

# Statistiques descriptives

## 1. Vocabulaire statistique

Pour pratiquer les sciences statistiques, on doit collecter un grand nombre de données. Pour cela, on effectue une **enquête statistique** sur un ensemble de personnes ou d'objets.

### ■ DÉFINITION : Population

L'ensemble sur lequel cette enquête est effectuée est appelé la **population**.

### ■ DÉFINITION : Individu

Chaque élément de cette population est appelé un **individu**.

Il est rare de pouvoir interroger tous les individus d'une population, donc on en interroge un sous-ensemble, qui permet une estimation des résultats de l'enquête.

### ■ DÉFINITION : Échantillon

Un tel sous-ensemble est appelé un **échantillon**.

On s'intéresse à l'ensemble des résultats d'une enquête statistique pour en tirer des informations.

### ■ DÉFINITION : Caractère

La question posée lors d'une enquête statistique porte sur une particularité de l'ensemble étudié, appelée le **caractère**.

### ■ DÉFINITION : Valeurs du caractère

Les différentes réponses possibles sont les **valeurs du caractère**. On les note  $x_i$ .



### Attention

Ces « valeurs » ne sont pas forcément numériques !

On peut regrouper le vocabulaire dans un tableau :

| Vocabulaire ensembliste | Vocabulaire statistique | Vocabulaire probabiliste |
|-------------------------|-------------------------|--------------------------|
| Ensemble                | Population              | Ensemble des issues      |
| Sous-ensemble           | Échantillon             | Événement                |
| Élément                 | Individu                | Issue                    |
| Propriété               | Caractère               | ×                        |

### Exemple :

On interroge tous les élèves de troisième du lycée franco-allemand de Buc.

- Population : l'ensemble des élèves de troisième.
- Individu : un élève de troisième.
- Échantillon : la classe de troisième C.
- Caractère : le nombre de frères et sœurs.
- Valeurs du caractère : 0, 1, 2, 3, ...

Il existe différents types de caractères :

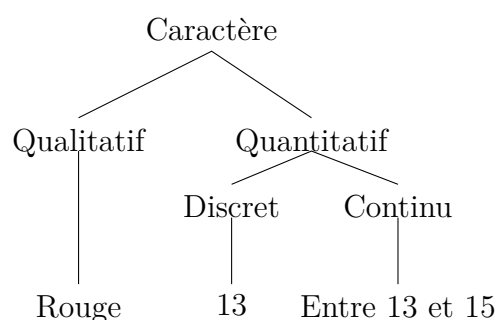
### ■ DÉFINITION : Caractères quantitatifs

Le caractère est dit **quantitatif** si on peut le mesurer, en associant un nombre à chaque valeur du caractère de la population. Il y a alors deux types de caractères quantitatifs possibles :

- On dit qu'il est quantitatif **discret** s'il peut prendre un nombre fini de valeurs (nombre de frères et sœurs...)
- On dit qu'il est quantitatif **continu** s'il peut prendre toutes les valeurs d'un intervalle de façon continue (taille, poids...)

### ■ DÉFINITION : Caractères qualitatifs

Le caractère est dit **qualitatif** si on ne peut pas le mesurer (couleur des yeux...).



### ■ DÉFINITION : Série statistique

On appelle **série statistique** l'ensemble des valeurs collectées.

### ■ DÉFINITION : Effectif

Pour une valeur donnée du caractère, le nombre de fois où elle apparaît est appelé l'**effectif** de cette valeur. Les effectifs sont notés  $n_i$ .

### ■ DÉFINITION : Fréquence

La **fréquence** d'une valeur du caractère est le quotient de l'effectif de la valeur par l'effectif total. Elle peut s'exprimer en pourcentage.

### Exemple :

On considère 250 familles. Le nombre d'engins motorisés par famille est donné par ce tableau :

|                 |    |     |    |    |
|-----------------|----|-----|----|----|
| Nombre d'engins | 0  | 1   | 2  | 3  |
| Effectif        | 25 | 150 | 65 | 10 |
| Fréquences      |    |     |    |    |

Compléter la ligne des fréquences.

Interpréter la fréquence de 1 :

.....

.....

