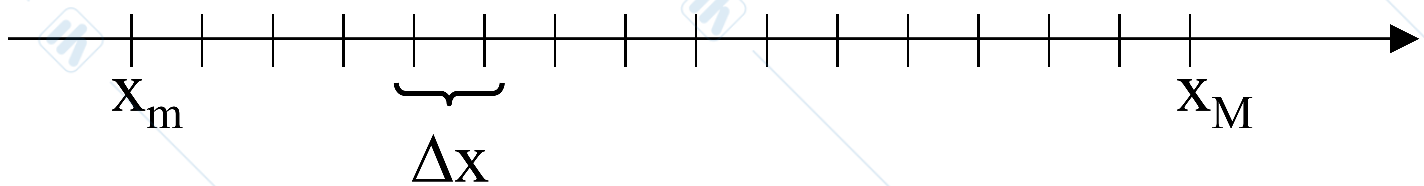
Istogrammi e confronto con la distribuzione normale

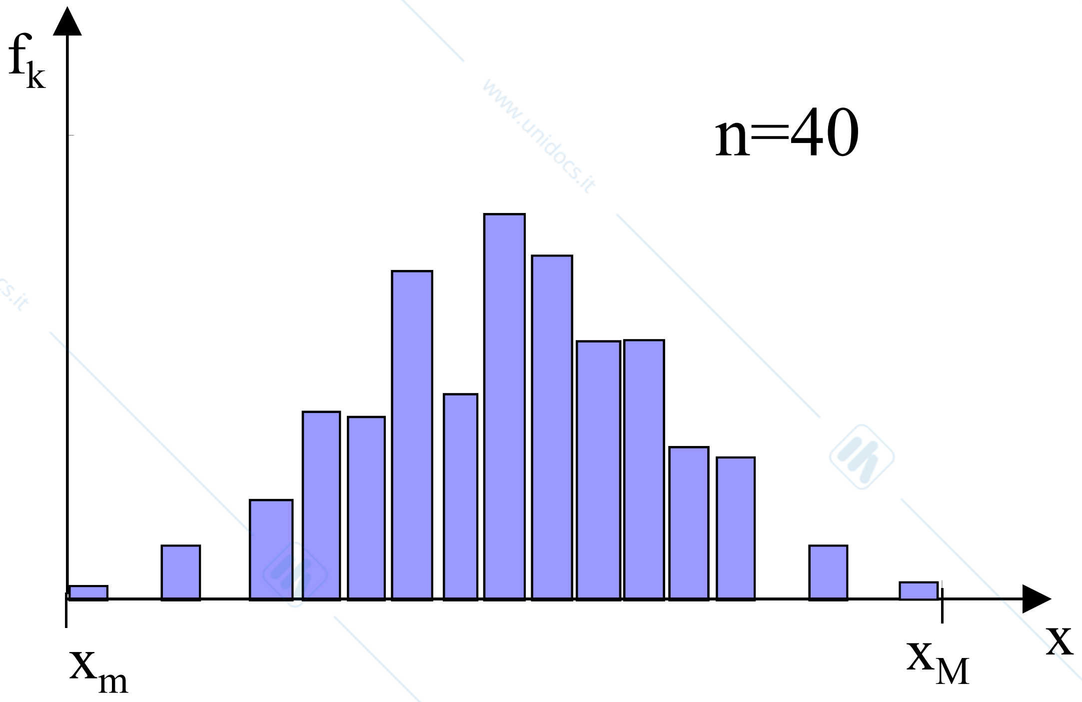
Supponiamo di effettuare per **n** volte la misurazione della stessa grandezza nelle stesse condizioni (es. la massa di un oggetto, la tensione di una pila, la lunghezza di un oggetto, ecc.): in generale, i risultati ottenuti non saranno tutti uguali tra loro a causa degli errori casuali.

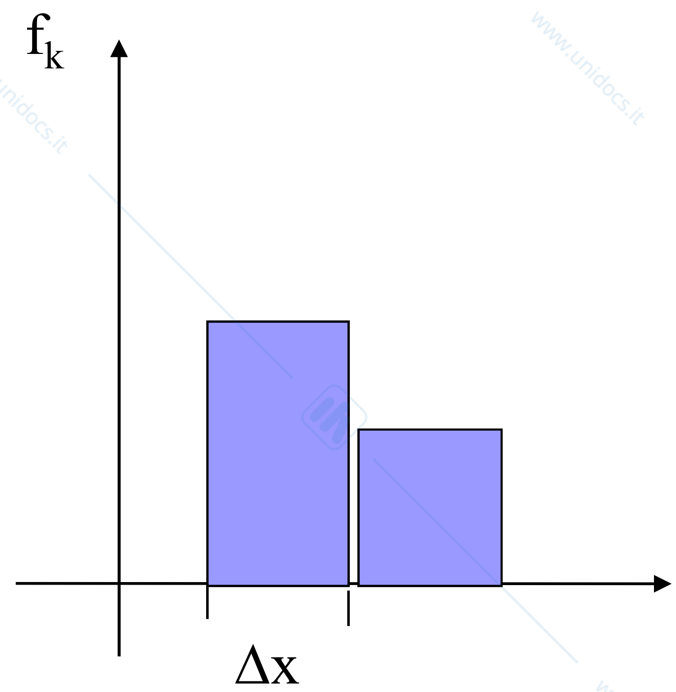
Consideriamo i valori minimi e massimi misurati ($\mathbf{x_m}$ e $\mathbf{x_M}$) e dividiamo l'intervallo compreso tra essi in *r* parti uguali di ampiezza

