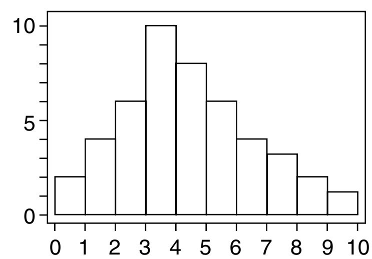
CHAPITRE 2

Statistique

1 Prérequis : « Pour démarrer » (page 32)

Exercice	Prérequis testés	Réponse	En complément
1	Retrouver les effectifs d'une série à partir de son histogramme. Localiser la classe médiane à partir des effectifs cumulés.	c	Autre exercice du même type : quelle est la classe médiane ?  Réponse : l'effectif est 46 ; la classe médiane est [4 ; 5[.
2	Calculer la valeur approchée de la moyenne d'une série dont les valeurs sont regroupées en classe.	c	Même question sur l'exercice ci-dessus. Réponse : la valeur approchée (au centième) de la moyenne est 4,37.
3	Calculer la moyenne d'une série à partir des moyennes de sous-séries.	b	Complément : quelle devrait être la moyenne des filles pour que la moyenne de la classe soit égale à 15 ? Réponse : 17.
4	Utiliser les propriétés de linéarité de la moyenne.	c	• Exercice du même type : que se passerait-il si le professeur augmentait toutes les notes de 30 % ? • Complément : que se passerait-il s'il ajoutait 2 à chaque note ?
5	Connaître la sensibilité des paramètres déjà rencontrés (moyenne, médiane, étendue) aux valeurs extrêmes.	a	Vérification à la calculatrice. Calculer la moyenne, la médiane et l'étendue de chacune des deux séries.

2 Objectifs

À la fin de ce chapitre, les élèves doivent savoir :
- reconnaître les différents types de variables et la nature des données d'une série statistique ;
- savoir ce qu'est une série chronologique ;
- savoir trouver une tendance dans l'évolution d'une série chronologique par lissage (à la main ou avec un tableur) ;
- 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) ;
- reconnaître l'intérêt et la pertinence de ces différentes mesures ;
- représenter une série statistique par un diagramme en boîte ;
- comparer deux séries à partir de leurs diagrammes en boîtes ;
- résumer une série statistique par un couple de valeurs (moyenne ; écart-type) ou (médiane ; écart interquartile) ;
- exploiter un tableau à double entrée ;
- trouver les fréquences marginales ou les fréquences conditionnelles à partir d'un tableau à double entrée ;
- faire le lien entre arbre et tableau à double entrée.

3 Difficultés et erreurs

3.1 Confusion dans la nature des données et dans leur représentation

• Des données en pourcentages peuvent provenir d'une variable qualitative. Cette variable détermine une partition de la population à partir de ses différentes modalités. Chaque pourcentage représente alors la part de la population pour chacune des modalités. La représentation des données peut être un graphique circulaire.

→ *Exercice 17 (page 48)*

• Une série de pourcentages peut être également une série de valeurs prises par une variable quantitative. (On ne saurait faire, dans ce cas, de graphique circulaire, car le total des pourcentages n'est pas égal à 100. Il n'a d'ailleurs aucun sens, quelle que soit sa valeur !)

→ *Exercice 39 (page 51)*

Ou encore une série chronologique.

→ *Exercice 18 (page 48)*

3.2 Histogramme à pas non constant

L'erreur la plus fréquente est de construire des rectangles avec des hauteurs proportionnelles aux effectifs (comme pour les histogrammes à pas constant) au lieu du rapport effectif / amplitude de l'intervalle c'est-à-dire à l'effectif par unité de la variable.

→ *Exercice 25 (page 49)*

3.3 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 dont le point de départ est le classement des valeurs de la série.

→ *Exercice 33 (page 50)*

Les valeurs données par les calculatrices ou les tableurs ne sont pas toujours égales à celles obtenues avec la définition du cours (imposée par le programme officiel).

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

1) Calcul de la médiane et partage de la série en deux sous-séries.

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

→ *Activité 1 (page 44)*

3.4 Confusion entre la variance et l'écart-type

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

La seule différence est la présence de la racine carrée pour l'écart-type. Les élèves ont tendance à l'oublier quand ils font le calcul à la main et que la question posée est de déterminer l'écart-type.

Par contre, les calculatrices donnent, en général, l'écart-type. Les élèves ont tendance à oublier d'élever au carré quand la question posée est de déterminer la variance.

→ *Exercice 40 (page 51)*

3.5 Différence entre σ_n et σ_{n-1}

Les calculatrices fournissent deux valeurs pour l'écart-type :

σ_n ou σ qui correspond à la formule

$$\sqrt{\frac{1}{N}\sum_{i=1}^{p} n_i\,(x_i - \bar{x})^2}$$

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 ou σ_n sur les calculatrices.

3.6 Problèmes de valeurs approchées

Supposons que l'écart-type d'une série statistique soit $s = 3{,}281$ et sa variance $s^2 = 10{,}764\,961$.

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 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.

Conclusion : il ne faut pas faire de calculs avec les valeurs déjà arrondies. Il faut garder le maximum de décimales tant que les calculs ne sont pas finis.

→ *Exercice 40 (page 51)*

3.7 Confusion entre rang et valeur

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

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

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

3.8 Difficultés dans les tableaux à double entrée

Tout tableau à double entrée ne fournit pas des effectifs marginaux ou des fréquences marginales. Même chose pour les fréquences conditionnelle.

Exemple : évolution de la répartition des étudiants dans l'enseignement supérieur (Source : *Tableaux de l'économie française 2000-2001*. INSEE.)

année / type d'enseignement	1990-1991	1998-1999
université	1 171 852	1 404 453
classes préparatoires	71 430	77 084
STS	199 084	234 300
autres	256 350	373 687
total	1 698 716	2 089 524

Le total par ligne n'a pas de sens.

4 Description des approches

4.1 Nature des données. Description graphique (page 34)

A. Raisons du choix et objectifs

Cette approche a pour but de faire réfléchir sur la nature des données d'une série statistique et sur le sens de la moyenne des valeurs d'une série statistique.

B. Corrigé

La moyenne des densités des 15 pays n'est pas la densité moyenne de l'Union Européenne à moins de calculer une moyenne pondérée (par la superficie de chacun des pays).

C. Scénario possible de mise en œuvre

Pendant un premier temps (5 minutes), certains élèves vont sûrement calculer la moyenne des 15 valeurs : 147 h/km^2. D'autres se poseront peut-être (en tout cas, on l'espère) des questions sur la possibilité d'ajouter des valeurs exprimées dans une unité quotient.

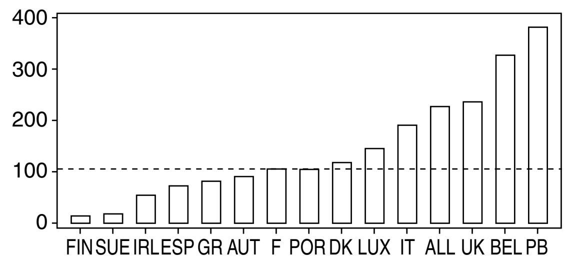
La mise en commun qui constituera le deuxième temps (15 minutes) devrait faire apparaître, des contre-exemples :

- avec la densité trouvée, connaissant la superficie de l'Union Européenne (2 890 000 km^2), la population de l'ensemble des 15 pays serait de 425 000 000 d'habitants. On espère que ceci paraîtra trop important à des élèves de 1re ES. La population réelle est de 370 000 000 d'habitants (et la densité de 128 h/km^2) ;
- si deux départements, un immense et désertique ayant une densité de 10 h/km^2, et un petit et surpeuplé ayant une densité de 390 h/km^2, sont réunis, la densité de l'ensemble des deux est-elle 200 h/km^2 ? On espère que des arguments sur la superficie seront évoqués. Dans le cas contraire, on prendra un exemple : 1 000 km^2 pour le premier, 100 km^2 pour le deuxième, la densité de l'ensemble est alors d'environ 44,5 h/km^2. En effet : $(1\,000 \times 10 + 100 \times 390) / 1\,100 \approx 44{,}5$.

Un dernier temps (5 minutes) permettra de conclure que la moyenne des valeurs d'une série statistique n'a pas toujours un sens concret même si elle a toujours un sens mathématique : la valeur a qui minimise $\frac{1}{N} \sum_{i=1}^{15} n_i (x_i - a)^2$ (voir page 41).

