LE MISURE DI TENDENZA CENTRALE

Dott. Giuseppe Di Martino
Scuola di Specializzazione in Igiene e Medicina Preventiva

OBIETTIVI

Individuare un indice che rappresenti significativamente un insieme di dati statistici

Esempio: Nella tabella seguente sono riportati i valori del tasso glicemico rilevati su 10 pazienti:

Paziente	Glicemia (mg/100cc)
1	x_1=103
2	x_2=97
3	x_3=90
4	x_4=119
5	x_5=107
6	x_6=71
7	x_7=94
8	x_8=81
9	x_9=92
10	x_{10}=96
Totale	**950**

Calcolo delle frequenze di ogni classe: assolute e relative percentuali

Classi di valori di glicemia	Frequenza assoluta	Frequenza relativa
70 —| 80	1	1 / 10 ·100% = 10 %
80 —| 90	2	2 / 10 ·100% = 20 %
90 —| 100	4	4 / 10 ·100% = 40 %
100 —| 110	2	2 / 10 ·100% = 20 %
110 —| 120	1	1 / 10 ·100% = 10 %
Totale	**10**	**100 %**

LE MISURE DI POSIZIONE

- Media aritmetica
- Mediana
- Moda
- Media armonica
- Media geometrica

LA MEDIA ARITMETICA

DEFINIZIONE: La <u>**media aritmetica**</u> è quel valore che avrebbero tutte le osservazioni se non ci fosse la variabilità (casuale o sistematica).

Più precisamente, è quel valore che sostituito a ciascun degli **n** dati ne fa rimanere costante la somma.

LA MEDIA ARITMETICA

Dato un insieme di n elementi $\{x_1, x_2, \dots x_n\}$

Si dice **media aritmetica semplice** di n numeri il numero che si ottiene dividendo la loro somma per n.

$$\overline{x} = \frac{x_1 + x_2 + \dots x_n}{n}$$

LA MEDIA ARITMETICA

Formalmente possiamo esprimere la media aritmetica semplice attraverso la seguente formula:

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

LA MEDIA ARITMETICA

Nell'Esempio in esame si ha:

$$\bar{x} = \frac{\sum_{\mathrm{i=1}}^{\mathrm{n}} x_i}{\mathrm{n}} = \frac{950}{10} = 95mg/100cc$$

Esempio Riportiamo i tempi di sopravvivenza (mesi) di 19 pazienti con cancro dell'addome

Mesi di sopravvivenza (x_i)	Frequenza (f_i)
8,5	2
9,2	4
7,3	8
6,8	2
10,1	3
Totale	**19**

$x_i \cdot f_i$
17
36,8
58,4
13,6
30,3
156,1

LA MEDIA ARITMETICA PESATA

Si dice **media aritmetica pesata** di n numeri:

$$\frac{x_1 \cdot p_1 + x_2 \cdot p_2 + \ldots + x_m \cdot p_m}{p_1 + p_2 + \ldots + p_m}$$

Dove i pesi p_j sono le frequenze assolute di ogni modalità

LA MEDIA ARITMETICA PESATA

Nell’esempio precedente la media aritmetica (ponderata) è data da:

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n} = \frac{156,1}{19} = 8,2 mesi}$$

LE PROPRIETA' DELLA MEDIA ARITMETICA

- ✓ compresa tra il minimo dei dati e il massimo dei dati

- ✓ la somma degli scarti dalla media aritmetica <u>è sempre uguale a zero</u>

$$\sum (x_i - \bar{x}) = 0$$

- ✓ la somma degli scarti al quadrato dalla media aritmetica assume valore minimo

$$\sum (x_i - \bar{x})^2 < \sum (x_i - z)^2$$

- ✓ la media dei valori $k \cdot x_i$ è pari a la media aritmetica · k (dove k è un numero reale qualsiasi)

- ✓ la media dei valori $x_i \pm h$ è pari a media aritmetica ± h (dove h è un numero reale qualsiasi)

