

## Cours 1 : Définition

### I. Définition

- Population : ensemble d'éléments (individus) faisant l'objet d'une étude statistique.  
*Ex* : - ensemble d'élèves d'un amphi (on étudie les notes, l'âge, la taille, le sexe...)  
 - ensemble des pièces fabriquées par une machine (on étudie la taille, le poids...)
- Echantillon : partie d'une population  
*Ex* : - le groupe A est un échantillon de la population de l'amphi.  
 - 100 pièces tirées au hasard est un échantillon
- Unité statistique (élément, individu) : C'est un élément de la population dont on veut étudier un certain caractère
- Caractère : peut être :  
 - quantitatif : âge, note, taille, poids  
 - qualificatif : sexe, couleurs, qualité, ...
- Variable statistique : Ce sont les valeurs prises par un caractère quantitatif :  
 - discrète : prend une valeur isolées (nombre d'individus = N)  
 - continue : prend toutes les valeurs d'un intervalle (la longueur d'une pièce usinées : valeur  $\in ]5,6[cm$ )
- Série statistique : C'est l'ensemble des valeurs prises par la variable statistique sur l'ensemble de la population.

### II. Série statistiques d'un caractère quantitatif discret

- On considère un échantillon de taille n.  
 On considère les valeurs possibles de la variable (du caractère x) :  $x_1, x_2, \dots, x_p$
- Série statistique : ensemble des n valeurs  $x_i$  (sur un dé :  $x_i \in \{1,2,3,4,5,6\}$ )  
 exemple de série  $\{3,1,2,6,1,3,5,\dots\}$  avec n = 100 valeurs.
- Effectif total : n
- Effectif partiel de la valeur x : C'est le nombre de fois n que la variable x prend la valeur  $x_i$ . On appelle cela aussi la fréquence absolue de  $x_i$ .  

$$n_1 + n_2 + \dots + n_p = n \text{ (effectif total)}$$
- Fréquence relative de x :  $f_i = \frac{n_i}{n} \Rightarrow f_1 + f_2 + \dots + f_p = 1$

## Statistique

- Etendue de la série : écart entre la plus grande et la plus faible des valeurs de  $x$ .

Exemple :

100 familles de 4 enfants et on étudie le caractère = nombre de garçons.  
On a la série de statistique : 3,1,3,2,0,4,1,3... avec 100 valeurs

- Effectif total : 100 cas, échantillon de 100 familles
- Caractère : nombre de garçon variable stat quantitatif et discrète  $x_i \in \{0,1,2,3,4\}$
- Etendue de la série :  $4 - 0 = 4$
- On compte les valeurs trouvées :

$x_i$	0	1	2	3	4
$n_i$	7	20	43	25	5

- Effectif partiel de  $x=2$  : 43 familles ont 2 garçons.
- Fréquence relative de  $x_2$  :  $f_2 = \frac{43}{100} = 43\%$

### III. Séries statistiques d'un caractère quantitatif continu

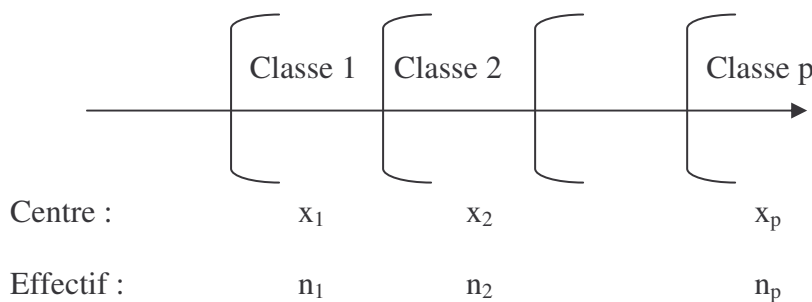
Exemple de poids de nourrisson :

classe	1	2	3	4	5	6	7	8
limites	2,2-2,5	2,5-2,8	2,8-3,1	3,1-3,4	3,4-3,7	3,7-4,0	4,0-4,3	4,3-4,6
centre	2,35	2,65	2,95	3,25	3,55	3,85	4,15	4,45
effectif	5	11	24	40	42	20	13	6
fréq. relat. %	3,1	6,8	14,9	24,8	26,1	12,4	8,1	3,7

→ Ici le nombre  $p$  de valeurs de  $x_i$  prises par la variable est infini. (parfois  $p$  est fini mais très grand ( $p=100000$ ) donc on considère quand même que le caractère est continu).

