Chapitre 5

Statistiques à une variable

Sommaire

5.1 Généralités

5.1.1 Définitions

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

Définition 5.1. *Population* : C'est l'ensemble sur lequel porte une étude statistique ; une *sous-population* est alors un *sous-ensemble* de la population ; on dit aussi *échantillon*.

Définition 5.2. *Individu* : C'est un élément de la population.

Définition 5.3. *Caractère* : C'est ce qu'on observe chez l'individu.

Définition 5.4. *Modalités* : Ce sont les différentes valeurs prises par le caractère.

Définition 5.5. *Quantitative* ou *qualitative* : La série statistique est dite *quantitative* quand les modalités sont des nombres et *qualitative* sinon.

Définition 5.6. *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, ...(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) 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.

On a aussi :

Définition 5.7. *Effectif d'une valeur* : C'est le nombre de fois que la valeur d'un caractère (la modalité) revient dans la série.

Définition 5.8. *Fréquence d'une valeur* : C'est l'effectif de la modalité divisé par l'effectif total; elle est comprise entre 0 et 1 et n'a pas d'unité.

Définition 5.9. *Classes de valeurs* : S'il y a trop de valeurs 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 Présentation d'une série statistique

Les données d'une série statistique quantitative sont présentées :

1^er^ cas : Par la liste de tous ses p éléments x_1, x_2, $\dots x_p$; on les notera parfois sous la forme d'un ensemble $S = \{x_1, x_2, \dots x_p\}$.

2^e^ cas : Par le tableau des effectifs n_1, n_2, ..., n_p ou des fréquences f_1, f_2, ..., f_p de chacune des p modalités x_1, x_2, $\dots x_p$.

Modalités	x_1	x_2	$\dots$	x_p	Total
Effectif	n_1	n_2	$\dots$	n_p	N

ou

Modalités	x_1	x_2	$\dots$	x_p	Total
Fréquences	f_1	f_2	$\dots$	f_p	1

3e cas : Par le tableau des effectifs n_1, n_2, ..., n_p ou des fréquences f_1, f_2, ..., f_p de chacune des p classes dans lesquelles sont rangées les individus de la série $[\alpha_1\,;\alpha_2[$, $[\alpha_2\,;\alpha_3[$, ..., $[\alpha_p\,;\alpha_{p+1}]$.

Classes	$[\alpha_1\,;\alpha_2[$	$[\alpha_2\,;\alpha_3[$	...	$[\alpha_p\,;\alpha_{p+1}]$	Total
Effectif	n_1	n_2	...	n_p	N

ou

Classes	$[\alpha_1\,;\alpha_2[$	$[\alpha_2\,;\alpha_3[$	...	$[\alpha_p\,;\alpha_{p+1}]$	Total
Effectif	f_1	f_2	...	f_p	1

Les centres des classes sont $c_1 = \frac{\alpha_1+\alpha_2}{2}$, $c_2 = \frac{\alpha_2+\alpha_3}{2}$, ..., $c_p = \frac{\alpha_p+\alpha_{p+1}}{2}$.

5.1.3 Paramètres statistiques

Pour faire parler ces (souvent longues) séries, il est nécessaire de les résumer : on produit alors *des* statistiques. Tout résumé met en évidence certaines caractéristiques de la série mais engendre une *perte d'information*, toutes les données n'étant plus accessibles.

Les résumés d'une série statistique quantitative sont, en général, de deux natures : les *mesures centrales* qui visent à remplacer toutes les valeurs de la série par une seule et les *mesures de dispersion* qui visent à renseigner comment les valeurs sont distribuées autour de la valeur centrale.

Certains résumés vont traditionnellement ensemble car la façon de les obtenir sont proches. On a ainsi d'une part le groupe « moyenne, variance, écart-type » et d'autre part le groupe « médiane, quartiles, valeurs extrêmes ».
Ces résumés sont l'objet des paragraphes suivants.

5.2 Moyenne, variance et écart-type

5.2.1 Une mesure centrale : la moyenne

La mesure centrale de ce groupe de paramètres est la moyenne arithmétique.

