

L'étude d'un problème statistique peut se décomposer en quatre étapes :

- recueil de données
- classement de ces données (statistique descriptive),
- réduction de ces données (statistique descriptive),
- analyse de ces données visant à la déduction de prévisions (statistique inférentielle).

## I RECUEIL DE DONNEES

### 1° Etude statistique

Une étude statistique descriptive s'effectue sur une **population** (des personnes, des villes, des voitures: : :) dont les éléments sont des **individus (unités statistiques)** et consiste à observer et étudier un même aspect sur chaque individu, nommé **caractère** (taille, nombre d'habitants, consommation). Les valeurs prises par le caractère sont appelées **modalités**

### 2° caractère étudié

Il existe deux types de caractère :

**Caractère quantitatif** : Les modalités sont numériques. le caractère peut "se mesurer".

On distingue alors deux types de caractère quantitatif :

- **discret** : c'est un caractère quantitatif qui ne prend qu'un nombre fini de valeurs.

Par exemple : le nombre d'enfants d'un couple, notes d'élèves.

- **continu** : c'est un caractère quantitatif qui, théoriquement, peut prendre toutes les valeurs d'un intervalle de l'ensemble des nombres réels. Ses valeurs sont alors regroupées en classes.

Par exemple : la taille d'un individu, distances, salaire.

**Caractère qualitatif** : les modalités ne sont pas numériques. Le caractère ne peut pas se mesurer. par exemple : la profession, la couleur des yeux, la nationalité.

## II: CLASSEMENT DE DONNEES

### 1° Série statistique

- Le caractère étudié prend un certain nombre de valeurs que l'on note  $x_1, x_2, x_3, \dots, x_n$
- L'**effectif** de la valeur  $x_i$  est le nombre d'apparitions de  $x_i$ . : on le note  $n_i$ .
- L'**effectif total** est la somme des effectifs :  $N = n_1 + n_2 + n_3 + \dots + n_n = \sum n_i x_i$
- La **fréquence**  $f_i$  de la valeur  $x_i$  est le quotient  $\frac{n_i}{N}$ . (La somme des fréquences vaut 1)

En général une série statistique à caractère discret se présente sous la forme :

Valeurs (modalités)	$x_1$	$x_2$	.....	$x_n$
Effectifs	$n_1$	$n_2$	.....	$n_n$
Fréquences	$f_1$	$f_2$	.....	$f_n$

En général une série statistique à caractère continu se présente sous la forme :

classes	$[x_1 ; x_2 [$	$[x_2 ; x_3 [$	.....	$[x_n ; x_{n+1} [$
Effectifs	$n_1$	$n_2$	.....	$n_n$
Fréquences	$f_1$	$f_2$	.....	$f_n$

- La **fréquence cumulée** jusqu'à la valeur  $x_k$  est:  $f_1 + f_2 + \dots + f_k$ .

### 2° Exemples de représentation graphique :

**Diagramme circulaire**, Caractère qualitatif. On divise un disque en secteurs circulaires dont l'aire est proportionnelle à la fréquence correspondante.

**diagramme en bâtons** Caractère quantitatif discret On représente en abscisses :les modalité du caractère (las valeurs prises par le caractère et en ordonnées :les effectifs. Chaque couple  $(x_i, n_i)$  est représenté par un bâton. En joignant les sommet des bâtons, on obtient **le polygone des effectifs**

**Histogramme** Caractère quantitatif continu On représente en abscisse les modalités prises par la caractère .

### III REDUCTION DES DONNEES (CARACTERE QUANTITATIF)

#### 1° Paramètre d'une série

Lorsqu'une série comporte un grand nombre de valeurs, on cherche à la résumer, si possible, à l'aide de quelques nombres significatifs appelés paramètres.

- L'**étendue**.
- Le **mode** ou la **classe modale**.
- La **médiane** ou la **classe médiane**.
- Les **quartiles**, les **déciles**, les **centiles**.
- La **moyenne**

#### 2° Le mode et l'étendue

Ce sont les paramètres les plus simple à étudier

**Le mode** : caractéristique de *position*

C'est la modalité la plus représentée. Quand le caractère est quantitatif continu on parle de : **classe modale**

**L'étendue** : caractéristique de *dispersion*

C'est la différence entre la plus grande valeur du caractère et la plus petite de ses valeurs.

#### 3° Moyenne d'une série statistique : caractéristique de *position*

- La moyenne, désignée par  $\bar{x}$ , est le quotient de la somme des valeurs du caractère (pondérées par les effectifs)

$$\text{par l'effectif total : } \bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{N} = \frac{\sum_{i=1}^p n_i x_i}{N}$$

La valeur moyenne peut aussi se calculer avec les fréquences :  $\bar{x} = f_1 x_1 + f_2 x_2 + \dots + f_n x_n = \sum_{i=1}^p f_i x_i$ .

- Une moyenne élaguée est obtenue en supprimant des valeurs de la variable qui paraissent aberrantes.
- Pour le calcul de la moyenne, on peut partager la population en deux ou plusieurs parties et calculer la moyenne de chacune des parties : on obtient la moyenne générale en calculant la moyenne de ces moyennes partielles  $\bar{x}_1$  et  $\bar{x}_2$  affectées chacune d'un coefficient égal à l'effectif de chaque partie considérée,  $n_1$  et  $n_2$  :

$$\bar{x} = \frac{n_1 \times \bar{x}_1 + n_2 \times \bar{x}_2}{N}$$

- Si l'on multiplie toutes les valeurs  $x_i$  par une constante, alors la moyenne est multipliée par cette constante :

$$k\bar{x} = k \bar{x}.$$

- Si l'on ajoute un même nombre à toutes les valeurs  $x_i$ , alors la moyenne est augmentée de cette constante :

$$\bar{x} + a = \bar{x} + a.$$

#### 4° Médiane : caractéristique de *position*.

On appelle médiane d'une série statistique une valeur  $M_e$  du caractère qui partage la population en deux parties de même effectif. Les éléments de la première partie correspondent à des valeurs du caractère inférieures à  $M_e$ , ceux de la deuxième partie correspondent à des valeurs du caractère supérieures à  $M_e$ .

#### Exemple :

Au dernier devoir de mathématiques, les sept premières notes sont 8, 10, 9, 9, 12, 12 et 16

.On les classe en ordre croissant : 8 9 9 10 12 12

Il y a trois notes supérieures à 10 et trois inférieures

Soit la série de douze notes suivantes 8 9 9 9 10 10 | 12 12 12 16 16 17 On a :  $M_e = \frac{10 + 12}{2}$

Pour calculer une médiane on utilise les effectifs cumulés croissants ou décroissants

#### 5° Quartiles, déciles, centile : caractéristique de *dispersion*.

Les quartiles sont trois valeurs du caractère a, b et c qui partage la population en quatre parties de même effectif

1° partie : les modalités prises sont inférieures à a :

2° partie les modalités prises sont comprises entre a et b :

3° partie les modalités prises sont comprises entre b et c :

4° partie les modalités prises sont supérieures à c.

Les déciles partage la série en dix parties de même effectif.