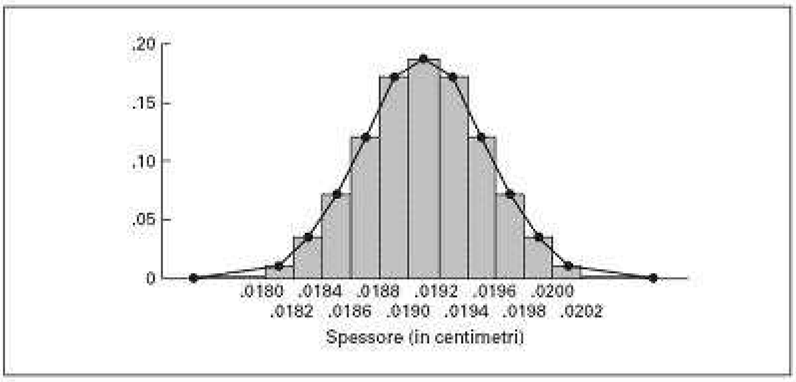
La distribuzione normale

Alla curva gaussiana `e legata l’idea di “curva per descrivere i dati”. Quando si effettuano tante misurazioni di una stessa grandezza con un certo stru-mento, si avranno risultati differenti dovuti all’inevitabile imprecisione sia dello strumento sia dell’operato della persona che utilizza lo strumento. Se si rappresentano le misure ottenute su un grafico, e se il numero di misurazioni è molto elevato, la curva che si ottiene è proprio la curva di



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

La curva ha un massimo attorno alla media dei valori misurati ed è più omeno stretta a seconda della dispersione dei valori attorno alla media. La dispersione si misura con la deviazione standard.

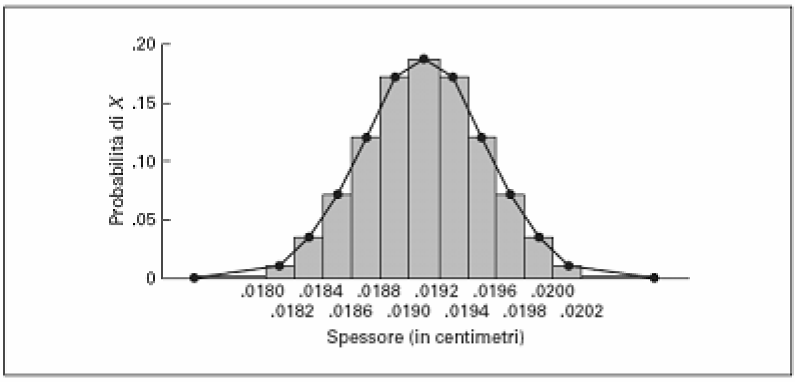
<u>Distribuzioni di probabilità continue</u>

Alcuni tipici fenomeni continui sono l’altezza, il peso, le variazioni giornaliere nei prezzi di chiusura di un’azione, il tempo che intercorre fra gli arrivi di aerei presso un aeroporto, il tempo necessario per servire un cliente in un negozio.

La distribuzione normale (o distribuzione Gaussiana) è la distribuzione continua più utilizzata in statistica

Molte variabili statistiche che osserviamo nella realtà hanno una distribuzione con caratteristiche simili a quelle della distribuzione normale.

Consideriamo ad esempio lo spessore misurato in centimetri di 10 000 rondelle di ottone prodotte da una grande società metallurgica. Il fenomeno aleatorio continuo di interesse, lo spessore delle rondelle, si distribuisce approssimativamente come una normale.



La distribuzione normale

Utilizzeremo il simbolo f(X) per denotare l'espressione matematica di una funzione di densità di probabilità. Nel caso della distribuzione normale la funzione di densità di probabilità normale è data dalla seguente espressione:

$$f(X) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(1/2)[(X-\mu)/\sigma]^2}$$

dove

e= costante matematica approssimata da 2.71828

π= costante matematica approssimata da 3.14159

μ = valore atteso della popolazione (media aritmetica)

σ= scarto quadratico medio della popolazione

X= valori assunti dalla variabile aleatoria, $-\infty<X<+\infty$

Per un insieme di dati con distribuzione normale:

• approssimativamente il 68.26% apparterrà all'intervallo ($\mu - \sigma$, $\mu + \sigma$)

• approssimativamente il 95.44% apparterrà all'intervallo ($\mu - 2\sigma$, $\mu + 2\sigma$)

• approssimativamente il 99.73% apparterrà all'intervallo ($\mu - 3\sigma$, $\mu + 3\sigma$)

Valutazione dell'ipotesi di normalità

Non tutti i fenomeni continui sono distribuiti normalmente e non tutti seguono una distribuzione che può essere approssimata adeguatamente con una normale. È quindi importante verificare la plausibilità dell'ipotesi di normalità, cioè di accertare se in effetti un insieme di dati può provenire da una distribuzione normale. Dal punto di vista pratico il problema è di valutare la bontà di adattamento del modello normale a un insieme di dati, problema che deve essere affrontato ancora prima di applicare altre metodologie di analisi.

Due sono gli approcci esplorativi di carattere descrittivo che possono essere adottati:

1. Il confronto fra le caratteristiche dei dati e le proprietà di un'eventuale distribuzione normale sottostante

2. La costruzione di un normal probability plot