Esempio Lunghezza (cm) in un campione di 66 neonati

55.9	51.3	53.0	50.5	54.9	53.4	53.7	50.0	53.8	52.5	55.6
47.9	54.3	56.0	51.8	54.1	55.6	57.6	53.3	51.1	54.3	52.3
55.3	52.4	56.3	53.7	54.4	54.5	52.5	52.7	51.4	55.5	52.7
57.4	51.7	50.8	49.4	52.0	53.7	54.8	53.5	49.5	50.4	56.4
48.5	53.1	49.5	53.2	53.1	52.6	54.3	54.9	53.7	55.2	51.7
51.4	51.0	52.6	52.8	59.3	56.4	51.5	58.9	52.3	54.6	53.8

la media aritmetica dei 66 valori di lunghezza è:

(55.9+51.3+53.0+50.5+54.9+53.4+…+53.8)/66

= 3517.500/66

= **53.295**

MEDIA ARITMETICA PER DATI RAGGRUPPATI IN CLASSI

Valore centrale della classe X_i	f_i	%	$X_i f_i$
48.0	2	3.03	96.00
49.5	3	4.55	148.50
51.0	12	18.18	612.00
52.5	15	22.73	787.50
54.0	14	21.21	756.00
55.5	10	15.15	555.00
57.0	5	7.58	285.00
58.5	4	6.06	234.00
60.0	1	1.52	60.00
	66	**100**	**3534.00**

$$\bar{x} = \frac{48.0 \times 2 + 49.5 \times 3 + \ \dots 60.0 \times 1}{2 + 3 + \ \dots 1} = \frac{3534.0}{66} = 53.545$$

IMPORTANTE

La **media aritmetica** è la misura di posizione più usata ma a volte, altre misure come la **mediana** e la **moda** si dimostrano più utili.

Si consideri un campione di valori di VES (*velocità di eritrosedimentazione*, mm/ora) misurati in 7 pazienti:

{8, 5, 7, 6, 35, 5, 4}

In questo caso, la media che è = 10 mm/ora non è un valore tipico della distribuzione: **soltanto un valore su 7 è superiore alla media!**

Limite della media aritmetica:
è notevolmente influenzata dai valori estremi della distribuzione.

Esempio: Età alla morte di 5 soggetti

$x_1 = 34$ anni $x_2 = 70$ anni $x_3 = 74$ anni

$x_4 = 64$ anni $x_5 = 68$ anni

La media aritmetica è pari a:

$$\boxed{\bar{x} = (34 + 70 + 74 + 64 + 68) / 5 = 62 anni}$$

LA MEDIANA

DEFINIZIONE

La **mediana (Me)** è quell'osservazione che bipartisce la distribuzione in modo tale da lasciare al "di sotto" lo stesso numero di termini che lascia al "di sopra"

L'idea che è alla base della **mediana** è di cercare un numero che sia più grande di un 50% delle osservazioni e più piccolo del restante 50%

LA MEDIANA

Ritornando all'Esempio della Glicemia, per il calcolo della mediana è necessario disporre i dati in ordine crescente:

71, 81, 90, 92, 94, 96, 97, 103, 107, 119

$$\boxed{Me = (94+96)/2 = 95 \text{ mg/100 cc}}$$

LA MEDIANA

Il fatto che mediana e media aritmetica in questo caso coincidano non è casuale in quanto la distribuzione è simmetrica.
Ma, in generale, ciò non avviene!

Vantaggio nell'uso della mediana: non è influenzata dalle osservazioni aberranti o estreme.

Fasi operative per il calcolo della mediana

1) ordinamento crescente dei dati

2) se il numero di dati n è dispari, la mediana corrisponde al dato che occupa la **(n+1)/2** esima posizione

3) se il numero di dati n è pari, la mediana è data dalla **media aritmetica** dei due dati che occupano la posizione **n/2** e quella **(n/2)+1**

LA MODA

DEFINIZIONE

La **Moda (Mo)** è l'osservazione che si verifica con maggiore frequenza in una data distribuzione

NB: Si possono avere anche più valori modali

LA MODA

Mesi sopravvivenza (x_i)	Frequenze	Frequenze Cumulate	Cum %
6,8	2	2	10.5
7,3	8	10	52.6
8,5	2	12	63.1
9,2	4	16	84.2
10,1	3	19	100
Totale	**19**		

Media aritmetica= 8,2 mesi
Mediana= 7,3 mesi
Moda=7,3 mesi

In presenza di una distribuzione di frequenze è necessario considerare le frequenze cumulate

