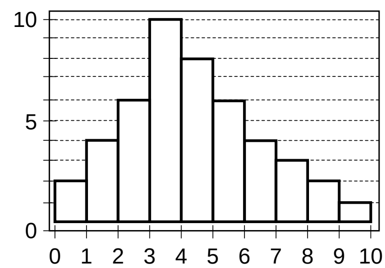
Chapitre

1 Statistique

1 Prérequis : « Pour démarrer » (page 8)

Exercice	Prérequis testés	Réponse	En complément
1	Localiser la classe médiane d'une série.	c	• Rappeler la définition de la médiane, de la classe médiane. • Proposer l'exemple suivant.  Réponse : l'effectif est 46. La classe médiane est [4; 5[.
2	Calculer une valeur approchée de la moyenne d'une série dont les valeurs sont regroupées en classe.	c	Même question avec la série précédente. Réponse : une valeur approchée (au centième) de la moyenne est 4,37.
3	Calculer la moyenne d'une série à partir des moyennes de sous-groupes.	b	• Rappeler la formule vue en seconde. Proposer d'autres questions, par exemple : Quelle devrait être la moyenne des filles pour que la moyenne de la classe soit égale à 15 ?
4	Utiliser les propriétés de linéarité de la moyenne.	c	• Rappeler les propriétés de linéarité de la moyenne. • Envisager d'autres situations : multiplier les notes par 1,5 ; ajouter 2 à toutes les notes.
5	Connaître la sensibilité des paramètres connus (moyenne médiane, étendue) aux valeurs extrêmes.	a	• Vérifier avec la calculatrice. • Proposer d'autres situations pour rappeler l'influence des valeurs extrêmes sur la moyenne, la médiane.

2 Objectifs

• Représenter une série statistique par un diagramme en boîte.
• Comparer deux séries à l'aide de leurs diagrammes en boîte.
• Caractériser une série statistique par des mesures de tendance centrale (moyenne, médiane), de dispersion (étendue, variance, écart-type, écart interquartile) et de position (quartiles, déciles).
• Connaître l'intérêt et la pertinence de ces différentes mesures.
• Apprécier l'influence des valeurs extrêmes et les conséquences d'une transformation affine des données sur les différents paramètres de tendance centrale et de dispersion.

3 Difficultés et erreurs

3.1 Calcul des quartiles

• Si la moyenne, l'écart-type et la variance sont donnés par des formules, la détermination des quartiles est moins immédiate. Elle nécessite un algorithme de classement des valeurs de la série.

→ *Exercices 18 à 20 (page 20)*

• Les valeurs données par les calculatrices ou les tableurs ne sont pas toujours identiques à celles obtenues avec la définition du cours.

Par exemple, les calculatrices Casio ou Texas déterminent les quartiles en deux étapes :

* calcul de la médiane et partage de la série en deux sous-séries ;

* Q1 est la médiane de la première sous-série
Q3 est la médiane de la deuxième sous-série.

→ *Activité 1 (page 16)*

3.2 Confusion entre variance et écart-type

La variance et l'écart-type sont obtenus à partir du même calcul : $\frac{1}{N}\sum_{i=1}^{p} n_i(x_i - \bar{x})^2$.

Les élèves confondent souvent les deux valeurs et donnent l'une à la place de l'autre.

→ *Exercices 29 à 31 (page 21)*

3.3 Différence entre σ_n et σ_{n-1}

Les calculatrices fournissent deux valeurs pour l'écart-type : σ ou σ_n qui correspond à la formule $\sqrt{\frac{1}{N}\sum_{i=1}^{p} n_i(x_i - \bar{x})^2}$ et σ_{n-1} ou s qui correspond à la formule $\sqrt{\frac{1}{N-1}\sum_{i=1}^{p} n_i(x_i - \bar{x})^2}$.

L'écart-type, noté s dans le cours (comme le demande le programme officiel) correspond donc au résultat noté σ et σ_n des calculatrices.

3.4 Problèmes de valeurs approchées

Supposons que l'écart-type d'une série statistique soit $s = 3{,}281$ et sa variance $s^2 = 10{,}764961$. Une valeur approchée de s (au dixième) est 3,3 et une valeur approchée de s^2 (au dixième) est 10,8. Les élèves ont tendance à prendre pour valeur approchée de s^2, le carré de la valeur approchée de s, c'est-à-dire $3{,}3^2$ (qui est égal à 10,89), ils obtiennent alors 10,9.

Les élèves doivent donc éviter de faire des calculs à partir de valeurs déjà arrondies.

→ *Exercice 43 (page 22)*

3.5 Confusion entre rang et valeur

Une série a 24 valeurs que l'on peut classer dans l'ordre croissant : $x_1, x_2, \ldots, x_{24}$.

Le premier quartile Q_1 est la valeur de rang $\frac{24}{4}$ donc $Q_1 = x_6$.

Certains élèves écrivent $Q_1 = 6$, c'est-à-dire confondent le rang de la valeur et la valeur elle-même.

→ *Exercice 56 et 61 (page 24)*

4 Description des approches

4.1 Diagramme en boîte (page 10)

A. Raisons du choix et objectifs

La médiane et l'étendue d'une série statistique ne donnent pas d'indications sur la répartition des valeurs de cette série.

Le choix de deux séries de même effectif ($n = 30$) ayant la même médiane (mesure de tendance centrale) et la même étendue (mesure de dispersion) mais dont les valeurs ne sont « visiblement » pas réparties de la même façon, doit amener les élèves à réfléchir à des moyens de mesurer et visualiser ces différences.

B. Corrigé

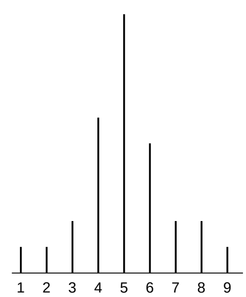
• Argument intuitif

Pour la série A, beaucoup de valeurs sont situées autour de la valeur centrale (médiane 5).

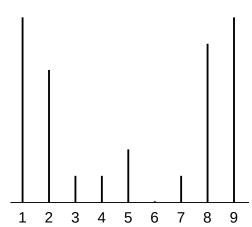
Pour la série B, au contraire, beaucoup de valeurs sont situées près du minimum ou du maximum.

• Argument graphique

Les diagrammes en bâtons de ces deux séries permettent de visualiser l'argument précédent :



Série A



Série B

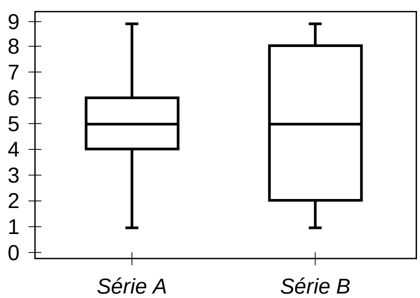
• Arguments quantitatifs

Pour la série A, 21 des 30 valeurs sont comprises entre 4 et 6, c'est-à-dire très proches de la médiane. Pour la série B, seulement 3 valeurs sont dans cet intervalle.
Environ la moitié des valeurs de la série B (16 sur 30) sont situées entre 2 et 8 alors que plus de la moitié des valeurs de la série A sont situées entre 4 et 6.

C. Scénario possible de mise en œuvre

Après un temps de recherche individuelle (10 minutes), les élèves peuvent travailler par groupes de 4 (15 minutes) avec comme consigne de se mettre d'accord sur un moyen de rendre compte de la différence observée. La mise en commun (20 minutes) permettra de valider tous les arguments cités précédemment et d'introduire l'idée de « moitié centrale » (l'intervalle interquartile) que l'on doit déterminer par le calcul des quartiles et illustrer par un diagramme en boîte.

