

Chapitre 3 : Statistiques à deux variables quantitatives.

Eliot ROY

2022-2023

Points du programme.

Contenus.

- Nuage de points associé à une série statistique à deux variables quantitatives.
- Ajustement affine.

Capacités attendues.

- Représenter un nuage de points.
- Déterminer et utiliser un ajustement affine pour interpoler ou extrapoler des valeurs inconnues.
- Représenter un nuage de points en effectuant un changement de variable donné (par exemple u^2 , $\frac{1}{t}$, $\frac{1}{\sqrt{n}}$, $\log(y)$...) afin de conjecturer une relation de linéarité entre deux nouvelles variables.

Commentaires.

- Les ajustements affines peuvent être réalisés graphiquement au jugé : l'appréciation de leur qualité peut faire l'objet d'une discussion au sein de la classe.
- La méthode des moindres carrés est présentée : recherche d'une droite d'équation $y = ax + b$ réalisant le minimum de $\Sigma(y_i - (ax_i + b))^2$ pour le nuage de points (x_i, y_i) .
- Privilégier des contextes réels en lien avec les enseignements de spécialité.
- Les élèves sont entraînés à exercer leur esprit critique sur la pertinence, au regard des données et de la situation étudiée, d'une modélisation par ajustement affine et sur les limites des extrapolations faites dans ce cadre.

Algorithmique.

- Automatiser le calcul de $\Sigma(y_i - (ax_i + b))^2$.
- Rechercher un couple (a, b) minimisant cette expression parmi un ensemble fini de couples proposés par les élèves ou générés par balayage, tirage aléatoire...

I Définition.

Définition 1 : Une série statistique à deux variables X et Y est la donnée d'une liste de couples $((x_1, y_1), (x_2, y_2), \dots, (x_N, y_N))$. Pour chaque $i \in \llbracket 1, N \rrbracket$, x_i est la valeur du caractère X de l'individu i , et y_i est la valeur du caractère Y de l'individu i .

La série statistique à deux variables peut être rangée dans un tableau :

Valeurs des x_i	x_1	x_2	...	x_N
Valeurs des y_i	y_1	y_2	...	y_N

Exemple : On étudie la taille (en cm) et la masse (en kg) d'un groupe de 10 élèves du Lycée Valentine Labbé. On obtient l'observation suivante sur l'échantillon de 10 personnes :

Taille (en cm)	178	175	182	183	177	182	185	190	183	182
Masse (en kg)	71	71	78	75	70	78	78	82	82	77

Remarque : Lorsque, dans une population, on n'étudie qu'un seul caractère dont les valeurs sont relevées à des dates différentes, on obtient une série statistique à deux variables, où le temps est l'une des deux variables. Une telle série est dite chronologique.

Une série statistique à deux variables est aussi appelée une série statistique double.

II Représentation d'un nuage de points.

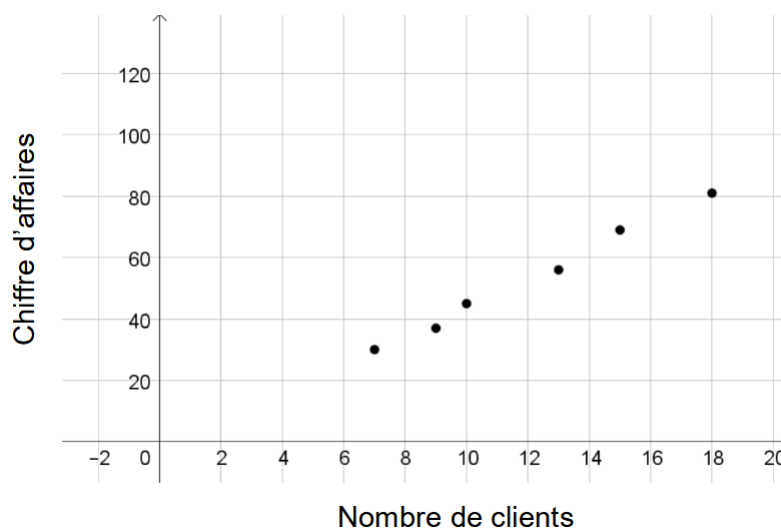
Définition 2 : Considérons une série statistiques $((x_1, y_1), (x_2, y_2), \dots, (x_N, y_N))$.

