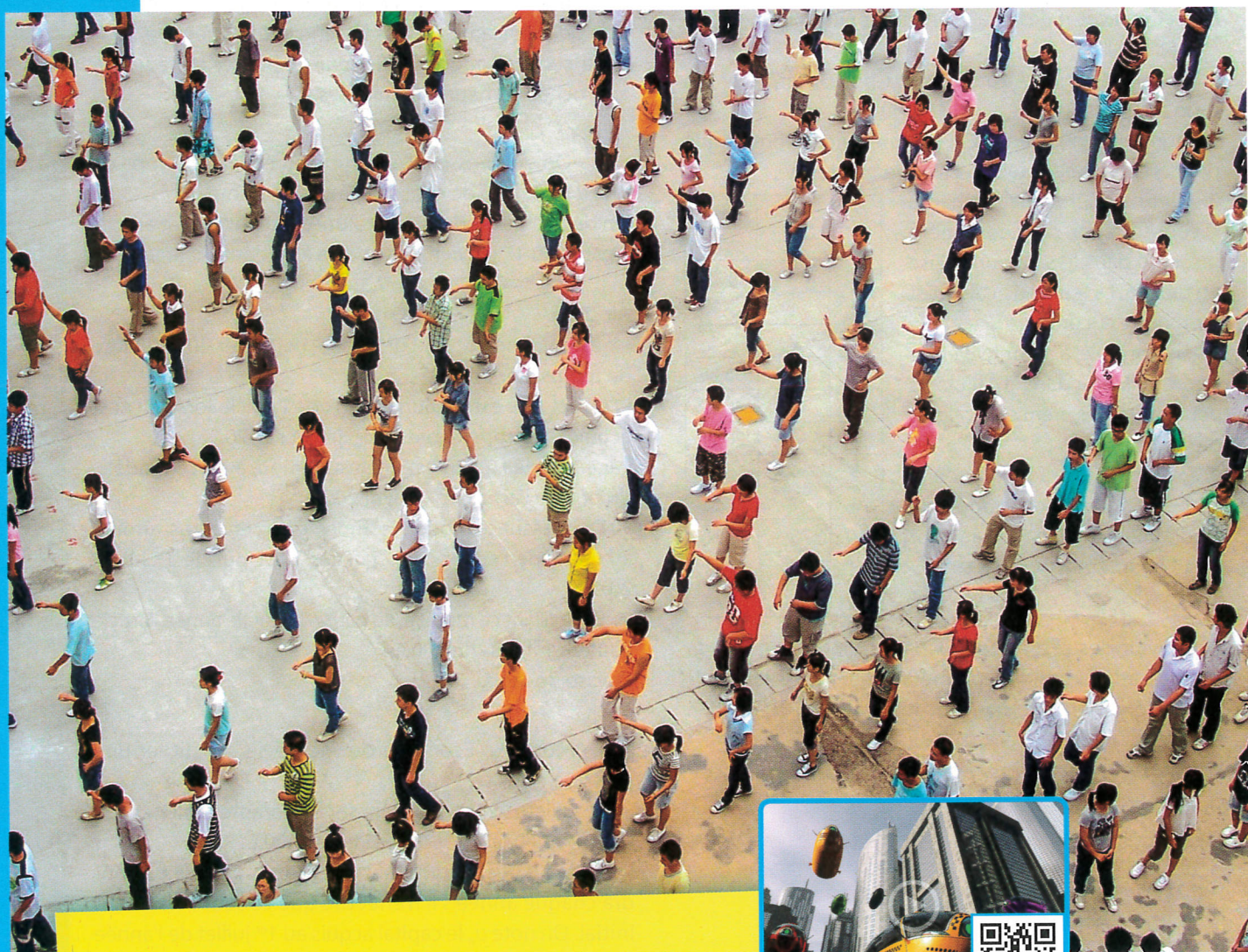


5

Statistiques à deux variables quantitatives

CAPACITÉS

- Représenter un nuage de points.
- Déterminer et utiliser un ajustement affine pour interpoler ou extrapoler des valeurs inconnues.
- Représenter un nuage de points en effectuant un changement de variable donné afin de conjecturer une relation de linéarité entre de nouvelles variables.



