Statistique - Seconde

https://www.xm1math.net

xm1math.net

1. Vocabulaire

- **Population** : c'est l'ensemble étudié.
- **Individu** : c'est un élément de la population.
- **Effectif total** : c'est le nombre total d'individus.
- **Caractère** : c'est la propriété étudiée.
  Il y a deux sortes de caractères :
  - les caractères quantitatifs que l'on peut mesurer avec des nombres.
    On distingue les caractères quantitatifs **discrets** qui ne peuvent prendre qu'un nombre fini de valeurs (notes à un devoir...) et les caractères quantitatifs **continus** dont on regroupe les valeurs par intervalles (taille, durée d'écoute...).
  - les caractères qualitatifs (profession, marque de voiture ...).

Exemple(s)

On considère l'exemple suivant (qui servira pour les prochains paragraphes) : les 17 élèves d'une classe sont séparés en deux groupes.

- Les notes obtenues à un devoir par le **groupe A** sont : 4 ; 4 ; 9 ; 9 ; 9 ; 15 ; 15 ; 17 ; 17
- Les notes obtenues à un devoir par le **groupe B** sont : 6 ; 6 ; 8 ; 8 ; 10 ; 10 ; 12 ; 12

- La population étudiée est l'ensemble des 17 élèves
- Le caractère étudié est la note obtenue
- L'effectif total est égal à 17

2. Séries statistiques

a) Classement des données

Définition

On appelle **série statistique** la donnée simultanée des valeurs du caractère étudié (notées x_i), rangées dans l'ordre croissant, et des effectifs (notés n_i) de ces valeurs.

Exemple(s)

Pour le groupe A, le regroupement des notes en série statistique est :

valeur	4	9	15	17
effectif	2	3	2	2

Pour le groupe B, le regroupement des notes en série statistique est :

valeur	6	8	10	12
effectif	2	2	2	2

Définition

L'effectif cumulé croissant d'une valeur x est la somme des effectifs des valeurs y tels que $y \leqslant x$.
L'effectif cumulé décroissant d'une valeur x est la somme des effectifs des valeurs y tels que $y > x$.

Exemple(s)

Pour le groupe A de l'exemple, le tableau des effectifs cumulés croissants et décroissants est :

valeur	4	9	15	17
effectif cumulé croissant	2	5	7	9
effectif cumulé décroissant	7	4	2	0

2. Séries statistiques

Une série statistique peut aussi être définie par les fréquences, plutôt que par les effectifs.

Définition

Dans une population d'effectif total N, la **fréquence** de la valeur x_i d'effectif n_i est $f_i = \dfrac{n_i}{N}$.

Exemple(s)

Pour le groupe B de l'exemple, le tableau des fréquences est :

valeur	6	8	10	12
fréquence en pourcentage	25%	25%	25%	25%

b) Représentation graphique

Pour les caractères quantitatifs discrets, on peut utiliser le **diagramme en bâton** basé sur le principe suivant :
Dans un repère orthogonal, pour chaque valeur de la série statistique on trace un trait vertical dont la hauteur correspond à l'effectif (dans l'unité choisie).

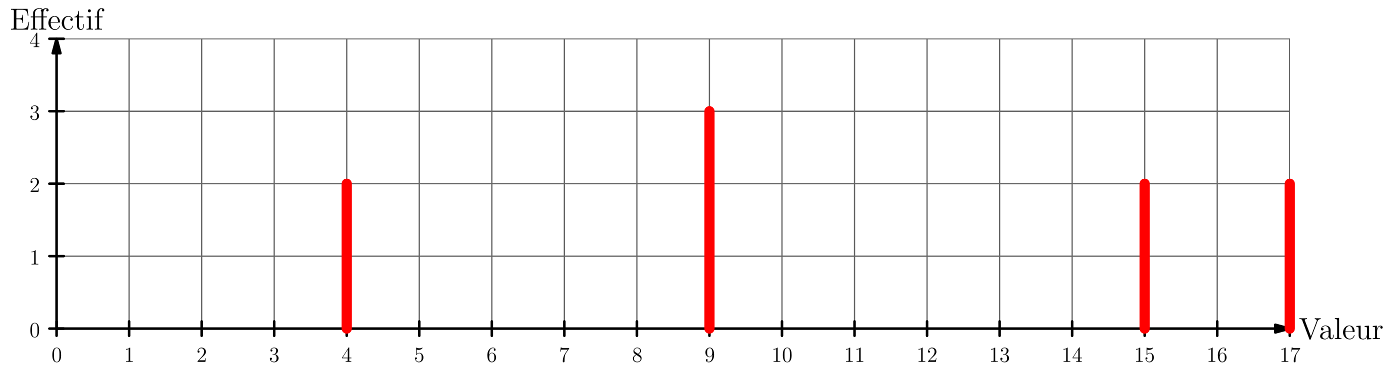
2. Séries statistiques

Exemple(s)

