

2.2.2) Cas d'une v.s. continue:

Dans le cas d'une v.s. continue, pour établir le tableau statistique, il faut effectuer au préalable une répartition

en classes des données. * Une classe est définie par la donnée de ses extrémités inférieure et supérieure : par convention $[e_{i-1}, e_i[$.

Cela nécessite de définir le nombre de classes et l'amplitude associée à chaque classe.

* La $i^{\text{ème}}$ classe ^{par exemple} est donnée par $[e_{i-1}, e_i[$.

où e_{i-1} : l'extrémité inférieure.

e_i : l'extrémité supérieure

$a_i = (e_i - e_{i-1})$: amplitude de la $i^{\text{ème}}$ classe.

• le centre de cette classe est $x_i = \frac{e_i + e_{i-1}}{2}$

Règles : 1) En général, on choisit des classes de même amplitude.

2) Le choix du nombre de classes et de leurs amplitudes se fait en fait de l'effectif total (n) de la population.

3) Toute diminution du nombre de classes et toute augmentation ~~de la~~ de l'amplitude, conduit à une perte d'information.

4) Règle de Sturges : $k = \text{nombre de classes} \approx 1 + (3,3 \log n)$.

et l'amplitude est $a = \frac{\max x_i - \min x_i}{\text{nombre de classes}} = \frac{\text{étendue de la série}}{\text{nombre de classes}}$

• Le tableau statistique d'une v.s. continue se présente sous la forme suivante :

classes	centre x_i	n_i	b_i
$[e_0, e_1[$	x_1	n_1	b_1
$[e_1, e_2[$	x_2	n_2	b_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$[e_{k-1}, e_k[$	x_k	n_k	b_k

• Règle : Les modalités d'une v.s. continues sont les centres des classes $x_i = \frac{e_i + e_{i-1}}{2}$

- Pour représenter graphiquement une v.s. continue on utilise l'histogramme : qui constitue une généralisation du diagramme en bâtons à la notion de classes.

⑩ Deux cas sont possibles :

a/ Séries statistiques à classes égales (même amplitude)

• Chaque classe est représentée par un rectangle dont la base est l'amplitude et la hauteur est proportionnelle à la fréquence (absolue ou relative).

Exemple 3: On considère la répartition des ouvriers d'une entreprise selon le salaire mensuel (en m DA). Les résultats sont donnés dans le tableau suivant:

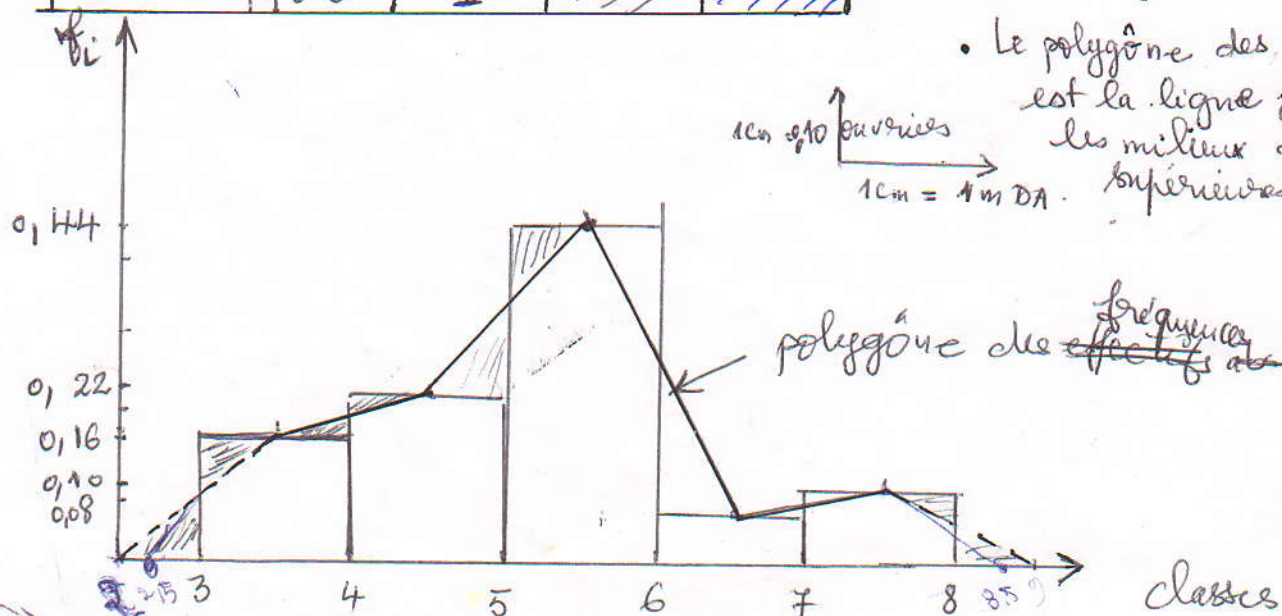
Classes Salaire (en m DA)	n_i	f_i	$F_i \nearrow$	$F_i \searrow$
$[3, 4[$	16	0,16	0,16 = $F(4)$	1 = $F(3)$
$[4, 5[$	22	0,22	0,38 = $F(5)$	0,84 = $F(4)$
$[5, 6[$	44	0,44	0,82 = $F(6)$	0,62 = $F(5)$
$[6, 7[$	08	0,08	0,90 = $F(7)$	0,18 = $F(6)$
$[7, 8[$	10	0,10	1 = $F(8)$	0,10 = $F(7)$
Total	100	1		

