

# Séquence 8 : Séries statistiques à deux variables

## I) Séries statistiques à deux variables

Définition :

Une série statistique à deux variables est une série statistique étudiant simultanément deux caractères sur un même échantillon de  $n$  individus extraits d'une population.

On peut présenter une série statistique à deux variables à l'aide d'un tableau statistique de la forme :

Valeurs du 1 <sup>er</sup> caractère	$x_1$	$x_2$	...	$x_n$
Valeurs du 2 <sup>nd</sup> caractère	$y_1$	$y_2$	...	$y_n$

Exemple :

Dans ce cours nous considérerons l'exemple ci-dessous.

On a mesuré par échographie la taille d'un fœtus humain en fonction du nombre de semaines de grossesse.

Les résultats sont donnés dans le tableau ci-après.

Nombre de semaines $x_i$	6	10	14	18	22	26	30
Taille du fœtus $y_i$ (en cm)	2	7	16	24	32	36	39

Les deux caractères statistiques étudiés sont le nombre de semaines et la taille du fœtus.

Remarque :

Quand sur une population, on étudie un seul caractère dont les valeurs sont des relevées à des dates différentes, on forme une série statistique à deux variables où le temps est une des deux variables.

Une telle série est dite chronologique.

## II) Nuage de points

Définition :

Soit deux variables  $X$  et  $Y$  étudiées sur un même échantillon de  $n$  individus extraits d'une population.

On présente cette série statistique dans le tableau suivant :

Valeurs du 1 <sup>er</sup> caractère $X$	$x_1$	$x_2$	...	$x_n$
Valeurs du 2 <sup>nd</sup> caractère $Y$	$y_1$	$y_2$	...	$y_n$

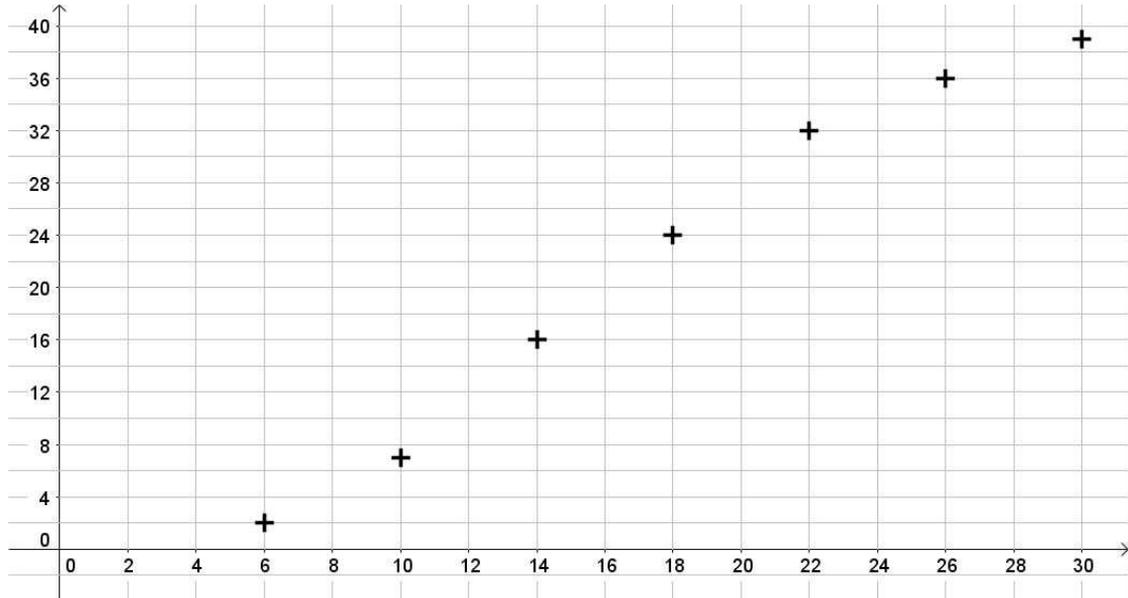
Dans un repère du plan, le nuage de points associé à cette série est l'ensemble des points de coordonnées  $(x_i; y_i)$  où l'entier naturel  $i$  prend toutes les valeurs entre 1 et  $n$ .

Soit  $\bar{x}$  la moyenne des valeurs prises par le premier caractère et  $\bar{y}$  la moyenne des valeurs prises par le deuxième caractère. On appelle point moyen du nuage de points de la série statistique, le point  $G$  de coordonnées  $(\bar{x}; \bar{y})$ .

Exemple :

Le nuage de points de la série statistique à deux variables est constitué des 7 points :

(6;2) , (10;7) , (14;16) , (18;24) , (22;32) , (26;36) , (30;39).



Calculons la moyenne de chaque série statistique :

$$\bar{x} = \frac{6 + 10 + 14 + 18 + 22 + 26 + 30}{7} = 18$$

$$\bar{y} = \frac{2 + 7 + 16 + 24 + 32 + 36 + 39}{7} = \frac{156}{7}$$

Le point moyen  $G$  a pour coordonnées  $\left(18; \frac{156}{7}\right)$ .

### III) Ajustement affine

**Théorème (méthode des moindres carrés) :**

Soit deux variables  $X$  et  $Y$  étudiées sur un même échantillon de  $n$  individus extraits d'une population.

On présente cette série statistique dans le tableau suivant :

Valeurs du 1 <sup>er</sup> caractère $X$	$x_1$	$x_2$	...	$x_n$
Valeurs du 2 <sup>nd</sup> caractère $Y$	$y_1$	$y_2$	...	$y_n$

Soit un nuage de  $n$  points  $(x_i; y_i)$ ,  $i$  entier naturel allant de 1 à  $n$ . Il existe une unique droite d'équation  $y = ax + b$  passant par le point moyen  $G$  du nuage de points telle que la somme des carrés des écarts des valeurs observées  $y_i$  à cette droite soit minimale.