Pour le groupe A de l'exemple dont la série statistique était

valeur	4	9	15	17
effectif	2	3	2	2

le diagramme en bâtons donne :



2. Séries statistiques

Pour les caractères quantitatifs discrets, on peut aussi utiliser le **polygone des effectifs** qui se trace en reliant les points successifs de coordonnées (valeur,effectif) pour obtenir une ligne brisée.

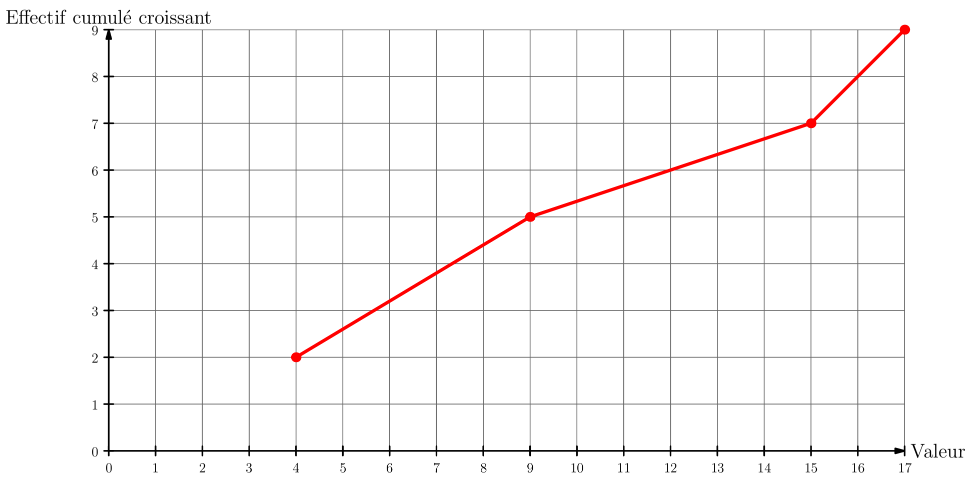
En utilisant les effectifs cumulés croissants ou décroissants, on obtient le polygone des effectifs cumulés croissants ou décroissants.

Exemple(s)

Pour le groupe A de l'exemple dont le tableau des effectifs cumulés croissants était

valeur	4	9	15	17
effectif cumulé croissant	2	5	7	9

le polygone des effectifs cumulés croissants donne :



3. Moyenne et écart-type d'une série statistique

Définition

On appelle moyenne de la série

valeur	x_1	x_2	$\cdots$	x_k
effectif	n_1	n_2	$\cdots$	n_k

, le réel $\overline{x} = \dfrac{n_1x_1 + n_2x_2 + \cdots + n_kx_k}{N}$

(N représente l'effectif total et k est le nombre de valeurs prises par le caractère)

Exemple(s)

- Pour la série du groupe A

valeur	4	9	15	17
effectif	2	3	2	2

, la moyenne est

$$\overline{x_A} = \frac{2 \times 4 + 3 \times 9 + 2 \times 15 + 2 \times 17}{9} = 11$$

- Pour la série du groupe B

valeur	6	8	10	12
effectif	2	2	2	2

, la moyenne est

$$\overline{x_B} = \frac{2 \times 6 + 2 \times 8 + 2 \times 10 + 2 \times 12}{8} = 9$$

3. Moyenne et écart-type d'une série statistique

Propriété(s)

- *En utilisant les fréquences, on a :* $\overline{x} = f_1x_1 + f_2x_2 + \cdots + f_kx_k$.
- *Si on ajoute à toutes les valeurs d'une série statistique le même nombre b, on augmente la moyenne de cette série de b.*
- *Si les valeurs d'une série statistique sont multipliées ou divisées par un même nombre a, la moyenne de cette série est aussi multipliée ou divisée par a.*
- *Si une population d'effectif N est composée d'une partie d'effectif N_A et de moyenne $\overline{x_A}$ et d'une autre partie d'effectif N_B et de moyenne $\overline{x_B}$, alors la moyenne de la population totale est telle que :* $\overline{x} = \dfrac{N_A\overline{x_A} + N_B\overline{x_B}}{N}$