Voti ordinati (x_i)	Frequenze (f_i)	Freq. Cum. (F_i)	Freq.Cum. (F_i %)
18	2 (10.5)	2	10.5
20	4 (21.0)	2+4 = 6	31.5
22	8 (42.1)	6+8 = 14	73.6
24	2 (10.5)	14+2 = 16	84.1
27	2 (10.5)	16+2 = 18	94.6
30	1 (5.4)	18+1 = 19	100
Totale	**19**		

Voti ordinati	Frequenze	Freq.Cum. F_i	Freq.Cum. F_i%
18	2 (10.5)	2	10.5
20	4 (21.0)	6	31.5
22	8 (42.1)	14	73.6
	2 (10.5)	16	84.1
	2 (10.5)	18	94.6
	1 (5.4)	19	100
	19 (100.0)		

La Mediana

I QUANTILI

✓Generalizzano la mediana

✓L'idea alla base di un **quantile(p)**, dove $p \in [0;1]$, è di cercare un numero che sia più grande p% dei dati osservati e più piccolo del restante (1-p%) dei dati.

QUANTILI

I quantili con p uguale a 0,25; 0,50 e 0,75 vengono chiamati rispettivamente il primo, il secondo e il terzo **quartile**.

Dividono la popolazione in quattro parti uguali.
Si osservi che il 2 quartile coincide con la mediana.

I quantili con p = 0,01; … ; 0,99 si chiamano **percentili**.

QUALE MISURA DI POSIZIONE USARE?

A quale misura di tendenza centrale ci riferiamo?

- Il proprietario di una ditta afferma "Lo stipendio mensile nella nostra ditta è **2.700** euro"
- Il sindacato dei lavoratori dice che "lo stipendio medio è di **1.700** euro".
- L'agente delle tasse dice che "lo stipendio medio è stato di **2.200** euro".

Queste risposte diverse sono state ottenute tutte dai dati della seguente tabella.

Stipendio mensile	N° di lavoratori
1.300	2
1.700	22
2.200	19
2.600	3
6.500	2
9.400	1
23.000	1

Media aritmetica = 2.700€

Mediana = 2.200€

Moda = 1.700€

INTERPRETAZIONE DELLE MISURE DI POSIZIONE

- La **media aritmetica** indica che, se il denaro fosse distribuito in modo che ciascuno ricevesse la stessa somma, ciascun dipendente avrebbe avuto 2.700 euro

- La **moda** ci dice che la paga mensile più comune è di 1.700.euro

- La moda si considera spesso come il valore tipico dell'insieme di dati poiché è quello che si presenta più spesso. **Non tiene però conto degli altri valori** e spesso in un insieme di dati vi è **più di un valore** che corrisponde alla definizione di moda.

- La **mediana** indica che circa metà degli addetti percepiscono meno di 2.200.euro, e metà di più.

- La mediana **non è influenzata dai valori estremi** eventualmente presenti ma solo dal fatto che essi siano sotto o sopra il centro dell'insieme dei dati.

RELAZIONE TRA MEDIA MODA E MEDIANA

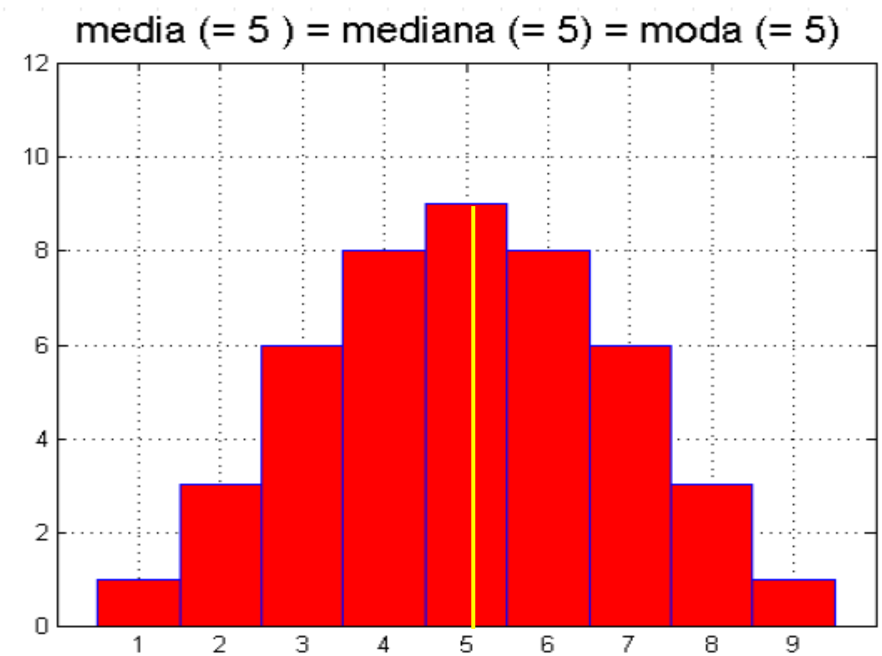
In una distribuzione perfettamente **simmetrica**, la media, la mediana e la moda hanno lo stesso valore. In una distribuzione **asimmetrica**, la media si posiziona nella direzione dell'asimmetria. Nelle distribuzioni di dati biologici, l'asimmetria è quasi sempre verso destra (asimmetria positiva, verso i valori più elevati), e quindi la media è maggiore della mediana o della moda

