9

Statistiques à deux variables

Le prix du paquet de cigarettes et le nombre de paquets de cigarettes vendus font régulièrement l'objet de statistiques qui peuvent aider à lutter contre le tabagisme.

Existe-t-il un lien de corrélation entre le prix d'un paquet de cigarettes et le nombre de paquet vendus ?

↪ Exercice 33 p. 238

VIDÉO

Lutter contre le tabagisme
lienmini.fr/maths-c09-01

Pour prendre un bon départ

EXO Prérequis lienmini.fr/maths-c09-02

Les rendez-vous Sésamath

1 Calculer les caractéristiques d'une série statistique à une variable (1)

1. Calculer la moyenne, la médiane, la variance et l'écart-type de la série statistique suivante.

Valeurs de x_i	25	37	54	62	71	78

2. Vérifier les résultats à l'aide du menu **Stats** de la calculatrice.

2 Calculer les caractéristiques d'une série statistique à une variable (2)

Dans le tableau suivant on a noté la pointure des élèves d'une classe de 3e

Pointures x_i	36	37	38	39	40	41	43	44	45
Effectifs n_i	1	3	6	6	6	4	3	2	1

À l'aide de la calculatrice déterminer la moyenne, la médiane, la variance et l'écart-type de la série statistique.

3 Tracer une droite

On considère un repère orthogonal dont les axes sont gradués de telle sorte qu'une unité sur l'axe des abscisses correspond à 4 unités sur l'axe des ordonnées.

1. Dans ce repère, tracer la droite d d'équation $y = 4x + 12$.

2. Le point M(11 ; 66) appartient-il à d ?

3. Quelle est l'ordonnée du point de d d'abscisse 25 ?

4. Quelle est l'abscisse du point de d d'ordonnée 204 ?

4 Déterminer l'équation d'une droite donnée

Dans un repère orthogonal du plan, on considère les points A et B de coordonnées respectives (– 2 ; 11) et (5 ; 1).

Déterminer l'équation réduite de la droite (AB).

5 Étudier une fonction exponentielle

On considère la fonction f définie sur $\mathbb{R}$ par $f(x) = e^{0,5x+6}$.

Et on appelle $\mathscr{C}_f$ sa courbe représentative dans un repère orthogonal $(O\,;\vec{i}\,,\vec{j})$.

1. Représenter $\mathscr{C}_f$ à l'aide de la calculatrice graphique.

2. Calculer $f(2)$.

3. Résoudre l'équation $f(x) = 2\,500$.

Physique 30 min Thème 9

1 Nuage de points, point moyen

A ▸ Loi d'ohm

Lors d'un TP de Physique le professeur a demandé à une élève de mesurer conjointement, la tension *U* aux bornes d'un conducteur ohmique et, l'intensité *I* du courant qui le traverse. Elle a trouvé les valeurs ci-contre.

I (en mA)	0	4	7	15	25	30
U(en V)	0	1	2	4	6	7

Ces valeurs conjointes forment **une série statistique double à deux variables** *I* et *U*.

1. Dans un repère orthogonal du plan, placer les points $M_1(0 ; 0)$, $M_2(4 ; 1)$, $M_3(7 ; 2)$, $M_4(15 ; 4)$, $M_5(25 ; 6)$ et $M_6(30 ; 7)$ associés à cette série statistique. (Attention, il ne faut pas les relier !)

Cet ensemble de points est appelé **nuage de points** de la série statistique.

2. Calculer la moyenne $\overline{x}$ des valeurs de la variable *X*, puis la moyenne $\overline{y}$ des valeurs de la variable *Y*. Placer dans le repère le point G de coordonnées ($\overline{x}$; $\overline{y}$). Ce point est appelé **point moyen** du nuage.

3. Les points du nuage forment-ils un dessin particulier (droite, arc de cercle, parabole, ou autre courbe caractérisant une fonction usuelle) ?

B ▸ D'autres statistiques à deux variables

Faire de même avec chacune des séries statistiques à deux variables suivantes.

1. Un élève a obtenu chaque mois lors du premier semestre les notes x_i en mathématiques et les notes y_i en anglais.

Notes x_i en maths	13	8	14	10	7	6
Notes y_i en anglais	17	20	14	6	11	10

2. Une entreprise étudie l'évolution conjointe des quantités produites x_i (en tonnes) et du coût moyen y_i de fabrication d'une tonne de produit (en milliers d'euros).

Quantité x_i (en tonnes)	0,1	0,5	1	1,5	2	3	3,5	4
Coût moyen y_i (en milliers d'euros)	8	5	3	1,5	1	2	4	7

▶ **Remarque :** L'étude de séries statistiques à deux variables peut mener, lorsque cela a du sens, à réfléchir à un éventuel lien de corrélation ou de cause à effet entre les deux variables.

↪ Cours 1 p. 224

TICE 10 min

2 Ajustement affine avec un tableur

Le tableau ci-contre présente l'évolution du budget publicitaire et du chiffre d'affaires d'une société au cours des 9 dernières années.

Budget publicitaire x_i (en millions d'euros)	2,5	3	3,4	3,9	4	5	5,5	6,2	7
Chiffre d'affaire y_i (en millions d'euros)	13	18	17	19	19	24	25	31	36

1. a) Ouvrir un tableur et entrer les valeurs du tableau ci-dessus dans les colonnes A et B.
b) Insérer un graphique **Nuage de points**.

2. Que peut-on dire de ces points ?

3. Effectuer un clic droit sur un point du nuage. Sélectionner **Ajouter une courbe de tendance** dans le menu contextuel. Dans la fenêtre qui s'affiche ensuite, sélectionner **linéaire.**

Coup de pouce (On pourra aussi éventuellement activer les cases **Afficher l'équation** et **Afficher le coefficient de détermination** qui donnent des informations utiles sur la droite et la série statistique.)

▶ **Remarque :** La droite obtenue ajuste au mieux ce type de nuage.
On dit qu'on vient de réaliser un **ajustement affine** du nuage.

↪ Cours 2 p. 226

TICE 15 min

3 Ajustement affine avec GeoGebra

Une youtubeuse a lancé sa chaîne de décoration en septembre 2019. Depuis cette date, elle a essayé de noter, de temps en temps, le nombre d'abonnés à sa chaîne. Les résultats sont présentés dans le tableau ci-dessous dans lequel chaque mois, où elle a relevé le nombre d'abonnés, est numéroté par son rang depuis le mois de septembre, considéré comme rang initial 0.

Mois	Novembre	Décembre	Janvier	Mars	Mai	Juin	Juillet	Septembre
Rang du mois x_i	2	3	4	6	8	9	10	12
Nombre d'abonnés n_i	56	61	72	95	150	207	312	560

TICE Ajustement affine lienmini.fr/maths-c09-03

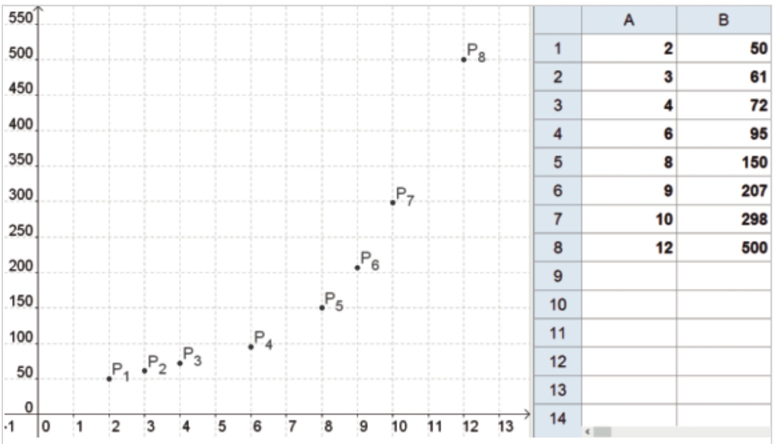
1. Télécharger le fichier **GeoGebra**.

