

# La distribuzione normale - Parte II: Un esempio pratico

Michele Nichelatti, Maurizio Nordio, Umberto Maggiore, Maurizio Postorino, Aurelio Limido, Anteo Di Napoli

a nome del Comitato Scientifico SIN-RIDT

## THE NORMAL DISTRIBUTION - PART II: A PRACTICAL EXAMPLE

**Abstract.** The creatinine measure obtained in 300 healthy males is presented as an example of normal distribution, together with the concept of standardized Gaussian curve.

**Key words:** Normal distribution, Gaussian curve, Standardized Gaussian curve

**Conflict of interest:** None.

**Financial support:** None.

Accettato: 18 Maggio 2014

## La creatinina

Se misuriamo la creatinina emetica in un numero sufficientemente grande di soggetti, ci aspettiamo di trovare dei valori raccolti entro un certo intervallo. Come esempio, in un campione reale di 300 maschi adulti sani, sono state osservate le seguenti quantità:

- Media:  $\bar{x} = 1.11$  mg/dL
- Deviazione *standard*:  $s = 0.10$  mg/dL
- Mediana: 1.10 mg/dL
- Moda: 1.05 mg/dL
- Minimo: 0.85 mg/dL
- Massimo: 1.40 mg/dL

L'istogramma fornito in Figura 1 rappresenta la frequenza dei valori osservati.

La distribuzione dei dati relativi alla creatinina è normale (ovvero "gaussiana": lo abbiamo visto nello scorso numero), come si può verificare con uno qualsiasi dei vari *test* statistici disponibili allo scopo, eppure c'è qualcosa di diverso rispetto alla distribuzione normale vista nel numero scorso: lì la media della distribuzione era fissata a 0 e la deviazione *standard* aveva sempre un valore pari a 1; qui, invece, la media della creatinina è 1.11 e la deviazione *standard* è 0.10.

Per capire cosa stia succedendo ai nostri dati, e come questi possano essere espressi, esprimiamo lo stesso istogramma non con i dati della frequenza, ma con quelli della densità di probabilità.

Esprimere il nostro istogramma come densità (Fig. 2) anziché come frequenza significa, in buona sostanza, sostituire ai valori della frequenza quei valori che imporrebbero alla curva "a campana" ideale, che si potrebbe tracciare dai dati, di avere un'area pari a 1, esattamente:

## La densità di probabilità

Come facciamo allora a dire che la curva che otteniamo da questi dati è la stessa curva che abbiamo visto la volta scorsa? La risposta è abbastanza semplice: la curva che vediamo ora non è la stessa curva che abbiamo visto nel numero scorso, tuttavia possiamo ottenerla, con dei passaggi molto semplici. Infatti, semplificando molto il ragionamento e tralasciando parecchi aspetti matematici che al momento non ci interessano, la curva normale della densità di probabilità per la distribuzione della creatinina (cioè la curva tracciata in rosso in Fig. 2), così come ogni curva normale che si può ottenere da altri dati, si scrive partendo dall'equazione generale:

$$y = \frac{1}{s\sqrt{2\pi}} \exp\left(-\frac{(x-\bar{x})^2}{2s^2}\right), \quad (1)$$

che si trasforma in:

$$y = \frac{1}{0.1\sqrt{2\pi}} \exp\left(-\frac{(x-1.11)^2}{0.02}\right), \quad (2)$$

visto che per i nostri dati, come già visto, risultano una media

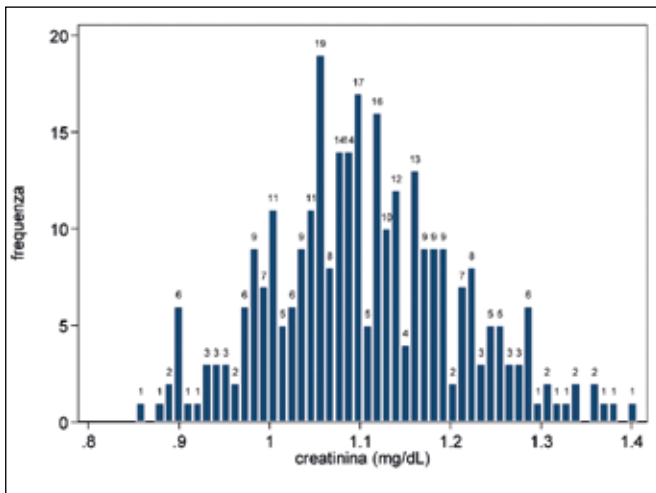


Fig. 1 - Distribuzione di frequenza della misura della creatinina in 300 maschi adulti sani.

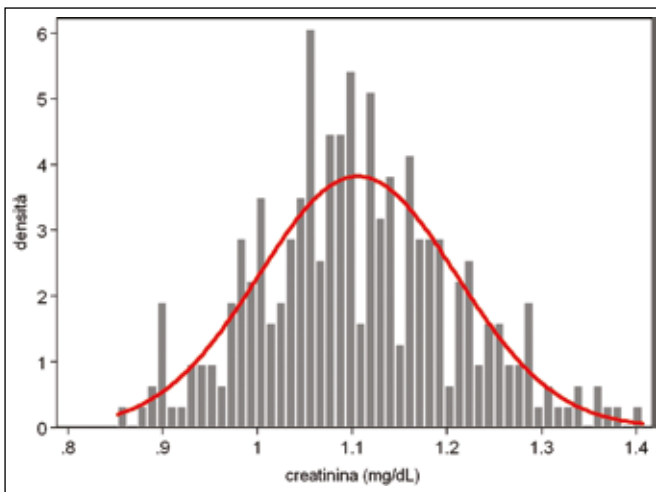


Fig. 2 - Istogramma dei valori della creatinina, avente per ordinata la densità (e non la frequenza) dei valori. La curva tracciata in rosso è la distribuzione dei valori, ottenuta in modo da avere un'area pari a 1; tale distribuzione è normale.

