STATISTICA MEDICA

Prof.ssa Marta Di Nicola

N.P.D. 3° Blocco 2° piano

0871-3554007

marta.dinicola@unich.it

http://www.biostatistica.unich.it

LE MISURE DI VARIABILITA’

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.*

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

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

LE MISURE DI VARIABILITÀ

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

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.

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

Limiti del campo di variazione:

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

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

1ª idea → $\boxed{\sum_{i=1}^{n} (x_i - \bar{x})}$

2ª idea → $\boxed{\sum_{i=1}^{n} |x_i - \bar{x}|}$

3ª idea → $\boxed{\sum_{i=1}^{n} (x_i - \bar{x})^2}$

LA DEVIANZA

DEFINIZIONE: La somma dei quadrati degli scarti dalla media aritmetica

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

Esempio 9 Valori del tasso glicemico in 10 soggetti

X_i (glicemia mg/100cc)	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
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).

Mesi sopravvivenza (x_i)	Frequenze		
6,8	2	$(6.8-8.2)^2*2$	
7,3	8		
8,5	2		
9,2	4		
10,1	3		
Totale	**19**		

LA VARIANZA

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

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

LA DEVIAZIONE STANDARD

DEFINIZIONE: La radice quadrata della varianza

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

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, $\bar{x}$:

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

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

$(81-81.4)^2 = 0.16$ $(78-81.4)^2 = 11.56$

$(79-81.4)^2 = 5.76$ $(80-81.4)^2 = 1.96$

$(82-81.4)^2 = 0.36$ $(87-81.4)^2 = 31.36$

$(83-81.4)^2 = 2.56$ $(82-81.4)^2 = 0.36$

$(80-81.4)^2 = 1.96$ $(82-81.4)^2 = 0.36$

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

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

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

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

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

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

SCARTO INTERQUARTILE (IQR)

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

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

IL COEFFICIENTE DI VARIAZIONE

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

La variabilità guarda alle differenze tra le unità sperimentali. E' pero 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.

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× 100 = 7,4%

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

CV = 4,5/16,0 × 100 = 28,1 %

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

CV = 13,8/35,0 × 100 = 39,4 %

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

BOX-PLOT

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