

Statistiques à deux variables

Définition 1. Les statistiques à deux variables consistent à étudier deux caractères simultanément. Dans le cas où les deux caractères sont d'ordre quantitatif la série de leurs valeurs se note $(x_i; y_i)_{1 \leq i \leq n}$ où n est l'effectif total.

Exemple 1. Étude simultanée du poids et de la taille des individus d'une population.

Définition 2. On appelle nuage de points associé à une série statistique $(x_i; y_i)_{1 \leq i \leq n}$, l'ensemble des points de coordonnées $(x_i; y_i)$ dans un repère orthogonal.

Exercice 1. On considère la série statistique à deux variables suivante :

x_i	1	3	4	5	7
y_i	2	4	2	4	3

Représenter le nuage de points associé à cette série statistique.

Définition 3. On appelle point moyen du nuage de points associé à une série statistique $(x_i; y_i)_{1 \leq i \leq n}$ le point $G(\bar{x}; \bar{y})$ avec $\bar{x} = \frac{\sum x_i}{n}$ et $\bar{y} = \frac{\sum y_i}{n}$.

Exercice 2. Déterminer les coordonnées du point moyen du nuage de l'exercice précédent.

Définition 4. On appelle droite d'ajustement d'un nuage de points toute droite le remplaçant « au mieux ».

Exemple 2. Les droites d'équations $y = 3$ et $y = \frac{1}{2}x + 1$ sont des droites d'ajustement du nuage de points de l'exercice précédent.

Propriété 1. Pour toute droite d'ajustement passant par le point moyen du nuage, la somme des écarts verticaux des points par rapport à cette droite est nulle.

Exercice 3. Vérifier la propriété pour les droites d'ajustement précédentes.

Définition 5. On appelle droite de régression de y en x ou droite d'ajustement de y en x par la méthode des moindres carrés d'un nuage de points la droite qui minimise la somme des carrés des écarts verticaux.

Propriété 2. La droite de régression de y en x d'un nuage de points passe par le point moyen.

Propriété 3. La droite de régression de y en x d'un nuage de points a pour équation $y = ax + b$ avec $a = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$ et $b = \bar{y} - a\bar{x}$.

Exercice 4. Déterminer la droite de régression de y en x du nuage de points de l'exercice précédent.