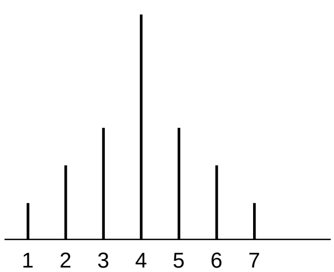
Pour la série A : $Q_1 = 4$ et $Q_3 = 6$.
Pour la série B : $Q_1 = 2$ et $Q_3 = 8$.



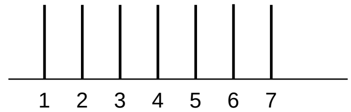
4.2 Mesures de dispersion (page 12)

A. Raisons du choix et des objectifs

Les deux séries proposées ont même moyenne 4 et même étendue 6. Les valeurs de chacune des deux séries sont réparties symétriquement par rapport à la moyenne :



Série A



Série B

La majorité des valeurs de la série A sont proches de la moyenne, ce n'est pas le cas de la série B qui est donc plus dispersée.
L'objectif de cette approche est de mesurer cette dispersion en prenant en compte les écarts de toutes les valeurs par rapport à la moyenne.

B. Corrigé

Pour la série A, effectif : 18, moyenne : 4, étendue : 6.
Pour la série B, effectif : 14, moyenne : 4, étendue : 6.

Calcul des écarts

Série A

valeur x_i	1	2	3	4	5	6	7
effectif n_i	1	2	3	6	3	2	1
écart $x_i - \bar{x}$	– 3	– 2	– 1	0	1	2	3
$(x_i - \bar{x})^2$	9	4	1	0	1	4	9
$\lvert x_i - \bar{x}\rvert$	3	2	1	0	1	2	3

Série B

valeur x_i	1	2	3	4	5	6	7
effectif n_i	2	2	2	2	2	2	2
écart $x_i - \bar{x}$	– 3	– 2	– 1	0	1	2	3
$(x_i - \bar{x})^2$	9	4	1	0	1	4	9
$\lvert x_i - \bar{x}\rvert$	3	2	1	0	1	2	3

Pour la série A :

$$\sum_{i=1}^{7} n_i\lvert x_i - \bar{x}\rvert = 20 \ ; \frac{1}{N}\sum_{i=1}^{7} n_i\lvert x_i - \bar{x}\rvert \approx 1{,}1;$$

$$\sum_{i=1}^{7} n_i(x_i - \bar{x})^2 = 40 \ ; \frac{1}{N}\sum_{i=1}^{7} n_i(x_i - \bar{x})^2 \approx 2{,}2 \ ;$$

Pour la série B :

$$\sum_{i=1}^{7} n_i\lvert x_i - \bar{x}\rvert = 24 \ ; \frac{1}{N}\sum_{i=1}^{7} n_i\lvert x_i - \bar{x}\rvert \approx 1{,}7;$$

$$\sum_{i=1}^{7} n_i(x_i - \bar{x})^2 = 56 \ ; \frac{1}{N}\sum_{i=1}^{7} n_i(x_i - \bar{x})^2 \approx 4.$$

Conclusion : pour les deux calculs : écart absolu moyen et écart quadratique moyen (variance) on obtient des valeurs supérieures pour la série B.

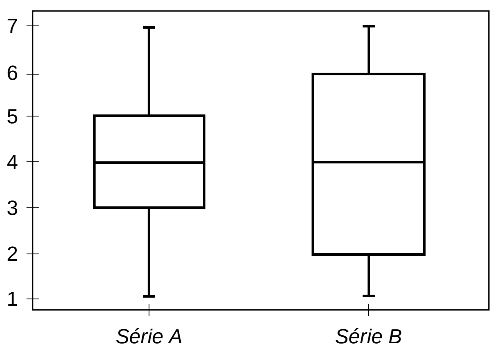
Remarque : l'écart absolu moyen $\frac{1}{N}\sum_{i=1}^{7} n_i|x_i - \bar{x}|$ est plus « naturel » mais la variance $\frac{1}{N}\sum_{i=1}^{7} n_i \ (x_i - \bar{x})^2$ a des propriétés mathématiques qui la rendent d'un usage plus courant.

C. Scénario possible de mise en œuvre

On peut reprendre le même déroulement que dans l'approche précédente.
Remarque : si on ne demande pas explicitement un moyen de mesurer la dispersion qui prenne en compte les écarts de chacune des valeurs de la série par rapport à la moyenne, on risque d'obtenir des procédures faisant intervenir les quartiles.

Pour la série A : $Q_1 = 3$, $Q_3 = 5$, et $Q_3 - Q_1 = 2$
pour la série B : $Q_1 = 2$, $Q_3 = 6$, et $Q_3 - Q_1 = 4$.

Le calcul des écarts interquartiles amène donc à la même conclusion : B est plus dispersée que A ; ce qu'on peut encore illustrer par les diagrammes en boîtes.



4.3 Transformation affine des données (page 14)

A. Raisons du choix et des objectifs

Les élèves connaissent les propriétés de linéarité de la moyenne : la moyenne augmente de deux points.
Les mesures de dispersion sont-elles également augmentées de deux unités ?

B. Corrigé

Pour la série de départ : moyenne $= 7{,}8$, $Q_1 = 5$, $Q_3 = 10$ donc $Q_3 - Q_1 = 5$, $s^2 = 5{,}76$ donc $s \approx 2{,}4$.
Pour la série transformée : moyenne $= 9{,}8$, $Q_1 = 7$, $Q_3 = 12$ donc $Q_3 - Q_1 = 5$, $s^2 = 5{,}76$ donc $s \approx 2{,}4$.
Les mesures de dispersion ne sont pas affectées par la transformation.
Remarque : il en est de même pour l'étendue $(11 - 4 = 13 - 7 = 7)$.

C. Scénario possible de mise en œuvre

Dans un premier temps, on peut demander aux élèves de proposer des hypothèses.
Après recensement des propositions, ils peuvent passer à une vérification sur l'exemple donné puis à la démonstration.
On peut prolonger cette approche en demandant ce qui se passerait si on multipliait les notes par 1,25.
Réponse : la nouvelle moyenne est à peu près la même $(9{,}75)$ mais les mesures de dispersion sont augmentées (les écarts se creusent !).

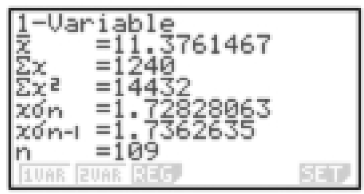
5 Activités

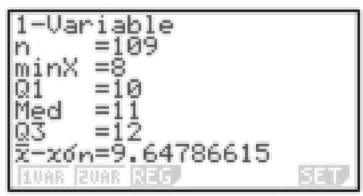
5.1 Calculs statistiques et diagramme en boîte (page 16)

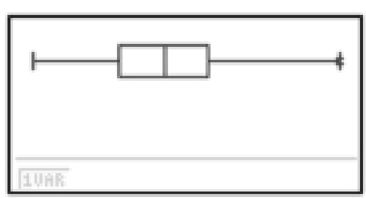
A. Notion utilisée

Utilisation de la calculatrice pour déterminer les paramètres de position moyenne (médiane, quartiles) et de dispersion (variance, écart-type) et pour construire le diagramme en boîte d'une série statistique.

