

A Le programme

L'étude et la comparaison de séries statistiques menées en classe de Seconde se poursuivent avec la mise en place de nouveaux outils dans l'analyse de données. L'objectif est de faire réfléchir les élèves sur des données réelles, riches et variées (issues, par exemple, de fichiers mis à disposition par l'INSEE).

Contenus	Capacités attendues	Commentaires
Statistique descriptive, analyse de données. Caractéristiques de dispersion : variance, écart-type. Diagramme en boîte.	<ul style="list-style-type: none"> • Utiliser de façon appropriée les deux couples usuels qui permettent de résumer une série statistique : « moyenne, écart-type » et « médiane, écart interquartile ». • Étudier une série statistique ou mener une comparaison pertinente de deux séries statistiques à l'aide d'un logiciel ou d'une calculatrice. 	On utilise la calculatrice ou un logiciel pour déterminer la variance et l'écart-type d'une série statistique. Des travaux réalisés à l'aide d'un logiciel permettent de faire observer des exemples d'effets de structure lors du calcul de moyennes.

B Notre point de vue

Nous avons regroupé dans ce chapitre la partie du programme relative aux statistiques descriptives et l'analyse de données. Les notions nouvelles concernant ce chapitre sont peu nombreuses : variance, écart-type et diagramme en boîte.

Comme le demande le programme officiel, le plus souvent possible, les données utilisées sont des données récentes issues de sites officiels : INSEE, INED, OCDE, EuroStat...

L'**activité 1** « Taux de chômage dans la zone euro » et l'**activité 2** « Visualisation de la répartition des taux » permettent la découverte de l'écart interquartile, des diagrammes en boîte et de leur utilisation. Ces notions constituent l'essentiel de la première page de cours.

L'**activité 3** « Notes trimestrielles » permet la découverte de nouvelles caractéristiques de dispersion : la variance et l'écart-type. L'**activité 4** « Embouteillage à la chaîne » utilise la calculatrice pour comparer la dispersion de deux séries statistiques à l'aide de leurs écart-types. Ces notions sont reprises dans la deuxième page de cours.

À l'issue de ce chapitre, les élèves ont à leur disposition les deux couples usuels qui permettent de résumer une série statistique : « moyenne - écart-type » et « médiane - écart interquartile ». Les deux pages de cours se terminent chacune par un paragraphe présentant succinctement quand et comment utiliser chacun de ces couples.

Une fiche TICE permet de faire le point sur les techniques à connaître pour effectuer des calculs statistiques sur une calculatrice ou avec un tableur. Le programme officiel demandant d'utiliser la calculatrice ou un logiciel pour déterminer la variance et l'écart-type d'une série statistique, la connaissance de ces fonctionnalités est indispensable aux élèves.

Le programme officiel ayant comme objectif « de faire réfléchir les élèves sur des données réelles, riches et variées (issues, par exemple, de fichiers mis à disposition par l'Insee) », l'étude de ce type de données au volume souvent important nécessite la fourniture de fichiers et l'utilisation d'outils TICE adaptés, c'est le cas dans certains exercices proposés.

Le **TP1** permet de travailler sur des données de plus grande importance issues d'un fichier de l'INSEE ; le **TP2** permet de découvrir le lissage par la méthode des moyennes mobiles.

Les notions abordées dans le chapitre 7

- Écart interquartile
- Diagramme en boîte
- Variance et écart-type
- Comparaison de séries statistiques

C Réactiver les savoirs

Voir manuel page 285 et le site www.bordas-index.fr pour les corrigés détaillés.

D Activités

Activité 1 Taux de chômage dans la zone euro

Cette activité introduit la notion d'intervalle et d'écart interquartile. Un des objectifs essentiels est de faire découvrir à l'élève que l'intervalle interquartile contient environ 50 % des valeurs et que l'étendue de cet intervalle est donnée par l'écart interquartile.

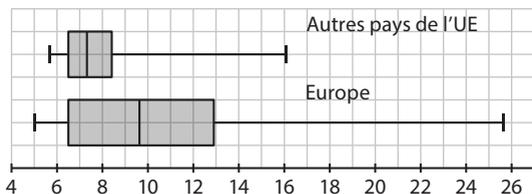
On peut noter que les valeurs de la série sont ordonnées.

1. La médiane est la 10^e valeur, soit 9,7.
2. a. Le premier quartile est la 5^e valeur, soit 6,5.
Le troisième quartile est la 15^e valeur, soit 13.
L'intervalle interquartile est [6,5 ; 13].
- b. L'écart interquartile est 13 - 6,5, soit 6,5.
3. a. L'affirmation est vraie, c'est la définition de la médiane.
- b. L'affirmation est vraie, car 6,5 correspond au 1^{er} quartile, soit 25 % de l'effectif environ et 13 correspond au 3^e quartile, soit 75 % de l'effectif environ.

Activité 2 Visualisation de la répartition des taux

Cette activité permet d'introduire la notion de diagramme en boîte et de découvrir comment comparer deux séries à l'aide de ces diagrammes.

1. a. Non, ces quatre intervalles sont tous d'amplitude différente.
 - b. Voir graphique
 - c. Voir graphique.
 - d. La France se situe entre la médiane et le troisième quartile de la série.
 - e. Les données sont plus concentrées avant le premier quartile et plus dispersées après le troisième quartile.
2. Les valeurs de la série sont ordonnées.
La médiane est la 5^e valeur, soit 7,3.
Le premier quartile est la 3^e valeur, soit 6,5.
Le troisième quartile est la 7^e valeur, soit 8,5.



3. On peut constater que les deux séries ont le même premier quartile.

En revanche, l'écart interquartile ainsi que l'étendue sont nettement plus importants dans la zone euro que dans les autres pays de l'UE.

Dans les deux cas, les données sont plus dispersées au-delà du 3^e quartile, et sont plus concentrées avant le 1^{er} quartile.