Définition 5.10 (Moyenne arithmétique)**.** La *moyenne arithmétique* d'une série statistique quantitative $S = \{x_1, x_2, \ldots, x_n\}$ est le nombre, noté $\overline{x}$ défini par :

1er cas : $\overline{x} = \frac{x_1+x_2+\ldots+x_p}{p}$

2e cas : $\overline{x} = \frac{n_1x_1+n_2x_2+\ldots+n_px_p}{N}$ ou $\overline{x} = f_1x_1 + f_2x_2 + \ldots + f_px_p$

3e cas : $\overline{x} = \frac{n_1c_1+n_2c_2+\ldots+n_pc_p}{N}$ ou $\overline{x} = f_1c_1 + f_2c_2 + \ldots + f_pc_p$ ($\overline{x}$ est alors une estimation de la moyenne).

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 l'agrégation de ces deux séries. Elle a le défaut d'être sensible aux valeurs extrêmes.

5.2.2 Des mesures de dispersion : la variance et l'écart-type

Les mesures de dispersion de ce groupe de paramètres sont la variance et l'écart-type.

Définition 5.11 (Variance et écart-type)**.** La *variance* V est le réel positif défini par :

1er cas : $V = \frac{x_1^2+x_2^2+...+x_p^2}{p} - \overline{x}^2$

2e cas : $V = \frac{n_1x_1^2+n_2x_2^2+...+n_px_p^2}{N} - \overline{x}^2$ ou $\overline{x} = f_1x_1^2 + f_2x_2^2 + \ldots + f_px_p^2 - \overline{x}^2$

3e cas : $V = \frac{n_1c_1^2+n_2c_2^2+...+n_pc_p^2}{N} - \overline{x}^2$ ou $\overline{x} = f_1c_1^2 + f_2c_2^2 + \ldots + f_pc_p^2 - \overline{x}^2$ (V est alors une estimation de la variance).

L'écart-type σ est le réel positif défini par : $\sigma = \sqrt{V}$.

Plus la variance (ou l'écart-type) est éloignée de 0 et plus les données sont dispersées autour de la moyenne.
Comme la moyenne, la variance et l'écart-type sont sensibles aux valeurs extrêmes.

5.3 Médiane, quartiles, valeurs extrêmes

5.3.1 Une mesure centrale : la médiane

La mesure centrale de ce groupe de paramètres est la 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 :

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

Définition 5.12 (Médiane dans le cadre scolaire)**.** Soit une série statistique quantitative comportant n données : $S = \{x_1, x_2, \ldots, x_i, \ldots, x_n\}$ telles que $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$.

- Si n est impair, la $\left(\frac{n+1}{2}\right)^{\text{e}}$ donnée de la série est la médiane.
- Si n est pair, tout nombre compris entre le $\frac{n}{2}^{\text{e}}$ 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 :
$$m = \frac{\left(\frac{n}{2}\right)^{\text{e}} + \left(\frac{n}{2}+1\right)^{\text{e}}}{2}$$

Enfin la médiane d'une série continue ou regroupée en classes est la valeur qui correspond à une fréquence cumulée de 50 %.

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.

5.3.2 Des mesures de dispersion : les valeurs extrêmes et les quartiles

Les mesures de dispersion de ce groupe de paramètres sont les valeurs extrêmes

Valeurs extrêmes

Définition 5.13. Les *valeurs extrêmes* d'une série quantitative sont ses valeurs *minimale*, x_{min}, et *maximale*, x_{max}.
L'*étendue e* est la différence entre les valeurs extrêmes de la série : $e = x_{\text{max}} - x_{\text{min}}$.

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 5.14 (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 : $S = \{x_1, x_2, \ldots, x_i, \ldots, x_n\}$ telles que $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$. Alors :*

- *La donnée de rang $\frac{1}{4}n$ (ou sa valeur approchée par excès à l'entier supérieur si $\frac{1}{4}n$ n'est pas un entier) convient toujours comme premier quartile.*
- *La donnée de rang $\frac{3}{4}n$ (ou sa valeur approchée par excès à l'entier supérieur si $\frac{3}{4}n$ n'est pas un entier) convient toujours comme troisième quartile.*

On l'admettra.
Comme la médiane, les quartiles ne sont pas sensibles aux valeurs extrêmes.

Exemples.

