1 Conventions

Dans toute la suite, on étudiera une population, notée E, et une variable statistique quantitative X définie sur E.

Exemple 1. Si on étudie par exemple le nombre de poupées Barbue que possèdent les élèves de la classe de 1$^{\text{ère}}$ ES3
▷ E est l'ensemble des élèves de la classe
▷ X est la fonction qui, à un élément de E, associe le nombre de poupées Barbue qu'il ou elle possède.

Si E possède n éléments, on notera $\mathrm{V} = \{x_1, x_2, \cdots, x_{n-1}, x_n\}$ l'ensemble **ordonné par valeurs croissantes** des valeurs prises par X.
Notez bien que certains éléments de V peuvent être égaux : en effet, deux élèves différents peuvent avoir le même nombre de poupées Barbue.

2 Médiane

2.1 Définition

Définition 2. La médiane M_e est un nombre tel que :
▷ **au moins** 50% des éléments de V sont inférieurs à M_e,
▷ **au moins** 50% des éléments de V sont supérieurs à M_e.
On distingue deux cas selon la parité de n :
▷ **Si n est impair** : on sépare V en deux groupes distincts de même effectif. M_e est la valeur centrale, à la fois maximum de la *partie basse* et minimum de la *partie haute* ;
▷ **Si n est pair** : on sépare V en deux groupes distincts de même effectif. M_e est la valeur centrale, demi-somme du maximum de la *partie basse* et du minimum de la *partie haute* ;

Exemples 3. – $\mathrm{V} = \{1, 2, 2, 5, 5, 8, 8, 9, 37\}$
▷ $\mathrm{M}_e = 5$
– $\mathrm{V} = \{1, 2, 2, 5, 8, 8, 9, 37\}$
▷ $\mathrm{M}_e = \dfrac{5+8}{2} = 6,5$

3 Quartiles

3.1 L'idée

Pour avoir une idée un peu plus précise de la série statistique étudiée, on voudrait séparer notre population en 4 groupes au lieu de 2.

On a donc envie de calculer les médianes des parties basses et hautes.

On a également comme cahier des charges d'avoir au moins 25% des valeurs prises par x inférieures au premier quartile Q_1 et au moins 75% des valeurs prises par x inférieures au troisième quartile Q_3.

3.2 Expérimentons

Exemple 4. $\mathrm{V} = \{1, 2, 3, 4, 5, 6, 7, 8\}$
On peut séparer l'effectif en quatre groupes de même effectif égal à 25% de l'effectif total donc ça « colle »

$$\mathrm{V} = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

On peut prendre
– $\mathrm{Q}_1 = \dfrac{2+3}{2} = 2,5$ et 25% des effectifs ont une valeur inférieure à Q_1
– $\mathrm{M}_e = \dfrac{4+5}{2} = 4,5$ et 50% des effectifs ont une valeur inférieure à M_e
– $\mathrm{Q}_3 = \dfrac{6+7}{2} = 6,5$ et 75% des effectifs ont une valeur inférieure à Q_3
et Q_1 et Q_3 sont bien les médianes respectives des parties basses et hautes.

Exemple 5. $\mathrm{V} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$
On peut séparer l'effectif en parties hautes et basses

$$\mathrm{V} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

On peut prendre

– $Q_1 = 3$ et $3/_{10} \simeq 33\% (\geqslant 25\%)$ des effectifs ont une valeur inférieure à Q_1
– $M_e = \dfrac{5+6}{2} = 5,5$ et 50% des effectifs ont une valeur inférieure à M_e
– $Q_3 = 8$ et $8/_{10} = 80\% (\geqslant 75\%)$ des effectifs ont une valeur inférieure à Q_3
et Q_1 et Q_3 sont bien les médianes respectives des parties basses et hautes.

