

Corso di Statistica

Gli indici di tendenza centrale
Lezione due

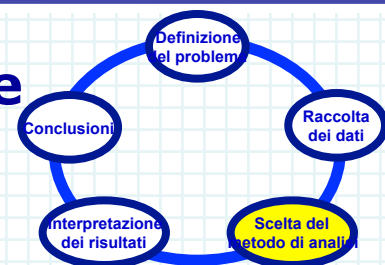
Maria Spano

maria.spano84@gmail.com



Gli indici sintetici

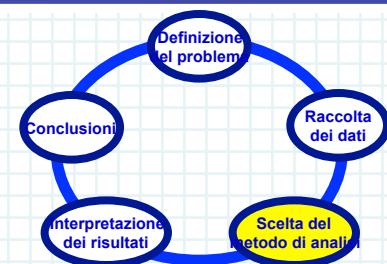
- **Tendenza centrale**
- **Forma**
- **Variabilità**



- ✓ Consentono il passaggio da una pluralità di informazioni ad un'unica misura numerica;
- ✓ Sintetizzano l'intera distribuzione in un singolo valore, consentendo così confronti nel tempo, nello spazio o tra circostanze differenti;
- ✓ In alcuni casi, consentono di verificare se le conseguenze di una determinata azione abbiano prodotto il risultato desiderato, in quale direzione e con quale intensità.

Gli indici sintetici

- Posizione
- Forma
- Variabilità



✓ Indici assoluti



Dipendono dalla natura della variabile che si sta esaminando e sono espressi nella stessa unità di misura della variabile.

✓ Indici relativi

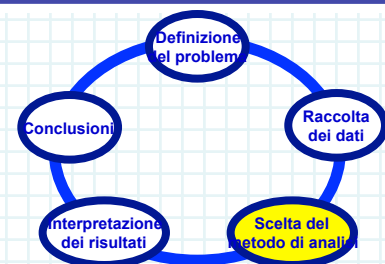


Sono svincolati dall'unità di misura perché costruiti come *rapporti tra indici assoluti* o *tra indici assoluti e loro valori estremi*. Sono, quindi, *numeri puri*, utili per confrontare fenomeni omogenei.

✓ Indici normalizzati



Sono particolari indici relativi che *variano in un intervallo finito*, generalmente in $[0, 1]$ oppure in $[-1, +1]$. Sono, quindi, di immediata interpretazione.



La media aritmetica

Esempio: distribuzione unitaria semplice

unità	età
1	35
2	37
3	59
4	54
5	44
6	38
7	62
8	71
9	56
10	60
11	33
12	46
13	41
14	53
15	38
16	55
17	50
18	63
19	35
20	51
totale	981

$$M = \frac{981}{20} = 49,05$$

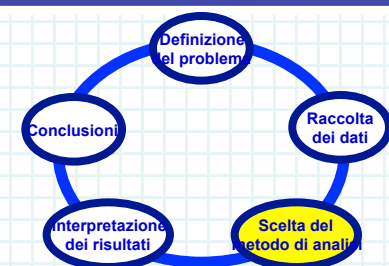
$$M = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^k x_i}{n}$$

La media aritmetica

Esempio: distribuzione di frequenze

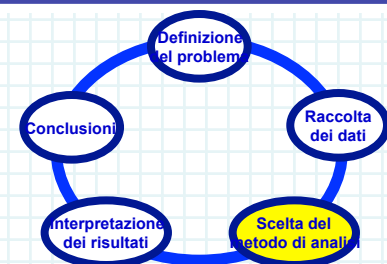
$$M = \frac{\sum_{i=1}^k x_i \cdot n_i}{n}$$

$$M = \sum_{i=1}^k x_i \cdot \frac{n_i}{n}$$



Età studenti del Corso (x_i)	Frequenze assolute (n_i)	$x \cdot n$	Freq. Relative (f_i)	$x \cdot f$
18	2	36	0,011	0,2
19	44	836	0,232	4,4
20	66	1320	0,347	6,9
21	32	672	0,168	3,5
22	18	396	0,095	2,1
23	13	299	0,068	1,6
24	9	216	0,047	1,1
25	6	150	0,032	0,8
Totale	190	3925	1,000	20,7

$$M = \frac{3925}{190} = 20,7$$



La media aritmetica

Esempio: distribuzione in classi

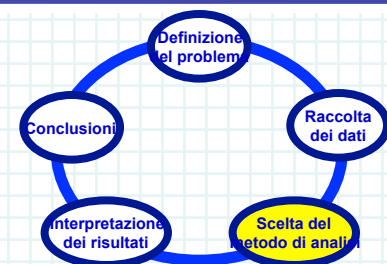
Tempo per raggiungere la Facoltà (in min.)	Frequenze assolute (n_i)	valori centrali (c)	$c * n$
0 - 20	84	10	840
20 - 40	81	30	2430
40 - 60	44	50	2200
60 - 120	18	90	1620
<i>Totale</i>	<i>227</i>		<i>7090</i>

