Statistiques à deux variables
Problème de recherche : la méthode des moindres carrés

On considère le nuage de quatre points suivant :

x_i	0	1	2	3
y_i	1	3	4	8

Pour une droite d'équation $y = ax + b$, on mesure l'écart avec le nuage par la somme des carrés des écarts verticaux :

$$S(a,b) = \sum_{i=1}^{4} \left(y_i - ax_i - b\right)^2.$$

La droite des **moindres carrés** est celle qui rend $S(a,b)$ minimale.

La mission

1. Calculer S pour la droite passant par le premier et le dernier point du nuage.
2. On fixe a. Démontrer que $S(a,b)$ est minimale pour $b = \bar{y} - a\bar{x}$, où $\bar{x}$ et $\bar{y}$ sont les moyennes des x_i et des y_i. Interpréter géométriquement.
3. On prend $b = \bar{y} - a\bar{x}$. Démontrer que S est alors une fonction du second degré de a, puis déterminer la valeur de a qui la rend minimale.
4. Donner l'équation de la droite des moindres carrés et la valeur minimale de S. Comparer avec la droite d'ajustement donnée par la calculatrice.

Pour aller plus loin

Pourquoi n'utilise-t-on pas plutôt la somme des écarts $\sum(y_i - ax_i - b)$? Montrer que cette somme est nulle pour une infinité de droites.

Coups de pouce

Bloqué ? Des indices progressifs sont disponibles en ligne. N'en ouvrir qu'un à la fois, et seulement après avoir cherché !

Éléments de correction

1. Une première droite

La droite passant par $(0;1)$ et $(3;8)$ a pour équation $y = \frac{7}{3}x + 1$. Les écarts sont 0, $3 - \frac{10}{3} = -\frac{1}{3}$, $4 - \frac{17}{3} = -\frac{5}{3}$ et 0 : $S = \frac{1}{9} + \frac{25}{9} = \frac{26}{9} \approx 2{,}89$.

2. Le choix de b

Pour a fixé, on note $c_i = y_i - ax_i$. Alors $S = \sum(c_i - b)^2 = 4b^2 - 2b\sum c_i + \sum c_i^2$: c'est un polynôme du second degré en b, de coefficient dominant $4 > 0$, minimal en $b = \frac{2\sum c_i}{8} = \frac{\sum c_i}{4} = \bar{y} - a\bar{x}$. Cela signifie que $\bar{y} = a\bar{x} + b$: la meilleure droite passe par le **point moyen** $G(\bar{x};\bar{y}) = (1{,}5;4)$.

3. Le choix de a

Avec $b = \bar{y} - a\bar{x}$: $y_i - ax_i - b = (y_i - \bar{y}) - a(x_i - \bar{x}) = Y_i - aX_i$. Ici :

X_i	$-1{,}5$	$-0{,}5$	0,5	1,5	
Y_i	-3	-1	0	4	
X_iY_i	4,5	0,5	0	6	$\sum = 11$
X_i^2	2,25	0,25	0,25	2,25	$\sum = 5$

avec $\sum Y_i^2 = 9 + 1 + 0 + 16 = 26$. Donc $S = \sum(Y_i - aX_i)^2 = 5a^2 - 22a + 26$, polynôme du second degré en a, minimal en $a = \frac{22}{10} = 2{,}2$.

4. La droite des moindres carrés

$a = 2{,}2$ et $b = 4 - 2{,}2 \times 1{,}5 = 0{,}7$: $y = 2{,}2x + 0{,}7$, avec $S = 5 \times 2{,}2^2 - 22 \times 2{,}2 + 26 = 1{,}8$, contre 2,89 pour la droite de la question 1. C'est exactement la droite donnée par la calculatrice : la calculatrice utilise la méthode des moindres carrés.

Pour aller plus loin

$\sum(y_i - ax_i - b) = 4(\bar{y} - a\bar{x} - b)$ est nulle pour toutes les droites passant par G : les écarts positifs et négatifs se compensent. Les carrés empêchent cette compensation.

Enrichissement historique : Adrien-Marie Legendre publie la méthode des moindres carrés en 1805. Carl Friedrich Gauss la publie en 1809 et affirme l'utiliser depuis 1795 ; il s'en était servi pour retrouver la position de l'astéroïde Cérès, découvert en 1801.

Ce qu'on a utilisé du chapitre

- Point moyen d'un nuage, droite d'ajustement par la méthode des moindres carrés.
- Minimum d'une fonction du second degré (rappel de Première).