

# LES STATISTIQUES

## I) Le vocabulaire des statistiques

Les statistiques sont un domaine des mathématiques au développement assez récent mais dont les applications sont nombreuses et variées : leur but est de déduire des lois et des comportements généraux (de façon la plus fiable possible) à partir de l'étude d'un nombre limité d'exemples (un échantillon).

On les rencontre bien sûr dans l'étude des comportements humains (sondages, ...) et économiques, mais aussi dans l'industrie (fiabilité d'une machine, ...), en biologie ...

Une étude statistique comporte généralement quatre étapes :

- le recueil des données : sondages, recensements, enquêtes, mesures ...
- la présentation des résultats : tableaux, graphiques ...
- le calcul de paramètres statistiques : moyenne, médiane ...
- l'exploitation des données : savoir tirer des conclusions des calculs précédents.

Les statistiques fonctionnent avec un vocabulaire spécifique, que nous allons étudier avec l'exemple ci-après :

On interroge 1 000 personnes à la sortie d'un supermarché. Voici les quatre questions posées et les réponses obtenues :

Question 1 : Venez-vous régulièrement dans ce magasin ?

OUI	NON
912	88

Question 2 : Combien de temps venez-vous d'y passer ?

Temps ( en mn )	moins de 15	de 15 à 30	de 30 à 45	de 45 à 60	plus de 60
nb de personnes	62	363	465	87	23

Question 3 : Etes-vous satisfait du choix qui vous est offert ?

très satisfait	satisfait	moyennement satisfait	peu satisfait
418	250	276	56

Question 4 : Combien de fois par semaine venez-vous habituellement ? (en moyenne)

nb de fois	0	1	2	3	4	5
nb de réponses	88	253	287	241	98	33

Une étude statistique s'effectue sur un ensemble appelé **population**, dont les éléments sont des **individus** et consiste à observer, étudier un aspect sur chaque individu, nommé **caractère**. Il en existe deux types :

- Les **caractères qualitatifs** (client régulier de la question 1, satisfaction de la question 3, ...)

Dans l'étude de la satisfaction de la question 3, *Très Satisfait, Satisfait, ...* sont les **modalités** du caractère.

- Les **caractères quantitatifs** que l'on peut mesurer.

- Le *nombre de venue dans le magasin par semaine* est un **caractère quantitatif discret** : il ne peut prendre que des **valeurs** isolées 0, 1, 2, 3 ...
- le *temps passé dans le magasin* est un **caractère quantitatif continu** : il peut prendre toutes les valeurs de l'intervalle  $[0;+\infty[$ , les valeurs sont alors regroupées en **classes** ( $[0;15[$ ,  $[15;30[$  ...)

## II) Compréhension des données

Dans l'exemple précédent :

- 1) Combien de personnes viennent régulièrement au magasin ?
- 2) Combien de personnes au total sont interrogées ?
- 3) Combien sont satisfaits du magasin ?
- 4) Combien de personnes y sont restées entre 15 et 30 mn ?
- 5) Combien de personnes y sont restées moins de 30 mn ?
- 6) Combien de personnes y sont restées plus de 30 mn ?
- 7) Combien de personnes y sont restées entre 15 et 45 mn ?
- 8) Combien de personnes y sont restées au plus 15 mn ?
- 9) Combien de personnes y sont restées au moins 15 mn ?
- 10) Combien de personnes viennent en moyenne 2 fois par semaine ?
- 11) Combien de personnes viennent en moyenne plus de 2 fois ?
- 12) Combien de personnes viennent en moyenne au plus 2 fois ?
- 13) Comment expliquer le 88 de la question 4 ? Où le retrouve-t-on ?

### III) Représentations graphiques

#### 1) Le diagramme en bâtons

Ce diagramme sert à représenter un caractère qualitatif ou un caractère quantitatif discret.

##### Exemple

Nombre de livres lus par an sur une population de 100 personnes :

Nb de livres	0	1	2	3	4	5	6	7	8	9	10
effectifs	13	24	20	15	7	8	5	2	3	1	2

Construire le diagramme en bâtons correspondant.

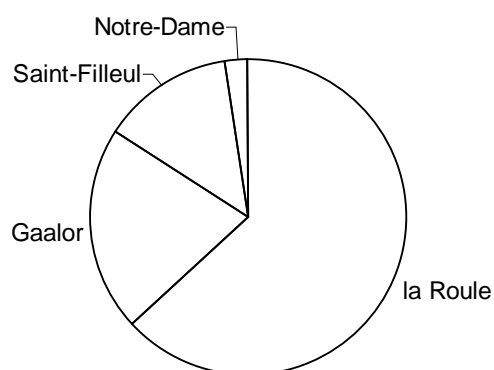
##### Remarque

On joint parfois les sommets de deux bâtons consécutifs par un segment de droite.