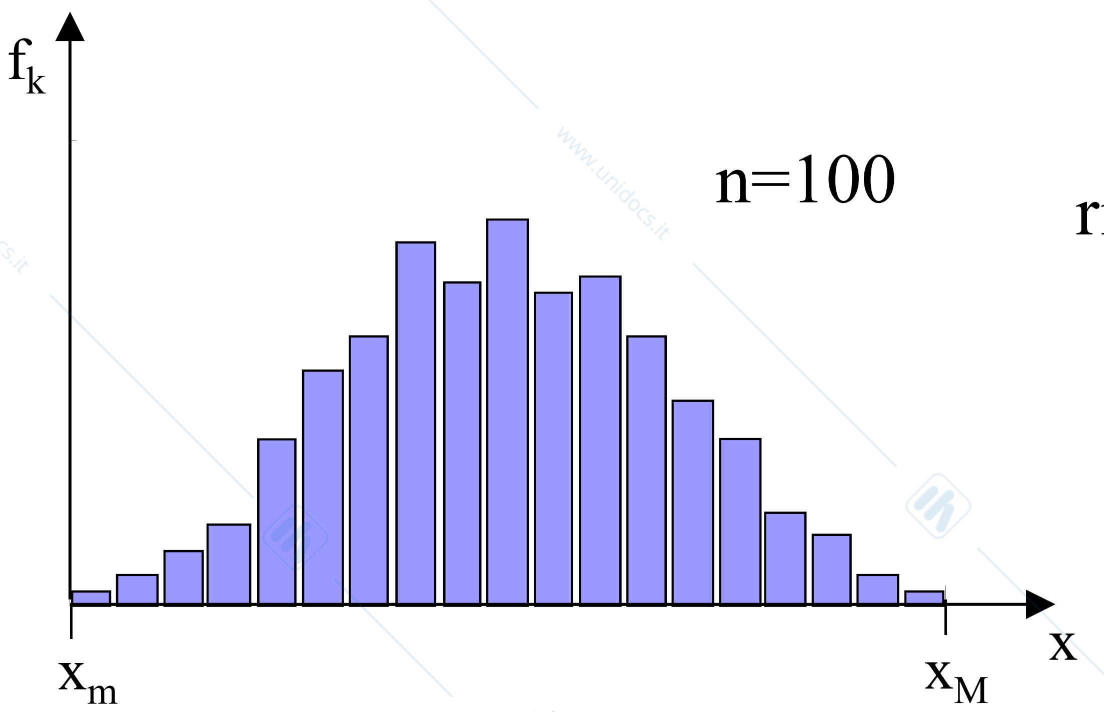
$$\Delta x = \frac{x_M - x_m}{r}$$

Consideriamo un sistema di assi cartesiani e riportiamo sulle ascisse i valori delle misure ottenute e in ordinata il numero di risultati della misura che cadono nel sotto-intervallo