2. Que pouvez-vous dire du nuage de points de cette série statistique double ?

3. Dans la colonne C du tableur, calculer le logarithme népérien de chaque valeur n_i de la série du nombre d'abonnés (colonne B)

4. Sélectionner les colonnes A et C puis créer le nuage de points associé à l'aide d'un clic droit **Créer une liste de points.**

Coup de pouce **Remarque :** Ces points étant placés beaucoup plus bas dans le repère, il faut changer l'échelle à l'aide d'un clic droit au milieu du graphique, et sélectionner **axeX : axeY,** puis, **1 : 1.**



Qu'observe-t-on?

5. Utiliser alors l'icone , puis sélectionner tous les points du nuage pour obtenir la droite de régression.

6. Lire dans la fenêtre **Algèbre** l'équation réduite de la droite. On obtient ainsi une relation (ou fonction) qui lie (approximativement) les valeurs de la colonne C que l'on nommera y_i avec les valeurs x_i de la colonne A.

7. Sachant que $y_i = \ln(n_i)$, en déduire une relation (ou fonction) qui lie (approximativement) les valeurs y_i avec les valeurs x_i.

8. Représenter la courbe de cette fonction sur le graphique en utilisant la barre de saisie : « f(x)=......

↪ Cours 2 p. 226

05 min

4 Interpolation et extrapolation

Dans l'activité précédente, on a vu que le nombre d'abonnés n d'une youtubeuse, après x mois en partant de septembre 2019, pouvait être ajusté par la relation $n = e^{0,23x+3,36}$.

1. Elle n'a pas relevé le nombre d'abonnés sur sa chaîne au mois d'août, 11 mois après son lancement. À l'aide de l'ajustement obtenu, donner une estimation du nombre d'abonnés au mois d'août.

2. Si la tendance se poursuit, combien d'abonnés peut-elle espérer avoir en janvier 2025 ?

↪ Cours 4 p. 230

Cours

1 Généralités

Définition Série statistique à deux variables quantitatives

Lorsque l'on étudie conjointement deux caractères (ou variables) x et y sur une même population de taille n, on associe à chaque individu de la population un couple $(x_i\,;y_i)$, où x_i et y_i sont les valeurs respectives des variables x et y prises par l'individu « numéro i » (où i est un nombre entier entre 1 et n, ou parfois entre 0 et $n-1$).
On appelle série statistique double (x,y) l'ensemble des couples $(x_i\,;y_i)$ associés à chaque individu de la population.
On la présente en général dans un tableau.

Valeurs x_i de la variable x	x_1	x_2	x_3	x_4	...	x_i	...	x_n
Valeurs y_i de la variable y	y_1	y_2	y_3	y_4	...	y_i	...	y_n

Remarque La liste de valeurs associées à la variable x est une série statistique simple dont on peut calculer la moyenne $\overline{x}$. Il en va de même pour les valeurs de y dont la moyenne est $\overline{y}$:

$$\overline{x}=\frac{x_1+x_2+\ldots\ldots\ldots+x_n}{n} \quad \text{et} \quad \overline{y}=\frac{y_1+y_2+\ldots\ldots\ldots+y_n}{n}$$

Exemple

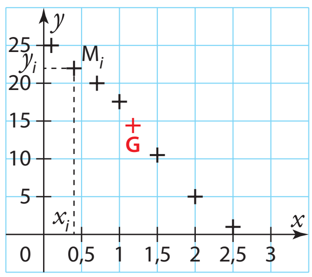
Une équipe de météorologues envoie un ballon sonde à partir du sol de leur station située à environ 100 m d'altitude. Le ballon est équipé d'un thermomètre et d'un altimètre afin de relever simultanément la température atmosphérique et l'altitude atteinte.

Altitude x (en km)	0,1	0,4	0,7	1	1,5	2	2,5
Température y (en °C)	25	22	20	17,5	10,5	5	1

Les données relevées forment une série statistique double (x, y).
Ici, $x_1=0{,}1$ et $y_1=25$; $x_2=0{,}4$ et $y_2=22$; etc…

Définition Nuage de points

À chaque couple $(x_i\,;y_i)$ de la série statistique double (x,y) on peut associer le point M_i de coordonnées $(x_i\,;y_i)$ dans un repère.
L'ensemble de ces points est appelé nuage de points associé à la série statistique double (x, y).



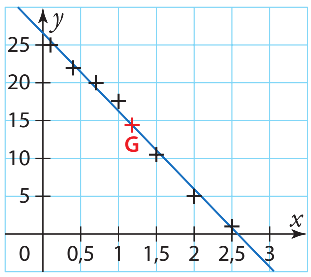
Définition Point moyen

On appelle point moyen du nuage, le point G de coordonnées $(\overline{x}\,;\overline{y})$

Exemple

Le point moyen de la série à deux variables (altitude et température) précédente est G(1,2 ; 14,4) car :

$$\overline{x}=\frac{0{,}1+0{,}4+\ldots+2{,}5}{7}\approx 1{,}2 \quad \text{et} \quad \overline{y}=\frac{25+22+\ldots+1}{7}\approx 14{,}4$$



On cherche s'il existe un lien entre ces deux variables, l'altitude et la température. On va donc essayer de trouver une courbe qui « approche au mieux » le nuage, c'est-à-dire une courbe qui passe au plus près des points du nuage.
On dit que l'on a effectué un **ajustement**.
Cette courbe d'ajustement, si elle existe, représente alors une fonction f qui permet quasiment d'exprimer la variable y en fonction de la variable x sous la forme $y=f(x)$.
Ici, les points sont presqu'alignés donc on peut ajuster le nuage par une droite : on a donc quasiment une relation du type $y=ax+b$ entre les deux variables x et y de la série statistique.

EXOS Méthodes lienmini.fr/maths-c09-04 Les rendez-vous Sésamath

Exercices résolus

Méthode 1 Représenter une série statistique double

Énoncé

Chaque année une association organise une tombola. Elle achète un certain nombre de lots puis vend des billets de tombola par carnets de 10. Le nombre de lots achetés et le nombre de carnets vendus lors des 9 dernières tombolas est donné dans le tableau ci-contre

Nombre de lots x_i	38	26	20	13	8	34	32	41
Nombre de carnets y_i	69	64	60	59	55	67	68	70

1. Représenter cette série sous la forme d'un nuage de points dans un repère orthonormé.
2. Calculer les coordonnées du point moyen de ce nuage et le placer dans le repère.
3. Quelle forme présente le nuage de points ?
4. Tracer « au jugé » la courbe associée.

Solution

1. **1** On obtient la représentation graphique ci-dessous.

