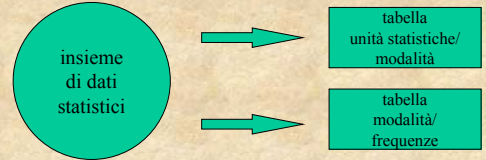


INDICI DI TENDENZA CENTRALE

MODA
MEDIA
MEDIANA

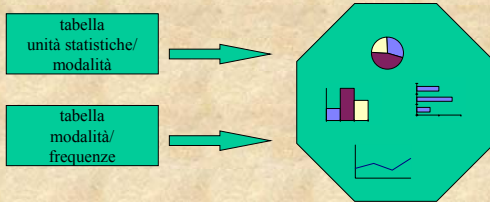
Al fine di semplificare la lettura e l'interpretazione di un fenomeno oggetto di un'indagine statistica, i dati possono essere:

- organizzati in una tabella



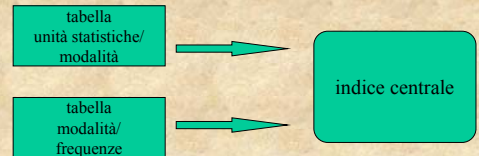
Al fine di semplificare la lettura e l'interpretazione di un fenomeno oggetto di un'indagine statistica, i dati possono essere:

- rappresentati mediante un grafico



Al fine di semplificare la lettura e l'interpretazione di un fenomeno oggetto di un'indagine statistica, i dati possono essere:

- elaborati e sintetizzati in un unico dato



La sintesi dei dati in un unico dato, se da un lato semplifica l'interpretazione del fenomeno, dall'altro comporta la perdita di tutte le informazioni tranne una.

E' IMPORTANTE QUINDI SCEGLIERE QUALE INFORMAZIONE CONSERVARE



L'informazione da conservare deve descrivere in che modo si è manifestato il fenomeno, deve quindi essere legata alle modalità del carattere indagato.

L'indice centrale deve allora sintetizzare le modalità.

INDICI CENTRALI PER CARATTERI QUALITATIVI

1° CASO

Il carattere si manifesta con modalità sempre diverse (o con ripetizioni non significative).

Non esiste un modo per sintetizzare i dati in un unico dato.

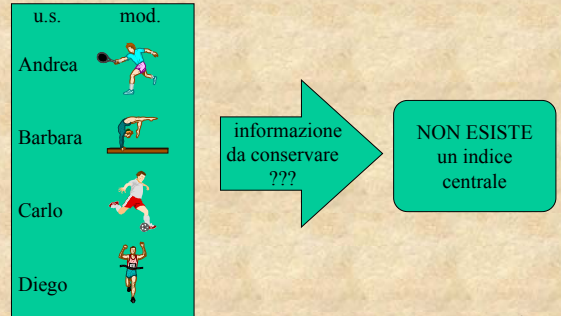
2° CASO

Il carattere si manifesta con modalità che a volte si ripetono.

E' possibile descrivere il fenomeno scegliendo di conservare solo l'informazione relativa alla modalità più frequente, la **MODA**.

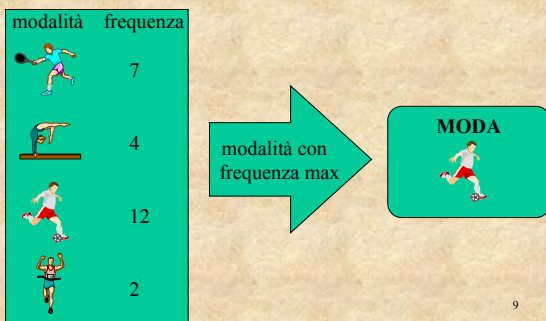
7

Esempio 1: quattro amici alla domanda "quale sport preferisci?" hanno dato quattro risposte diverse:



8

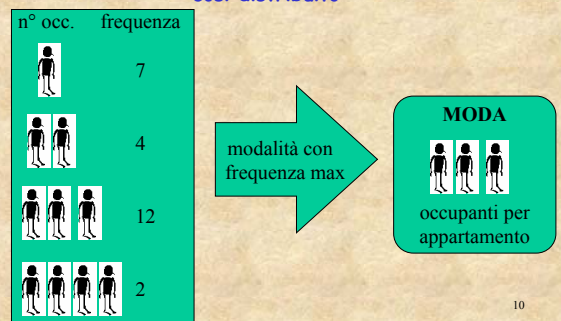
Esempio 2: tra i 4 corsi offerti da un centro sportivo, 25 ragazzi hanno optato di iscriversi a uno dei corsi secondo la seguente tabella:



9

E' evidente che la MODA può essere determinata anche nel caso di caratteri quantitativi.

Esempio: in un condominio con 25 appartamenti, il numero di occupanti ogni singolo appartamento è così distribuito:



10

ALTRI INDICI CENTRALI PER CARATTERI QUANTITATIVI

Le modalità di un carattere quantitativo, essendo espresse da numeri, si prestano ad essere sintetizzate anche nel caso di estrema eterogeneità del fenomeno.

E' possibile infatti:

- eseguire un'operazione opportunamente scelta e calcolare una **MEDIA**
- ordinarle in modo crescente o decrescente ed individuare la **MEDIANA**

11

Calcolare una **MEDIA** significa sintetizzare le modalità di un carattere quantitativo in modo da conservare una grandezza che da esse dipenda.



12

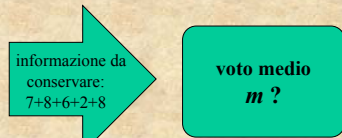
Esempio 1: si vuole valutare il profitto in matematica di uno studente che in 5 prove abbia riportato i seguenti voti:

u.s.	mod.
prova	voto
prima	7
seconda	8
terza	6
quarta	2
quinta	8



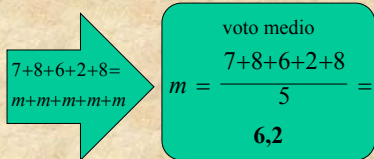
Il voto medio può essere calcolato in modo che, sostituito ai singoli voti, esso non alteri il punteggio totale raggiunto dallo studente. Si vuole cioè conservare la SOMMA dei voti.

u.s.	mod.
prova	voto
prima	7
seconda	8
terza	6
quarta	2
quinta	8



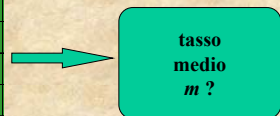
La media che conserva la somma è la MEDIA ARITMETICA

u.s.	mod.
prova	voto
prima	7
seconda	8
terza	6
quarta	2
quinta	8



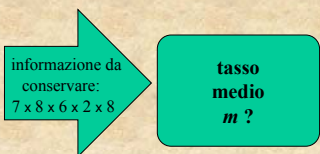
Esempio 2: si vuole valutare il tasso medio di promozione nelle 5 prove previste in un concorso, ad ognuna delle quali si può accedere solo avendo superato quella precedente. In tabella sono riportati i tassi percentuali di promozione nelle singole prove.

u.s.	mod.
prova	% prom.
prima	7
seconda	8
terza	6
quarta	2
quinta	8



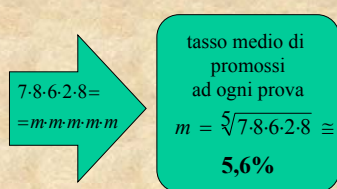
Il tasso medio può essere calcolato in modo che, sostituito ai singoli tassi di promozione, esso non alteri la percentuale totale di candidati che ha superato il concorso. Si vuole cioè conservare il PRODOTTO dei tassi.