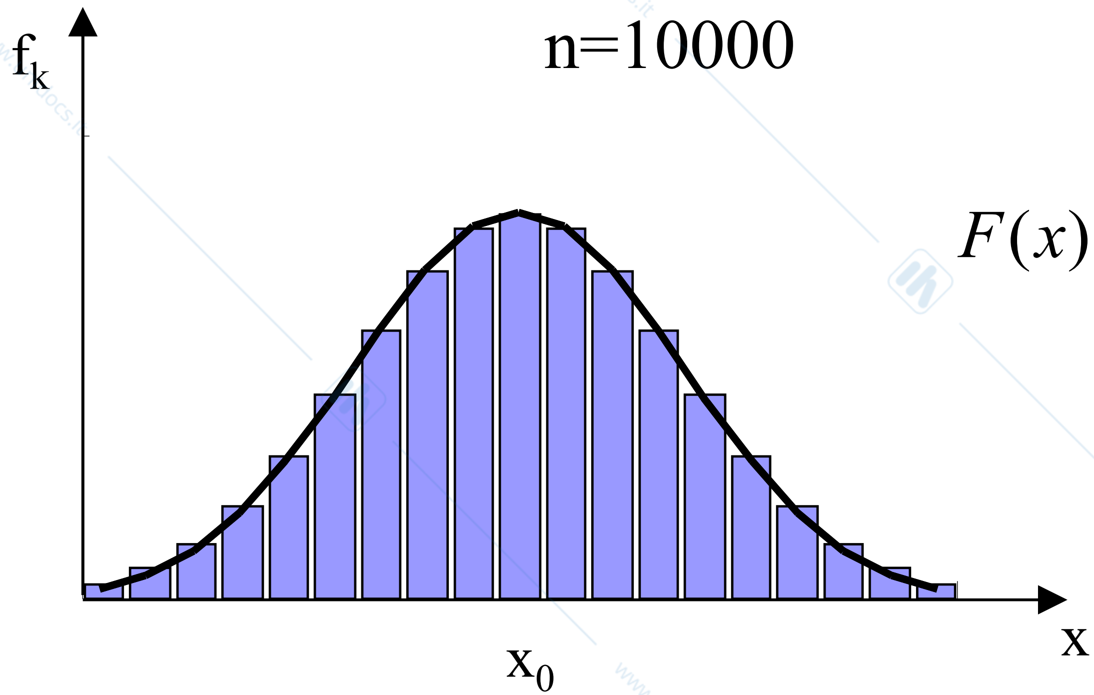






f_k è il n^{ro} di misure che rientrano nell'intervallo Δx

Funzione di distribuzione normale o funzione di Gauss



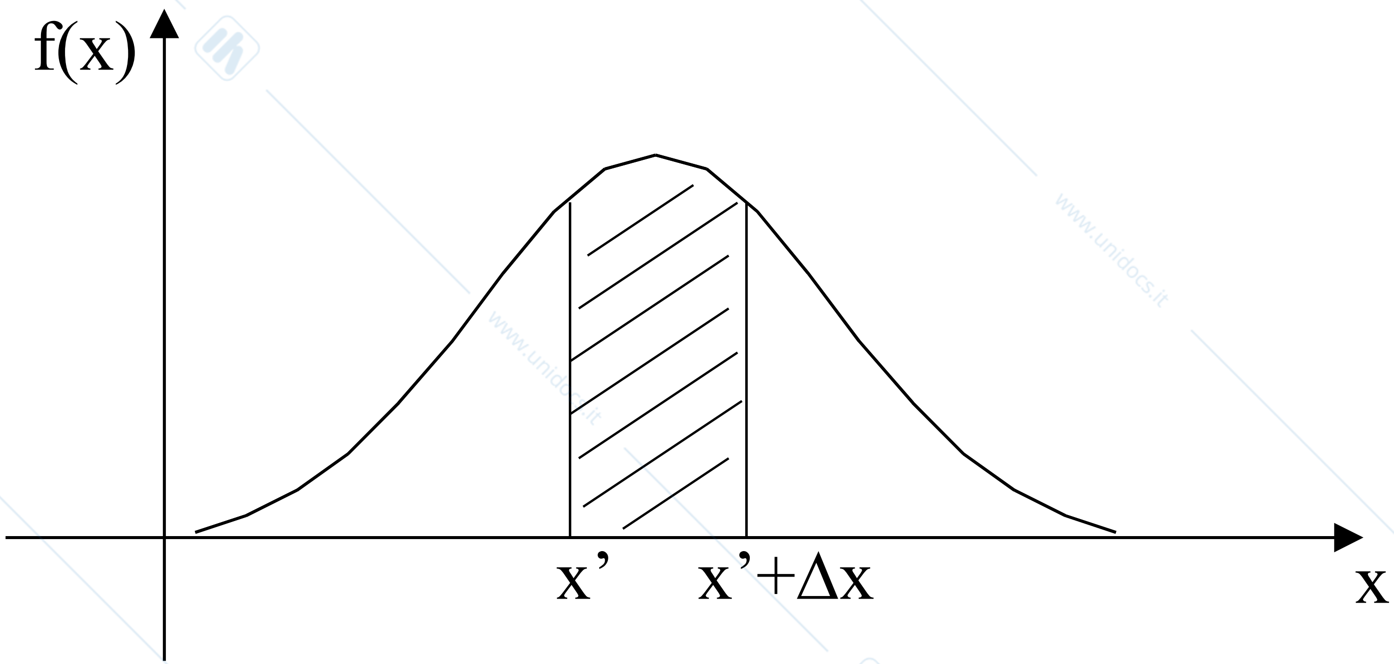
$$F(x) = \frac{n}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-x_0)^2}{2\sigma^2}\right]$$

Rappresenta , per un determinato valore di σ e di x_0 che variano di caso in caso, la distribuzione delle misure per una estesa classe di grandezze fisiche.

N.B. F(x)/n fornisce la probabilità di ottenere un dato valore x della grandezza in seguito ad una misura.