DISTRIBUZIONE SIMMETRICA

Le osservazioni equidistanti dalla mediana (coincidente in questo caso col massimo centrale) presentano la stessa frequenza relativa

Un esempio importante è fornito dalla ***distribuzione normale***

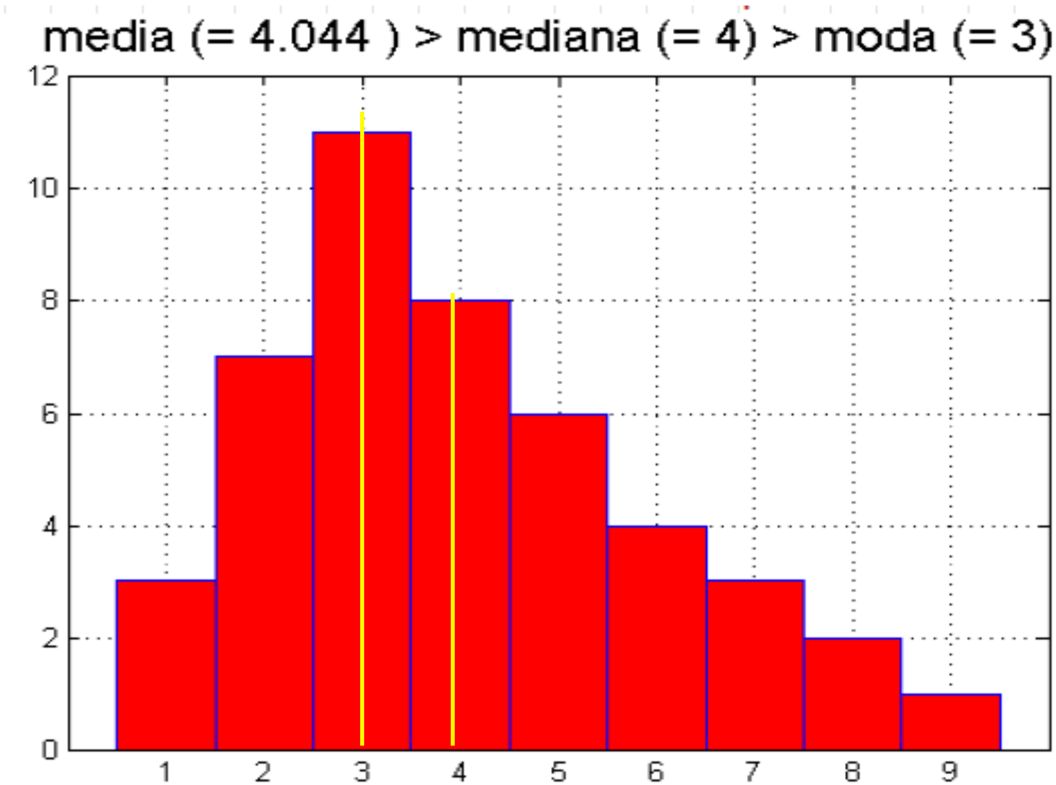
Media = Mediana = Moda



DISTRIBUZIONE ASIMMETRICA POSITIVA

La curva di frequenza ha una coda più lunga a destra del massimo centrale

Media > Mediana > Moda



DISTRIBUZIONE ASIMMETRICA NEGATIVA

La curva di frequenza ha una coda più lunga a sinistra del massimo centrale

Media < Mediana < Moda