CHAPITRE 11

Statistiques

D'un siècle à un autre

Le 7 janvier 2009, à Marseille, les palmiers étaient sous la neige. Pourtant, l'année 2009 fait partie des dix années les plus chaudes depuis 1850.

C'est sur l'étude de séries historiques que reposent les études sur le changement climatique. Ainsi, on a pu établir que la température moyenne à la surface de la Terre a augmenté de près d'un degré depuis 1860.

Ces analyses de données utilisent de nombreux outils mathématiques inventés par John Tukey.

En savoir plus sur **John Wilder Tukey**

→ Chercheurs d'hier p. 279

Rappels & Questions-tests

Compléments numériques

Moyenne d'une série

On considère une série quantitative d'effectif total N.

Valeur	x_1	x_2	......	x_p
Effectif	n_1	n_2	......	n_p
Fréquence	f_1	f_2	......	f_p

- **La moyenne $\overline{x}$** peut se calculer à partir :
– des effectifs : $\overline{x} = \dfrac{n_1x_1 + n_2x_2 + \ldots + n_px_p}{N}$
avec $N = n_1 + n_2 + \ldots + n_p$;
– des fréquences : $\overline{x} = f_1x_1 + f_2x_2 + \ldots + f_px_p$.

- On dit que $\overline{x}$ est la moyenne des valeurs **pondérées** par leurs effectifs (ou leurs fréquences).

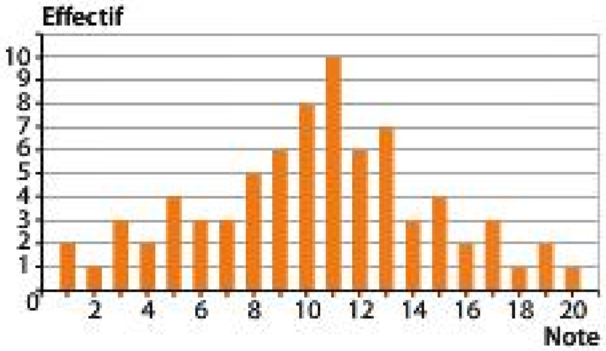
- Si la série est regroupée en classes, on fait l'hypothèse que la répartition des valeurs est uniforme à l'intérieur de chaque classe. On calcule alors une valeur approchée de la moyenne en affectant l'effectif (ou la fréquence) de chaque classe au centre de celle-ci.

Médiane et quartiles

On suppose la série précédente rangée dans l'ordre croissant. Les N valeurs sont écrites l'une après l'autre, une même valeur étant répétée autant de fois que l'indique son effectif : $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_N$.

- **La médiane Me** est le nombre tel que :
— si N est impair, $N = 2k + 1$, alors $Me = x_{k+1}$;
— si N est pair, $N = 2k$, alors $Me = \dfrac{x_k + x_{k+1}}{2}$.
- **Les quartiles Q_1 et Q_3** sont les valeurs que l'on calcule de la manière suivante :
Q_1 (resp. Q_3) est la valeur dont le rang est le premier entier supérieur ou égal à $\dfrac{N}{4}$ (resp. $\dfrac{3N}{4}$).

1 Le relevé des notes de mathématiques dans un jury du bac S est donné par le diagramme suivant.



a) Calculez la note moyenne arrondie à 0,01 près.
b) Quel est le pourcentage, arrondi à 1 % près, des candidats qui ont obtenu au moins la note 10 ?

2 Voici la répartition des salaires bruts mensuels (en milliers d'euros) dans une entreprise.

Salaire	[1 ; 2[	[2 ; 3[	[3 ; 4[	[4 ; 5[
Fréquence	0,48	0,34	0,16	0,02

a) Donnez une estimation du salaire moyen.
b) Le directeur financier constate que les promotions de fin d'année ont modifié les seules fréquences des tranches [1 ; 2[et [2 ; 3[. Le salaire moyen est passé à 2 300 €.
On note x la fréquence de la première classe.
Quelle est la nouvelle répartition des salaires ?

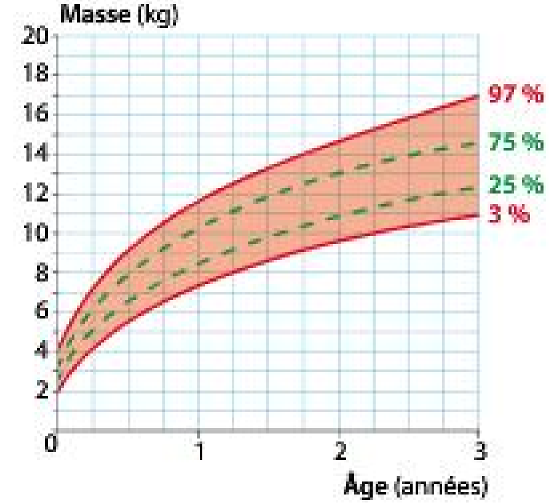
3 On reprend la série de l'exercice 1.
a) Calculez la note médiane.
b) Déterminez les quartiles Q_1 et Q_3.
c) Quel est le pourcentage (arrondi à 1 % près) des candidats qui ont obtenu une note dans l'intervalle [8 ; 13] ?

4 Pendant une semaine, Chloé a noté chaque jour le nombre de ses nouvelles amies sur Facebook.
Elle a obtenu une série de sept valeurs, toutes différentes, dont voici les paramètres :
- valeur mini : 2 ;
- médiane : 11 ;
- moyenne : 9 ;
- quartile Q_1 : 3 ;
- quartile Q_3 : 13 ;
- étendue : 15.

Pouvez-vous trouver, dans l'ordre croissant, le nombre de nouvelles amies avec qui Chloé a été en contact chaque jour de cette semaine ?

Voir les corrigés p. 363

Activité 1 LA MASSE DES BÉBÉS



Les courbes ci-contre sont issues de l'étude statistique de la masse des bébés entre 0 et 3 ans.

Pour chaque bébé, on indique son âge et sa masse par un point du graphique. Pour 3 % de ces bébés, ce point est situé sous la courbe inférieure rouge.

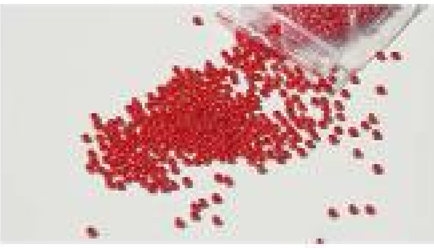
1 Quelle information donne la courbe supérieure ?

2 a) Quelle masse lit-on pour un bébé de trois mois sur la courbe repérée par 25 % ?
b) À quel paramètre statistique correspond-elle ?

3 a) Quel est le troisième quartile de la série des masses des enfants de 2 ans ?
b) Dans quel intervalle se situe la médiane de la série des masses d'un enfant d'un an ? de deux ans ?
c) À quel âge 75 % des bébés pèsent-ils 8 kg ou moins ?

Activité 2 APPROCHE D'UNE MESURE DE DISPERSION TICE

Une entreprise utilise deux machines, M_1 et M_2, pour remplir des sachets de perles. Voici la composition des sachets issus d'un contrôle sur des échantillons provenant de M_1 et M_2.



Nombre de perles : x_i	108	109	110	111	112	113	114	115
M_1 : Nombre de sachets	3	3	10	33	95	39	16	1
M_2 : Nombre de sachets	1	5	16	79	30	26	25	18

1 Calculez pour chaque machine le nombre moyen (noté $\overline{x}$) de perles par sachet.

2 La direction souhaite obtenir des sachets contenant un nombre de perles le plus proche possible de 112. Les machines sont réglées pour des moyennes de 112, mais il subsiste des écarts.

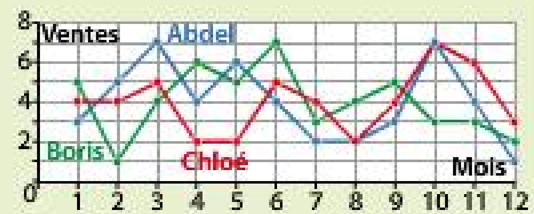
L'objectif est de savoir quelle machine a des résultats s'écartant le moins possible de la moyenne.
Pour cela, on procède pour chaque série, de la manière suivante :
- on mesure chacun des écarts $(x_i - \overline{x})$, puis on les élève au carré pour obtenir $(x_i - \overline{x})^2$;
- on calcule alors la moyenne pondérée de ces carrés.

Appliquez cette méthode en utilisant un tableur. outil 9

Problème ouvert

Refaites cet exercice après le chapitre ... Est-il plus facile ?

Le directeur d'une concession automobile offre tous les ans au meilleur de ses trois agents un séjour d'une semaine aux Antilles. Il décide de récompenser celui qui réalise les ventes les plus régulières.
Les profils de vente sont donnés sur le graphique.
Quel est le vendeur qui sera choisi ?



1 Diagramme en boîte

1.1 Médiane et quartiles

On considère une série statistique X, à caractère quantitatif, d'effectif N, dont toutes les données sont rangées dans l'ordre croissant : $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_N$.

La **médiane Me** et les **quartiles Q_1 et Q_3** sont des paramètres de position qui permettent un partage de la population suivant le schéma ci-dessous.

Animation

Exemple. Voici le relevé, dans l'ordre croissant, du rythme cardiaque au repos, exprimé en pulsations par minute, des 34 élèves d'une classe de 1re S.

Rythme cardiaque	59	63	70	71	73	77	79	80	84	86	88	91
Effectif	2	1	5	2	5	4	2	2	4	5	1	1

Voir les méthodes de calcul dans les rappels p. 266.

L'effectif est pair, $34 = 2 \times 17$, donc $\text{Me} = \dfrac{x_{17} + x_{18}}{2}$ soit $\text{Me} = \dfrac{77 + 77}{2} = 77$.

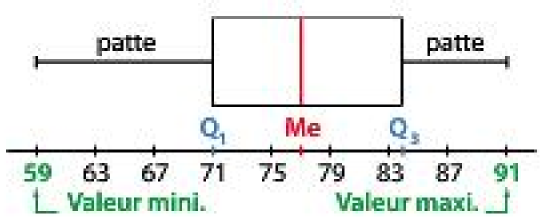
Les quartiles sont $Q_1 = x_9 = 71$ et $Q_3 = x_{26} = 84$.

Remarque. $Q_1 = 71$. Or 10 valeurs sur 34 sont inférieures ou égales à 71 soit environ 29 %, donc ce pourcentage dépasse 25 %. Ce phénomène résulte de la prise en compte de la valeur 71 plusieurs fois. Il illustre la définition de Q_1 par un partage des valeurs du type (au moins 25 % ; au moins 75 %). Il en est de même pour la définition de Me et de Q_3.

1.2 Représentation par un diagramme en boîte

La répartition des données peut être représentée par un diagramme dit en boîte, qui résume le caractère étudié par les valeurs extrêmes, la médiane et les quartiles.

Le diagramme en boîte ci-dessous correspond à l'exemple précédent.



Note
Ce type de diagramme, dû au statisticien américain John Wilder Tukey (1915-2000), est aussi appelé **diagramme en boîte à moustaches** ou **diagramme à pattes**.

On lit également la répartition des données rangées dans l'ordre croissant :

— au moins 50 % des valeurs sont dans l'intervalle interquartile $[Q_1 ; Q_3]$;

— au moins 25 % des valeurs ne dépassent pas Q_1 alors qu'au moins 25 % dépassent Q_3.

Un tel diagramme illustre la dispersion de la moitié des valeurs autour de la médiane.

Les pattes représentent la plage des valeurs restantes limitées par les valeurs minimale et maximale ; ainsi la boîte ne tient pas compte directement des valeurs extrêmes.

2 Paramètres de dispersion

2.1 Étendue et écart interquartile

Nous connaissons déjà deux paramètres de dispersion : l'étendue et l'écart interquartile.

- **L'étendue** est la différence entre les valeurs extrêmes : $e = x_{max} - x_{min}$.

Cet indicateur grossier ne tient compte que des valeurs extrêmes et non des intermédiaires.

- **L'écart interquartile** est la différence entre les quartiles Q_3 et Q_1 : $E_I = Q_3 - Q_1$.

Il mesure la dispersion des 50 % des valeurs qui **entourent la médiane**.

2.2 Variance et écart-type

On considère une série statistique X dont les valeurs x_i ont pour effectifs respectifs n_i ; on la note : $X = (x_i ; n_i)$.

Valeur	x_1	x_2	---	x_p
Effectif	n_1	n_2	---	n_p

Nous allons définir un nouveau paramètre qui permet de **mesurer la dispersion des valeurs x_i autour de la moyenne $\overline{x}$**.

Pour cela, on calcule la moyenne des carrés des écarts $(x_i - \overline{x})$ (voir activité 2 p. 267).

Définition 1

$(x_i ; n_i)$ est une série statistique d'effectif total N.

- **La variance** de cette série est le nombre :

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

- **L'écart-type** de cette série est le nombre $s = \sqrt{V}$.

