

Chapitre 2 - Statistiques descriptives

Parcours 2 : Comment étudier le climat ?

A) Séries statistiques

1) Recueil de données

Définition 1 : Établir une **statistique**, c'est relever pour tous les « **individus** » d'une « **populations** » la valeur d'un « **caractère** ».

Définition 2 :

Le caractère peut être **qualitatif**, quand elle prend des valeurs non numériques ;

quantitatif, quand elle prend des valeurs numériques.

Exemples :

Population	Individu	Caractère	Type du caractère
Lycée du BdA	Elève	Couleur des yeux	qualitatif
Classe	Elève	Note	quantitatif
Union européenne	pays	Nombre de commune	quantitatif

Remarque :

Quand il est quantitatif, un caractère peut être :

- discret, quand il prend un nombre fini de valeurs ;
- continu, quand il prend toute valeur comprise entre deux nombres donnés.

B) Caractères de position

1) La moyenne

Pour résumer une série statistique très importante (ou comparer deux séries statistiques), on peut calculer la moyenne.

Définition 3 :

La **moyenne** d'une série statistique est le nombre réel qui pourrait remplacer toutes les valeurs de la série sans changer leur somme.

Avec la série de données et les **effectifs correspondants** :

Valeurs	x_1	x_2	x_3	...	x_p
Effectifs	n_1	n_2	n_3	...	n_p

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

Avec la série de données et les **fréquences correspondantes** :

Valeurs	x_1	x_2	x_3	...	x_p
Effectifs	f_1	f_2	f_3	...	f_p

$$\text{Moyenne} = \bar{x} = f_1 \times x_1 + f_2 \times x_2 + \dots + f_p \times x_p$$

Remarque :

Si les données sont regroupées par classes, le calcul est le même que précédemment en prenant le centre de la classe comme valeur pour le calcul.

2) Les quartiles

Pour une comparaison différente, on peut déterminer les quartiles.

Définition 4 :

La **médiane** est un nombre réel qui sépare la moitié des valeurs les plus petites de la moitié des valeurs les plus grandes.

La détermination de la médiane nécessite le classement des données

- Si la série est de taille impaire ($2n + 1$), la médiane est la valeur du terme de rang $n + 1$.
- Si la série est de taille paire ($2n$), la médiane est la demi-somme des valeurs des termes de rang n et $n + 1$.

Remarque : La médiane n'est pas sensible aux valeurs extrêmes (contrairement à la moyenne).

Définition 5 :

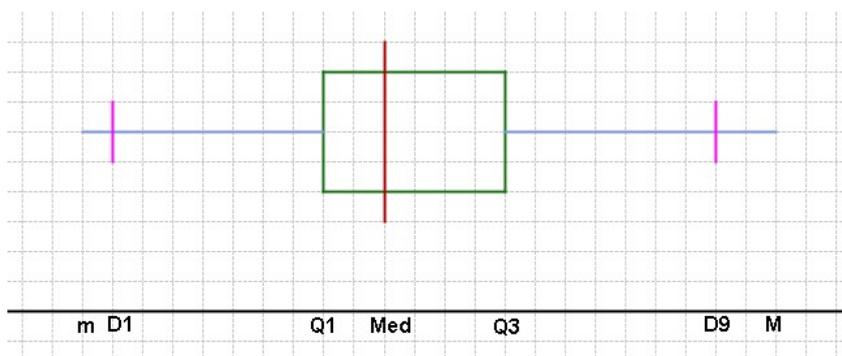
Les **quartiles** d'une série statistique sont les trois valeurs q_1 , q_2 et q_3 du caractère qui partagent la population en quatre parties de même effectif.

- **Premier quartile** q_1 : C'est la plus petite valeur de la série, telle qu'au moins 25% des données lui sont inférieures ou égales.
- **Deuxième quartile** q_2 : C'est la médiane.
- **Troisième quartile** q_3 : C'est la plus petite valeur de la série, telle qu'au moins 75% des données lui sont inférieures ou égales.

Remarque :

Dans de nombreuses disciplines, il est d'usage de présenter ce résumé sous une forme graphique :

- le **diagramme en boîte** (de TUCKEY)
- ou **boîte à moustaches**



C) Caractères de dispersion

1) L'étendue

Définition 6 :

L'**étendue** est la différence entre la valeur la plus grande et la valeur la plus petite du caractère étudiée.

L'étendue ne renseigne pas énormément sur la dispersion des valeurs puisque seules les deux valeurs extrêmes interviennent dans son calcul.

2) L'écart type

On pourrait faire la moyenne des écarts entre chaque valeur de la série et la moyenne. Mais les valeurs supérieures à la moyenne annuleraient les valeurs inférieures à la moyenne.

Définition 7 :

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

$$V = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2}{N}$$

2. L'**écart type** est la racine carrée de la variance : $s = \sqrt{V}$.

L'écart type est plus souvent utilisé que la variance car s'exprime dans la même unité que celle des valeurs de la série.

3) L'écart interquartile

Définition 8 :

L'**écart interquartile** est (comme son nom l'indique) l'écart entre Q_1 et Q_3 : $Q_3 - Q_1$

