

Chapitre 4- Séries statistiques à deux variables quantitatives

Table des matières

Chapitre 4 - Séries statistiques à deux variables quantitatives	1
1. Le nuage de points.....	2
2. Le point moyen	3
3. L'ajustement affine	4
3.1. La droite « au jugé » ou droite de Mayer.....	4
3.2. La méthode des moindres carrés.....	5
4. Interpolation et extrapolation.....	5

L'objectif de ce chapitre est d'étudier les éventuelles relations entre **deux variables statistiques**.

Quand on étudie une entreprise, on a souvent envie de savoir si deux grandeurs sont liées : le budget publicité influence-t-il le chiffre d'affaires ? L'ancienneté d'un salarié influence-t-elle son salaire ? Le nombre d'heures de formation influence-t-il la productivité ? C'est exactement ce que permettent d'étudier les séries statistiques à deux variables quantitatives : on observe deux nombres (x et y) pour chaque individu (un mois, un salarié, un produit...), et on cherche s'il existe une tendance générale, un lien approximativement linéaire entre les deux.

Une **série statistique à deux variables** $(x_i ; y_i)$ est constituée d'une liste de n couples de valeurs $(x_1 ; y_1) ; (x_2 ; y_2) ; \dots ; (x_n ; y_n)$.

1. Le nuage de points

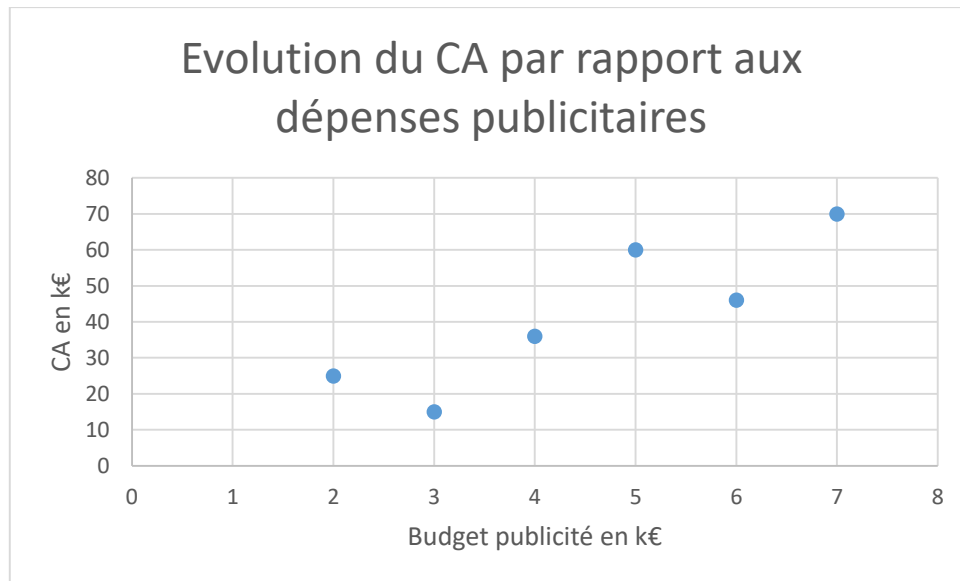
Pour chaque individu i , on relève un couple de valeurs $(x_i ; y_i)$. Si on place chaque couple sur un repère (x en abscisse, y en ordonnée), on obtient un nuage de points.

De ce fait, le **nuage de points** de cette série est l'ensemble des points du plan de coordonnées $(x_1 ; y_1) ; (x_2 ; y_2) ; \dots ; (x_n ; y_n)$.

Application pratique :

Une entreprise a relevé, sur 6 mois, son budget publicité x_i (en k€) et le chiffre d'affaires y_i réalisé (en k€) :

Mois	1	2	3	4	5	6
Budget pub. x_i (k€)	2	3	4	5	6	7
Chiffre d'affaires y_i (k€)	25	15	36	60	46	70



En observant ce nuage de points, on peut vérifier s'il y a **dépendance ou indépendance** entre les deux variables. Effectivement on observe une liaison forte et positive que l'on peut mesurer grâce au **coefficient de corrélation (hors programme)**.

