# Chapitre 5

# Statistiques

## Sommaire



## 5.1 Rappels de Seconde

### 5.1.1 Vocabulaire

**Définition 5.1.** Une série statistique est un ensemble d'observations collectées et on a les définitions suivantes :

**Population :** C'est l'ensemble sur lequel porte une étude statistique ;

**Individu :** C'est un élément de la population ;

**Caractère :** C'est ce qu'on observe chez l'individu ;

**Modalité :** Ce sont les différentes valeurs prises par le caractère ;

**Quantitative ou qualitative :** La série statistique est dite *quantitative* quand les modalités sont des nombres et *qualitative* sinon ;

**Discrète ou continue :** Dans le cas d'une série quantitative, celle-ci est dite *discrète* si les modalités sont limitées à un ensemble fini de valeurs et *continue* si les modalités peuvent prendre n'importe quelle valeur dans un intervalle.

**Exemples.**

- On peut s'intéresser à une classe (population), comportant des élèves (individus) et observer leur nombre de frères et sœurs (caractère) qui peuvent être $0, 1, 2, \ldots$ (modalités), ces données formant alors une série statistique quantitative discrète.
- On peut s'intéresser à une chaîne d'usine produisant des bras de suspension pour voiture (population), et observer sur chaque pièce (individu) ses dimensions exactes (caractère) qui peuvent varier entre 500 et 750 mm (modalités), ces données formant alors une série statistique quantitative continue.
- On peut s'intéresser à la population française (population dont on prendra un échantillon) comportant des individus (individus) et estimer leur intention de vote (caractère) pouvant être n'importe lequel des candidats se présentant (modalités), ces données formant alors une série statistique qualitative.

**Définition 5.2.** On a aussi :

**Effectif d'une modalité :** C'est le nombre de fois que la modalité d'un caractère revient dans la série ;

**Fréquence d'une modalité :** C'est l'effectif de la modalité divisé par l'effectif total ; elle est comprise entre 0 et 1 ;

**Classes :** S'il y a trop de modalités différentes, elles sont rangées par *classe* (intervalle), l'effectif de la classe étant alors le nombre de modalités appartenant à cet intervalle.

### 5.1.2 Mesures centrales

Elles visent à résumer la série par une seule valeur représentative, d'une certaine manière, de toutes les valeurs de la série.

#### Moyenne arithmétique

**Définition 5.3** (Moyenne arithmétique)**.** La *moyenne arithmétique* d'une série statistique quantitative $S = \{x_1, x_2, \ldots, x_n\}$ est le nombre, noté $\overline{x}$ :

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

*Remarque.* De la définition, on peut déduire que $n\overline{x} = x_1 + x_2 + \ldots + x_n$, ce qui peut s'interpréter de la manière suivante : « La somme de toutes les valeurs de la série est inchangée si on remplace chaque valeur par $\overline{x}$ ».

La moyenne a des avantages calculatoires : si l'on connaît les moyennes et les effectifs de deux séries (ou deux sous-séries), on peut obtenir la moyenne de la série constituée de l'agrégation de ces deux séries. Elle a le défaut d'être sensible aux valeurs extrêmes.

#### Médiane

**Définition** (Médiane dans le cas général)**.** On appelle *médiane* d'une série statistique quantitative tout nombre $m$ tel que :
- la moitié au moins des valeurs de la série est inférieure à $m$
- la moitié au moins des valeurs de la série est supérieure à $m$

*Remarques.*
- Rappel : mathématiquement « inférieur » et « supérieur » signifient, en français, « inférieur ou égal » et « supérieur ou égal ».

- On admettra qu'un tel nombre existe toujours.
- La médiane partage la série en deux sous-séries ayant *quasiment* le même effectif; *quasiment* car si plusieurs valeurs de la série sont égales à la médiane, les données inférieures à la médiane et les données supérieures à la médiane ne seront pas forcément en nombre égal.
- Il faut comprendre la médiane comme « la valeur du milieu ».

Plusieurs valeurs peuvent parfois convenir pour la médiane, aussi convient-on de prendre, dans le cadre scolaire [1], les valeurs, uniques, suivantes :

**Définition 5.4** (Médiane dans le cadre scolaire)**.** Soit une série statistique quantitative comportant $n$ données : $S = \{x_1, x_2, \dots, x_i, \dots, x_n\}$ telles que $x_1 \leqslant x_2 \leqslant \dots \leqslant x_n$.