Activité 3 Notes trimestrielles

Cette activité permet de mesurer la dispersion des valeurs d'une série statistique autour de sa moyenne et de découvrir les notions de variance et d'écart-type.

1. a. La moyenne des notes d'Alicia est 12,5 ; celle des notes de Bertrand est 12,5.

Ordonnons les deux séries de notes : Alicia : 7 - 11 - 15 - 17. Thomas : 9 - 12 - 14 - 15.

La médiane des notes d'Alicia est 13, celle des notes de Thomas est 13.

- b. La moyenne et la médiane sont les mêmes pour les deux élèves. Par contre, la dispersion de leurs résultats est différente : les notes d'Alicia semblent plus dispersées que celles de Thomas.

2. a. Tableau complété :

Note	11	7	17	15	11
Écart avec la moyenne	-1,5	-5,5	4,5	2,5	-1,5
Carré de l'écart	2,25	30,25	20,25	6,25	2,25

Somme des carrés des écarts	59
Moyenne des carrés des écarts	14,75

On obtient une variance de 14,75.

- b. L'écart-type est environ 3,84.

3. a. Chacune des valeurs $(\bar{x} - 14)^2, \dots, (\bar{x} - 12)^2$ mesure le carré des écarts entre chaque note et la moyenne du trimestre ; en divisant la somme de ces carrés par l'effectif, soit 4, on obtient la moyenne.

- b. La variance est :

$$V = \frac{1}{4} ((12,5 - 14)^2 + (12,5 - 9)^2 + (12,5 - 15)^2 + (12,5 - 12)^2) \\ = \frac{21}{4}, \text{ soit } 5,25.$$

L'écart-type est $\sqrt{5,25}$, soit environ 2,29.

4. L'écart-type des notes d'Alicia est plus élevé que celui des notes de Thomas, cela correspond au fait que les notes d'Alicia sont plus dispersées que celles de Thomas.

Activité 4 Dispersion de deux séries statistiques

Cette courte activité a pour objectif de comparer la dispersion de deux séries statistiques à l'aide des valeurs de leurs écart-types, ces écart-types étant obtenus avec l'aide de la calculatrice.

1. À l'aide de la calculatrice, on obtient une moyenne de 102,56 pour la chaîne 1 et une moyenne de 102,58 pour la chaîne 2.

2. a. L'écart-type du premier échantillon est bien 1,08.

Chaîne 1	Chaîne 2
$\bar{x}=102,56$ $\Sigma x=5128$ $\Sigma x^2=525986$ $Sx=1,090964748$ $\sigma x=1,08$ $n=50$	$\bar{x}=102,58$ $\Sigma x=5129$ $\Sigma x^2=526205$ $Sx=1,213697336$ $\sigma x=1,201499064$ $n=50$

b. L'écart-type du second échantillon est 1,20 (arrondi au centième).

c. Le réglage de la chaîne 1 est meilleur que celui de la chaîne 2.

E Exercices

Pour démarrer

1. La série est déjà ordonnée et possède cinq valeurs.

Premier quartile : 2^e valeur, soit 9.

Troisième quartile : 4^e valeur, soit 15.

2. L'écart interquartile est 15 - 9, soit 6.

2 Exercice corrigé, voir p. 285 du manuel.

3 1. Série ordonnée :

11 - 12 - 13 - 13 - 13 - 14 - 14 - 15 - 19 - 21 - 22 - 22 - 23 - 24 - 24 - 26 - 28.

2. La médiane est la 9^e valeur, soit 19.

3. Le 1^{er} quartile est la 5^e valeur soit 13, il y a donc 25 % des personnes dont l'âge est inférieur ou égal à 13. De même, le 3^e quartile est la 13^e valeur, soit 23.

4. L'écart interquartile est donc 23 - 13 = 10.

4 1.

Note	1	2	3	4	5	6	7	8	9	10
Effectif cumulé croissant	2	3	4	9	13	22	26	29	34	35

2. La médiane est la 18^e valeur, soit 6.

3. Le premier quartile est la 9^e valeur, soit 4 ; le troisième quartile est la 27^e valeur, soit 8.

4. L'écart interquartile est 8 - 4, soit 4.

5 Série 1 : médiane 39, étendue 11, écart interquartile 4.

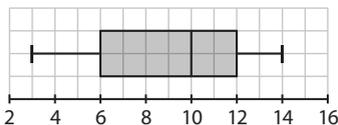
Série 2 : médiane 37, étendue 9, écart interquartile 7.

6 1. 1^{er} quartile : 12 ; 3^e quartile : 17 ; écart interquartile : 5.

2. Environ 50 % des valeurs car 14 est la médiane de la série.

3. Environ 50 % des valeurs car [12 ; 17] est l'intervalle interquartile.

7

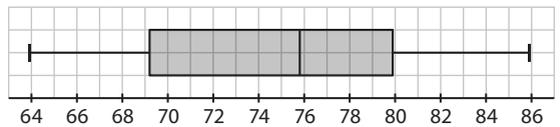


8 Exercice corrigé, voir p. 285 du manuel.

9 1. a. **Vrai**, car entre le minimum et la médiane d'une série, il y a environ 50 % des valeurs.

b. **Faux** : l'étendue est 85,99 - 63,90, soit 22,09 ; l'écart interquartile est 79,90 - 69,19, soit 10,71.

3. On peut choisir comme échelle 1 cm = 2 unités, on obtient alors le diagramme suivant :



10 1. Service 1 : médiane 6, étendue 12, écart interquartile 7. Service 2 : médiane 5,5, étendue 14, écart interquartile 3,5.

2. Le service 2 a des meilleurs résultats dans l'ensemble.

11 La série rouge possède le plus grand écart interquartile et la série noire le plus petit.

12 1. La moyenne est $\frac{102 + 106}{2}$, soit 104. La variance est $\frac{102^2 + 106^2}{2} - 104^2$, soit 4 et l'écart-type est 2.

