

LEZIONI DI STATISTICA MEDICA

Prof. SIMONE ACCORDINI

Lezione n.6

- Indici di posizione: moda, mediana, percentili

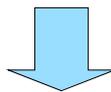
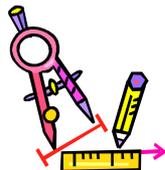


*Sezione di Epidemiologia & Statistica Medica
Università degli Studi di Verona*

“un qualsiasi insieme di dati porta in sé una
certa quantità di informazione”

OBIETTIVO:
riassumere l'informazione raccolta in modo
SINTETICO ed EFFICACE

diversi strumenti e possibilità
offerti dalla statistica



**STATISTICHE DI
BASE**



La distribuzione di una variabile è compiutamente descritta da tre **CARATTERISTICHE FONDAMENTALI**:

- ✓ **tendenza centrale o posizione**
- ✓ **dispersione o variabilità**
- ✓ **forma**

Le **misure descrittive sintetiche** sono chiamate:

STATISTICHE
(\bar{x} , s , p)



quando sono calcolate su un **CAMPIONE**
(si esprimono con lettere dell'alfabeto latino)

PARAMETRI
(μ , σ , π)



quando descrivono una **POPOLAZIONE**
(si esprimono con lettere dell'alfabeto greco)



INDICI DI POSIZIONE

(measures of location or central tendency)

1. MODA
2. MEDIANA
3. MEDIA ARITMETICA



MODA

E' la scelta fatta dalla maggioranza della popolazione, lo stile che "tutti" seguono

in statistica non è diverso

Si definisce moda (*classe modale*) di una distribuzione di frequenza la modalità o il valore (*intervallo di classe*) della variabile a cui corrisponde la massima frequenza.

esempio:

X = tipo di parto
(50 neonati)

MODA

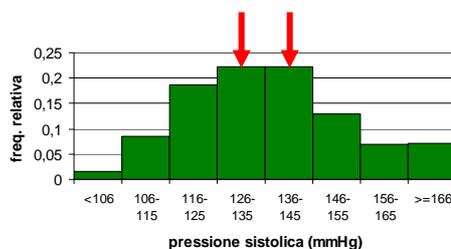
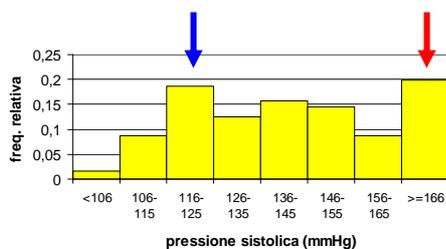
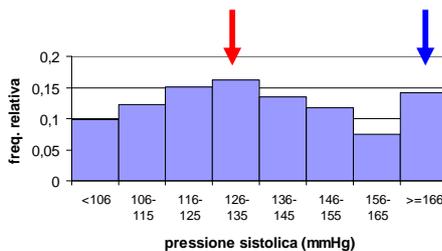
modalità x_i	frequenza assoluta n_i	frequenza relativa p_i	frequenza relativa percentuale p_i (%)
normale	35	0.70	70%
forcipe	1	0.02	2%
cesareo	14	0.28	28%
TOTALE	50	1.00	100%



MA LA MODA E' SEMPRE UN CORRETTO INDICE DI POSIZIONE?

Pressione sistolica in tre gruppi di maschi giapponesi: nativi, prima e seconda generazione di immigrati negli USA (Issei e Nisei).

Winkelstein et al. *Am J Epidemiol* 1975;102:502-13.



