N_i	P_i
41	3	3	2.1%
42	2	5	3.6%
43	6	11	7.9%
44	11	22	15.7%
45	8	30	21.4%
46	17	47	33.6%
47	21	68	48.6%
48	14	82	58.6%
49	17	99	70.7%
50	15	114	81.4%
51	10	124	88.6%
52	10	134	95.7%
53	5	139	99.3%
54	1	140	100.0%
Totale	140		

$q_1 \rightarrow 46$

$m=q_2 \rightarrow 48$

$q_3 \rightarrow 50$

25%

50%

75%

Distribuzione di frequenze della lunghezza dell'avambraccio (in cm)

Decili: definizione operativa basata sulle frequenze percentuali cumulate

	x_i	n_i	N_i	P_i	
	41	3	3	**2.1%**	
	42	2	5	**3.6%**	
	43	6	11	**7.9%**	10%
d_1 →	44	11	22	**15.7%**	20%
d_2 →	45	8	30	**21.4%**	
d_3 →	46	17	47	**33.6%**	30%
d_4 →	47	21	68	**48.6%**	40%
$m=d_5$ →	48	14	82	**58.6%**	60%;70%
$d_6;d_7$ →	49	17	99	**70.7%**	
d_8 →	50	15	114	**81.4%**	80%
	51	10	124	**88.6%**	
d_9 →	52	10	134	**95.7%**	90%
	53	5	139	**99.3%**	
	54	1	140	**100.0%**	
	Totale	**140**			

Distribuzione di frequenze della lunghezza dell'avambraccio (in cm)

Primo quartile: esempio

Distribuzione di frequenze dei voti di 150 studenti all'esame di statistica

x_i	n_i	N_i	P_i
18	3	3	2.0%
19	2	5	3.3%
20	10	15	10.0%
21	7	22	14.7%
22	11	***33***	22.0%
23	13	***46***	30.7%
24	29	75	50.0%
25	17	92	61.3%
26	8	100	66.7%
27	15	115	76.7%
28	10	125	83.3%
29	10	135	90.0%
30	15	150	100.0%
Totale	**150**		

37
25%

- Primo quartile:

$$h = 150 \cdot \frac{1}{4} = 37.5 \Rightarrow$$

$$\Rightarrow q_1 = y_{[37.5]} = y_{37} = 23$$

perché dalle frequenze cumulate si evince che il 37-esimo posto in graduatoria è occupato dal termine 23.

Mediana : esempio

Distribuzione di frequenze dei voti di 150 studenti all'esame di statistica

x_i	n_i	N_i	P_i
18	3	3	2.0%
19	2	5	3.3%
20	10	15	10.0%
21	7	22	14.7%
22	11	33	22.0%
23	13	46	30.7%
24	29	*75*	**50.0%**
25	17	*92*	61.3%
26	8	100	66.7%
27	15	115	76.7%
28	10	125	83.3%
29	10	135	90.0%
30	15	150	100.0%
Totale	150		

75

76

- Mediana:

$$h = 150 \cdot \frac{2}{4} = 75 \Rightarrow$$

$$\Rightarrow m = \frac{y_{75} + y_{76}}{2} = \frac{24 + 25}{2} = 24.5$$

perché dalle frequenze cumulate si evince che il 75-esimo posto in graduatoria è occupato dal termine 24 ed il 76-esimo dal termine 25.

Terzo quartile: esempio

Distribuzione di frequenze dei voti di 150 studenti all'esame di statistica

x_i	n_i	N_i	P_i
18	3	3	2.0%
19	2	5	3.3%
20	10	15	10.0%
21	7	22	14.7%
22	11	33	22.0%
23	13	46	30.7%
24	29	75	50.0%
25	17	92	61.3%
26	8	*100*	66.7%
27	15	*115*	76.7%
28	10	125	83.3%
29	10	135	90.0%
30	15	150	100.0%
Totale	150		

112
75%

- Terzo quartile:

$$h = 150 \cdot \frac{3}{4} = 112.5 \Rightarrow$$

$$\Rightarrow q_3 = y_{[112.5]} = y_{112} = 27$$

perché dalle frequenze cumulate si evince che il 112-esimo posto in graduatoria è occupato dal termine 27.

Primo decile: esempio

Distribuzione di frequenze dei voti di 150 studenti all'esame di statistica

x_i	n_i	N_i	P_i
18	3	3	2.0%
19	2	5	3.3%
20	10	*15*	10.0%
21	7	*22*	14.7%
22	11	33	22.0%
23	13	46	30.7%
24	29	75	50.0%
25	17	92	61.3%
26	8	100	66.7%
27	15	115	76.7%
28	10	125	83.3%
29	10	135	90.0%
30	15	150	100.0%
Totale	**150**		

15

16

Primo decile:

$$h = 150 \cdot \frac{1}{10} = 15 \Rightarrow$$

$$\Rightarrow d_1 = \frac{y_{15} + y_{16}}{2} = \frac{20+21}{2} = 20.5$$

perché dalle frequenze cumulate si evince che il 15-esimo posto in graduatoria è occupato dal termine 20 ed il 16-esimo dal termine 21.

