COURS • 4ÈME ANNÉE MATHÉMATIQUES

Chapitre 17 — Statistiques

◆ Sommaire

1. Série statistique double
2. Ajustement affine par la méthode des moindres carrés
3. Ajustements se ramenant à un ajustement affine

◆ Objectifs du chapitre

– Représenter le nuage de points d'une série statistique double et placer son point moyen.

– Calculer (à la calculatrice) variances, covariance et coefficient de corrélation linéaire, et juger de la pertinence d'un ajustement affine.

– Déterminer la droite de régression par la méthode des moindres carrés et l'utiliser pour des estimations.

– Effectuer un changement de variable ($z=\ln y$, $z=\dfrac{1}{y}$, $X=e^{kx}$...) pour se ramener à un ajustement affine, puis revenir à la variable initiale.

Le plan du cours suit celui du manuel officiel (Tome 2, chapitre 10 : *Statistiques*).

1 Série statistique double

◆ Définitions

Une série statistique double est une liste de couples (x_i, y_i), $1 \le i \le n$. Elle est représentée par un **nuage de points** $M_i(x_i ; y_i)$. Les séries (x_i) et (y_i) sont ses distributions marginales.

Le **point moyen** est $G(\overline{x} ; \overline{y})$, avec $\overline{x}=\dfrac{1}{n}\sum x_i$ et $\overline{y}=\dfrac{1}{n}\sum y_i$.

◆ Variance, covariance, corrélation

$$V(x)=\frac{1}{n}\sum x_i^2 - \overline{x}^2,\ cov(x,y)=\frac{1}{n}\sum x_i y_i - \overline{x}\,\overline{y},\ r=\frac{cov(x,y)}{\sigma_x \sigma_y}.$$

On a toujours $-1 \le r \le 1$. Plus $|r|$ est proche de 1, plus les points sont proches d'une droite ; en pratique, on considère un ajustement affine comme justifié lorsque $|r|$ est proche de 1 (par exemple $|r| \ge 0,9$). Le

signe de r est celui de la pente.

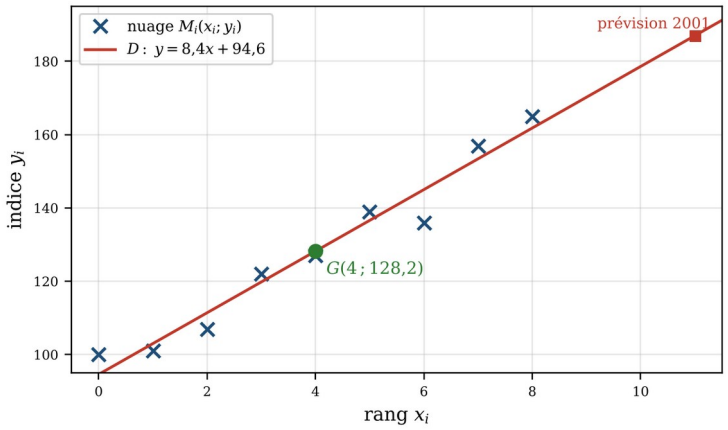
2 Ajustement affine par la méthode des moindres carrés

◆ Théorème

La droite de régression de y en x, qui minimise $\sum\left(y_i - (a x_i + b)\right)^2$, a pour équation

$$y = ax + b \text{ avec } a = \frac{cov(x, y)}{V(x)}, b = \overline{y} - a\,\overline{x}.$$

Elle passe par le point moyen G. La droite de régression de x en y est $x = a' y + b'$ avec $a' = \frac{cov(x, y)}{V(y)}$; on a $aa' = r^2$.



Nuage de points, point moyen G et droite de régression (données de l'exercice 5). Une prévision consiste à lire la droite en dehors de la plage des données (extrapolation).

3 Ajustements se ramenant à un ajustement affine

◆ Méthode — changement de variable

– Si le nuage a une allure **exponentielle** (croissance à taux constant), on pose $z = \ln y$. Un ajustement $z = ax + b$ donne $y = e^b e^{ax} = B e^{ax}$.

– Si y semble inversement proportionnel à une fonction affine de x, on pose $z = \frac{1}{y}$: $z = ax + b$ donne $y = \frac{1}{ax + b}$.

– Plus généralement, on choisit un changement $X = \varphi(x)$ ou $Y = \psi(y)$ pour lequel le nouveau coefficient de corrélation est plus proche de 1 en valeur absolue.

On revient toujours à la variable initiale avant de faire une prévision.

◆ Attention

Une droite de régression ne permet des prévisions raisonnables qu'au voisinage des données. Une extrapolation lointaine peut être absurde (temps négatif, charge supérieure à celle d'une flèche plus courte…) : il faut toujours confronter le modèle au sens physique du problème.