B. Corrigé







5.2 Déciles d'une série statistique (page 16)

A. Notions utilisées

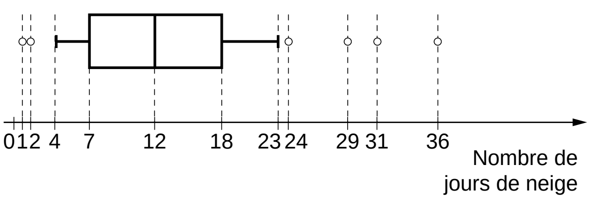
• Détermination de la médiane et des quartiles d'une série statistique.
• Calcul du premier et du neuvième décile d'une série statistique.
• Construction d'un diagramme en boîte limité au premier et au neuvième déciles.

B. Corrigé

Les valeurs rangées dans l'ordre croissant sont :
1 2 2 2 4 4 5 5 5 5 7 7 7 8 9 9 10 11 12 12 12 12 12 12 12 13 13 13 14 14 16 16 17 18 18 18 19 20 21 22 22 22 22 23 24 29 31 36.

L'effectif est $n = 49$ donc $\frac{n}{10} = 4{,}9$; $\frac{9n}{10} = 44{,}1$; $\frac{n}{4} = 12{,}25$; $\frac{3n}{4} = 36{,}75$.

D_1 est la 5$^{\text{ème}}$ valeur, D_9 la 45$^{\text{ème}}$, Q_1 la 13$^{\text{ème}}$ et Q_3 la 37$^{\text{ème}}$. La médiane est la 25$^{\text{ème}}$ valeur. Par conséquent $D_1 = 4$, $D_9 = 23$, $Q_1 = 7$ et $Q_3 = 18$.
La médiane est égale à 12.



5.3 Diagrammes en boîte de deux séries (page 17)

A. Notions utilisées

• Calcul des quartiles et de la médiane d'une série.
• Construction du diagramme en boîte d'une série.
• Comparaison de deux séries à l'aide de leurs diagrammes en boîte.

B. Corrigé

a) *Salaire des hommes*

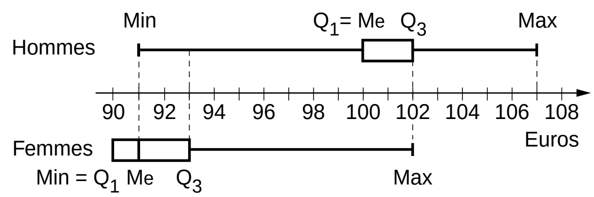
salaire	91	93	96	100	102	105	106	107
effectif	6	8	16	76	36	16	17	14

Salaire des femmes

salaire	90	91	93	96	98	102
effectif	50	33	18	14	3	3

Médianes et quartiles
Pour les hommes : $Q_1 = 100$ et $Q_3 = 102$.
La médiane est égale à 100.

Pour les femmes : $Q_1 = 90$ et $Q_3 = 93$.
La médiane est égale à 91.



b) Au moins 50 % des femmes ont un salaire inférieur ou égal au plus petit salaire masculin (en fait 73 % à cause des ex æquo). Au moins 25 % des hommes ont un salaire supérieur ou égal au plus gros salaire féminin (en fait 44 % à cause des ex æquo).
Il faut revenir à la distribution des salaires pour avoir les pourcentages précis mais l'intérêt du graphique est de donner l'idée de le faire.

5.4 Écart-type et écart interquartile (page 17)

A. Notions utilisées

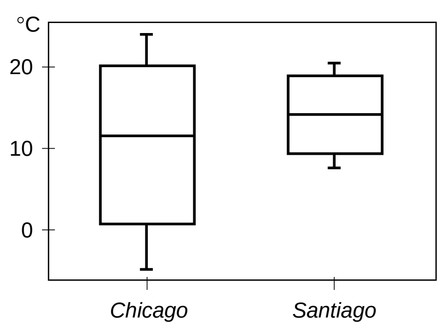
• Calcul des quartiles d'une série statistique.
• Construction du diagramme en boîte d'une série statistique.
• Comparaison de deux séries statistiques à l'aide de leurs diagrammes en boîte.
• Influence des valeurs aberrantes sur l'écart-type et l'intervalle interquartile.

B. Corrigé

■ **Comparer la dispersion des deux séries**

	minimum	Q_1	Me	Q_3	maximum	$Q_3 - Q_1$	s
Chicago	– 5	0	11,5	20	24	20	9,9
Santiago	7	9	13,5	17	20	8	4,4

Les températures sont plus homogènes (moins dispersées) à Santiago du Chili qu'à Chicago. C'est visible sur les diagrammes en boîtes (qui illustrent l'écart interquartile ou l'étendue). Le calcul de l'écart-type le confirme.



■ **Influence d'une valeur extrême**

	minimum	Q_1	Me	Q_3	maximum	$Q_3 - Q_1$	s
Série de départ	5,21	5,24	5,29	5,32	7,98	0,08	0,81
Série élaguée	5,21	5,24	5,29	5,31	5,33	0,07	0,04

L'écart interquartile est pratiquement inchangé quand on enlève la valeur suspecte. Au contraire l'écart-type (très sensible aux valeurs extrêmes) est fortement diminué (20 fois plus petit).

5.5 L'échelle du Q. I. (page 18)

A. Notion utilisée

Transformation affine des données.

B. Corrigé

a) $y_i = \dfrac{x_i - 80}{12} = \dfrac{x_i}{12} - \dfrac{20}{3}$.

b) $z_i = 15\left(\dfrac{x_i}{12} - \dfrac{20}{3}\right) + 100 = \dfrac{5}{4}x_i$.

c) Si le résultat est 92, le QI est 115 $\left(\dfrac{5}{4} \times 92 = 115\right)$.

Si le QI est 110, le résultat au test est $\left(\dfrac{4}{5} \times 110 = 88\right)$.

5.6 Fluctuation de l'écart-type entre séries de même taille (page 18)

A. Notion utilisée

Utilisation d'un tableur.

B. Corrigé

■ **Modification des valeurs de la série**

a) Si on remplace, par exemple 15 par 17, l'écart-type augmente (il devient égal à 5,36).

b) Si on remplace, par exemple 2 par 4, l'écart-type diminue (il devient égal à 4,85).

c) Si on remplace, par exemple 18 par 30, l'écart-type devient plus grand que 6 ($s = 7{,}9$).

d) Si on remplace, par exemple 18 par 15,1, l'écart-type devient plus petit que 4,6 ($s = 4{,}53$).

6 Corrigés des exercices et problèmes

Faire le point

• **Pour se tester**

1 c **2** b **3** c **4** a **5** b
6 b **7** c

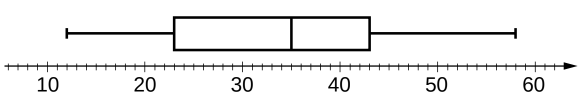
• **Vrai ou faux**

8 V **9** F **10** V **11** F **12** V
13 V **14** F **15** F **16** V **17** V

Exercices d'application

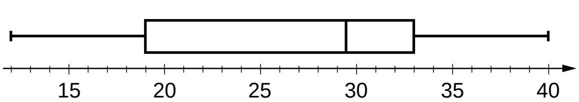
18 $Q_1 = 23$; $Q_3 = 43$; Me = 3 ;

minimum = 12 ; maximum = 58.



