CHAPITRE 11 Statistiques

ACTIVITÉS (page 267)

Activité 1

1 À un âge donné, 97 % des bébés ont une masse qui ne dépasse pas celle définie par la courbe supérieure.

2 **a)** Pour les enfants de trois mois : $M_1 \approx 5$ kg.
b) Cette masse correspond au premier quartile Q_1.

3 **a)** Pour les enfants deux ans : $Q_3 \approx 13$ kg.
b) Pour la série des enfants d'un an : Me $\in$ [8,5 ; 10,3].
Pour la série des enfants de deux ans : Me $\in$ [11 ; 13].
c) Les bébés de six mois sont tels que 75 % d'entre eux ont une masse de 8 kg ou moins.

Activité 2

1 Pour M_1 et M_2 on obtient la même moyenne : $\overline{x} = 112$.

2 Feuille de calcul pour M_1 :

	A	B	C
1	Nombre de perles: x_i	Nombre de sachets : n_i	Carrés des écarts à la moyenne: $(x_i - \overline{x})^2$
2	108	3	16
3	109	3	9
4	110	10	4
5	111	33	1
6	112	95	0
7	113	39	1
8	114	16	4
9	115	1	9
10	Nombre moyen de perles par sachet		Moyenne des carrés des écarts
11	112		1,3

Feuille de calcul pour M_2 :

	A	B	C
1	Nombre de perles: x_i	Nombre de sachets : n_i	Carrés des écarts à la moyenne: $(x_i - \overline{x})^2$
2	108	1	16
3	109	5	9
4	110	16	4
5	111	79	1
6	112	30	0
7	113	26	1
8	114	25	4
9	115	18	9
10	Nombre moyen de perles par sachet		Moyenne des carrés des écarts
11	112		2,46

Les moyennes pondérées des carrés des écarts $(x_i - \overline{x})^2$ pour M_1 et M_2 sont respectivement 1,3 et 2,46.
Ainsi M_1 procède à la mise en sachet la plus régulière.

PROBLÈME OUVERT

Sur l'année, les agents A, B et C ont vendu chacun 48 voitures donc leurs moyennes mensuelles de vente sont égales : $\overline{x} = 4$.
La comparaison des variances permet de déterminer le profil du vendeur « le plus régulier ».

On obtient : $V_A = \dfrac{7}{2}$, $V_B = \dfrac{8}{3}$ et $V_C = \dfrac{7}{3}$.
Ainsi la variance la plus petite, qui traduit la dispersion la plus faible autour de la moyenne, est V_C.
Chloé sera choisie !

EXERCICES Application (page 271)

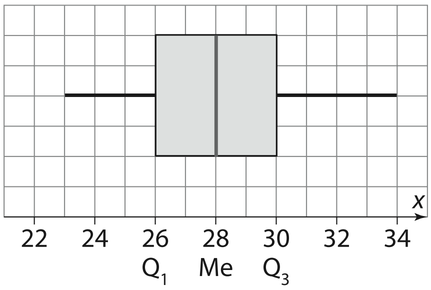
1 **1.** Répartition des vainqueurs suivant l'âge :

Âge	23	24	25	26	27	28	29	30	31	32	33	34
Effectif	1	4	4	3	5	10	6	3	3	2	2	2

L'âge le plus fréquent est 28 ans.

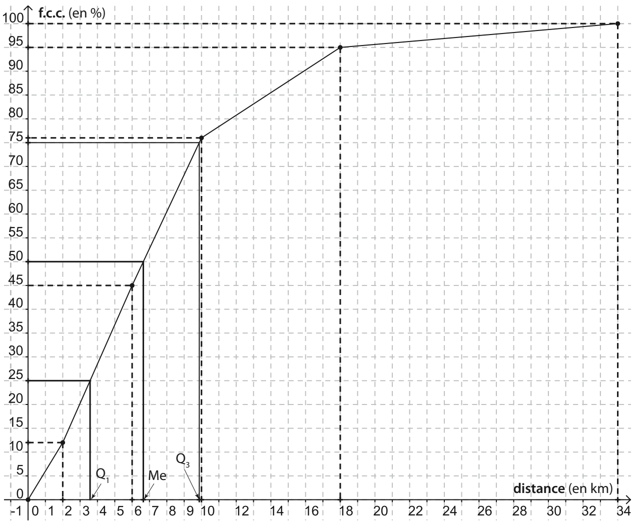
2. a) Me = 28 ; Q_1 = 26 ; Q_3 = 30.

b) Diagramme en boîte :



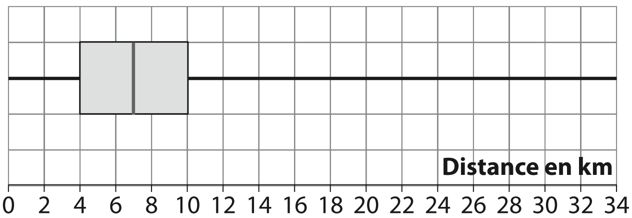
c) Au moins 50 % des vainqueurs du Tour de France ont entre 26 et 30 ans.
Au moins un quart des vainqueurs ont 26 ans ou moins.
Au moins un quart des vainqueurs ont 30 ans ou plus.

2 **1.** Courbe des f.c.c.



2. a) Graphiquement :
$Q_1 \approx 3{,}5$ km ; Me ≈ 6,6 km ; $Q_3 \approx 10$ km.

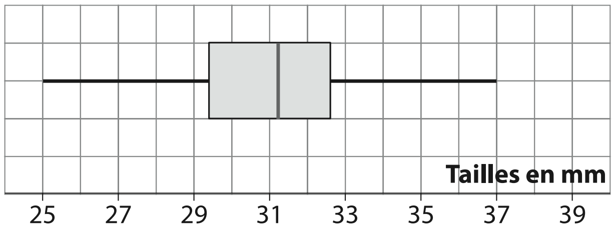
b) Diagramme en boîte :



Environ la moitié des élèves habitent entre 3,5 et 10 km du lycée.
Environ un quart des élèves habitent à moins de 3,5 km alors qu'un quart des élèves habitent à plus de 10 km.

3 **1.** Graphiquement :
$Q_1 \approx 29{,}4$ mm ; Me ≈ 31,2 mm ; $Q_3 \approx 32{,}6$ mm.

2. a) Diagramme en boîte :



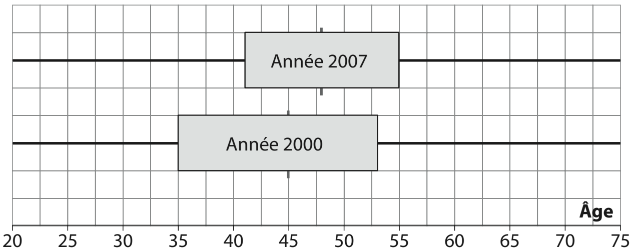
b) Environ la moitié des plumes ont une taille comprise entre 29,4 et 32,6 mm. Environ un quart des plumes mesurent moins de 29,4 mm alors qu'un quart des plumes mesurent plus de 32,6 mm.

4 *Corrigé dans le manuel.*

5 **1.** Graphiquement, on obtient :

Année 2000	$Q_1 \approx 35$	Me ≈ 45	$Q_3 \approx 53$
Année 2007	$Q'_1 \approx 41$	Me' ≈ 48	$Q'_3 \approx 55$

2. a) Diagramme en boîtes superposées :

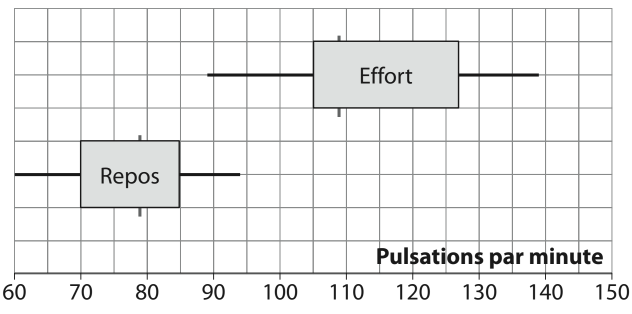


b) On constate que : $Q'_1 > Q_1$, Me' > Me et $Q'_3 > Q_3$.
Graphiquement, la boîte qui correspond à l'année 2007 est décalée vers la droite.
Les écarts interquartiles sont tels que $Q'_3 - Q'_1 = 14$ et $Q_3 - Q_1 = 18$. Ainsi la concentration autour de la médiane est plus importante en 2007 qu'en 2000.
On peut estimer qu'entre les années 2000 et 2007, la répartition des âges traduit une tendance au vieillissement de la population des chefs d'entreprises agricoles.

6 **1.** Les paramètres sont donnés ci-dessous :

	min	Q_1	Me	Q_3	max
Repos	60	70	79	85	94
Effort	89	105	109	127	139

D'où le diagramme en boîtes superposées :



2. a) Comparons les étendues et les écarts interquartiles :

	Étendue	Écart interquartile
Repos	34	15
Effort	50	22

L'étendue de la série des rythmes cardiaques avec effort est la plus importante. De même, la comparaison des écarts interquartiles indique que la dispersion des valeurs autour de la médiane est plus importante lors des tests avec effort.

b) Lors des tests au repos, environ :

50 % des élèves ont un rythme cardiaque compris entre 70 et 85 ; 25 % ne dépassent pas 70 pulsations par minute alors que 25 % dépassent 85 pulsations par minute.

– Lors des tests avec effort, environ :

la moitié des élèves ont un rythme cardiaque compris entre 105 et 127 ; un quart des élèves ne dépassent pas 105 pulsations par minute alors qu'un quart dépassent 127 pulsations par minute.

7 **1. a)** On constate que plus le nombre n de simulations augmente, plus la dispersion des fréquences, évaluée par l'étendue et l'écart interquartile, diminue.

b) On peut conjecturer que lorsque n devient grand, la fréquence tend à se stabiliser vers une valeur théorique qui correspond à la probabilité d'obtenir S = 7 lors du lancer de deux dés parfaits. (On retrouve la loi des grands nombres : voir Chapitre 12, Activité 2, page 293.)

2. a) Lors du lancer de deux dés parfaits, une issue est un couple du type $(x\,;\,y)$ où x est le numéro du dé 1 et y celui du dé 2 ; toutes ces issues sont équiprobables.