Aujourd'hui, la population de la Chine est la plus élevée au monde, talonnée par l'Inde.

Depuis 1960, elle a été multipliée de plus de 118 %. La population chinoise continue d'augmenter... et de vieillir. Si la Chine est aujourd'hui un pays à l'économie puissante, elle va devoir s'organiser pour répondre aux besoins de demain.

Combien d'habitants peut-on prévoir en Chine en 2025 ?



Découvrons les dynamiques de la population

1:05

lienmini.fr/10445-27

→ Pour le découvrir **Activité 3** p. 99

Pour retrouver les automatismes

Pour revoir les acquis de Première

Questions
Flash

Diaporama

15 diapositives
pour retrouver
ses automatismes



lienmini.fr/10445-28

1 Équation réduite de droite

- L'équation réduite d'une droite d non parallèle à l'axe des ordonnées est de la forme :

$$y = ax + b.$$

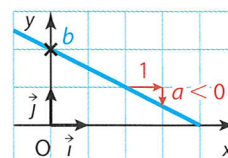
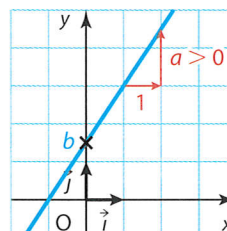
Coefficient directeur de d

Ordonnée à l'origine de d

- Si $A(x_A; y_A)$ et $B(x_B; y_B)$ sont deux points de d alors

$$a = \frac{y_B - y_A}{x_B - x_A}.$$

- $C(x_C; y_C) \in d$ si et seulement si $y_C = ax_C + b$.



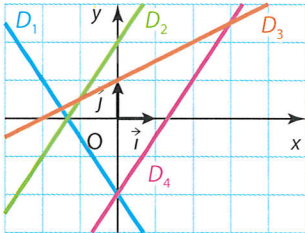
2 Moyenne d'une série de valeurs

La moyenne de n nombres réels x_1, x_2, \dots, x_n est : $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$.

Vérifier les acquis de Première

QCM Pour chacune des questions posées, indiquer la bonne réponse puis justifier.

Aide

	a	b	c	d	
1. On donne les points A(-1 ; 3) et B(0 ; -2). La droite (AB) a pour équation réduite :	$y = -5x + 2$	$y = -\frac{1}{5}x - 2$	$y = -5x - 2$	$y = -\frac{1}{5}x + \frac{14}{5}$	1
2. Parmi les droites représentées ci-dessous, celle qui a pour équation réduite $y = 1,5x - 2$ est : 	D_1	D_2	D_3	D_4	1
3. Soit D la droite d'équation réduite $y = -2,4x + 5,1$. Le point de D d'abscisse 3 a pour ordonnée :	12,3	-2,1	-0,875	2,7	1
4. Soit D la droite d'équation réduite $y = 0,2x - 6,2$. Le point de D d'ordonnée 4,3 a pour abscisse :	-5,34	52,5	9,5	10,3	1
5. Quelle est la moyenne de : 23, 26, 20, 45, 51, 39 ?	$\bar{x} = 31$	$\bar{x} = 32,5$	$\bar{x} = 27$	$\bar{x} = 34$	2

→ Voir **Corrigé** p. 324

OBJECTIF Représenter une série statistique à deux variables → Cours 18 p. 100

On a relevé la superficie x , exprimée en m^2 , et le prix de vente y , exprimé en milliers d'euros, de 8 appartements situés dans le quartier résidentiel d'une petite ville.