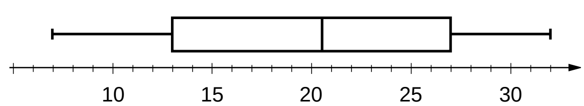
19 $Q_1 = 19$; $Q_3 = 33$; Me = 29,5 ;

minimum = 12 ; maximum = 40.



20 $Q_1 = 13$; $Q_3 = 27$; Me = 19,5 ;

minimum = 7 ; maximum = 32.

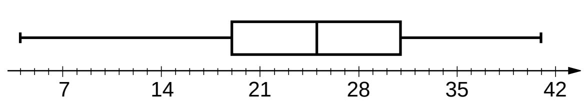


21 **a)**

valeur	4	12	19	25	31	41
effectif	2	3	4	6	8	1
effectif cumulé croissant	2	5	9	15	23	24

b) $Q_1 = 19$; $Q_3 = 31$; Me = 25 ; minimum = 4 ; maximum = 41.

c)



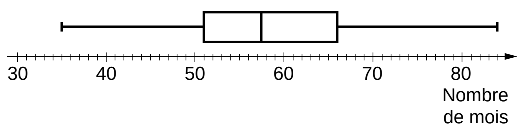
22

classe	[2 ; 8 [	[8 ; 10 [	[10 ; 12 [	[12 ; 20 [
effectif	7	12	14	7
effectif cumulé croissant	7	19	33	40

La médiane se situe dans la classe [10 ; 12 [;
le premier quartile se situe dans la classe [8 ; 10 [;
le deuxième quartile se situe dans la classe [10 ; 12 [.

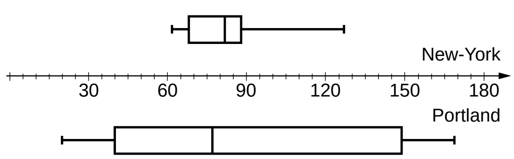
23 **a)** $Q_1 = 51$; $Q_3 = 66$; $Me = 57{,}5$;
minimum = 35 ; maximum = 84.

b)



24 **a)**

	minimum	Q_1	Me	Q_3	maximum
New York	64	68	82	88	127
Portland	20	40	77	149	169



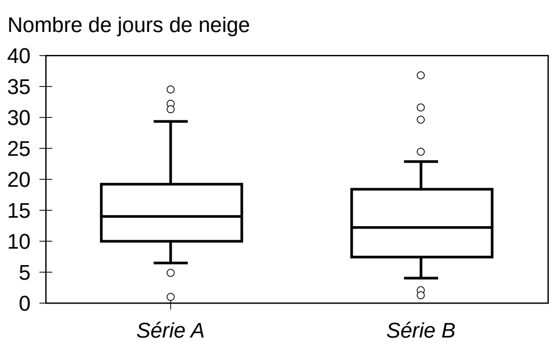
b) Les précipitations varient beaucoup d'un mois à l'autre à Portland (grande dispersion de valeurs). Elles sont plus régulières à New York.

Remarque : le total annuel est à peu près le même à New York (1 035 mm) qu'à Portland (1 076 mm).

25 **a)**

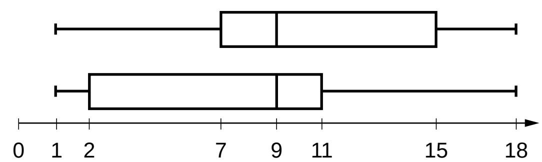
	minimum	Q_1	Me	Q_3	maximum	D_1	D_9
série A	1	10	14	19	34	6	29
série B	1	7	12	18	36	4	23

b)



c) Il y a moins de jours de neige dans la deuxième partie du siècle (deux en moyenne) bien que le maximum du siècle se trouve dans cette deuxième moitié.

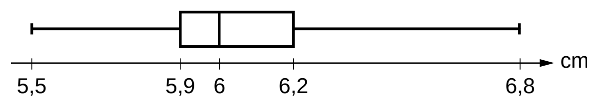
26 **a)** Série A : $Q_1 = 7$; $Q_3 = 15$; $Me = 9$.
Série B : $Q_1 = 2$; $Q_3 = 11$; $Me = 9$.



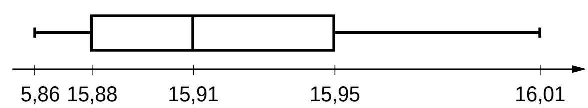
b) Les valeurs de la série A sont très concentrées entre 7 et 9 ainsi qu'au-delà de 15.
Les valeurs de la série B sont très concentrées entre 9 et 11 ainsi qu'avant 2.

27 **a)** $Q_1 = 5{,}9$; $Q_3 = 6{,}2$; $Me = 6$;

minimum = 5,5 ; maximum = 6,8.



28 La valeur suspecte est le maximum 19,81 ; presque quatre unités au-delà de la valeur précédente. Sans cette valeur, $Q_1 = 15{,}88$; $Q_3 = 15{,}95$;
$Me = 15{,}91$; minimum = 15,86 ; maximum = 16,01.



29 **a)** $\bar{x} \approx 2{,}372$; $s^2 \approx 0{,}005$; $s \approx 0{,}069$.

b) L'écart-type est exprimé dans la même unité que la valeur de la série (en mm).

30 **a)** moyenne : $\bar{x} \approx 23{,}292$.
b) variance : $s^2 \approx 83{,}623$;
écart-type : $s \approx 9{,}145$.

31

centre de classe	5	9	11	16
effectif	7	9	8	7

a) $\bar{x} \approx 10{,}2$.
b) $s^2 \approx 14{,}3$.
c) $s \approx 3{,}8$.

32 $s \approx 0{,}56$.

33 **a)** $\bar{x} \approx 32{,}1$ élèves.

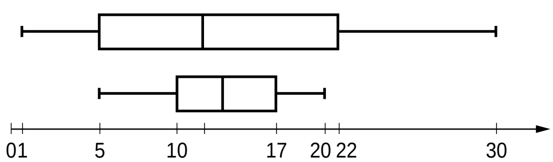
b) $s \approx 4{,}2$ élèves.

34 $\bar{x} \approx 18{,}5$; $s \approx 5{,}5$; $Q_3 - Q_1 = 22 - 16 = 6$.

35 **a)** $\bar{x} \approx 10\,200$.

b) $s \approx 1\,346$.

36 **a)**



La première série est beaucoup plus dispersée que la deuxième.

b) Série S_1 : $Q_3 - Q_1 = 22 - 5 = 17$;
Série S_2 : $Q_3 - Q_1 = 17 = 10 = 7$.

c) Série S_1 : $s \approx 9{,}2$; Série S_2 : $s \approx 3{,}9$.

37

a) Série S_1 : $Q_3 - Q_1 = 15 - 5 = 10$; $s \approx 12{,}9$.

b) Série S_2 : $Q_3 - Q_1 = 13 - 5 = 8$; $s \approx 4{,}5$.

c) L'écart interquartile est beaucoup plus robuste (c'est-à-dire moins sensible) que l'écart-type aux valeurs extrêmes ou aberrantes de la série.

38 **a)** moyenne $\bar{x} \approx 385{,}1$; $s \approx 39{,}3$.

b) $\bar{x} - 2s \approx 306{,}5$; $\bar{x} + 2s \approx 463{,}7$.
3 valeurs sur 35 ne sont pas dans l'intervalle $[\bar{x} - 2s\ ;\ \bar{x} + 2s]$, donc $\frac{32}{35}$ soit 91,4 % environ des valeurs se trouvent dans cet intervalle.