On peut aussi écrire : $V = \dfrac{\sum_{i=1}^{p} n_i(x_i - \overline{x})^2}{N}$.
Le numérateur se lit : « somme de $i = 1$ jusqu'à p de $n_i(x_i - \overline{x})^2$ ».

Remarque. La variance V et l'écart-type s sont deux nombres positifs tels que $V = s^2$.

Variance et fréquence. Comme $\dfrac{n_i}{N} = f_i$, on peut aussi calculer la variance à partir de la série $(x_i ; f_i)$ des fréquences.

$$V = f_1(x_1 - \overline{x})^2 + f_2(x_2 - \overline{x})^2 + \ldots + f_p(x_p - \overline{x})^2.$$

Là aussi on peut écrire : $V = \sum_{i=1}^{p} f_i(x_i - \overline{x})^2$.

Théorème 1

La variance de la série $X = (x_i ; n_i)$ est telle que : $V = \dfrac{n_1x_1^2 + n_2x_2^2 + \ldots + n_px_p^2}{N} - \overline{x}^2$.

Ainsi, V est « la moyenne des carrés des valeurs x_i moins le carré de la moyenne $\overline{x}$ ».

Démonstration. Par définition de la variance, $NV = n_1(x_1 - \overline{x})^2 + n_2(x_2 - \overline{x})^2 + \ldots + n_p(x_p - \overline{x})^2$.

Dans l'expression de NV, on remplace chaque $(x_i - \overline{x})^2$ par $(x_i^2 - 2x_i\overline{x} + \overline{x}^2)$:

$NV = n_1(x_1^2 - 2x_1\overline{x} + \overline{x}^2) + n_2(x_2^2 - 2x_2\overline{x} + \overline{x}^2) + \ldots + n_p(x_p^2 - 2x_p\overline{x} + \overline{x}^2)$

$NV = (n_1x_1^2 + n_2x_2^2 + \ldots + n_px_p^2) - (2n_1x_1\overline{x} + 2n_2x_2\overline{x} + \ldots + 2n_px_p\overline{x}) + (n_1\overline{x}^2 + n_2\overline{x}^2 + \ldots + n_p\overline{x}^2)$

$NV = (n_1x_1^2 + n_2x_2^2 + \ldots + n_px_p^2) - 2\overline{x}(n_1x_1 + n_2x_2 + \ldots + n_px_p) + \overline{x}^2(n_1 + n_2 + \ldots + n_p)$.

Or par définition de la moyenne, $n_1x_1 + n_2x_2 + \ldots + n_px_p = N\overline{x}$ et $n_1 + n_2 + \ldots + n_p = N$ donc

$NV = (n_1x_1^2 + n_2x_2^2 + \ldots + n_px_p^2) - 2\overline{x}N\overline{x} + \overline{x}^2N = (n_1x_1^2 + n_2x_2^2 + \ldots + n_px_p^2) - 2N\overline{x}^2 + N\overline{x}^2$

$NV = (n_1x_1^2 + n_2x_2^2 + \ldots + n_px_p^2) - N\overline{x}^2$.

D'où, en divisant par N : $V = \dfrac{n_1x_1^2 + n_2x_2^2 + \ldots + n_px_p^2}{N} - \overline{x}^2$.

Exemple. Les notes de mathématiques d'Anna et Brahim lors des dix contrôles réalisés au cours de l'année scolaire sont représentées sur le diagramme ci-après.

On vérifie qu'Anna et Brahim ont la même moyenne : $\overline{x_A} = \overline{x_B} = 11{,}4$.

Examinons la dispersion des résultats par rapport à la moyenne en comparant les écarts-types.

La variance de la série de notes d'Anna est, d'après la définition 1 :

$$V = \frac{\sum_{i=1}^{p} n_i(x_i - \overline{x})^2}{N}$$

$$V_A = \frac{(8-11{,}4)^2 + 2\times(9-11{,}4)^2 + (11-11{,}4)^2 + 3\times(12-11{,}4)^2 + 2\times(13-11{,}4)^2 + (15-11{,}4)^2}{10} = 4{,}24.$$

Pour la variance de la série de notes de Brahim, utilisons la formule du théorème 1 :

$$V = \frac{\sum_{i=1}^{p} n_i x_i^2}{N} - \overline{x}^2$$

$$V_B = \frac{2\times 7^2 + 10^2 + 4\times 11^2 + 13^2 + 16^2 + 17^2}{10} - 11{,}4^2 = 9{,}64.$$

Note

Dans la pratique, l'utilisation de la calculatrice permet d'obtenir rapidement ces divers paramètres.

Ainsi les écarts-types sont $s_A = \sqrt{V_A} \approx 2{,}06$ et $s_B = \sqrt{V_B} \approx 3{,}10$.

La comparaison $s_A < s_B$ indique que les notes d'Anna sont moins dispersées autour de la moyenne et donc que ses résultats sont plus réguliers.

Le calcul nous confirme une première impression donnée par le graphique.

Remarque. Si les données sont exprimées avec une unité (mètre, seconde, gramme…) alors la variance V s'exprime en unité « carré » (m^2, s^2, g^2, …). L'écart-type $s = \sqrt{V}$ s'écrit avec la même unité que les données.

3 Résumé d'une série statistique

Résumer une série, c'est indiquer la **répartition des données** en utilisant différents indicateurs. Deux questions peuvent alors être posées :

- Autour de quelle valeur centrale les données sont-elles réparties ?
- Quelle est l'importance de la dispersion des données autour de cette valeur centrale ?

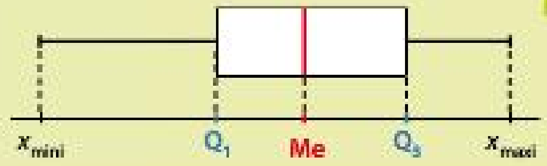
On utilise habituellement un paramètre de position indiquant une tendance centrale et un paramètre de dispersion.

Ainsi pour résumer une série on peut déterminer puis, interpréter suivant l'étude désirée, l'un des couples définis dans le tableau ci-dessous. → *Exercice 34, page 283*

Paramètre de tendance centrale	Paramètre de dispersion	Propriété
médiane : Me	écart interquartile : $E_I = Q_3 - Q_1$	peu sensible aux valeurs extrêmes
moyenne : $\overline{x}$	écart-type : s	sensible aux valeurs extrêmes

OBJECTIF 1 Utiliser un diagramme en boîte

- Un diagramme en boîte illustre la répartition des valeurs d'une série ordonnée dans l'ordre croissant suivant le schéma :



EXERCICE RÉSOLU A Construire et interpréter un diagramme en boîte

Le tableau suivant indique l'espérance de vie (en années) dans les pays africains en 2007 (source : *Word Factbook 2007-2008*).

Espérance de vie	32	38	39	40	41	42	44	47	48	50	51	52	53
Nombre de pays	1	1	1	1	2	1	1	1	2	4	5	1	2

Espérance de vie	54	55	57	58	59	61	62	63	70	72	76	77
Nombre de pays	5	2	2	2	3	2	1	1	1	2	2	1

1. Représenter cette série par un diagramme en boîte.

2. À partir du diagramme, présentez la répartition des pays africains suivant l'espérance de vie de leur population.

Méthode

1. On s'assure que la série est rangée dans l'ordre croissant.

- On calcule la médiane.

N est l'effectif total de la série :
– si N est impair, $N = 2k + 1$, alors $Me = x_{k+1}$;
– si N est pair, $N = 2k$, alors $Me = \frac{x_k + x_{k+1}}{2}$

- On calcule les quartiles.

On cherche le premier entier j (resp. k) tel que :
$$j \geqslant \frac{N}{4} \text{ (resp. } k \geqslant \frac{3N}{4}\text{)}.$$
Alors $Q_1 = x_j$ et $Q_3 = x_k$

- On repère ensuite sur un axe les valeurs remarquables : valeur minimale, Q_1, Me, Q_3, valeur maximale.

On trace alors la boîte dont la longueur correspond à celle de $[Q_1 ; Q_3]$, et les pattes.

2. On indique que :
- la boîte représente environ 50 % des valeurs autour de la médiane ;
- les pattes représentent les valeurs en dehors de l'intervalle interquartile $[Q_1 ; Q_3]$.

Solution

1. • D'après le tableau, la série est rangée dans l'ordre croissant.

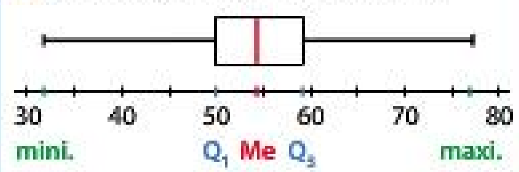
- L'effectif total est N = 47.

Il est impair, 47 = 2 × 23 + 1, donc $Me = x_{24}$.
La 24^e valeur est 54 d'où Me = 54.

- $\frac{N}{4} = 11{,}75$ donc le premier entier supérieur ou égal à 11,75 est 12.

Ainsi, $Q_1 = x_{12}$.
La 12^e valeur est 50 donc $Q_1 = 50$.
De même $Q_3 = x_{36} = 59$.

- D'où le diagramme en boîte suivant.



2. Ainsi, dans environ 50 % des pays africains, l'espérance de vie se situe entre 50 et 59 ans, dans 25 % des pays elle ne dépasse pas 50 ans et dans les 25 % restants elle est supérieure à 59 ans.

Mise en pratique

1 Le tableau indique l'âge des vainqueurs des derniers Tours de France.

28	27	33	25	29	34	33	32	31
30	29	28	28	24	32	31	30	29
28	27	29	28	28	28	25	31	24
23	28	27	34	25	24	29	30	26
29	28	27	26	25	24	28	27	26

1. Quel est l'âge le plus fréquent de la série ?

2. a) Déterminez la médiane et les quartiles.

b) Représentez la distribution des âges par un diagramme en boîte.

c) Commentez cette répartition.

2 Le tableau indique les fréquences cumulées croissantes (f.c.c.) donnant la répartition des élèves d'un lycée suivant la distance d (en km) entre le domicile et ce lycée.

Distance d (en km)	< 2	< 6	< 10	< 18	< 34
f.c.c. (en %)	12	45	76	95	100

1. Tracez la courbe représentative des f.c.c.

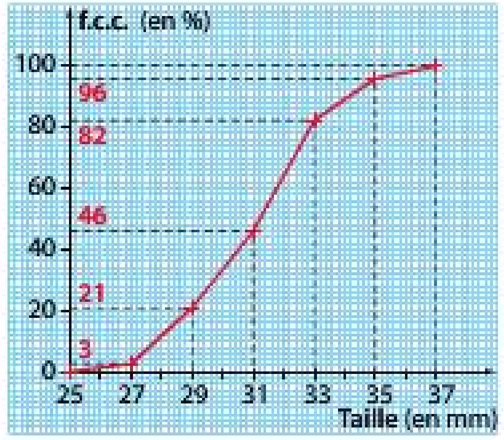
2. a) Placez sur le graphique les valeurs qui représentent une estimation de la médiane et des quartiles.

b) Tracez le diagramme en boîte associé, puis résumez la répartition des élèves suivant d.

3 **La plume du peintre**

Un ornithologue a mesuré la taille (en mm) des plumes du peintre qu'on trouve aux extrémités des ailes de la bécasse des bois.

La répartition des tailles est représentée par la courbe des fréquences cumulées croissantes suivante :

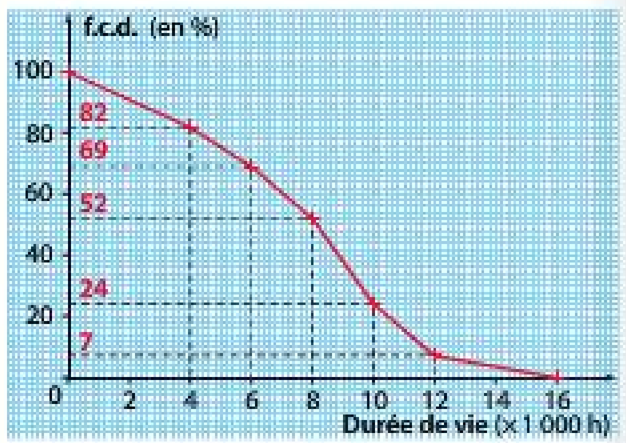


1. Donnez par lecture graphique une estimation des quartiles et de la médiane.

2. a) Tracez le diagramme en boîte associé.

b) Déduisez-en une présentation résumant la répartition des tailles des plumes.

4 Un test de durée de vie, sur un lot d'ampoules basse consommation, est résumé par la courbe des fréquences cumulées décroissantes suivante :



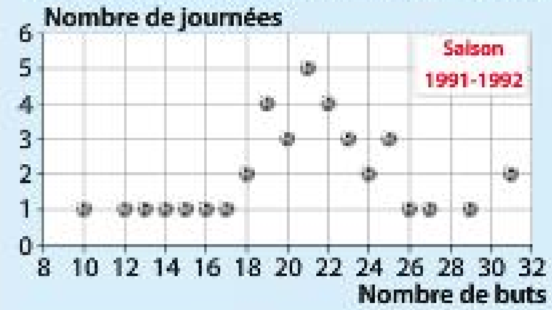
1. Quelle est la durée de vie que dépassent 75 % du lot d'ampoules ? 50 % ? 25 % ?