**Remarque :** La somme des fréquences de toutes les valeurs de la série statistique est 1.

2. Caractères qualitatifs

Les séries statistiques peuvent être représentées grâce à des diagrammes circulaires, semi-circulaires, rectangles... C'est bien là l'unique information que l'on peut obtenir dans le cas des caractères qualitatifs. L'aire de chaque valeur du caractère est proportionnelle à l'effectif de cette valeur. Les fréquences permettent d'obtenir la part du diagramme qui devra être consacrée à chaque valeur du caractère.

On considère les données statistiques suivantes :

- Population : Élèves du lycée franco-allemand
- Échantillon : Classe de 3<sup>e</sup>
- Caractère : Couleurs des cheveux

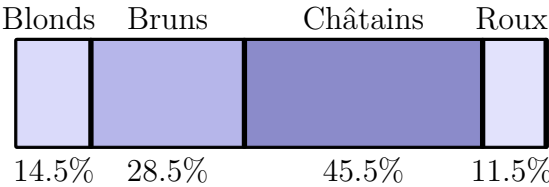
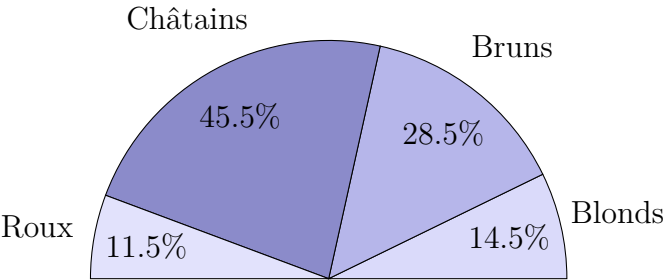
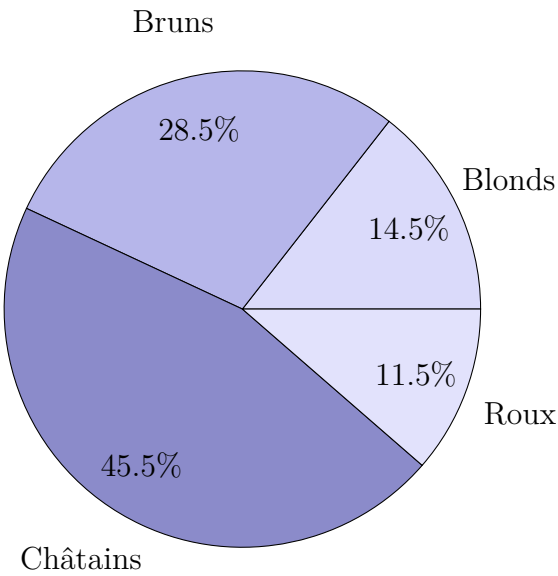
Après une enquête, on a obtenu la série statistique suivante.

| Valeurs du caractère, $x_i$ | Blonds | Bruns | Châtains | Roux | Autres | Total |
|-----------------------------|--------|-------|----------|------|--------|-------|
| Effectifs, $n_i$            | 29     | 57    | 91       | 23   | 0      | 200   |

Compléter le tableau suivant :

| Valeurs du caractère, $x_i$                                   | Blonds | Bruns | Châtains | Roux | Total |
|---------------------------------------------------------------|--------|-------|----------|------|-------|
| Effectifs, $n_i$                                              | 29     | 57    | 91       | 23   | 200   |
| Fréquences $f_i$                                              |        |       |          |      |       |
| Part d'un diagramme circulaire<br>$f_i \times 360^\circ$      |        |       |          |      |       |
| Part d'un diagramme semi-circulaire<br>$f_i \times 180^\circ$ |        |       |          |      |       |
| Part d'un rectangle de 10 cm<br>$f_i \times 10$               |        |       |          |      |       |

On obtient les diagrammes suivants :



3. Caractères quantitatifs

A. Caractères quantitatifs discrets

C'est dans le cas des caractères quantitatifs discrets que l'on peut obtenir le plus d'informations. Nous allons définir les caractéristiques que vous devez savoir donner. On travaillera dorénavant avec l'exemple suivant :

- Population : Élèves du lycée franco-allemand
- Échantillon : Classe de 3<sup>e</sup>
- Caractère : Nombre de frères et sœurs

Comme d'habitude, on écrit la série statistique sous forme d'un tableau :

| Nombre de frères et sœurs, noté $x_i$ | Nombre d'élèves, notés $n_i$ |  |  |
|---------------------------------------|------------------------------|--|--|
| 0                                     | 2                            |  |  |
| 1                                     | 11                           |  |  |
| 2                                     | 11                           |  |  |
| 3                                     | 3                            |  |  |
| 4                                     | 1                            |  |  |
| 5                                     | 2                            |  |  |
| Total                                 | 30                           |  |  |

On peut compléter le tableau de l'exemple précédent par une colonne dans laquelle on calcule la fréquence. On notera dans cette colonne les  $f_i$  et on utilise la relation donnée dans la page 1.

a. Étendue

■ **DÉFINITION** : Étendue  
L'**étendue** d'une série statistique est le nombre positif égal à la différence entre la plus grande valeur du caractère et la plus petite valeur du caractère.

