

Statistiques descriptives

16

Statistiques/
Probabilités

1 Vocabulaire :

1.1 Introduction

Une étude statistique commence par un recueil de données.

Prenons un exemple classique avec des notes :

Dans le tableau suivant sont regroupées les notes obtenues par les élèves d'une classe de seconde lors d'un contrôle.

4	5	6	6	6	8	8	9	10	11	11	11	12	12	12	12	13
13	14	14	16	16	16	16	16	16	16	17	17	17	17	19	19	19

Remarque : A bien regarder !

Toutes les définitions qui suivent vont prendre exemple sur cette série. Gardez bien les données sous les yeux !



Cours en vidéo

1.2 Définitions fondamentales

Définition : Vocabulaire de base

- **Série statistique** : Ensemble des valeurs collectées.
- **Population** : Ensemble sur lequel porte l'étude statistique.
- **Individus** : Éléments qui composent la population.
- **Caractère étudié** : Propriété que l'on observe sur les individus.
- **Caractère qualitatif** : Un caractère peut être qualitatif si on étudie une qualité des individus : (situation de famille, sexe, couleur des yeux, type d'habitation...).
- **Caractère quantitatif** : Un caractère peut être quantitatif si on mesure une quantité des individus.
- **Caractère continu** : Un caractère est dit continu lorsqu'il peut prendre théoriquement toutes les valeurs d'un intervalle (taille, temps d'écoute...); dans ce cas, les valeurs sont regroupées en intervalles appelés des classes.
- **Caractère discret** : Un caractère est dit discret lorsqu'il ne peut prendre qu'un nombre fini de valeurs. (Année de naissance, nombre de frères et sœurs, ...)
- **Modalité** : Les différentes valeurs obtenues sont appelées valeurs du caractère ou modalités.
- **Effectif** : Pour une modalité (ou valeur) du caractère, on appelle effectif le nombre d'individus de la population ayant cette valeur.
- **Effectif total** : Nombre total d'individus de la population (ou de l'échantillon).

Explications en
vidéo

Méthode : Maîtriser le vocabulaire de statistiques. (Niveau *)

Dans notre exemple de début de cours avec le relevé de notes, déterminez, la série statistique, la population, les individus, le caractère, sa nature et les trois premières modalités de la série, avec leur effectif et l'effectif total de la série.



Correction pdf

1.3 Notations :**Définition :** Tableau des effectifs

On regroupe souvent les données dans un tableau des effectifs :

On appelle souvent x_1, x_2, \dots, x_p les modalités, et n_1, n_2, \dots, n_p leur effectif correspondant.

Modalités	x_1	x_1	...	x_p
Effectif	n_1	n_2	...	n_p

L'effectif total, souvent noté N , vaut : $N = n_1 + n_2 + \dots + n_p$.

Exemple : Exemple en vidéo

On étudie le nombre d'enfants résidant dans chaque foyer d'une ville :

Nombre d'enfants	0	1	2	3	4	5	6
Nombre de foyers	290	170	155	95	43	20	10



Exemple en vidéo

Exemple : Application à notre exemple :

En reprenant le relevé de notes, la valeur x_1 du caractère est 4, et n_1 son effectif est 1.

La valeur x_2 du caractère est 5, n_2 son effectif est 1.

la valeur x_3 du caractère est 6, n_3 son effectif est 3 etc..

La plus grande modalité est 19, son effectif est 3 :

Notes	4	5	6	...	19
Effectif	1	1	3	...	3

1.4 Fréquence :**Définition**

- **Fréquence :**

Pour une valeur du caractère, on appelle **fréquence** le quotient de l'effectif de cette valeur par l'effectif total.

- **Calculer une fréquence :**

La fréquence peut être exprimée en pourcentage, en décimal ou en fraction :

$$\text{fréquence} = \frac{\text{effectif de la valeur}}{\text{effectif total}}$$



Vidéo
complémentaire

- **Notations :**

On note souvent f_1, f_2, \dots, f_p les fréquences respectives des modalités x_1, x_2, \dots, x_p , donc :

$$f_1 = \frac{n_1}{N}, \quad f_2 = \frac{n_2}{N}, \quad \dots, \quad f_p = \frac{n_p}{N}.$$

- **Propriété :**

On en déduit que : $0 \leq f_1 \leq 1$, $0 \leq f_2 \leq 1, \dots$, $0 \leq f_p \leq 1$ et $f_1 + f_2 + \dots + f_p = 1$

Méthode : Calculer des fréquences. (Niveau *)

Compléter le tableau des effectifs et fréquences, à 10^{-2} près, à partir du relevé de début de cours :

Notes	4	5	6	8	9	10	11	12	13	14	16	17	19
Eff.													
Fréq.													