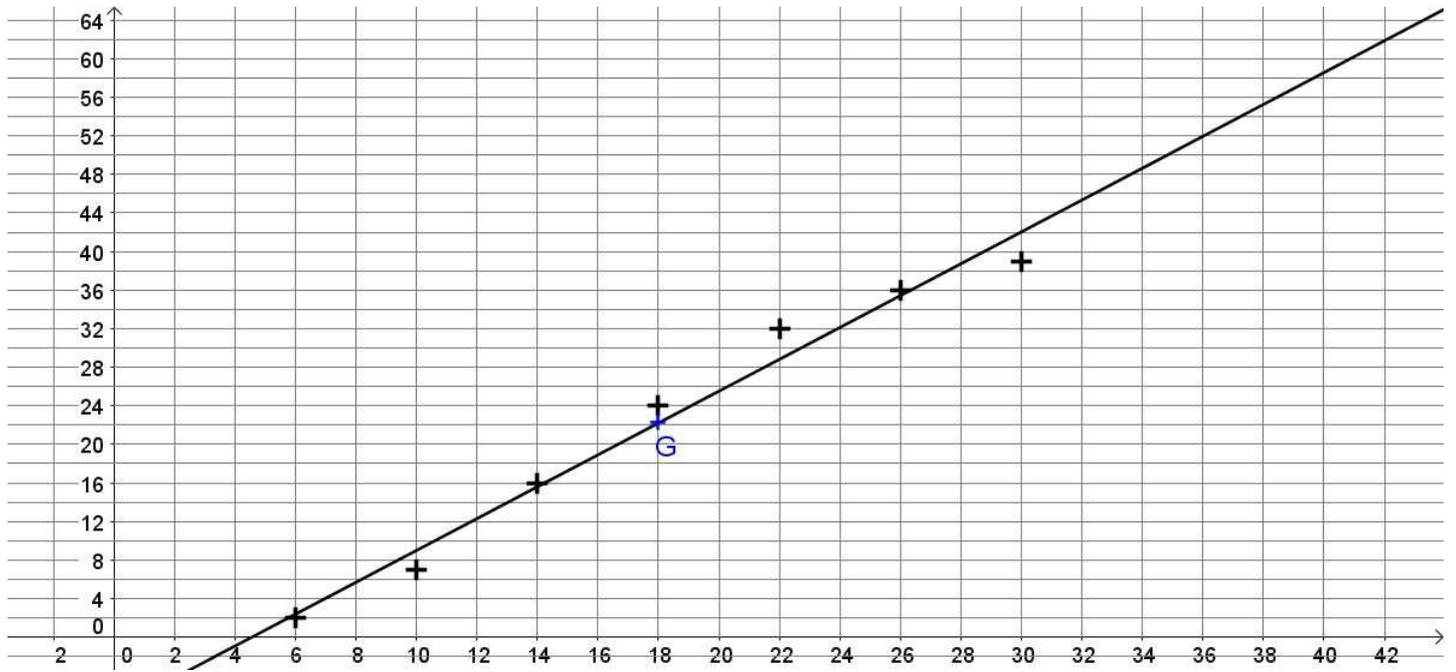
Autrement dit,  $\sum_{i=1}^n (y_i - (ax_i + b))^2$  est minimale.

Cette droite s'appelle la droite de régression de  $y$  en  $x$  ou encore la droite d'ajustement de  $y$  en  $x$ .

Remarques :

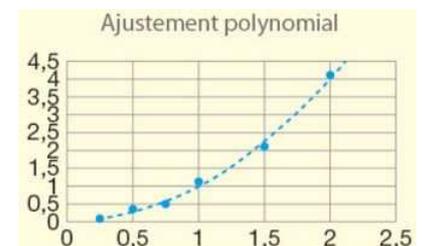
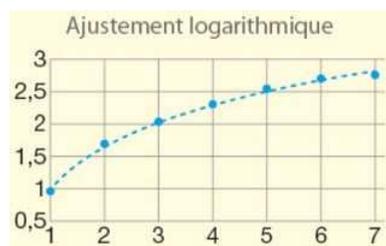
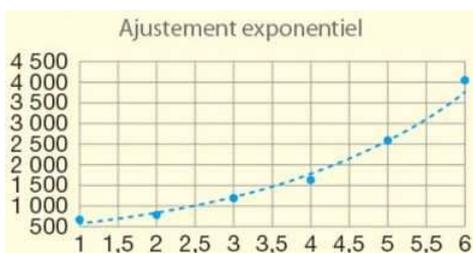
- Une droite d'ajustement d'un nuage de points est une droite passant "au plus près" du nuage de points associé à cette série.
- Un ajustement affine permet de faire des estimations.
- Si les estimations sont réalisées dans le domaine d'étude de la série, ce sont des interpolations; sinon ce sont des extrapolations (prévisions pour les années à venir par exemple).

### Exemple :



La droite d'ajustement de  $y$  en  $x$  par la méthode des moindres carrés a pour équation  $y = 1.65x - 7.44$ , les coefficients étant donnés à 0.01 près.

Dans certains cas, l'ajustement affine ne semble pas l'ajustement le plus adapté : un autre ajustement à l'aide d'une courbe (exponentielle, logarithmique, polynomiale ...) peut être envisagé.



Lorsque l'ajustement du nuage de points par une courbe est judicieux, on peut effectuer un changement de variables afin de se ramener à un ajustement affine, ce qui nous permet d'utiliser les notions vues précédemment.