u.s.	mod.
prova	% prom.
prima	7
seconda	8
terza	6
quarta	2
quinta	8



La media che conserva il prodotto è la MEDIA GEOMETRICA

u.s.	mod.
prova	% prom.
prima	7
seconda	8
terza	6
quarta	2
quinta	8



PRINCIPALI TIPI DI MEDIA

tipo di MEDIA	grandezza da conservare	formula $m = \dots$
ARITMETICA	$x_1 + x_2 + \dots + x_n = m + m + \dots + m$	$m = \frac{x_1 + x_2 + \dots + x_n}{n}$
GEOMETRICA	$x_1 \cdot x_2 \cdot \dots \cdot x_n = m \cdot m \cdot \dots \cdot m$	$m = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$
ARMONICA	$\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} = \frac{1}{m} + \frac{1}{m} + \dots + \frac{1}{m}$	$m = \left[\frac{1}{n} \left(\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} \right) \right]^{-1}$
QUADRATICA	$x_1^2 + x_2^2 + \dots + x_n^2 = m^2 + m^2 + \dots + m^2$	$m = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}}$

19

A volte può essere utile sintetizzare le modalità di un carattere quantitativo in modo da conservare solo quella che occupa la posizione centrale secondo un ordinamento crescente o decrescente.

L'indice centrale avente questa proprietà è la **MEDIANA**



20

Esempio: si vuole valutare il punteggio del giocatore avente una posizione *media* in classifica, dopo una gara tra cinque concorrenti che hanno totalizzato i seguenti punteggi:

u.s.	mod.
giocat.	punti
Andrea	7
Barbara	8
Carlo	6
Diego	2
Enza	8



punteggio medio m ?

21

Dopo avere compilato la classifica, si sintetizza il comportamento dei concorrenti conservando solo il punteggio del giocatore che occupa la posizione centrale, in questo caso la terza.

u.s.	mod.
giocat.	punti
Barbara	8
Enza	8
Andrea	7
Carlo	6
Diego	2

informazione da conservare: punteggio centrale

punteggio mediano
7

22

Nel caso di un numero pari di unità statistiche, la **MEDIANA** si determina solitamente calcolando la media aritmetica delle due modalità che occupano la posizione centrale

u.s.	mod.
giocat.	punti
Barbara	8
Enza	8
Andrea	7
Carlo	6

informazione da conservare: somma punteggi centrali

punteggio mediano
 $(8+7)/2 = 7,5$

23

classi di modalità

come si calcolano gli indici centrali se le modalità di un carattere quantitativo continuo sono ripartite in classi?

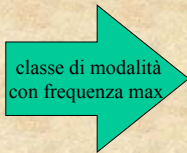
• Esempio: popolazione residente in Italia nel 1991

età	freq
[0,5)	4,9
[5,15)	11,0
[15,20)	7,6
[20,40)	30,1
[40,60)	25,3
[60,75)	14,4
[75,105)	6,7

24

La MODA conserva la modalità con frequenza più elevata.
 Nel caso di classi di modalità conserva la classe con frequenza più elevata.

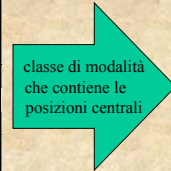
età	freq
[0,5)	4,9
[5,15)	11,0
[15,20)	7,6
[20,40)	30,1
[40,60)	25,3
[60,75)	14,4
[75,105)	6,7



CLASSE MODALE
[20,40)

La MEDIANA conserva la modalità che occupa la posizione centrale.
 Nel caso di classi di modalità e frequenze percentuali si conserva la classe cui appartengono la 50^{ma} e la 51^{ma} posizione

età	freq	freq.cum.
[0,5)	4,9	4,9
[5,15)	11,0	15,9
[15,20)	7,6	23,5
[20,40)	30,1	53,6
[40,60)	25,3	78,9
[60,75)	14,4	93,3
[75,105)	6,7	100,0



CLASSE MEDIANA
[20,40)