Secondo decile: esempio

Distribuzione di frequenze dei voti di 150 studenti all'esame di statistica

x_i	n_i	N_i	P_i
18	3	3	2.0%
19	2	5	3.3%
20	10	15	10.0%
21	7	*22*	14.7%
22	11	*33*	22.0%
23	13	46	30.7%
24	29	75	50.0%
25	17	92	61.3%
26	8	100	66.7%
27	15	115	76.7%
28	10	125	83.3%
29	10	135	90.0%
30	15	150	100.0%
Totale	150		

30,31; 20%

- Secondo decile:

$$h = 150 \cdot \frac{2}{10} = 30 \Rightarrow$$

$$\Rightarrow d_2 = \frac{y_{30} + y_{31}}{2} = \frac{22+22}{2} = 22$$

perché dalle frequenze cumulate si evince che il 30-esimo ed il 31-esimo posti in graduatoria sono occupati dal termine 22.

Terzo decile: esempio

Distribuzione di frequenze dei voti di 150 studenti all'esame di statistica

x_i	n_i	N_i	P_i
18	3	3	2.0%
19	2	5	3.3%
20	10	15	10.0%
21	7	22	14.7%
22	11	***33***	22.0%
23	13	***46***	30.7%
24	29	75	50.0%
25	17	92	61.3%
26	8	100	66.7%
27	15	115	76.7%
28	10	125	83.3%
29	10	135	90.0%
30	15	150	100.0%
Totale	**150**		

45,46;
30%

- **Terzo decile:**

$$h = 150 \cdot \frac{3}{10} = 45 \Rightarrow$$

$$\Rightarrow d_3 = \frac{y_{45} + y_{46}}{2} = \frac{23+23}{2} = 23$$

perché dalle frequenze cumulate si evince che il 45-esimo ed il 46-esimo posti in graduatoria sono occupati dal termine 23.

Quarto decile: esempio

Distribuzione di frequenze dei voti di 150 studenti all'esame di statistica

x_i	n_i	N_i	P_i
18	3	3	2.0%
19	2	5	3.3%
20	10	15	10.0%
21	7	22	14.7%
22	11	33	22.0%
23	13	*46*	30.7%
24	29	*75*	50.0%
25	17	92	61.3%
26	8	100	66.7%
27	15	115	76.7%
28	10	125	83.3%
29	10	135	90.0%
30	15	150	100.0%
Totale	150		

60,61;
40%

- Quarto decile:

$$h = 150 \cdot \frac{4}{10} = 60 \Rightarrow$$

$$\Rightarrow d_4 = \frac{y_{60} + y_{61}}{2} = \frac{24 + 24}{2} = 24$$

perché dalle frequenze cumulate si evince che il 60-esimo ed il 61-esimo posti in graduatoria sono occupati dal termine 24.

Sesto decile: esempio

Distribuzione di frequenze dei voti di 150 studenti all'esame di statistica

x_i	n_i	N_i	P_i
18	3	3	2.0%
19	2	5	3.3%
20	10	15	10.0%
21	7	22	14.7%
22	11	33	22.0%
23	13	46	30.7%
24	29	*75*	50.0%
25	17	*92*	61.3%
26	8	100	66.7%
27	15	115	76.7%
28	10	125	83.3%
29	10	135	90.0%
30	15	150	100.0%
Totale	150		

90,91;
60%

- Sesto decile:

$$h = 150 \cdot \frac{6}{10} = 90 \Rightarrow$$

$$\Rightarrow d_6 = \frac{y_{90} + y_{91}}{2} = \frac{25 + 25}{2} = 25$$

perché dalle frequenze cumulate si evince che il 90-esimo ed il 91-esimo posti in graduatoria sono occupati dal termine 25.

Settimo decile: esempio

Distribuzione di frequenze dei voti di 150 studenti all'esame di statistica

x_i	n_i	N_i	P_i
18	3	3	2.0%
19	2	5	3.3%
20	10	15	10.0%
21	7	22	14.7%
22	11	33	22.0%
23	13	46	30.7%
24	29	75	50.0%
25	17	92	61.3%
26	8	***100***	66.7%
27	15	***115***	76.7%
28	10	125	83.3%
29	10	135	90.0%
30	15	150	100.0%
Totale	150		

105,106;
70%

- Settimo decile:

$$h = 150 \cdot \frac{7}{10} = 105 \Rightarrow$$

$$\Rightarrow d_7 = \frac{y_{105} + y_{106}}{2} = \frac{27+27}{2} = 27$$

perché dalle frequenze cumulate si evince che il 105-esimo ed il 106-esimo posti in graduatoria sono occupati dal termine 27.