Dans notre situation, on voit que les points sont globalement alignés : plus le budget publicité augmente, plus le chiffre d'affaires semble élevé. On va chercher à résumer cette tendance par une droite : c'est l'ajustement affine (partie 3).

2. Le point moyen

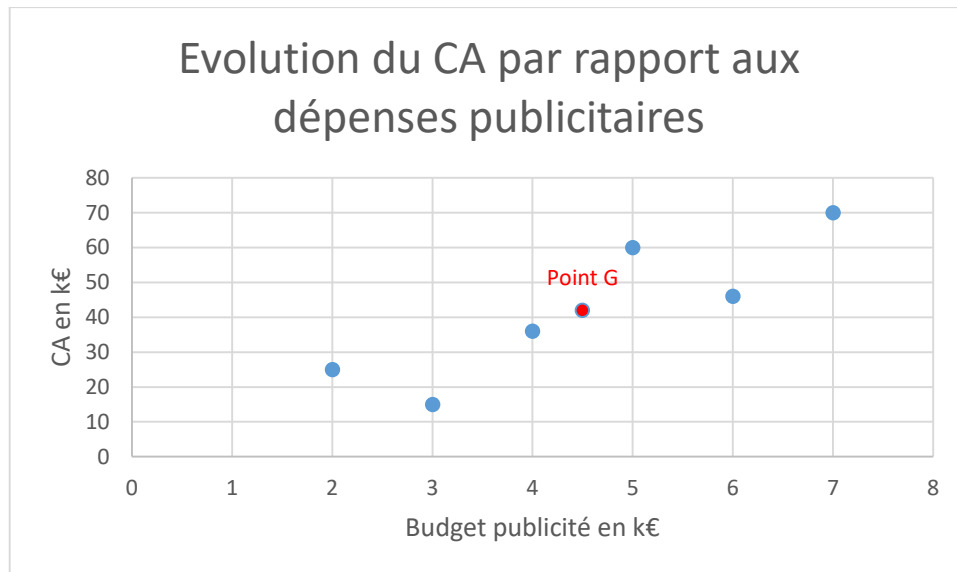
Le point moyen $G(\bar{x} ; \bar{y})$ est le « centre de gravité » du nuage de points. Ses coordonnées sont simplement les moyennes des x_i et des y_i :

$$\bar{x} = (x_1 + x_2 + \dots + x_n) / n \quad \bar{y} = (y_1 + y_2 + \dots + y_n) / n$$

Sur notre exemple :

$$\bar{x} = (2+3+4+5+6+7) / 6 = 4,5 \quad \bar{y} = (25+15+36+60+46+70) / 6 = 42$$

Propriété importante : la future droite d'ajustement doit toujours passer par le point moyen G. C'est une bonne façon de vérifier qu'on n'a pas fait d'erreur de calcul.



3. L'ajustement affine

Ajuster le nuage, c'est trouver l'équation d'une droite $y = ax + b$ qui "résume" au mieux le nuage de points. Il existe plusieurs méthodes ; aucune démonstration mathématique n'est exigée, seule l'utilisation compte.

3.1. La droite « au jugé » ou droite de Mayer

Ces deux méthodes sont rapides, mais approximatives : deux personnes différentes n'obtiendront pas exactement la même droite.

- **Droite au jugé** : on trace à la main, à l'œil, une droite qui semble passer « au milieu » du nuage.
- **Droite de Mayer** : méthode un peu plus rigoureuse — on partage le nuage en deux groupes de tailles égales (les premiers points / les derniers points), on calcule le point moyen de chaque groupe, puis on trace la droite qui relie ces deux points moyens.