39

interventions quotidiennes	0	1	2	3	4	5	6
effectif	84	105	72	59	28	15	2
fréquences	23 %	28,8 %	19,7 %	16,2 %	7,7 %	4,1 %	0,5 %

b) $\bar{x} \approx 1{,}7$; $\bar{s} \approx 1{,}4$.

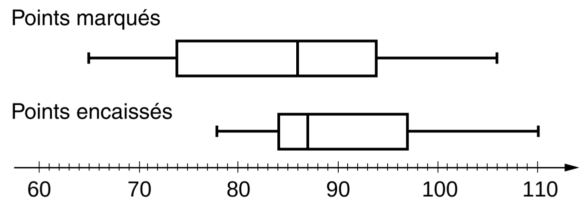
c) $\bar{x} - s \approx 0{,}3$; $\bar{x} + s \approx 3{,}1$.

$84 + 45 = 129$ valeurs ne sont pas dans l'intervalle $[\bar{x} - s\ ;\ \bar{x} + s]$, donc $\frac{236}{365}$, soit 64,7 % environ des valeurs sont dans cette intervalle.
Avec les fréquences 28,8 % + 19,7 % + 16,2 % = 64,7 %.

40 **a)** Points marqués $\bar{x} \approx 87$; $s \approx 13$.
Points encaissés $\bar{x} \approx 91$; $s \approx 9$.

b)

	minimum	Q_1	Me	Q_3	maximum
Points marqués	65	74	86	94	106
Points encaissés	78	84	87	97	110



c) Points marqués : $Q_3 - Q_1 = 94 - 74 = 20$.
Points encaissés : $Q_3 - Q_1 = 97 - 84 = 13$.

d) La série des points marqués est plus dispersée que celle des points encaissés.

41 $\bar{x} \approx 74{,}3$ minutes ; $s \approx 9{,}2$ minutes.

42

	moyenne	variance	écart-type
x	40	311,67	17,65
y	125	2805,03	52,95

43 **a)** Les nouvelles valeurs sont :
9 14 19 25 29 30 35 36 38 40 40 42 44 46 46 47 47 49 50 51 51 52 52 54 55 55 56 44 52 52.

b)

	moyenne	variance	écart-type
série obtenue	42	152	12,3
série d'origine	54,2	1,52	1,23

44

y	0	1	2	3	4	5	6	7
effectif	12	27	52	110	63	35	10	7

a) $\bar{y} \approx 3{,}2$; $s_y \approx 1{,}5$.
b) $\bar{x} = \bar{y} + 257$ donc $\bar{x} \approx 260{,}2$.
$s_x \approx s_y$ donc $s_x \approx 1{,}5$.

45 **a)** $\bar{x} \approx 55{,}87$ euros ; $s \approx 6{,}14$ euros.

b) $\bar{x} \approx 56{,}17$ euros ; $s \approx 6{,}14$ euros.

c) $\bar{x} \approx 58{,}66$ euros ; $s \approx 6{,}45$ euros.

46 Il suffit d'augmenter toutes les notes de deux points.

47

poids x	166	167	167,5	168	169	170,5	171	172	173
$y = x - 170$	– 4	– 3	– 2,5	– 2	– 1	0,5	1	2	3
effectif	1	6	12	21	36	48	34	24	18

Pour la série transformée : moyenne $\bar{y} \approx 0{,}15\text{g}$; $s_y \approx 1{,}68\text{g}$; $\bar{y} - 2s_y \approx -3{,}2$; $\bar{y} + 2s_y \approx 3{,}5$.
Une seule valeur n'est pas dans l'intervalle $[\bar{y} - 2s_y ; y + 2s_y]$ donc $\dfrac{199}{200}$, soit 99,5 % des valeurs sont dans cet intervalle.

48

centre de classe x	2,5	7,5	12,5	17,5	22,5	27,5
$y = \dfrac{x}{2{,}5}$	1	3	5	7	9	11
effectif	12	52	46	18	7	10

$\bar{y} \approx 4{,}807$; $s_y \approx 2{,}544$; $s_y^2 \approx 6{,}473$.
$\bar{x} = 2{,}5\bar{y}$; $s_x = 2{,}5 \cdot s_y$; $s_x^2 = 2{,}5^2 \cdot s_y^2$.
$\bar{x} \approx 12{,}0$; $s_x \approx 6{,}4$; $s_x^2 \approx 40{,}5$.

49 **a)** $y = x \cdot 10^{11} - 432\,567\,000$
Les valeurs de la nouvelle série sont : 189 ; 127 ; 156 ; 370 ; 433 ; 238.
b) et **c)**

	moyenne	écart-type
série y	252,2	1,23
série de départ	0,004 325 672 522	0,000 000 000 012 3

50 **a)** Les valeurs de la nouvelle série sont :
-2 3 12 -5 3 9 -9 7 1 2 -3 -4 2 -4 -1.

b) Les quartiles de cette série sont $Q_3 = 3$ et $Q_1 = -4$, l'écart interquartile est $Q_3 - Q_1 = 7$.

c) Pour la série initiale $Q_3 = (3 + 2000)$: $100 = 20{,}3$ et $Q_1 = (-4 + 2000)$: $100 = 19{,}6$ et l'écart interquartile est

$$Q_3 - Q_1 = 20{,}3 - 19{,}6 = 0{,}07.$$

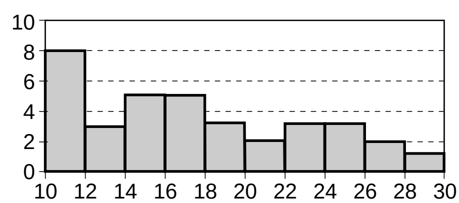
51 Il faut enlever 11,2 à chaque note et diviser le résultat par 1,8 $\left(y = \dfrac{x - 11{,}2}{1{,}8}\right)$.

52 **a)** Wladimir : $\dfrac{12{,}8 - 11{,}5}{2{,}5} = 0{,}52$

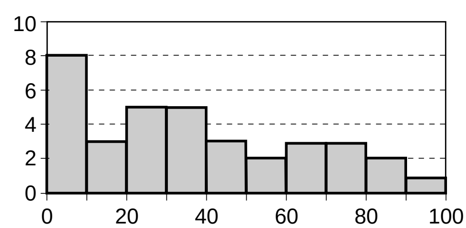
Mafalda : $\dfrac{86 - 79}{9} \approx 0{,}78$

b) Mafalda se situe à 0,78 écarts-types au dessus de la moyenne tandis que Wladimir est à 0,52 écarts-types seulement au dessus de la moyenne. Mafalda est donc mieux classée que Wladimir.

53 **a)**



b)



54 **a)** $\bar{x} \approx 38{,}9$; $s \approx 4{,}6$.

b) $38{,}9 - 2 \times 4{,}6 = 29{,}7$; $38{,}9 + 2 \times 4{,}6 = 4{,}1$.
Trois valeurs n'appartiennent pas à l'intervalle $[\bar{x} - 2s ; \bar{x} + 2s]$ donc $\dfrac{32}{35}$, soit 91,4 % environ des valeurs appartiennent à cet intervalle.

c) $\bar{y} = 0$; $s_y = 1$.

d) $x \in [\bar{x} - 2s ; \bar{x} + 2s] \iff y \in [-2\ 2]$.
Le pourcentage est le même que celui obtenu en **b)**.