Exemple 6. $V = \{1,2,3,4,5,6,7,8,9,10,11\}$

On peut séparer l'effectif en parties haute et basse

$$V = \{1,2,3,4,5,6,7,8,9,10,11\}$$

On peut prendre
– $Q_1 = 3$ et $3/_{11} \simeq 27\% (\geqslant 25\%)$ des effectifs ont une valeur inférieure à Q_1
– $M_e = \dfrac{4+5}{2} = 4,5$ et 50% des effectifs ont une valeur inférieure à M_e
– $Q_3 = 9$ et $9/_{11} = 82\% (\geqslant 75\%)$ des effectifs ont une valeur inférieure à Q_3
et Q_1 et Q_3 sont bien les médianes respectives des parties basses et hautes.
Un cas plus problématique a été gardé pour la fin...

Exemple 7. $V = \{1,2,3,4,5,6,7,8,9\}$

On peut séparer l'effectif en parties haute et basse avec la médiane au milieu :

$$V = \{1,2,3,4,5,6,7,8,9\}$$

et prendre $Q_1 = \dfrac{2+3}{2} = 2,5$, mais...
... seulement $2/_9 \simeq 22\%$ des effectifs ont une valeur inférieure à Q_1, ce qui est contraire à notre cahier des charges.
Dans ce cas particulier, on va inclure la médiane dans les parties basses et hautes pour éviter cet écueil
Cela donne
– Partie Basse $= \{1,2,3,4,5\}$ donc $Q_1 = 3$ et $3/_9 \simeq 33\% \geqslant 25\%$ des effectifs ont une valeur inférieure à Q_1, ce qui convient.
– Partie Haute $= \{5,6,7,8,9\}$ donc $Q_3 = 7$ et $7/_9 \simeq 78\% \geqslant 75\%$ des effectifs ont une valeur inférieure à Q_1, ce qui convient.
Il est aisé de constater que ces observations vont se répéter par cycle de longueur 4.

C'est quand V a un nombre d'éléments égal à un multiple de 4 plus 1 que nous devons être prudents.

D'ailleurs, toutes les machines à calculer ne s'accordent pas sur le calcul des quartiles.

La méthode que je vous propose est la plus cohérente et sera en accord avec la détermination graphique des quartiles que nous verrons bientôt.

Elle permet également d'avoir une définition rigoureuse comme nous allons le voir.

3.3 Définissons

Définition 8. **Le premier quartile** est obtenu en prenant la médiane de la sous-série contenant les observations dont le rang est strictement inférieur à celui de la médiane (*la partie basse*) pour autant qu'au moins 25% des observations soient inférieures ou égales à cette valeur.
Sinon, il faut inclure la médiane dans la partie basse.

Définition 9. **Le troisième quartile** est obtenu en prenant la médiane de la sous-série contenant les observations dont le rang est strictement supérieur à celui de la médiane (*la partie haute*) pour autant qu'au moins 75% des observations soient inférieures ou égales à cette valeur.
Sinon, il faut inclure la médiane dans la partie haute.

4 Fonction de répartition

4.1 Définition

Nous reprenons les notations des sections précédentes.

Définition 10. Nous noterons F la fonction définie, pour tout réel x, par la **proportion** des valeurs observées qui sont **inférieures ou égales à x**

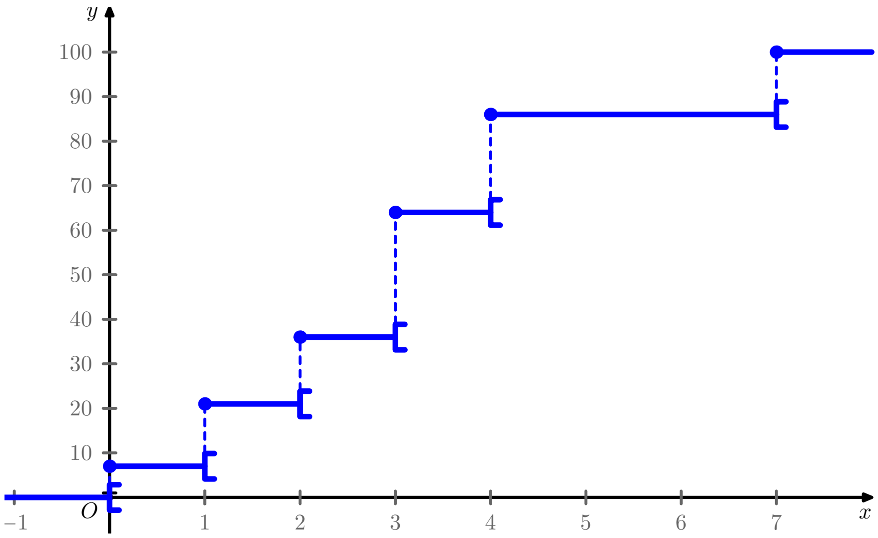
4.2 Un exemple

Exemple 11. Considérons la série ordonnée :