Ottavo decile: esempio

Distribuzione di frequenze dei voti di 150 studenti all'esame di statistica

x_i	n_i	N_i	P_i
18	3	3	2.0%
19	2	5	3.3%
20	10	15	10.0%
21	7	22	14.7%
22	11	33	22.0%
23	13	46	30.7%
24	29	75	50.0%
25	17	92	61.3%
26	8	100	66.7%
27	15	***115***	76.7%
28	10	***125***	83.3%
29	10	135	90.0%
30	15	150	100.0%
Totale	**150**		

120,121;
80%

- Settimo decile:

$$h = 150 \cdot \frac{8}{10} = 120 \Rightarrow$$

$$\Rightarrow d_8 = \frac{y_{120} + y_{121}}{2} = \frac{28+28}{2} = 28$$

perché dalle frequenze cumulate si evince che il 120-esimo ed il 121-esimo posti in graduatoria sono occupati dal termine 28.

Nono decile: esempio

Distribuzione di frequenze dei voti di 150 studenti all'esame di statistica

x_i	n_i	N_i	P_i
18	3	3	2.0%
19	2	5	3.3%
20	10	15	10.0%
21	7	22	14.7%
22	11	33	22.0%
23	13	46	30.7%
24	29	75	50.0%
25	17	92	61.3%
26	8	100	66.7%
27	15	115	76.7%
28	10	125	83.3%
29	10	*135*	90.0%
30	15	*150*	100.0%
Totale	150		

135

136

Nono decile:

$$h = 150 \cdot \frac{9}{10} = 135 \Rightarrow$$

$$\Rightarrow d_9 = \frac{y_{135} + y_{136}}{2} = \frac{29+30}{2} = 29.5$$

perché dalle frequenze cumulate si evince che il 135-esimo posto in graduatoria è occupato dal termine 29 ed il 136-esimo dal termine 30.

Decili: esempio

Distribuzione di frequenze dei voti di 150 studenti all'esame di statistica

x_i	n_i	N_i	P_i
18	3	3	2.0%
19	2	5	3.3%
20	10	15	**10.0%**
21	7	22	14.7%
22	11	33	22.0%
23	13	46	30.7%
24	29	75	**50.0%**
25	17	92	61.3%
26	8	100	66.7%
27	15	115	76.7%
28	10	125	83.3%
29	10	135	**90.0%**
30	15	150	100.0%
Totale	150		

$d_1=20.5$

d_2

d_3

d_4

$m=d_5=24.5$

d_6

d_7

d_8

$d_9=29.5$

20%

30%

40%

60%

70%

80%

Mediana per le distribuzioni di frequenze con modalità raggruppate in classi

Il calcolo della mediana avviene in **due fasi**:

1. Si individua, prima, la **classe mediana**, con la stessa regola stabilita in precedenza, trattando le classi alla stregua di modalità di un carattere discreto
2. All'interno di questa si calcola la **mediana** con la formula

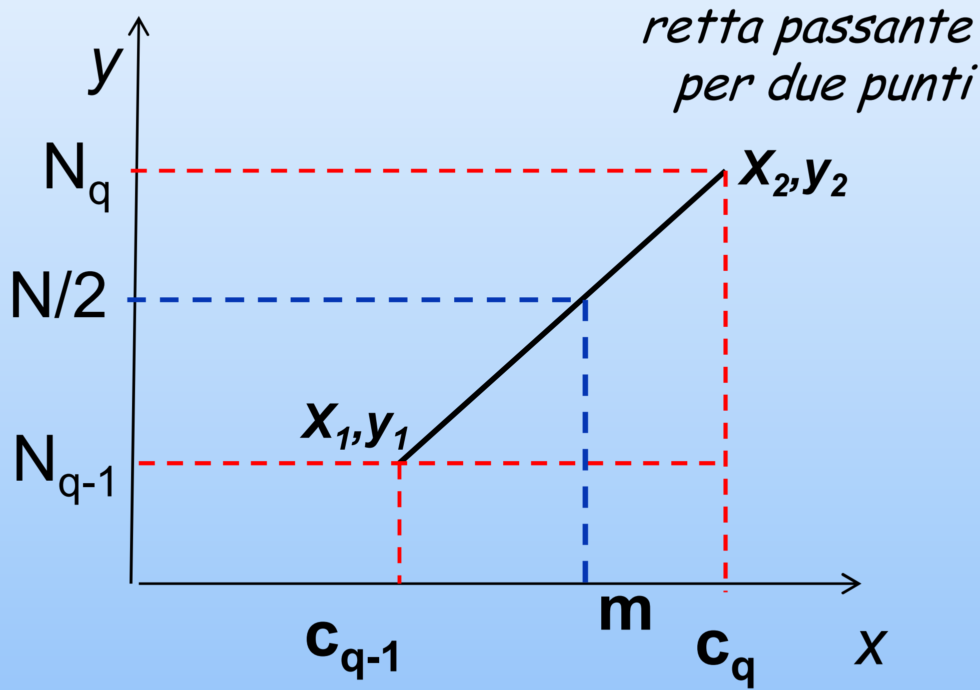
$$m = c_{q-1} + \frac{N/2 - N_{q-1}}{N_q - N_{q-1}}(c_q - c_{q-1})$$

La formula si basa sull'ipotesi di uniforme distribuzione delle unità nelle classi.

c_{q-1} e c_q sono gli estremi della classe mediana

N_{q-1} e N_q sono le frequenze cumulate nelle classi $q-1$ e q.

