Ajustement affine d'un nuage de points

Définition

On considère deux séries statistiques de même ffectif $x_1, x_2, \ldots x_n$ et $y_1, y_2, \ldots y_n$.
On appelle M_i le point de coordonnées $(x_i\, y_i)$ pour $1 \leqslant i \leqslant n$.
On appelle nuage de points l'ensemble des points correspondants.

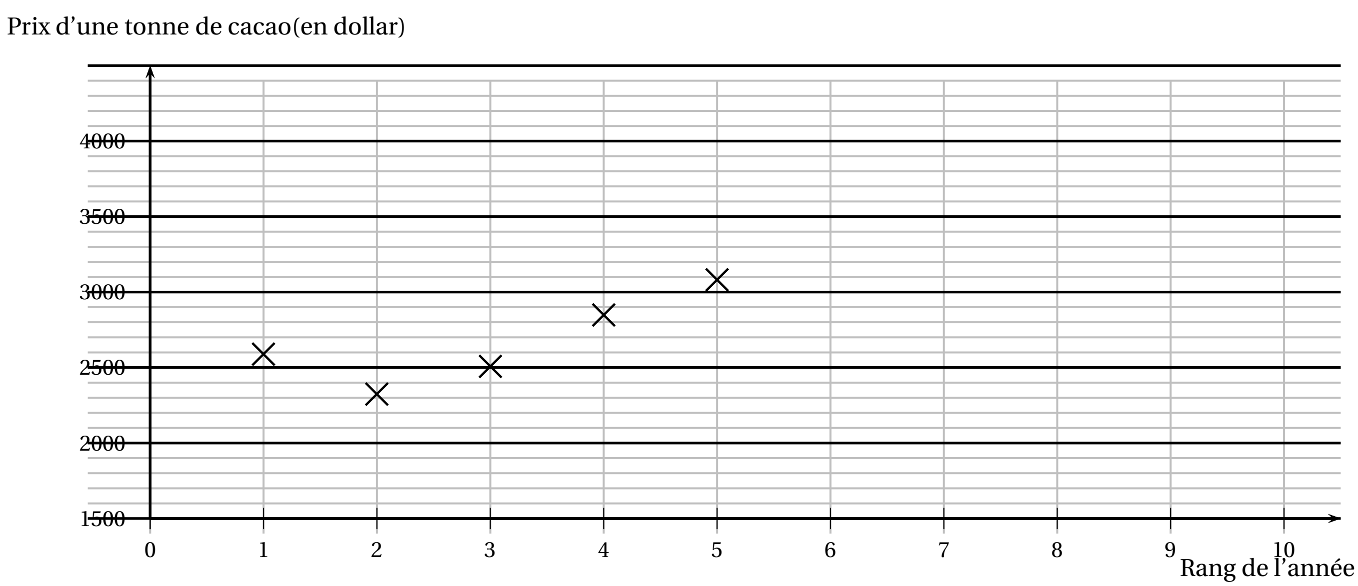
Exemple, d'après Bac Métropole STMG juin 2017 :

Le tableau suivant donne le prix moyen en dollar US de la tonne du cacao en provenance de la Côte d'Ivoire au 1er janvier des années 2011 à 2015.

Année	2011	2012	2013	2014	2015
Rang de l'année : x_i	1	2	3	4	5
Prix (en dollar) d'une tonne de cacao : y_i	2 589,70	2 324,85	2 507,55	2 847,85	3 081,45

Source : INSEE

Représenter le nuage de points de coordonnées $(x_i\ ;\ y_i)$, pour i variant de 1 à 5, est représenté ci-dessous.



1. Non, ces points ne sont pas alignés .
2. On peut considèrer que ces points sont à peu près alignés ; dans la suite, on va voir une méthode pour trouver une droite qui passe « proche » de tous les points, qu'on appelle droite d'ajustement affine.

Méthode de Mayer pour l'ajustement affine

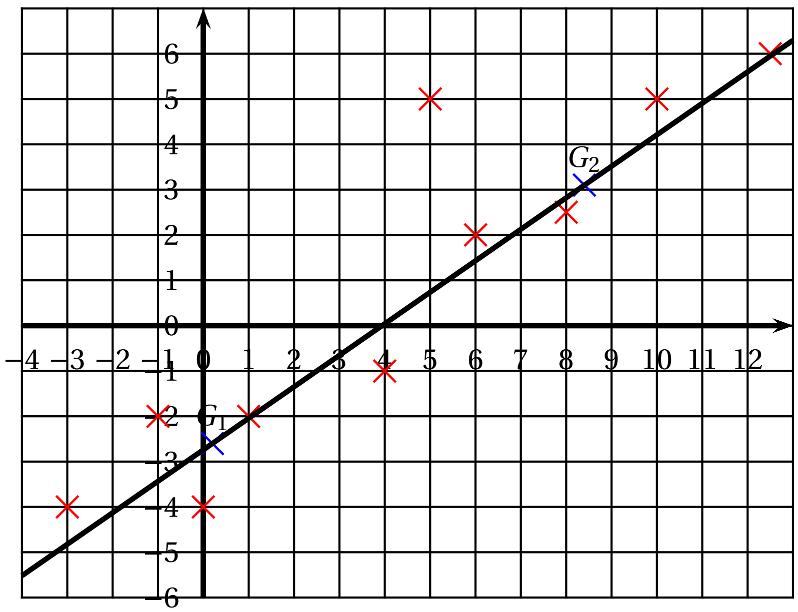
Après avoir classé les points du nuage de points dans l'ordre croissant des abscisses, on les sépare en deux groupes de taille égale (dans le cas ou le nombre de points est impair, un groupe aura un point de plus).
On détermine le **point moyen** (c'est-à-dire le point dont l'abscisse est la moyenne des valeurs en abscisse et l'ordonnée est la moyenne des valeurs en ordonnée) de chacun des deux groupes et on les place sur le graphique.
Pour finir, on trace la droite passant par les points moyens G_1 et G_2 pour obtenir la **droite d'ajustement de Mayer** .