- S'il y a $n = 29$ données dans la série, rangées dans l'ordre croissant :
  - $\frac{1}{4} \times 29 = 7{,}25$ donc la huitième (valeur approchée par excès de 7,25) donnée de la série convient comme premier quartile ;
  - $\frac{3}{4} \times 29 = 21{,}75$ donc la vingt-deuxième (valeur approchée par excès de 21,75) donnée de la série convient comme troisième quartile.

2. Ce sont aussi ces quartiles que prennent les statisticiens

- 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 interquartiles de la manière suivante :

Définition 5.15. 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 $[Q_1 ; Q_3]$.

5.4 Représentations graphiques

Si les mesures centrales et les mesures de dispersion ont pour but de résumer une série statistique en quelques nombres, les représentations graphiques, elles, visent à la visualiser.

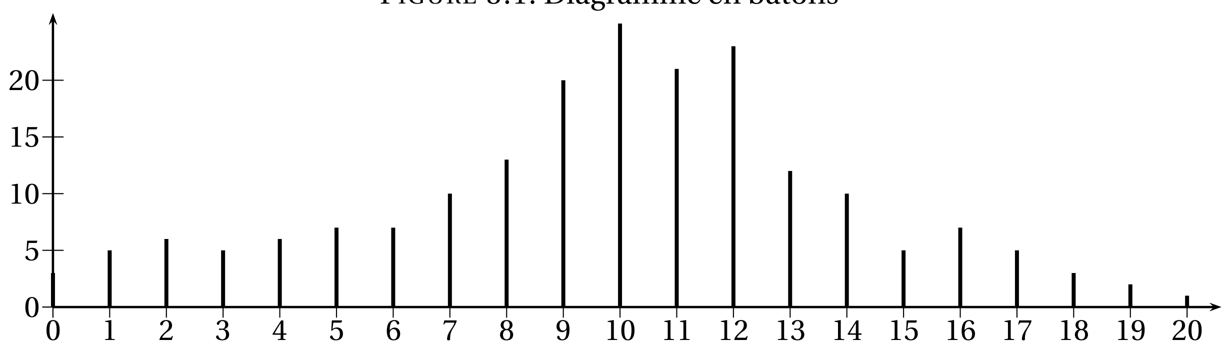
5.4.1 Diagramme à bâtons

On considère la série :

Valeurs x_i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Effectifs 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 à bâtons de la figure 5.1, de la présente page.

FIGURE 5.1: Diagramme en bâtons



5.4.2 Diagrammes basés sur la fréquence

Les séries statistiques peuvent aussi être représentées en diagrammes circulaires, semi-circulaires, rectangulaires, etc. L'aire de chaque modalité devra être proportionnelle à l'effectif de cette modalité. Les fréquences permettent d'obtenir assez facilement la part du diagramme qui devra être consacrée à chaque modalité.
Ainsi si on considère la série suivante :

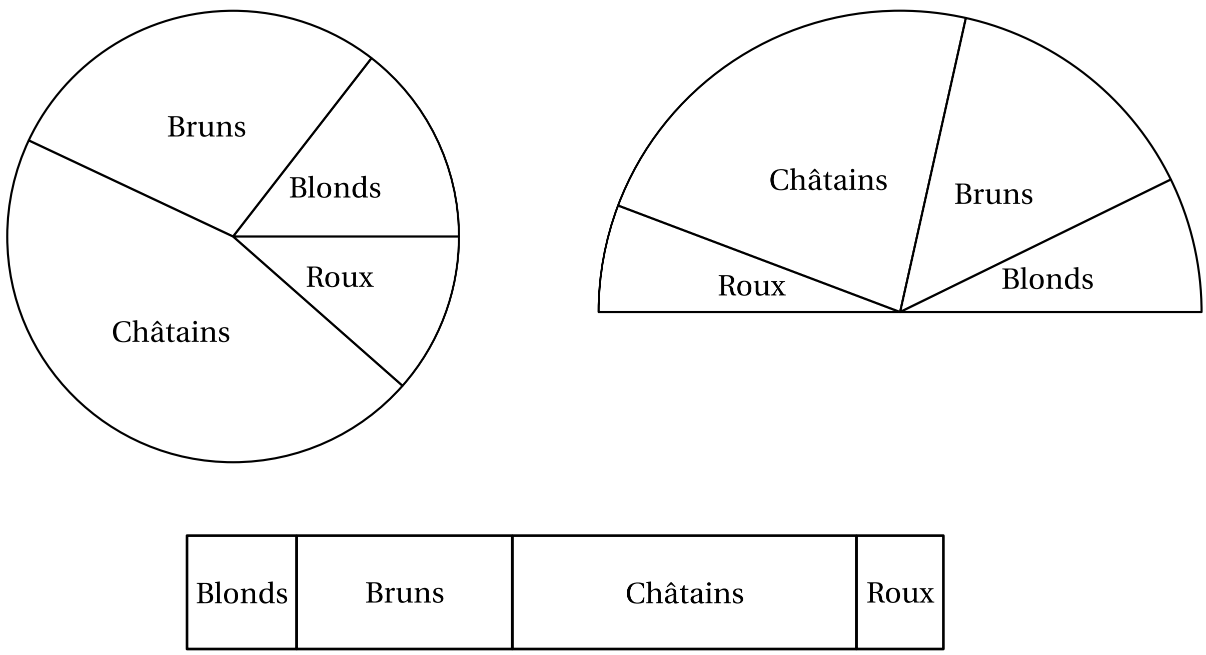
Données	Blonds	Bruns	Châtains	Roux
Effectifs n_i	25	57	91	23

