Régression en utilisant la méthode des moindres carrées

$$\sum_{i=1}^{n} {d_i}^2 = \sum_{i=1}^{n} (y_{icalc} - y_i)^2 = min$$

Fonction polynomiale :

$$f(x) = y = a_0 + a_1x^1 + a_2x^2 + \cdots + a_mx^m \quad (m\; l'ordre)$$

$$a_0 + a_1{x_1}^1 + a_2{x_1}^2 + a_3{x_1}^3 + \cdots + a_m{x_1}^m - y_1 = d_1$$

$$a_0 + a_1{x_2}^1 + a_2{x_2}^2 + a_3{x_2}^3 + \cdots + a_m{x_2}^m - y_2 = d_2$$

.

.

$$a_0 + a_1{x_n}^1 + a_2{x_n}^2 + a_3{x_n}^3 + \cdots + a_m{x_n}^m - y_n = d_n$$

$$\Rightarrow \textstyle\sum_{i=1}^{n} {d_i}^2 = \sum_{i=1}^{n} [a_0 + a_1{x_i}^1 + a_2{x_i}^2 + a_3{x_i}^3 + \cdots + a_m{x_i}^m - y_i]^2 = min$$

$= H(a_0, a_1, a_2, .., a_m)$ doit être minimiser ⇒ On dérive /inconnus $(a_0, a_1, a_2, .., a_m)$

$$\frac{\partial H}{\partial a_0} = 2\sum_{i=1}^{n} (a_0 + a_1{x_i}^1 + a_2{x_i}^2 + a_3{x_i}^3 + \cdots + a_m{x_i}^m - y_i) = 0$$

$$\frac{\partial H}{\partial a_1} = 2\sum_{i=1}^{n} (a_0 + a_1{x_i}^1 + a_2{x_i}^2 + a_3{x_i}^3 + \cdots + a_m{x_i}^m - y_i)x_i = 0$$

$$\frac{\partial H}{\partial a_2} = 2\sum_{i=1}^{n} (a_0 + a_1{x_i}^1 + a_2{x_i}^2 + a_3{x_i}^3 + \cdots + a_m{x_i}^m - y_i){x_i}^2 = 0$$

.

.

$$\frac{\partial H}{\partial a_m} = 2\sum_{i=1}^{n} (a_0 + a_1{x_i}^1 + a_2{x_i}^2 + a_3{x_i}^3 + \cdots + a_m{x_i}^m - y_i){x_i}^m = 0$$

$$\Rightarrow na_0 + a_1\sum_{i=1}^{n} x_i + a_2\sum_{i=1}^{n} {x_i}^2 +..+ a_m\sum_{i=1}^{n} {x_i}^m = \sum_{i=1}^{n} y_i$$

$$a_0\sum_{i=1}^{n} x_i + a_1\sum_{i=1}^{n} {x_i}^2 + a_2\sum_{i=1}^{n} {x_i}^3 +..+ a_m\sum_{i=1}^{n} {x_i}^{m+1} = \sum_{i=1}^{n} y_ix_i$$

$$a_0\sum_{i=1}^{n} {x_i}^2 + a_1\sum_{i=1}^{n} {x_i}^3 + a_2\sum_{i=1}^{n} {x_i}^4 +..+ a_m\sum_{i=1}^{n} {x_i}^{m+2} = \sum_{i=1}^{n} y_i{x_i}^2$$

.

.

$$a_0\sum_{i=1}^{n} {x_i}^m + a_1\sum_{i=1}^{n} {x_i}^{m+1} +..+ a_m\sum_{i=1}^{n} {x_i}^{2m} = \sum_{i=1}^{n} y_i.{x_i}^m$$

Procédure

1/ Choisir l'ordre le plus petit '1'

$$y = a_0 + a_1 x$$

$$\begin{cases} na_0 + a_1 \sum_{i=1}^{n} x_i = \sum y_i \ ..\boxed{1} \\ a_0 \sum x_i + a_1 \sum {x_i}^2 = \sum y_i . x_i \ ..\boxed{2} \end{cases}$$

$\boxed{1}$

$$a_0 = \frac{\sum y_i}{n} - a_1 \frac{\sum x_i}{n} \Rightarrow \boxed{a_0 = \bar{y} - a_1 \bar{x}}$$

