

Objectif

Introduire un nouveau diagramme statistique : le diagramme en boîte.

Note

Ce graphique est appelé un diagramme en boîte.

Objectif

Introduire un nouvel indicateur statistique pour rendre compte des différences entre deux séries.

1

Espérance de vie à la naissance

Les tableaux ci-dessous indiquent l'espérance de vie à la naissance des hommes et des femmes de divers pays du monde (source : OMS, 2003).

Femmes

Japon	Canada	Australie	France	Royaume-Uni	États-Unis	Chine	Brésil	Égypte	Russie	Inde
84,7	83,8	83	84,8	80,5	79,8	72,7	72,3	69	72,1	62

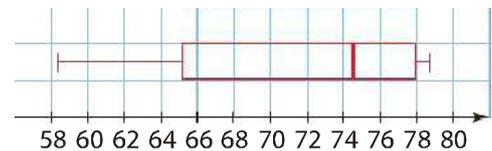
Hommes

Japon	Canada	Australie	France	Royaume-Uni	États-Unis	Chine	Brésil	Égypte	Russie	Inde
78	78,7	78,9	78,1	75,8	74,6	69,6	65,7	65,3	58,4	60,1

On se propose d'utiliser un nouveau diagramme statistique pour comparer deux séries.

- Déterminer les médianes et les quartiles de chacune des deux séries ci-dessus.
- Paul a représenté ci-contre les données trouvées pour l'une des séries.

Reproduire ce graphique et tracer un graphique analogue pour l'autre série, en choisissant la même échelle.



- Comparer les diagrammes pour commenter la différence entre les espérances de vie à la naissance des femmes et des hommes dans ces pays.



Lila et Tom sont amis. Ils ont souscrit, il y a un an, le même abonnement pour leur téléphone portable à 25 € par mois (SMS illimités). Cependant, au bout d'un an, Tom a payé 73 € de plus que Lila.

On se propose de montrer l'utilité d'un nouvel indicateur statistique.

- Le conseiller de leur opérateur s'étonne :
 - « Vous téléphonez chacun en moyenne 1 h 53 min environ par mois » ;
 - « Pour chacun de vous, la moitié de vos appels mensuels durent 1 h 50 min 30 s ou moins et l'autre moitié 1 h 50 min 30 s ou plus » ;
 - « Là encore, c'est surprenant, le quart de vos appels mensuels ne dépasse pas 1 h 45 min 30 s ; les trois quarts ne dépassent pas 1 h 59 min 30 s ».
 À quels indicateurs statistiques le conseiller fait-il allusion ?

- Le conseiller décide d'éditer le détail des durées d'appels mensuels. Elles sont résumées dans la feuille de calcul ci-contre.

- Quelle formule faut-il saisir en cellule B16 ?
- Compléter la phrase : « En cellules C3 à C14, on calcule les carrés des écarts entre ... et ... ».
- Dans la cellule C16, saisir la formule :

$$=SOMME(C3:C14)/12$$

- Compléter la cellule D16 et reproduire les mêmes calculs pour Lila.

- Expliquer pourquoi l'indicateur σ , appelé **écart-type**, permet de distinguer les deux séries. Que montre-t-il ?

	A	B	C	D	E
1		Tom			Lila
2	Mois	Durée x_i (en min)	$(x_i - \bar{x})^2$		Durée x_i (en min)
3	1	111			120
4	2	96			119
5	9	137			105
6	4	106			121
7	5	112			119
8	6	139			120
9	7	95			106
10	8	120			110
11	9	119			110
12	10	105			111
13	11	106			104
14	12	110			105
15		Moyenne \bar{x}	Variance V	$\sigma = \sqrt{V}$	Moyenne \bar{z}
16					

Médiane, quartiles et diagramme en boîte

1 Des paramètres de tendance centrale : médiane, quartiles

a) La médiane

DÉFINITION La liste des N données est **rangée par ordre croissant**.

- Si la série est de taille **impaire** ($N = 2n + 1$), la **médiane** est la donnée de rang $n + 1$.
- Si la série est de taille **paire** ($N = 2n$), la **médiane** est la demi-somme des données de rang n et de rang $n + 1$.

EXEMPLE : La longueur médiane des lancers de javelot présentés dans ce tableau est $Me = 41$ m.

Longueur (en m)	37	39	40	41	42	43	44	48
Effectif	4	3	4	3	2	3	5	2

En effet, l'effectif total est 26 ($26 = 2 \times 13$), donc la médiane est la demi-somme des 13^e et 14^e longueurs. Or ces deux longueurs sont égales à 41 m.

Note

La médiane et les quartiles ne sont pas sensibles aux valeurs extrêmes.

b) Les quartiles

DÉFINITIONS La liste des N données est **rangée par ordre croissant**.

- Le **premier quartile** est la plus petite donnée Q_1 de la liste telle qu'**au moins un quart** des données de la liste sont **inférieures ou égales à Q_1** .
- Le **troisième quartile** est la plus petite donnée Q_3 de la liste telle qu'**au moins les trois quarts** des données de la liste sont **inférieures ou égales à Q_3** .

EXEMPLE : Pour la série des lancers de l'exemple ci-dessus :

- $26 : 4 = 6,5$, donc Q_1 est la 7^e longueur, à savoir $Q_1 = 39$ m ;
- $3 \times (26 : 4) = 19,5$, donc Q_3 est la 20^e longueur, à savoir $Q_3 = 44$ m.

2 Un paramètre de dispersion : l'écart interquartile

DÉFINITION L'**écart interquartile** est la différence entre le troisième quartile et le premier quartile : $Q_3 - Q_1$.

EXEMPLE : Dans l'exemple du paragraphe 1, l'écart interquartile est $44 \text{ m} - 39 \text{ m} = 5 \text{ m}$.

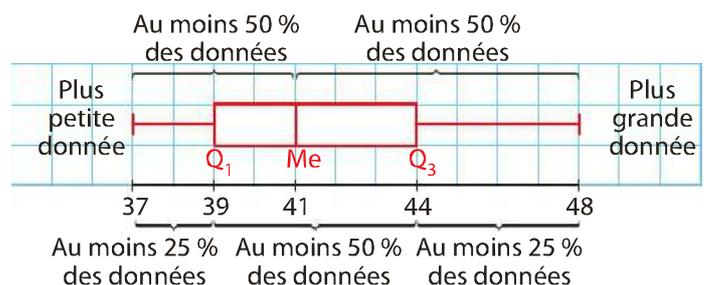
3 Diagramme en boîte

Note

Le diagramme en boîte (*box plot*) ou en boîte à moustaches (*box and whiskers plot*) est dû au statisticien américain John Tukey (1915-2000).

Pour visualiser ces paramètres et la répartition des données autour de la médiane, on utilise un **diagramme en boîte**. Voici celui de l'exemple précédent.

Ce diagramme est souvent utilisé pour comparer plusieurs séries.



Exercice résolu 1 Comparer des séries avec la calculatrice

► Voir aussi exercice 15 page 162

Énoncé

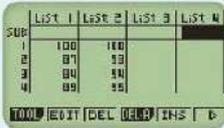
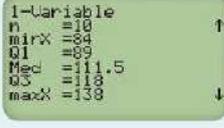
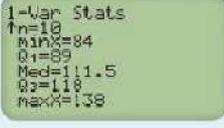
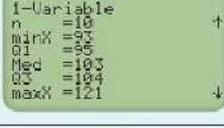
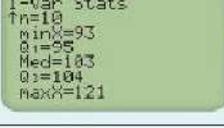
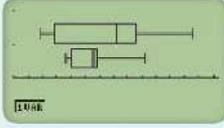
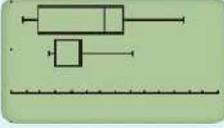
On a relevé l'évolution du cours moyen du dollar (USD) et de la livre (GBP) par rapport à l'euro (EUR) depuis son introduction en 1999.

On a retenu l'indice 100 en 1999 (source : ECB 2010).

Année	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
USD	100	87	84	89	106	117	117	118	129	138
GBP	100	93	94	95	105	103	104	103	104	121

Avec la calculatrice, afficher la médiane, les quartiles et le diagramme en boîte de chacune de ces séries.

Solution

Commentaires	Casio	TI
<ul style="list-style-type: none"> On entre les données. 	MENU 2 (STAT) 	STAT 1 (Edit) 
<ul style="list-style-type: none"> On affiche les paramètres. 	F2 (CALC) F6 (Set) 1 Var XList : List 1 1 Var Freq : 1 F1 (1 VAR) 	STAT ► (CALC) 1 (1-Var Stats) 2nd L1 
	F6 (Set) 1 Var XList : List 2 1 Var Freq : 1 F1 (1 VAR) 	STAT ► (CALC) 1 (1-Var Stats) 2nd L2 
<ul style="list-style-type: none"> On affiche les diagrammes en boîte. <p>Régler la fenêtre graphique (ici, $75 \leq X \leq 150$, pas 5 et $-1 \leq Y \leq 2$).</p>	EXIT EXIT F1 (GRPH) F6 (SET) Graph Type : Box XList : List 1 Frequency : 1 Recommencer pour Graph 2 EXIT F4 (SEL) Stat Graph 1 : Draw On Stat Graph 2 : Draw On F6 (DRAW) 	2nd STATPLOT ENTER Renseigner : On ENTER Type :  XList : L1 Fréq : 1 Recommencer avec Plot 2 GRAPH 

Conclusion : Durant cette période, l'indice médian est de 103 pour la livre et 111,5 pour le dollar. Les variations de cet indice sont plus resserrées autour de la médiane pour la livre (écart interquartile : 9 points) que pour le dollar (écart interquartile : 29 points).