2. La moyenne est $\frac{5 + 3 + 9 + 7}{4}$, soit 6. La variance est $\frac{5^2 + 3^2 + 9^2 + 7^2}{4} - 6^2$, soit 5 et l'écart-type est $\sqrt{5}$, soit environ 2,24.

13 On utilise la calculatrice, on obtient : $\bar{x} = 113,37$ et $\sigma \approx 6,80$.

14 Exercice corrigé, voir p. 286 du manuel.

15 1.

Classe	[0 ; 2[[2 ; 4[[4 ; 6[[6 ; 8[
Centre	1	3	5	7
Effectifs	10	7	5	2

2. $\bar{x} \approx 2,92$ et $\sigma \approx 1,96$.

16 1. L'effectif est 228. 2. $\bar{x} \approx 12,07$ et $\sigma \approx 1,17$.

17 **Fichiers associés sur www.bordas-index.fr et sur le manuel numérique Premium :**

733200_chap07_exercice17_correction.xlsx (Excel)

733200_chap07_exercice17_Correction.ods (OpenOffice).

1. **=MOYENNE(A1:J3)**

2. Réponse B : **=ECARTYPEP(A1:J3)**

18 1. On obtient :

– une moyenne de 7 et un écart-type de 1,41 environ pour Marion ;

– une moyenne de 7 et un écart-type de 2,16 environ pour Matthias.

2. C'est Marion qui a les résultats les plus réguliers.

Pour s'entraîner

19

Taille	1,70	1,71	1,72	1,73	1,74	1,75
Cumulé croissant	4	8	10	13	16	19

Taille	1,76	1,77	1,80	1,82	1,83
Cumulé croissant	25	28	31	32	33

1. Médiane : 1,75 ; étendue : $1,83 - 1,70 = 0,13$; écart interquartile : $1,76 - 1,72 = 0,04$.

2. $[1,72 ; 1,76]$ est l'intervalle interquartile. Ainsi, il y a environ 50% des tailles comprises entre 1,72 et 1,76 (valeur exacte 51,5%).

20 1. La médiane est entre la 6^e et la 7^e valeur, le premier quartile est la 3^e valeur et le troisième quartile est la 9^e valeur. On peut proposer comme série :

2 - 2 - 4 - 4 - 6 - 7 - 9 - 10 - 11 - 12 - 14 - 15.

21 Exercice corrigé, voir p. 286 du manuel.

22 Fichier associé sur www.bordas-index.fr et sur le manuel numérique Premium :

733200_chap07_exercice22_correction.alg (AlgoBox).

1. 16 valeurs, 1^{er} quartile rang 4, 17 valeurs, 1^{er} quartile rang 5.

2.

Sinon q prend la valeur $q + 1$
Afficher q

3.

Casio	Texas
<pre>=====Q1 ?>N# Int (N/4)→Q# Frac (N/4)→R# If R=0# Then Q# Else Q+1# IfEnd</pre>	<pre>PROGRAM: Q1 :Prompt N :ent(N/4)→Q :reste(N,4)→R :If R=0 :Then :Disp Q :Else :Disp Q+1 :End</pre>

23 Vrai, quand toutes les valeurs sont doublées, leur rang est inchangé.

24 Vrai, le 1^{er} quartile est la 71^e valeur et le 3^e quartile est la 211^e valeur, soit 141 valeurs.

25 Faux, les valeurs des quartiles d'une série peuvent ne pas être entières, de ce fait l'écart interquartile peut lui aussi ne pas être entier.

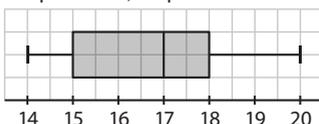
En revanche, l'écart interquartile contient un nombre entier de valeurs.

26 1.

Âge	14	15	16	17	18	19	20
Cumulé croissant	2	23	43	59	77	89	92

Médiane 17, 1^{er} quartile 15, 3^e quartile 18.

2.



27 Exercice corrigé, voir p. 286 du manuel.

28 1.

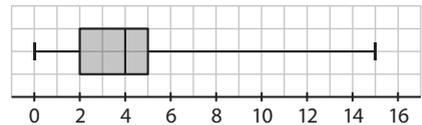
Nombre Hors-jeu	0	1	2	3	4	5	6	7
Effectif	1	4	13	11	9	10	2	4
Cumulé croissant	1	5	18	29	38	48	50	54

Nombre Hors-jeu	8	9	10	11	12	13	14	15
Effectif	4	2	1	1	0	0	1	1
Cumulé croissant	58	60	61	62	62	62	63	64

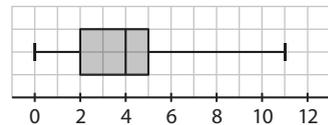
Médiane 4, 1^{er} quartile 2, 3^e quartile 5.

2. Écart interquartile 3.

3.



4. La médiane et les quartiles sont inchangés. Il y a peu de modifications sur le graphique, seule la moustache correspondant au maximum a été modifiée.



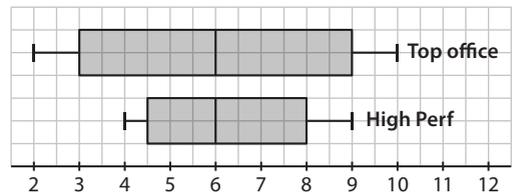
29 Faux, la boîte contient environ (et non pas exactement) 50% des valeurs : par exemple si l'effectif est impair, il n'est pas possible d'avoir exactement 50% des valeurs.

30 1. Faux, il y a seulement environ 25% des valeurs.

2. Vrai, car $[25 ; 30]$ correspond à l'intervalle interquartile.

3. Faux, il y a environ 25% des valeurs inférieures à 25.

31 1.