On obtient ainsi le **polygone des effectifs**.

#### 2) Le diagramme circulaire

Au XIX<sup>e</sup> siècle, Rouen possédait 4 sources pour l'approvisionnement en eau potable : La Roule (Darnétal), Gaalor (Nord de la Porte Bouvreuil), Saint-Filleul (Ouest de Rouen), Notre-Dame (voisine de Gaalor). Voici la répartition du volume d'eau journalier de ces sources :



Armoiries de la ville de Rouen

- Sachant que les différents volumes d'eau sont, par ordre croissant : 2, 14, 21, 63, attribuer à chaque secteur la valeur correspondante (sur 100 %).
- On sait que le volume total est de 1 900 m<sup>3</sup> par jour.  
Expliquer pourquoi et comment on peut en déduire les volumes d'eau des différentes sources.

Voici les dépenses d'investissement d'une commune :

Poste	Voirie	Social	Remboursement des emprunts	Enseignement - Sport Culture - Loisirs	Urbanisme Environnement
Dépenses	9 592 836	2 838 411	4 962 561	5 456 333	8 204 686

Expliquer comment représenter ce tableau sous forme de diagramme circulaire où, les angles au centre (et donc les aires) des secteurs sont proportionnels aux effectifs ou aux fréquences.

#### 3) Regroupement en classes - Histogrammes

Dans le cas de caractères quantitatifs et lorsque les données sont nombreuses, on les regroupe en classes.

Dans un histogramme, les effectifs (ou les fréquences) et les aires des rectangles sont proportionnels.

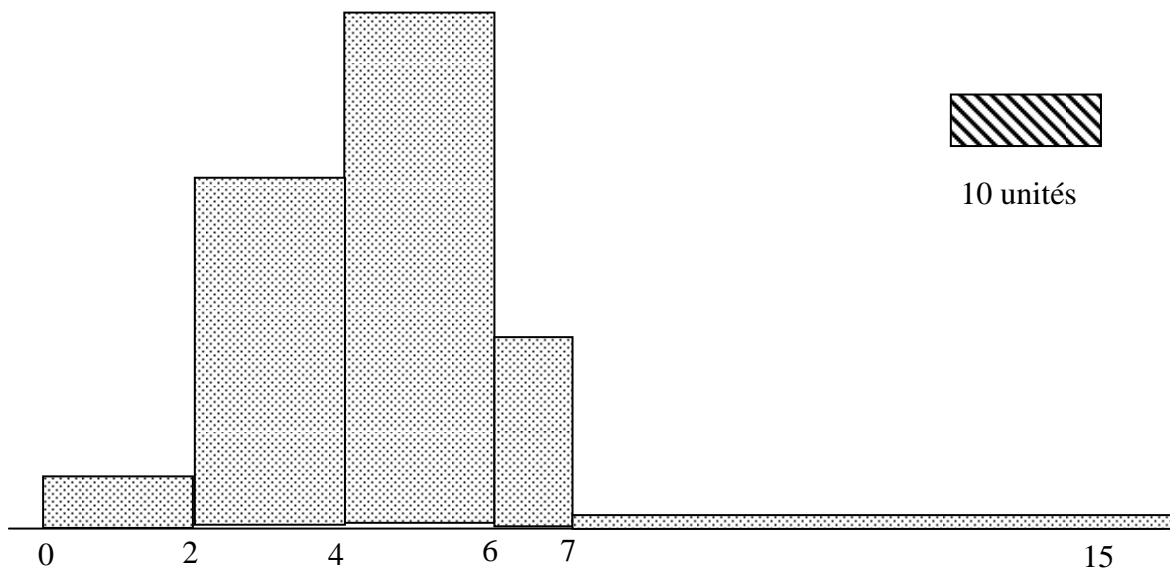
##### Exemple 1

Représenter par un histogramme la répartition d'un groupe d'élèves suivant la taille :

Taille ( en cm )	[150;155[	[155;160[	[160;165[	[165;170[	[170;175[	[175;180]
effectif	5	12	28	23	13	9

### Exemple 2

L'exemple suivant donne la largeur de 200 rues de Rouen considérées à la fin du XVIIIème siècle dont on donne un histogramme.



