

1 Vocabulaire des statistiques

Définitions :

- *Population* : ensemble étudié ;
- *individu* : élément de la *population* ;
- *caractère* : propriété étudiée.
Un *caractère* peut être :
 - *qualitatif* : les valeurs du *caractère* ont une valeur descriptive (ex. : ville d'origine) ;
 - *quantitatif* : les valeurs du *caractère* désignent des quantités (ex. : taille).Un *caractère quantitatif* peut être considéré comme :
 - *discret* : valeurs peu nombreuses (ex. : nombre d'enfants) ;
 - *continue* : valeurs infinies ou très nombreuses (ex. : taille).Les valeurs d'un caractère sont notées x_i , on supposera qu'elles sont au nombre de k dans la série.
- *classe* : partie de la *population* de valeurs de *caractères* proches ;
- *effectif* pour la valeur x_i : nombre d'individus d'une classe (noté n_i) ou d'une population (effectif total noté N).
- *fréquence* pour la valeur x_i : $f_i = \frac{n_i}{N}$ ou en pourcentage $f_i = \frac{n_i}{N} \times 100$;
- *effectif cumulé croissant (ECC)* pour la valeur x_i : somme des effectifs des valeurs inférieures ou égales à x_i ;
- *effectif cumulé décroissant (ECD)* pour la valeur x_i : somme des effectifs des valeurs supérieures ou égales à x_i ;

2 Représentations graphiques

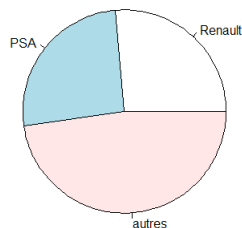
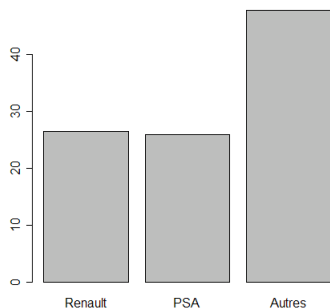
2.1 Caractères qualitatifs ou quantitatifs discrets

On utilise un diagramme *en bâtons* ou *en barres* ou *circulaire*.

Exemple :

Ventes de voitures neuves en août 1996 :

Renault	PSA	étrangères
26,4%	25,9 %	47,7%

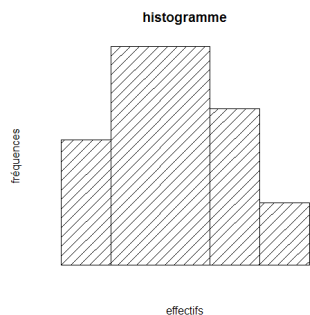


2.2 Caractères quantitatifs continus

2.2.1 histogrammes

Propriété :

Dans un *histogramme* l'effectif d'une classe est proportionnel à l'aire des rectangles.



Exemple :

prix	$[0 ; 500[$	$[500 ; 1000[$	$[1000 ; 1500[$	$[1500 ; 2000[$	$[2000 ; 2500[$
nombre d'articles	4	14	6	5	2

2.2.2 Polygone des effectifs cumulés

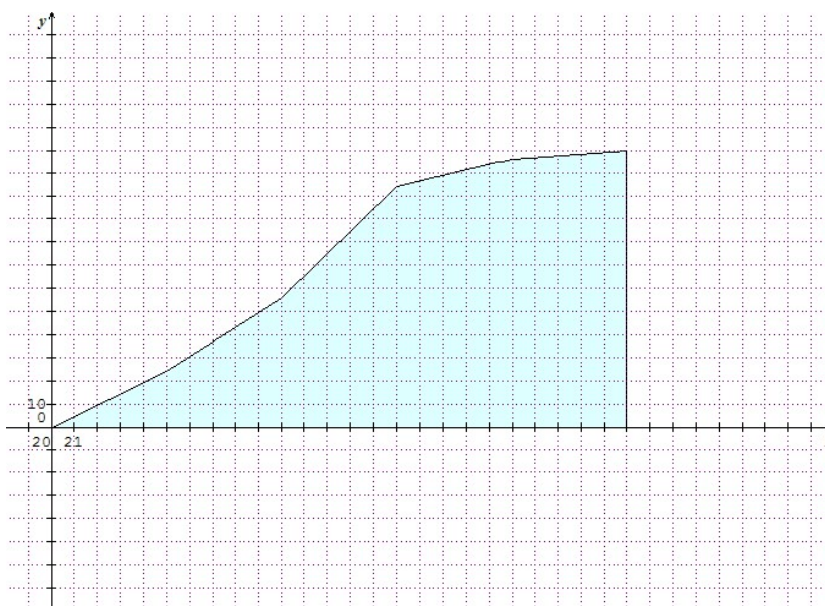
Méthode de contruction :

Lorsque la série est regroupée en classes $[a_i; b_i[$, le polygone des effectifs cumulés croissants est formé des segments reliant les points ayant pour abscisse l'extrémité b_i de chaque classe et pour ordonnée l'effectif cumulé croissant de la classe correspondante.