Exemple : soient les deux séries données dans le tableau ci-dessous :

x_i	-3	-1	0	1	4	5,5	6	8	10	12,5
y_i	-4	-2	-4	-2	-1	0	2	2,5	5	6

On sépare la série statistique en deux séries de cinq points.

1. Placer les points sur le graphique ci-dessous.



2. • L'abscisse du premier point moyen est $x_{G_1} = \dfrac{-3+(-1)+0+1+4}{5} = \dfrac{1}{5} = 0,2$. L'ordonnée du premier point moyen est $y_{G_1} = \dfrac{-4+(-2)+(-4)+(-2)+(-1)}{5} = \dfrac{-13}{5} = -2,6$.
   Donc $\boxed{G_1\,(0,2\,;\,-2,6)}$

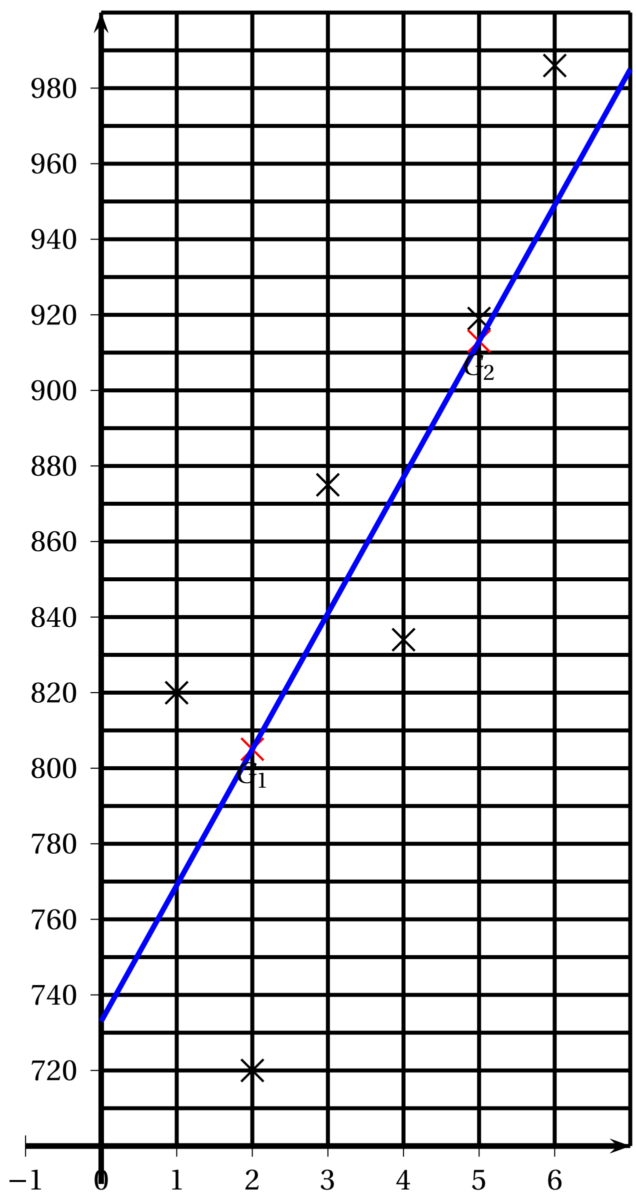
   De même : $x_{G_2} = \dfrac{5,5+6+8+10+12,5}{5} = \dfrac{42}{5} = 8,4$ et $y_{G_2} = \dfrac{0+2+2,5+5+6}{5} = \dfrac{15,5}{5} = 3,1$.
   Donc $\boxed{G_2(8,4\,;\,3,1)}$
3. Voir graphique.
4. L'équation de la droite de Mayer est de la forme $y = ax + b$.
   Le coefficient directeur est $a = \dfrac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = \dfrac{3,1-(-2,6)}{8,4-0,2} = \dfrac{5,7}{8,2} \approx 0,695$.
   Le droite passe par G_1 (ou G_2).
   Les coordonnées de G_1 vérifient cette équation.
   $-2,6 = 0,695 \times 0,2 + b$ donc $b = -2,6 - 0,695 \times 0,2 = -2,739$.
   L'équation de la droite de Mayer est alors : $\boxed{y = 0,695 - 2,739}$

Exercice I

Voici le nombre d'entrées faites par un musée pendant les six premiers mois de l'année.

x_i	1	2	3	4	5	6
y_i	820	720	875	834	919	986

1. Nuage de points :



2. • Le premier point moyen a pour coordonnées : $G_1\,(2\ ;\ 805)$.
   • Le deuxième point moyen a pour coordonnées : $G_2\,(5\ ;\ 913)$.
3. Par lecture graphique, on peut estimer le nombre de visiteurs le septième jour à 984 environ.
4. Trouver l'équation de la droite de Mayer.
   Le coefficient directeur de la droite de Mayer est : $a = \dfrac{913-805}{5-2} = \dfrac{108}{3} = 36$.
   L'équation de cette droite est $y = 36x + b$.
   Elle passe par G_1 donc $805 = 36 \times 2 + b$ d'où $b = 805 - 36 \times 2 = 805 - 72 = 733$.
   L'équation est $\boxed{y = 36x + 733}$.
   Pour $x = 7$, on trouve $36 \times 7 + 733 = \boxed{985}$.