2. $\overline{x} = \dfrac{38+26+20+13+8+34+32+41}{8} = 26{,}5$

et $\overline{y} = \dfrac{69+64+60+59+55+67+68+70}{8} = 64$

Le point moyen a donc pour coordonnées G (26,5 ; 64).

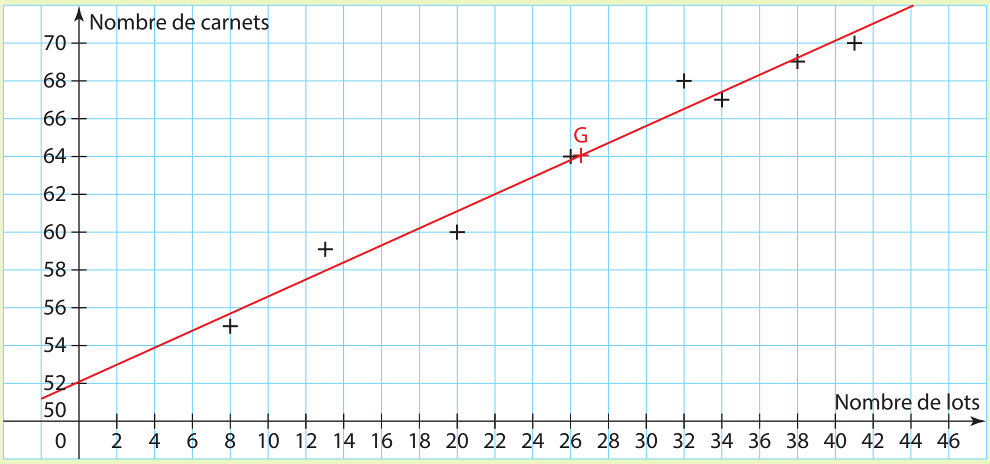
3. Le nuage de points a une forme « allongée, rectiligne, montante ».

4. La courbe usuelle qui se rapproche le plus de sa forme globale est une droite. **2**

Conseils & Méthodes

1 Étant donné les valeurs de la série y, il est plus pratique de commencer à graduer l'axe des ordonnées à partir de 50, afin de garder une échelle (1 carreau pour 2 unités) suffisamment précise.

2 Pour tracer la droite, on place sa règle sur le point moyen et on « ajuste » au mieux la « pente » de façon à passer, en moyenne, le plus proche possible de tous les points.



À vous de jouer !

1 Soit la série statistique suivante à deux variables x et y.

Valeurs de x	5	7	9	12	13	16
Valeurs de y	27	23	18	13	13	6

1. Représenter cette série sous la forme d'un nuage de points dans un repère orthogonal.
2. Calculer les coordonnées du point moyen de ce nuage et le placer dans le repère.
3. Quelle forme présente le nuage de points ?
4. Tracer « au jugé » la courbe associée.

2 Le tableau présente le chiffre d'affaires annuel (en millions d'euros) d'un fabricant de micro-processeurs entre 2002 (année de rang 0) et 2005 (année de rang 3).

Valeurs x_i	0	1	2	3
Valeurs y_i	12	19	31	58

1. Représenter cette série sous la forme d'un nuage de points dans un repère orthogonal.
2. Calculer les coordonnées du point moyen de ce nuage et le placer dans le repère.
3. Quelle forme présente le nuage de points ?
4. Tracer « au jugé » la courbe associée.

↪ Exercices 20 à 21 p. 236

Cours

2 Ajustement affine

Définition **Droite d'ajustement, ajustement affine**

Lorsque les points du nuage sont sensiblement alignés, on appelle droite d'ajustement une droite qui passe au plus près des points du nuage. On dit que cette droite réalise un ajustement affine du nuage de points.

Remarques

① La variable y peut alors s'exprimer de façon approchée sous la forme $y = ax + b$ (équation de la droite d'ajustement). Ainsi on estime qu'il existe un lien affine entre x et y.

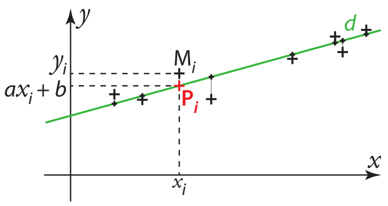
② Il est généralement admis que la droite d'ajustement passe par le point moyen G du nuage.

③ Pour obtenir cette droite il existe différentes méthodes l'une d'elles est présentée ci-après.

Propriété **Méthode des moindres carrés**

Soit un nuage de points $M_i(x_i\,;y_i)_{1\leqslant i\leqslant n}$, une droite d d'équation $y = ax + b$ qui ajuste le nuage, et les points $P_i(x_i\,;ax_i + b)$ sur la droite d. Il existe un réel a et un réel b tels que la somme $M_1P_1{}^2 + M_2P_2{}^2 + \ldots + M_nP_n{}^2$ est minimale.

Pour ces valeurs de a et de b, la droite d'équation $y = ax + b$ est appelée droite des moindres carrés associée au nuage de points $M_i(x_i\,;y_i)_{1\leqslant i\leqslant n}$ (ou droite de régression de y en x).



Démonstration

VIDÉO
Démonstration
lienmini.fr/maths-c09-05

Apprendre à démontrer p. 236

Remarque L'idée de la méthode des moindres carrés est de trouver une droite d passant « en moyenne » au plus près des points du nuage.
$M_iP_i = y_i - (ax_i + b)$ représente l'écart entre le point M_i du nuage et la droite.
Lorsqu'on fait la somme des carrés de ces écarts, on cherche à ce que le résultat soit le plus petit possible.
Il s'agit donc de trouver les nombres a et b qui minimisent cette quantité.

Propriété **Droite des moindres carrés et covariance**

La droite des moindres carrés a pour équation $y = ax + b$ avec $a = \dfrac{\text{cov}(x, y)}{\text{var}(x)}$ et $b = \overline{y} - a\overline{x}$.

$\text{cov}(x, y)$ est la covariance des séries statistiques x et y et se calcule de la façon suivante :

$$\text{cov}(x, y) = \frac{(x_1 - \overline{x})(y_1 - \overline{y}) + (x_2 - \overline{x})(y_2 - \overline{y}) + \ldots\ldots + (x_n - \overline{x})(y_n - \overline{y})}{n}$$

Remarques

① On peut obtenir les coefficients a et b de la droite des moindres carrés directement avec la calculatrice.

② La covariance $\text{cov}(x, y)$ peut-être positive ou négative. On peut la noter aussi σ_{xy} et pour les calculs on pourra aussi utiliser :

$$\text{cov}(x, y) = \sigma_{xy} = \frac{x_1y_1 + x_2y_2 + \ldots + x_ny_n}{n} - \overline{xy}$$

Définition **Coefficient de corrélation linéaire**

Le coefficient de corrélation linéaire d'une série statistique de variables x et y est le nombre r défini par :

$$r = \frac{\sigma_{xy}}{\sigma(x)\sigma(y)}$$

Propriété **Modèle de corrélation et ajustement affine**

Le coefficient de corrélation linéaire r vérifie : $-1 \leqslant r \leqslant 1$:

- **plus $|r|$ est proche de 1, plus l'ajustement affine est un bon modèle de corrélation entre les variables x et y.**
- **plus $|r|$ est proche de 0, moins l'ajustement n'a de sens.**
- **Si $|r| = 1$, c'est que la droite de régression passe par tous les points du nuage.**

EXOS Méthodes lienmini.fr/maths-c09-04 Les rendez-vous Sésamath

Exercices résolus

Méthode 2 Obtenir la droite de régression et le coefficient de corrélation linéaire à la calculatrice

Énoncé

On considère le nuage de points associé à la série statistique de la Méthode 1 p.225 À l'aide de la calculatrice, déterminer l'équation de la droite d'ajustement du nuage par la méthode des moindres carrés et le coefficient de corrélation linéaire.
(Les valeurs seront arrondies à 10^{-3}).

x_i	38	26	20	13	8	34	32	41
y_i	69	64	60	59	55	67	68	70

Conseils & Méthodes

Sur la calculatrice on peut aussi représenter le nuage de points. Cela permet de s'assurer qu'un ajustement affine a du sens.