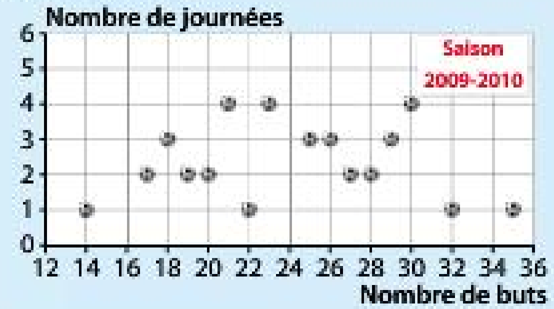
2. Tracez le diagramme en boîte associé.

3. Si vous étiez le responsable de fabrication, comment présenteriez-vous ces résultats sur la durée de vie ?

EXERCICE RÉSOLU B Comparer des séries

Les graphiques donnent la répartition du nombre de buts marqués par journée lors des saisons 1991-1992 (série 1) et 2009-2010 (série 2) de la Ligue 1 de football.





1. Représentez par un diagramme en boîtes superposées les deux séries.

2. Comparez les deux séries ; déduisez-en la saison où le nombre de buts marqués par journée est le plus régulier et celle où la répartition des journées montre une meilleure efficacité des buteurs.

Méthode

1. • On calcule les indicateurs nécessaires à la réalisation du diagramme en boîte de la série 1.

• On calcule les indicateurs de la série 2.

• On construit les diagrammes à partir du même axe. Ainsi les boîtes sont superposées ce qui facilite la comparaison graphique.

2. La comparaison peut se faire sur la valeur de tendance centrale, Me, ainsi que sur la dispersion des valeurs.

• On compare les médianes.

• On compare les étendues et les écarts interquartiles, qui traduisent la dispersion des valeurs.

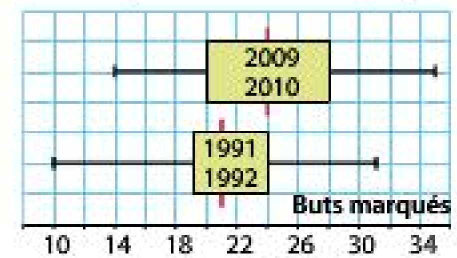
• On conclut quant aux questions posées.

Solution

1. • Pour la série 1, les paramètres sont : effectif N = 38 ; $x_{mini} = 10$; $x_{maxi} = 31$; $Q_1 = x_{10} = 19$; $Q_3 = x_{29} = 24$; $Me = \dfrac{x_{19} + x_{20}}{2} = \dfrac{21 + 21}{2} = 21$.

• De même, pour la série 2 : $N' = 38$; $x'_{mini} = 14$; $x'_{maxi} = 35$; $Q'_1 = 20$; $Me' = \dfrac{23 + 25}{2} = 24$; $Q'_3 = 28$.

• D'où le diagramme en boîtes superposées.



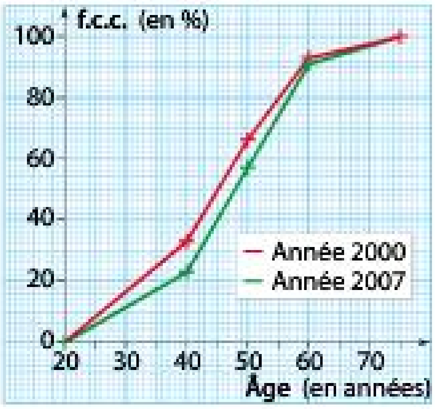
2. On constate que Me < Me'.
Lors de la moitié des journées de la saison 1991-1992, il a été marqué 21 buts au moins contre 24 buts lors de la saison 2009-2010.
L'étendue des deux séries est la même : 21.
Les écarts interquartiles sont :

$$Q_3 - Q_1 = 5 \text{ et } Q'_3 - Q'_1 = 8.$$

• Ainsi la série 1 est la moins dispersée autour de la médiane donc la saison 1991-1992 est celle où le nombre de buts marqués par journée est le plus régulier.
On remarque que $Me' = Q_3$. Ainsi, pour la série 2, la moitié des journées ont un bilan de 24 buts ou plus alors qu'un tel résultat n'est atteint que pour un quart des journées de la série 1. La répartition des journées en 2009-2010 montre une meilleure efficacité des buteurs.

Mise en pratique

5 Le graphique des fréquences cumulées croissantes indique la répartition des âges des chefs d'exploitations agricoles pour 2000 et 2007.



1. Estimez pour chacune des deux années les quartiles et la médiane de chacune des séries.

2. a) Tracez les diagrammes en boîte associés.

b) Comparez les deux séries. Commentez l'évolution de la répartition des âges.

6 On a relevé les rythmes cardiaques (en pulsations par minute) au repos puis à l'effort d'un groupe d'élèves.

Repos	74	84	70	69	70	79
Effort	133	105	94	89	103	121

Repos	86	85	83	79	86	73	69
Effort	133	106	106	111	133	89	105

Repos	70	85	77	77	88	76	63
Effort	127	108	109	98	126	139	111

Repos	84	82	90	94	77	89	60
Effort	128	130	114	108	106	100	119

1. Représentez par un diagramme en boîtes superposées les deux séries.

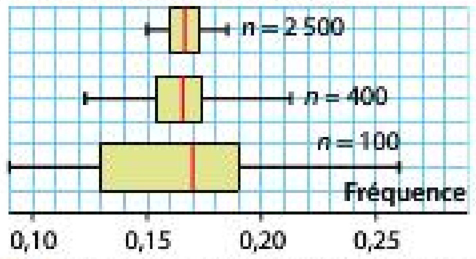
2. a) Comparez les étendues et la dispersion autour de la médiane de ces séries.

b) Présentez un résumé de la répartition des rythmes cardiaques au repos ou avec effort sur cet échantillon.

7 On lance deux dés équilibrés et on note la somme S des numéros apparus.
Une expérience consiste à répéter n fois cette épreuve ; on calcule alors la fréquence de l'issue « S = 7 » lors de ces n lancers.
On a simulé 100 fois cette expérience d'abord lorsque $n = 100$ puis $n = 400$, enfin $n = 2\,500$.
On a ainsi obtenu trois séries de fréquences dont la répartition est illustrée par le diagramme en boîtes superposées.

1. a) Comparez la dispersion des fréquences suivant les valeurs de n.

b) Quelle conjecture peut-on faire sur la fréquence lorsque n devient grand ?



2. a) Vérifiez à l'aide d'un tableau à double entrée que la probabilité de l'événement « S = 7 » est $\frac{1}{6}$.

b) Ce résultat est-il en accord avec les valeurs successives de la médiane ?

8 Une expérience consiste à lancer 200 fois trois pièces équilibrées et à examiner le nombre de sorties de Pile. On note « P = 1 » et « P = 2 » les issues donnant un seul Pile et exactement deux Pile.
L'expérience a été simulée 100 fois.
Voici le nombre de réalisations des issues « P = 1 » et « P = 2 » lors de la série de 100 expériences.

Issue « P = 1 ».

86	89	73	70	89	86	79	72	83	71
77	60	76	83	82	83	90	72	71	71
80	77	67	75	88	77	81	81	73	76
78	77	83	78	77	79	70	72	76	73
68	70	83	77	73	75	76	71	77	69
73	71	76	76	82	84	65	76	55	71
72	85	71	80	77	65	73	79	72	77
71	76	87	79	75	73	74	80	69	64
78	81	82	73	65	83	76	90	85	76
75	66	68	72	78	67	83	61	71	75

Issue « P = 2 ».

70	66	80	77	75	82	74	74	69	82
76	81	74	70	74	72	58	83	79	78
70	73	79	75	59	82	62	61	80	74
74	77	69	76	70	72	76	76	81	83
63	80	68	70	72	79	71	76	81	73
74	81	77	80	85	71	78	74	91	73
79	80	73	75	66	78	79	79	78	73
80	65	76	75	73	72	78	86	75	84
80	63	76	81	83	71	78	62	71	68
81	75	77	76	71	91	74	82	85	77

1. Utilisez la calculatrice pour déterminer les paramètres (mini – Q_1 – Me – Q_3 – maxi) de chacune des séries.

2. Représentez ces séries par un diagramme en boîtes superposées.

3. Comparez les deux séries. Quelle conjecture pouvez-vous émettre concernant la répartition des fréquences de ces issues ?

OBJECTIF 2 Utiliser le couple (moyenne ; écart-type)

- **Variance V d'une série statistique :**

$$V = \frac{\sum_{i=1}^{p} n_i(x_i - \overline{x})^2}{N} \quad \text{ou} \quad V = \frac{\sum_{i=1}^{p} n_i x_i^2}{N} - \overline{x}^2$$

n_i est l'effectif associé à la valeur x_i,
N est l'effectif total :
$N = n_1 + n_2 + \ldots + n_p$,
$\overline{x}$ est la moyenne.

- **Écart-type :** $s = \sqrt{V}$.
- Dans le cas d'une série regroupée en classes, on obtient une estimation de la moyenne et de l'écart-type en affectant les effectifs au centre de chaque classe.

EXERCICE RÉSOLU C Calculer l'écart-type et évaluer la dispersion

Les tableaux indiquent les durées d'ensoleillement, exprimées en heures, dans les villes de Brest et Strasbourg durant le mois de juin 2010 (source météociel).

- **Brest**

Durée en heures (x_i)	0	1	4	6	7	9	10	12	13	14
Nombre de jours (n_i)	2	3	4	3	2	3	3	2	4	4

- **Strasbourg**

Durée en heures (x_i)	0	1	4	5	6	8	12	13	14	15
Nombre de jours (n_i)	1	4	3	4	3	4	3	1	3	4

1. Quelle est la durée moyenne d'ensoleillement, $\overline{x}$, pour chacune des deux villes ?

2. Calculez l'écart-type s de chacune des séries.

3. Pour chacune des villes, quel est le pourcentage de valeurs dans $[\overline{x} - s\,;\overline{x} + s]$?
L'affirmation « en juin, il a fait aussi beau à Brest qu'à Strasbourg » est-elle fondée ?

Méthode

1. On calcule les moyennes des deux séries.
On utilise la formule écrite avec Σ :

$$\overline{x} = \frac{\sum_{i=1}^{p} n_i x_i}{N}.$$

2. Pour calculer l'écart-type $s = \sqrt{V}$, on détermine d'abord la variance V.

On utilise la formule : $V = \frac{\sum_{i=1}^{p} n_i x_i^2}{N} - \overline{x}^2$.

- On en déduit l'écart-type s.

Note
Les valeurs de $\overline{x}$ et s peuvent être obtenues à la calculatrice (voir page 280).

3. On exprime sous forme de pourcentage la proportion de valeurs dans $[\overline{x} - s\,;\overline{x} + s]$.

- $p = \frac{\text{effectif dans l'intervalle}}{\text{effectif total}}$.

- On compare les durées moyennes et les pourcentages, qui traduisent la dispersion des valeurs autour de la moyenne.

Solution

1. Pour Brest, la durée moyenne d'ensoleillement est :

$$\overline{x} = \frac{2 \times 0 + 3 \times 1 + \ldots + 4 \times 13 + 4 \times 14}{30} = 8 \text{ h}.$$

De même pour Strasbourg : $\overline{x} = 7{,}9$ h.

2. Pour Brest, la somme des carrés est :
$2 \times 0^2 + 3 \times 1^2 + \ldots + 4 \times 13^2 + 4 \times 14^2 = 2\,564$.
D'où le calcul de la variance :

$$V = \frac{2\,564}{30} - 8^2 \approx 21{,}47.$$

- Or $s = \sqrt{V}$ donc $s \approx 4{,}63$ h.

De même pour Strasbourg :

$$V = \frac{2\,605}{30} - 7{,}9^2 \approx 24{,}42 \text{ et } s \approx 4{,}94 \text{ h}.$$

3. Pour Brest, l'intervalle $[\overline{x} - s\,;\overline{x} + s]$ a pour bornes $\overline{x} - s \approx 3{,}37$ et $\overline{x} + s \approx 12{,}63$.
Ainsi, 17 valeurs sont dans cet intervalle.
D'où la proportion $p_B = \frac{17}{30} \approx 0{,}57$ soit 57 %.
De même, pour Strasbourg :
$\overline{x} - s \approx 2{,}96$, $\overline{x} + s = 12{,}84$ et $p_S \approx 57\,\%$.

- Les durées moyennes sont proches et la dispersion des valeurs autour d'elles est comparable, donc l'affirmation semble fondée.

Mise en pratique

9 **Tir à l'arc**

Lors d'un entraînement au tir à l'arc, Vân et Léa envoient 10 volées de 3 flèches sur une cible.

