

Chap.3 : Statistique descriptive. Analyse de données.

I. Vocabulaire

La statistique a pour but la collecte, l'analyse et l'interprétation des observations relatives à des phénomènes collectifs.

Une étude statistique comprend 4 parties:

- a) le recueil de données (enquête)
- b) la présentation des résultats (tableaux, diagrammes)
- c) le calcul des paramètres caractéristiques
- d) l'exploitation (domaine de l'économie).

Les premières études statistiques étaient démographiques : on en a conservé le vocabulaire.

L'ensemble sur lequel porte l'étude statistique s'appelle la **population**.

Un élément de cet ensemble est un **individu**.

L'étude statistique rend compte d'un aspect des individus de la population appelé **variable** ou **caractère**.

Si l'on peut mesurer cet aspect, la variable est de nature **quantitative**. Elle prend différentes **valeurs**.

Si les valeurs sont isolées, on dit que la variable est **discrète**. Si les valeurs appartiennent à un intervalle de \mathbb{R} , on dit que la variable est **continue**.

Si l'aspect ne se traduit pas par des nombres, la variable est **qualitative**. Elle prend différentes **modalités**.

Exemples :

① On étudie le prix d'un produit dans différents points de vente.

La population est

La variable est, il s'agit d'une variable

② On étudie la couleur des voitures vendues par un garage.

La population est

La variable est, il s'agit d'une variable

③ On étudie la taille des élèves d'une classe de seconde.

La population est

La variable est, il s'agit d'une variable

Le nombre d'individus pour lesquels la variable prend une valeur donnée est l'**effectif** de cette valeur.

La somme des effectifs de toutes les valeurs est l'**effectif total**. C'est le nombre d'individus de la population.

La **fréquence** d'une valeur est le quotient $\frac{\text{effectif de la valeur}}{\text{effectif total}}$.

Remarque : On résume en général une étude statistique dans un tableau.

| | | | | | |
|-----------|-------|-------|-------|-------|-------|
| Valeurs | x_1 | x_2 | | x_p | Total |
| Effectifs | n_1 | n_2 | | n_p | N |
| Fréquence | f_1 | f_2 | | f_p | 1 |

Exemple : l'étude de l'exemple ① précédent donne les résultats suivants :

35 ; 37 ; 40 ; 35 ; 36 ; 36 ; 39 ; 36 ; 35 ; 37 ; 37 ; 38 ; 36 ; 35 ; 40 ; 37 ; 36 ; 36 ; 38 ; 39 ; 35 ; 36 ; 37 ; 38 ; 36.

La liste de ces valeurs est appelée **série brute**.

| | | | | | | |
|---------------------------|----|----|----|----|----|----|
| Prix constaté en € | 35 | 36 | 37 | 38 | 39 | 40 |
| Nombre de points de vente | | | | | | |
| Fréquence | | | | | | |
| Fréquence en % | | | | | | |

On parle de **série dépouillée** quand on connaît la liste des valeurs prises et l'effectif de chacune d'elle.

II Présentation d'une série statistique.

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 .

La **fréquence cumulée croissante** de x_i est la somme des fréquences des valeurs inférieures (resp. supérieures) ou égales à x_i .

Exemple : Série ④ Ce tableau donne le nombre d'enfants âgés de 0 à 16 ans dans un échantillon de 100 familles.

| | | | | | | |
|--------------------------------|----|----|----|----|---|---|
| Nbre d'enfants | 0 | 1 | 2 | 3 | 4 | 5 |
| Effectif | 50 | 23 | 10 | 14 | 2 | 1 |
| Effectifs cumulés croissants | | | | | | |
| Fréquence cumulées croissantes | | | | | | |

Dresser le même tableau pour la série ①.

III Représentation graphique.

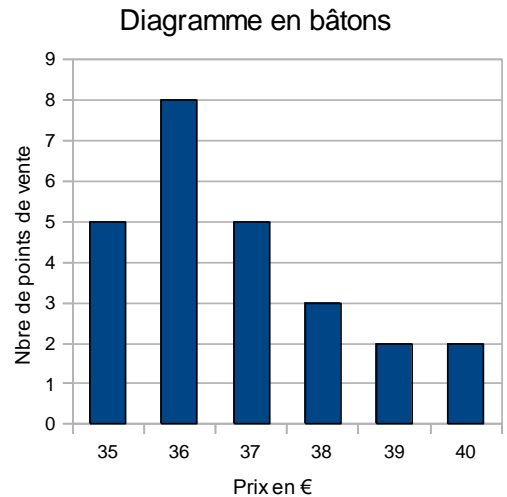
Il existe plusieurs méthodes pour représenter graphiquement une série statistique.