Le tableau indique les sommes correspondantes :

Dé 1 / Dé 2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

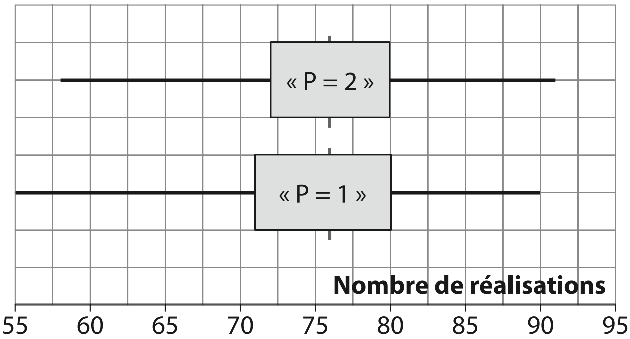
$P(S = 7) = \dfrac{6}{36} = \dfrac{1}{6}$.

b) Ce résultat est en accord avec la simulation où la médiane des séries de fréquences semble se stabiliser vers cette valeur $\dfrac{1}{6}$.

8 **1.** Tableau des différents paramètres :

	min	Q_1	Me	Q_3	max
P = 1	55	71	76	80	90
P = 2	58	72	76	80	91

2. Diagramme en boîtes superposées :



3. Ces deux séries ont la même médiane et des écarts interquartiles proches. Les valeurs extrêmes diffèrent sensiblement (effet de la fluctuation d'échantillonnage).

La fréquence médiane de réalisation de ces issues est $\dfrac{76}{200} = 0{,}38$. On peut conjecturer que la répartition des fréquences des issues P = 1 et P = 2 sont comparables, avec des fréquences médianes proches de 0,375.

Remarque : lors du lancer de trois pièces équilibrées, les probabilités d'obtenir 1 fois Pile ou 2 fois Pile sont égales à $\dfrac{3}{8}$ (voir Chapitre 13 : loi binomiale).

9 **1.** Indicateurs des deux séries :

	Vân	Léa
Moyenne	$\overline{x} = \dfrac{217}{30}$	$\overline{x'} = \dfrac{217}{30}$
Écart-type	$s \approx 1{,}94$	$s' \approx 1{,}41$

2. Vân et Léa ont la même moyenne ; la dispersion des tirs de Léa autour de la moyenne est plus faible que pour Vân $(s' < s)$ donc Léa est plus régulière.

10 *Corrigé dans le manuel.*

11 **1.** Durées moyennes d'attente, en secondes :

$$\overline{d_A} \approx 66\,;\ \overline{d_B} \approx 74.$$

2. Écarts-types, en secondes :

$$s_A \approx 44\,;\ s_B \approx 49.$$

3. L'opérateur A est le plus efficace puisqu'il offre :

– la durée d'attente moyenne la plus faible $(\overline{d_A} < \overline{d_B})$;

– une meilleure concentration des durées d'attente autour de la moyenne $(s_A < s_B)$.

12 **1.** On utilise les centres des classes pour obtenir une estimation de la SAU moyenne.

Nord-Pas-de-Calais : $\overline{x} \approx 82$ ha.

Picardie : $\overline{x'} \approx 130$ ha.

2. De même pour une estimation des écarts-types :

Nord-Pas-de-Calais : $s \approx 60$ ha.

Picardie : $s' \approx 90$ ha.

3. Les deux régions possèdent un nombre comparable d'exploitations agricoles (Nord-Pas-de Calais : 10 209 ; Picardie : 10 160). Notons qu'en Picardie :

– la SAU moyenne est plus grande : $\overline{x'} > \overline{x}$;

– la dispersion autour de la moyenne est plus importante : $s' > s$.

Ce phénomène est dû, en particulier, à la différence de répartition dans les classes extrêmes : en Picardie, 19 % des SAU ne dépassent pas 50 ha et 17 % dépassent 200 ha, alors que dans le Nord-Pas-de-Calais, les pourcentages correspondants sont respectivement 35 % et 3 %.

EXERCICES Activités de recherche (page 278)

17 **Scénario d'évolution**

• ***Les outils :***

– Représentations graphiques.

– Indicateurs statistiques.

• ***Les objectifs :***

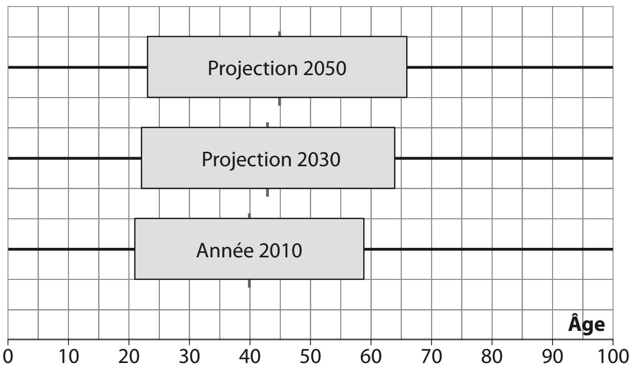
– Lire différents types de graphiques.

– Interpréter des paramètres statistiques.

– Comparer des séries de données.

1. Projection 2030 : $Q_1 \approx 22$, Me ≈ 43, $Q_3 \approx 64$.
Projection 2050 : $Q_1 \approx 23$, Me ≈ 45, $Q_3 \approx 66$.

2. Diagramme en boîtes superposées :



3. Les projections prévoient un vieillissement de la population avec :

– une stabilité de la population des jeunes (environ un quart de la population de moins de 22 ans) ;

– un recul de l'âge médian qui passera de 39,5 ans en 2010 à 43 ans en 2030 et à 45 ans en 2050 ;

– une plus grande dispersion dans la tranche centrale (environ 50 % de la population avait entre 21 et 58,5 ans en 2010 alors que la moitié de la population aura entre 22 et 64 ans en 2030, et entre 23 et 66 ans en 2050) ;

– une plus grande proportion de seniors (moins de 25 % de la population dépassait 60 ans en 2010 alors que le quart de la population dépassera 64 ans en 2030 et 66 ans en 2050).

18 **Effet de structure**

• ***Les outils :***

– Tableau de données à double entrée.

– Calculs de moyennes dans divers référentiels.

• ***Les objectifs :***

– Exploiter les données d'un tableau.

– Comparer des moyennes.

– Expliquer un effet de structure.

1. Dépense moyenne par habitant : $\overline{d_A} \approx 18$ € ; $\overline{d_B} \approx 14$ €.

2. a) Tableau des dépenses moyennes par personne et tranche d'âge :

Âge	[0 ; 20[	[20 ; 30[	[30 ; 50[	[50 ; 90[
A	8	17	19	24
B	8	17	19	24

b) Pour chaque tranche d'âge les moyennes sont de même niveau pour les villes A et B.

3. Pourtant, la dépense moyenne par personne dans la ville A est supérieure à celle dans la ville B.
Ce phénomène est dû, en particulier, à la plus grande proportion de personnes âgées (tranche [50 ; 90[) dans la ville A (36 %) que dans la ville B (7 %).

19 **Une propriété de la variance**

• ***Les outils :***

– Formule de la variance.

– Calcul algébrique.

– Variations d'un trinôme du second degré.

• ***Les objectifs :***

– Exprimer une fonction qui traduit la moyenne des carrés des écarts d'un réel t aux valeurs de la série.

– Prouver que la variance est le minimum de cette fonction.

1. Pour tout t de I,

$g(t) = f_1(x_1 - t)^2 + f_2(x_2 - t)^2 + \dots + f_p(x_p - t)^2$

$g(t) = f_1(x_1^2 - 2x_1t + t^2) + f_2(x_2^2 - 2x_2t + t^2) + \dots + f_p(x_p^2 - 2x_pt + t^2)$

$g(t) = (f_1 + f_2 + \dots + f_p)t^2 - 2(f_1x_1 + f_2x_2 + \dots + f_px_p)t + f_1x_1^2 + f_2x_2^2 + \dots + f_px_p^2.$

Or $f_1 + f_2 + \dots + f_p = 1$, $f_1x_1 + f_2x_2 + \dots + f_px_p = \overline{x}$

et $f_1x_1^2 + f_2x_2^2 + \dots + f_px_p^2 - \overline{x}^2 = V$

soit $f_1x_1^2 + f_2x_2^2 + \dots + f_px_p^2 = V + \overline{x}^2$;

donc $g(t) = t^2 - 2\overline{x}t + V + \overline{x}^2$.

2. a) L'expression de g est du type $g(t) = At^2 + Bt + C$ avec $A = 1$, $B = -2\overline{x}$ et $C = V + \overline{x}^2$.

g admet un extremum pour $t = -\dfrac{B}{2A} = \overline{x}$ (valeur dans I).

b) Comme $A > 0$, cet extremum est un minimum de valeur :
$g(\overline{x}) = \overline{x}^2 - 2\overline{x}^2 + V + \overline{x}^2 = V$.
Ainsi, V est le minimum de g, obtenu pour $t = \overline{x}$.

20 **Narration de recherche**

On a la relation $T = at + b$ avec T température en °F et t température en °C.

Les coefficients a et b sont tels que :

$\begin{cases} b = 32 \\ 100a + b = 212 \end{cases}$ soit $\begin{cases} a = 1{,}8 \\ b = 32 \end{cases}$. D'où $T = 1{,}8t + 32$.

• ***Deux pistes :***

– ***Piste 1 :*** écrire les deux séries avec la même unité (par exemple en °C) puis déterminer les moyennes et les écarts-types dans cette unité. C'est la méthode la plus évidente mais la plus fastidieuse.

– ***Piste 2 :*** calculer les indicateurs des deux séries dans leurs unités respectives puis trouver la correspondance pour obtenir, par exemple, les indicateurs de Londres en °C. Il faut donc conjecturer puis utiliser les formules : $\overline{T} = 1{,}8\overline{t} + 32$ et $s_T = 1{,}8s_t$.

(Pour les formules : voir l'exercice 52, page 288.)

• On obtient :

– pour Londres, $\overline{T} \approx 70{,}9$ °F et $s_T \approx 3{,}74$ °F

d'où $\overline{t} \approx 21{,}6$ °C et $s_t \approx 2{,}1$ °C.