In questo caso, la soluzione più comune consiste nell'utilizzare il **valore centrale delle classi**

$$M = \frac{7090}{227} = 31,2$$

La media aritmetica

M **m** $\bar{x}$



Media semplice:

$$M = \frac{\sum_{i=1}^k x_i}{n}$$

Media con dati organizzati in frequenze:

$$M = \frac{\sum_{i=1}^k x_i \cdot n_i}{n}$$

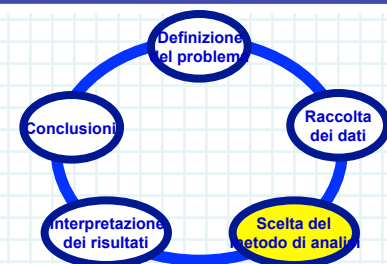
Media con le frequenze relative:

$$M = \sum_{i=1}^k x_i \cdot \frac{n_i}{n}$$

Media con dati organizzati in classi:

$$M = \frac{\sum_{i=1}^C c_i \cdot n_i}{n}$$

La media aritmetica: le proprietà



1. Criterio di internalità

La m.a. è sempre compresa tra il minimo e il massimo della distribuzione osservata:

2. La media come baricentro

La somma degli scarti dalla media è nulla:

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

3. Linearità della m.a.

Se la variabile **X** ha media **M**, allora la variabile **$Y=a+bX$** ha media pari a **$a+bM$** :

Questa proprietà implica che:

- .) Se si aggiunge o si sottrae una costante **a** alla variabile **X**, la media sarà modificata dello stesso ammontare (caso **b=1**)
- .) Se la variabile **X** è moltiplicata per un coefficiente **b** costante, la media risulterà moltiplicata per lo stesso ammontare (caso **a=0**)

La media aritmetica: le proprietà

4. Proprietà associativa della m.a.

La media di una variabile osservata in più gruppi può essere ottenuta come media delle medie dei singoli gruppi, tenuto conto della eventuale differente numerosità:

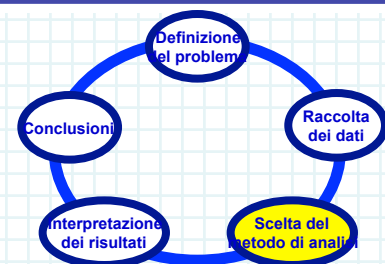
Data una popolazione su cui è definita una variabile X con media m , se dividiamo la popolazione in k gruppi, di numerosità $n_1, n_2, \dots, n_k$, si ha:

$$\mu = \mu_1 \cdot \frac{n_1}{n} + \mu_2 \cdot \frac{n_2}{n} + \dots + \mu_k \cdot \frac{n_k}{n}$$

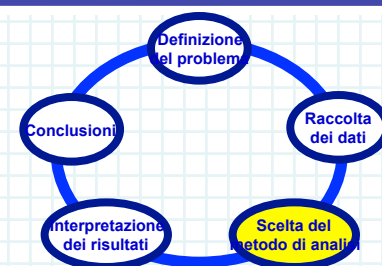
5. Minimizzazione dei quadrati degli scarti

La media aritmetica rende minima la somma degli scarti al quadrato:

$$\sum_i (x_i - \mu)^2 = \min$$



La media aritmetica ponderata



Esame	Crediti	Voto		Voto*Crediti	
		Stud. X	Stud. Y	Stud. X	Stud. Y
1	4	25	25	100	100
2	6	30	30	180	180
3	4	30	27	120	108
4	5	28	24	140	120
5	12	22	22	264	264
6	8	27	30	216	240
7	9	25	25	225	225
8	9	30	28	270	252
9	7	24	30	168	210
10	5	30	30	150	150
11	10	20	27	200	270
12	5	27	20	135	100
13	10	26	28	260	280
14	6	28	26	168	156
15	10	22	30	220	300
16	4	30	22	120	88
17	4	30	22	120	88
18	12	22	30	264	360
	130	476	476	3320	3491