**Si $n$ est impair :** Il n'y a qu'une médiane possible et c'est celle qui est située au rang $\frac{n+1}{2}$.

$$m = \left(\frac{n+1}{2}\right)^{\text{ième}}$$

**Si $n$ est pair :** Il y a plusieurs médianes possibles : tout nombre compris entre le $\frac{n}{2}^{\text{ième}}$ élément de la série et le suivant est **une** médiane. Dans le cadre scolaire **la** médiane sera la moyenne des deux données centrales de la série :

$$m = \frac{\left(\frac{n}{2}\right)^{\text{ième}} + \left(\frac{n}{2}+1\right)^{\text{ième}}}{2}$$

C'est cette médiane qui sera attendue systématiquement dans les exercices et les évaluations.

La médiane a l'avantage de ne pas être influencée par les valeurs extrêmes. Elle n'a aucun avantage pratique dans les calculs, puisque, pour connaître la médiane d'une série constituée de l'agrégation de deux séries, il faut nécessairement re-ordonner la nouvelle série pour trouver sa médiane, qui n'aura pas de lien avec les deux médianes des deux séries initiales.

### 5.1.3 Mesures de dispersion

Elles visent à indiquer comment les données de la série statistique sont dispersées par rapport aux mesures centrales.

#### Valeurs extrêmes

**Définition 5.5.** Les valeurs extrêmes d'une série quantitative sont ses valeurs *minimale* et *maximale* et l'*étendue* est la différence entre les valeurs extrêmes de la série.

1. Les statisticiens, eux, prennent n'importe quel nombre convenant parmi les médianes possibles ; sur des séries de grande taille, ils ont tous le même ordre de grandeur

### Quartiles

**Définition** (Quartiles dans le cas général)**.** Soit $S$ une série statistique quantitative.

- On appelle *premier quartile*, noté $Q_1$, tout réel tel que
  - au moins 25% des valeurs de la série ont une valeur inférieure ou égale à $Q_1$
  - au moins 75% des valeurs de la série ont une valeur supérieure ou égale à $Q_1$
- On appelle *deuxième quartile*, noté $Q_2$, tout réel tel que
  - au moins 50% des valeurs de la série ont une valeur inférieure ou égale à $Q_2$
  - au moins 50% des valeurs de la série ont une valeur supérieure ou égale à $m$
- On appelle *troisième quartile*, noté $Q_3$, tout réel tel que
  - au moins 75% des valeurs de la série ont une valeur inférieure ou égale à $Q_3$
  - au moins 25% des valeurs de la série ont une valeur supérieure ou égale à $Q_3$

*Remarques.*

- $Q_2$ est, par définition, la médiane de la série.
- On admettra que de tels nombres existent toujours.
- La médiane partage une série en deux sous-séries ayant quasiment le même effectif (environ 50 %) ; les premier, troisième quartiles et la médiane partageront une série en quatre sous-séries ayant quasiment le même effectif (environ 25 %).

Comme pour la médiane, selon le nombre $n$ de données dans la série, il y a parfois plusieurs possibilités. Aussi on convient de prendre, dans le cadre scolaire [2], systématiquement les nombres suivants :

**Définition 5.6** (Quartiles dans le cadre scolaire)**.** Soit $S$ une série statistique quantitative dont les données sont ordonnées dans l'ordre croissant. On appelle :

- *premier quartile*, noté $Q_1$, **la première valeur de la série** telle qu'au moins 25 % des valeurs de la série ont une valeur inférieure ou égale à $Q_1$ ;
- *troisième quartile*, noté $Q_3$, **la première valeur de la série** telle qu'au moins 75 % des valeurs de la série ont une valeur inférieure ou égale à $Q_3$.

Ce sont ces quartiles qui seront attendus systématiquement dans les exercices et les évaluations.

*Remarque.* Si l'on adopte le même type de définition pour le deuxième quartile on ne tombe pas forcément sur la valeur de la médiane telle que définie dans le cadre scolaire.
Par exemple la série $S = \{1 ; 2 ; 3 ; 4\}$ a pour médiane $m = \frac{2+3}{2} = 2,5$ et pour deuxième quartile $Q_2 = 2$ car c'est la première valeur de la série telle que au moins 50 % des valeurs de la série lui sont inférieures.

La propriété suivante permet de trouver aisément $Q_1$ et $Q_3$ :

**Propriété 5.1.** *Soit une série statistique quantitative comportant $n$ données : $S = \{x_1, x_2, \dots, x_i, \dots, x_n\}$ telles que $x_1 \leqslant x_2 \leqslant \dots \leqslant x_n$. Alors :*

- *La donnée de rang $\frac{1}{4}n$ (arrondi au supérieur si ce n'est pas un entier) convient toujours comme premier quartile.*
- *La donnée de rang $\frac{3}{4}n$ (arrondi au supérieur si ce n'est pas un entier) convient toujours comme troisième quartile.*

On l'admettra.