Voici les résultats de Vân :

7	8	3	8	9	9	4	9	3	5
7	8	7	9	6	4	10	9	10	8
7	7	6	8	7	6	9	10	8	6

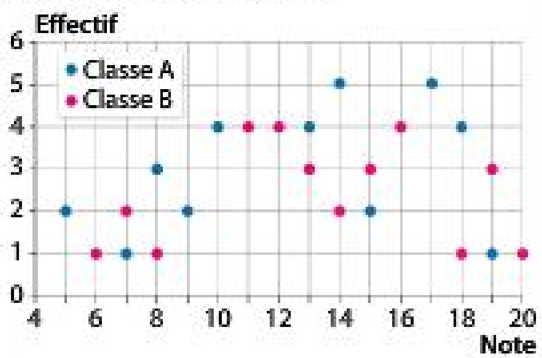
Voici les résultats de Léa :

4	8	8	7	9	7	7	9	6	5
8	9	8	8	7	7	7	6	9	5
4	8	9	7	9	8	7	8	6	7

1. Calculez la moyenne et l'écart-type de chacune des séries.

2. Quelle est la tireuse qui semble la plus régulière ? Justifiez.

10 La répartition des notes obtenues lors de la présentation des TPE par deux classes de 1[re] S, est représentée sur le graphique.



1. Pour chacune des deux classes calculez la note moyenne $\bar{x}$ et l'écart-type s.

2. Déterminez pour chacune des classes le pourcentage d'élèves dont la note appartient à l'intervalle $[\bar{x}-s\,;\bar{x}+s]$.

3. La préparation aux TPE a été commune et avant les épreuves les professeurs indiquent que les deux classes semblent d'un niveau comparable. Au vu des résultats, cette impression est-elle confirmée ?

11 Une enquête a testé les durées d'attente d (en s) chez deux opérateurs A et B, fournisseurs d'accès à internet (*Source* : DSL Valley).

d	[20 ; 30[	[30 ; 40[	[40 ; 50[	[50 ; 60[
A		13 %	41 %	19 %
B	5 %	15 %	25 %	11 %

d	[60 ; 80[	[80 ; 120[	[120 ; 180[	[180 ; 240[
A	11 %	6 %	4 %	6 %
B	18 %	12 %	7 %	7 %

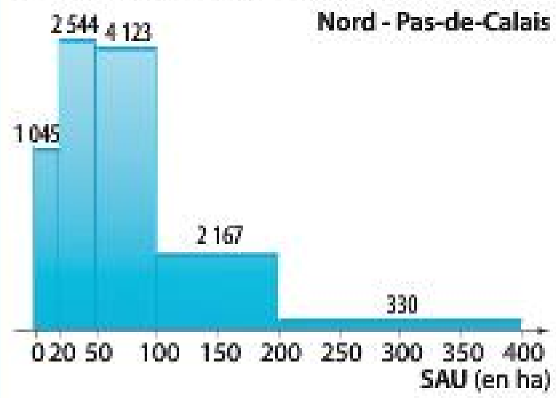
1. Estimez les durées moyennes d'attente chez chacun des opérateurs.

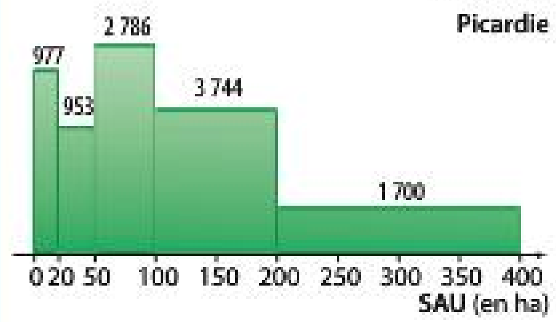
Aide
Utiliser le centre de chaque classe.

2. Calculez une estimation de l'écart-type pour chacune des séries.

3. Au vu de ces résultats quel opérateur semble posséder le service de relation avec les consommateurs le plus efficace ?

12 Les histogrammes représentent pour deux régions la répartition des exploitations agricoles suivant leur surface agricole utile (SAU en ha).





1. Donnez une valeur approchée de la SAU moyenne dans chaque région.

2. Calculez une estimation de l'écart-type de chacune des séries.

3. Utilisez ces indicateurs pour présenter les différences notables dans les structures des exploitations agricoles de ces deux régions.

Pour se tester

Exercices interactifs

13 Questions sur le cours

Complétez les propositions suivantes.

On ordonne dans l'ordre croissant les tailles (en cm) des 35 élèves d'une classe de première.

a) Le premier quartile est la taille de rang ……

b) Le troisième quartile est la taille de rang ……

c) Dans le diagramme en boîte associé à la série, la longueur de la boîte correspond à ……

d) Dans le diagramme en boîte associé à la série, la longueur totale des pattes correspond à ……

e) La variance est …… de l'écart-type.

f) La variance s'exprime en ……

14 Vrai ou faux

Vrai ou faux ? Justifiez votre réponse.

On considère une série de données (x_i) d'effectif N, de moyenne $\bar{x}$, d'écart-type s et de médiane Me.

a) La moyenne des carrés des valeurs x_i vaut $\bar{x}^2+s^2$.

b) Si $s=0$ alors la moyenne des carrés des valeurs x_i est égale au carré de leur moyenne.

c) Si $N=32$ et $\bar{x}=12$ alors $\sum_{i=1}^{32} x_i = 384$.

d) Si $N=10$, $\bar{x}=5$ et $\sum_{i=1}^{10} x_i^2 = 310$ alors $s^2=26$.

15 QCM Une seule réponse exacte

Pour chaque affirmation, une seule réponse est exacte. Identifiez-la en justifiant votre réponse.

Pour contrôler la qualité de ses analyses biologiques, un laboratoire procède au dosage d'un même échantillon d'urée tous les jours durant un mois. Ces tests ont pour but de détecter d'éventuelles erreurs de manipulation et d'examiner l'homogénéité des résultats donnés par un dispositif. Le tableau donne la série X des relevés de ces dosages en $g \times L^{-1}$.

0,31	0,30	0,30	0,30	0,29	0,28	0,31	0,28
0,30	0,31	0,29	0,31	0,36	0,32	0,33	0,30
0,28	0,32	0,30	0,30	0,30	0,29	0,29	0,28
0,35	0,30	0,29	0,32	0,27	0,31		

1. Les paramètres de la série X sont tels que :

a) Me = 0,31 b) $\bar{x}$ < Me

c) $s \approx 0{,}0004$ d) écart interquartile : $e = 0{,}02$

2. Les mesures hors de [Me − 1,5e ; Me + 1,5e] sont considérées comme « anormales ». Le laboratoire les élimine et utilise la série tronquée X′ pour donner les « limites de confiance » 0,271 et 0,328 entre lesquelles doivent se trouver toutes les données x'_i.

a) Une seule valeur est anormale.

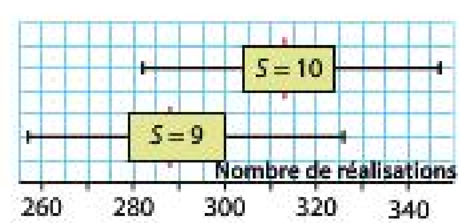
b) Toutes les valeurs x'_i sont dans les limites.

c) Le seuil d'alerte est atteint.

16 QCM Au moins une réponse exacte

Pour chaque affirmation, plusieurs réponses peuvent être exactes. Identifiez-les en justifiant votre réponse.

Une expérience consiste à lancer 2 500 fois trois dés et à relever le nombre de réalisations d'une somme de 9 et celle d'une somme de 10. On a simulé 100 fois cette expérience. Voici la distribution du nombre de réalisations de « S = 9 » et de « S = 10 » :



1. La série 1 associée à l'issue « S = 10 » est telle que :

a) $Q_1 = 304$. b) $Q_3 = 324$.

c) Dans 50 % des cas, il y a 313 réalisations ou moins.

d) Dans 25 % des cas, il y a au moins 304 réalisations.

2. La série 2 associée à l'issue « S = 9 » est telle que :

a) Me = 288. b) L'écart interquartile est 21.

c) Dans 25 % des cas il y a 279 réalisations ou plus.

d) Dans 75 % des cas il y a au plus 300 réalisations.

3. Les fréquences f_9 et f_{10} de réalisation de « S = 9 » et « S = 10 » lors des 100 expériences sont telles que :

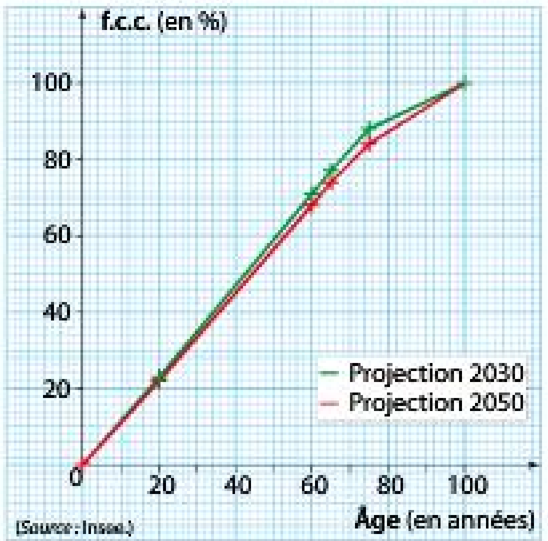
a) $0{,}102 < f_9 < 0{,}131$. b) $0{,}112 < f_{10} < 0{,}140$.

Voir les corrigés p. 366

Apprendre à chercher

17 Scénario d'évolution

En 2010, la structure de la population de la France métropolitaine suivant l'âge entre 0 et 100 ans est telle que $Q_1 = 21$, Me $= 39{,}5$ et $Q_3 = 58{,}5$.
Les projections sur la structure en 2030 puis en 2050 sont données par le graphique.



Objectif Construire un diagramme en boîtes superposées puis comparer la répartition de la population de 2010 avec celles des projections.

Pour construire un diagramme en boîte, on utilise les valeurs minimale et maximale du caractère ainsi que la médiane et des quartiles. Pour les projections, le graphique des fréquences cumulées croissantes permet d'estimer ces indicateurs.

1. Pour chacune des projections indiquez une valeur approchée des quartiles et de la médiane.

2. Représentez la situation par un diagramme en boîtes superposées. (On adoptera l'échelle : 1 cm → 10 ans.)

La comparaison peut alors porter sur trois catégories de la population définies à partir du premier quartile, de la boîte (tendance centrale) et du troisième quartile.

3. Présentez les évolutions de la répartition de la population française suivant ces projections.

18 Effet de structure

Voici les résultats d'une enquête dans deux villes A et B portant sur le montant des dépenses (en €) concernant 30 médicaments d'usage courant, non remboursés.

1. Comparez la dépense moyenne par habitant (arrondie à 1 € près) pour la ville A et la ville B.

	Âge	[0 ; 20[	[20 ; 30[	[30 ; 50[	[50 ; 90[
A	Population	2 034	2 024	2 915	3 940
	Dépenses	16 207	34 612	55 687	94 312
B	Population	2 850	1 943	1 741	524
	Dépenses	22 829	32 934	33 207	12 586

Objectif Pour chacune des villes, comparez la dépense moyenne par habitant. Pour chaque tranche d'âge, comparez la dépense moyenne par personne et par ville. Commentez les résultats.

Pour une ville, la dépense moyenne par habitant est le rapport : $\dfrac{\text{dépense totale}}{\text{population de la ville}}$.
De même on calcule la dépense par personne et par tranche d'âge en se référant à l'effectif d'une tranche.

2. a) Présentez dans un tableau, les dépenses moyennes par personne et par tranche d'âge.
(Les résultats seront arrondis à 1 € près.)

b) Comparez ces résultats pour les villes A et B.

3. Quel phénomène remarquez-vous ? Comment pouvez-vous l'expliquer ?

Commentaire
La répartition des personnes suivant l'âge influence le montant moyen des dépenses. Ce phénomène est appelé « effet de structure ».

19 Une propriété de la variance

Une série statistique d'effectif N, de moyenne $\overline{x}$ et de variance V est définie par sa distribution de fréquences.

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

On définit sur l'intervalle $I = [x_1 ; x_p]$ la fonction g par $g(t) = f_1(x_1 - t)^2 + f_2(x_2 - t)^2 + \ldots + f_p(x_p - t)^2$.
Ainsi $g(t)$ est la moyenne des $(x_i - t)^2$.

Objectif Prouvez que V est le minimum de g.

g est une fonction polynôme de degré 2 de la variable t. Pour trouver son extremum on peut penser à utiliser la forme développée de $g(t)$ puis à utiliser les résultats sur le sens de variation d'une telle fonction.

1. Prouvez que pour tout t de I, $g(t) = At^2 + Bt + C$ avec $A = 1$, $B = -2\overline{x}$ et $C = V + \overline{x}^2$.

2. a) Pour quelle valeur de t, g a-t-elle un extremum ?

b) Justifiez la nature et la valeur de cet extremum. Concluez.

Narration de recherche

→ L'objectif n'est pas seulement de résoudre le problème posé : vous devez aussi noter les différentes idées même lorsqu'elles n'ont pas permis de trouver la réponse. Expliquez pourquoi vous avez changé de méthode et ce qui vous a fait avancer, etc.

