# LA DISTRIBUZIONE NORMALE



## Premessa

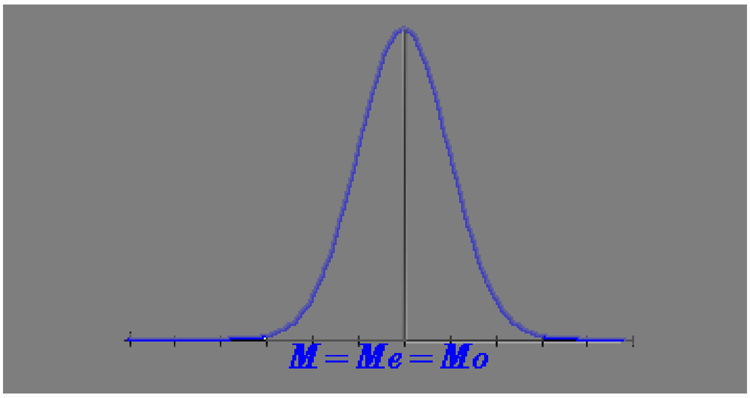
Un tipo molto importante di distribuzione di frequenza è quella normale. Questa distribuzione è particolarmente utile perché nella realtà molte distribuzioni empiriche si approssimano ad essa ed anche perché è di notevole importanza dal punto di vista teorico nel campo della statistica induttiva. Per i matematici è di grand'utilità il ragionare su distribuzioni basate su un numero di casi infinitamente" grande".
Invece di trattare distribuzioni empiriche d'aspetto frastagliato, come, ad esempio, quelle rappresentabili con un istogramma o con una spezzata, è possibile concepire curve lisce e continue, basate su un numero infinito di casi ed esprimere mediante equazioni matematiche relativamente semplici.
La distribuzione normale è una di questa curve. Essa è detta anche curva degli errori accidentali o di **Gauss.**

## Principali caratteristiche della curva normale

Un carattere quantitativo X, osservato in N unità del collettivo statistico, assume generalmente determinazioni $x_i$ (i=1…N) in tutto o in parte diverse. A volte si può ritenere, soprattutto se il carattere è distribuito in maniera omogenea, che, se non ci fosse l'influenza di fattori accidentali, sarebbe da attendersi un unico valore di x, che indichiamo con $\mu$. Le differenze $x_i-\mu$ sono assimilabili ad errori accidentali, cioè ad errori dovuti a cause numerose e non facilmente individuabili. I valori di tali errori, pur non essendo prevedibili, presentano alcune regolarità:

- Gli errori positivi sono altrettanto numerosi di quelli negativi (legge di simmetria),
- Gli errori piccoli, in valore assoluto, sono più frequenti di quelli grandi e quindi gli errori sono tanto più rari quanto più sono grandi.
- La rappresentazione di tali errori porta ad un diagramma simmetrico rispetto all'ordinata condotta per il punto di ascissa zero, col massimo su quest'ordinata e con andamento decrescente da entrambi i lati al crescere della differenza $|x_i-\mu|$.
- La curva normale è una curva continua e simmetrica, basata su un numero infinito di casi e quindi non potrà essere formata da distribuzioni di dati reali, ma solo

approssimata. Essa ha una forma campanulare, è unimodale ed ha la caratteristica che media, mediana e moda coincidono.



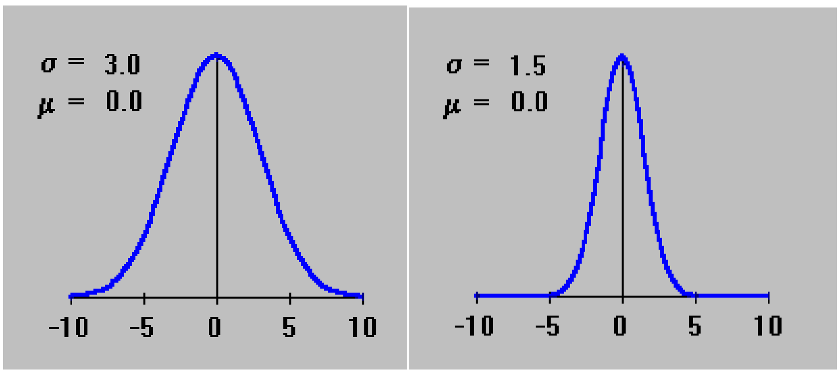

L'equazione della curva normale è la seguente:

