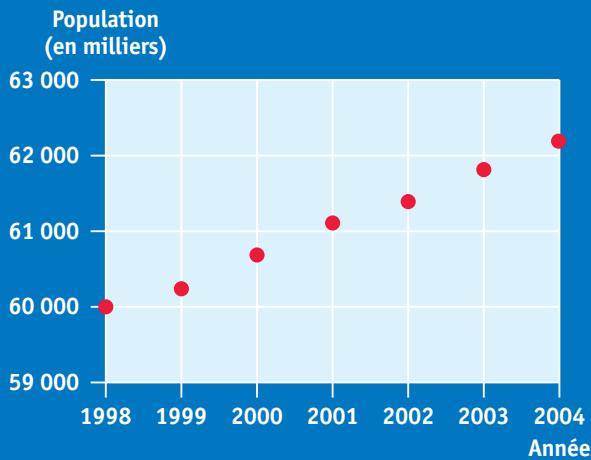


# Statistiques



**Évolution de la population en France**



Le graphique ci-contre indique l'évolution de la population française de 1998 à 2004. On constate que les points obtenus sur ce graphique sont presque alignés. On peut alors penser que ce type d'évolution va continuer pour les années futures. On trace une droite qui reflète approximativement la progression de la population française de 1998 à 2004. Si ce modèle reste valable pour la suite, on pourra alors estimer la population française métropolitaine en 2005, en 2006..., tant que ce modèle restera valable.

Nous allons étudier, dans ce chapitre, ce procédé fort utile dans le domaine de la prévision.

## Les notions abordées dans ce chapitre

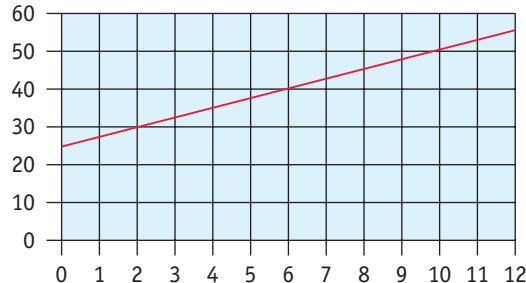
1. Séries statistiques simples : page 82
2. Séries statistiques doubles : page 86

# Testez-vous avant de commencer !

Pour chacune des questions suivantes, choisir la (ou les) bonne(s) réponse(s).

## A. Savez-vous... lire une représentation graphique ?

Le graphique ci-contre représente l'évolution du nombre de nuitées réservées dans les gîtes ruraux d'un département (en milliers) en fonction du rang de l'année (l'année de rang 1 étant 1991).



	A	B	C	D
1. Le nombre de nuitées pour 1994 est, en milliers	25	30	35	40
2. Il y a eu plus de 42 milliers de nuitées à partir de	1996	1997	1998	1999
3. Le nombre de nuitées a doublé par rapport à l'année 1990 à partir de	1994	1998	2000	2002

## B. Savez-vous... calculer des paramètres d'une série statistique ?

Soit la série statistique formée des valeurs suivantes : 1, 2, 4, 6, 9, 10, 17.

	A	B	C	D
4. La moyenne de la série est	6	6,5	7	9
5. L'écart type de la série est, à 0,1 près	5,1	5,5	10	16
6. La médiane de la série est	6	6,5	7	9

## C. Savez-vous... déterminer une équation de droite ?

Dans le plan rapporté au repère orthonormal  $(O ; \vec{i}, \vec{j})$ , soit les points  $A(2 ; 1)$ ,  $B(4 ; 4)$  et  $C(3 ; 2)$ .

	A	B	C	D
7. Le coefficient directeur de la droite $(AB)$ est	$\frac{2}{3}$	2	$\frac{3}{2}$	3
8. Une équation de la droite $(BC)$ est	$y = x$	$y = 2x - 4$	$y = -2x + 12$	$y = 2x + 4$
9. Une équation de la droite passant par $A$ et de coefficient directeur 3 est	$y = -x + 3$	$y = 3x$	$y = 3$	$y = 3x - 5$
10. La droite passant par $A$ et parallèle à $(BC)$ a pour équation	$y = 3x - 5$	$y = x - 1$	$y = -2x + 5$	$y = 2x - 3$

# 1. Séries statistiques simples

## Activité 1

### Étude d'une production

Lors de la fabrication d'une pièce, on étudie son diamètre pour un lot de 50 pièces. Les résultats sont consignés dans le tableau ci-dessous.

- 1.** Construire l'histogramme de cette série.
- 2.** Compléter le tableau donné en calculant les fréquences de chaque classe, puis les fréquences cumulées croissantes.
- 3.** Tracer le diagramme des fréquences cumulées croissantes de cette série.
- 4.** Calculer la moyenne  $\bar{x}$  et l'écart type  $\sigma$  de cette série statistique (on calculera d'abord le centre des classes).

Diamètre (en cm)	Effectif
[98,8 ; 99,2[	2
[99,2 ; 99,6[	5
[99,6 ; 100[	15
[100 ; 100,4[	22
[100,4 ; 100,8[	4
[100,8 ; 101,2[	2

## Activité 2

### Les vacances au soleil

Le tableau donne la répartition, selon le prix, de 1 000 billets « séjour » vendus par une agence de voyage pour une période donnée.

Destination	Malte	Baléares	Corse	Tunisie	Turquie	Grèce	Crète	Rép. Dominicaine	Égypte
Prix du séjour (en €)	550	600	650	700	750	775	800	875	900
Nombre de billets « séjour » vendus	30	75	150	130	210	175	150	60	20

- 1.** Construire le tableau des effectifs cumulés croissants de cette série.
- 2.** Déterminer la médiane de cette série.
- 3.** Le premier quartile  $Q_1$  est la 250<sup>e</sup> valeur de la série ordonnée des prix de séjour. Pourquoi ? Déterminer  $Q_1$ .
- 4.** Déterminer le troisième quartile  $Q_3$ .
- 5.** Construire le diagramme en boîte de cette série de prix.

## Activité 3

### Les pannes d'ordinateurs

Deux sociétés de vente de matériel informatique (Gauter et Mason) accordent un service après-vente gratuit pendant un an. Elles ont répertorié pendant cette année le nombre de dépannages fournis à chaque acheteur d'un micro-ordinateur.

Nombre $x_i$ de dépannages	0	1	2	3	4	5
Effectifs $n_i$ pour la société Gauter	27	45	29	5	2	1
Effectifs $n'_i$ pour la société Mason	129	36	31	30	25	4

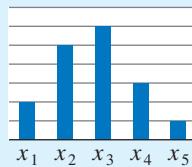
- 1.** Quel a été le nombre de ventes de micro-ordinateurs durant cette année pour l'entreprise Gauter ? pour l'entreprise Mason ?
- 2.** Déterminer, à 0,1 près, la moyenne du nombre de dépannages pour chacune de ces deux sociétés. Ceci incite-t-il à choisir plutôt l'une que l'autre ?
- 3.** Déterminer le nombre médian de dépannages pour chacune des sociétés. Quelle conséquence peut-on en tirer ?
- 4.** Calculer, à 0,1 près, l'écart type de chacune de ces deux séries statistiques. Que peut-on à présent en conclure ?

# Ce qu'il faut savoir

## ● Représentations graphiques

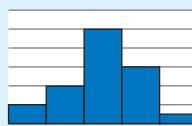
### • Diagramme en bâtons (ou en barres)

Il est formé de bâtons dont l'abscisse est la valeur  $x_i$  et la hauteur l'effectif  $n_i$  (ou la fréquence  $f_i$ ).



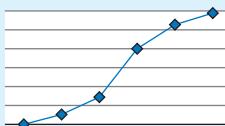
### • Histogramme

Lorsque les valeurs sont regroupées en classes, on représente la série par des rectangles ayant pour base chacune des classes et dont l'aire est proportionnelle à l'effectif de la classe (ou à la fréquence).



### • Diagramme cumulatif

Lorsque la série est regroupée en classes, le diagramme des effectifs cumulés croissants est formé des segments reliant les points ayant pour abscisse l'extrémité  $x_i$  de chaque classe, et pour ordonnée  $N_i$  l'effectif cumulé croissant de la valeur  $x_i$ .



## ● Médiane, quartiles et déciles

Les  $n$  valeurs de la série statistique sont supposées rangées dans l'ordre croissant.

### • La médiane est :

- la valeur centrale de la série si  $n$  est impair ;
- la demi-somme des deux valeurs centrales si  $n$  est pair.

• Le **1<sup>er</sup> quartile  $Q_1$**  (respectivement le **3<sup>e</sup> quartile  $Q_3$** ) est le plus petit nombre de la série tel qu'au moins 25 % (respectivement 75 %) des données soient inférieures ou égales à ce nombre.

• Le **1<sup>er</sup> décile  $D_1$**  (respectivement le **9<sup>e</sup> décile  $D_9$** ) est le plus petit nombre de la série tel qu'au moins 10 % (respectivement 90 %) des données soient inférieures ou égales à ce nombre.

## Détermination pratique

■  $Q_1$  : on calcule  $\frac{n}{4}$ , on détermine le plus petit entier  $p$  supérieur ou égal à  $\frac{n}{4}$ .  $Q_1$  est la  $p$ -ième valeur de la série ordonnée.

entier  $p$  supérieur ou égal à  $\frac{n}{4}$ .  $Q_1$  est la  $p$ -ième valeur de la série ordonnée.

■  $Q_3$  : on fait de même en remplaçant  $\frac{n}{4}$  par  $\frac{3n}{4}$ .

■  $D_1$  : on fait de même en remplaçant  $\frac{n}{4}$  par  $\frac{n}{10}$ .

L'écart interquartile est la différence  $Q_3 - Q_1$ .

## ● Moyenne, variance et écart type

Soit un caractère statistique quantitatif défini par le tableau :

Valeurs	$x_1$	...	...	$x_p$
Effectifs	$n_1$	...	...	$n_p$

La moyenne  $\bar{x}$  est donnée par la formule :

$$\bar{x} = \frac{1}{n} (n_1 x_1 + n_2 x_2 + \dots + n_p x_p) = \frac{1}{n} \sum_{i=1}^{i=p} n_i x_i.$$

La variance de cette série est le réel défini par :

$$V = \frac{1}{n} (n_1 (x_1 - \bar{x})^2 + n_2 (x_2 - \bar{x})^2 + \dots + n_p (x_p - \bar{x})^2).$$

L'écart type  $\sigma$  de cette série est  $\sigma = \sqrt{V}$ .

La détermination de l'écart type est toujours faite à la calculatrice.

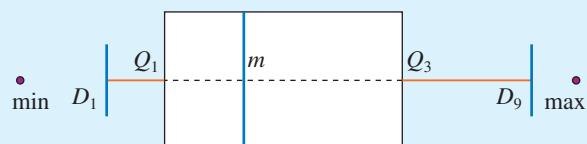
Dans le cas d'une série regroupée en classes, la moyenne et la variance se calculent par les formules données plus haut, où chaque valeur  $x_i$  est remplacée par le centre de la classe.

## ● Diagramme en boîte

C'est un diagramme regroupant la médiane, le premier et le troisième quartile. Les valeurs sont disposées sur un axe. L'abscisse du bord inférieur de la boîte est  $Q_1$ , celle du bord supérieur est  $Q_3$ , celle de la barre située à l'intérieur de la boîte est la médiane.

• Si on place le premier décile et le neuvième décile, on obtient le **diagramme en boîte à moustaches** de la série. Les moustaches sont les segments reliant les points d'abscisse  $D_1$  et  $Q_1$ , puis  $Q_3$  et  $D_9$ .

• On peut compléter ce diagramme par la valeur maximale et la valeur minimale de la série.