Dans le cas présenté, la médiane (109) aurait plus de sens : elle partage les pays en deux groupes, ceux à plus faible densité, ceux à plus forte densité. La France est à peu près à la médiane (107).

En complément, on peut demander comment on pourrait représenter cette distribution.



4.2 Lissage par les moyennes mobiles (page 36)

A. Raisons du choix et objectifs

Le but de cet approche est de différencier les variations entre deux valeurs consécutives d'une série chronologique et la tendance générale. On peut observer, par exemple, une baisse ponctuelle dans une tendance générale haussière.

B. Corrigé

Il se peut que les résultats du mois précédent aient été très faibles, soit exceptionnellement, soit parce que les mois correspondants sont traditionnellement faibles. La comparaison doit donc être plutôt effectuée sur les mêmes mois des années précédentes et le responsable des ventes a pu, dans ce cas, constater une hausse beaucoup plus modérée ou même une baisse.

C. Scénario possible de mise en œuvre

Après un temps de réflexion individuelle (5 minutes), les élèves sont mis par groupes de deux pour comparer

leurs hypothèses (5 minutes). Une mise en commun en classe complète (15 minutes) permettra de faire apparaître les notions de variations ponctuelles et de tendance à plus long terme.
On pourra proposer des exemples (si les élèves ne l'ont pas fait) : le chômage qui augmente chaque mois de septembre (arrivée des jeunes sur le marché du travail) même quand on est dans une période de diminution, les ventes des magasins de jouets qui croissent au mois de décembre même si on est en période de récession, etc.
La conclusion (5 minutes) portera sur l'intérêt de dégager la tendance globale d'une série chronologique en éliminant les variations périodiques ou exceptionnelles, ce qui est le but du lissage. On pourra même évoquer une deuxième étape (hors programme) avec les séries corrigées des variations saisonnières (comme celle du chômage, par exemple).
Bibliographie : *Méthodes Statistiques*, B. Grais, Dunod, 1992.

4.3 Diagramme en boîte (page 38)

A. Raisons du choix et objectifs

La médiane et l'étendue d'une série statistique ne donnent pas d'idée sur la répartition des valeurs entre le minimum et le maximum.
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.

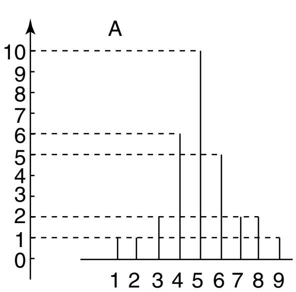
B. Corrigé

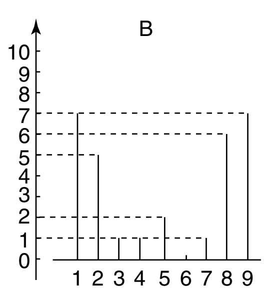
- Argument intuitif

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 barres de ces deux séries permettent de visualiser l'argument précédent :



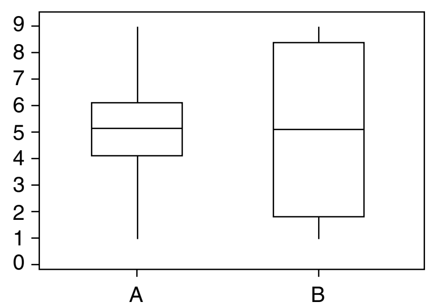


- Arguments quantitatifs

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

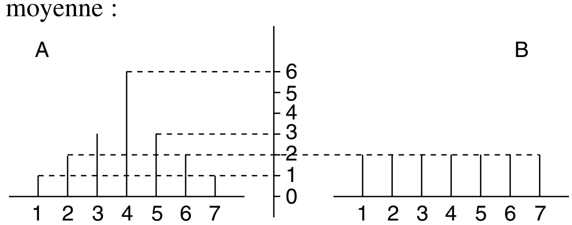
Après un temps de recherche individuelle (10 minutes), les élèves peuvent travailler par groupes de 4 (15 minutes) avec comme consigne de se mettre d'accord sur un argument.
La mise en commun (20 minutes) permettra de valider tous les arguments cités précédemment et d'introduire l'idée de « moitié centrale » (l'intervalle interquartile) que l'on peut déterminer par le calcul des quartiles et illustrer sur un diagramme en boîte.
Avec les définitions données dans le cours,
pour la série A : $Q_1 = 4$ et $Q_3 = 6$,
pour la série B : $Q_1 = 2$ et $Q_3 = 8$



4.4 Mesures de dispersion (page 40)

A. Raisons du choix et 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 :



La majorité des valeurs de la série A sont proches de la moyenne, ce n'est pas le cas pour 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 $x_i - \overline{x}$	−3	−2	−1	0	1	2	3
$(x_i - \overline{x})^2$	9	4	1	0	1	4	9
$\lvert x_i - \overline{x}\rvert$	3	2	1	0	1	2	3

$\sum_{i=1}^{18} n_i(x_i - \overline{x}) = 20$; $\frac{1}{N}\sum n_i \lvert x_i - \overline{x}\rvert \approx 1{,}1$;

$\sum_{i=1}^{18} (x_i - \overline{x})^2 = 40$; $\frac{1}{N}\sum_{i=1}^{18} (x_i - \overline{x})^2 \approx 2{,}2$.

Série B

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

$\sum_{i=1}^{14} n_i(x_i - \overline{x}) = 24$; $\frac{1}{N}\sum_{i=1}^{14} \lvert x_i - \overline{x}\rvert \approx 1{,}7$;

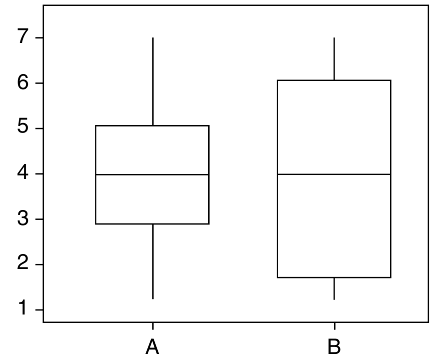
$\sum_{i=1}^{14} (x_i - \overline{x})^2 = 56$; $\frac{1}{N}\sum_{i=1}^{14} (x_i - \overline{x})^2 \approx 4$.

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

Remarque : l'écart absolu moyen $\frac{1}{N}\sum_{i=1}^{8} (x_i - \overline{x})$ est plus « naturel » mais la variance $\frac{1}{N}\sum_{i=1}^{8} (x_i - \overline{x})^2$ 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 : $Q_1 = 3$, $Q_3 = 5$, $Q_3 - Q_1 = 2$.
Pour la série B : $Q_1 = 2$, $Q_3 = 6$, $Q_3 - Q_1 = 4$.
Le calcul des écarts interquartiles amène donc à la même conclusion : B est plus dispersée que A ; ce qu'on peut encore illustrer sur le diagramme en boîte.



4.5 Tableaux à double entrée (page 42)

A. Raison du choix et objectifs

Cette approche a pour but de faire apparaître, sur un exemple, la différence entre la fréquence de A_i sachant B_j et de B_j sachant A_i.

B. Corrigé

1. Le pourcentage de garçons parmi les lauréats est $\frac{30}{55}$ soit environ 54,5 %.

Le pourcentage de filles parmi les lauréats est $\frac{25}{55}$ soit environ 45 %.

2. Le pourcentage de réussite parmi les garçons est $\frac{30}{55}$ soit environ 54,5 %.

Le pourcentage de réussite parmi les filles est $\frac{25}{35}$ soit environ 71,4 %.

3. Le pourcentage de garçons ayant réussi sur l'ensemble du lycée est $\frac{30}{90}$ soit environ 33 %.

Le pourcentage de filles ayant réussi sur l'ensemble du lycée est $\frac{25}{90}$ soit environ 28 %.

C'est le deuxième calcul qui permet de répondre que les filles réussissent mieux que les garçons dans ce lycée.

C. Scénario possible de mise en œuvre

Premier temps (10 minutes) : les élèves cherchent la réponse individuellement.
Deuxième temps (5 minutes) : le professeur demande oralement ceux qui pensent que les garçons réussissent mieux, ceux qui pensent le contraire.
Troisième temps (10 minutes) : si il y a des réponses divergentes, on commence par faire exposer sa méthode par un élève qui pense que les garçons réussissent mieux. On demande ensuite si d'autres méthodes ont été employées.

