

Nuages de points

On a relevé la superficie x , exprimée en m^2 , et le prix de vente y , exprimé en milliers d'euros, de 8 appartements situés dans le quartier résidentiel d'une petite ville.

Superficie x	61	72	86	93	105	122	143	184
Prix y	195	210	238	264	280	300	354	415

- Quels sont les deux caractères étudiés sur la population constituée des 8 appartements ?
- (a) Dans un repère orthogonal (unités graphiques : 1cm pour 20m^2 en abscisses et 1cm pour 50 milliers d'euros en ordonnée), placer les points $M_1(x_1, y_1), M_2(x_2, y_2), \dots, M_8(x_8, y_8)$ associés à chacun de ces 8 appartements.
(b) Quelle remarque peut-on faire sur la position de ces 8 points ?
- (a) Calculer \bar{x} et \bar{y} , moyennes respectives des superficies et des prix de vente des 8 appartements.
(b) Placer sur le graphique le point $G(\bar{x}, \bar{y})$. Où est-il situé par rapport aux points du nuage ? Ce point s'appelle le point moyen du nuage de points de la série statistique.

Méthode des moindres carrés

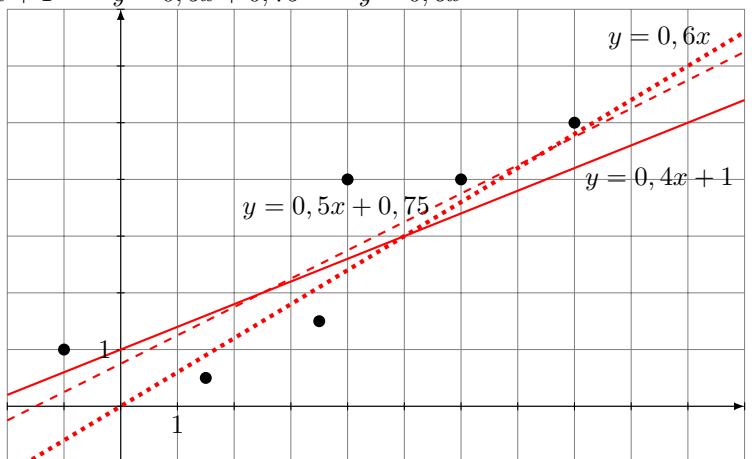
On décide de représenter graphiquement le nuage de points associé à la série statistique ci-dessous :

X	-1	1,5	3,5	4	6	8
Y	1	0,5	1,5	4	4	5

Un ajustement affine est une droite passant au plus près de tous les points.

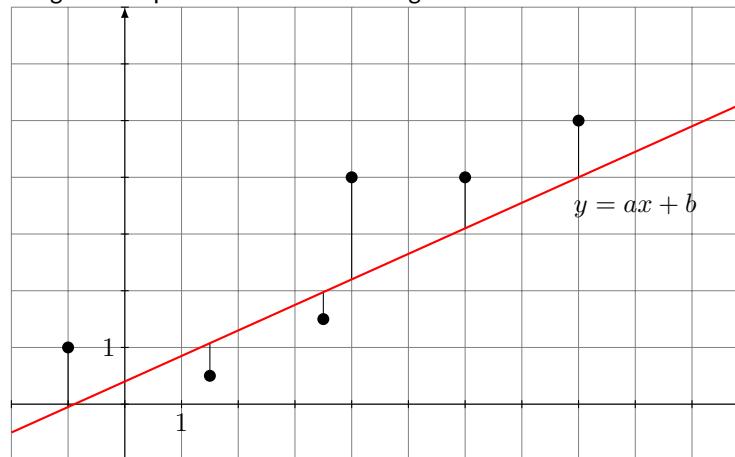
- Trois élèves proposent pour ajustement affine trois droites différentes :

$$y = 0,4x + 1 \quad y = 0,5x + 0,75 \quad y = 0,6x$$



- L'une de ces droites semble-t-elle donner un meilleur ajustement que les autres ?
- Proposer un ajustement qui semble encore meilleur.

- Pour estimer la qualité d'un ajustement affine, on utilise la méthode dite des moindres carrés. Pour cela, on calcule la somme des carrés des écarts entre les ordonnées des points du nuage et des points de la droite. Il s'agit ensuite de minimiser cette somme.



- Expliquer pourquoi le carré de l'écart des ordonnées entre le point d'abscisse x_i de la droite d'équation $y = ax + b$ et un point $(x_i; y_i)$ du nuage de points s'exprime par $(y_i - (ax_i + b))^2$
 - On veut écrire une fonction en langage Python, appelée `qualite`, qui renvoie la somme
- $$\sum_{i=1}^n (y_i - (ax_i + b))^2 = (y_1 - (ax_1 + b))^2 + (y_2 - (ax_2 + b))^2 + \dots + (y_n - (ax_n + b))^2$$
- lorsqu'on donne en paramètres des valeurs pour a et b . Compléter le programme suivant pour cela :

```
def qualite(a,b):
    X = [-1, 1.5, 3.5, 4, 6, 8]
    Y = [1, 0.5, 1.5, 4, 4, 5]
    S = 0
    for i in range(len(X)):
        S = S + (.....)**2
    return S
```

- Exécuter ce programme en l'appliquant aux équations de la question 1. Selon ce critère, quelle est la droite qui donne le meilleur ajustement ?