

SÉRIES STATISTIQUES À UNE VARIABLE

Objectifs

Connaissances (A SAVOIR)

- Regroupement par classes d'une série statistique.
- Représentation d'une série statistique par un diagramme en secteurs, en bâtons, en colonnes, à lignes brisées.
- Indicateurs de position : mode, classe modale, moyenne, médiane, quartiles.
- Indicateurs de dispersion : étendue, écart type, écart interquartile $Q_3 - Q_1$.
- Diagrammes en boîte à moustaches.



Capacités (A SAVOIR FAIRE)



- Recueillir et organiser des données statistiques.
- Organiser des données statistiques en choisissant un mode de représentation adapté à l'aide des fonctions statistiques d'une calculatrice ou d'un tableur.
- Extraire des informations d'une représentation d'une série statistique.
- Comparer et interpréter des séries statistiques à l'aide d'indicateurs de position et de dispersion calculés avec les fonctions statistiques d'une calculatrice ou d'un tableur.
- Construire le diagramme en boîte à moustaches associé à une série statistique avec ou sans TIC.
- Comparer et interpréter des diagrammes en boîte à moustaches.

La statistique était, historiquement, l'étude de l'ensemble des données numériques de la population et permettait d'en décrire l'état. Aujourd'hui son utilité s'est généralisée.

Une étude statistique consiste en un **recueil de données** grâce à une enquête, puis en une **organisation** de ces données sous forme de tableaux statistiques et de représentations graphiques. Cette organisation est suivie du **traitement** de ces informations (calcul de paramètres de position et de dispersion), et la dernière étape étant l'**analyse** de ces informations.

1. Représentation des données statistiques

1.1. Organisation dans un tableau

Les données sont organisées dans un tableau :

Caractère (=propriété étudiée)	Effectifs (n_i)
x_1	n_1
x_2	n_2
x_3	n_3
x_4	n_4
$N = n_1 + n_2 + n_3 + n_4$	

Valeurs prises par le caractère

Effectif total

Nombre de fois où l'on a la valeur du caractère considéré

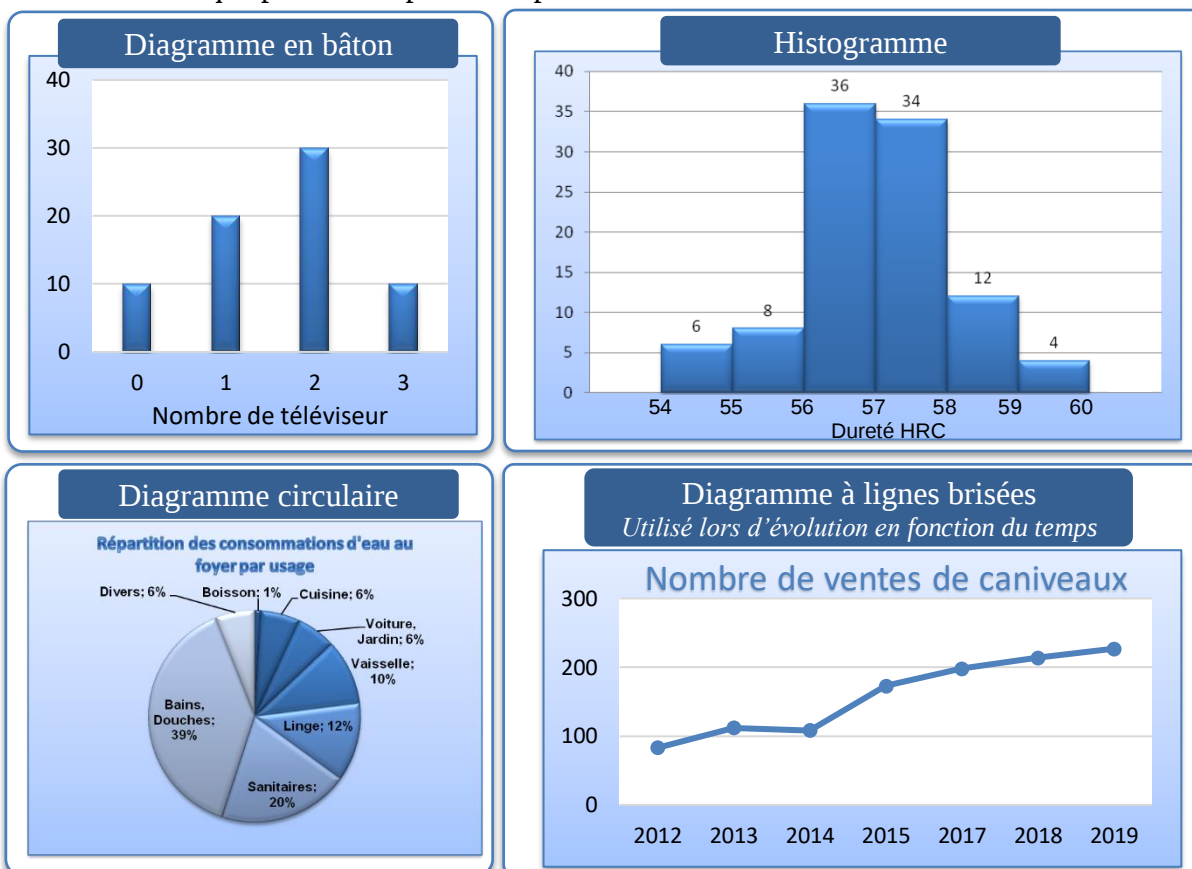
Lorsque les données sont nombreuses et que le caractère est mesurable de façon continue, on regroupe les données par classes : ce sont des intervalles notés entre crochets $[a ; b[$, a étant compris (**inclus**) dans l'intervalle et b non compris (**exclus**) de l'intervalle.