2. Ce sont aussi ces quartiles que prennent les statisticiens

**Exemples 5.1.**

- S'il y a $n = 29$ données dans la série, rangées dans l'ordre croissant :
  - $\frac{1}{4} \times 29 = 7,25 \approx 8$ donc la huitième donnée de la série convient comme premier quartile ;
  - $\frac{3}{4} \times 29 = 21,75 \approx 22$ donc la vingt-deuxième donnée de la série convient comme troisième quartile.
- S'il y a $n = 64$ données dans la série, rangées dans l'ordre croissant :
  - $\frac{1}{4} \times 64 = 16$ donc la seizième donnée de la série convient comme premier quartile ;
  - $\frac{3}{4} \times 64 = 48$ donc la quarante huitième donnée de la série convient comme troisième quartile.

#### Interquartiles

Une fois les premier et troisième quartiles disponibles, on définit l'écart et l'intervalle interquartile de la manière suivante :

**Définition 5.7.** Soit $S$ une série statistique quantitative et $Q_1$ et $Q_3$ ses premier et troisième quartiles. On appelle :
- *écart interquartile* la différence $Q_3 - Q_1$ ;
- *intervalle interquartile* l'intervalle $[Q_1 ; Q_3]$.

## 5.2 Découverte d'un nouvel indicateur de dispersion

### 5.2.1 Un problème

Soit $S_1$, $S_2$ et $S_3$ les trois séries de notes suivantes :
- $S_1 = \{2; 2; 4; 4; 6; 10; 14; 16; 16; 18; 18\}$
- $S_2 = \{2; 8; 9; 10; 10; 10; 11; 12; 18\}$
- $S_3 = \{2; 5; 5; 5; 5; 6; 6; 7; 9; 10; 11; 13; 14; 14; 15; 15; 15; 15; 18\}$

1. Pour chacune de ces séries, déterminer les valeurs extrêmes, la moyenne et la médiane. Que constate-t-on ?
2. Ces trois séries n'ont pas la même allure.
   (a) Déterminer les premier et troisième quartiles de chaque série. Témoignent-ils des allures de chaque série ?
   (b) Inventer des mesures, basées sur les distances entre les valeurs centrales de la série et chacun des nombres de la série qui peuvent être mesurées par :
   - la distance entre $a$ et $b$ est donnée par $b - a$ quand $a < b$, et par $a - b$ sinon
   - la distance entre $a$ et $b$ est donnée par $(a - b)^2$

### 5.2.2 Bilan et compléments

#### Variance, écart-type

**Définition 5.8.** Soit $S$ une série statistique quantitative comportant $n$ données : $S = \{x_1, x_2, \dots, x_i, \dots, x_n\}$.
On appelle :
- *moyenne arithmétique* de $S$, notée $\overline{x}$, le nombre $\overline{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$.
- *variance* de $S$, notée $V$, le nombre $V = \frac{(\overline{x} - x_1)^2 + \dots + (\overline{x} - x_n)^2}{n}$
- *écart-type* de $S$, noté $s$, le nombre $s = \sqrt{V}$.

*Remarque.* La variance est la moyenne des carrés des écarts à la moyenne. Elle mesure donc la dispersion des valeurs autour de la moyenne. Elle n'est pas très parlante car elle s'exprime dans le carré de l'unité du caractère.
L'écart-type a l'avantage de s'exprimer dans la même unité que le caractère.
L'écart-type permet de comparer la dispersion de deux séries, quand l'ordre de grandeur des données des deux séries est le même. Contrairement à l'écart interquartile, il tient compte de l'ensemble de la population.

On dispose d'une formule plus pratique pour calculer la variance :

**Théorème 5.2.** *Soit $S$ une série statistique quantitative comportant $n$ données : $S = \{x_1, x_2, \dots, x_i, \dots, x_n\}$, de moyenne arithmétique $\overline{x}$, alors :*

$$V = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - \overline{x}^2$$

La preuve sera faite en classe.