MISURE D'ORDINE PER UNA DISTRIBUZIONE

RANGO: posizione di un'osservazione x_i in una serie di dati ordinati in modo crescente

esempio: 0.8, 2.2, 0.1, 1.6, 1.9, 0.2, 1.0, 1.2, 1.9 (serie statistica)

0.1, 0.2, 0.8, 1.0, 1.2, 1.6, 1.9, 1.9, 2.2 (serie ordinata)

rango 5

RANGO PERCENTILE (R_p): proporzione di osservazioni che hanno un valore $\leq x_k$

$$R_p = \frac{\text{rango}(x_k)}{n+1} * 100$$



Esempio: nelle seguenti tabelle si riportano le osservazioni ordinate del peso per n soggetti

$n = 5$

PESO (kg)	53	55	60	61	63
-----------	----	----	----	----	----

Rango = 3

$$R_p = [3 / (5+1)] * 100 = 50\%$$



$n = 59$

PESO (kg)	53	55	60	61	63	65	92
-----------	----	----	----	----	----	----	-------	----

Rango = 3

$$R_p = [3 / (59+1)] * 100 = 5\%$$



MEDIANA

- valore centrale di una serie **ORDINATA** di dati ($R_p = 50\%$)
- le osservazioni vengono separate dal valore mediano in due parti con uguale frequenza



Esempio:

campione di 5 unità ($n = 5$)
 $X =$ altezza



175 cm



160 cm



185 cm



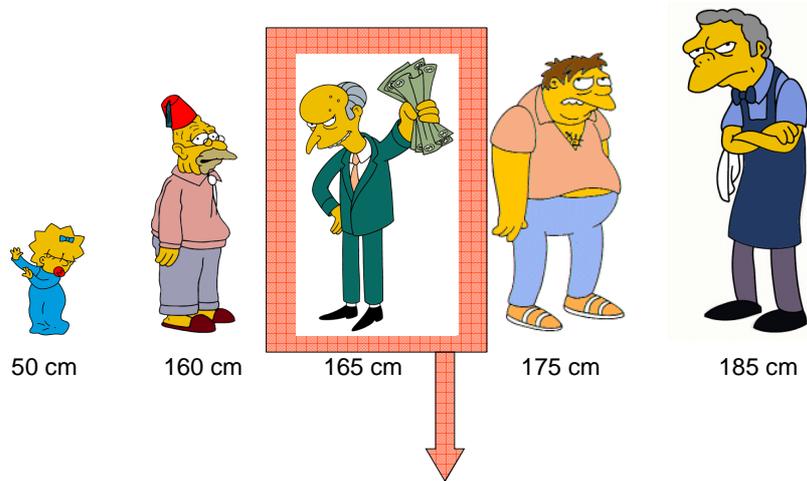
50 cm



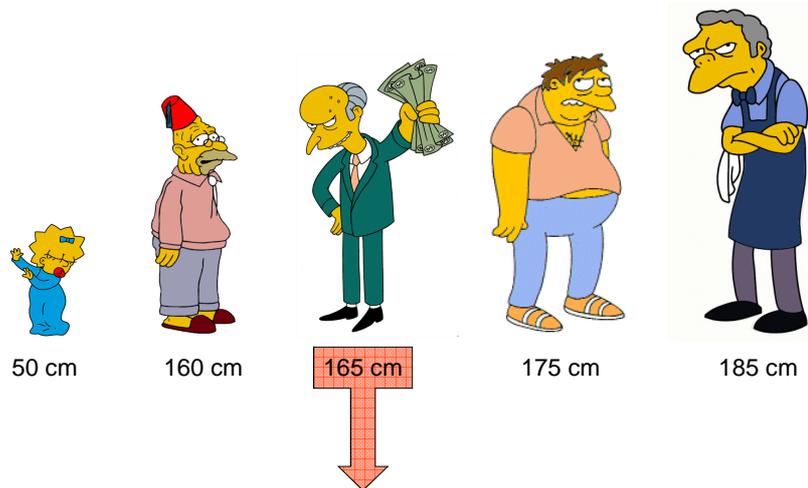
165 cm



1. ordino le unità secondo un valore crescente di altezza



2. identifico l'unità centrale nella serie ordinata di dati



3. la mediana è il **VALORE** che la variabile altezza assume sull'unità che divide il campione in due parti con uguale frequenza



K-ESIMO PERCENTILE

è la modalità o il valore (x_k) della variabile tale che il k% delle osservazioni risultano essere $\leq x_k$