– pour Stuttgart, $\overline{t} \approx 21{,}7$ °C et $s_t \approx 4{,}1$ °C.

Conclusion : les deux villes ont des températures moyennes du même ordre mais la dispersion autour de la moyenne est plus faible à Londres qu'à Stuttgart. C'est à Londres que le climat est le plus tempéré.

21 Narration de recherche

On appelle x_i la note au devoir n° i.

• Sur les huit premières notes, la moyenne est 9 et la variance 3^2 donc :

$$\sum_{i=1}^{8} x_i = 72 \,;\ \frac{\sum_{i=1}^{8} x_i^2}{8} - 9^2 = 3^2 \text{ d'où } \sum_{i=1}^{8} x_i^2 = 720.$$

• Sur les dix notes, les objectifs sont $\overline{x} = 10$ et $V \leqslant 12$.

Ces conditions se traduisent par :

$$\sum_{i=1}^{10} x_i = 100 \text{ d'où } x_9 + x_{10} = 28 \qquad \textbf{(1)} \,;$$

$$\frac{\sum_{i=1}^{10} x_i^2}{10} - 10^2 \leqslant 12 \text{ soit } \frac{720 + x_9^2 + x_{10}^2}{10} - 100 \leqslant 12$$

d'où $x_9^2 + x_{10}^2 \leqslant 400$ **(2)**.

• ***Deux pistes :***

– ***Piste 1 :*** traitement algébrique

On pose $x_9 = t$ avec $0 \leqslant t \leqslant 20$.

D'après **(1)**, $x_{10} = 28 - t$ donc la condition **(2)** équivaut à $t^2 + (28 - t)^2 \leqslant 400$ soit $t^2 - 28t + 192 \leqslant 0$.

Les racines de ce trinôme du 2nd degré sont 12 et 16 ;

il est négatif entre ses racines donc $12 \leqslant t \leqslant 16$.

Ainsi, les seules possibilités pour le couple d'entiers $(x_9 \,;\, x_{10})$ sont : (12 ; 16) ; (13 ; 15) ; (14 ; 14) ; (15 ; 13) ; (16 ; 12).

On vérifie alors que ces cinq couples sont effectivement solutions du problème.

– ***Piste 2 :*** traitement avec un tableur (CALC)

Dans un tableau à double entrée où les valeurs sont des entiers variant de 0 à 20, on fait afficher les couples solutions lorsque les conditions **(1)** et **(2)** sont vérifiées simultanément.

– Saisir dans la plage A2:A22 les entiers (x_9) de 0 à 20 ;

– Saisir dans la plage B1:V1 les entiers (x_{10}) de 0 à 20 ;

– Compléter la cellule B2 par l'instruction :

=SI($A2+B$1=28 ET ($A2)^2+(B$1)^2<=400;"sol";"")

Remarque : si les conditions (1) et (2) sont vérifiées, on affiche « sol » sinon on n'affiche rien.

– Recopier cette formule vers la droite puis vers le bas dans la plage de cellules B2:V20.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	0																					
3	1																					
4	2																					
5	3																					
6	4																					
7	5																					
8	6																					
9	7																					
10	8																					
11	9																					
12	10																					
13	11																					
14	12																	sol				
15	13																sol					
16	14															sol						
17	15														sol							
18	16													sol								
19	17																					
20	18																					
21	19																					
22	20																					

22 TP – Étudier une série de notes

Voir page 280.

23 TP – Étudier une répartition de salaires

1. A. a) Salaire moyen dans l'entreprise A : $S_A = 2\,010$ €.

b) Dans l'entreprise B, $S_B = \dfrac{M_B}{60}$ d'où $M_B = 60S_B$.

Comparer S_B et S_A équivaut à comparer $60S_B$ et $60S_A$ soit M_B et $60S_A$.

B. a) Feuille de calcul.

Saisir en B2 : $\boxed{\text{=60-A2}}$;

en C2 : $\boxed{\text{=A2*2940+B2*1650}}$;

en D2 : $\boxed{\text{=C2-60*2010}}$.

b) Vue d'écran

	A	B	C	D
1	Entreprise B Cadres	Entreprise B Ouvriers	Masse salariale M_B	Comparaison : $M_B - 60S_A$
2	1	59	100 290	-20 310
3	2	58	101 580	-19 020
4	3	57	102 870	-17 730
5	4	56	104 160	-16 440
6	5	55	105 450	-15 150
7	6	54	106 740	-13 860
8	7	53	108 030	-12 570
9	8	52	109 320	-11 280
10	9	51	110 610	-9 990
11	10	50	111 900	-8 700
12	11	49	113 190	-7 410
13	12	48	114 480	-6 120
14	13	47	115 770	-4 830
15	14	46	117 060	-3 540
16	15	45	118 350	-2 250
17	16	44	119 640	-960
18	**17**	**43**	**120 930**	**330**
19	18	42	122 220	1 620
20	19	41	123 510	2 910
21	20	40	124 800	4 200
22	21	39	126 090	5 490
23	22	38	127 380	6 780
24	23	37	128 670	8 070
25	24	36	129 960	9 360

Dès que l'entreprise B emploie au moins 17 cadres, le salaire moyen S_B est supérieur à S_A.

2. a) $S_B > S_A \Leftrightarrow \dfrac{2\,940x + (60 - x)\,1\,650}{60} > 2\,010$.

b) $S_B > S_A \Leftrightarrow x > \dfrac{21\,600}{1\,290}$.

Or x est un entier inférieur à 60, donc $17 \leqslant x < 60$.

$S_B > S_A$ signifie que l'entreprise B emploie au moins 17 cadres.

c) La proportion minimale de cadres chez B telle que $S_B > S_A$ est alors $p_0 = \dfrac{17}{60}$ soit $p_0 \approx 28{,}3\,\%$.

EXERCICES Entraînement (page 283)

DE TÊTE

24 **1.** Série des précipitations en mm :
$Q_1 = x_4 = 0$; Me $= x_8 = 0{,}2$; $Q_3 = x_{12} = 2{,}4$.
2. Série des températures maximales en °C :
$Q_1 = 19$; Me $= 20$; $Q_3 = 21$.
3. $\overline{P} \approx 3$ mm.
4. $\overline{T} \approx 20$ °C.

25 **1.** Effectif : $n = \dfrac{\sum x}{\overline{x}} = 12$.
2. Moyenne des carrés : $\dfrac{\sum x^2}{n} = \dfrac{1\,560}{12} = 130$.
3. Écart-type :
$s = \sqrt{\text{moyenne des carrés} - \text{carré de la moyenne}}$
$s = \sqrt{130 - 121} = 3$.

UTILISER LA NOTATION Σ

26 **a)** $1 + 3 + 5 + 7 + 9 + 11 + 13 + 15 + 17 = 81$.
Remarque : on peut utiliser la méthode de calcul de la somme de n termes consécutifs d'une suite arithmétique de raison 2 (voir chapitre 5).
b) $1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10 = 55$.
Remarque : $1 + 2 + 3 + \ldots + n = \dfrac{n(n+1)}{2}$ (voir chapitre 5).
c) $1 + \dfrac{1}{2} + \left(\dfrac{1}{2}\right)^2 + \left(\dfrac{1}{2}\right)^3 + \left(\dfrac{1}{2}\right)^4 + \left(\dfrac{1}{2}\right)^5 + \left(\dfrac{1}{2}\right)^6 = \dfrac{127}{64}$.
Remarque : $1 + q + q^2 + \ldots + q^n = \dfrac{1 - q^{n+1}}{1 - q}$ $(q \neq 1)$ (voir chapitre 5).

27 **a)** $S = \sum_{i=0}^{p} 5^i$. **b)** $T = \sum_{j=1}^{n} j^2$. **c)** $U = \sum_{k=1}^{p} 10^{-k}$.

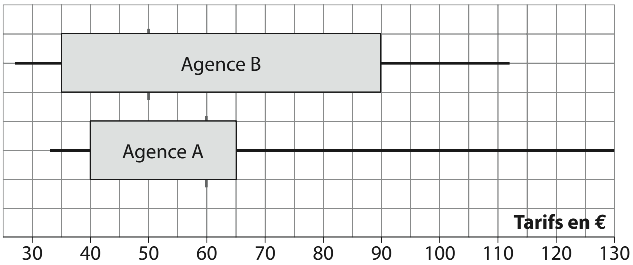
28 **a)** $S = \sum_{k=2}^{25} \sqrt{k}$. **b)** $T = \sum_{j=0}^{125} (2j + 1)$. **c)** $U = \sum_{i=1}^{10} 2^i$.

29 *Corrigé dans le manuel.*

PARAMÈTRES D'UNE SÉRIE

30 **1.** Diagramme en boîtes superposées :

	min	Q_1	Me	Q_3	max
Agence A	33	40	60	65	130
Agence B	27	35	50	90	112



2. a) Agence B. **b)** Agence B.
c) Agence A. **d)** Agence A.
3. La dispersion de 50 % des tarifs autour de la médiane est plus faible pour A que pour B.
Pour les tarifs hors de la zone centrale (pattes), les tarifs de A sont plus dispersés que ceux de B.

31 **1.** Pour les entreprises de 11 à 20 salariés, le chiffre d'affaires est $1{,}125 \times 16\,000 = 18\,000$ M€.
Elles réalisent 14 % du CAT d'où $0{,}14 \times \text{CAT} = 18\,000$.
Donc CAT $\approx 128\,571$ M€.
2. Les entreprises de plus de 20 salariés réalisent 38 % du CAT donc leur chiffre d'affaires moyen est :
$$\frac{0{,}38 \times 128\,571}{9\,000} \approx 5{,}429 \text{ M€}.$$
3. Les entreprises de 0 à 10 salariés réalisent, par complément, 48 % du CAT donc leur chiffre d'affaires moyen est :
$$\frac{0{,}48 \times 128\,571}{346\,000} \approx 0{,}178 \text{ M€}.$$
4. Tableau des chiffres d'affaires moyens :

Nombre de salariés	0 à 10	11 à 20	> 20
CA moyen en M€	0,178	1,125	5,429

La comparaison des chiffres d'affaires moyens peut se faire en évaluant les rapports suivants :
$$\frac{1{,}125}{0{,}178} \approx 6{,}3 \text{ et } \frac{5{,}429}{0{,}178} \approx 30{,}5.$$
On peut donc estimer que les chiffres d'affaires moyens des trois catégories sont dans une échelle : 1 – 6 – 30.

