

CORRÉLATION ET CAUSALITÉ

— TERMINALE — MATHS COMPLÉMENTAIRES —

1. STATISTIQUES À DEUX VARIABLES

Une série à deux variables est une liste de couples $(x_i; y_i)$, $1 \leq i \leq n$.

- **nuage de points** : les points $M_i(x_i; y_i)$;
- **point moyen** : $G(\bar{x}; \bar{y})$ avec $\bar{x} = \frac{1}{n} \sum x_i$, $\bar{y} = \frac{1}{n} \sum y_i$.

$$V(x) = \frac{1}{n} \sum (x_i - \bar{x})^2$$
$$\text{cov}(x, y) = \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})$$

La covariance est positive quand x et y varient plutôt dans le même sens.

2. AJUSTEMENT AFFINE

Droite des moindres carrés

La droite $y = ax + b$ qui minimise $\sum (y_i - ax_i - b)^2$:

$$a = \frac{\text{cov}(x, y)}{V(x)} \quad b = \bar{y} - a\bar{x}$$

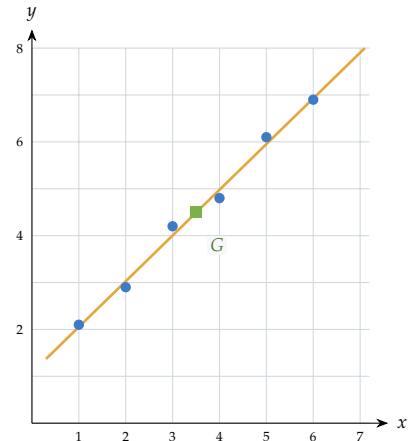
Elle **pass**e par le point moyen G .

Coefficient de corrélation linéaire

$$r = \frac{\text{cov}(x, y)}{\sigma(x)\sigma(y)} \in [-1; 1]$$

$|r|$ proche de 1 (en pratique $|r| \geq 0,9$) : ajustement pertinent. Le signe de r est celui de a .

Exemple : $G(3,5; 4,5)$, $y = 0,98x + 1,08$, $r \approx 0,997$. Prédiction année 8 : 8,9 M€ (extrapolation).



Corrélation n'est pas causalité

Un $|r|$ proche de 1 montre que les variables évoluent **ensemble**, pas que l'une **provoque** l'autre. Méfiance envers un **facteur caché** (glaces et noyades : la chaleur).

3. AJUSTEMENT ET CHANGEMENT DE VARIABLE

Nuage d'allure **exponentielle** (croissance en pourcentage) : on pose $z_i = \ln y_i$. Si le nuage $(x_i; z_i)$ est rectiligne, on l'ajuste par $z = ax + b$, puis :

$$y = e^{ax+b} = k e^{ax} \quad \text{avec } k = e^b$$

e^a est le coefficient multiplicateur par unité de x . D'autres changements sont possibles : $X = x^2$ (distance de freinage $d \approx 0,006 v^2$), $X = \frac{1}{x}$...

Exemple : culture de bactéries

t_i	0	1	2	3	4	5
N_i	2,3	3,1	4,6	6,8	9,9	14,6
z_i	0,83	1,13	1,53	1,92	2,29	2,68

$z \approx 0,375t + 0,793$ ($r \approx 0,999$), donc $N \approx 2,21 e^{0,375t}$.
 $e^{0,375} \approx 1,45$: +45 % par heure ; à $t = 6$: $N \approx 20,9$.

4. MÉTHODES TYPES

Ajustement affine

1. calculer le point moyen G ;
2. a et b à la calculatrice ;
3. vérifier $|r|$ proche de 1 ;
4. prévoir en remplaçant x .

Ajustement exponentiel

1. poser $z_i = \ln y_i$;
2. ajuster $z = ax + b$;
3. revenir à $y = e^b e^{ax}$;
4. interpréter e^a (taux).

Corrélation ou causalité ?

1. une corrélation forte ne prouve rien ;
2. chercher un facteur caché ;
3. seule une expérience contrôlée établit une cause.

À RETENIR

🕒 La droite des moindres carrés passe par $G(\bar{x}; \bar{y})$.

📐 $a = \frac{\text{cov}(x, y)}{V(x)}$,
 $b = \bar{y} - a\bar{x}$.

📊 $|r|$ proche de 1 :
ajustement pertinent.

📈 Allure exponentielle :
 $z = \ln y$, puis
 $y = e^b e^{ax}$.

💡 Corrélation $\neq$ causalité ;
extrapolation prudente.