

## COURS 02 : Suite du chapitre 1 - Séries statistiques à une seule dimension

### II. Représentation d'une série statistique à une seule dimension

#### 1. La variable statistique quantitative

##### 1.1 Définition de la variable statistique

Une variable statistique (ou caractère statistique) est ce qui est observé ou mesuré sur les individus d'une population statistique. On peut la définir mathématiquement comme une application  $X$  de  $\Omega$  dans  $\mathbb{R}$  (si  $X$  est continue)  $\mathbb{N}$  (si  $X$  est discrète).

$$\begin{aligned} X : \Omega &\rightarrow \mathbb{R} \text{ (ou } \mathbb{N}) \\ w &\rightarrow X(w) \end{aligned}$$

$X(w)$  représente les modalités de la variable statistique étudiée et ces modalités sont les valeurs de la variable la statistique notées :  $x_1, x_2, \dots$

La variable statistique est notée  $X, Y, Z, \dots$  et ces valeurs correspondantes  $x_1, x_2, \dots, y_1, y_2, \dots, z_1, z_2, \dots$

- Si la variable statistique est quantitative continue, les valeurs de  $X(w) \in \mathbb{R}$  ;
- Si la variable statistique est quantitative discrète, les valeurs de  $X(w) \in \mathbb{N}$  ;
- Si la variable statistique est qualitative, on peut coder  $X(w)$  en  $0, 1, 2, \dots$  mais les opérations mathématiques sur ces valeurs n'ont aucun sens.

**Exemple :** On tire un échantillon de 20 ménages et on compte le nombre d'enfants dans chaque ménage, on obtient les résultats suivants : 0 1 3 1 4 5 5 5 6 7 2 1 4 4 4 5 2 1 1 2

Population statistique = 20 ménages

$$\Omega = \{w_1, w_2, \dots, w_{20}\}$$

Caractère statistique  $X$  représente le nombre d'enfants

Type du caractère  $X$  est quantitatif

Modalités de ce caractère  $X = 0, 1, 2, 3, 4, 5, 6, 7$

Le ménage n° 1 est composé de 4 enfants :  $X(w_1) = 4$

Le ménage n° 2 n'a pas d'enfants :  $X(w_2) = 0$

....

Le ménage n° 20 est composé de 6 enfants :  $X(w_{20}) = 6$

##### 1.2 Description de la série statistique d'une variable quantitative continue

Soit  $X$  une variable statistique quantitative continue. La variable statistique peut être présentée par un tableau comme elle peut être présentée par des graphiques.

Pour mieux représenter une variable quantitative continue ; on prend toutes les valeurs d'un intervalle de l'ensemble des nombres réels et on regroupe ces valeurs en  $k$  classes notées  $[e_i, e_{i+1}[$ . Par exemple la taille d'un individu, le nombre d'heures passées devant la télévision.

On peut définir le centre de la classe  $i$ , notée  $ci$

$$c_i = \frac{e_{i+1} + e_i}{2} \text{ pour tout } i = 1, K$$

et l'amplitude de la classe  $i$ , notée  $a_i$

$$a_i = e_{i+1} - e_i \text{ pour tout } i = 1, K$$

### 1.2.1 Représentation par des tableaux statistiques

Dans un tableau statistique on trouve les effectifs, les fréquences et aussi les effectifs cumulés et les fréquences cumulées des valeurs de la variable statistique  $X$ .

#### a. Effectif (ou fréquence absolue)

On appelle **effectif** d'une classe  $[e_i, e_{i+1}[$  le nombre de fois ou les valeurs  $X$  de cette classe apparaît dans la population statistique étudiée. Ce nombre est noté  $n_i$ . L'effectif est parfois appelé fréquence absolue.

