Statistiques

I] Séries statistiques :

Définition : Établir une **statistique**, c'est relever pour tous les **individus** d'une **population** la valeur d'un **caractère**.

Définition : Le caractère peut être :
- **qualitatif**, quand elle prend des valeurs non numériques ;
- **quantitatif**, quand elle prend des valeurs numériques.

Exemple : On étudie les élèves d'une classe.
- Caractère qualitatif : couleur des yeux (bleus, marron, verts, ...).
- Caractère quantitatif : taille en cm (150, 162, 175, ...).

II] Indicateurs de position :

Pour étudier une série statistique, on a besoin d'outils.
Un de ceux-ci est le **paramètre de position** : moyenne, médiane, quartile.

1 - Moyenne :

On considère la série statistique composée de p valeurs et donnée par le tableau ci-dessous :

Valeur	x_1	x_2	x_3	...	x_{p-1}	x_p
Effectif	n_1	n_2	n_3	...	n_{p-1}	n_p
Fréquence	f_1	f_2	f_3	...	f_{p-1}	f_p

Définition : La **moyenne** de cette série statistique est le réel noté $\overline{x}$ défini par

$$\overline{x} = \frac{n_1x_1 + n_2x_2 + ... + n_px_p}{n_1 + n_2 + ... + n_p} = \frac{n_1x_1 + n_2x_2 + ... + n_px_p}{N}$$

en notant $N = n_1 + n_2 + \ldots + n_p$ l'effectif total de la série.

Propriété : On peut également calculer la moyenne à l'aide des fréquences :

$$\overline{x} = x_1f_1 + x_2f_2 + \ldots + x_pf_p$$

Propriété (linéarité de la moyenne) : Soit a et b deux nombres réels.
- Si dans une série, on multiplie toutes les valeurs par a, alors **la moyenne est multipliée par a**.
- Si dans une série, on ajoute b à toutes les valeurs, alors on **ajoute b à la moyenne**.

Exemple : Voici les notes d'un contrôle sur 20.

Note	8	10	12
Effectif	3	4	3

1) Calculer la moyenne des notes obtenues.

$$\overline{x} = \frac{3 \times 8 + 10 \times 4 + 12 \times 3}{3 + 4 + 3} = \frac{24 + 40 + 36}{10} = \frac{100}{10} = 10$$

Avec les fréquences : $f_1 = \frac{3}{10} = 0{,}3$; $f_2 = \frac{4}{10} = 0{,}4$; $f_3 = \frac{3}{10} = 0{,}3$

$$\overline{x} = 8 \times 0{,}3 + 10 \times 0{,}4 + 12 \times 0{,}3 = 2{,}4 + 4 + 3{,}6 = 10$$

2) On ajoute deux points à toutes les notes. Calculer la nouvelle moyenne.
A l'aide de la propriété sur la linéarité de la moyenne on a : $\overline{x'} = \overline{x} + 2 = 10 + 2 = 12$

3) On multiplie toutes les notes par 1,3. Calculer la nouvelle moyenne.
A l'aide de la propriété sur la linéarité de la moyenne on a : $\overline{x''} = \overline{x} \times 1{,}3 = 10 \times 1{,}3 = 13$

2 - Médiane :

Définition : Dans une série de valeurs rangées par ordre croissant, la **médiane** est une valeur qui partage la série en deux parties de même effectif.

Méthodologie : En pratique, on adopte la démarche suivante pour déterminer la médiane M d'une série statistiques d'effectif total N :

- On range d'abord les N valeurs du caractère par ordre croissant.
- Si N est impair, M est la valeur centrale de la série.
- Si N est pair, M est la moyenne des deux valeurs « centrales » de la série.