55 **a)**

	moyenne	écart-type
section A	9,5	3,9
section B	7	4,7

b) Si on appelle x_i les notes de la série B, la transformation $y_i = \dfrac{(x_i - 7)}{4,7}$ suivie de la transformation $z_i = 3,9y_i + 9,5$ ramène la moyenne à 9,5 et l'écart-type à 3,9.

On applique donc la transformation affine suivante aux notes de la section B : $z_i = \dfrac{3,9}{4,7}(x_i - 7) + 9,5$.

	moyenne	écart-type
x	7	4,7
y	0	1
z	9,5	3,9

56 **1. a)** $n = 9$.

b) 18 – 21 – 22 – 23 – 24 – 25 – 26 – 28 – 34

2. a) $\dfrac{n}{4} = \dfrac{9}{4} = 2,25$.

b) Q_1 est la valeur de rang 3 ; $Q_1 = 22$.

3. a) $\dfrac{3n}{4} = 6,75$

b) Q_3 est la valeur de rang 7 ; $Q_3 = 26$.

57 **1. a)** $M_e = 26$.

b) $Q_1 = 15$; $Q_3 = 32$.

c) Mini $= 5$; Maxi $= 40$.

2. a) $Q_3 - Q_1 = 32 - 15 = 17$.

b) Au moins 25 % des valeurs sont inférieures ou égales à 15.
Au moins 75 % des valeurs sont inférieures ou égales à 32.

58 **1. a)**

somme x_i	2	3	4	5	6	7	8	9	10	11	12
effectif n_i	28	64	76	102	148	170	142	106	72	74	18
$n_i x_i$	56	192	304	510	888	1 190	1 136	954	720	814	216

b) $\Sigma n_i \cdot x_i = 6\,980$.

$\bar{x} = \dfrac{\Sigma n_i x_i}{1000} = 6,98$.

2.

$n_i^2.x_i^2$	112	576	1 216	2 550	5 328
8 330	9 088	8 586	7 200	8 954	2 592

$\Sigma n_i x_i = 54\,532$.

$s^2 = \dfrac{\Sigma n_i \cdot x_i^2}{\text{N}} - \bar{x}^2 = \dfrac{54\,532}{1\,000} - 6,98^2 \approx 5,81$.

59 **a)** La série la plus dispersée est S_3 car il y a beaucoup de valeurs éloignées du centre de la distribution. La série la moins dispersée est S_2 car il y a beaucoup de valeurs proches du centre de la distribution.

Remarque : ces 3 séries ont même effectif (20) même moyenne (5,5) et même médiane (5,5).

b) $s_1 \approx 2,87$; $s_2 \approx 1,57$; $s_3 \approx 3,47$.

60 **1. a)**

x_i	168	170	175	177	178	179	180	181	182	183	185	188
y_i	– 12	– 10	– 5	– 3	– 2	– 1	0	1	2	3	5	8
effectif	1	1	3	4	5	8	9	7	3	2	1	1

b) $\bar{y} \approx -0,778$;
valeur arrondie au dixième : $\bar{y} \approx -0,8$.
c) $y_i = x_i - 180$
donc $\bar{y} = \bar{x} - 180$ ou $\bar{x} = \bar{y} + 180$
$\bar{x} \approx 180 - 0,8 \approx 179,2$.

2. $s_y \approx 3,3$ (arrondi au dixième), $s_x = s_y$ donc $s_x \approx 3,3$.

61 Les valeurs de la série ordonnées dans l'ordre croissant sont :

valeur	2	4	7	9	10	12	15	18	19	21	36
rang	1	2	3	4	5	6	7	8	9	10	11

L'effectif de la série est $n = 11$.
Le rang du premier quartile est le plus petit entier supérieur à $\dfrac{11}{4}$ c'est-à-dire 2,75, c'est donc la troisième valeur : $Q_1 = 7$.
Le rang du troisième quartile est le plus entier supérieur à $\dfrac{3 \times 11}{4}$ c'est-à-dire 8,25, c'est donc la neuvième valeur : $Q_3 = 19$.

Exercices d'approfondissement

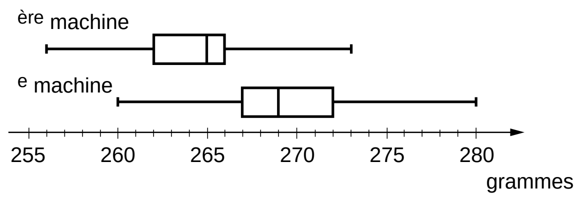
62 **1. a)** Échantillon 1 : $Q_1 = 262$; $Q_3 = 266$.

Échantillon 2 : $Q_1 = 267$; $Q_3 = 272$.

b) Échantillon 1 : $Q_3 - Q_1 = 4$.

Échantillon 2 : $Q_3 - Q_1 = 5$.

2. a)



b) La dispersion est une caractéristique de chaque machine. Les deux machines semblent toutes les deux du même niveau de précision. Par contre la deuxième a tendance à produire des paquets trop lourds. La première machine semble la plus appropriée.

63 **a)** Les effectifs des classes de collège sont peu dispersés, ceux de seconde encore moins (mais ils sont plus élevés). On observe une plus grande dispersion en première et en terminale.

b) Ceci est dû aux différentes séries de première et terminale.

64 **a)** et **b)**

série	moyenne	écart-type	coefficient de variation
S	11	5,85	53,2 %
S_1	110	58,5	53,2 %
S_2	1100	585	53,2 %

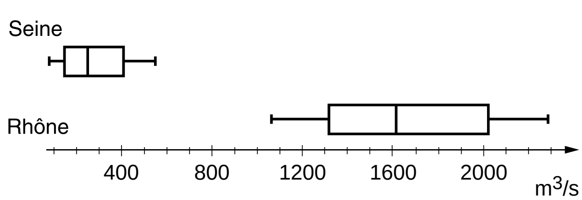
À l'échelle près, on obtient le même diagramme en boîte. Le vérifier à la calculatrice.

65 **1.** et **2.**

fleuve	moyenne	écart-type	coefficient de variation
Seine	258,3	154,6	59,9 %
Rhône	1639	406	24,8 %

Le Rhône a un plus fort débit et de grosses variations dans l'année, mais relativement à son débit, c'est la Seine la plus irrégulière.

3.



66 **1.** $\bar{x} \approx 11{,}6$; $s \approx 2{,}2$.

Intervalle 1 : [6,2; 8,4[

Intervalle 2 : [8,4; 10,6[

Intervalle 3 : [10,6; 12,8[

Intervalle 4 : [12,8; 15[

Intervalle 5 : [15; 17[

Dans le premier intervalle : Alsace.

Dans le deuxième intervalle : Auvergne, Centre, Franche-Comté, Ile de France, Limousin, Rhône-Alpes.

Dans le troisième intervalle : Aquitaine, Bourgogne, Bretagne, Champagne-Ardennes, Corse, Lorraine, Midi-Pyrénées, Basse-Normandie, Pays-de-Loire, Poitou-Charentes.

Dans le quatrième intervalle : Haute-Normandie, Picardie.

Dans le cinquième intervalle : Languedoc-Roussillon, PACA, Nord-Pas-de-Calais.

67 **1.** L'écart absolu moyen est égal à $\frac{38}{7} \approx 5{,}43$.