On appelle effectif total de la population, noté  $N$ , la somme des  $k$  effectifs particuliers  $n_i$  correspondant à chacune des valeurs, soit :

$$n_1 + n_2 + \dots + n_K = \sum_{i=1}^K n_i = N$$

L'effectif  $n_i$  qui correspond à la classe  $[e_i, e_{i+1}[$  est égal au nombre d'unités statistiques qui se retrouvent dans cette classe. Par exemple : **[20 ans, 22 ans[** et l'effectif correspondant = 20  
**Cela signifie : 20 étudiants sont âgés entre 20 et 21 ans révolus. (car 22 n'appartient pas à la classe).**

#### b. Fréquence (ou fréquence relative)

On appelle fréquence relative de la modalité ou de la classe  $[e_i, e_{i+1}[$ , notée  $f_i$ , le rapport de l'effectif  $n_i$  associé à cette classe de valeurs de  $X$  et de l'effectif total  $N$  de la population observée. On peut dire aussi que la fréquence  $f_i$  qui correspond à la classe  $[e_i, e_{i+1}[$  est égale à la proportion d'unités statistiques qui se retrouvent dans cette classe.

$$f_i = \frac{n_i}{N} \text{ pour tout } i = 1, \dots, K$$

Ce rapport est égal au pourcentage d'individus présentant classe  $[e_i, e_{i+1}[$ , par rapport à l'ensemble de la population observée.  $f_i$  est toujours comprise entre 0 et 1.

On peut aussi exprimer les fréquences en pourcentage %, noté  $p_i$  est toujours comprise entre 0 et 100%.

Pour une série statistique à  $K$  modalités, on a :

$$\sum_{i=1}^K f_i = f_1 + f_2 + \dots + f_k = 1$$

Et  $\sum_{i=1}^K p_i = p_1 + p_2 + \dots + p_k = 100\%$

On nomme le tableau qui contient les valeurs de la variable statistique X, ces effectifs, ces fréquences et ces pourcentages un **TABLEAU STATATISTIQUE de la série** (xi; ni).  
Un tableau statistique doit avoir un titre clair et un numéro et une source.

c. Effectifs cumulés

Dans une série statistique où les valeurs de la variable étudiée X peuvent être rangées selon un ordre croissant (ou décroissant) on peut définir les effectifs cumulés. Un effectif cumulé correspond à l'effectif d'une modalité auquel on ajoute les effectifs de toutes les modalités précédentes. Il peut être **croissant**  $\vec{N}$  comme il peut être **décroissant**  $\overleftarrow{N}$ . On cumule les **ni**.

d. Fréquences cumulées

Dans une série statistique où les valeurs de la variable étudiée X peuvent être rangées selon un ordre croissant on peut également définir des fréquences cumulées. Une fréquence cumulée correspond à la fréquence d'une modalité auquel on ajoute les fréquences de toutes les modalités précédentes.

Elle peut être **croissante**  $F_i$  comme il peut être **décroissante**  $G_i$ . On cumule les **fi**.

On peut aussi exprimer les fréquences cumulées en pourcentage %, et elle est toujours comprise entre 0 et 100%.

**Remarque :**

**Lorsqu'il y a trop de valeurs discrètes, ou lorsque le caractère de la population est de nature continue, on regroupe les valeurs en classes [ei,ei+1[.**

Exemple :

salaire en euros	ni	fi	Fi	Gi
[0,1000[	200	0,5	0,5	1
[1000,1500[	120	0,3	0,8	0,5
[1500,2500[	60	0,15	0,95	0,2
[2500,3000[	20	0,05	1	0,05
total	400			

Sur les graphiques pour x=0 : (F,G) = (0,1)

Pour x = 1000 : (F,G) = (0,5 ; 0,5)

Pour X=3000 : (F,G) = (1,0)

Attention à l'alignement de Fi et de Gi !!!

F(x) = P (X< x) et G(x) = 1 – F(x)

## 1.2.2 La représentation Graphique d'une variable statistique continue

### a- HISTOGRAMME

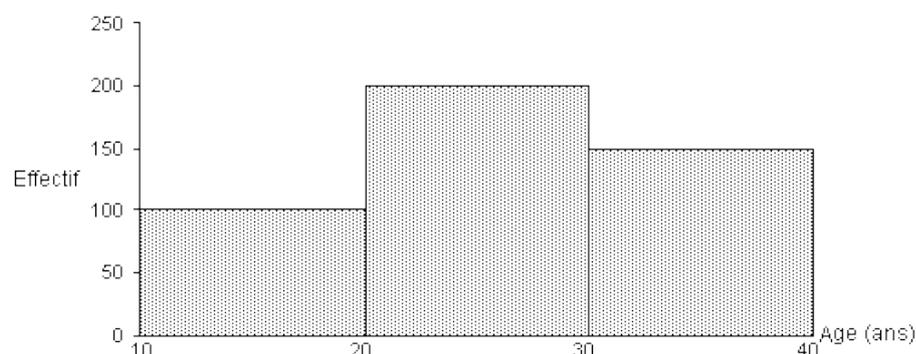
Le diagramme représentant la série de la variable statistique continue est un histogramme :