Correction pdf

S'évaluer : QCM n°1



1.5 Effectif cumulé :

Définition

Effectif cumulé :

Pour une valeur x d'une série statistique quantitative, l'effectif cumulé croissant (respectivement décroissant) de x est la somme des effectifs des valeurs inférieures (respectivement supérieures) ou égales à x .



Vidéo de cours

Méthode : Calculer des Effectifs Cumulés. (Niveau *)

Compléter le tableau :

Note	4	5	6	8	9	10	11	12	13	14	16	17	19
Effectif	1	1	3	2	1	1	3	4	2	2	7	4	3
ECC													



Correction pdf

Combien d'élèves ont une note inférieure ou égale à 10 ?

Remarque

La dernière valeur des ECC est l'effectif total, puisque toutes les modalités sont inférieures ou égales à la valeur maximum de la série.

1.6 Fréquence cumulée :**Définition****Effectif cumulé :**

Pour une valeur x d'une série statistique quantitative, la fréquence cumulée croissante (respectivement décroissante) de x est la somme des fréquences des valeurs inférieures (respectivement supérieures) ou égales à x .

Méthode : Calculer des Effectifs Cumulés. (Niveau *)

Compléter le tableau à 10^{-2} près :

Note	4	5	6	8	9	10	11	12	13	14	16	17	19
Effectif	1	1	3	2	1	1	3	4	2	2	7	4	3
FCC													



Correction pdf

1.7 Exemple d'une série continue**Exemple**

On a interrogé en 2021 un échantillon de 4812 Français concernant la durée hebdomadaire d'écoute de la télévision (en heures).

Le caractère étudié, à savoir la durée d'écoute, est quantitatif continu : il peut prendre théoriquement toutes les valeurs de l'intervalle $[0 ; 50]$.

Les données sont regroupées en classes $[0 ; 10]$, $[10 ; 15[$, $[15 ; 20[$, $[20 ; 30[$ et $[30 ; 50]$.

Durée	$[0 ; 10[$	$[10 ; 15[$	$[15 ; 20[$	$[20 ; 30[$	$[30 ; 50]$
Effectif	972	924	826	1069	1021

Remarque : Bien pratique!

Il y a beaucoup de situations où le statisticien est obligé de ranger les données en classes pour étudier la série. Soit pour des séries continues, ou pour de grands nombres de modalités différentes. Pour les revenus, la taille, l'âge, .. d'une population par exemple.

Cela posera des difficultés dans l'étude de ces séries. Nous en reparlerons.

S'évaluer : QCM n°2





Vidéo de cours

2 Représentations graphiques

2.1 Séries à caractère quantitatif discret

Propriété : Représenter un diagramme en bâtons :

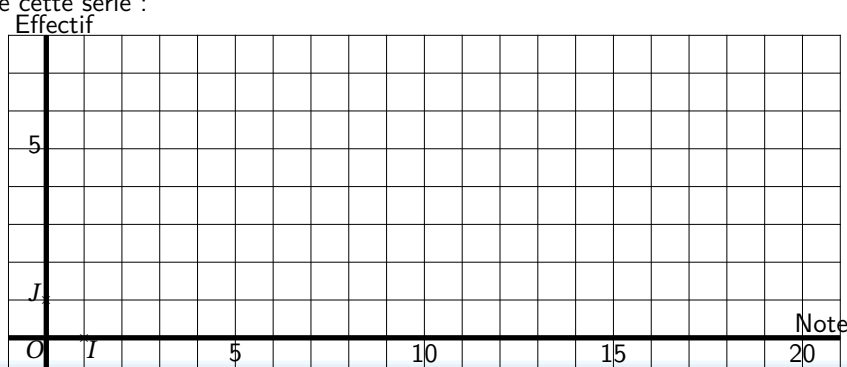
Dans un **diagramme en bâtons**, on représente une série statistique discrète par des segments dont la hauteur est proportionnelle à l'effectif de la valeur qu'ils représentent.