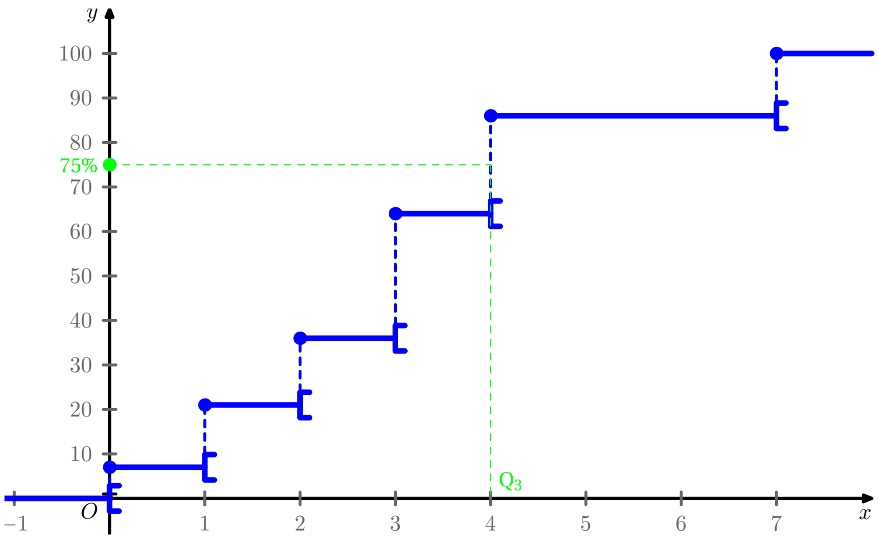
$$V = \{0,1,1,2,2,3,3,3,3,4,4,4,7,7\}$$

- Si $x \in]-\infty; 0[$, alors $F(x) = 0$
- Si $x \in [0; 1[$, alors $F(x) = 1/_{14} \simeq 0,07$
- Si $x \in [1; 2[$, alors $F(x) = 3/_{14} \simeq 0,21$
- Si $x \in [2; 3[$, alors $F(x) = 5/_{14} \simeq 0,36$
- Si $x \in [3; 4[$, alors $F(x) = 9/_{14} \simeq 0,64$
- Si $x \in [4; 7[$, alors $F(x) = 12/_{14} \simeq 0,86$
- Si $x \in [7; +\infty[$, alors $F(x) = 1$

4.3 représentation graphique



La courbe obtenue va nous permettre de **lire** les valeurs des différents quartiles



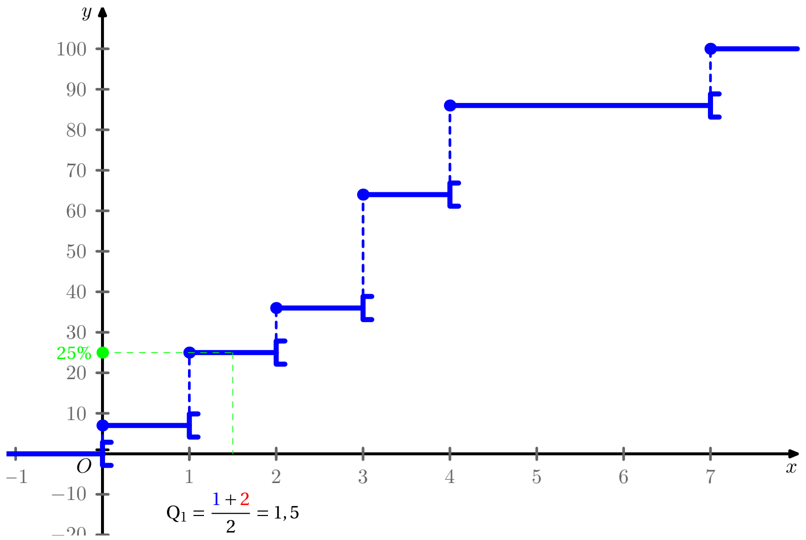
On vérifie en utilisant la définition en termes de médianes

