



## Applicazione di metodi statistici allo screening ed alla diagnosi degli errori congeniti del metabolismo



M. Cassanello, U. Caruso.

Laboratorio per lo studio degli errori congeniti del metabolismo,  
Clinica Pediatrica, Istituto G. Gaslini, Genova

Il metodo più semplice ed immediato per rappresentare in modo sintetico un insieme di osservazioni relative ad un certa variabile e' l'utilizzo della:

### DISTRIBUZIONE DI FREQUENZA

### DISTRIBUZIONE DI FREQUENZA

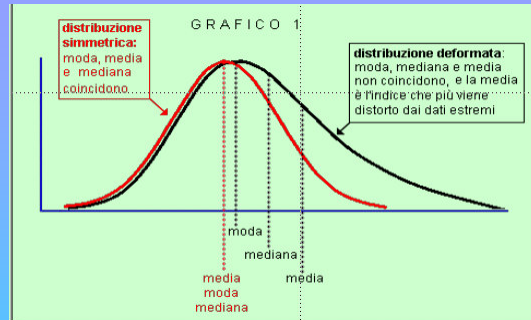


Insieme dei possibili valori (divisi in intervalli di classe) di una variabile in associazione alla frequenza rilevata all'interno della popolazione

- Frequenza semplice / assoluta
- Frequenza relativa / percentuale
- Frequenza relativa / percentuale cumulata

Phe $\mu\text{M}$	Fr assoluta	Fr %	Fr Cum %
57.4	9	0.06	90.3
57.5	12	0.08	90.38
57.6	11	0.08	90.45
57.7	17	0.12	90.57
57.8	9	0.06	90.63
57.9	16	0.11	90.74
58	9	0.06	90.81
58.1	14	0.1	90.9
58.2	7	0.05	90.95
58.3	7	0.05	91
58.4	15	0.1	91.1
58.5	19	0.13	91.23
58.6	19	0.13	91.36
58.7	10	0.07	91.43
58.8	4	0.03	91.46
58.9	9	0.06	91.52
59	12	0.08	91.61

riportando in un sistema di assi cartesiani i risultati in termini di concentrazione e frequenze associate, si ottiene una curva che può essere simmetrica o deformata da cui si individuano i parametri fondamentali:



**Misure di Posizione**

- Media
- Mediana
- Moda
- Percentili/quantili

**Misure di Variabilità**

- Range
- Deviazione standard
- Varianza
- Coeff. di variazione

**Forma distribuzione**

- Simmetria
- Curtosi
- Rappresentazioni grafiche

**Misure di posizione o indici di tendenza centrale**

- Utili ad individuare la tendenza centrale mostrata dai dati.
- Sono le misure più appropriate per sintetizzare l'insieme delle osservazioni.
- Non sono sufficienti a caratterizzare la distribuzione dei dati osservati.



## Misure di posizione o indici di tendenza centrale

### Caratteristiche:

#### MEDIA:

$$\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$$

- E' l'indice più utilizzato, il suo calcolo si basa su tutte le osservazioni e si presenta come un punto di equilibrio /baricentro della distribuzione.
- Il principale difetto: assume significato solo quando si riferisce a distribuzioni di frequenze con andamento "normale", gaussiano cioè quando la variabilità non e' elevata.
- In caso contrario e' fortemente influenzata dai valori estremi (valori anomali o outliers) e quindi non rappresenta più adeguatamente la "centralità".

## Misure di posizione o indici di tendenza centrale

### Caratteristiche:

**MEDIANA:** • E' il valore che occupa la posizione centrale in un insieme **ORDINATO** di dati.

- E' una misura **ROBUSTA** in quanto poco influenzata dalla presenza di dati anomali
- Come la media e' la misura di tendenza centrale nella statistica parametrica, la mediana e' l'indice di posizione impiegato in quasi tutti i test non parametrici.

## Misure di posizione NON centrale

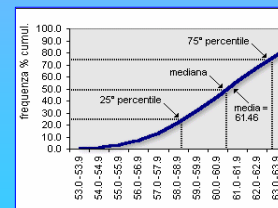
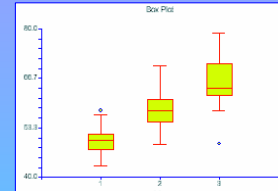
### Caratteristiche:

#### QUANTILI:

Sono indici di posizione non centrali che dividono una distr. di frequenza in un certo numero di parti uguali dopo aver ordinato i dati, es:

• **QUARTILI:** sono in numero di 3, il secondo quartile ( $Q_2$ ) coincide con la mediana, utilizzati ad esempio nelle rappresentazioni grafiche (Box Whiskers Plot).

