

Statistique à deux variables

DURÉE ESTIMÉE

20 minutes

OBJECTIFS DU CHAPITRE

Calculer les coordonnées du point moyen d'un nuage

Déterminer l'équation de la droite des moindres carrés

Interpréter le coefficient de corrélation linéaire

Réaliser un ajustement par changement de variable

Utiliser un ajustement pour interpoler ou extrapoler

1. Série statistique à deux variables

Définition

Soit deux caractères quantitatifs x et y observés simultanément sur n individus. La **série statistique à deux variables** obtenue est la liste des n couples :

$$(x_1; y_1), (x_2; y_2), \dots, (x_n; y_n)$$

Définition

Dans un repère du plan, l'ensemble des points $M_i(x_i; y_i)$ pour $i \in \{1, 2, \dots, n\}$ est appelé **nuage de points** associé à la série.

Définition

Le **point moyen** du nuage est le point $G(\bar{x}; \bar{y})$ dont les coordonnées sont les moyennes respectives des deux caractères :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{et} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Exemple

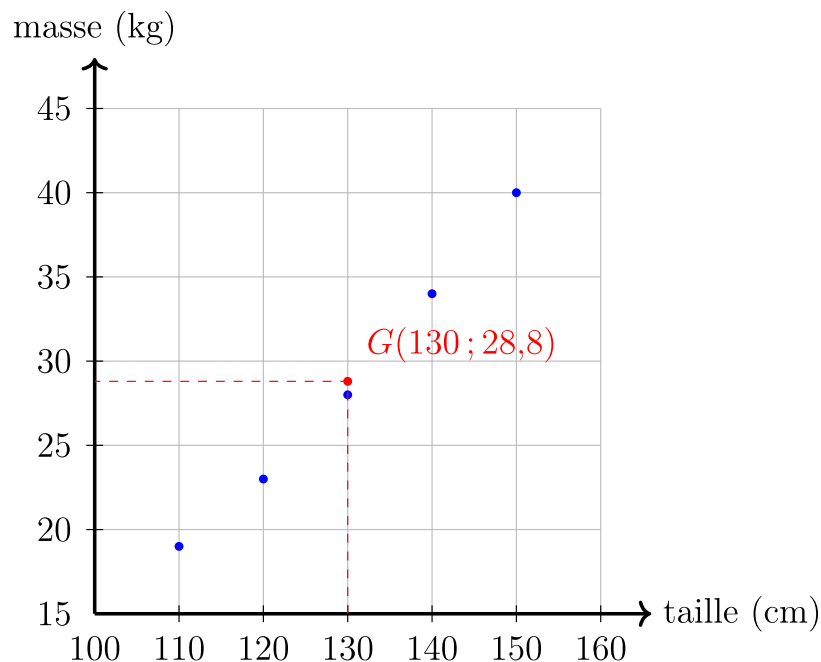
On relève la taille x (en cm) et la masse y (en kg) de 5 enfants :

- $(110; 19), (120; 23), (130; 28), (140; 34), (150; 40)$

Les coordonnées du point moyen sont :

$$\bar{x} = \frac{110 + 120 + 130 + 140 + 150}{5} = 130 \quad \text{et} \quad \bar{y} = \frac{19 + 23 + 28 + 34 + 40}{5} = 28,8$$

Le point moyen est donc $G(130; 28,8)$.



2. Ajustement affine — Droite des moindres carrés

Lorsque le nuage de points présente une forme allongée, on peut chercher à modéliser la liaison entre x et y par une fonction affine $y = ax + b$. On parle alors d'**ajustement affine**.



Définition

On appelle **droite des moindres carrés** (ou **droite de régression de y en x**) la droite d'équation $y = ax + b$ qui minimise la somme des carrés des écarts verticaux entre les points du nuage et la droite, c'est-à-dire la quantité :

$$S(a, b) = \sum_{i=1}^n (y_i - (ax_i + b))^2$$



Propriété

La droite des moindres carrés passe toujours par le point moyen $G(\bar{x}; \bar{y})$ du nuage.