32 **1.** Taux moyen : $\overline{t} \approx 49{,}8$; écart-type : $s \approx 1{,}8$.
2. La 1re condition est vérifiée : $50 \in [49{,}5\,;\,50{,}1]$.
La 2e condition consiste à étudier le pourcentage de fromages dans l'intervalle $J = [46{,}2\,;\,53{,}4]$. Or dans l'intervalle $[47\,;\,53]$ contenu dans J, il y a déjà 91 % des fromages analysés donc la condition 2 est vérifiée.
L'appellation « 50 % de matière grasse » est acceptée.

33 *Corrigé dans le manuel.*

34 **1. a) b)** Feuille de calcul

2. a) Vue d'écran des résultats :

	A	B	C	D	E	F	G	H	I	J
1	Population		Critère	Q_1	Me	Q_3	$Q_3 - Q_1$		Moyenne	Ecart-type
2	16									
3	18		Totalité	236	464	1 128	892		1 659	4 507
4	19									
5	32		P < 50 000	236	463	1 126	890		1 575	3 756
6	33									
7	36		50 < P < 30 000	241	466	1 125	885		1 422	2 759
8	38									
9	42		50 < P < 15 000	238	463	1 084	846		1 268	2 127

b) Les valeurs de la médiane et des quartiles sont peu modifiées. Ces paramètres sont peu sensibles à la suppression de quelques valeurs extrêmes.

Il n'en est pas de même des valeurs de la moyenne et de l'écart-type qui sont ici influencées par la suppression des valeurs les plus grandes.

c) La population médiane pour les communes du Nord est d'environ 460 habitants.

Environ un quart des communes ont moins de 240 habitants et un quart plus de 890 habitants. Environ la moitié des communes ont entre 240 et 890 habitants.

d) Le couple (moyenne ; écart-type) est inadapté à la situation.

A contrario, le couple (médiane ; écart interquartile) donne une meilleure information sur la répartition des communes (voir question **c**)).

35 **1.** Formule de la variance utilisée :

Variance = moyenne des carrés – carré de la moyenne.

$$V = \frac{\sum_{i=1}^{N} x_i^2}{N} - \bar{x}^2.$$

2. a) Tableau des différentes étapes :

Variables	Init.	E_1	E_2	E_3	E_4	E_5
i	1	1	2	3	4	5
$x[i]$		12	16	7	11	9
Somme	0	12	28	35	46	55
Somme carrés	0	144	400	449	570	651

b) $V = \frac{651}{5} - \left(\frac{55}{5}\right)^2 = 9{,}2.$

3. Avec une calculatrice :

$s \approx 3{,}03315017$ d'où $V = s^2 \approx 9{,}2$.

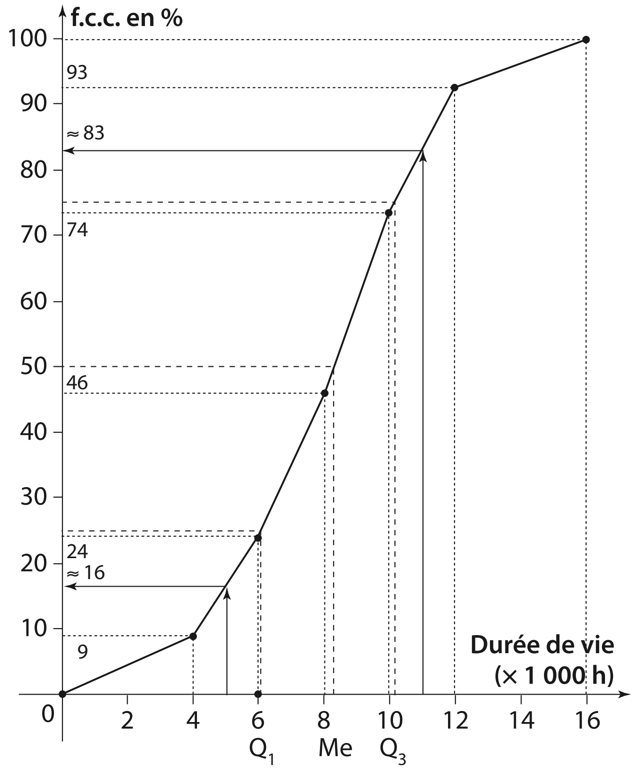
Remarque : pour obtenir la valeur exacte de V à partir des résultats de la calculatrice : $V = \frac{651}{5} - 11^2 = 9{,}2$.

D'où la validation du programme proposé.

CHOISIR UN RESUMÉ ADAPTÉ

36 *Corrigé dans le manuel.*

37 **1.** Courbe des f.c.c.



2. a) Graphiquement, en milliers d'heures :

$Q_1 \approx 6$, Me $\approx 8{,}3$, $Q_3 \approx 10{,}1$.

b) Tableau des fréquences :

d	[0 ; 4[	[4 ; 6[	[6 ; 8[	[8 ; 10[	[10 ; 12[	[12 ; 16[
Fréq.	0,09	0,15	0,22	0,28	0,19	0,07

D'où, en milliers d'heures : $\bar{d} \approx 8$; $s \approx 3$.

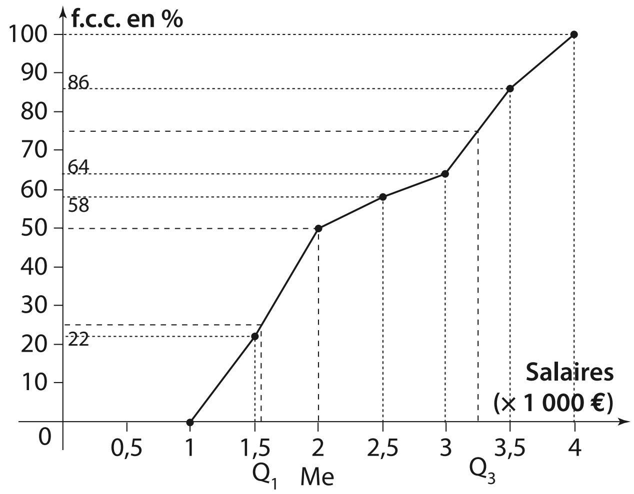
c) L'intervalle $[\bar{x} - s\,;\bar{x} + s]$ correspond à [5 ; 11].

Graphiquement, 5 a pour image une f.c.c. d'environ 16 %, et de même, 11 a pour image environ 83 %.

Ainsi le pourcentage d'ampoules ayant une durée de vie comprise entre 5 000 et 11 000 heures peut être estimé à 83 % – 16 % = 67 %.

On retrouve ainsi l'affirmation du fabricant.

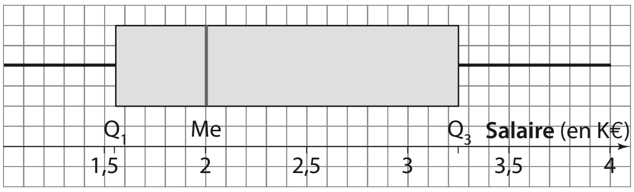
38 **1.** Courbe des f.c.c.



Graphiquement, en milliers d'euros :

$Q_1 \approx 1{,}55$, Me ≈ 2, $Q_3 \approx 3{,}25$.

D’où le diagramme en boîte :



2. Salaire moyen et écart-type, en milliers d’euros :

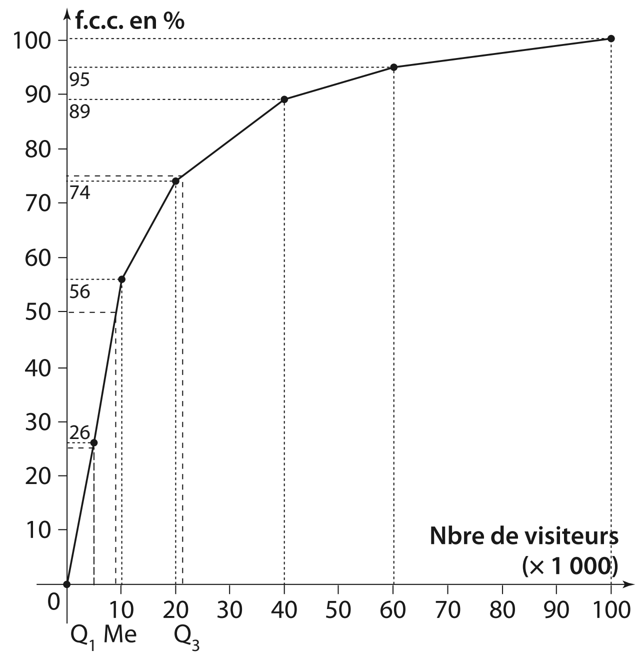
$$\overline{x} \approx 2{,}35 \,;\; s \approx 0{,}91.$$

3. Pour résumer cette série, la médiane et l’intervalle interquartile semblent bien adaptés du point de vue d’un salarié : il peut se situer sur l’échelle des salaires.

La moitié des salariés reçoivent 2 000 € ou plus et environ 50 % des salaires sont compris entre 1 550 € et 3 250 €.

39 **1.** Tableau et courbe des f.c.c.

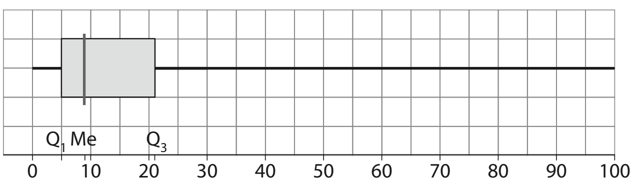
***v* : nombre de visiteurs**	< 5	< 10	< 20	< 40	< 60	< 100
f.c.c.	0,26	0,56	0,74	0,89	0,95	1



Graphiquement, en milliers de visiteurs :

$$Q_1 \approx 5,\; Me \approx 9,\; Q_3 \approx 21.$$

D’où le diagramme en boîte :



3. À partir de l’histogramme, on obtient : $\overline{x} \approx 17{,}3$, $s \approx 19{,}6$.

4. Le couple (médiane ; écart interquartile) est le plus pertinent pour traduire la répartition de ces musées suivant le nombre de visiteurs en 2008.

La moitié des petits musées ont accueilli moins de 9 000 visiteurs.