Exercice résolu 1 Comparer deux séries

► Voir aussi exercice 25 page 164

Énoncé

Voici le relevé des poids nets de 30 paquets de biscuits, pris pour les uns dans une unité de fabrication industrielle et pour les autres dans une fabrique artisanale.

• Tableau 1

Masse (en g)	198	199	200	201	202
Effectif	3	2	18	5	2

• Tableau 2

Masse (en g)	198	199	200	201	202
Effectif	5	7	7	6	5

- a) Calculer, dans chacun des cas, la moyenne et l'écart-type.
b) Ces résultats permettent-ils de savoir d'où provient chaque tableau ?

Solution

a) **Tableau 1** : $\bar{x}_1 = \frac{3 \times 198 + 2 \times 199 + \dots + 2 \times 202}{30}$, soit $\bar{x} \approx 200,03$ g.

$V_1 = \frac{3 \times (198 - \bar{x})^2 + \dots + 2 \times (202 - \bar{x})^2}{30}$, soit $V_1 \approx 0,899$, donc $\sigma_1 \approx 0,95$ g.

Tableau 2 : De même, on obtient $\bar{x}_2 \approx 199,97$ g et $\sigma_2 \approx 1,33$ g.

b) Les moyennes \bar{x}_1 et \bar{x}_2 sont à peu près égales ; elles ne permettent donc pas de distinguer les deux tableaux.

En revanche, σ_2 est supérieur à σ_1 , ce qui indique que le tableau 2 provient probablement de la fabrique artisanale.

Pour le tableau 1, les robots qui fabriquent les biscuits font moins d'erreurs de dosage : σ_1 est plus proche de zéro.

Note

- La calculatrice note $x\sigma n$ (Casio) ou σx (TI) l'écart-type. Elle n'affiche pas la variance mais $V = \sigma^2$.
- Avec un tableur, on utilise la formule =ECARTYPEP(plage de données) pour obtenir σ .

Exercice résolu 2 Résumer une série statistique

► Voir aussi exercice 33 page 165

Énoncé

On a relevé la taille d'un million d'adultes de sexe masculin, de moins de 30 ans.

Taille (en cm)	[130 ; 135[[135 ; 140[[140 ; 145[[145 ; 150[[150 ; 155[[155 ; 160[
Effectif	8	63	411	2 071	7 000	29 000

Taille (en cm)	[160 ; 165[[165 ; 170[[170 ; 175[[175 ; 180[[180 ; 185[[185 ; 190[
Effectif	60 870	111 300	170 233	197 171	180 000	126 893

Taille (en cm)	[190 ; 195[[195 ; 200[[200 ; 205[[205 ; 210[[210 ; 215[
Effectif	70 504	30 662	10 436	2 779	599

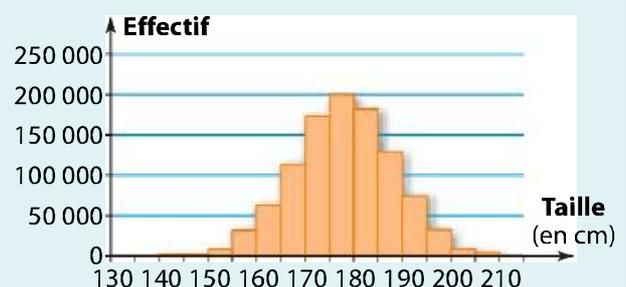
Réaliser un histogramme de cette série et expliquer pourquoi l'utilisation du couple (moyenne ; écart-type) semble pertinente.

Solution

D'après la symétrie de l'histogramme ci-contre, on peut prévoir que $Me \approx \bar{x}$.

Dans ce cas-là, on privilégie le couple (moyenne ; écart-type) pour résumer cette série.

Avec la calculatrice, on trouve $\bar{x} \approx 178$ cm et $\sigma \approx 10,1$ cm.



Moyenne, variance et écart-type

Dans toute la suite, on considère la série statistique suivante.

Valeur	x_1	x_2	...	x_p	Total
Effectif	n_1	n_2	...	n_p	$N = n_1 + n_2 + \dots + n_p$

1 Un indicateur de tendance centrale : la moyenne

Note

Pour une série regroupée en classes, on obtient une valeur approchée de la moyenne de la série en prenant pour x_i , les centres (ou milieux) des classes.

DÉFINITION La **moyenne** de cette série statistique est le nombre réel, noté \bar{x} , tel que :

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{N}$$

EXEMPLE : La moyenne de la série de dons présentée ci-contre est :

$$\bar{x} = \frac{12 \times 10 + 17 \times 15 + \dots + 5 \times 50}{55} = 21$$

Le don moyen est de 21 €.

Don (en euros)	10	15	20	30	50
Nombre de donateurs	12	17	10	11	5

Remarque : La moyenne \bar{x} est très sensible aux valeurs extrêmes.

2 Des paramètres de dispersion : variance, écart-type

Note

L'avantage de l'écart-type est de s'exprimer dans la même unité que les x_i .

DÉFINITIONS • La **variance** de cette série statistique est le nombre réel, noté V , tel que :

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

• L'**écart-type** d'une série statistique est le nombre réel, noté σ , tel que $\sigma = \sqrt{V}$.

EXEMPLE : Pour l'exemple précédent, la variance est :

$$V = \frac{12 \times (10 - 21)^2 + 17 \times (15 - 21)^2 + \dots + 5 \times (50 - 21)^2}{55} = \frac{7\,170}{55}$$

soit $V \approx 130,3$.

L'écart-type est $\sigma = \sqrt{\frac{7\,170}{55}}$, soit $\sigma \approx 11,4$ €.

Remarque : Dans le cas où c'est la distribution des fréquences qui est connue :

$$\bar{x} = f_1 \times x_1 + f_2 \times x_2 + \dots + f_p \times x_p$$

$$\text{et } V = f_1 (x_1 - \bar{x})^2 + f_2 (x_2 - \bar{x})^2 + \dots + f_p (x_p - \bar{x})^2$$

Valeur	x_1	x_2	...	x_p
Fréquence	f_1	f_2	...	f_p

3 Résumé d'une série statistique

On résume souvent une série statistique par un paramètre de tendance centrale associé à un paramètre de dispersion.

Deux choix sont couramment proposés : **le couple (moyenne ; écart-type)**, qui a l'inconvénient d'associer deux paramètres sensibles aux valeurs extrêmes, et **le couple (médiane ; écart interquartile)**, qui n'a pas ce défaut mais dont la détermination est moins pratique.

C Tirer profit d'une erreur

Elliot et Naima ont à faire l'exercice suivant : calculer l'effectif total, la moyenne, la médiane, les quartiles et l'écart-type de la série ci-dessous.

Valeur	3	16	25	27	42	55
Effectif	16	9	3	13	15	8

Voici les résultats obtenus par Elliot à l'aide de sa calculatrice, et les remarques de son professeur.

A la calculatrice, j'ai obtenu les résultats suivants :
 $N = 6$, $\bar{x} = 28$, $Me = 26$,
 $Q_1 = 16$, $Q_3 = 42$.
Il faut tenir compte des effectifs

- 5 a)** Comment le professeur sait-il que les effectifs n'ont pas été pris en compte ?
b) Reprendre l'exercice demandé et indiquer les valeurs correctes.

6 Naima a obtenu les écrans de calculatrice ci-dessous.

```
1-Var Stats
x̄=10.04761905
Σx=1688
Σx²=19822
Sx=4.139497965
σx=4.127159642
↓n=168
```

```
1-Var Stats
↑n=168
minX=3
Q1=8
Med=9
Q3=15
maxX=16
```

Imaginer l'erreur commise par Naima.

7 Ce tableau donne le nombre d'opérations quotidiennes dans une clinique vétérinaire durant une année.

Opération	0	1	2	3	4	5	6
Effectif	84	105	72	59	28	15	2

Avec la calculatrice, calculer \bar{x} , Me , Q_1 , Q_3 .

Conseils

Lorsque les valeurs d'une série statistique sont rentrées dans la liste 1 de la calculatrice, et les effectifs dans la liste 2, ne pas oublier d'indiquer ce choix à la calculatrice.

D Utiliser un tableur

Les résultats d'une enquête de satisfaction réalisée auprès de 20 joueurs d'un nouveau jeu vidéo sont présentés dans la feuille de calcul ci-dessous (de 0 : pas du tout satisfait à 10 : excellent).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Joueur	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	Note attribuée	2	5	9	1	10	5	6	8	8	7	2	0	3	7	10	9	7	5	9	6
3	Moyenne	Les formules à saisir dans les cellules de la colonne B sont :																			
4	Ecart type	B3	=MOYENNE(B2:U2)	B4	=ECARTYPEP(B2:U2)	B5	=MEDIANE(B2:U2)														
5	Médiane	B6	=QUARTILE(B2:U2;1)	B7	=QUARTILE(B2:U2;3)																
6	Premier quartile																				
7	Troisième quartile																				

- 8** Reproduire et compléter la feuille de calcul ci-dessus.
9 Dans la feuille de calcul ci-dessous des effectifs sont associés aux valeurs.

	A	B	C	D	E	F
1	Valeur	12	13	14	15	16
2	Effectif	3	2	6	5	8
3	On recopie vers le bas	12	13	14	15	16
4		12	13	14	15	16
5		12		14	15	16
6				14	15	16
7				14	15	16
8				14		16
9						16
10						16

- a)** Utiliser la fonction « recopier vers le bas » pour se ramener à un tableau du type de l'exercice précédent (effectifs tous égaux à 1).
b) Déterminer alors les indicateurs statistiques \bar{x} , σ , Me , Q_1 et Q_3 .