On a alors :

x_i	Blonds	Bruns	Châtains	Roux	Total
n_i	29	57	91	23	200
Fréquence f_i	$\frac{29}{200} = 0,145$	0,285	0,455	0,115	1
Part d'un diagramme circulaire	$0,145 \times 360 = 52,2\,^{\circ}$	102,6 °	163,8 °	41,4 °	360 °
Part d'un diagramme semi-circulaire	$0,145 \times 180 = 26,1\,^{\circ}$	51,3 °	81,9 °	20,7 °	180 °
Part d'un rectangle de 10 cm	$0,145 \times 10 = 1,45\,\text{cm}$	2,85 cm	4,55 cm	1,15 cm	10 cm
Fréquence en pourcentage	$0,145 = \frac{14,5}{100} = 14,5\,\%$	28,5 %	45,5 %	11,5 %	100 %

On obtient les diagrammes de la figure 5.2, de la présente page.

FIGURE 5.2: Diagrammes circulaire, semi-circulaire et rectangulaire



5.4.3 Diagramme en boîte

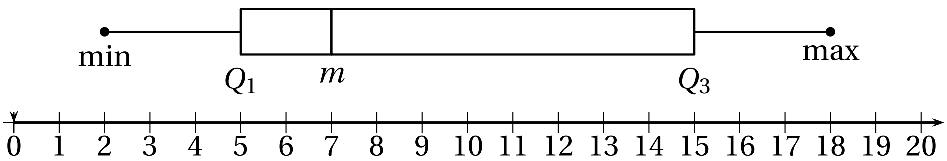
On peut représenter graphiquement les valeurs extrêmes, les quartiles et la médiane par un *diagramme en boîte*, appelé aussi *boîte à moustaches*, conçu de la manière suivante :

- au centre une boîte 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.

La figure 5.3, page suivante en est un exemple.

Ce type de diagramme permet une interprétation visuelle et rapide de la dispersion des séries statistiques. Il permet également d'apprécier des différences entre des séries (lorsqu'elles ont des ordres de grandeurs comparables).

FIGURE 5.3: Exemple de diagramme en boîte



Remarques.

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

5.5 Utilisation de la calculatrice

On peut retrouver tous ces paramètres statistiques en utilisant les listes d'une calculatrice. On trouvera dans la figure 5.4, de la présente page la façon d'utilisation la calculatrice selon le modèle utilisé, pour obtenir ces paramètres.

FIGURE 5.4: Paramètres stastistiques à la calculatrice

1^er^ cas

Calculatrice Casio

- Dans le menu **STAT**, entrer les valeurs de la variable dans la 1ère colonne (**LIST1**).
- Pour définir la liste des valeurs, utiliser la commande **CALC** (F2), puis choisir **SET** pour définir les listes.

Choisir **LIST1** pour **1VAR XLIST** et **1** pour **1VAR FREQ**. Valider les réglages (EXE).

- Pour obtenir les paramètres, choisir **1VAR** (F1) (la touche ▸ permet d'afficher les autres paramètres).
- Pour effacer une liste, placer le curseur sur une valeur de la liste à effacer, puis choisir **DEL-A** (F4) et valider (F1).