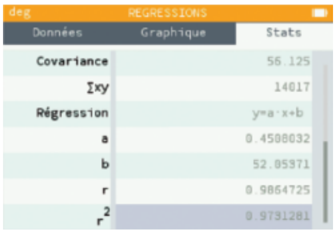
Solution

TI-83 Premium CE

Appuyer sur la touche stats de la calculatrice Sélectionner le menu **ÉDIT 1:Modifier** puis entrer les valeurs de la série statistique dans les colonnes L1 et L2.
Appuyer de nouveau sur la touche stats et sélectionner le menu **CALC 4 :RégLin(ax+b)**.
S'assurer qu'on a **Xliste : L1** et **Yliste : L2.** On obtient alors l'équation de la droite de régression de y en x, ainsi que le coefficient de corrélation linéaire *r*.

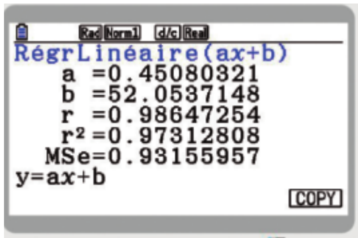
NUMWORKS

Appuyer sur la touche de la calculatrice. Sélectionner le menu Régressions puis entrer les valeurs de la série statistique dans les colonnes X1 et Y1. Appuyer sur la touche . Puis sélectionner le menu **Stats** en vous déplaçant avec la flèche . On obtient alors toutes les caractéristiques de la série statistique, et notamment, en se déplaçant vers le bas , l'équation de la droite de régression, ainsi que le coefficient de corrélation linéaire *r*.



CASIO GRAPH 90+E

Appuyez sur la touche MENU de la calculatrice Sélectionner le menu Statistique EXE puis entrer les valeurs de la série statistique dans les colonnes.
List1 et **List2** et appuyer sur F2
Vérifier les réglages SET : **2varXliste : list1 et 2var Yliste : list2.**
Appuyer sur la touche REG puis sur F1 X puis sur F1 de nouveau : on obtient alors l'équation de la droite de régression de y en x, ainsi que le coefficient de corrélation linéaire *r* .



À vous de jouer !

3 On considère le nuage de points associé à la série statistique de l'exercice 1 p. 225

Valeurs de x	5	7	9	12	13	16
Valeurs de y	27	23	18	13	13	6

À l'aide de la calculatrice, déterminer l'équation de la droite d'ajustement du nuage par la méthode des moindres carrés et le coefficient de corrélation linéaire.

4 Dans un pays, on a relevé tous les dix ans le PIB et la consommation d'électricité par habitant.

PIB par habitant (en milliers d'euros)	5	10	16	24	31
Consommation électrique par habitant (en MWh)	1	2,3	4	7	8,9

À l'aide de la calculatrice, déterminer l'équation de la droite d'ajustement du nuage par la méthode des moindres carrés et le coefficient de corrélation linéaire.

➥ Exercices 22 à 25 p. 237

Cours

3 Ajustement exponentiel

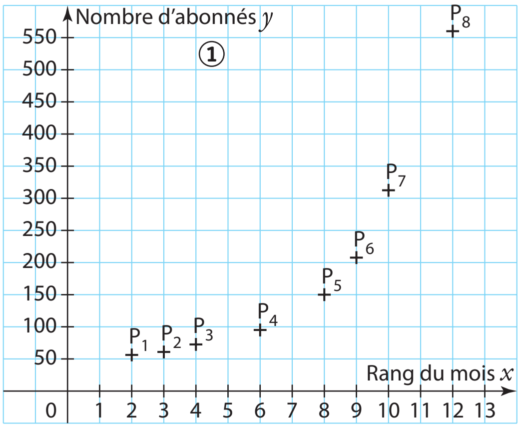
Parfois le nuage de points de la série statistiques (x,y) a une forme générale qui ressemble à la **courbe d'une fonction exponentielle $x \mapsto e^{ax+b}$**
Pour déterminer l'expression de la fonction dont la courbe représentative approche au mieux le nuage de points on effectue un **changement de variable, en posant $y' = \ln(y)$** .
Le nuage de points de la série statistique (x, y') présente alors une **forme rectiligne** que l'on peut ajuster par la **droite de régression de y'** en x, d'équation $y' = ax + b$.
On a alors $\ln(y) = ax + b$ ce qui équivaut à $y = e^{ax+b}$.
Cette équation représente une courbe exponentielle qui ajuste au mieux le nuage.
On dit alors qu'on a réalisé un **ajustement exponentiel**.

Exemple

On a étudié l'évolution presque tous les mois du nombre d'abonnés à la chaîne d'une youtubeuse depuis l'ouverture de celle-ci en septembre 2019.

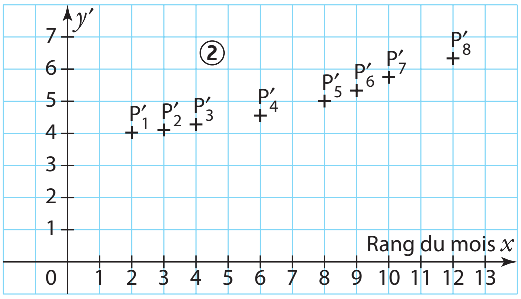
Mois	Nov.	Déc.	Janv.	Mars	Mai	Juin	Juillet	Sept.
Rang du mois x_i	2	3	4	6	8	9	10	12
Nombre d'abonnés y_i	56	61	72	95	150	207	312	560

• Les points $P_i(x_i ; y_i)$ de la série statistique (x, y) ne présentent pas une forme rectiligne.
Un ajustement affine du nuage n'est donc pas pertinent.
En revanche, on observe une forme de nuage caractéristique d'une évolution exponentielle (graphe ①).



• Si on considère maintenant les variables x et $y' = \ln(y)$, on obtient le tableau de valeurs ci-dessous et le nuage suivant (graphe ②).