## 1. Séries statistiques simples

### Énoncé 1

**OBJECTIF :**

Tracer un diagramme en bâtons

Déterminer des pourcentages, la moyenne et l'écart type

Un atelier conditionne du fil sous forme de bobines. Une série de 50 mesures a donné le tableau suivant, où  $x_i$  désigne le poids d'une bobine en grammes.

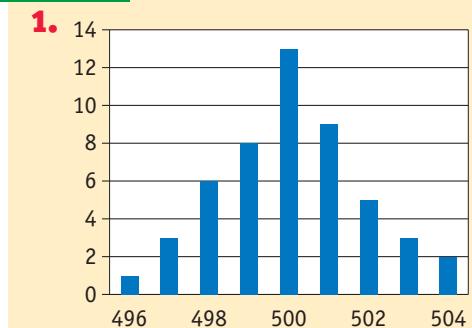
$x_i$	496	497	498	499	500	501	502	503	504
Effectif $n_i$	1	3	6	8	13	9	5	3	2

1. Tracer le diagramme en bâtons de cette série.
2. Quel est le pourcentage de bobines de poids compris entre 497 et 503 grammes ?
3. Calculer la moyenne et l'écart type de cette série de mesures ( $\approx 10^{-2}$  près).

### SOLUTION

**Méthode**

Pour calculer un pourcentage, on divise l'effectif partiel par l'effectif total.



2. Le nombre de bobines dont le poids est compris entre 497 et 503 grammes est  $50 - (1 + 2)$ , soit 47. Le pourcentage cherché est donc égal à  $\frac{47}{50}$ , soit 94 %.
3. Avec la calculatrice, on trouve la moyenne  $\bar{x}$  et l'écart type  $\sigma$  de cette série :  $\bar{x} = 500,06$  et  $\sigma \approx 1,80$ .

### Et maintenant, exercez-vous !

Une usine de tissage fabrique des rouleaux de tissu. Pour contrôler la production, le fournisseur mesure les longueurs de 200 rouleaux de tissu d'un échantillon. Un rouleau est conforme si sa longueur appartient à  $[49,6 ; 50,4]$ . On obtient :

Longueur	49,5	49,6	49,7	49,8	49,9	50	50,1	50,2	50,3	50,4	50,5	50,6
Effectif	3	4	10	25	38	42	39	23	7	4	2	3

1. Tracer le diagramme en bâtons de cette série.
2. Déterminer le pourcentage de rouleaux conformes.
3. Calculer la moyenne et l'écart type de cette série.

### Énoncé 2

**OBJECTIF :**

Tracer un histogramme

Déterminer la moyenne et l'écart type

Le responsable d'un magasin a relevé, pendant une semaine, le montant en euros des achats de chacun de ses clients.

Montant des achats	[10 ; 30[	[30 ; 50[	[50 ; 70[	[70 ; 90[	[90 ; 110[	[110 ; 130[
Nombre de clients	20	25	54	60	32	17

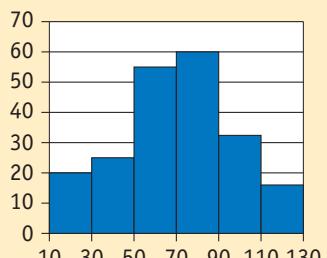
1. Construire l'histogramme de cette série.
2. Calculer, à  $10^{-2}$  près, la moyenne et l'écart type de cette série.

### SOLUTION

**Méthode**

Pour calculer la moyenne et l'écart type d'une série donnée en classes, on détermine d'abord le centre de chaque classe.

1. Voici l'histogramme de cette série.
2. Les centres des classes sont : 20, 40, 60, 80, 100 et 120. La calculatrice donne la moyenne  $\bar{x}$  et de l'écart type  $\sigma$  du montant des achats :  $\bar{x} \approx 70,58$  et  $\sigma \approx 27,31$ .



**Et maintenant, exercez-vous !**

Dans un garage, une enquête a été menée sur un échantillon de 50 véhicules quand ceux-ci se présentent pour le premier changement de l'embrayage. On a obtenu la série suivante, en fonction du kilométrage effectué avant l'opération.

Montant des achats	[70 ; 80[	[80 ; 90[	[90 ; 100[	[100 ; 110[	[110 ; 120[	[120 ; 130[	[130 ; 140[
Nombre de clients	4	3	5	7	16	9	6

1. Construire l'histogramme de cette série.
2. Calculer, à  $10^{-2}$  près, la moyenne et l'écart type des kilométrages effectués.

**Énoncé 3****OBJECTIF :**

Calculer la médiane et les quartiles

Tracer un diagramme en boîte

Dans le jeu « Kéno », on remplit un bulletin de 10 numéros pris parmi les nombres de 1 à 70. On tire au sort 20 numéros. On s'intéresse au nombre de numéros gagnants sur un bulletin. Sur 10 000 bulletins, on a obtenu les résultats suivants.

Nombre de numéros gagnants	0	1	2	3	4	5	6	7	8	9	10
Effectif	254	1 253	2 521	2 922	1 962	822	220	41	5	0	0

1. Calculer la médiane et les quartiles de cette série.
2. Construire le diagramme en boîte de cette série.

**SOLUTION**

1. On construit le tableau des effectifs cumulés croissants.

$x_i$	0	1	2	3	4	5	6	7	8	9	10
Effectifs cumulés croissants	254	1 507	4 028	6 950	8 912	9 734	9 954	9 995	10 000	10 000	10 000

La 5 000<sup>e</sup> et la 5 001<sup>e</sup> valeur de la série ordonnée sont égales à 3, donc la médiane est 3. Puisque  $\frac{10\ 000}{4} = 2\ 500$ ,  $Q_1$  est la 2 500<sup>e</sup> valeur de la série ordonnée,

soit  $Q_1 = 2$ . De même,  $Q_3$  est la 7 500<sup>e</sup> valeur de la série ordonnée, soit  $Q_3 = 4$ .

2. Voici le diagramme en boîte de cette série.

**Méthode**

Pour déterminer la médiane et les quartiles d'une série, on se sert des effectifs cumulés croissants.

**Et maintenant, exercez-vous !**

On donne le nombre des demandes de carte nationale d'identité traitées par jour, pour chacun des jours où le service est ouvert, dans une commune, pour un mois donné : 5, 11, 7, 8, 17, 6, 11, 12, 8, 20, 10, 11, 3, 6, 9, 10, 11, 5, 7, 6, 4, 5.

1. Calculer la médiane et les quartiles de cette série.
2. Construire le diagramme en boîte de cette série.
3. On estime que l'organisation du service est efficace si, pendant au moins la moitié des jours où le service est ouvert, le nombre de demandes traitées par jour est compris entre 6 et 11. L'organisation est-elle satisfaisante ?

**Vrai – Faux**

- A. La moyenne est toujours supérieure à la médiane.
- B. L'écart type est toujours positif.
- C. Le premier quartile de la série 1, 2, 3, 4, 5 est 1.
- D. Dans un diagramme en boîte, la boîte contient 50 % des valeurs.

**Activité 1****La prime d'ancienneté**

Les cadres d'une entreprise ont reçu des primes différentes selon leur ancienneté. Sept d'entre eux comparent le montant de leur prime. Leurs observations sont reportées dans le tableau ci-dessous, où l'ancienneté est exprimée en années et la prime en euros.

Cadre	n° 1	n° 2	n° 3	n° 4	n° 5	n° 6	n° 7
Ancienneté	2	8	11	17	20	22	26
Prime	270	390	460	590	650	710	830

Le tableau définit deux séries statistiques, celle donnant l'ancienneté de ces cadres, formée des valeurs  $x_1, x_2, \dots, x_7$  et celle donnant la prime de ces cadres, formée des valeurs  $y_1, y_2, \dots, y_7$ .

1. À quoi est égal  $x_1 ? x_2 ? y_1 ? y_2 ?$
2. Calculer la moyenne  $\bar{x}$  de la série statistique des anciennetés  $x_i$ .
3. Calculer la moyenne  $\bar{y}$  de la série statistique des primes  $y_i$ .
4. Dans un repère orthogonal du plan, placer les six points  $M_1(x_1 ; y_1), M_2(x_2 ; y_2), \dots$  On prendra comme unités 1 cm pour deux ans en abscisses et 1 cm pour 100 euros en ordonnées.
5. Placer le point  $G$  de coordonnées  $(\bar{x} ; \bar{y})$ . Quelle est sa position par rapport aux points déjà placés ?
6. a) Tracer la droite  $(M_1 M_7)$ .
- b) Cette droite passe-t-elle par les autres points ?
- c) Cette droite peut-elle renseigner sur la prime que peut obtenir un cadre ayant une ancienneté de 30 ans ?

**Activité 2****Vente de logiciels**

Une entreprise, spécialisée dans la vente de logiciels éducatifs, souhaite faire une étude sur l'évolution de son chiffre d'affaires. Les relevés des chiffres d'affaires semestriels des trois années antérieures sont donnés dans le tableau suivant.

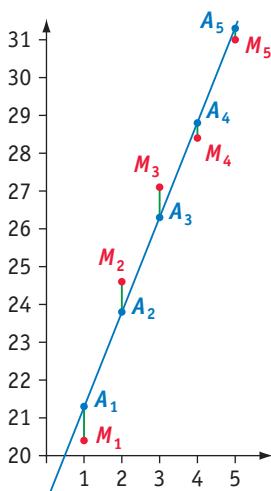
Semestres	1 <sup>er</sup> semestre 2004	2 <sup>e</sup> semestre 2004	1 <sup>er</sup> semestre 2005	2 <sup>e</sup> semestre 2005	1 <sup>er</sup> semestre 2006
Rang du semestre	1	2	3	4	5
Chiffres d'affaires semestriels (en milliers d'euros)	20,4	24,6	27,1	28,4	31

Ce tableau définit deux séries statistiques simples, la série des rangs  $x_i$  des semestres et la série  $y_i$  des chiffres d'affaires semestriels de l'entreprise.

1. Dans un repère orthogonal du plan, placer les cinq points  $M_1(x_1 ; y_1), M_2(x_2 ; y_2), \dots$  On prendra comme unités 1 cm sur l'axe des abscisses et 1 cm pour 2 milliers d'euros en ordonnées, et on débutera la graduation de l'axe des ordonnées à 20 milliers d'euros.
2. Calculer la moyenne  $\bar{x}$  de la série des valeurs  $x_i$ .
3. Calculer la moyenne  $\bar{y}$  de la série des chiffres d'affaires  $y_i$ .
4. Placer le point  $G$  de coordonnées  $(\bar{x} ; \bar{y})$ .

**Vocabulaire**

La droite  $d$  donnée par la calculatrice est appelée droite des moindres carrés.



**5.** Les points  $M_1, \dots, M_5$  sont presque alignés. On va essayer diverses droites pour trouver une droite passant au plus près de ces points.

a) Tracer la droite  $(M_1M_2)$ . Est-ce une droite acceptable ?

b) Tracer la droite  $(M_1M_5)$ . Est-ce une droite acceptable ?

c) À l'aide de la calculatrice, on trouve la droite  $d$  d'équation  $y = 2,5x + 18,8$ .

Tracer cette droite  $d$ .

Est-elle préférable aux deux droites précédentes ?

d) À l'aide de la droite  $d$ , faire une prévision pour le chiffre d'affaires au second semestre 2006.

