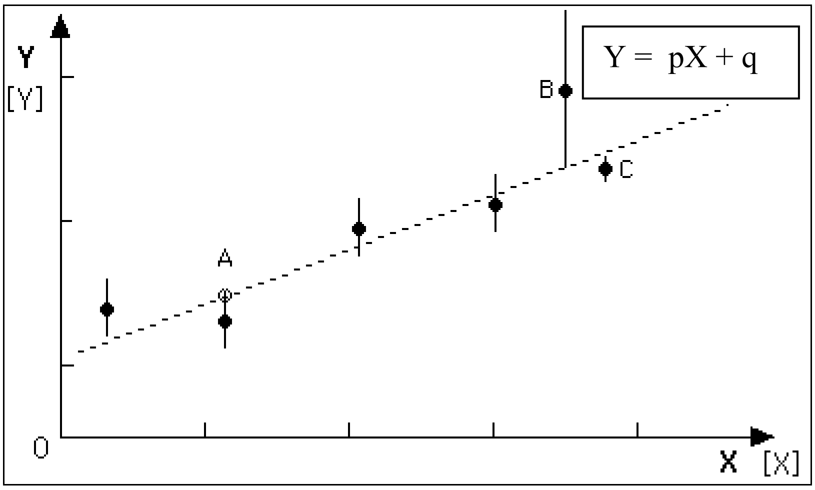
IL METODO DEI MINIMI QUADRATI

Vediamo come si presenterebbe un grafico delle misure avendo tratteggiato la retta: **Y = p X + q** della quale per il momento ancora non conosciamo i valori di **p** e **q**.



Poiché i punti su questo grafico rappresentano delle misure (quindi affette da incertezze) la retta (funzione analitica) non passerà per tutti i punti sperimentali anche se la nostra schematizzazione della legge fisica fosse corretta. Per chiarire meglio il ruolo delle incertezze ogni punto è stato rappresentato con un segmento la cui semilarghezza è pari all'incertezza della misura. Il nostro scopo è quello di determinare la retta che meglio approssima tutti i dati sperimentali.

Come esempio esaminiamo in dettaglio il punto A e calcoliamo la distanza del valore Y_i misurato direttamente (cerchio pieno) da quello stimato a partire dalla conoscenza di X_i (cerchio vuoto). Essa potrebbe aiutarci per stimare i parametri **p** e **q** (poiché le Y_i fluttuano intorno ai valori $\mathbf{p\,X_i + q}$, più è piccola la distanza e migliore sarà stata la nostra stima dei parametri).

Dovendo dare una valutazione complessiva della distanza di tutti i punti dalla retta potremmo sommare tutte le distanze delle Y_i dalle $\mathbf{pX_i + q}$: $Y_i - (pX_i + q)$.
Per evitare compensazioni di differenze positive e negative è meglio utilizzare il modulo della differenza o meglio ancora la somma dei quadrati delle distanze[1]:

$$U = \sum_{i=1,N} \left[Y_i - (pX_i + q)\right]^2$$

Tuttavia se esaminiamo i punti B e C notiamo che mentre la distanza dalla retta del punto B è superiore a quella del punto C, l'incertezza della misura del punto B è molto più elevata e quindi quest'ultima deve influenzare la nostra stima meno di quanto non debba fare il punto C che è meglio determinato.
Per questo motivo, qualora le diverse determinazioni di Y abbiano incertezze diverse, viene definita la variabile:

$$U = \sum u_i^2 = \sum \left(\frac{Y_i - (pX_i + q)}{\sigma_i} \right)^2$$

[1] In realtà la scelta della forma della funzione U si basa su argomentazioni ben più solide di queste. La loro discussione, tuttavia è oltre gli scopi di questo corso.

dove u_i rappresenta ancora la distanza di Y_i da $\mathbf{pX_i + q}$ ma, essendo divisa per la deviazione standard, viene attribuito un maggior peso alle misure con incertezza minore; si comporta cioè come una variabile ridotta (standardizzata).