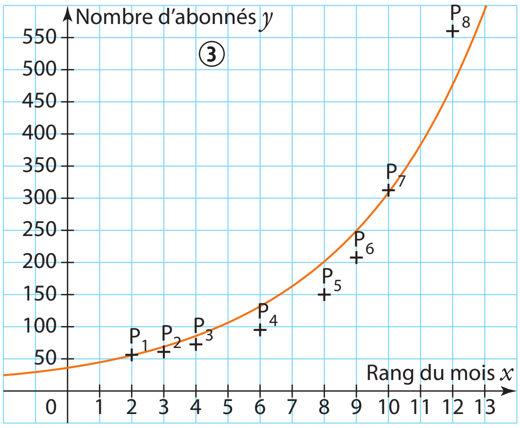
x_i	2	3	4	6	8	9	10	12
$y'_i = \ln(y_i)$	4	4,1	4,3	4,6	5	5,3	5,7	6,3



• Les points semblent quasi alignés : on peut donc effectuer un ajustement affine de la série statistique (x, y').
La méthode des moindres carrés donne :

$$y' = 0{,}23x + 3{,}38$$

On a alors :
$\ln(y) = 0{,}23x + 3{,}38 \Leftrightarrow y = e^{0{,}23x + 3{,}38}$
La courbe d'équation : $y = e^{0{,}23x + 3{,}38}$ ainsi obtenue ajuste au mieux le nuage de points P_i de la série statistique (x, y) initiale (graphe ③).



Remarque Cette méthode d'ajustement ne fonctionne que si la variable y ne prend que des valeurs positives.
Dans les autres cas il faudra réaliser un changement de variable un peu différent.

Méthode 3 Effectuer un ajustement exponentiel

Énoncé

On a relevé, pendant un an, sur différents parcours de même longueur, la vitesse moyenne x_i des véhicules et le nombre d'accidents mortels y_i au total sur l'année, pour i entier variant de 1 à 5.

Vitesse x_i (en km·h^{-1})	30	50	80	90	100
Nombre d'accidents y_i	1	6	41	66	102

1. **Représenter ces données dans un repère orthogonal d'unités bien choisies.**

2. **On pose $y'_i = \ln(y_i)$ pour tout entier i de 1 à 5. Calculer les valeurs de y'_i arrondies au dixième.**

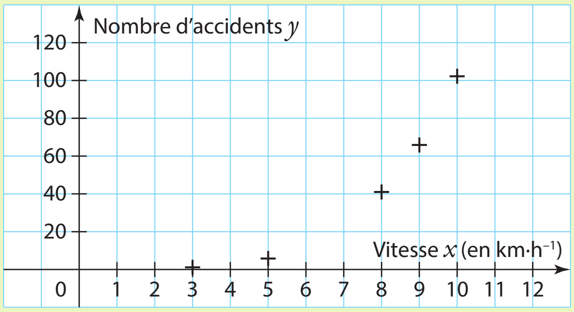
3. **Représenter le nuage de points $(x_i\,;y'_i)$ dans un autre repère et vérifier que sa forme peut être ajustée par une droite.**

4. **Déterminer l'équation de la droite de régression de y' en x. (les coefficients seront arrondis à 10^{-3} près).**

5. **En déduire une expression de y en fonction de x.**

Solution

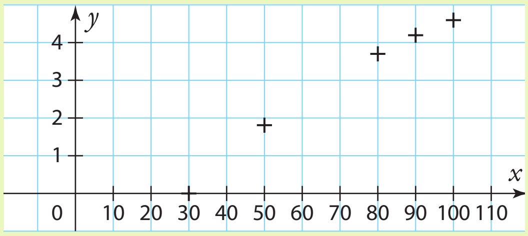
1. On obtient le nuage suivant. 1



2. On obtient le tableau suivant.

x_i	30	50	80	90	100
y'_i	0	1,8	3,7	4,2	4,6

3. On obtient la représentation graphique ci-contre 2
On remarque que les points $(x_i\,;y'_i)$ sont presqu'alignés.



4. Avec la calculatrice, on trouve $y' = 0{,}066x - 1{,}745$ et $r \approx 0{,}993$

5. Comme $y' = \ln(y)$, on en déduit que $\ln(y) = 0{,}066x - 1{,}745$ et donc $y = e^{0{,}066x - 1{,}7425}$ 3

Conseils & Méthodes

1 La forme du nuage obtenu invite à chercher un ajustement exponentiel.

2 Les valeurs y'_i étant beaucoup plus petites que les valeurs y_i, on sera la plupart du temps obligé de changer l'échelle de l'axe des ordonnées et donc d'utiliser un autre repère orthogonal.

3 On pourra éventuellement utiliser les propriétés de la fonction exponentielle pour écrire : $y = e^{0{,}066x} \times e^{-1{,}745}$. Puis, avec $e^{1{,}745} \approx 0{,}175$, on a : $y = 0{,}175e^{0{,}066x}$

À vous de jouer !

5 On considère la série statistique à deux variables suivante.

Valeurs x_i	0	1	2	3
Valeurs y_i	12	19	31	58

1. On pose $y'_i = \ln(y_i)$ pour tout entier i de 0 à 3. Calculer les valeurs y'_i.
2. Représenter le nuage de points $(x_i\,;y'_i)$ dans un repère et vérifier que sa forme peut être ajustée par une droite.
3. Déterminer l'équation de la droite de régression de y' en x. (Les coefficients seront arrondis à 10^{-3} près)
4. En déduire une expression de y en fonction de x.

6 On mesure l'évolution au cours de temps x, en heures, du taux de saturation y de monoxyde de carbone d'un patient intoxiqué qui reçoit un traitement à base d'oxygène.

Temps x_i (en heures)	0	0,5	1	1,5	2	2,5	3
Taux de saturation y_i (en %)	50	38	27	16	8	5	3

1. Représenter le nuage de points $(x_i\,;y_i)$ dans un repère orthogonal d'unités bien choisies.
2. On pose $z_i = \ln(y_i)$ pour tout entier i de 1 à 6. Calculer les valeurs z_i.
3. Déterminer l'équation de la droite de régression de z en x. (Les coefficients seront arrondis à 10^{-3} près).
4. En déduire une expression de y en fonction de x.

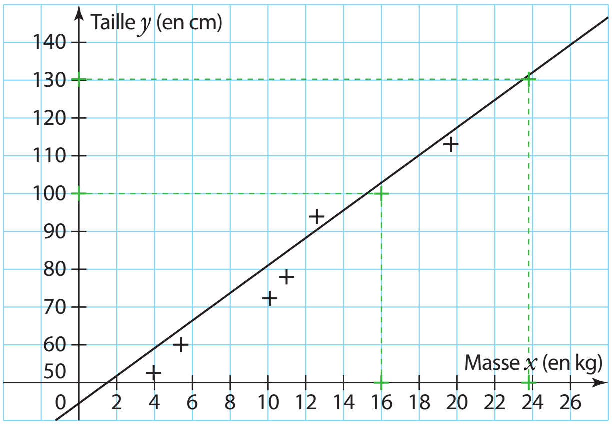
↪ Exercices 26 à 28 p. 237

4 Interpolations et extrapolations

Lorsqu'on a pu réaliser un ajustement (affine ou exponentiel) satisfaisant pour le nuage de points d'une série statistique à deux variables x et y, alors on peut considérer, si cela a du sens dans la situation étudiée, que les quantités x et y ne sont pas complètement indépendantes l'une de l'autre et ont un lien qui peut être modélisé par une fonction f (affine ou exponentielle), qui correspond à l'ajustement trouvé, et telle que $y = f(x)$.
Cela permet d'estimer (graphiquement ou algébriquement) une valeur de la quantité y qui pourrait correspondre à une valeur théorique de la quantité x, non relevée initialement dans l'étude statistique. Et réciproquement on peut aussi estimer une valeur de x qui pourrait correspondre à une valeur donnée de y. Dans le cas où les valeurs choisies restent dans le « domaine d'étude » défini par l'échantillon, on dit que l'on réalise une **interpolation**. Dans le cas contraire, il s'agit alors d'une **extrapolation**.