$$Y = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

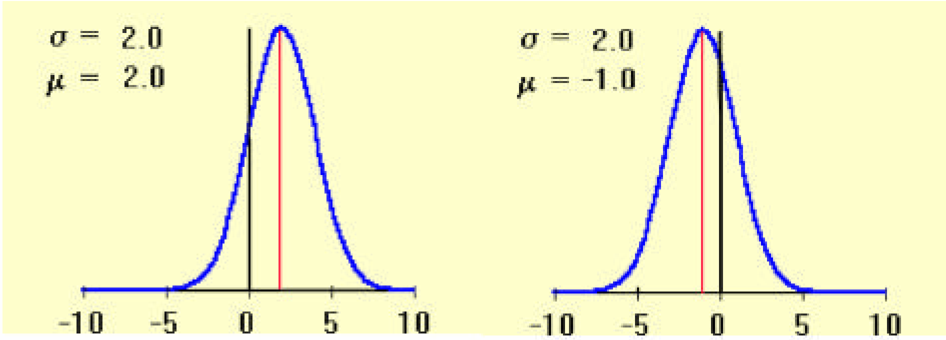
in cui y è l'altezza della curva per ogni dato valore di x.

Poiché π ed e sono costanti, per conoscere l'esatta forma della distribuzione è necessario conoscere la media aritmetica μ e lo scostamento quadratico medio σ. Si avranno pertanto diverse curve normali quanti sono le possibili combinazioni di medie e di scarti quadratici medi.

Per esempio, due curve con medie uguali e scarto diverso differiranno per quanto riguarda la forma più o meno appuntita: se minore è lo scarto quadratico medio, la curva è più appuntita



Se varia la media a parità di scarto quadratico medio la curva trasla rispetto all'asse delle x mantenendo inalterata la forma.

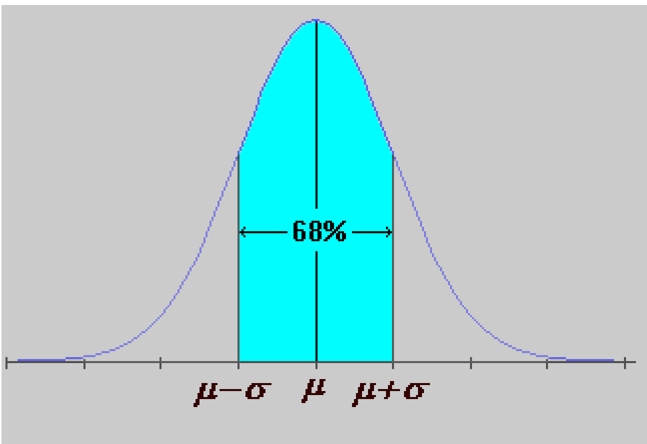



Si può dimostrare che:

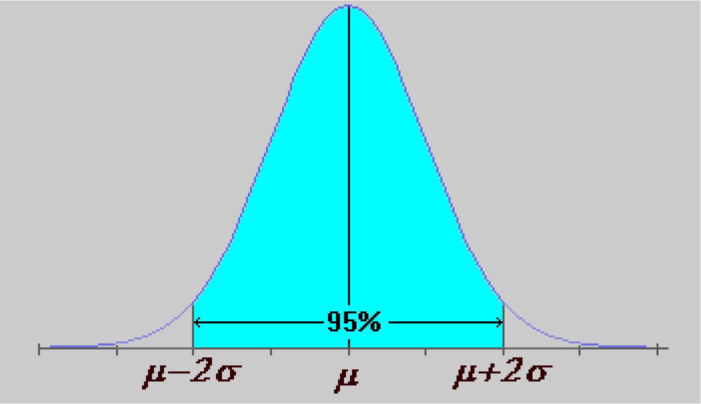
$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx = 1$$

cioè la funzione è di densità, in quanto l'area delimitata dalla curva e l'asse delle x è 1 cioè il 100%.

La variabile normale è un esempio classico di variabile continua che assume tutti i valori compresi tra $-\infty$ a $+\infty$. Pertanto, ha più senso determinare anziché l'ordinata, la percentuale dei casi compresi in un certo intervallo.
Un problema di questo tipo è facilitato da una delle proprietà della curva normale: qualunque siano la media e lo scostamento, l'area tra la media e un'ordinata posta ad una determinata distanza, definita in termini di scarto quadratico medio, è costante. Nella distribuzione normale si dimostra che i punti di flesso della curva hanno per ascissa i valori $\mu-\sigma$ e $\mu+\sigma$. Inoltre nell'intervallo $\mu-\sigma$, $\mu+\sigma$ la curva comprende il 68,27% dei casi osservati;



