# Séquence 2 : Statistiques

## 1. Généralités

### 1.1 Vocabulaires

#### 1.1.1 Population - Variable

Effectuer une étude statistique consiste à collecter, organiser et exploiter des informations sur un ensemble appelé population, délimité par une propriété commune.

Cette population est constituée d'individus ou unités statistiques, qui peuvent être des objets, des idées, des êtres vivants…

La propriété étudiée est appelée variable ou caractère.

Le caractère est qualitatif lorsque les valeurs prises ne sont pas des nombres, et quantitatif, lorsque les valeurs prises sont des nombres.

Un caractère quantitatif peut être discret si les valeurs prises sont isolées, ou continu s'il peut prendre toutes les valeurs possibles d'un intervalle.

#### 1.1.2 Effectifs – Fréquence- Classes

L'effectif total est le nombre d'individus de la population.

On note en général $x_1$, $x_2$, …, $x_n$ les valeurs prises par la variable étudiée et $n_i$ le nombre d'individus sur lesquels on a observé la valeur $x_i$. $n_i$ est appelé effectif de la valeur $x_i$ de la variable.

La série statistique ainsi définie se note ($x_i$ , $n_i$).

L'effectif total est alors $N = \sum_{i=1}^{n} n_i = n_1 + n_2 + \ldots + n_n$.

Le rapport $\frac{n_i}{N} = f_i$ est appelé fréquence de $x_i$

On a :

- Quel que soit i, $0 \leq f_i \leq 1$
- $$\sum_{i=1}^{n} f_i = f_1 + f_2 + .. + f_n = \sum_{i=1}^{n} \frac{n_i}{N} = \frac{\sum_{i=1}^{n} n_i}{N} = 1$$
- $100.f_i$ donne le pourcentage des individus ayant le caractère $x_i$.

### 1.1.3 Regroupement en classe

Lorsque le caractère est continu, on ne peut pas considérer chaque valeur séparément, on regroupe alors ces valeurs par classe.

De même lorsque l'effectif est assez important, il est plus commode de regrouper les valeurs par classe.

Exemple :

La population étudiée est l'ensemble des élèves d'une classe. Le caractère étudié est la note obtenue lors d'un certain examen.

Les notes obtenues sont :

12 12 14 5 8 8 9 16 15 7 6 10 10 12 9 9 10 7 6 10 11 9 7 9 11

Écrivons cette série de notes dans l'ordre croissant :

5 6 6 7 7 7 8 8 9 9 9 9 9 10 10 10 10 11 11 12 12 12 14 15 16

On voit que 1 élève a eu 5, deux ont eu 6, …. On peut réécrire cette série sous forme de tableau :

| Notes | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| effectifs | 1 | 2 | 3 | 2 | 5 | 4 | 2 | 3 | 0 | 1 | 1 | 1 |

### 1.1.4 Effectifs cumulés – Fréquences cumulées

Considérons une série à caractère quantitatif $x_i$. On ordonne les valeurs dans l'ordre croissant :

$x_1 < x_2 <\ldots< x_{k.}$

Si $n_i$ est l'effectif de la valeur $x_i$, on appelle effectif cumulé croissant jusqu'à la $i^e$ valeur le nombre : c'est le nombre des individus présentant une modalité inférieure à $x_i$ .

| Notes | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| effectifs | 1 | 2 | 3 | 2 | 5 | 4 | 2 | 3 | 0 | 1 | 1 | 1 |
| Effectif cumulé | 1 | 3 | 6 | 8 | 13 | 17 | 19 | 22 | 22 | 23 | 24 | 25 |

Ce tableau nous donne le nombre d'élèves qui ont eu une note inférieure à une note donnée. Par exemple, 6 élèves ont eu une note inférieure ou égale à 7, 13 n'ont pas eu la moyenne…

On définit de même

la fréquence cumulée croissante : $\sum_{j=1}^{i} f_j = f_1 + f_2 + \ldots + f_i$ où N est l'effectif total de la population

| Notes | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| effectifs | 1 | 2 | 3 | 2 | 5 | 4 | 2 | 3 | 0 | 1 | 1 | 1 |
| Fréquences | 0,04 | 0,08 | 0,12 | 0,08 | 0,2 | 0,16 | 0,08 | 0,12 | 0 | 0,04 | 0,04 | 0,04 |
| Fréquences cumulées | 0,04 | 0,12 | 0,24 | 0,32 | 0,52 | 0,68 | 0,76 | 0,88 | 0,88 | 0,92 | 0,96 | 1 |

## 2. Diagrammes

Un diagramme est une représentation graphique de la série. Il permet de visualiser ensemble les données statistiques.