Exemples :

On donne la série de valeurs :
13 ; 7 ; 8 ; 9 ; 10 ; 5 ; 4

On range dans l'ordre croissant :
4 ; 5 ; 7 ; 8 ; 9 ; 10 ; 13

Série de 7 valeurs, effectif impair donc la médiane est la valeur centrale (la $4^{ème}$).
M = 8

On donne la série de valeurs :
12 ; 11 ; 7 ; 5 ; 3 ; 15 ; 6 ; 9

On range dans l'ordre croissant :
3 ; 5 ; 6 ; 7 ; 9 ; 11 ; 12 ; 15

Série de 8 valeurs, effectif pair donc la médiane est la moyenne des deux valeurs centrales (la $4^{ème}$ et la $5^{ème}$: 7 et 9).

$$M = \frac{7 + 9}{2} = 8$$

3 - Quartiles, écart interquartile :

Définitions : On considère que les valeurs d'une série statistique sont rangées dans l'ordre croissant.

- Le **1^{er} quartile**, noté Q_1, est la plus petite valeur de la série telle qu'au moins 25 % des données soient inférieures ou égales à Q_1.
- Le **$2^{ème}$ quartile**, noté Q_2, est la médiane.
- Le **$3^{ème}$ quartile**, noté Q_3, est la plus petite valeur de la série telle qu'au moins 75 % des données soient inférieures ou égales à Q_3.
- On appelle **intervalle interquartile** l'intervalle $[Q_1 ; Q_3]$.

- On appelle **écart interquartile** la différence $Q_3 - Q_1$.

Exemple : On considère la série suivante de 20 notes (déjà rangées) :

5 ; 5 ; 7 ; 7 ; 8 ; 8 ; 8 ; 9 ; 9 ; 9 ; 9 ; 10 ; 10 ; 10 ; 11 ; 11 ; 12 ; 13 ; 14 ; 16.

1) Calculer le 1er quartile Q_1.
On a : Effectif total N = 20 .
On calcule : 25 % de N = 0,25 × 20 = 5.
On cherche la plus petite valeur telle qu'au moins 5 données lui soient inférieures ou égales.

$$\underbrace{5 ; 5 ; 7 ; 7 ; 8}_{5 \text{ valeurs}} ; 8 ; 8 ; 9 ; 9 ; 9 ; 9 ; 10 ; 10 ; 10 ; 11 ; 11 ; 12 ; 13 ; 14 ; 16.$$

Le 1er quartile Q_1 est 8.

2) Calculer le 3ème quartile Q_3.
On a : Effectif total N = 20 .
On calcule : 75 % de N = 0,75 × 20 = 15.
On cherche la plus petite valeur telle qu'au moins 15 données lui soient inférieures ou égales.

$$\underbrace{5 ; 5 ; 7 ; 7 ; 8 ; 8 ; 8 ; 9 ; 9 ; 9 ; 9 ; 10 ; 10 ; 10 ; 11}_{15 \text{ valeurs}} ; 11 ; 12 ; 13 ; 14 ; 16.$$

Le 3ème quartile Q_3 est 11.

3) Calculer la médiane (2ème quartile).
On a un nombre pair de valeur.
On calcule la moyenne des deux valeurs au centre : la 10ème et la 11ème.
La 10ème et la 11ème valeurs sont 9 et 9, donc : $M = \frac{9+9}{2} = 9$

La médiane est 9.

4) Calculer l'intervalle interquartile.
$[Q1 ; Q3] = [8 ; 11]$

5) Calculer l'écart interquartile.
$Q_3 - Q_1 = 11 - 8 = 3$

4 - Diagramme en boîte :

Définition : Pour résumer les différentes valeurs que l'on a déterminées, on réalise un diagramme appelé **diagramme en boîte** dans lequel figurent :

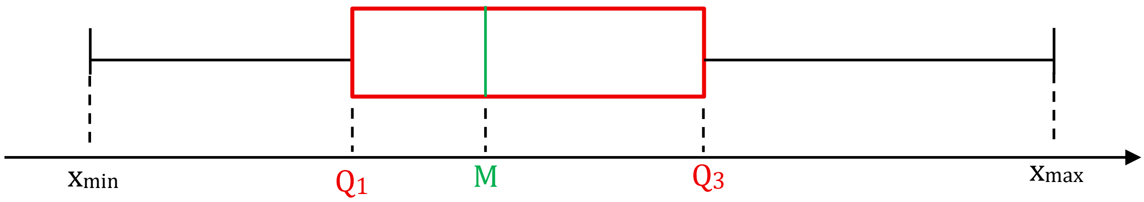
- les valeurs minimum et maximum ;
- les quartiles Q_1 et Q_3 ;
- la médiane Q_2.