Conseils

Dans la formule des quartiles :

$$= \text{QUARTILE}(\text{plage}; \text{numéro du quartile})$$

ne pas oublier le numéro du quartile ;

dans celle de l'écart-type : $=\text{ECARTYPEP}(\text{plage})$

ne pas oublier le deuxième « P » (pour population).

Accompagnement personnalisé

Pour comprendre

A Élaborer une fiche synthèse

Voici un extrait de la fiche synthèse de Coralie sur le diagramme en boîte.

N est l'effectif total.

- $Me = \begin{cases} \text{donnée de rang } n + 1 & \text{si } N = 2n + 1 \\ \dots & \text{si } N = 2n. \end{cases}$
- Au moins $\frac{1}{4} \times N$ données sont inférieures ou égales à Q_1 .
- Au moins ... données sont inférieures ou égales à Q_3 .

Au moins 25 % des données

1 Recopier et compléter cette fiche synthèse.

2 Les prix moyens à la vente, en euros, du litre de Sans Plomb 95 et du kilogramme de banane, en France métropolitaine en 2009, sont indiqués ci-dessous.

	Janv.	Fév.	Mars	Avr.	Mai	Juin
SP 95	1,101	1,114	1,141	1,168	1,207	1,27
Banane	1,6	1,64	1,75	1,90	1,79	1,74

	Juil.	Août	Sept.	Oct.	Nov.	Déc.
SP 95	1,234	1,273	1,239	1,236	1,265	1,255
Banane	1,68	1,58	1,55	1,57	1,54	1,53

Tracer le diagramme en boîte de chacune de ces séries statistiques. Commenter les graphiques obtenus.

Conseils

Il faut savoir calculer Me , Q_1 et Q_3 mais aussi savoir les interpréter.

B Résumer une série par le couple d'indicateurs approprié

Voici quelques indications sur le choix du couple le plus approprié.

Lorsque l'**histogramme** représentant la série présente une certaine **symétrie** (voir exercice résolu 2 p. 159), la moyenne \bar{x} et la médiane Me sont proches l'une de l'autre, on peut se contenter de résumer la série par le couple $(\bar{x}; \sigma)$.

- Lorsque la moyenne \bar{x} et la médiane Me sont différentes, on donne en général les deux couples $(\bar{x}; \sigma)$ et $(Me; Q_3 - Q_1)$. Si, de plus, la série est de grande taille (plus de 100 termes), on cherche à comprendre cette différence en observant l'asymétrie de l'histogramme.

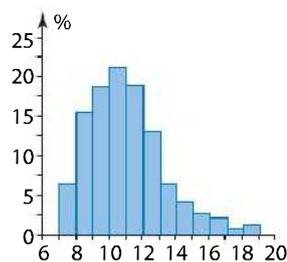
- Le couple $(Me; Q_3 - Q_1)$ convient à toutes les distributions et est en particulier peu sensible aux valeurs extrêmes.

3 Cet histogramme représente une série de 500 notes obtenues à un concours.

On sait que :

$\bar{x} = 11,2$ et $Me = 10,1$.

Quelle caractéristique de l'histogramme explique cette différence entre \bar{x} et Me ?

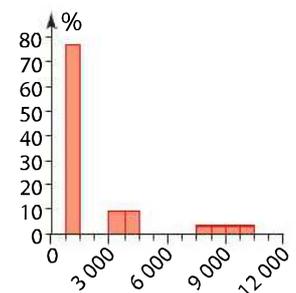


4 Cet histogramme représente la répartition des 400 salaires pratiqués dans une entreprise.

On sait que :

$\bar{x} = 1\,900$ et $Me = 1\,000$.

Quelle caractéristique de cette distribution des salaires explique cette différence entre \bar{x} et Me ?

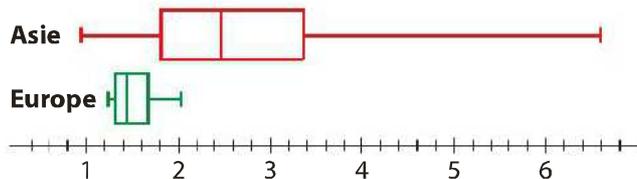


Exercices de base

Pour créer des automatismes

Médiane, quartiles et diagramme en boîte

10 Les diagrammes en boîte ci-dessous, extraits d'un magazine, illustrent les taux de fécondité en Europe et en Asie en 2009.



- Pour chaque diagramme, lire une valeur approchée du minimum, du maximum, des quartiles et de la médiane.
- Commenter les données obtenues.

11 Construire le diagramme en boîte d'une série qui correspond aux données suivantes :

```
1-Var Stats
↑n=50
minX=12
Q1=18
Med=26
Q3=31
maxX=45
```

12 Lors d'un entraînement de basket, chacun des 12 joueurs de l'équipe effectue 10 lancers francs. Le nombre de paniers réussis est indiqué ci-dessous :

Joueur	1	2	3	4	5	6	7	8	9	10	11	12
Paniers réussis	7	3	5	4	5	8	1	6	5	6	4	2

- Déterminer la médiane et les quartiles de cette série.
- Construire son diagramme en boîte.

► **Aide** : Les valeurs sont-elles rangées par ordre croissant ? Le nombre de données est-il pair ou impair ?

13 Jérémie a relevé le nombre de spams reçus sur son adresse hotmail au cours des 15 derniers jours. Il a obtenu la série suivante :

3 7 4 21 13 5 22 12 3 12 6 21 33 15 20

- Déterminer la médiane et les deux quartiles de cette série.
- Construire le diagramme en boîte de cette série.

14 La feuille de calcul ci-dessous résume les performances sur 100 mètres d'un athlète.

	A	B	C	D	E	F	G	H	I	J	K
1 Performance (en secondes)	12,3	11,9	11,8	12	12,1	11,7	12,2	11,7	11,5	12,3	
2 Médiane											
3 Premier quartile											
4 Troisième quartile											

Quelle formule faut-il saisir en cellule B2 ? B3 ? B4 ?

15 Une commune du littoral surveille l'état de sa plage, et chaque matin, un employé va compter le nombre de méduses échouées. Voici le bilan de ses relevés de juin et juillet.



Nombre de méduses	0	3	4	6	7	11	12	15	16	52
Nombre de jours	10	13	8	4	2	7	5	7	4	1

- Déterminer la médiane et les quartiles de cette série.
- En relisant son tableau, l'employé réalise qu'un jour il a inversé deux chiffres, et noté 52 au lieu de 25. Cela change-t-il les valeurs déterminées au a) ?
- Construire le diagramme en boîte de la série rectifiée.

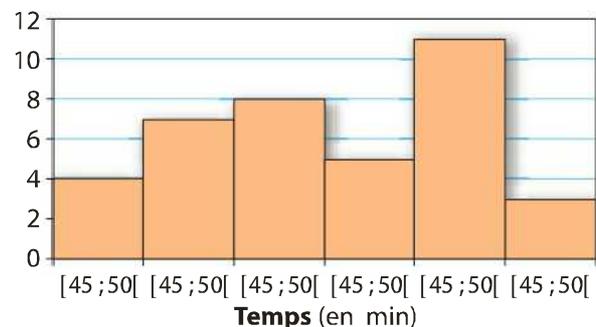
► **Conseil** : se reporter à l'exercice résolu 1, page 157.

16 Lors d'un devoir commun, les notes de tout l'établissement ont été regroupées en classes.

Note	[0;4[[4;8[[8;10[[10;12[[12;16[[16;20[
Effectif	21	46	117	123	86	7

Dans quelles classes sont la médiane et les quartiles ?

17 L'histogramme ci-dessous donne les durées en minutes du trajet Lyon-Saint-Étienne (59 km) relevées par des automobilistes sur une journée.



- Déterminer la classe médiane et les classes de chacun des quartiles.
- Construire le diagramme en boîte en remplaçant chaque paramètre par le centre de la classe auquel il appartient.

► **Aide** : Commencer par reproduire les données du graphique dans un tableau.

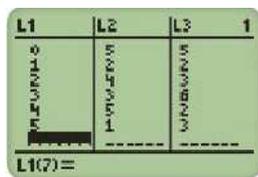
18 Le fichier pluie.xls, à télécharger sur le site compagnon (www.nathan.fr/lycee/hyperbole), contient les relevés pluviométriques du 23 octobre 2010 de stations météo françaises. Les valeurs nulles ont été écartées. En voici un extrait.

	A	B
1	Station météo	Pluie (en mm)
2	Abbeville	5,4
3	Agen	4
4	Albi	2

- a) Utiliser un tableur pour déterminer le minimum, le maximum, les quartiles et la médiane de cette série.
- b) Construire le diagramme en boîte correspondant.

19 On a relevé le nombre d'essais par match de deux équipes de rugby lors d'une saison.

Ces résultats ont été saisis à la calculatrice de la façon suivante :



- liste 1 : le nombre d'essais ;
- liste 2 : le nombre de matchs pour l'équipe 1 ;
- liste 3 : le nombre de matchs pour l'équipe 2.

Utiliser la calculatrice pour répondre aux questions suivantes.

- a) Durant cette saison, combien de matchs ont été joués par chaque équipe ?
- b) Déterminer, pour chaque équipe, la médiane et les quartiles.
- c) Représenter les diagrammes en boîte associés.

20 Les tableaux ci-dessous indiquent, pour 100 SMS, le nombre de lignes d'écritures utilisées pour l'envoi d'un SMS, suivant le type de téléphone.