Exemple(s)

Pour notre exemple :
la moyenne du groupe A, composée de $N_A = 9$ élèves, était $\overline{x_A} = 11$ et la moyenne du groupe B, composée de $N_B = 8$ élèves, était $\overline{x_B} = 9$.
On peut en déduire que la moyenne globale de la classe est

$$\overline{x} = \frac{N_A\overline{x_A} + N_B\overline{x_B}}{N} = \frac{9 \times 11 + 8 \times 9}{17} \approx 10,06$$

3. Moyenne et écart-type d'une série statistique

Définition

- **La variance** V d'une série statistique d'effectif total N est la moyenne des carrés des écarts à la moyenne : $V = \dfrac{n_1\left(x_1-\overline{x}\right)^2+n_2\left(x_2-\overline{x}\right)^2+\cdots+n_k\left(x_k-\overline{x}\right)^2}{N}$.
- **L'écart-type** σ d'une série statistique est la racine carrée de la variance : $\sigma=\sqrt{V}$.
  Il sert à mesurer la dispersion des valeurs de la série autour de la moyenne.

Remarque(s)

Pour des petites séries, on utilise ce genre de tableau pour effectuer les calculs :

valeur	x_1	x_2	$\cdots$	x_k
écart à la moyenne				
carré de l'écart				
effectif				

Le calcul de la variance se fait alors en calculant la moyenne des carrés de l'écart (avant-dernière ligne) en tenant compte des effectifs que l'on rappelle à la dernière ligne.

3. Moyenne et écart-type d'une série statistique

Exemple(s)

Dans le groupe A, la moyenne était de 11. Donc,

valeur	4	9	15	17
écart à la moyenne	-7	-2	4	6
carré de l'écart	49	4	16	36
effectif	2	3	2	2

Variance $V_A = \dfrac{2\times 49+3\times 4+2\times 16+2\times 36}{9} \approx 23,8$

Écart-type $\sigma_A \approx \sqrt{23,8} \approx 4,9$

Dans le groupe B, la moyenne était de 9. Donc,

valeur	6	8	10	12
écart à la moyenne	-3	-1	1	3
carré de l'écart	9	1	1	9
effectif	2	2	2	2

Variance $V_B = \dfrac{2\times 9+2\times 1+2\times 1+2\times 9}{8} = 5$

Écart-type $\sigma_B = \sqrt{5} \approx 2,2$

Propriété(s)

- *L'écart-type est de même unité que les valeurs de la série statistique.*
- *Si on ajoute à toutes les valeurs d'une série statistique le même nombre b, l'écart-type reste inchangé.*
- *Si les valeurs d'une série statistique sont multipliées ou divisées par un même nombre strictement positif a, l'écart-type de cette série est aussi multiplié ou divisé par a.*
- *Si les valeurs d'une série statistique sont multipliées ou divisées par un même nombre strictement négatif a, l'écart-type de cette série est multiplié ou divisé par $(-a)$*

4. Médiane et écart interquartile d'une série statistique

Principe

- La médiane M d'une série statistique partage la population en deux parties de telle façon qu'au moins 50% des valeurs du caractère soient inférieurs ou égaux à M et qu'au moins 50% des valeurs du caractère soient supérieurs ou égaux à M.
- Le principe théorique des quartiles est de partager la population en quatre parties de même effectif.
- Il existe plusieurs manières de déterminer pratiquement la médiane et les quartiles. Les calculatrices, les tableurs et les logiciels de statistique n'utilisent pas tous la même méthode. Une seule méthode pratique sera présentée ici.

Détermination pratique

En écrivant les valeurs du caractère par ordre croissant de telle façon que chaque valeur apparaisse un nombre de fois égal à son effectif :

- **Médiane :**
  - si l'effectif total est impair, la médiane M est la valeur du caractère située au milieu de la liste ;
  - si l'effectif total est pair, la médiane M est la demi-somme des deux valeurs du caractère situées au milieu de la liste.
- **Quartiles :** en partageant la liste en deux sous-séries de même effectif (si l'effectif total est impair, on te tient pas compte de la médiane)
  - le premier quartile Q_1 est la médiane de la sous-série inférieure ;
  - le troisième quartile Q_3 est la médiane de la sous-série supérieure ;
  - l'écart interquartile est défini par $Q_3 - Q_1$ et sert à mesurer la dispersion des valeurs.