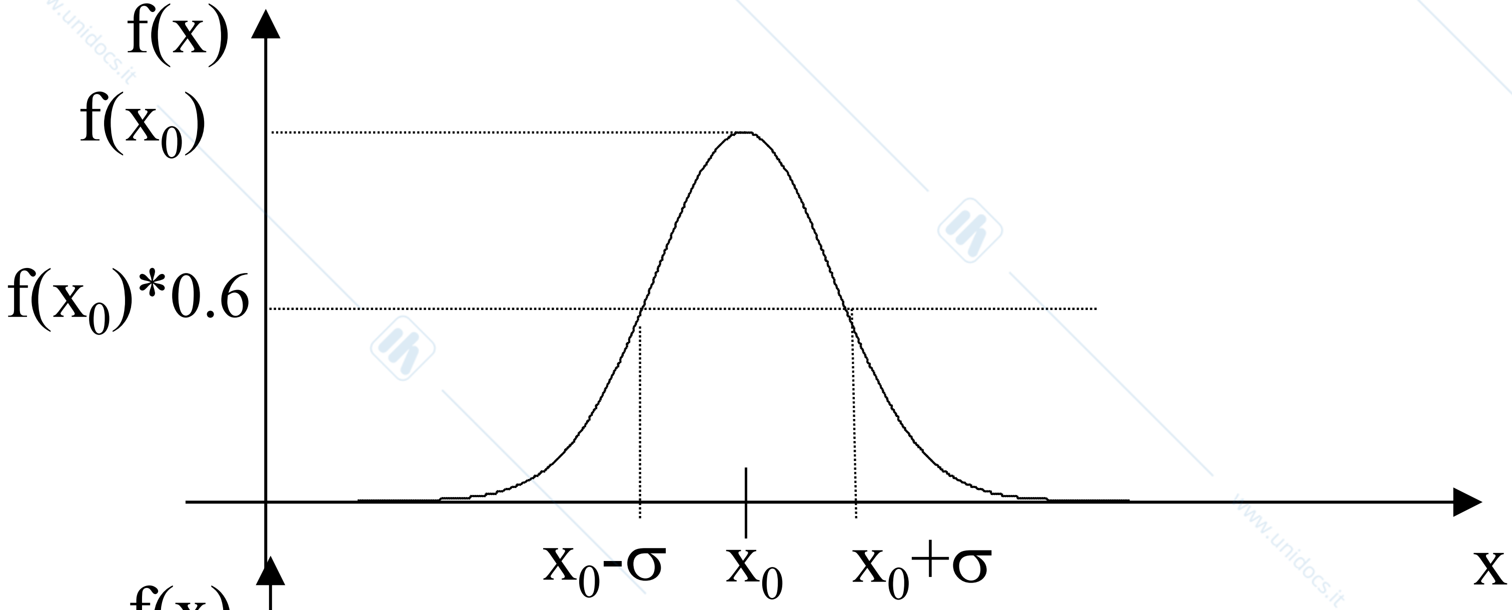
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-x_0)^2}{2\sigma^2}\right]$$ è una densità di probabilità

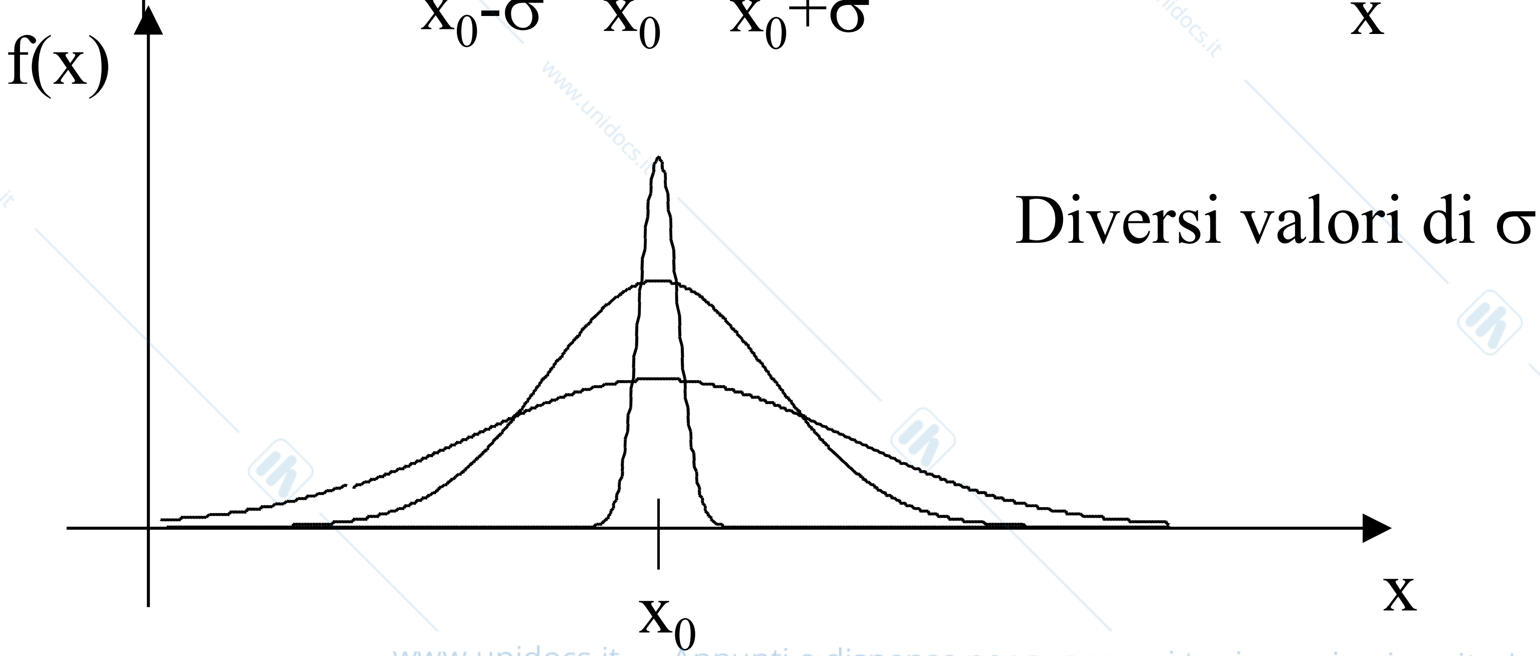
per cui $\int_{x'}^{x'+\Delta x} f(x)dx$ rappresenta la probabilità che la misura dia un valore della grandezza x compreso tra x' e x'+Δx



