

## STATISTICA MEDICA



*Dott.ssa Marta Di Nicola*

*N.P.D. 3° Blocco 2° piano*

*0871-3554007*

*m.dinicola@unich.it*

<http://www.biostatistica.unich.it>

Dott.ssa Marta Di Nicola

## STATISTICA DESCRITTIVA



### LE MISURE DI VARIABILITA'

Dott.ssa Marta Di Nicola

In assenza di variabilità in una popolazione la statistica non sarebbe necessaria: un singolo *elemento* o unità campionaria sarebbe sufficiente a determinare tutto ciò che occorre sapere su una popolazione. Ne consegue, perciò, che nel presentare informazioni su un campione non è sufficiente fornire semplicemente una misura della *media* ma servono informazioni sulla **variabilità**.

Dott.ssa Marta Di Nicola

**Esempio** Si considerino inizialmente, le seguenti due distribuzioni di valori riferiti all'età di 10 individui:



Soggetti	I gruppo	II gruppo
1	20aa	10aa
2	30aa	25aa
3	40aa	40aa
4	50aa	55aa
5	60aa	70aa
Tot	200aa	200aa
Media Aritmetica	$200aa/5=40aa$	$200aa/5=40aa$

Dott.ssa Marta Di Nicola

### LE MISURE DI VARIABILITÀ



- ✓ Campo di variazione (range);
- ✓ Devianza;
- ✓ Varianza;
- ✓ Deviazione Standard;
- ✓ Coefficiente di variazione (variabilità relativa).

Dott.ssa Marta Di Nicola

### IL CAMPO DI VARIAZIONE O RANGE



**DEFINIZIONE:** Il **Campo di variazione o Range** corrisponde alla differenza fra la modalità più piccola e la modalità più grande della distribuzione.

$$R = X_{\max} - X_{\min}$$

Dott.ssa Marta Di Nicola

### Limiti del campo di variazione:



- ✓ è troppo influenzato dai valori estremi;
- ✓ tiene conto dei due soli valori estremi, trascurando tutti gli altri.

Dott.ssa Marta Di Nicola

Occorre allora un indice di dispersione che consideri tutti i dati (e non solo quelli estremi), confrontando questi con il loro valor medio.

1<sup>a</sup> idea  $\longrightarrow \sum_{i=1}^n (x_i - \bar{x})$

2<sup>a</sup> idea  $\longrightarrow \sum_{i=1}^n |x_i - \bar{x}|$

3<sup>a</sup> idea  $\longrightarrow \sum_{i=1}^n (x_i - \bar{x})^2$

Dott.ssa Marta Di Nicola

### LA DEVIANZA

**DEFINIZIONE:** La somma dei quadrati degli scarti dalla media aritmetica

$$\sum_{i=1}^k (x_i - \bar{x})^2 f_i$$

Dott.ssa Marta Di Nicola

### Esempio 9 Valori del tasso glicemico in 10 soggetti

X <sub>i</sub> (glicemia mg/100cc)	x <sub>i</sub> - $\bar{x}$	(x <sub>i</sub> - $\bar{x}$ ) <sup>2</sup>
103	+8	64
97	+2	4
90	-5	25
119	+24	576
107	+12	144
71	-24	576
96	+1	1
$\bar{x} = 95$	94	1596

**La quantità 1596 esprime la Devianza della distribuzione (Dev).**

Dott.ssa Marta Di Nicola

### LA VARIANZA

**DEFINIZIONE:** La somma dei quadrati degli scarti dalla media aritmetica divisi per la numerosità

$$\frac{\sum_{i=1}^k (x_i - \bar{x})^2 f_i}{n}$$

Dott.ssa Marta Di Nicola

### LA DEVIAZIONE STANDARD

**DEFINIZIONE:** La radice quadrata della varianza

$$\sqrt{\frac{\sum_{i=1}^k (x_i - \bar{x})^2 f_i}{n - 1}}$$

Dott.ssa Marta Di Nicola



Calcolare la **deviazione standard (DV)** delle seguenti 10 osservazioni (mm):

