

## Partie 1 : Statistique et probabilités

# Statistiques à deux variables

- ❖ Depuis 1975, la température moyenne sur le globe augmente dangereusement. Pour de nombreux scientifiques, ce phénomène est directement lié à l'augmentation des émissions de CO<sub>2</sub> due à l'activité humaine. Ils prédisent pour les vingt ans à venir, un réchauffement moyen de 0,2 °C par décennie.

**Comment ont-ils pu faire cette prévision ?**

Ils faut des données depuis un certain nombre d'années sur : les émissions de CO<sub>2</sub> et la température moyenne sur Terre. Ici, on a donc 2 variables, et on essaie de voir le lien qu'elles ont entre elles. Une fois qu'on a trouvé le lien entre les deux variables qui fonctionne par exemple de 1975 à 2010, on peut prédire ce qui se passera plus tard.

On est dans les statistiques à deux variables, et le lien s'appelle un ajustement.

### 1 Je vérifie mes acquis

Pour chaque question, choisir la (ou les) bonne(s) réponse(s).

1. Voici les notes obtenues par Thomas en mathématiques : 12,5 ; 15 ; 13,5 ; 17 ; 12,5 ; 13,5. Sa moyenne est :

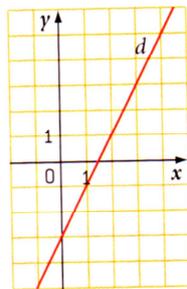
- a. 13,5    b. 14    c. 14,5    d. 15

2. Dans un repère (O ; I, J), la droite d'équation  $y = -2x + 7$  passe par :

- a. A(0 ; 7)    b. B(0 ; -2)    c. C(-2 ; 7)    d. D(2 ; 3).

3. Une droite  $d$  est représentée ci-après. Cette droite  $d$  passe par :

- a. A(1 ; 2)  
b. B(1 ; 1)  
c. C(2 ; 1)  
d. D(0 ; -3)



4. La droite  $d$  de l'exercice 3 a pour équation réduite :

- a.  $y = -3x + 2$     b.  $y = 2x$   
c.  $y = x - 3$     d.  $y = 2x - 3$

5. Dans un repère (O ; I, J), soient A(1 ; 1,5) et B(3 ; 0,5). Le coefficient directeur de la droite (AB) est égal à :

- a. 2    b. 0,5    c. -0,5    d. -2

6. On reprend l'exercice 5. L'ordonnée à l'origine de la droite (AB) est égale à :

- a. 2    b. 0,5    c. -0,5    d. -2

7. Parmi les expressions suivantes, indiquer celles qui définissent une fonction affine :

- a.  $f(x) = \frac{1}{3}x$     b.  $f(x) = \frac{1}{3x}$   
c.  $f(x) = 2 - 5x$     d.  $f(x) = 3x^2 - 4$

#### Réponses

1. **b** ( $84/6 = 14$ )

2. **a** (on remplace  $x$  par 0, on a  $y=7$ )

3. **c et d**

4. **d** ( $y=ax+b$  ;  $b=-3$  ; et  $a=2$ )

5. **c**  $a = \frac{y_B - y_A}{x_B - x_A} = \frac{0,5 - 1,5}{3 - 1} = \frac{-1}{2} = -0,5$

6. **a**

$a=-0,5$  donc l'équation est  $y = -0,5x + b$

Comme A est sur la droite (AB) alors A vérifie l'équation de la droite donc

$$y_A = ax_A + b$$

$$\text{donc } 1,5 = -0,5 \times 1 + b$$

$$1,5 = -0,5 + b$$

$$1,5 + 0,5 = b \text{ donc } b=2$$

7. **a et c.**

## 2 Activités

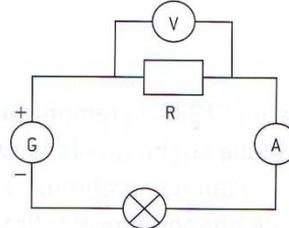
### 2.1 Activité 1 : Ohm, sweet Ohm

Répondre sur une feuille à carreaux

**Objectif :** représentation et analyse d'une série statistique double



Pour étudier le lien entre la tension  $U$  mesurée aux bornes d'un conducteur ohmique et l'intensité  $I$  du courant électrique qui le parcourt, on a réalisé le montage suivant :



Le générateur utilisé est un générateur à basse fréquence et permet de régler la tension fournie au circuit.

Grâce à ce montage, on a fait les relevés suivants :

Tension $U$ (en V) : $u_i$	0	3	5,1	8	10,3	12
Intensité du courant $I$ (en mA)	0	13,8	22,7	37,0	47,0	54,1

On se propose de construire la caractéristique de ce dipôle.

**1. a.** Dans un repère orthogonal, placer les différents points correspondants à chaque couple de mesures  $(I, U)$ . (On choisira des unités adaptées pour chacun des axes.)

Le **nuage de points** obtenus représente la **série statistique à deux variables**  $I$  et  $U$  où  $I$  est l'intensité du courant et  $U$  la tension aux bornes du dipôle.