N.B. $\int_{-\infty}^{\infty} f(x)dx = 1$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-x_0)^2}{2\sigma^2}\right]$$





Il valore più probabile della misura

La misurazione di una grandezza non dà mai il suo valore 'vero'.

E' possibile dimostrare che il valore **più probabile** della grandezza è fornito dalla **media aritmetica** $<x>$ degli n valori $x_1, x_2, \ldots, x_n$ ottenuti nella sua misurazione ripetuta

$$< x > = \frac{x_1 + x_2 + \ldots + x_n}{n} = \sum_{i=1}^{n} \frac{x_i}{n}$$

Qualora il numero **n** delle misurazioni sia sufficientemente grande in modo che l'istogramma visto prima possa essere approssimato dalla funzione di Gauss, il valore medio $<x>$ coincide con il valore x_0, per cui la gaussiana assume il suo valore massimo.

Il valore più probabile della misura

In realtà, il ‘valore medio’ sarà vicino al valore centrale x_0, tanto più quanto maggiore è il numero n delle misure eseguite.

Il valor medio <x> rappresenta la miglior stima possibile del valore vero della grandezza.

La migliore stima per la larghezza σ della distribuzione di Gauss è data dalla **deviazione standard**, definita dalla relazione:

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \langle x \rangle)^2}$$

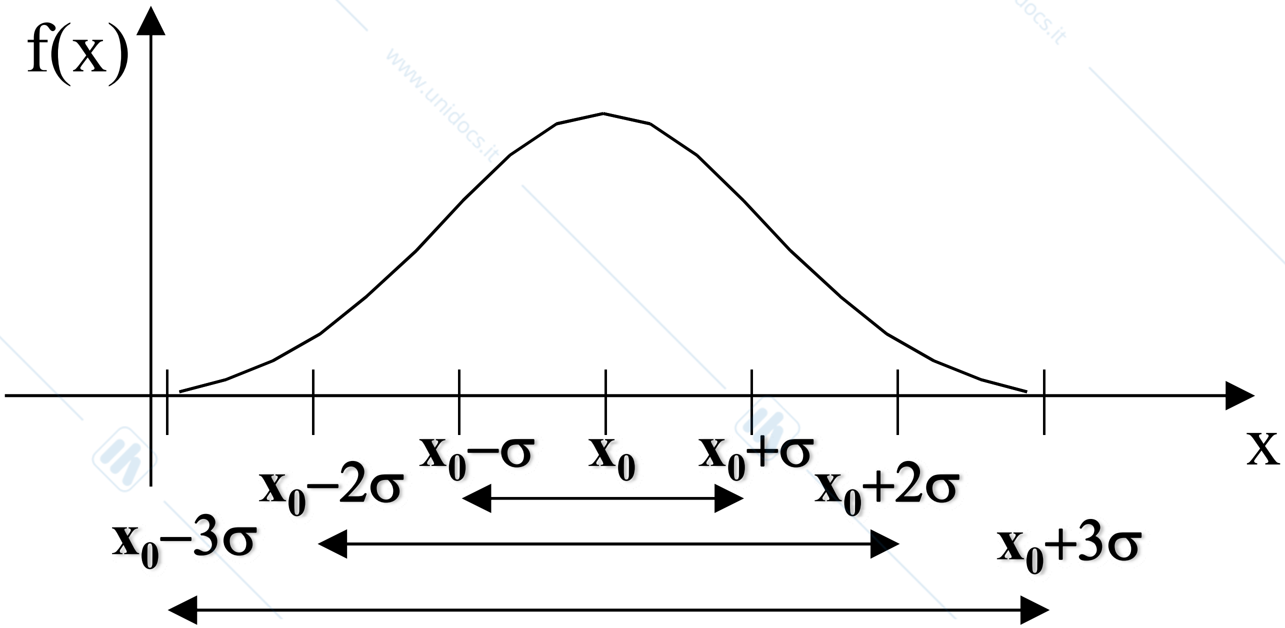
La deviazione standard $$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \langle x \rangle)^2}$$

