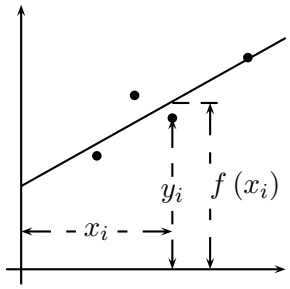
# Regressione




Obiettivo della regressione è a partire da un insieme di punti dati, trovare la funzione che meglio approssimi l'andamento dei dati. questo si può fare mediante qualsiasi tipo di funzione matematica, ma in questo caso ci si occupa della regressione mediante polinomi, in particolare polinomi di 1° grado, cioè rette.

Siano dati un insieme di $n$ punti di coordinate $P_i(x_i, y_i)$.

$$P_1 = x_1, y_1 \,,\, P_2 = x_2, y_2 \,,\, \ldots \,,\, P_n = x_n, y_n$$

Si considera la funzione approssimante $y(x) = f(x) = ax + b$ in cui i parametri incogniti sono $a$ e $b$. Per ogni punto possiamo trovare il corrispondente scarto, dato dalla relazione:

$$s_i = f(x_i) - y_i$$

Di solito si considera il quadrato di tale scarto in modo da evitare che la somma di questi scarti (funzione che vogliamo minimizzare) non diventi zero per somma di valori di segno opposto; si avrà così lo scarto della regressione dato da:

$$S = \sum_{i=1}^{n} s_i^2$$

Questo scarto è funzione dei parametri $a$ e $b$ ed è maggiore o uguale a 0, in quanto somma di quadrati:

$$S(a, b) = \sum_{i=1}^{n} [f(x_i) - y_i]^2 = \sum_{i=1}^{n} [ax_i + b - y_i]^2 \geq 0$$

Essendo una funzione di secondo grado, per trovare il minimo di questa si deve avere che entrambe le derivate prime sia contemporaneamente uguali a zero (mentre le derivate seconde non vengono considerate, in quanto è una funzione crescente, e il punto a derivata prima nulla può essere solo un minimo):

$$\min S(a, b) \quad \Rightarrow \quad \begin{cases} \dfrac{\partial S}{\partial a} = 0 = 2\displaystyle\sum_{i=1}^{n} [(ax_i + b - y_i)\, x_i] \\ \dfrac{\partial S}{\partial b} = 0 = 2\displaystyle\sum_{i=1}^{n} [(ax_i + b - y_i)\, 1] \end{cases}$$

Si ottiene quindi:

$$\begin{cases} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i & = & \sum_{i=1}^{n} x_i y_i & \Rightarrow & a\overline{x^2} + b\overline{x} & = & \overline{xy} \\ a\sum_{i=1}^{n} x_i + n\,b & = & \sum_{i=1}^{n} y_i & \Rightarrow & a\overline{x} + b & = & \overline{y} \end{cases}$$

dove la seconda parte è stata ottenuta dividendo entrambi i membri per il numero id punti considerato $n$. Moltiplicando la seconda equazione per $-\overline{x}$, valore medio di $x$, e sommandola alla prima si ottiene:

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

quindi si ottengono i valori per i parametri $a$ e $b$:

$$a = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2} \qquad b = \overline{y} - a\overline{x} = \frac{\overline{y}\left(\overline{x^2} - \overline{x}^2\right) - \overline{x}\left(\overline{xy} - \overline{x} \cdot \overline{y}\right)}{\overline{x^2} - \overline{x}^2} = \frac{\overline{y} \cdot \overline{x^2} - \overline{x} \cdot \overline{xy}}{\overline{x^2} - \overline{x}^2}$$