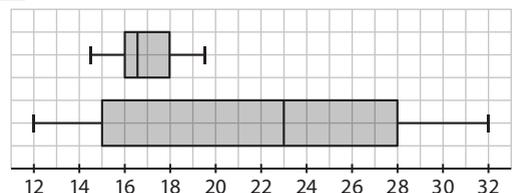
2. La médiane est la même pour les deux séries, mais la série associée à la marque TopOffice est beaucoup plus dispersée.

3. Sans informations complémentaires, comme par exemple la moyenne, il est difficile de conseiller une marque plutôt qu'une autre. Si on souhaite un nombre de pannes le plus constant possible sur toute les machines, on peut conseiller la marque HighPerf.

32 Exercice résolu, voir p. 174 du manuel.

33 Exercice corrigé, voir p. 286 du manuel.

34 1.



2. Ces diagrammes montrent que les températures sont beaucoup plus dispersées dans les champs. On peut supposer que les arbres permettent de maintenir une température plus stable.

Les diagrammes montrent aussi que les températures sont globalement plus basses en forêt, on peut émettre l'hypothèse que les arbres aident à conserver la fraîcheur.

35 1. Affirmation vraie.

2. Réciproque : « Si deux séries ont les mêmes diagrammes en boîte, alors ces deux séries sont identiques. » Cette réciproque est fausse.

36 1. « Avec la boule de 79 mm, j'ai réussi de très bons lancers mais également de très mauvais » : cette affirmation correspond à la série qui a la plus grande étendue, donc la série 3.

« Avec la boule de 71 mm, la moitié de mes lancers était à moins de 16 cm du cochonnet » : cette affirmation correspond à la série qui a une médiane de 16, donc la série 1.

« La boule de 75 mm avec laquelle je suis le plus régulier » : cette affirmation correspond à la série pour lesquelles les valeurs sont le moins dispersées, donc la série 2.

37 1. Faux. 2. Faux. 3. Vrai. 4. Faux.

38 Moyenne 16,2 ; écart-type 0,6.

39 Moyenne environ 37,7 ; écart-type environ 1,8.

40 2. Moyenne environ 58,0 ; écart-type environ 68,8.

41 Fichiers associés sur www.bordas-indice.fr et sur le manuel numérique Premium :

733200_chap07_exercice41_correction.xlsx (Excel)

733200_chap07_exercice41_correction.ods (OpenOffice).

1. `=MOYENNE(A1:H1)`

2. L'élève a utilisé la formule `=ECARTYPE(A1:H1)`

au lieu de la formule `=ECARTYPEP(A1:H1)`

42 Fichiers associés sur www.bordas-indice.fr et sur le manuel numérique Premium :

733200_chap07_exercice42_correction.xlsx (Excel)

733200_chap07_exercice42_correction.ods (OpenOffice).

1. `=MOYENNE(A1:I2)` 2. `=ECARTYPEP(A1:I2)`

3. La moyenne et l'écart-type sont sensiblement les mêmes, la série est très dispersée.

43 Exercice résolu, voir p. 175 du manuel.

44 1. Affirmation vraie.

2. Réciproque : « Si l'écart-type d'une série est nul, toutes les valeurs de cette série sont nulles ».

Cette réciproque est fausse. Il suffit que les valeurs de la série soient toutes identiques.

45 Faux. Un écart-type est toujours positif.

46 Faux : par exemple la série 1 - 2 - 3 a pour écart-type 0,82 environ alors que la série 2 - 2 - 3 a pour écart-type 0,47 environ.

47 1. Le salaire moyen est 2,3. C'est le même pour les deux séries.

2. Écart-type PME A : environ 1,647 ; PME B environ 0,999.

3. La répartition des salaires est plus homogène dans la PME B que dans la PME A.

48 1. $\bar{x} \approx 3,587$ et $\sigma \approx 0,925$.

2. $\bar{x} \approx 4,587$ et $\sigma \approx 1,489$.

3. Les annonces de la rubrique « Immobilier - Ventes » ont en moyenne une ligne de plus que les annonces de la rubrique « Ventes de véhicules ».

Cela peut se comprendre car, pour un bien immobilier (en particulier pour une maison), il y a beaucoup de précisions à mettre dans l'annonce : surface du terrain, nombre de pièces, situation ...

En comparant les deux écart-types, on peut constater que les valeurs de la 2^e série sont plus dispersées que celles de la 1^{re} série.

49 1.	Année	2009	2011	2013
	Moyenne	26668	24310	22912
	Écart-type	5808	2832	3131

2. Les moyennes vont en diminuant mais restent du même ordre de grandeur. En revanche, l'écart-type de la série en 2009 est environ deux fois plus élevé que les deux autres écarts-types. Les données de la première série sont nettement plus dispersées.

50 Exercice résolu, voir p. 176 du manuel.

51 Exercice corrigé, voir p. 286 du manuel.

52 Fichier associé sur www.bordas-indice.fr et sur le manuel numérique Premium :

733200_chap07_exercice52_Correction.alg (Algobox)

Entrée	Saisir M Saisir S
Traitement	Afficher « Borne inférieure », $M-2*S$ Afficher « Borne supérieure », $M+2*S$

53 1. L'affirmation est fausse.

Par exemple, les séries 7 - 10 - 11 - 12 et 8 - 9 - 10 - 13 sont différentes, pourtant elle ont la même moyenne et le même écart-type.

2. La réciproque est : « Si deux séries ont exactement les mêmes valeurs, alors ces deux séries ont exactement la même moyenne et le même écart-type. »

Cette réciproque est vraie.

54 Vrai. Il suffit de revenir à la définition de la variance.

55 Faux, par exemple la série 8 - 9 - 10 - 11 - 12 a cinq valeurs et un écart-type de 1,4 environ, alors que la série 8 - 10 - 12 a trois valeurs et un écart-type de 1,6 environ.