Calculatrice TI

- Dans le menu **STATS**, choisir **1:EDIT** (1 ou entrer).
- Saisir les valeurs de la variable dans la 1ère colonne **L1**, puis mettre en mémoire (entrer).
- Pour définir la liste des valeurs, utiliser la commande **CALC** (▸), choisir **1:STATS 1-Var**, valider (entrer).

Taper **L1** (2nde 1 , 2nde 2).

- Pour afficher les paramètres, valider (entrer). (▸ permet d'afficher les autres paramètres).
- Pour effacer une liste, placer le curseur sur le nom de la liste à effacer, puis choisir **annul** et valider (entrer).

2e et 3e cas

Calculatrice Casio

- Dans le menu **STAT**, entrer les valeurs de la variable dans la 1ère colonne (**LIST1**) et les effectifs dans la 2ème colonne (**LIST2**).
- Pour définir les listes, utiliser la commande **CALC** (F2), puis choisir **SET** pour définir les listes.

Choisir **LIST1** pour **1VAR XLIST** et **LIST2** pour **1VAR FREQ**. Valider les réglages (EXE).

- Pour obtenir les paramètres, choisir **1VAR** (F1) (▸ permet d'afficher les autres paramètres).
- Pour effacer une liste, placer le curseur sur une valeur de la liste à effacer, puis choisir **DEL-A** (F4) et valider (F1).

Calculatrice TI

- Dans le menu **STATS**, choisir **1:EDIT** (1 ou entrer).
- Saisir les valeurs de la variable dans la 1ère colonne **L1**, les effectifs dans la 2ème colonne **L2**, et mettre en mémoire (entrer).
- Pour définir les listes, utiliser la commande **CALC** (▸), choisir **1:STATS 1-Var**, valider (entrer).

Taper **L1,L2** (2nde 1 , 2nde 2).

- Pour afficher les paramètres, valider (entrer). (▸ permet d'afficher les autres paramètres).
- Pour effacer une liste, placer le curseur sur le nom de la liste à effacer, puis choisir **annul** et valider (entrer).

5.6 Exercices

EXERCICE 5.1.
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 ?

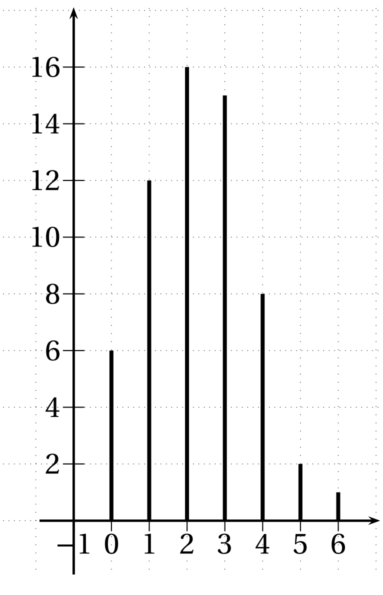
On gardera en tête l'allure de cette série « canonique » où les données sont parfaitement réparties qui pourra servir de référence pour décrire les autres séries : plus on s'éloigne de ce cas particulier, plus on pourra parler « d'irrégularité » de dispersion.

EXERCICE 5.2.
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.
Une société de transport possédant 50 autocars a fait un relevé statistique sur 60 jours du nombre de cars en panne. Les résultats sont donnés dans le graphique ci-dessous.



1. À l'aide du graphique, compléter les effectifs du tableau :

Nombre de cars en panne	0	1	2	3	4	5	6
Nombre de jours (effectifs)							
Effectif cumulé croissant							

2. (a) Calculer les effectifs cumulés croissants du tableau.
   (b) Déterminer alors :
   - la médiane m et interpréter le résultat ;
   - les premier et troisième quartile Q_1 et Q_3.

   (c) Retrouver ces résultats en utilisant la calculatrice.

EXERCICE 5.4.
On a réalisé une enquête sur le temps en secondes que doit attendre un abonné qui contacte, par téléphone, son fournisseur d'accès Internet.
Cette enquête a concerné 200 abonnés et a donné les résultats suivants :

Temps	2	5	10	15	20	25	30	35	45	50	60
Abonnés	5	6	8	20	60	41	21	15	10	9	5

1. Calculer la moyenne $\overline{x}$ et l'écart-type σ (arrondi au dixième) de cette série en utilisant la calculatrice.
2. Quel est le pourcentage des temps qui se trouvent dans l'intervalle $[\overline{x}-2\sigma\,;\,\overline{x}+2\sigma]$?