### 2.1.1 Diagramme à bandes. Diagramme à bâtons

On porte en abscisses les valeurs de la variable x et en ordonnées les effectifs. Les effectifs sont représentés par des rectangles (bandes) verticales de longueurs proportionnelles aux effectifs. On peut remplacer les bandes par des segments : on obtient un diagramme en bâtons.

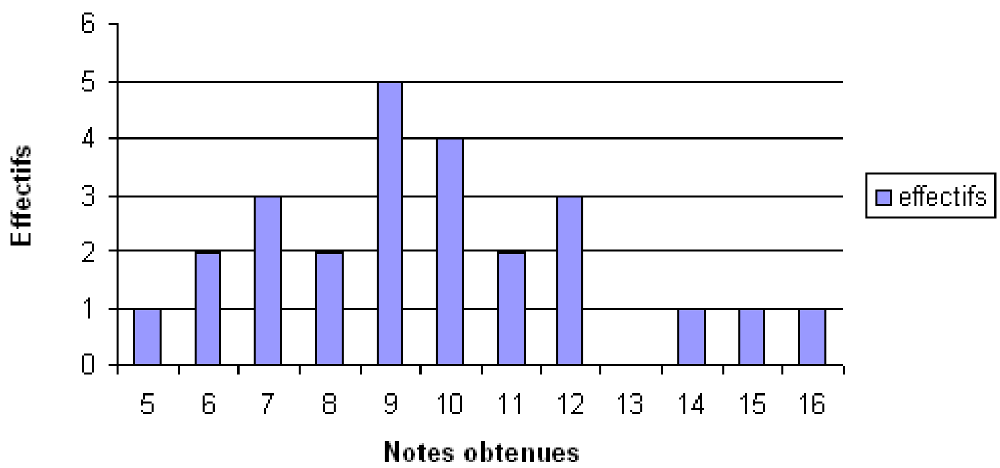



Diagramme à une seule bande.

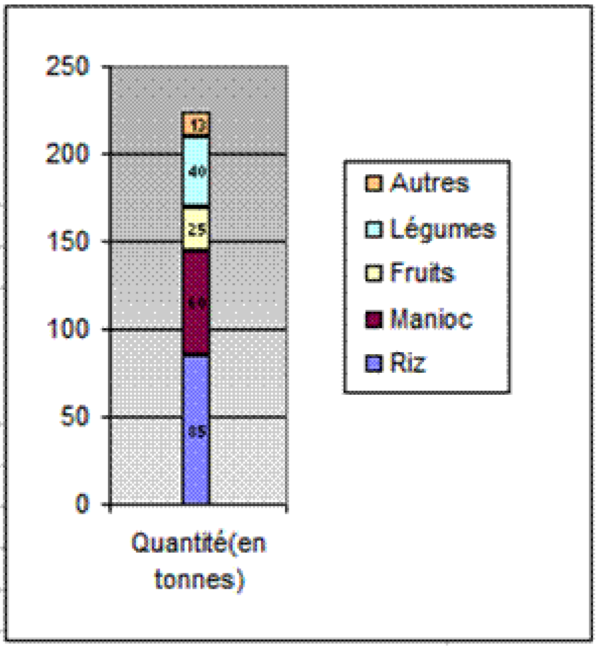
La longueur d'une bande est partagée proportionnellement aux effectifs ou aux fréquences.

Exemple : Voici la production agricole annuelle d'une certaine commune rurale .

| Produit | Riz | Manioc | Fruits | Légumes | Autres |
|---|---|---|---|---|---|
| Quantité (en tonne) | 85 | 60 | 25 | 40 | 13 |

### 2.1.2 Diagramme en secteur

C'est un diagramme de même type que le diagramme à une seule bande. Le disque est partagé en secteurs dont les angles sont proportionnels aux effectifs.