Dernier temps (5 minutes) : on conclut sur les notions de fréquences conditionnelles en insistant sur l'ensemble de référence sur lequel on calcule les pourcentages (cf. chapitre 1 sur les pourcentages).

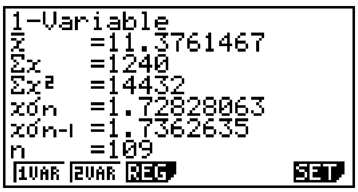
5 Activités

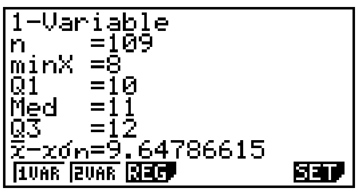
5.1 Calculs statistiques et diagramme en boîte (page 44)

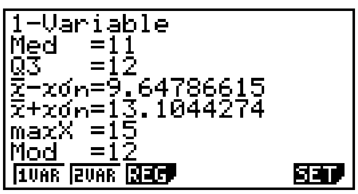
A. Notions utilisées

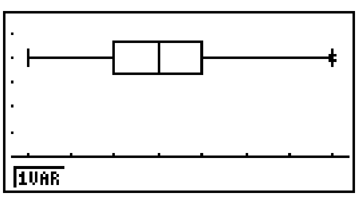
Utilisation de la calculatrice pour déterminer les paramètres de position (médiane, quartiles) et de dispersion (variance, écart-type) et pour construire le diagramme en boîte d'une série statistique.

B. Corrigé









5.2 Déciles d'une série statistique (page 44)

A. Notions utilisées

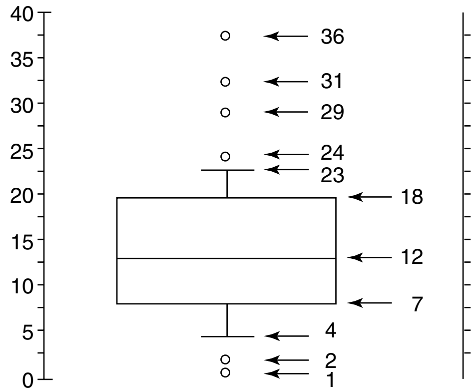
- Détermination de la médiane et des quartiles d'une série statistique.
- Calcul du premier et du dernier décile d'une série statistique.
- Construction d'un diagramme en boîte limité au premier et au dernier déciles.

B. Corrigé

Les valeurs rangées dans l'ordre croissant sont :
1 2 2 2 4 4 5 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 29 31 36.

L'effectif est $n = 49$ donc $\frac{n}{10} = 4{,}9$; $\frac{9n}{10} = 44{,}1$; $\frac{n}{4} = 12{,}25$; $\frac{3n}{4} = 34{,}75$.

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



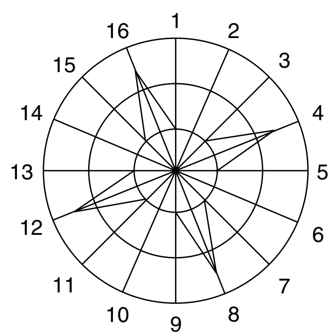
5.3 Lisser une série chronologique par les moyennes mobiles d'ordre 4 (page 45)

A. Notions utilisées

- Moyennes mobiles.
- Lissage par les moyennes mobiles.

B. Corrigé

1. On observe que, chaque année, le quatrième trimestre est beaucoup plus important que les trois autres.
Ceci peut se visualiser sur un graphique en étoile.
Il ne serait donc pas pertinent de calculer des moyennes mobiles d'ordre 3 car certaines comporteraient un de ces trimestre, et d'autres pas. On aurait peu de chances d'obtenir un bon lissage.

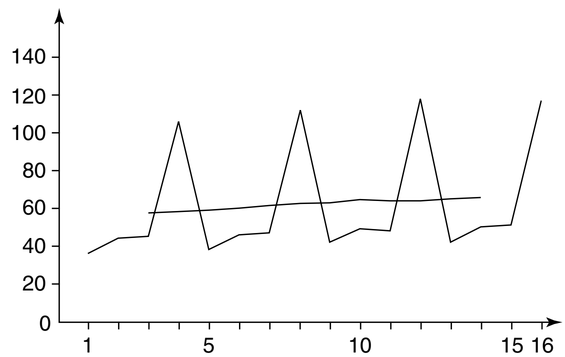


2. et **3.** On calcule la série des moyennes mobiles d'ordre 4 (MM4) puis la série des moyennes mobiles centrées d'ordre 2 (MMC2).

année	1997				1998			
trimestre	1	2	3	4	1	2	3	4
ventes	36	44	45	106	38	46	47	112
MM4		57,75	58,25	58,75	59,25	60,75	61,75	62,5
MMC4			58	58,5	59	60	61,25	62,125

année	1999				2000			
trimestre	1	2	3	4	1	2	3	4
ventes	42	49	48	118	42	50	51	118
MM4	62,75	64,25	64,25	64,5	65,25	65,25		
MMC4	62,625	63,5	64,25	64,375	64,875	65,25		

4.



5.4 Diagrammes en boîtes de deux séries (page 45)

A. Notions utilisées

- Calcul des quartiles et de la médiane d'une série.
- Construire le diagramme en boîte d'une série.
- Comparer deux séries par leur diagramme en boîte.

B. Corrigé

Salaires des hommes

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

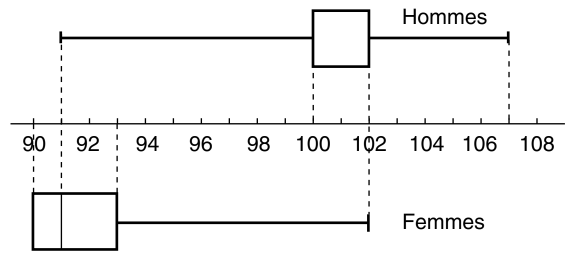
Salaires des femmes

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

Médiane 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.



Conclusion : 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 des graphiques est de donner l'idée de le faire !

5.5 Sensibilité des paramètres d'une série aux valeurs extrêmes (page 46)

A. Notions utilisées

- Paramètres de tendance centrale : moyenne, médiane.
- Paramètres de dispersion : étendue, écart-type, écart interquartile.
- Pertinence et robustesse aux valeurs extrêmes de ces différents paramètre.

B. Corrigé

1.

série	moyenne	médiane	étendue	écart interquartile	écart-type
S_1	14,1	14	48	4	5,97
S_2	13,6	14	17	4	4,15

2. On constate que la moyenne, l'écart-type et surtout l'étendue sont plus sensibles à la présence de valeurs exceptionnellement grandes (ou petites). La médiane et l'écart interquartile, au contraire, sont plus robustes c'est-à-dire moins affectés par les valeurs extrêmes (ici, ils sont même inchangés).

3. Un candidat à l'embauche sera plus intéressé par la médiane des salaires à moins qu'il ne soit candidat au poste de PDG ! (auquel cas il sera plus intéressé par le salaire maximum).

5.6 Visualiser l'effet de structure

A. Notions utilisées

- Effet de structure.
- Effet des pondérations dans le calcul de la moyenne d'une variable dans une population à partir des moyennes de cette variable dans des sous-populations.

B. Corrigé

Par exemple :

salaires moyens

	A	B	C	D	E	F
1	catégorie	1	2	3	4	total
2	effectif	100	200	300	400	1000
3	salaire	180	240	260	320	
4						
5	*salaire moyen de l'entreprise A*				272	
6						
7	catégorie	1	2	3	4	total
8	effectif	280	190	350	180	1000
9	salaire	200	260	280	340	
10						
11	*salaire moyen de l'entreprise B*				264,6	
12						

6 Corrigés des exercices et problèmes

Faire le point

- **Pour se tester**

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

6 a **7** b

- **Vrai ou faux**

8 V **9** V **10** V **11** F **12** V

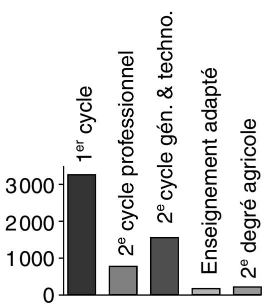
13 V **14** F **15** V

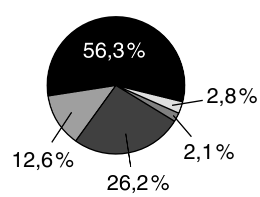
Exercices d'application

16 **a)** La variable étudiée est le type d'orientation des étudiants.