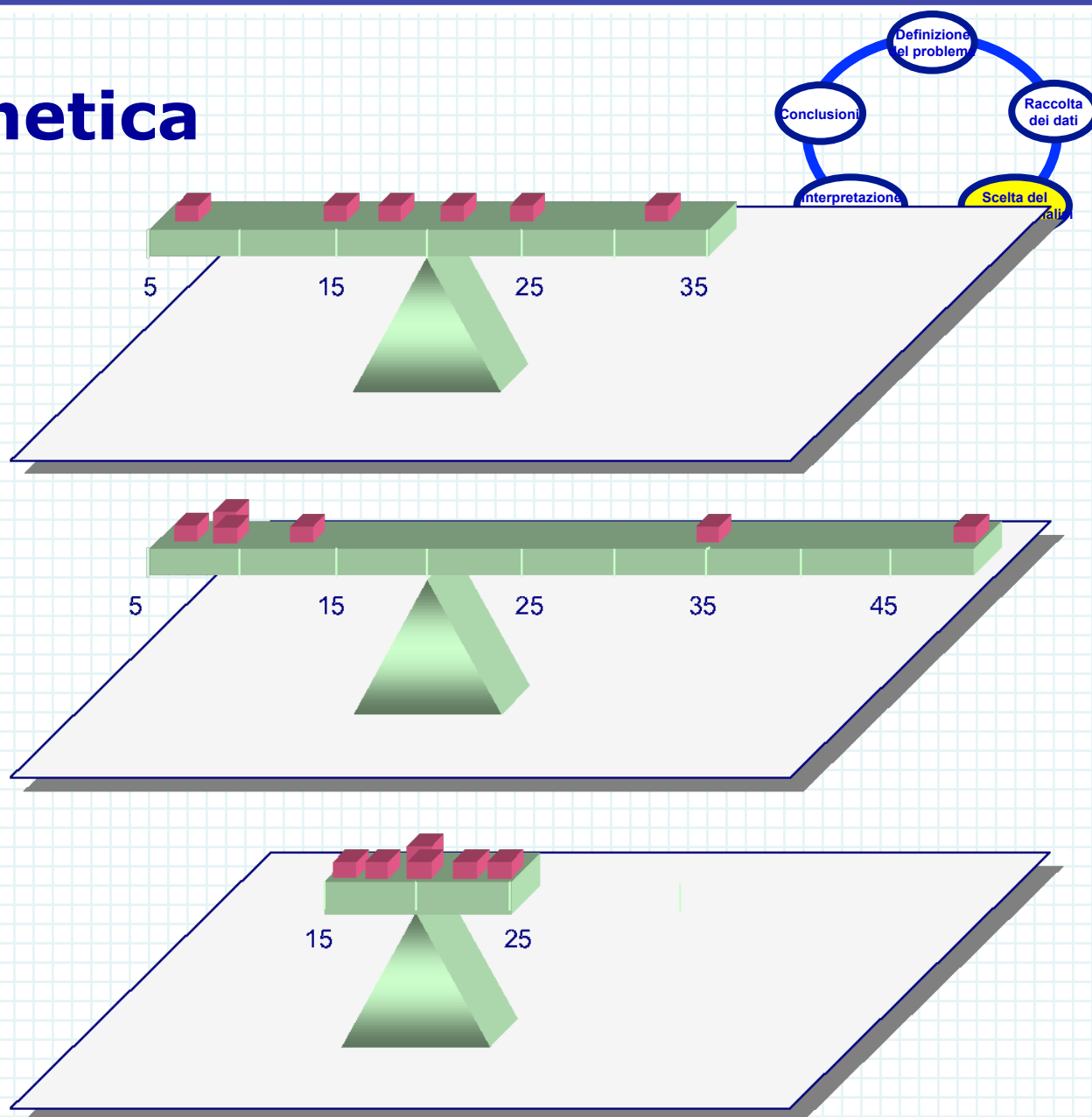
$$\mu(X) = \frac{\sum_{i=1}^n x_i \cdot p_i}{\sum_i p_i}$$

$$= \frac{3320}{130} = 25,5$$

$$\mu(Y) = \frac{\sum_{i=1}^n y_i \cdot p_i}{\sum_i p_i}$$

$$= \frac{3491}{130} = 26,9$$

La media aritmetica



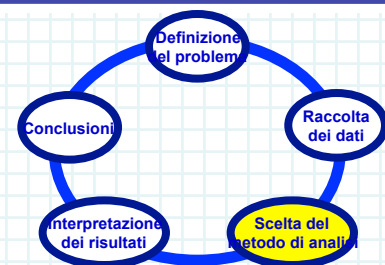
Le medie “robuste”

La mediana

La mediana, **Me**, è il valore assunto dall’unità statistica che occupa la posizione centrale della distribuzione ordinata in modo non decrescente.

E’ un indice “robusto” in quanto non dipende da variazioni che si verificano nelle code della distribuzione (dove si possono trovare i c.d. “valori anomali”)

La mediana è il valore assunto dall’unità statistica che divide il collettivo in due parti di uguale numerosità: una parte formata dalle unità che presentano una modalità inferiore o uguale a quella dell’unità centrale e una parte formata dalle unità che presentano una modalità superiore o uguale a quella dell’unità centrale



Le medie “robuste”

I passi per il calcolo della mediana

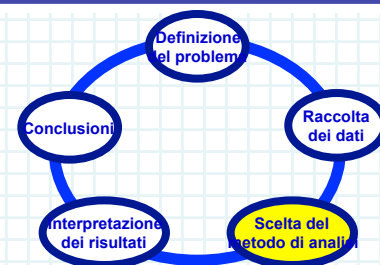
1. Si ordina la distribuzione in modo non decrescente
2. Si calcola la posizione mediana

$$Pos(Me) = \frac{n+1}{2}$$

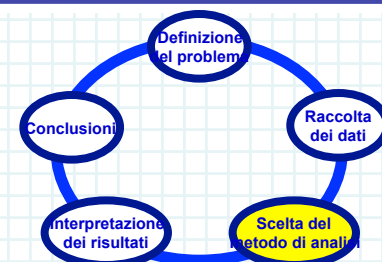
Posizione mediana $\left\{ \begin{array}{l} n \text{ dispari:} \\ n \text{ pari:} \end{array} \right.$

$$Pos(Me) = \frac{n}{2} ; \frac{n}{2} + 1$$

3. Si osserva il valore che occupa la posizione mediana



La mediana



La mediana è l'osservazione che, nella serie ordinata dei dati, lascia alla sua destra il 50% delle osservazioni e alla sinistra il 50% delle osservazioni.

