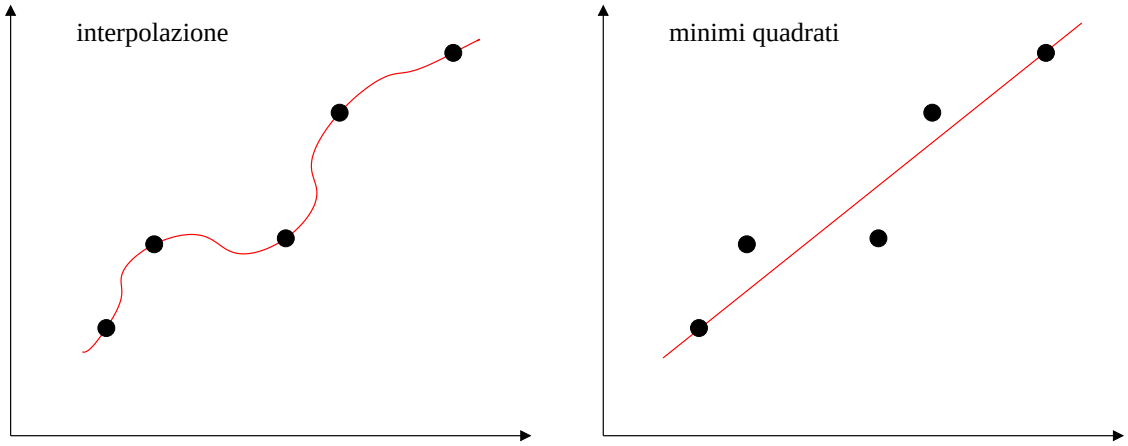
Capitolo 11

I minimi quadrati

Il problema con cui abbiamo a che fare adesso consiste nel determinare con buona approssimazione una curva (funzione) che descriva il fenomeno a cui i dati appartengono. Lo scopo è quello di conoscere l'andamento del fenomeno anche negli intervalli fra i punti di osservazione.

Consideriamo due tipi fondamentali di approssimazione:

- Interpolazione
- Approssimazione ai minimi quadrati



11.1 Il metodo dei minimi quadrati

L'approssimazione ai minimi quadrati è una tecnica di ottimizzazione volta a determinare una funzione analitica che approssimi un insieme di dati senza

necessariamente passare per i dati stessi (interpolazione), o meglio che si avvicini più possibile ad un'interpolazione di un insieme di dati (tipicamente punti del piano). Infatti, se i dati provengono da misure sperimentali e sono quindi affetti da errore oppure se non sono molto precisi (poche cifre significative) allora è opportuno approssimare ai minimi quadrati anziché interpolare.

In particolare la funzione trovata deve essere quella che **minimizza la somma dei quadrati delle distanze dai punti**.

11.1.1 Caso lineare

Molto spesso, un esperimento richiede la misura di molte coppie di valori di due grandezze di cui si vuol verificare l'esistenza di una **dipendenza** funzionale di tipo **lineare**. Consideriamo 2 variabili x ed y legate da una **relazione lineare**, del tipo:

$$y = A + Bx$$

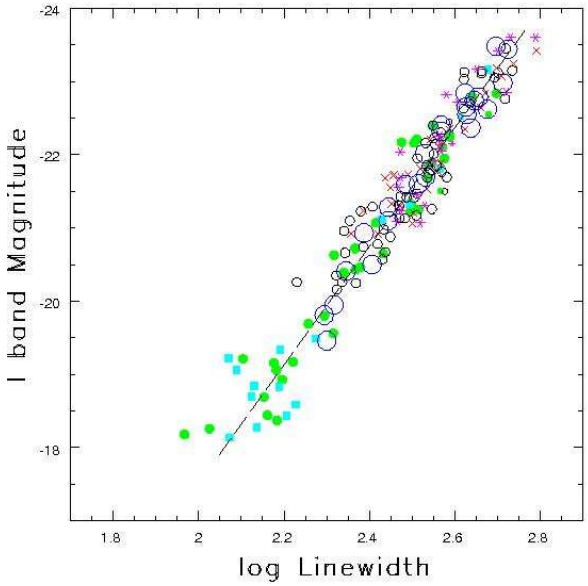
dove A eB sono due costanti.