**b.** Comment semblent être disposés les différents points du nuage ?

**2. a.** Calculer  $I_{\text{moy}}$  et  $U_{\text{moy}}$  les moyennes respectives des intensités et des tensions relevées ; puis placer le point  $G$  de coordonnées  $(I_{\text{moy}}; U_{\text{moy}})$ .

Ce point  $G$  est appelé le **point moyen du nuage de points**.

**b.**  $O$  étant l'origine du repère, tracer la droite  $(OG)$ .

Que représente cette droite pour le conducteur ohmique ?

Déterminer l'équation réduite de la droite  $(OG)$ . (Arrondir les coefficients à 0,01 près.)

**c.** Recopier et compléter le tableau ci-dessous.

Intensité $I$ (en mA)	0	13,8	22,7	37,0	47,0	54,1
Ordonnées des points de $(OG)$ correspondants : $\gamma_i$						
Écarts $\gamma_i - u_i$						

Vérifier que ces écarts sont inférieurs à 0,15 V.

Cette droite  $(OG)$  « passe au plus près » des différents points du nuage.

On dit que c'est **une droite d'ajustement du nuage**.

Son équation réduite donne une relation approchée entre les deux caractères étudiés.

**3. a.** Quelle relation peut-on en déduire entre  $U$  et  $I$  ?

Quelle loi d'électricité reconnaît-on ici ? Estimer la résistance de dipôle ohmique.

**b.** Estimer alors la tension  $U$  pour une intensité  $I$  de 30 mA ; puis pour 65 mA.

**c.** Estimer l'intensité  $I$  pour une tension  $U$  de 6 V ; puis pour 15 V.

## A retenir :

En calculant les **moyennes** on détermine les coordonnées du **point moyen G**.

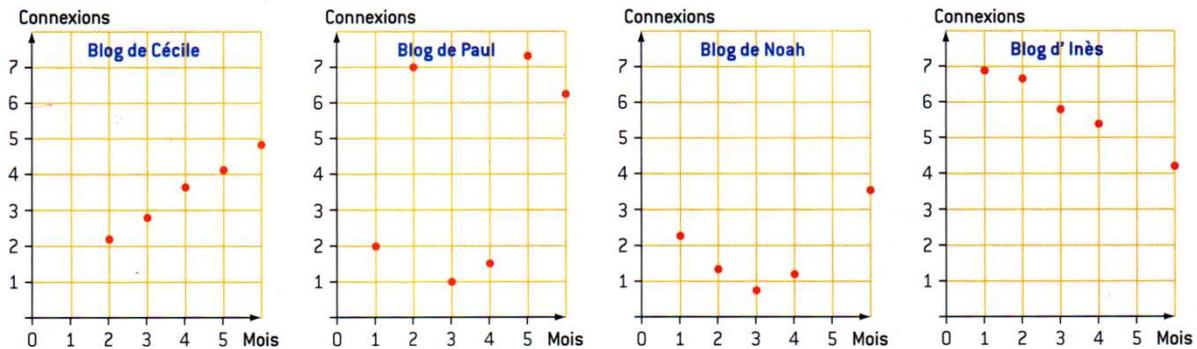
La droite qui passe au **plus près** des différents points du nuage et qui passe **par G** s'appelle la **droite d'ajustement**, on parle **d'ajustement affine**.

## 2.2 Activité 2 : Prévoir des connexions

*Répondre sur une feuille à carreaux*

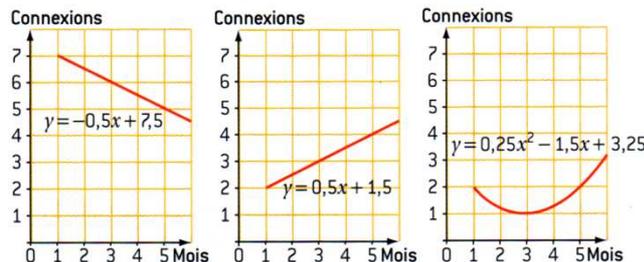
### **Objectif :** étude de séries chronologiques et comprendre la notion d'ajustement et son utilisation

Cécile, Paul, Noah et Inès, quatre élèves de Terminale Bac Pro, ont chacun créé leur blog. Mois après mois, depuis leur création, chaque lycéen a noté le nombre de connexions mensuelles sur son blog. Voici les quatre nuages de points donnant le nombre de connexions (exprimé en centaines) selon les mois.



Le mois noté « 1 » correspond au mois de janvier. On dit que janvier est le mois de **rang 1**.

1. Quel caractère lit-on sur l'axe des abscisses ? Sur l'axe des ordonnées ?
2. a. Quel mois Cécile a-t-elle créé son blog ?  
b. Quel mois Noah et Inès ont-ils oublié de noter le nombre de connexions ?  
c. Parmi ces quatre blogs, lequel a-t-il été le plus visité au mois d'avril ?
3. En observant la forme des différents nuages, dire pour quel blog il ne semble pas exister de lien entre le nombre de connexions et les rangs des mois.
4. a. Pour chacun des trois autres blogs, choisir la courbe ci-dessous qui semble passer au « plus près » des différents points du nuage.