La répartition selon une variable continue ne peut être réalisée que si on détermine des intervalles sur lesquels on quantifiera le nombre de sujets ayant une valeur comprise dans cet intervalle.

L'**histogramme** sera représenté par des rectangles accolés les uns aux autres, la surface de chaque rectangle devant être proportionnelle à l'effectif de sujets compris dans l'intervalle. Pour cela, il est nécessaire de définir l'amplitude des intervalles de regroupement, puis de déterminer la hauteur du rectangle pour conserver cette proportionnalité.

Afin de simplifier la lecture du graphique, il est fortement conseillé d'utiliser des intervalles de même amplitude.

Exemple : une étude réalisée auprès de 400 sujets ayant un âge compris entre 10 et 40 ans : 100 personnes ont une valeur comprise dans l'intervalle [10 ; 20[, 200 dans l'intervalle [20 ; 30[ et 150 dans l'intervalle [30 ; 40[. Le second rectangle sera 2 fois plus haut que le premier et le troisième une fois et demi, car la base du rectangle est de même longueur.



Graphique 1 : répartition de 400 sujets selon leur âge.

Il ne faut pas oublier que l'on ne peut représenter l'information par un histogramme que si on fixe les bornes minimale et maximale.

Si les intervalles n'avaient pas été de même taille (amplitudes inégales), il aurait fallu modifier la hauteur afin que la surface des rectangles conserve la proportionnalité. Cela rend souvent la lecture du graphique trop compliquée.

Lorsqu'au moins deux classes ont des amplitudes différentes, la hauteur proportionnelle à l'effectif ne permet plus de construire un histogramme. En effet, la surface de chaque rectangle n'est plus proportionnelle à l'effectif (conformément à la définition). D'où la nécessité de corriger les fréquences (ou les effectifs).

$$f_i^c = \frac{f_i}{a_i} \text{ et } n_i^c = \frac{n_i}{a_i}$$

avec  $f_i^c$  : fréquences corrigées et  $n_i^c$  : effectifs corrigés

## b- POLYGONES DES EFFECTIFS OU DES FREQUENCES

On obtient le polygone des effectifs (ou des fréquences) en reliant les milieux des segments supérieurs de chaque rectangle de l'histogramme (l'aire du polygone des fréquences est l'aire de la surface par la ligne polygonale et l'axe des abscisses)

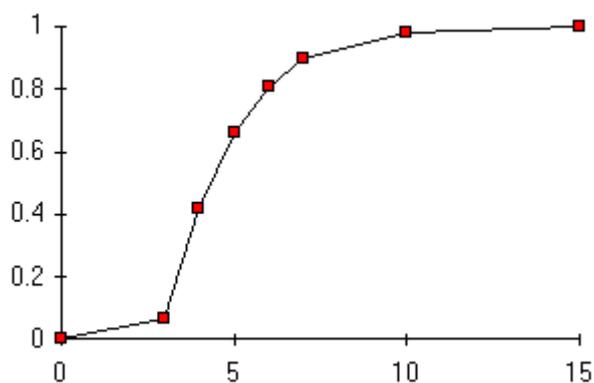
Courbes des fréquences cumulées ( ou Polygone des fréquences cumulées)

c-

On appelle courbes des fréquences cumulées croissantes la représentation graphique de la fonction de la répartition de la variable  $X$ . Les données étant groupées en classes, la fréquence cumulée croissante  $F_i$  associée à la classe n°  $i$  représente la proportion d'individus de la population pour lequel la variable prend une valeur strictement inférieure à la limite supérieure  $e_i$  de la classe n°  $i$

En pratique, la courbe des fréquences cumulées croissantes est obtenue en joignant, dans un système d'axes orthogonaux, les points d'abscisses  $e_i$  (extrémité de la classe n°  $i$ ) et d'ordonnée  $F_i$  correspondante. Joindre les points  $(e_i, F_i)$  revient à faire l'hypothèse d'une répartition uniforme des individus à l'intérieur des classes.

