

## 2<sup>nde</sup> – ChXVI – Statistiques et échantillonnage

### I- Vocabulaire et notation

Une étude statistique commence par un recueil de données concernant un caractère sur les individus d'une population.

Le caractère étudié prend un certain nombre de valeurs.

Si ces valeurs sont des nombres, il s'agit d'un caractère quantitatif.

Sinon, il s'agit d'un caractère qualitatif.

Dans le cas d'un caractère quantitatif, notons  $x_1, x_2, \dots, x_p$  les  $p$  valeurs que peut prendre le caractère.

On note  $n_1, n_2 \dots n_p$  les effectifs correspondant à chacune des valeurs  $x_1; x_2 \dots x_p$ .

L'effectif total est le nombre d'individu, qu'on notera  $N$ .

$N = n_1 + n_2 + \dots + n_p$  que l'on note

$$N = \sum_{i=1}^p n_i \quad (\text{la somme des } n \text{ indice } i \text{ pour } i \text{ allant de } 1 \text{ à } p)$$

La moyenne  $\bar{x}$  est la somme des  $n_i x_i$  divisée par

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

La médiane est la valeur centrale quand toutes les valeurs sont rangées par ordre croissant ou décroissant.

Exemple : pour 5 valeurs, la médiane est la troisième, pour 11, la 6<sup>ème</sup>.

S'il y a un nombre pair de valeurs, on fait la moyenne des deux valeurs centrales.

La fréquence d'une valeur est le quotient  $f_i = \frac{n_i}{N}$

$$= \frac{\text{nombre d'individus ayant le caractère } x_i}{\text{nombre total d'individus}}$$

On a  $0 \leq f_i \leq 1$  Une fréquence est un nombre compris entre 0 et 1.

(comme une probabilité)

On peut demander des fréquences sous forme décimale, fractionnaire, ou en pourcentage.

#### Exemple 1 :

On étudie la taille en cm d'un groupe de 5 individus dont les tailles sont 155 ; 164 ; 170 ; 170 et 185. C'est un caractère .....

#### Valeurs prises par le caractère :

$x_1 = \dots ; x_2 = \dots ; x_3 = \dots$  et  $x_4 = \dots$

Ici :  $p = 4$  car le caractère peut prendre 4 valeurs différentes.

#### Effectifs correspondant à chacune des valeurs :

$n_1 = \dots$  car ..... individu mesure  $x_1 = \dots$

$n_2 = \dots$  car ..... individu mesure  $x_2 = \dots$

$n_3 = \dots$  car ..... individus mesurent  $x_3 = \dots$

$n_4 = \dots$  car ..... individu mesure  $x_4 = \dots$

$N = \dots$

#### Calcul de la moyenne :

$$\bar{x} = \frac{\sum_{i=1}^p n_i x_i}{N} = \frac{n_1 x_1 + n_2 x_2 + n_3 x_3 + n_4 x_4}{N}$$

= .....

= .....

#### Evaluation de la médiane :

La médiane est la valeur « centrale »

On range les valeurs par ordre croissant au

décroissant : .....

Celle du milieu (= la 3<sup>ème</sup> sur 5) est la médiane :

.....

#### Le mode :

Le mode est la valeur la plus représentée (= dont l'effectif est le plus grand) Ici, c'est .....

#### Calcul des fréquences :

Valeurs du caractère	Effectifs	Fréquence (décimale)	Fréquence (en %)
$x_1 =$	$n_1 =$	$f_1 =$	$f_1 =$
$x_2 =$	$n_2 =$	$f_2 =$	$f_2 =$
$x_3 =$	$n_3 =$	$f_3 =$	$f_3 =$
$x_4 =$	$n_4 =$	$f_4 =$	$f_4 =$
Total :	$N =$	1	100%

Avec des regroupements par classes :

Lorsque les données sont en très grand nombre, on peut les **regrouper par classes** : ce sont des intervalles [ a ; b [ de même **amplitude** b – a.

Le **centre d'une classe** [ a ; b [ est le nombre  $\frac{a + b}{2}$

*Exemple 2* : On a réalisé un test de course à pied pour le 35 élèves d'une classe de seconde durant 20 minutes. Voici les distances (en km) parcourues par les élèves durant ce laps de temps :  
 3,2 – 4,1 – 2,2 – 5,3 – 3,4 – 3,6 – 4,2 – 4,7 – 2,5 – 3,4 – 4 – 4,2 – 5,2 – 2,6 – 3 – 4,8 – 5 – 4,3 – 3 – 3,5 – 3,7 – 3,8 – 3,8 – 4,4 – 2,8 – 4,5 – 4,2 – 4,4 – 3,8 – 3,7 – 3,8 – 4,8 – 4 – 3,8 – 2,9

On regroupe ces valeurs en classes d'amplitude 0,5, la première classe étant [ 2 ; 2,5 [

Classe	Centre $c_i$ de la classe	Effectif $n_i$	$n_i \times c_i$
[ 2 ; 2,5 [			
Total :			