**6. a)** Calculer les ordonnées des cinq points  $A_1, A_2, A_3, A_4, A_5$  de la droite  $d$ , d'abscisses respectives 1, 2, 3, 4 et 5.

b) Tracer les segments  $[A_1M_1], [A_2M_2], [A_3M_3], [A_4M_4]$  et  $[A_5M_5]$ .

c) La distance  $A_1M_1$  est égale à  $\sqrt{(21,3 - 20,4)^2} = 0,9$ . Montrer que la somme des carrés des distances  $A_1M_1^2 + A_2M_2^2 + A_3M_3^2 + A_4M_4^2 + A_5M_5^2$  est égale à 2,34.

d) Le même calcul pour les droites  $(M_1M_2)$  et  $(M_1M_5)$  donne respectivement 62,49 et 4,365. Dire à l'aide du dessin pourquoi cette somme est plus faible dans le cas de la droite  $d$ .

## Ce qu'il faut savoir

Quand on étudie deux caractères statistiques sur une même population, on obtient une série statistique double. Si les valeurs prises par le premier caractère sont  $x_1, x_2, \dots, x_n$  et celles prises par le second caractère sont  $y_1, y_2, \dots, y_n$ , alors les valeurs prises par cette série sont les couples  $(x_1 ; y_1), (x_2 ; y_2), \dots, (x_n ; y_n)$ .

### NUAGE DE POINTS

L'ensemble des points  $\{M_1, M_2, \dots, M_n\}$ , où  $M_i$  a pour coordonnées  $(x_i ; y_i)$ ,  $M_2 (x_2 ; y_2)$ , ...,  $M_n (x_n ; y_n)$  dans un repère du plan, est le **nuage de points** de la série.

Le **point moyen** du nuage de points  $\{M_1, M_2, \dots, M_n\}$  est le point  $G$  de coordonnées  $(\bar{x} ; \bar{y})$ , où  $\bar{x}$  est la moyenne de la série  $(x_i)$  et  $\bar{y}$  est la moyenne de la série  $(y_i)$ .

### Ajustement affine

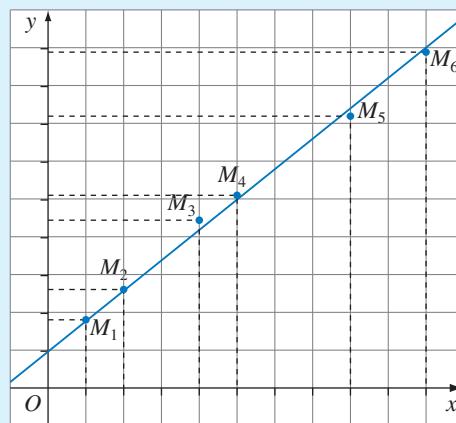
Quand les points d'un nuage sont sensiblement alignés, on peut construire une droite passant au plus près de ces points. On dit que cette droite réalise un **ajustement affine** du nuage de points.

On peut déterminer cette droite :

- soit **graphiquement** : on trace la droite au jugé, le plus près possible des points ;
- soit par la **méthode des moindres carrés** : la droite

est alors appelée droite d'ajustement de  $y$  en  $x$  par la méthode des moindres carrés, ou **droite de régression de  $y$  en  $x$** . Cette droite passe toujours par le point moyen du nuage.

Les coefficients de l'équation de cette droite s'obtiennent sur les calculatrices à l'aide du menu STAT (voir page 91).



### SÉRIE CHRONOLOGIQUE

Une série chronologique est une série qui porte sur des observations dépendant du temps.

*Exemple :* La série des chiffres d'affaires mensuels d'un hypermarché sur une année est une série chronologique. Dans ce cas, un ajustement affine peut permettre de faire des prévisions.

## Énoncé 1

## OBJECTIF :

Construire un nuage de points et un point moyen  
Faire une estimation à l'aide d'une droite d'ajustement

Le tableau suivant donne le chiffre d'affaires (en milliers d'euros) d'une entreprise de 1998 à 2005.

Année	1998	1999	2000	2001	2002	2003	2004	2005
Rang $x_i$	1	2	3	4	5	6	7	8
Chiffre d'affaires $y_i$	1 660	1 810	1 980	2 170	2 350	2 480	2 650	2 800

1. Représenter, dans un repère orthogonal, le nuage de points de cette série. On graduera l'axe des ordonnées à partir de 1 400.

2. Calculer les coordonnées du point moyen  $G$  de ce nuage. Placer  $G$  sur le graphique.

3. On choisit pour droite d'ajustement de ce nuage la droite  $d$  d'équation :

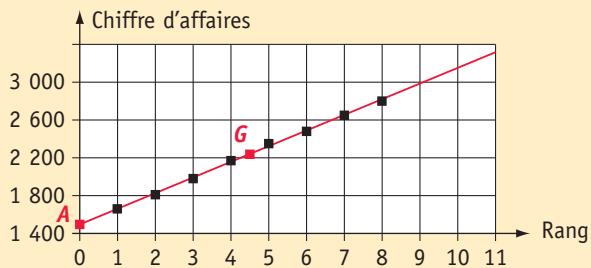
$$y = 165x + 1 495.$$

Montrer que  $d$  passe par  $G$ , puis tracer  $d$ .

4. Calculer la valeur du chiffre d'affaires prévu pour 2008 à l'aide de l'ajustement choisi.

## SOLUTION

1.



2. Le point moyen  $G$  de ce nuage a pour coordonnées  $(\bar{x} ; \bar{y})$ , où  $\bar{x}$  est la moyenne des  $x_i$  et  $\bar{y}$  la moyenne des  $y_i$ .

À l'aide de la calculatrice, on trouve  $\bar{x} = 4,5$  et  $\bar{y} = 2 237,5$ .

D'où  $G(4,5 ; 2 237,5)$ .

3.  $d$  passe par  $G$  car  $165 \times 4,5 + 1 495 = 2 237,5$ . Un autre point de  $d$  est le point  $A$  de coordonnées  $(0 ; 1 495)$ , car  $165 \times 0 + 1 495 = 1 495$ .

On peut alors tracer  $d$  à l'aide des points  $A$  et  $G$ .

4. Le rang de l'année 2008 est 11. On obtient le chiffre d'affaires prévisible en 2008 en remplaçant  $x$  par 11 dans l'équation de  $d$  :

$$y = 165 \times 11 + 1 495 = 3 310.$$

Le chiffre d'affaires estimé pour 2008 est de 3 310 milliers d'euros.

## Méthode

Pour effectuer une estimation de la variable  $y$  pour une valeur donnée de  $x$  à l'aide d'un ajustement affine, on remplace  $x$  dans l'équation réduite de la droite d'ajustement et on calcule  $y$ .

## Et maintenant, exercez-vous !

Une chaîne d'hôtels a effectué une analyse établissant un lien entre le taux d'occupation des chambres (en %) et le montant des frais de publicité (en milliers d'euros).

Frais de publicité $x_i$	30	27	32	35	35	22	24	35
Taux d'occupation $y_i$	52	45	67	55	76	48	32	72

1. Représenter le nuage de points de cette série dans un repère orthogonal. Choisir 1 cm pour 2,5 milliers d'euros et 1 cm pour 10 % en ordonnées.

2. Déterminer les coordonnées du point moyen  $G$  du nuage. Placer  $G$ .

3. On choisit comme droite d'ajustement de ce nuage la droite  $\Delta$  passant par  $G$  et par le point  $P(35 ; 72)$ . Tracer  $\Delta$ .

4. Déterminer graphiquement le montant des frais de publicité laissant espérer un taux d'occupation de 80 %.

**Énoncé 2****OBJECTIF :**

Construire un nuage de points

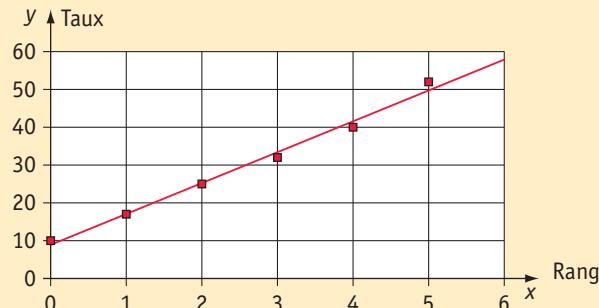
Effectuer un ajustement par la méthode des moindres carrés

Faire une estimation à l'aide de la droite d'ajustement

On étudie le taux d'équipement en micro-ordinateurs connectés à Internet des ménages français. Les résultats de cette étude sont donnés ci-dessous.

Année	2000	2001	2002	2003	2004	2005
Rang de l'année $x_i$	0	1	2	3	4	5
Taux $y_i$ en pourcentage	10	17	25	32	40	52

1. Représenter le nuage de points  $M_i(x_i ; y_i)$  dans un repère orthogonal. Unités : 2 cm pour une année et 1 cm pour un taux de 5 %.
2. Déterminer une équation de la droite d'ajustement  $d$  de  $y$  en  $x$  par la méthode des moindres carrés (on arrondira les coefficients à 0,01 près). Tracer  $d$ .
3. On réalise un ajustement du nuage à l'aide de la droite  $d$ .
  - a) Déterminer une estimation du taux d'équipement des ménages français en 2007.
  - b) À partir de quelle année peut-on estimer qu'au moins 90 % des ménages seront équipés en micro-ordinateurs connectés à Internet ?

**SOLUTION****1.**

2. Une équation de la droite d'ajustement de  $y$  en  $x$  par la méthode des moindres carrés est  $y = 8,17x + 8,90$ .

3. a) 2007 est l'année de rang 7.

Pour  $x = 7$ , on obtient  $y = 8,17 \times 7 + 8,9 \approx 66$ .

On peut estimer à 66 % le taux d'équipement des ménages en 2007.

b) On doit résoudre l'inéquation  $8,17x + 8,90 \geq 90$ .

Ceci équivaut à  $8,17x \geq 81,1$ , soit encore  $x \geq \frac{81,1}{8,17}$ . Puisque  $\frac{81,1}{8,17} \approx 9,9$ , c'est

à partir de l'année de rang 10, c'est-à-dire en 2010 qu'au moins 90 % des ménages seront équipés.

**Et maintenant, exercez-vous !**

Dans une entreprise créée en 2000, on appelle  $y$  le pourcentage de salariés payés au S.M.I.C. par rapport au nombre total de salariés en fonction du nombre  $x$  d'années écoulées depuis 2000.

<b><math>x</math></b>	0	1	2	3	4	5
<b><math>y</math></b>	8,6	10,6	10,8	12,6	13	14,3

1. Représenter le nuage de points de cette série dans un repère orthogonal. Unités : 2 cm pour une année et 1 cm pour 1 % en ordonnées.

2. Déterminer une équation de la droite d'ajustement  $d$  de  $y$  en  $x$  par la méthode des moindres carrés (on arrondira les coefficients à 0,01 près).

3. On suppose que la droite  $d$  réalise un ajustement du nuage de points. Estimer quel sera le pourcentage de salariés payés au S.M.I.C. en 2007.

# UTILISATION DES CALCULATRICES

## Séries statistiques simples

On considère la série statistique définie de la façon suivante.

$x_i$	1	2	4	7
$n_i$	2	5	11	4

On va calculer à l'aide d'une calculatrice les divers paramètres de cette série.

### 1. Saisie des données

#### Texas

- a) On appuie sur la touche **STAT**, puis on choisit le menu **EDIT**, suivi de **ENTER**.

- b) Dans l'éditeur de listes, on tape chaque valeur  $x_i$  dans la colonne L1, puis chaque effectif  $n_i$  dans la colonne L2.

Pour passer d'une valeur à l'autre, on utilise les flèches verticales du curseur.  
Pour changer de liste, on se sert des flèches horizontales du curseur.

L1	L2	L3	$\Sigma$
1	2	-----	
2	5	-----	
4	11	-----	
7	4	-----	
-----	-----	-----	
<b>L2(4) = 4</b>			

#### Casio Graph

- a) Ouvrir le menu **STAT**, en sélectionnant l'icône **STAT**, suivi de **EXE**.