Il metodo dei minimi quadrati consiste nella minimizzazione di U al variare dei parametri **p** e **q**: quanto più U è piccolo, tanto più la retta stimata passa nelle vicinanze delle coppie di punti sperimentali. I valori dei parametri che minimizzano U costituiscono la nostra stima.
Si ottengono imponendo che le derivate parziali di U rispetto ai parametri **p** e **q** siano nulle:

$$\frac{\partial U}{\partial p} = \sum \frac{\partial}{\partial p}\left(\frac{Y_i - (pX_i + q)}{\sigma_i}\right)^2 = -2\sum X_i \frac{Y_i - (pX_i + q)}{\sigma_i^2} =$$

$$= -2\sum \frac{X_i Y_i}{\sigma_i^2} + 2p\sum \frac{X_i^2}{\sigma_i^2} + 2q\sum \frac{X_i}{\sigma_i^2}$$

$$\frac{\partial U}{\partial q} = \sum \frac{\partial}{\partial q}\left(\frac{Y_i - (pX_i + q)}{\sigma_i}\right)^2 = -2\sum X_i \frac{Y_i - (pX_i + q)}{\sigma_i^2} =$$

$$= -2\sum \frac{Y_i}{\sigma_i^2} + 2p\sum \frac{X_i}{\sigma_i^2} + 2q\sum \frac{1}{\sigma_i^2}$$

ponendo $\frac{\partial U}{\partial p} = 0$ si ottiene: $\qquad \sum \frac{X_i Y_i}{\sigma_i^2} = p_S \sum \frac{X_i^2}{\sigma_i^2} + q_S \sum \frac{X_i}{\sigma_i^2}$

e da $\frac{\partial U}{\partial q} = 0$ si ottiene: $\qquad \sum \frac{Y_i}{\sigma_i^2} = p_S \sum \frac{X_i}{\sigma_i^2} + q_S \sum \frac{1}{\sigma_i^2}$

dove p_S e q_S rappresentano le stime rispettivamente di **p** e **q.**

Si è così ottenuto un sistema di due equazioni nelle due incognite p_S e q_S dalla cui risoluzione si ottengono:

$$p_S = \frac{\sum \frac{X_i Y_i}{\sigma_i^2} \sum \frac{1}{\sigma_i^2} - \sum \frac{X_i}{\sigma_i^2} \sum \frac{Y_i}{\sigma_i^2}}{\sum \frac{X_i^2}{\sigma_i^2} \sum \frac{1}{\sigma_i^2} - \sum \frac{X_i}{\sigma_i^2} \sum \frac{X_i}{\sigma_i^2}} \qquad q_S = \frac{\sum \frac{X_i^2}{\sigma_i^2} \sum \frac{Y_i}{\sigma_i^2} - \sum \frac{X_i Y_i}{\sigma_i^2} \sum \frac{X_i}{\sigma_i^2}}{\sum \frac{X_i^2}{\sigma_i^2} \sum \frac{1}{\sigma_i^2} - \sum \frac{X_i}{\sigma_i^2} \sum \frac{X_i}{\sigma_i^2}} \qquad \textbf{A)}$$

I valori di p_S e q_S sono affetti da un'incertezza statistica dovuta alle incertezze delle singole misure di Y_i ma, poiché p_S e q_S sono funzioni lineari delle variabili aleatorie Y_i [2] è possibile, applicando la formula di propagazione delle incertezze in modo esatto, stimare "facilmente" la loro varianza ottenendo:

---

[2] le X_i non sono v.a. perché abbiamo supposto trascurabile la loro varianza (i segmenti esprimenti l'incertezza sono solo nella direzione verticale); anche le σ_i non sono v.a. nella misura in cui le riteniamo note anziché stimate dai dati

$$\sigma^2(p_S) = \frac{\sum \frac{1}{\sigma_i^2}}{\left(\sum \frac{X_i^2}{\sigma_i^2}\sum\frac{1}{\sigma_i^2} - \sum\frac{X_i}{\sigma_i^2}\sum\frac{X_i}{\sigma_i^2}\right)} \qquad \sigma^2(q_S) = \frac{\sum \frac{X_i^2}{\sigma_i^2}}{\left(\sum \frac{X_i^2}{\sigma_i^2}\sum\frac{1}{\sigma_i^2} - \sum\frac{X_i}{\sigma_i^2}\sum\frac{X_i}{\sigma_i^2}\right)}$$

Nella pratica non sempre sarà possibile effettuare un così elevato numero di misure per ogni valore X_i da poter stimare con sufficiente approssimazione le varianze delle Y_i (anzi all'interno di questo corso non ne avremo quasi mai la possibilità).
Tuttavia spesso sarà possibile, se per tutte le misure si è usato lo stesso metodo e gli stessi strumenti, ritenere che le varianze delle Y_i siano tutte uguali fra loro.