Les coefficients a et b s'obtiennent à l'aide d'une **calculatrice** ou d'un **tableur** (mode statistique deux variables, fonction « régression linéaire » ou « DROITEREG »).



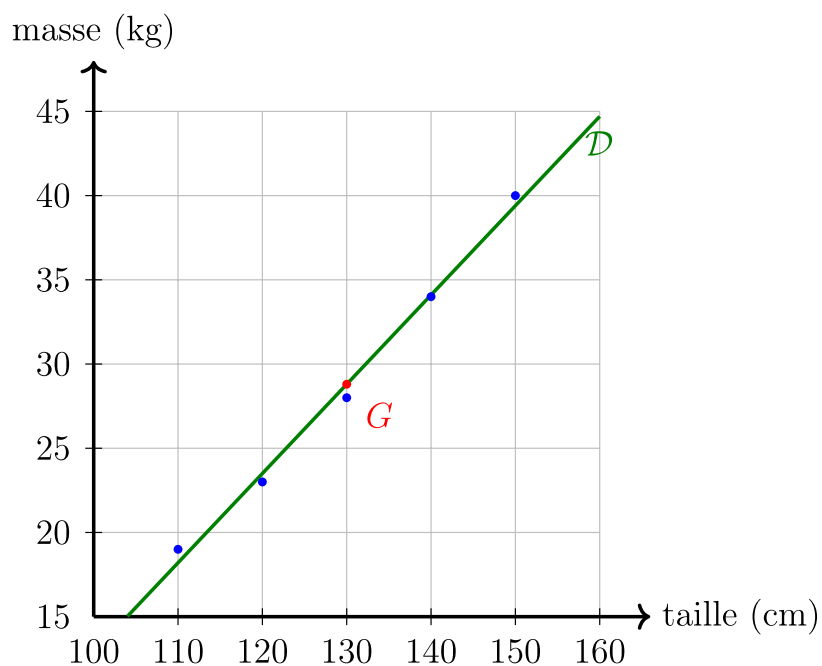
Exemple

Avec les données précédentes (taille / masse), la calculatrice donne l'équation de la droite des moindres carrés :

$$y = 0,53x - 40,1$$

On vérifie que cette droite passe (aux arrondis près) par le point moyen $G(130; 28,8)$:

$$0,53 \times 130 - 40,1 = 68,9 - 40,1 = 28,8$$



3. Coefficient de corrélation linéaire

Définition

Le **coefficient de corrélation linéaire** de la série, noté r , est un nombre réel compris entre -1 et 1 , qui mesure la qualité de l'ajustement affine.

Propriété

Soit r le coefficient de corrélation linéaire d'une série statistique à deux variables.

- $-1 \leq r \leq 1$
- Plus $|r|$ est proche de 1 , plus le nuage est aligné et plus l'ajustement affine est pertinent.
- Si $r > 0$, y tend à croître avec x ; si $r < 0$, y tend à décroître avec x .
- Si $|r|$ est proche de 0 , l'ajustement affine n'est **pas** pertinent (les variables peuvent toutefois être liées par une autre relation).

Remarque

Un coefficient de corrélation r proche de 1 ou -1 indique une **forte corrélation** entre les deux variables, mais **n'implique pas une relation de cause à effet**. La distinction entre corrélation et causalité est essentielle.

Exemple

Avec les données précédentes, la calculatrice donne $r \approx 0,997$. L'ajustement affine est donc particulièrement pertinent.

4. Ajustement par changement de variable

Lorsque le nuage de points n'est pas aligné, mais présente une forme régulière (courbe exponentielle, logarithmique, hyperbole, etc.), on peut effectuer un **changement de variable** qui se ramène à un ajustement affine.

Propriété

Voici quelques changements de variables classiques :

- Si $y = A \cdot e^{kx}$ (modèle exponentiel), on pose $z = \ln(y)$: alors $z = \ln(A) + kx$ est affine en x .
- Si $y = A \cdot x^k$ (modèle puissance), on pose $X = \ln(x)$ et $Z = \ln(y)$: alors $Z = \ln(A) + kX$ est affine en X .
- Si $y = \frac{a}{x} + b$ (modèle hyperbolique), on pose $X = \frac{1}{x}$: alors $y = aX + b$ est affine en X .