- b) Dans l'éditeur de listes, on tape chaque valeur  $x_i$  dans la colonne List1, suivi de **EXE**, puis chaque effectif  $n_i$  dans la colonne List2, suivi de **EXE**.

List 1	List 2	List 3	List 4
1	1	2	
2	2	5	
3	4	11	
4	1	4	
5			

Pour changer de liste, on se sert des flèches horizontales du curseur.

### 2. Calcul des paramètres statistiques usuels

- a) À l'aide de la touche **STAT**, on choisit le menu **CALC**, puis on sélectionne **1 – Var Stats**, suivi de **ENTER**.

- b) **1 – Var Stats** apparaît à l'écran ; on tape alors :

L1 **L1** L2 (L1 et L2 sont les touches secondes des touches **1** et **2**), suivie de **ENTER**.

1-Var Stats L1,L2
-------------------

- c) Les paramètres s'affichent :  $\bar{x}$  est la moyenne,  $\sigma_x$  est l'écart type,  $n$  est l'effectif total, Med est la médiane, Q1X et Q3X sont les quartiles, MinX est la plus petite valeur et MaxX est la plus grande valeur.

- a) Activer le menu **CALC** en appuyant sur **F2**.

- b) Dans le menu **SET**, choisir **List1** pour **1 VarX List** et **List2** pour **1 Var Freq**.

Si chaque effectif vaut 1, choisir 1 pour **1 Var Freq**.

1Var XList :List1
1Var Freq :List2
2Var XList :List1
2Var YList :List2
2Var Freq :1

- c) On tape **EXIT** pour revenir au menu.

- d) On sélectionne le menu **1Var** avec la touche **F1**.

Les paramètres s'affichent :  $\bar{x}$  est la moyenne,  $x\sigma$  est l'écart type,  $n$  est l'effectif total, Med est la médiane,  $Q_1$  et  $Q_3$  sont les quartiles,  $\min X$  est la plus petite valeur et  $\max X$  est la plus grande valeur.

### 3. Effacement des données en vue d'une nouvelle étude

Dans l'éditeur de listes, pour effacer une liste, placer le curseur sur le nom de la liste, puis taper sur la touche **CLEAR**, suivie de **ENTER**.

Dans l'éditeur de listes, pour effacer une liste, placer le curseur sur un élément de la liste, sélectionner **DEL-A** (touche **F4**), puis **YES**.

#### Attention

Les calculatrices donnent dans certains cas des résultats erronés pour les quartiles  $Q_1$  et  $Q_3$ . Cela tient au fait que la définition de ces paramètres n'est pas uniformisée dans le monde. Il faut bien se souvenir que, pour ces deux paramètres, la calculatrice ne donne qu'une indication...

## Séries statistiques à deux variables

On considère la série statistique à deux variables définie par le tableau.

### 1. Saisie des données

$x_i$	1	2	3	4	5
$y_i$	2,8	5,3	6,2	8,5	11

#### Texas

- a) On appuie sur la touche **STAT**, puis on choisit le menu **EDIT**, suivi de **ENTER**.
- b) Dans l'éditeur de listes, on tape chaque valeur  $x_i$  dans la colonne L1, puis chaque valeur  $y_i$  dans la colonne L2.

### 2. Détermination des paramètres statistiques de la série

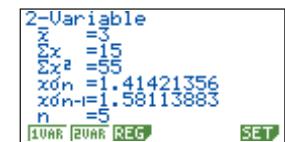
- a) À l'aide de la touche **STAT**, on choisit le menu **CALC**, puis on sélectionne **2 – Var Stats**, suivi de **ENTER**.
- b) **2 – Var Stats** apparaît à l'écran ; on tape alors :  
L1 ↵ L2 (L1 et L2 sont les touches secondes des touches 1 et 2), suivie de **ENTER**.
- c) Les paramètres s'affichent :  $\bar{x}$  est la moyenne des  $x_i$  et  $\bar{y}$  la moyenne des  $y_i$ .

### 3. Équation de la droite de régression

- a) À l'aide de la touche **STAT**, on choisit le menu **CALC**, puis on sélectionne **LinReg(ax+b)**, suivi de **ENTER**.
- b) **LinReg(ax+b)** apparaît à l'écran ; on tape alors : L1 ↵ L2 (L1 et L2 sont les touches secondes des touches 1 et 2), suivie de **ENTER**.
- c) Les paramètres de l'équation de la droite de régression de  $y$  en  $x$  s'affichent.

#### Casio Graph

- a) Ouvrir le menu **STAT**, en sélectionnant l'icône **STAT**, suivi de **EXE**.
- b) Dans l'éditeur de listes, on tape chaque valeur  $x_i$  dans la colonne **List1**, suivi de **EXE**, puis chaque valeur  $y_i$  dans **List2**, suivi de **EXE**.



- a) Activer le menu **CALC** en appuyant sur **F2**.
- b) Dans le menu **SET**, choisir **List1** pour **2 VarX List** et **List2** pour **2 VarYList**.
- c) On tape **EXIT** pour revenir au menu.
- d) On sélectionne le menu **2Var** avec la touche **F2**. Les paramètres s'affichent :  $\bar{x}$  est la moyenne des  $x_i$  et  $\bar{y}$  la moyenne des  $y_i$ .



- a) Activer le menu **CALC** en appuyant sur **F2**.
- b) Dans le menu **SET**, choisir **List1** pour **2 VarX List** et **List2** pour **2 VarYList**.
- c) On tape **EXIT** pour revenir au menu.
- d) On choisit le menu **REG** avec la touche **F3**.
- e) On choisit le premier type proposé, symbolisé par **X** (touche **F1**).

L'écran **LinearReg** affiche les paramètres de l'équation de la droite de régression de  $y$  en  $x$ .

### 4. Représentation graphique du nuage de points

- a) Activer le menu **STAT PLOTS** en appuyant sur la touche **STATPLOT**.
- b) On sélectionne le graphe que l'on va utiliser, en général **Plot 1**, suivi de **ENTER**.
- c) Dans l'éditeur, choisir **On**, le type de graphe **Scatter**, puis **L1** pour **Xlist**, **L2** pour **Ylist** et la forme des points avec **Mark**.
- d) La touche **WINDOW** permet de définir la fenêtre d'affichage. La fonction **ZoomStat** redéfinit la fenêtre en fonction des données.
- e) La touche **GRAPH** permet d'obtenir le nuage de points.

- a) Activer le menu **GRPH** en appuyant sur **F1**.
- b) Dans le menu **SET**, l'en-tête est le nom du graphique, soit **StatGraph1** en général. On choisit **Scatter** dans la rubrique **Graph Type**, **List1** pour **XList** et **List2** pour **YList**.



- c) On tape **EXIT** pour revenir au menu.
- d) Le menu **GPH1** permet d'obtenir le nuage de points. On peut parcourir celui-ci avec la fonction **Trace**.

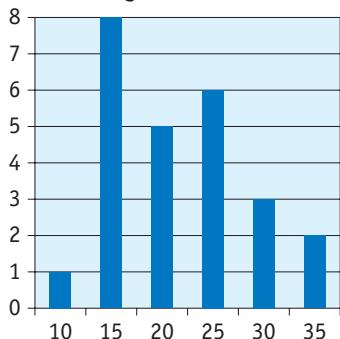
## Pour s'entraîner

### Séries statistiques simples : graphiques

- 1** Tracer le diagramme en bâtons de la série statistique suivante.

$x_i$	1	2	3	4	5
$n_i$	3	5	4	3	1

- 2** On donne le diagramme en barres suivant.

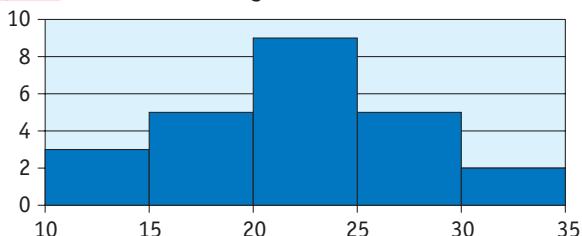


À quelle série statistique correspond ce diagramme ?

- 3** Tracer l'histogramme de la série donnée par le tableau suivant.

Classes	[0 ; 2[	[2 ; 4[	[4 ; 6[	[6 ; 8[
Effectifs	10	7	5	2

- 4** On donne l'histogramme suivant.



Quelle est la série statistique associée à cet histogramme ?

- 5** Le responsable des ventes d'un fournisseur de matériel électronique a noté le niveau de la vente journalière pendant plusieurs jours ouvrables consécutifs.

Nombre d'unités $x_i$ vendues par jour	0	1	2	3	4
Nombre de jours où l'on a vendu $x_i$	8	12	19	6	5

Tracer le diagramme en bâtons des fréquences.

- 6** Le tableau suivant donne la répartition des 355 magasins d'une enseigne de sport en France selon leur nombre d'employés.

Nombre d'employés	1	2	3	4	5	6	7
Effectifs	25	115	124	72	10	6	3

Tracer le diagramme en bâtons des fréquences.

- 7** La série suivante donne le nombre de jours de neige par année à Paris, de 1900 à 1948.

Les 49 valeurs de cette série ne sont pas classées par ordre chronologique mais par ordre croissant :

1 – 5 – 6 – 6 – 6 – 7 – 7 – 8 – 8 – 9 – 10 – 10 – 11 – 11 – 11 – 12 – 12 – 12 – 13 – 13 – 13 – 14 – 14 – 14 – 14 – 15 – 16 – 17 – 17 – 17 – 18 – 18 – 18 – 18 – 19 – 19 – 20 – 20 – 20 – 20 – 23 – 26 – 29 – 29 – 31 – 32 – 32 – 34.

- 1. a)** Calculer, à 0,1 près, le nombre moyen  $\bar{x}$  de jours de neige par année, à Paris, sur la période 1900 à 1948.

- b)** Déterminer la médiane, le premier et le troisième quartile de cette série.

- 2.** Les nombres de jours de neige par an, à Paris, ont également été relevés de 1949 à 1997. On fournit ci-dessous les caractéristiques de cette nouvelle série.

Minimum	Premier quartile	Médiane	Troisième quartile	Maximum	Moyenne
1	7	12	18	36	13,3

- a)** Donner l'écart interquartile de chacune de ces deux séries.

- b)** Construire le diagramme en boîte de chacune des deux séries.

- c)** Comparer ces deux diagrammes en boîte.

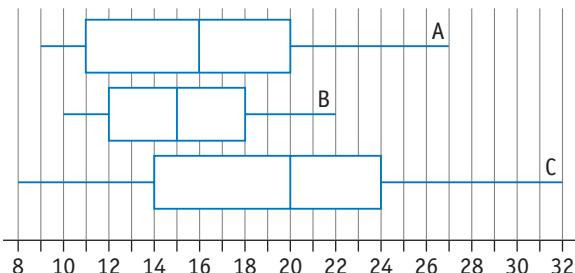
- 3.** À propos de ces relevés météorologiques, divers commentaires ont été relevés dans la presse, dont celui-ci : « Des hivers de moins en moins neigeux au cours du siècle. »

Que peut-on penser de ce commentaire ?

- 8** Une fabrique de boules de pétanque conçoit des boules de compétition de différents diamètres : 71 mm, 75 mm et 79 mm.

Un champion régional décide d'acheter des boules, mais hésite sur le diamètre. Pour faire son choix, il place un cochonnet à 9 mètres, pointe 200 fois avec chacune des boules de différents diamètres et mesure la distance au cochonnet.