Exemple :

Prix de vente moyen d'un article constaté dans 120 magasins.

Prix	$[20; 25[$	$[25; 30[$	$[30; 35[$	$[35; 40[$	$[40; 45[$
Effectif	24	32	48	12	4
E.C.C.	24	56	104	116	120



3 Caractéristiques d'une série statistique

3.1 Médiane

Définition :

La *médiane* d'une série statistique est la valeur du caractère qui sépare la série *ordonnée* en deux sous séries de même effectif.

Exemples :

- On considère la série : 1 ; 5 ; 6 ; 6 ; 7 ; 8 ; 9 ; 10 ; 10. L'effectif total est 9. La valeur centrale est donc la valeur de rang 5. La médiane est donc 7.
- On considère la série : 1 ; 5 ; 6 ; 6 ; 7 ; 8 ; 9 ; 10. L'effectif total est 8. Il n'y a donc pas une unique valeur centrale mais deux. On choisit donc pour médiane la moyenne entre ces deux valeurs centrales, soit les valeurs de rang 4 et 5. La médiane est donc $\frac{6+7}{2} = 6,5$.

3.2 Quartiles

Définition :

- Le premier quartile noté Q_1 de la série statistique est la plus petite valeur telle qu'au moins 25% des valeurs lui soient inférieures ou égales ;
- le troisième quartile noté Q_3 de la série statistique est la plus petite valeur telle qu'au moins 75% des valeurs lui soient inférieures ou égales.

Détermination pratique :

On suppose la série ordonnée dans l'ordre croissant des valeurs du caractère. Soit N l'effectif total.

- Si $\frac{N}{4}$ est un entier alors Q_1 est la valeur de rang $\frac{N}{4}$ et Q_3 est la valeur de rang $\frac{3N}{4}$;
- si $\frac{N}{4}$ n'est pas un entier, alors Q_1 est la valeur dont le rang suit $\frac{N}{4}$ et Q_3 est la valeur dont le rang suit $\frac{3N}{4}$.

3.3 Moyenne

Définition :

La *moyenne* est donnée par :

$$\frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{N}$$

Remarque :

Dans le cas d'un *caractère continue* (donc de valeurs regroupées en *classes*), on prend pour x_i le centre $\frac{a_i+b_i}{2}$ des classes $[a_i; b_i[$.

Propriété :

La moyenne $\bar{x}$ est aussi donnée par :

$$\bar{x} = f_1x_1 + f_2x_2 + \dots + f_kx_k$$

Preuve :

$$\begin{aligned}\frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{N} &= \frac{n_1}{N}x_1 + \frac{n_2}{N}x_2 + \dots + \frac{n_k}{N}x_k \\ &= f_1x_1 + f_2x_2 + \dots + f_kx_k\end{aligned}$$

3.4 Étendue

Définition :

L'étendue d'une série statistique est la différence entre la plus grande valeur et la plus petite.

3.5 L'écart interquartile

C'est un critère de dispersion de la série.

L'intervalle interquartile est $[Q_1 ; Q_3]$.

La signification de l'intervalle interquartile est que les 50% « centraux » des valeurs de la série se trouvent dans cet intervalle.

L'écart interquartile est $Q_3 - Q_1$.

3.6 Choix des indicateurs

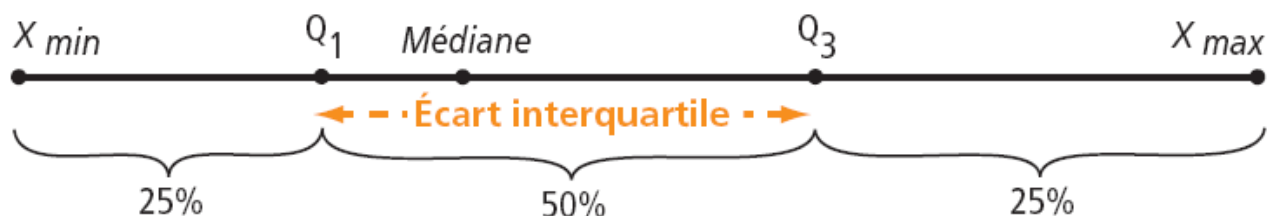
En général, on choisit un indicateur de position et un indicateur de dispersion :

- **Moyenne , étendue**

Le défaut du premier couple est qu'il est sensible aux valeurs extrêmes.

- **Médiane et intervalle interquartile**

Ce couple **médiane – intervalle interquartile** est souvent représenté graphiquement par un **diagramme en boîte**, appelé parfois « **boîte à moustaches** » ou « **boîte à pattes** ».



Ce couple est moins sensible aux valeurs extrêmes.