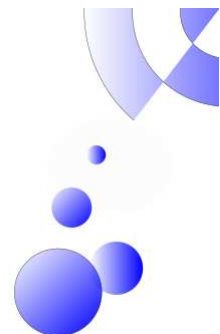
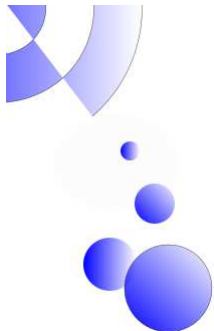




Table des Matières

| | |
|-----------------------------------------|---|
| I. Introduction | 1 |
| II. Nuage de points et point moyen | 1 |
| III. Approximation d'un nuage de points | 2 |





I. Introduction

Activité 1

Le tableau suivant donne l'implication des citoyens dans les sciences participatives liées à la biodiversité entre 2011 et 2017 :

| | | | | | | | |
|--------------|--------|--------|--------|--------|--------|--------|--------|
| année | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 |
| participants | 21 143 | 30 363 | 35 209 | 30 940 | 35 403 | 45 251 | 53 732 |

biodiversité - les chiffres clés - 2018 - ministère du développement durable

- Faire un graphique de la situation, l'origine du repère aura pour coordonnées (2010 ; 0).
Échelle :
 - en abscisse 1cm pour une unité,
 - en ordonnée 1cm pour 10000 participants.
- Quel nombre de participants peut-on prévoir en 2025 ? Expliquer la démarche et votre modèle.

Pour tout le cours, on considère une série statistique à deux variables discrètes x et y prenant n valeurs :

| | | | | | | |
|-----|-------|-------|-----|-------|-----|-------|
| x | x_1 | x_2 | ... | x_i | ... | x_n |
| y | y_1 | y_2 | ... | y_i | ... | y_n |

II. Nuage de points et point moyen

Définition

Soient une série statistique à deux variables discrètes, x et y . On appelle nuage de point la représentation des points $M(x_i ; y_i)$ dans un repère $(O; \vec{i}, \vec{j})$.

Définition

Le point moyen G du nuage est le point de coordonnée $G(\bar{x}; \bar{y})$.
 \bar{x} est la moyenne des valeurs de la variable x et \bar{y} est la moyenne des valeurs de la variable y .

III. Approximation d'un nuage de points

Si le nuage de points est assez allongé, on cherche à approcher les points du nuage par une droite δ qui devra être la plus proche possible des points du nuage.

Activité 2

Soit la série statistique de variables discrètes x et y :

| | | | | | | | |
|-----|----|----|----|----|----|----|----|
| x | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| y | 21 | 30 | 35 | 31 | 35 | 45 | 54 |

1. Faire le nuage de points dans un repère orthogonal $(O; \vec{i}, \vec{j})$.

Échelle :

- en abscisse 1cm pour une unité,
- en ordonnée 1cm pour 5 unités et l'axe des ordonnées comme à l'unité 20.

2. On approche le nuage de points par deux droites d_1 et d_2 dont on donne les équations :

- $d_1 : y = 4,6x + 17,4$
- $d_2 : y = 4,5x + 18,0$

3. Tracer les droites sur le graphique précédent.

4. (a) Compléter le tableau des erreurs absolues au carré en ordonnées pour chacune des droites :

| | | | | | | | | |
|-------------------|----|----|----|----|----|----|----|--------|
| x | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
| y | 21 | 30 | 35 | 31 | 35 | 45 | 54 | |
| y_{d_1} | | | | | | | | |
| y_{d_2} | | | | | | | | sommes |
| $(y_{d_1} - y)^2$ | | | | | | | | |
| $(y_{d_2} - y)^2$ | | | | | | | | |

(b) Quelle droite approxime le mieux le nuage ?

Activité 3

La méthode des moindres carrés

$(O; \vec{i}, \vec{j})$ est un repère orthonormé du plan.