Superficie : x_i (en m^2)	61	72	86	93	105	122	143	184
Prix de vente : y_i (en milliers d'euros)	195	210	238	264	280	300	354	415

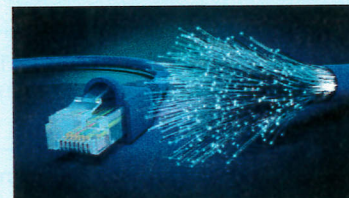


- Quels sont les deux caractères étudiés sur la population constituée des 8 appartements ?
- Dans un repère orthogonal (unités graphiques : 1 cm pour 20 m^2 sur l'axe des abscisses et 1 cm pour 50 milliers d'euros), placer dans ce repère les points $M_1(x_1; y_1), M_2(x_2; y_2), \dots, M_8(x_8; y_8)$ associés à chacun des 8 appartements.
 - Quelle remarque peut-on faire concernant la position de ces 8 points ?
- Calculer \bar{x} et \bar{y} , moyennes respectives des superficies et des prix de vente des 8 appartements.
 - Placer sur le graphique le point $G(\bar{x}; \bar{y})$. Où est-il situé par rapport aux points du nuage ? Ce point s'appelle le **point moyen du nuage de points** de la série statistique.

OBJECTIF Apprécier la qualité d'un ajustement affine et effectuer un ajustement selon la méthode des moindres carrés → Cours 2 p. 102

Une entreprise de fabrication de fibres optiques, destinées aux opérateurs téléphoniques, a mis sur le marché un nouveau produit qui a beaucoup de succès. L'évolution du nombre de milliers d'acquéreurs depuis 18 mois est résumée dans le tableau ci-dessous :

Trimestre : x_i	1	2	3	4	5	6
Nombre de milliers d'acquéreurs : y_i	155	232	327	472	636	720



- Dans un repère orthogonal (unités graphiques : 1 cm pour 1 trimestre sur l'axe des abscisses et 1 cm pour 100 milliers d'acquéreurs sur l'axe des ordonnées), construire le nuage de points $M_i(x_i; y_i)$ de cette série statistique.

Construire ensuite la droite (M_1M_6) , la droite Δ_1 d'équation $y = 120x + 10$ et la droite Δ_2 d'équation $y = 110x + 35$.

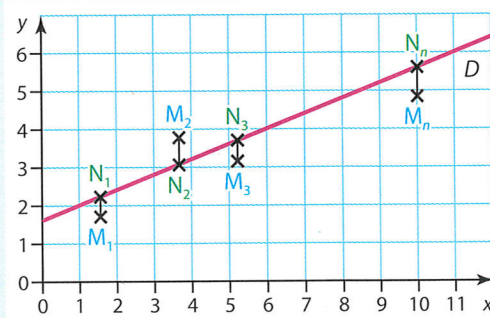
- Que dire de ces droites par rapport au nuage de points ?
- Déterminer l'équation réduite $y = ax + b$ de la droite (M_1M_6) .

2. On appelle somme des carrés des écarts verticaux la somme $S = (M_1N_1)^2 + (M_2N_2)^2 + (M_3N_3)^2 + \dots + (M_nN_n)^2$ où M_1, M_2, \dots, M_n sont les points du nuage de points d'une série statistique et N_1, N_2, \dots, N_n les points de D de mêmes abscisses respectives que M_1, M_2, \dots, M_n .

- Ouvrir le fichier Excel : lienmini.fr/10445-29.

Dans les cellules C10 à C15, saisir les valeurs du coefficient directeur a et de l'ordonnée à l'origine b obtenues à la question 1. pour chacune des trois droites.

- Dans les cellules D3 à I3, saisir une formule que l'on copiera vers le bas, permettant de compléter jusqu'à la ligne 8.
- Dans les cellules E10, E12 et E14, saisir la formule permettant d'afficher la somme S .
- Si l'on doit choisir la droite passant le plus près des points du nuage, laquelle choisir ? Justifier.



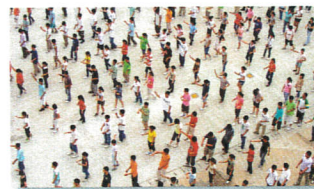
3 Évolution de la population en Chine

DÉMOGRAPHIE

OBJECTIF Utiliser un ajustement affine pour faire des estimations → Cours 2 p. 102

On donne dans le tableau ci-dessous le nombre d'habitants en Chine de l'année 2010 à l'année 2018.