# 5.3 Représentations graphiques

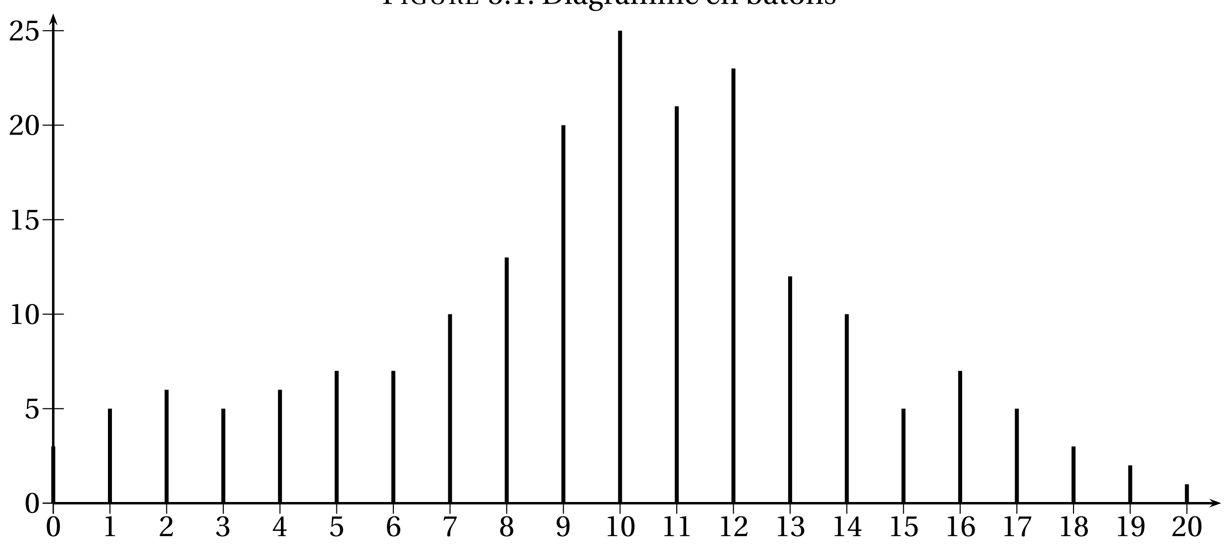
## 5.3.1 Diagramme à bâtons

Si on considère la série :

| $x_i$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_i$ | 3 | 5 | 6 | 5 | 6 | 7 | 7 | 10 | 13 | 20 | 25 | 21 | 23 | 12 | 10 | 5 | 7 | 5 | 3 | 2 | 1 |

On obtient le diagramme en bâtons de la figure 5.1, de la présente page, où les modalités sont en abscisse et les effectifs en ordonnée.

FIGURE 5.1: Diagramme en bâtons



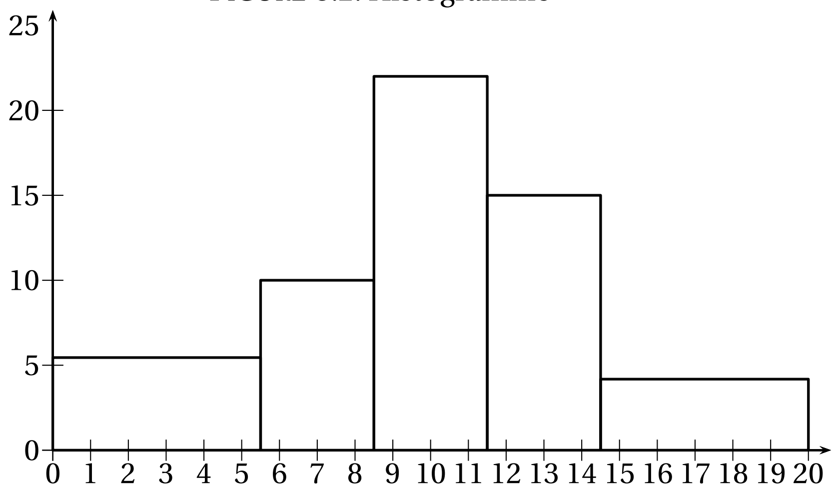
### 5.3.2 Histogramme

Si les données sont regroupées en classes (intervalles), la série peut-être représentée par un histogramme où chaque rectangle a son *aire* proportionnelle à l'effectif (ou à la fréquence) de la classe. Par exemple, la même série que précédemment mais regroupée en classe donne :

| $x_i$ | $[0;5]$ | $]5;8]$ | $]8;11]$ | $]11;14]$ | $]14;20]$ |
|---|---|---|---|---|---|
| $n_i$ | 30 | 30 | 66 | 45 | 23 |

Elle donne l'histogramme de la figure 5.2, de la présente page, les calculs pour obtenir les hauteurs des rectangles étant détaillés sous le graphique.