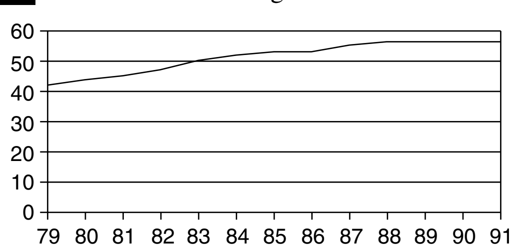
b) C'est une variable qualitative (on a pris ici 5 modalités). Les données obtenues sont des effectifs.

17 La variable étudiée est le type d'enseignement suivi par les élèves du second degré en France en 1998-1999. Cette variable est qualitative. Elle a 5 modalités.

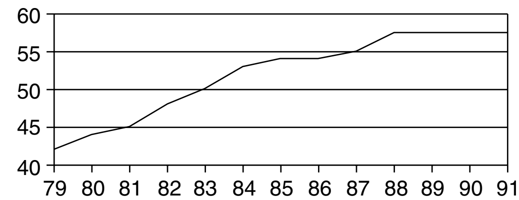




18 Avec l'origine à 0.



Avec l'origine à 40.



19 **a)** La variable étudiée est la température moyenne mensuelle à Paris.

b) Cette série de 12 valeurs n'est pas une série chronologique. La valeur du mois de mars n'a pas été observée un mois après celle du mois de février. On ne peut pas obtenir une 13e valeur.

c) Les météorologues calculent la moyenne annuelle (sur la période de 30 ans) en faisant la moyenne (non pondérée) des 12 moyennes mensuelles.

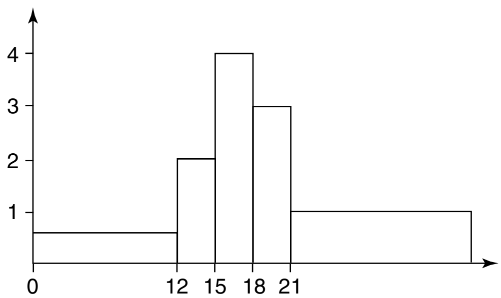
20 **a)** Variables quantitatives discrètes : a. (voir info à la suite de l'exercice 16) b. i.
Variables quantitatives continues : c. d. e. g. j. k. l. m.

b) Taux ou pourcentage : c. l. m.
Moyenne : e.
Indice : j.
Effectifs : b. g. i. k. (on peut aussi dire a. et d. mais ce sont plutôt des mesures).
Info : L'indice Dow-Jones était initialement une moyenne (non pondérée) des cours des 30 plus grandes valeurs industrielles américaines. Compte tenu des changements (entrées-sorties) dans la liste des 30, et des divisions d'actions, il a été transformé en indice.
Remarque : les variables c. g. l. peuvent conduire à la construction d'un cartogramme, c'est-à-dire à la représentation codée sur une carte des pays ou des régions dont les valeurs sont semblables, à partir de la répartition des valeurs en différents intervalles (*cf.* exercice 73).

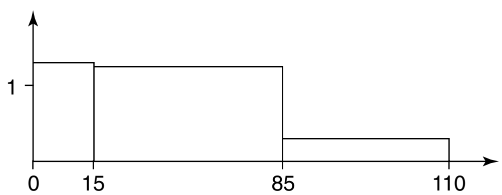
21 Variables qualitatives :
f. (élémentaire, secondaire, supérieur : l'ordre est lié au nombre d'années d'études) ;
h. (employé, cadre moyen, cadre supérieur, profession libérale : l'ordre est contestable !) ;
n. (très satisfait, satisfait, moyennement satisfait, mécontent, très mécontent : l'ordre est naturel !)
Remarque : ce n'est pas parce qu'on remplace les modalités précédentes par des codes : 1, 2, 3, 4, 5 par exemple, que la variable devient quantitative pour autant.

22 Séries chronologiques : d. e. j. m.

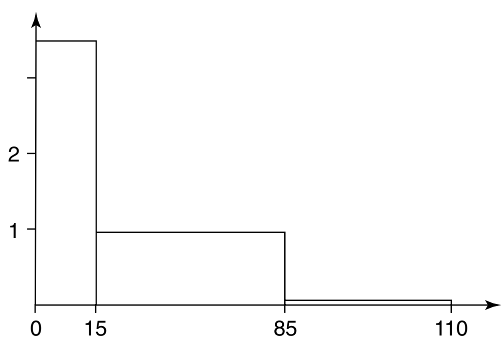
23



24 **a)**



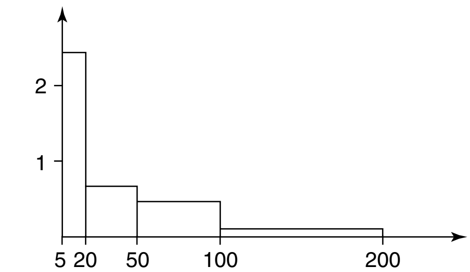
b)



25

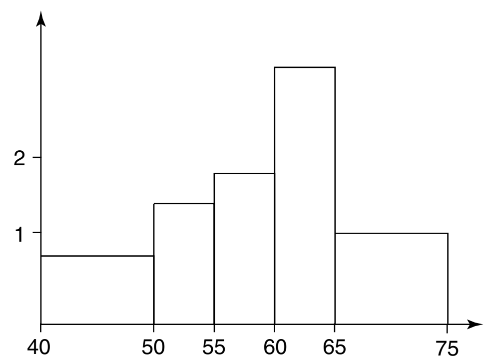
classe	[11 ; 17[	[17 ; 19[	[19 ; 21[	[21 ; 23[	[23 ; 25[	[25 ; 35[
hauteur du rectangle	3	5	6	5	2	1
amplitude de l'intervalle	6	2	2	2	2	10
effectif	9	5	6	5	2	5

26

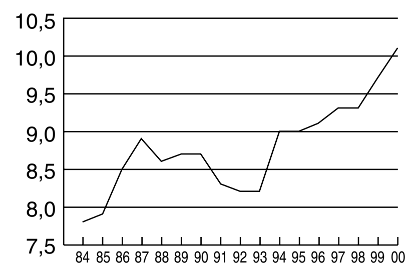


27 **a)** Ce graphique correspond à un histogramme dont les classes sont [42 ; 44[[44 ; 46[, …, [74 ; 75[. L'amplitude commune est égale à 2 minutes.

b)

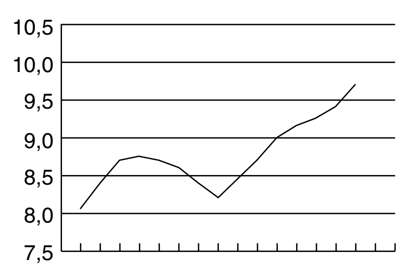


28 **a)**



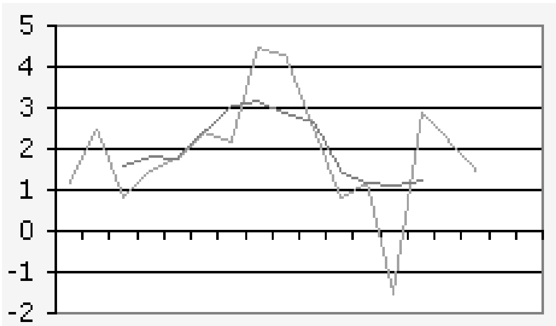
b) Les moyennes mobiles d'ordre 3 sont : 8,07 ; 8,43 ; 8,67 ; 8,73 ; 8,67 ; 8,57 ; 8,4 ; 8,23 ; 8,47 ; 8,73 ; 9,03 ; 9,13 ; 9,23 ; 9,43 ; 9,7.

c)

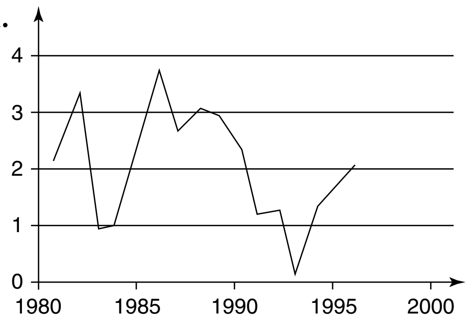


d) Le graphique des moyennes mobiles fait apparaître une tendance à la hausse.