Dans un repère du plan, l'ensemble des points $M_1(x_1, y_1), \dots, M_N(x_N, y_N)$ constitue le **nuage de points** associé à cette série statistique.

Exemple : Le tableau suivant présente l'évolution du nombre de clients et du chiffre d'affaires d'une micro-entreprise (en milliers d'euros) au cours des six derniers mois :

Nombre de clients	7	9	10	13	15	18
Chiffre d'affaires	30	37	45	56	69	81

On peut tracer le nuage de points associé à cette série statistique :



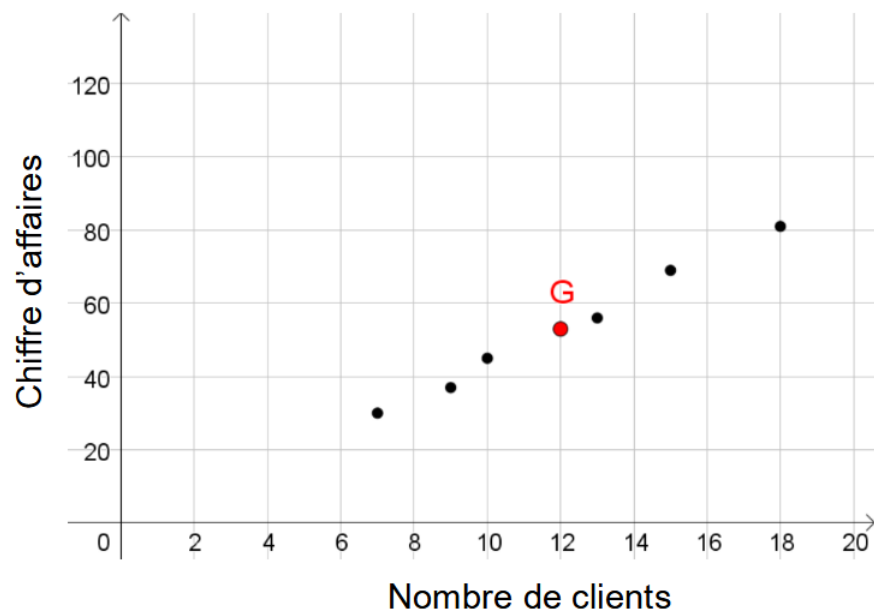
Définition 3 : Le **point moyen** du nuage de points de cette série statistique est le point $G(\bar{x}, \bar{y})$ où $\bar{x}$ est la moyenne de la série statistique $(x_1, x_2, \dots, x_N)$ et $\bar{y}$ est la moyenne de la série statistique $(y_1, y_2, \dots, y_N)$.

Exemple : Reprenons l'exemple précédent sur le chiffre d'affaires d'une micro-entreprise.

Le point moyen G a pour coordonnées $(12; 53)$:

$$\bar{x} = \frac{7+9+10+13+15+18}{6} = 12.$$

$$\bar{y} = \frac{30+37+45+56+69+81}{6} = 53.$$



Objectifs : Représenter un nuage de points.
Exercices 19, 20, 21, 23 page 109.
Exercices 38, 39, 40, 41, 42 page 112.

III Ajustement affine.

III.1 Méthode des moindres carrés.