Environ 50 % des petits musées ont reçu entre 5 000 et 21 000 visiteurs.

Seulement un quart des petits musées ont accueilli plus 21 000 visiteurs.

COMPARAISON DE SÉRIES

40 **1. a)** Tableau des fréquences :

***d* (× 10³ km)**	[60 ; 70[	[70 ; 80[	[80 ; 90[	[90 ; 100[	[100 ; 110[
Fréquence (en %)	4	6	10	14	18
***d* (× 10³ km)**	[110 ; 120[	[120 ; 130[	[130 ; 140[	[140 ; 150[	[150 ; 160[
Fréquence (en %)	26	10	8	3	1

b) $\overline{d} \approx 107$; $s \approx 19$ (en milliers de km).

2. $\overline{d'} \approx 115{,}3$; $s' \approx 18{,}7$ (en milliers de miles).

3. a) $\overline{d'} \approx 186$; $s' \approx 30$ (en milliers de km).

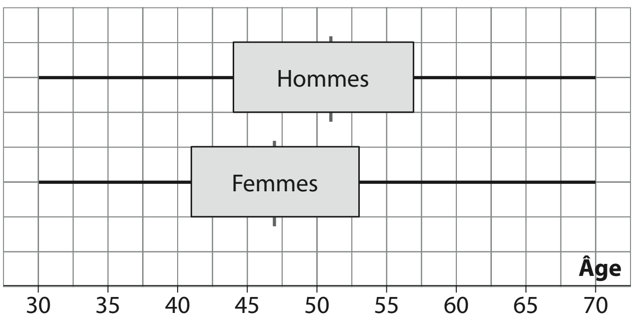
b) On constate que $\overline{d} < \overline{d'}$ et $s < s'$.

On en déduit que, à Londres, les taxis roulent davantage avant que la compagnie les renouvelle. De plus, la dispersion par rapport à la moyenne est également plus élevée dans la compagnie londonienne. On peut donc penser que le parc des taxis londoniens est plus ancien.

41 **1. a)** Estimation graphique des indicateurs en années :

	Q_1	Me	Q_3
Médecins H	44	51	57
Médecins F	41	47	53

b) Diagramme en boîtes superposées :



2. a) Tableau des fréquences suivant les tranches d’âge :

Âge	[30 ; 35[	[35 ; 40[	[40 ; 45[	[45 ; 50[
Médecins H	5 %	8 %	13 %	20 %
Médecins F	10 %	13 %	17 %	21 %
Âge	[50 ; 55[	[55 ; 60[	[60 ; 65[	[65 ; 70[
Médecins H	22 %	22 %	8 %	2 %
Médecins F	20 %	14 %	4 %	1 %

Estimation de l’âge moyen et de l’écart-type :

$$\overline{a}_H \approx 50 \,;\; s_H \approx 8{,}5 \,;\; \overline{a}_F \approx 47 \,;\; s_F \approx 8{,}5.$$

b) Graphiquement, on détermine le pourcentage de médecins dans l’intervalle indiqué.

Hommes : I = [41,5 ; 58,5] ; $p_H \approx 66$ %.

Femmes : J = [38,5 ; 55,5] ; $p_F \approx 63$ %.

3. On constate que :

– pour chaque catégorie, les âges médian et moyen sont proches ;

– la dispersion autour de la médiane mesurée par l’écart interquartile est du même ordre pour les deux catégories ;

– la dispersion autour de la moyenne mesurée par l'écart-type est aussi du même ordre.
Ainsi les répartitions des médecins Hommes et Femmes sont comparables, avec un rajeunissement de la population féminine d'environ 3 ans en moyenne.

Remarque : on peut évoquer ce type de répartition en parlant de distribution « normale » des fréquences. Ce thème sera abordé en classe de terminale.

42 **1. a)** Justification des calculs :

• films français :

$p_1 = \dfrac{41}{41 + 67 + 8} \approx 35\,\%$; de même $p_{18} \approx 37\,\%$.

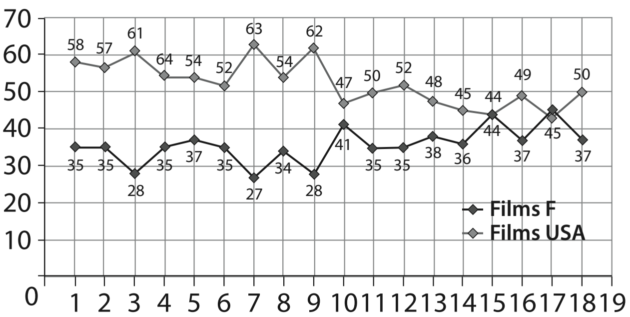
• films américains :

$p'_1 = \dfrac{67}{41 + 67 + 8} \approx 58\,\%$; de même $p'_{18} \approx 50\,\%$.

b) Tableau des parts respectives en % :

Année	1	2	3	4	5	6	7	8	9
France	35	35	28	35	37	35	27	34	28
États-Unis	58	57	61	54	54	52	63	54	62
Année	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
France	41	35	35	38	36	44	37	45	37
États-Unis	47	50	52	48	45	44	49	43	50

2. Représentation graphique



3. La part des films américains est en général supérieure à la part des films français. Cependant, la tendance est à un rapprochement de ces parts (baisse : États-Unis ; hausse : France) et donc à la diminution des écarts constatés.

Remarque : le problème d'ajustement affine est ici posé.

43 **a)** Faux.

b) Faux.

c) Vrai.

d) Vrai.

Remarque : pour les réponses **a)** et **b)**, mettre en évidence des contre-exemples.

POUR ALLER PLUS LOIN

44 **1.** Sur l'année 2007, 10 % des revenus annuels salariaux des hommes (respectivement des femmes) ne dépassent pas 2 872 € (respectivement 1 770 €).
De même, 90 % des revenus des hommes (respectivement des femmes) ne dépassent pas 37 259 € (respectivement 28 236 €).

2. a) Pour les hommes, le 1[er] quartile correspond à un revenu mensuel de $\dfrac{11\,340}{12} = 945$ €. Donc la proposition est vraie.

b) Pour les femmes, le 3[e] quartile correspond à un revenu mensuel de $\dfrac{21\,340}{12} \approx 1\,805$ €. Donc la proposition est vraie.

3. a) Tableau des écarts salariaux relatifs E en % :

	D_1	Q_1	Me	Q_3	D_9
Écart E	38	33	18	16	24

b) Ce tableau indique que l'écart relatif E, très important pour les bas salaires associés à D_1, a tendance à diminuer jusqu'au niveau des salaires associés à Q_3 puis à augmenter pour les hauts salaires associés à D_9.
Pour synthétiser, les écarts relatifs sont très importants dans les classes extrêmes et sont atténués pour les classes de la zone centrale.

Remarque : Une étude plus fine sur la totalité des déciles voire sur les centiles est cependant nécessaire pour préciser la tendance observée.

45 **1.** Tableau des indicateurs :

	St-Denis	Paris	Lille	Le Havre
D_1	6	11	8	8
Q_1	15	24	20	19
Me	33	40	36	38
Q_3	51	56	52	57
D_9	67	79	69	81

• St-Denis (Réunion) : environ 10 % de la population a moins de 6 ans et 10 % plus de 67 ans. La moitié de la population a moins de 33 ans.

• Paris : environ 10 % de la population a moins de 11 ans et 10 % plus de 79 ans. La moitié de la population a plus de 40 ans.

• Lille : environ 25 % de la population a moins de 20 ans, 50 % a entre 20 et 52 ans, et 10 % plus de 69 ans.

• Le Havre : environ 25 % de la population a moins de 19 ans, 50 % a entre 19 et 57 ans, et 10 % plus de 81 ans.

2. On peut noter que :

• St-Denis possède la population la plus jeune alors qu'à Paris la population des jeunes est sous-représentée ;

• Paris et Le Havre possèdent la plus forte proportion de personnes âgées de plus de 80 ans alors que cette catégorie est sous-représentée à St-Denis et à Lille ;

• Le Havre et St-Denis sont les villes où la dispersion est la plus importante ; Paris et Lille possèdent la plus faible dispersion autour de l'âge médian.

46 **1.** $\overline{y} = \dfrac{y_1 + y_2 + \ldots + y_N}{N} = \dfrac{ax_1 + ax_2 + \ldots + ax_N}{N}$

$\overline{y} = a\left(\dfrac{x_1 + x_2 + \ldots + x_N}{N}\right)$

soit : $\overline{y} = a\overline{x}$ **(1)**.

$V_y = \dfrac{(y_1 - \overline{y})^2 + (y_2 - \overline{y})^2 + \ldots + (y_N - \overline{y})^2}{N}$

$$V_y = \frac{(ax_1 - a\overline{x})^2 + (ax_2 - a\overline{x})^2 + \ldots + (ax_N - a\overline{x})^2}{N}$$

$$V_y = \frac{a^2[(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_N - \overline{x})^2]}{N}$$

soit : $V_y = a^2 V_x$ **(2)**.

2. • Prix moyen du gazole

On utilise la formule **(1)** pour le calcul en euros.

– France : 1,17 € ;

– Japon : $\frac{1}{113{,}90} \times 112{,}50 \approx 0{,}99$ € ;

– États-Unis : $0{,}72 \times 0{,}82 \approx 0{,}59$ € ;

Le gazole a le prix moyen le plus bas aux États-Unis.

• Écart-type

On sait que $V_y = a^2V_x$; or ici $a > 0$ donc $s_y = as_x$.

– France : 0,04 € ;

– Japon : $\frac{1}{113{,}90} \times 6 \approx 0{,}05$ € ;

– États-Unis : $0{,}72 \times 0{,}03 \approx 0{,}02$ €.

C'est au Japon que la dispersion des prix par rapport au prix moyen est la plus forte.

47 **1.** On traite une instruction conditionnelle : si le terme de rang i de la « liste x » est dans l'intervalle $[m - s ; m + s]$, alors la variable « compte » augmente de 1.

2. $m \approx 11{,}4$; $s \approx 4{,}6$. Test : 66,7 %.

