INTERPOLAZIONE DI DATI

Acquisire dati derivanti da osservazioni o esperimenti è una delle componenti fondamentali dell'attività scientifica. I dati tuttavia devono essere sottoposti ad una disamina attenta, per comprendere se fra essi si possano stabilire delle correlazioni o delle relazioni di tipo funzionale.

Una delle operazioni matematiche basilari per raggiungere questo fine è l'interpolazione.

Supponiamo per semplicità di avere a che fare con dati relativi a due sole variabili, x e y , di cui abbiamo solo n coppie di valori x_i e y_i , $i = 1...n$, acquisite mediante una attività di misura (ricordiamo gli espermenti del biennio: $x =$volume, $y =$massa, oppure $x =$lunghezza del pendolo, $y =$periodo ecc...) o mediante rilevamenti di altro genere.

L'obiettivo è stabilire se vi sia una relazione di tipo funzionale fra il secondo e il primo elemento di ogni coppia. Rappresentare questi dati in un diagramma cartesiano (ad ogni coppia (x_i, y_i) corrisponde il punto P_i del piano di cui x_i e y_i sono le coordinate) può essere di aiuto, ma per una comprensione più precisa occorre svolgere un'operazione di interpolazione dei dati.

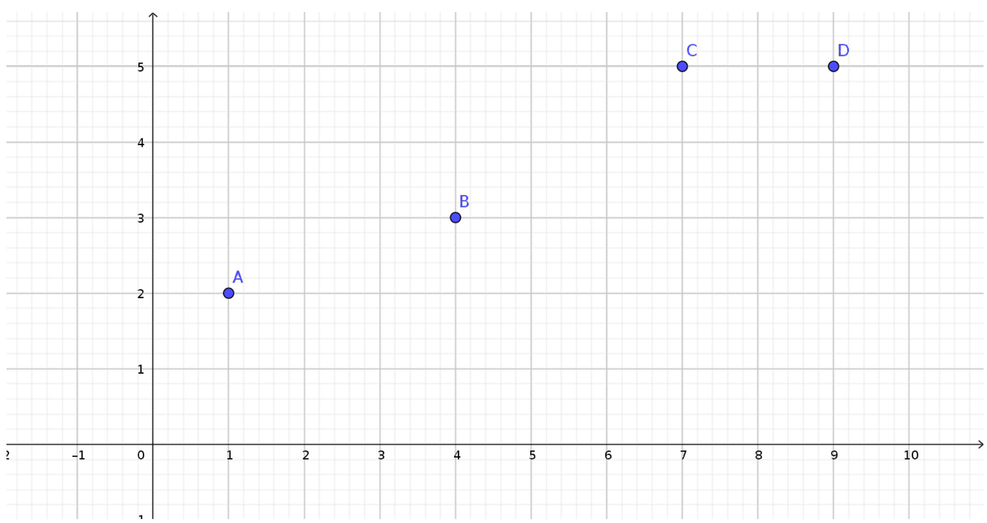
- interpolare i punti può voler dire trovare una funzione, in genere di tipo polinomiale, al cui grafico questi punti appartengano;

- tenendo conto che i dati risultanti da osservazioni o esperimenti sono spesso caratterizzati da incertezze, interpolare può voler dire che si cerca una funzione che, pur non passando necessariamente dai punti P_i li approssimi "al meglio".

Partiamo da un semplice esempio, che aiuti l'intuizione, successivamente faremo un discorso più generale.

Supponiamo di avere acquisito in un'attività sperimentale quattro coppie di dati relativi a due grandezze x e y : $A = (1, 2)$, $B = (4, 3)$, $C = (7, 5)$ e $D = (9, 5)$.

Rappresentiamo i dati in un riferimento cartesiano:



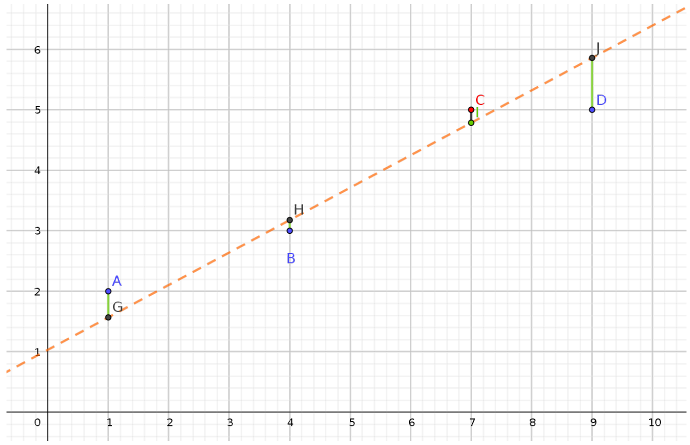
La posizione dei punti potrebbe far pensare ad una relazione lineare, ma qual è la retta rappresentativa? scelta con quale criterio?