EXERCICE 5.5.
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.

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.

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

EXERCICE 5.6.
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 5.7.
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 5.8.
Dans la chaîne de production d'une entreprise montant un certain modèle de matériel pour la téléphonie mobile, 40 opérateurs réalisent le même assemblage.
Une étude statistique sur le temps mis pour effectuer cet assemblage par chacun des opérateurs a permis d'obtenir le tableau suivant :

Temps d'assemblage (en secondes)	[105 ; 110[	[110 ; 115[	[115 ; 120[	[120 ; 125[	[125 ; 130]
Nombre d'opérateurs	2	6	12	14	6

1. (a) Déterminer la proportion d'opérateurs (en %) qui mettent entre 110 et 120 secondes pour réaliser cet assemblage.

(b) Déterminer la proportion des opérateurs (en %) qui mettent au moins 120 secondes pour réaliser cet assemblage.

2. Dans cette question, on fait l'approximation suivante : toutes les valeurs d'une même classesont égales au centre de la classe. On note la moyenne $\overline{x}$ de cette série statistique et σ son écart-type.
Vérifier, avec votre calculatrice, qu'en arrondissant à 10^{-1}, on a $\overline{x} \simeq 119,5$ et $\sigma \simeq 5,3$.

EXERCICE 5.9.
Le service contrôle qualité d'une entreprise teste 1 000 pièces en métal pour vérifier si leur diamètre est proche de 50 mm. Les résultats sont donnés ci-dessous.

Diamètre	[49,4 ; 49,6[	[49,6 ; 49,8[	[49,8 ; 50[	[50 ; 50,2[	[50,2 ; 50,4[	[50,4 ; 50,6[	[50,6 ; 50,8]
Effectif	10	110	200	320	250	80	30

1. Calculer le diamètre moyen $\overline{x}$ des pièces testées, puis l'écart-type σ de la série. Arrondir au dixième.

2. La machine est considérée comme bien réglée lorsqu'au moins 95 % des pièces testées ont un diamètre dans l'intervalle $[\overline{x} - 2\sigma ; \overline{x} + 2\sigma]$.

   (a) Déterminer cet intervalle. On arrondira les bornes au dixième.

   (b) La machine est-elle bien réglée ?

EXERCICE 5.10.
On a relevé dans un supermarché les prix des pains vendus au poids.

Pains courants (en €) :

2	2	2,1	2,18	2,24	2,44	2,5	2,75	2,75	2,96	2,98
3,1	3,22	3,34	3,34	3,34	3,4	3,68	3,7	3,75	3,78	
3,78	3,78	3,81	3,81	3,81	3,96	4,2	5,2			

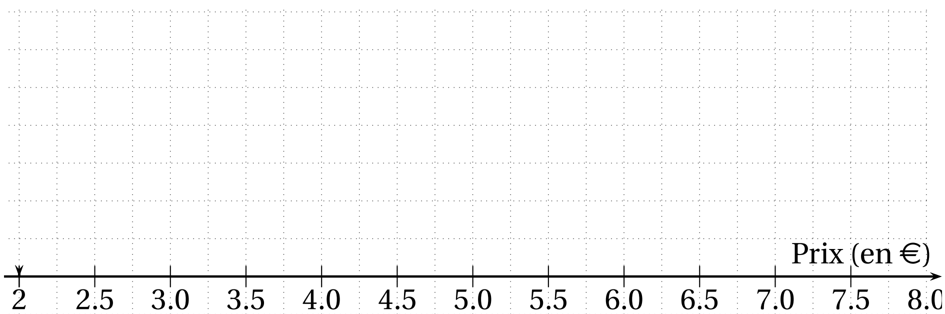
Pains spéciaux (en €) :

3,78	3,78	3,8	3,98	4,07	4,07	4,07	4,07	4,1	4,31	4,35
4,5	4,6	4,6	4,6	5,48	5,48	5,82	5,85	4,35	5,9	

1. (a) Pour chacune des deux séries statistiques, déterminer, sans calculatrice, l'étendue, la médiane, les premier et troisième quartiles et l'écart interquartile.

