CHAP 02 Statistiques à deux variables

Compétences

- Représenter un nuage de points et déterminer le point moyen d'une série statistique à deux variables
- Réaliser un ajustement affine par la méthode des moindres carrés, à la calculatrice
- Utiliser un ajustement pour effectuer des interpolations ou des extrapolations
- Changer de variable pour linéariser un nuage de points de forme non affine

Fil rouge *coloration drone*

Pendant un vol, le pilote relève à intervalles réguliers la durée de vol écoulée et le niveau de batterie restant du drone.
Peut-on modéliser cette relation par une fonction affine, afin de prédire l'autonomie restante du drone à tout instant du vol ?

I. Intro

Idée

Existe-t-il un **lien** (statistique, c'est-à-dire *mathématique*, et non de *cause à effet*) entre deux caractères d'une population ? Pour cela, on étudie *simultanément* les deux caractères et on présente les résultats sous forme de tableaux ou de graphiques.

II. Tableaux de données, nuages de points

II.1. Exemples

Exemple 2.1

Une société a mis au point un nouveau matériel destiné aux PME de logistique et mène un enquête dans la région de Provence-Alpes-Côte d'Azur auprès de 500 entreprises aptes à recevoir ce matériel, pour déterminer à quel prix chacune de ces entreprises accepterait d'acquérir ce nouveau matériel :

Prix proposé en milliers d'euros x_i	40	36	32	28	24	20	16	12	10	8
Nb d'entreprises disposées à acheter à ce prix y_i	60	70	130	210	240	340	390	420	440	500

Exemple 2.2

Pour des véhicules légers de la gamme 9 à 11 chevaux de puissance fiscale, roulant en palier (ou en descente), on a relevé les consommations moyenne et les vitesses correspondantes :

Vitesse en km/h x_i	10	20	30	40	50	60	70	80	90
Consommation en L/100 km y_i	16,5	11,5	9	7,5	6,8	6,6	7	7,5	9

Exemple 2.3

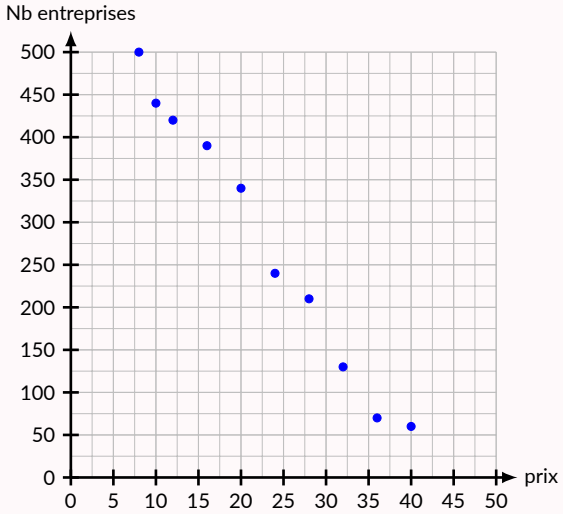
Les chiffres d'affaires trimestriels d'une grande entreprise ont été, pour les 12 derniers trimestres :

Rang du trimestre x_i	1	2	3	4	5	6	7	8	9	10	11	12
C.A. en millions d'euros y_i	300	450	130	200	280	410	200	250	320	500	210	250

Licence Ouverte Etalab 2.0

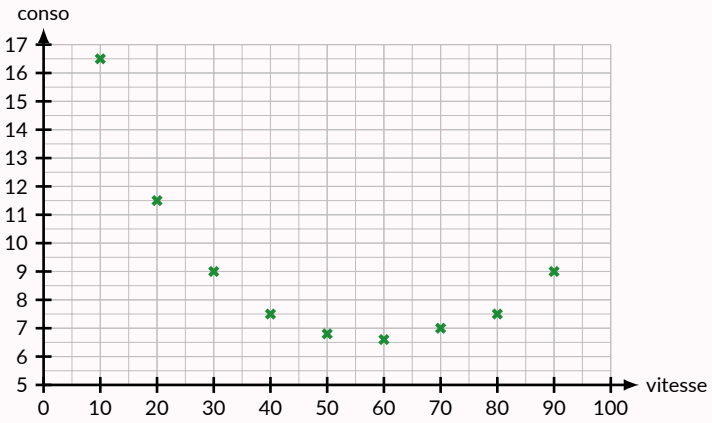
▸ Illustration 2.1

Si on place dans un repère les points $M_i(x_i\ ;\ y_i)$, le graphique obtenu est appelé nuage de points de la série statistique (on ne doit pas relier ces points !).



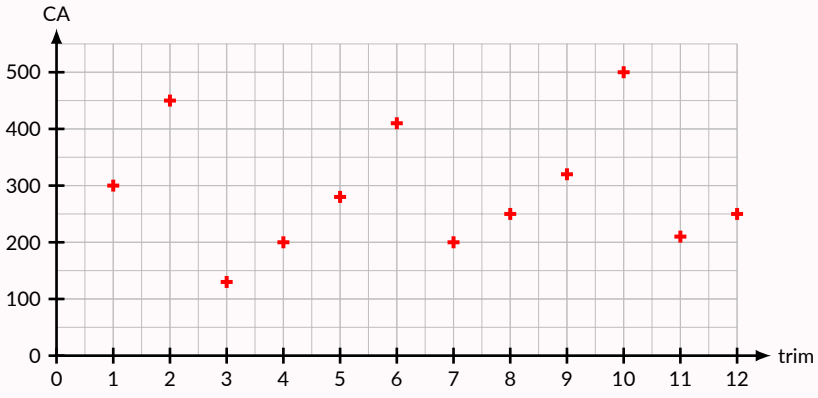
▸ Illustration 2.2

On obtient le nuage de points suivants :



▸ Illustration 2.3

On obtient le nuage de points suivants :



▸ Remarques 2.1

- Le nuage de l'exemple 1 a une forme allongée.
- Le nuage de l'exemple 2 a la forme d'une parabole.
- Le nuage de l'exemple 3 est constitué de points dispersés partout dans le plan.

