

I. VOCABULAIRE

Réaliser une étude statistique consiste à classer les **individus** d'une **population** en fonction d'un **caractère** (ou **variable**).

Exemple :

- Classer les **élèves** d'une **classe** en fonction de leur **âge**.
- Classer les **voitures** garées sur un **parking** en fonction de leur **couleur**.
- Classer les **joueurs** d'une **équipe de foot** en fonction de leur **poste**.
- Classer des **forfaits** d'un **opérateur téléphonique** en fonction de leur **prix**.

Le caractère étudié peut être **qualitatif** (couleur, poste des joueurs de foot...) ou **quantitatif** (âge, prix...).

Exemple :

Âge (caractère)	14	15	16	17	TOTAL
Effectif	1	27	5	2	35
Effectif cumulé croissant	1	28	33	35	
Effectif cumulé décroissant	35	34	7	2	
Fréquence (%)	0,029 (2,9)	0,771 (77,1)	0,143 (14,3)	0,057 (5,7)	1 (100)

Pour chaque valeur du caractère, on peut indiquer l'**effectif** (le nombre d'individus) ou la **fréquence** (la proportion d'individu par rapport à la totalité de la population).

Cette fréquence peut s'exprimer sous la forme :

- d'un nombre décimal entre 0 et 1 (Exemple : 0,057)
- d'un pourcentage (Exemple : 5,7%)
- d'une fraction (Exemple : $\frac{2}{35}$)

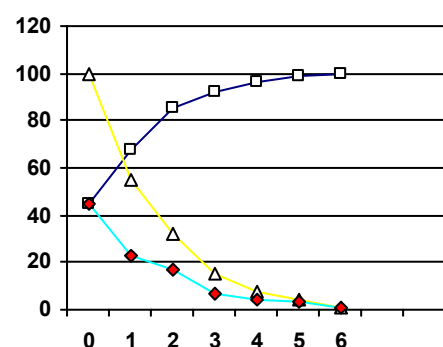
En règle générale :

Valeur du caractère	x_1	x_2	...	x_p	TOTAL
Effectif	n_1	n_2	...	n_p	$\sum_{i=1}^p n_i = N$
Effectif cumulé croissant	n_1	$n_1 + n_2$...	$n_1 + n_2 + \dots + n_p$	
Effectif cumulé décroissant	$n_1 + n_2 + \dots + n_p$	$n_2 + \dots + n_p$...	n_p	
Fréquence	$f_1 = \frac{n_1}{N}$	$f_2 = \frac{n_2}{N}$...	$f_p = \frac{n_p}{N}$	1

II. REPRESENTATIONS GRAPHIQUES**a. Courbe des effectifs (croissants, décroissants) :**

On considère le tableau suivant qui donne le nombre d'enfants dans un effectif de 100 familles :

Nombre d'enfants	0	1	2	3	4	5	6
Effectif	45	23	17	7	4	3	1
ECC	45	68	85	92	96	99	100
ECD	100	55	32	15	8	4	1



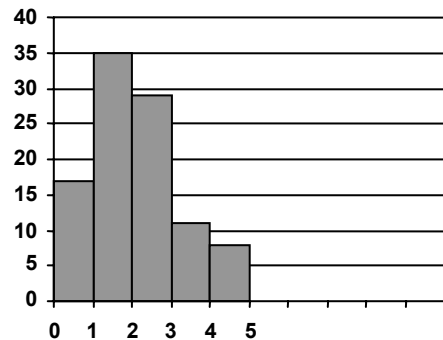
On a représenté cette série sous forme d'une **courbe** →

b. Histogramme :

On considère le tableau suivant qui donne le temps passé quotidiennement devant la télévision :

Temps de TV	0 à 1h	1h à 2h	2h à 3h	3h à 4h	4h à 5h
Fréquence	17	35	29	11	8

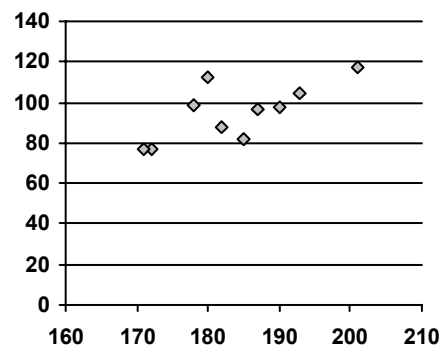
On a représenté cette série sous forme d'un **histogramme** →

**c. Nuage de points :**

On considère le tableau suivant qui donne la taille et le poids de 10 joueurs de rugby :