Posizioni occupate dalle unità statistiche	Variabile x
1	19
2	22
3	25
4	26
5	27

$$\text{Posizione mediana: } Pos(Me) = \frac{n+1}{2}$$

Mediana: 25

(Media=23,8)

Posizioni occupate dalle unità statistiche	Variabile x
1	19
2	22
3	25
4	26

$$\text{Posizione mediana } Pos(Me) = \frac{n}{2} ; \frac{n}{2} + 1$$

Mediana: 23,5

(Media=23)

La mediana

Esempio

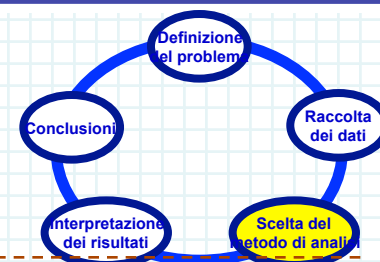
1. Si ordina la distribuzione in modo non decrescente
2. Si calcola la posizione mediana
3. Si osserva il valore che occupa la posizione mediana

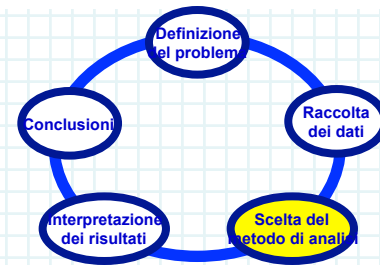
Posizione mediana:

n dispari: $P(Me) = \frac{n+1}{2}$

n pari: $P(Me) = \frac{n}{2} ; \frac{n}{2} + 1$

Età studenti del Corso	Frequenze assolute (n_i)	Freq. Relative (f_i)	Freq. percentuali (p_i)	Freq. ass. cumulate (N_i)	Freq. rel. cumulate (F_i)	Freq. % cumulate (P_i)
18	2	0,011	1,1	2	0,011	1,1
19	44	0,232	23,2	46	0,242	24,2
20	66	0,347	34,7	112	0,589	58,9
21	32	0,168	16,8	144	0,758	75,8
22	18	0,095	9,5	162	0,853	85,3
23	13	0,068	6,8	175	0,921	92,1
24	9	0,047	4,7	184	0,968	96,8
25	6	0,032	3,2	190	1,000	100,0
	190	1,000	100,0			





La mediana per dati raggruppati in classi

I passi per il calcolo della mediana

1. Si ordina la distribuzione in modo non decrescente

2. Si calcola la posizione mediana

$$P(Me) = \frac{n+1}{2}$$

Posizione mediana $\left\{ \begin{array}{l} n \text{ dispari:} \\ n \text{ pari:} \end{array} \right.$

$$P(Me) = \frac{n}{2} ; \frac{n}{2} + 1$$

3. Si osserva la classe mediana

$$Me = L_{inf} + \frac{\frac{N}{2} - N_{Cl_{(Prec)}}}{n_{Cl_{(Med)}}} \times c \quad ; \quad Me = L_{inf} + \frac{0,5 - F_{Cl_{(Prec)}}}{f_{Cl_{(Med)}}} \times c$$

**ampiezza
della classe = c**

La mediana

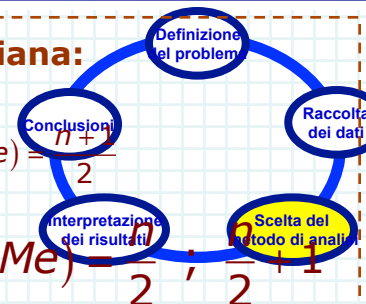
Esempio

1. Si ordina la distribuzione in modo non decrescente
2. Si calcola la posizione mediana
3. Si osserva il valore che occupa la posizione mediana

Posizione mediana:

n dispari: $P(Me) = \frac{n+1}{2}$

n pari: $P(Me) = \frac{n}{2}, \frac{n}{2} + 1$



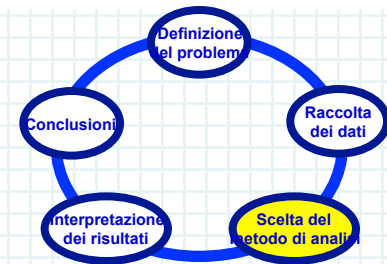
Per dati raggruppati in classi:

$$Me = L_{inf} + \frac{\frac{N}{2} - N_{Cl(Prec)}}{n_{Cl(Med)}} \times c$$

Tempo per raggiungere la Facoltà (in min.)	Frequenze assolute (n_i)	Freq. Relative (f_i)	Freq. percentuali (p_i)	Freq. ass. cumulate (N_i)	Freq. rel. cumulate (F_i)	Freq. % cumulate (P_i)
0- 20	84	0,370	37,0	84	0,370	37,0
20- 40	81	0,357	35,7	165	0,727	72,7
40- 60	44	0,194	19,4	209	0,921	92,1
>60	18	0,079	7,9	227	1,000	100,0
	227	1,000	100,0			

$$Me = 20 + \frac{114 - 84}{81} (40 - 20)$$

I quartili



Il primo quartile, Q_1 , è il valore tale che il 25% delle osservazioni è più piccolo di Q_1 e il 75% è più grande di Q_1

Il terzo quartile, Q_3 , è il valore tale che il 75% delle osservazioni è più piccolo di Q_3 e il 25% è più grande di Q_3

Posizioni occupate dalle unità statistiche	Variabile x	
1	18	
2	19	
3	20	
4	21	
5	22	
6	26	
7	27	
8	28	