Voici les diagrammes en boîte élagués aux déciles représentant ce test. Les extrémités du diagramme sont respectivement le premier et le neuvième décile. Voici quelques sensations du joueur après le test : « Avec la boule de 79 mm, j'ai réussi de très bons lancers mais également de très mauvais. Avec la boule de 71 mm, j'ai eu de très bonnes sensations, la moitié de mes lancers étaient à moins de 16 cm du cochonnet et j'en ai réussis de très beaux. Mais ma préférence va à la boule de 75 mm avec laquelle je suis plus régulier. »



Associer à chaque type de boule le diagramme en boîte correspondant.

## Séries statistiques simples : calculs de paramètres

- 9** On a relevé chaque jour pendant un mois le nombre de pannes survenues sur les distributeurs bancaires d'une ville.

Nombre de pannes	0	1	2	3	4	5
Nombre de jours	5	7	12	4	1	1

Calculer la moyenne et l'écart type de cette série.

- 10** On donne la série statistique.

$x_i$	1	3	4	5	7	8	10
$n_i$	3	13	14	16	24	17	8

- a.** Calculer les effectifs cumulés croissants.  
**b.** Déterminer la médiane de cette série.  
**c.** Calculer la moyenne et l'écart type de cette série.

- 11** Une entreprise fabrique des verres et les conditionne dans des paquets. Tous les paquets contiennent le même nombre de verres.

Sur un échantillon de 49 paquets, les nombres de verres défectueux par paquets sont les suivants :

7 - 5 - 5 - 4 - 4 - 4 - 9 - 7 - 9 - 2 - 7 - 8 - 7 - 8 - 4 - 4 - 9 - 10 - 5 - 10 - 6 - 4 - 5 - 6 - 1 - 2 - 5 - 7 - 8 - 0 - 6 - 0 - 1 - 5 - 2 - 0 - 5 - 2 - 3 - 3 - 4 - 1 - 3 - 10 - 1 - 0 - 10 - 2 - 7.

- a.** Déterminer le nombre médian de verres défectueux par paquet.

- b.** Calculer une valeur approchée à  $10^{-2}$  près de la moyenne et de l'écart type de cet échantillon.

- 12** Une entreprise spécialisée dans la décoration intérieure d'appartements réalise une enquête sur la durée de ses interventions. Les résultats à l'issue de deux mois d'enquête sont regroupés dans ce tableau.

Durée des interventions (en jours)	1	2	3	4	5
Effectif	250	220	154	78	25

- a.** Construire le diagramme en bâtons des effectifs.  
**b.** Calculer les fréquences correspondant aux durées des interventions.  
**c.** Calculer la durée médiane des interventions.  
**d.** Calculer la durée moyenne des interventions.  
**e.** L'affirmation suivante est-elle exacte ?  
« Plus de trois fois sur cinq, notre intervention a une durée inférieure à la durée moyenne de nos interventions. »

- 13** On considère la série statistique donnant les durées d'intervention, en minutes, d'un réparateur de postes de télévision sur un mois.

Durée	[0 ; 15[	[15 ; 30[	[30 ; 45[	[45 ; 60[
Nombre	32	88	47	25
	[60 ; 75[	[75 ; 90[	[90 ; 105[	[105 ; 120[
	13	7	5	3

- 1. a)** Calculer la durée moyenne d'intervention de ce réparateur.  
**b)** Calculer l'écart type de cette série.  
**2.** On suppose que ce réparateur ne fait plus toutes les réparations dont la durée dépasse 60 minutes.  
**a)** Déterminer la moyenne de la nouvelle série obtenue.  
**b)** Calculer l'écart type de cette nouvelle série. Comparer avec celui calculé à la première question.

- 14** Un site de vente aux enchères sur Internet désire réaliser une étude statistique de sa clientèle. Les responsables de l'étude utilisent un échantillon de 3 000 clients, parmi les plus réguliers du site. Les résultats concernant l'âge des clients sont donnés dans le tableau.

Classe	[13 ; 18[	[18 ; 20[	[20 ; 25[	[25 ; 30[
Effectif	148	210	815	959
	[30 ; 35[	[35 ; 45[	[45 ; 55[	[55 ; 70[
	559	221	57	31

# EXERCICES et PROBLÈMES

- Construire l'histogramme de cette série.
- Déterminer, à 0,1 près, l'âge moyen  $m$  des 3 000 clients du site de vente aux enchères.
- Déterminer, à 0,1 près, l'écart type  $\sigma$  de la série des âges des clients.
- Peut-on estimer que le pourcentage des individus qui ont un âge appartenant à la plage  $[m - \sigma ; m + \sigma]$  est supérieur ou égal à 75 % ?

- 15** Le tableau suivant donne la répartition des 1 300 salariés d'une entreprise en fonction de leur salaire mensuel (exprimé en euros) et de leur sexe.

	[1000 ; 1500]	[1500 ; 2000]	[2000 ; 2500]	[2500 ; 3000]
Hommes	440	200	50	10
Femmes	400	180	15	5

Tous les résultats seront donnés à  $10^{-2}$  près.  
Déterminer la moyenne des variables suivantes :

- le salaire des hommes ;
- le salaire des femmes ;
- le salaire du personnel.

## Nuages de points

*Dans les exercices 16 à 20, représenter le nuage de points  $M_i(x_i ; y_i)$ , puis calculer les coordonnées du point moyen du nuage.*

- 16** Le tableau suivant donne l'évolution des bénéfices  $y_i$  réalisés, en milliers d'euros, depuis le nouvel aménagement d'une crêperie en 1999 en fonction du nombre d'années  $x_i$  écoulées depuis l'aménagement.

Années	1999	2000	2001	2002	2003
$x_i$	1	2	3	4	5
$y_i$	0,96	1,69	2,03	2,31	2,55

- 17** En prévision du lancement d'un nouveau produit, une société a effectué une enquête auprès de clients éventuels pour fixer le prix de vente de ce produit. Le tableau donne le nombre  $y_i$  d'acheteurs éventuels en fonction du prix  $x_i$  de vente en euros.

$x_i$	9	10	11	12	13	14	15	16
$y_i$	120	100	90	70	60	50	40	30

- 18** Le tableau suivant donne la taille moyenne  $y_i$  d'un jeune enfant (en cm) selon son âge  $x_i$  (en mois).

$x_i$	6	9	12	15	18	21	24	27	30	33
$y_i$	66	71	74	77	80	83	85	88	90	92

- 19** Le tableau suivant donne l'évolution de la température  $y_i$  (en degrés Celsius) en fonction de l'année  $x_i$  depuis 1974 en un lieu donné.

$x_i$	1974	1978	1982	1986	1990	1994	1998
$y_i$	19,12	19,7	19,62	20	20,6	20,88	20,92

- 20** Le tableau suivant donne les hauteurs  $y_i$  (en centimètres) d'une plante mesurée tous les trois jours à midi du 1<sup>er</sup> au 16 juillet.

Jour $x_i$	1	4	7	10	13	16
$y_i$	6,5	8,4	12	15,4	19,7	

La hauteur mesurée le jour 16 a été effacée, mais on sait que le point moyen a pour ordonnée 14,5.  
Compléter le tableau.

## Ajustement affine

- 21** L'entreprise REVET'DUR fabrique un revêtement de sols anti-abrasifs pour les industries. Il se présente sous la forme d'un produit plastique liquide qui se solidifie au bout de quelques heures.

Le PDG de cette entreprise désire faire une estimation de ses ventes afin d'étudier ses investissements futurs. Dans le tableau ci-dessous est reporté la production des dix dernières années en tonnes.

Année	1996	1997	1998	1999	2000
Rang	1	2	3	4	5
Production en tonnes	49	48	51	56	58
	2001	2002	2003	2004	2005
	6	7	8	9	10
	57	61	65	66	68

- a.** Représenter graphiquement cette série chronologique. On prendra pour unités 1 cm pour une année en abscisses, et 1 cm pour 5 tonnes en ordonnées en débutant la graduation de l'axe des ordonnées à 45.

- b.** Déterminer les coordonnées du point moyen  $G$  du nuage de points.

- c.** Placer  $G$ , puis tracer la droite d'ajustement de ce nuage d'équation  $y = 2,2x + 45,8$ .

- d.** Déterminer, par le calcul, la production prévisible en 2006 et 2007.

- 22** Une étude de l'Insee a permis de dresser le tableau suivant, où  $x_i$  représente le rang de l'année et  $y_i$  représente l'indice du chiffre d'affaires hors taxe (la base 100 étant pour l'année de rang 1).

$x_i$	1	2	3	4	5	6	7	8	9	10
$y_i$	100	97	91	106	121	127	136	158	167	182

**1.** Représenter le nuage de points  $M_i$  de coordonnées  $(x_i ; y_i)$  associé à cette série statistique dans un repère orthogonal. On prendra comme unités graphiques : 1 cm pour 1 unité sur l'axe des abscisses et 1 cm pour 5 unités sur l'axe des ordonnées en commençant à 70.

**2.** Calculer les coordonnées du point moyen  $G$  du nuage.

**3.** On admet que la droite  $\Delta$  d'équation :

$$y = \frac{553}{55}x + 73,2 \text{ est un ajustement affine du nuage.}$$

**a)** Vérifier que le point  $G$  appartient à la droite  $\Delta$ .

**b)** Tracer  $\Delta$ .

**4.** On utilise cet ajustement et on admet que l'année de rang 1 correspond à 1996.

**a)** Déterminer graphiquement l'indice prévisible du chiffre d'affaires en 2006. Vérifier le résultat par le calcul.

**b)** Déterminer à partir de quelle année le chiffre d'affaires doublera par rapport à celui de 1996.

**23** Deux élèves de BTS ont créé un site Internet durant leur cycle d'études. Ils ont relevé sur le tableau suivant le nombre de visiteurs par mois de leur site, depuis sa création le 1<sup>er</sup> novembre 2004 jusqu'à la fin du mois de juin 2005.

Mois	novembre	décembre	janvier	février
Rang $x_i$	1	2	3	4
Nombre $y_i$ de visiteurs	322	325	328	327
mars	avril	mai	juin	
5	6	7	8	
334	332	335	337	

**1.** Représenter le nuage de points  $A_i$  de coordonnées  $(x_i ; y_i)$  dans un repère orthogonal d'unités : 2 cm en abscisses, 2 cm pour 5 personnes en ordonnées.

On commencera la graduation à 315 sur l'axe des ordonnées et on graduera l'axe des abscisses jusqu'à 11.

**2.** Déterminer les coordonnées du point moyen  $G$  de ce nuage. Placer  $G$  sur le graphique.

**3.** On choisit pour ajustement affine du nuage la droite  $d$  passant par  $G$  et de coefficient directeur 2.

**a)** Déterminer une équation de la droite  $d$ .

**b)** Construire la droite  $d$ .

**4.** On suppose que le nombre de visiteurs évolue en suivant cet ajustement.

**a)** Déterminer graphiquement une estimation du nombre de visiteurs au mois d'août 2005. Cette lecture devra être justifiée par un tracé en pointillé.

**b)** Déterminer par un calcul une estimation du nombre de visiteurs au mois d'octobre 2005.

**c)** Déterminer le pourcentage d'augmentation du nombre de visiteurs entre le mois de novembre 2004 et le mois d'octobre 2005. Justifier la réponse.

**24** Une entreprise étudie l'évolution, à partir de 1998, du pourcentage de cadres parmi ses employés. Le tableau suivant donne, pour les années indiquées, le nombre  $x$  d'années écoulées depuis 1998 ainsi que le pourcentage  $y$  de cadres parmi les employés.