25-esimo percentile = I° quartile

→ 25% delle osservazioni è $\leq x_{0.25}$

50-esimo percentile = II° quartile = mediana

→ 50% delle osservazioni è $\leq x_{0.5}$

75-esimo percentile = III° quartile

→ 75% delle osservazioni è $\leq x_{0.75}$



Calcolo del k-esimo percentile - I (dati disponibili a livello individuale)

- Si determina il **rango** (posizione) dell'osservazione corrispondente al **k-esimo rango percentile**

$$\text{rango} = (n+1) * \text{k-esimo rango percentile} / 100$$

- Utilizzando la distribuzione delle frequenze cumulate, si individua l'**osservazione** (x_k) corrispondente al k-esimo rango percentile

esempio:

	Valori	Freq.	Cumul.
	0.1	1	1
	0.2	1	2
	0.8	1	3
	1.0	1	4
MEDIANA	1.2	1	5
	1.6	1	6
	1.9	2	8
	2.2	1	9
	Totale	9	

Calcolo della mediana:

(rango percentile = 50)

$$1. \text{ rango} = (9+1) * 50 / 100 \\ = 10 / 2 = 5$$

$$2. \text{ mediana: } x_{0.5} = 1.2$$



Calcolo della mediana

- numero di osservazioni dispari

⇒ esiste un'unica osservazione centrale con rango: $(n+1)/2$

⇒ la mediana è data da: $x_{0,5} = x_{(n+1)/2}$

- numero di osservazioni pari

⇒ le osservazioni centrali sono due con rango: $n/2$ e $n/2 + 1$

⇒ la mediana è data da: $x_{0,5} = (x_{n/2} + x_{(n/2+1)})/2$

(se X è quantitativa)



esempio:

n = 9 0.1, 0.2, 0.8, 1.0, **1.2**, 1.6, 1.9, 1.9, 2.2

$$\text{rango} = (9+1)/2 = 5$$

$$\Rightarrow x_{0,5} = 1.2$$

n = 8 0.1, 0.2, 0.8, **1.0**, **1.2**, 1.6, 1.9, 2.2

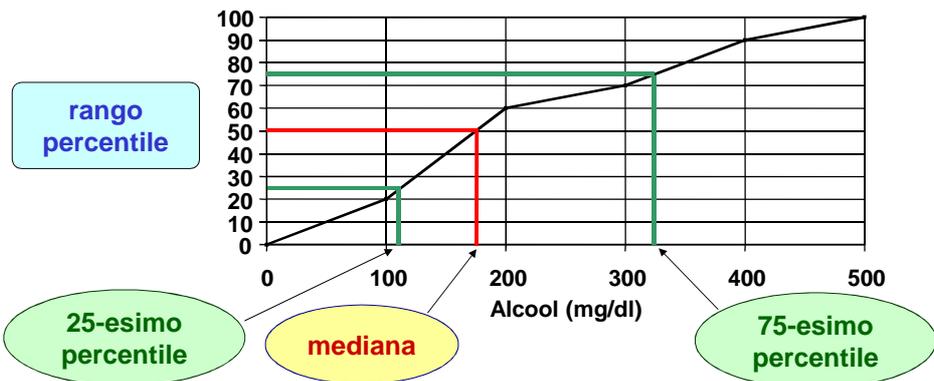
$$\text{rango: } 8/2 = 4 \text{ e } 8/2 + 1 = 5$$

$$\Rightarrow x_{0,5} = (1.0+1.2)/2 = 1.1$$



Calcolo del k -esimo percentile - II (dati raggruppati in intervalli - metodo grafico)

esempio: distribuzione del livello ematico di alcool nella circolazione sanguigna
in un gruppo di 250 soggetti



Il 25-esimo percentile (1° quartile) è circa pari a 110 mg/dl

La **mediana** (50-esimo percentile) è circa pari a 175 mg/dl

Il 75-esimo percentile (3° quartile) è circa pari a 320 mg/dl