

Répéter en solo

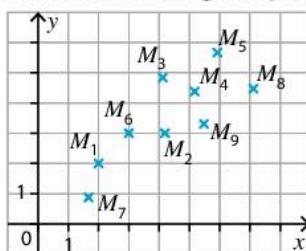
Carte mentale

Représenter une série statistique à deux variables

- Deux variables statistiques x et y sont étudiées simultanément.
- Pour i allant de 1 à n , la variable x prend les valeurs x_i et la variable y prend les valeurs y_i .

variable x	x_1	x_2	x_n
variable y	y_1	y_2	y_n

- Dans un repère $(O ; I, J)$, l'ensemble des points $M_i(x_i ; y_i)$ constitue un **nuage de points**.



- Le point moyen du nuage de points est :

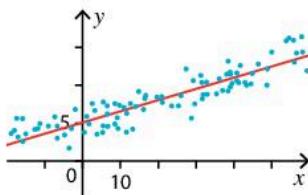
$$G\left(\frac{x_1 + x_2 + \dots + x_n}{n} ; \frac{y_1 + y_2 + \dots + y_n}{n}\right),$$

soit $G(\bar{x} ; \bar{y})$.

Séries statistiques à deux variables

Réaliser un ajustement affine

- Réaliser, lorsque cela est possible, un ajustement affine (droite de régression) signifie tracer une droite qui « approche au mieux » tous les points du nuage.



Le point moyen $G(\bar{x} ; \bar{y})$ appartient à la droite de régression.

- La **droite des moindres carrés** est un ajustement affine d'équation $y = mx + p$, avec $m = \frac{\text{cov}(x ; y)}{\sigma^2(x)}$ et $p = \bar{y} - m\bar{x}$.

Étudier la corrélation et la causalité

- Coefficient de corrélation linéaire r :

$$r = \frac{\text{cov}(x ; y)}{\sigma(x)\sigma(y)} \text{ et } -1 \leq r \leq 1$$

avec $\text{cov}(x ; y)$ la **covariance des variables x et y** .

$$\begin{aligned} \text{cov}(x ; y) &= \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{n} \\ &= \frac{x_1 y_1 + x_2 y_2 + \dots + x_n y_n - \bar{x} \bar{y}}{n} \end{aligned}$$

- Lorsque r est très proche de 1 ou de -1, la corrélation entre les séries x et y est forte.
- Lorsque r est proche de 0, le nuage de points ne peut pas être « ajusté au mieux » par une droite.
- Une forte corrélation ne signifie pas qu'il y a un lien de cause à effet (**causalité**) entre les valeurs des deux séries.



Réaliser un autre type d'ajustement

Lorsque les points ne paraissent pas globalement alignés et lorsque le coefficient de corrélation linéaire des séries x et y est proche de 0, le nuage de points ne peut pas être « ajusté au mieux » par une droite. Il se peut qu'une autre courbe puisse l'ajuster au mieux.

On procède pour cela à un changement de variable pour déterminer l'équation de la courbe.