	A	B	C	D	E	F	G	H
1	Avec un smartphone							
2	Nombre de lignes	1	2	3	4	5	Plus de 5	
3	Nombre de SMS	15	12	18	22	18	15	
4	Sans smartphone							
5	Nombre de lignes	1	2	3	4	5	Plus de 5	
6	Nombre de SMS	27	19	23	21	4	6	

- a) Utiliser la calculatrice ou le tableur pour déterminer la médiane et les quartiles de ces deux séries statistiques.
- b) Que se passe-t-il pour ces indicateurs si l'on remplace la donnée « Plus de 5 » par la valeur 6 ?
- c) Construire alors le diagramme en boîte de ces séries.

21 On a utilisé un moteur de comparaison de prix sur Internet pour étudier les prix proposés pour un certain type de cartouche d'encre d'imprimante. La série des prix en euros est donnée par le diagramme en tige et feuille ci-après. Déterminer la médiane et les quartiles de cette série.

12	89
13	
14	269
15	669
16	4799
17	1367778999
18	288999
19	0223579
20	0
21	99
22	059

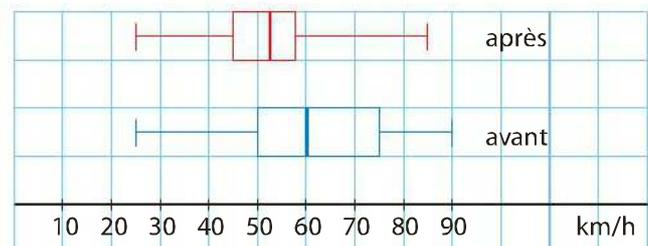
Lecture : le premier prix est 12,8 € ; puis 12,9 € ; il n'y a pas de prix entre 13 et 14 euros et le prix suivant est 14,2 €...

22 Afin de lutter contre les excès de vitesse lors de la traversée de son village, un maire a fait installer un radar indicateur de vitesse pour inciter les automobilistes à respecter la vitesse maximale autorisée de $50 \text{ km} \cdot \text{h}^{-1}$.



Les vitesses des véhicules entrant dans le village ont été relevées sur une journée avant la pose du radar, puis après la mise en service de celui-ci.

Les séries obtenues sont représentées par les diagrammes en boîte ci-dessous.



En utilisant les pourcentages définis par les quartiles et la médiane, décrire le comportement des automobilistes avant et après la mise en place du radar.

Moyenne, variance et écart-type

23 Lors d'une saison sur douze meetings d'athlétisme, un sauteur à la perche a réussi les performances suivantes :

Hauteur du saut (en m)	5,40	5,50	5,55	5,60
Effectif	1	5	4	2

Sans utiliser le menu STAT de la calculatrice, calculer la moyenne, la variance et l'écart-type de cette série.

Aide : Utiliser les formules du cours page 158 et les exemples présentés.



Exercices

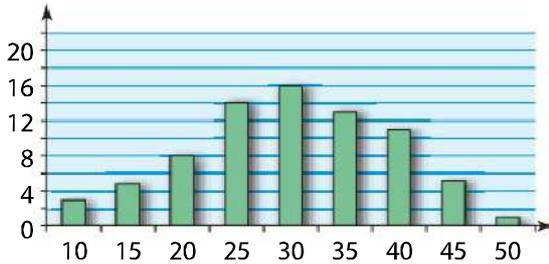
24 Sur l'écran de sa calculatrice, Lily a obtenu les données suivantes :

```
1-Var Stats
x̄=6.322857143
sx=221.3
sx²=1427.34
∑x=9089702365
σx=8958908461
n=35
```

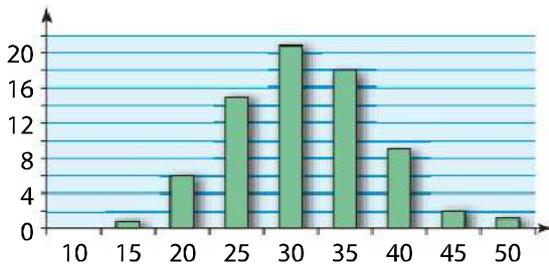
Déterminer une valeur approchée de la variance V au centième près.

25 Voici les diagrammes en bâtons de deux séries.

Série 1



Série 2



- a) Par lecture graphique :
- estimer la moyenne de chaque série ;
 - conjecturer laquelle des deux séries a le plus grand écart-type.
- b) Établir le tableau des valeurs et effectifs de chaque série, et contrôler par le calcul les réponses données à la question précédente.

► **Conseil :** se reporter à l'exercice résolu 1, page 159.

26 Durant le quinquennat d'un président de la République, des sondages mesurant sa cote de popularité sont effectués tous les six mois.

On interroge toujours le même nombre d'individus.

Sondage	1	2	3	4	5	6	7	8	9	10
Cote de popularité (en %)	52	53	49	38	35	31	39	42	45	49

Calculer la moyenne et l'écart-type de la cote de popularité de ce président sur ce quinquennat.

27 Calculer la moyenne, la variance et l'écart-type de chacune des séries dans les cas particuliers suivants.

- a) La série comporte 31 termes identiques égaux à 15.
 b) La série comporte 20 termes égaux à 32 et 20 termes égaux à 48.

28 Une série statistique est détaillée dans le tableau ci-après.

Valeur	5	9	14	19	35	41
Fréquence	0,08	0,11	0,30	0,27	0,13	

- a) Quelle est la fréquence de la valeur 41 ?
 b) Calculer la moyenne et l'écart-type de cette série.

29 Dans une région de France, le SAMU a dressé le bilan des interventions pour des accidents corporels graves sur la route durant l'année.

Nombre d'interventions par jour	0	1	2	3	4	5
Nombre de jours	13	70	115	109	61	35

- a) Représenter cette série par un diagramme en bâtons.
 b) Calculer l'arrondi au dixième de la moyenne et de l'écart-type.

30 Un artisan verrier fabrique une plaque de verre coloré pour la réalisation d'un vitrail. L'épaisseur idéale pour son travail est 2,1 mm.

La série suivante donne l'épaisseur de cette plaque, en millimètres, mesurée en différents points.

2,2	2,5	2,1	1,9	2,3
2,2	1,8	2,5	1,8	1,7

- a) Calculer la moyenne de la série. Le verrier peut-il être satisfait de son travail ?
 b) Donner une valeur approchée au centième de la variance et de l'écart-type de la série.
 En quelles unités sont donnés ces résultats ?
 c) Quelle indication cela apporte-t-il sur la qualité de la plaque réalisée ?

31 Le tableau suivant donne les fourchettes acceptables de deux paramètres sanguins.

Bilan lipidique		
Prélèvement à jeûn	Homme	Femme
TG : Triglycérides	0,45 à 1,75 g/L	0,35 à 1,4 g/L
Cholestérol total	1,5 à 2,5 g/L	

On suppose que les intervalles ont été obtenus en calculant, sur un échantillon représentatif, l'intervalle $[\bar{x} - 2\sigma ; \bar{x} + 2\sigma]$, celui-ci contenant environ 95 % des données obtenues sur l'échantillon.

Travaux pratiques

Pour expérimenter et modéliser

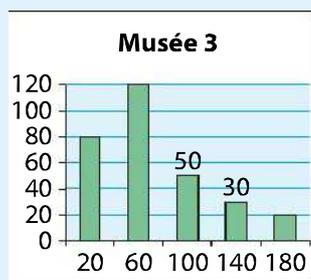
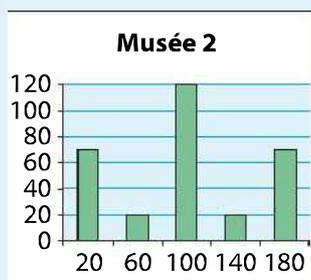
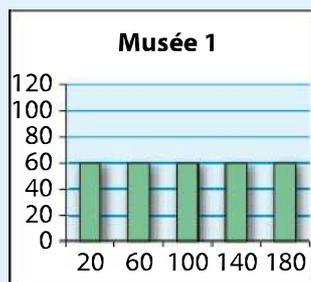
37 Deux situations

OBJECTIF Utiliser le couple d'indicateurs $(\bar{x}; \sigma)$ ou $(Me; Q_3 - Q_1)$ approprié.

1. Fréquentation de trois musées

Ces graphiques illustrent la fréquentation de trois musées de province sur 300 jours de l'année.

Le nombre de visiteurs est indiqué en abscisse et le nombre de jours en ordonnée.



- Sans calcul, prévoir la comparaison de \bar{x} et Me à la vue du diagramme en bâtons pour chaque musée. Expliquer.
- Pour chaque musée, calculer \bar{x} , σ , Me, Q_1 et Q_3 .
- Résumer chaque série par le ou les couples qui vous semblent appropriés.

2. En médecine

Voici des données concernant l'âge et la durée du séjour à l'hôpital de 397 hommes insuffisants cardiaques.

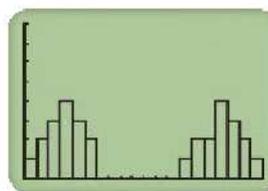
Indicateur	Âge	Durée du séjour
Moyenne	69,22	14,07
Écart-type	14,92	15,14
Médiane	71	10
Min-Max	22-100	0-99
Q_1	60,75	5
Q_3	80	16,25

- Graduer un axe de 0 à 120 avec 1 cm pour 20. Représenter sur un même schéma les diagrammes en boîte de la série des âges et de la série des durées de séjour.
- Analyser ces représentations et indiquer ce qui vous guide vers un couple d'indicateurs pour résumer chacune des ces séries.