Exemples

① Le nuage de points ci-contre représente la masse x (en kg) et la taille y (en cm) d'un enfant, relevés par son médecin traitant lors des 6 dernières visites médicales.
La forme du nuage a permis de réaliser un ajustement affine.
La droite d'ajustement obtenue permet de modéliser de façon approchée le lien entre les deux variables.
Ainsi on peut estimer graphiquement que si l'enfant avait été voir son médecin, entre la 5^e et la 6^e visite, au moment où il mesurait environ 100 cm, la pesée aurait probablement donné une valeur autour de 16 kg.

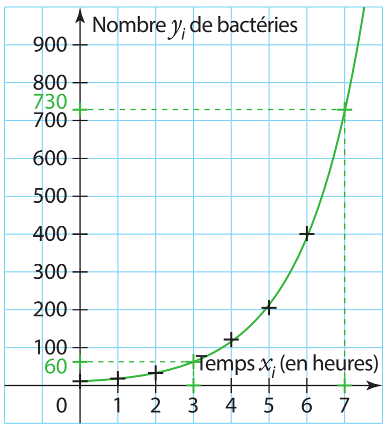


Les deux valeurs $y \approx 100$ et $x \approx 16$ font partie du domaine d'étude qui va de $y_{min} \approx 53$ à $y_{max} \approx 113$ d'une part, et de $x_{min} \approx 4$ à $x_{max} \approx 20$.
Alors l'estimation que nous venons de faire s'appelle une **interpolation**. Nous avons interpolé la valeur 16 kg à partir de la valeur 100 cm (on aurait tout aussi bien pu interpoler la valeur 100 cm à partir de la valeur 16 kg).
Si l'on souhaite faire une estimation en dehors du domaine d'étude (par exemple pour $y \approx 130$ on estime $x \approx 24$) on réalise une **extrapolation**.
Nous pouvons extrapoler (imaginer) la masse d'environ 24 kg de l'enfant lorsqu'il mesurera 130 cm.

② Dans un laboratoire, on étudie l'évolution d'une population de bactéries anaérobies.
Le nombre y_i de bactéries est relevé à différents instants x_i en heures.

Temps x_i (en heures)	0	1	2	4	5	6
Nombres y_i de bactéries	10	18	33	121	205	400

Après avoir représenté le nuage de points $(x_i\,;\,y_i)$ et constaté qu'il présentait une forme « ajustable » par une courbe « exponentielle », on a réalisé un ajustement exponentiel avec la méthode de la page précédente et on a tracé la courbe de la fonction obtenue. Cette courbe permet de modéliser de façon approchée le lien entre les deux variables.
Ainsi, graphiquement, on peut estimer le nombre de bactéries qu'il y avait au bout de 3 heures : soit environ 60 (interpolation).
En supposant que l'évolution observée se prolongera d'heures en heures on peut estimer le nombre de bactéries qu'il y aura au bout de 7 heures : environ 730 (extrapolation).



Exercices résolus

Méthode 4 Effectuer des interpolations et des extrapolations

Énoncé

Pour l'achat d'une grosse quantité de ballons de handball, un fabricant propose un tarif dégressif selon la quantité d'articles commandés.
Le tableau ci-contre présente un relevé des prix proposés.

Nombre de ballons x_i	100	500	1000	2000
Prix unitaire y_i (en €)	19,90	19	17,90	15,50

1. **Déterminer l'équation de la droite de régression de y en x et le coefficient de corrélation r.**
2. **D'après le coefficient r, l'ajustement affine de cette série statistique semble-t-il un bon choix ?**
3. **Déterminer le prix unitaire que devrait proposer le fabricant pour un achat de 1500 ballons.**
4. **Quelle quantité de ballons faudrait-il acheter pour obtenir un prix unitaire de 12 € ?**

Solution

1. et **2.** Avec la calculatrice on obtient : $y = -0{,}0023x + 20{,}16$. Et le coefficient de corrélation $r \approx -0{,}99978$. **1** Ce coefficient de corrélation est très proche de – 1, ce qui signifie que les points sont presque tous alignés et que l'ajustement affine de cette série statistique est donc particulièrement recommandé et très précis.

3. On peut effectuer une interpolation en utilisant l'équation de la droite d'ajustement. On obtient alors $y = -0{,}0023 \times 1\,500 + 20{,}16 = 16{,}71$.
Ainsi pour un achat de 1500 ballons, le prix unitaire que devrait proposer le fabricant pourrait être d'environ 16,70 €. **2**

4. Dans cette question, on sort du domaine d'étude (entre 15,50 € et 19,90 € d'après le tableau), on va donc réaliser une extrapolation, toujours à partir de l'équation de la droite d'ajustement. On a alors :

$$12 = -0{,}0023x + 20{,}16 \Leftrightarrow x = \frac{12 - 20{,}16}{-0{,}0023} \approx 3548\,.$$

Pour obtenir un prix unitaire de 12 €, il faudrait probablement acheter une quantité de ballons proche de 3550. **2**

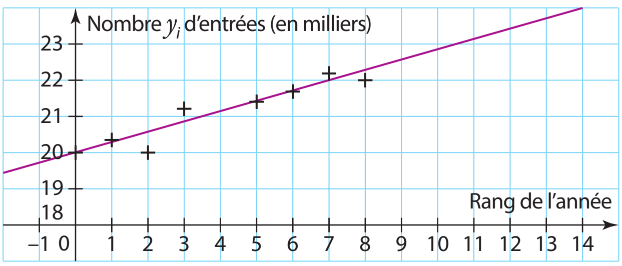
Conseils & Méthodes

1 Plus r est proche de -1, plus sa valeur absolue est proche de 1 : c'est l'indicateur d'une très bonne corrélation linéaire entre les deux variables étudiées.

2 Les résultats d'interpolation et d'extrapolation ne sont que des valeurs indicatives qui découlent de beaucoup d'approximations et d'ajustements successifs, donc ils doivent toujours être donnés avec beaucoup de précautions (utiliser le subjonctif et de termes comme « environ », « probablement »…)

À vous de jouer !

7 Le nuage de points $(x_i\,;y_i)$ suivant donne le nombre total d'entrées annuelle y_i (en milliers) d'un cinéma municipal selon l'année x_i depuis 2010 (année de rang 0).
Ce nuage a été ajusté par la droite d représentée en violet.
1. Il manque les données de 2014. Estimer graphiquement le nombre d'entrées que le cinéma a totalisé cette année-là.
2. Si la tendance observée se poursuit encore pendant plusieurs années, en quelle année le cinéma peut-il espérer atteindre les 24 000 spectateurs par an ?



8 Dans une école primaire, chaque année depuis 2014 (année de rang 0), on a relevé le nombre de livres empruntés à la bibliothèque de l'école.

Rang de l'année x_i	0	1	2	3	4	5
Nombre de livres y_i	201	187	191	162	163	134

1. Déterminer l'équation de la droite de régression de y en x et le coefficient de corrélation r.
2. D'après le coefficient r, l'ajustement affine de cette série statistique semble-t-il un bon choix ?
3. Peut-on estimer le nombre de livres qui seront empruntés à la bibliothèque en 2021 ?
4. Si la tendance se poursuit selon cet ajustement affine pendant les prochaines années, en quelle année pourrait-on avoir moins de 65 livres empruntés ?