3. a) Programme modifié :

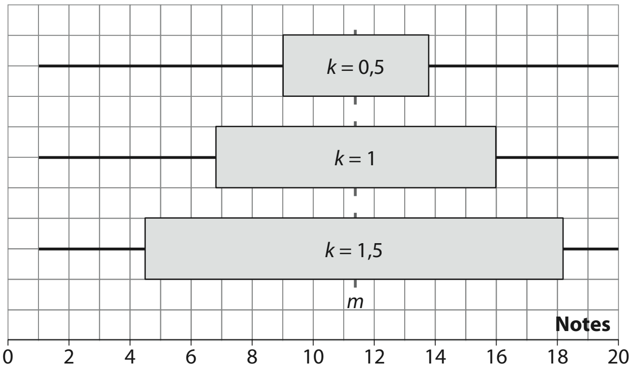
```
VARIABLES
  x EST_DU_TYPE LISTE
  i EST_DU_TYPE NOMBRE
  n EST_DU_TYPE NOMBRE
  m EST_DU_TYPE NOMBRE
  s EST_DU_TYPE NOMBRE
  k EST_DU_TYPE NOMBRE
  compte EST_DU_TYPE NOMBRE
  pourcentage EST_DU_TYPE NOMBRE
DEBUT_ALGORITHME
  i PREND_LA_VALEUR 1
  compte PREND_LA_VALEUR 0
  pourcentage PREND_LA_VALEUR 0
  AFFICHER "saisissez l'effectif total n"
  LIRE n
  AFFICHER "saisissez la moyenne m"
  LIRE m
  AFFICHER "saisissez l'écart type s"
  LIRE s
  AFFICHER "saisissez le coefficient k"
  LIRE k
  POUR i ALLANT_DE 1 A n
    DEBUT_POUR
    AFFICHER "saisissez la valeur numéro"
    AFFICHER i
    LIRE x[i]
    SI (x[i]>=m-k*s ET x[i] <=m+k*s) ALORS
      DEBUT_SI
      compte PREND_LA_VALEUR compte+1
      FIN_SI
    FIN_POUR
  pourcentage PREND_LA_VALEUR (compte/n)*100
  AFFICHER pourcentage
FIN_ALGORITHME
```

b) Tests : $k = 0{,}5 \rightarrow 37$ %

$k = 1 \rightarrow 66{,}7$ % (résultat de la question **2.**)

$k = 1{,}5 \rightarrow 85{,}2$ %

4. Diagramme en boîtes :



Les boîtes sont centrées sur la moyenne m et ont pour demi-longueurs respectives $0{,}5s$, s et $1{,}5s$; elles sont contenues les unes dans les autres.

Cela était prévisible vu que si l'on donne à k des valeurs de plus en plus grandes, les intervalles $[m - ks ; m + ks]$ sont inclus les uns dans les autres.

48 Glycémie moyenne : $\overline{G_A} \approx 1{,}11$; $\overline{G_B} \approx 1{,}36$.

Écart-type : $s_A \approx 0{,}16$; $s_B \approx 0{,}19$.

Dans le groupe A, la glycémie moyenne baisse de façon significative pour atteindre une valeur à la limite supérieure de la norme ; la dispersion des valeurs autour de la moyenne est moindre. Il n'en est pas de même pour le groupe B.

On peut donc penser que le médicament est efficace.

49 Pour tout entier i tel que $1 \leqslant i \leqslant N$, $y_i = x_i - \overline{x}$.

$$\overline{y} = \frac{y_1 + y_2 + \ldots + y_N}{N} = \frac{x_1 - \overline{x} + x_2 - \overline{x} + \ldots + x_N - \overline{x}}{N}$$

$$\overline{y} = \frac{x_1 + x_2 + \ldots + x_N}{N} - \frac{N\overline{x}}{N} = \overline{x} - \overline{x} \text{ donc } \overline{y} = 0.$$

$$V_y = \frac{(y_1 - \overline{y})^2 + (y_2 - \overline{y})^2 + \ldots + (y_N - \overline{y})^2}{N} = \frac{y_1^2 + y_2^2 + \ldots + y_N^2}{N}$$

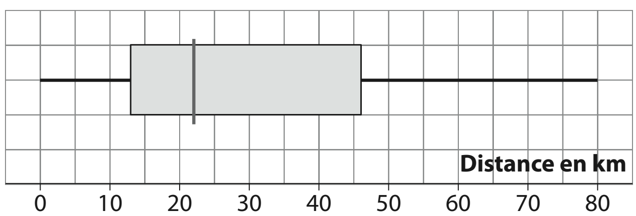
$$V_y = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_N - \overline{x})^2}{N} \text{ donc } V_y = V_x.$$

50 Graphiquement, on détermine une estimation (en km) des quartiles et de la médiane de la série :

$$Q_1 \approx 13, \text{ Me} \approx 22 \text{ et } Q_3 \approx 46.$$

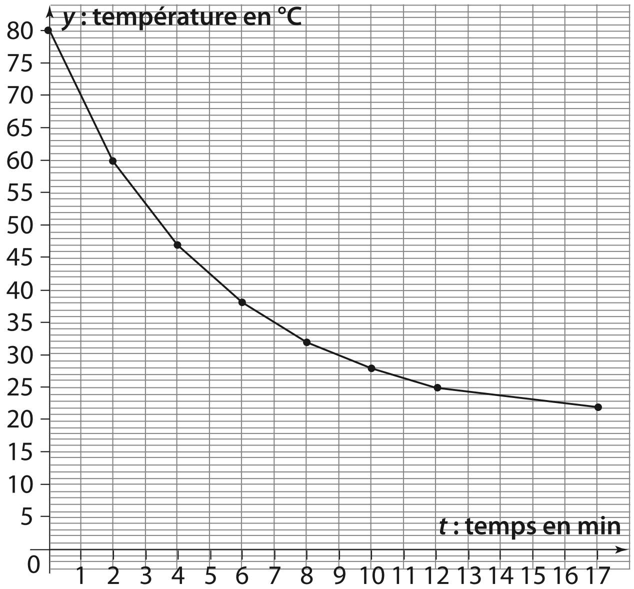
Remarque : Si N est l'effectif total, Q_1 correspond à un e.c.d. de $\frac{3N}{4}$ et Q_3 à un e.c.d. de $\frac{N}{4}$. La méthode d'interpolation linéaire (voir l'exercice 51) permet le calcul effectif de ces estimations.

Diagramme en boîte :



Interprétation : près de la moitié des salariés de cette entreprise parcourent moins de 22 km, un quart moins de 13 km et un quart plus de 46 km ; environ la moitié des salariés parcourent entre 13 et 46 km ; les salariés éloignés de plus de 50 km de leur lieu de travail représentent moins du quart de l'effectif.

51 **1.** Représentation graphique :



2. a) $f(t) = at + b$.

$a = \dfrac{y_B - y_A}{t_B - t_A}$ d'où $a = -0{,}6$; $b = y_A - at_A$ d'où $b = 32{,}2$.

Ainsi : $f(t) = -0{,}6t + 32{,}2$, $x \in [12\,;\,17]$.

b) Estimation de la température après 15 min :

$f(15) = 23{,}2$ (en °C).

3. a) La fonction affine est représentée sur l'intervalle [4 ; 6] par le segment [CD] avec C(4 ; 47) et D(6 ; 38).

On obtient : $g(t) = -4{,}5t + 65$.

b) On résout l'équation $g(t) = 40$.

$t = \dfrac{25}{4{,}5} = \dfrac{50}{9}$ (en min). D'où le temps d'attente :

5 min + $\dfrac{5}{9}$ min soit environ 5 min 33 s.

52 **1.** $\overline{x} = f_1x_1 + f_2x_2 + \ldots + f_px_p$.

$\overline{y} = f_1 \times (ax_1 + b) + f_2 \times (ax_2 + b) + \ldots + f_p \times (ax_p + b)$

$\overline{y} = af_1x_1 + f_1b + af_2x_2 + f_2b + \ldots + af_px_p + f_pb$

$\overline{y} = a(f_1x_1 + f_2x_2 + \ldots + f_px_p) + (f_1 + f_2 + \ldots + f_p)b$

Or $f_1x_1 + f_2x_2 + \ldots + f_px_p = \overline{x}$ et $f_1 + f_2 + \ldots + f_p = 1$,

donc : $\overline{y} = a\overline{x} + b$ **(1)**.

2. $V_y = f_1y_1^2 + f_2y_2^2 + \ldots + f_py_p^2 - \overline{y}^2$

$V_y = f_1 \times (ax_1 + b)^2 + f_2 \times (ax_2 + b)^2 + \ldots + f_p \times (ax_p + b)^2 - (a\overline{x} + b)^2$

$V_y = a^2f_1x_1^2 + 2abf_1x_1 + b^2f_1 + a^2f_2x_2^2 + 2abf_2x_2 + f_2b^2 + \ldots + a^2f_px_p^2 + 2abf_px_p + b^2f_p - a^2\overline{x}^2 - 2ab\overline{x} - b^2$

$V_y = a^2(f_1x_1^2 + f_2x_2^2 + \ldots + f_px_p^2 - \overline{x}^2) + 2ab(f_1x_1 + f_2x_2 + \ldots + f_px_p - \overline{x}) + b^2(f_1 + f_2 + \ldots + f_p - 1)$

Or $f_1x_1^2 + f_2x_2^2 + \ldots + f_px_p^2 - \overline{x}^2 = V_x$,

$f_1x_1 + f_2x_2 + \ldots + f_px_p - \overline{x} = 0$ et $f_1 + f_2 + \ldots + f_p - 1 = 0$,

donc : $V_y = a^2V_x$ **(2)**.

3. La relation affine entre les deux séries est définie par $y_i = \dfrac{5}{9}x_i - \dfrac{160}{9}$.

D'après **(1)** : $\overline{y} = \dfrac{5}{9}\overline{x} - \dfrac{160}{9}$ avec $\overline{x} = 67$;

donc $\overline{y} \approx 19{,}4$ (arrondi à 0,1 °C près).

D'après **(2)** : $V_y = \left(\dfrac{5}{9}\right)^2 V_x$; d'où $s_y = \dfrac{5}{9}s_x$.

Or $s_x = 5$ donc $s_y \approx 2{,}8$ (arrondi à 0,1 °C près).

53 **1. a)** On utilise la formule **(1)** de l'exercice 52.