FIGURE 5.2: Histogramme

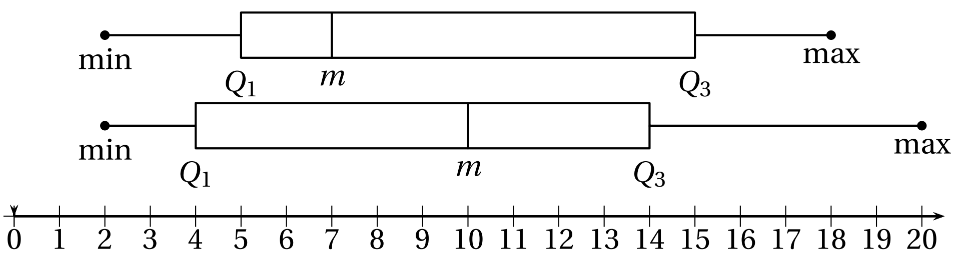



| $x_i$ | $[0;5,5]$ | $]5,5;8,5]$ | $]8,5;11,5]$ | $]11,5;14,5]$ | $]14,5;20]$ |
|---|---|---|---|---|---|
| $n_i$ | 30 | 30 | 66 | 45 | 23 |
| Aire = $n_i$ | 30 | 30 | 66 | 45 | 23 |
| Largeur = amplitude de la classe | 5,5 | 3 | 3 | 3 | 5,5 |
| Hauteur = $\dfrac{\text{Aire}}{\text{Largeur}}$ | 5,45 | 10 | 22 | 15 | 4,18 |

### 5.3.3 Diagramme en boite

On peut représenter graphiquement les valeurs extrêmes, les quartiles et la médiane par un diagramme en boite, appelé aussi boite à moustaches, conçues de la manière suivante :

- au centre une boite allant du premier au troisième quartile, séparée en deux par la médiane ;
- de chaque côté une moustache allant du minimum au premier quartile pour l'une, et du troisième quartile au maximum pour l'autre.



Ces diagrammes permettent une interprétation visuelle et rapide de la dispersion des séries statistiques. Ils permettent également d'apprécier des différences entre des séries (lorsqu'elles ont des ordres de grandeurs comparables).

*Remarques.*

- La hauteur des boites est arbitraire (on les fait parfois proportionnelles à l'effectif total de la série).
- La boite contient les 50% des données centrales.

## 5.4 Exercices

**EXERCICE 5.1.**
Dans chaque cas, calculer la moyenne et la médiane de la série, et conseiller le narrateur sur la meilleure stratégie pour minimiser son résultat auprès de ses parents qui ne connaissent rien en statistique :

1. « Je n'ai eu que 8 sur 20 au contrôle de statistiques. Nous sommes 10 en classe. La meilleure note est 19. Ensuite il y a un 10, quatre 9, un 8 (moi) et trois 2. »
2. « Encore un 8 ! Cette fois les notes sont 2, 3, 4, 5, 7, 8 (moi), 9, 9, 18, 19. »
3. « Toujours un 8 ! Cette fois il y a eu trois 7 et un 19, 18, 12, 11, 10, 8 (moi) et 2. »

**EXERCICE 5.2.**
On donne la série suivante :

11, 12, 13, 4, 17, 5, 13, 13, 5, 6, 6, 10, 10, 8, 9, 9, 11, 11, 14, 5, 14, 9, 9, 15, 7, 8, 15.

1. Déterminer la moyenne de la série.
2. Déterminer la médiane et les quartiles de la série.
3. Représenter le diagramme en boîte correspondant.
4. Quel est l'écart interquartile de la série ?
5. Quel est l'intervalle interquartile de la série ?

**EXERCICE 5.3.**
Dans une classe, les notes sont les suivantes :

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20.

1. Déterminer la moyenne de la série.
2. Déterminer la médiane et les quartiles de la série.
3. Représenter le diagramme en boîte correspondant.
4. Que remarque-t-on sur ce diagramme ? Pouvait-on s'y attendre ?

*On gardera en tête l'allure de cette série « canonique » où les données sont parfaitement réparties qui pourra servir de référence pour décrire les autres séries : plus on s'éloigne de ce cas particulier, plus on pourra parler « d'irrégularité » de dispersion.*

**EXERCICE 5.4.**
On a relevé le prix de vente d'un CD et le nombre de CD vendus chez différents fournisseurs. Les résultats forment une série statistique à une variable donnée par le tableau ci-dessous.

| | | | | | |
|---|---|---|---|---|---|
| Prix de vente (en €) | 15 | 16 | 17 | 18 | 19 |
| Nombre de CD vendus | 83 | 48 | 32 | 20 | 17 |

1. Quelles sont les différentes valeurs de la série.
2. Donner la fréquence correspondant à chacune de ces valeurs.
3. Donner la moyenne et la médiane de la série. Que représentent ces nombres ?
4. Représenter la série par un diagramme semi-circulaire.