Spiegazione della formula della mediana per le distribuzioni di frequenze con modalità raggruppate in classi basata sull'equazione della retta passante per due punti



$$\frac{y-y_1}{y_2-y_1}=\frac{x-x_1}{x_2-x_1}$$

$$(x_1,y_1)=(c_{q-1},N_{q-1});$$

$$(x_2,y_2)=(c_q,N_q)$$

$$\frac{y-N_{q-1}}{N_q-N_{q-1}}=\frac{x-c_{q-1}}{c_q-c_{q-1}}$$

$$x=c_{q-1}+\frac{c_q-c_{q-1}}{N_q-N_{q-1}}\left(y-N_{q-1}\right)$$

$$y=\frac{N}{2}\Rightarrow m=c_{q-1}+\frac{N/2-N_{q-1}}{N_q-N_{q-1}}(c_q-c_{q-1})$$

Mediana per le distribuzioni di frequenze con modalità raggruppate in classi: calcolo

$$m = c_{q-1} + \frac{N/2 - N_{q-1}}{N_q - N_{q-1}}(c_q - c_{q-1})$$

Distribuzione di frequenze degli studenti di un corso di Statistica secondo l'età:

Classe	n_i	N_i
19-21	31	31
21-24	45	76
24-27	5	81
27-30	1	82
Totale	82	

$$h = 82 \cdot \frac{1}{2} = 41 \Rightarrow$$

La classe mediana è 21-24

- La mediana è data da

$$m = 21 + \frac{82/2 - 31}{76 - 31}(24 - 21) = 21.67$$

Quartili per le distribuzioni di frequenze con modalità raggruppate in classi

Il primo e terzo quartile si calcolano con lo stesso procedimento descritto per la mediana:

Le formule si basano sull'ipotesi di uniforme distribuzione delle unità nelle classi.

Primo quartile

$$q_1 = c_{q-1} + \frac{N/4 - N_{q-1}}{N_q - N_{q-1}}(c_q - c_{q-1})$$

c_{q-1} e c_q sono gli estremi della classe del primo quartile

N_{q-1} e N_q sono le frequenze cumulate nelle classi $q-1$ e q

Terzo quartile

$$q_3 = c_{q-1} + \frac{N(3/4) - N_{q-1}}{N_q - N_{q-1}}(c_q - c_{q-1})$$

c_{q-1} e c_q sono gli estremi della classe del terzo quartile

N_{q-1} e N_q sono le frequenze cumulate nelle classi $q-1$ e q

Primo quartile per le distribuzioni di frequenze con modalità raggruppate in classi: calcolo

$$q_1 = c_{q-1} + \frac{N/4 - N_{q-1}}{N_q - N_{q-1}}(c_q - c_{q-1})$$

Distribuzione di frequenze degli studenti di un corso di Statistica secondo l'età:

Classe	n_i	N_i
19-21	31	31
21-24	45	76
24-27	5	81
27-30	1	82
Totale	82	

$$h = 82 \cdot \frac{1}{4} = 20{,}5 \Rightarrow |h| = 20 \Rightarrow$$

La classe del primo quartile è 19-21

- Il primo quartile è dato da

$$q_1 = 19 + \frac{82/4 - 0}{31 - 0}(21 - 19) = 20.32$$

Terzo quartile per le distribuzioni di frequenze con modalità raggruppate in classi: calcolo

$$q_3 = c_{q-1} + \frac{N(3/4) - N_{q-1}}{N_q - N_{q-1}}(c_q - c_{q-1})$$

Distribuzione di frequenze degli studenti di un corso di Statistica secondo l'età:

Classe	n_i	N_i
19-21	31	31
21-24	45	76
24-27	5	81
27-30	1	82
Totale	82	

$$h = 82 \cdot \frac{3}{4} = 61{,}5 \Rightarrow |h| = 61 \Rightarrow$$

La classe del terzo quartile è 21-24

- Il terzo quartile è dato da

$$q_3 = 21 + \frac{82(3/4) - 31}{76 - 31}(24 - 21) = 23.03$$

Media aritmetica

La **media aritmetica** è quel valore di sintesi che sostituito alle modalità lascia inalterata la loro somma

- Insieme alle percentuali e ai grafici, la media aritmetica è lo strumento statistico più largamente utilizzato
- La media aritmetica di una **distribuzione statistica disaggregata** si calcola come la somma dei termini x_1, x_2, …, x_N divisa per N

$$\mu = \frac{x_1 + x_2 + \cdots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i .$$

Proprietà della media aritmetica (1)

La media aritmetica presenta le seguenti proprietà:

- **1)** Il prodotto $N \cdot \mu$ è uguale al totale del carattere nella distribuzione

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \Rightarrow N\mu = \sum_{i=1}^{N} x_i$$

- **2) Internalità:** Se a e b sono il minimo e il massimo dell'insieme x_1, x_2, …, x_N, la media aritmetica è compresa tra queste due quantità

$$a \le \mu \le b$$

Proprietà della media aritmetica (2)