Exemple :

Dans l'exemple précédent, l'étendue est .....

b. Mode

■ **DÉFINITION** : Mode  
On appelle **mode** d'une série statistique une valeur du caractère qui a le plus grand effectif.

Exemple :

Dans l'exemple précédent, .....

Exemple : Moyenne de la série précédente  
Quelle est le nombre moyen de frères et sœurs dans la classe interrogée ?  
.....  
.....

Exemple : Médiane de la série précédente  
Quelle est une médiane du nombre de frères et sœurs dans la classe interrogée ? Comment interpréter ce résultat ?  
.....  
.....  
.....  
.....

## c. Moyenne

### ■ DÉFINITION :

La **moyenne** d'une série statistique est égale au quotient de la somme de ses valeurs par l'effectif total.

Exemple :

On lance deux dés et on note le total des points obtenus : 10 ; 12 ; 5 ; 2 ; 3 ; 6 ; 7 ; 11

Calculer la moyenne des points obtenus. ....

### ■ PROPRIÉTÉ :

La moyenne est la valeur qui, substituée à chaque valeur du caractère de la série, redonne la même somme.

Exemple : Notion de moyenne pondérée

Lily joue au Volley-ball. Elle a participé à tous les matchs de la saison. La capitaine de son équipe a noté le nombre de points marqués par Lily lors de chaque match :

3 - 2 - 5 - 5 - 6 - 0 - 3 - 4 - 3 - 6 - 6 - 4 - 2 - 5 - 4 - 4 - 4 - 6 - 4 - 4

- Quel(le) est...
  - la population étudiée ? .....
  - le caractère étudié ? .....
  - l'effectif total de la série ? .....
  - l'étendue de la série ? .....

- Compléter le tableau ci-contre.

- Lily fait le calcul :  $\frac{0 + 1 + 2 + 3 + 4 + 5 + 6}{7} = 3$   
Expliquer pourquoi ce nombre n'est pas le nombre moyen de points de Lily.

| Nombre de buts | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|----------------|---|---|---|---|---|---|---|
| Effectif       |   |   |   |   |   |   |   |

- Une autre joueuse effectue le calcul ci-dessous.

$$\frac{0 + 2 + 2 + 3 + 3 + 3 + 4 + 4 + 4 + 4 + 4 + 4 + 4 + 5 + 5 + 5 + 6 + 6 + 6 + 6}{20}$$

Combien trouve-t-elle ? .....

Que signifie ce nombre ?

- La capitaine de l'équipe calcule plus simplement cette moyenne. Comment a-t-elle fait ?

Exemple :

Déterminer une médiane de la série de buts marqués par Lily. Que signifie ce nombre ?

## ■ DÉFINITION :

La **moyenne pondérée** d'une série statistique est égale à la somme des produits de chaque valeur du caractère par son effectif, divisée par l'effectif total.

### Exemple :

Dans l'exemple des frères et sœurs, compléter la dernière colonne du tableau en effectuant le produit de chaque valeur du caractère par son effectif, puis calculer le nombre moyen de frères et sœurs des élèves de la classe de 3<sup>e</sup> interrogée.

**Remarque :** Une moyenne n'est pas nécessairement égale à l'une des valeurs de la série.

Elle prend en compte les valeurs extrêmes d'une série.

## ■ PROPRIÉTÉ :

Une moyenne est toujours comprise entre la plus petite et la plus grande valeur du caractère.

Dans la suite, on considère les notes obtenues en mathématiques par Manon, Julien et Séverine au dernier trimestre. Les DM sont coefficient 1, les interrogations écrites coefficient 3 et les contrôles coefficient 5. Cela signifie que la note du contrôle compte cinq fois plus qu'une note de DM, ou qu'une note de contrôle équivaut à cinq fois la même note en DM ! Pour comprendre l'importance des coefficients, observons l'exemple suivant :

### Exemple :

Manon a obtenu

9 au contrôle, 20 à l'interrogation écrite, 7 et 8 aux DM.