↪ Exercices 30 à 31 p. 238

Exercices résolus

EXOS Méthodes lienmini.fr/maths-c09-04 Les rendez-vous Sésamath

Méthode 5 Calculer par la méthode des moindres carrés

Cours 2 p. 226

Énoncé

Le tableau ci-contre représente la masse x (en kg) et la taille y (en cm) d'un enfant, relevés par son médecin traitant lors des 6 dernières visites médicales.

Masse x_i (en kg)	4	5,4	10,2	11	12,6	19,8
Taille y_i (en cm)	53	61	72	78	94	113

1. **Calculer la covariance de x et y puis la variance de x.**
2. **En déduire l'équation de la droite de régression de y en x.**
3. **Représenter le nuage de points $(x_i\,;y_i)$ dans un repère puis tracer la droite de régression.**

Solution

1. On commence par calculer les valeurs des moyennes $\overline{x}$ et $\overline{y}$:

$$\overline{x} = \frac{4+5{,}4+10{,}2+11+12{,}6+19{,}8}{6} = \frac{63}{6} = 10{,}5 \quad \text{et} \quad \overline{y} = \frac{53+61+72+78+94+113}{6} = \frac{471}{6} = 78{,}5$$

Puis on peut s'aider d'un tableau pour détailler les étapes du calcul.

Masse x_i (en kg)	4	5,4	10,2	11	12,6	19,8
Taille y_i (en cm)	53	61	72	78	94	113
Écarts $x_i - \overline{x}$	−6,5 **1**	−5,1	−0,3	0,5	2,1	9,3
Écarts $y_i - \overline{y}$	−25,5	−17,5	−6,5	−0,5	15,5	34,5
Produit des écarts $(x_i - \overline{x})(y_i - \overline{y})$	165,75	89,25	1,95	−0,25	32,55	320.85

La covariance est la moyenne des produits des écarts :

$$\text{cov}(x,y) = \frac{165{,}75+89{,}25+1{,}95-0{,}25+32{,}55+320{,}85}{6} = \frac{610{,}1}{6} \approx 101{,}68$$

La variance de x est la moyenne des carrés des écarts $(x_i - \overline{x})^2$:

$$\text{var}(x) = \frac{(-6{,}5)^2+(-5{,}1)^2+(-0{,}3)^2+0{,}5^2+2{,}1^2+9{,}3^2}{6} = \frac{159{,}5}{6} \approx 26{,}58$$

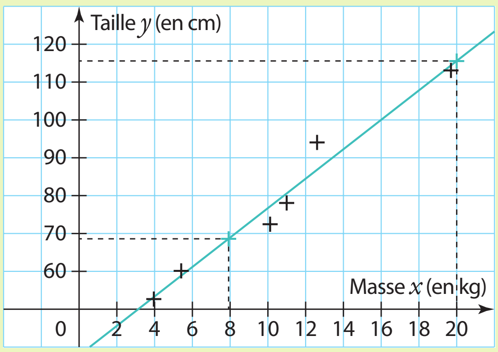
2. On en déduit que le coefficient directeur de la droite de régression de y en x est : $a = \dfrac{\text{cov}(x,y)}{\text{var}(x)} \approx \dfrac{101{,}68}{26{,}58} \approx 3{,}83$

et l'ordonnée à l'origine $b \approx 78{,}5 - 3{,}83 \times 10{,}5 \approx 38{,}29$.

Donc l'équation de la droite de régression de y en x est :
$y = 3{,}89x + 38{,}29$.

3. On peut tracer la droite à l'aide d'un tableau comme ci-contre **2**.

x	8	20
y	69,5	116



Conseils & Méthodes

1 $x_1 = 4$ et $\overline{x} = 10{,}5$ donc l'écart est égal à $4 - 10{,}5 = -6{,}5$.

2 Pour tracer la droite on fait un tableau de valeurs en choisissant deux valeurs de x relativement éloignées pour améliorer la précision du tracé et en calculant le y correspondant à l'aide de l'équation de la droite.

À vous de jouer !

9 On considère la série statistique à deux variables x et y suivante.

Valeurs de x	5	7	9	12	16
Valeurs de y	27	23	18	13	6

1. Calculer la covariance de x et y puis la variance de x.
2. En déduire l'équation de la droite de régression de y en x.
3. Représenter le nuage de points $(x_i\,,y_i)$ dans un repère puis tracer la droite de régression.

10 On a relevé à certaines latitudes x, la température y à la surface de la mer.

Latitudes x_i (en °)	−45	−20	0	30
Températures y_i (en °C)	12,2	12,9	13,4	14,1

1. Calculer la covariance de x et y puis la variance de x.
2. En déduire l'équation de la droite de régression de y en x.
3. Représenter le nuage de points $(x_i\,;y_i)$ dans un repère puis tracer la droite de régression.

Exercices 38 à 40 p. 240

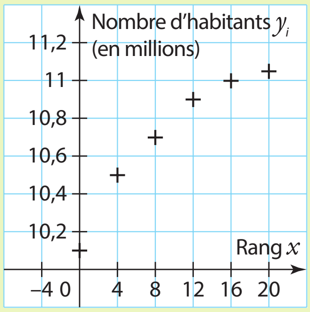
Exercices résolus

Méthode 6 D'autres ajustements

Énoncé

Le tableau ci-contre donne le nombre d'habitants y_i (avec i entier de 1 à 6) en Grèce, en millions d'habitants, tous les 4 ans depuis 1990 (année de rang x_1=0).

Années	1990	1994	1998	2002	2006	2010
Rang x_i	0	4	8	12	16	20
Nombre d'habitants y_i (en millions)	10,1	10,5	10,7	10,9	11	11,05



1. **Au vu du nuage de points $(x_i\,;y_i)$ ci-contre, un ajustement affine semble-t-il pertinent ?**

2. **On pose, pour tout entier i compris entre 1 et 6, $y'_i = e^{y_i}$. Calculer les valeurs y'_i.**

3. **Représenter le nuage de points $(x_i\,;y'_i)$. Vérifier qu'il peut être ajusté par une droite.**

4. **Déterminer la droite de régression de y' en x.**
En déduire une expression de y en fonction de x.

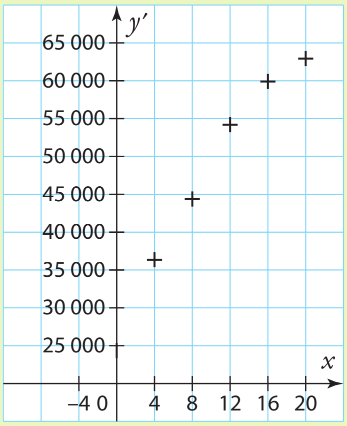
Solution

1. Les points $(x_i\,;y_i)$ ne semblent pas suivre une tendance « rectiligne », mais plutôt légèrement incurvée, l'ajustement affine ne semble donc pas le mieux adapté à la situation.

