# SÉRIES STATISTIQUES A DEUX VARIABLES

→ Il s'agit d'étudier simultanément DEUX caractères d'une "population".

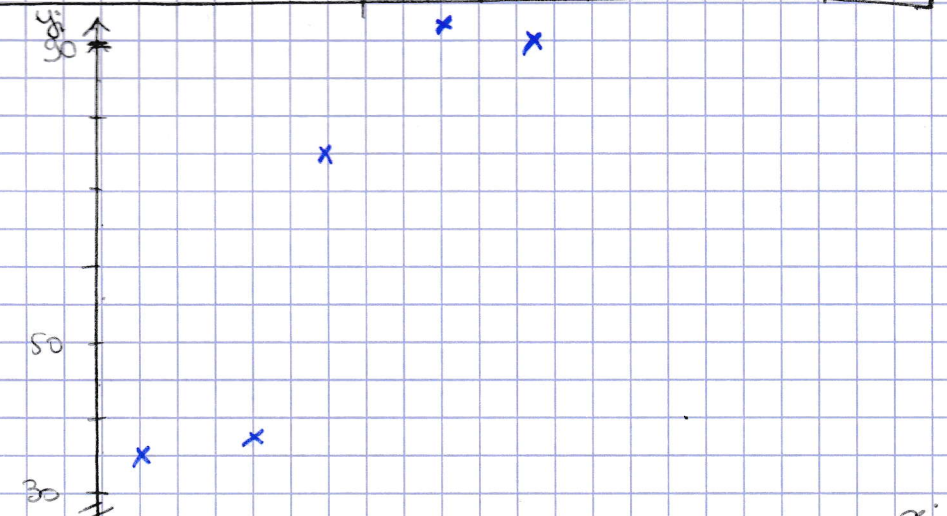
## I. NUAGE DE POINTS

Exemple : Une grande surface s'intéresse au lien existant entre ses dépenses de publicité et son CA.

Elle réalise les observations suivantes :

| | | | | | |
|---|---|---|---|---|---|
| Dépenses de pub (k€) $x_i$ | 0,5 | 2 | 2,9 | 4,5 | 5,6 |
| CA (k€) $y_i$ | 35 | 37 | 75 | 92 | 90 |

Graphiquement :




↳ L'ensemble des points $M_i$ de coordonnées $(x_i ; y_i)$ constitue le NUAGE de points représentant la série statistique.

Une fois le nuage dessiné, on cherche à trouver une fonction $f$ telle que la courbe d'équation $y = f(x)$ passe "le plus près possible" des points du nuage → c'est le problème de l'AJUSTEMENT

Ici, il semble qu'on puisse approcher le nuage par une droite (car le nuage est aplati/allongé) → Ajustement AFFINE

Mais on peut avoir aussi un ajustement parabolique, logarithmique, exponentiel .... ou ne pas en avoir !

⚠ Corrélation ≠ causalité.
(évoluer conjointement)

**Définition** : On appelle POINT MOYEN d'un nuage de $n$ points $M_i$ de coordonnées $(x_i, y_i)$, le point $G$ de coordonnées

$$\boxed{(x_G, y_G) = (\bar{x}, \bar{y})}$$

où $\bar{x}$ = moyenne des $x_i$
$\bar{y}$ = moyenne des $y_i$ → à déterminer à la calculatrice

ex précédent : $G : (3,1 ; 65,8)$

# II. AJUSTEMENT AFFINE

## A. Méthodes graphiques

### 1) Ajustement "à la règle"

→ Imprécis ⇒ Non utilisé !

### 2) Méthode de MAYER

Non utilisée ! Principe :

→ on divise le nuage en deux "sous-nuages"

→ on calcule les 2 points moyens $G_1$ et $G_2$

→ La droite de Mayer est celle qui relie $G_1$ et $G_2$

$(G_1G_2)$ a pour équation $y = ax + b$

avec $a = \dfrac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}}$ et $b$ obtenu en remplaçant les coordonnées de $G_1$ et $G_2$

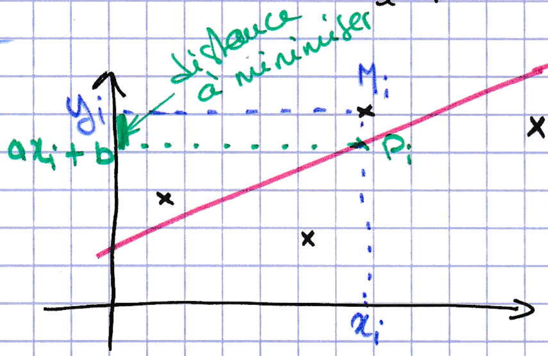
Cette méthode manque aussi de précision.

## B. Méthode des moindres carrés

### 1- Principe

→ On cherche la droite qui pourrait remplacer "au mieux" le nuage de points, càd d'être la + proche possible de tous les points.

Déf° : On appelle DROITE de RÉGRESSION de $y$ en $x$ la droite D telle que la somme $\sum_{i=1}^{n} M_iP_i^2 = \sum_{i=1}^{n} [y_i - (ax_i+b)]^2$ soit minimale.




$M_i$ = observations, de coordonnées $(x_i, y_i)$

$P_i$ = Points théoriques, de coord $(x_i ; ax_i+b)$

→ Distance entre $M_i$ et $P_i$ $= M_iP_i$

→ Comme certaines sont <0, on prend les carrés

→ On veut les distances les + petites = les MOINDRES CARRÉS (d'où le nom de la méthode)

Beaucoup + rare :

Déf° : On appelle Droite de régression de $x$ en $y$ la droite D' telle que la somme $\sum_{i=1}^{n} M_iQ_i = \sum_{i=1}^{n} [x_i - (ay_i+b)]^2$ soit minimale

(Principe inverse : on veut minimiser les écarts horizontaux)

### 2- Equation de la droite de régression

On montre que la droite de régression D de $y$ en $x$ a pour équation $y = ax+b$ avec $a = \frac{\sigma_{xy}}{(\sigma_x)^2} = \frac{\text{Covariance de } x \text{ et } y}{\text{Variance de } x}$

et $\bar{y} = a\bar{x} + b$.

EN PRATIQUE, a et b sont DONNÉS par le menu statistiques de la calculatrice. (⇒ formules inutiles !)

Pour la régression de $x$ en $y$ : équation $x = a'y+b'$ → on inverse $x$ et $y$ dans la calculatrice.

Dans l'exemple du début du cours $a \simeq 12,77$ et $b \simeq 26,22$,
donc on a $y = 12,77x + 26,22$.

## C _ Coefficient de corrélation linéaire

→ Il permet d'apprécier la qualité d'un ajustement affine.
Un tel ajustement est justifié lorsque le nuage est aplati/allongé.

On a $\boxed{r = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y}}$   EN PRATIQUE, il est donné par la calculatrice.

Propriétés : • $r \in \mathbb{R}$, il est du même signe que a (si $<0$, corrélation négative, la droite "descend", et si $r>0$, corrélation positive, et la droite "monte")

• On a TOUJOURS $\boxed{-1 \leq r \leq 1}$

• Il existe une "bonne" corrélation entre x et y lorsque $|r|$ est "suffisamment" voisin de 1.

• $r^2 = aa'$

Remarques : – selon les secteurs, on choisit pour quelles valeurs de r la corrélation est jugée suffisante pour effectuer un ajustement affine par les moindres carrés (c'est à dire à la calculatrice).
– Forte corrélation ≠ causalité.