A l'aide de l'histogramme, remplir le tableau suivant résumant le nombre de rues ayant la caractéristique donnée:

Inférieures à 2m	Entre 2m et 4m	Entre 4m et 6m	Entre 6m et 7m	Entre 7m et 15m

### IV Paramètres de position

Ce sont des valeurs qui permettent d'avoir un ordre de grandeur de la série étudiée, il y en a trois principales :

#### 1) Moyenne

Soit  $x_1, x_2, \dots, x_p$  les valeurs prises par un caractère. A chaque valeur  $x_i$ , on associe l'effectif  $n_i$ .  $N$  est l'effectif total. La **moyenne** de la série statistique  $(x_i, n_i)$  est le nombre réel  $m$  ou  $\bar{x}$  défini par :

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{N} = \frac{\sum_{i=1}^p n_i x_i}{N}$$

#### Exemple

Reprendre les données de l'exemple III) 1) plus haut et calculer le nombre moyen de livres lus.

#### 2) Médiane

Lorsque le caractère étudié est ordonné, la **médiane** est une valeur  $Me$  du caractère qui partage la population en deux sous-ensembles de même effectif.

#### Exemples

Une série de valeurs est la suivante : 12, 9, 10, 16, 8, 11, 12.

Après avoir ordonné la série (8, 9, 10, 11, 12, 12, 16), la médiane est 11

Si nous ajoutons la valeur 5 : (5, 8, 9, 10, 11, 12, 12, 16), une médiane est alors toute valeur entre 10 et 11

Quelle est la médiane de l'exemple du III) 1)

#### 3) Mode

On appelle **mode** la (ou les) valeur(s) du caractère dont l'effectif est le plus grand.

Dans le cas d'un regroupement en classes, on appelle **classe modale** la classe dont l'effectif est le plus grand. Dans le cas du III) 1), le mode est

## V) Paramètre de dispersion

Ce sont des grandeurs qui permettent d'évaluer l'amplitude de la série étudiée.

L'**étendue** d'une série statistique est la différence entre la plus grande valeur et la plus petite valeur prises par le caractère (quantitatif). Dans l'exemple précédent, cette étendue est

L'**écart interquartile** est égal à la différence  $Q_3 - Q_1$  où

$Q_1$  est le premier quartile, c'est-à-dire la plus petite valeur de la série telle qu'au moins 25% des données soient inférieures ou égales à ce nombre  $Q_1$ .

$Q_3$  est le troisième quartile, c'est-à-dire la plus petite valeur de la série telle qu'au moins 75% des données soient inférieures ou égales à ce nombre  $Q_3$ .

Dans l'exemple du III) 1) le premier quartile est , le troisième quartile est et l'écart interquartile est

## VI) Exercice

L'étude d'une table de décès de la ville de Rouen pour l'année 1819 permet de présenter le tableau suivant :

Age du décès	Moins de 1 an	Entre 1 an et 5 ans	Entre 5 ans et 15 ans	Entre 15 ans et 30 ans	Entre 30 ans et 50 ans	Entre 50 ans et 75 ans	Plus de 75 ans
Effectif	622	347	258	316	392	760	222

- représenter par un diagramme convenable cette série statistique
- déterminer la classe modale
- pour la moyenne, on ne peut pas en donner une valeur exacte (il faudrait pour cela une variable discrète) mais on peut en donner une estimation en prenant comme valeur le centre de chaque classe. Calculer ainsi cette moyenne (il n'y a pas eu de décès d'une personne de plus de 100 ans à Rouen en 1819).
- Détermination de la médiane :

Il y a deux méthodes, l'une graphique, l'autre se fait par interpolation linéaire. Dans les deux cas, comme pour les paramètres précédents, on ne peut en trouver qu'une estimation.

On détermine tout d'abord les fréquences cumulées croissantes et décroissantes :

Age du décès	Moins de 1 an	Entre 1 an et 5 ans	Entre 5 ans et 15 ans	Entre 15 ans et 30 ans	Entre 30 ans et 50 ans	Entre 50 ans et 75 ans	Plus de 75 ans
Effectif							
Fréquence							
Fréquences Cumulées Croissantes							

On trace sur un même graphique le **polygone des fréquences cumulées croissantes** :

- l'axe des ordonnées doit être gradué jusqu'à l'effectif total
- Pour le polygone des effectifs cumulés croissants, on place les points (0;0), (1;??), ... (on pourrait faire de même pour le polygone des fréquences cumulées décroissantes).

Méthode graphique :

L'abscisse du point d'ordonnée 0,5 est une estimation de la médiane.

Utilisez le polygone des fréquences cumulées croissantes pour déterminer la médiane ainsi que le premier et le troisième quartile puis l'écart interquartile.