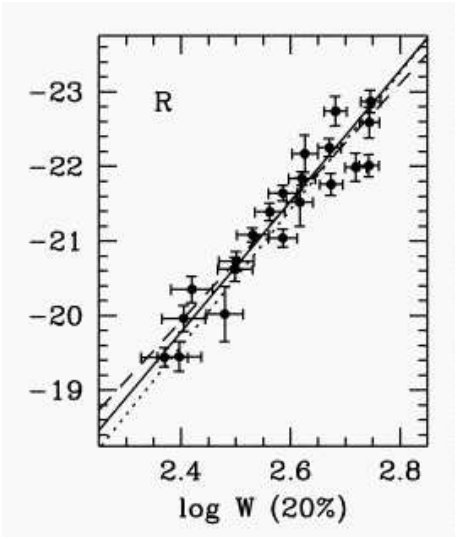
In questa formulazione x rappresenta la *variabile indipendente*, y rappresenta la *variabile dipendente*, A è l'intercetta e B la pendenza o coefficiente angolare.

Esempi:

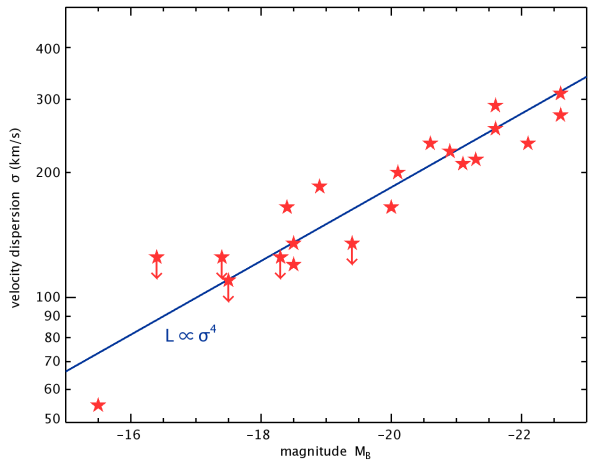
- in un moto uniformemente accelerato la velocità v dipende linearmente dal tempo t: $v = v_o + at$ dove a è l'accelerazione costante;

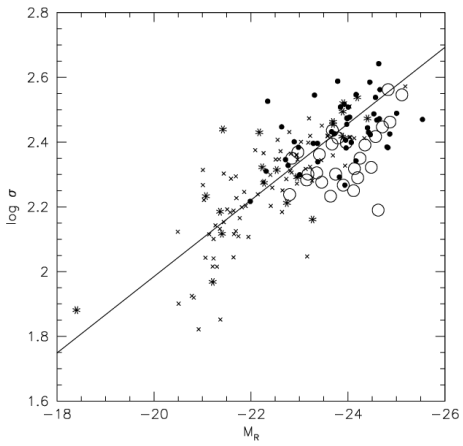
- la relazione Tully-Fisher osservata per le galassie a spirale;





- la relazione Faber-Jackson osservata per le galassie ellittiche





Si abbiano a disposizione di N misure $x_1, x_2, \ldots x_N$ e le corrispondenti N misure $y_1, y_2, \cdots y_N$. A causa dell'inevitabile verificarsi di errori di misura, i punti (x_i, y_i) non si disporranno esattamente sulla relazione prevista tra le due grandezze fisiche, ma cadranno in prossimità alla retta che la descrive.