4. Médiane et écart interquartile d'une série statistique

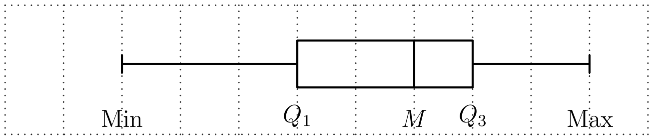
Exemple(s)

- Groupe A :
  - Détermination de la médiane : $4 ; 4 ; 9 ; 9 ; \overline{9} ; 15 ; 15 ; 17 ; 17$
    La médiane est $M = 9$ (valeur située au milieu, car l'effectif total est impair)
  - Détermination de Q_1 et Q_3 : on partage la liste en deux sous-séries de même effectif (on ne tient pas compte de la médiane car l'effectif total est impair) :
    $\underbrace{4 ; \underline{4} ; \underline{9} ; 9}$; 9 ; $\underbrace{15 ; \underline{15} ; \underline{17} ; 17}$
    Les effectifs des deux sous-séries étant pairs, Q_1 est la demi-somme de la sous-série inférieure et Q_3 est la demi-somme de la sous-série supérieure :
    $Q_1 = \frac{4+9}{2} = 6,5$; $Q_3 = \frac{15+17}{2} = 16$
    L'écart interquartile est $Q_3 - Q_1 = 16 - 6,5 = 9,5$
- Groupe B :
  - Détermination de la médiane : $6 ; 6 ; 8 ; \overline{8} ; \overline{10} ; 10 ; 12 ; 12$
    La médiane est $M = \frac{8+10}{2} = 9$ (demi-somme des deux valeurs situées au milieu, car l'effectif total est pair)
  - Détermination de Q_1 et Q_3 : on partage la liste en deux sous-séries de même effectif :
    $\underbrace{6 ; \underline{6} ; \underline{8} ; 8}$; $\underbrace{10 ; \underline{10} ; \underline{12} ; 12}$
    Les effectifs des deux sous-séries étant pairs, Q_1 est la demi-somme de la sous-série inférieure et Q_3 est la demi-somme de la sous-série supérieure :
    $Q_1 = \frac{6+8}{2} = 7$; $Q_3 = \frac{10+12}{2} = 11$
    L'écart interquartile est $Q_3 - Q_1 = 11 - 7 = 4$

4. Médiane et écart interquartile d'une série statistique

Diagramme en boîtes

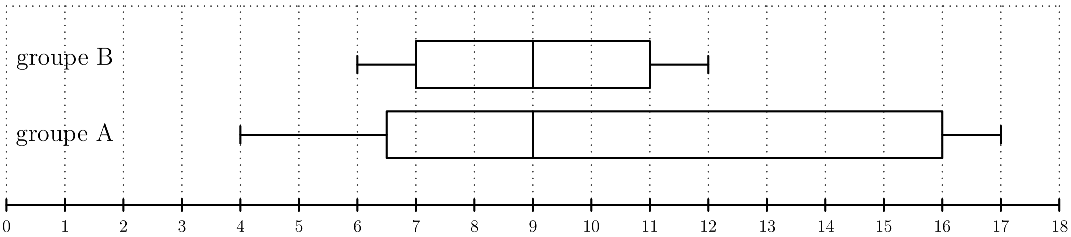
Le diagramme en boîtes d'une série statistique se construit de la façon suivante :



- 25% des valeurs du caractère sont comprises entre Min et Q_1 ;
- 25% des valeurs du caractère sont comprises entre Q_1 et M ;
- 25% des valeurs du caractère sont comprises entre M et Q_3 ;
- 25% des valeurs du caractère sont comprises entre Q_3 et Max.

Exemple(s)

- Pour le groupe A, on a $Min = 4$; $Q_1 = 6,5$; $M = 9$; $Q_3 = 16$; $Max = 17$
- Pour le groupe B, on a $Min = 6$; $Q_1 = 7$; $M = 9$; $Q_3 = 11$; $Max = 12$



5. Exemple d'étude d'un caractère quantitatif continu

Exemple(s)

On a posé la question suivante à 34 élèves : « Combien de temps avez-vous consulté votre téléphone portable hier ? ». Voici les résultats obtenus :

temps en minutes	$[0,30[$	$[30,60[$	$[60,120[$	$[120,180[$
nombre d'élèves	12	8	10	4

Remarque : les intervalles $[0,30[$, $[30,60[$ ... sont appelés **classes** du caractère.

