
1 Nuages de points

Définition.

Lorsqu'on étudie deux caractères x et y liés à une même population de taille n , on associe à chaque individu un couple $(x_i; y_i)$ et on présente les résultats dans un tableau. On appelle série statistique double l'ensemble des couples $(x_i; y_i)$ pour i allant de 1 à n .

Exemple.

On relève pendant plusieurs jours la température et l'altitude grâce à un ballon sonde. On obtient :

| | | | | | | | |
|-----------------------|-----|-----|-----|------|------|---|-----|
| Altitude en km | 0,1 | 0,4 | 0,7 | 1 | 1,5 | 2 | 2,5 |
| Température en degrés | 25 | 22 | 20 | 17,5 | 10,5 | 5 | 1 |

Définition.

On appelle nuage de points associé à une série statistique à deux variables $(x; y)$ l'ensemble des points $M_i(x_i; y_i)$

Définition.

On appelle point moyen le point $G(\bar{x}; \bar{y})$ avec \bar{x} moyenne de la série statistique x et \bar{y} moyenne de la série statistique y

Exemple.

Calculer le point moyen de la série précédente.

2 Ajustement affine

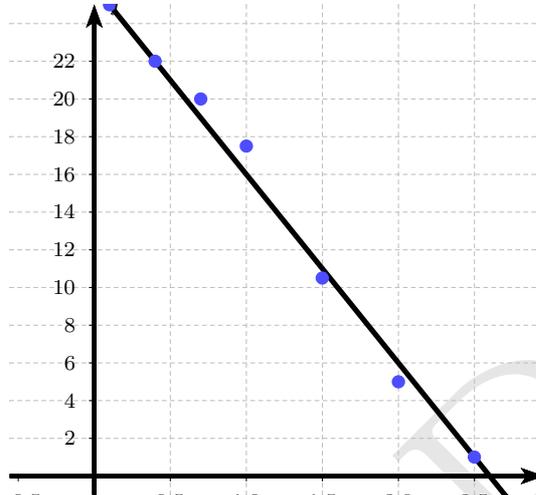
2.1 Droite des moindres carrés

Définition.

Appliquer un ajustement affine, c'est trouver la droite qui se rapproche le plus d'un maximum de points du nuage.

Exemple.

Quand on représente les points de l'exemple précédent, on obtient :



On remarque que plusieurs droites sont possibles, il faut donc essayer de trouver celle qui s'ajuste le mieux.

Propriété.

Soit une droite D d'équation $y = ax + b$ qui approche un nuage de points d'une série statistique. Soient M_i les points du nuage et P_i les points de la droite de même abscisse. On appelle droite des moindres carrés associée à ce nuage, la droite D avec a et b qui minimisent $M_1P_1^2 + M_2P_2^2 + \dots + M_nP_n^2$.

Définition.

- On appelle variance de la série statistique x , le réel $V(x) = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2$
- On appelle écart-type d'une série statistique x , le réel $\sigma(x) = \sqrt{V(x)}$
- On appelle covariance de la série statistique double $(x; y)$ le nombre réel, noté $cov(x; y)$, égal à : $cov(x; y) = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y})$

Propriété.

La droite des moindres carrés associée au nuage de points de la série double $(x; y)$:

- passe par le point moyen G de la série double
- a pour équation : $y = a(x - \bar{x}) + \bar{y}$ avec $a = \frac{cov(x; y)}{V(x)}$

2.2 Se ramener à un ajustement affine

Définition.

On appelle coefficient de corrélation linéaire entre x et y le réel $r = \frac{\text{cov}(x; y)}{\sigma(x)\sigma(y)}$

Propriété.

- $-1 \leq r \leq 1$
- Plus $|r|$ s'approche de 1 , et plus l'ajustement affine est indiqué
- Plus $|r|$ s'approche de 0 , et moins l'ajustement affine est approprié
- Si $|r| = 1$, le nuage de points est une droite .

Remarque.

Dans certaines situations , il est possible de se ramener à un ajustement affine en faisant un changement de variables .