20 Histoire de climat

Il existe une correspondance de type affine entre les mesures en degré Fahrenheit et celles en degré Celsius définie par 32 °F ↦ 0 °C et 212 °F ↦ 100 °C.
Voici le relevé des températures maximales à Londres (en °F) et à Stuttgart (en °C) au mois d'août 2010.

Londres (en °F)	75	74	74	68	73	71	71	76	77
Stuttgart (en °C)	28	22	21	22	17	17	23	22	23

Londres (en °F)	67	73	69	64	65	73	78	70	71	71	71
Stuttgart (en °C)	26	26	20	21	22	20	18	19	20	22	26

Londres (en °F)	76	74	71	75	65	70	65	69	67	67	69
Stuttgart (en °C)	29	30	24	21	21	28	24	16	17	13	16

Dans quelle ville le climat a-t-il été le plus tempéré durant cette période ? Justifiez.

21 Un esprit calculateur

Dominique, élève de première S se désole en examinant ses résultats. Sur les huit premiers contrôles, sa moyenne est exactement de 9.
Le professeur qui l'encourage, lui indique que pour le bilan annuel il lui reste encore deux devoirs en classe.
Il loue ses efforts réguliers et lui signale que l'écart-type sur ses huit premiers contrôles est exactement de 3.

L'élève se fixe alors les objectifs suivants :
- atteindre 10 de moyenne au bilan annuel ;
- ne pas trop dégrader la régularité de ses résultats en augmentant au maximum de 3 la variance.

Le professeur note ses copies en points entiers.
Quelles notes doit viser Dominique à ses deux derniers contrôles pour atteindre ces objectifs ?

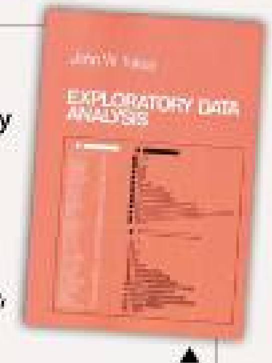
Eux aussi, ils ont cherché avant de trouver !

Chercheurs d'hier

→ Voici quelques mathématiciens importants qui ont travaillé dans le domaine des probabilités et statistiques.

John Wilder Tukey
(1915-2000)

Mathématicien et statisticien américain, John Wilder Tukey a travaillé à la théorie des nombres et à la modélisation informatique en développant des théories sur l'analyse des données. Il est à l'origine du mot *bit* (*binary digit*), unité élémentaire d'information ne prenant que deux valeurs, 0 et 1. Pour Tukey, la statistique est tout à la fois une science, un art et une technique pour inférer du particulier au général. L'inventeur du diagramme en boîte déclarait en 1977 : « The greatest value of a picture is when it forces us to notice what we never expected to see. »



▲ Dans *Exploratory Data Analysis*, publié en 1977, Tukey présente notamment le principe de la « boîte à moustaches », ou diagramme en boîte.

Sur le Web
https//www.bibmath.net/quotidien/statistiques.php3

Utiliser sa calculatrice

→ Pour calculer les principaux paramètres

TP 22 Étudier une série de notes

Voici les notes des élèves d'une classe de 1re S lors d'un devoir.

1. Faites afficher les principaux paramètres de la série.

2. Représentez la série par un diagramme en boîte.

12 – 8 – 6 – 7 – 12 – 13 – 4 – 9 – 13 –
18 – 9 – 14 – 12 – 19 – 13 – 9 – 12 – 6
– 17 – 11 – 17 – 13 – 16 – 17 – 14 – 19
– 7 – 15 – 12 – 10 – 16 – 13 – 13 – 14

Avec une Casio

1. Sélectionnez le menu STAT.

- Entrez les données dans List1.
- Sélectionnez l'icône CALC (F2).
- Utilisez l'instruction SET (F6).

1VAR 2VAR REG SET

Puis choisissez :
– la liste des données : List1 ;
– la fréquence : 1 (données une à une).
Validez par EXE .
Pour l'affichage des paramètres : 1VAR (F1), puis utilisez la touche de direction ▼.

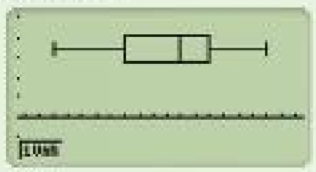
2. Revenir à la liste initiale par EXIT .

- Sélectionnez l'icône GRPH (F1).
- Utilisez l'instruction SET (F6) pour définir le type de graphique.

GPH1 GPH2 GPH3 SEL SET

Puis choisissez :
– le type de graphique : Box (F2) ;
– la liste : List1 ;
– la fréquence : 1.
Validez par EXE .

- Pour afficher le diagramme en boîte, appuyez sur F1 (icône GPH1).

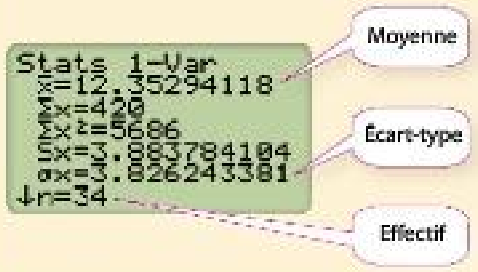


Avec une TI

1. Appuyez sur stats puis, dans le menu EDIT, sélectionnez l'option 1.

- Entrez les données dans L1.
- Appuyez sur stats puis, dans le menu CALC, sélectionnez l'option 1.

- Indiquez la liste utilisée : 2nde 1 (L1). Validez par entrer .

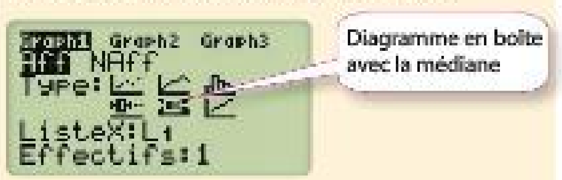


- Pour accéder aux autres paramètres, utilisez la touche de direction ▼.

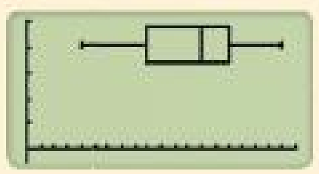
2. Appuyez sur 2nde f(x) (graph stats).

- Validez l'option 1 en appuyant sur entrer .
- Activez l'affichage ; choisissez l'icône du diagramme en boîte (avec la médiane) puis appuyez sur entrer .

Choisissez la liste L1, puis les effectifs 1.



- Appuyez sur fenêtre ; réglez X de 0 à 20.
- Appuyez sur graphe .



Utiliser un tableur

→ Pour simuler un effet de structure

TP 23 Étudier une répartition de salaires

COMPÉTENCES

TICE	Mathématiques
■ Utiliser un tableur	■ Calculer une moyenne
■ Définir un protocole	■ Mettre en inéquation
■ Réaliser et interpréter une simulation	■ Résoudre une inéquation

Deux entreprises A et B emploient deux types de personnel : des cadres et des ouvriers.
Dans l'entreprise A, il y a quatre cadres et seize ouvriers ; le salaire moyen des cadres est 3 090 € et celui des ouvriers 1 740 €.
L'entreprise B emploie soixante personnes et les salaires moyens des cadres et des ouvriers sont inférieurs respectivement de 150 € et 90 € à ceux de l'entreprise A. Cependant, le directeur financier de B annonce que le salaire moyen sur l'ensemble de ses employés est supérieur à celui dans l'entreprise A. Est-ce possible ?

1. Simuler avec un tableur

A. Mise au point d'un protocole

On désigne par S_A et S_B les salaires moyens et M_A et M_B les masses salariales dans les entreprises A et B.

a) Quel est le salaire moyen S_A ?

b) Justifiez que le problème revient à comparer la masse salariale $M_B = 60S_B$ avec $60S_A$.

B. Simulation suivant la structure de l'entreprise B

a) Ouvrez une feuille de calcul et complétez-la pour obtenir la présentation suivante.

	A	B	C	D
1	Entreprise B Cadres	Entreprise B Ouvriers	Masse salariale M_B	Comparaison : M_B . $60S_A$
2	1			

outil 9

Utilisez les formules qui conviennent pour remplir les cellules B2, C2 et D2.

b) Étendez ces formules vers le bas pour obtenir les masses salariales M_B suivant la répartition du personnel dans l'entreprise B.
Existe-t-il des répartitions pour lesquelles l'affirmation du directeur financier semble vraie ?
Concluez.

2. Démontrer

On note x le nombre de cadres dans l'entreprise B.

a) Traduisez par une inéquation la condition $S_B > S_A$.

b) Résolvez cette inéquation puis concluez.

c) À partir de quel pourcentage de cadres, la structure du personnel de l'entreprise B permet-elle d'obtenir un salaire moyen supérieur à celui de l'entreprise A.

Commentaire

L'influence de la répartition du personnel (cadre/ouvrier) sur le calcul du salaire moyen dans l'entreprise est appelé « effet de structure ».

Entraînement

Compléments numériques

DE TÊTE

24 Voici pour les quinze premiers jours de septembre 2010 le relevé des précipitations P et celui des températures maximales T pour une station météo de la ville de Lille.

Précipitations (en mm)	0,2	0,2	0,2	0	0	2,4	14,9
Température (en °C)	19	20	21	20	21	21	22

Précipitations (en mm)	16,5	0,4	0,2	2	0,4	0	5,1	0
Température (en °C)	16	21	20	24	20	20	19	18

1. Indiquez la médiane et les quartiles de la série des précipitations.

2. Même question avec la série des températures.

3. Quelle est à 1 mm près la précipitation moyenne sur cette période ?

4. Quelle est, à 1 °C près, la température maximale moyenne durant cette période ?

25 Sur l'écran d'une calculatrice sont affichés les résultats suivants obtenus lors du traitement d'une série de notes.

```
1-Variable
x̄   =11
Σx  =132
Σx² =1560
```

1. Quel est l'effectif ?

2. Quelle est la moyenne des carrés ?

3. Quel est l'écart-type de la série ?

UTILISER LA NOTATION Σ

La notation $\sum_{i=1}^{p} a_i$ remplace l'écriture $a_1 + a_2 + \ldots + a_p$.

Par exemple : $\sum_{i=1}^{5} i^3 = 1^3 + 2^3 + 3^3 + 4^3 + 5^3$

26 Explicitez puis calculez les sommes suivantes :

a) $\sum_{i=0}^{8} (2i+1)$ b) $\sum_{k=1}^{10} k$ c) $\sum_{j=0}^{6} \left(\frac{1}{2}\right)^j$

27 Écrivez les sommes suivantes en utilisant Σ :

a) $S = 1 + 5 + 5^2 + \ldots + 5^p$ b) $T = 1^2 + 2^2 + 3^2 + \ldots + n^2$

c) $U = 10^{-1} + 10^{-2} + 10^{-3} + \ldots + 10^{-p}$

28 Écrivez les sommes suivantes en utilisant Σ :

a) $S = \sqrt{2} + \sqrt{3} + \ldots + \sqrt{25}$ b) $T = 1 + 3 + 5 + \ldots + 251$

c) $U = 2 + 4 + 8 + \ldots + 1\,024$

29 X est la suite des dix premiers nombres impairs.

1. Écrivez leur moyenne $\overline{x}$ avec Σ. Combien vaut $\overline{x}$?

2. Écrivez leur variance V avec Σ. Combien vaut V ?

3. On ajoute 1 à chaque valeur, on obtient la série X'. Que valent sa moyenne $\overline{x}'$ et sa variance V' ?

PARAMÈTRES D'UNE SÉRIE

30 Voici, suivant les activités proposées, les tarifs en euros de deux agences qui organisent des week-ends de canyoning sur différents sites.

- **Agence A :** 33-65-35-60-45-60-62-65-65-55-130-40-100-65-40-35-120-55-80-35-80.
- **Agence B :** 34-70-35-55-90-50-50-35-105-27-50-112-48-55-112-70-35-105-50.

1. Représentez sur un même graphique chacune des séries par un diagramme en boîte.

2. Dites à quelles agences correspondent les informations suivantes :

a) Trois quarts des tarifs dépassent 35 €.

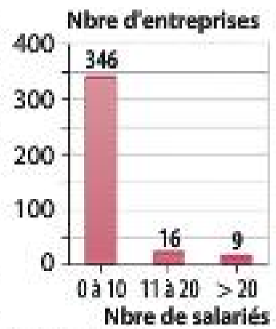
b) La moitié des tarifs ne dépasse pas 50 €.

c) Un quart des tarifs dépasse 65 €.

d) La moitié des tarifs est entre 40 € et 65 €.

3. Commentez la dispersion des deux séries à partir de la comparaison des deux diagrammes.

31 La répartition des entreprises du bâtiment suivant leur type (0 à 10 salariés ; 11 à 20 ; plus de 20) est donnée sur le diagramme où les effectifs sont indiqués en milliers.



(Source : Capeb.)

