Chapitre III : Statistique descriptive et analyse de données

I Construction d'une série

I.1 Exemples

Exemple 1. On regarde la répartition de la population de Montataire en fonction de l'âge (données 2007, *http ://www.cartesfrance.fr*).

Compléter le tableau suivant.

Valeur (ans)	$[0;14]$	$[15;29]$	$[30;44]$	$[45;59]$	$[60;74]$	> 75
Effectif	2665	2669	2502	2391	1356	683
Effectif cumulé	2665		7836	10227	11583	12266
Fréquence	$0,217$	$0,218$		$0,195$	$0,111$	$0,057$
Fréquence cumulée	$0,217$	$0,435$	$0,639$		$0,944$	1
Fréquence (en %)	$21,7$	$21,8$		$19,5$		$5,7$
Fréquence cumulée (en %)	$21,7$	$43,5$	$63,9$		$94,4$	100

Combien de personnes ont entre 45 et 59 ans ?

....

Combien de personnes ont moins de 30 ans à Montataire ?

....

Quelle est la population totale de Montataire ?

....

Exemple 2. La taille des élèves de la classe.

Valeur (en cm)	< 160	$[160; 170[$	$[170; 180[$	> 180
Effectif				
Effectif cumulé				
Fréquence				
Fréquence cumulée				
Fréquence (en %)				
Fréquence cumulée (en %)				

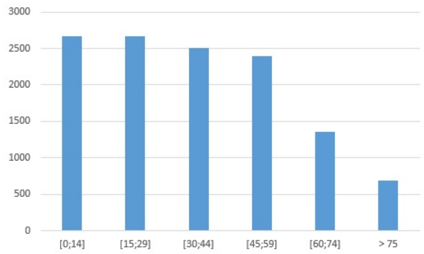
I.2 Généralisation

On s'intéresse à un caractère **quantitatif** (désigné par un nombre, exemple la taille) ou **qualitatif** (désigné par un mot, exemple la couleur des cheveux). On désigne par x_1, x_2, $\dots$, x_p les p valeurs différentes de ce caractère. On note n_1, n_2, $\dots$, n_p les effectifs associés et $N = n_1 + n_2 + \cdots + n_p$ l'effectif total.

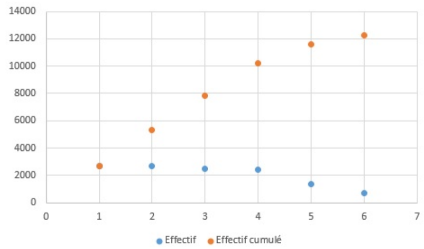
Valeur	x_1	x_2	$\dots$	x_p
Effectif	n_1	n_2	$\dots$	n_p
Effectif cumulé	n_1	$n_1 + n_2$	$\dots$	$n_1 + n_2 + \cdots + n_p = N$
Fréquence	$\frac{n_1}{N}$	$\frac{n_2}{N}$	$\dots$	$\frac{n_p}{N}$
Fréquence cumulée	$\frac{n_1}{N}$	$\frac{n_1+n_2}{N}$	$\dots$	$\frac{n_1+\cdots+n_p}{N} = 1$
Fréquence (en %)	$\frac{n_1}{N} \times 100$	$\frac{n_2}{N} \times 100$	$\dots$	$\frac{n_p}{N} \times 100$
Fréquence cumulée (en %)	$\frac{n_1}{N} \times 100$	$\frac{n_1+n_2}{N} \times 100$	$\dots$	100

II Représentations graphiques

Diagramme en bâtons



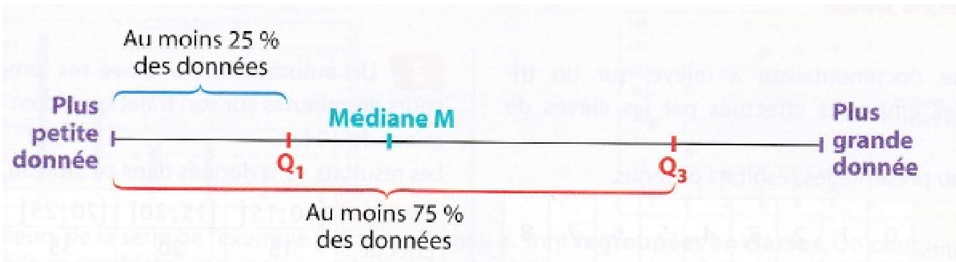
En nuage de points



III Caractéristiques de position, dispersion

Définition III.1

- La **moyenne**, notée $\bar{x}$ est donnée par
$$\bar{x} = \frac{n_1 x_1 + \cdots + n_p x_p}{N} \qquad \text{ou encore} \qquad \bar{x} = f_1 x_1 + \ldots f_p x_p.$$
- La **médiane** est la plus petite valeur pour laquelle la fréquence cumulée dépasse $0,5$ (ou 50%).
- Le **premier quartile** $\mathbf{Q_1}$ est la plus petite valeur pour laquelle la fréquence cumulée dépasse $0,25$ (ou 25%).
- Le **troisième quartile** $\mathbf{Q_2}$ est la plus petite valeur pour laquelle la fréquence cumulée dépasse $0,75$ (ou 75%).
- **L'étendue** est la différence entre la plus grande valeur et la plus petite valeur : $x_p - x_1$.
- **L'écart inter-quartile** est la différence entre les deux quartiles $\mathbf{Q_3} - \mathbf{Q_1}$.



Exemple 1. Puisque l'on a des classes, on va choisir pour représentant le milieu de la classe :

Classe (ans)	$[0;14]$	$[15;29]$	$[30;44]$	$[45;59]$	$[60;74]$	> 75
Représentant	$\frac{0+14}{2} = 7$	$\frac{29+15}{2} = 22$	$\frac{30+44}{2} = 37$	$\frac{45+59}{2} = 52$	$\frac{60+74}{2} = 67$	82

La moyenne vaut
$$\bar{x} = \frac{2665 \times 7 + 2669 \times 22 + 2502 \times 37 + 2391 \times 52 + 1356 \times 67 + 683 \times 82}{12266} \simeq 35,96.$$

La classe médiane est $[30;44]$, on peut donc prendre pour médiane le représentant de la classe, soit 37 ans.
La classe du premier quartile est $[15;29]$, le premier quartile vaut donc $\mathbf{Q_1} = 22$ ans. La classe du troisième quartile est $[45;59]$, le troisième quartile vaut donc $\mathbf{Q_3} = 52$ ans.
L'étendue vaut dans ce modèle $82 - 0 = 82$ ans et l'écart inter-quartile $\mathbf{Q_3} - \mathbf{Q_1} = 52 - 22 = 30$ ans.