Top chrono

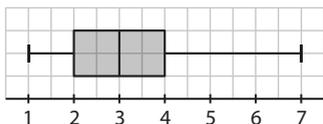
56 1.	Nombre d'employés	1	2	3	4	5	6	7
	Effectif cumulé	40	145	259	331	343	353	355

Médiane : 178^e valeur, soit 3. 1^{er} quartile : 89^e valeur, soit 2. 3^e quartile : 267^e valeur, soit 4.

L'écart interquartile est donc 2.

2. Étendue 6, écart interquartile 2.

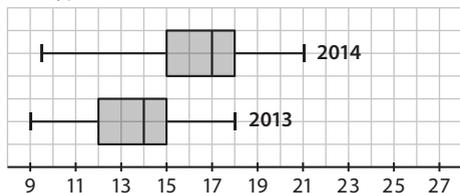
3.



57 1. Nombre annuel moyen de visiteurs : environ 8,43 millions.

2. L'écart-type est environ 0,765.

58



Les températures sont environ 3 °C plus hautes en 2014 qu'en 2013, mais leur répartition a sensiblement la même allure, en particulier l'écart interquartile est le même.

59 1. Pour le tireur A, la moyenne est 29 et l'écart-type environ 14,69.

Pour le tireur B, la moyenne est 29 et l'écart-type environ 13,75.

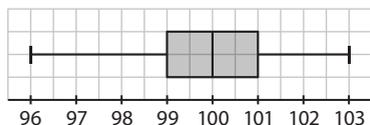
2. Les deux tireurs ont le même score moyen.

En revanche, le tireur A a des résultats plus dispersés.

Le tireur le plus régulier est le tireur B.

Revoir les points essentiels

60 1.



61 1. $\bar{x} \approx 1752,7$ et $\sigma \approx 475,1$.

2. C'est dans la première PME qu'il y a le moins d'inégalités salariales.

Travaux pratiques

TPI Les populations par communes selon l'INSEE

L'objectif est de fournir aux élèves un fichier de l'INSEE à l'état brut et de les aider à analyser les informations que l'on peut extraire de cette quantité importante de résultats.

Le fichier original est disponible à l'adresse :

http://www.insee.fr/fr/ppp/bases-de-donnees/recensement/populations-legales/pages2014/zip/HIST_POP_COM_RP12.zip

Les questions posées proposent quelques pistes pour effectuer cette analyse, mais, étant donnée la richesse des informations

fournies, il est évident que ce ne sont pas les seules possibles.

Fichiers associés sur www.bordas-indice.fr et sur le manuel numérique Premium :

733200_chap07_TP1.xlsx (Excel)

733200_chap07_TP1.ods (OpenOffice).

733200_chap07_TP1corrige.xlsx (Excel)

733200_chap07_TP1corrige.ods (OpenOffice).

A. Calcul de quelques paramètres

1. La médiane est 432.

L'écart interquartile est $1\ 065 - 195$, soit 870.

2. Il faut une graduation allant de 0 à environ 2 250 000. Au mieux, on peut prendre 1 cm pour 100 000 sur une feuille A4 en format « Paysage ».

Dans ces conditions, la valeur du 3^e quartile sera représentée par une distance d'environ 0,001 cm ! Cette valeur est encore plus petite pour la médiane et le 1^{er} quartile.

Sur le diagramme, toutes ces valeurs sont donc superposées avec l'origine de la graduation.

3. a. On obtient une moyenne d'environ 1 734 et un écart-type d'environ 14 741.

b. On constate que les données sont énormément dispersées (l'écart-type étant environ 8,5 fois supérieur à la moyenne).

4. a. La moyenne est environ 4 fois supérieure à la médiane.

b. L'étendue, 2 240 621, est environ 152 fois supérieure à l'écart-type. À noter que l'écart-type est 17 fois supérieur à l'écart interquartile et que l'étendue est égale à plus de 2 575 fois l'écart interquartile.

5. Si on enlève la ville de Paris, la médiane et les quartiles sont inchangés. En revanche, on obtient une moyenne d'environ 1 673 et un écart-type d'environ 8 954.

La moyenne a baissé mais pas énormément (de moins de 4 %), l'écart-type a lui beaucoup baissé (de presque 40 %) et l'étendue encore plus (presque 62 %).

B. Étude de quelques données

1. Le pourcentage de la population habitant Paris est environ 3,5 %.

2. Le pourcentage de la population habitant dans l'une des 36 villes les plus peuplées est environ 14,7 %, alors que cela représente 0,1 % des communes.

3. a. La population de 75 % des communes les moins peuplées est 10 024 486.

Ce résultat est à comparer avec celui de la question précédente : il y presque autant d'habitants dans les 36 villes les plus peuplées que dans 75 % des communes les moins peuplées.

b. Le pourcentage correspondant est environ 15,8 %.

4. L'affirmation est vraie.

C. Évolution de la population

1. a. La médiane est 356.

L'écart interquartile est $734 - 188$, soit 546.

b. La moyenne est environ 1 269 et l'écart-type environ 16 523.

2. a. Le pourcentage de la population habitant Paris est environ 6 %.

b. Population des 36 villes les plus peuplées : 9 068 518.

Le pourcentage correspondant est environ 19,5 %.

3. La population moyenne était moins importante en 1962, en revanche l'écart-type était encore plus important qu'en 2012. Le 1^{er} quartile est inchangé, en revanche, la médiane et le 3^e quartile étaient moins importants en 1962.

La proportion de la population habitant Paris ou une des 36 plus grandes villes était encore plus importante en 1962 qu'en 2012.

D. Comparaison de deux populations

1. a. Il y a 112 communes répertoriées.

La médiane est 8 313.

L'écart interquartile est : 20 172 - 4 179, soit 15 993.

b. La moyenne est environ 16 654 et l'écart-type environ 22 411.

2. La série possède beaucoup moins de valeurs : 112.

Les valeurs de la médiane et des quartiles sont nettement plus importantes (environ 20 fois plus grandes) que celles obtenues pour la métropole.