Poniamo che vi sia una retta di equazione $y = mx + q$ che "passi attraverso" i punti. Dobbiamo trovare i valori di m e q che meglio descrivono la relazione funzionale fra le ascisse e le ordinate di questi punti.

Definito (x_i, y_i) un generico punto fra quelli dati, definisco residuo r_i la differenza fra l'ordinata del punto e l'ordinata del punto appartenente alla retta $y = mx + q$ avente la stessa ascissa:

$$r_i = y_i - mx_i - q$$

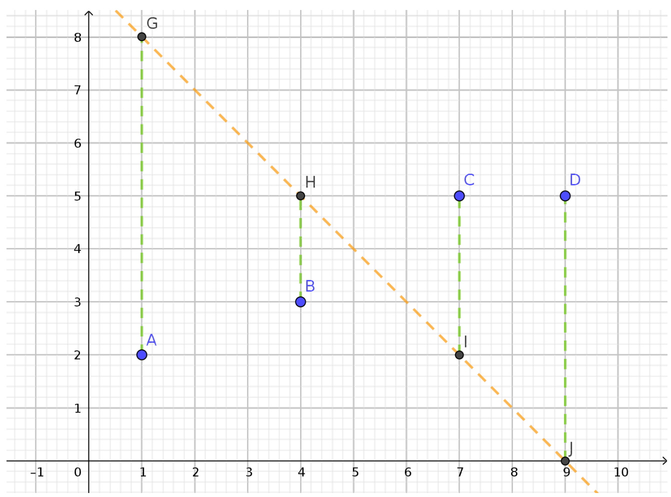
Ovviamente se il punto P_i appartiene alla retta, il relativo residuo è nullo.



Definisco residuo totale la somma dei residui di tutti i punti.

Si potrebbe pensare che la retta che meglio descrive la relazione fra le ascisse e le ordinate dei punti della distribuzione sia quella che minimizza il residuo totale.

Questa scelta però porta a risultati privi di ragionevolezza. Esaminando i punti A, B, C e D del nostro esempio, otterremmo che la retta disegnata qua sotto (in arancione per chi vede il colore) potrebbe essere la retta che meglio congiunge i punti, infatti ha residuo totale nullo. Ma è evidente a tutti che la retta disegnata non approssima assolutamente l'andamento dei punti A, B,C e D.



Il motivo di questo risultato inaccettabile deriva dal fatto che, potendo essere il residuo sia positivo che negativo, vi può essere un residuo totale molto piccolo o anche nullo pur in presenza di residui singoli ingenti, che evidenziano un discostamento significativo della retta dai punti che ad essa dovrebbero essere vicini.

Occorre quindi considerare solo residui con segno positivo: possiamo scegliere il valore assoluto, che però presenta qualche controindicazione nelle operazioni di calcolo, opure il passaggio al quadrato. Scegliamo questa seconda opzione.

Diamo quindi una nuova definizione di residuo e di residuo totale:
$r_i = (y_i - mx_i - q)^2$
$r_{tot} = \sum_{i=1}^{n} r_i = \sum_{i=1}^{n} (y_i - mx_i - q)^2$
L'obiettivo è trovare i valori di m e di q che rendano minimo r_{tot}.

Torniamo ai nostri quattro punti A , B, C e D, e calcoliamo il residuo totale r_{tot}:

$$r_{tot} = (2 - m \cdot 1 - q)^2 + (3 - m \cdot 4 - q)^2 + (5 - m \cdot 7 - q)^2 + (5 - m \cdot 9 - q)^2$$

Sviluppando i quadrati dei trinomi:
$r_{tot} = 4 + m^2 + q^2 - 4m - 4q + 2mq + 9 + 16m^2 + q^2 - 24m - 6q + 8mq + 25 + 49m^2 + q^2 - 70m - 10q + 14mq + 25 + 81m^2 + q^2 - 90m - 10q + 18mq$
(e meno male che sono solo quattro punti!)