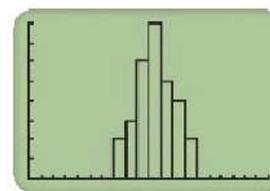
38 La lecture

OBJECTIF Savoir interpréter deux types de graphiques et choisir le résumé pertinent.

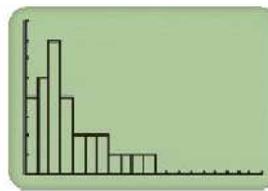
On a demandé aux 30 élèves de quatre classes d'un lycée de dire le nombre de livres qu'ils ont lus au cours de l'année 2010 (le nombre de livres lus va de 0 à 20). Les données ont permis de réaliser quatre diagrammes en boîte et quatre histogrammes pour lesquels le repère est identique.



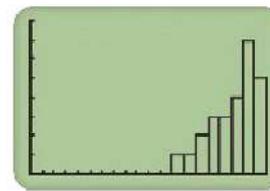
H1



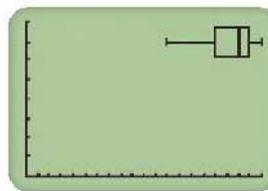
H2



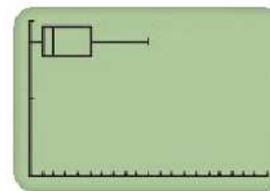
H3



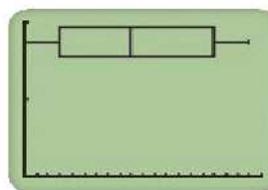
H4



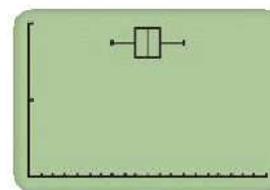
B1



B2



B3



B4

- Associer chaque histogramme au diagramme en boîte correspondant à la même série.
- Parmi ces quatre classes, l'une est une classe littéraire L1 aimant beaucoup la littérature, la deuxième est également une classe littéraire L2, mais dans laquelle une moitié est peu intéressée par la lecture, la troisième est une classe scientifique S1 homogène composée d'élèves sérieux mais pas particulièrement grands lecteurs, et la quatrième est une classe scientifique S2 d'élèves dont plus de la moitié lit au plus deux livres par an.

- a) Donner la moyenne et l'écart-type de la série des taux de triglycérides pour l'homme et pour la femme.
- b) Donner la moyenne et l'écart-type de la série des taux de cholestérol.
- c) On suppose que les échantillons d'hommes et de femmes utilisés pour obtenir les fourchettes en ce qui concerne les triglycérides étaient de même taille. Un patient ayant un taux de 1,55 g/L dont on ignore le sexe peut-il être considéré comme ayant un taux normal ? Même question pour un taux de 0,37 g/L.

32 Un artisan chocolatier et une usine de fabrication de chocolat produisent du chocolat noir à 70 % de cacao. Voici les relevés des pourcentages de cacao de 50 échantillons de chocolat pris chez chaque producteur.

Tableau 1

Pourcentage de cacao	68	69	70	71	72	73
Effectif	4	7	22	20	5	2

Tableau 2

Pourcentage de cacao	68	69	70	71	72	73
Effectif	6	12	13	14	7	8

- a) Calculer, dans chacun des cas, la moyenne et l'écart-type.
- b) Ces résultats permettent-ils de savoir d'où provient chaque tableau ?

33 On a relevé les températures minimales du mois de septembre données par 151 stations météo françaises. Pour améliorer la lisibilité de ces résultats, on les a regroupées dans le tableau suivant en classes d'amplitude 1 °C :

t]-2 ; -1]]-1 ; 0]]0 ; 1]]1 ; 2]]2 ; 3]]3 ; 4]
n	1	4	6	6	19	14

t]4 ; 5]]5 ; 6]]6 ; 7]]7 ; 8]]8 ; 9]]9 ; 10]
n	24	13	9	7	7	7

t]10 ; 11]]11 ; 12]]12 ; 13]]13 ; 14]]14 ; 15]]15 ; 16]
n	7	2	2	2	3	1

- a) Construire l'histogramme de cette série. Expliquer pourquoi le couple (moyenne ; écart-type) peut convenir pour résumer cette série.
- b) Donner une estimation de la moyenne et de l'écart-type de cette série en considérant les centres des classes.

► **Conseil :** se reporter à l'exercice résolu 2, page 159.

34 On note S_1 la série suivante :

5 54 55 58 63 64 70 72

S_2 est la série obtenue en supprimant la valeur 5.

S_3 est la série obtenue en remplaçant la valeur 5 par la valeur 50.

- a) Pour chacune de ces séries, calculer les trois mesures de dispersion (l'étendue, l'écart-type et l'écart interquartile).
- b) Calculer les deux mesures de tendance centrale (la moyenne et la médiane).
- c) Quelle caractéristique de ces différentes mesures est mise en évidence par cet exemple ?

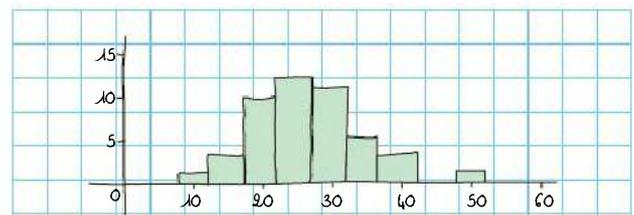
35 Pour rédiger

Lire l'énoncé et la copie d'un élève, puis rédiger la réponse en tenant compte des remarques du professeur.

Énoncé

Expliquer quel peut être le bon couple résumé pour décrire la série suivante :

x_i	10	15	20	25	30	35	40	50
n_i	1	3	10	12	11	5	3	1



Ici, un diagramme en bâtons serait plus adapté.

L'histogramme est assez symétrique.

Alors je calcule la moyenne et l'écart-type.

Ce n'est pas une obligation.

36 Les résultats d'une enquête d'opinion sur les efforts consentis par les nations pour réduire les gaz à effet de serre sont indiqués ci-dessous.

	A	B	C	D	E	F	G	H
1	Résultats : de 0 (pas du tout satisfait) à 10 (très satisfait)							
2	4	5	2	4	4	2	3	5
3	5	1	0	5	6	7	3	4
4	5	3	4	4	6	5	5	3
5	Moyenne							
6	Ecart-type							
7	Médiane							
8	Premier quartile							
9	Troisième quartile							
10	Ecart interquartile							

- a) Reproduire et compléter cette feuille de calcul.
- b) Choisir le couple d'indicateurs le plus approprié pour décrire les résultats de cette enquête.

Associer les graphiques à chaque classe.

c) Expliquer pourquoi il semble peu judicieux de résumer certaines de ces séries (à préciser) par le couple (moyenne ; écart-type).

d) Dire à quelles classes les commentaires suivants s'appliquent.

- ❶ 50 % des élèves lisent moins de 6 livres par an.
- ❷ 90 % des élèves lisent au plus 7 livres par an.
- ❸ Au moins 50 % des élèves lisent entre 3 et 16 livres par an.
- ❹ Au moins 50 % des élèves lisent plus de 18 livres par an.

e) Sans calcul, préciser, pour chaque classe, si la moyenne est supérieure, inférieure ou à peu près égale à la médiane. Vérifier par le calcul.

f) Proposer d'autres commentaires pertinents pour chaque classe.

39 Égalité des chances

OBJECTIF Observer et comprendre un effet de structure.

À l'époque du président Jimmy Carter (1977-1981), l'administration américaine était très sensible au respect des égalités (hommes-femmes ; noirs-blancs). Les résultats aux examens des étudiants d'une université de la côte Ouest sont donnés dans le tableau ci-dessous.

	A	B	C	D
1	Université des Lettres et des Sciences			
2	dirigée par un recteur			
3	Faculté des Lettres	Faculté des Sciences		
4	dirigée par un doyen		dirigée par un doyen	
5	Étudiants présents	100	500	600
6	Étudiants reçus	50	400	450
7	Étudiantes présentes	400	100	500
8	Étudiantes reçues	200	80	280

a) Suite à la parution de ces résultats, le recteur de l'université, scandalisé par les différences entre étudiantes et étudiants, écrit à ses doyens pour les sommer de s'expliquer sur de telles disparités.

Argumenter les propos du recteur par des chiffres.

b) À la réception de la requête du recteur, les doyens furent très surpris.

Expliquer pourquoi par des chiffres.

c) Quelqu'un a-t-il tort ?

Info

L'effet de structure, bien connu des statisticiens, s'explique par le fait que des variations sur les sous-populations entraînent parfois des variations contraires sur la population entière.

40 Pollution aux nitrates

OBJECTIF Se familiariser avec les courbes en cloche.

Une espèce de poissons de rivière est réputée pour être sensible aux nitrates.

En 2010, des scientifiques ont effectué une étude sur ces poissons. Ils ont pesé 250 spécimens. Les résultats sont présentés dans le tableau ci-dessous.



Masse (en grammes)	Effectif
[90 ; 92[1
[92 ; 94[2
[94 ; 96[3
[96 ; 98[4
[98 ; 100[7
[100 ; 102[14
[102 ; 104[20
[104 ; 106[27
[106 ; 108[30
[108 ; 110[32

Masse (en grammes)	Effectif
[110 ; 112[30
[112 ; 114[26
[114 ; 116[20
[116 ; 118[14
[118 ; 120[8
[120 ; 122[4
[122 ; 124[3
[124 ; 126[2
[126 ; 128[2
[128 ; 130[1

1. a) Représenter cette série statistique par un histogramme.

b) Tracer la courbe obtenue en reliant les milieux des côtés supérieurs des rectangles.