La moyenne est presque 10 fois plus élevée. L'écart-type a lui aussi augmenté, mais dans une proportion moindre.

Bien qu'encore très dispersées, les données ont une répartition plus homogène que celle de la métropole. Il est par exemple possible de tracer un diagramme en boîte avec une échelle où les différents paramètres ne sont pas tous superposés.

TP2 Lissage par la méthode des moyennes mobiles

Fichiers associés sur www.bordas-indice.fr et sur le manuel numérique Premium :

733200_chap07_TP2.xlsx (Excel)

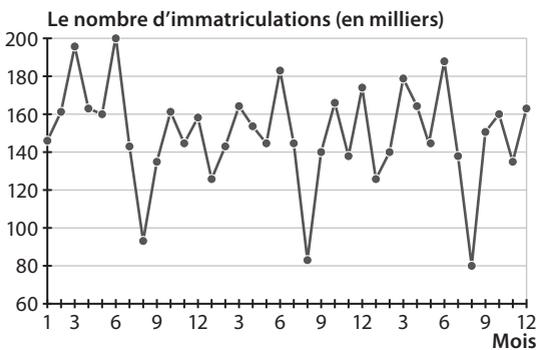
733200_chap07_TP2.ods (OpenOffice).

733200_chap07_TP2corrigé.xlsx (Excel)

733200_chap07_TP2corrigé.ods (OpenOffice).

A. Étude de la série chronologique

1. On obtient le graphique :



2. Le nombre d'immatriculations varie beaucoup d'un mois à l'autre. Les « pics » sont situés les mois de juin et les « creux » les mois d'août.

B. Lissage des données

1. a. Formule dans la cellule D3 : **=MOYENNE(C2:C4)**

b. et c. Voir énoncé.

d. La courbe de lissage d'ordre 3 à des variations moins importantes que la courbe initiale.

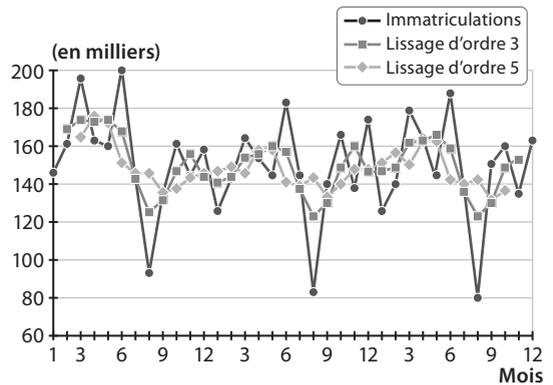
La méthode de lissage permet d'atténuer les variations locales dues aux fluctuations mensuelles et fait mieux apparaître la tendance générale.

2. a. Formule dans la cellule E4 : **=MOYENNE(C2:C6)**

Recopier cette formule vers le bas jusqu'en E35.

	A	B	C	D	E
1	Année	Mois	Nombre d'immatriculations	Lissage d'ordre 3	Lissage d'ordre 5
2	2012	1	145726		
3	2012	2	161335	167293	
4	2012	3	194818	172766	164837
5	2012	4	162146	172374	175398
6	2012	5	160159	173612	171604

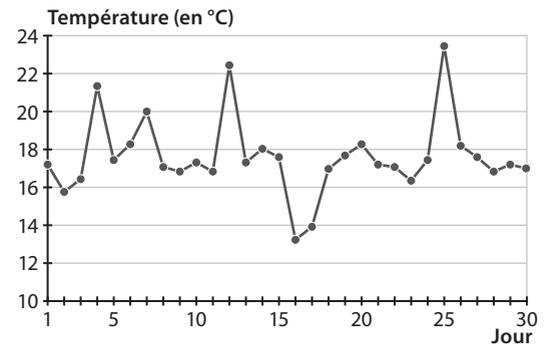
b.



c. Le nombre d'immatriculations durant les années 2012-2014 est relativement stable. On peut éventuellement observer une légère baisse la première année.

C. Application : lissage de températures

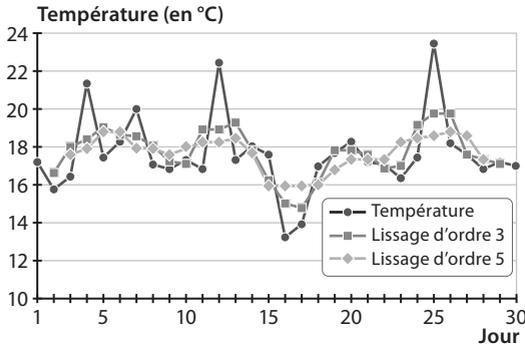
1. Comme dans la partie A, on place dans la colonne A les numéros d'ordre des jours du mois et dans la colonne B les températures relevées. On sélectionne la plage de données (zone B2:B31) et on représente graphiquement la série.



2. Dans la colonne C, on calcule les valeurs obtenues par la méthode de lissage d'ordre 3 et dans la colonne D, les valeurs obtenues par la méthode de lissage d'ordre 5.

	A	B	C	D
1	Jour	Température	Lissage d'ordre 3	Lissage d'ordre 5
2	1	17,2		
3	2	15,8	16,50	
4	3	16,5	17,87	17,66
5	4	21,3	18,43	17,88

3.



Une fois de plus, on peut constater que la méthode de lissage permet d'atténuer les variations locales dues aux fluctuations journalières et fait mieux apparaître la tendance générale suivie par la température.

Pour approfondir

62 Histogramme A – Diagramme en boîte 2.

Histogramme B – Diagramme en boîte 3.

Histogramme C – Diagramme en boîte 1.

63 1.

Âge	[20 ; 30[[30 ; 40[[40 ; 50[[50 ; 60[[60 ; 70[
Centre	25	35	45	55	65
Effectif	6	8	16	16	4

On obtient une moyenne de 45,8 et un écart-type d'environ 11,3.