   (b) Comparer la dispersion des valeurs de ces deux séries par rapport à leur médiane.

2. (a) Construire sur le graphique ci-dessous les diagrammes en boîtes des deux séries.



   (b) Les phrases suivantes sont-elles vraies ou fausses ?

- Les trois quarts des prix des pains courants sont inférieurs aux prix des pains spéciaux.
- La moitié des prix des pains spéciaux sont inférieurs au quart des prix des pains courants.
- Plus du quart des prix des pains spéciaux ont supérieurs à tous les prix des pains courants.

EXERCICE 5.11.
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

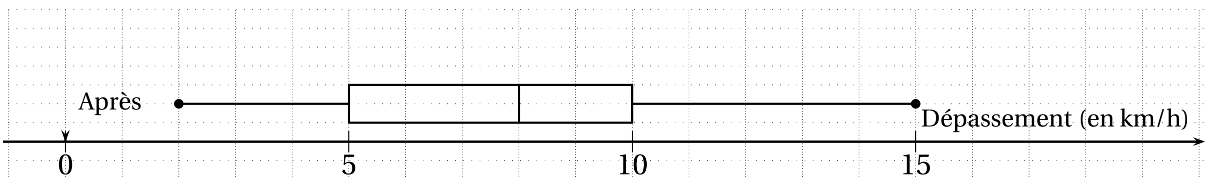
Le but de l'exercice est de comparer la dispersion des notes en Maths et en Histoire-Géographie.

1. Calculer $\overline{x}$ et $\overline{x'}$ les moyennes respectives de Maths et d'Histoire-Géographie.
2. Calculer médiane m_e et quartiles Q_1 et Q_3 en Maths.
3. Calculer médiane m'_e et quartiles Q'_1 et Q'_3 en Histoire-Géographie.
4. Représenter les diagrammes en boîte des notes en Maths et en Histoire-Géographie.
5. Interpréter les résultats obtenus.

EXERCICE 5.12.
En ville la vitesse est limitée à 50 km/h.
Les autorités effectuent une enquête dans une zone où il semble y avoir des excès de vitesse.
Un dispositif est mis en place pour mesurer les dépassements de vitesse autorisée.
On relève un échantillon de 125 mesures effectuées sur des véhicules en excès de vitesse.
Le tableau indique les résultats de cette enquête.

Dépassement (en km/h)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Nombre de véhicules	5	6	5	5	12	7	6	7	6	4	5	6	5	5	11	8	6	7	6	3

1. (a) Quel est le pourcentage de véhicules pour lesquels le dépassement de la vitesse autorisée est supérieur (ou égal) à 10 km/h ?
   (b) Déterminer la médiane et les premier et troisième quartiles de cette série.
2. Des aménagements urbains destinés à ralentir les véhicules et à prévenir les conducteurs sont mis en place. Quelques temps après, on effectue alors une nouvelle étude sur 125 véhicules en excès de vitesse. On donne le diagramme en boîte de cette nouvelle série (les extrémités correspondent au minimum et au maximum).



   (a) Déterminer Q_1, m et Q_3.
   (b) Après les aménagements, chacune des affirmations suivantes est-elle vraie ? Justifier.

- 75 % des automobilistes en excès de vitesse dépassent la vitesse autorisée d'au plus 10 km/h.
- 50 % des automobilistes en excès de vitesse dépassent la vitesse autorisée d'au moins 8 km/h.
- La médiane a diminué de 2 %.

3. (a) Construire le diagramme en boîte correspondant à la série avant les aménagements au-dessus de celui correspondant à « après ».
   (b) Quelles comparaisons peut-on faire entre les séries « avant » et « après » les aménagements ?
   (c) Les aménagements ont-ils été efficaces ?

5.7 Travaux dirigés

EXERCICE 5.13.
Une entreprise d'un pays de l'Union Européenne emploie 400 tehcniciens et 500 cadres en 2014. Le salaire mensuel net moyen d'un technicien est 1 000 € et celui d'un cadre est de 1 400 €.

