

Laboratorio borra

Tommaso Miliani

02/12/24

1 Regressione lineare (fit lineare)

$$y = a + bx$$

Rappresenta il modello che lega due grandezza x, y

σ_x

$\sigma_{yi} = \sigma_y$

Ipotesi :

trascurabili

posso anche non conoscerlo

Errori casuali dominanti.

Le probabilità in funzione di tutte le variabili in gioco

$$P(x_i, y_i, \dots, y_n) = P(x_i) \cdot P(y_i) \cdot \dots \cdot P(y_n);$$

P è quindi la massima verosimiglianza per cui la probabilità di $P(y_i, \dots, y_n)$ deve essere massima.

Nella Gaussiana si ha che il vero valore di Y è dato dal modello $Y = a + bx$ per cui sostituendo nella gaussiana si ottiene la probabilità:

$$P(y_i, \dots, y_n) \propto \exp\left(-\frac{\sum_{i=1}^n (y_i - a - bx_i)^2}{2\sigma_y^2}\right)$$

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - a - bx_i)^2}{\sigma_y^2} \text{ tale che } \begin{cases} \frac{\partial \chi^2}{\partial a} = 0 \\ \frac{\partial \chi^2}{\partial b} = 0 \end{cases}$$

1. Conosciuti i σ_y che io conosco dai punti sperimentali che deviano rispetto alla retta sperimentale. Quello che vedo è che a questo punto sulla retta c'è una certa deviazione standard e quindi procedo a costruire una σ_y a partire dai risultati ottenuti sul grafico dagli esperimenti.

$$\bar{\sigma}_y^2 = \sum_{i=1}^n \frac{(y_i - a - bx_i)^2}{N - 2}$$

E' una sorta di media di tutti i singoli σ_{yi} . N-2 è presente e torna poiché i punti che stanno su una retta e voglio verificarlo ho bisogno di almeno due punti (da due punti passa una ed una sola retta). Quindi i primi due li uso per verificare che esiste effettivamente una legge lineare che mi vincola i miei punti sul grafico e quindi $N \geq 3$.

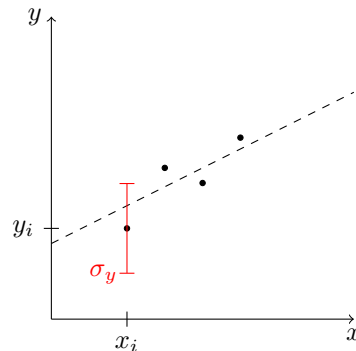
Conosciuti i σ_y il mio caso migliore è che $\bar{\sigma}_y = \sigma_y$ e che quindi i miei σ_y siano confrontabili con lo scarto dal modello lineare.

Altrimenti $\bar{\sigma}_y \gg \sigma_y$ allora siamo nel seguente in figura (le barre rosse sono molto piccole ma le misure si discostano troppo dal modello). Se invece $\bar{\sigma}_y \ll \sigma_y$ allora il mio σ_y è sovrastimato. e le bande rosse sono molto grandi. Nell'ultimo caso se non conosco σ_y allora posso dire che $\sigma_y = \bar{\sigma}_y$. Questo è il caso più comune. Ora si trova l'incertezza di a e b

$$a = a(x_i, y_i) \pm \sigma_a$$

$$b = b(x_i, y_i) \pm \sigma_b$$

Figura 1: Misure rispetto al modello lineare



Gli x_i non hanno errore mentre gli y_i hanno un errore ben definito e quindi, attraverso il metodo delle derivate parziali, si ottiene che:

$$\sigma_a = \sqrt{\sum_{i=1}^n \left| \frac{\partial a}{\partial y_i} \right|^2} \sigma_y^2 = \sigma_y \sqrt{\frac{\sum_{i=1}^n x_i^2}{N \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}}$$

$$\sigma_b = \sqrt{\sum_{i=1}^n \left| \frac{\partial b}{\partial y_i} \right|^2} \sigma_y^2 = \sigma_y \sqrt{\frac{N}{N \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}}$$

Il coefficiente di correlazione lineare è dato da:

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

Nel caso generale dunque \bar{x} e \bar{y} sono la media di tutte le misure e se queste stanno tutte sulla retta modello allora anche loro due stanno sulla retta modello e si calcolano come:

$$\bar{y} = a + b\bar{x}$$

$$\bar{y} = \sum_{i=1}^n \frac{y_i}{N} = \frac{1}{N} \sum_{i=1}^n (a + bx_i)$$

Da qui sostituendo ne deriva che:

$$\sum_{i=1}^n (y_i - \bar{y}) = \sum_{i=1}^n (a + bx_i - a - b\bar{x})$$

$$= b \sum_{i=1}^n (x_i - \bar{x})$$

Date queste sostituzioni si ottiene che

$$r = \frac{b \sum_{i=1}^n (x_i - \bar{x})^2}{|b| \sum_{i=1}^n (x_i - \bar{x})^2} =$$

$$\frac{b}{|b|} = \pm 1$$

Adesso se ci sono almeno due punti che differiscono di molto (potrebbero non essere i soli) e controllo come contribuiscono ad r . (nel caso ideale $r = 1$ mentre nella realtà più è vicino ad 1 e meglio è il nostro modello).

2 Metodo grafico per fit lineari

x	y	σ_y
x_i	y_i	σ_{y_i}

Cerco prima la retta con la pendenza minore (rossa) che sta dentro le barre d'errore e poi quella con pendenza maggiore (blu) rispetto a quello modello.

a_{best}, b_{best}	retta nera tratteggiata
a_{min}, b_{max}	retta blu
a_{max}, b_{min}	retta rossa

A questo punto:

$$\delta_a = \max\{|a_{best} - a_{min}|, |a_{best} - a_{max}|\}$$

$$\delta_b = \max\{|b_{best} - b_{min}|, |b_{best} - b_{max}|\}$$

Si arriva dunque a definire a come la distanza tra la retta modello e quella con la pendenza minore mentre $b = \frac{\Delta y}{\Delta x}$. Se un punto dista dalle rette possibili più della sua barra di errore allora va scartato (questo differisce dal criterio di Cheuvenet).

Figura 2: Modello lineare non conforme con i dati trovati

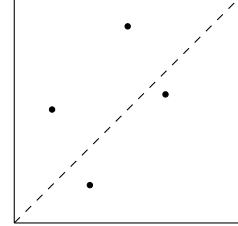


Figura 3: Grafico per il fit