Accorpando i termini simili, si ottiene:

$$r_{tot} = 147m^2 - 188m + 42mq + 4q^2 - 30q + 63$$

L'espressione ottenuta è quadratica nella variabile m. Vediamo che il coefficiente di m^2 è positivo, quindi, considerando q come un parametro, l'equazione rappresenta una famiglia di parabole con la concavità verso l'alto, che avranno il minimo in corrispondenza del vertice. Quindi il valore di m che corrisponde al minimo sarà:

$$m = -\frac{(-188 + 42q)}{2 \cdot 147} = \frac{94 - 21q}{147}$$

Da questa equazione si ricava la relazione:

$$147m = 94 - 21q$$

In modo analogo, possiamo pensare a r_{tot} come ad un'espressone quadratica nella variabile q , in cui m riveste il ruolo di parametro:

$$r_{tot} = 4q^2 + 42mq - 30q + 147m^2 - 188m + 63$$

Anche in questo caso la famiglia di parabole rappresentata ha la concavità rivolta verso l'alto, e quindi i valori che minimizzano r_i sono i vertici. I valori di q che corrispondono al minimo sono:

$$q = -\frac{42m - 30}{2 \cdot 4} = \frac{15 - 21m}{4}$$

Da cui si ricava la relazione:

$$4q = 15 - 21m$$

Le due relazioni che abbiamo trovato possono essere messe a sistema:

$$\begin{cases} 147m + 21q = 94 \\ 21m + 4q = 15 \end{cases}$$

Risolvendolo (per favore, non la sostituzione!) si ricava $m = \frac{61}{147} \cong 0,415$ e $q = \frac{231}{147} \cong 1,571$.

Provate adesso a scrivere le coordinate dei punti su un foglio elettronico (le ascisse in una colonna, le ordinate in un'altra colonna), tracciate il grafico e fate comparire la linea di tendenza. Qual è la sua equazione?

La retta che abbiamo trovato viene definita retta dei minimi quadrati o retta di regressione.

L'operazione che abbiamo fatto si chiama interpolazione lineare dei dati.

ESERCIZI

1. Prova a interpolare linearmente i punti $P_1 = (1,4)$, $P_2 = (3,3)$, $P_3 = (6,1)$, $P_3 = (4,2)$. Confronta il risultato con quello che dà un foglio elettronico.

2. Prova a interpolare i punti A, B, C e D con una parabola (invece di introdurre la retta $y = mx + q$ dovrai introdurre una generica parabola di equazione $y = ax^2 + bx + c$, le equazioni da mettere a sistema saranno tre...). I calcoli sono pesanti, ma puoi darli in pasto a qualche programma tipo Wolfram Alpha. Anche in questo caso puoi fare il confronto con la linea di tendenza di secondo grado data dal foglio elettronico.

RETTA DEI MINIMI QUADRATI

Generalizziamo adesso ciò che abbiamo fatto nell'esempio dei quattro punti. Se i punti sono n e si vuole trovare l'equazione della retta di equazione $y = mx + q$ che meglio li interpoli, possiamo definire il residuo totale:

$$r_{tot} = \sum_{i=1}^{n} r_i = \sum_{i=1}^{n} (y_i - mx_i - q)^2$$

e cercare per quali valori di m e n esso sia minimo.

$$\sum_{i=1}^{n} (y_i - mx_i - q)^2 = \sum_{i=1}^{n} \left(y_i^2 + m^2x_i^2 + q^2 - 2mx_iy_i - 2qy_i + 2mqx_i\right)$$

$$\sum_{i=1}^{n}\left(y_i^2+m^2x_i^2+q^2-2mx_iy_i-2qy_i+2mqx_i\right)=\sum_{i=1}^{n}y_i^2+m^2\sum_{i=1}^{n}x_i^2+\sum_{i=1}^{n}q^2-2m\sum_{i=1}^{n}x_iy_i-2q\sum_{i=1}^{n}y_i+2mq\sum_{i=1}^{n}x_i$$

Si deve tener conto che:

$$\sum_{i=1}^{n}q^2=nq^2$$

$\sum_{i=1}^{n}q^2=nq^2, \sum_{i=1}^{n}y_i=n\overline{y}$, ove con $\overline{y}$ si è denotata la media delle n ordinate dei punti. Similmente $\sum_{i=1}^{n}x_i=n\overline{x}$.

Quindi il residuo totale è $r_{tot}=\sum_{i=1}^{n}y_i^2+m^2\sum_{i=1}^{n}x_i^2+nq^2-2m\sum_{i=1}^{n}x_iy_i-2qn\overline{y}+2mqn\overline{x}$.