Année	1998	1999	2000	2001	2002	2003	2004	2005
$x$	1	2	3	4	5	6	7	8
$y$	11,9	14,2	15,8	18,1	19,6	20,3	21,2	22,9

**1.** Dans un repère orthogonal  $(O ; \vec{i}, \vec{j})$  d'unité graphique 1 cm, représenter le nuage des points  $M$  de coordonnées  $(x ; y)$ . On graduera l'axe des ordonnées à partir de 10.

**2.** On nomme  $G$  le point moyen du nuage de points.

**a)** Calculer les coordonnées du point  $G$  et placer ce point sur le graphique.

**b)** Tracer sur le graphique une droite  $d$  passant par  $G$  qui réalise un bon ajustement affine du nuage de points.

**c)** Déterminer graphiquement une équation de la droite  $d$ .

**3.** On réalise, à l'aide de la droite  $d$ , un ajustement affine du nuage représenté. Utiliser l'équation de la droite  $d$  pour estimer :

**a)** le pourcentage de cadres parmi les employés de l'entreprise en 2006 ;

**b)** à partir de quelle année le pourcentage de cadres parmi les employés de l'entreprise dépasserait 30 %.

**25** Le tableau ci-après présente l'évolution de l'emploi dans l'éducation et dans la santé en France de 1968 à 1996. Par exemple, on peut lire qu'en 1968, 4,3 % de la population active de 1968 travaille dans l'éducation.

**1.** Quel est l'effectif, arrondi en millions, de la population active en France en 1968 ? en 1996 ?

**2.** Construire le nuage de points associé à la série statistique  $(x_i ; y_i)$  dans un repère orthogonal. On choisira sur l'axe des abscisses 0,2 cm pour une unité et sur l'axe des ordonnées 1 cm pour 1 %.

**3.** On note  $G$  le point moyen du nuage formé par ces cinq points.

**a)** Calculer les coordonnées de  $G$ . Placer  $G$  sur le graphique.

# EXERCICES et PROBLÈMES

**b)** On choisit pour ajustement affine du nuage la droite  $\Delta$  de coefficient directeur 0,123 et passant par  $G$ . Déterminer une équation de  $\Delta$  et tracer la droite  $\Delta$  sur le graphique.

**4. a)** Construire le nuage de points associé à la série statistique  $(x_i ; z_i)$  dans le même repère que précédemment. On représentera les points du deuxième nuage d'une couleur différente du premier.

**b)** On choisit pour ajustement affine du nuage la droite  $\Delta'$  d'équation  $y = 0,248x + 3,53$ . Tracer la droite  $\Delta'$  sur le graphique.

**5. a)** Déterminer graphiquement l'année à partir de laquelle le nombre d'emplois dans la santé dépasse celui dans l'éducation.

**b)** En utilisant les ajustements affines des questions **3** et **4**, déterminer, par le calcul, une estimation de l'année à partir de laquelle il y aura 1,5 fois plus d'emplois dans la santé que dans l'éducation.

Quelles seront alors les parts de l'emploi dans la santé et dans l'éducation ?

Année	Rang $x_i$ de l'année	Éducation		Santé et action sociale	
		en milliers	Part $y_i$ de l'emploi (en %)	en milliers	Part $z_i$ de l'emploi (en %)
1968	1	860	4,3	730	3,7
1975	8	1 180	5,6	1 140	5,4
1982	15	1 310	6,1	1 610	7,5
1989	22	1 550	7,0	2 050	9,2
1996	29	1 730	7,9	2 300	10,5

## Ajustement par la méthode des moindres carrés

**26** On a réalisé un ajustement d'une série statistique à deux variables  $x_i$  et  $y_i$  par la méthode des moindres carrés.

On a trouvé pour équation de cette droite :

$$y = 2,4x + 5,1.$$

On sait que la moyenne des  $x_i$  est 1,5.

Calculer les coordonnées du point moyen du nuage.

**27** Le point moyen du nuage d'une série statistique à deux variables  $x$  et  $y$  a pour coordonnées  $(2 ; 5)$ . Le coefficient directeur de la droite d'ajustement de  $y$  en  $x$  par la méthode des moindres carrés est 10.

Déterminer une équation de cette droite.

**28** La série statistique suivante donne l'évolution du pourcentage  $y_i$  des salariés travaillant à temps

partiel par rapport au total des salariés d'une entreprise en fonction du rang  $x_i$  de l'année.

Année	1992	1994	1995	1998	1999	2001	2002	2003
$y_i$	2	4	5	8	9	11	12	13
$x_i$	8,9	10,2	10,5	12,2	12,3	13,2	13,8	14,9

Déterminer, à l'aide de la calculatrice, une équation de la droite d'ajustement de  $y$  en  $x$  par la méthode des moindres carrés (on arrondira les coefficients à 0,1 près).

**29** Pour un produit, on a relevé au cours des huit derniers mois le prix de vente  $x_i$  au kilogramme (en euros) et les quantités achetées  $y_i$  en milliers de tonnes.

Déterminer, à l'aide de la calculatrice, une équation de la droite d'ajustement de  $y$  en  $x$  par la méthode des moindres carrés (on arrondira les coefficients à  $10^{-2}$  près).

**30** **1.** Dans un journal économique de juin 2001, un journaliste commente ainsi la baisse continue du nombre de reprises d'entreprises :

- en 1999, 43 300 sociétés ont été reprises ;
- en 2000, ce chiffre a baissé de 3 %.

Calculer le nombre de sociétés reprises en 2000. On donnera le résultat arrondi à la centaine la plus proche.

**2.** Le tableau ci-dessous représente le nombre de reprises d'entreprises en France (indice 100 pour l'année 1987) :  $x_i$  est le rang de l'année et  $y_i$  est l'indice du nombre de reprises.

Année	1993	1994	1995	1996	1997	1998	1999	2000
$y_i$	1	2	3	4	5	6	7	8
$x_i$	80	83	78	78	78	73	72	70

**a)** Représenter par un nuage de points  $M_i(x_i ; y_i)$  cette série. On prendra comme unités : en abscisses 1 cm pour une année en commençant au rang 0, et en ordonnées 1 cm pour dix points d'indice.

**b)** Déterminer une équation de la droite d'ajustement  $d$  de  $y$  en  $x$  par la méthode des moindres carrés. On arrondira les coefficients trouvés à 0,01 près.

**c)** On admet que la droite  $d$  réalise un ajustement affine convenable du nuage de points et que l'évolution du nombre de reprises d'entreprises ne sera pas modifiée dans les années à venir.

Déterminer par le calcul en quelle année l'indice descendra pour la première fois au-dessous de 50.

**31** Le tableau suivant donne le prix d'une tonne de matière première en milliers d'euros au 1<sup>er</sup> janvier de chaque année.

Année	2002	2003	2004	2005
Rang de l'année $x_i$	0	1	2	3
Prix d'une tonne en milliers d'euros $y_i$	6,48	5,74	5,19	5,01

**1.** Représenter le nuage de points associé à la série statistique  $(x_i ; y_i)$ , le plan étant rapporté à un repère orthogonal (unités graphiques : 1 cm pour une année sur l'axe des abscisses, 2 cm pour un millier d'euros sur l'axe des ordonnées).

**2.** On envisage un ajustement affine pour modéliser l'évolution du prix de cette matière première.

a) Déterminer une équation de la droite d'ajustement de  $y$  en  $x$  obtenue par la méthode des moindres carrés et la tracer sur le graphique précédent (les calculs seront effectués à la calculatrice et les résultats seront donnés à  $10^{-3}$  près).

b) En supposant que cet ajustement affine reste valable pour les années suivantes, quel serait le prix d'une tonne de matière première au 1<sup>er</sup> janvier 2007 ?

**32** En prévision du lancement d'un nouveau produit, une société a effectué une enquête auprès de clients éventuels pour fixer le prix de vente de ce produit. Les résultats sont donnés dans le tableau ci-dessous.

Prix $x_i$ de vente en euros	9	10	11	12	13	14	15	16
Nombre $y_i$ d'acheteurs éventuels	120	100	90	70	60	50	40	30

**1.** Représenter graphiquement le nuage de points  $M_i(x_i ; y_i)$ . Unités : 1 cm pour 1 sur l'axe des abscisses ; 1 cm pour 10 sur l'axe des ordonnées.

**2. a)** Déterminer une équation de la droite de régression de  $y$  en  $x$  par la méthode des moindres carrés. Tracer cette droite.

**b)** Estimer graphiquement le prix maximal pour qu'il y ait au moins 20 acheteurs potentiels.

**c)** Calculer le nombre d'acheteurs que l'on peut prévoir si le prix est fixé à 8 euros. Quelle serait alors la recette ?

**33** Une usine fabrique en très grand nombre des pièces métalliques. La fabrication est vérifiée journalièrement par prélèvement de pièces produites, dont on mesure la longueur.

En fin de journée, on calcule la longueur moyenne des pièces prélevées durant la journée et le contremaître juge que la fabrication est valable lorsque cette longueur moyenne est comprise entre 7,45 cm et 7,55 cm. Dans ce cas, on considère que la machi-

ne est bien réglée, sinon on doit procéder à son réglage.

Fin novembre, la machine est bien réglée et fabrique donc des pièces de longueur valable au début du mois de décembre. Durant les 8 premiers jours du mois de décembre, on a obtenu, en fin de journée, les résultats suivants.

Jour $x_i$	1	2	3	4	5	6	7	8
Longueur moyenne $y_i$	7,5	7,5	7,49	7,49	7,48	7,47	7,47	7,46

**1.** Représenter dans un repère orthogonal le nuage de points  $M_i(x_i ; y_i)$  associé à la série statistique ci-dessus. On utilisera comme unités 1 cm pour un jour en abscisses et 1 cm pour 0,01 cm en ordonnées en commençant à graduer l'axe à partir de 7,44.

**2. a)** Déterminer une équation de la droite d'ajustement de  $y$  en  $x$  par la méthode des moindres carrés. On arrondira les coefficients trouvés à 0,01 près.

**b)** Tracer la droite  $d$ .

**3.** On admet que la droite  $d$  représente l'évolution, dans le temps, de la longueur des pièces fabriquées par la machine. En utilisant le graphique, déterminer à partir de quel jour du mois de décembre la machine a besoin d'un nouveau réglage. Vérifier le résultat par calcul.

**34** En 2004, une caisse de retraite propose à ses adhérents un barème de rachat d'un trimestre de cotisation des années antérieures selon le tableau.

Âge de l'adhérent en années	54	55	56	57	58
Rang $x_i$	0	1	2	3	4
Montant $y_i$ du rachat d'un trimestre de cotisation en euros	2 229	2 285	2 340	2 394	2 449

(Source : CARMF, Mai 2004)

**a.** Calculer l'augmentation en pourcentage du rachat d'un trimestre entre un salarié de 54 ans et un salarié de 58 ans. On donnera le résultat arrondi à l'unité.

**b.** Représenter le nuage de points associé à la série statistique  $(x_i ; y_i)$  dans un repère orthogonal. Sur l'axe des abscisses, on placera 0 à l'origine et on choisira 2 cm pour une unité ; sur l'axe des ordonnées, on placera 2 200 à l'origine et on choisira 1 cm pour 20 euros.

**c.** Dans cette question, les calculs effectués à la calculatrice ne seront pas justifiés.

Le nuage de points permet de penser qu'un ajustement affine est justifié. Donner une équation de la

# EXERCICES et PROBLÈMES

droite de régression ( $D$ ) de  $y$  en  $x$ , obtenue par la méthode des moindres carrés. Représenter la droite ( $D$ ) dans le repère précédent.

**d.** Quel serait avec cet ajustement affine le montant du rachat d'un trimestre pour un salarié âgé de 60 ans ?

**e.** En fait, le montant de rachat d'un trimestre pour un salarié âgé de 60 ans est de 2 555 euros et le montant du rachat d'un trimestre après 60 ans est calculé de la façon suivante : à partir de 60 ans, le montant du rachat baisse de 3 % par an. Calculer le montant du rachat d'un trimestre pour un salarié ayant 65 ans.