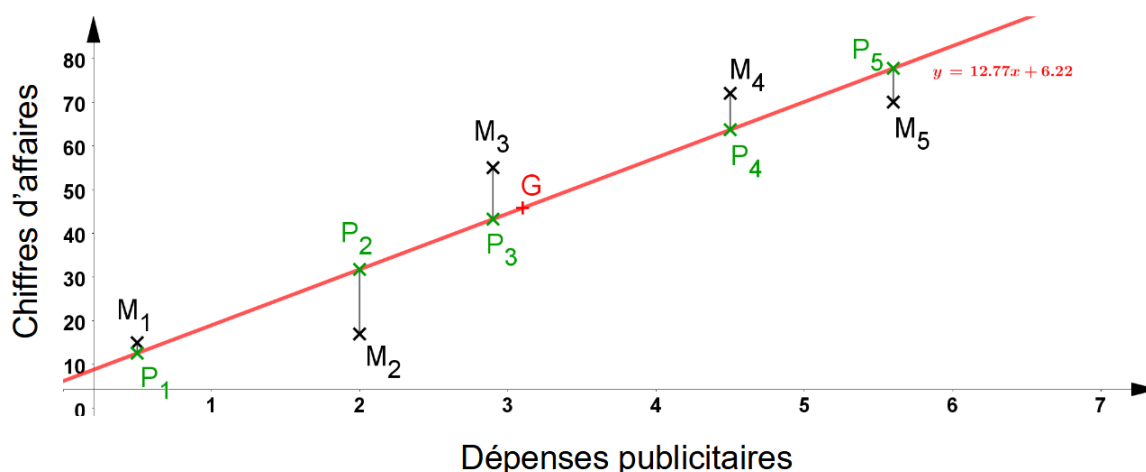
Considérons le nuage de points correspondant à la série statistique double $((x_1, y_1), (x_2, y_2), \dots, (x_N, y_N))$, où y est représenté en fonction de x . Un ajustement affine du nuage de points consiste à approcher le mieux possible le nuage de points grâce à une droite affine. On utilise pour cela la méthode des moindres carrés.

Définition 4 : La droite des moindres carrés de la série $((x_1, y_1), (x_2, y_2), \dots, (x_N, y_N))$ est la droite d'équation $y = ax + b$ où a et b sont choisis de manière à minimiser la quantité $\sum_{k=1}^N (y_k - (ax_k + b))^2$.

Exemple : Voici les résultats d'une enquête réalisée auprès d'une entreprise spécialisée dans l'électroménager. Les données sont exprimées en millions d'euros.

Dépenses publicitaires	0,5	2	2,9	4,5	5,6
Chiffres d'affaires	15	17	55	72	70

Sur le graphique ci-contre, on a représenté le nuage de points correspondant à cette série statistique, ainsi que le point moyen $G(3, 1; 45, 8)$.



Pour tracer la droite des moindres carrés, on a cherché à passer au plus près de chacun des points M_i , en minimisant la somme des carrés des écarts entre chaque point M_i et P_i . Autrement dit, on a cherché à rendre minimale la quantité $S = M_1P_1^2 + M_2P_2^2 + \dots + M_NP_N^2$.

$M_3P_3^2 + M_4P_4^2 + M_5P_5^2$. La droite affine qui minimise cette quantité est la droite d'équation $y = 12,77x + 6,22$.

Remarque : Les coefficient a et b de la droite des moindres carrés $y = ax + b$ s'obtiennent à l'aide de la calculatrice ou du tableur.

Propriété 5 : La droite d'ajustement de y en x par la méthode des moindres carrés passe par le point moyen du nuage de points de la série statistique double.

Objectifs : Déterminer un ajustement affine.
Exercices 24, 25, 27, 28 page 110.
Exercice 43 page 112 Exercices 44, 46 page 113.