Licence Ouverte Etalab 2.0

II.2. Ajustement d'un nuage

▸ **Intro**

Le nuage étant dessiné, on peut essayer de trouver une fonction f dont la courbe passe « le plus près possible » des points du nuage : c'est le problème de l'**ajustement**.
Lorsqu'une telle fonction f existe on dit qu'il y a une **corrélation** entre les deux caractères.

▸ **Exemples 2.4**

— Le nuage de l'exemple 1 pourrait être ajusté par une droite.

— Le nuage de l'exemple 2 pourrait être ajusté par une parabole.

— Le nuage de l'exemple 3 ne peut être ajusté tel qu'il est par une courbe.

II.3. Point moyen

▸ **Définition 2.1**

Le point moyen d'un nuage de points $M_i(x_i\,;y_i)$ est le point $G(\overline{x}\,;\overline{y})$.

▸ **Exemples 2.5**

$G_1(22{,}6\,;280)$; $G_2(50\,;9{,}04)$ et $G_3(6{,}5\,;291{,}67)$.

▸ **Calculatrice 2.1**

Les calculatrices (menu CALC STATS) permettent de déterminer les coordonnées du point moyen, via CALC $\overline{x}$ et CALC $\overline{y}$.

III. Ajustement affine

III.1. Au « jugé »

▸ **Méthode/Propriété 3.1**

On peut tracer, au **jugé** une droite qui passe « au plus près » des points du nuage.
On démontre que les « meilleures » droites d'ajustement sont celles qui passent par le point moyen du nuage.

III.2. Par la méthode des moindres carrés

▸ **Méthode/Propriété 3.2**

On utilise la calculatrice ou un logiciel qui donne :

— le coefficient de corrélation r qui permet de contrôler la validité de l'ajustement qui ($|r|$ « proche » de 1) ;

— l'équation réduite de la droite de régression de y en x sous la forme $y = ax + b$.

On vérifie que le point moyen est bien sur la droite d'ajustement.

▸ **Calculatrice 3.1**

Les calculatrices (menu CALC STATS) permettent de déterminer une équation de la droite de régression, via (en général) l'option CALC `linreg`.

▸ **Méthode pratique**

Une fois l'équation de la droite des moindres carrés obtenue, il *suffit* de déterminer deux points de cette droite (on rappelle qu'elle passe par le point moyen !) pour la tracer.
On peut essayer de prendre comme abscisses de ces points les valeurs minimale et maximale de la variable X.

Licence Ouverte Etalab 2.0

III.3. Utilisation de l'ajustement

▸ Méthode et utilisation

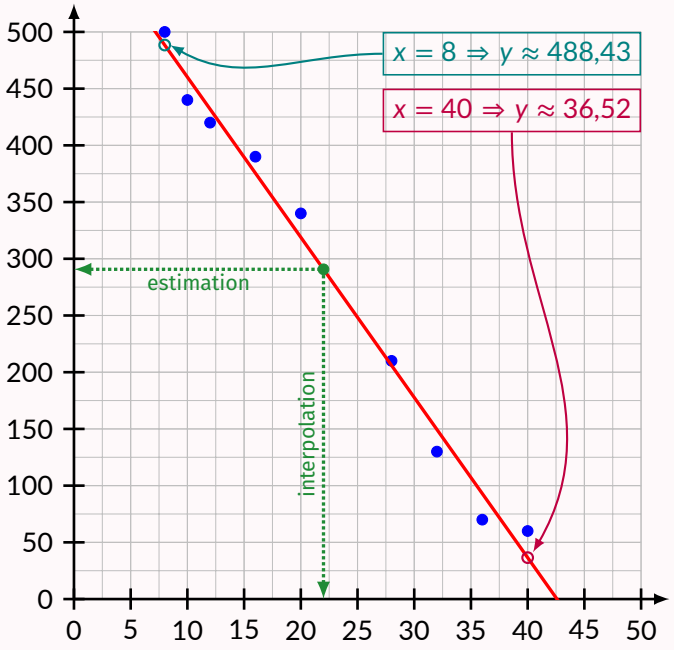
On peut utiliser l'ajustement pour faire des prévisions (« futur ») ou faire de l'interpolation (« passé »).

▸ Illustration 3.1

On considère les données :

Valeurs x	40	36	32	28	20	16	12	10	8
Valeurs y	60	70	130	210	340	390	420	440	500

La calculatrice donne $\begin{cases} a \approx -14{,}12 \\ b \approx 601{,}41 \end{cases}$. Donc, pour $x = 22$, $y \approx a \times 22 + b \approx 290{,}72$.



III.4. « Transformation » d'un nuage, changement de variable(s)

▸ Idée

On peut « transformer » l'un des deux paramètres (ou les deux) de la série à l'aide d'une certaine fonction :

- on obtient un nouveau nuage de points (généralement de forme *allongée*) ;
- on détermine l'équation de la droite d'ajustement ;
- on revient aux variables initiales (en *retournant* la fonction).

▸ Exemples 3.1

On transforme avec « ln », on retourne avec « exp » (et inversement).
On transforme avec « $\sqrt{}$ », on retourne avec « ^{2} » (et inversement).

Licence Ouverte Etalab 2.0