Ch 3 Statistique descriptive ; Analyse de données (1)

1°/ Vocabulaire : rappels
2°/ Présentation d'une série statistique
3°/ Paramètres d'une série statistique

1°/ Vocabulaire : rappels

Vocabulaire : Lorsque l'on mène une étude statistique :

- On doit d'abord définir le ***caractère*** que l'on veut étudier. Ce caractère peut être ***qualitatif*** (qualité), comme la couleur des yeux, ou ***quantitatif*** (quantité) comme la taille par exemple.
- On doit délimiter la ***population*** étudiée, c'est-à-dire l'ensemble des individus qui ont participé à l'enquête.
- On doit déterminer ***l'effectif total***, c'est-à-dire le nombre total d'individus qui ont participé (les 32 élèves), puis les ***effectifs*** selon chaque valeur proposée.

Une fois le recueil de données effectué, on peut « travailler » avec les données et déterminer :

- la ***fréquence*** de chaque valeur, c'est-à-dire le quotient $\frac{effectif\ de\ la\ valeur}{effectif\ total}$. Par commodité, on l'exprimera généralement en pourcentages.
- Si le caractère est quantitatif, on peut calculer ***l'étendue*** de la série, c'est-à-dire la différence entre la plus grande valeur et la plus petite valeur.

Dans la suite du chapitre, les séries étudiées seront des séries dont le caractère est quantitatif.

2°/ Présentation d'une série statistique

2.1 Effectifs cumulés

Définition : On note x_i une valeur prise par un caractère quantitatif.
L'**effectif cumulé** croissant (resp. décroissant) de x_i est la somme des effectifs des valeurs inférieures (resp. supérieures) ou égales à x_i.

Application : calculer les effectifs cumulés croissants d'une série

Le tableau ci-dessous a été dressé après recensement du nombre d'enfants dans chacun des foyers d'un village. Compléter la ligne des effectifs cumulés croissants (E.C.C.) :

Nombre d'enfants x_i	0	1	2	3	4	5	6	7
Effectif n_i	290	170	155	95	43	27	20	10
E.C.C.								

Le tableau des effectifs cumulés croissants est intéressant dans le sens où il donne directement le nombre de foyer ayant au plus 2 enfants, par exemple.

2.2 Fréquences cumulées

Définition : On note x_i une valeur prise par un caractère quantitatif. La **fréquence cumulée croissante** (resp. décroissante) de x_i est la somme des fréquences des valeurs inférieures (resp. supérieures) ou égales à x_i.

Application : calculer les fréquences cumulées croissantes d'une série

Reprenons l'exemple précédent. Compléter les lignes des fréquences, puis des fréquences cumulées croissants (F.C.C.) :

Nombre d'enfants x_i	0	1	2	3	4	5	6	7
Effectif n_i	290	170	155	95	43	27	20	10
Fréquence f_i (en %)								
F.C.C. (en %)								

Grâce à ce tableau, nous pouvons dire que ……% des familles de ce village ont au plus deux enfants.

3°/ Paramètres d'une série statistique

Dans ce paragraphe, on considère une série statistique de taille N et on note n_i l'effectif de chaque variable x_i ($1 \leq i \leq p$).

3.1 Paramètres de position (ou de tendance centrale)

Définition : La ***moyenne*** d'une série statistique est le nombre, noté $\bar{x}$, défini par :

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \cdots + n_px_p}{N}$$

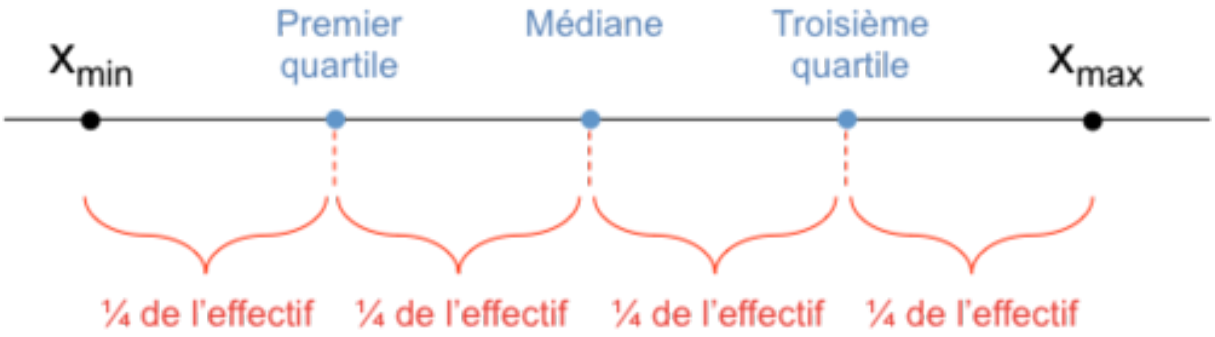
Application :calculer la moyenne d'une série statistique

Dans un lycée, il y a 6 classes de 34 élèves, 3 classes de 32 élèves, 9 classes de 31 élèves et une classe de 28 élèves. Calculer le nombre moyen d'élèves par classe.

Définition : La **médiane** d'une série statistique est le nombre, noté Me, tel que 50% au moins des individus ont une valeur du caractère inférieur à la médiane. Autrement dit, en considérant que la liste des données est ordonnée par **ordre croissant** :

- si la série est de taille impaire (N=2n+1), la médiane est la donnée de rang n + 1 ;
- si la série est de taille paire (N=2n+1), la médiane est la demi-somme des données de rang n et n + 1.

Remarque : pour déterminer la médiane d'une série statistique, il est souvent utile de calculer les E.C.C.



- Le **1er quartile Q_1** est la plus petite valeur de la série ordonnée telle que au moins 25 % (1/4) des données lui soient inférieures ou égales.
- Le **3ème quartile Q_3** est la plus petite valeur de la série ordonnée telle que au moins 75 % (3/4) des données lui soient inférieures ou égales.

Propriété : Environ 50% des données d'une série ordonnée sont comprises entre les quartiles Q_1 et Q_3.

Application : déterminer la médiane et les quartiles d'une série statistique

Voici les relevés des notes données par un jury à l'épreuve de mathématiques du baccalauréat.

Après avoir calculé les effectifs cumulés croissants, déterminer la note médiane et les quartiles de cette série.

Note x_i	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Effectif n_i	1	2	2	4	6	5	4	5	11	5	6	6	5	4	3	5	2	1
E.C.C																		

3.2 Paramètres de dispersion

- L'**étendue** d'une série est la différence entre la plus grande et la plus petite valeur de la série, notées x_{max} et x_{min}.
- L'intervalle $[Q_1; Q_3]$ est appelé l'**intervalle interquartile**.
- Q_1-Q_3 est appelé l'**écart interquartile**.
- Le **mode** est la valeur dont l'effectif est le plus important.

L'analyse de ces indicateurs permet de dire si une série est plus ou moins **dispersée** qu'une autre.