

LEZIONI DI STATISTICA MEDICA

Dott. SIMONE ACCORDINI

Lezione n.7

- Misure di posizione: media aritmetica



Sezione di Epidemiologia & Statistica Medica
Università degli Studi di Verona

MEDIA ARITMETICA

La media aritmetica di una distribuzione di frequenza è la somma dei **valori osservati** diviso il **numero totale delle osservazioni**

→ **misura dell'intensità media del fenomeno**

Formalmente: siano (x_1, x_2, \dots, x_n) le osservazioni della variabile X su un campione di n unità statistiche, allora

$$\bar{x} = \sum_{i=1}^n x_i / n = (x_1 + x_2 + \dots + x_n) / n$$

esempio:
(8 osservazioni)

5	16	13	27	11	5	13	13
↑	↑	↑	↑	↑	↑	↑	↑
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8

$$\bar{x} = (5+16+13+27+11+5+13+13)/8 = 103/8 = 12.9$$



MEDIA ARITMETICA PONDERATA - I

Se più unità statistiche presentano lo stesso valore

⇒ la media aritmetica può essere calcolata moltiplicando quel valore per la frequenza con cui compare nella distribuzione

$$\bar{x} = \frac{\sum_{i=1}^k x_i n_i}{n} = \frac{x_1 n_1 + x_2 n_2 + \dots + x_k n_k}{n}$$



k = numero di valori differenti osservati

x_i = i -esimo valore osservato

n_i = frequenza corrispondente al valore x_i → PESO



Nel caso di variabili quantitative continue, i dati sono spesso organizzati in classi per una migliore sintesi descrittiva

Distribuzione del FEV₁
(cl/sec) in 54 soggetti
maschi di età 20-44 anni
(indagine *European
Community Respiratory
Health Survey* - ECRHS)



	n_i
[223 – 270.25)	0
[270.25 – 317.5)	3
[317.5 – 364.75)	9
[364.75 – 412)	10
[412 – 459.25)	14
[459.25 – 506.5)	8
[506.5 – 553.75)	7
[553.75 – 601]	3
TOTALE	54



Distribuzione del FEV₁ (cl/sec) in 54 soggetti maschi di età 20-44 anni (ECRHS)

Per il calcolo della media ponderata, le osservazioni di una classe si assumono coincidenti con il **valore centrale della classe**

punto centrale della 1^a classe:
(223+270.25)/2

	x_i	n_i	$x_i n_i$
[223 – 270.25)	246.625	0	0
[270.25 – 317.5)	293.875	3	881.625
[317.5 – 364.75)	341.125	9	3070.125
[364.75 – 412)	388.375	10	3883.750
[412 – 459.25)	435.625	14	6098.750
[459.25 – 506.5)	482.875	8	3863
[506.5 – 553.75)	530.125	7	3710.875
[553.75 – 601]	577.375	3	1732.125
TOTALE		54	23240.25

$$\bar{x} = (x_1 n_1 + x_2 n_2 + \dots + x_8 n_8) / n = 23240.25 / 54 = 430.375 \text{ cl/s}$$



MEDIA ARITMETICA PONDERATA - II

$$\bar{x} = \frac{\bar{x}_1 n_1 + \bar{x}_2 n_2 + \dots + \bar{x}_k n_k}{n_1 + n_2 + \dots + n_k}$$



k = numero di gruppi

\bar{x}_i = media aritmetica per il gruppo i-esimo

n_i = dimensione campionaria del gruppo i-esimo → PESO

\bar{x} = media aritmetica complessiva

esempio: valore medio dell'altezza nei maschi e nelle femmine matricole della Facoltà di Medicina (A.A. 95/96)

Sesso	n_i	\bar{x}_i
maschi	34	177
femmine	91	166.1
Totale	125	

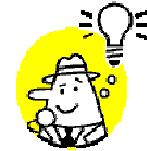
$$\bar{x} = \frac{177 * 34 + 166.1 * 91}{125} = 169.1 \text{ cm}$$



La media aritmetica gode di diverse proprietà, le due principali dal punto di vista applicativo sono legate al concetto di **SCARTO**:

PRIMA PROPRIETA' DELLA MEDIA ARITMETICA

la **somma algebrica** degli scarti delle osservazioni dalla loro media aritmetica è pari a **zero**



$$\sum_{i=1}^n (x_i - \bar{x}) = \underbrace{(x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_n - \bar{x})}_{\text{scarto (distanza) della prima osservazione dalla media}} = 0$$

scarto (distanza) della prima osservazione dalla media

⇒ media aritmetica = **punto 'centrale' della distribuzione**



Verifica empirica della prima proprietà:

FEV ₁	x _i	n _i	(x _i - \bar{x})n _i
[223 - 270.25]	246.625	0	0
[270.25 - 317.5]	293.875	3	-409.5
[317.5 - 364.75]	341.125	9	-803.25
[364.75 - 412]	388.375	10	-420
[412 - 459.25]	435.625	14	73.5
[459.25 - 506.5]	482.875	8	420
[506.5 - 553.75]	530.125	7	698.25
[553.75 - 601]	577.375	3	441
TOTALE		54	0

$$-409.5 = (293.875 - 430.375) \times 3$$

scarti negativi

scarti positivi

$$\bar{x} = 430.375 \text{ cl/s}$$

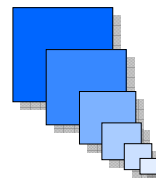
media ponderata

la somma algebrica degli scarti delle osservazioni dalla loro media aritmetica è pari a zero



SECONDA PROPRIETA' DELLA MEDIA ARITMETICA

la somma del quadrato degli scarti delle osservazioni da un qualsiasi valore A è **minima** quando A è pari alla media aritmetica



$$\sum_{i=1}^n (x_i - \bar{x})^2 \leq \sum_{i=1}^n (x_i - A)^2$$

QUALE MISURA DI POSIZIONE UTILIZZARE?



TIPO DI VARIABILE	OPERAZIONI CONSENTITE	MODA	MEDIANA	MEDIA
nominale	= ≠	Sì	No	No
ordinale	= ≠ < >	Sì	Sì	No
quantitativa	= ≠ < > - + (/ *)	Sì	Sì	Sì

CONFRONTO TRA LE MISURE DI POSIZIONE PER UNA VARIABILE QUANTITATIVA

esempio:

durata della degenza ospedaliera
di 9 individui (in giorni)



CAMPIONE 4 5 12 3 4 4 95 8 6

Moda = 4

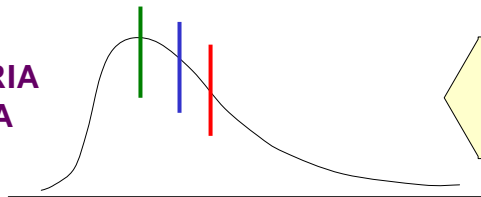
Mediana = 5

Media = 15.7 (senza il soggetto *outlier* sarebbe 5.7)

La media aritmetica è poco "robusta" in presenza di
valori anomali (outliers)!

RELAZIONE TRA MODA, MEDIANA E MEDIA ARITMETICA

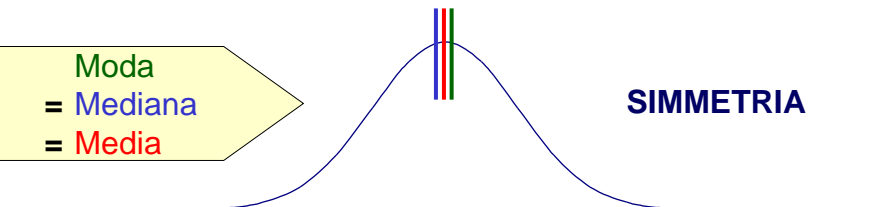
**ASIMMETRIA
POSITIVA**



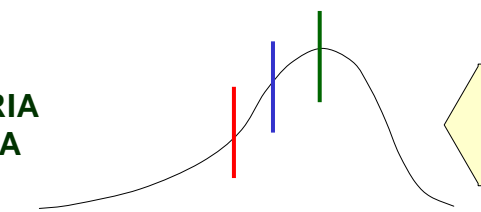
Moda
< Mediana
< Media

Moda
= Mediana
= Media

SIMMETRIA



**ASIMMETRIA
NEGATIVA**



Moda
> Mediana
> Media

CONFRONTO TRA LE MISURE DI POSIZIONE PER UNA VARIABILE QUANTITATIVA

MODA

Buona misura quando un valore ha una frequenza relativa molto elevata



MEDIANA

Buona misura con distribuzioni asimmetriche (es. *tempo di sopravvivenza*)

MEDIA ARITMETICA

Buona misura con distribuzioni simmetriche (es. *molti parametri biologici*)

Facile da trattare matematicamente

Utilizza tutta l'informazione contenuta nei dati



Dipende dal raggruppamento arbitrario dei dati

Difficile da trattare matematicamente

E' inaffidabile in caso di distribuzioni asimmetriche