- **3) La somma algebrica degli scarti dalla media aritmetica è uguale a zero** $\sum_{i=1}^{N}(x_i-\mu)=0$

dimostrazione

$$\sum_{i=1}^{N}(x_i-\mu)=\sum_{i=1}^{N}x_i-\sum_{i=1}^{N}\mu=\sum_{i=1}^{N}x_i-N\mu=\sum_{i=1}^{N}x_i-\sum_{i=1}^{N}x_i=0$$

$$\sum_{i=1}^{N}(x_i-c)=0\Rightarrow\sum_{i=1}^{N}x_i-\sum_{i=1}^{N}c=0;\ \sum_{i=1}^{N}x_i-Nc=0;\quad c=\frac{\sum_{i=1}^{N}x_i}{N}=\mu$$

Proprietà utilizzate

Sommatoria di una costante $\sum_{i=1}^{N}\mu=N\mu$

Dissociativa $\sum_{i=1}^{N}(x_i-c)=\sum_{i=1}^{N}x_i-\sum_{i=1}^{N}c$

La somma algebrica degli scarti dalla media aritmetica è uguale a zero: esempio

1	6	4	1

$$\sum_{i=1}^{4}(x_i-\bar{x})=(1-3)+(6-3)+(4-3)+(1-3)=$$
$$-2+3+1-2=0$$

Voto esame di matematica (x_i)	n_i	$x_i n_i$	$(x_i$-$x)$	$(x_i$-$x)ni$
18	5	90	18 - 24,76 = -6,76	-33,8
20	7	140	20 - 24,76 = -4,76	-33,32
23	8	184	23 - 24,76 = -1,76	-14,08
26	8	208	26 - 24,76 = 1,24	9,92
27	10	270	27 - 24,76 = 2,24	22,4
28	7	196	28 - 24,76 = 3,24	22,68
30	5	150	30 - 24,76 = 5,24	26,2
Totale	**50**	1238		**0**

$$\sum_{i=1}^{7}(x_i-\bar{x})n_i=0$$

Proprietà della media aritmetica (3)

- **4) La somma degli scarti al quadrato dei valori x_1, x_2, …, x_N da una costante c è minima quando c è uguale alla media aritmetica** $\sum_{i=1}^{N}(x_i-\mu)^2 = \min$

dimostrazione

$$S = \sum_{i=1}^{N}(x_i-c)^2$$

$$\frac{dS}{dc} = -2\sum_{i=1}^{N}(x_i-c) = 0 \Rightarrow \quad c = \frac{\sum_{i=1}^{N} x_i}{N} = \mu$$

$$\frac{d^2S}{dc^2} = 2N > 0$$

La somma degli scarti dalla media aritmetica al quadrato è un minimo

1	6	4	1

$$\sum_{i=1}^{4}(x_i-\mu)^2=(1-3)^2+(6-3)^2+(4-3)^2+(1-3)^2=4+9+1+4=18$$

Se al posto di 3 metto un qualsiasi altro valore,

questa somma sarà sempre >18

Proprietà della media aritmetica (4)

- **5) linearità** : Se ogni singolo termine della distribuzione, x_i , viene sottoposto alla trasformazione

$$a + bx_i,$$

con a costante qualsiasi e $b \neq 0$, la nuova media è legata a quella originaria dalla medesima trasformazione $a + b\mu$

dimostrazione

$$\frac{\sum_{i=1}^{N}(a+bx_i)}{N} = \frac{\sum_{i=1}^{N} a + \sum_{i=1}^{N} bx_i}{N} = \frac{Na + b\sum_{i=1}^{N} x_i}{N} = a + b\frac{\sum_{i=1}^{N} x_i}{N} = a + b\mu$$

Proprietà utilizzate

Sommatoria di una costante $\sum_{i=1}^{N} a = Na$

Moltiplicazione per una costante $\sum_{i=1}^{N} bx_i = b\sum_{i=1}^{N} x_i$

Proprietà della media aritmetica (5)

- **6) associatività** Se un collettivo statistico di N unità viene suddiviso in L sottoinsiemi disgiunti aventi numerosità $N^{(1)}$, $N^{(2)}$, …, $N^{(L)}$ e medie $\mu^{(1)}$, $\mu^{(2)}$, …, $\mu^{(L)}$, allora la media aritmetica del collettivo può essere così calcolata

$$\mu = \frac{1}{N}\left(\mu^{(1)} \cdot N^{(1)} + \mu^{(2)} \cdot N^{(2)} + \cdots + \mu^{(L)} \cdot N^{(L)}\right) = \frac{1}{N}\sum_{l=1}^{N} \mu^{(l)} \cdot N^{(l)}$$

Media aritmetica per una distribuzione disaggregata: calcolo

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i .$$

Serie storica dei divorzi in Italia nel periodo 2003-2009:

Anno	2003	2004	2005	2006	2007	2008	2009
N. di divorzi	43856	45097	47036	49534	50669	54351	54456