L'obiettivo è quello di trovare la retta che meglio si adatta alle misure, cioè in pratica trovare la miglior stima per le costanti A e B utilizzando le N coppie di dati a disposizione (x_i, y_i). Si faccia la seguente ipotesi: *l'errore su una delle variabili sia trascurabile* (es: la variabile indipendente x) *rispetto a quello dell'altra.*

Se A e B fossero noti $\longrightarrow$ per ogni x_i potremmo calcolare il valore vero del corrispondente y_i:

$$y_i^* = A + B\,x_i$$

A causa degli errori di misura in generale si avrà $y_i \neq y_i^*$. La quantità $s_i = y_i - y_i^*$ rappresenta lo scarto del generico y_i dal rispettivo valore teorico y_i^*. Sia σ_i l'errore quadratico medio della generica misura y_i. In generale, ad ogni misura y_i corrisponderà un σ_i diverso.

La probabilità che le operazioni di misura forniscano il valore osservato y_i è data da:

$$P(y_i) \propto \frac{1}{\sigma_i} e^{\frac{-s_i^2}{2\sigma_i^2}} \propto \frac{1}{\sigma_i} e^{\frac{-(y_i - A - B\,x_i)^2}{2\sigma_i^2}}$$

La probabilità composta di ottenere l'insieme delle misure $y_1, y_2, \cdots y_N$ (eventi compatibili indipendenti) è data dal prodotto:

$$P(y_1, y_2, \cdots y_N) = P(y_1) \times P(y_2) \times \cdots P(y_N) \propto \prod_{i=1}^{N} \frac{1}{\sigma_i} e^{\frac{-\chi^2}{2}}$$

in cui l'esponente è espresso da:

$$\boxed{\chi^2 = \sum_{i=1}^{N} \frac{(y_i - A - B\,x_i)^2}{\sigma_i^2}}$$

Al variare dei parametri A e B, la retta $y = A + Bx$ traslerà verso il basso o verso l'alto, o cambierà pendenza. Di conseguenza anche gli scarti assumeranno valori ogni volta diversi. Alcuni s_i diventeranno più grandi, altri più piccoli, altri ancora cambieranno di segno, e così via.

Infatti, il valore assoluto di un certo s_i non è, di per sè, significativo, mentre ciò che conta è il valore dello scarto s_i rispetto all'errore statistico σ_i da cui è affetta la determinazione y_i. Al tendere a zero dei σ_i, ovvvero per misure infinitamente precise, la retta dovrebbe passare esattamente per y_i e quindi gli scarti dovrebbero essere nulli.

Il requisito che deve soddisfare la **miglior retta** è quello di rendere il più possibile piccoli i valori assoluti degli scarti divisi per i rispettivi errori. È chiaro che rendere piccolo uno scarto implica renderne altri grandi, quindi occorre un compromesso.

Si dimostra che, in base ad un principio generale detto di **massima verosimiglianza**, la miglior retta è quella che *rende minima la somma dei quadrati degli scarti divisi per gli errori.*

Ma perchè minimizzare proprio la somma dei quadrati degli scarti? La somma degli scarti non è adatta a quantificare l'aderenza della retta agli N punti. Gli scarti possono essere infatti positivi o negativi e la loro somma può essere piccola in valore assoluto anche per rette palesemente inadatte a descrivere gli N punti. Per esempio, siano dati i tre punti allineati $(1,1)$, $(2,2)$, $(3,3)$. Ovviamente la miglior retta è $y = x$; la somma degli scarti è nulla. Ma è nulla anche per qualunque retta passi per $(2,2)$, quindi di equazione $y = m(x-2)+2$. È dunque opportuno utilizzare valori positivi.

La somma dei moduli degli scarti è una scelta plausibile. Tuttavia la somma dei quadrati anziché dei valori assoluti si accorda in modo naturale con la media aritmetica: la media aritmetica di una sequenza di numeri gode della proprietà di rendere minima la somma dei quadrati degli scarti.

Ciò equivale a trovare la coppia di costanti A e B tali da rendere massima la probabilità composta $P(y_1, y_2, \cdots y_N)$, ovvero minima la quantità χ^2.