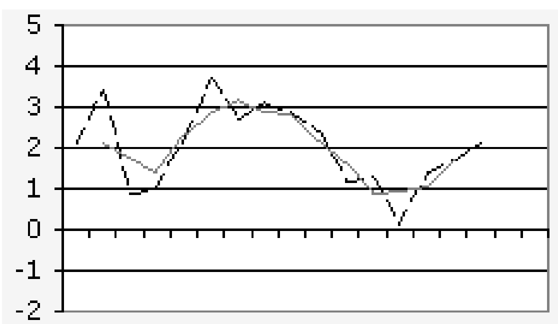
29 Les moyennes mobiles d'ordre 5 sont : 1,56 ; 1,8 ; 1,74 ; 2,48 ; 3,04 ; 3,18 ; 2,86 ; 2,66 ; 1,46 ; 1,18 ; 1,12 ; 1,2.



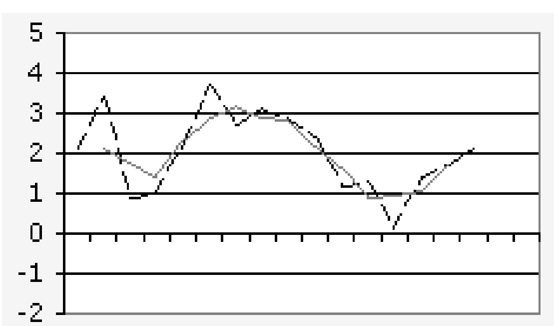
30 **1.**



2. et **3.** Les moyennes mobiles d'ordre 3 sont : 2,13 ; 1,77 ; 1,4 ; 2,33 ; 2,9 ; 3,17 ; 2,9 ; 2,8 ; 2,17 ; 1,63 ; 0,87 ; 0,93 ; 1,07 ; 1,73.



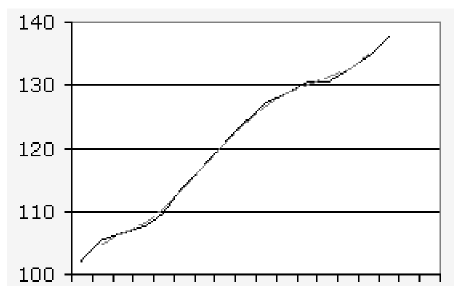
4. a) et **b)** Les moyennes mobiles d'ordre 5 sont : 1,94 ; 2,26 ; 2,12 ; 2,56 ; 2,94 ; 2,96 ; 2,46 ; 2,18 ; 1,58 ; 1,28 ; 1,14 ; 1,32.



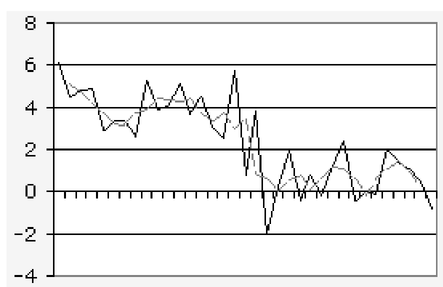
31 **a)** et **b)**

année	indice	MM3
1981	102,10	
1982	105,57	104,73
1983	106,52	106,56
1984	107,59	108,06
1985	110,06	110,59
1986	114,13	113,80
1987	117,22	117,40
1988	120,85	120,81
1989	124,35	124,18
1990	127,34	126,85
1991	128,87	128,92
1992	130,54	130,03
1993	130,67	131,24
1994	132,50	132,64
1995	134,75	134,95
1996	137,58	

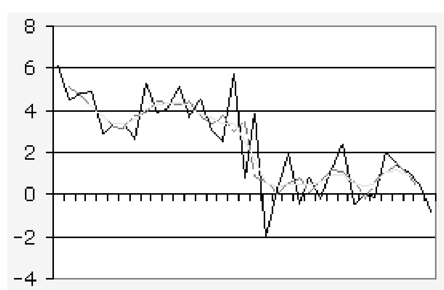
c)



32 **a)**

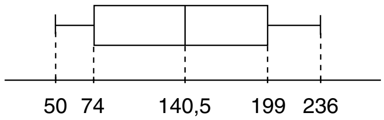


b)



33 **a)** $M_e = 140{,}5$; $Q_1 = 74$; $Q_3 = 199$.

b)

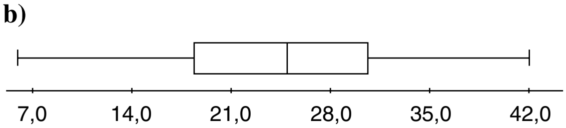


34 **a)**

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

$Q_1 = 19$; $Q_3 = 31$; $M_e = 25$; minimum = 4 ; maximum = 41.

b)



c) $Q_3 - Q_1 = 31 - 19 = 12$

35

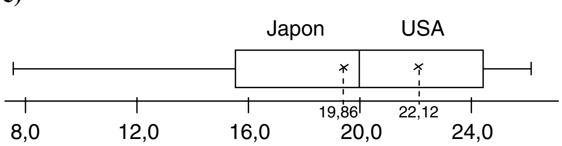
classe	[2 ; 8[	[8 ; 10[	[10 ; 12[	[12 ; 20[
effectif	7	12	14	7
effectif cumulé croissant	7	19	33	40

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[.

36 **a)** $M_e = 20{,}13$; $Q_1 = 15{,}41$; $Q_3 = 24$.

b) Minimum = 6,28 ; maximum = 26,77.

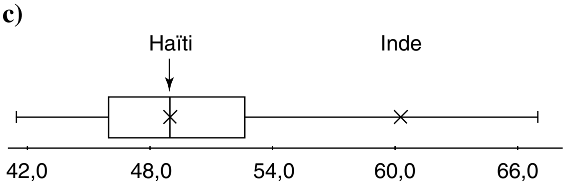
c)



37 **a)** $M_e = 49$; $Q_1 = 15{,}41$; $Q_3 = 53$.

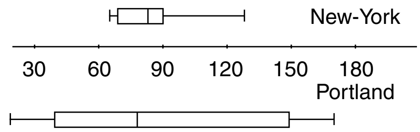
b) Minimum = 39 ; maximum = 69.

c)



38 **a)**

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

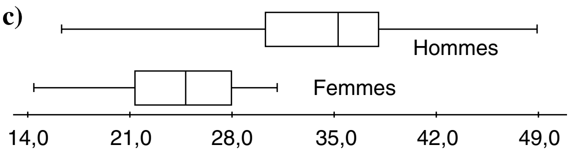


b) • Les précipitations varient beaucoup d'un mois à l'autre à Portland (grande dispersion de valeurs). Elles sont plus régulières à New-York.
• Remarque : le total annuel est à peu près le même dans les deux villes : 1 035 mm à New-York et 1 076 mm à Portland.

39 **a)** $M_e = 36$; $Q_1 = 30$; $Q_3 = 39$

b) $M_e = 25$; $Q_1 = 21$; $Q_3 = 28$

c)



40 **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 les valeurs de la série (le mm).

41 Moyenne : $\bar{x} \approx 23{,}292$; variance : $s^2 \approx 83{,}623$; écart-type : $s \approx 9{,}145$.

42

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) $s \approx 3{,}8$.

43 **a)** Série S_1 – étendue : 47 ; $s \approx 12{,}9$; $Q_3 - Q_1 = 15 - 5 = 10$; $M_e = 10{,}5$; $\bar{x} = 13{,}5$.
Série S_2 – étendue : 13 ; $s \approx 4{,}5$; $Q_3 - Q_1 = 13 - 5 = 8$; $M_e = 9$; $\bar{x} = 9{,}44$.

b) L'écart interquartile et la médiane sont beaucoup plus robustes c'est-à-dire moins sensibles aux valeurs extrêmes ou aberrantes.

44 Moyenne : $\bar{x} \approx 27{,}84$; écart-type : $s \approx 6{,}41$.

45 $s \approx 0{,}56$.

46 **a)**

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

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

c) $\bar{x} - s = 0{,}3$; $\bar{x} + s = 3{,}1$.
84 + 45 c'est-à-dire 129 valeurs ne sont pas dans l'intervalle $[\bar{x} - s\,;\,\bar{x} + s]$, donc $\dfrac{236}{365}$ soit 64,7 % des valeurs environ sont dans cet intervalle.
Avec les fréquences :
28,8 % + 19,8 % + 16,2 % = 64,8 %.