Année : x_i	2010	2011	2012	2013	2014	2015	2016	2017	2018
Nombre d'habitants en milliards : y_i	1,338	1,344	1,351	1,357	1,364	1,371	1,379	1,386	1,393



1. Construire le nuage de points de cette série statistique (unités graphiques : 1 cm pour 1 année sur l'axe des abscisses en prenant pour origine 2008 et 1 cm pour 0,01 milliard d'habitants sur l'axe des ordonnées en prenant pour origine 1,3 milliard).
2. Soit D la droite d'équation $y = 0,007x - 12,6$. Construire la droite D sur le graphique précédent. Cette droite réalise-t-elle un bon ajustement affine du nuage ?
3. a. Déterminer par le calcul le nombre d'habitants prévisible en Chine en 2025.
b. Déterminer par le calcul l'année à partir de laquelle on peut estimer que le nombre d'habitants en Chine va dépasser 1,45 milliard.
4. Retrouver graphiquement les résultats obtenus aux deux questions précédentes.

4 Temps de montage d'un véhicule

INDUSTRIE

OBJECTIF Faire un changement de variable pour se ramener à un ajustement affine → Cours 2 p. 102

On a relevé durant 6 années consécutives le temps de montage moyen d'un véhicule donné. Voici les résultats obtenus :

Année : x_i	2013	2014	2015	2016	2017	2018
Temps de montage en heures : y_i	24,8	21,3	18,5	16,7	15,5	14,8



1. Construire le nuage de points $M(x_i; y_i)$ de cette série dans un repère orthogonal (unités graphiques : 1 cm pour 1 année sur l'axe des abscisses en prenant pour origine 2010 et 1 cm pour 1 heure sur l'axe des ordonnées en prenant pour origine 14 heures).
Un ajustement affine de ce nuage est-il envisageable ?
2. On décide de remplacer les valeurs de y par leur racine carrée ; on définit ainsi une nouvelle variable : $z = \sqrt{y}$.
a. Recopier et compléter le tableau suivant (résultats arrondis à 10^{-2} près) :

x_i	2013	2014	2015	2016	2017	2018
z_i						
- b. Dans un nouveau repère orthogonal (unités graphiques : 1 cm pour 1 année sur l'axe des abscisses en prenant pour origine 2010 et 1 cm pour 0,1 sur l'axe des ordonnées en prenant pour origine 3,8), construire le nuage de points $N(x_i; z_i)$.
Un ajustement affine est-il envisageable avec ce nouveau nuage de points ?
- c. À l'aide de la calculatrice, déterminer une équation de la droite d'ajustement de z en x par la méthode des moindres carrés (arrondir les coefficients à 10^{-2} près).
- d. En déduire une expression de y en fonction de x puis estimer le temps de montage moyen de ce modèle de véhicule en 2025.

1 Séries statistiques à deux variables

A Définition

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 ^{er} caractère	x_1	x_2	...	x_n
Valeurs du 2 nd caractère	y_1	y_2	...	y_n

REMARQUE • Dans ce chapitre, la calculatrice sera indispensable. Par exemple, pour saisir les données du tableau statistique.

B Nuage de point d'une série statistique à deux variables et point moyen du nuage de points

DÉFINITION Dans un repère orthogonal du plan, l'ensemble des points $M_i(x_i; y_i)$ pour i allant de 1 à n est appelé le **nuage de points** de la série statistique.

EXEMPLE • Le tableau suivant présente l'évolution du nombre de clients et du chiffre d'affaires en milliers d'euros d'une micro-entreprise au cours des 6 mois :

Nbre de clients : x_i	7	9	10	13	15	18
Chiffre d'affaires : y_i	30	37	45	56	69	81

On place les points de coordonnées $(x_i; y_i)$ pour obtenir le nuage de points.

→ Voir **Exercice résolu 1**

DÉFINITION Soit \bar{x} la moyenne des valeurs prises par le 1^{er} caractère et \bar{y} celle des valeurs prises par le 2nd caractère. On appelle **point moyen du nuage de points** de la série statistique, le point G de coordonnées $(\bar{x}; \bar{y})$.