In questo caso, considerando che $\sigma_i^2 = \sigma^2$, la funzione da minimizzare è:

$$U = \sum u_i^2 = \sum\left(\frac{Y_i - (pX_i + q)}{\sigma}\right)^2$$

e con pochi passaggi si ottengono (è un esercizio che si raccomanda caldamente) le equazioni:

$$p_S = \frac{N\sum X_i Y_i - \sum X_i \sum Y_i}{N\sum X_i^2 - \sum X_i \sum X_i} \qquad q_S = \frac{\sum X_i^2 \sum Y_i - \sum X_i Y_i \sum X_i}{N\sum X_i^2 - \sum X_i \sum X_i} \qquad \textbf{B)}$$

che si sarebbero potute ricavare dalle **A** ponendo le $\sigma_i^2 = \sigma^2$.

Anche in questo caso la formula di propagazione delle incertezze ci consente di calcolare le incertezze da associare alle stime dei parametri:

$$\sigma^2(p_s) = \sum_j \left(\frac{\partial p_S}{\partial Y_j}\right)^2 \sigma^2(Y_j) = \sum_j \left(\frac{\partial p_S}{\partial Y_j}\right)^2 \sigma^2 = \sum_j \left(\frac{NX_j - \sum X_i}{N\sum X_i^2 - \sum X_i \sum X_i}\right)^2 \sigma^2 =$$

$$= \frac{\sum_j \left[NX_j - \sum X_i\right]^2}{\left[N\sum X_i^2 - \sum X_i \sum X_i\right]^2}\sigma^2 = \frac{\sum_j \left[N^2 X_j^2 + \sum X_i \sum X_i - 2NX_j \sum X_i\right]}{\left[N\sum X_i^2 - \sum X_i \sum X_i\right]^2}\sigma^2 =$$

$$= \frac{N^2 \sum X_j^2 + N\sum X_i \sum X_i - 2N\sum X_j \sum X_i}{\left[N\sum X_i^2 - \sum X_i \sum X_i\right]^2}\sigma^2 =$$

$$= \frac{N\left[N\sum X_i^2 - \sum X_i \sum X_i\right]}{\left[N\sum X_i^2 - \sum X_i \sum X_i\right]^2}\sigma^2 = \frac{N\sigma^2}{N\sum X_i^2 - \sum X_i \sum X_i}$$

Analogamente si può ricavare: $\sigma^2(q_S) = \dfrac{\sum X_i^2 \sigma^2}{N\sum X_i^2 - \sum X_i \sum X_i}$

Ora non resta che stimare dai dati a disposizione la varianza delle Y_i che abbiamo supposto essere indipendente da Y_i e pari a σ^2.

Per questo scopo consideriamo la quantità $\dfrac{\sum\left[Y_i - (pX_i + q)\right]^2}{N}$; essa rappresenta la media aritmetica del quadrato degli scarti delle Y_i dai valori $\mathbf{pX_i + q}$ attesi in base alle X_i; in altre parole se N tendesse ad infinito sarebbe la solita definizione di varianza.
Poiché non conosciamo **p** e **q** ma solo le loro stime p_S e q_S una stima non distorta di σ^2 è:

$$\sigma^2 = \frac{\sum\left[Y_i - \left(p_S X_i + q_S\right)\right]^2}{N-2}$$

(N-2 perché questa volta ci sono due relazioni che legano fra loro le v.a. Y_i).

••• Generalizziamo quanto esaminato finora al caso di una funzione di ordine più elevato: schematizziamo il nostro fenomeno con una grandezza Y che dipende dalla grandezza X e da un certo numero di parametri non misurati o non misurabili $q_1, q_2, \ldots q_m$:

$$Y = f(X; q_1, q_2, \ldots, q_m)$$

Per ognuna delle N misure della grandezza X supponiamo di aver effettuato una serie di misure del corrispondente valore assunto dalla grandezza Y e di averne determinata l'incertezza:

X_1	Y_1	σ_1
X_2	Y_2	σ_2
...	...	...
X_N	Y_N	σ_N

Analogamente al caso particolare precedente si minimizza la quantità

$$U = \sum\left(\frac{Y_i - f\left(X_i; q_1, q_2, \cdots, q_m\right)}{\sigma_i}\right)^2$$

derivandola rispetto agli m parametri q; dal sistema di m equazioni in m incognite si ricavano le stime dei parametri q. In pratica è raro che il metodo dei minimi quadrati venga utilizzato per funzioni più complesse della parabola perché si preferiscono in quel caso altri metodi di stima dei parametri.

••• Un caso particolare di applicazione dei minimi quadrati è rappresentato dalla stima del coefficiente angolare di una retta passante per l'origine:

$$Y = \mathbf{p}\, X$$

Non è possibile utilizzare i risultati già ottenuti ponendo semplicemente $q_S = 0$ ma occorre ricavare nuovamente la stima p_S di **p** perché nel caso in cui la retta non sia forzata a passare per l'origine il coefficiente angolare potrebbe essere diverso.