Julien a obtenu

12 au contrôle, 11 à l'interrogation écrite et 13 et 14 aux DM.

et Séverine a obtenu

15 au contrôle, 4 à l'interrogation écrite, 16 et 17 aux DM.

Compléter les tableaux de la feuille suivante avec les valeurs de chacune des séries.

Pour chaque élève, répondre aux questions suivantes :

- Pour Manon :
  - Étendue :
  - Moyenne :
- Pour Julien :
  - Étendue :
  - Moyenne :
- Pour Séverine :
  - Étendue :
  - Moyenne :

Quelle appréciation écrire sur le bulletin de chacun ?

## d. Médiane

### ■ DÉFINITION :

Une **médiane** est un nombre qui sépare **une série ordonnée** en deux séries de même effectif.

On peut donc dire que

- la moitié des valeurs de la série sont inférieures à la médiane.
- la moitié des valeurs de la série sont supérieures à la médiane.

### ■ PROPRIÉTÉ :

- Si l'effectif total de la série est impair, une médiane est la valeur de la série qui, une fois ordonnée, est au milieu de cette série. Soit  $N$  l'effectif total de la série. Une médiane est la valeur du terme de rang  $\frac{N+1}{2}$  de la série ordonnée.
- Si l'effectif total de la série est pair, une médiane est un nombre qui se trouve entre les deux valeurs du milieu de la série. C'est, par exemple, la moyenne de ces deux valeurs de la série, c'est-à-dire la moyenne des valeurs  $x_{\frac{N}{2}}$  et  $x_{\frac{N}{2}+1}$ .

#### Exemple :

Un élève a obtenu les résultats suivants pendant le trimestre :

$$11 - 16 - 18 - 8 - 12$$

- Quelle est la moyenne de la série ?
- Quelle est une médiane de la série ? Comment interpréter ce nombre ?

#### Exemple :

Un élève a obtenu les résultats suivants pendant le trimestre :

$$11 - 16 - 18 - 8 - 12 - 13$$

- Quelle est la moyenne de la série ?
- Quelle est une médiane de la série ? Comment interpréter ce nombre ?

#### Exemple :

On considère les notes de Séverine, Julien et Manon.

Déterminer une médiane de chaque série.

— Manon :

— Julien :

— Séverine :

## B. Caractères quantitatifs continus

Pour les caractères quantitatifs continus, on sépare les valeurs du caractère en intervalles de même taille.

### ■ DÉFINITION :

Les intervalles dans lesquels on regroupe les valeurs sont appelés les **classes**.  
Leur taille est appelée l'**amplitude de la classe**.

On considère l'enquête suivante :

- Population : Élèves du lycée franco-allemand
- Échantillon : Classe de 3<sup>e</sup>
- Caractère : Taille, en cm

On a obtenu les résultats suivants :

156 - 159 - 168 - 161 - 156 - 176 - 174 - 162 - 158 - 161 - 174 - 161 - 151 - 166 - 167 - 176 - 153 - 173 - 166 - 175 -  
166 - 161 - 175 - 160 - 174 - 166 - 161 - 175 - 172 - 160

Regrouper cette série statistique par classes d'amplitude 5 cm et calculer les fréquences en % arrondies à l'unité :

| Tailles<br>$x_i$    | $150 \leq t < 155$ | $155 \leq t < 160$ | $160 \leq t < 165$ | $165 \leq t < 170$ | $170 \leq t < 175$ | $175 \leq t < 180$ | Total |
|---------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|-------|
| Effectifs<br>$n_i$  |                    |                    |                    |                    |                    |                    |       |
| Fréquences<br>$f_i$ |                    |                    |                    |                    |                    |                    |       |
| Fréquences<br>en %  |                    |                    |                    |                    |                    |                    |       |

## Histogramme des effectifs

On peut associer aux classes un **histogramme**, dans lequel la légende est **indispensable**.

Si les classes ont toutes la même amplitude, alors la hauteur des rectangles est proportionnelle à l'effectif de chaque classe, comme dans la figure de gauche ci-dessous. Sinon, dans cet exemple, en regroupant les deux premières classes, on obtiendra la représentation de droite :