Méthode : Représenter un diagramme en bâtons. (Niveau *)

Compléter le tableau à 10^{-2} près :

On continue à travailler avec les données de l'exemple sur les notes. Voici le diagramme

en bâtons de cette série :



Correction pdf

Propriété : Représenter un diagramme circulaire :

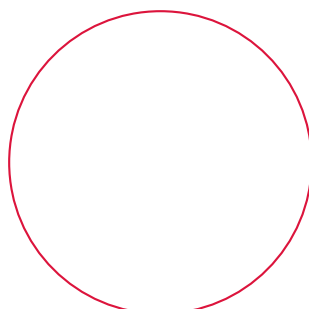
Un **Diagramme circulaire**, aussi appelé camembert (uniquement en France et par les non-matheux!!), permet de représenter un petit nombre de valeurs (ou de classes) par des angles proportionnels à la fréquence (ou l'effectif) de ces valeurs.

Méthode : Représenter un diagramme circulaire . (Niveau *)

Dans une compétition d'athlétisme, quatre pays s'affrontent : la France, l'Allemagne, la Suède et la Norvège. On note le pourcentage de médailles obtenues par chacun des pays :

Pays	France	Allemagne	Suède	Norvège
Pourcentage de médailles	25 %	10 %	40 %	25 %

Représenter le diagramme circulaire associé à cette série statistique



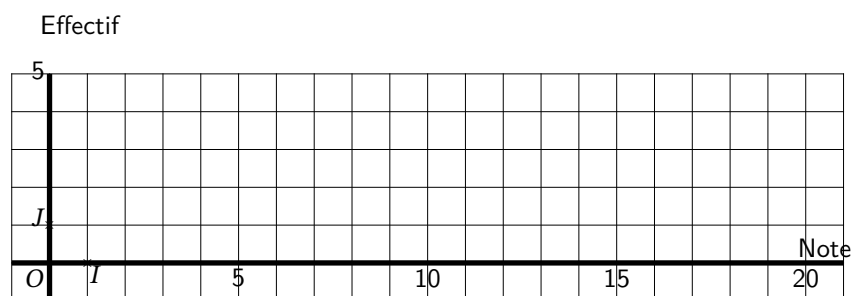
Correction pdf

Propriété : Nuage de points :

Dans un **nuage de points**, on représente une série statistique discrète par des points dont les abscisses sont les valeurs du caractère, et les ordonnées sont les effectifs correspondants, parfois reliés par des segments.

Méthode : Représenter un diagramme en bâtons. (Niveau *)

On travaille toujours avec les données de l'exemple sur les notes. Représenter le nuage de points de cette série :



Correction pdf

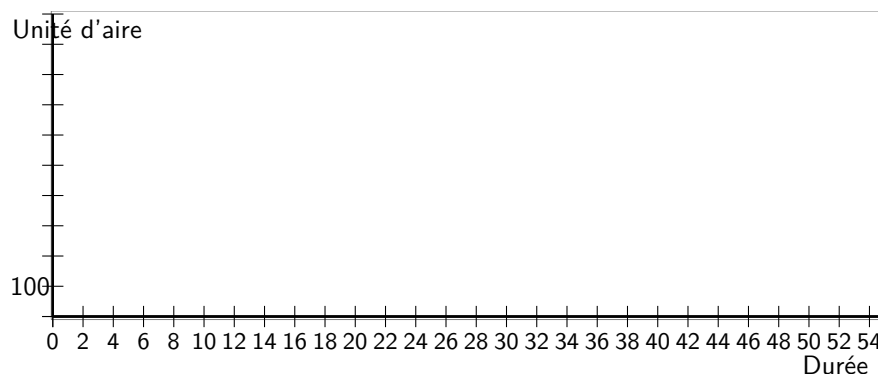
2.2 Séries à caractère quantitatif continu**Propriété : Histogramme**

Dans un **histogramme**, on représente une série statistique continue par des rectangles dont la **largeur** correspond à l'**amplitude** de chaque classe et dont l'**aire** est **proportionnelle** à l'**effectif** de la classe.

Méthode : Représenter un histogramme. (Niveau *)

On travaille avec les données de l'exemple sur la durée d'écoute de la télévision, proposé page 3.