Rappresenta l'incertezza in una misura e fornisce il cosiddetto **limite di confidenza** della misura. Si dimostra infatti che:

vi è il **68.7 % di probabilità** che il risultato di una misura differisca meno di **σ** dal valore vero

il **95.4 % di probabilità** che la misura cada entro **2σ** dal valore vero

il **99.7 % di probabilità** che la misura cada entro **3σ** dal valore vero



Si dimostra inoltre che l'incertezza nella stima del valor medio (che rappresenta la miglior stima e quindi la migliore approssimazione al valore vero della grandezza misurata), è data dalla relazione:

$$\sigma_{<x>} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - <x>)^2}$$

che definisce **la deviazione standard della media** (errore medio empirico della media aritmetica): con il **68.7 %** di probabilità, il valore medio <x> si discosta da quello vero per meno di $\boldsymbol{\sigma_{<x>}}$, con il **95.4 %** <x> si discosta per meno di $\boldsymbol{2\sigma_{<x>}}$, con il **99.7 %** <x> si discosta per meno di $\boldsymbol{3\sigma_{<x>}}$.

$$<x> - k \cdot \sigma_{<x>} \leq x \leq <x> + k \cdot \sigma_{<x>} \qquad \left| x - <x> \right| \leq k \cdot \sigma_{<x>}$$

Quindi, la probabilità che $\left|x - <x>\right| \le k \cdot \sigma_{<x>}$

è data dalla probabilità di Gauss

$$\int_{x_0 - k\sigma_{<x>}}^{x_0 + k\sigma_{<x>}} f(x)\,dx$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-\left(x - x_0\right)^2}{2\sigma^2}\right]$$

Riassumendo, si potrà indicare come valore di x

$$\boxed{x = \langle x \rangle \pm \Delta x}$$

dove

$$\langle x \rangle = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \Delta x = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - \langle x \rangle)^2}$$

N.B.: Nel caso in cui l'errore Δx calcolato come sopra sia inferiore all'errore strumentale, si usa quest'ultimo come errore massimo.

Metodo dei minimi quadrati

Uno dei più interessanti tipi di esperimento riguarda la misura di parecchi valori di **due** diverse variabili fisiche, per investigare la relazione matematica tra le due variabili.

Es. lasciamo cadere un corpo da una certa altezza. Tale corpo sarà soggetto all'accelerazione di gravità g. Nel caso in cui per il tempo t=0 esso abbia una velocità iniziale $v=v_0$, la sua velocità v dovrebbe essere una funzione lineare del tempo t,

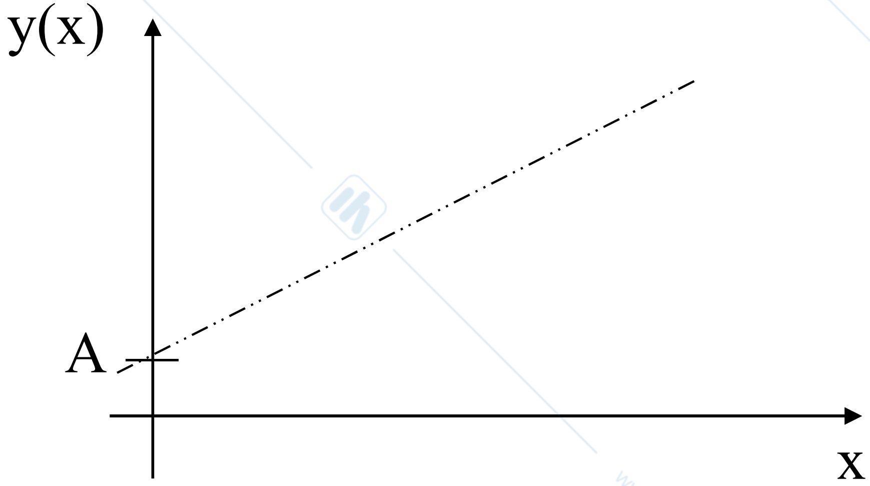
$$v(t) = v_0 + gt$$

tale relazione è **lineare** del tipo

$$y(x) = A + Bx$$

dove A e B sono costanti.