Pour un caractère qualitatif, on peut faire :

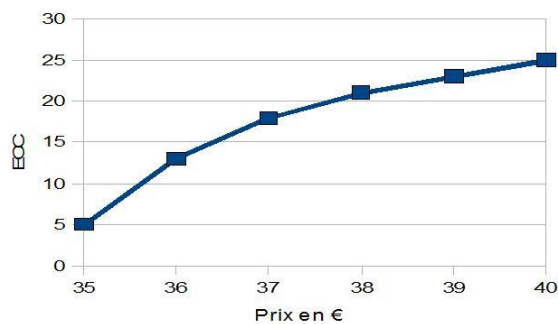
- un diagramme à bâtons : il est constitué de segments de droite verticaux dont les hauteurs sont égales aux effectifs ou aux fréquences de chaque modalité.
- un diagramme circulaire

Pour un caractère quantitatif, on peut faire :

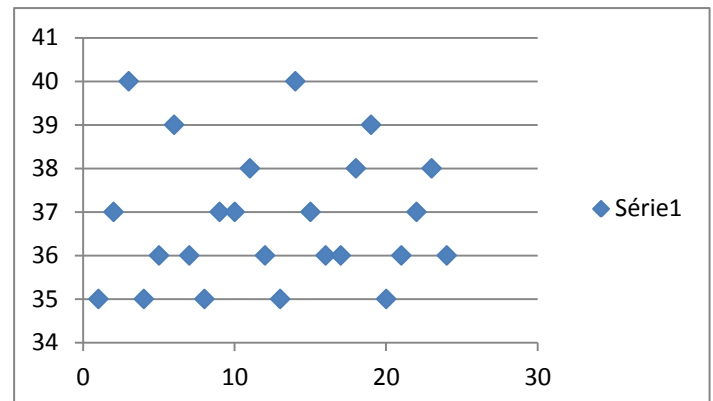
- Les diagrammes en bâtons :



- Les courbes de fréquences ou d'effectifs cumulés :



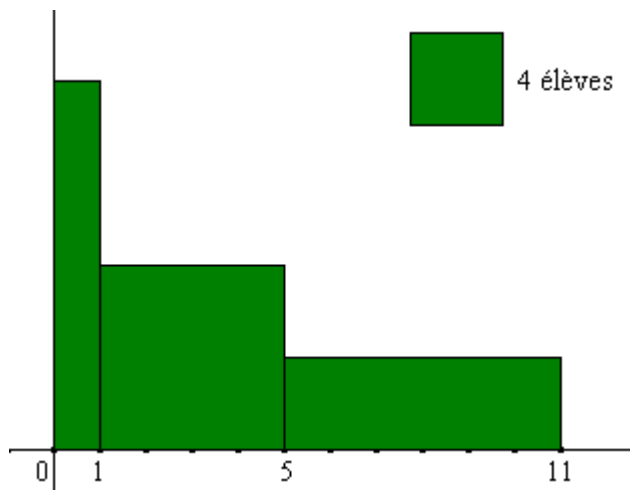
- Les nuages de points :



- Les histogrammes : lorsque la série est à variable continue, elle peut comporter de très nombreuses valeurs différentes, on les regroupe alors par classe pour obtenir l'**histogramme**.

Exemple : Série ⑤ : On s'intéresse à la distance entre le domicile et le lycée pour les élèves d'une classe. Les données sont toutes différentes, on décide donc de les **regrouper par classe** :

| | | | |
|------------------|---------|---------|----------|
| distance (en km) | [0 ; 1[| [1 ; 5[| [5 ; 11[|
| nombre d'élèves | 8 | 16 | 12 |



Dans le cas d'un histogramme, les effectifs (ou les fréquences) et les aires des rectangles sont proportionnels.

Lorsque les classes n'ont pas la même amplitude, on ne peut pas proposer d'unité sur l'axe des ordonnées.

Construction de l'histogramme représentant l'exemple :

On choisit 0,5 cm pour 1 km sur l'axe des abscisses et 1 cm² représente 4 élèves.

Activité 3 p 246 : Regroupement par classe.

IV Paramètres des séries quantitatives.

1°) Paramètres de position.

a) Moyenne

La **moyenne** notée **m** ou \bar{x} est le quotient de la somme des valeurs par l'effectif total. On l'obtient aussi en faisant la somme des produits des valeurs par leur fréquence.

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{n_1 + n_2 + \dots + n_p} = f_1x_1 + f_2x_2 + \dots + f_px_p.$$

Exemples :

La moyenne de la série ① précédente est

$$\bar{x} = \frac{\dots \times \dots + \dots \times \dots + \dots \times \dots + \dots \times \dots + \dots \times \dots + \dots \times \dots}{\dots + \dots + \dots + \dots + \dots + \dots} \approx$$

ou $\bar{x} = \dots \times \dots + \dots \times \dots + \dots \times \dots + \dots \times \dots + \dots \times \dots + \dots \times \dots \approx$

Calculer la moyenne de la série ④

Dans le cas de regroupement par classes, on applique la formule aux centres des classes.