Durée	[0 ; 10[[10 ; 15[[15 ; 20[[20 ; 30[[30 ; 50]
Effectif	972	924	826	1069	1021



Correction pdf

Propriété : Polygone d'effectifs ou de fréquences cumulés

- Le **polygone des effectifs cumulés croissants** (respectivement **décroissants**) d'une série statistique continue est la ligne brisée qui joint les points du plan dont les abscisses sont les bornes de chaque classe et dont les ordonnées sont les effectifs cumulés croissants (respectivement décroissants) de ces valeurs.
- Le **polygone des fréquences cumulées croissantes** (respectivement **décroissantes**) d'une série statistique

continue est la ligne brisée qui joint les points du plan dont les abscisses sont les bornes de chaque classe et dont les ordonnées sont les fréquences cumulées croissantes (respectivement décroissantes) de ces valeurs.

Méthode : Représenter un polygone des effectifs cumulés croissants : (Niveau *)

On travaille avec les données de l'exemple sur la durée d'écoute de la télévision. La situation est toujours celle de l'exemple sur le temps d'écoute de la télévision. Le tableau des effectifs cumulés croissants est le suivant :

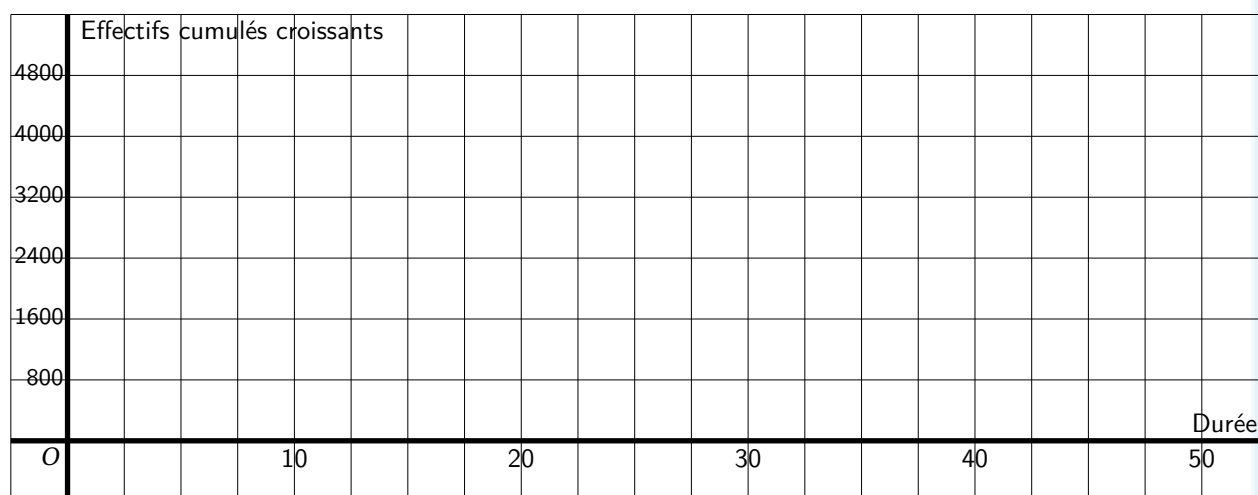
Durée	0	10	15	20	30	50
ECC	0	972	1896	2722	3791	4812



Correction pdf

Construire le polygone des effectifs cumulés croissants.

Combien de personnes regardent moins 20 heures la télévision ?



S'évaluer : QCM n°3



3 Paramètres de position :

3.1 La Moyenne :

Définition : Moyenne d'une série :

On considère une série statistique donnée par le tableau suivant :

Valeur	x_1	x_2	x_3	x_p
Effectif	n_1	n_2	n_3	n_p
Fréquence	f_1	f_2	f_3	f_p



Vidéo de cours

La **moyenne** de cette série statistique est le réel noté \bar{x} défini par

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{n_1 + n_2 + \dots + n_p} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{N}$$

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

Propriété : Avec les fréquences :

On peut également calculer la moyenne à l'aide des fréquences : $\bar{x} = x_1 f_1 + x_2 f_2 + \dots + x_p f_p$.