Remarques:

- L'aire de chaque rectangle est égale à $f_i \times a$ et

d'aire de tout l'histogramme est $\sum f_i \cdot a = a \sum f_i = a$ (unité de surface)

- Le polygone des fréquences est la ligne joignant les milieux de côtés supérieurs des rectangles



Histogramme du salaire mensuel des ouvriers d'entreprise

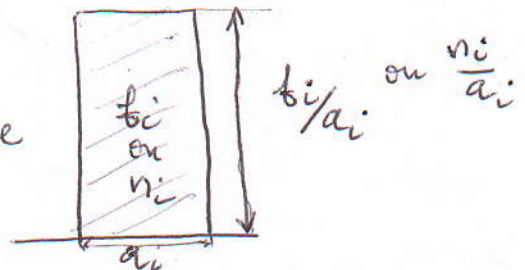
b/ Séries statistiques à classes inégales (amplitudes différentes)

Chaque classe est représentée par un rectangle dont la base est égale à l'amplitude et la hauteur égale à la fréquence divisée par l'amplitude (fréquences corrigées)

• Rq. 1) Dans ce cas, l'aire de chaque rectangle

$$\text{est : } a_i \times \frac{n_i}{a_i} = n_i$$

est proportionnelle à la fréquence ou l'effectif de la classe.



2°) En pratique, pour simplifier les calculs, on choisit l'amplitude unité $a = \text{pgcd}(a_i)$ ou bien "a" l'amplitude la plus répétée.

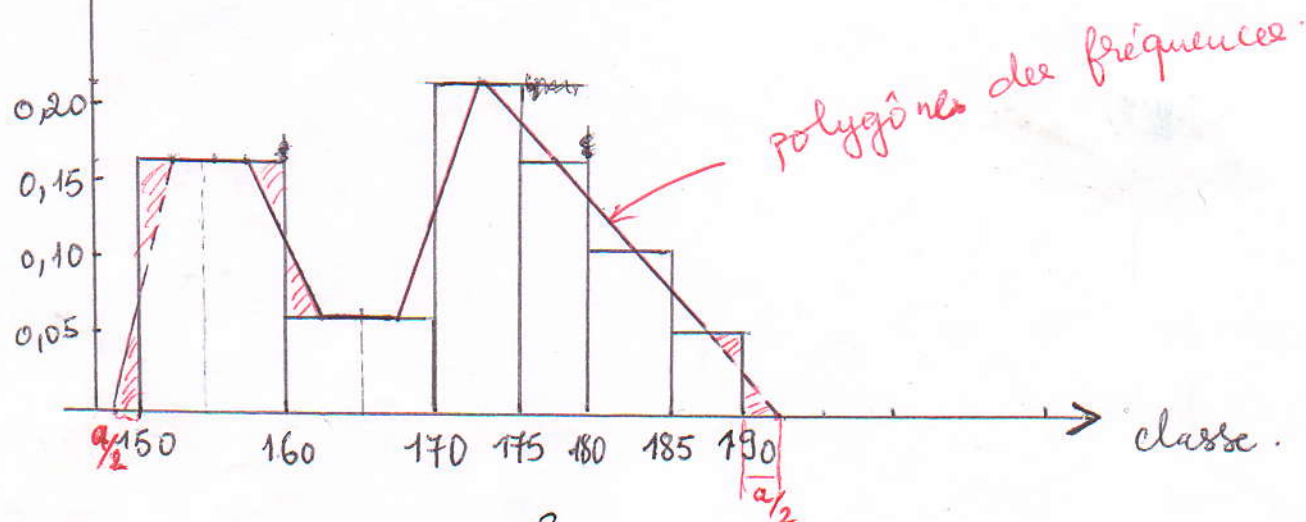
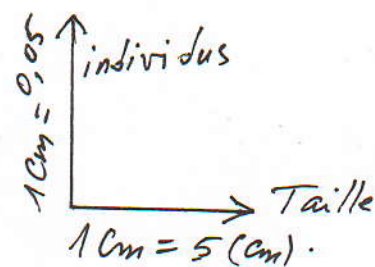
Exemple 4: La répartition d'un gpe d'individus par taille (cm) est donnée par le tableau suivant:

Taille (cm)	n_i	b_i	a_i	$b_i^c = \frac{b_i \times a}{a_i}$ $\frac{b_i}{a_i} \times a$
[150, 160[	33 33	0,33	10	0,165
[160, 170[	14 14	0,14	10	0,07
[170, 175[	21	0,21	05	0,21
[175, 180[	16	0,16	05	0,16
[180, 185[	11	0,11	05	0,11
[185, 190[	05	0,05	05	0,05
Total:	100	1	////	////

$a = 5 \text{ cm}$
 $= \text{pgcd}(10, 5)$
 a est aussi l'amplitude la plus répétée.

La représentation graphique de cette série par l'histogramme est la suivante:

$$f_{ic} = \frac{f_i \times a}{a_i}$$



L'aire de tout l'histogramme est $\sum_i a f_i = a$ (unité de surface)
i.e. surface = 5 cm^2 .

Courbe Cumulative:

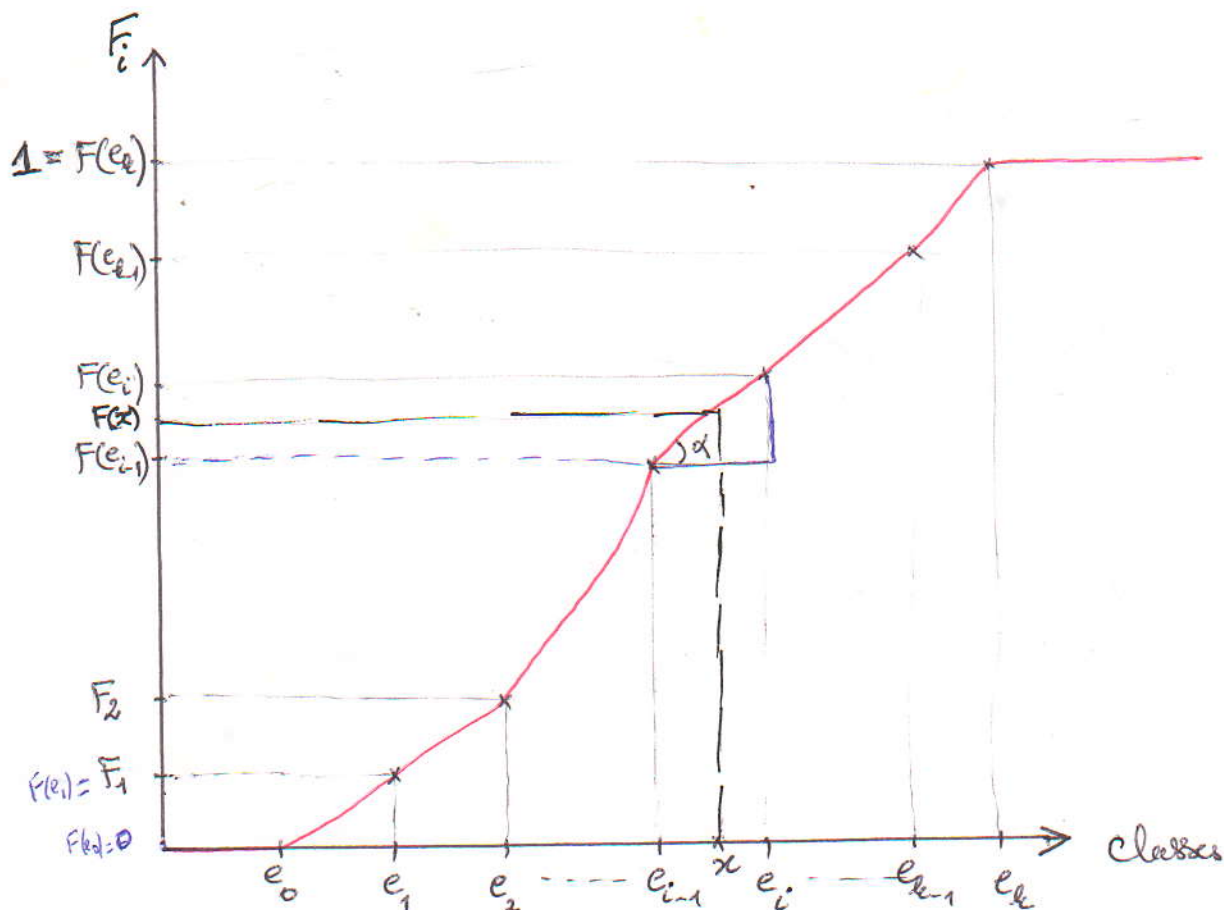
La fréquence relative cumulée F_i est la somme des fréquences relatives de chaque classe correspondant aux valeurs de la v.s. continue inférieure à l'extrémité supérieure e_i de la i -ème classe: $[e_{i-1}, e_i[: F_i = f_1 + f_2 + \dots + f_i = \sum_{j=1}^i f_j$.