Pour la Série ⑤ :

$$x = \frac{\dots \times \dots + \dots \times \dots + \dots \times \dots}{\dots + \dots + \dots} \approx$$

b) Médiane

La **médiane** d'une série statistique partage cette série en deux parties de telle sorte que :

- au moins la moitié des valeurs sont inférieures ou égales à la médiane ;
- au moins la moitié des valeurs sont supérieures ou égales à la médiane.

Remarque : Pour trouver la médiane, on range les valeurs de la variable dans l'ordre croissant et choisit la valeur « centrale » si l'effectif est impair ou la moyenne des valeurs « centrales » si l'effectif total est pair.

Exemples :

- La médiane de la série 4 – 5 – 7 – 9 – 15 est **7** car la valeur « centrale » (la 3^e) est 7.
- La médiane de la série 5 – 8 – 9 – 14 est **8,5** car les valeurs « centrales » (les 2^e et 3^e) sont 8 et 9 et leur moyenne est 8,5.
- L'effectif total de la série ① est 25. La médiane est donc la valeur. La médiane est

Interprétation : Activité 2 p 245 du livre : La moyenne ne fait pas tout...

c) Les quartiles.

Définitions : Soit une série statistique dont les données sont rangées par ordre croissant.

Le premier quartile est la plus petite donnée Q_1 de la liste telle qu'au moins un quart des données de la liste sont inférieures ou égales à Q_1 .

Le troisième quartile est la plus petite donnée Q_3 de la liste telle qu'au moins les trois quarts des données sont inférieures ou égales à Q_3 .

Exemple : calculer le premier et le troisième quartile de la série ① .

2°) Paramètres de dispersion.

On appelle **étendue** d'une série statistique la différence entre la plus grande valeur du caractère et la plus petite valeur.

$Q_3 - Q_1$ est appelé **écart interquartile**.

Exemple : Pour la série ① ,

Devoir Maison : n °50 p 260.

3°) Détermination graphique de la médiane et des quartiles.

La courbe des fréquences ou des effectifs cumulés croissants permet de déterminer graphiquement la médiane et les quartiles.

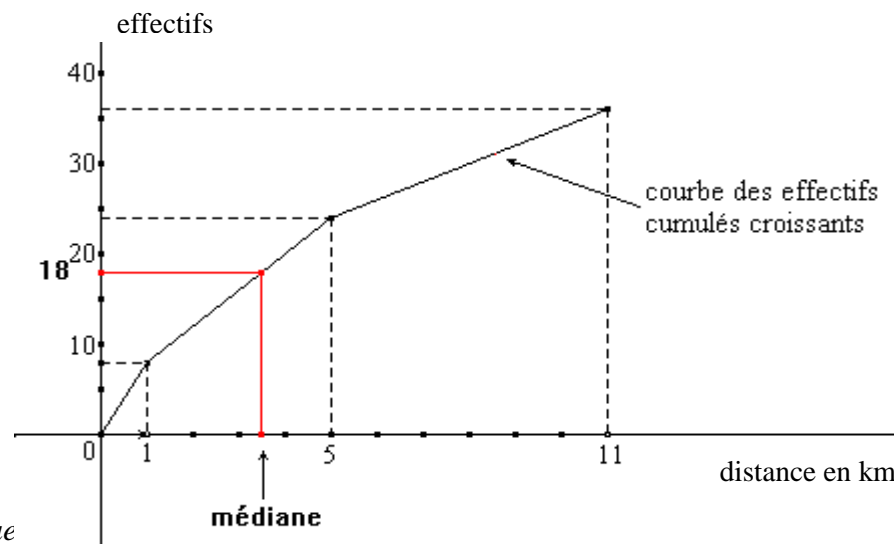
Méthode : revenons sur l'exemple de Série ⑤ .

On construit tout d'abord le tableau des effectifs cumulés croissants (ou celui des fréquences cumulées croissantes)

| | | | |
|----------------------------|---------|---------|----------|
| distance (en km) | [0 ; 1[| [1 ; 5[| [5 ; 11[|
| effectifs cumulés | 8 | 24 | 36 |
| fréquences cumulées (en %) | 22,2 | 66,7 | 100 |

On place dans un repère orthogonal les points (0 ; 0), puis (1 ; 8), (5 ; 24) et (11 ; 36).

On admet que la répartition dans chaque classe est uniforme, ainsi on joint ces points par des segments.



$\frac{36}{2} = 18$ La médiane est l'abscisse du point de la courbe d'ordonnée 18.

Ici, la distance médiane est environ 3,5 km.

Remarque : on obtient la même médiane en utilisant les fréquences cumulées croissantes. La médiane est l'abscisse du point de la courbe d'ordonnée 50%. Q_1 est l'abscisse du point de la courbe d'ordonnée 25%.

4°) Utilisation de la calculatrice.

Les calculatrices sont capables de calculer les caractéristiques de position et de dispersion d'une série statistique.

Pour les TI : page 326 du livre.

Pour les Casio : page 329 du livre.

Devoir Maison : TD 2 p 254.