S'évaluer : QCM n°4



Bonus : Vidéos complémentaires



Le cours



Moyenne d'une série
rangée par classe



Moyenne calculée
à partir des fréquences

3.2 Médiane

Définition

Définition

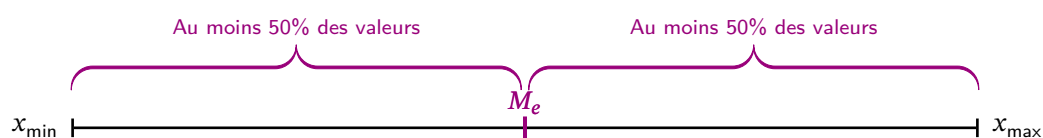
La **médiane** M d'une série statistique est un réel qui partage cette série en deux parties telles que :

- Au moins 50 % des valeurs sont inférieures ou égales à la médiane ;
- Au moins 50 % des valeurs sont supérieures ou égales à la médiane.



Vidéo de cours

Illustration



Propriété

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 pair, M est la moyenne des deux *valeurs centrales* de la série.
- Si N est impair, M est la valeur centrale de la série.

Méthode : Calculer la médiane d'une série simple : (Niveau *)

Benjamin a obtenu ces notes ce trimestre-ci en mathématiques :
8 ; 8 ; 13 ; 5 ; 7 ; 13 et 12.
Déterminer une médiane de cette série.



Correction
MathALEA

Méthode : Calculer la médiane d'une série rangée dans un tableau d'effectifs : (Niveau **)

On a réalisé 49 lancers d'un dé à 8 faces.
Les résultats sont inscrits dans le tableau ci-dessous :

Scores	1	2	3	4	5	6	7	8
Nombre d'apparitions	2	5	11	3	7	6	6	9

Déterminer une médiane de cette série.



Correction
MathALEA

S'évaluer : QCM n°5



Médiane ou moyenne ?

Remarque : Ne pas confondre !

Attention de bien comprendre les mécanismes de construction des deux paramètres de position : médiane et moyenne. Ce sont tous les deux des paramètres de position d'une série, c'est à dire qu'ils donnent une valeur centrale à la série.

Mais leur interprétation est à manipuler avec prudence.

Particulièrement, la moyenne est très sensible à des valeurs extrêmes.



Méthode : Différencier médiane et moyenne (Niveau **)

La série représente la répartition des salaires dans une entreprise :

salaire en € x_i	1200	1500	1800	2000	2200	25000
Nombre de salariés n_i	2	4	9	6	3	1

Calculer et comparer le salaire moyen et le salaire médian.



Correction

S'évaluer : QCM n°6



3.3 Quartiles

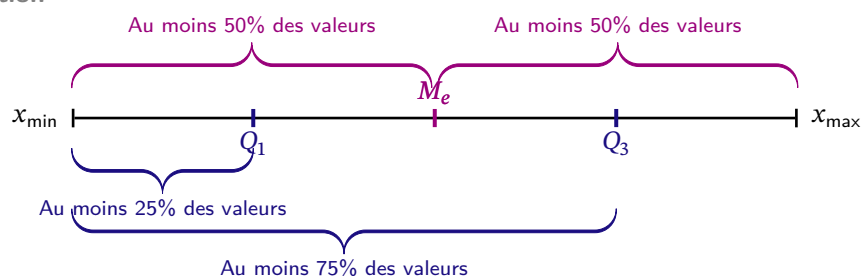
Définition

On considère une série statistique.

- Le premier **quartile** 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 troisième **quartile** Q_3 est la plus petite valeur de la série telle qu'au moins 75 % des données soient inférieures ou égales à Q_3 .



Illustration



Vidéo
complémentaire

Méthode : Déterminer les quartiles d'une série. (Niveau *)

Soit une série statistique d'effectif total N :

En pratique, on calcule le quart de l'effectif, soit $\frac{N}{4}$, puis $\frac{3N}{4}$.

- Si quotient est un nombre entier, il donne respectivement le **rang** (attention, pas sa **valeur**) de Q_1 ou de Q_3 .
- Si le quotient n'est pas un nombre entier, le **rang** du quartile est arrondi par excès à son entier supérieur.

Déterminer les deux quartiles de la série de notes donnée en début de cours.

.....

.....

.....