La media annua dei divorzi è data da

$$\mu = \frac{43856 + 45097 + 47036 + 49534 + 50669 + 54351 + 54456}{7}$$

$$= 49286.$$

Media aritmetica: verifica delle proprietà 3 e 4

- Proprietà 3: uguaglianza a zero della somma degli scarti

$$(43856-49286)+(45097-49286)+(47036-49286)+$$
$$+(49534-49286)+(50669-49286)+(54351-49286)+$$
$$+(54456-49286)=0.$$

- Proprietà 4: la somma dei quadrati degli scarti dalla media aritmetica è:

$$(43856-49286)^2+(45097-49286)^2+(47036-49286)^2+$$
$$+(49534-49286)^2+(50669-49286)^2+(54351-49286)^2+$$
$$(54456-49286)^2=106452438.$$

quantità inferiore a quella che si ottiene sostituendo a 49286 un altro numero qualsiasi.

Media aritmetica: verifica della proprietà 6

Quotazioni di borsa di un titolo azionario in 8 sedute successive:

12.8, 13.0, 13.4, 13.4 , 13.6, 13.5, 13.6, 13.7

- Proprietà 6: se suddividiamo la distribuzione data nelle due seguenti:

  A. 12.8, 13.0, 13.4, 13.4 , 13.6 $\mu_A = 13.240$

  B. 13.5, 13.6, 13.7 $\mu_B = 13.600$

la media aritmetica della distribuzione iniziale

$$\mu = \frac{12.8 + 13.0 + 13.4 + 13.4 + 13.6 + 13.5 + 13.6 + 13.7}{8} = 13.375.$$

può essere ottenuta come

$$\mu = \frac{13.240 \cdot 5 + 13.600 \cdot 3}{8} = 13.375.$$

Media aritmetica per le distribuzioni di frequenze

Modalità singole

$$\mu = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \ldots + x_k \cdot n_k}{N} = \frac{1}{N}\sum_{i=1}^{k} x_i \cdot n_i$$

$$= x_1 \cdot f_1 + x_2 \cdot f_2 + \cdots + x_k \cdot f_k = \sum_{i=1}^{k} x_i \cdot f_i$$

Modalità raggruppate in classi

$$\mu = \frac{\overline{x}_1 \cdot n_1 + \overline{x}_2 \cdot n_2 + \cdots + \overline{x}_k \cdot n_k}{N} = \frac{1}{N}\sum_{i=1}^{k} \overline{x}_i \cdot n_i$$

dove $\overline{x}_i = \frac{c_{i-1} + c_i}{2}$ è il **valore centrale** della generica classe.

Media aritmetica per una distribuzione di frequenze a modalità singole: calcolo

$$\mu = \frac{1}{N}\sum_{i=1}^{k} x_i \cdot n_i$$

Distribuzione di frequenze della lunghezza dell'avambraccio (in cm)

x_i	n_i	$x_i \cdot n_i$
41	3	123
42	2	84
43	6	258
44	11	484
45	8	360
46	17	782
47	21	987
48	14	672
49	17	833
50	15	750
51	10	510
52	10	520
53	5	265
54	1	54
Totale	140	6682

- La media aritmetica della distribuzione è data da:

$$\mu = \frac{41 \cdot 3 + 42 \cdot 2 + \cdots + 54 \cdot 1}{140}$$

$$= \frac{6682}{140} = 47.729$$

Media aritmetica per una distribuzione di frequenze a modalità raggruppate in classi: calcolo

$$\mu = \frac{1}{N}\sum_{i=1}^{k} \overline{x}_i \cdot n_i$$

Distribuzione di frequenze degli studenti di un corso di Statistica secondo l'età:

Classe	$\overline{x}_i$ Valore centrale	n_i	$\overline{x}_i \cdot n_i$
19-21	20.0	31	620
21-24	22.5	45	1012.5
24-27	25.5	5	127.5
27-30	28.5	1	28.5
Totale		82	1788.5

- La media aritmetica della distribuzione è data da:

$$\mu = \frac{20.0 \cdot 31 + 22.5 \cdot 45 + 25.5 \cdot 5 + 28.5 \cdot 1}{82}$$

$$= \frac{1788.5}{82} = 21.8$$

Media armonica

La media armonica di una distribuzione statistica disaggregata, $x_1, x_2, \ldots, x_N$, i cui termini sono tutti diversi da 0, è data dal rapporto tra N e la somma dei reciproci dei termini:

$$\mu_a = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_N}} = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}}.$$

La media armonica di una distribuzione statistica di frequenze, i cui termini sono tutti diversi da 0, è

$$\mu_a = \frac{N}{\frac{n_1}{x_1} + \frac{n_2}{x_2} + \cdots + \frac{n_k}{x_k}} = \frac{N}{\sum_{i=1}^{k} \frac{n_i}{x_i}}.$$

Media armonica per una distribuzione disaggregata: calcolo

$$\mu_a = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}}.$$

x_i	$1/x_i$
12.80	0.0781
13.00	0.0769
13.40	0.0746
13.40	0.0746
13.50	0.0741
13.60	0.0735
13.60	0.0735
13.70	0.0730
Totale	0.5984