Per prima cosa determiniamo i punti in cui le derivate parziali di χ^2 rispetto ad A e B si annullano:

$$\frac{\partial \chi^2}{\partial A} = -2 \sum_{i=1}^{N} \frac{(y_i - A - Bx_i)}{\sigma_i^2} = 0$$

$$\frac{\partial \chi^2}{\partial B} = -2 \sum_{i=1}^{N} \frac{(y_i - A - Bx_i)}{\sigma_i^2} x_i = 0$$

Queste due equazioni possono essere riscritte nella forma:

$$\begin{cases} \displaystyle\sum_{i=1}^{N} \frac{y_i}{\sigma_i^2} - A \sum_{i=1}^{N} \frac{1}{\sigma_i^2} - B \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} = 0 \\ \displaystyle\sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2} - A \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} - B \sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2} = 0 \end{cases}$$

Si tratta di un sistema di due equazioni lineari nelle due incognite A e B, che ammette una sola soluzione.

Definendo:

$$S_x = \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}; \qquad S_y = \sum_{i=1}^{N} \frac{y_i}{\sigma_i^2}; \qquad S_{xx} = \sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2};$$

$$S_{xy} = \sum_{i=1}^{N} \frac{x_i \cdot y_i}{\sigma_i^2}; \qquad S = \sum_{i=1}^{N} \frac{1}{\sigma_i^2}$$

i coefficienti A e B sono dati da:

$$\boxed{A = \frac{S_{xx} \cdot S_y - S_x \cdot S_{xy}}{S \cdot S_{xx} - S_x \cdot S_x}} \qquad \boxed{B = \frac{S \cdot S_{xy} - S_x \cdot S_y}{S \cdot S_{xx} - S_x \cdot S_x}}$$

Nell' ipotesi che *le N misure della variabile y siano descrivibili dalla distribuzione di Gauss con errore quadratico medio* σ_y (cioè $\sigma_i = \sigma_y \; \forall i = 1, 2, \cdots N$), le equazioni si semplificano in quanto:

$$S_x = \frac{1}{\sigma_y^2} \sum_{i=1}^{N} x_i; \qquad S_y = \frac{1}{\sigma_y^2} \sum_{i=1}^{N} y_i; \qquad S_{xx} = \frac{1}{\sigma_y^2} \sum_{i=1}^{N} x_i^2$$

$$S_{xy} = \frac{1}{\sigma_y^2} \sum_{i=1}^{N} x_i \cdot y_i; \qquad S = \sum_{i=1}^{N} \frac{1}{\sigma_y^2} = \frac{N}{\sigma_y^2}$$

Semplificando i fattori σ_y^2 che compaiono ovunque ricaviamo:

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

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

Dividendo ogni termine per N^2 e dalla definizione di media aritmetica di una grandezza otteniamo:

$$A = \frac{\overline{x^2}\,\overline{y} - \overline{x} \cdot \overline{xy}}{\overline{x^2} - \overline{x}^2}$$

$$B = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2}$$

La retta risultante che ha come intercetta sull'asse delle Y il parametro A e come coefficiente angolare il parametro B, entrambi determinati a partire dai dati sperimentali di x e y è detta **retta dei minimi quadrati o retta di regressione di** y **su** x. Si verifica facilmente che la retta dei minimi quadrati così determinata passa attraverso il punto di coordinate:

$$\tilde{x} = \frac{\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \qquad \tilde{y} = \frac{\sum_{i=1}^{N} \frac{y_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}$$

Utilizzando un'analogia meccanica, esso rappresenta il baricentro degli N punti di coordinate $(x_1, y_1), (x_2, y_2) \cdots (x_N, y_N)$, nei quali siano concentrate le masse $\frac{1}{\sigma_1^2}, \frac{1}{\sigma_2^2}, \cdots, \frac{1}{\sigma_N^2}$.

Nel caso di misure y_i tutte affette dalla stessa indeterminazione statistica σ_y, i valori $\tilde{x}$ e $\tilde{y}$ coincidono semplicemente con le medie aritmetiche delle due serie di misure. In altre parole, la retta di regressione passa per il punto avente come coordinate i valori medi delle due serie di misure.