Tassement du béton (en cm)	Centres de classes (x_i)	Effectifs (n_i)
$[2 ; 2,5[$	$\frac{2 + 2,5}{2} = 2,25$	6
$[2,5 ; 3[$	$\frac{2,5 + 3}{2} = 2,75$	7
$[3 ; 3,5[$	3,25	7
$[3,5 ; 4[$	3,75	82
$[4 ; 4,5[$	4,25	85
$[4,5 ; 5[$	4,75	8
$[5 ; 5,5[$	5,25	5
		200

Lorsque l'on utilise un regroupement par classe, on est amené à calculer le **centre de classe** qui est le **milieu** de l'intervalle de la classe.

1.2. Représentations graphiques

Une série statistique peut être représentée par :



Attention, la représentation de données peut fausser leur interprétation si celle-ci n'est pas correctement adaptée ou réalisée.

2. Indicateurs de position

2.1. Mode

Le mode est la **valeur de la variable** statistique qui a **le plus grand effectif** (il peut y avoir plusieurs modes).

2.2. Classe modale

Dans le cas où la distribution statistique est faite à l'aide de classes, la classe modale est **la classe** qui contient **le plus grand effectif**.

Dans ce cas, le mode est alors le centre de la classe modale.

2.3. La moyenne

C'est la valeur théorique qui pourrait remplacer toutes les valeurs de la série sans changer la somme totale. Elle se calcule à l'aide de la relation suivante :

$$\bar{x} = \frac{\text{somme de toutes les valeurs}}{\text{nombre total de valeurs}}$$

La moyenne est aussi donnée par les fonctions statistiques de la calculatrice ou le tableur.

Dans le cas des séries où les caractères sont rangés par classes, on convient par approximation que toutes les valeurs appartenant à une même classe sont égales à la **valeur du centre de classe** x_i .



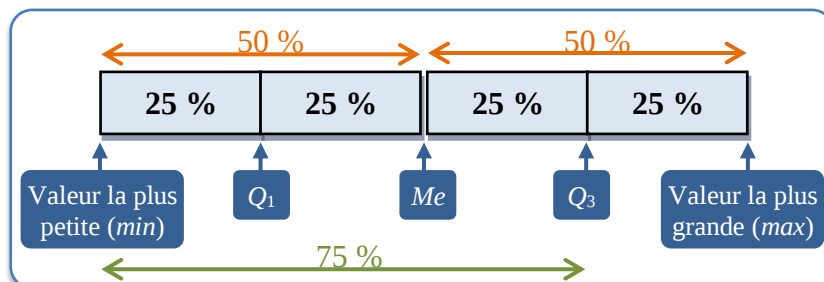
2.4. Médiane

La médiane Me d'une série statistique est le nombre qui partage cette série en **deux** séries de **même effectif**. La moitié des effectifs (50 %) a donc une valeur du caractère **en dessous ou égale** de la valeur médiane et l'autre moitié (50 %) **au-dessus ou égale**.

La médiane est donnée par les fonctions statistiques de la calculatrice ou le tableur.