**EXERCICE 5.5.**
Dans deux entreprises $A$ et $B$, les employés sont classés en deux catégories : ouvriers et cadres.
Le tableau 5.1, de la présente page, donne la répartition des employés en fonction de leur catégorie professionnelle et de leur salaire mensuel net, en euros.

TABLE 5.1: Salaires des entreprise $A$ et $B$ de l'exercice 5.5

**Entreprise** $A$

| Salaires | 1 500 | 2 500 | 3 500 |
|---|---|---|---|
| Ouvriers | 114 | 66 | 0 |
| Cadres | 0 | 8 | 12 |

**Entreprise** $B$

| Salaires | 1 500 | 2 500 | 3 500 |
|---|---|---|---|
| Ouvriers | 84 | 42 | 0 |
| Cadres | 0 | 12 | 12 |

1. (a) Calculer la moyenne des salaires de tous les employés de l'entreprise $A$
   (b) Calculer la moyenne des salaires des ouvriers de l'entreprise $A$
   (c) Calculer la moyenne des salaires des cadres de l'entreprise $A$
2. Faire les mêmes calculs pour l'entreprise $B$
3. Le P.D.G. de l'entreprise $A$ dit à celui de l'entreprise $B$ : « Mes employés sont mieux payés que les vôtres. »
   « Faux »répond ce dernier, « mes ouvriers sont mieux payés et mes cadres également. »
   Expliquer ce paradoxe.

**EXERCICE 5.6.**
Le recensement de 1 999 a permis d'observer l'existence en France de dix communes de plus de 200 000 habitants. Le tableau ci-dessous donne la liste de ces communes et de leurs populations en milliers d'habitants.

| | |
|---|---|
| Paris | 2 116 |
| Marseille | 798 |
| Lyon | 445 |
| Toulouse | 391 |
| Nice | 341 |
| Nantes | 269 |
| Strasbourg | 264 |
| Montpellier | 225 |
| Bordeaux | 215 |
| Rennes | 206 |

1. Quelle est l'étendue de cette série statistique ? Que devient l'étendue quand on retire Paris de la liste ?
2. Comparer moyenne et médiane des populations de ces dix villes. Que constate-ton ? Comment expliquer ceci ?
3. Faire de même si on retire Paris de la liste. Que constatez-vous maintenant ?
4. Que choisiriez-vous entre « moyenne étendue »et « médiane étendue »pour résumer cette série statistique ? Expliquez votre choix.

**EXERCICE 5.7.**
Les tableaux dont il est question dans cet exercice sont ceux de la page de la présente page.

1. Le tableau 1 présente la répartition des salaires mensuels dans une entreprise (*source : DoC TICE-MEN*).
   Calculer la moyenne des salaires de cette entreprise et déterminer la médiane.
2. Une erreur a été commise dans les relevés : les effectifs correspondant aux salaires de 1 100 € et 1 400 € ont été permutés. Les données exactes sont celles du tableau 2.
   Comparer les moyenne et médiane de cette série à celles de la précédente. Les variations étaient-elles prévisibles ?
3. Rêvons … 35 personnes ont un salaire de 3 400 € et l'effectif total est inchangé. Utiliser le tableau 3 pour imaginer une répartition des effectifs telle que la médiane ne soit pas modifiée : comment va varier la moyenne ?
   Quelle répartition des effectifs respectant ces contraintes donnera la moyenne des salaires la plus élevée (calculer alors cette moyenne) ? Quelle répartition des effectifs respectant ces contraintes donnera la moyenne des salaires la moins élevée (calculer alors cette moyenne) ?
   Mathias prétend qu'avec 35 personnes ayant un salaire de 3 400 €, la moyenne est obligatoirement plus élevée. A-t-il raison ?