11.1.2 Indeterminazione sulla misura di y

Abbiamo ipotizzato che ogni misura y_i si distribuisca attorno al valor vero y_i^* secondo una distribuzione gaussiana di larghezza σ_y, la stessa per tutti le N determinazioni.

In questo caso, è importante notare che le soluzioni per i parametri A e B **non** dipendono da σ_y, ma solo dai dati sperimentali per x e y. In altre parole la soluzione al problema non richiede una conoscenza a priori della indeterminazione statistica della grandezza y.

Di fatto, una buona stima dell'errore quadratico medio è ottenibile sempre a partire dai soli dati sperimentali:

$$\sigma_y = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - A - Bx_i)^2}$$

Se le due costanti teoriche A e B sono approssimate dalle nostre miglior stime dei dati sperimentali i gradi di libertà del sistema si riducono da N a $N - 2$, per cui al fine di non sottostimare σ_y, è più corretto utilizzare la formula:

$$\sigma_y = \sqrt{\frac{1}{N-2} \sum_{i=1}^{N} (y_i - A - Bx_i)^2}$$

La quantità σ_y stima la distanza media dei punti dalla retta di regressione. Se σ_y è circa uguale all' incertezza attesa δy, i dati sono consistenti con la

relazione lineare stabilita, se invece σ_y è molto maggiore di δy ci sono motivi per dubitare della relazione lineare che si presume legare le variabili x e y.

11.1.3 Indeterminazione sui parametri A e B

I parametri A e B risultano affetti da incertezza poiché lo sono i valori y_i. Utilizzando le formule di propagazione degli errori statistici è possibile stimare σ_A e σ_B in termini di σ_i:

$$\sigma_A^2 = \sum_{i=1}^{N} \left(\frac{\partial A}{\partial y_i}\right)^2 \sigma_i^2 \qquad \sigma_B^2 = \sum_{i=1}^{N} \left(\frac{\partial B}{\partial y_i}\right)^2 \sigma_i^2$$

• Primo caso: Errori tutti uguali $\sigma_1 = \sigma_2 = \cdots \sigma_N = \sigma_y$

$$\sigma_A^2 = \sigma_y^2 \frac{\sum_{i=1}^{N} x_i^2}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2} = \frac{\sigma_y^2 \overline{x^2}}{N(\overline{x^2} - \overline{x}^2)}$$

$$\sigma_B^2 = \sigma_y^2 \frac{N}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2} = \frac{\sigma_y^2}{N(\overline{x^2} - \overline{x}^2)}$$

• Secondo caso: Errori diversi $\sigma_1 \neq \sigma_2 \neq \cdots \sigma_N$

In questa circostanza la terminologia appropriata è quella di **minimi quadrati ponderati**.

Riprendiamo le espressioni generali per i parametri A e B in funzione delle quantità S, ed esplicitando solo quelle che contengono una dipendenza dalle y_i:

$$A = \frac{S_{xx} \cdot \sum_{i=1}^{N} \frac{y_i}{\sigma_i^2} - S_x \cdot \sum_{i=1}^{N} \frac{x_i \cdot y_i}{\sigma_i^2}}{S \cdot S_{xx} - S_x \cdot S_x}$$

$$B = \frac{S \cdot \sum_{i=1}^{N} \frac{x_i \cdot y_i}{\sigma_i^2} - S_x \cdot \sum_{i=1}^{N} \frac{y_i}{\sigma_i^2}}{S \cdot S_{xx} - S_x \cdot S_x}$$

Le derivate parziali sono:

$$\frac{\partial A}{\partial y_i} = \frac{S_{xx} \cdot \frac{1}{\sigma_i^2} - S_x \cdot \frac{x_i}{\sigma_i^2}}{S_{xx} \cdot S - S_x^2}$$

$$\frac{\partial B}{\partial y_i} = \frac{S \cdot \frac{x_i}{\sigma_i^2} - S_x \cdot \frac{1}{\sigma_i^2}}{S_{xx} \cdot S - S_x^2}$$