• **PERCENTILI:** sono in numero di 99, si intende il livello di misura al di sotto del quale cade una determinata percentuale della distribuzione, sono significativi nel caso di curve non gaussiane.



## Misure di dispersione o indici di variabilità

• Quando si parla di variabilità in una distribuzione, generalmente ci si riferisce alla variazione dei dati attorno ad una misura di tendenza centrale.

• Descrivono la dispersione delle misure; insieme agli indici di posizione permettono di caratterizzare la popolazione osservata.

• Esistono numerosi indici utilizzabili come misure di variabilità. Quelli di più comune impiego sono:

- \_ l'intervallo di variazione (range)
- \_ il coefficiente di variazione (C.V.)
- \_ la deviazione standard (DS)

## Misure di dispersione o indici di variabilità

### Caratteristiche:

#### **RANGE:**

Indice piuttosto grezzo che misura la dispersione totale dei dati tra il massimo e il minimo delle osservazioni, quindi non caratterizza completamente l'intera distribuzione.

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

#### **C.V. %:**

Indice di variabilità relativa e misura la dispersione % rispetto alla media, e' un numero puro il che rende possibile il confronto tra due variabili su scale differenti.

$$CV = \frac{S}{\bar{x}} \times 100$$

## Misure di dispersione o indici di variabilità

### Caratteristiche:

#### **DEV.ST.:**

Definita come scarto quadratico medio e' la radice quadrata della varianza, consente di misurare la dispersione delle singole osservazioni intorno alla media.

$$DS = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

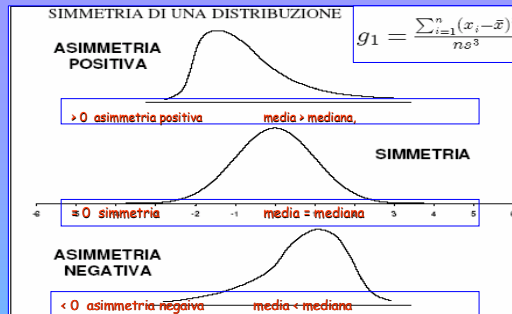
Si rivela molto utile per quantificare l'intervallo entro il quale si distribuiscono le varie misure.

E' un ottimo indice di variazione dei dati quando essi sono distribuiti normalmente e rappresenta probabilmente l'indice di variazione usato più comunemente

## Indici di forma

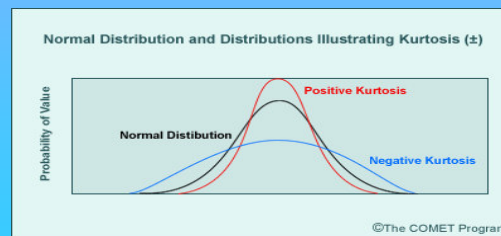
### Simmetria (*skewness*):

Indica l'assenza o presenza di specularità di una distribuzione rispetto al suo asse di simmetria e fornisce indicazioni sulla normalità della curva.



### Curtosi:

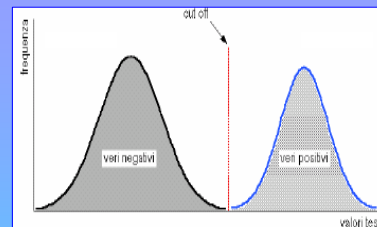
Fa riferimento al maggiore o minore accentramento delle misure in prossimità del suo massimo,



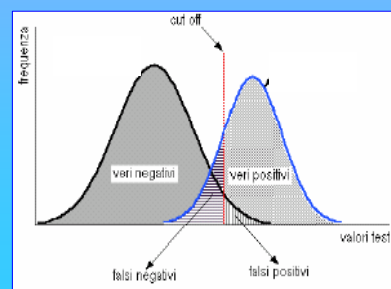
## Test quantitativi

I test di tipo quantitativo rappresentano lo strumento base nelle operazioni di *screening*, eseguite cioè su popolazioni presuntivamente sane allo scopo di identificare la presenza di malattie subcliniche.

Nella condizione ideale le popolazioni pos/neg dovrebbero essere separate individuando facilmente il valore soglia



Nella realtà esiste una zona di sovrapposizione fra le distribuzioni dei valori dei soggetti risultati negativi o positivi.



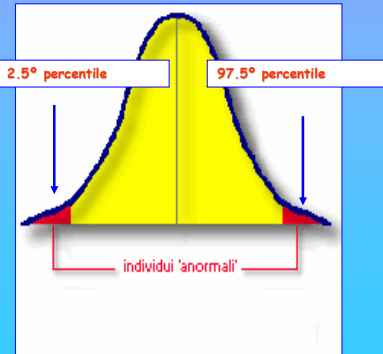
## Intervallo di riferimento

In generale, secondo le linee guida, l'intervallo di riferimento accettato e' quello compreso fra il :

**2.5° percentile**  
(Limite inferiore)

**97.5° percentile**  
(Limite superiore)

Se tale distribuzione segue una curva Gaussiana allora il limite della normalità corrisponde al 95 % delle osservazioni ed e' calcolato con media e dev.st. :



95 % dei dati rientra entro l'intervallo :

**MEDIA ± 2DS**

99.7 % dei dati rientra entro l'intervallo :

**MEDIA ± 3DS**

Generalmente la distribuzione delle misurazioni biologiche non e' Gaussiana!!

