2 Séries statistiques à deux variables

On étudie simultanément deux caractères d'une population.
La série statistique à deux variables est donnée par des couples de valeur $(x_i ; y_i)$.
On la représente graphiquement par un **nuage de points**.

1. Point moyen

On appelle **point moyen** d'un nuage de n points $M_i\ (x_i ; y_i)$ le point G de coordonnées :

$$x_G = \overline{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n)\ ;\ y_G = \overline{y} = \frac{1}{n}(y_1 + y_2 + \dots + y_n).$$

2. Ajustement affine

Lorsque le nuage a une forme allongée, il est possible de tracer une droite D au voisinage de ses points.
On dit alors que l'on a un **ajustement affine**.

3. Ajustement affine par la méthode des moindres carrés

Par la méthode des moindres carrés, on détermine deux droites, appelées **droites de régression**.
La droite de régression de y en x a pour équation $y = ax + b$.
La droite de régression de x en y a pour équation $x = a'y + b'$.
Ces deux droites passent par le **point moyen *G*** du nuage de points.
Les équations de ces droites sont obtenues avec une calculatrice ou avec un logiciel (voir *Fiche méthode 27*).
Le **coefficient de corrélation *r*** donné par la calculatrice permet de comparer la qualité de deux ajustements. Plus ce coefficient est **proche de 1 ou – 1**, meilleur est l'ajustement.
Les droites de régression permettent d'obtenir **des estimations** :
– pour obtenir une estimation de ***y*** pour une valeur de ***x*** donnée, on utilise l'équation de la droite de régression **de *y* en *x***.
– pour obtenir une estimation **de *x*** pour une valeur de ***y*** donnée, on utilise l'équation de la droite de régression ***x* en *y***.

Comment représenter une série statistique à deux variables et déterminer le point moyen du nuage avec un logiciel ou une calculatrice graphique ?

On utilise ici le logiciel Sine Qua Non ou une calculatrice (Casio ou TI).

Exemple.
Une entreprise a relevé le coût total de production mensuel (en k€), noté y, en fonction de la production x (en tonnes).

Production x (en tonnes)	1	2	4	6	8	10
Coût total y (en k€)	36,3	38,5	44,6	48,4	51,1	54,2

1. Représenter le nuage de points $M_i(x_i\ ;\ y_i)$ dans un repère orthogonal.
2. Déterminer les coordonnées du point moyen G de la série.

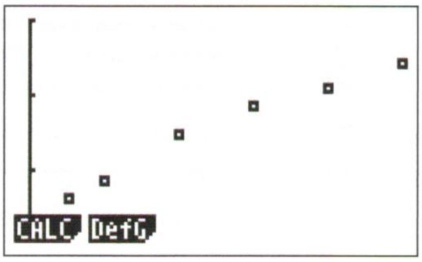
Avec une calculatrice Casio Graph 35+

• *Dans* MENU STATS EXE, *on sélectionne* ▶ *en tapant* F6, DEL-A *avec* F4, YES *avec* EXE.

On entre les valeurs x_i *dans List 1, les valeurs* y_i *dans List 2.*

• *On sélectionne* ▶ *par* F6 *puis* GRAPH *par* F1, *puis on choisit* GPH1 *avec* F1.

On obtient le nuage de points.



• *Pour obtenir les coordonnées de G taper* EXIT *2 fois.*

Sélectionner CALC *avec* F2 *puis* Set *avec* F6.

Sur 2VarX-List, *on sélectionne List1 ;*

sur 2VarY-List, *on sélectionne List2 ;* EXE.

F2 *pour* 2VAR.

On lit $\overline{x} = 5{,}166\ldots$, *puis* $\overline{y} = 45{,}516\ldots$.

Avec une calculatrice TI 82 stats.fr ou 83 Plus

• *On tape* Stats *puis on sélectionne* 4:Effliste .

On tape 2nde 1 , 2nde 2 *(pour L1, L2) puis* entrer

(on a effacé les listes précédentes.)

• *On tape* Stats *puis on sélectionne* 1:Edit entrer .