$\bar{x} = 1.11$  e una deviazione *standard* incidentalmente, osserviamo che la curva normale che otteniamo dei nostri dati ha un massimo nel punto di coordinate  $x = 1.11, y = 3,989$ . Come vediamo, allora, l'equazione definita nello scorso numero, cioè:

$$y = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad (3)$$

non è altro che l'equazione (2), in cui si impone che la media diventi  $\bar{x} = 0$  e che la deviazione *standard* diventi  $s = 1$ . Questa curva di equazione (2) è detta *curva gaussiana standardizzata* (o anche *normalizzata*, in alcuni testi).

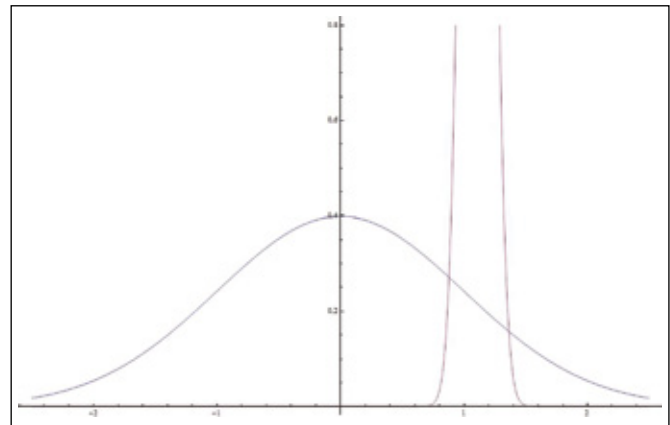


Fig. 3 - Grafico di confronto tra la curva gaussiana data dall'equazione (3) in blu e quella data dall'equazione (2) in rosso. L'area sottesa dalle curve è uguale a 1 in entrambi i casi.

In Figura 3 viene mostrato il confronto tra le curve dall'equazione (3), cioè la curva standardizzata e quella dell'equazione (2), ovvero la curva dei dati. Entrambe le curve sottendono un'area pari a 1, ovvero, usando il linguaggio statistico appropriato,

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = 1, \quad (4)$$

$$\int_{-\infty}^{+\infty} \frac{1}{0.1\sqrt{2\pi}} \exp\left(-\frac{(x - 1.11)^2}{0.02}\right) dx = 1. \quad (5)$$

In pratica, quindi, la standardizzazione della nostra curva gaussiana della creatinina consiste nello "schiacciare" la curva rossa di Figura 3 fino ad abbassarne il vertice e allargarne la base e poi nel traslarla a sinistra, in modo che il punto di massimo abbia ascissa uguale a zero.

Ovviamente le operazioni necessarie per la normalizzazione cambiano decisamente in funzione della distribuzione dei dati. Per la creatinina abbiamo trovato dati molto concentrati attorno alla media (infatti la deviazione *standard* era piccola rispetto al valore della media), con un coefficiente di variazione molto piccolo, pari a:

$$CV = \frac{s}{\bar{x}} = \frac{0.1}{1.11} = 0.09;$$

se invece avessimo ottenuto una distribuzione dei nostri dati con una deviazione *standard* molto larga e un valore massimo della distribuzione relativamente piccolo, avremmo dovuto fare l'operazione opposta, cioè "tirare" verso l'alto la curva, in modo da restringerne la base.

In pratica queste operazioni di aggiustamento della curva passano attraverso una trasformazione della variabile della

curva originale, che diventa la nuova variabile della curva standardizzata, così da avere, in generale:

$$\frac{x - \mu}{\sigma} \rightarrow X.$$

### Commento

Quanto visto in questo numero della rivista è un esempio pratico, in cui si sono saltati alcuni passaggi del ragionamento per motivi di spazio e per non invocare dei formalismi matematici troppo complicati. Abbiamo però visto che, quando i dati sono sufficientemente numerosi e quando sono distribuiti in modo più o meno omogeneo attorno alla media, è possibile ricavarne la curva a campana che li descrive con una buona approssimazione, anche se, come nel nostro caso, l'istogramma originale che descrive la frequenza delle misure (Fig. 1) non ha un aspetto particolarmente "pulito" e poteva ricordare una curva a campana solo in modo approssimativo.

Comprendere questo passaggio dall'istogramma dei dati (di un campione o di un'intera popolazione) alla curva a campana che li descrive è un passaggio cruciale nella comprensione della distribuzione normale; la cosa interessante che vedremo proseguendo il nostro discorso sarà che una distribuzione normale può generarsi anche da distribuzioni che nulla hanno a che vedere con la normalità.

### Riassunto

La creatinina misurata in 300 maschi sani è utilizzata per introdurre il concetto di distribuzione normale, assieme al concetto di curva Gaussiana standardizzata.

**Parole chiave:** Distribuzione normale, Curva Gaussiana, Curva Gaussiana standardizzata

**Dichiarazione di conflitto di interessi:** Gli Autori dichiarano di non avere conflitto di interesse.

**Contributi economici agli Autori:** Gli Autori dichiarano di non aver ricevuto sponsorizzazioni economiche per la preparazione dell'articolo.

#### *Indirizzo degli Autori:*

Dr. Michele Nichelatti  
 Servizio di Biostatistica  
 Dipartimento di Ematologia e Oncologia  
 Ospedale Niguarda Ca' Granda  
 Piazza Ospedale Maggiore 3  
 20162 Milano  
[michele.nichelatti@ospedaleniguarda.it](mailto:michele.nichelatti@ospedaleniguarda.it)