En faisant cela, on **effectue un ajustement** de chaque nuage de points.

b. En déduire, pour chaque blog, une relation approchée du nombre de connexions mensuelles  $C$  en fonction du rang du mois  $n$ .

5. a. Déterminer alors une valeur approchée de la donnée manquante pour Noah et Inès.

b. Donner une estimation du nombre de connexions sur chaque blog, pour le mois d'août.

## A retenir :

**Dans certains cas, on ne peut pas faire d'ajustement (blog de Paul).**

**Faire un ajustement permet de faire une estimation de ce qui va se passer plus tard (mois d'août par exemple).**

### 3 Résumé :

#### 3.1 La notion de statistique à 2 variables :

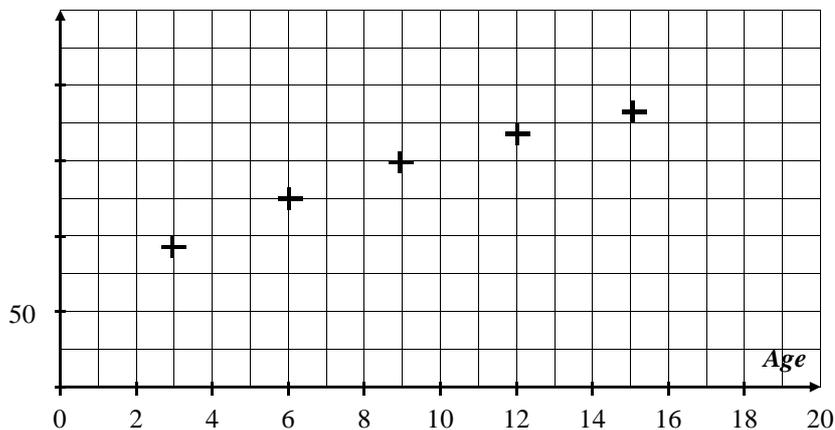
Lorsque l'on étudie conjointement deux caractères quantitatifs  $x$  et  $y$ , on obtient alors une série statistique quantitative à deux variables.

Dans un repère orthogonal, l'ensemble des  $n$  points  $M_i$  de coordonnées  $(x_i ; y_i)$  forment le nuage de points représentant cette série statistique à deux variables.

Exemple 1 : on étudie la taille d'un nourrisson et son âge. Les données de cette série double sont :

Age (en mois) : $x_i$	3	6	9	12	15
Taille (en cm) $y_i$	59	65	70	74	77

Le nuage de points représentant cette série double est : (à compléter)



#### Cas particulier : série chronologique

On dit qu'une série est « chronologique », lorsque l'on étudie les valeurs prises par un caractère  $Y$  au cours du temps.

ATTENTION : si le temps est exprimé qualitativement (par exemple les mois de l'année), on doit alors définir un caractère quantitatif  $X$ , par exemple le rang du mois.

Exemple 2 : Pour étudier les dépenses de santé suivant les mois (janvier, février, etc.), on associera à chaque mois son rang (1, 2, 3,...).

#### 3.2 La notion de point moyen :

Le point moyen d'un nuage de  $n$  points  $M_i$  de coordonnées  $(x_i, y_i)$  est le point  $G$  de coordonnées  $(\bar{x}; \bar{y})$ , où  $\bar{x}$  et  $\bar{y}$  sont les moyennes respectives des abscisses  $x_i$  et des ordonnées  $y_i$ .

Exemple : calculer les coordonnées du point moyen G de l'exemple 1 :

.....  
.....  
.....  
.....

Placer G sur le graphique de l'exemple 1.

Remarque : En général, on fait apparaître le point moyen sur le nuage de points.

### 3.3 La notion d'ajustement affine :

Effectuer un ajustement affine d'un nuage de points, c'est déterminer une fonction affine  $f : x \rightarrow ax+b$  dont la droite représentative  $d$  passe « au plus près » de tous les points du nuage.

- **Une droite d'ajustement passe toujours par G le point moyen du nuage.**
- **L'équation réduite  $y = ax + b$  de cette droite d'ajustement donne de façon approchée une relation entre les ordonnées et les abscisses des points du nuage.**

#### Remarques :

- ❖ Cette relation sera d'autant plus précise que la forme du nuage de points est allongée.
- ❖ Suivant la forme du nuage de points, on peut l'ajuster par d'autres types de fonctions, ou ne pas pouvoir effectuer d'ajustement si le nuage est trop « dispersé ».

Exemple (suite de l'exemple 1) : La forme du nuage de points est très allongée, on effectue donc un

.....

Tracer la droite d'ajustement  $d$  en la faisant passer « au plus près » des points et par le point G.

Cette droite passe par G (..... ; ..... ) et le point I (0 ; 55,5).

Calculer le coefficient directeur  $a$  de la droite  $d$ . Déterminer alors l'équation réduite de la droite  $d$ .