47 **a)** $\bar{x} \approx 18{,}5$ **b)** $s \approx 5{,}5$; $Q_3 - Q_1 = 22 - 16 = 6$.

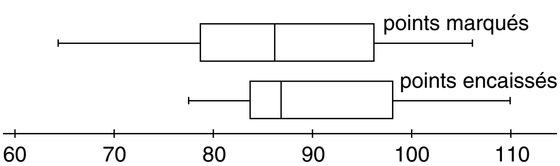
48 **a)** $\bar{x} \approx 10\,200$ **b)** $s \approx 11\,346$.

49 **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 $\dfrac{32}{35}$ soit 91,4 % environ des valeurs se trouvent dans cet intervalle.

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

b)

	minimum	Q_1	M_e	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) On note plus de variations dans le nombre de points marqués que dans le nombre de points encaissés.

51 **a)** L'écart absolu moyen est égal à $\dfrac{38}{7}$ soit environ 5,43.

b) $\dfrac{1}{n}\sum_{i=1}^{n} |x_i - 17| = \dfrac{37}{7}$ soit environ 5,29.

c) $\dfrac{1}{n}\sum_{i=1}^{n} |x_i - M_e| = \dfrac{36}{7}$ soit environ 5,14.

52 **a)** et **b)** Distribution de la couleur de cheveux :

	noir	châtain	blond	roux
effectif	82	70	65	24
fréquence	34 %	29 %	27 %	10 %

Distribution de la couleur des yeux :

	bleu	vert	marron
effectif	72	67	102
fréquence	29,9 %	27,8 %	42,3 %

c) Distribution de la couleur des yeux pour les individus ayant les cheveux noirs.

bleu	vert	marron
8,5 %	18,3 %	73,2 %

d) Distribution de la couleur des cheveux pour les individus ayant les yeux bleus.

noir	châtain	blond	roux
9,7 %	20,8 %	55,6 %	13,9 %

53 **a)** Pourcentage d'hommes du groupe O :
$100 \times \dfrac{200}{1\,000} = 20\,\%$. 20 % d'hommes du goupe O.

b) Pourcentage de femmes du groupe AB :
$100 \times \dfrac{15}{1\,000} = 1{,}5\,\%$. 1,5 % de femmes du groupe AB.

c) Distribution des groupes sanguins chez les hommes :

AB	A	O	B
4,7 %	46,7 %	37,4 %	11,2 %

Exemple : $\dfrac{25}{535} \times 100 \approx 4{,}7\,\%$.

Distribution des groupes sanguins chez les femmes :

AB	A	O	B
3,2 %	43 %	43 %	10,8 %

Exemple : $\dfrac{15}{465} \times 100 \approx 3{,}2\,\%$.

54 **a)** 13,9 % est le pourcentage des élèves de sixième qui sont de niveau 3 en lecture et 2 en calcul.
4 % est le pourcentage des élèves de sixième qui sont de niveau 2 en lecture et 3 en calcul.

b) Les pourcentages marginaux correspondent à la répartition des niveaux en calcul et en lecture.

Répartition des niveaux en calcul :

niveau	1	2	3	4
pourcentage	33 %	48,2 %	15,4 %	3,5 %

Répartition des niveaux en lecture :

niveau	1	2	3	4
pourcentage	14,7 %	50,7 %	26,6 %	8,1 %

c) Proportion d'élèves maîtrisant uniquement les connaissances de base en calcul parmi ceux qui ne maîtrisent pas les compétences de base en lecture : 4,7 % / 14,7 % = 32 %.

d) Proportion d'élèves maîtrisant uniquement les connaissances de base en lecture parmi ceux qui ne maîtrisent pas les compétences de base en calcul : 20,2 % / 33 % = 61,2 %.

55 **a)** 10,2 % n'est pas le total des 3 pourcentages de la colonne « masculin » mais leur moyenne pondérée c'est-à-dire le taux de chômage pour les hommes (de plus de 15 ans).

b) 10,2 % des hommes (de plus de 15 ans) étaient au chômage ;
11,2 % des 25-49 ans et 11,8 % de la population était au chômage.

56 **a)** Distributions des fréquences marginales.

filles	garçons
55,8 %	44,2 %

général	technologique	professionnel
64,1 %	29,6 %	6,3 %

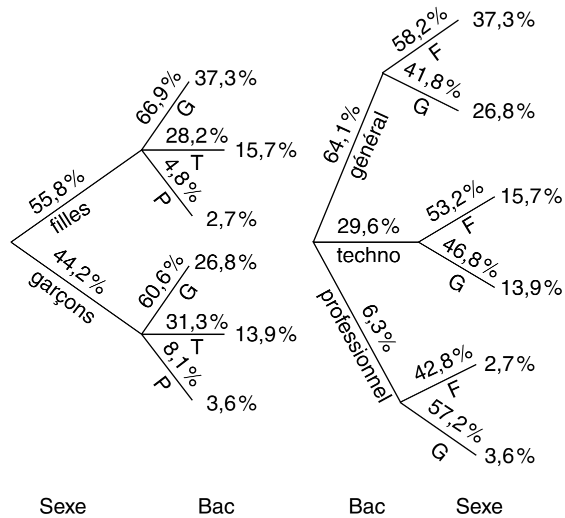
b) Distributions de fréquences conditionnelles :

bac \ sexe	filles	garçons	total
général	58,2 %	41,8 %	100 %
technologique	53,2 %	46,8 %	100 %
professionnel	42,8 %	57,2 %	100 %

bac \ sexe	filles	garçons
général	66,9 %	60,6 %
technologique	28,2 %	31,3 %
professionnel	4,8 %	8,1 %
total	100 %	100 %

c)

bac \ sexe	filles	garçons	total
général	37,3 %	26,8 %	64,1 %
technologique	15,7 %	13,9 %	29,6 %
professionnel	2,7 %	3,6 %	6,3 %
total	55,8 %	44,2 %	100 %



57 **a)** Avec les coefficients 3, 1, 4 et 2 la moyenne du candidat est de 8. Il a échoué.

b) Avec les coefficients 1, 3, 2 et 4 la moyenne du candidat serait de 10,2. Il aurait réussi.

c) Quels que soient les coefficients, sa moyenne sera toujours comprise (strictement) entre 6 et 12. Il ne peut donc pas avoir de mention.

58 **a)** t = 0,9 ‰ × 20 % + 2,2 ‰ × 65 % + 56 ‰ × 15 % = 10,01 ‰.

b) t' = 2,7 ‰ × 52 % + 6,6 ‰ × 47 % + 168 ‰ × 2 % = 7,866 ‰.
Le taux de mortalité global est inférieur au précédent bien qu'il soit le triple dans chaque tranche d'âge. C'est une conséquence de la structure des deux populations.

59 **a)** Mélange 50-50 : l'augmentation est de (50 % × 20 % + 50 % × 10 %) = 15 %.

b) Mélange 20-80 : l'augmentation est de (20 % × 20 % + 80 % × 10 %) = 18 %.
Mélange 80-20 : l'augmentation est de (80 % × 20 % + 20 % × 10 %) = 20 %.

c) La moyenne sera toujours entre 10 % et 20 %.

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

b) 18 ; 21 ; 22 ; 23 ; 24 ; 25 ; 26 ; 28 ; 34.

2. a) $\frac{n}{4} = \frac{9}{4} = 2{,}25$. Donc $i = 3$.

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

3. a) $\frac{3n}{4} = 6{,}75$. Donc $i = 7$.

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

61 **1. a)** Me = 26.

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

c) Min = 5 ; Max = 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.