Les entreprises employant de 11 à 20 salariés ont un chiffre d'affaires moyen de 1,125 million d'euros ; elles réalisent 14 % du chiffre d'affaires total (CAT) des entreprises du bâtiment. Les entreprises de plus de 20 salariés réalisent 38 % du CAT.

1. Calculez le CAT arrondi à 1 million d'euros près.

2. Quel est le chiffre d'affaires moyen (à 1 millier d'euros près) des entreprises de plus de 20 salariés ?

3. Calculez le chiffre d'affaires moyen pour les entreprises de 0 à 10 salariés.

4. Comparez les différents chiffres d'affaires moyens.

32 Une coopérative laitière fabrique un fromage qui doit contenir, selon l'étiquette, 50 % de matière grasse. Un organisme de contrôle de qualité prélève 100 fromages afin d'analyser leur taux de matière grasse.

Voici les résultats de l'analyse :

Taux	[45 ; 47[	[47 ; 49[	[49 ; 51[	[51 ; 53[	[53 ; 55[
Effectif	6	25	45	21	3

1. Calculez une valeur approchée du taux moyen $\overline{t}$ et de l'écart-type s.

2. Une production de fromage peut être vendue sous l'appellation « 50 % de matière grasse » si les deux conditions suivantes sont remplies :

- 50 appartient à l'intervalle $[\overline{t} - 0{,}3 ; \overline{t} + 0{,}3]$.
- Plus de 90 % des fromages analysés appartiennent à l'intervalle $[\overline{t} - 2s ; \overline{t} + 2s]$.

Que pensez-vous de la production de cette coopérative ?

33 Une entreprise A emploie 5 techniciens et 20 ouvriers. Une entreprise B emploie 42 personnes réparties entre techniciens et ouvriers. Pour chacune des entreprises le salaire moyen des techniciens est 2 760 € et celui des ouvriers 1 680 € mais dans l'entreprise B le salaire moyen est inférieur de 36 € à celui de A.

1. Calculez le salaire moyen dans l'entreprise A.

2. a) Déterminez la répartition des employés dans l'entreprise B.

b) Comparez les fréquences des catégories techniciens dans les entreprises A et B. Le résultat était-il prévisible ?

→ *Apprendre à chercher, exercice 18.*

34 **Sensibilité aux valeurs extrêmes** TICE

Téléchargez sur le site Nathan le fichier Excel de la population du département du Nord (62).

L'objectif est d'examiner l'évolution des différents indicateurs lorsqu'on modifie la série en éliminant quelques valeurs extrêmes.

1. a) Ouvrez la feuille des communes et copiez en colonne A d'une nouvelle page, la colonne J donnant la population totale de chacune des 895 communes.

b) Triez la série obtenue dans l'ordre croissant.

2. a) Complétez la feuille de calcul où P désigne la population d'une commune.

	A	B	C	D	E	F	G	H	I	J
1	Population		Critère	Q1	Me	Q3	Q3 - Q1		Moyenne	Écart-type
2	16									
3	18		Totalité							
4	18									
5	32		P<50 000							
6	33									
7	35		50<P<30 000							
8	38									
9	42		50<P<15 000							

b) Comment évoluent la médiane et les quartiles ? En est-il de même pour la moyenne et l'écart-type ?

c) Présentez la répartition des communes du Nord en utilisant la définition de la médiane et des quartiles.

d) Pour résumer la série de la population des communes du Nord est-il préférable d'utiliser le couple (moyenne ; écart-type) ou le couple (médiane ; écart interquartile) ? Justifiez votre choix.

35 **Calcul de la variance** ALGORITHMIQUE

Avec Algobox, Chloé a écrit l'algorithme suivant afin de calculer la variance d'une série de notes.

```
VARIABLES
  x EST_DU_TYPE LISTE
  i EST_DU_TYPE NOMBRE
  n EST_DU_TYPE NOMBRE
  somme EST_DU_TYPE NOMBRE
  somme_carrés EST_DU_TYPE NOMBRE
  v EST_DU_TYPE NOMBRE
DEBUT_ALGORITHME
  i PREND_LA_VALEUR 1
  somme PREND_LA_VALEUR 0
  somme_carrés PREND_LA_VALEUR 0
  AFFICHER "Saisissez l'effectif total"
  LIRE n
  POUR i ALLANT_DE 1 A n
    DEBUT_POUR
    AFFICHER "Saisissez la valeur numéro"
    AFFICHER i
    LIRE x[i]
    somme PREND_LA_VALEUR somme+x[i]
    somme_carrés PREND_LA_VALEUR somme_carrés+pow(x[i],2)
    FIN_POUR
  v PREND_LA_VALEUR somme_carrés/n-pow(somme/n,2)
  AFFICHER "La variance est : "
  AFFICHER v
FIN_ALGORITHME
```

On se propose de vérifier cet algorithme en le testant pas à pas, sur la série de notes : 12-16-7-11-9.

1. Quelle formule permet ici de calculer la variance ?

2. a) Complétez le tableau indiquant la valeur des variables lors des différentes étapes.

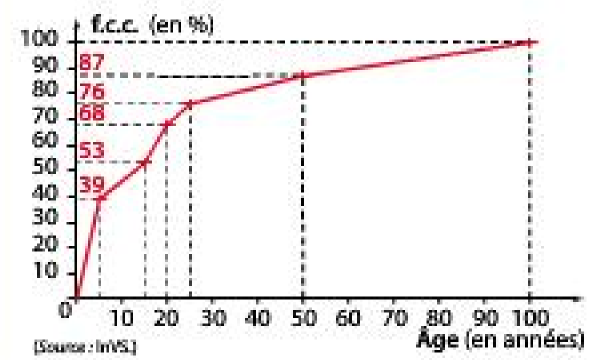
Variable	Initialisation	Étape 1	Étape 2	…
i	1	1	2	…
$x[i]$	non déclaré	12	16	…
somme	0	12	28	…
somme_carrés	0	144	400	…

b) Quelle valeur sera affichée pour V ?

3. Déterminez à la calculatrice la variance de la série. Comparez avec le résultat obtenu. Concluez.

CHOISIR UN RÉSUMÉ ADAPTÉ

36 La répartition des cas d'infections à méningocoques suivant l'âge, en France en 2007, est représentée par la courbe des fréquences cumulées croissantes.



(Source : InVS.)

1. Graphiquement, indiquez une valeur approchée de la médiane et des quartiles.

2. a) Dressez le tableau des fréquences.

b) Déduisez-en une estimation de la moyenne et de l'écart-type.

3. Quels indicateurs vous semblent pertinents pour présenter la répartition de ces cas ?

37 Voici les résultats d'un test sur la durée de vie (en milliers d'heures) d'un lot de composants électroniques.

d	< 4	< 6	< 8	< 10	< 12	< 16
f.c.c. en %	9	24	46	74	93	100

1. Tracez la courbe des f.c.c.

2. a) Déterminez graphiquement une valeur approchée de la médiane et des quartiles.

b) Calculez une estimation de la durée de vie moyenne et de l'écart type.

3. Le fabricant indique que 2 composants sur 3 de ce type ont une durée de vie entre 5 000 et 11 000 h.
Quel couple d'indicateurs permet de confirmer ceci.

38 La répartition des salaires mensuels bruts, exprimés en milliers d'euros, dans une petite entreprise est indiquée dans le tableau :

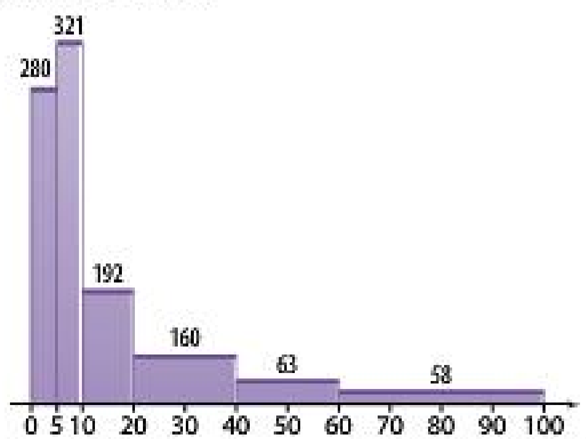
Salaire	[1 ; 1,5[	[1,5 ; 2[	[2 ; 2,5[	[2,5 ; 3[	[3 ; 3,5[	[3,5 ; 4[
Effectif	11	14	4	3	11	7

1. Tracez la courbe des f.c.c. puis déduisez-en le diagramme en boîte de la série.

2. Calculez une estimation du salaire moyen et de l'écart-type.

3. Quels indicateurs choisiriez-vous pour résumer cette série ? Justifiez.

39 La répartition des « petits musées » suivant le nombre de visiteurs en 2008 est représentée par l'histogramme ci-dessous.



1. Tracez la courbe des fréquences cumulées croissantes. (Les f.c.c. seront arrondies à 1 % près.)

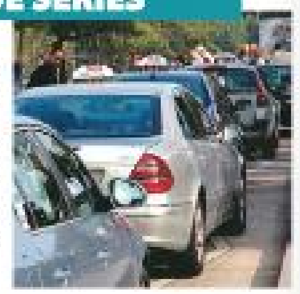
2. Déduisez-en le diagramme en boîte de la série.

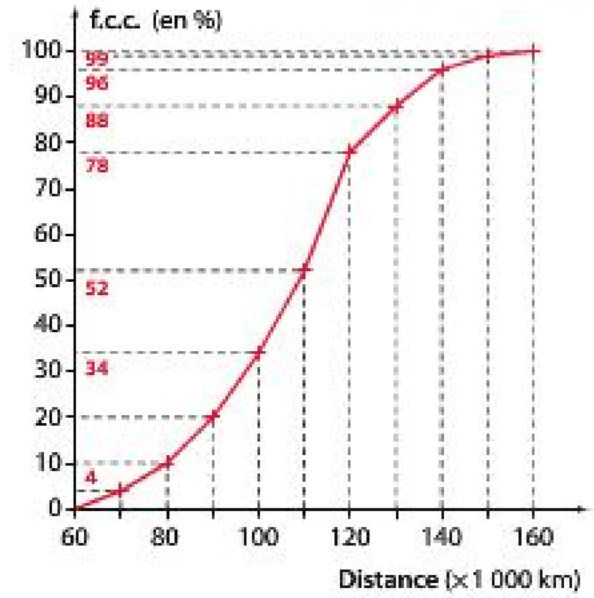
3. Calculez une estimation du nombre moyen de visiteurs et l'écart-type.

4. Quel couple d'indicateurs vous semble le plus pertinent pour traduire la répartition de ces musées ?

COMPARAISON DE SÉRIES

40 Une compagnie de taxis parisiens a relevé les distances *d* (en milliers de km) parcourues par ses véhicules avant qu'elle ne s'en sépare. Voici son diagramme des fréquences cumulées croissantes.





Une compagnie de taxis londonienne a relevé de même les distances *d'* (en milliers de miles) parcourues par ses véhicules avant qu'elle ne s'en sépare.

d'	[70 ; 80[	[80 ; 90[	[90 ; 100[	[100 ; 110[	[110 ; 120[
Effectif	4	5	11	15	20

d'	[120 ; 130[	[130 ; 140[	[140 ; 150[	[150 ; 160[
Effectif	18	12	7	2

1. a) Dressez le tableau des fréquences des distances *d*.

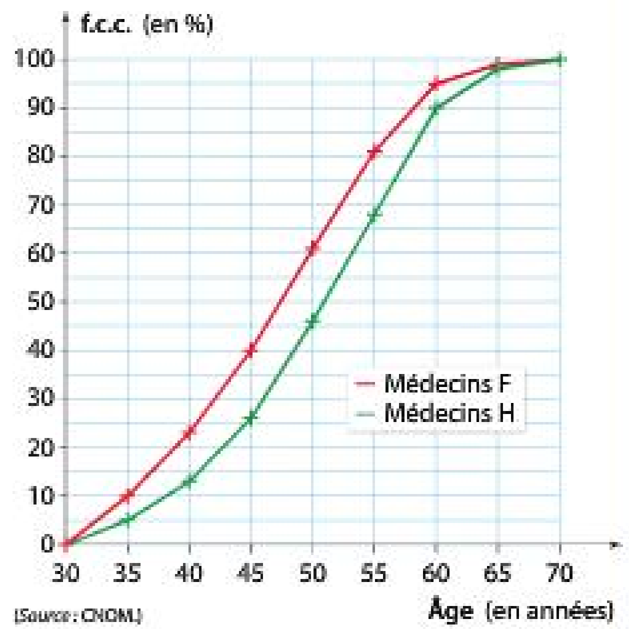
b) Déduisez-en une estimation de la distance moyenne $\overline{d}$ et de l'écart-type *s*.

2. Calculez une approximation de la distance moyenne $\overline{d'}$ et de l'écart-type *s'*.

3. a) Sachant que 1 mile vaut 1 609 m, comparez les moyennes et les écarts-types des deux séries.

b) Commentez l'état de ces deux parcs de taxis.
NB : les calculs seront arrondis à 1 millier de km près.