La quantità da minimizzare è $U = \sum\left(\frac{Y_i - pX_i}{\sigma_i}\right)^2$.

$$\frac{\partial U}{\partial p} = \sum \frac{\partial}{\partial p}\left(\frac{Y_i - pX_i}{\sigma_i}\right)^2 = -2\sum X_i \frac{Y_i - pX_i}{\sigma_i} = -2\sum \frac{X_i Y_i}{\sigma_i^2} + 2p\sum\frac{X_i^2}{\sigma_i^2}.$$

Ponendo $\frac{\partial U}{\partial p} = 0$ si ottiene $\sum \frac{X_i Y_i}{\sigma_i^2} = p_S \sum \frac{X_i^2}{\sigma_i^2}$ da cui: $p_S = \frac{\sum \frac{X_i Y_i}{\sigma_i^2}}{\sum \frac{X_i^2}{\sigma_i^2}}$

In questo caso è facile verificare che p_S sia effettivamente un minimo in quanto $\frac{\partial^2 U}{\partial p^2} > 0$ (nel caso di 2 o più parametri la verifica è meno banale).

Applicando la formula di propagazione delle incertezze in modo esatto è al solito possibile stimare la varianza della stima di **p** che risulta essere una combinazione lineare delle v.a. Y_i:

$$\sigma^2(p_S) = \sum_{j=1,N}\left(\frac{\partial}{\partial Y_j}\frac{\sum\frac{X_iY_i}{\sigma_i^2}}{\sum\frac{X_i^2}{\sigma_i^2}}\right)^2\sigma_j^2 = \frac{\sum_{j=1,N}\left(\frac{\partial}{\partial Y_j}\sum\frac{X_iY_i}{\sigma_i^2}\right)^2\sigma_j^2}{\left(\sum\frac{X_i^2}{\sigma_i^2}\right)^2} = \frac{\sum_{j=1,N}\left(\frac{X_j}{\sigma_j^2}\right)^2\sigma_j^2}{\left(\sum\frac{X_i^2}{\sigma_i^2}\right)^2} =$$

$$= \frac{\sum_{j=1,N}\frac{X_j^2}{\sigma_j^2}}{\left(\sum\frac{X_i^2}{\sigma_i^2}\right)^2} = \frac{1}{\sum\frac{X_i^2}{\sigma_i^2}}$$

Anche in questo caso, qualora le varianze delle Y_i non siano note ma sia possibile ritenerle tutte uguali fra loro, col procedimento già visto si potrà ottenere la stima del parametro **p**:

Dalla relazione $p_S = \frac{\sum\frac{X_iY_i}{\sigma_i^2}}{\sum\frac{X_i^2}{\sigma_i^2}}$ ponendo $\sigma_i^2 = \sigma^2$ si ottiene $p_S = \frac{\sum X_iY_i}{\sum X_i^2}$

e da $\sigma^2(p_S) = \frac{1}{\sum\frac{X_i^2}{\sigma_i^2}}$ si ottiene $\sigma^2(p_S) = \frac{\sigma^2}{\sum X_i^2}$.

Dai dati a disposizione stimiamo quindi la varianza delle Y_i che abbiamo supposto essere indipendente da Y_i e pari a $\sigma^2 = \frac{\sum(Y_i - p_SX_i)^2}{N-1}$

(N-1 perché questa volta c'è una sola relazione che lega fra loro le v.a. Y_i).

••• Un caso notevole: la **media pesata**.

Può capitare che una grandezza fisica venga misurata più volte con metodi o strumenti diversi. In questo caso occorrerà stimare il valor vero della misura a partire da una serie di stime di medie e varianze.

Vediamo come il metodo dei minimi quadrati possa permetterci di ricavare la formula della media pesata.

In questo caso, per usare la notazione precedente, la funzione da studiare è **Y = q** mentre i dati a nostra disposizione sono N stime di Y_i e σ_i^2.

Al solito consideriamo la funzione $U = \sum u_i^2 = \sum\left(\frac{Y_i - q}{\sigma_i}\right)^2$;

$$\frac{\partial U}{\partial q} = \frac{\partial}{\partial q}\sum\left(\frac{Y_i - q}{\sigma_i}\right)^2 = -2\sum\frac{Y_i - q}{\sigma_i^2};$$

ponendo $\frac{\partial U}{\partial q} = 0$ si ottiene $$q_s = \frac{\sum \frac{Y_i}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}}.$$

La nostra migliore stima è dunque una media pesata dove i pesi sono l'inverso dei quadrati delle incertezze corrispondenti.
Va notato che alla stima del valor vero contribuiscono maggiormente le misure con varianze più piccole cioè $\frac{1}{\sigma_i^2}$ più grande (si attribuisce un significato maggiore alle misure meglio determinate).