→ On constitue des classes en divisant l'étendue de la série en un certain nombre d'intervalles.

Remarque : L'orientation des crochets est totalement arbitraire et communément admise.



## Statistique

→ Exemple des nourrissons : cf feuille photocopiée.

Les pesées faites à 10 grammes près et donnent des poids entre 2,240 et 4,490 kg.

- 8 classes d'amplitudes (étendue) : 0,3 kg

Attention : borne inférieure de la classe 1 <  $x_{\min}$  et borne supérieure de la classe 8 >  $x_{\max}$  mais pas forcément obligatoire.

- Fréquence relative :  $\frac{n_i}{n} = \frac{n_i}{n_1 + n_2 + \dots + n_p} = 161$

- Remarque :  $\sum_{i=0}^{i=p} f_i = 1$  ; peut varier de 0,01 selon les arrondis.

### IV. Série statistiques d'un caractère qualitatif

- On groupe les résultats en autant de classes que de modalités du caractère.

*Exemple* : étude sur les couleurs de fleurs : 3 couleurs possibles ( 3 modalités) : rouge, jaune, bleu : 3 classes différentes.

➡ Variable ordinale : si les modalités peuvent être ordonnées :

- Taille vestimentaire : M, L , XL , ...

- Préférence + ou – grande.

➡ - Variable nominale : on ne peut ordonner les modalités :

- Sexe : M ou F, couleurs, professions

➡ Variable dichotomique : Si seulement 2 modalités : Exemple : Oui / Non

### V. Représentation graphique des séries statistiques

#### 1) *Caractère discret*

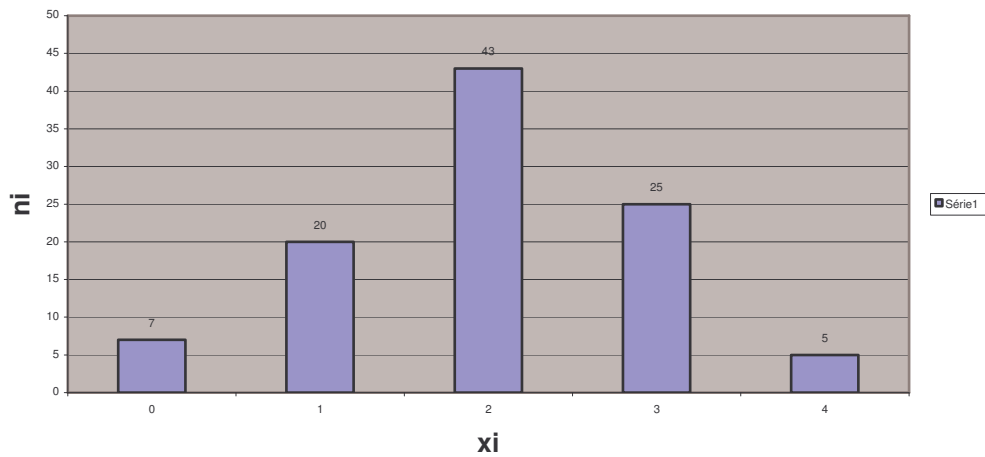
- diagramme en bâtons :