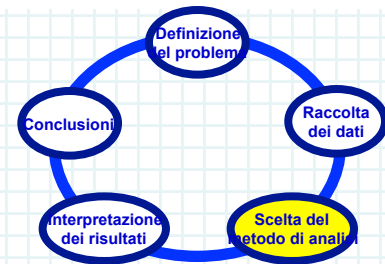


Q1: 19,5



Q3: 26,5

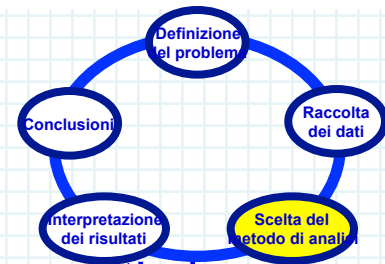
I quartili



- $\frac{N}{4}$ non è un numero intero: Q1 è l'elemento che occupa il posto $\left[\frac{N}{4} \right] + 1$ nella successione ordinata ($\left[\frac{N}{4} \right]$ è la parte intera di $\frac{N}{4}$)

Posizioni occupate dalle unità statistiche	Variabile x
1	20
2	21
3	23
4	25
5	26

I quartili

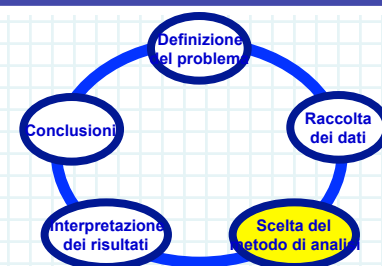


- $\frac{N}{4}$ è un numero intero: Q1 è la media aritmetica degli elementi che occupano il posto $\frac{N}{4}$ ed il posto $\frac{N}{4} + 1$ nella successione ordinata

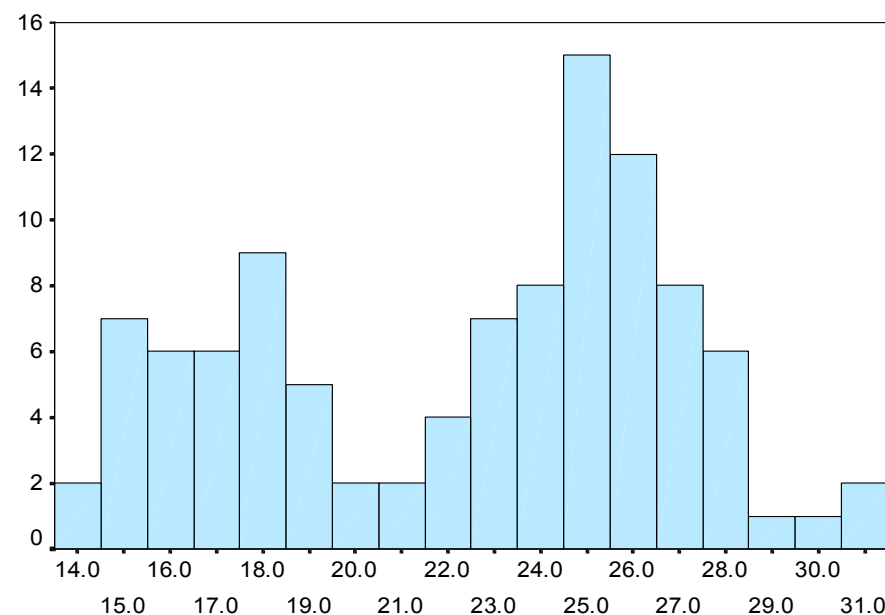
Posizioni occupate dalle unità statistiche	Variabile x
1	20
2	21
3	23
4	25

La moda

La moda è il valore più frequente in un insieme di dati



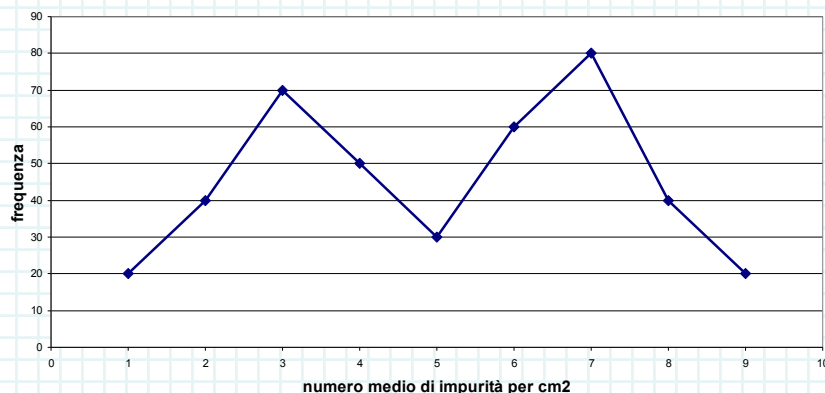
Variabile x	Frequenze assolute
20	2
22	2
25	5
26	10
<i>totale</i>	19



Reddito pro capite 1997 (in milioni di lire)