Exemple



La courbe cumulative (ou polygone des fréquences cumulées) est obtenue en portant les points dont les abscisses représentent la borne supérieure de chaque classe et les ordonnées les fréquences cumulées correspondantes, puis en reliant ces points par des segments de droite. Son équivalent dans la théorie probabiliste est la fonction de répartition.

**Remarque : On peut présenter les fréquences cumulées croissantes et décroissantes sur le même graphique.**

**Ici je les présente sur le tableau en prenant un exemple.**

*Si le temps me le permet je commence la variable discrète par un exemple !!!*

2.1.1 La description statistique d'une variable statistique discrète

**Exemple des ménages :** Tableau statistique correspondant :

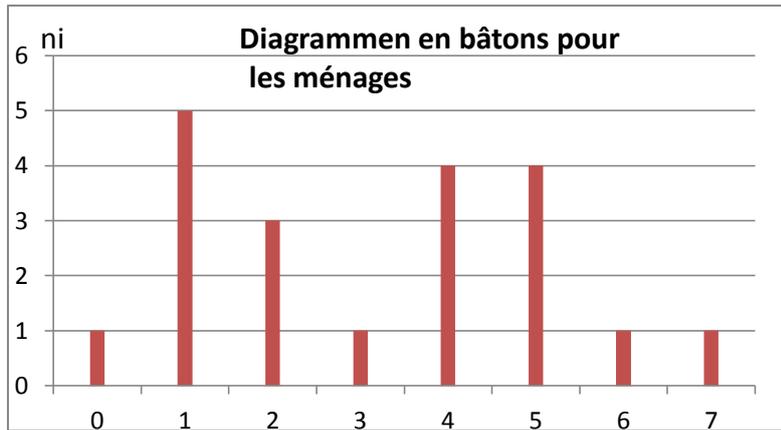
nombre d'enfants $x_i$	nombre de manages $n_i$	Fréquence relative $f_i$	effectifs cumulés croissants $\vec{N}$	fréquences relatives croissantes $F_i$	fréquences relatives d croissantes $G_i$
0	1	0,05	1	0,05	<b>1,00</b>
1	5	0,25	6	0,30	0,95
2	3	0,15	9	0,45	0,70
3	1	0,05	10	0,50	0,55
4	4	0,20	14	0,70	0,50
5	4	0,20	18	0,90	0,30
6	1	0,05	19	0,95	0,01
7	1	0,05	<b>20</b>	<b>1,00</b>	0,05
TOTAL	<b>20</b>				

Attention à l'alignement de  $F_i$  et de  $G_i$  !!!

$$F(x) = P(X < x) \text{ et } G(x) = 1 - F(x)$$

Les tableaux statistiques doivent être accompagnés des représentations graphiques car généralement les représentations graphiques sont plus faciles et rapides à interpréter.

2.1.2.1 Diagramme en bâtons



2.1.2.2. Polygone : En reliant les sommets des bâtons, nous obtenons un polygone.

2.1.2.3 Diagrammes cumulatifs

On appelle diagramme cumulatif, la courbe représentative de la fonction de répartition.

On appelle de la fonction de répartition d'un caractère X, l'application de IR dans [0,1]

