

# Statistique

« Il est prouvé que fêter les anniversaires est bon  
pour la santé.

Les statistiques montrent que les personnes qui en fêtent  
le plus deviennent les plus vieilles. »

Den Hartog

## Introduction

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.

# I. Vocabulaire

On a demandé aux élèves d'un collège le sport qu'ils pratiquaient.

On a ensuite regroupé les résultats dans un tableau.

Sport	football	basket	tennis	handball	danse
nombre d'élèves	106	52	37	89	102

Une étude statistique s'effectue sur un ensemble appelé **population**.

C'est ici l'ensemble des élèves d'un collège.

Les éléments de la population sont des **individus**.

Ici, chaque élève est un individu.

Une étude statistique consiste à observer, étudier un aspect sur chaque individu nommé **caractère**.

Ici, le caractère est le sport pratiqué.

Le caractère a différentes **valeurs** (ou **modalités**).

Ici les valeurs sont « football », « basket », « tennis », « handball » et « danse ».

Il existe deux types de caractère :

- Les **caractères qualitatifs** : les valeurs de la série ne sont pas des nombres.

C'est le cas ici : « football », « basket », « tennis », « handball » et « danse ».

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

## A. L'effectif

**L'effectif** de chaque valeur est le nombre d'individus qui la composent.

Ici, l'effectif de la valeur « football » est 106, l'effectif de la valeur « basket » est 52, etc.

**L'effectif total** est la somme des effectifs de toutes les valeurs.

Ici, l'effectif total est  $106 + 52 + 37 + 89 + 102 = 386$ .

## B. La fréquence

La **fréquence** d'une valeur d'une série statistique est le quotient de l'effectif de cette valeur par l'effectif total.

Ici, l'effectif de la valeur « football » est 106, sa fréquence est égale à  $\frac{106}{386} \approx 0,27$ .

### Remarques

- Les fréquences sont toujours inférieures à 1 et la somme des fréquences est égale à 1.
- Les fréquences sont souvent exprimées en pourcentage (il suffit de multiplier la fréquence par 100).

## II. Représentation graphique d'une série statistique

### A. Diagramme en bâton

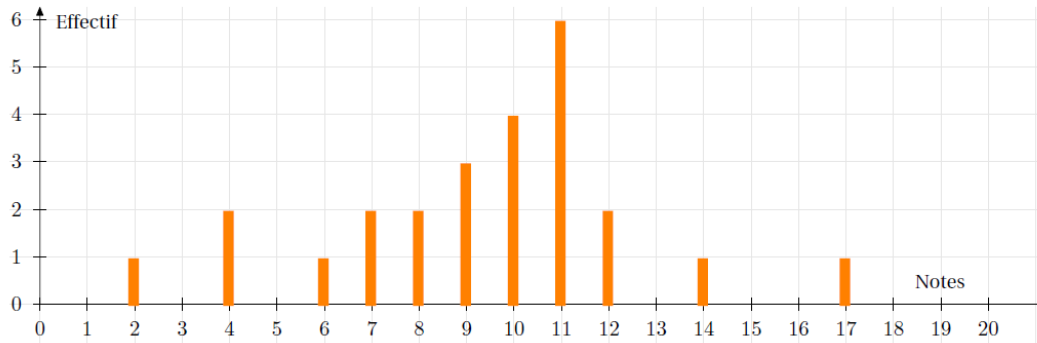
#### Utilisation

Ce diagramme sert à représenter un caractère qualitatif ou un caractère quantitatif discret (un nombre fini de valeurs).

la hauteur de chaque bâton est proportionnelle à l'effectif (ou à la fréquence) associé à chaque valeur.

**Exemple :** diagramme en bâtons représentant la série des notes obtenues par une classe à un contrôle.

Notes	2	4	6	7	8	9	10	11	12	14	17	Total
Effectif	1	2	1	2	2	3	4	6	2	1	1	25
Fréquence %	4	8	4	8	8	12	16	24	8	4	4	100



## B. Diagramme circulaire

### Utilisation

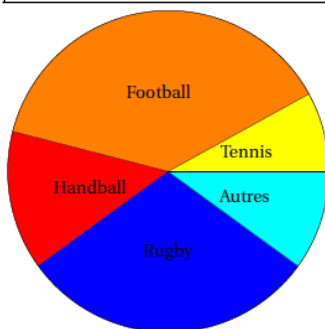
Ce diagramme est utilisé pour tout type de caractère.

La mesure de chaque secteur angulaire est proportionnelle à l'effectif (ou à la fréquence) associé.

**Exemple :** diagramme circulaire représentant la répartition des adhérents à un club sportif.

Sport	Tennis	Football	Handball	Rugby	Autres	Total
Effectif	4	19	7	15	5	50
Fréquence %	8	38	14	30	10	100
Mesure de l'angle en degrés	28,8	136,8	50,4	108	36	360

$\times 3,6$



## C. Histogramme

### Utilisation

Lorsque le caractère étudié est quantitatif et lorsque les valeurs sont regroupées en classe.