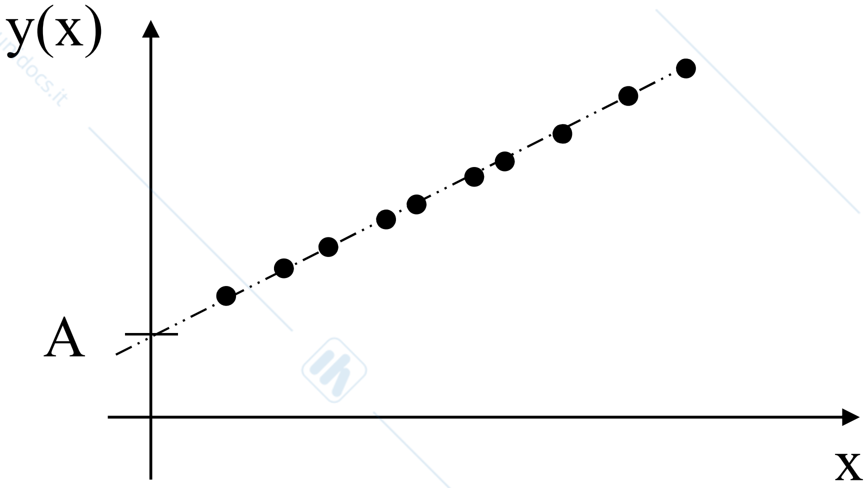
Se le due variabili sono in relazione come $y(x) = A + Bx$, allora un grafico di y in funzione di x dovrebbe essere una linea retta che ha pendenza (coefficiente angolare) B e interseca l'asse y in $y=A$.



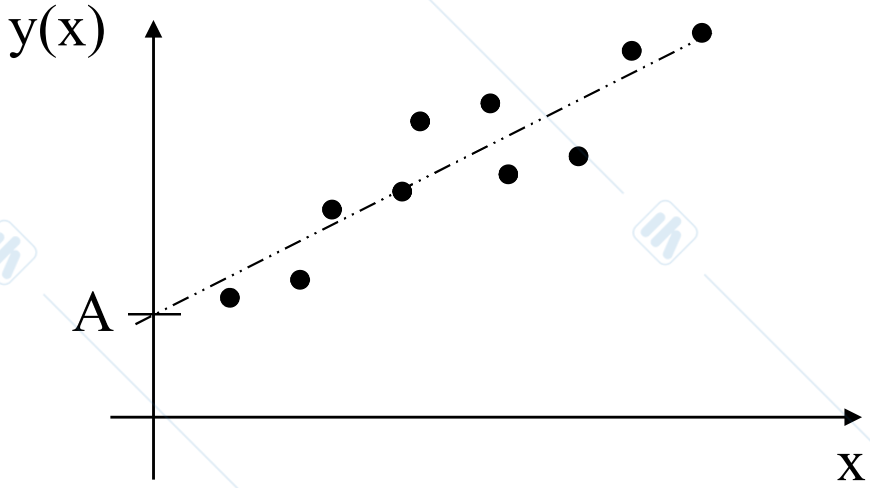
N.B. I punti $y_1, y_2, \ldots, y_n$ **non** sono N misure della stessa grandezza

Se si misurassero N diversi valori di $x_1, x_2, \ldots\ldots, x_n$ e i valori corrispondenti $y_1, y_2, \ldots, y_n$ e se le misure non fossero soggette ad incertezze, allora ciascuno dei punti (x_i, y_i) dovrebbero giacere esattamente sulla retta

In realtà, essendo presenti delle ‘incertezze’, i punti potranno risultare ‘sparpagliati’ intorno alla retta.



Caso ideale



Caso reale

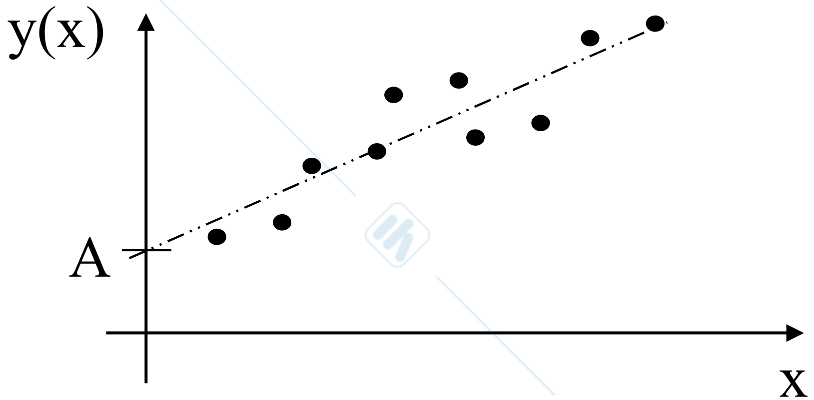
Se prendiamo ‘per garantito’ che **y** e **x** soddisfano una relazione lineare, ci si può porre il problema di trovare la miglior retta per interpolare un insieme di punti misurati (x_1, y_1), (x_2, y_2),….. (x_n, y_n), cioè trovare un **fit lineare (regressione lineare o curva dei minimi quadrati per una retta)**. Si dovranno trovare quindi le migliori stime dei coefficienti della retta A e B.

