Objectifs visés par l'enseignement des statistiques et probabilités à l'occasion de résolutions de problèmes…

→ **dans le cadre de l'analyse de donnée**s, rendre les élèves capables

- de déterminer et interpréter des résumés d'une série statistique ;
- de réaliser la comparaison de deux séries statistiques à l'aide d'indicateurs de position et de dispersion, ou de la courbe
- des fréquences cumulées ;

→ **dans le cadre de l'échantillonnage**

- faire réfléchir les élèves à la conception et la mise en oeuvre d'une simulation ;
- sensibiliser les élèves à la fluctuation d'échantillonnage, aux notions d'intervalle de fluctuation et d'intervalle de confiance et à l'utilisation qui peut en être faite.

CONTENUS	CAPACITES ATTENDUES	COMMENTAIRES
Statistique descriptive, analyse de données Caractéristiques de position et de dispersion - médiane, quartiles ; - moyenne. - Utiliser un logiciel (par exemple, un tableur) ou une calculatrice pour étudier une série statistique.	- Passer des effectifs aux fréquences, calculer les caractéristiques d'une série définie par effectifs ou fréquences. - Calculer des effectifs cumulés, des fréquences cumulées. - Représenter une série statistique graphiquement (nuage de points, histogramme, courbe des fréquences cumulées).	L'objectif est de faire réfléchir les élèves sur des données réelles, riches et variées (issues, par exemple, d'un fichier mis à disposition par l'INSEE), synthétiser l'information et proposer des représentations pertinentes.
Échantillonnage Notion d'échantillon. Intervalle de fluctuation d'une fréquence au seuil de 95%*. Réalisation d'une simulation.	- Concevoir, mettre en oeuvre et exploiter des simulations de situations concrètes à l'aide du tableur ou d'une calculatrice. - Exploiter et faire une analyse critique d'un résultat d'échantillonnage.	Un échantillon de taille n est constitué des résultats de n répétitions indépendantes de la même expérience. À l'occasion de la mise en place d'une simulation, on peut : - utiliser les fonctions logiques d'un tableur ou d'une calculatrice, - mettre en place des instructions conditionnelles dans un algorithme. L'objectif est d'amener les élèves à un questionnement lors des activités suivantes : - l'estimation d'une proportion inconnue à partir d'un échantillon ; - la prise de décision à partir d'un échantillon.

***** L'intervalle de fluctuation au seuil de 95%, relatif aux échantillons de taille n, est l'intervalle centré autour de p, proportion du caractère dans la population, où se situe, avec une probabilité égale à 0, 95, la fréquence observée dans un échantillon de taille n. Cet intervalle peut être obtenu, de façon approchée, par simulation.
Le professeur peut indiquer aux élèves le résultat suivant, utilisable dans la pratique pour des échantillons de taille n > 25 et des proportions p du caractère comprises entre 0, 2 et 0, 8 : si f désigne la fréquence du caractère dans l'échantillon, f appartient à l'intervalle $[p - 1/\sqrt{n}\ ;\ p + 1/\sqrt{n}]$ avec une probabilité d'au moins 0, 95. Le professeur peut faire percevoir expérimentalement la validité de cette propriété mais **elle n'est pas exigibl**e.

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 :

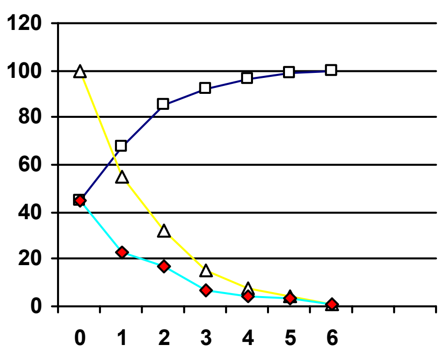
Valeur du caractère	x_1	x_2	…	x_p	TOTAL
Effectif	n_1	n_2	…	n_p	$\sum_{i=1}^{p} ni = N$
Effectif cumulé croissant	n_1	$n_1 + n_2$	…	$n_1 + n_2 + \ldots + n_p$	
Effectif cumulé décroissant	$n_1 + n_2 + \ldots + n_p$	$n_2 + \ldots + n_p$	…	n_p	
Fréquence	$f_1 = \frac{n_1}{N}$	$f_2 = \frac{n_2}{N}$	…	$f_p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



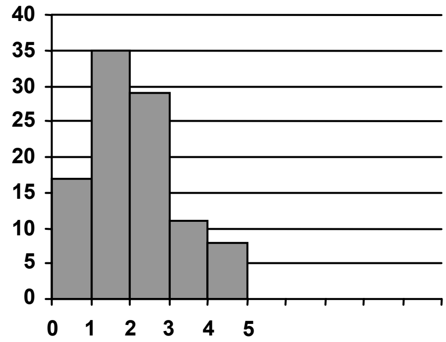
On a représenté cette série sous forme d'une **courbe** →

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** →

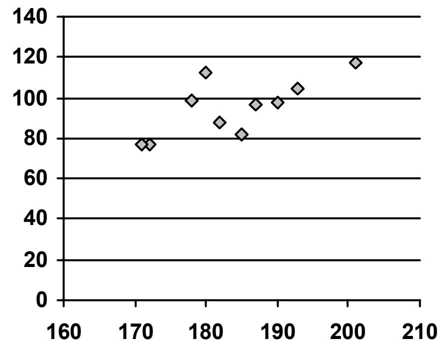


c. Nuage de points :

On considère le tableau suivant quidonne 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 :

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

Exemple :

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

Alors $\overline{x} = \dfrac{13 + 11 + 12 + 13 + 10 + 19 + 11 + 7}{8} = \dfrac{96}{8} = 12$

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 $\sum_{i=1}^{p} ni = N$. Alors la moyenne de cette série statistique est :

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

Exemple :

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

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

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 1er et 3ème quartile, puisque le 2ème quartile est la médiane.

Dans l'exemple précédent :

1er quartile : 26 (¼ de 22 = 5,5 → 6ème valeur)
2ème quartile = médiane : 28
3ème quartile : 31 (¾ de 22 = 16,5 → 17ème valeur)

Définition mathématique :

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

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