41 La répartition des médecins en activité en 2008 en France métropolitaine suivant le sexe et l'âge est indiquée sur le graphique.
En vert : la courbe des f.c.c. des médecins hommes.
En rouge : la courbe des f.c.c. des médecins femmes.



1. a) Pour chacune des séries indiquez une valeur approchée des paramètres suivants : Q_1, Me, Q_3.

b) À partir de ces paramètres construisez le diagramme en boîtes superposées.

2. a) Calculez une estimation de l'âge moyen et de l'écart-type de chacune des populations de médecins.

b) Estimez à l'aide du graphique les pourcentages de chaque population dont l'écart d'âge à la moyenne ne dépasse pas l'écart-type.

3. Comparez à partir des résultats précédents la répartition des médecins hommes et femmes.

42 La répartition en millions de spectateurs de 1992 (année 1) à 2009 (année 18) pour chacun des types de films est donnée par le tableau suivant.

Année	1	2	3	4	5	6	7	8	9
Films français	41	47	35	46	51	52	47	50	47
Films américains	67	76	76	70	74	78	108	83	103
Autres	8	10	13	14	12	19	16	21	16

Année	10	11	12	13	14	15	16	17	18
Films français	77	64	61	75	64	84	65	86	74
Films américains	87	92	90	94	80	83	87	82	100
Autres	23	28	22	27	32	22	26	22	27

On se propose de dresser le tableau indiquant le pourcentage des entrées, réalisé par chacun des types de films, français et américains.

Année	1	2	3	…	18
Films français : part en %	35	…	…	…	37
Films américains : part en %	58	…	…	…	50

1. a) Indiquez les calculs qui permettent de compléter la première colonne et la dernière colonne.

Note
Les pourcentages sont arrondis à 1 % près.

b) Complétez le tableau des parts respectives des entrées pour les films français et américains.

2. Représentez sur un même graphique les parts des films français et américains suivant les années.

3. Comparez les parts respectives des deux cinémas. Quelle semble être l'évolution des écarts constatés ?

43 **Implication et quantificateurs** LOGIQUE

Dans chacun des cas, indiquez si l'implication (P) ⇒ (Q) est vraie ou fausse. Justifiez.
On donne une série statistique comprenant *n* valeurs $x_1, x_2 \dots x_n$.

a) (P) : $\sum_{i=1}^{n} x_i = 0$;
(Q) : Il existe un entier *i* tel que $x_i = 0$.

b) (P) : $x_1 \times x_2 \times \dots \times x_n = 0$;
(Q) : pour tout entier *i* tel que $1 \leq i \leq n$, $x_i = 0$.

c) (P) : $\sum_{i=1}^{n} x_i^2 = 0$;
(Q) : pour tout entier *i* tel que $1 \leq i \leq n$, $x_i = 0$.

d) (P) : $x_1^2 \times x_2^2 \times \dots \times x_n^2 = 0$;
(Q) : Il existe un entier *i* tel que $x_i = 0$.

POUR ALLER PLUS LOIN

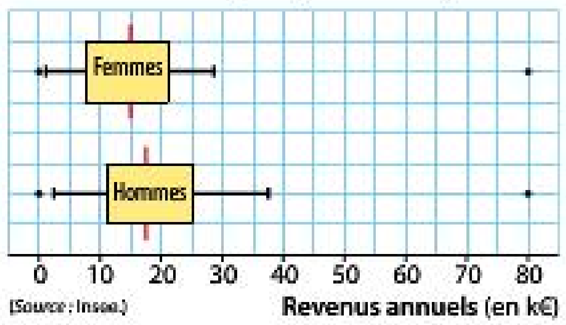
Avec cette définition, qui dépasse le cadre du programme, vous pouvez résoudre **les exercices 44 et 45**.

Définition. Les déciles D_1, D_2, …, D_9 sont les valeurs qui partagent la population en dix groupes représentant 10 % de la population. Ils sont analogues aux quartiles lors du partage de la population en tranches de 25 %.

44 Le tableau donne la répartition des revenus annuels salariaux (en euros), par sexe, en France, en 2007.

	D_1	Q_1	Me	Q_3	D_9
H	2 872	11 340	17 748	25 881	37 259
F	1 770	7 645	14 472	21 653	28 236

La situation est illustrée par le diagramme en boîtes superposées où les extrémités des pattes correspondent aux déciles D_1 et D_9. On indique par des points les valeurs inférieures à D_1 ou supérieures à D_9.



1. Traduisez les renseignements concernant D_1 et D_9.

2. Les propositions sont-elles vraies ou fausses ?

a) Trois quarts des hommes ont un salaire mensuel d'au moins 945 €.

b) Trois quarts des femmes ont un salaire mensuel d'au plus 1 805 €.

3. L'inégalité entre les salaires des hommes et des femmes peut être mesurée, en pourcentage, par l'écart relatif E entre les salaires.

$$E = \frac{\text{salaire homme} - \text{salaire femme}}{\text{salaire homme}} \times 100.$$

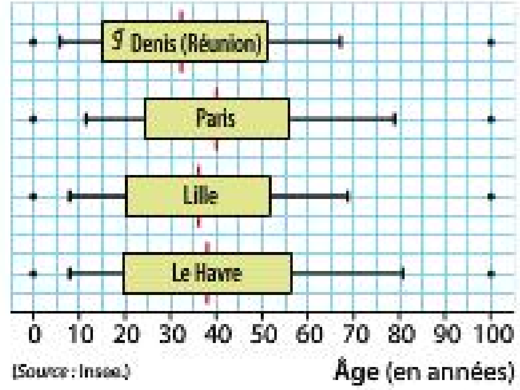
a) Calculez cet écart pour D_1, Q_1, Me, Q_3, D_9.

b) Commentez l'évolution de cet écart en fonction du montant des salaires.

45 La répartition de la population suivant l'âge, dans quatre villes françaises, est représentée par un diagramme en boîtes superposées dont les extrémités sont les déciles.

1. À partir des indicateurs lus sur le diagramme, comparez la répartition des habitants de ces villes en fonction de l'âge.

2. Présentez une synthèse de vos comparaisons.



La rue du maréchal-Leclerc à Saint-Denis de la Réunion.

ROC Restitution organisée de connaissances

46 **Prérequis.** X désigne une série statistique (x_i) de N données, alors sa moyenne et sa variance sont définies par :

$$\overline{x} = \frac{\sum_{i=1}^{N} x_i}{N} \text{ et } V_x = \frac{\sum_{i=1}^{N} (x_i - \overline{x})^2}{N}.$$

1. Démonstration

On considère la nouvelle série statistique Y de même effectif que X dont les valeurs sont définies par :

$$y_i = ax_i \quad (a \in \mathbb{R}).$$

Prouvez que sa moyenne et sa variance sont telles que :

$$\overline{y} = a\overline{x} \quad \text{et} \quad V_y = a^2 V_x.$$

2. Application

Une enquête a été réalisée dans divers pays en octobre 2010 sur le prix du litre de gazole.

	Prix moyen	Écart-type
France	1,17 €	0,04 €
Japon	112,50 ¥	6 ¥
États-Unis	0,82 $	0,03 $

(*Source* : FIA.)

La parité des monnaies est alors la suivante :

1 $ ⟼ 0,72 € ; 1 € ⟼ 113,90 ¥.

Dans quel pays :
– le prix moyen du gazole est-il le plus bas ?
– la dispersion des prix est-elle la plus forte ?

47 **Les boîtes autour de la moyenne** ALGORITHMIQUE

On donne une série statistique dont on a calculé la moyenne m et l'écart-type s. L'algorithme suivant, écrit avec Algobox a pour but de déterminer le pourcentage de valeurs dans l'intervalle $[m - s ; m + s]$.

```
1    VARIABLES
2      x EST_DU_TYPE LISTE
3      i EST_DU_TYPE NOMBRE
4      n EST_DU_TYPE NOMBRE
5      m EST_DU_TYPE NOMBRE
6      s EST_DU_TYPE NOMBRE
7      compte EST_DU_TYPE NOMBRE
8      pourcentage EST_DU_TYPE NOMBRE
9    DEBUT_ALGORITHME
10     i PREND_LA_VALEUR 1
11     compte PREND_LA_VALEUR 0
12     pourcentage PREND_LA_VALEUR 0
13     AFFICHER "saisissez l'effectif total n"
14     LIRE n
15     AFFICHER "saisissez la moyenne m"
16     LIRE m
17     AFFICHER "saisissez l'écart type s"
18     LIRE s
19     POUR i ALLANT_DE 1 A n
20       DEBUT_POUR
21       AFFICHER "saisissez la valeur numéro"
22       AFFICHER i
23       LIRE x[i]
24       SI (x[i]>=m-s ET x[i]<=m+s) ALORS
25         DEBUT_SI
26         compte PREND_LA_VALEUR compte+1
27         FIN_SI
28       FIN_POUR
29     pourcentage PREND_LA_VALEUR (compte/n)*100
30     AFFICHER pourcentage
31   FIN_ALGORITHME
```

1. Que signifient les instructions de la ligne 24 à 27 ?

2. Testez cet algorithme pour la série de notes :
14-17-6-8-13-8-10-13-12-
16-12-7-13-6-10-12-14-20-
4-19-1-18-6-8-13-13-15.

Indication

Les valeurs de m et de s seront obtenues à la calculatrice et arrondies à 0,1 près.

3. a) Modifiez l'algorithme afin qu'il calcule le pourcentage de valeurs dans l'intervalle $[m - ks ; m + ks]$ où k est un coefficient positif fixé.

b) Testez-le sur la série indiquée avec $k = 0{,}5$, $k = 1$ puis $k = 1{,}5$.

4. Représentez la situation par un diagramme en boîtes superposées associées à chacun des intervalles $[m - ks ; m + ks]$ lorsque $k = 0{,}5$, $k = 1$ et $k = 1{,}5$. Que constatez-vous ?

Prendre toutes les initiatives

48 **Efficacité d'un médicament**

L'objet de l'étude statistique est la mesure de la glycémie à jeun G exprimée en $g \cdot L^{-1}$. La norme est l'intervalle [0,7 ; 1,1].

Un laboratoire qui teste un médicament hypoglycémiant donne à un groupe A le médicament et à un groupe B un placebo (il a la même apparence mais ne contient pas de produit actif).

Les deux groupes A et B ont la même moyenne 1,6 $g \cdot L^{-1}$ et pour écart-type 0,3 $g \cdot L^{-1}$.

Voici les résultats des analyses après un mois de test :

G	[0,8 ; 0,9[	[0,9 ; 1[	[1 ; 1,1[	[1,1 ; 1,2[	[1,2 ; 1,3[
A	7	15	34	20	8
B	1	1	3	15	25

G	[1,3 ; 1,4[	[1,4 ; 1,5[	[1,5 ; 1,6[	[1,6 ; 1,7[	[1,7 ; 1,8[
A	8	7	1	0	0
B	15	14	13	7	6

Comparez ces résultats et indiquez si le médicament vous semble efficace.

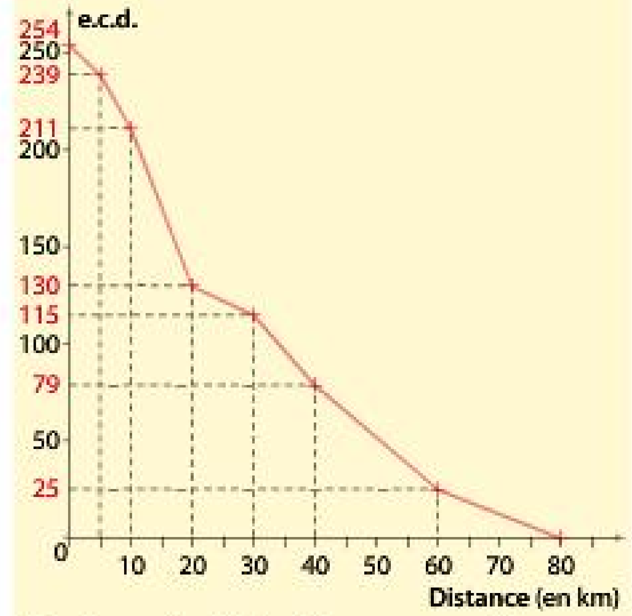
49 **Une propriété de la moyenne et de la variance**

X est une série statistique contenant N données (x_i) de moyenne $\overline{x}$ et de variance V_x. On considère la série Y de même effectif dont les valeurs sont $y_i = x_i - \overline{x}$.

Que valent sa moyenne $\overline{y}$ et sa variance V_y ?

50 **Résumé d'une enquête**

Une enquête sur la distance du trajet entre le domicile et le lieu de travail des salariés d'une entreprise est résumée par la courbe des effectifs cumulés décroissants.



Représentez la série des distances par un diagramme en boîte. Interprétez ce diagramme pour résumer l'enquête.

Approfondissement

Compléments numériques

51 Un café à bonne température

Une personne verse du café (à 80 °C) dans une tasse.
Le tableau indique l'évolution de la température y suivant le temps t de refroidissement.