2. a) $\frac{1}{7}\sum_{i=1}^{7} |x_i - 17| = \frac{37}{7} \approx 5{,}29$.

Cette valeur inférieure à l'écart absolu moyen.

b) $\frac{1}{7}\sum_{i=1}^{7} |x_i - 16| = \frac{36}{7} \approx 5{,}14$.

68 **a)** La moyenne est égale à 4,93. Elle appartient donc à l'intervalle [4,88; 5,12].

b) L'écart-type est égal à 0,20. Il appartient donc à l'intervalle [0,13; 0,25].

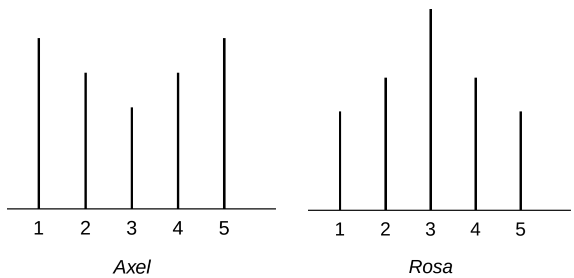
c) Rien ne permet de dire que la machine soit déréglée. On poursuit donc la production.

69 **a)**

	moyenne	médiane	écart-type	étendue	écart interquartile
Axel	3	3	4	1,55	3
Rosa	3	3	4	1,26	2

Les mesures de tendance centrale sont égales. La dispersion, mesurée par l'étendue ou l'écart interquartile est légèrement plus grande pour Axel mais cela ne suffit pas pour différencier les deux élèves.

b)



c) Rosa peut être qualifiée d'élève « moyen » car ses notes sont assez groupées autour de la moyenne.
Axel a des notes irrégulières, soit proches du maximum, soit proches du minimum.
Une appréciation du style « Du bon et du moins bon » lui irait très bien !

70 **a)**

7	1
8	25
9	357
10	12238
11	23446
12	357
13	2

b) Le premier quartile est 9,50,
le troisième quartile est 11,40.

71 **1. a)** Ce graphique correspond à un histogramme dont les classes ont une amplitude commune de 2 minutes.

b) L'avantage de ce type de graphique est, qu'en plus de montrer la forme de la distribution des valeurs, il n'y a aucune perte d'information, c'est-à-dire que les valeurs sont encore disponibles et lisibles sur le graphique.

2. L'étendue est $75 - 42 = 33$ minutes. Une valeur approchée (au dixième) de l'écart-type est 8,1 minutes.

72 **1. a)**

C.A.	[100; 150[	[150; 200[	[200; 250[	[250; 300[
effectif	40	50	60	70
effectif cumulé croissant	40	90	150	220

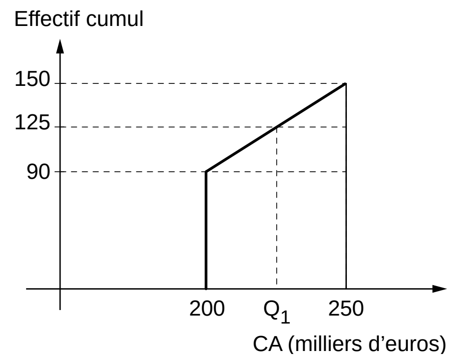
C.A.	[300; 350[	[350; 400[	[400; 450[	[450; 500[
effectif	100	80	60	40
effectif cumulé croissant	320	400	460	500

b) L'effectif est égal à 500. Le premier quartile Q_1 se trouve dans la classe [200; 250[car $\frac{500}{4} = 125$.

c) Le troisième quartile Q_3 se trouve dans la classe [350; 400[car $\frac{3}{4} \times 500 = 375$.
La médiane Me se trouve dans la classe [300; 350[car $\frac{500}{2} = 250$.

2. a) Pour calculer Q_1, il faut faire l'hypothèse que les valeurs sont réparties uniformément, c'est-à-dire que les accroissements de la variable sont proportionnels aux accroissements des effectifs cumulés.

C.A.	200	Q_1	250
effectif cumulé croissant	90	125	150



On a donc d'après la propriété de Thalès :

$$\frac{Q_1 - 200}{125 - 90} = \frac{250 - 200}{150 - 90}.$$

b) $Q_1 \approx 229{,}2$.

c) $Me \approx 315$; $Q_3 \approx 384{,}4$.

73 **a)**

$$\sum_{i=1}^{n}(a_i+b_i) = (a_1+b_1)+(a_2+b_2)+\ldots+(a_n+b_n)$$
$$= (a_1+a_2+\ldots+a_n)+(b_1+b_2+\ldots+b_n)$$
$$= \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i.$$

b) $\sum_{i=1}^{n} \lambda a_i = \lambda a_1 + \lambda a_2 + \ldots + \lambda a_n$

$$= \lambda(a_1 + a_2 + \ldots + a_n)$$
$$= \lambda \sum_{i=1}^{n} a_i.$$

74 $\displaystyle\sum_{i=1}^{p} n_i(x_i - \bar{x}) = \sum_{i=1}^{p} n_i x_i - \sum_{i=1}^{p} n_i \bar{x}$

$$= \mathrm{N}\bar{x} - \bar{x}\sum_{i=1}^{p} n_i = \mathrm{N}\bar{x} - \bar{x}\mathrm{N}$$
$$= 0.$$

75 **1. a)** $\bar{y} = a\bar{x}$.

b) $\displaystyle \mathrm{V}' = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$

$$= \frac{1}{n}\sum_{i=1}^{n}(ax_i - a\bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} a^2(x_i - \bar{x})^2$$
$$= \frac{1}{n}a^2\sum_{i=1}^{n}(x_i - \bar{x})^2 \text{ donc } \mathrm{V}' = a^2\mathrm{V}.$$

c) $s' = |a|s = as$ car $a > 0$.

2. a) Si $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$ et $a > 0$

alors $ax_1 \leqslant ax_1 \leqslant \ldots \leqslant ax_n$

c'est-à-dire $y_1 \leqslant y_2 \leqslant \ldots < y_n$.

b) Par conséquent le rang de Q_1' est le même que le rang de Q_1 donc $\mathrm{Q}_1' = a\mathrm{Q}_1$.

c) De même ${\mathrm{Q}_3}' = a\mathrm{Q}_3$,
donc $\mathrm{Q}_3' - \mathrm{Q}_1' = a(\mathrm{Q}_3 - \mathrm{Q}_1)$.

76 $\displaystyle\sum_{i=1}^{n}\sum_{k=1}^{n}(x_i - x_k)^2 = \sum_i\sum_k(x_i^2 - 2x_ix_k + x_k^2)$

$$= \sum_i\left(nx_i^2 - 2x_i\sum_k x_k + \sum_k x_k^2\right)$$
$$= n\sum_i x_i^2 - 2n\bar{x}\sum_i x_i + n\sum_k x_k^2$$
$$= 2n\sum_i x_i^2 - 2n^2\,\bar{x}^2$$
$$= 2n^2\left(\frac{1}{n}\sum_i x_i^2 - \bar{x}^2\right)$$

$= 2n^2\mathrm{V}$ où V est la variance de la série.

77 La moyenne de la nouvelle série est :

$$\bar{x'_n} = \frac{40 \times 15{,}75 + 18}{41} = \frac{648}{41} \approx 15{,}8.$$

$$\mathrm{V} = \frac{1}{40}\sum_{i=1}^{40} x_i^2 - \bar{x}^2 \text{ où } \bar{x} = 15{,}75.$$

$$\sum_{i=1}^{40} x_i^2 = 40\left(\mathrm{V} + \bar{x}^2\right)\ ;\ \sum_{i=1}^{41} x_i^2 = 40\left(\mathrm{V} + \bar{x}^2\right) + 18^2$$

pour la nouvelle série.