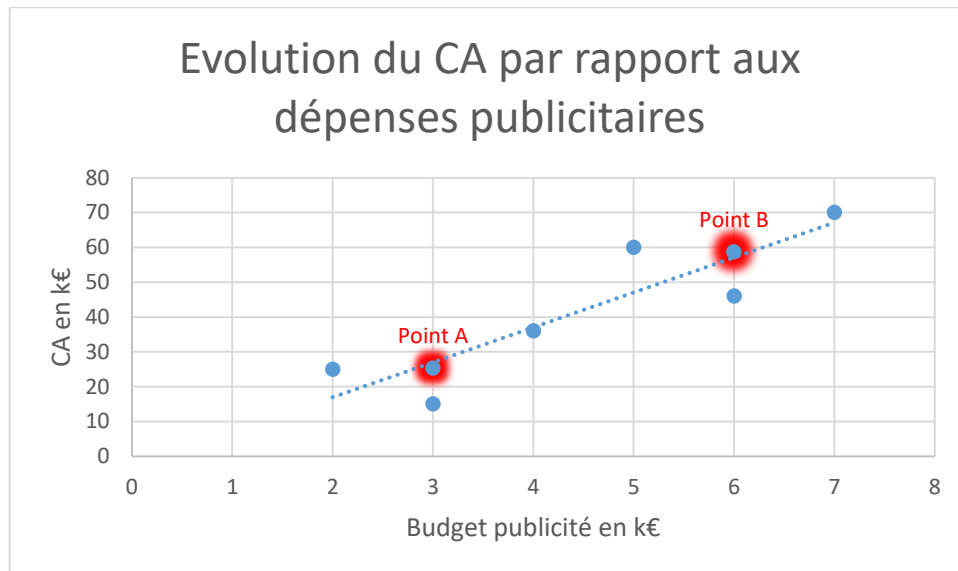
Mois	1	2	3	4	5	6
Budget pub. x_i (k€)	2	3	4	5	6	7
Chiffre d'affaires y_i (k€)	25	15	36	60	46	70

Point A : les mois 1, 2 et 3

$$\bar{x} = (2+3+4) / 3 = 3 \quad \bar{y} = (25+15+36) / 3 = 25,33$$

Point B : les mois 4, 5 et 6

$$\bar{x} = (5+6+7) / 3 = 6 \quad \bar{y} = (60+46+70) / 3 = 58,67$$



3.2. La méthode des moindres carrés

C'est la méthode la plus utilisée, car elle donne un résultat unique. Elle consiste à choisir la droite $y = ax + b$ qui minimise la somme des carrés des écarts verticaux entre les points observés et la droite :

$$\text{Minimiser } \sum_i (y_i - (a \cdot x_i + b))^2$$

En pratique, on ne fait jamais ce calcul « à la main » : on utilise la calculatrice (régression linéaire) ou un tableur.

Avec un tableau (Excel, google sheets...)

- =PENTE(plage_y ; plage_x) → donne le coefficient directeur a
- =ORDONNEE.ORIGINE(plage_y ; plage_x) → donne l'ordonnée à l'origine b
- Clic droit sur le nuage de points → Ajouter une courbe de tendance → cocher « Afficher l'équation » : Excel affiche directement $y = ax + b$

Sur notre exemple, le tableur donne (arrondi) :

$$y = 9,71x - 1,97$$

4. Interpolation et extrapolation

Une fois l'équation de la droite trouvée, on peut estimer des valeurs de y pour des x qu'on n'a pas observés.

- Interpoler : estimer une valeur de x qui se situe à l'intérieur de la plage des données observées (ici, entre 2 et 7). C'est relativement fiable, car on reste dans une zone où le lien linéaire a été vérifié.

- Extrapoler : estimer une valeur de x qui se situe en dehors de la plage observée (par exemple $x = 15$). C'est plus risqué : rien ne garantit que la tendance observée entre 2 et 7 continue au-delà.

Esprit critique : un ajustement affine n'est qu'un modèle simplifié. Une extrapolation trop lointaine peut donner des résultats absurdes (chiffre d'affaires négatif, croissance infinie...). Il faut toujours questionner la pertinence du résultat au regard du contexte réel de gestion.

Application pratique :

Suite de l'exemple, on vous demande de retrouver $x = 10$

$$y = 9,71 \times 10 - 1,97 = 95,13$$

Avec un budget de 10k€, le CA estimé serait de 95,13k€.