On entre les valeurs x_i *dans L1, les valeurs* y_i *dans L2.*

• *On tape* 2nde f(x) *pour graph stats puis* 1 .

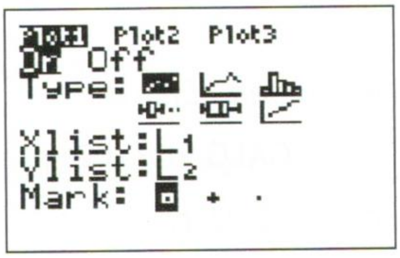
• *On sélectionne sur l'écran ci-dessous :*

– On ,

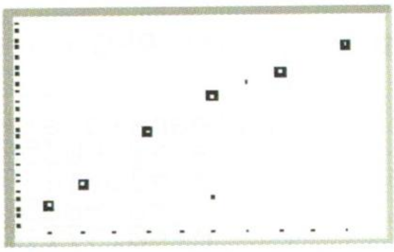
– *le type de représentation,*

– *L1 pour ListeX et L2 pour ListeY,*

– *le type de marque.*



• *On tape* ZOOM 9 *et on obtient le nuage de points.*



• *Pour obtenir les coordonnées de G :*

on tape Stats *, on sélectionne* CALC *puis* 2: Stats2Var

et on tape 2nde 1 , 2nde 2 entrer .

On lit $\overline{x} = 5{,}166\ldots$, $\overline{y} = 45{,}516\ldots$

Comment réaliser un ajustement affine par la méthode des moindres carrés avec une calculatrice ou un logiciel ?

On utilise ici le logiciel Sine Qua Non ou une calculatrice (Casio ou TI).

Exemple.

Une entreprise a relevé le coût total de production mensuel (en k€), noté y, en fonction de la production x (en tonnes).

Production x (en tonnes)	1	2	4	6	8	10
Coût total y (en k€)	36,3	38,5	44,6	48,4	51,1	54,2

1. Déterminer le coefficient de corrélation de la série.
2. Donner une équation de la droite D, droite de régression de y en x obtenue par la méthode des moindres carrés.
3. Estimer à l'aide de cette droite le coût total correspondant à une production de 12 tonnes.

Avec une calculatrice Casio Graph 35+

- *On tape* MENU STAT EXE *,*
- *On efface les listes précédentes en sélectionnant* ▶ *par* F6 *,* DEL-A *par* F4 *et* YES *par* F1 *.*
- *On entre les x_i dans List 1, les y_i dans List 2.*
- *On sélectionne* ▶ *par* F6 *puis* CALC *par* F2 *et* Set *par* F6 *.*

Sur la ligne 2VarXList *, on sélectionne List1 ;*

sur la ligne 2VarYList *, on sélectionne List2 ;*

sur la ligne 2VarFreq, on sélectionne 1 puis EXE *.*

- *On obtient les résultats en sélectionnant* REG *par* F3 *puis* X *par* F1 *et* ax+b *par* F1 *:*

```
LinearReg(ax+b)
    a =2.00356164
    b =35.1649315
    r =0.98851473
    r²=0.97716137
   MSe=1.42689041
y=ax+b
```

Remarque :

Pour obtenir une équation de la droite de régression de x en y,

sur la ligne 2VarXList *, on remplace List1 par List2 ;*

et sur la ligne 2VarYList*, on remplace List2 par List1 ;*

et on obtient a′ et b′ de l'équation $x = a'y + b'$.

▶ ▶ ▶

Avec une calculatrice TI 82 stats.fr ou 83 Plus

• *On tape* [stats] *et on sélectionne* 4 : Effliste [2nde] [1] [,] [2nde] [2] [entrer] *pour supprimer les listes précédentes.*

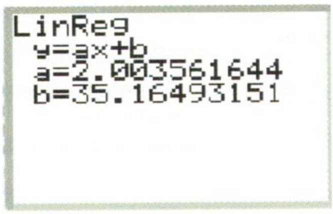
On tape [stats] *, on sélectionne* 1 : Edite [entrer] .

• *On entre les* x_i *dans L1, les* y_i *dans L2.*

On tape [stats] *, on sélectionne* CALC . *On sélectionne* 4 : Reglin (ax + b) .