2. Estimer, à l'aide d'un tableau ou de la calculatrice, la moyenne et l'écart-type de cette série. Donner des valeurs approchées par excès au centième près.

3. Déterminer chacun des intervalles ci-dessous, en remplaçant chaque borne obtenue par la borne la plus proche de la classe dans laquelle elle se trouve. Calculer le pourcentage de poissons dans l'intervalle obtenu.

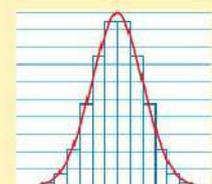
a) $[\bar{x} - \sigma ; \bar{x} + \sigma]$;

b) $[\bar{x} - 2\sigma ; \bar{x} + 2\sigma]$;

c) $[\bar{x} - 3\sigma ; \bar{x} + 3\sigma]$.

Info

Lorsqu'une distribution a l'allure d'une **courbe en cloche**, le pourcentage de données dans l'intervalle $[\bar{x} - \sigma ; \bar{x} + \sigma]$ est toujours approximativement de 68 %, dans $[\bar{x} - 2\sigma ; \bar{x} + 2\sigma]$ est approximativement de 95 %.



Exercices d'entraînement

Pour développer des compétences

Critiquer, argumenter

41 Nombre de salariés

Ce tableau donne des informations sur certaines entreprises.

Nombre de salariés	Nombre d'entreprises
0 salarié	1 745 754
1 à 9 salariés	983 994
10 à 49 salariés	163 453
50 à 249 salariés	26 397
250 à 1 999 salariés	5 073
2 000 salariés ou plus	453
Total	2 925 124

Champ : activités marchandes hors agriculture.

Source : Insee, REE (Répertoire des entreprises et des établissements), 2007.

Expliquer les raisonnements des trois élèves ci-dessous :

- **Thomas** : « On ne peut pas calculer le nombre moyen de salariés par entreprise à l'aide de ce tableau ».
- **Lily** : « Seul le couple (médiane ; écart interquartile) est pertinent pour ce tableau ».
- **Karen** : « On peut calculer la moyenne et l'écart-type pour les PME (< 250 salariés) mais les résultats seront imprécis ».

42 Le meilleur régime

Deux régimes amaigrissants ont été testés sur deux groupes de personnes d'âges différents. Les résultats sont les suivants.

Âge (ans)	Perte de poids moyenne (kg)		Effectif	
	Régime 1	Régime 2	Régime 1	Régime 2
[20 ; 30[2,5	3,5	40	10
[30 ; 50]	1,2	1,6	10	90

- Quels résultats utiliseriez-vous pour faire la promotion du régime 1 ?
- Quels résultats utiliseriez-vous pour faire la promotion du régime 2 ?
- Pour un client entre 20 et 50 ans, quel régime choisir ?

43 Le pessimiste

Pierre dit : « La majorité des gens sont plus stupides que la moyenne ».

Marine lui répond : « Ce que tu dis est absurde ».

Ce que dit Pierre est-il vraiment absurde dans toute population ?

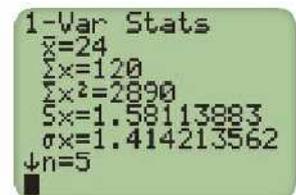
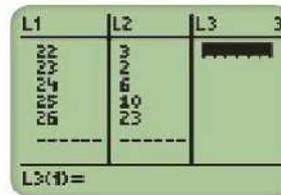
Si oui, expliquer pourquoi, et si non, trouver un contre-exemple.

Pour répondre, on pourra considérer qu'on mesure l'intelligence par une « note » comme cela se fait par exemple avec le QI.

44 Incohérences

Un élève a saisi les listes 1 (des valeurs) et 2 (des effectifs) sur sa calculatrice.

Sa calculatrice affiche les résultats des calculs.



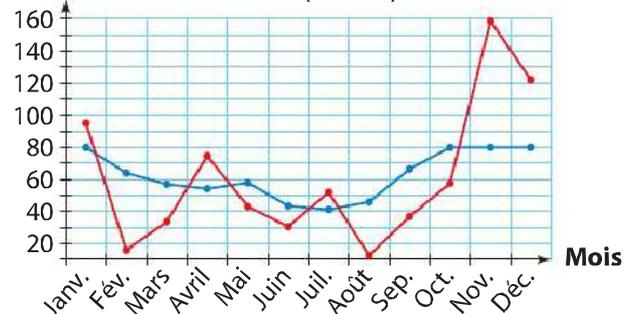
- Trouver au moins deux incohérences entre les écrans.
- Saisir les données de l'écran 1 afin d'obtenir un écran de calcul correct.

Mener des raisonnements

45 Il pleut, il pleut...

Le graphique ci-dessous donne le relevé (en bleu) de la pluviométrie moyenne mensuelle en sud Vendée depuis 1947 et (en rouge) celui de la pluviométrie mensuelle de l'année 2009.

Pluviométrie mensuelle (en mm)

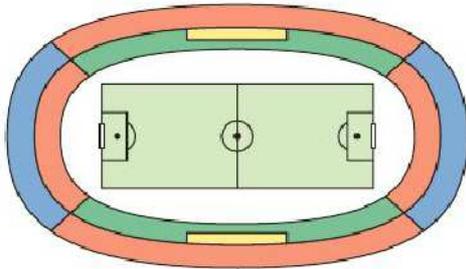


- Donner en utilisant ce graphique une estimation de la médiane et des quartiles pour chacune des deux séries.
- Donner également une estimation de la moyenne et de l'écart-type pour chacune des deux séries.
- Comment peut-on expliquer les différences constatées entre les résumés des deux séries ?

46 Stade Maracanã

Le mythique stade Maracanã (Rio de Janeiro, Brésil) compte 87 101 places.

Lors d'un match de football entre le Brésil et l'Argentine, le prix des places dépend de la couleur de la zone.



Le prix des places et leur nombre sont indiqués dans le tableau ci-dessous :

Zone	Nombre de places	Prix (en Réal R\$)
bleue	20 904	10
orange	46 164	22
verte	17 420	34
jaune	2 613	70

- a) Calculer les indicateurs \bar{x} , σ , Me, Q_1 et Q_3 .
- b) Quel est le couple d'indicateurs le plus approprié pour résumer le prix des places ?

47 objectif **Bac** **Étude sur les saumons**

Un navire de pêche, affrété par des scientifiques, effectue des prélèvements de saumons en Atlantique Nord pour les étudier. Un banc de 63 saumons a été capturé. On souhaite savoir si ces saumons sont plutôt sauvages ou plutôt issus d'un élevage d'où ils se seraient échappés. Les saumons ont été mesurés ; les résultats sont consignés dans le tableau ci-dessous :

Taille (en cm)	116	117	118	119	120	121	122	123	124	125
Effectif	2	0	1	5	5	5	4	4	5	4
Taille (en cm)	116	117	118	119	120	121	122	123	124	125
Effectifs cumulés croissants	2	2	3	8	13	18	22	26	31	35

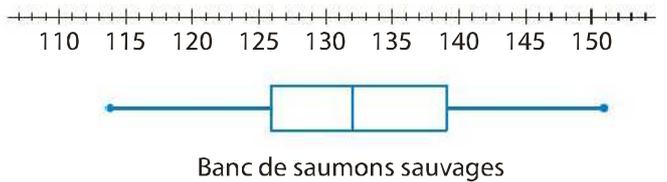
Taille (en cm)	126	127	128	129	130	131	132	133	134
Effectif	2	3	2	5	6	3	4	2	1
Taille (en cm)	126	127	128	129	130	131	132	133	134
Effectifs cumulés croissants	37	40	42	47	53	56	60	62	63

- 1. a) Donner la médiane, les premier et troisième quartiles de cette série, en détaillant votre démarche.
- b) Quelle est l'étendue de cette série ?

2. Construire le diagramme en boîte correspondant en utilisant l'échelle indiquée ci-dessous.

3. Le diagramme en boîte correspondant à un banc de saumons sauvages est tracé ci-dessous.

Peut-on dire que les saumons capturés sont plutôt issus d'un élevage ? Pourquoi ?



48 Vente par correspondance

1. Anita a relevé dans le tableau ci-dessous les prix de 218 paires de sandales pour femme d'un catalogue de vente par correspondance. Elle les a regroupés par tranche de 10 euros.

Prix (en €)]0 ; 10]]10 ; 20]]20 ; 30]]30 ; 40]]40 ; 50]
Nombre de sandales	3	16	28	43	47

Prix (en €)]50 ; 60]]60 ; 70]]70 ; 80]]80 ; 90]]90 ; 100]
Nombre de sandales	22	24	15	10	7

Prix (en €)]100 ; 110]]110 ; 120]]120 ; 130]]130 ; 140]
Nombre de sandales	1	0	0	2

On utilisera les valeurs centrales des classes pour effectuer les calculs.

a) Représenter sur la calculatrice le diagramme en bâtons de la série des prix.

b) Déterminer une valeur approchée au centième par excès de la moyenne et de l'écart-type.

2. Léo s'est intéressé aux appareils photos numériques. Il a relevé les prix dans le même catalogue en les classant, dans le tableau ci-dessous, par tranche de 40 euros.

Prix (en €)]0 ; 40]]40 ; 80]]80 ; 120]]120 ; 160]
Nombre d'appareils	5	7	10	17

Prix (en €)]160 ; 200]]200 ; 240]]240 ; 280]]280 ; 320]
Nombre d'appareils	9	3	12	3

Prix (en €)]320 ; 360]]360 ; 400]]400 ; 440]
Nombre d'appareils	0	2	4

Exercices

On utilisera les valeurs centrales des classes pour effectuer les calculs.