Soit un nuage de n points $M_i(x_i; y_i)_{1 \leq i \leq n}$, G le point moyen de coordonnées $(\bar{x}; \bar{y})$ et une droite δ d'équation réduite $y = mx + p$.

Le but de l'activité est de déterminer la droite qui minimise la l'erreur des carrés des écarts en ordonnées soit déterminer m et p tels que la somme $\sum_{i=1}^n (y_i - (mx_i + p))^2$ soit la plus petite possible.

Soient les fonctions f et g :

$$\begin{aligned} f: \mathbb{R} &\rightarrow \mathbb{R} \\ p &\mapsto f(p) = \sum_{i=1}^n (y_i - (mx_i + p))^2 \end{aligned}$$

$$\begin{aligned} g: \mathbb{R} &\rightarrow \mathbb{R} \\ m &\mapsto g(m) = \sum_{i=1}^n (y_i - (mx_i + p))^2 \end{aligned}$$

1. Propriété de la fonction f :

- Donner la nature de la fonction f ?
- Justifier que le minimum de f est atteint pour $f'(p) = 0$ soit $\sum_{i=1}^n (y_i - (mx_i + p)) = 0$.
- En déduire que les coordonnées de G vérifient $y = mx + p$.
Ainsi G appartient à la droite δ et $p = \bar{y} - m\bar{x}$.

2. Propriété de la fonction g :

- Montrer que

$$g(m) = \sum_{i=1}^n (y_i - (mx_i + p))^2 = \sum_{i=1}^n ((y_i - \bar{y}) - m(x_i - \bar{x}))^2 = m^2 \sum_{i=1}^n (x_i - \bar{x})^2 - 2m \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) + \sum_{i=1}^n (y_i - \bar{y})^2$$

- Donner la nature de la fonction g .

- Montrer que le minimum de g est atteint en $m = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$

$$\text{Remarque : } \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}}{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} = \frac{\sigma_{xy}}{\sigma_x^2}.$$

σ_x^2 est la variance de la variable x .

☞ Définition

Soit une série statistique à deux variables x et y .
On appelle covariance de x et y , notée σ_{xy} le nombre :

$$\sigma_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}$$

☞ Définition

Soit une série statistique à deux variables x et y , on note \bar{x} et \bar{y} la moyenne respective de la variable x et de la variable y .

$(O; \vec{i}, \vec{j})$ est un repère orthogonal du plan. On appelle droite de régression linéaire par la méthode des moindres carrés y en x , la droite qui minimise l'erreur absolue en ordonnée au carré, cette droite a pour équation $y = mx + p$ avec :

- $m = \frac{\sigma_{xy}}{\sigma_x^2}$
- $p = \bar{y} - m\bar{x}$

☞ Remarque

La droite δ déterminée par la méthode des moindres carrés passe par le point moyen G du nuage.

☞ Exercice 1

On reprend l'activité 1 :

Le tableau suivant donne l'implication des citoyens dans les sciences participatives liées à la biodiversité entre 2011 et 2017 :

| année | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 |
|--------------|--------|--------|--------|--------|--------|--------|--------|
| participants | 21 143 | 30 363 | 35 209 | 30 940 | 35 403 | 45 251 | 53 732 |

1. Déterminer la droite de régression linéaire par la méthode des moindres carrés.
Vérification par la calculatrice.
2. Avec ce modèle, déterminer l'implication des citoyens dans les sciences participatives liées à la biodiversité en 2030.

☞ Exercice 2

Soit la série statistique à deux variables x et y qui donne l'évolution du nombre d'habitants de la population française (en milliers) suivant le temps :

| Année x | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
|----------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Population y | 62 765 | 63 070 | 63 375 | 63 697 | 63 697 | 64 300 | 64 468 | 64 639 | 64 737 | 64 821 | 64 124 |

Source : Insee

1. À l'aide de la calculatrice, déterminer la droite de régression linéaire par la méthode des moindres carrés.
2. Avec ce modèle, déterminer la population française en 2030.