2.5. Les quartiles

Le quartile Q_1 , la médiane Me et le quartile Q_3 partagent les valeurs ordonnées de la série en **quatre** parties **égales**.



- Q_1 est la **plus petite** donnée de la série pour laquelle **au moins** 25 % des données (soit $\frac{1}{4}$ des données) sont égales ou inférieures à Q_1 (Q_1 est obligatoirement une donnée de la série).
- Q_3 est la **plus petite donnée** de la série pour laquelle **au moins** 75 % des données (soit $\frac{3}{4}$ des données) sont égales ou inférieures à Q_3 (Q_3 est obligatoirement une donnée de la série).

Les quartiles sont donnés par les fonctions statistiques de la calculatrice ou le tableur.

3. Indicateurs de dispersion

3.1. Étendue

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

3.2. Écart-type

L'écart-type σ est un indicateur de dispersion associé à la **moyenne**.

Il traduit une **moyenne d'écarts** autour de la valeur moyenne $\bar{x}$.

Plus l'écart-type est grand, plus la dispersion des données est **grande**.

L'écart-type est donné par les fonctions statistiques de la calculatrice ou le tableur.

3.3. Écart interquartile $Q_3 - Q_1$

L'écart interquartile $Q_3 - Q_1$ est un indicateur de dispersion associé à la **médiane**.

C'est l'écart maximal entre les valeurs de la moitié centrale de la série.

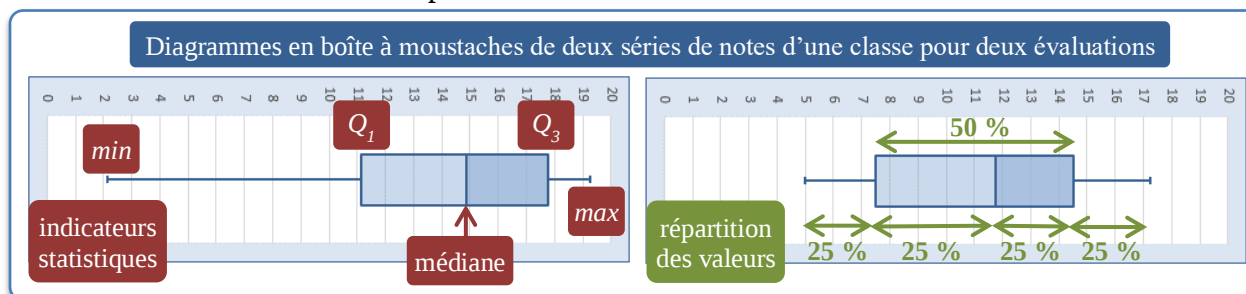
Plus l'écart interquartile $Q_3 - Q_1$ est grand, plus la dispersion des données est **grande**.

4. Diagramme en boîte à moustaches

Le diagramme en boîte à moustaches représente certains indicateurs d'une série.

La « boîte » est limitée par les quartiles Q_1 et Q_3 , et contient la **médiane**.

Les « moustaches » sont limitées par **les valeurs extrêmes**.



5. Application à la prise de mesures

5.1. Notion de chiffres significatifs

Le nombre de chiffres significatifs d'une valeur indique sa **précision** : plus la valeur comporte de chiffres significatifs, plus sa **précision est grande**.

Tous les chiffres d'un nombre sont significatifs, **sauf les zéros placés à gauche du premier chiffre non nul** (en revanche, les zéros écrits à la fin d'un calcul sont significatifs) :

Nombre	Chiffres significatifs
0,105	3
0,1050	4
10,105	5
0,055	2
0,0450	3
10,022	5

5.2. Notion d'incertitude de mesure

La mesure de la valeur d'une grandeur physique **n'est jamais exacte**.

La **moyenne** d'une série de mesures indépendantes est le **meilleur estimateur** de la valeur de la grandeur étudiée.

Le résultat d'une série de mesure s'écrit : $X = \bar{x} \pm U(x)$

avec :

- $\bar{x}$ la valeur moyenne des mesures ;

- $U(x)$ est l'incertitude type qui caractérise la dispersion des valeurs autour de la valeur moyenne $\bar{x}$.

On peut estimer avec un bon niveau de confiance que la vraie valeur est positionnée dans un encadrement appelé intervalle de confiance : $\bar{x} - U(x) \leq X \leq \bar{x} + U(x)$

L'incertitude associée à une mesure effectuée avec un **instrument** peut s'évaluer à partir d'indications fournies **par le constructeur**.