*Exemple* : famille de 4 enfants

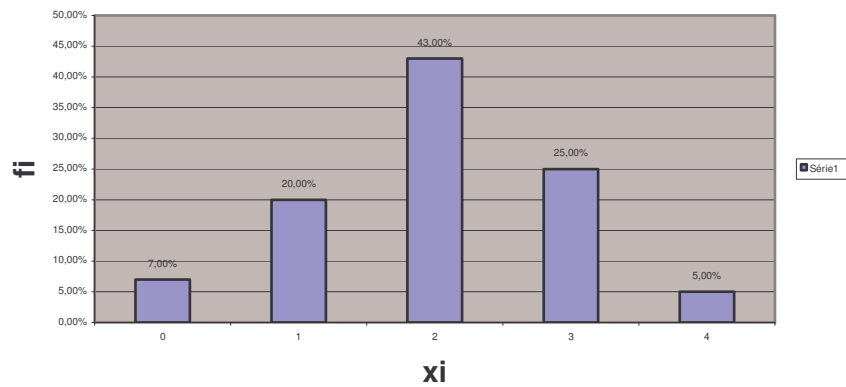
$n_i$  en fonction de  $x_i$

$f_i$  en fonction de  $x_i$

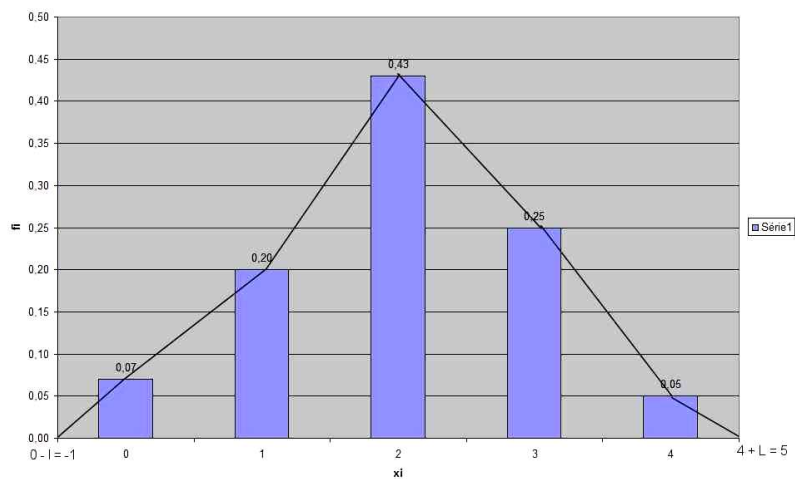
## Statistique



ou



- Polynôme des effectifs (fréquence absolue) : on joint les extrémités des sommets des bâtons.



*Remarque* : Si les valeurs de  $x_i$  sont équidistantes on prolonge le polygone à gauche et à droite.

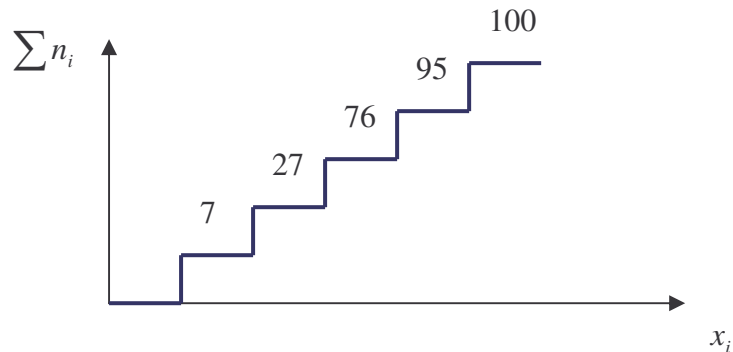
- Polygone des fréquences absolues: On remplace  $n_i$  par  $f_i$  sur l'axe des ordonnées.

## Statistique

- Diagramme cumulatif

- Effectif cumulé jusqu'à la  $i^{\text{ème}}$  valeur ( $x_i$ ) de caractère  $n_1+n_2+\dots+n_i$
- Fréquence cumulative cumulée :  
On va tracer  $f_1+f_2+\dots+f_i$