EXEMPLE • Dans la situation ci-dessus, $\bar{x} = (7 + 9 + 10 + 13 + 15 + 18) : 6 = 12$ et $\bar{y} = (30 + 37 + 45 + 56 + 69 + 81) : 6 = 53$.

Le point moyen G du nuage de points a pour coordonnées (12 ; 53).

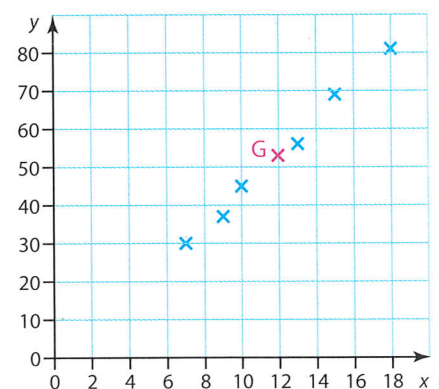
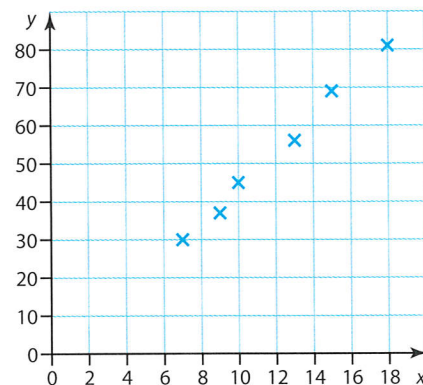
REMARQUE • La calculatrice détermine les coordonnées du point moyen d'un nuage de points, une fois les données du tableau saisies dans la calculatrice :

	Sur TI	Sur Casio
Instructions	<p>listes stats puis aller jusqu'à la colonne CALC, sélectionner 2 : Stats 2 Var, descendre jusqu'à « Calculer »</p> <p>puis <small>précédent</small> entrer</p>	<p>F2 : CALC puis F2 : LEVAR</p>

→ Voir **Exercice résolu 2**

Notation

Lorsqu'on donne une **série statistique à deux variables**, on note généralement x_1, x_2, \dots, x_n les valeurs prises par le 1^{er} caractère et y_1, y_2, \dots, y_n celles prises par le 2nd caractère. Les valeurs prises par cette série sont alors les couples $(x_1; y_1), (x_2; y_2), \dots, (x_n; y_n)$.



Exercice résolu

1

Représenter un nuage de points

Une personne court sur un tapis roulant dont la vitesse peut être modifiée. On a relevé sa fréquence cardiaque en battements par minute selon l'intensité du travail fourni, exprimée en kilojoules. Voici les résultats obtenus :

Intensité du travail fourni : x_i (en kJ)	10	12	15	22	34	45	53	60
Fréquence cardiaque : y_i (en battements·min ⁻¹)	69	80	88	97	114	126	145	158

Construire le nuage de points représentant cette série statistique dans un repère orthogonal. On prendra comme unités graphiques 1 cm pour 10 kJ sur l'axe des abscisses et 1 cm pour 20 battements·min⁻¹ sur l'axe des ordonnées en prenant pour origine 60 battements·min⁻¹.

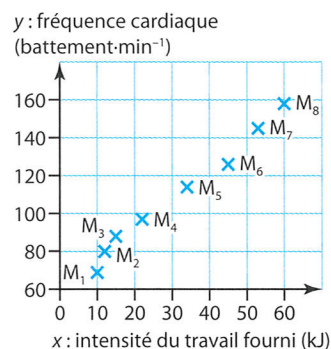
Solution

On place successivement les points de coordonnées (10 ; 69), (12 ; 80), (15 ; 88), (22 ; 97), (34 ; 114), (45 ; 126), (53 ; 145), (60 ; 158).

Méthode

Pour représenter un nuage de points

- 1 On **construit** les deux axes orthogonaux. On les **gradue** en faisant attention aux unités imposées et aux éventuelles indications sur le début des graduations.
- 2 On **légende** les axes (axe des abscisses : 1^{er} caractère étudié ; axe des ordonnées : 2nd caractère étudié).
- 3 On **place** les points $M_i(x_i ; y_i)$.
Attention : on ne les relie pas entre eux !