2. On obtient le tableau ci-contre **1**

x_i	0	4	8	12	16	20
$y'_i = e^{y_i}$	24 343	36 316	44356	54 176	59 874	62 944

3. On obtient le nuage de points ci-contre. Les points $(x_i\,;y'_i)$ sont moins incurvés et semblent plus alignés :



4. À la calculatrice, on obtient :

$y' = 1\,954x + 27\,466$.
Or $y' = e^y$, donc $y = \ln(y')$, c'est-à-dire
$y = \ln(1\,954x + 27466)$.

Conseils & Méthodes

1 Lorsque le nuage présente une forme incurvée non rectiligne il est très difficile de savoir s'il existe une courbe de corrélation entre les deux variables et surtout quelle courbe pourrait correspondre car plusieurs courbes de fonctions « usuelles » pourraient convenir. Le changement de variable proposé ici permet un assez bon ajustement du nuage avec une courbe « logarithme » mais qui n'est peut-être pas le meilleur.

À vous de jouer !

11 On considère la série statistique $(x\,;y)$ suivante.

x_i	30	45	65	100	120	165	200	250
y_i	1,2	2	2,6	3,1	3,5	3,7	3,8	3,85

1. Représenter sur la calculatrice le nuage de points $(x_i\,;y_i)$.
2. On pose, pour tout entier i compris entre 1 et 8, $y'_i = e^{y_i}$; calculer les valeurs y'_i.
3. Représenter le nuage de points $(x_i\,;y'_i)$.
Vérifier qu'il peut être ajusté par une droite.
4. Déterminer la droite de régression de y' en x.
En déduire une expression de y en fonction de x.

12 TICE La capacité éolienne raccordée en France métropolitaine depuis 2001 (année de rang 0) forme une série statistique à deux variables.

Rang de l'année x_i	1	2	3	4	5	6	7
Puissance y_i (en MW)	94	219	751	2 252	4 573	6 714	8 157

1. Représenter le nuage de points de la série statistique $(x;y)$ sur la calculatrice ou sur un tableur.
2. On pose, pour tout entier i compris entre 0 et 7, $y'_i = \sqrt{y_i}$. Calculer les valeurs y'_i.
3. Déterminer le coefficient de corrélation linéaire de la série $(x\,;y')$ et en déduire que le nuage de points $(x_i\,;y'_i)$ peut être approchée de façon pertinente par une droite.
4. Déterminer la droite de régression de y' en x.
En déduire une expression de y en fonction de x.

↪ Exercices 45 à 47 p. 241

Exercices apprendre à démontrer

VIDÉO
Démonstration
lienmini.fr/maths-c09-05

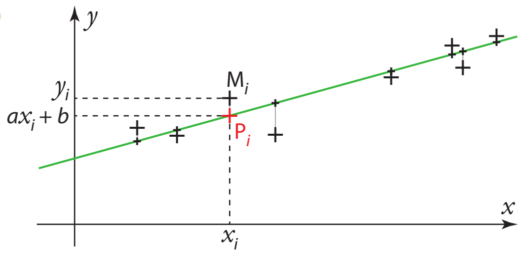
La propriété à démontrer

La droite des moindres carrés associée au nuage de points $M_i(x_i\,;y_i)_{1\leqslant i\leqslant n}$ passe par le point moyen G du nuage.

On utilisera la définition de la droite des moindres carrés.

Comprendre avant de rédiger

- La droite des moindres carrés d'équation $y = ax + b$ passe par le point moyen $G(\overline{x}\,;\overline{y})$ si et seulement si on a $\overline{y} = a\overline{x} + b$.
- L'équation de la droite des moindres carrés est caractérisée par les réels a et b pour lesquels la somme $S = M_1P_1^2 + M_2P_2^2 + \ldots + M_nP_n^2$ est minimale.
- Montrons donc que les réels a et b pour lesquels S est minimale vérifient l'égalité $b = \overline{y} - a\overline{x}$.



Rédiger

Étape 1
Exprimer en fonction de a et b la somme $S = M_1P_1^2 + \ldots + M_nP_n^2$.

On utilise la formule
$MP = \sqrt{(x_P - x_M)^2 + (y_P - y_M)^2}$

Étape 2
On étudie la somme S en fonction des valeurs de b.

Étape 3
On a, d'une part, $f_i'(b) = 1$ et d'autre part, pour toute fonction f_i, la dérivée de f_i^2 est égale à $2f_i'f_i$.
On peut donc dériver S.

Étape 4
On détermine la valeur de b pour laquelle S est minimale.

Pour cela il faut étudier le signe de la dérivée S' pour tout réel b.

Étape 5
On conclut à l'aide de la valeur de b sur l'appartenance du point moyen G à la droite des moindres carrés.

La démonstration rédigée

Comme, pour tout entier i de 1 à n, M_i et P_i ont la même abscisse x_i, alors on a : $M_iP_i = \sqrt{(ax_i + b - y_i)^2}$ et donc $M_iP_i^2 = (ax_i + b - y_i)^2$.
Donc $S = (ax_1 + b - y_1)^2 + (ax_2 + b - y_2)^2 + \ldots + (ax_n + b - y_n)^2$.

S peut alors être considérée comme une fonction de variable b et se présente comme une somme de carrés de fonctions affines $f_i : b \mapsto (ax_i + b - y_i)$ dérivables pour tout réel b.

Ainsi, S est dérivable pour tout réel b, et on a :
$S'(b) = 2(ax_1 + b - y_1) + 2(ax_2 + b - y_2) + \ldots + 2(ax_n + b - y_n)$
$S'(b) = 2(ax_1 + b - y_1 + ax_2 + b - y_2 + \ldots + ax_n + b - y_n)$
$S'(b) = 2(ax_1 + ax_2 + \ldots + ax_n + n \times b - y_1 - y_2 - \ldots - y_n)$
$S'(b) = 2nb + 2(ax_1 + ax_2 + \ldots + ax_n - y_1 - y_2 - \ldots - y_n)$

On résout d'abord $S'(b) = 0$:
$2nb + 2(ax_1 + ax_2 + \ldots + ax_n - y_1 - y_2 - \ldots - y_n) = 0$
$\Leftrightarrow b = \dfrac{y_1 + y_2 + \ldots + y_n}{n} - \dfrac{a(x_1 + x_2 + \ldots + x_n)}{n} \Leftrightarrow b = \overline{y} - a\overline{x}$.
Avec le même raisonnement, on peut montrer que :
$S'(b) > 0 \Leftrightarrow b > \overline{y} - a\overline{x}$ et aussi que $S'(b) < 0 \Leftrightarrow b < \overline{y} - a\overline{x}$.
Cela signifie que la fonction S admet un minimum lorsque $b = \overline{y} - a\overline{x}$.

Cette valeur de b est l'ordonnée à l'origine de la droite des moindres carrés. Or $b = \overline{y} - a\overline{x} \Leftrightarrow \overline{y} = a\overline{x} + b$.
Les coordonnées du point moyen G vérifient l'équation de la droite d des moindres carrés associée au nuage de points $M_i(x_i\,;y_i)_{1\leqslant i\leqslant n}$.

Pour s'entraîner

Une valeur du tableau ci-contre a été effacée, mais on sait que le nuage de points $(x_i\,;y_i)$ a pu être ajusté par la droite des moindres carrés d'équation $y = -1{,}1x + 3{,}35$.
Déterminer la valeur manquante.

x_i	4	5		8
y_i	-1	-2	-5	-5