La moda

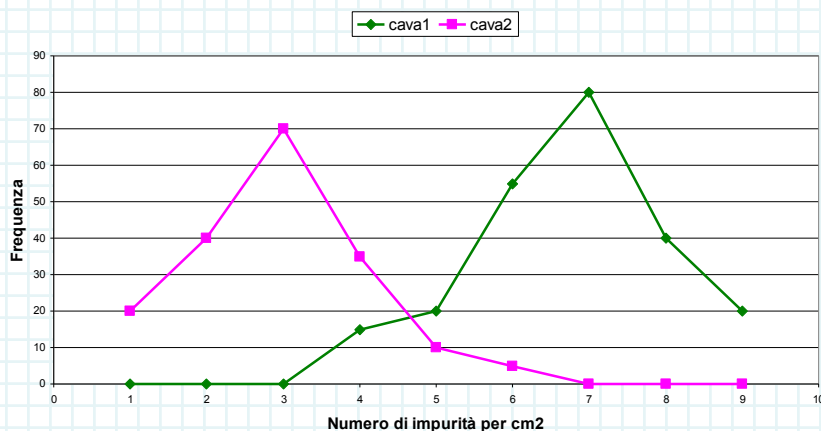
Un esempio: un'impresa produttrice di vasellame vuole controllare la qualità della creta utilizzata nella lavorazione. Viene rilevato il numero medio di impurità per cm² su 410 pezzi



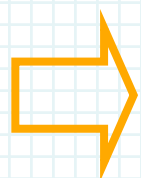
numero medio di impurità per cm ²	frequenza
1	20
2	40
3	70
4	50
5	30
6	60
7	80
8	40
9	20
<i>totale</i>	<i>410</i>

La moda

Osservazione: la creta utilizzata proviene da due diverse cave



numero medio di impurità per cm ²	cava1	cava2
1	0	20
2	0	40
3	0	70
4	15	35
5	20	10
6	55	5
7	80	0
8	40	0
9	20	0
	230	180



La presenza di due mode in una distribuzione può essere dovuta alla presenza di due gruppi di unità distinti rispetto ad una variabile non osservata

Altri indici “robusti”

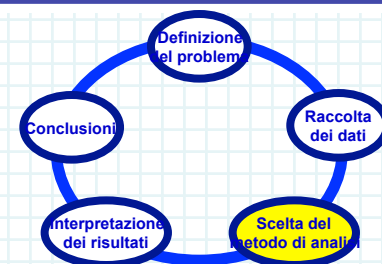
✓ I quartili

✓ I decili

✓ I percentili

✓ I quantili

✓ La moda



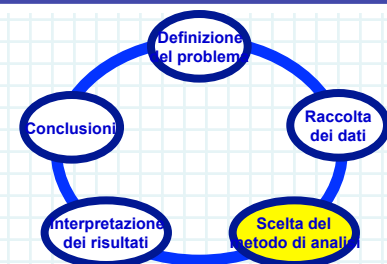
Caratteri, informazione e indici

Variabili quantitative ➡ Media, Mediana, Moda

Mutabili ordinabili ➡ Mediana, Moda

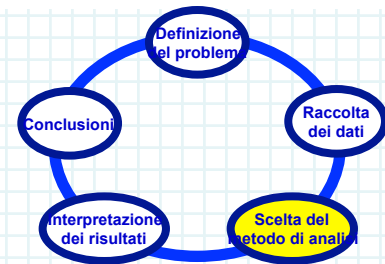
Mutabili sconnesse ➡ Moda

Qualche considerazione



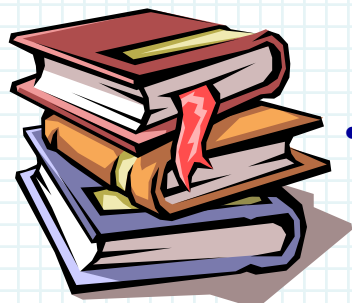
- ✓ La scelta dell'indice di tendenza centrale dipende dal tipo e dalle caratteristiche della distribuzione;
- ✓ Più che individuare l'indice "migliore in assoluto" (che non esiste), è importante anche valutare le differenze tra le diverse misure, che possono fornire ulteriori, importanti informazioni anche, ad esempio, sulla *forma* della distribuzione;
- ✓ Volendo comunque definire delle caratteristiche dei diversi indici di posizione, possiamo dire che: (*Piccolo, 2004*)

Qualche considerazione

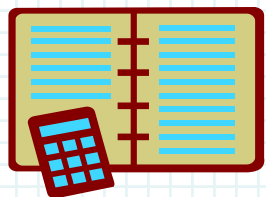


- ☞ La moda è utile quando occorre “minimizzare gli scontenti”, e quindi in tutte quelle situazioni in cui il consenso e il numero delle singole unità ha significato per la decisione. In breve, la moda è un indice per governare;
- ☞ La mediana minimizza i costi complessivi ed è resistente ai valori estremi. Quindi, la mediana è un indice per decisioni che implicano costi elevati nei casi estremi;
- ☞ La media aritmetica è il baricentro dei dati e propone, quindi, un valore che equi-ripartisce il fenomeno tra le unità statistiche, pervenendo così a decisioni nelle quali contano, a parità numerica, gli estremi molto più dei valori centrali. Quindi, la media aritmetica è un indice di equilibrio generale.