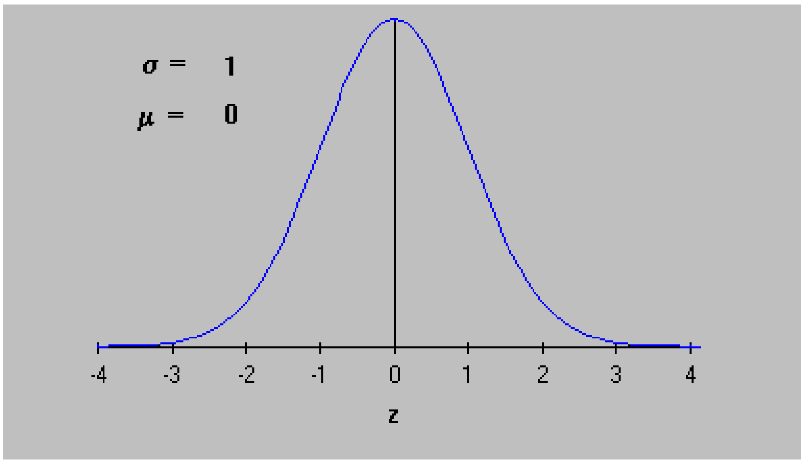
nell'intervallo tra μ-2σ, μ+2σ la curva comprende il 95,45% dei casi osservati e tra μ-3σ, μ+3σ il 99,73%. Oltre l'area racchiusa tra la curva e l'asse delle ascisse è quasi trascurabile nell'intervallo tra μ-2σ, μ+2σ la curva comprende il 95,45% dei casi osservati e tra μ-3σ, μ+3σ il 99,73%. Oltre l'area racchiusa tra la curva e l'asse delle ascisse è quasi trascurabile




Naturalmente è sempre possibile, con un semplice procedimento, determinare l'area sottesa alla porzione di curva delimitata da due qualsiasi ordinate, senza che le distanze dalla media siano multiple esatti dello scarto quadratico medio.

## La curva normale standardizzata

Essendo infinite le curve normali al variare della media e dello scarto quadratico medio è sorta la necessità di costruire una curva unica a cui fare riferimento e di cui sono state calcolate le aree, tale curva è chiamata normale standardizzata ed è come si



può dimostrare una distribuzione con media 0 e scarto 1, vale a dire con valore massimo all’origine degli assi e flessi nei punti –1, +1.

Per ottenere la normale standardizzata è necessario effettuare un’opportuno cambiamento di variabile del tipo:

$$Z = \frac{x - \mu}{\sigma}$$

Si avrà allora: $M(Z) = \sum_{i=1}^{N} \frac{\left(\frac{x-\mu}{\sigma}\right)}{N} = \sum_{i=1}^{N} \left(\frac{x-\mu}{\sigma}\right)\frac{1}{N} = 0$ per una nota proprietà della media aritmetica ( $\sum (x - \mu) = 0$

$$VAR(Z) = \frac{\sum_{i=1}^{N} \left(Z - \bar{Z}\right)^2}{N} = \sum_{i=1}^{N} \frac{\left(\frac{x-\mu}{\sigma}\right)^2}{N}$$

essendo $\bar{Z} = 0$ si avrà:

$$\sum_{i=1}^{N} \left(\frac{x-\mu}{\sigma}\right)^2 \frac{1}{N} = \frac{\sum_{i=1}^{N} (x-\mu)^2}{\sigma^2} \frac{1}{N} = \frac{\sigma^2}{\sigma^2} = 1$$

Essendo l’espressione evidenziata in rosso la varianza di x.

Graficamente la trasformazione della variabile originaria x in variabile standardizzata Z si ottiene passando dal sistema cartesiano ortogonale al sistema con asse Z con la curva al valore massimo nell’origine e flessi a –1, +1.
Fatta la trasformazione, si possono usare delle apposite tavole nelle quali sono riportate le porzioni di aree per tutte le ordinate. I valori della tabella indicano la proporzione dei casi che cadono nell’intervallo delimitato dalla media (cioè il punto di ascissa 0) e dell’ordinata Z. Con l’uso di questa tavola si può trasformare ciascuna curva normale in modo tale che sia possibile calcolare la percentuale dei casi sottostante ogni porzione della curva.

$$Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Poiché π ed e sono costanti, per conoscere l’esatta forma della distribuzione è necessario conoscere  la media aritmetica μ e lo scostamento quadratico medio σ

Si può dimostrare che:

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

cioè la funzione è di densità, in quanto l’area delimitata dalla curva e l’asse delle x è 1 cioè il 100%.