Temps (en min)	0	2	4	6	8	10	12	17
Température (en °C)	80	60	47	38	32	28	25	22

On se propose d'utiliser ces résultats pour estimer dans les mêmes conditions :

- la température du café au bout d'un quart d'heure ;
- le temps d'attente nécessaire pour que la personne puisse déguster ce café à 40 °C.

1. Placez dans un repère orthogonal les points définis par le tableau puis joignez les points successifs.

Principe : on considère qu'une approximation de la loi de refroidissement est obtenue en joignant chacun des points définis dans le tableau par un segment de droite.

2. a) Déterminez la fonction affine représentée par le segment [AB] avec A(12 ; 25) et B(17 ; 22).

b) Déduisez-en une estimation de la température du café au bout d'un quart d'heure.

3. a) Déterminez la fonction affine représentée par le segment qui convient à une température de 40 °C.

b) Déduisez-en une estimation du temps d'attente nécessaire pour une dégustation du café à 40 °C.

Commentaire

La méthode employée s'appelle « l'interpolation linéaire ». Elle peut être utilisée pour déterminer une valeur approchée des quartiles ou de la médiane à partir de la courbe des fréquences cumulées.

52 Effet d'un changement affine

X désigne une série statistique dont on connaît la distribution des fréquences $(x_i ; f_i)$ pour $1 \leqslant i \leqslant p$, la moyenne $\overline{x}$ et la variance V_x.

On considère la série Y dont la distribution des fréquences est $(y_i ; f_i)$ avec pour $1 \leqslant i \leqslant p$:

$$y_i = ax_i + b \quad (a \in \mathbb{R}^*, b \in \mathbb{R}).$$

On note $\overline{y}$ et V_y sa moyenne et sa variance.

1. Démontrez que $\overline{y} = a\overline{x} + b$.

2. Démontrez que $V_y = a^2 V_x$.

3. Application

La série des températures maximales relevées à Londres en septembre 2010 a pour moyenne 67 °F et pour écart-type 5 °F.

La correspondance entre les mesures x en degrés Fahrenheit et les mesures y en degrés Celsius x est :

$$y = \frac{5x - 160}{9}.$$

Déduisez-en la moyenne et l'écart-type de la série lorsque les températures sont exprimées en °C.

53 La stratégie de la péréquation

Une entreprise décide de modifier sa politique salariale. Les salaires sont exprimés en milliers d'euros. Le salaire brut moyen est $\overline{x} = 2{,}4$, l'écart-type $s = 1{,}2$ et les différents salaires se répartissent entre 1,2 et 6.

La direction envisage une transformation affine de la série des salaires afin d'obtenir une moyenne de 2,5.

Elle étudie trois stratégies :

Piste 1 : on ajoute 100 euros à tous les salaires.

Piste 2 : on multiplie tous les salaires par $k = \frac{25}{24}$.

Piste 3 : on opère une transformation $x \mapsto y = ax + b$ de façon que la moyenne soit 2,5 et l'écart-type 1,1.

Dans les justifications, on utilisera les formules sur la moyenne et la variance vues à l'exercice 52.

1. On s'intéresse aux pistes 1 et 2.

a) Justifiez que le salaire moyen est bien 2,5.

b) Calculez alors les écarts-types correspondants.

c) Quels sont les effets sur les salaires les plus bas ? sur les plus élevés ?

2. On s'intéresse à la piste 3.

a) Déterminez les coefficients a et b.

b) Dans quel intervalle seront alors les salaires ?

3. Des trois pistes indiquez celle qui vous paraît rendre plus homogène la répartition des salaires.

54 Électricité issue des combustibles fossiles TICE

On se propose d'étudier l'évolution de la production d'électricité à partir des combustibles fossiles (charbon, pétrole, gaz) de 1989 à 2009 dans trois pays européens (France, Allemagne, Royaume Uni).

On prendra comme origine l'année 1989 (rang 0).

Les fréquences sont exprimées en pourcentage de la production totale dans chaque pays.

(Source : perspective.usherbrooke)

Année	0	1	2	3	4	5	6
FR	9,8	9,2	10,3	8,9	6	5,9	6,2
D	65	66,1	65,8	63,3	64	64,2	63,7
UK	66,1	66,6	67,6	63,9	63,8	66	65,9

Année	7	8	9	10	11	12	13
FR	6,9	6,2	8,3	7,8	7,8	7,2	8,2
D	63,7	62,5	63,9	62,1	62,3	61,8	62,1
UK	66,2	66,7	67,3	68,8	72,3	71,9	72,3

Année	14	15	16	17	18	19	20
FR	8,7	8,5	9,4	8,5	8,8	8,8	8,8
D	62,9	60,5	61,1	60,1	60,2	60,1	59,6
UK	73	74,2	73,1	74,3	74,1	74,1	74,2

1. Complétez une feuille de calcul dans un tableur.

	A	B	C	D
1	Année	FR	D	UK
2	0	9,8	65	66,1
3	1	9,2	66,1	66,6

2. a) Représentez sur le même graphique les fréquences pour chacun des pays.

Aide

Sélectionnez les quatre colonnes ; utilisez un diagramme du type XY avec points reliés et affichage des étiquettes de données.

b) Commentez le sens de variation des fréquences dans chaque pays.

3. La droite passant par le premier et le dernier point de chaque série donne une approximation des données.

On se propose d'estimer la fréquence en 2015 de chaque classe à partir des trois droites.

	A	B	C	D
24	Coefficient directeur			
25				
26	Valeur estimée en 2015			

a) Faites afficher en B24, C24 et D24 le coefficient directeur de chaque droite.

b) Précisez l'expression des fonctions affines représentées par ces droites.

c) Les valeurs estimées en 2015 sont les images par ces fonctions affines du rang de l'année 2015.

Faites afficher les résultats en B26, C26, D26.

Comparez alors les tendances prévisibles d'ici 2015.

55 Condition nécessaire et suffisante LOGIQUE

Dans chacun des cas, indiquez si la condition (Q) est nécessaire et suffisante pour la proposition (P). Justifiez.

On donne une série statistique comprenant n valeurs, $x_1, x_2, \ldots, x_n$, de moyenne $\overline{x}$, de médiane Me et de variance V.

1. (P) : $\overline{x} = 0$; (Q) : pour tout i, $x_i = 0$.

2. (P) : $V = 0$; (Q) : pour tout i, $x_i = \overline{x}$.

3. (P) : Me $= \overline{x}$; (Q) : $\sum_{i=1}^{n} x_i = n\text{Me}$.

Prendre toutes les initiatives

56 Effet d'une absence

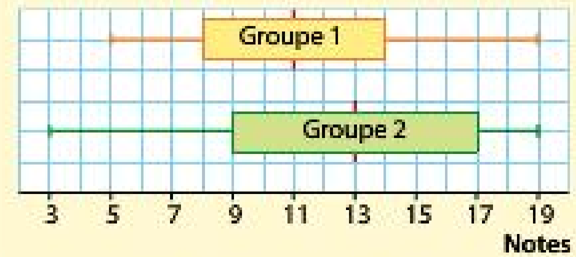
Lors de l'évaluation d'une séance de travaux pratiques le professeur de physique relève les résultats suivants d'un groupe d'élèves : présents 15, absent 1, moyenne 11, écart-type 4.

L'élève absent obtient 19 lors d'une séance de rattrapage.

Quelles sont les nouvelles valeurs de la moyenne et de l'écart-type de ce groupe ?

57 Un jeu de boîtes

Le diagramme illustre la dispersion des résultats de deux groupes de 15 élèves d'une classe de 1re S. Les boîtes sont centrées sur la moyenne $\overline{x}$ et la longueur est celle de l'intervalle $[\overline{x} - s ; \overline{x} + s]$.



Construisez la boîte représentant les résultats de la totalité de la classe.

58 Une simulation de la réponse à un QCM

Un QCM contient cinq questions avec pour chacune quatre réponses proposées dont une seule est exacte. Chaque bonne réponse rapporte 1 point, une fausse 0.

Jeff envisage de répondre au hasard à chaque question.

1. a) Créez une feuille de calcul dans un tableur pour simuler cette expérience.

b) Procédez à la simulation des réponses à 100 QCM.

2. a) Calculez la moyenne $\overline{x}$ et l'écart-type s de la série de notes obtenues.

b) Quel est le pourcentage de notes supérieures à 3 ?

c) Quel est le pourcentage de notes dans $[\overline{x} - s ; \overline{x} + s]$?

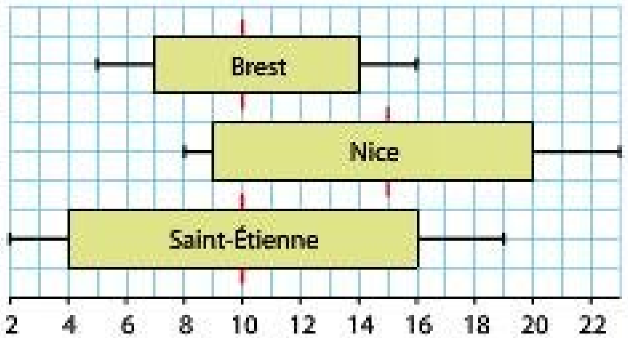
Que pensez-vous de la stratégie de réponse au hasard ?

Travail en autonomie

→ **Les exercices suivants permettent de revoir les principales méthodes de résolution abordées dans le chapitre. Faites ces exercices à votre rythme, en utilisant si besoin les** ***coups de pouce*** **page 381.**

A Un climat tempéré

Le diagramme ci-dessous représente la répartition des températures moyennes (t en °C) relevées chaque mois de l'année 2010, à Saint-Étienne, Nice et Brest.



1. Indiquez pour chacune d'elles les paramètres des séries de relevés des températures moyennes.

2. Comparez ces trois séries. 1

B Des tests contradictoires ?

Voici les résultats des tests de deux jeux vidéos auprès d'un échantillon de 1 600 adolescents.

• Filles	Avis positifs	Avis négatifs
Test du jeu A	102	306
Test du jeu B	89	223

• Garçons	Avis positifs	Avis négatifs
Test du jeu A	467	309
Test du jeu B	67	37

1. Parmi les filles, quel est le jeu le plus apprécié ? 2

2. Même question, à partir du groupe des garçons.

3. On considère l'échantillon complet.

a) Quel est le jeu le plus apprécié ? 3

b) Expliquez le phénomène observé. 4

C L'argent de poche

Une enquête a établi la répartition de l'argent de poche (en €) que reçoivent chaque mois les lycéens de plus de quinze ans.
Les fréquences sont exprimées en pourcentage.

On suppose la répartition uniforme à l'intérieur de chaque classe.

1. Calculez une estimation de :

a) la somme moyenne reçue $\overline{x}$;

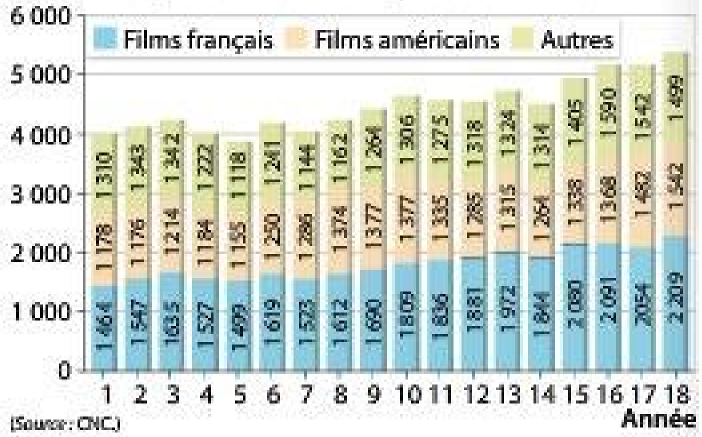
b) de l'écart type s. 5

2. Déterminez le pourcentage de lycéens qui reçoivent une somme comprise entre $\overline{x} - s$ et $\overline{x} + s$. 6

D Au cinéma

Le diagramme en barres empilées indique la répartition des films exploités en France de 1992 (année 1) jusqu'en 2009 (année 18).

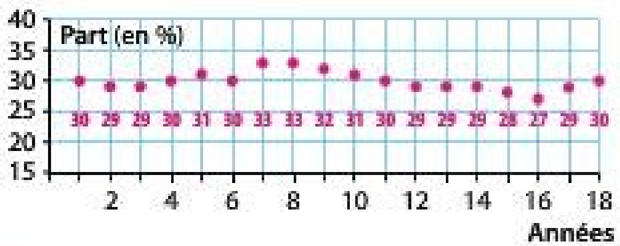
Répartition des films exploités en France



(Source : CNC.)

1. Dressez un tableau donnant la part respective (à 1 % près) des productions française et américaine dans les films exploités en France durant cette période. 7

2. Le graphique représente la part de la production américaine.



a) Représentez la part de la production française par un diagramme du même type.

b) Comparez l'évolution des parts respectives de chaque production durant cette période.

3. Sur le graphique obtenu en **2. a)**, la droite passant par le premier point et le dernier point indique une tendance. Si cette tendance se poursuit, vers quelle année la part de la production française atteindra-t-elle 45 % ? 8