*Diagramme cumulatif des effectifs :*



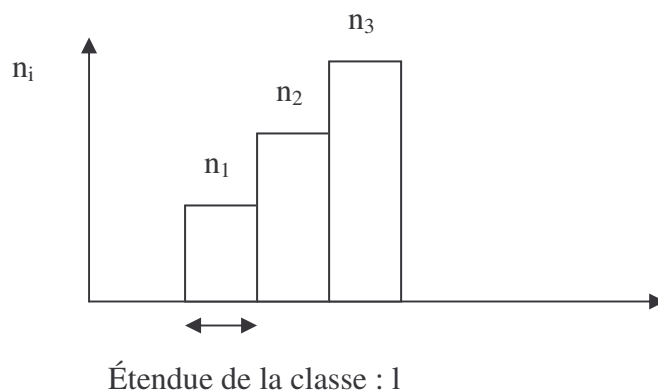
Si on veut obtenir le diagramme cumulatif des fréquences relatives, il suffit de remplacer  $\sum n_i$  par  $\sum f_i$  et, de façon générale :

$$\left. \begin{aligned}
 F(x) &= 0 \text{ si } x < x_i \\
 &= \sum_{j=1}^i f_j \text{ si } x_i \leq x < x_i + i \\
 &= 1 \text{ si } x \geq x_p
 \end{aligned} \right\} \text{ Fonction définie sur } \mathbb{R}$$

### 2) caractère continu

- Diagramme en bâton impossible, on a trop de valeurs.

a) *Histogramme :*



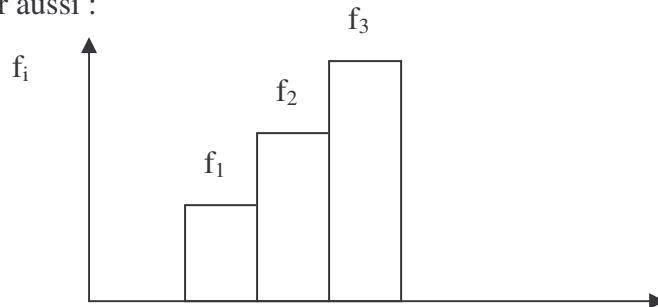
**Remarque :** Aire d'un rectangle :  $l \cdot n_i$

## Statistique

aire totale :  $n_1.l+n_2.l+\dots+n_p.l = \left( l \cdot \sum_{i=1}^p n_i \right)$

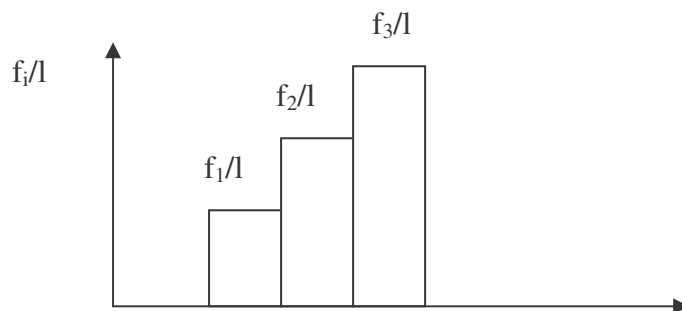
→ Aire totale :  $l.n$

On peut tracer aussi :



Ici, l'aire totale =  $L$ .

Enfin, on peut aussi tracer :



Ici, l'aire totale = 1.

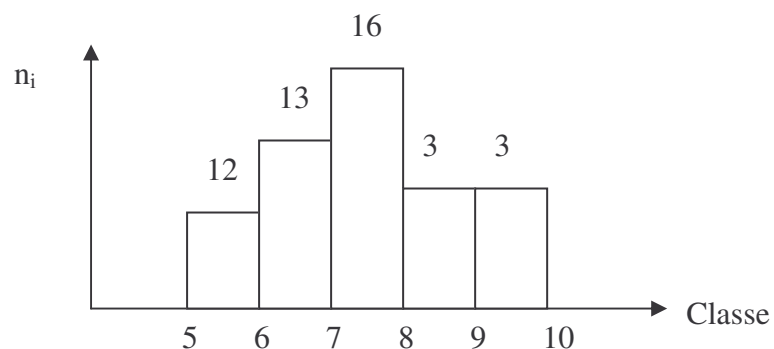
b) Classes inégales :

Sur un exemple :

classe	5-6	6-7	7-8	8-10
$n_i$	12	13	16	6

On veut garder la propriété sur les aires :