Exemple

On souhaite ajuster y en fonction de x par un modèle exponentiel à partir de la série :

- $(1; 2,7), (2; 7,4), (3; 20,1), (4; 54,6)$

On pose $z = \ln(y)$ et on calcule la nouvelle série $(x_i; z_i)$:

- $(1; 0,99), (2; 2,00), (3; 3,00), (4; 4,00)$

L'ajustement affine de z en x donne $z = x - 0,01$, soit $\ln(y) \approx x$, donc $y \approx e^x$.

5. Interpolation et extrapolation



Définition

Soit $y = f(x)$ un ajustement obtenu à partir d'un nuage de points pour lequel les valeurs de x sont comprises entre $x_{\min}$ et $x_{\max}$.

- On appelle **interpolation** l'utilisation de l'ajustement pour estimer y pour une valeur de x comprise dans l'intervalle $[x_{\min}; x_{\max}]$.
- On appelle **extrapolation** l'utilisation de l'ajustement pour estimer y pour une valeur de x située **en dehors** de l'intervalle $[x_{\min}; x_{\max}]$.



Remarque

L'extrapolation doit être maniée avec prudence : rien ne garantit que le modèle reste valide en dehors du domaine d'observation. Par exemple, l'extrapolation d'une croissance exponentielle de population sur plusieurs siècles donne des résultats irréalistes.



Exemple

Reprenons les données taille/masse, ajustées par $y = 0,53x - 40,1$.

- **Interpolation** : pour un enfant de taille $x = 125$ cm, on estime sa masse à
 $y \approx 0,53 \times 125 - 40,1 = 26,1$ kg.
- **Extrapolation** : pour une taille de 180 cm (hors plage observée), le modèle prédit
 $y \approx 0,53 \times 180 - 40,1 = 55,3$ kg, mais cette prédiction est moins fiable car aucune donnée

n'a été observée à cette taille.

Les questions essentielles

1. Comment calculer les coordonnées du point moyen d'un nuage ?

Le point moyen G a pour coordonnées les moyennes $\bar{x}$ et $\bar{y}$ des deux variables. On calcule séparément la moyenne des x_i et celle des y_i , puis on note $G(\bar{x} ; \bar{y})$.

Voir la fiche méthode : [Calculer les coordonnées du point moyen d'un nuage](#)

2. Comment déterminer l'équation de la droite des moindres carrés ?

On saisit les valeurs en mode statistique deux variables sur la calculatrice, puis on lance la régression linéaire pour lire la pente a et l'ordonnée à l'origine b . On vérifie ensuite que la droite passe par le point moyen G .

Voir la fiche méthode : [Déterminer l'équation de la droite des moindres carrés](#)

3. Comment interpréter le coefficient de corrélation linéaire r ?

On compare $|r|$ à 1 : plus $|r|$ est proche de 1, plus l'ajustement affine est pertinent. Le signe de r indique le sens de la liaison (croissante si $r > 0$, décroissante si $r < 0$).

Voir la fiche méthode : [Interpréter le coefficient de corrélation linéaire](#)

4. Comment ajuster un nuage non aligné par changement de variable ?

On choisit un modèle (exponentiel, puissance, hyperbolique) selon la forme du nuage, on applique le changement de variable adapté ($\ln(y)$, $\ln(x)$, $\frac{1}{x}$...) pour se ramener à une régression affine, puis on revient aux variables initiales.

Voir la fiche méthode : [Réaliser un ajustement par changement de variable](#)

5. Quelle est la différence entre interpolation et extrapolation ?

L'interpolation utilise l'ajustement pour estimer y pour une valeur de x comprise dans la plage observée. L'extrapolation l'utilise hors de cette plage : elle est moins fiable et doit être maniée avec prudence.

Voir la fiche méthode : [Utiliser un ajustement pour interpoler ou extrapoler](#)