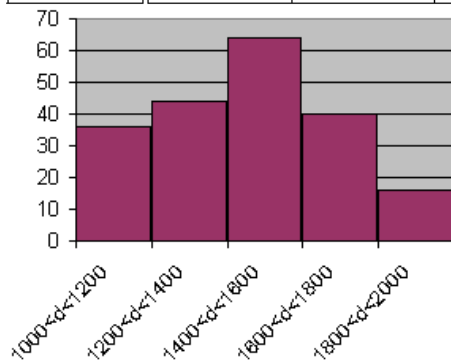
L'aire de chaque rectangle est alors proportionnelle à l'effectif (ou à la fréquence) associé à chaque classe. Lorsque les classes ont la même **amplitude** (largeur des intervalles), c'est la hauteur de chaque rectangle qui est proportionnelle à l'effectif.

### Remarque

Lorsque le caractère est quantitatif discret et lorsque les données sont nombreuses, on peut les regrouper en classe.

**Exemple :** histogramme représentant la répartition des salaires dans une entreprise.

Salaires	$1000 \leq S < 1200$	$1200 \leq S < 1400$	$1400 \leq S < 1600$	$1600 \leq S < 1800$	$1800 \leq S < 2000$	Total
Effectif	36	44	64	40	16	200
Fréquence	0,18	0,22	0,32	0,2	0,08	1



## D. Nuage de points

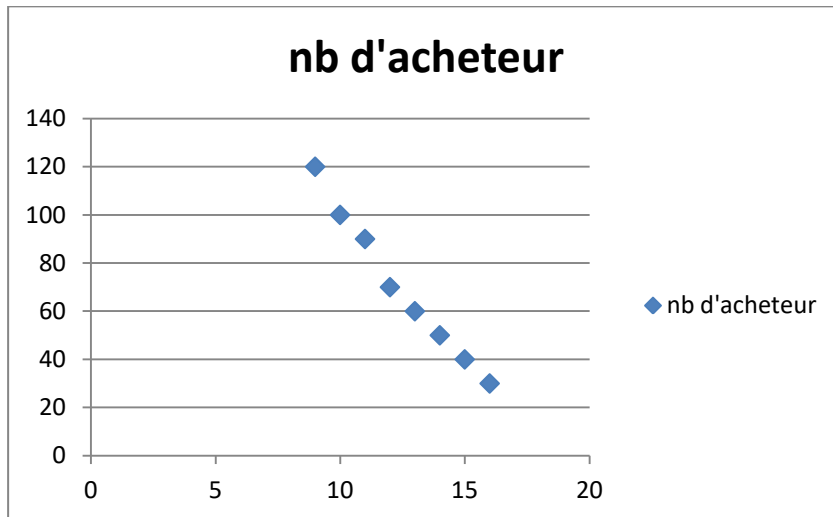
### Utilisation

Lorsque le caractère étudié est quantitatif discret.

Chaque couple de valeurs est représenté par un point dans un repère orthogonal.

**Exemple :** nuage de points représentant le nombre d'acheteur potentiel d'un produit en fonction de son prix de vente.

Prix (€)	9	10	11	12	13	14	15	16
Nb d'acheteur	120	100	90	70	60	50	40	30



### III. Paramètres d'une série statistique

#### A. Paramètre de position : la moyenne

La **moyenne** d'une série statistique est le quotient de la somme de toutes les valeurs de cette série par l'effectif total. On la note  $\bar{m}$ .

**Exemple 1** : Marie a indiqué dans le tableau suivant le temps qu'elle a passé devant la télévision la semaine dernière.

Jour	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche
Temps (en min)	61	53	114	60	42	126	132

Le temps quotidien moyen est :

$$\bar{m} = \frac{61+53+114+60+42+126+132}{7} = \frac{588}{7} = 84 \text{ (en min)}$$

La **moyenne pondérée** d'une série de données est la somme des produits des valeurs par leur effectif, divisée par l'effectif total.

**Exemple 2 :** La taille en mètre de 100 requins blancs.

Taille (m)	1,5	2	2,5	3	3,5	4	4,5
Effectif	8	10	25	32	19	4	2

La taille moyenne est :

$$\bar{m} = \frac{1,5 \times 8 + 2 \times 10 + 2,5 \times 25 + 3 \times 32 + 3,5 \times 19 + 4 \times 4 + 4,5 \times 2}{100} \approx 2,82 \text{ (en m)}$$

**Exemple 3 :** Un supermarché a relevé les dépenses (en €) de ses clients en 2h un jour donné.

Dépenses (en €)	[0 ; 30[	[30 ; 60[	[60 ; 100[	[100 ; 120[
Effectif	12	25	42	67
Milieu de classe	15	45	80	110

Pour déterminer la moyenne, on détermine les milieux des classes de la série puis on effectue le calcul.

$$\bar{m} = \frac{15 \times 12 + 45 \times 25 + 80 \times 42 + 110 \times 67}{146} \approx 82,43 \text{ (en €)}$$

## A. Paramètre de dispersion : l'étendue

**L'étendue** d'une série statistique est égale à la différence entre la plus grande et la plus petite des données de la série.

### Remarque

L'étendue est une mesure de **dispersion** des valeurs : plus l'étendue est grande, plus les valeurs sont dispersées.

**Exemple 1 : masse des baguettes.**

$$e = 260 - 235 = 25 \text{ (en g)}$$

**Exemple 2 : connexion internet.**

$$e = 300 - 40 = 260 \text{ (en min)}$$