Ordiniamo questa espressione secondo le m decrescenti:

$$r_{tot}=m^2\sum_{i=1}^{n}x_i^2-2m\sum_{i=1}^{n}x_iy_i+2mqn\overline{x}+\sum_{i=1}^{n}y_i^2-2qn\overline{y}+nq^2$$

e, raccogliamo m fra il secondo e il terzo termine:

$$r_{tot}=m^2\sum_{i=1}^{n}x_i^2+2m\left(qn\overline{x}-\sum_{i=1}^{n}x_iy_i\right)+\sum_{i=1}^{n}y_i^2-2qn\overline{x}+nq^2$$

r_{tot}è una forma di secondo grado in cui il coefficiente di m^2è $\sum_{i=1}^{n}x_i^2$, cioè una quantità positiva. Graficamente possiamo associarla ad una parabola (ad una famiglia di parabole, se consideriamo il parametro q) con la concavità verso l'alto.

Il minimo di questa quantità si trova in corrispondenza del vertice:

$$m=-\frac{2\left(qn\overline{x}-\sum_{i=1}^{n}x_iy_i\right)}{2\sum_{i=1}^{n}x_i^2}$$

Semplificando i 2 e arrangiando diversamente questa equazione possiamo scrivere:

$$m\sum_{i=1}^{n}x_i^2+qn\overline{x}=\sum_{i=1}^{n}x_iy_i$$

In modo analogo, possiamo ordinare l'espressione di r_{tot}secondo le potenze decrescenti della variabile q, considerando m come parametro, ottenendo:

$$r_{tot}=nq^2+2mqn\overline{x}-2qn\overline{y}+m\sum_{i=1}^{n}x_i^2+2m\sum_{i=1}^{n}x_iy_i+\sum_{i=1}^{n}y_i^2$$

Poiché il coefficiente del termine quadratico è positivo, anche in questo caso possiamo pensare al minimo della parabola in corrispondenza del vertice:

$$q=-\frac{2\left(mn\overline{x}-n\overline{y}\right)}{2n}=-m\overline{x}+\overline{y}$$

Relazione che possiamo scrivere nel modo seguente:

$$m\overline{x}+q=\overline{y}$$

Le due relazioni che abbiamo trovato messe a sistema ci restituiscono i valori di$m$ e q che rendono minimo il valore di r_{tot}:

$$\begin{cases}m\sum_{i=1}^{n}x_i^2+qn\overline{x}=\sum_{i=1}^{n}x_iy_i\\ m\overline{x}+q=\overline{y}\end{cases}$$

Utilizzando il metodo di Cramer abbiamo:

$$m = \frac{\begin{vmatrix} \sum\limits_{i=1}^{n} x_i y_i & n\overline{x} \\ \overline{y} & 1 \end{vmatrix}}{\begin{vmatrix} \sum\limits_{i=1}^{n} x_i^2 & n\overline{x} \\ \overline{x} & 1 \end{vmatrix}} = \frac{\sum\limits_{i=1}^{n} x_i y_i - n\overline{x} \cdot \overline{y}}{\sum\limits_{i=1}^{n} x_i^2 - n\overline{x}^2}$$

$$q = \frac{\begin{vmatrix} \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i y_i \\ \overline{x} & \overline{y} \end{vmatrix}}{\begin{vmatrix} \sum\limits_{i=1}^{n} x_i^2 & n\overline{x} \\ \overline{x} & 1 \end{vmatrix}} = \frac{\overline{y} \cdot \sum\limits_{i=1}^{n} x_i^2 - \overline{x} \cdot \sum\limits_{i=1}^{n} x_i y_i}{\sum\limits_{i=1}^{n} x_i^2 - n\overline{x}^2}$$

Questi sono i coefficienti della retta dei minimi quadrati.

A dispetto della apparente complessità di queste formule, data una distribuzione di n punti del piano, è sufficiente calcolare $\overline{x}$, $\overline{y}$, $\sum\limits_{i=1}^{n} x_i^2$ e $\sum\limits_{i=1}^{n} x_i y_i$ per ottenere l'equazione della migliore retta interpolante.

ESERCIZI

1. Trova l'equazione della retta che meglio interpola i seguenti dati:

x	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9

2. Un esperimento di laboratorio condotto in una prima classe del liceo scientifico produce la seguente tabella, nella quale sono riportate le lunghezze del filo di un pendolo e i quadrati dei rispettivi periodi di oscillazione:

$l\ (m)$	0,610	0,632	0,660	0,722
$T^2 (s^2)$	2,46	2,59	2,69	2,92