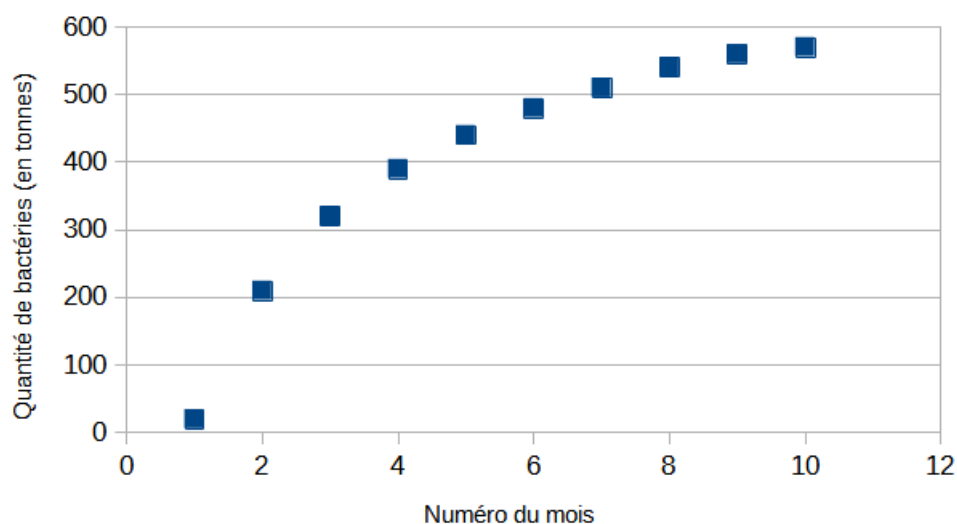
III.2 Ajustement par changement de variable.

Dans certains cas, il est nécessaire de réaliser un changement de variables avant de procéder à l'ajustement affine.

Exemple : Chez un producteur d'huîtres, une bactérie appelée *Vibrio estuarianus* est apparue à partir du mois d'août 2019. Le tableau ci-dessous donne la quantité y_i (exprimée en tonnes) d'huîtres affectées par cette bactérie, en fonction de x_i qui représente le numéro du mois depuis l'apparition de la bactérie. Le numéro 1 correspond au mois d'août 2019, le numéro 2 correspond au mois de septembre 2019, et ainsi de suite.

x_i	1	2	3	4	5	6	7	8	9	10
y_i	20	210	320	390	440	480	510	540	560	570

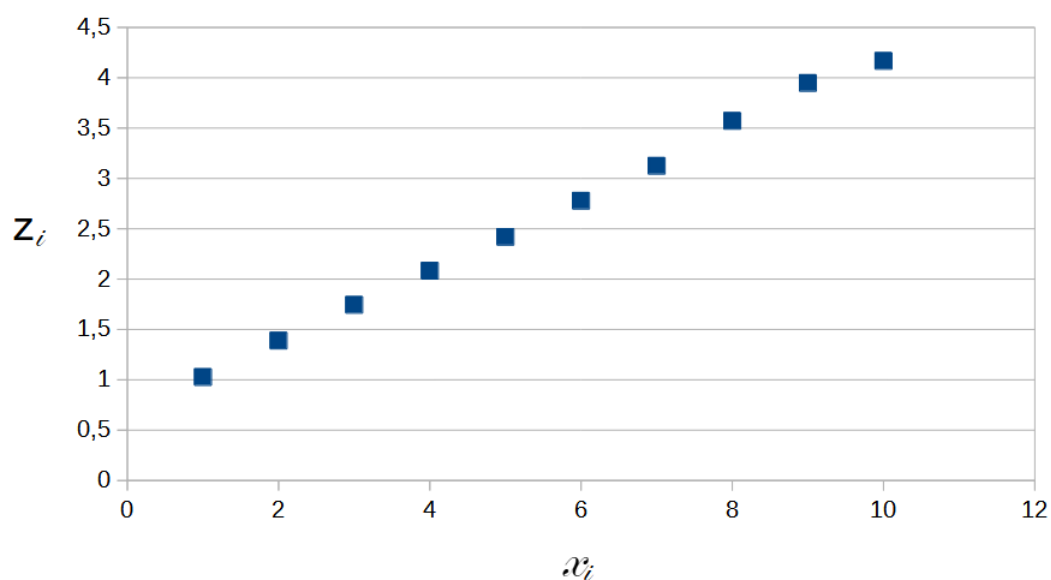
En représentant graphiquement ce nuage de points, on obtient le graphique ci-dessous :



La représentation graphique ne témoigne pas d'une croissance affine : on va donc procéder à un changement de variable, en posant $z_i = \frac{750}{750 - y_i}$.