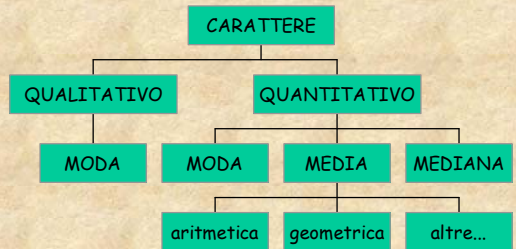
Per calcolare la MEDIA ARITMETICA si determina dapprima la media di ogni singola classe e successivamente si calcola la media ponderata di queste.

età	media	freq
[0,5)	2,5	4,9
[5,15)	10,0	11,0
[15,20)	17,5	7,6
[20,40)	30,0	30,1
[40,60)	50,0	25,3
[60,75)	67,5	14,4
[75,105)	90,0	6,7



MEDIA ARITMETICA
circa 40 anni

RIEPILOGO



Come scegliere l'indice più opportuno?

- ognuno degli indici visti per i caratteri quantitativi presenta dei vantaggi e degli svantaggi
- la scelta dipende da quale tipo di informazione si vuole ricavare dall'insieme di dati

Esempio: la retribuzione dei dipendenti di una ditta è distribuita secondo la seguente tabella

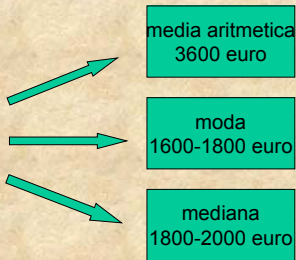
Lo stipendio è indicato in migliaia di euro mensili.

classe di reddito	n° dipend. = freq.
[1.4 , 1.6)	40
[1.6 , 1.8)	250
[1.8 , 2.0)	225
[2.0 , 5.0)	60
[5 , 10)	15
[10 , 100)	6
[100 , 200)	4

Qual è il REDDITO MEDIO ?

Qual è l'indice che trasmette l'informazione più rappresentativa dei dati?

classe di reddito	n° dipend. = freq.
[1.4 , 1.6)	40
[1.6 , 1.8)	250
[1.8 , 2.0)	225
[2.0 , 5.0)	60
[5 , 10)	15
[10 , 100)	6
[100 , 200)	4



31



MEDIA

VANTAGGIO:

- conserva un'informazione "globale", formata cioè col contributo di tutti i dati

SVANTAGGIO:

- distorce la percezione del fenomeno quando vi siano dati eccezionalmente grandi o piccoli (a meno che non siano simmetrici tra loro)

Nell'esempio è evidente che prevale lo svantaggio. L'alto stipendio dei dirigenti e degli amministratori porta ad un valore numericamente esatto ma che dice poco rispetto alla reale distribuzione dei dati.

32



MODA

VANTAGGI:

- non è influenzata da valori delle modalità che si presentino solo in casi eccezionali e quindi non significativi in termini di andamento *medio* del fenomeno
- è immediatamente individuabile anche da un diagramma intensità/frequenze

SVANTAGGIO:

- perdita totale delle informazioni relative a tutte le modalità diverse da quella avente maggiore frequenza

Nell'esempio si perde l'informazione relativa alla classe di reddito successiva a quella modale, avente anch'essa un'alta frequenza.

33



MEDIANA

VANTAGGIO:

- non è influenzata da valori delle modalità che si presentino solo in casi eccezionali e quindi non significativi in termini di andamento *medio* del fenomeno

SVANTAGGIO:

- se i dati tendono a concentrarsi intorno a valori che non occupano la posizione centrale, la mediana rappresenta solo uno *spartiacque*

Nell'esempio è forse l'indice *migliore* perché la classe mediana non è solo uno spartiacque, ma ha un'elevata frequenza: non è solo il 50% dei dipendenti a guadagnare meno di 2000 euro al mese, ma addirittura più dell'85%.

34