→ Voir **Exercices 20 à 27** p. 106

Exercice résolu

2

Déterminer les coordonnées du point moyen d'un nuage de points

Le tableau ci-dessous donne la production en France de véhicules de marque française entre 2013 et 2018 :

Année : x_i	2013	2014	2015	2016	2017	2018
Nombre de véhicules produits : y_i (en milliers)	1 445	1 503	1 656	1 753	1 908	1 937

Source : https://ccfa.fr/wp-content/uploads/2018/09/analyse_statistiques_2018_fr.pdf.

Calculer les coordonnées du point moyen G de cette série statistique.

Solution

$$\bar{x} = \frac{2013 + 2014 + 2015 + 2016 + 2017 + 2018}{6} = 2\,015,5$$

Donc G(2 015,5 ; 1 700,3).

Méthode

Pour déterminer les coordonnées du point moyen d'un nuage de points

- 1 On **calcule** \bar{x} en additionnant toutes les valeurs prises par le 1^{er} caractère et en divisant le résultat par le nombre de valeurs.
- 2 On **calcule** \bar{y} de la même façon mais en travaillant avec le 2nd caractère.
- 3 On **conclut** en donnant les coordonnées de G (vérifier avec la calculatrice).

$$\bar{y} = \frac{1\,445 + 1\,503 + 1\,656 + 1\,753 + 1\,908 + 1\,937}{6} \approx 1\,700,3$$

→ Voir **Exercices 28 à 34** p. 107

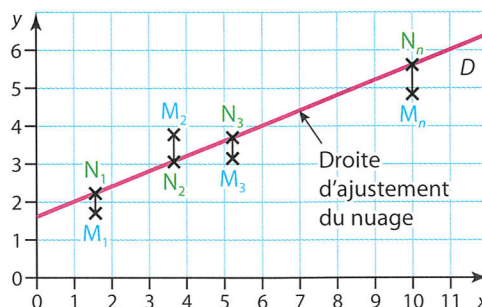
2 Ajustement affine

On cherche s'il existe une relation, certes approximative, entre les deux variables x et y d'une **série statistique**. L'allure du **nuage de points** représentant la série statistique permet d'avoir une idée de la réponse. En effet, la disposition des points peut parfois suggérer la courbe d'une fonction f . Cela signifie alors que le lien approximatif entre x et y est de la forme $y = f(x)$.

On est donc amené à chercher une courbe « passant au plus près » des points du nuage : on parle alors d'**ajustement** du nuage.

Dans le cas particulier où les points du nuage sont proches de l'alignement, la courbe cherchée est une droite et on dit alors que cette droite réalise un **ajustement affine du nuage**.

Si la forme d'un nuage de points $M_i(x_i; y_i)$ suggère un ajustement affine, on peut trouver un grand nombre de droites ajustant ce nuage.



La **méthode des moindres carrés** consiste à déterminer la droite rendant la somme $(M_1N_1)^2 + (M_2N_2)^2 + (M_3N_3)^2 + \dots + (M_nN_n)^2$ la plus petite possible.

On admettra que cette droite **existe** et qu'elle est **unique**.

DÉFINITION La droite obtenue par la méthode des moindres carrés est appelée la **droite d'ajustement de y en x par la méthode des moindres carrés**.

REMARQUE • Pour obtenir les coefficients a et b de l'équation réduite $y = ax + b$ de cette droite, on utilise la calculatrice (ou un tableur).