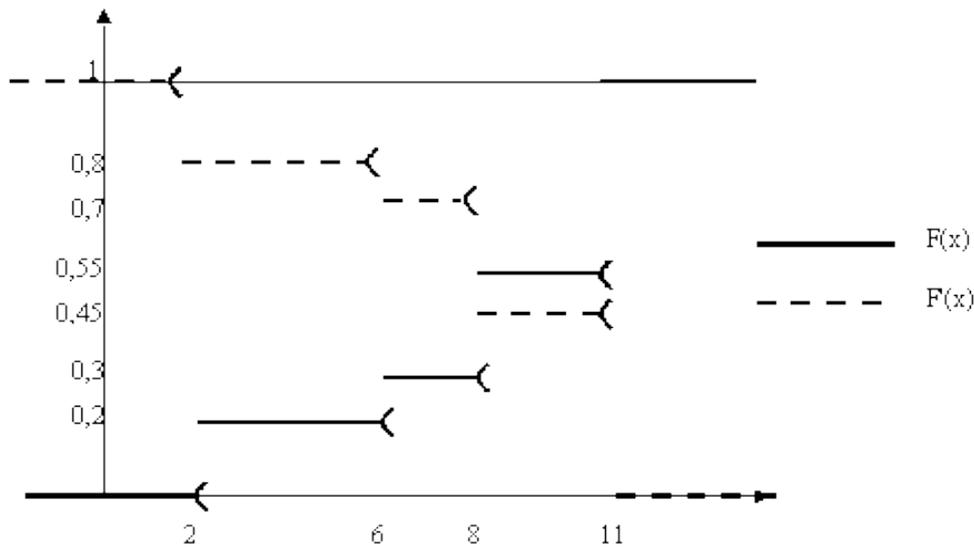
$$F : \mathbb{R} \rightarrow [0,1]$$

$$x \rightarrow F(x) = P(X < x)$$

$F(x)$  : la proportion des individus dont la valeur du caractère est strictement inférieure à x ;

Remarque :

- Le plus souvent diagramme cumulatif est obtenu à partir des fréquences cumulées croissantes. Dans le cas d'une variable discrète, la courbe se présente sous forme d'escaliers car la fonction F est constante par intervalle ;
- Chaque segment de cette courbe en escalier est ouvert à gauche et fermé à droite (sauf le dernier)

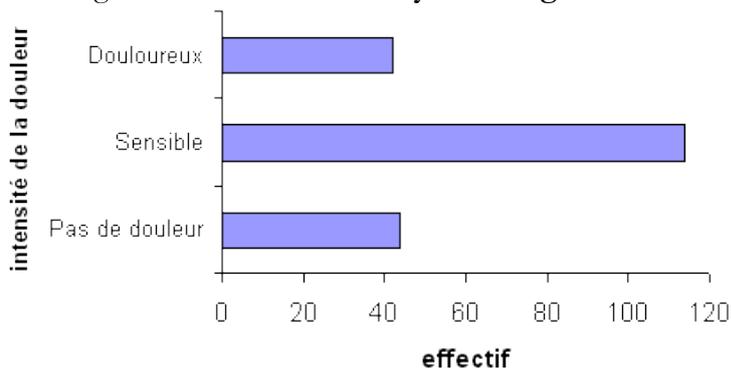


### 1.3 Description de la série statistique d'une variable qualitative

Lorsque les variables sont qualitatives ordinales ou quantitatives discrètes, cela doit se traduire sur le graphique. On ne peut notamment pas utiliser un diagramme en secteur pour ce type de variable.

Intensité de douleur à la mise au sein : diagramme en bâtons (horizontal ou vertical / effectifs ou pourcentages).

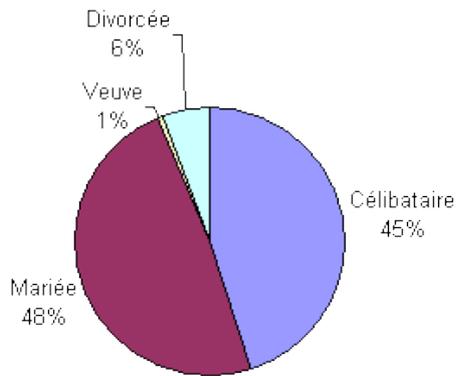
#### 2.3.1 diagramme en bâtons ou Tuyaux d'orgue.



**Graphique n°** : répartition des 200 femmes selon l'intensité de douleur à la mise au sein.

#### 2.3.2 diagramme en secteur

Statut marital : diagramme en secteurs



Graphique n° : répartition des 200 femmes selon leur statut marital.

Remarque : il peut exister plusieurs façons de représenter une même variable, notamment selon les effectifs ou le pourcentage. Le choix doit se faire pour que le graphique soit le plus compréhensible possible.