I limiti dell'intervallo di riferimento sono una funzione delle caratteristiche della distribuzione

I centili e quindi l'intervallo di riferimento possono essere stimati attraverso due metodi statistici:

- Metodo parametrico
- Metodo NON - parametrico

• Metodo parametrico

Teorie e calcoli basati su curva Gaussiana

\_ Trasformazione dei valori non simmetrici in una distr. Normale attraverso una  $F(x)$ : logaritmica, esponenziale etc.

\_ Calcolo dei limiti di rif. a seconda del livello decisionale adottato:

MEDIA  $\pm$  3DS

MEDIA  $\pm$  4DS

\_ Ritrasformazione dei limiti per ottenere l'intervallo finale.

• Metodo NON - parametrico

Non c'e' alcuna assunzione riguardo la  $f(x)$  matematica che regola la distribuzione dei dati misurati.

\_ E' il metodo di scelta raccomandato dal **NCCLS** (National Committee for Clinical Laboratory Standards) e dal **IFCC** (International Federation of Clinical Chemistry).

\_ Calcolo dei centili dopo aver individuato la distribuzione che meglio si adatta ai dati

## Livello decisionale e programmi di *screening* neonatale

- Il livello decisionale o *cut - off* nei programmi di *screening* neonatale deve essere periodicamente definito da ogni singolo laboratorio .
- Si basa sull'analisi statistica di almeno 5000-10000 campioni
- Generalmente i valori (IRT, AA, AC, TSH) non seguono propriamente una distribuzione simmetrica per cui il *cut - off* e' definito in base a procedure di analisi statistica non parametrica (centili).
- I programmi di *screening* neonatale dovrebbero caratterizzarsi per un indice di richiamo non superiore all' 1% e la scelta del *cut - off* e' di solito pari al 1° e al 99° centile.

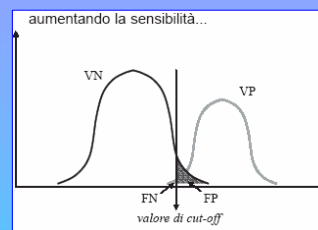
## La scelta del centile al quale fissare il *cut - off* e' funzione:

- dei livelli di efficacia del programma
- un buon compromesso fra alta sensibilità e specificità
- costi economici e sociali

### **SENSIBILITA'**

capacita' del test di individuare in una popolazione i soggetti malati

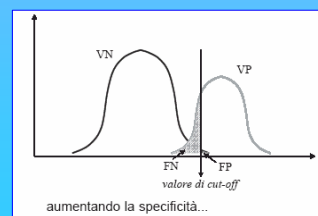
Test molto sensibile → Pochi falsi negativi



### **SPECIFICITA'**

capacita' del test di individuare in una popolazione i soggetti sani

Test molto specifico → Pochi falsi positivi



## Verifica della normalità e *goodness of fit*

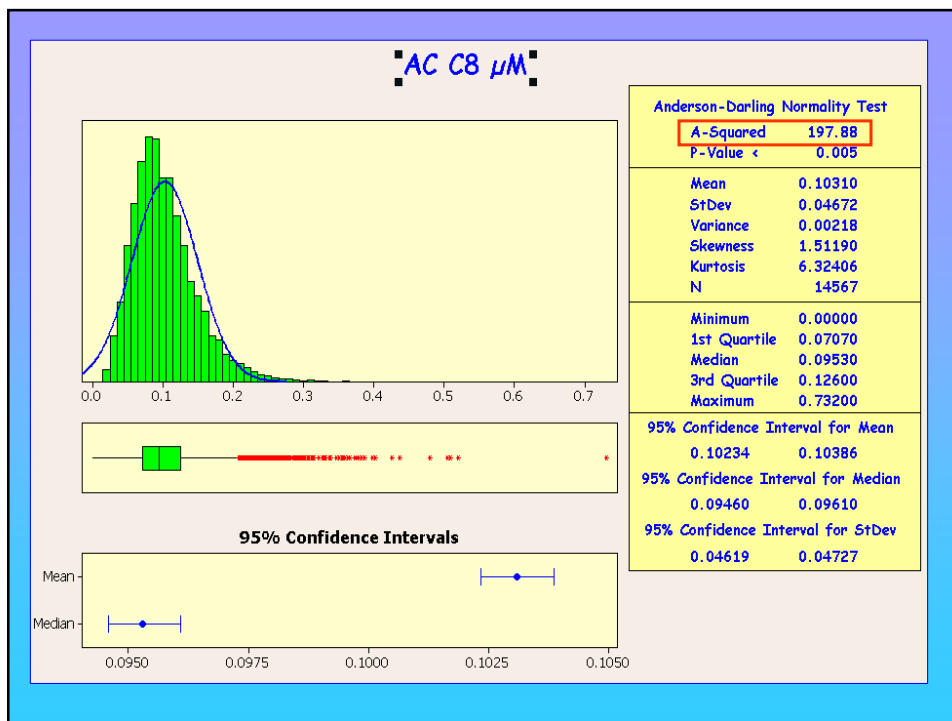
- Calcolo degli indici di posizione e dispersione
- Calcolo dei coefficienti di Skewness e curtosi
- Impiego del mezzo grafico:

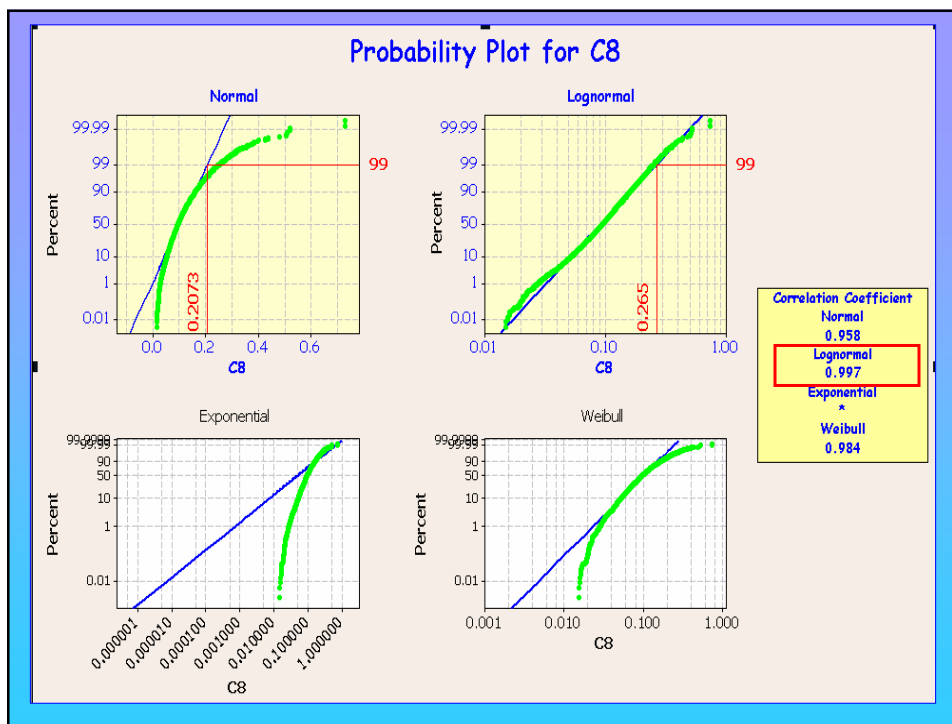
**ISTOGRAMMA , BOX AND WHISKERS PLOT, PROBABILITY PLOT**

- Test Non Parametrici: per verificare la conformità dei dati ad un modello teorico (*goodness of fit*)

**Kolmogorov - Smirnov    Anderson - Darling**

permettono di individuare la funzione che si adatta meglio ai dati per non incorrere in valutazioni errate nella scelta del cut - off affidabile .





## Bibliografia

- Natiol Committee for Clinical Laboratory Standards, How to define and determine reference intervals in the clinical laboratory: approved guideline, NCCLS document C28-A . ISBN 1-56238-269-1.
- Solberg HE. Approved recommendations 1987 on the theory of reference values: Part 5. Statistical treatment of collected reference values: determination of reference limits. J Clin Chem Clin Biochem 1987;25:645-56. Clin Chim Acta 1987;170:513-32.
- Solberg HE. Establishment and use of reference values. In: Burtis CA, Ashwood ER, editors. Tietz textbook of clinical chemistry. 2nd edn. Philadelphia: Saunders; 1994.
- Sasse E A. Objective evaluation of data in screening for disease. Clin Chim Acta 315 (2002) 17-30.
- Poulsen O.M. Calculation and application of coverage intervals for biological reference values. Clin Chim Acta 1987;165,111-118.
- Bourke G.J. Interpretazione ed uso delle statistiche mediche.
- Matthews D.E. Farewell V.T. Using and understanding medical statistics. 3rd revised edition.

Grazie per l'attenzione !!