• *On entre L1, L2 avec* [2nde] [1] [,] [2nde] [2] [entrer] .

Cela donne le résultat :



si r^2 *et r ne sont pas affichés, on tape* [2nde] [0] *et on sélectionne* CorrelAff *dans la liste.*

Remarque :

Pour obtenir une équation de la droite de régression de x en y,

Sélectionner 4 : Reglin (ax + b) *, taper* [2nde] [2] [,] [2nde] [1] *pour L2, L1.*

On obtient les coefficients a′ et b′ de l'équation $x = a'y + b'$.

De cette manière, on trouve les résultats suivants :

1. *Le coefficient de corrélation est* $r = 0{,}988515$.
r est proche de 1, l'ajustement affine est justifié.

2. *La droite D de régression de y en x a pour équation :*

$$y = 2\,x + 35{,}16.$$

3. *Pour une production de 12 tonnes, on peut estimer que :*

$$y = 2 \times 12 + 35{,}16 \quad \text{soit} \quad y = 59{,}16.$$

Exercices :

18 **C** Dans cet exercice, les calculs seront effectués à 10^{-3} près.
L'étude du coût de maintenance annuel d'une installation de chauffage dans un immeuble de bureaux, en fonction de l'âge de l'installation, a donné les résultats suivants.

Âge x_i (en années)	1	2	3	4	5	6
Coût y_i (en k€)	7,55	9,24	10,74	12,84	15,66	18,45

1. Représenter le nuage de points $M_i(x_i\,;y_i)$ dans un repère orthogonal (unités graphiques : 2 cm en abscisse, 1 cm en ordonnée).
Peut-on envisager un ajustement affine de ce nuage ?

2. a) Déterminer le coefficient de corrélation linéaire de la série statistique double $(x_i\,;y_i)$.
Le résultat obtenu confirme-t-il l'observation faite à la question **1.** ?

b) Déterminer, par la méthode des moindres carrés, une équation de la droite de régression D de y en x.

Tracer D dans le même repère que celui de la question **1.**

c) En admettant que l'évolution du coût constaté pendant six ans se poursuive les années suivantes, donner une estimation du coût de maintenance de l'installation lorsqu'elle aura huit ans.

19 Le nombre d'internautes en France est donné (en millions) dans le tableau suivant :

Année	2001	2003	2005	2007	2009	2011
x : rang de l'année	1	3	5	7	9	11
y : nombre d'internautes (en millions)	12,86	20,67	25,07	29,55	33,64	39,36

1. Donner le coefficient de corrélation linéaire entre les séries x et y. Arrondir le résultat au centième.

2. On envisage un ajustement affine. Donner une équation de la droite de régression de y en x obtenue par la méthode des moindres carrés. Arrondir les coefficients au centième.

3. En utilisant l'équation précédente, estimer le nombre d'internautes en 2015, en arrondissant le résultat au demi-million.

20 En 2013, une caisse de retraite propose à ses adhérents un barème de rachat d'un trimestre de cotisation des années antérieures selon le tableau suivant.

Âge de l'adhérent (en années)	56	57	58	59	60
Rang x_i	0	1	2	3	4
Montant y_i du rachat d'un trimestre de cotisation (en euros)	3 906	3 994	4 081	4 167	4 251

1. Donner une équation de la droite de régression D de y en x, obtenue par la méthode des moindres carrés.

2. Quel serait avec cet ajustement affine le montant du rachat d'un trimestre pour un salarié âgé de 62 ans ?

21 **C** Une machine fabrique en grande série des billes d'acier.
La moyenne des diamètres des billes produites en une semaine varie au cours du temps. La fabrication est jugée valable tant que cette moyenne reste dans l'intervalle [3,25 ; 3,32].
La semaine numérotée 0 correspond à celle du réglage initial. Des contrôles hebdomadaires effectués lors des quatre premières semaines de fonctionnement ont donné les résultats suivants.

Semaine s	0	1	2	3	4
Moyenne m	3,32	3,32	3,31	3,29	3,27

1. a) Calculer le coefficient de corrélation de la série statistique.

b) Déterminer une équation de la droite d'ajustement de s en m par la méthode des moindres carrés.