2. Dans le premier garage, le total des âges de tous les acheteurs est 2 290 (ce résultat peut être lu sur la calculatrice). Dans le second garage, le total des âges de tous les acheteurs est $30 \times 39 = 1\,170$.

L'âge moyen de tous les acheteurs de la marque est $\frac{2\,290 + 1\,170}{50 + 30}$, soit 43,25.

64 1. a. Le salaire moyen des employés de l'entreprise Deschamps est 1 870,37 € et le salaire moyen des employés de l'entreprise Laville est 1 833,33 €.

b. Le salaire moyen des cadres de l'entreprise Deschamps est 3 166,67 € et le salaire moyen des cadres de l'entreprise Laville est 3 000 €.

2. Le salaire moyen de tous les salariés de l'entreprise Deschamps est 2 000 € et le salaire moyen de tous les salariés de l'entreprise Laville est 2 020 €.

3. On a vu dans la question 1. que le salaire moyen des employés ainsi que celui des cadres est plus élevé dans l'entreprise Deschamps, le PDG de cette entreprise dit vrai.

On a aussi vu dans la question 2. que le salaire moyen de tous les salariés est plus élevé dans les entreprises Laville, le PDG de cette entreprise dit lui aussi vrai.

Ce paradoxe est dû à la répartition des effectifs dans les différentes classes : il y a proportionnellement plus de cadres dans l'entreprise Laville (environ 19 %) que dans l'entreprise Deschamps (environ 11 %).

Les cadres étant en moyenne mieux payés que les employés, cela fait augmenter le salaire moyen de l'ensemble des salariés de l'entreprise.

65 1. • Classe de Delphine : moyenne 7,29, écart-type 1,70.

• Classe de Vincent : moyenne 15,46, écart-type 3,13.

Pour pouvoir comparer, il faut que les deux devoirs soient comptés sur le même nombre de points.

• Classe de Vincent : moyenne 7,73/10, écart-type 1,56.

2. L'affirmation de Delphine ne prend pas en compte le fait que les deux devoirs ne sont pas comptés sur le même nombre de points.

3. • Classe de Delphine : coefficient de variation 23,3 %.

• Classe de Vincent : coefficient de variation 20,2 %.

C'est dans la classe de Vincent que les résultats sont le moins dispersés.

66 Fichiers associés sur www.bordas-indice.fr et sur le manuel numérique Premium :

733200_chap07_exercice66.xlsx (Excel),

733200_chap07_exercice66.ods (OpenOffice).

733200_chap07_exercice66_correction.xlsx (Excel),

733200_chap07_exercice66_correction.ods (OpenOffice).

On commence par calculer les différents paramètres statistiques.

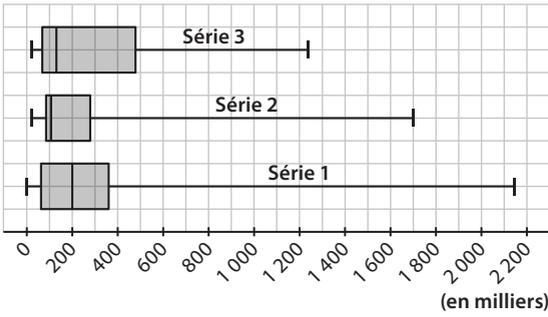
	26/02/2014 au 04/03/2014	30/04/2014 au 06/05/2014	24/12/2014 au 30/12/2014
Moyenne	309 892	290 472	300 339
Écart-type	447 427	431 622	348 859
Minimum	46 187	30 582	31 519
Q_1	65 298	79 326	58 731
Médiane	204 877	108 431,5	128 706,5
Q_3	347 439	166 177	414 369
Maximum	2 151 921	1 715 283	1 241 370
Étendue	2 105 734	1 684 701	1 209 851
Interquartile	282 141	86 851	355 638

On peut observer que l'écart-type est à chaque fois supérieur à la moyenne. Pour les séries 1 et 2, l'écart-type est même presque égal à 1,5 fois la moyenne.

Cela signifie que ces séries ont des valeurs très dispersées.

Ces observations sont confirmées par les valeurs des étendues et des écarts interquartile.

On peut ensuite tracer les diagrammes en boîte de ces trois séries.



Ils confirment les observations déjà faites.

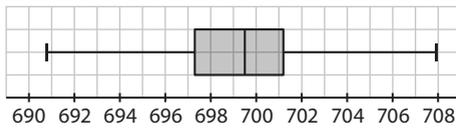
Les valeurs sont très dispersées après la médiane et beaucoup plus concentrées avant. C'est ce qui explique que la moyenne est à chaque fois beaucoup plus élevée que la médiane.

À noter que si l'on supprime la valeur maximum de la série 1, son diagramme devient presque symétrique et l'écart-type est divisé par 3.

La fréquentation de la semaine du 26/02 est fortement influencée par la valeur maximale.

67 1. Médiane : 699,55 ; premier quartile : 697,3 ; troisième quartile : 701,2.

Écart interquartile : 3,9.

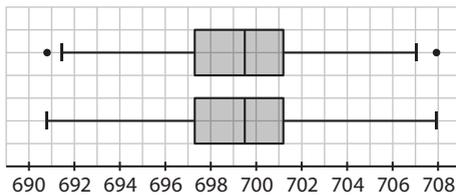


2. Intervalle : [691,45 ; 707,05].

a. Il y a deux valeurs aberrantes, les deux valeurs extrêmes 690,8 et 707,9.

b. Il y a 5 % de valeurs aberrantes.

3. a.



b. Comme on pouvait s'y attendre, seules les deux extrémités du diagramme sont modifiées pour faire apparaître les deux valeurs aberrantes.