La variance de la nouvelle série est :

$\displaystyle \mathrm{V}' = \frac{1}{41}\sum_{i=1}^{41} x_i^2 - \bar{x'}^2$, on obtient :

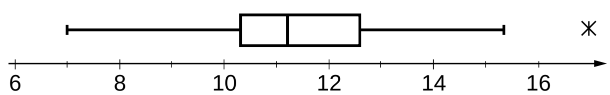
$\mathrm{V}' \approx 127{,}921$ et $s' \approx 11{,}31$.

Problèmes de synthèse

78 **a)**

minimum	Q_1	Me	Q_3	maximum
7	10,3	11,2	12,7	16,5

$12{,}7 + 1{,}5 \times (12{,}7 - 10{,}3) = 16{,}3$
et $10{,}3 - 1{,}5 \times (12{,}7 - 10{,}3) = 6{,}7$.
La valeur 16,5 peut donc être considérée comme aberrante.



79 **1. a)** La moyenne théorique est 4,5.

b) La moyenne de la série S_1 est de 4,67 (la médiane est 4,5).

c) L'arrondi au dixième de l'écart-type de la série S_1 est 2,9.

2. a) Les 25 moyennes sont :
4,5 – 5,75 – 1,25 – 4,25 – 4,25 – 5 – 6,5 – 6,25
3,75–6,5–3,75 – 5,25 – 4,5 – 3 – 7 – 2,75
2,5 – 4,5 – 5,75 – 5,5 – 4,25 – 3,75 – 5 – 5,75
5,5.

b) La moyenne de cette série S_2 est encore 4,67 ce qui était prévisible d'après les propriétés de la moyenne.

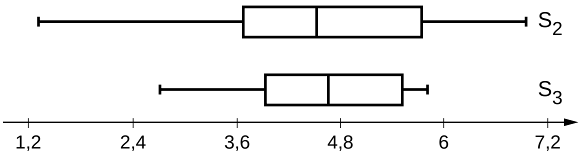
c) L'écart-type de cette série S_2 est égal à 1,4. Il est inférieur à celui de la série S_1 (environ la moitié), ce qui était prévisible car il y a moins de variations dans les moyennes de 4 valeurs que dans les valeurs individuelles.

3. a) Les 10 moyennes sont :
4,6–3,4–5,8 – 5,4 – 4,7 – 4,7 – 2,9 – 5,5 – 3,9
5,8.

b) La moyenne est encore 4,67 ce qui est prévisible.

c) L'écart-type de cette série S_3 est égal à 0,96. Il est inférieur à celui de la série S_1 (environ 3 fois moins), et à celui de la série S_2 ce qui était prévisible puisqu'il y a moins de variations dans des moyennes de 10 valeurs que dans des moyennes de 4 valeurs.

4. a)



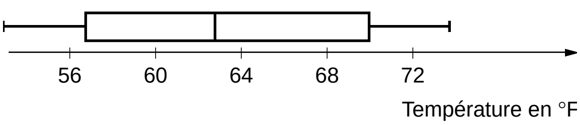
b) Cette étude illustre les fluctuations des moyennes dans des échantillons d'effectif *n*. Ces variations diminuent quand *n* augmente. Plus précisément, leur écart-type varie en fonction inverse de la racine carrée en *n*.

80 **a)**

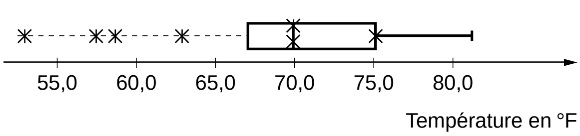
189	33
189	4444444555
189	6667
189	
190	
190	
190	
190	677777777777
190	8888888999
191	0000000

b) Le graphique en tiges et feuilles montre une distribution « bi-modale », c'est-à-dire avec deux maximums relatifs. Dans cette situation, les mesures de tendance centrale sont des valeurs peu caractéristiques.
Cette distribution peut éventuellement indiquer que l'expérience a été mal conduite.

81 **1.**

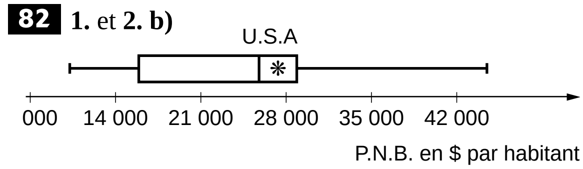


2. a)



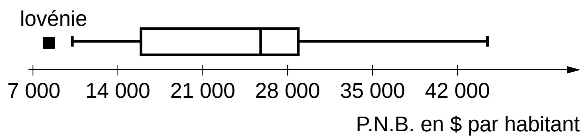
b) Presque tous les incidents (6 sur 7) ont eu lieu pour des températures inférieures ou égales à la médiane.
On peut aussi remarquer que le pourcentage de tirs sans incident en dessous de 66°F est égal à 0 % !
Pour finir, la température prévue (29°F) est très éloignée des températures pour lesquelles les observations ont été faites, ce qui rend toujours les extrapolations très hasardeuses. Malheureusement, les experts qui, après de longues discussions, ont autorisé le tir, ne se sont pas appuyés uniquement sur des considérations statistiques.

82 **1.** et **2. b)**



2. a) Quatre pays de l'UE avaient un PNB supérieur à celui des États-Unis en 1996.

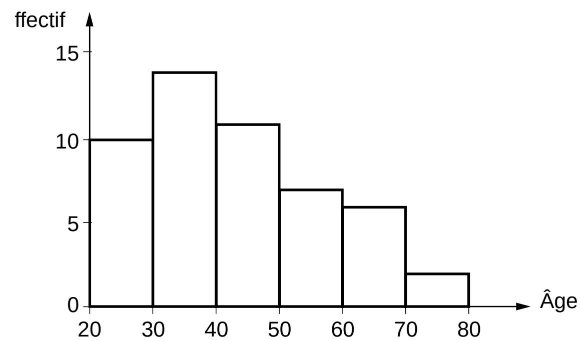
3.



83 **1.** Analyse par âge

a) $Q_1 = 32$ ans ; $Q_3 = 52$ ans.

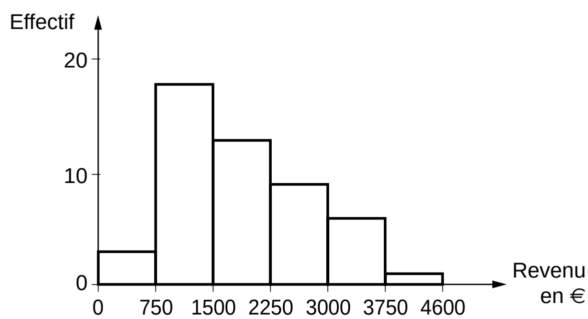
b)



2. Analyse par revenu mensuel

a) $Q_3 = 1100$ € ; $Q_3 = 2500$ €

b)



3. La moitié de la clientèle est âgée de 32 à 52 ans.
La moitié de la clientèle a un revenu mensuel compris entre 1100 et 2500 euros. Les histogrammes montrent que la classe d'âge la plus représentée est celle des 30-40 ans et le revenu mensuel le plus courant se situe entre 750 et 1500 euros.