	A	B	C	D	E	F
1		Techniciens		Cadres		Entreprise
2		Effectif	Salaire moyen	Effectif	Salaire moyen	Salaire moyen
3	2014	400	1 000,00 €	500	1 400,00 €	1 222,22 €
4	2015		1 100,00 €		1 500,00 €	

1. Quelle formule peut-on saisir dans la cellule **F3** pour calculer le salaire moyen dans l'entreprise ?
2. À la suite d'une restructuration, les effectifs sont modifiés. Les salaires mensuels moyens augmentent tous, passant à 1 100 € pour les techniciens et à 1 500 € pour les cadres. Néanmoins, le salaire moyen dans l'entreprise diminue. À l'aide du tableur, rechercher un exemple illustrant ce paradoxe.
3. Comment l'expliqueriez-vous ?

EXERCICE 5.14.
Le gérant d'une station-service relève les prix du diesel dans les 50 stations-services de sa région pour lui permettre de fixer le prix qu'il va afficher.

1. (a) Récopier sur tableur le tableau ci-contre.
   (b) Indiquer la formule à saisir en **C2** pour calculer la fréquence de chaque prix.
   (c) Indiquer la formule à saisir en **D2**, puis en **D3** pour calculer les fréquences cumulées croissantes.
   (d) Compléter le tableau en calculant toutes les fréquences et toutes les fréquences cumulées croissantes.
   (e) Construire la courbe des fréquences cumulées croissantes.

2. (a) Les fonctions :
   - $\boxed{\textbf{=MEDIANE(plage)}}$,
   - $\boxed{\textbf{=QUARTILE(plage ; 1)}}$
   - $\boxed{\textbf{=QUARTILE(plage ; 3)}}$

   calculent, respectivement, la médiane, le premier et le troisième quartile, mais seulement lorsque les données sont présentées comme dans le 1er cas [3].
   Ici nous sommes dans le 2e cas aussi, déterminer ces paramètres statistiques « à la main ».

   (b) Préciser ce que calcule la formule saisie en **F9** : $\boxed{\textbf{=SOMMEPROD(A2 :A10 ; C2 :C10)}}$.

   (c) Comparer la médiane et la moyenne.

	A	B	C	D	E	F
1	Prix	Effectif	Fréquence	F.C.C.		
2	1,22	2				Q1
3	1,23	5				
4	1,26	6				Médiane
5	1,27	8				
6	1,31	10				Q3
7	1,35	7				
8	1,38	6				Moyenne
9	1,41	4				
10	1,45	2				

EXERCICE 5.15 (Uniformisation par péréquation)**.**
Les notes sur 100 accordées par deux jurys différents lors d'un concours sont présentées dans le tableau suivant.

Note	4	5	8	9	10	11	12	13	15	16	18
Effectif (jury A)	0	1	2	5	11	12	19	23	15	8	4
Effectif (jury B)	1	0	3	11	28	24	14	10	5	3	1

1. Calculer, avec la calculatrice, la moyenne $\overline{x}_\text{A}$ et la variance V_A des notes accordées par le jury A.
   Calculer de même la moyenne $\overline{x}_\text{B}$ et la variance V_B des notes accordées par le jury B.
   Arrondir ses résultats à 10^{-1} près.

2. On désire harmoniser les notations et, pour ce faire, on décide de modifier les notes du jury B de façon à obtenir une moyenne et une variance identiques à celles du jury A.
   Pour cela, afin d'obtenir les notes finales y'_i de la série statistique du jury B, on fait subir aux notes initiales y_i, un changement de variable affine sous la forme $y'_i = ay_i + b$.
   En notant $\overline{x'}_\text{B}$ la moyenne et V'_B la variance des notes finales accodées par le jury B, on admet alors que :
   $$\overline{x'}_\text{B} = a\overline{x}_\text{B} + b \qquad \text{et} \qquad V'_\text{B} = a^2 V_\text{B}$$
   Déterminer les réels positifs a et b de façon à ce que $\overline{x'}_\text{B} = \overline{x}_\text{A}$ et $V'_\text{B} = V_\text{A}$.
   Arrondir a et b à 10^{-1} près.
   L'opération effectuée ainsi s'appelle une péréquation.

3. On décide d'admettre tous les candidats dont la note est supérieure à 10.
   Un candidat a obtenu 9 avec le jury B. Sera-t-il admis après péréquation ?

3. On peut y arriver avec le tableur dans les autres cas mais ce n'est pas simple.