Une fois les données du tableau saisies sur la calculatrice :

	Sur TI	Sur Casio
Instructions	<p>puis aller jusqu'à la colonne CALC, sélectionner 4 : RégLin(ax+b), descendre jusqu'à Calculer puis entrer.</p>	<p>F2 : CALC puis F3 : REG F1 : XY puis F1 : ax+b</p>

→ Voir **Exercice résolu 3**

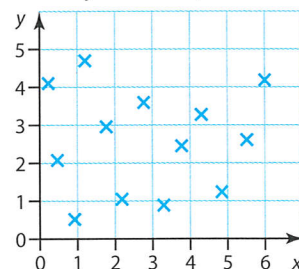
PROPRIÉTÉ (admise) La **droite d'ajustement de y en x par la méthode des moindres carrés passe par le point moyen** du nuage de points de la série statistique.

REMARQUE • En déterminant l'équation de la droite d'ajustement $y = ax + b$ (et donc un lien approximatif entre x et y), on va pouvoir faire des estimations de valeurs de y correspondant à des valeurs de x données et vice versa.

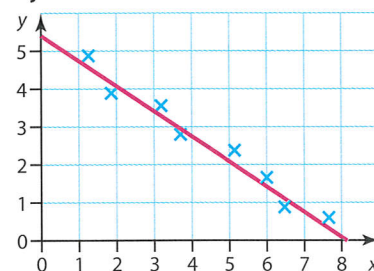
Ces estimations peuvent se faire soit graphiquement après avoir tracé la droite, soit par le calcul.

→ Voir **Exercice résolu 4**

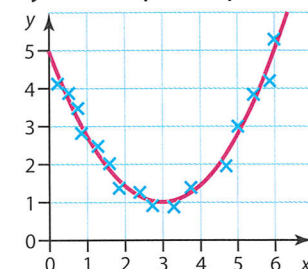
Pas d'ajustement particulier



Ajustement affine



Ajustement par une parabole



Exercice résolu

3

Déterminer un ajustement affine

Le tableau ci-dessous donne l'évolution du SMIC horaire brut de 2015 à 2019.

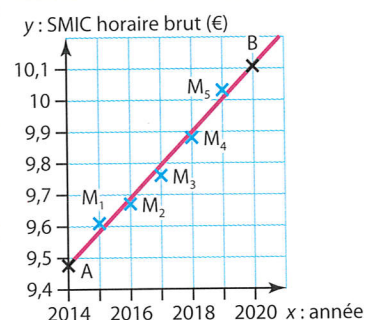
Année : x_i	2015	2016	2017	2018	2019
SMIC horaire : y_i (en €)	9,61	9,67	9,76	9,88	10,03

Source : <https://www.insee.fr/fr/statistiques/1375188>

1. Représenter le nuage de points de la série statistique dans un repère orthogonal. On prendra comme unités graphiques 1 cm pour 1 an sur l'axe des abscisses en prenant pour origine 2014 et 10 cm pour 1 € sur l'axe des ordonnées en prenant pour origine 9,4 €.
2. Déterminer l'équation réduite de la droite Δ d'ajustement de y en x par la méthode des moindres carrés. Les coefficients a et b seront arrondis à 10^{-3} près.
3. Représenter la droite Δ sur le graphique de la question 1.

Solution

1. Voir ci-contre.
2. La calculatrice donne $a \approx 0,105$ et $b \approx -201,995$.
L'équation réduite de la droite Δ est donc :
$$y = 0,105x - 201,995$$
3. Pour représenter la droite Δ , on détermine les coordonnées de deux points lui appartenant. En prenant par exemple $x = 2\,014$, on obtient : $y = 9,475$ et en prenant $x = 2\,020$, on obtient : $y = 10,105$.
Donc la droite Δ passe par A(2 014 ; 9,475) et par B(2 020 ; 10,105).



→ Voir Exercices 35 à 39 pp. 107-108

Exercice résolu

4

Utiliser un ajustement affine pour interpoler ou extrapoler des valeurs inconnues

On prend l'énoncé de l'Exercice résolu 3 : on souhaite faire des estimations du SMIC horaire brut en supposant que la droite Δ obtenue donne une approximation satisfaisante de l'évolution de y en fonction de x jusqu'en 2030.

1. Déterminer, par le calcul, le SMIC horaire brut estimé pour l'année 2025.
2. Déterminer, par le calcul, à partir de quelle année on peut estimer que le SMIC horaire brut dépassera 10,90 €.