2. En déduire un pronostic pour la valeur maximale du temps séparant deux réglages successifs.

22 Afin de mesurer l'évolution de l'utilisation du vélo, une communauté urbaine organise le comptage régulier des vélos en plusieurs points de l'agglomération. Le tableau ci-dessous indique le nombre moyen, sur un mois, de vélos comptés par jour.

Mois	Mars 2005	Juin 2005	Déc. 2005	Juin 2006	Déc. 2006	Juin 2007
Rang x_i du mois	0	3	9	15	21	27
Nombre moyen y_i de vélos comptés par jour (en milliers)	3,9	4,4	5,1	6,4	7,1	7,6

1. Représenter le nuage de points $M_i\ (x_i\ ;\ y_i)$ dans un repère orthogonal.
On prendra pour unités graphiques : en abscisse, 1 centimètre pour représenter 3 mois et en ordonnées, 1 centimètre pour représenter 1 millier.

2. Déterminer les coordonnées du point moyen G et le placer sur la représentation graphique.

3. Déterminer, en utilisant la calculatrice, l'équation de la droite d'ajustement affine de y en x obtenue par la méthode des moindres carrés. On arrondira les coefficients obtenus à 10^{-2} près.
Tracer la droite d'ajustement sur la représentation graphique.

4. À l'aide de l'ajustement réalisé, déterminer une estimation du nombre moyen de vélos que l'on devait prévoir par jour au mois de décembre 2007 (on arrondira le résultat à 10^{-1}).

5. On sait qu'en décembre 2007, le nombre moyen de vélos observés a été en fait de 7600.
Déterminer, en pourcentage, l'erreur commise dans l'estimation précédente.

23 **C** Sur un parcours donné, la consommation y d'une voiture est donnée en fonction de sa vitesse moyenne x par le tableau suivant :

x (en km/h)	80	90	100	110	120
y (en L/100 km)	4	4,8	6,3	8	10

1. La consommation est-elle proportionnelle à la vitesse moyenne ? Justifier la réponse.

2. a) Représenter le nuage de points correspondant à la série statistique $(x_i\ ;\ y_i)$ dans un repère orthogonal du plan (on prendra 2 cm pour 10 km/h sur l'axe des abscisses et 1 cm pour 1 L sur l'axe des ordonnées).

b) Déterminer les coordonnées du point moyen G du nuage et le placer sur le graphique.

c) À l'aide d'une calculatrice, donner une équation, sous la forme $y = ax + b$, de la droite d'ajustement affine de y en x par la méthode des moindres carrés et tracer cette droite (on arrondira a au millième et b au centième).

d) En utilisant cet ajustement, estimer la consommation aux 100 km (arrondie au dixième) de la voiture pour une vitesse de 130 km/h.

3. La forme du nuage permet d'envisager un ajustement exponentiel.
On pose $z = \ln(y)$ et on admet que la droite d'ajustement obtenue pour les cinq points $(x\ ;\ z)$ du nuage par la méthode des moindres carrés, a pour équation :

$$z = 0{,}023\,4x - 0{,}508\,0.$$

a) Écrire y sous la forme $y = A\,\mathrm{e}^{Bx}$ (donner A et B arrondis à 10^{-4}).

b) Tracer, sur le même graphique, la courbe d'équation $y = A\,\mathrm{e}^{Bx}$ pour x élément de l'intervalle $[80\ ;\ 120]$.

c) En utilisant cet ajustement, estimer la consommation aux 100 km (arrondie au dixième) de la voiture, pour une vitesse de 130 km/h.

4. Des deux valeurs obtenues dans les questions **2. d)** et **3. c)**, pour la consommation à une vitesse de 130 km/h, laquelle vous semble la plus proche de la consommation réelle ?
Expliquer votre choix.