Numéro	1	2	4	6	8	9	10	12	14	15
Taille (cm)	180	178	201	187	193	172	185	182	171	190
Poids (Kg)	112	99	117	97	105	77	82	88	77	98

On a représenté cette série sous forme d'un **nuage de points** →

**II. MOYENNE****a. Moyenne simple :**

Soit x un caractère qui prend les valeurs x_1, x_2, \dots, x_p , alors la moyenne de cette série statistique est :

$$\bar{x} = \frac{\sum_{i=1}^p x_i}{p}$$

Exemple :

Voici les 8 notes (sur 20) d'un élève ce trimestre : 12 ; 15 ; 12 ; 13 ; 10 ; 19 ; 11 ; 7

$$\text{Alors } \bar{x} = \frac{12 + 15 + 12 + 13 + 10 + 19 + 11 + 7}{8} = \frac{96}{8} = 12$$

b. Moyenne pondérée (coefficientée) :

Soit x un caractère qui prend les valeurs x_1, x_2, \dots, x_p . Les effectifs respectifs de chaque valeur sont n_1, n_2, \dots, n_p .

L'effectif total est donc $\sum_{i=1}^p n_i = N$. Alors la moyenne de cette série statistique est :

$$\bar{x} = \frac{\sum_{i=1}^p n_i x_i}{N}$$

Exemple :

Dans une classe, 7 élèves ont eu 12, 3 élèves ont eu 15, 5 élèves ont eu 9 et 1 seul élève a eu 18.

$$\text{Alors la moyenne de la classe est } \bar{x} = \frac{7 \times 12 + 3 \times 15 + 5 \times 9 + 1 \times 18}{7 + 3 + 5 + 1} = \frac{192}{16} = 12$$

III. MEDIANE, QUARTILES, ETENDUE

Cette série statistique porte sur l'âge des joueurs de l'équipe de France championne d'Europe en 2000 :

ÂGE	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	TOTAL
EFFECTIF	0	1	2	0	1	0	4	0	5	2	1	2	2	0	1	1	22

Ce qui signifie qu'il y a :

- 1 joueur âgé de 21 ans.
- 2 joueurs âgés de 22 ans.
- 4 joueurs âgés de 26 ans.
- Aucun joueur âgé de 33 ans...

a. Médiane :

C'est la valeur (de l'âge) qui se trouve au MILIEU de la série, qui la partage en deux séries d'effectif égal.

Réécrivons tous les ages par ordre croissant :

21 22 22 24 26 26 26 26 28 28 **28** | **28** 28 29 29 30 31 31 32 32 34 35

11 joueurs
Médiane = **28**
11 joueurs

La médiane de cette série statistique est de 28 ans.

Remarques :

- Dans le cas où l'effectif de la série est impair, la « ligne de partage » est située juste sur une valeur : C'est la valeur médiane.
- Dans le cas où l'effectif de la série est pair (dans notre exemple), la « ligne de partage » est située juste entre deux valeurs de la série. Si ces deux valeurs sont différentes, on prend leur moyenne pour valeur médiane.

Définition mathématique :

Soit une série de N données est rangée dans l'ordre croissant

- Si N est impair ($N = 2n + 1$) alors la médiane est la $(n + 1)$ -ème donnée.
- Si N est pair ($N = 2n$) alors la médiane est la moyenne entre les n -ème et $(n + 1)$ -ème données.

b. Quartiles :

C'est les valeurs qui partagent la série en quatre séries d'effectif égal.

En fait, on ne détermine que le 1^{er} et 3^{ème} quartile, puisque le 2^{ème} quartile est la médiane.

Dans l'exemple précédent :

- 1^{er} quartile : 26 ($\frac{1}{4}$ de 22 = 5,5 \rightarrow 6^{ème} valeur)
- 2^{ème} quartile = médiane : 28
- 3^{ème} quartile : 31 ($\frac{3}{4}$ de 22 = 16,5 \rightarrow 17^{ème} valeur)

Définition mathématique :

Soit une série de N données est rangée dans l'ordre croissant

- Le 1^{er} quartile est **le plus petit nombre Q_1** tel que **au moins un quart** des données sont inférieures ou égales à Q_1 .
- Le 3^{ème} quartile est **le plus petit nombre Q_3** tel que **au moins trois quarts** des données sont inférieures ou égales à Q_3 .

c. Caractéristiques de dispersion :

Deux séries de données peuvent avoir des moyennes et médianes très proches, tout en étant constituées des données très différentes. Pour les comparer, on calcule deux **caractéristiques de dispersion** :

L'étendue : c'est la différence entre la valeur la plus haute et la valeur la plus basse de la série.

L'écart interquartile : c'est la différence $Q_3 - Q_1$.