

Statistiques à deux variables

Prélude - Historique et utilité

En 1805, Adrien-Marie Legendre publie la méthode des moindres carrés pour calculer l'orbite des comètes; Carl Friedrich Gauss affirme l'utiliser depuis 1795. Cette méthode permet de trouver la droite qui « colle » le mieux à un nuage de points. En gestion, on l'utilise pour prévoir des ventes, des coûts ou une demande à partir des données passées.

I Nuage de points et point moyen

Définition

Une **série statistique à deux variables** est une liste de couples $(x_i ; y_i)$. Dans un repère, les points $M_i(x_i ; y_i)$ forment le **nuage de points**.

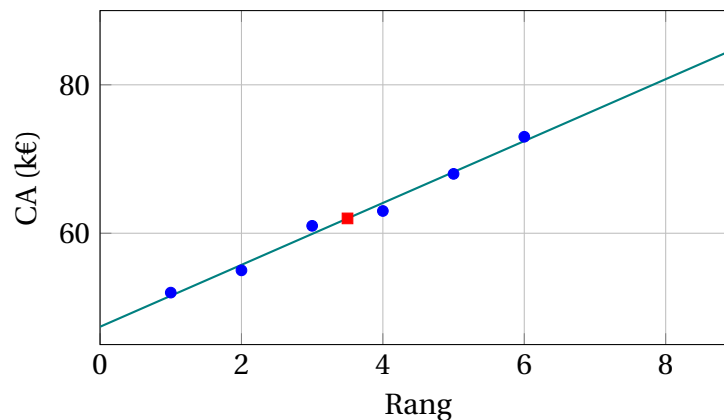
Le **point moyen** est $G(\bar{x} ; \bar{y})$.

Exemple

Chiffre d'affaires d'une PME (en milliers d'euros) :

Année	2019	2020	2021	2022	2023	2024
Rang x_i	1	2	3	4	5	6
CA y_i	52	55	61	63	68	73

$$\bar{x} = 3,5 \text{ et } \bar{y} = \frac{372}{6} = 62 : G(3,5 ; 62).$$



II Ajustement affine

Définition

Lorsque le nuage a une forme « allongée » proche d'une droite, on réalise un **ajustement affine** : on approche le nuage par une droite d'équation $y = ax + b$.

La calculatrice ou le tableur fournit la **droite des moindres carrés** (droite de régression de y en x), qui passe par le point moyen G .

Exemple

Pour la PME, la calculatrice donne $y = 4,17x + 47,4$. On vérifie : $4,17 \times 3,5 + 47,4 \approx 62$: la droite passe par G.

- **Interpolation** : estimation à l'intérieur des données.
- **Extrapolation** : prévision en dehors des données. En 2027 ($x = 9$), on prévoit $4,17 \times 9 + 47,4 \approx 84,9$ milliers d'euros.

III Changement de variable

Méthode

Si le nuage de points $(x_i ; y_i)$ n'est pas allongé selon une droite, on peut parfois transformer une variable (par exemple $z = \frac{1}{x}$, $z = \sqrt{x}$, $z = x^2$ ou $z = \log(y)$) pour obtenir un nuage $(z_i ; y_i)$ proche d'une droite. On réalise alors un ajustement affine $y = az + b$, puis on revient à la variable de départ.

Exemple

Le coût moyen de fabrication y (en €) d'un objet selon la quantité produite x (en centaines) :

x_i	1	2	4	5	10
y_i	62	36	23	20,5	15
$z_i = \frac{1}{x_i}$	1	0,5	0,25	0,2	0,1

Le nuage $(x_i ; y_i)$ est courbé, mais le nuage $(z_i ; y_i)$ est presque aligné. L'ajustement donne $y \approx 52z + 10$, soit $y \approx \frac{52}{x} + 10$.

Pour 800 objets ($x = 8$) : $y \approx \frac{52}{8} + 10 = 16,5$ €.

Remarque

Une corrélation entre deux variables ne signifie pas que l'une est la cause de l'autre : les ventes de glaces et les coups de soleil augmentent ensemble... à cause du beau temps.