## Ajustements non affines

**35** Le tableau suivant donne la distance de freinage nécessaire à une automobile circulant sur une route humide pour s'arrêter :  $x_i$  représente la vitesse de l'automobile en km/h et  $d_i$  la distance de freinage en mètres.

$x_i$	30	40	50	60	70	80	90	100	110	120
$d_i$	18	26	40	58	76	98	120	148	180	212

**1. a)** Représenter le nuage de points  $N_i(x_i ; y_i)$  dans un repère orthogonal, en prenant comme unités 1 cm pour 10 km/h en abscisses et 1 cm pour 25 m en ordonnées.

**b)** Tracer une droite d'ajustement de ce nuage de points.

**2.** On se propose d'améliorer cet ajustement. Pour cela, on définit la variable  $y_i = \sqrt{d_i}$ .

**a)** Donner dans un tableau les valeurs de  $y_i$  en fonction de  $x_i$ , arrondies au centième le plus proche.

**b)** Représenter le nuage de points  $M_i(x_i ; y_i)$  dans un repère orthogonal, avec pour unités graphiques,

en abscisses 1 cm pour 10 km/h et en ordonnées 1 cm pour une unité.

**c)** Déterminer une équation de la droite d'ajustement  $d$  de  $y$  en  $x$  par la méthode des moindres carrés, en arrondissant les coefficients à 0,01 près. Tracer cette droite sur le graphique précédent.

**3. a)** En utilisant l'équation de la droite  $d$ , déterminer une estimation de  $y$  si la vitesse de l'automobile était de 140 km/h. En déduire la distance de freinage, à 1 m près, correspondant à cette vitesse.

**b)** À l'aide de la droite d'ajustement tracée en **1b**, estimer graphiquement la distance de freinage à 140 km/h.

**36** Une entreprise fabrique un équipement. Le prix unitaire  $y_i$  de ce produit est fonction du nombre  $n_i$  d'unités produites. On a relevé les résultats suivants.

$n_i$	1	2	3	4	5	6	7	8	9	10
$y_i$	1 501	811	544	436	375	320	287	252	233	210

**1. a)** Représenter le nuage de points de coordonnées  $(n_i ; y_i)$ . On prendra pour unités 1 cm pour une unité en abscisses et 1 cm pour 100 euros en ordonnées.

**b)** Un ajustement affine vous paraît-il justifié ? Pourquoi ?

**2. a)** On pose  $x_i = \frac{1}{n_i}$ . Calculer les valeurs de  $x_i$  à  $10^{-2}$  près par défaut, puis représenter le nuage de points de coordonnées  $(x_i ; y_i)$ . On prendra 10 cm pour une unité en abscisses et 1 cm pour 100 euros en ordonnées.

**b)** Déterminer une équation de la droite de régression de  $y$  en  $x$  par la méthode des moindres carrés (arrondir les coefficients à  $10^{-2}$  près).

**c)** En déduire l'expression de  $y$  en fonction de  $n$ .

45 min

## Pour aller plus vite

Traiter ces exercices en 45 minutes.

**1.** Soit la série statistique définie par le tableau.

$x_i$	0	2	3	4	6
$n_i$	5	11	9	4	1

**a)** Quelle est la médiane de cette série ?

**b)** Déterminer le premier quartile et le troisième quartile de cette série.

**c)** Quelle est la moyenne de cette série ?

**d)** Quel est l'écart type de cette série ?

**2.** Soit la série statistique double.

$x_i$	1	4	5	7	10	15
$y_i$	3,5	8	9,5	15	20	28

**a)** Représenter graphiquement le nuage de points associé à cette série.

**b)** Déterminer les coordonnées du point moyen du nuage de points associé.

**c)** Déterminer une équation de la droite d'ajustement de ce nuage par la méthode des moindres carrés.

**d)** À l'aide de cet ajustement, estimer la valeur de  $y$  pour  $x$  égal à 15.

## Pour approfondir

- 37** Le tableau suivant donne, sur six semaines consécutives, le chiffre d'affaires  $C_i$  d'un grand magasin (en millions d'euros) et les dépenses publicitaires  $D_i$  de ce magasin (en milliers d'euros).

$C_i$	10	11	13	14	15	16
$D_i$	1,7	1,9	2	2,2	2,3	2,5

1. Représenter le nuage de points  $(D_i ; C_i)$  associé à cette série statistique. On prendra en abscisses 10 cm pour unité, en commençant la graduation à 1,5 et en ordonnées 1 cm pour unité, en commençant la graduation à 8.

2. Placer le point moyen du nuage.

3. a) Déterminer une équation de la droite d'ajustement de  $y$  en  $x$  par la méthode des moindres carrés (on arrondira les coefficients à 0,01 près). On exprimera  $D$  en fonction de  $C$ .

b) Tracer cette droite sur le graphique du 1.

4. Dans cette question, on reprend les mêmes données, mais on échange les rôles de  $C$  et  $D$ .

a) Sur un nouveau graphique, représenter le nuage de points de coordonnées  $(C_i ; D_i)$ .

b) Placer le point moyen de ce nouveau nuage.

c) Déterminer une équation de la droite d'ajustement de  $y$  en  $x$  par la méthode des moindres carrés (on arrondira les coefficients à 0,01 près). On exprimera  $C$  en fonction de  $D$ .

d) Tracer cette droite sur le graphique du 4a.

e) À l'aide de l'expression trouvée en 4b, calculer  $D$  en fonction de  $C$  (on arrondira les coefficients à 0,01 près).

f) Tracer la droite dont on a trouvé l'équation en 4d sur le graphique du 1.

g) Que constate-t-on pour ces deux droites ?

Ces deux droites sont appelées **droites de régression de la série** : la première droite trouvée est la droite de régression de  $y$  en  $x$ , la seconde est la droite de régression de  $x$  en  $y$ .

**38** On a relevé, pour chaque année entre 2000 et 2004, le budget de la France  $x_i$  (en milliards d'euros) et le nombre de mariages  $y_i$  en France (en milliers).

Année	2000	2001	2002	2003	2004
$y_i$	253,8	260,9	266,3	273,8	283,7
$x_i$	298	288	279	276	259

a. Représenter le nuage de points de coordonnées  $(x_i ; y_i)$ . On prendra en abscisses 1 cm pour 5 unités en commençant la graduation à 250 et en ordonnées 1 cm pour 5 unités en commençant la graduation à 250.

b. Un ajustement affine du nuage est-il envisageable ?

c. Déterminer une équation de la droite d'ajustement de  $y$  en  $x$  par la méthode des moindres carrés.

d. Peut-on estimer le nombre de mariages lorsque le budget de la France sera de 400 milliards d'euros ? de 600 milliards d'euros ?

e. Commenter les résultats trouvés en d.

**39** On a relevé en France de 1982 à 2003 le montant  $x_i$  du S.M.I.C. horaire brut en euros et la consommation annuelle moyenne d'alcool  $y_i$ , en litres d'alcool pur par personne.

Année	1982	1985	1988	1991	1994	1997	2000	2003
$x_i$	2,90	3,97	4,34	4,98	5,42	6,01	6,41	7,19
$y_i$	18,5	17,3	16,4	15,7	15,5	15	14,2	14

a. Représenter le nuage de points de coordonnées  $(x_i ; y_i)$ . On prendra en abscisses 1 cm pour une unité et en ordonnées 1 cm pour une unité en commençant la graduation à 12.

b. Un ajustement affine du nuage est-il envisageable ?

c. Déterminer une équation de la droite d'ajustement de  $y$  en  $x$  par la méthode des moindres carrés.

d. Peut-on estimer la consommation d'alcool en France lorsque le S.M.I.C. sera de 12 euros ?

e. À partir de quelle valeur du S.M.I.C. la consommation d'alcool en France aura-t-elle cessé ? Pourquoi alors le gouvernement ne porte pas le S.M.I.C. à ce niveau comme mesure de santé publique ?

**40** Un magasin d'électroménager vend des aspirateurs depuis le 1er janvier 1998.

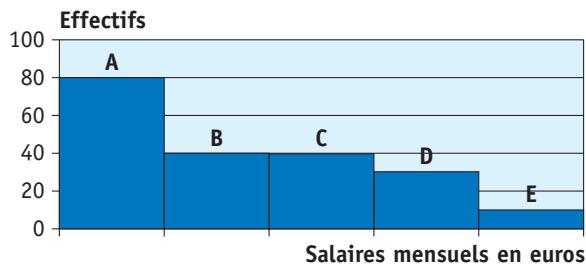
Son directeur nous a fourni les renseignements ci-dessous.

Année	1998	1999	2000	2001	2002	2003
Nombre $y_i$ d'aspirateurs vendus	594	670	770	830	930	1 000

À l'aide d'un ajustement affine que l'on justifiera, donner une estimation du nombre d'aspirateurs que le magasin peut espérer vendre en 2008.

## Pour le Bac

**Énoncé A** Voici la répartition des salaires dans une entreprise. On dénombre cinq classes de salaires différentes.



Par exemple, les salariés appartenant à la classe A touchent un salaire mensuel compris entre 1 000 euros inclus et 1 400 euros exclu.

- Recopier et compléter à l'aide du diagramme le tableau suivant.

Salaires mensuels en euros	[1 000 ; 1 400[	[1 400 ; 1 800[	[1 800 ; 2 200[	[2 200 ; 2 600[	[2 600 ; 3 000[
Effectifs					

- Déterminer le nombre de salariés de l'entreprise.
- Déterminer le salaire mensuel moyen dans cette entreprise.
- Déterminer l'écart type de la série des salaires.

D'après Bac STT CG-IG, 2002.

**Énoncé B** Le tableau suivant donne le prix (exprimé en euros) d'une machine de 1999 à 2004.

Année	1999	2000	2001	2002	2003	2004
Rang $x_i$	1	2	3	4	5	6
Prix $y_i$	18 300	18 900	19 800	20 400	21 000	21 900

- Construire, dans un repère orthogonal, le nuage de points  $M_i$  de coordonnées  $(x_i ; y_i)$  associé à cette série statistique. On prendra sur l'axe des abscisses 2 cm pour unité, sur l'axe des ordonnées 1 cm pour un millier d'euros en commençant à graduer à partir de 10 000.
- Déterminer les coordonnées du point moyen du nuage, et le placer sur le graphique.
- Déterminer une équation de la droite d'ajustement de  $y$  en  $x$  par la méthode des moindres carrés (arrondir les coefficients trouvés à 0,01 près). Tracer cette droite.
- En utilisant l'ajustement précédent, déterminer graphiquement, puis par le calcul, une estimation du prix de la machine en 2005.

D'après Bac STT ACA-ACC, 2005.

**Énoncé C** Un opérateur de radiotéléphonie est amené chaque année à réaliser des investissements considérables pour améliorer son réseau.

Le tableau suivant donne les investissements réalisés par cet opérateur de 1998 à 2002, ainsi que le nombre d'abonnés obtenu.

Années	1998	1999	2000	2001	2002
Investissements $x_i$ en milliards d'euros	1	1,1	1,2	1,3	1,4
Nombre d'abonnés $y_i$ en milliers	90	100	105	110	112

- Représenter le nuage de points  $M_i(x_i ; y_i)$  dans un repère orthogonal d'unités graphiques 2 cm pour 0,1 milliard d'euros en abscisses, et 5 cm pour 10 milliers d'abonnés en ordonnées. On commencera la graduation de l'axe des abscisses à 1 et celle des ordonnées à 80.