Dove e come studiare

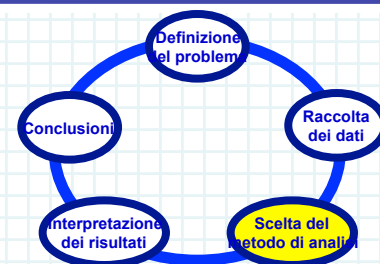


- Libro di testo: D. Piccolo (2004) – Statistica per le decisioni – Il Mulino.
Cap. 4 (escluso pagine da 78 a 81 e paragrafo 4.8)
- Libro di testo: S. Borra, A. Di Ciaccio (2008) – Statistica – Metodologie per le scienze economiche e sociali – McGraw-Hill.
Cap. 3 (escluso paragrafo 3.3)



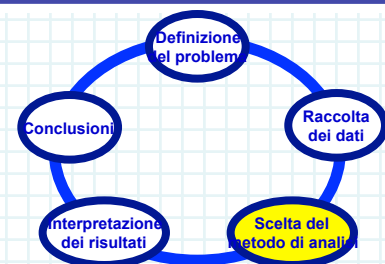
File “esercizi indici sintetici.pdf”

Esercizio n. 1 – punto 5
Esercizio n. 2
Esercizio n.3 – punto *c*
Esercizio n. 4
Esercizio n. 5 – punto *a*
Esercizio n. 7 – punto *c*
Esercizio n. 12 – punto *b*



Determinare la media aritmetica della seguente successione di numeri:

12, 15, 19, 23, 28



La media aritmetica è : $M = \frac{\sum_{i=1}^k x_i}{n}$

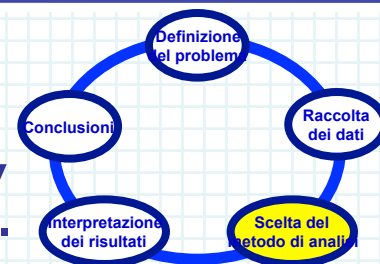
**Quando la
frequenza
delle unità
statistiche è
pari ad 1**

$$\mu = \frac{12+15+19+23+28}{5} = \frac{97}{5} = 19,4$$



è il numero di unità statistiche considerate

Nella tabella seguente è riportata la distribuzione delle famiglie per numero di componenti in un dato comune, calcolare il numero medio di componenti delle famiglie.



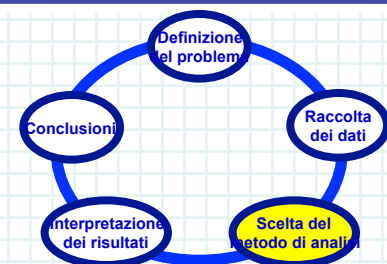
Numero di componenti	Famiglie
x_i	n_i
1	153
2	225
3	335
4	564
5	346
6	133
7	75
8	49
Totale	1880

N.B.
$$M = \frac{\sum_{i=1}^k x_i \cdot n_i}{n}$$

$$\mu = \frac{7309}{1880} = 3,88$$

$x_i \cdot n_i$
153
450
1005
2256
1730
798
525
392
7309

Calcolare l'età media degli iscritti ad un'associazione sportiva, sulla base dei dati che si evincono dalla tabella seguente:



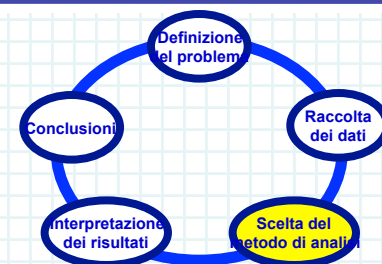
Classi di età	Iscritti
3-15	115
15-25	156
25-40	130
40-50	110
50-60	90
Oltre 60	38
Totale	639

$$\bar{x}_i = x_i + x_{i+1} / 2$$

x_i	n_i	$x_i n_i$
9	115	1035
20	156	3120
32,5	130	4225
45	110	4950
55	90	4950
65	38	2470
	639	20750

$$\mu = 20750 / 639 = 32,47$$

Data la seguente tabella, determinarne la moda.

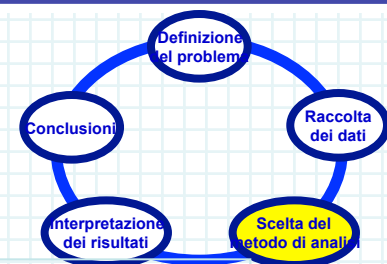


Numero di componenti	Famiglie
x_i	n_i
1	153
2	225
3	335
4	564
5	346
6	133
7	75
8	49
Totale	1880

La moda di una distribuzione è la modalità del carattere cui corrisponde la massima frequenza.

Scorrendo lungo la colonna delle frequenze, la moda è la modalità 4 poiché ad essa corrisponde la frequenza massima $n_i = 564$

La tabella seguente riporta la distribuzione di 100 fumatori per classi di età, determinare l'età modale.



Classi di età	Fumatori
30-33	2
34-37	3
38-41	9
42-45	19
46-49	29
50-53	17
54-57	10
58-61	7
62-65	4
Totale	100

Per variabili continue si distingue:

- Se le classi di modalità hanno uguale ampiezza, la moda cade in quella con maggiore frequenza;
- Se le classi di modalità hanno diversa ampiezza, la moda cade nella classe con maggiore densità di frequenza;

$$Mo = L_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) c$$

Dove:

L_1 è il confine inferiore della classe modale;

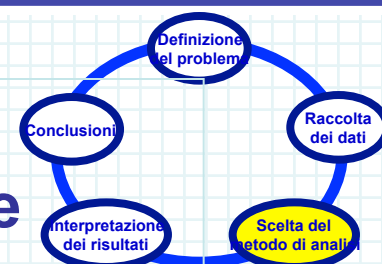
Δ_1 è l'eccesso della frequenza modale sulla frequenza della classe immediatamente inferiore;