On obtient alors le tableau et le graphique ci-dessous :

x_i	1	2	3	4	5	6	7	8	9	10
z_i	1,03	1,39	1,74	2,08	2,42	2,78	3,13	3,57	3,95	4,17



La nouvelle représentation graphique est alors mieux adaptée pour réaliser un ajustement affine.

Objectifs : Représenter un nuage de points en effectuant un changement de variables.

Exercices 36, 37 page 111.

III.3 Interpolation et extrapolation.

Définition 6 : A partir de l'ajustement affine, on peut faire des estimations par :

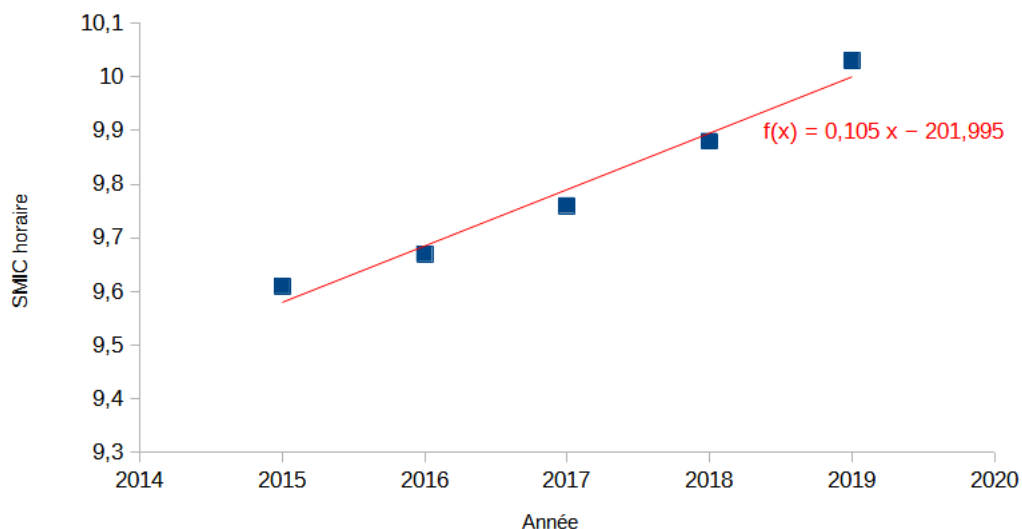
- **interpolation**, lorsqu'on estime une valeur inconnue dans le domaine d'étude ;
- **extrapolation**, lorsqu'on estime une valeur inconnue en dehors du domaine d'étude fourni par l'échantillon.

Exemple : Le tableau ci-dessous donne l'évolution du SMIC horaire brut de 2015 à 2019.

Année	2015	2016	2017	2018	2019
SMIC horaire (en euros)	9,61	9,67	9,76	9,88	10,03

On souhaite estimer le SMIC horaire brut pour l'année 2025, qui ne figure pas dans la série statistique. On va donc procéder par extrapolation en s'aidant d'un ajustement affine du nuage de points. On considérera que l'ajustement affine est une approximation satisfaisante de l'évolution du SMIC jusqu'en 2030.

A l'aide d'un tableur, on obtient le graphique ci-dessous :



La droite d'ajustement a pour équation $y = 0,105x - 201,995$.

Pour estimer la valeur qu'aura le SMIC en 2025, on calcule donc la valeur de y pour $x = 2025$:

$y = 0,105 \times 2025 - 201,995 = 10,63$: le SMIC horaire brut est donc estimé à 10,63 € en 2025.

Objectifs : Utiliser un ajustement affine pour interpoler ou extrapoler.
Exercices 29, 31 page 110 Exercice 32 page 111 Exercice 47 page 113
Exercices 33, 34 page 111 Exercice 48 page 113 Exercices 49, 50, 51, 52 page 114