$$V = \{0,1,1,2,2,3,3,3,3,4,4,4,7,7\}$$

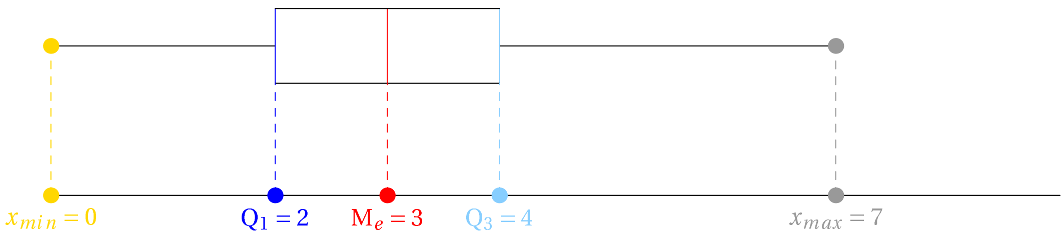
On a bien
- $Q_1 = 2$
- $M_e = \dfrac{3+3}{2} = 3$
- $Q_3 = 4$

4.4 Cas pathologique

Si une des droites horizontales rencontrent « l'escalier »selon un segment horizontal, on s'adapte :



4.5 Boîte à moustaches



5 Écart-type

5.1 Introduction

Un autre moyen de caractériser une série statistique est de mesurer « l'éloignement »de l'ensemble des valeurs x_i par rapport à la moyenne $\overline{x}$: comment faire ?

Exemple 12. Reprenons la série précédente :

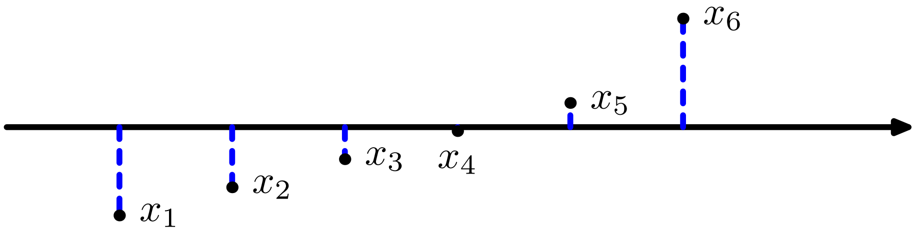
Valeurs x_i	0	1	2	3	4	7
Effectifs n_i	1	2	2	4	3	2

Calculez la moyenne $\overline{x}$ de cette série.

Complétez le tableau suivant proposant trois façons de « mesurer »pour chaque valeur « l'éloignement »par rapport à $\overline{x}$.

$x_i - \overline{x}$						
$\lvert x_i - \overline{x}\rvert$						
$(x_i - \overline{x})^2$						

Calculez dans chacun des trois cas l'éloignement moyen.



$\overline{x} \simeq 3,14$

$x_i - \overline{x}$	$-3,14$	$-2,14$	$-1,14$	$-0,14$	$0,86$	$3,86$
$\lvert x_i - \overline{x}\rvert$	3,14	2,14	1,14	0,14	0,86	3,86
$(x_i - \overline{x})^2$	9,86	4,58	1,30	0,02	0,74	14,90

$$\begin{aligned} E_1 &= \frac{(-3,14\times 1)+(-2,14\times 2)+(-1,14\times 2)+(-0,14\times 4)+(0,86\times 3)+(3,86\times 2)}{12} \\ &= 0 \end{aligned}$$

$$\begin{aligned} E_2 &= \frac{\lvert -3,14\rvert\times 1+\lvert -2,14\rvert\times 2+\lvert -1,14\rvert\times 2+\lvert -0,14\rvert\times 4+\lvert 0,86\rvert\times 3+\lvert 3,86\rvert\times 2}{12} \\ &\simeq 3,74 \end{aligned}$$

$$\begin{aligned} E_3 &= \frac{(-3,14)^2\times 1+(-2,14)^2\times 2+(-1,14)^2\times 2+(-0,14)^2\times 4+(0,86)^2\times 3+(3,86)^2\times 2}{12} \\ &\simeq 16,58 \end{aligned}$$