68 1. On a le tableau :

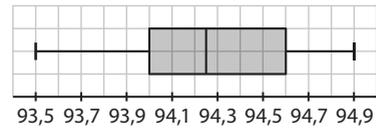
Valeurs	93,5	93,6	93,7	93,8	93,9	94	94,1	94,2
Effectifs	1	1	3	1	4	5	6	4

Valeurs	94,3	94,4	94,5	94,6	94,7	94,8	94,9
Effectifs	2	2	6	4	7	3	1

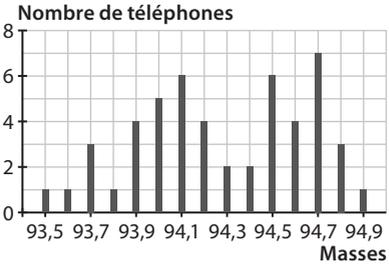
2. Médiane : 94,25 ; 1^{er} quartile : 94 ; 3^e quartile : 94,6 ; écart interquartile : 0,6.

Moyenne : 94,28 ; écart-type : environ 0,365.

3.



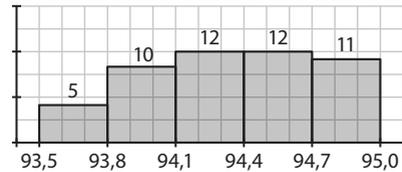
4.



5. On peut constater que les données ne sont pas réparties de façon homogène : il y a beaucoup de valeurs concentrées autour de 94,1 et de 94,6.

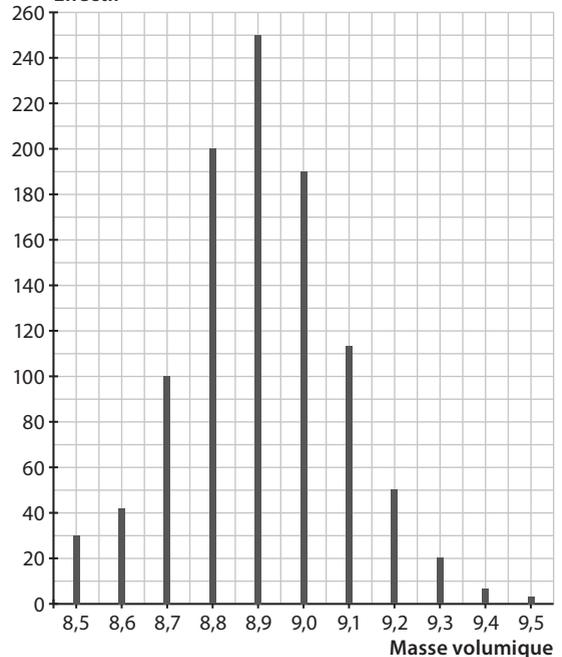
Cette constatation n'a pas pu être faite avant, car ni les paramètres statistiques ni le diagramme en boîte ne le permettaient.

6. L'histogramme donne l'impression que les données sont réparties équitablement. Le choix de ce rangement en classes n'est donc pas pertinent.

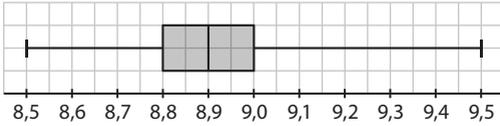


69 1.

Effectif



2.



3. Médiane : 8,9 ; écart interquartile : 0,2.

64,2 % des élèves ont une mesure dans l'intervalle [8,7 ; 9,1]

4. a. Moyenne : $\frac{8\ 884,8}{997}$, soit environ 8,912

b. Écart-type : 0,175 à 0,001 près par défaut.

c. L'intervalle $[\mu - 2\sigma ; \mu + 2\sigma]$, soit [8,56 ; 9,26], contient environ 94,8 % des valeurs.

L'intervalle $[\mu - 3\sigma ; \mu + 3\sigma]$, soit [8,38 ; 9,44], contient environ 99,4 % des valeurs.

70 1. Calcul des différents paramètres statistiques :

Groupe M

Moyenne : 14,4 ; écart-type : 1,40.

Médiane : 14,25 ; premier quartile : 13,5 ; troisième quartile : 15.

Étendue : 6 ; écart interquartile : 1,5.

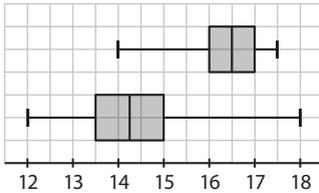
Groupe P

Moyenne : 16,27 ; écart-type : 0,86.

Médiane : 16,5 ; premier quartile : 16 ; troisième quartile : 17.

Étendue : 3,5 ; écart interquartile : 1.

En prenant comme échelle 1 unité = 1 cm, on obtient les diagrammes suivants :



2. Au vu des deux diagrammes, il semble que le médicament fasse baisser de manière significative la pression artérielle alors que le placebo, lui, semble ne pas avoir eu d'effet (même psychologique).

En revanche, les valeurs sont beaucoup plus dispersées chez les personnes ayant pris le médicament.

Les valeurs des moyennes et des écart-types confirment ces observations.

71 Salaire moyen en 2014 : environ 1 387,06 €.

Situation en 2015 :

Catégorie	Ouvrier	Ouvrier qualifié	Cadre
Salaire	1 212	1 363,5	1 515
Effectifs	56	39	16

Catégorie	Cadre supérieur	Dirigeant
Salaire	3 030	3 787,5
Effectifs	3	1

Salaire moyen en 2015 : environ 1375,36 €.

Tous les salaires ont augmenté de 1 %, pourtant le salaire moyen des salariés de l'entreprise a baissé.

Ce paradoxe est dû à la répartition des effectifs dans les différentes catégories : en 2014, les ouvriers représentent environ 47 % des salariés de l'entreprise et les ouvriers qualifiés 34 %, alors qu'en 2015, les ouvriers représentent environ 66 % des salariés de l'entreprise et les ouvriers qualifiés 46 %.

Ces salariés (ouvriers et ouvriers qualifiés) qui ont un salaire inférieur au salaire moyen et qui sont en plus grande proportion dans l'entreprise en 2015 font baisser le salaire moyen.