Dopo lunghi ma banali passaggi i risultati finali sono:

$$\boxed{\sigma_A^2 = \frac{S_{xx}}{S_{xx} \cdot S - S_x^2}} \qquad \boxed{\sigma_B^2 = \frac{S}{S_{xx} \cdot S - S_x^2}}$$

Abbiamo in questo modo determinato la miglior linea retta che approssima un certo insieme di dati sperimentali, rendendo minima la somma dei quadrati delle distanze, misurate nella direzione dell'asse delle y, dei punti dalla retta. Tali distanze sono comunemente denominate **residui**. La retta così determinata è denominata **retta di regressione di x su y**.

Avremmo anche potuto, in modo analogo, minimizzare la somma dei residui misurati nella direzione dell'asse delle x. In altre parole, rappresentata la retta come:

$$x = A + B\,y$$

minimizzare la quantità

$$\chi^2 = \sum_{i=1}^{N} \frac{(x_i - A - B\,y_i)^2}{\sigma_i^2}$$

La retta così determinata è denominata **retta di regressione di y su x**. In generale le due rette di regressione non coincidono. Quanto più la distribuzione dei punti è prossima ad essere rettilinea, tanto più le due rette di regressione tendono a coincidere.

Una terza possibilità sarebbe quella di rendere minima la somma dei quadrati dei residui ortogonalmente dai punti dalla retta stessa. La retta così determinata è denominata **retta di regressione ortogonale**. Queste considerazioni mostrano che la risposta al problema fornita dal metodo dei minimi quadrati non è in generale univoca, potendo variare a seconda della quantità che si intende minimizzare. In pratica è vero che, se la distribuzione dei dati sperimentali ben corrisponde ad una distribuzione lineare, le diverse rette di regressione sono contenute entro l'indeterminazione entro cui ciascuna di esse è determinata.

Spesso, anche quando la relazione funzionale tra le due grandezze x e y è di tipo non lineare, ci si può ricondurre al caso lineare mediante opportuni cambiamenti di variabili. Esempi significativi di possibile frequente applicazione:

• $y = A + B\,x^p$

Se p è noto, possiamo porre $z = y$ e $w = x^p$, ottenendo $z = A + Bw$

• $y = B\,x^p$

Se p è incognito, possiamo porre $z = \log y$ e $w = \log x$, $a = \log B$, $b = p$, ottenendo $z = a + bw$

• $y = Ae^{Bx}$, con A e B da determinarsi.

Poniamo $z = \log y$, $w = x$, con $a = \log A$, $b = B$, ottenendo $z = a + bw$

Il cambiamento di variabili impone molta cautela e attenzione. Infatti se tutte le misure y_i sono affette dalla stessa indeterminazione σ_y, ciò non è ugualmente valido per i valori $z_i = \log y_i$. Infatti dalla propagazione degli errori sappiamo che $\sigma_{z_i} = \dfrac{\sigma_y}{y_i}$, che dipende dal valore del corrispondente y_i.

La conversione al caso lineare è utile non solo per la possibilità di applicare il metodo dei minimi quadrati appena descritto, ma anche per verificare graficamente la relazione lineare tra due grandezze fisiche.

11.2 Generalizzazione del metodo dei minimi quadrati

Il metodo dei minimi quadrati, introdotto supponendo una relazione lineare tra x e y, può essere applicato in modo più generale.

Consideriamo, per semplicità, il caso di due variabili x e y. Si abbiano a disposizione N coppie di dati $(x_1, y_1), (x_2, y_2), \cdots (x_N, y_N)$. Supponiamo inoltre che l'errore su x sia trascurabile e che le y_i siano affette da indeterminazioni statistiche σ_i.