Histogramme

Pour la représentation graphique d'un caractère continu, on utilise généralement un **histogramme** dont le principe est le suivant : dans un repère orthogonal on porte en abscisse les valeurs des bornes, puis pour chaque classe on trace un rectangle dont **l'aire est proportionnelle à l'effectif**.

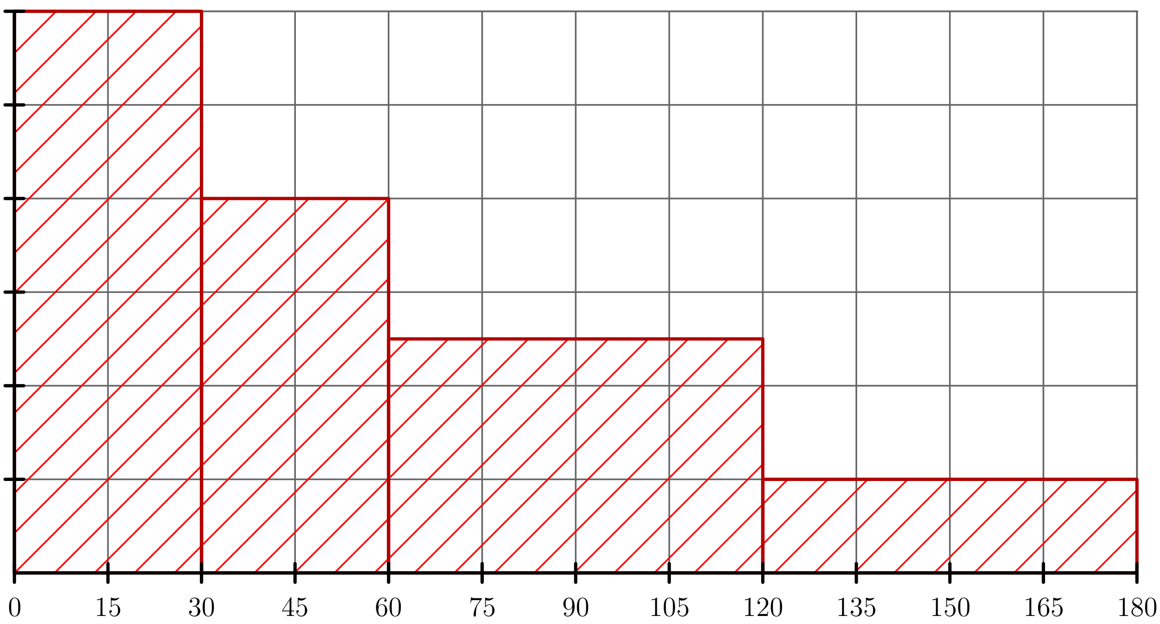
Pour déterminer la hauteur de chaque rectangle en respectant les unités données, on peut utiliser un tableau. Pour l'exemple, cela donne : (unités : en abscisse 1 cm représente 15 min et 1 cm^2 représente 1 élève)

temps en minutes	$[0,30[$	$[30,60[$	$[60,120[$	$[120,180[$
aire du rectangle en cm^2	12	8	10	4
largeur du rectangle en cm	2	2	4	4
hauteur du rectangle en cm	6	4	2,5	1

5. Exemple d'étude d'un caractère quantitatif continu

Histogramme

Il n'y a plus alors qu'à tracer :



5. Exemple d'étude d'un caractère quantitatif continu

Moyenne et écart-type pour des valeurs regroupées en classes

Pour calculer la moyenne d'une série statistique d'un caractère quantitatif continu, on prend comme valeur du caractère **le milieu de chaque classe**.

Pour l'exemple,

- Calcul de la moyenne :

valeur (milieu de chaque intervalle)	15	45	90	150
effectif	12	8	10	4

Moyenne $\overline{x} = \dfrac{12 \times 15 + 8 \times 45 + 10 \times 90 + 4 \times 150}{34} = 60$

- Calcul de la variance et de l'écart-type :

valeur (milieu de chaque intervalle)	15	45	90	150
écart à la moyenne	-45	-15	30	90
carré de l'écart	2025	225	900	8100
effectif	12	8	10	4

Variance $V = \dfrac{12 \times 2025 + 8 \times 225 + 10 \times 900 + 4 \times 8100}{34} \approx 1985,3$

Écart-type $\sigma \approx \sqrt{1985,3} \approx 44,6$

Fin du chapitre

xm1math.net