- Madame Armand propose un ajustement graphique à l'aide de la droite  $d$  passant par les points  $A(1,1 ; 100)$  et  $B(1,3 ; 110)$  du nuage. Déterminer une équation de cette droite.

- Madame Pons propose d'ajuster le nuage par la courbe représentative  $\mathcal{C}$  de la fonction  $f$  définie sur l'intervalle  $[1 ; 1,6]$  par  $f(x) = a - \frac{b}{x}$ , où  $a$  et  $b$  sont des réels.

Pour cela, on pose  $X = -\frac{1}{x}$ .

- Recopier et compléter le tableau :

$X_i$					
$y_i$	90	100	105	110	112

On arrondira les valeurs de  $X_i$  à 0,01 près.

- Déterminer une équation de la droite d'ajustement de  $y$  en  $X$  par la méthode des moindres carrés. On arrondira les coefficients trouvés à 0,01 près.

- En déduire l'expression de  $y$  en fonction de  $x$ .
- En 2003, l'opérateur a augmenté ses investissements de 0,2 milliard d'euros. Le nombre d'abonnés observé a été de 118 000.

- Calculer l'estimation du nombre d'abonnés en 2003 avec chacun des modèles proposés par Madame Armand ou Madame Pons.

- En considérant la valeur effectivement observée en 2003, quel modèle vous paraît le plus approprié ?

D'après Bac STT ACA-ACC, 2004.

# 1. La méthode des moindres carrés sur tableau

L'objectif de cette activité est d'utiliser le tableur pour tracer un nuage de points, de déterminer une équation de la droite d'ajustement de ce nuage par la méthode des moindres carrés, de tracer cette droite et d'estimer des valeurs à l'aide de cet ajustement.

Monsieur Martin tient une librairie dont le chiffre d'affaires ne cesse de croître depuis 2000.

Année	2000	2001	2002	2003	2004	2005
Rang $x_i$ de l'année	1	2	3	4	5	6
Chiffre d'affaires $x_i$ en milliers d'euros	92	94,5	96,5	99	101,7	106

## 1. Étude statistique de cette série

**1.** Entrer les valeurs  $x_i$  dans la colonne A et les valeurs  $y_i$  dans la colonne B.

**2.** Pour représenter le nuage de points de cette série, sélectionner la plage de valeurs, puis, avec l'assistant graphique, utiliser le type « Nuage de points » et le premier sous-type de graphique.

**3.** Les points étant sensiblement alignés, on va déterminer une équation de la droite de régression de  $y$  en  $x$ .

**a)** Rendre la fenêtre graphique active en cliquant sur le graphique (zone de traçage), puis dans le menu **Graphique**, choisir « Ajouter une courbe de tendance », et sélectionner « Type de régression linéaire » : la droite cherchée s'affiche.

**b)** Pour obtenir une équation de cette droite, il suffit de cliquer (droit) sur la droite tracée, de sélectionner « Format de la courbe de tendance », puis « Options », et de cocher le critère : afficher l'équation sur le graphique.

**4.** Pour déterminer une estimation du chiffre d'affaires en 2006, on utilise la fonction **Tendance** : on accède à cette fonction avec l'assistant « Coller une fonction » **fx**, puis en choisissant le menu **Statistiques**. On entre dans **Y\_connus** la plage des valeurs  $y_i$ , dans **X\_connus** la plage des valeurs  $x_i$  et dans les **X\_nouveaux** la valeur 7.

Calculer cette valeur estimée, puis faire de même pour les années 2007 et 2008.

## 2. Comparaison de différentes droites d'ajustement

Dans cette partie, on veut comparer la « qualité » de différentes droites d'ajustement de ce nuage. Pour cela, on calcule pour chaque valeur  $x_i$  le « résidu »  $y_i - \hat{y}_i$ , où  $\hat{y}_i$  est la valeur calculée à l'aide de l'équation de la droite choisie.

A	B	C	D	E	G	H	I	J	K
1			Moindres carrés		Droite d1			Droite d2	
2	Rang	Chiffre d'affaires	$\hat{y}_i$	Carrés des écarts	$\hat{y}_i$	Carrés des écarts		$\hat{y}_i$	Carrés des écarts
3	1	92	91,5616	0,19219456	92,7				
4	2	94,5	94,2502	0,06240004		95,4			
5	3	96,5	96,9398	0,19254544		98,1			
6	4	99	99,6274	0,39363076		100,8			
7	5	101,7	102,3116	0,379456		103,5			
8	6	106	105,0046	0,99082115		106,2			
		somme	2,21104796	somme					

**1.** On considère d'abord la droite de régression déterminée par le tableur.

**a)** Calculer les valeurs  $\hat{y}_i$  dans la colonne D à l'aide de l'équation donnée par le tableur.

**b)** Calculer les carrés des résidus  $(y_i - \hat{y}_i)^2$  dans la colonne E, puis la somme de ces carrés des résidus en E9.

**2.** Faire de même avec la droite  $d_1$  suggérée par le graphique, celle d'équation  $y = 2,7x + 90$ .

**3.** On considère à présent la droite  $d_2$  passant par les deux points extrêmes du nuage.

**a)** Déterminer une équation de  $d_2$ .

**b)** Calculer la somme des carrés des résidus si l'on prend  $d_2$  comme droite d'ajustement.

**4.** Une calculatrice donne pour droite d'ajustement la droite  $d_3$  d'équation  $y = 2,68x + 88,87$ . Calculer la somme des carrés des résidus pour cette droite  $d_3$ .

**5.** Comparer tous les résultats trouvés dans cette partie.

## 2. Lissage par la méthode des moyennes mobiles sur tableur

L'objectif de cette activité est d'utiliser le tableur pour lisser une série chronologique, c'est-à-dire éliminer les variations ponctuelles d'une série chronologique afin de pouvoir dégager une tendance à plus long terme. Un cas d'application remarquable de cette méthode est l'étude des cours d'une action à la bourse.

### 1. Étude d'une série chronologique

Voici, en euros, les diverses valeurs prises par une action lors du mois de septembre 2005 à la bourse de Paris (jours ouvrables).

18,04	18,04	18,12	18,45	18,67	18,80
18,68	18,91	18,66	18,55	18,27	18,25
17,77	17,35	17,59	18,03	17,72	17,37
17,17	17,30	17,42	17,53		

1. Entrer dans la colonne A d'une feuille de calcul les numéros d'ordre des jours ouvrables du mois.
2. Saisir dans la colonne B les cours de cette action pour le mois.
3. Pour représenter cette série chronologique, sélectionner tout d'abord la plage des données, c'est-à-dire A2:B23. Dans l'assistant graphique, choisir Nuage de points, puis Nuage de points reliés par une courbe.

### 2. Lissage des données

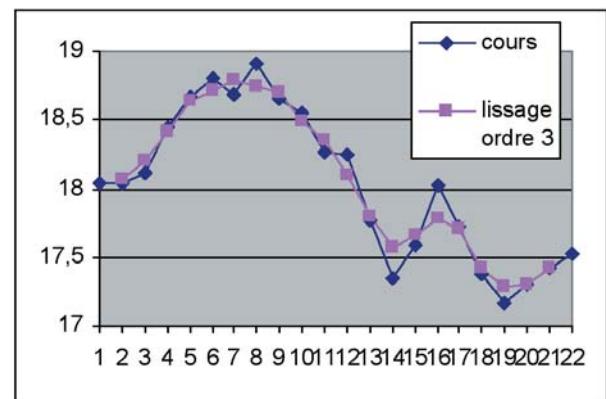
Les variations journalières du cours de l'action gênent pour dégager une tendance claire.

1. Dans un premier temps, on va remplacer la valeur de l'action du jour 2 par la moyenne des valeurs des jours 1, 2 et 3, puis remplacer la valeur de l'action du jour 3 par la moyenne des valeurs des jours 2, 3 et 4... jusqu'à l'avant-dernière valeur. Quelle formule doit-on écrire dans la cellule C3 pour obtenir la moyenne des valeurs écrites en B1, B2 et B3 ?
2. Recopier cette formule vers le bas jusqu'en C22 : on obtient dans la colonne C la série des valeurs lissées par la **méthode des moyennes mobiles d'ordre 3** (on arrondira les valeurs trouvées à deux décimales).

3. Pour représenter la série initiale ainsi que la série obtenue par lissage, sélectionner la plage A2:C23 et opérer comme dans la **partie 1**.

4. Effectuer à présent un lissage par moyennes mobiles d'ordre 5 : on remplace la valeur du jour 3 par la moyenne des valeurs des jours 1 à 5, et ainsi de suite. Calculer les valeurs dans la colonne D, puis représenter les trois séries obtenues.

5. Comparer les trois courbes obtenues.



### 3. Application

On a mesuré la température en un lieu donné à la même heure durant le mois de septembre 2005. On a obtenu les résultats suivants, en degrés.

17,2	15,8	16,5	21,3	17,5	18,3	20	17,1
16,9	17,4	16,8	22,5	17,4	18	17,6	13,2
13,9	17	17,6	18,4	17,2	17,1	16,5	17,4
23,5	18,2	17,7	16,9	17,2	17,1		

1. Représenter cette série chronologique à l'aide du tableur.
2. Effectuer un lissage d'ordre 3 des valeurs par la méthode des moyennes mobiles, puis un lissage d'ordre 5 (arrondir à deux décimales).
3. Représenter les séries obtenues. Commenter les graphiques.

# Et pour finir testez-vous !

Pour chacune des questions suivantes, choisir la (ou les) bonne(s) réponse(s).

## A. Savez-vous... étudier une série statistique à une variable ?

La série statistique suivante donne le nombre de tomates abîmées dans 800 cagettes de 8 tomates.

Nombre de tomates abîmées $x_i$	0	1	2	3	4	5	6	7	8
Nombre de cagettes $n_i$	194	215	172	126	54	27	9	2	1

	A	B	C	D
1. La moyenne de cette série est, à 0,1 près	1	1,5	1,7	2
2. L'écart type de cette série est, à 0,1 près	0,5	1	1,5	2
3. La médiane de cette série est	1	1,1	1,5	1,7
4. Le premier quartile de cette série est	0,1	0,2	0,5	1
5. Le troisième quartile de cette série est	2	2,2	2,5	3
6. Le neuvième décile de cette série est	4	4,5	5	6

## B. Savez-vous... étudier une série statistique à deux variables ?

Le tableau suivant donne l'évolution du montant du S.M.I.C. horaire brut en euros.

Année	2001	2002	2003	2004	2005
Rang de l'année $x_i$	0	1	2	3	4
S.M.I.C. $y_i$	6,67	6,82	7,19	7,61	8,03

On considère le nuage de points  $M_i$  de coordonnées  $(x_i ; y_i)$ .

	A	B	C	D
7. Pour cette série statistique	Cette série est une série chronologique	Le point $A(2003 ; 7,19)$ est un point du nuage	L'année de rang 10 est 2010	Un ajustement affine du nuage est envisageable
8. Le point moyen du nuage de points associé à cette série a pour coordonnées, à 0,01 près	(2 ; 7,19)	(2 ; 7,26)	(2,5 ; 7,25)	(2 ; 7)
9. Une équation de la droite d'ajustement de ce nuage par la méthode des moindres carrés est (à 0,01 près)	$y = 0,40x + 7$	$y = 0,30x + 6,45$	$y = 0,35x + 6,75$	$y = 0,35x + 6,56$
10. En 2006, on peut prévoir un montant du S.M.I.C. horaire brut en euros égal à	8,15	8,31	8,43	8,50
11. Si cet ajustement reste valable, le S.M.I.C. horaire brut dépassera 10 € en	2009	2010	2011	2012