Méthodologie : Pour construire un diagramme en boîte, on suit les étapes suivantes :

1. **Tracer un axe gradué :**
   - On choisit un axe horizontal.
   - On le gradue en fonction des valeurs de la série (de la valeur minimum à la valeur maximum).
2. **Placer les cinq valeurs importantes sur l'axe :**
   - la valeur minimum ;
   - le 1^er^ quartile Q_1 ;
   - la médiane Q_2 ;
   - le 3^ème^ quartile Q_3 ;
   - la valeur maximum.
3. **Tracer la boîte :**
   - On trace un rectangle (la « boîte ») entre Q_1 et Q_3 .
   - À l'intérieur de cette boîte, on trace un segment vertical à la position de la médiane Q_2.
4. **Tracer les moustaches :**
   - On trace un segment (une « moustache ») entre la valeur minimum et Q_1.
   - On trace un autre segment entre Q_3 et la valeur maximum.

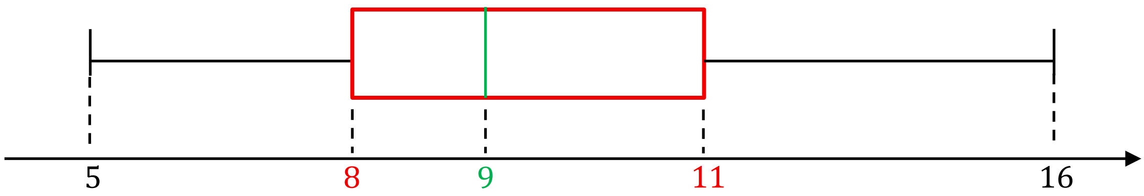
Ainsi, le diagramme en boîte est composé :

- d'une **boîte** délimitée par Q_1 et Q_3 , coupée par la médiane ;
- de **deux moustaches** reliant la boîte aux valeurs extrêmes (minimum et maximum).



Exemple : (à partir de la série précédente)

- Minimum = 5
- $Q_1 = 8$
- Médiane Q_2 est 9
- $Q_3 = 11$
- Maximum = 16



III] Indicateurs de dispersion :

1 - Etendue :

Définition : L'étendue d'une série statistique est la différence entre sa plus grande et sa plus petite valeur.

Exemple : Pour la série de 20 notes.

- $x_{min} = 5$
- $x_{max} = 16$

Etendue = 16 – 5 = 11

2 - Variance et écart-type :

Définitions :

- La **variance** est la moyenne des carrés des écarts de chaque valeur avec à la moyenne.

$$V = \frac{n_1(x_1 - \overline{x})^2 + n_2(x_2 - \overline{x})^2 + \ldots + n_3(x_3 - \overline{x})^2}{N}$$

- **L'écart type** est la racine carrée de la variance : $\sigma = \sqrt{V}$

Remarque : On utilise plus souvent l'écart type car il s'exprime dans la même unité que celle des valeurs de la série.

- Plus **l'écart-type est proche de 0**, plus les valeurs de la série sont **homogènes**.
- À l'inverse, **plus l'écart-type est grand**, plus les valeurs sont **hétérogènes**.

Exemple : On reprend la série,

Note	8	10	12
Effectif	3	4	3

On a déjà trouvé la moyenne : $\overline{x} = 10$.

1) Calculer la variance :

$$V = \frac{3(8-10)^2 + 4(10-10)^2 + 3(12-10)^2}{10} = \frac{3 \times 4 + 4 \times 0 + 3 \times 4}{10} = \frac{12 + 0 + 12}{10} = \frac{24}{10} = 2{,}4.$$

2) Calculer l'écart-type :

$$\sigma = \sqrt{2{,}4} \approx 1{,}55.$$