De manière analogue: $N_i = \sum_{j=1}^i n_j$.

Soit X une v.s. continue et $[e_0, e_1[$, $[e_1, e_2[$, ..., $[e_{k-1}, e_k[$ ses classes.

On associe à chaque extrémité supérieure sa fréq. cumulée: $F_1 = F(e_1)$, $F_2 = F(e_2)$ --- $F_i = F(e_i)$, $F(e_k) = 1$
 $F(e_0) = 0$.

La courbe cumulative est représentée sous la forme suivante



soit $x \in [e_{i-1}, e_i[$, Déterminons la valeur de la fct de répartition empirique au pt x ie $F(x)$.

$$\text{on a } \tan \alpha = \frac{F(e_i) - F(e_{i-1})}{e_i - e_{i-1}} = \frac{F(x) - F(e_{i-1})}{x - e_{i-1}}$$

$$= \frac{b_i}{a_i} = \frac{F(x) - F(e_{i-1})}{x - e_{i-1}} \Rightarrow (x - e_{i-1}) \frac{b_i}{a_i} = F(x) - F(e_{i-1})$$

On aura :

$$F(x) = F(e_{i-1}) + \frac{x - e_{i-1}}{a_i} \times b_i$$

D'où la fct de répartition empirique d'une v.s. continue de classes $[e_0, e_1[, \dots, [e_{k-1}, e_k[$

est $F: \begin{matrix} \mathbb{R} & \longrightarrow & [0,1] \\ x & \longmapsto & F(x) \end{matrix}$

$$F(x) = \begin{cases} 0 & \text{si } x \leq e_0 \\ F(e_{i-1}) + \frac{x - e_{i-1}}{a_i} f_i & \text{si } x \in [e_{i-1}, e_i[\\ 1 & \text{si } x \geq e_k \end{cases}$$

Exemple 5: Reprenons l'exemple 3: répartition des ouvriers selon le salaire mensuel (en M.DA).

Classes	f_i	$F_i \uparrow$ $F_i = F(x)$	$F_i \downarrow$
$[3, 4[$	0,16	0,16 = $F(3)$	1 = $F(3)$
$[4, 5[$	0,22	0,38 = $F(4)$	0,84 = $F(4)$
$[5, 6[$	0,44	0,82 = $F(5)$	0,62 = $F(5)$
$[6, 7[$	0,08	0,90 = $F(6)$	0,18 = $F(6)$
$[7, 8[$	0,10	1 = $F(7)$	0,10 = $F(7)$
TOTAL	1	///	///