Sur une calculatrice, ces opérations peuvent être menées rapidement.

Sur TI

1

- $\boxed{\text{STAT}}\boxed{1}$
- rentrez les x_i en [L1]
- rentrez les n_i en [L2]
- pour la moyenne : $\boxed{\text{STAT}}$[CALC]1 Var [L1]$\boxed{,}$[L2]$\boxed{\text{ENTER}}$et on lit $\overline{x}$
- On retourne sur $\boxed{\text{STAT}}\boxed{1}$et on se place sur [L3]
- On tape [L1]$\boxed{-}\boxed{3}\boxed{,}\boxed{1}\boxed{4}$
- pour l'écart moyen : $\boxed{\text{STAT}}$[CALC]1 Var [L3]$\boxed{,}$[L2]$\boxed{\text{ENTER}}$et on lit $\overline{x}$

5.2 Quelle mesure choisir ?

Écart algébrique moyen

Définition 13. On appelle **écart algébrique moyen** le nombre :

$$l_m = \frac{1}{N}\sum_{i=1}^{p} n_i(x_i - \overline{x}).$$

Ce nombre est toujours nul et ne permet pas de distinguer deux séries.

Démonstration.
$$\begin{aligned} l_m &= \frac{1}{N}\sum_{i=1}^{p} n_i(x_i-\overline{x}) \\ &= \frac{1}{N}\sum_{i=1}^{p} n_i x_i - \frac{1}{N}\sum_{i=1}^{p} n_i\overline{x} \quad \text{(disributivité)} \\ &= \overline{x} - \overline{x}\frac{1}{N}\sum_{i=1}^{p} n_i \quad \text{(on factorise par } \overline{x}\text{)} \\ &= \overline{x} - \overline{x}\frac{1}{N}N \quad \text{(par définition de N)} \\ &= \overline{x} - \overline{x} \\ &= 0 \end{aligned}$$
□

Écart absolu moyen

Définition 14. On appelle **écart absolu moyen** le nombre :

$$e_m = \frac{1}{N}\sum_{i=1}^{p} n_i\lvert x_i - \overline{x}\rvert.$$

Ce nombre fournit un très bon paramètre de dispersion mais il n'a pas d'application en statistique mathématique entre autres raisons parce que la valeur absolue se prête peu aux calculs.

5.3 Variance

On s'intéresse alors à la moyenne pondérée des nombres $(x_i - \overline{x})^2$ qui a permis de formuler de nombreuses propriétés en statistique et en probabilité, vous le verrez au fur et à mesure de vos études.

Définition 15. On appelle **variance** d'une série quelconque à caractère quantitatif discret le nombre :

$$V = \frac{1}{N} \sum_{i=1}^{p} n_i (x_i - \overline{x})^2 = \sum_{i=1}^{p} f_i (x_i - \overline{x})^2$$

L'écart-type de cette série est $s = \sqrt{V}$

Dans notre exemple 27, la variance vaut 16,58 et l'écart-type $\sqrt{16,58} \simeq 4,07$

On retrouve ce résultat sur les machines noté souvent sous la forme `xσ n`
On est amené à considérer la racine carrée de la variance pour avoir un résultat exprimé dans la même unité que le caractère étudié : l'écart-type et donc homogène à une sorte de « distance à la moyenne »ce qui est cohérent avec la démarche expérimentale.

Conclusion
On peut donc résumer de façon satisfaisante une série statistique en donnant :
– sa médiane et son écart-intercartile ;
– sa moyenne et son écart-type.