– *Piste 1 :* $t_i = x_i + 0{,}1$. D'où $\overline{t} = \overline{x} + 0{,}1$ soit $\overline{t} = 2{,}5$.

– *Piste 2 :* $z_i = \dfrac{25}{24}x_i$. D'où $\overline{z} = \dfrac{25}{24}\overline{x}$ soit $\overline{z} = 2{,}5$.

b) On utilise la formule **(2)** de l'exercice 52 :

$V_y = a^2V_x$ d'où, lorsque $a > 0$, $s_y = as_x$.

– *Piste 1 :* $t_i = x_i + 0{,}1$. D'où $s_t = s_x$ soit $s_t = 1{,}2$.

– *Piste 2 :* $z_i = \dfrac{25}{24}x_i$. D'où $s_z = \dfrac{25}{24}s_x$ soit $s_z = 1{,}25$.

c) La piste 1 conduit à augmenter, en pourcentage, davantage les salaires les plus bas que les salaires les plus hauts. Par exemple :

Salaire (× 1 000 €)	1,2	6
Augmentation relative en %	$\dfrac{0{,}1}{1{,}2} \times 100 \approx 8{,}3$	$\dfrac{0{,}1}{6} \times 100 \approx 1{,}7$

La piste 2 conduit à augmenter d'un montant plus important les salaires les plus hauts. Par exemple :

Salaire (× 1 000 €)	1,2	6
Augmentation (× 1 000 €)	$\dfrac{1}{24} \times 1{,}2 = 0{,}05$	$\dfrac{1}{24} \times 6 = 0{,}25$

2. a) La relation affine est du type $y = ax + b$ avec $a > 0$.

Les coefficients a et b sont tels que : $\begin{cases} \overline{y} = a\overline{x} + b \\ s_y = as_x \end{cases}$.

On résout le système $\begin{cases} 2{,}4a + b = 2{,}5 \\ 1{,}2a = 1{,}1 \end{cases}$

D'où : $a = \dfrac{11}{12}$ et $b = \dfrac{3}{10}$.

Ainsi : $y = \dfrac{11}{12}x + \dfrac{3}{10}$.

b) $1{,}2 \leqslant x \leqslant 6$ et la fonction affine $g : x \mapsto \dfrac{11}{12}x + \dfrac{3}{10}$ est strictement croissante sur [1,2 ; 6] ;

donc $g(1{,}2) \leqslant y \leqslant g(6)$ soit $1{,}4 \leqslant y \leqslant 5{,}8$.

Ainsi, les salaires, exprimés en milliers d'euros, sont dans l'intervalle [1,4 ; 5,8].

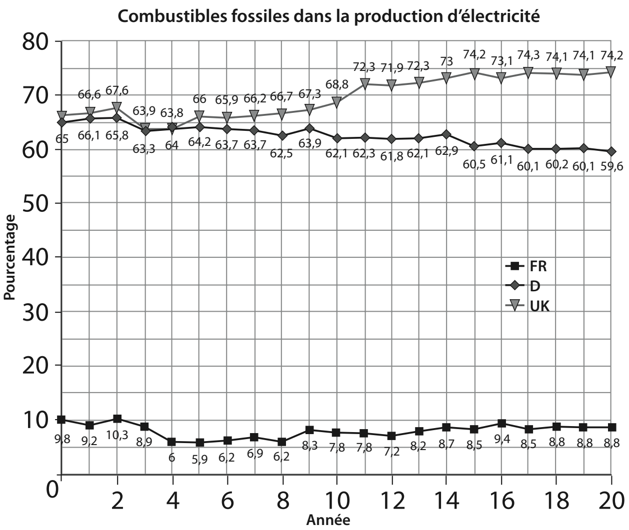
Remarque : la piste 3 conduit à une augmentation des bas salaires et une baisse des plus hauts.

3. La piste 3 rend la répartition des salaires plus homogène. Dans les trois cas, le salaire moyen est le même mais la piste 3 conduit à l'écart-type le plus faible, donc la dispersion des salaires autour de la moyenne est la plus faible.

54 **1.** Feuille de calcul

	A	B	C	D
1	Année	FR	D	UK
2	0	9,8	65	66,1
3	1	9,2	66,1	66,6
4	2	10,3	65,8	67,6
5	3	8,9	63,3	63,9
6	4	6	64	63,8
7	5	5,9	64,2	66
8	6	6,2	63,7	65,9
9	7	6,9	63,7	66,2
10	8	6,2	62,5	66,7
11	9	8,3	63,9	67,3
12	10	7,8	62,1	68,8
13	11	7,8	62,3	72,3
14	12	7,2	61,8	71,9
15	13	8,2	62,1	72,3
16	14	8,7	62,9	73
17	15	8,5	60,5	74,2
18	16	9,4	61,1	73,1
19	17	8,5	60,1	74,3
20	18	8,8	60,2	74,1
21	19	8,8	60,1	74,1
22	20	8,8	59,6	74,2
23				
24	Coefficient directeur	-0,050	-0,270	0,405
25				
26	Valeur estimée en 2015	8,500	57,980	76,630

2. a) Représentation graphique :



b) Graphiquement, la tendance d'évolution des fréquences semble être :

FR : stabilité ; D : décroissance ; UK : croissance.

3. a) On saisit en B24 : =(B$22-B$2)/20 puis on recopie vers la droite jusqu'en D24.

b) Fonctions affines associées

FR : $y = -0{,}05x + 9{,}8$;

D : $y = -0{,}027x + 65$;

UK : $y = 0{,}405x + 66{,}4$.

c) On saisit en B26 : =B$24*26+B$2 puis on recopie vers la droite jusqu'en D26.

Estimations en 2015 arrondies à 0,1 % près :

FR	D	UK
8,5 %	58 %	76,6 %

On prévoit une quasi stabilité en France avec une fréquence d'environ 8,5 %, une légère baisse en Allemagne et une légère hausse au Royaume-Uni.

55 **1.** (Q) $\Rightarrow$ (P) mais (P) $\not\Rightarrow$ (Q).

Contre-exemple : $x_1 = -2$, $x_2 = 1$, $x_3 = 1$, $x_4 = 0$.

(Il existe i tel que $x_i \neq 0$ et $\overline{x} = 0$.)

Ainsi (Q) est suffisante pour (P) mais n'est pas nécessaire.

2. • (P) $\Rightarrow$ (Q)

En effet : si V = 0, $\dfrac{\sum_{i=1}^{p}(x_i-\overline{x})^2}{n} = 0$, alors pour tout entier i tel que $1 \leqslant i \leqslant n$, $(x_i - \overline{x})^2 = 0$ d'où $x_i = \overline{x}$.

• (Q) $\Rightarrow$ (P)

En effet : si pour tout entier i tel que $1 \leqslant i \leqslant n$, $x_i = \overline{x}$, alors $V = \dfrac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n} = 0$.

Ainsi (Q) est nécessaire et suffisante pour (P).

3. • (P) $\Rightarrow$ (Q)

En effet : si Me $= \overline{x}$, alors $\sum_{i=1}^{n} x_i = n\overline{x} = n\text{Me}$.

• (Q) $\Rightarrow$ (P)

En effet : si $\sum_{i=1}^{n} x_i = n\text{Me}$, alors $n\overline{x} = n\text{Me}$; or $n \neq 0$, donc Me $= \overline{x}$.

Ainsi (Q) est nécessaire et suffisante pour (P).

56 On désigne par $\overline{x}$ et s respectivement la moyenne et l'écart-type du groupe complet.

• $\overline{x} = \dfrac{\text{total des notes}}{\text{effectif}}$ donc $\overline{x} = \dfrac{15 \times 11 + 19}{16} = 11{,}5$.

• La variance $V = s^2$ du groupe complet est telle que :

s^2 = moyenne des carrés – carré de la moyenne.

La variance de la série des notes des 15 élèves présents est

$v = 4^2 = \dfrac{\sum_{i=1}^{15} x_i^2}{15} - 11^2$ d'où $\sum_{i=1}^{15} x_i^2 = 2\,055$.

Ainsi la somme des carrés des notes du groupe complet est $2\,055 + 19^2 = 2\,416$.

On en déduit que $s^2 = \dfrac{2\,416}{16} - 11{,}5^2 = 18{,}75$ d'où

$s = \sqrt{18{,}75} = \sqrt{\dfrac{75}{4}} = \dfrac{5\sqrt{3}}{2} \approx 4{,}33$.

57 On note $\overline{x_1}$, $\overline{x'_2}$ et $\overline{x}$ les moyennes respectives du groupe 1, du groupe 2 et de la classe ; de même on désigne par s_1, s'_2 et s les écarts-types correspondants.

• Par lecture graphique : $\overline{x_1} = 11$, $\overline{x'_2} = 13$.

Or $\overline{x} = \dfrac{15\overline{x_1} + 15\overline{x'_2}}{30} = \dfrac{\overline{x_1} + \overline{x'_2}}{2}$, donc $\overline{x} = 12$.

• Par lecture graphique : $s_1 = 3$, $s'_2 = 4$.

Vu que $s_1^2 = \dfrac{\sum_{i=1}^{15} x_i^2}{15} - \overline{x}_1^2$, on a $\sum_{i=1}^{15} x_i^2 = 15\left(\overline{x}_1^2 + s_1^2\right)$;

de même : $\sum_{j=1}^{15} x'^2_j = 15\left(\overline{x'_2}^2 + s'^2_2\right)$.

D'où la somme des carrés de toutes les valeurs :

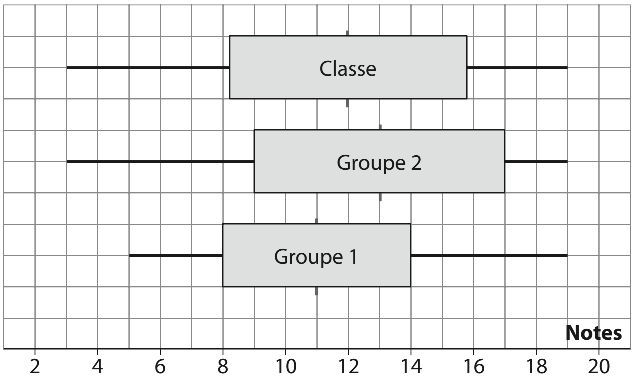
$\sum_{i=1}^{15} x_i^2 + \sum_{j=1}^{15} x'^2_j = 15(11^2 + 3^2) + 15(13^2 + 4^2) = 4\,725$.