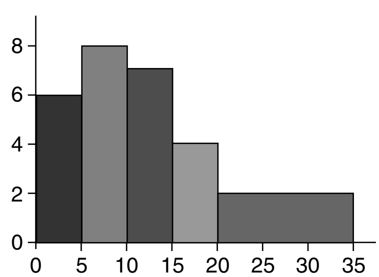
62 **1.**

classe	[0 ; 5[	[5 ; 10[	[10 ; 15[	[15 ; 20[	[20 ; 25[	[25 ; 30[	[30 ; 35[
effectif	6	8	7	4	3	2	1

2. a)

classe	[0 ; 5[	[5 ; 10[	[10 ; 15[	[15 ; 20[	[20 ; 35[
effectif	6	8	7	4	6
largeur de la classe	5	5	5	5	15
hauteur du rectangle	6	8	7	4	2

b)



63 **1. a)**

x_i	2	3	4	5	6	7	8	9	10	11	12
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) $\Sigma n_i x_i = 6\,980$.

$$\bar{x} = \frac{\Sigma n_i x_i}{1\,000} = \frac{6\,980}{1\,000} = 6{,}98.$$

2.

x_i	2	3	4	5	6	7	8	9	10	11	12
$n_i x_i^2$	112	576	1 216	2 550	5 328	8 330	9 088	8 586	7 200	8 954	2 592

$$\sum_{i=1}^{p} n_i x_i^2 = 54\,532.$$

$$s^2 = \frac{\sum_{i=1}^{p} n_i x_i^2}{N} - \bar{x}^2 = \frac{54\,532}{1\,000} - 6{,}98^2 \text{ soit } s^2 \approx 5{,}81.$$

64 **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.

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

65 Les valeurs de la série ordonnée 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 $\frac{11}{4}$ c'est-à-dire 2,75, c'est donc la troisième valeur : $Q_1 = 7$.

Le rang du troisième quartile est le plus petit entier supérieur à $3 \times \frac{11}{4}$ c'est-à-dire 8,25, c'est donc la neuvième valeur : $Q_3 = 19$.

Exercices d'appronfondissement

66

série	moyenne	écart-type	coefficient de variation
S	11	5,85	53,2 %
S_1	110	58,5	53,2 %
S_2	1 100	585	53,2 %

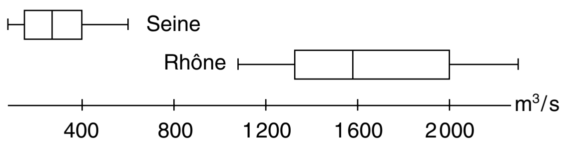
À l'échelle près, on obtient le même diagramme en boîte. Le vérifier à la calculatrice.

67 **1.**

fleuve	moyenne	écart-type	coefficient de variation
Seine	258,3	154,6	59,9 %
Rhône	1 639	406	24,8 %

2. 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.



68 **a)** Les effectifs des classes de collège sont très homogènes. Ceux de 2$^{\text{nde}}$ encore plus (et ils sont plus élevés !). On observe une plus grande hétérogénéité en 1$^{\text{re}}$ et en T$^{\text{ale}}$.

b) Ceci est dû aux différentes séries de 1$^{\text{re}}$ et T$^{\text{ale}}$.

69 **a)** La moyenne est égale à 4,93. Elle appartient 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 se soit déréglée. On poursuit la production.

70 **1. a)**

	agriculteurs	artisans commerçants chefs d'entreprise	professions libérales cadres supérieurs	professions intermédiaires
santé	1,9 %	6,3 %	43,7 %	12,5 %
IUT	3,7 %	8,7 %	25,5 %	20,1 %

	employés	ouvriers	retraités inactifs	indéterminés
santé	5,9 %	5,1 %	4,9 %	19,7 %
IUT	14,5 %	16,9 %	6,4 %	4,2 %

b) Les répartitions sont très différentes suivant les filières.

2. a) Pour la filière « santé », le pourcentage d'étudiants dont les parents sont cadres supérieurs est : $\dfrac{421\,620}{1\,302\,117} = 32{,}4\ \%$.

Le pourcentage d'étudiants dont les parents sont ouvriers est $\dfrac{145\,894}{1\,302\,117} = 11{,}2\ \%$.

b) Les enfants de cadres supérieurs sont donc surreprésentés dans la filière « santé » car il n'y a pas trois fois plus de cadres supérieurs que d'employés dans la population française. C'est plutôt le contraire puisqu'il y avait 29,6 % d'ouvriers et 9,6 % de cadres supérieurs en1989.

71 **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

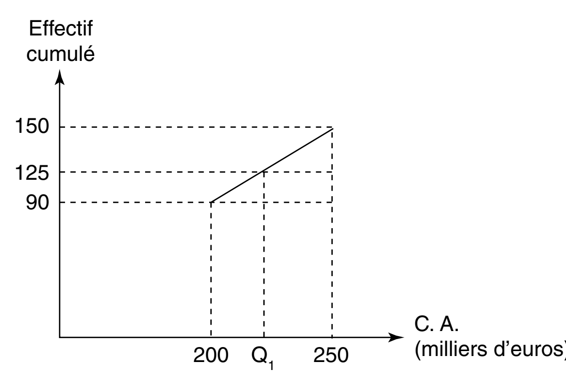
b) L'effectif est égal à 500. Le premier quartile Q_1 se trouve dans la troisième classe car $\dfrac{500}{4} = 125$.

c) Le troisième quartile Q_3 se trouve dans la classe [350 ; 400[car $3 \times \dfrac{500}{4} = 375$.

La médiane M_e se trouve dans la classe [300 ; 350[car $\dfrac{500}{2} = 250$.

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

chiffre d'affaires	200	Q_1	250
effectif cumulé croissant	90	125	150

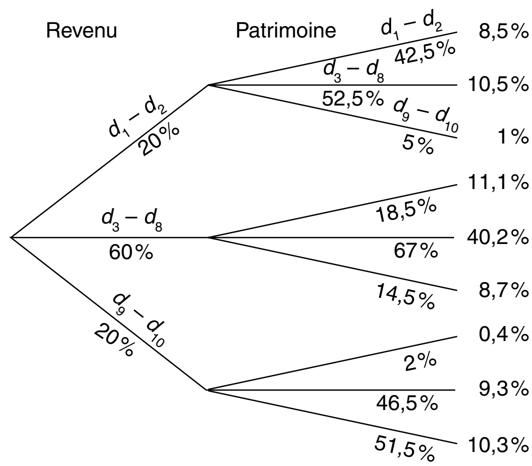


On a donc d'après la propriété de Thalès :
$$\frac{Q_1 - 200}{125 - 90} = \frac{250 - 200}{125 - 90}.$$

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

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

72 a)



b)

décile de revenu	décile de patrimoine			total
	1er et 2^{e}	3^{e} à 8^{e}	9^{e} et 10^{e}	
1er et 2^{e}	42,5 %	52,5 %	5 %	100
3^{e} à 8^{e}	18,5 %	67 %	14,5 %	100
9^{e} et 10^{e}	2 %	46,5 %	51,5 %	100
ensemble	20	60	20	

73 1. a) Le point de départ de cette étude statistique est une variable qualitative à deux modalités (chômage ou pas) étudiée sur la population française. À partir de cettte variable on a construit sur l'ensemble des 22 régions françaises une variable quantitative continue qui est le taux de chômage. Les 22 valeurs de la série statistique obtenue sont alors des pourcentages.

b) La moyenne est environ 11,64 et l'écart-type 2,18.

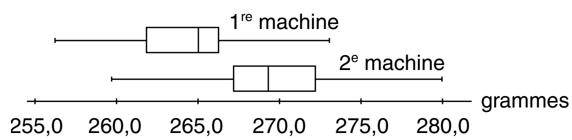
c) Intervalle 1 : [6,1 ; 8,3[Intervalle 2 : [8,3 ; 10,6[
Intervalle 3 : [10,6 ; 12,8[Intervalle 4 : [12,8 ; 15[
Intervalle 5 : [15 ; 17,1[.

d) 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.

74 1. a) Échantillon 1 : $Q_1 = 262$; $Q_3 = 266$.
Échantillon 2 : $Q_1 = 272$; $Q_3 = 267$.

b) Échantillon 1 : $Q_3 - Q_1 = 4$.
Échantillon 2 : $Q_3 - Q_1 = 5$.

2. a)



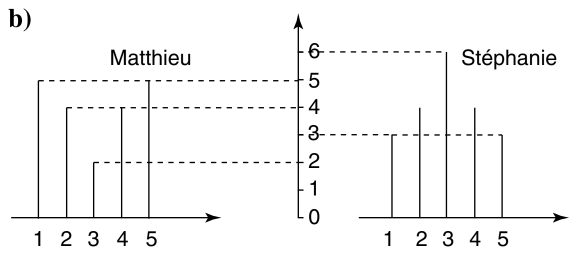
b) La dispersion est une caractéristique de chaque machine. Les deux machines semblent toutes les deux du même niveau de précision. Par contre la deuxième à tendance à produire des paquets trop lourds. Il faudrait effectuer un réglage.