TABLE 5.2: Données de l'exercice 5.7

**Tableau 1**

| Salaire en € | Effectif | Effectif cumulé croissant |
|---|---|---|
| 1 100 | 18 | 18 |
| 1 200 | 15 | 33 |
| 1 300 | 20 | 53 |
| 1 400 | 10 | 63 |
| 1 500 | 25 | 88 |
| 1 600 | 12 | 100 |
| 1 700 | 4 | 104 |
| 1 800 | 5 | 109 |
| 1 900 | 3 | 112 |
| 2 000 | 2 | 114 |
| 2 100 | 6 | 120 |
| 2 200 | 7 | 127 |
| 2 300 | 0 | 127 |
| 2 400 | 2 | 129 |
| 2 500 | 0 | 129 |
| 2 600 | 3 | 132 |
| 2 700 | 0 | 132 |
| 2 800 | 3 | 135 |
| 2 900 | 0 | 135 |
| 3 000 | 0 | 135 |
| 3 100 | 3 | 138 |
| 3 200 | 0 | 138 |
| 3 300 | 5 | 143 |
| 3 400 | 8 | 151 |

**Tableau 2**

| Salaire en € | Effectif | Effectif cumulé croissant |
|---|---|---|
| 1 100 | **10** | |
| 1 200 | 15 | |
| 1 300 | 20 | |
| 1 400 | **18** | |
| 1 500 | 25 | |
| 1 600 | 12 | |
| 1 700 | 4 | |
| 1 800 | 5 | |
| 1 900 | 3 | |
| 2 000 | 2 | |
| 2 100 | 6 | |
| 2 200 | 7 | |
| 2 300 | 0 | |
| 2 400 | 2 | |
| 2 500 | 0 | |
| 2 600 | 3 | |
| 2 700 | 0 | |
| 2 800 | 3 | |
| 2 900 | 0 | |
| 3 000 | 0 | |
| 3 100 | 3 | |
| 3 200 | 0 | |
| 3 300 | 5 | |
| 3 400 | 8 | |

**Tableau 3**

| Salaire en € | Effectif | Effectif cumulé croissant |
|---|---|---|
| 1 100 | | |
| 1 200 | | |
| 1 300 | | |
| 1 400 | | |
| 1 500 | | |
| 1 600 | | |
| 1 700 | | |
| 1 800 | | |
| 1 900 | | |
| 2 000 | | |
| 2 100 | | |
| 2 200 | | |
| 2 300 | | |
| 2 400 | | |
| 2 500 | | |
| 2 600 | | |
| 2 700 | | |
| 2 800 | | |
| 2 900 | | |
| 3 000 | | |
| 3 100 | | |
| 3 200 | | |
| 3 300 | | |
| 3 400 | **35** | **151** |

**EXERCICE 5.8.**
Des salariés d'une entreprise se sont réunis dans un restaurant où ils sont les seuls clients pour fêter le changement de leur grille de salaire : désormais ils touchent tous 1 700 € par mois.

1. Quelle est la moyenne et la médiane de la série constituée par leurs salaires.
2. Bill Gates, qui a un salaire bien plus grand, entre dans le restaurant : que devient la moyenne et la médiane de la série constituée par leurs salaires et celui de Bill Gates.

**EXERCICE 5.9.**
D'après Wikipédia, le salaire médian des salariés de 25 à 55 ans en France en 2008 était de 1 655 € nets et le salaire moyen de 2 069 € nets.
Conjecturer ce qui peut expliquer cette différence.

**EXERCICE 5.10.**
Trois séries statistiques, comportant 10 données chacune, ont les paramètres suivants :

- Série A : Minimum 10; maximum 50; moyenne 28; médiane 20.
- Série B : Minimum 10; maximum 50; moyenne 30; médiane 30.
- Série C : Minimum 10; maximum 50; moyenne 21,5; médiane 25.

Conjecturer pour chacune de ces séries comment peuvent être réparties les données.

**EXERCICE 5.11.**
Le tableau est le relevé de trois séries de notes obtenues en mathématiques dans des classes de seconde (à effectifs très réduits) lors d'un contrôle sur les statistiques.

| Classe 1 | Classe 2 | Classe 3 |
|---|---|---|
| 2 | 2 | 8 |
| 5 | 3 | 8 |
| 6 | 4 | 9 |
| 7 | 4 | 9 |
| 9 | 4 | 10 |
| 10 | 9 | 10 |
| 10 | 10 | 10 |
| 12 | 12 | 12 |
| 13 | 12 | 13 |
| 14 | 12 | 15 |
| 15 | 12 | 15 |
| 15 | 12 | 16 |
| 16 | 13 | 17 |
| 19 | 13 | 18 |

1. Déterminer la note médiane de chaque classe.
2. **SANS LA CALCULER**, conjecturer pour chaque série si la moyenne sera supérieure, inférieure ou proche de la médiane.

**EXERCICE 5.12.**
Dans une classe de 30 élèves, la moyenne des 20 filles est 11,5 et la moyenne des 10 garçons est 8,5.
Donner la moyenne de classe.

**EXERCICE 5.13.**
On donne la série suivante : 4, 5, 5, 5, 6, 6, 7, 8, 8, 9, 9, 9, 9, 10, 10, 11, 11, 11, 12, 13, 13, 13, 14, 14, 15, 15, 17.