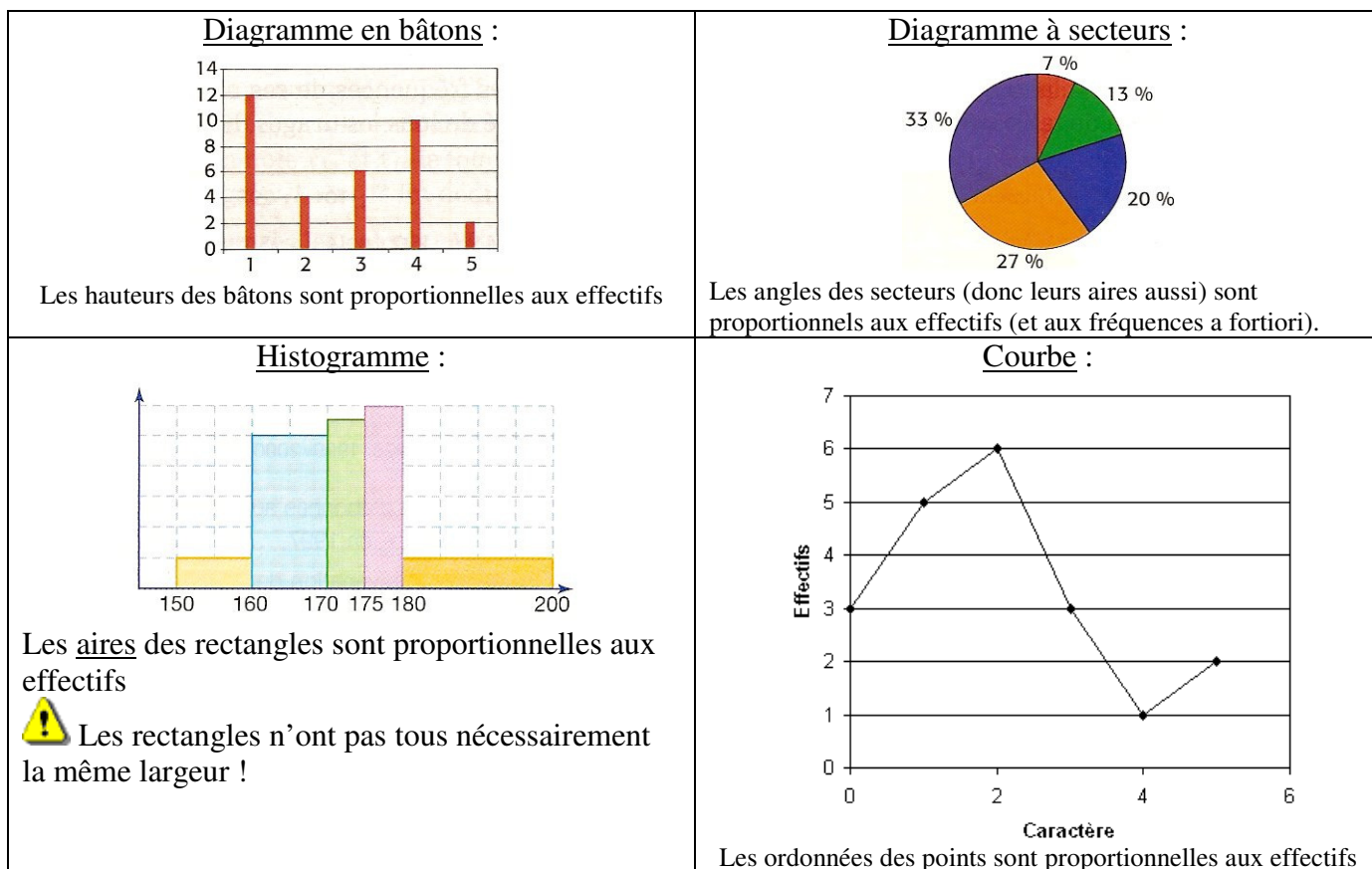
$\text{Moyenne} = \frac{\sum_{i=1}^5 n_i \times c_i}{N} = \frac{\dots\dots\dots}{\dots\dots\dots} \approx \dots\dots\dots$	(la moyenne se calcule avec le centre des classes et les effectifs correspondants)
<b>Médiane :</b>	Ici, on prend pour médiane le centre de la classe où se situe l'individu « central » (le 18 <sup>ème</sup> , sur 35 individus)
<b>Classe modale :</b>	C'est la classe dont l'effectif est le plus grand.

*Note* : une **moyenne élaguée** est une moyenne obtenue en ôtant à la série statistique certaines valeurs, jugées aberrantes.

