Distribuzione normale o di Gauss

E' la funzione continua più usata nelle applicazioni teoriche e pratiche. Essa è rappresentata da una curva simmetrica di forma campanulare, talvolta chiamata "*curva degli errori accidentali*". Mostra graficamente il numero degli scarti tra osservazioni reali e loro valore teorico, quando tali scarti sono casuali, cioè non dipendono dall'azione di fattori sistematici.

Le stime campionarie di un parametro della popolazione da cui proviene il campione sono descritte dalla funzione esponenziale:

$$\boxed{f(z) = \frac{1}{\sigma\sqrt{2\pi}} . e^{-\frac{z^2}{2\sigma^2}}}$$

Molte distribuzioni empiriche seguono la curva normale (caratteri antropometrici e biometrici).

La curva normale si usa anche per distribuzioni non normali, ma che possono approssimarsi a tale forma con opportune trasformazioni di variabile: **Ad esempio, una distribuzione asimmetrica rispetto alla variabile x può diventare quasi normale quando invece della variabile x si considera una sua trasformata del tipo radq(x) oppure x^2 oppure log x o 1/x, eccetera.**

Ha grande importanza nella teoria statistica per le sue numerose proprietà matematiche:**In teoria dei campioni, si dimostra che, anche quando la variabile di base non ha una distribuzione normale, la media campionaria, attraverso la quale si cerca di stimare la media vera della variabile di base, segue approssimativamente una distribuzione normale.**

Con forma esplicita degli scarti rispetto alla media, la funzione che descrive la distribuzione normale è:

$$\boxed{f(x) = \frac{N}{\sigma\sqrt{2\pi}} . e^{-\frac{(x-M)^2}{2\sigma^2}}}$$

Introducendo una nuova variabile t chiamata "unità standard" o "scarto ridotto" oppure "scarto standardizzato" e definita dalla relazione: $\boxed{t = \frac{x - M}{\sigma}}$

l'equazione della "*curva normale standardizzata*" diventa:

$$\boxed{f(t) = \frac{N}{\sqrt{2\pi}} . e^{-\frac{t^2}{2}}}$$

La distribuzione della variabile normale standardizzata t si ottiene dalla distribuzione della variabile di base x con una trasformazione di quest'ultima che trasferisce l'*origine dei valori nel punto medio* della distribuzione di base ed assume come nuova *unità di misura la deviazione standard* della variabile di base.

La trasformata t è una variabile caratterizzata dal fatto di avere media uguale a zero e deviazione standard uguale all'unità.

La normale come approssimazione della binomiale

La funzione normale può essere trovata anche come forma limite della funzione binomiale per valori di n abbastanza grandi. In

$$\boxed{P(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}}$$

Alcune proprietà della curva normale

La somma di tutti i valori di P(x) compresi tra x = - inf. ed x = + inf. è pari all'unità e vale la relazione:

$$\boxed{\int_{-\infty}^{+\infty} P(x).dx = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}}.e^{-\frac{(x-M)^2}{2\sigma^2}}.dx =}$$

$$= \frac{1}{\sigma\sqrt{2\pi}}.\int_{-\infty}^{+\infty} e^{-\frac{(x-M)^2}{2\sigma^2}}.dx = 1$$

oppure, nella forma standard in cui t = (x-M)/σ e dx = σdt:

$$\boxed{\int_{-\infty}^{+\infty} P(t).dt = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}.e^{-\frac{t^2}{2}}.dt = \frac{1}{\sqrt{2\pi}}.\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}}.dt = 1}$$

La curva normale è simmetrica rispetto alla verticale passante per la media della distribuzione per cui

$$\boxed{\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{(x-M)^2}{2\sigma^2}}.dx = \frac{2}{\sigma\sqrt{2\pi}}\int_{-\infty}^{0} e^{-\frac{(x-M)^2}{2\sigma^2}}.dx = \frac{2}{\sigma\sqrt{2\pi}}\int_{0}^{+\infty} e^{-\frac{(x-M)^2}{2\sigma^2}}.dx = 1/2}$$

e per la variabile ridotta standardizzata

$$\boxed{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}}.dt = \frac{2}{\sqrt{2\pi}}\int_{-\infty}^{0} e^{-\frac{t^2}{2}}.dt = \frac{2}{\sqrt{2\pi}}\int_{0}^{+\infty} e^{-\frac{t^2}{2}}.dt = 1/2}$$

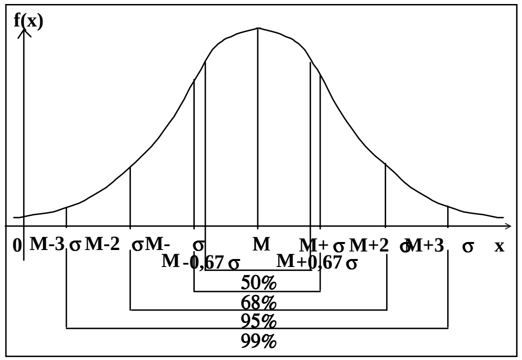
La distribuzione normale è completamente determinata dalla sua media M e dalla sua deviazione standard σ

- **Ha *un solo massimo* per x = M;**
- **Ha *due punti di flesso* per x = M ± σ;**
- **f(x) —› 0 per x —› ± inf. cioè *l'asse x è un asintoto*;**
- **Media, mediana e moda *coincidono*.**

Apposite tavole numeriche danno le ordinate della curva corrispondenti ad un dato valore dell'ascissa.

Esse danno anche la misura delle aree comprese sotto la curva a sinistra ed a destra di certe ordinate.

In generale le tavole sono preparate utilizzando i valori della variabile espressi in *unità standard*.



La metà dell'intera area è compresa nell'intervallo M± 0,676, il quale è anche chiamato "*errore probabile*".
Circa il 68% dell'area è nell'intervallo M± σ
Circa il 95% insiste sull'intervallo M± 2.σ
Circa il 99% circa ricade nell'intervallo M±3.σ.