Si sappia o si supponga che la relazione tra x e y sia espressa dalla funzione:

$$y = f(x, c_1, c_2, \cdots c_p)$$

e si vogliano determinare i valori dei parametri $c_1, c_2, \cdots c_p$ in corrispondenza dei quali la curva meglio approssima i punti sperimentali. Il metodo dei minimi quadrati permette di trovare tali valori; essi sono tali da rendere minima la quantità:

$$\boxed{\chi^2 = \sum_{i=1}^{N} \frac{[y_i - f(x, c_1, c_2, \cdots c_p)]^2}{\sigma_i^2}}$$

I valori dei parametri si ottengono risolvendo il sistema di equazioni:

$$\frac{\partial \chi^2}{\partial c_1} = \frac{\partial \chi^2}{\partial c_2} = \cdots \frac{\partial \chi^2}{\partial c_p} = 0$$

In pratica la soluzione di tale sistema di equazioni può essere assai laboriosa a seconda della particolare forma della funzione. Difficilmente si impiegano metodi analitici, più spesso il minimo di χ viene determinato numericamente con l'utilizzo di appositi programmi fatti operare al computer (e.g. ricavare il gradiente di χ e muoversi a passi verso valori sempre più piccoli).

Cosa cambia se anche le viariabili indipendenti x sono affette da errore? Indichiamo con σ_x e σ_y gli errori statistici su x e y rispettivamente. Applicando il principio della **massima verosimiglianza** si dimostra che i migliori valori dei parametri c_i sono quello che minimizzano l'espressione:

$$\chi^2 = \sum_{i=1}^{N} \frac{[y_i - f(x, c_1, c_2, \cdots c_p)]^2}{\delta_i^2}$$

avendo posto

$$\delta_i^2 = \sigma_{y_i}^2 + \left(\frac{\partial f}{\partial x}\right)^2_{x=x_i} \sigma_{x_i}^2$$

dove i singoli δ_i dipendono, in generale, essi stessi dai parametri c_i.

Possiamo osservare che δ_i^2 è proprio la varianza degli scarti $s_i = y_i - f(x_i)$. Infatti dalla propagazione degli errori statistici si ricava:

$$\sigma_{s_i}^2 = \sigma_{y_i}^2 + \left(\frac{\partial f}{\partial x}\right)^2 \sigma_{x_i}^2$$

In altre parole tutto avviene come se non ci fossero errori sulla variabile indipendente x a patto di sostituire all'errore σ_{y_i}, l'**errore efficace** δ_i, che ha il significato di errore quadratico medio dello scarto s_i.

D'altra parte si vede che per $\sigma_{x_i} \to 0$, si ha che $\delta_i \to \sigma_{y_i}$.

Riportandoci al caso di partenza, e cioè che la relazione $y = f(x, c_1, c_2, \cdots c_p)$ sia lineare:

$$y = A + Bx$$

l'errore efficace vale semplicemente:

$$\delta_i = \sqrt{(\sigma_{y_i})^2 + B^2(\sigma_{x_i})^2}$$

e quindi dobbiamo minimizzare la quantità

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - A - Bx_i)^2}{\sigma_{y_i}^2 + B^2\sigma_{x_i}^2}$$

Poiché il parametro B compare anche al denominatore le formule precedenti per calcolare i valori di A e B diventano molto più complicate. In questi casi in genere si usa per la derivata di $f(x, c_1, c_2, \cdots, c_p)$, (ovvero per B se la funzione è lineare) un valore approssimato, per esempio derivato graficamente, oppure ricavato da una prima stima provvisoria dei parametri effettuata senza tener conto delle σ_{x_i}.

11.3 Coefficiente di correlazione lineare

Supponiamo di aver misurato N coppie di valori (x_1,y_1), ..., (x_N,y_N) che pensiamo siano legate da una relazione lineare $y = A + Bx$. Abbiamo visto che possiamo calcolare A e B e avere anche una stima delle incertezze sulle misure. Però non sempre è possibile avere una stima delle incertezze in anticipo per poter capire se due grandezze sono legate o meno dalla relazione lineare. Abbiamo quindi bisogno di una quantità che ci dica quanto le due grandezze sono legate o "correlate".