- a) Représenter sur la calculatrice le diagramme en bâtons de la série des prix.
- b) Déterminer une valeur approchée au centième par excès de la moyenne et de l'écart-type.
3. a) Calculer et comparer les pourcentages de chacun des deux articles dans les intervalles :
- $[\bar{x} - \sigma ; \bar{x} + \sigma]$
 - $[\bar{x} - 2\sigma ; \bar{x} + 2\sigma]$
- b) Les deux articles sont de nature très différente. En est-il de même pour la répartition des prix ?

Communiquer à l'écrit, à l'oral

49 Le chat de Geluck



Quel est l'indicateur statistique auquel le chat fait allusion ?

50 En anglais

This table shows the annual wages, in £1 000, of the workers in a factory.

Wages (£ 1,000)	Frequency
10-20	100
20-30	80
30-40	40
40-60	35
60-80	20
80-120	5

- a) Estimate the median wage and the interquartile range.
- b) Estimate the mean wage and standard deviation.
- c) Which of the two couples do you think is more representative of wages at this factory ?

Info

Frequency est un faux ami et signifie effectif !

51 Groupes de niveau

Un organisme de formation d'adultes a organisé un test de compréhension d'un texte en français. Les scores vont de 0 à 10.

Pour un premier groupe de 60 stagiaires, les résultats sont les suivants :

Score	4	6	7	8	9	10
Effectif	5	22	18	10	4	1

Pour un autre groupe de 20 stagiaires dont la langue maternelle n'est pas le français, on obtient :

Score	0	1	2	3	4	6
Effectif	1	2	5	9	1	2

- a) Choisir un résumé et une représentation graphique pour chacune des deux séries.
- b) Cela a-t-il un sens de s'intéresser à la moyenne de la série obtenue en réunissant les deux séries ?

52 Les « bonnes » académies

Le tableau ci-dessous permet de comparer les résultats au baccalauréat de 1993 et 2003.

Quelles sont les bonnes académies ? Argumenter à l'aide de résumés statistiques pertinents.

Académie	1993			
	Général	Techno.	Profession.	Total
Aix-Marseille	36,7	12,3	5,6	54,6
Amiens	29,6	15,0	5,7	50,3
Besançon	33,0	13,7	6,3	53,0
Bordeaux	35,7	11,6	4,3	51,6
Caen	32,6	14,6	6,1	52,7
Clermont-Ferrand	35,8	13,4	7,4	56,6
Corse	38,3	9,9	5,1	53,3
Créteil	27,8	13,1	5,5	46,4
Dijon	33,1	15,0	5,6	53,7
Grenoble	37,9	14,3	5,8	58,0
Lille	30,3	15,7	7,2	53,2
Limoges	35,0	14,6	7,7	57,3
Lyon	34,0	13,0	6,3	53,3
Montpellier	34,1	12,3	4,2	50,6
Nancy-Metz	33,9	16,2	7,0	57,1
Nantes	37,1	14,4	6,8	58,3
Nice	35,3	11,1	4,5	50,9
Orléans-Tours	34,9	14,0	5,7	54,6
Paris	55,3	11,4	4,8	71,5
Poitiers	36,1	14,0	5,7	55,8
Reims	32,4	12,4	6,2	51,0
Rennes	40,5	17,9	6,2	64,6
Rouen	30,4	14,5	6,9	51,8
Strasbourg	30,6	12,3	6,0	48,9
Toulouse	39,9	13,9	5,8	59,6
Versailles	35,0	14,7	5,0	54,7
Métropole	34,9	13,9	5,9	54,7

Académie	2003			
	Général	Techno.	Profession.	Total
Aix-Marseille	34,3	17,3	11,0	62,6
Amiens	28,7	16,9	11,5	57,0
Besançon	32,2	17,8	12,4	62,5
Bordeaux	34,3	16,9	11,5	62,7
Caen	31,3	16,6	13,9	61,8
Clermont-Ferrand	34,9	17,7	13,2	65,7
Corse	35,7	17,2	7,0	59,9
Créteil	29,5	18,3	9,2	57,0
Dijon	32,5	18,5	12,7	63,7
Grenoble	36,8	19,3	11,2	67,3
Lille	29,0	19,0	11,8	59,7
Limoges	34,2	20,0	13,4	67,6
Lyon	34,8	16,3	11,5	62,5
Montpellier	32,5	16,6	9,8	58,8
Nancy-Metz	31,1	18,8	12,7	62,5
Nantes	34,1	17,5	12,7	64,3
Nice	35,7	15,2	9,9	60,8
Orléans-Tours	32,5	17,9	10,7	61,1
Paris	58,1	14,2	8,5	80,9
Poitiers	32,0	18,8	13,0	63,7
Reims	30,8	17,7	12,6	61,1
Rennes	36,1	22,2	13,2	71,4
Rouen	29,2	17,4	13,0	59,6
Strasbourg	31,1	16,6	11,8	59,4
Toulouse	34,7	19,2	11,9	65,8
Versailles	36,7	18,2	9,3	64,2
Métropole	33,7	17,9	11,4	62,9

Source : MEN-DEP.

53 Les deux bars

Les gérants de deux bars comparent leurs ventes d'expressos et de jus de fruits sur une journée.

Bar 1	Effectif	Prix (en euros)
Expresso	4	4
Jus de fruits	66	1,80

Bar 2	Effectif	Prix (en euros)
Expresso	66	2
Jus de fruits	4	0,90

Commentaire du gérant du bar 1 :

« Je ne comprends pas, nous avons vendu le même nombre de produits, je vends les miens deux fois plus chers que les tiens, et pourtant, en moyenne tes ventes sont supérieures ».

Expliquer ces commentaires et ce paradoxe.

54 Efficacité devant le but

On a relevé le nombre de buts marqués par deux joueurs durant une année.

Joueur A						
Nombre de buts par match	0	1	2	3	4	5
Effectif	12	10	0	6	8	4

Joueur B						
Nombre de buts par match	0	1	2	3	4	5
Effectif	0	12	19	7	3	0

- Calculer la médiane de chaque série.
- Un diagramme en boîte n'est pas très significatif dans cette situation. Pourquoi ?
- À l'aide de la calculatrice, calculer la moyenne et l'écart-type de chaque série.
- Utiliser les résultats précédents pour commenter les qualités et les défauts de chaque buteur.

S'initier à la logique**55 Implications et réciproques**

x est une valeur d'une série statistique.

On considère les propositions suivantes :

- A : « x est strictement supérieur à la médiane » ;
- B : « x est supérieur ou égal à la médiane » ;
- C : « Au moins 50 % des données de la série sont strictement inférieures à x » ;
- D : « Au plus 50 % des données de la série sont supérieures ou à égales à x ».

1. Pour chacune des propositions suivantes, dire si elle est vraie ou fausse.

- A implique C
- A implique D
- B implique C
- B implique D

2. Envisager la réciproque de chacune des propositions ci-dessus et dire si elle est vraie ou fausse.

56 Quantificateurs universel, existentiel

Les propositions suivantes sont-elles vraies ou fausses ? Justifier en utilisant des exemples ou des contre-exemples.

- Quelle que soit la série statistique, la moyenne est supérieure à l'écart-type.
- Quelle que soit la série statistique, le premier quartile est strictement inférieur à la médiane.
- Il existe une série statistique pour laquelle la moyenne est égale à l'écart-type.
- Il existe une série statistique pour laquelle l'écart interquartile est égal à l'étendue.



QCM

57 Pour chaque question, une seule réponse est exacte. Laquelle ?

Les temps de parcours en minutes de 50 candidats à une course ont été regroupés dans le tableau de valeurs suivant :

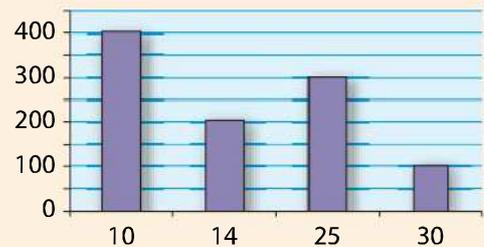
Valeur x_i	7	9	13	14	19	22	25	28	30
Effectif n_i	3	8	6	8	6	5	4	8	2

- Le troisième quartile de cette série est :
 a) 26,5 b) 22 c) 25
- Si on supprime la valeur 30, l'écart interquartile sera-t-il :
 a) le même ? b) plus petit ? c) plus grand ?
- La moyenne de cette série statistique est égale à :
 a) 18,56 b) 17,82 c) 19

58 Pour chaque question, une seule réponse est exacte. Laquelle ? Justifier.

Le diagramme en bâtons ci-contre représente une série.

- L'écart interquartile est :
 a) 15 b) 11 c) 20
- La variance de cette série est :
 a) 65,12 b) 57,41 c) 57,47
- Si on supprime la valeur 10, l'écart-type :
 a) ne change pas b) augmente c) diminue



Exercices interactifs

Vrai-Faux

59 Dire pour chaque affirmation si elle est vraie ou fausse.

- Pour avoir les informations nécessaires à la réalisation d'un diagramme en boîte, il faut ordonner les valeurs de la série étudiée.
- Dans une série, on a toujours :
 $Q_1 < Me < Q_3$
- Dans un diagramme en boîte, la médiane est toujours au milieu de la boîte.
- Il y a au moins 50 % des valeurs d'une série dans l'intervalle $[Q_1 ; Q_3]$.
- Il y a exactement 25 % des valeurs d'une série comprise entre Q_3 et la valeur maximale.