Quotazioni di borsa di un titolo azionario in 8 sedute successive:

12.8, 13.0, 13.4, 13.4 , 13.6, 13.5, 13.6, 13.7

- La media armonica della distribuzione è data da

$$\mu_a = \frac{8}{\frac{1}{12.8} + \frac{1}{13.0} + \cdots + \frac{1}{13.7}} = \frac{8}{0.5984} = 13.369.$$

Media armonica per una distribuzione di frequenze a modalità raggruppate in classi: calcolo

$$\mu_a = \frac{N}{\sum_{i=1}^{k} \frac{n_i}{\bar{x}_i}}.$$

Distribuzione di frequenze degli studenti di un corso di Statistica secondo l'età:

Classe	Valore centrale	n_i	$\frac{n_i}{\bar{x}_i}$
19-21	20.0	31	1.55
21-24	22.5	45	2.00
24-27	25.5	5	0.20
27-30	28.5	1	0.04
Totale		82	3.78

- La media armonica della distribuzione è data da:

$$\mu_a = \frac{82}{\frac{31}{20.0} + \frac{45}{22.5} + \frac{5}{25.5} + \frac{1}{28.5}} = \frac{82}{3.78} = 21.7$$

Media geometrica

La media geometrica di una distribuzione statistica disaggregata, x_1, x_2, …, x_N, in cui tutti i termini sono maggiori di 0, è data dalla radice N-esima del prodotto dei termini:

$$\mu_g = \sqrt[N]{x_1 \cdot x_2 \cdot \dots \cdot x_N} = \exp\left[\frac{1}{N}\sum_{i=1}^{N}\ln(x_i)\right]$$

La media geometrica di una distribuzione statistica di frequenze, i cui termini sono tutti diversi da 0, è

$$\mu_g = \sqrt[N]{x_1^{n_1} \cdot x_2^{n_2} \cdot \dots \cdot x_k^{n_k}} = \exp\left[\frac{1}{N}\sum_{i=1}^{k}n_i\ln(x_i)\right]$$

Media geometrica per una distribuzione disaggregata: calcolo

$$\mu_g = \exp\left[\frac{1}{N}\sum_{i=1}^{N}\ln(x_i)\right]$$

x_i	$\ln(x_i)$
12.80	2.549
13.00	2.565
13.40	2.595
13.40	2.595
13.50	2.603
13.60	2.610
13.60	2.610
13.70	2.617
Totale	20.744

Quotazioni di borsa di un titolo azionario in 8 sedute successive:

12.8, 13.0, 13.4, 13.4 , 13.6, 13.5, 13.6, 13.7

- La media geometrica della distribuzione è data da

$$\mu_g = \sqrt[8]{12.8 \cdot 13.0 \cdot \cdots \cdot 13.7} = 13.372$$

$$\mu_g = \exp\left[\frac{\ln(12.8) + \ln(13.0) + \cdots + \ln(13.7)}{8}\right]$$

$$= \exp\left(\frac{20.744}{8}\right) = \exp(2.593) = 13.370$$

Media geometrica per una distribuzione di frequenze a modalità singole: calcolo

$$\mu_g = \exp\left[\frac{1}{N}\sum_{i=1}^{k} n_i \ln(x_i)\right]$$

Distribuzione di frequenze della lunghezza dell'avambraccio (in cm)

x_i	n_i	$n_i \cdot \ln(x_i)$
41	3	11.141
42	2	7.475
43	6	22.567
44	11	41.626
45	8	30.453
46	17	65.087
47	21	80.853
48	14	54.197
49	17	66.161
50	15	58.680
51	10	39.318
52	10	39.512
53	5	19.851
54	1	3.989
Totale	140	540.912

- La media geometrica è data da:

$$\mu_g = \sqrt[140]{41^3 \cdot 42^2 \cdot \cdots \cdot 54^1}$$

$$= \exp\left[\frac{3\cdot\ln(41) + 2\cdot\ln(42) + \cdots + 1\cdot\ln(54)}{140}\right]$$

$$= \exp\left(\frac{540.912}{140}\right) = 47.639$$

Media quadratica

La media quadratica di una distribuzione statistica disaggregata, $x_1, x_2, \ldots, x_N$, è la radice quadrata della media aritmetica dei quadrati dei termini della distribuzione:

$$\mu_q = \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_N^2}{N}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}\,.$$

La media geometrica di una distribuzione statistica di frequenze è

$$\mu_q = \sqrt{\frac{x_1^2 n_1 + x_2^2 n_1 + \cdots + x_k^2 n_k}{N}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2 n_i}\,.$$

Media quadratica per una distribuzione disaggregata: calcolo

$$\mu_q = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}\,.$$

x_i	x_i^2
12.80	163.84
13.00	169.00
13.40	179.56
13.40	179.56
13.50	182.25
13.60	184.96
13.60	184.96
13.70	187.69
Totale	1431.82

Quotazioni di borsa di un titolo azionario in 8 sedute successive:

12.8, 13.0, 13.4, 13.4 , 13.6, 13.5, 13.6, 13.7

- La media quadratica della distribuzione è data da

$$\mu_q = \sqrt{\frac{12.8^2 + 13.0^2 + \cdots + 13.7^2}{8}}$$

$$= \sqrt{\frac{1431.82}{8}} = 13.378.$$

Media quadratica per una distribuzione di frequenze: calcolo

$$\mu_q = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2 n_i}.$$

Distribuzione di frequenze della lunghezza dell'avambraccio (in cm) :

x_i	n_i	$x_i^2 \cdot n_i$
41	3	5043
42	2	3528
43	6	11094
44	11	21296
45	8	16200
46	17	35972
47	21	46389
48	14	32256
49	17	40817
50	15	37500
51	10	26010
52	10	27040
53	5	14045
54	1	2916
Totale	140	320106

- Media quadratica:

$$\mu_q = \sqrt{\frac{41^2 \cdot 3 + 42^2 \cdot 2 + \cdots + 54^2 \cdot 1}{140}}$$

$$= \sqrt{\frac{320106}{140}} = 47.817$$

Medie analitiche ponderate

Media aritmetica ponderata

Siano $x_1, x_2, \ldots, x_k$ le osservazioni e $w_1, w_2, \ldots, w_k$ i rispettivi pesi. Allora, la **media aritmetica ponderata** di $x_1, x_2, \ldots, x_k$ è data dal rapporto tra la somma delle osservazioni moltiplicate per i rispettivi pesi e la somma dei pesi

$$\mu = \frac{x_1 \cdot w_1 + x_2 \cdot w_2 + \cdots + x_k \cdot w_k}{w_1 + w_2 + \cdots + w_k} = \frac{\sum_{i=1}^{k} x_i w_i}{\sum_{i=1}^{k} w_i}$$

Media aritmetica ponderata: esempio

$$\mu = \frac{\sum_{i=1}^{k} x_i w_i}{\sum_{i=1}^{k} w_i}$$

Dimensioni medie e il numero di famiglie per ripartizione territoriale (dati Istat):

Ripartizione territoriale	N. medio di componenti	N. famiglie	$x_i\, w_i$
Italia Nord-Occidentale	2.38	6217200	14796936
Italia Nord-Orientale	2.49	4232010	10537705
Italia Centrale	2.55	4242199	10817607
Italia Meridionale	2.92	4748274	13864960
Italia Insulare	2.77	2370993	6567651
Italia		21810676	56584859

- **Ampiezza media della famiglia per l'Italia**

$$\mu = \frac{\sum_{i=1}^{k} x_i w_i}{\sum_{i=1}^{k} w_i} = \frac{56.584.859}{21.810.676} = 2{,}59$$

Altre medie analitiche ponderate

media armonica

$$\mu_a = \frac{\sum_{i=1}^{k} w_i}{\sum_{i=1}^{k} \frac{w_i}{x_i}}$$

media geometrica

$$\mu_g = \sqrt[\sum_{i=1}^{k} w_i]{x_1^{w_1} \cdot x_2^{w_2} \cdot \ldots \cdot x_k^{w_k}}$$

$$= \exp\left[\frac{1}{\sum_{i=1}^{k} w_i} \sum_{i=1}^{k} w_i \cdot \ln(x_i)\right]$$

media quadratica

$$\mu_q = \sqrt{\frac{\sum_{i=1}^{k} x_i^2 \cdot w_i}{\sum_{i=1}^{k} w_i}}$$

Valore centrale

Sia data una distribuzione disaggregata x_1, x_2, …, x_N. Sia y_1, y_2, …, y_N la corrispondente distribuzione dei termini ordinati, con $y_1 \le y_2 \le \ldots \le y_N$.

Il **valore centrale** della distribuzione è la media aritmetica dei valori estremi

$$m_c = \frac{y_1 + y_N}{2}$$

Valore centrale: calcolo

$$m_c = \frac{y_1 + y_N}{2}$$

Previsione da parte di 14 economisti della variazione media percentuale dei prezzi al consumo per il prossimo anno:

2.1, 2.2, 1.8, 2.4, 2.5, 2.8, 2.1, 2.2, 2.1, 1.9, 1.8, 2.4, 2.9, 2.4

- Valore centrale
  - Fase 1: ordinamento dei termini

1.8, 1.8, 1.9, 2.1, 2.1, 2.1, 2.2, 2.2, 2.4, 2.4, 2.4, 2.5, 2,8, **2.9**

- Fase 2: calcolo

$$m_c = \frac{1.8 + 2.9}{2} = 2.35$$

Valore centrale per una distribuzione di frequenze a modalità singole: calcolo

$$m_c = \frac{y_1 + y_N}{2}$$

Distribuzione di frequenze della lunghezza dell'avambraccio (in cm)

Modalità	Frequenza
41	3
42	2
43	6
44	11
45	8
46	17
47	21
48	14
49	17
50	15
51	10
52	10
53	5
54	1
Totale	140

- Valore centrale:

$$m_c = \frac{41 + 54}{2} = 47.5$$