81 79 82 83 80 78 80 87 82 82

1. Si calcoli la media,  $\bar{x}$ :

$$\bar{x} = \frac{\sum x}{n} = \frac{814}{10} = 81.40$$

2. Si calcolino gli scarti dalla media sottraendo da ciascun valore la media; si elevi al quadrato tale quantità (il quadrato elide il segno -):

Dott.ssa Marta Di Nicola



$$\begin{aligned} (81-81.4)^2 &= 0.16 & (78-81.4)^2 &= 11.56 \\ (79-81.4)^2 &= 5.76 & (80-81.4)^2 &= 1.96 \\ (82-81.4)^2 &= 0.36 & (87-81.4)^2 &= 31.36 \\ (83-81.4)^2 &= 2.56 & (82-81.4)^2 &= 0.36 \\ (80-81.4)^2 &= 1.96 & (82-81.4)^2 &= 0.36 \end{aligned}$$

3. Si sommino tali quantità: la somma è pari a 56.4. La somma  $\sum (x - \bar{x})^2$  è detta **somma dei quadrati degli scarti** o, più semplicemente, **somma dei quadrati**.

Dott.ssa Marta Di Nicola



4. Si divida tale quantità per il numero di osservazioni meno 1:

$$\frac{\text{somma dei quadrati}}{(n-1)} = \frac{\sum (x - \bar{x})^2}{n-1} = \frac{56.4}{9} = 6.27$$

5. La deviazione standard è la radice quadrata di tale valore:

$$DS = \sqrt{6.27} = 2.50 \text{ mm}$$

Quindi la **deviazione standard** del campione di 10 unità estratto dalla popolazione è pari a 2.50 mm.

Dott.ssa Marta Di Nicola



### SCARTO INTERQUARTILE

**Scarto interquartile = (3° quartile) - (1° quartile)**

E' molto più *resistente* della varianza in presenza di poche osservazioni estreme. Per questo motivo è usato soprattutto nelle situazioni in cui si sospetta la possibile presenza di osservazioni anomale.

Dott.ssa Marta Di Nicola

## IL COEFFICIENTE DI VARIAZIONE



$$C.V. = \frac{\text{(deviazione standard)}}{\text{(media aritmetica)}}$$

La variabilità guarda alle differenze tra le unità sperimentali. E' però evidente che il significato pratico delle differenze può dipendere dal livello del fenomeno considerato.

Può quindi essere interessante disporre di una qualche misura di variabilità *aggiustata* in qualche maniera per tenere conto del livello del fenomeno.

Dott.ssa Marta Di Nicola

## Esempio



Data la media e la deviazione standard di campioni di (a) neonati, (b) bambini di tre anni e (c) bambini di 10 anni, dobbiamo chiederci se la **variabilità relativa** si modifica con l'età.

(a) Neonati  $\bar{x} = 3,1$  Kg; DS = 0,23 Kg

$$CV = 0,23/3,1 \times 100 = 7,4\%$$

(b) Bambini di 3 anni  $\bar{x} = 16,0$  Kg; DS = 4,5 Kg

$$CV = 4,5/16,0 \times 100 = 28,1\%$$

(c) Bambini di 10 anni  $\bar{x} = 35,0$  Kg; DS = 13,8 Kg

$$CV = 13,8/35,0 \times 100 = 39,4\%$$

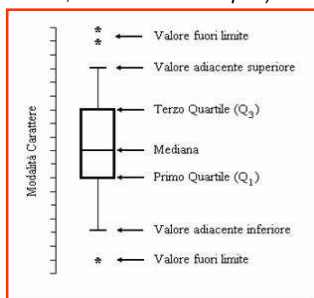
Osservando i tre valori del CV, si può notare che la **variabilità relativa** aumenta con l'età.

Dott.ssa Marta Di Nicola

## BOX-PLOT



Il nome deriva dall'inglese (*box and whiskers plot* spesso, anche in italiano, abbreviato in *boxplot*).



Dott.ssa Marta Di Nicola