Definiamo quindi il **coefficiente di correlazione lineare**:

$$r = \frac{\sigma_{xy}}{\sigma_x\, \sigma_y}$$

dove:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}} \qquad \sigma_y = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{N}}$$

sono gli scarti quadratici medi delle quantità x e y, mentre:

$$\sigma_{xy} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{N}$$

è il termine di **covarianza**.

Si tenga ben presente che σ_x e σ_y hanno un significato un po' diverso dal solito: se facciamo N misure della stessa quantità x, ci aspettiamo che σ_x sia piccolo. In questo caso invece poiché abbiamo N misure diverse, non è affatto detto che σ_x o σ_y siano piccoli.

Quindi sostituendo si ottiene:

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

Poiché la covarianza soddisfa la "disuguaglianza di Schwarz":

$$|\sigma_{xy}| \leq \sigma_x \sigma_y$$

si vede chiaramante che il coefficiente di correlazione è un termine compreso fra -1 e 1:

$$-1 \leq r \leq 1$$

Adesso supponiamo che i punti (x_i,y_i) si trovino esattamente sulla retta $y = A + Bx$. Avremo quindi che:

$$y_i = A + Bx_i$$

e inoltre:

$$\bar{y} = A + B\bar{x}$$

Sottraendo membro a membro queste relazioni si ottiene:

$$y_i - \bar{y} = B(x_i - \bar{x})$$

e sostituendo nell'espressione del coefficiente di correlazione:

$$r = \frac{B\sum_{i=1}^{N}(x_i - \bar{x})^2}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 B^2 \sum_{i=1}^{N}(x_i - \bar{x})^2}} = \frac{B}{|B|} = \pm 1$$

Quindi, in pratica, nel caso ideale in cui i punti giacciono perfettamente su una retta il coefficiente di correlazione vale +1 o -1, a seconda della pendenza positiva o negativa della retta. Ovviamente questo normalmente non accade, ma se due quantità sono ben correlate avremo un r molto vicino a +1 o -1.

Se al contrario, supponiamo che non vi sia alcuna correlazione fra le variabili x e y, allora qualunque sia il valore di y_i, ogni valore di x_i ha la stessa probabilità di trovarsi sopra $\bar{x}$ o sotto $\bar{x}$. E quindi i termini nella somma:

$$\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$

hanno la stessa probabilità di essere positivi o negativi, mentre i termini al denominatore di r sono tutti positivi. Al limite per N che tende all'infinito, il coefficiente di correlazione sarà pari a zero. Di solito si ha un numero finito di dati e quindi r non verrà mai zero, ma comunque ci si aspetta che sia piccolo quando non è presente alcuna correlazione fra x e y.

È evidente che il semplice calcolo di r non implica che si possa rispondere con sicurezza alla domanda se via o non via sia correlazione, proprio per il fatto che in caso di correlazione r è vicino a ± 1, mentre in caso di assenza di correlazione r è vicino a 0.

Supponiamo che le due quantità siano non correlate: al limite per N che tende all'inifito avremo che r varrà zero. Siccome N è normalmente un numero finito, allora conviene calcolare la probabilità che r sia maggiore di un valore specifico r_0:

$$P_N(|r| \geq r_0)$$

La seguente tabella mostra le probabilità percentuali di r in funzione di N e r_0:

In pratica, dato un set di N misure di due quantità x e y, si calcola il valore del coefficiente di correlazione lineare e si assume che sia r_0 e poi si può usare la tabella qui riportata per calcolare la probabilitèche le due quantità siano non correlate. Se questa probabilità è bassa potremo ragionevolmente dire che essere sono ben correlate, e viceversa.

	r_0										
N	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
3	100	94	87	81	74	67	59	51	41	29	0
6	100	85	70	56	43	31	21	12	6	1	0
10	100	78	58	40	25	14	7	2	0.5		0
20	100	67	40	20	8	2	0.5	0.1			0
50	100	49	16	3	0.4						0