Δ_2 è l'eccesso della frequenza modale sulla frequenza della classe immediatamente superiore;

c è l'ampiezza della classe modale;

Classi di età	Fumatori
30-33	2
34-37	3
38-41	9
42-45	19
46-49	29
50-53	17
54-57	10
58-61	7
62-65	4
Totale	100

Le classi hanno uguale ampiezza, quindi la classe modale è la quinta, ad essa corrisponde la frequenza massima (29).



$$Mo = L_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) c$$

Dove:

$$L_1 = 46$$

$$\Delta_1 = 10$$

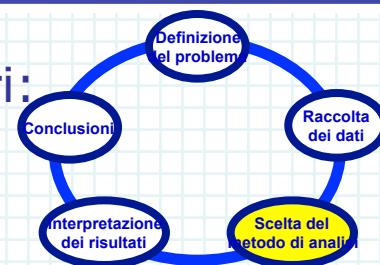
$$\Delta_2 = 12$$

$$c = 3$$

$$Mo = 46 + \left(\frac{10}{10 + 12} \right) * 3 = 47,36$$

Determinare la mediana della seguente successione di numeri:

12,15,19,23,28



1° step: ordinare in ordine crescente la successione di numeri (distribuzione di frequenze).

2° step: verificare se la numerosità è in numero pari o dispari

Siccome i dati sono in ordine crescente e sono in numero dispari (5), la mediana è individuata dal posto centrale:

$$C = (n+1)/2$$

$$C = (5+1)/2 = 3$$



$$\text{Me} = 19$$

Data la seguente tabella, determinare la mediana.

x_i	n_i
7	4
8	6
13	3
15	1
Totale	14

Siccome n è pari (14) è necessario calcolare i posti centrali:

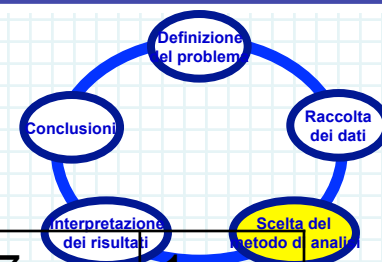
$$C_1 = n/2 ; C_2 = (n/2)+1$$



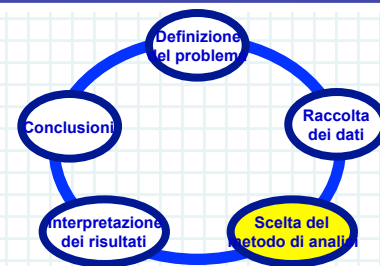
$$C_1 = 14/2 = 7$$

$$C_2 = (14/2)+1 = 8$$

La mediana è compresa tra il 7° e 8° posto, poiché entrambi si riferiscono alla modalità 8 si ha : $Me = 8$



7	1
7	1
7	1
7	1
8	1
8	1
8	1
8	1
8	1
8	1
13	1
13	1
13	1
15	1



La tabella seguente riporta la distribuzione delle persone partecipanti ad un viaggio organizzato per classi di età:

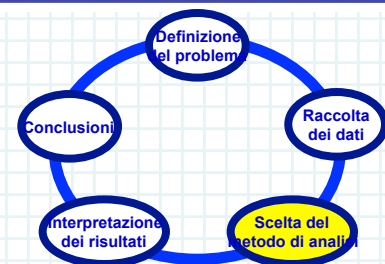
Determinare l'età mediana.

Classi di età	n. di partecipanti	Frequenze cumulate
0-13	22	22
14-44	65	87
45-64	80	167
65 e oltre	40	207
Totale	207	

$$Me = L_{\text{inf}} + \frac{\frac{N}{2} - N_{Cl_{(Prec)}}}{n_{Cl_{(Med)}}} \times c$$

$$Me = 45 + [(207/2 - 87) / 80] \times 19 = 48,91$$

La classe mediana è la classe 45-46, in quanto è la prima a cui corrisponde una frequenza cumulata superiore a $n/2$ (103,5)



La tabella seguente riporta la distribuzione di 125 atleti per classi di altezza (in cm):

Determinare primo, secondo e terzo quartile

Classi di altezze	n. di atleti	Frequenze cumulate
171-175	14	14
176-180	18	32
181-185	28	60
186-190	33	93
191-195	17	110
196-200	15	125
Totale	125	

La classe che contiene il primo quartile è la seconda essendo: $125/4 = 31,25$, ed essendo la sua frequenza cumulata pari a 32.

$$Q_1 = 176 + [(31,25 - 14) / 18] * 4 = 179,83$$

Q_2 ????

Q_3 ????

Riepilogo

Gli indici sintetici: tendenza centrale

- ✓ Media (nel caso di una distribuzione unitaria semplice, di una distribuzione di frequenza, di una distribuzione in classi)
- ✓ Moda (nel caso di una distribuzione unitaria semplice, di una distribuzione di frequenza, di una distribuzione in classi)
- ✓ Mediana (nel caso di una distribuzione unitaria semplice, di una distribuzione di frequenza, di una distribuzione in classi)
- ✓ Quartili (nel caso di una distribuzione unitaria semplice, di una distribuzione di frequenza, di una distribuzione in classi)
- ✓ Decili, percentili