1. Représenter le diagramme en boîte correspondant.
2. Quel est l'écart interquartile de la série ?
3. Quel est l'intervalle interquartile de la série ?

**EXERCICE 5.14.**
Le tableau ci-dessous donne une répartition des salaires mensuels en euros des employés dans une entreprise.

| Salaire | [1 000 ; 1 200[ | [1 200 ; 1 500[ | [1 500 ; 2 000[ | [2 000 ; 3 000[ | [3 000 ; 10 000[ |
|---|---|---|---|---|---|
| Effectif | 326 | 112 | 35 | 8 | 3 |

1. Quel est le nombre d'employés de l'entreprise ?
2. Quel est le nombre d'employés touchant un salaire mensuel supérieur ou égal à 1 200 € ?
3. Représenter les données par un histogramme.
4. Quel est le salaire moyen des employés de l'entreprise ?

**EXERCICE 5.15.**
Un entomologiste a fait des relevés sur la taille de 50 courtilières adultes.
33, 35, 36, 36, 37, 37, 37, 38, 38, 38, 39, 39, 39, 39, 40, 40, 40, 40, 40, 41, 41, 41, 41, 41, 41, 41, 42, 42, 42, 42, 42, 42, 43, 43, 43, 43, 44, 44, 44, 44, 45, 45, 45, 46, 46, 47, 47, 48, 48, 50.

1. Organiser les relevés dans le tableau d'effectifs suivant :

| Valeur | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Effectif | | | | | | | | | | | | | | | | | | |
| Effectif cumulé croissant | | | | | | | | | | | | | | | | | | |

2. Représenter les données par un diagramme à bâtons. Un diagramme circulaire serait-il intéressant ?
3. Calculer la moyenne de la série. Déterminer sa médiane. Déterminer les premier et troisième quartiles.
4. Construire le diagramme en boîte correspondant.
5. On regroupe les données en classes. Compléter le tableau des effectifs suivants :

| Valeur | [33 ; 37[ | [37 ; 40[ | [40 ; 42[ | [42 ; 44[ | [44 ; 47[ | [47 ; 51[ |
|---|---|---|---|---|---|---|
| Effectif | | | | | | |

Dessiner l'histogramme correspondant.

**EXERCICE 5.16.**
Le tableau suivant donne les effectifs des notes obtenues dans une classe en Mathématiques et en Histoire-Géographie :

| Notes | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Maths | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 3 | 4 | 4 | 1 | 3 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| H.-G. | 0 | 1 | 0 | 0 | 2 | 0 | 1 | 2 | 1 | 1 | 4 | 2 | 2 | 0 | 3 | 2 | 1 | 0 | 1 | 0 | 1 |

Le but de l'exercice est de comparer la dispersion des notes en Maths et en Histoire-Géographie.

1. Utilisation des quartiles
   (a) Calculer la médiane et l'écart interquartile en Maths.
   (b) Calculer la médiane et l'écart interquartile en Histoire-Géographie.
   (c) Représenter les diagrammes en boîte des notes en Maths et en Histoire-Géographie.
   (d) Interpréter ces résultats.
2. Utilisation des écarts-types
   (a) Calculer la moyenne et l'écart-type en Maths.
   (b) Calculer la moyenne et l'écart-type en Histoire-Géographie.
   (c) Interpréter ces résultats.

**EXERCICE 5.17.**
Le tableau suivant donne les effectifs des notes obtenues dans une classe en Mathématiques et en Sciences de la vie et de la terre :

| Note | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Maths | 0 | 1 | 0 | 2 | 0 | 1 | 3 | 0 | 5 | 0 | 2 | 1 | 1 | 0 | 0 | 0 | 2 | 1 | 1 | 0 | 2 |
| SVT | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 4 | 4 | 1 | 0 | 3 | 1 | 4 | 2 | 0 | 1 | 0 | 0 | 0 | 1 |

1. (a) Déterminer les rangs des quartiles et de la médiane des deux séries puis donner leurs valeurs.
   (b) Faire le diagramme en boite des deux séries sur une même échelle puis commenter le diagramme.
2. (a) Déterminer la moyenne des notes de Mathématiques et la comparer à la médiane en expliquant ce qui cause cet écart.
   (b) Déterminer la moyenne des notes de Sciences de la vie et de la terre. L'écart avec la moyenne est-il le même ? Pourquoi ?
   (c) Déterminer les écarts types des deux séries (arrondis au dixième) et commenter vos résultats.