60 Dire pour chaque affirmation si elle est vraie ou fausse. Justifier.

Dans une grande ville, on a observé les résultats suivants :

Prix d'une baguette (en €)	0,70	0,75	0,80	0,85	0,90	0,95
Nombre de boulangeries	5	3	10	2	4	1

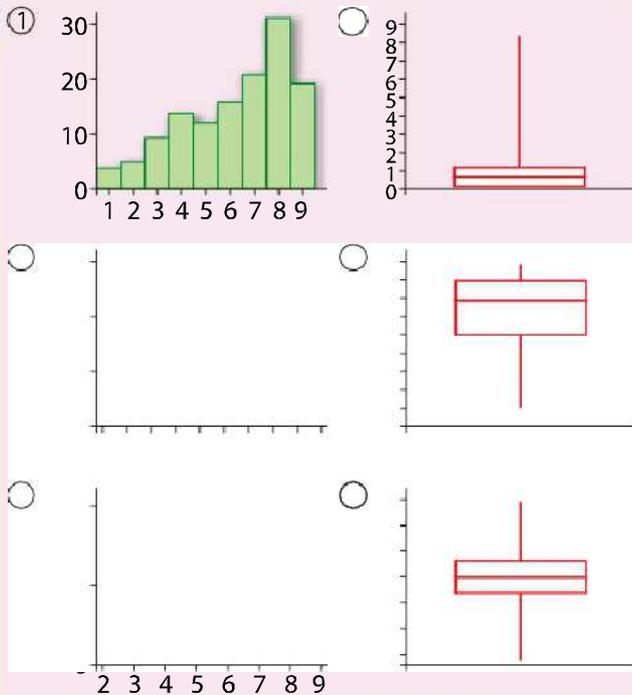
- Le prix moyen d'une baguette est 0,825 €.
- Le prix médian est inférieur au prix moyen.
- Au moins 75 % des prix de la baguette sont inférieurs ou égaux à 0,85 €.
- Si on enlève les deux valeurs extrêmes, l'écart-type est modifié.

Exercices interactifs

Se préparer au contrôle

61 D'un type de graphique à l'autre

Associer à chaque histogramme le diagramme en boîte qui lui correspond.



Conseils

- Quels indicateurs peut-on lire sur les diagrammes en boîte ?
- Peut-on estimer leur valeur sur les histogrammes ?

62 Un tableau de fréquences

On considère la série statistique suivante :

Valeur	1	3	4	7	8
Fréquence	0,02	0,15	0,07	0,1	0,03

Valeur	11	12	17	25	31
Fréquence	0,27	0,11	0,05	0,09	0,11

- À l'aide de la calculatrice, donner la médiane et les quartiles de cette série.
- Vérifier les résultats du a) par un calcul.
- Construire le diagramme en boîte.

Conseils

On peut utiliser les fréquences cumulées croissantes. Dans ce cas, à quelle fréquence cumulée croissante va correspondre Q_1 ?

63 Résumé (moyenne ; écart-type)

Le tableau suivant donne la répartition du nombre d'interventions graves quotidiennes d'un service de pompiers sur une année.

Intervention	0	1	2	3	4	5	6
Effectif	10	20	62	168	57	28	20

- Donner l'arrondi au dixième de la moyenne \bar{x} et de l'écart-type σ de cette série.
- Calculer le pourcentage, arrondi au dixième, des valeurs comprises dans l'intervalle $[\bar{x} - \sigma ; \bar{x} + \sigma]$.

Conseils

- Voir le cours page 158.

64 Regroupement des données en classes

40 élèves d'une classe ont chronométré leur temps de trajet (en minutes) pour venir au lycée ce jour-là.

10	12	5	15	10	7	18	12
9	20	16	35	7	24	11	4
22	39	26	51	31	21	9	15
43	11	18	22	23	12	28	17
12	9	37	37	21	1	7	14

- Calculer la moyenne et l'écart-type de cette série.
- Comme on envisage d'étendre l'enquête aux 1 500 élèves du lycée, il semble plus rapide de regrouper ces données en classes d'amplitude 10, la première classe étant alors $[0 ; 10[$. Effectuer ce regroupement et calculer la moyenne et l'écart-type à partir de ces nouvelles données.
- Comparer les résultats obtenus au a) et au b).

Conseils

- Présenter les données dans un tableau d'effectifs.

65 Utilisation du tableur

25 lycéens ont testé un jeu vidéo.

Chaque score est inscrit dans les cellules A2 à A26 de cette feuille de calcul.

	A	B	C
1	Score	Médiane	
2	315	Premier quartile	
3	326	Troisième quartile	
4	521	Ecart interquartile	
5	632	Moyenne	
6	541	Ecart-type	
7	201		
8	103		
9	412		

Quelles formules faut-il saisir dans les cellules de la colonne C ?

Exercices d'approfondissement

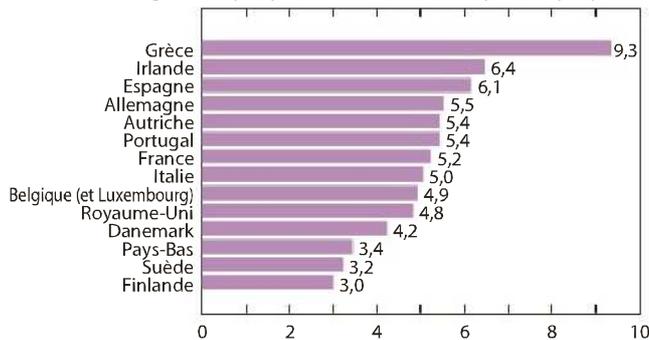
Pour aller plus loin

66 Avec un guide (1)

Voici deux tableaux concernant la consommation de tabac dans l'Union européenne.

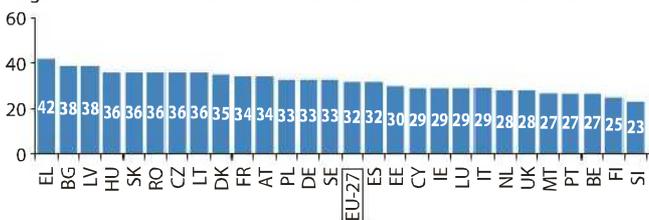
Graphique 1 : consommation de tabac en 1996

Nombre de cigarettes par personne de 15 ans et plus, et par jour



Graphique 2 : tobacco users

Regular or occasional use of combustible and/or non-combustible tobacco



1. a) La présentation des données du graphique 1 suggère fortement l'utilisation d'un des résumés du cours. Lequel ? Réaliser la représentation graphique associée à ce résumé, pour ce graphique.

b) Interpréter, en une phrase faisant référence à la situation étudiée, chacune des valeurs utilisées pour réaliser cette représentation.

2. a) Peut-on deviner quelle est la variable étudiée dans le deuxième graphique ?

b) Construire un histogramme en regroupant les données par classes d'amplitude 3, la première classe commençant à 23.

D'après cet histogramme, semble-t-il justifié d'utiliser le couple (moyenne ; écart-type) pour résumer la série ?

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

d) Était-il tout à fait certain que l'on trouverait la même valeur que la moyenne de l'Union européenne ? Expliquer.

Guide de résolution

- Bien identifier la variable représentée dans le premier graphique.
- Les graphiques séparés de leurs contextes sont parfois incomplets. La liste des abréviations suggère l'Europe (à 27), ce qui permet de dater approximativement les données. Les données sont-elles des effectifs ou des pourcentages ? S'agit-il exclusivement de fumeurs ?

67 Avec un guide (2)

Une machine produit des pièces dont le diamètre doit être de 5 cm. Bien sûr, on observe malgré tout de petites différences sur les diamètres. Pour savoir si la machine est bien réglée et fonctionne correctement, on prélève un échantillon de 40 pièces et on obtient la liste suivante :

4,9	5	5,2	4,7	4,8
5,1	4,5	5,2	4,9	4,8
4,9	4,9	4,9	5,3	5
4,8	4,8	4,9	5,1	5,3
4,5	4,9	4,9	5	4,8
4,8	5,3	4,8	5,1	5
5,1	4,8	4,7	5	5
4,8	4,6	4,7	4,9	4,7

On sait que, dans des conditions normales de fonctionnement :

- environ 95 % des données de l'échantillon appartiennent à l'intervalle $[\bar{x} - 2\sigma ; \bar{x} + 2\sigma]$, où $\bar{x} = 5$ est la moyenne théorique et σ est l'écart-type de l'échantillon ;
- environ 68 % des données de l'échantillon appartiennent à l'intervalle $[\bar{x} - \sigma ; \bar{x} + \sigma]$.

Peut-on supposer que la machine est bien réglée ?

Guide de résolution

Rentrer la série des données à la calculatrice pour obtenir l'écart-type σ .
Déterminer les intervalles $[\bar{x} - 2\sigma ; \bar{x} + 2\sigma]$ $[\bar{x} - \sigma ; \bar{x} + \sigma]$.
Ordonner la liste à la calculatrice.
Compter combien de données sont situées dans chacun des intervalles ainsi que les pourcentages qu'elles représentent de l'échantillon total. Conclure.

68 Coefficient de variation en lémurie

En mission à Madagascar, des scientifiques ont étudié deux espèces de lémuriens : des lémurs catta et des lémurs vari.

Ils ont relevé les masses suivantes (en kg).

• **Lémurs catta :**

3,1 - 4 - 3,2 - 3,9 - 4 - 3

• **Lémurs vari :**

4,8 - 5,4 - 4,6 - 5,3 - 4,4 - 5,7

1. a) Déterminer les moyennes et les écarts-types de chaque espèce de lémuriens.

b) Quelle est la série la plus dispersée ?