Or s^2 = moyenne des carrés – carré de la moyenne,

donc $s^2 = \dfrac{4\,725}{30} - 12^2 = 13{,}5$.

Ainsi : $s = \sqrt{13{,}5} \approx 3{,}7$.

• D'où la construction de la boîte (type : $[\overline{x} - s\,;\, \overline{x} + s]$) associée aux résultats de la classe.



58 **1. a) b)** Feuille de calcul

• *Protocole*

Pour chaque question, on code 1 la réponse exacte.

On choisit au hasard un entier entre 1 et 4 (compris) ; si l'on obtient 1 la réponse est juste et rapporte 1 point, sinon elle est fausse et rapporte 0 point. La note obtenue est la somme des résultats aux cinq questions.

• *Saisie des instructions*

	A	B	C	D	E	F	G	H	I	J
1	QCM n°	Q1	Q2	Q3	Q4	Q5	Note		Note >3	intervalle [$\bar{x}$-s;$\bar{x}$+s]
2	1	0	1	1	0	1	3		0	0
3	2	0	0	0	0	1	1		0	1
95	94	0	0	0	1	0	1		0	1
96	95	0	0	1	0	1	2		0	1
97	96	0	0	0	0	0	0		0	0
98	97	1	0	1	1	0	3		0	0
99	98	0	0	0	0	1	1		0	1
100	99	0	0	0	0	0	0		0	0
101	100	0	0	1	0	0	1		0	1
102								Moyenne	Pourcentage de notes >3	Pourcentage de notes dans l'intervalle
103								1,28	1	67
104								Ecart type		
105								0,95		

– Saisir en B2 =SI(ALEA.ENTRE.BORNES(1;4)=1;1;0) puis copier vers la droite jusqu'en F2.

– Saisir en G2 =SOMME(B2:F2).

– Copier vers le bas la plage A2:G2 jusqu'à la ligne 101.

2. a) Pour le calcul de la moyenne et de l'écart-type :

– saisir en H103 =MOYENNE(G2:G101) ;

– saisir en H105 =ECARTYPEP(G2:G101).

Résultats sur la simulation présentée : $\bar{x} = 1{,}28$; $s = 0{,}95$.

b) Relever en colonne I les notes supérieures à 3 (si la note est supérieure à 3, on code 1, sinon 0).

– Saisir en I2 =SI(G2>3;1;0) puis copier vers le bas jusqu'en I101.

– Saisir en I103 =SOMME(I2:I101).

Le pourcentage de notes supérieures à 3 est ici 1 %.

c) Relever en colonne J les notes dans l'intervalle $[\bar{x}-s\,;\bar{x}+s]$ (si la note est dans l'intervalle, on code 1, sinon 0).

– Saisir en J2

=SI((G2>=H$103-H$105)ET(G2<=H$103+H$105);1;0)

puis copier vers le bas jusqu'en J101.

– Saisir en J103 =SOMME(J2:J101)

Le pourcentage de notes dans l'intervalle est ici 67 %.

La stratégie de réponse au hasard n'est pas du tout efficace. Sur la simulation présentée, seulement 1 % des notes dépasse 3 et dans deux cas sur 3 environ, la note est dans l'intervalle $[\bar{x} - s\,;\,\bar{x} + s]$ et donc ne peut prendre que les valeurs 1 ou 2.

Remarque : l'étude de la loi binomiale (chapitre 13) permet d'apporter une réponse probabiliste à ce type de problème.

La fréquence observée doit être proche de la valeur $\frac{5}{4}$ et l'écart-type proche de $\frac{\sqrt{15}}{4}$.

EXERCICES Travail en autonomie (page 290)

A **1.** Tableau des indicateurs

	Saint-Étienne	Nice	Brest
mini	2	8	5
Q_1	4	9	7
Me	10	15	10
Q_3	16	20	14
max	19	23	16

2. Voici quelques éléments de comparaison :

• Saint-Étienne présente la plus grande étendue des températures et la plus grande dispersion autour de la médiane.

• Brest présente le relevé de températures le plus homogène : étendue des températures la plus petite et une faible dispersion autour de la médiane (identique à celle de Saint-Étienne).

• Nice est la ville la plus chaude : plus du quart des températures relevées sont au-dessus des températures maximales des deux autres villes.

Le quart des températures les plus basses est contenu dans l'intervalle [Q_1 ; Me] de Brest ou Saint-Étienne.

La médiane est presque au niveau des troisièmes quartiles de Brest et Saint-Étienne.

B **1.** Fréquences des avis positifs pour les filles :

$f_A = \frac{102}{408} = 0{,}250$; $f_B = \frac{89}{312} \approx 0{,}285$.

Le jeu B est le plus apprécié des filles.

2. Pour les garçons :

$f'_A = \frac{467}{776} \approx 0{,}602$; $f'_B = \frac{67}{104} \approx 0{,}644$.

Le jeu B est le plus apprécié des garçons.

3. a) Tableau des résultats sur l'échantillon complet :

	Avis positif	Avis négatif	Total
Jeu A	569	615	1 184
Jeu B	156	260	416
Total	725	875	1 600

$F_A = \dfrac{569}{1\,184} \approx 0{,}481$; $F_B = \dfrac{156}{416} = 0{,}375$.

Pour l'échantillon complet, le jeu A est le plus apprécié.

b) C'est un effet de structure !

– Une grosse proportion de garçons (88 %) a testé le jeu A.

– Chez les garçons, la proportion des avis positifs sur chacun des jeux est bien plus élevée que chez les filles.

– On teste beaucoup plus le jeu A que le jeu B, et la proportion de garçons est plus importante que celle des filles, donc la répartition des avis positifs est influencée par la structure de l'échantillon complet.

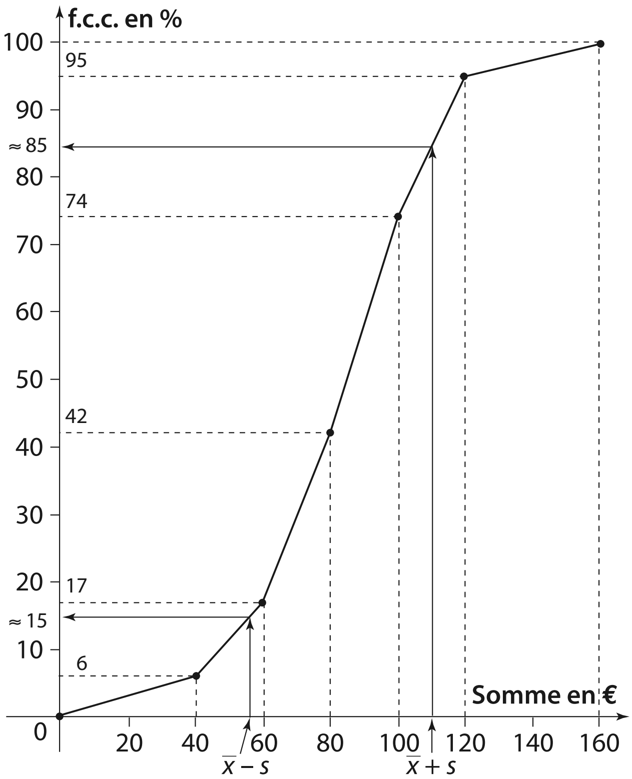
C **1. a)** Estimation de la somme moyenne reçue : $\bar{x} \approx 83$ €.

b) Estimation de l'écart-type : $s \approx 27$ €.

2. L'intervalle $[\bar{x} - s\,;\, \bar{x} + s]$ est assimilé à I = [56 ; 110].

Une lecture graphique de la courbe des f.c.c. permet d'obtenir le pourcentage de valeurs dans I.

$p \approx 85\,\% - 15\,\%$ soit $p \approx 70\,\%$.



Remarque : la méthode d'interpolation linéaire vue dans l'exercice 51 permet le calcul des f.c.c. images de $\bar{x} - s$ et $\bar{x} + s$.

– Sur [40 ; 60] : $f(x) = \dfrac{11}{20}x - 16$ et $f(56) = 14{,}8$.

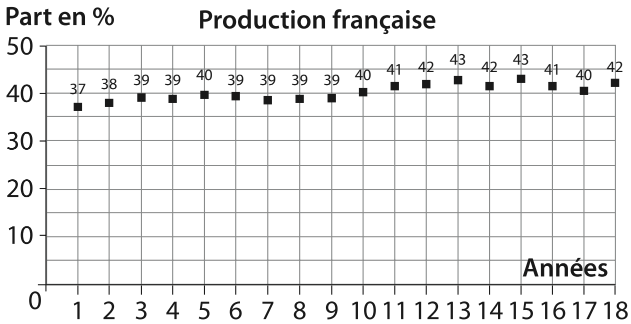
– Sur [100 ; 120] : $g(x) = \dfrac{21}{20}x - 31$ et $g(110) = 84{,}5$.

Les valeurs trouvées correspondent à celles obtenues graphiquement.

D **1.** Tableau des parts des productions respectives :

Année	1	2	3	4	5	6	7	8	9
France	37	38	39	39	40	39	39	39	39
États-Unis	30	29	29	30	31	30	33	33	32
Année	10	11	12	13	14	15	16	17	18
France	40	41	42	43	42	43	41	40	42
États-Unis	31	30	29	29	29	28	27	29	29

2. a) Représentation graphique de la part de la production française.



b) La part de la production américaine semble se stabiliser autour de 30 %. La tendance semble indiquer une légère croissance de la part de la production française.

3. La tendance peut être modélisée par la droite qui passe par les points A(1 ; 37) et B(18 ; 42).

La droite (AB) représente la fonction affine définie par :

$f(x) = \dfrac{5}{17}x + \dfrac{624}{17}$.

Remarque : avec ce modèle, la croissance de la production française est mise en évidence.

On résout l'inéquation : $f(x) \geqslant 45$.

$\dfrac{5}{17}x + \dfrac{624}{17} \geqslant 45 \Leftrightarrow x \geqslant \dfrac{141}{5}$.

Ainsi, à partir de l'année de rang 29 c'est-à-dire à partir de 2020, on peut estimer que la part de la production française atteindra 45 %.