Propriétés de linéarité de la moyenne :

- 1- Si on multiplie toutes les valeurs  $x_i$  d'une série statistique par un même nombre a, La moyenne est aussi multipliée par a.**
- 2- Si on ajoute un même nombre b à toutes les valeurs  $x_i$  d'une série statistique, la nouvelle moyenne vaut  $\bar{x} + b$ .**
- 3- Si la population est séparée en deux sous-groupes disjoints, d'effectifs p et q, et si les moyennes de ces sous-groupes sont respectivement  $m_1$  et  $m_2$ , la moyenne du groupe est  $m = \frac{p \times m_1 + q \times m_2}{p + q}$**

## II- Les représentations graphiques.



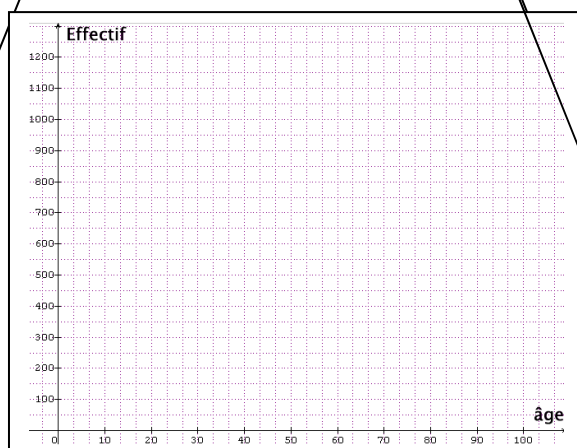
## III- Les effectifs cumulés et fréquences cumulées.

Exemple : Une enquête a été réalisée sur un échantillon représentatif de 1200 personnes regardant le journal télévisé de 20 heures, pour connaître leur âge. En voici les résultats :

Age	Effectif	Effectif cumulé croissant	Effectif cumulé décroissant
[ 0 ; 20 [	120		1200
[ 20 ; 40 [	300		
[ 40 ; 60 [	420		
[ 60 ; 80 [	270		
[ 80 ; 100 [	90	1200	
Total :	1200	1200	1200

Nombre total de personnes de moins de 40 ans

Nombre total de personnes de moins de 80 ans



Nombre total de personnes de 20 ans et plus

Nombre total de personnes de 60 ans et plus

Le graphique des effectifs cumulés croissants et celui des effectifs cumulés décroissants se coupent à la médiane.

Note : Pour les fréquences cumulées, on remplace les effectifs par les fréquences.

#### IV – Fluctuation d'échantillonnage

Une **expérience aléatoire** est une expérience dont on ne peut pas prédire le résultat.

**Exemples** : Lancer une pièce de monnaie et s'intéresser à la face visible. Lancer un dé et s'intéresser au résultat de la face supérieure. Tirer une boule dans une urne contenant des boules de diverses couleurs et noter la couleur de la boule tirée. Tirer une carte dans un jeu de cartes et noter la carte obtenue.

**Définition** : **un échantillon de taille n** est la série statistique formée des n résultats obtenus lorsqu'on répète n fois l'expérience.

**Application** : A l'aide des touches Ran# ou Rnd (« random ») et Int ou Ent (partie entière) de la calculatrice, fournir un échantillon de taille 30 en simulant un lancer de dé à six faces, dont on noterait le nombre de points (1 à 6) de la face supérieure :

**Ran#** fournit un nombre au hasard sans l'intervalle [ 0 ; 1 [  
**Int** donne la **partie entière** du nombre (= l'entier qui le précède)

Pour obtenir au hasard un entier entre 1 et 6, la formule est :

$$\boxed{\text{Int} ( \text{Ran\#} \times 6 ) + 1} \text{ ou } \boxed{\text{Int} ( \text{Ran\#} \times 6 + 1 )}$$

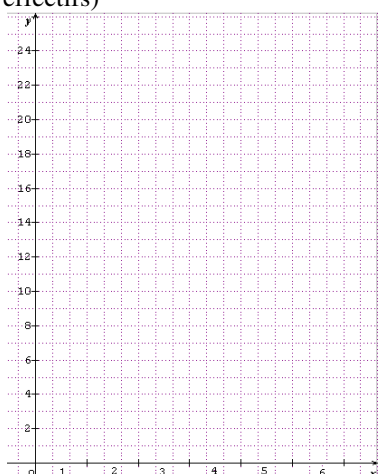
En effet : Ran# donne un nombre décimal dans [0 ; 1[  $0 \leq x < 1$   
 Ran#×6 donne un nombre décimal dans [0 ; 6[  $0 \leq 6x < 6$   
 Ran#×6 + 1 donne un nombre décimal dans [1 ; 7[  $1 \leq 6x + 1 < 7$   
 Int (Ran#×6 + 1) donne donc un entier entre 1 inclus et 7 exclu, soit entre 1 et 6.

Indiquer vos résultats :


La distribution des fréquences associée à un échantillon est la liste des fréquences des issues de l'échantillon :

issues	1	2	3	4	5	6	Total
Effectifs							30
Fréquences (arrondies à 10 <sup>-2</sup> )							1
Secteur en °							360

Représenter vos résultats sur un histogramme :  
 (indiquer les effectifs)



Un diagramme circulaire :  
 (indiquer les fréquences en pourcentages)