$\boxed{2}$

$$(\bar{y} - a_1 \overline{x}) \sum x_i + a_1 \sum {x_i}^2 = \sum y_i . x_i$$

$$\bar{y}\sum x_i - a_1 \bar{x} \sum x_i + a_1 \sum (x_i)^2 = \sum x_i y_i$$

$$\frac{a_1(\sum(x_i)^2 - \bar{x}\sum x_i)}{n} = \frac{\sum x_i y_i - \bar{y}\sum x_i}{n}$$

$$a_1 \overline{(x_\iota)^2} - (\bar{x})^2 = \overline{xy} - \bar{y}\bar{x} \Rightarrow \boxed{a_1 = \frac{\overline{xy} - \bar{y}\bar{x}}{\overline{x^2} - (\bar{x})^2}}$$

2/ Evaluation

Si $(d_i)^2 = min \approx 0$ (Arrêt)

Sinon passer a l'ordre supérieur

Ordre 2 :

$$y = a_0 + a_1 x + a_2 x^2$$

$$\begin{cases} na_0 + a_1 \sum x_i + a_2 \sum (x_i)^2 = \sum y_i \\ a_0 \sum x_i + a_1 \sum x_i^2 + a_2 \Sigma (x_i)^3 = \sum x_i y_i \\ a_0 \sum (x_i)^2 + a_1 \sum x_i^3 + a_2 \sum x_i^4 = \sum x^2 y \end{cases}$$

Retour à l'évaluation.

Exercice :

Soit les mesures suivantes :

X	-1	1	2	3	5
Y	0	4	9	16	36

- Ajuster aux données une courbe de régression.

Solution :

1/

$\boldsymbol{y = ax + b \;\; ou \; y = a_1x + a_0}$

Méthode 1 :

$$a_1 = \frac{\overline{xy} - \bar{y}\bar{x}}{\overline{x^2} - (\bar{x})^2} = \frac{50 - (2 * 13)}{8 - 4} = \frac{50 - 26}{4} = 6$$

$$a_0 = \bar{y} - a_1\bar{x} = \; 13 - 6 * 2 = 1$$

Méthode 2 :

$$\begin{cases} na_0 + a_1 \sum_{i=1}^{n} x_i = \sum y_i \\ a_0 \sum x_i + a_1 \sum {x_i}^2 = \sum y_i \,.\, x_i \end{cases}$$

$$\Rightarrow \begin{cases} 5a_0 + 10a_1 = 65 \\ 10a_0 + 40a_1 = 250 \end{cases} \Rightarrow \; a_0 = 1, \; a_1 = 6$$

2/ L'évaluation :

x	y	y_{calc}	D^2
-1	0	-5	25
1	4	7	9
2	9	13	16
3	16	19	9
5	36	16	400

$\boldsymbol{\sum d^2 \gg 0}$

Donc on passe à l'ordre supérieur :

$y = a_0 + a_1x + a_2x^2$

$$\begin{cases} na_0 + a_1\sum x_i + a_2\sum(x_i)^2 = \sum y_i \\ a_0\sum x_i + a_1\sum x_i^2 + a_2\Sigma(x_i)^3 = \sum x_i y_i \\ a_0\sum(x_i)^2 + a_1\sum x_i^3 + a_2\sum x_i^4 = \sum x^2 y \end{cases}$$

$$\Rightarrow \begin{cases} 5a_0 + 10a_1 + 40a_2 = 65 \\ 10a_0 + 40a_1 + 160a_2 = 250 \\ 40a_0 + 160a_1 + 724a_2 = 1084 \end{cases} \Rightarrow \begin{cases} a_0 + 2a_1 + 8a_2 = 13 \\ a_0 + 4a_1 + 16a_2 = 25 \\ 10a_0 + 40a_1 + 181a_2 = 271 \end{cases}$$

On utilise une des méthodes de résolution de système d'équations linéaires, exp Gauss Jordan ou Pivot de Gauss.

On trouve $a_0 = 1, \; a_1 = 2, a_2 = 1$, donc $y = x^2 + 2x + 1$

L'évaluation :

x	y	y_{calc}	D^2
-1	0	0	0
1	4	4	0
2	9	9	0
3	16	16	0
5	36	36	0

Donc $\sum d^2 = 0$, Le modèle $y = x^2 + 2x + 1$ est accepté