75 a)

	moyenne	médiane	écart-type	étendue	écart interquartile
Matthieu	3	3	4	1,55	3
Stéphanie	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 Matthieu mais cela ne suffit pas pour caractériser les deux élèves.

b)



c) Matthieu a des notes très irrégulières, soit proches du maximum, soit proches du minimum. Une appréciation du style « Du bon et du moins bon » lui irait très bien !
Stéphanie peut être qualifiée de « moyen » car ses notes sont assez groupées autour de la moyenne (3).

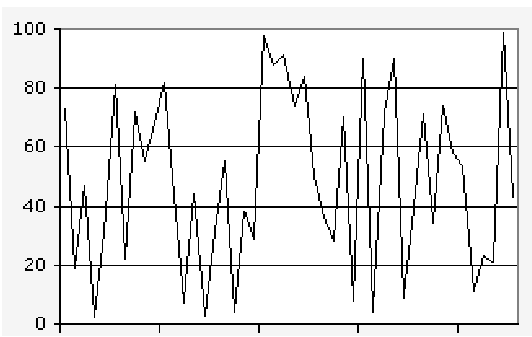
76 a) Pour le régime 1 :
Sur l'ensemble des personnes ayant participé au test, la perte de poids avec le régime 1 a été de 2,240 kg alors qu'elle n'a été que de 1,790 kg avec le régime 2.

b) Pour le régime 2 :
Sur l'ensemble des personnes entre 20 et 30 ans ayant participé au test, la perte de poids avec le régime 2 a été plus importante (1 kg en moyenne) qu'avec le régime 1.
Sur l'ensemble des personnes entre 30 et 50 ans ayant participé au test, la perte de poids avec le régime 2 a été également plus importante (0,4 kg en moyenne) qu'avec le régime 1.

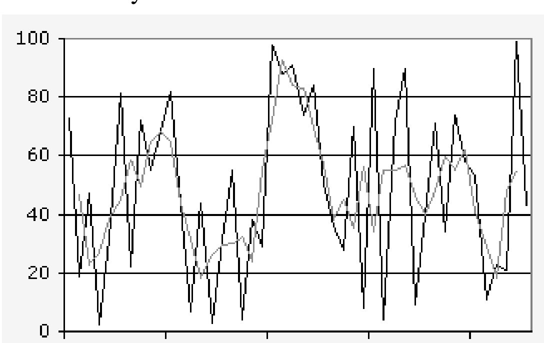
c) Le client se moque de l'effet de structure et quel que soit son âge, il choisira le régime numéro 2.

77 Supposons qu'après le plan social il reste deux cadres au salaire de 47 250 euros et 50 ouvriers au salaire de 16 500 euros. Le salaire moyen dans l'entreprise est de 17 682 euros au lieu de 15 297 euros ce qui représente une augmentation de plus de 15 %.

78 Série brute



Série des moyennes mobiles d'ordre 3



Conclusion : aucune tendance n'apparaît dans la série brute ni dans la série lissée.

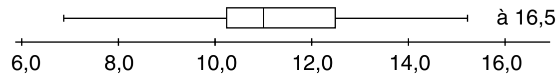
Problème de synthèse

79 **a)**

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

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

b)

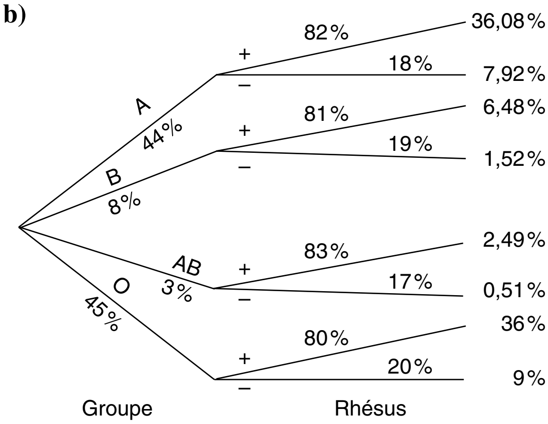


80 Une valeur est aberrante si elle dépasse $Q_3 + 1,5\,(Q_3 - Q_1)$ soit $230 + 1,5\,(230 - 70)$ c'est-à-dire 470.
Il n'y a donc pas de valeur aberrante.

81 **a)**

groupe / rhésus	A	B	AB	O	total
+	36,08 %	6,48 %	2,49 %	36 %	81,05 %
–	7,92 %	1,52 %	0,51 %	9 %	18,95 %
total	44 %	8 %	3 %	45 %	100 %

b)



82 **1. a)** La moyenne théorique est de 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 pour la série S_1 est de 2,9.

2. a) Les 25 moyennes sont : 4,50 ; 5,75 ; 1,25 ; 4,25 ; 4,25 ; 5,00 ; 6,50 ; 6,25 ; 3,75 ; 6,50 ; 3,75 ; 5,25 ; 4,50 ; 3,00 ; 7,00 ; 2,75 ; 2,50 ; 4,50 ; 5,75 ; 5,50 ; 4,25 ; 3,75 ; 5,00 ; 5,75 ; 5,50.

b) La moyenne de cette série S_2 est encore 4,67 ce qui était prévisible.

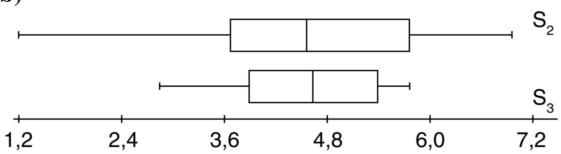
c) L'écart-type de cette série S_2 est égal à 1,4.

d) Il est inférieur à celui de la série S_1 (environ la moitié), ce qui était prévisible puisqu'il y a moins de variation dans des moyennes de 4 valeurs que dans des valeurs individuelles.

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

La moyenne est encore 4,67 ce qui était prévisible.
L'écart-type de 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.

b)



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

83

	moyenne	médiane	d_1	d_9	d_9/d_1
1985	8 281	6 950	3 489	13 675	3,92
1995	9 545	7 526	3 823	16 559	4,33

1. La moyenne est plus grande que la médiane car les gros salaires « tirent la moyenne vers le haut.

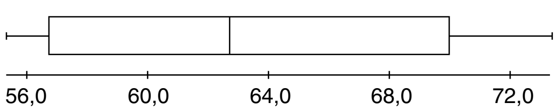
2. a) L'augmentation est de 15,3 % pour la moyenne et de 8,3 % pour la médiane.

b) L'augmentation est de 9,6 % pour d_1 et de 21,1 % pour d_9.
Ceci signifie que les gros salaires augmentent plus vite que les bas salaires.

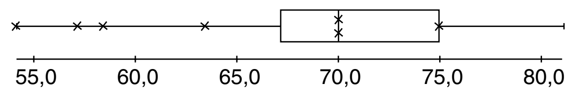
c) L'augmentation annuelle moyenne pour d_1 est : 0,9 %.
L'augmentation annuelle moyenne pour d_9 est : 1,9 %.

d) Le rapport interdécile a augmenté entre 1985 et 1995 (4,33 > 3,92). L'éventail des salaires s'est élargi pendant cette période.

84 **1.**



2. a)

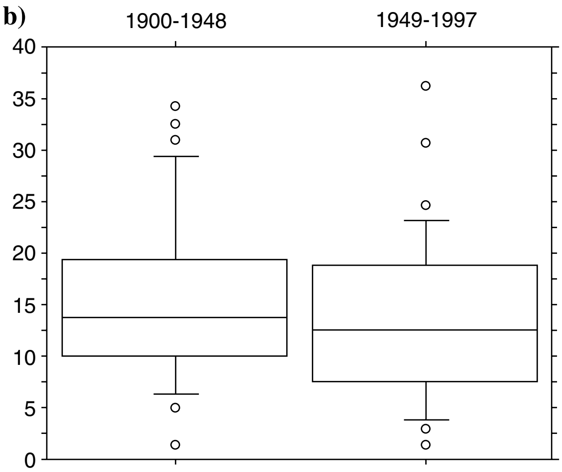


b) Presque tous les incidents (7 sur 8) ont eu lieu pour des températures inférieures ou égales à la médiane.
On peut aussi remarquer que le pourcentage de tirs sans incident en dessous de 66° F est égal à 0 % !
Pour finir, la température prévue (29° F) est très éloignée des températures pour lesquelles 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.

85 **a)**

	min	Q_1	M_e	Q_3	max	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.