24 On a relevé mois après mois, le coût d'amortissement d'une pompe hydraulique.

Mois x	1	2	3	4	5	6
Coût C (en €)	400	300	270	220	180	150

1. L'allure du nuage des points de la série $(x\,;C)$ conduit à poser $y = \ln C$.

a) Dresser le tableau de la série statistique $(x\,;y)$ en prenant des valeurs décimales arrondies à 10^{-3} près.

b) Calculer le coefficient de corrélation linéaire de cette série. (On en donnera une valeur décimale arrondie à 10^{-3} près.)

c) Justifier la pertinence d'un ajustement affine.

2. Déterminer une équation de la forme :

$$y = ax + b,$$

où a et b désignent des nombres réels, de la droite de régression de y en x.
(On prendra pour valeurs de a et b leurs valeurs décimales arrondies à 10^{-3} près.)

3. À partir du résultat de la question **2.**, déterminer l'expression de C, en fonction de x sous la forme $C = \alpha\,\beta^x$, où α et β sont des réels dont on donnera des valeurs approchées à 10^{-2} près.

4. Utiliser l'expression précédente pour évaluer le coût d'amortissement de la pompe au mois numéro 7, à 10^{-2} (€) près.

Correction :

18 **1.** Oui.

2. a) On obtient $\boldsymbol{r = 0{,}991}$ ce qui confirme l'observation de la question **1.**
b) La calculatrice donne : $a = 2{,}167$ et $b = 4{,}827$. Une équation de D est : $\boldsymbol{y = 2{,}167\,x + 4{,}827}$.
c) Pour 8 ans on aurait un coût de : **22,163 k€.**

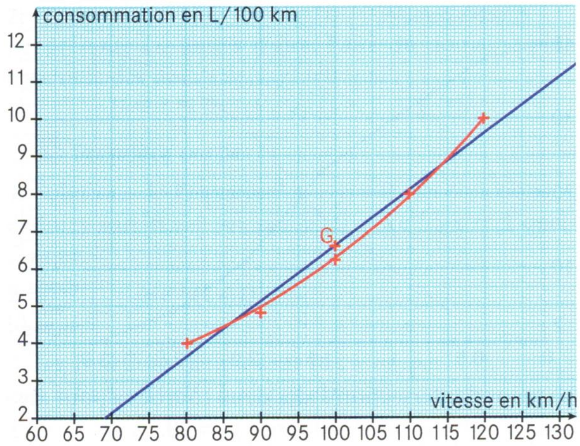
21 **1. a)** La calculatrice donne $\boldsymbol{r = -0{,}948}$;
$a = -69{,}15$; $b = 230{,}33$.
b) La droite d'ajustement a pour équation :

$$\boldsymbol{s = -69{,}15\,m + 230{,}33}.$$

2. Pour $m = 3{,}25$:
$s = -69{,}15 \times 3{,}25 + 230{,}33 \approx 5{,}59$.
Le temps maximal entre deux réglages est **5 semaines.**

22 **1.** $\dfrac{80}{4} \neq \dfrac{90}{4{,}8}$ donc la consommation **n'est pas** proportionnelle à la vitesse moyenne.

2. a)



b) Les coordonnées de G sont : $x_G = \mathbf{100}$; $y_G = \mathbf{6{,}62}$.
c) La calculatrice donne : $a = 0{,}152$ et $b = -8{,}58$.
Une équation de la droite d'ajustement est :
$\boldsymbol{y = 0{,}152\,x - 8{,}58}$.
d) Pour 130 km/h la consommation serait :
$C = 0{,}152 \times 130 - 8{,}58$ soit $\boldsymbol{C = 11{,}2}$ **litres.**

3. a) $z = 0{,}0234\,x - 0{,}508$ et $z = \ln y$
d'où $y = \mathrm{e}^{0{,}0234x - 0{,}508}$ soit $y = \mathrm{e}^{-0{,}508} \times \mathrm{e}^{0{,}0234x}$
donc : $\boldsymbol{y = 0{,}6017\ \mathrm{e}^{0{,}0234x}}$.
c) La consommation serait $C' = 0{,}6017 \times \mathrm{e}^{0{,}0234 \times 130}$ soit $\boldsymbol{C' = 12{,}6}$ L.

4. La valeur la plus proche de la consommation réelle est **12,6 L.**