Solution

1. Pour $x = 2025$, on a $y = 0,105 \times 2025 - 201,995 = 10,63$. En 2025, le SMIC horaire brut est donc estimé à 10,63 €.
2. On résout l'inéquation $0,105x - 201,995 > 10,9$. Elle est équivalente à : $x > \frac{10,9 + 201,995}{0,105}$ soit $x > \frac{14\,193}{7}$. Or $\frac{14\,193}{7} \approx 2\,027,6$. C'est donc à partir de l'année 2028 que l'on peut estimer que le SMIC horaire brut dépassera 10,90 €.

Méthode

Pour utiliser un ajustement affine pour interpoler ou extrapoler des valeurs inconnues

- 1 Si l'énoncé donne une valeur de x , on remplace x par cette valeur dans $y = ax + b$, et on calcule y .
- 2 Si l'énoncé donne une valeur de y , on remplace y par cette valeur dans $y = ax + b$, et on résout l'équation (ou l'inéquation) pour obtenir x .

→ Voir Exercices 40 à 46 p. 108

1 Séries statistiques à deux variables

● Une **série statistique à deux variables**, x et y , se présente sous la forme d'un tableau statistique :

Valeurs du 1 ^{er} caractère : x_i	x_1	x_2	...	x_n
Valeurs du 2 nd caractère : y_i	y_1	y_2	...	y_n

● Le **nuage de points** de la série statistique est l'ensemble des points $M_i(x_i; y_i)$ pour i allant de 1 à n .

● Le **point moyen du nuage** est le point $G(\bar{x}; \bar{y})$ où :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad \text{et} \quad \bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n}.$$

2 Ajustement affine

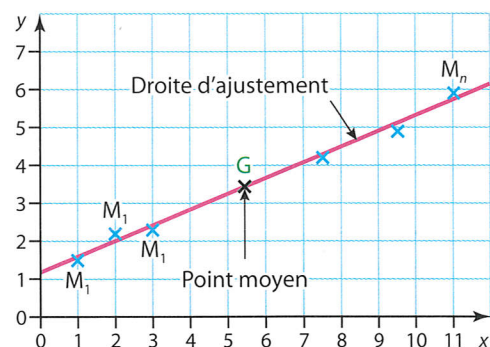
● Dans le cas particulier où les points du nuage sont proches de l'alignement, on peut réaliser **un ajustement affine du nuage de points**, c'est-à-dire trouver une droite passant au plus près des points du nuage. Il en existe une multitude.

● La **droite d'ajustement de y en x par la méthode des moindres carrés** est la droite passant au plus près des points du nuage.

● Elle passe par le **point moyen G** .

● Son équation $y = ax + b$ permet d'obtenir une relation approximative entre y et x .

● Les **valeurs de a et b** sont obtenues grâce à la **calculatrice** ou le **tableur**.



3 Avec la calculatrice

	Sur TI	Sur Casio
Saisir les données du tableau statistique	<p><small>stats</small> puis dans la colonne ÉDIT, sélectionner 1 : Modifier ... pour rentrer les valeurs du 1^{er} caractère dans L_1 et celles du 2nd caractère dans L_2.</p>	<p><small>MENU</small> <small>STAT</small> <small>EXE</small></p> <p>Rentrer les valeurs du 1^{er} caractère dans L_1 et celles du 2nd caractère dans L_2.</p>
Lire les coordonnées du point moyen du nuage de points	<p><small>stats</small> puis aller jusqu'à la colonne CALC, sélectionner 2 : Stats 2 Var, descendre jusqu'à Calculer puis <small>entrée</small>.</p>	<p>F2 : CALC puis F2 : VAR.</p>
Déterminer l'équation de la droite d'ajustement de y en x par la méthode des moindres carrés	<p><small>stats</small> puis aller jusqu'à la colonne CALC, sélectionner 4 : RégLin(ax+b), descendre jusqu'à Calculer puis <small>entrée</small>.</p>	<p>F2 : CALC puis F3 : REG F1 : X puis F1 : Y.</p>