Correction

S'évaluer : QCM n°7



4 Paramètres de dispersion

4.1 Ecart inter quartile :

Définition

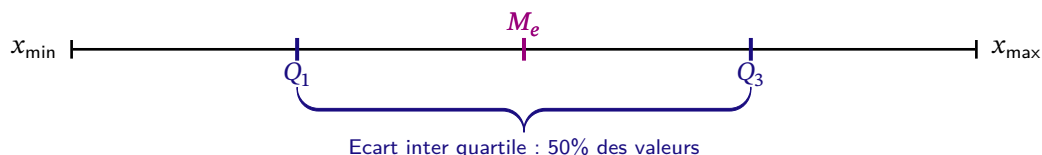
On appelle écart **inter-quartile** la différence entre le troisième et le premier quartiles :

$$Q_3 - Q_1$$



Vidéo de cours

Illustration



Propriété : Association paramètre de position et de dispersion :

En pratique, on associe l'**écart inter quartile**, paramètre de dispersion, à la **médiane**, paramètre de position.

Définition : Boîte à moustaches

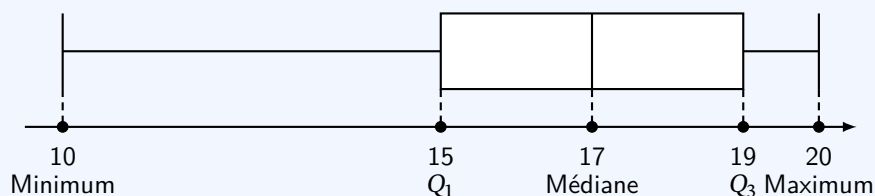
Il est commode d'illustrer la médiane et les quartiles d'une série par un diagramme, appelé **diagramme en boîte**, ou boîte à moustaches.

Pour une série donnée, on construit :

- Un **axe** permettant de repérer les valeurs de la variable étudiée,
- Un **rectangle** dont la longueur est égale à l'**écart inter-quartile** $Q_3 - Q_1$,
- Un **trait** représentant la **médiane**.
- Deux **traits** représentant les valeurs extrêmes de la série.

Exemple

Imaginons une série statistique, dont la valeur minimum vaille 10, la valeur maximum 20, $Q_1 = 15$, $Me = 17$, $Q_3 = 19$, on peut construire cette boîte à moustache :



S'évaluer : QCM n°8



4.2 Variance et écart-type :



Vidéo de cours



Introduction en pdf

Définition

La variance V d'une série statistique de moyenne \bar{x} dont les valeurs caractères sont $x_1, x_2, x_3, \dots, x_k$ et les effectifs

correspondants sont $n_1, n_2, n_3, \dots, n_k$ est égale à

$$V = \frac{n_1 \times (x_1 - \bar{x})^2 + n_2 \times (x_2 - \bar{x})^2 + \dots + n_k \times (x_k - \bar{x})^2}{n_1 + n_2 + \dots + n_k}$$

L'écart-type σ d'une série statistique de variance V est égal à : $\sigma = \sqrt{V}$

Méthode : Déterminer la variance et l'écart-type d'une série. (Niveau **)

Dans un service de maintenance, on a répertorié le nombre d'interventions par jour sur un mois. Calculer le nombre moyen d'interventions par jour puis compléter le tableau, afin de déterminer la variance et l'écart-type de la série.

Retrouver ces 3 résultats à la calculatrice. (seule compétence exigible !!)

Nombre d'interventions x_i	3	5	6	7	8	9
Nombre de jours n_i	2	4	9	6	3	1
$x_i - \bar{x}$						
$(x_i - \bar{x})^2$						
$n_i(x_i - \bar{x})^2$						



Correction

Remarque : Bonne nouvelle !!

La variance n'est utilisée à notre niveau que comme un outil qui permet de calculer l'écart-type. Son calcul étant fastidieux, on se contente d'une valeur donnée à la calculatrice.

Propriété : Bilan pour étudier une série :

Il faut associer à un paramètre de **position**, un paramètre de **dispersion**.

- On associe à la médiane, l'écart inter quartile, qui est égal à la différence entre le troisième et le premier quartiles.
- On associe à la moyenne, l'écart-type dont la définition est donnée plus loin.

C'est l'association de ce **couple de paramètre position-dispersion**, qui permet de caractériser le mieux une série statistique.

S'évaluer : QCM n°9