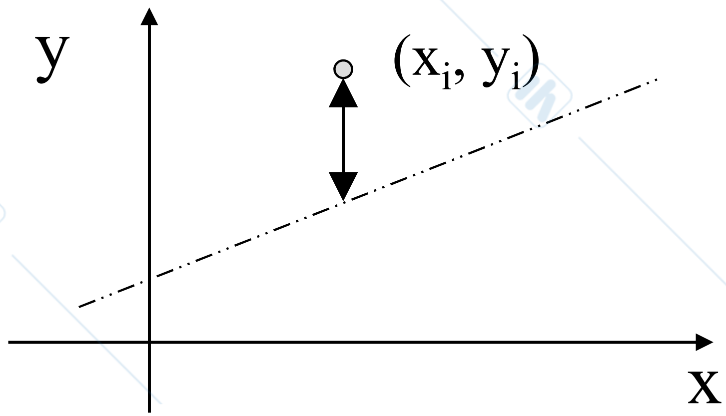
Il procedimento è il seguente

si eseguono **n** misure corrispondenti alle coppie $(x_1, y_1), (x_2, y_2), \ldots\ldots (x_n, y_n)$. Sapendo che la relazione è lineare si calcolano gli scarti $v_i = y_i - (A+Bx_i)$, si calcolano i quadrati e si sommano

$$\Phi = \sum_{i=1}^{n} \left(y_i - A - Bx_i\right)^2$$

Si cercano i valori di A e B per cui Φ sia la minima possibile. Questo equivale a rendere minimi i quadrati delle distanze dei punti (x_i, y_i) dalla retta, misurate nella direzione dell'asse y





Per far questo differenziamo Φ rispetto ad A e B e poniamo le derivate uguali a zero:

$$\begin{cases} \dfrac{\partial \Phi}{\partial A} = -2\sum_{i=1}^{n}\left(y_i - A - Bx_i\right) = 0 \\ \dfrac{\partial \Phi}{\partial B} = -2\sum_{i=1}^{n} x_i\left(y_i - A - Bx_i\right) = 0 \end{cases}$$

Queste due equazioni possono essere riscritte come equazioni simultanee per A e B:

$$\begin{cases} A \cdot n + B\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ A \cdot \sum_{i=1}^{n} x_i + B\sum_{i=1}^{n} x_i^{\,2} = \sum_{i=1}^{n} x_i y_i \end{cases} \longleftarrow \text{'equazioni normali'}$$

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

$$\begin{cases} A\cdot n + B\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ A\cdot\sum_{i=1}^{n} x_i + B\sum_{i=1}^{n} x_i^{\,2} = \sum_{i=1}^{n} x_i y_i \end{cases}$$

$$A = \frac{\left(\sum_{i=1}^{n} x_i^{\,2}\right)\cdot\left(\sum_{i=1}^{n} y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i\, y_i\right)}{n\left(\sum_{i=1}^{n} x_i^{\,2}\right) - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Migliori stime per le costanti A e B

$$B = \frac{n\left(\sum_{i=1}^{n} x_i\, y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^{\,2}\right) - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$A = \frac{\left(\sum_{i=1}^{n} x_i^{\,2}\right)\cdot\left(\sum_{i=1}^{n} y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i\, y_i\right)}{n\left(\sum_{i=1}^{n} x_i^{\,2}\right) - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$Y=A+Bx$$

Migliori stime per le costanti A e B

$$B = \frac{n\left(\sum_{i=1}^{n} x_i\, y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^{\,2}\right) - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Migliori stime per le incertezze di A e B (σ_A, σ_B)

$$\sigma_A = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\left(y_i - A - Bx_i\right)^2 \frac{\sum_{i=1}^{n} x_i^{\,2}}{n\left(\sum_{i=1}^{n} x_i^{\,2}\right) - \left(\sum_{i=1}^{n} x_i\right)^2}}$$

$$\sigma_B = \sqrt{\frac{n}{n-2}\sum_{i=1}^{n}\left(y_i - A - Bx_i\right)^2 \frac{1}{n\left(\sum_{i=1}^{n} x_i^{\,2}\right) - \left(\sum_{i=1}^{n} x_i\right)^2}}$$

Nel caso in cui la presumibile relazione tra x ed y non è lineare, in generale

$$y = f(x)$$

ciò che si fa è sempre minimizzare la somma degli scarti quadratici

$$\Phi = \sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2$$

ricavandosi un sistema di n equazioni in cui compaiono gli n parametri della funzione f(x)