Aire totale =  $47 \cdot 1 = 47$

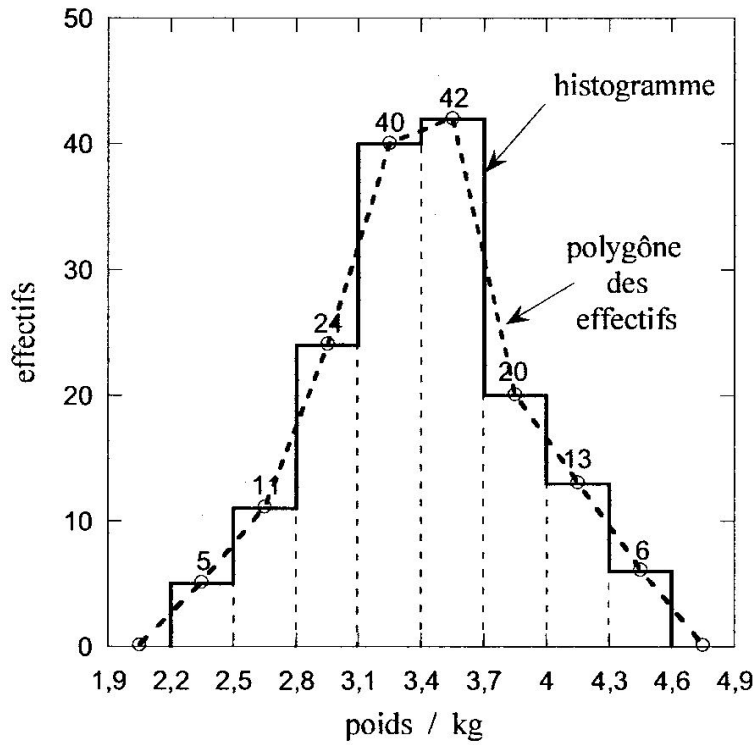


## Statistique

On divise l'effectif de la classe d'étendue 2 par 2.

c) *Polynôme des effectifs (fréquences absolues)*

Ligne brisée joignant es milieux des sommets des rectangles.

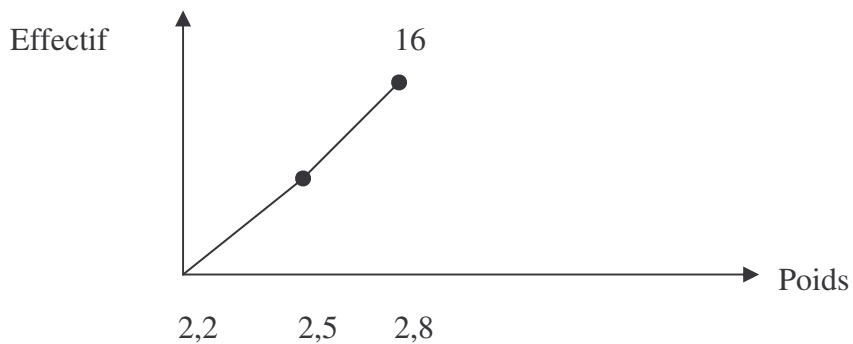


→ Ligne brisée joignant les points  $(x_i, n_i)$  où  $x_i$  est le centre

d) *Polynôme des effectifs cumulés*

Exemple de poids de nourrisson :

classe	1	2	3	4	5	6	7	8
limites	2,2-2,5	2,5-2,8	2,8-3,1	3,1-3,4	3,4-3,7	3,7-4,0	4,0-4,3	4,3-4,6
centre	2,35	2,65	2,95	3,25	3,55	3,85	4,15	4,45
effectif	5	11	24	40	42	20	13	6
fréq. relat. %	3,1	6,8	14,9	24,8	26,1	12,4	8,1	3,7



## Statistique

*Remarque :*

On n'a plus un diagramme en escalier comme dans le cas discret. Dans chaque classe, il y a plusieurs individus qu'on suppose équirépartis dans la classe : l'information sur la valeur cumulée se trouve à droite de la classe.

e) *Polygone des fréquences relatives cumulées*

On remplace dans le point d'avant  $n_i$  par  $f_i = \frac{n_i}{n}$ .