Calcoliamo la varianza della media pesata col solito metodo:

$$\sigma^2(q_s) = \sum_j \left(\frac{\partial q_s}{\partial Y_j}\right)^2 \sigma_j^2 = \sum_j \left(\frac{\partial}{\partial Y_j} \frac{\sum \frac{Y_i}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}}\right)^2 \sigma_j^2 = \frac{\sum_j \left(\frac{1}{\sigma_j^2}\right)^2 \sigma_j^2}{\left(\sum \frac{1}{\sigma_i^2}\right)^2} = \frac{\sum_j \frac{1}{\sigma_j^2}}{\left(\sum \frac{1}{\sigma_i^2}\right)^2} = \frac{1}{\sum \frac{1}{\sigma_i^2}}.$$

Cioè la varianza della media pesata è data dall'inverso della somma degli inversi delle singole varianze (e quindi è inferiore alla più piccola di esse).

Bisogna prestare attenzione al valore delle misure prima di utilizzare la media pesata: la teoria si basa sul fatto che le diverse misure si riferiscono alla stessa grandezza fisica e siano esenti da effetti sistematici.

Pertanto, qualora le misure non siano compatibili fra di loro (cioè la loro differenza sia significativamente più grande della somma quadratica delle loro incertezze), non è lecito calcolare la media pesata.

Esempio

Viene misurato il periodo di oscillazione di un pendolo verticale in funzione della massa applicata all'estremità della molla che lo costituisce.
Esiste una relazione lineare fra il quadrato del periodo di oscillazione e la massa; valutare col metodo dei minimi quadrati la costante di proporzionalità fra T^2 e M:

$$T^2 = p\,M + q.$$

Sono state ottenute le seguenti 6 coppie di misure:

i	M [kg]	T^2 [s^2]			
1	0,400	0,388	N	=	6
2	0,480	0,460	ΣX_i	=	3,6 kg
3	0,560	0,531	ΣY_i	=	3,378 s^2
4	0,640	0,600	$\Sigma X_i Y_i$	=	2,123 kg s^2
5	0,720	0,664	ΣX_i^2	=	2,272 kg^2
6	0,800	0,735	ΣY_i^2	=	1,985 s^4
			denominatore di p_s e q_s	=	0,672 kg^2

A partire dalle misure si calcolano (è indispensabile una calcolatrice, meglio se programmabile o già programmata per stimare la retta di regressione) le quantità:

$p_s = (0{,}8639 \pm 0{,}0069)\ s^2/kg$

$q_s = (0{,}0453 \pm 0{,}0042)\ s^2$

Se graficaste i 6 punti e tracciaste la miglior retta "a occhio" trovereste che la pendenza e l'intercetta ricavate col solito modo sono sostanzialmente uguali ai parametri ottenuti col metodo dei minimi quadrati. A cosa corrisponderebbe σ?

ALTRO ESEMPIO

Calcolare la media pesata delle tre misure:

(127,4 ± 1,2) μm
(129,1 ± 2,3) μm
(131,2 ± 2,5) μm

$$\overline{X}_p = \frac{\frac{127,4}{1,2^2}+\frac{129,1}{2,3^2}+\frac{131,2}{2,5^2}}{\frac{1}{1,2^2}+\frac{1}{2,3^2}+\frac{1}{2,5^2}} \pm \frac{1}{\sqrt{\frac{1}{1,2^2}+\frac{1}{2,3^2}+\frac{1}{2,5^2}}} = \frac{88,472+24,405+20,992}{1,043} \pm 0,979 = (128,29 \pm 0,98)\,\mu\text{m}$$

per le 2 misure più distanti $\quad t = \frac{127,4-131,2}{\sqrt{1,2^2+2,5^2}} = -\frac{3,8}{2,77} = -1,37 \Rightarrow |t| < 3$

e quindi è lecita l'applicazione della formula per la media pesata

Se non fosse lecita significherebbe che una o più incertezze sono state valutate in difetto e quindi non è possibile utilizzarle come peso: si deve usare la media aritmetica

In questo esempio la media aritmetica fornirebbe

$$\overline{X} = \frac{127,4+129,1+131,2}{3} \pm \frac{1,380}{\sqrt{3}} = 129,233 \pm 1,099 = (129,2 \pm 1,1)\,\mu\text{m}$$