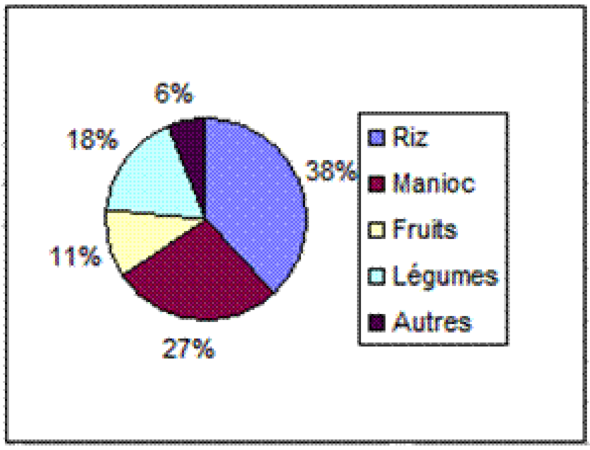




Diagramme à une seule bande.

Diagramme à secteur

### 2.1.3 Histogrammes

Cas d'une série continue ou série classée.

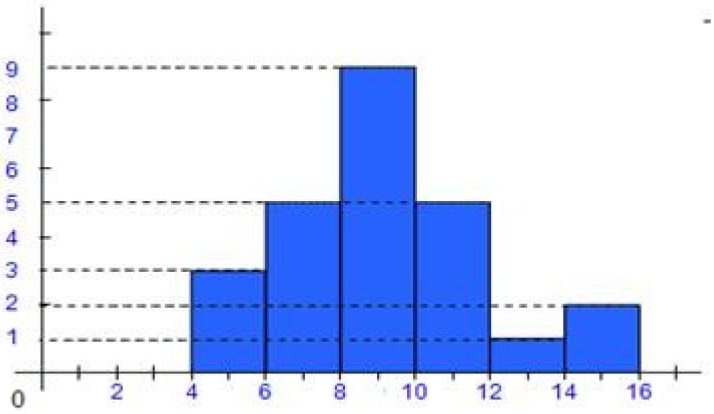
On porte en abscisses les valeurs de la variable x et en ordonnées les effectifs. L'effectif est représenté par un rectangle dont la base est égale à l'étendue de la classe et la hauteur proportionnelle à l'effectif.

Exemple

Dans l'exemple précédent, regroupons les notes en classes d'amplitude 2

On obtient le tableau suivant :

| Notes | [4 ; 6 ] | ]6 ; 8 ] | ] 8 ; 10 ] | ] 10 ; 12 ] | ] 12 ; 14 ] | ]14 ; 16 ] |
|---|---|---|---|---|---|---|
| Effectifs | 3 | 5 | 9 | 5 | 1 | 2 |




### 2.1.4 Polygones des effectifs

En reliant les extrémités des bâtons, on obtient le polygone des effectifs.

Dans le cas des histogrammes, on prend les centres des classes.

# 3. Caractéristiques d'une série statistique

Un caractéristique est une grandeur qu'on utilise pour résumer une série statistique.

On distingue deux sortes de caractéristique : caractéristiques de position et caractéristiques de dispersion.

### 3.1.1 Caractéristiques de positions

**a) Le mode :**

Le mode ( ou dominante) est la valeur la plus fréquente de la variable. C'est la variable qui a le plus grand effectif.

Le mode est défini même si la variable est qualitative.

Pour une série classée, dont les classes sont d'égal effectif, la classe modale est la classe qui correspond au plus grand effectif.

Si une série peut posséder un seul mode on dit qu'elle est unimodale. Si elle en possède plusieurs, on dit qu'elle est plurimodale.

**b) La moyenne :**

La moyenne arithmétique d'une série statistique est égale à la somme des valeurs du caractère divisées par leur nombre.

i- Cas des données énumérées : $\bar{x} = \dfrac{x_1 + x_2 + ... + x_k}{N} = \dfrac{1}{N}\sum_{i=1}^{k} x_i$

ii- Si la série est donnée par sa distribution d'effectifs, les valeurs $x_1$, $x_2$, …, $x_n$ ayant respectivement pour effectifs $n_1$, $n_2$, ..,$n_k$ , alors :

$$\bar{x} = \frac{n_1x_1+n_2x_2+\ldots+n_kx_k}{N} = \frac{\sum_{i=1}^{k} n_ix_i}{\sum_{i=1}^{k} n_i}$$

iii- Cas où les valeurs sont regroupées en classes : les $n_i$ valeurs de la i-ème classe sont supposées groupées au centre $x_i$ de la classe. On revient ainsi au cas précédent.

Remarque :

$$\bar{x} = \frac{n_1x_1+n_2x_2+\ldots+n_kx_k}{N} = \frac{n_1}{N}x_1 + \frac{n_2}{N}x_2 + \ldots + \frac{n_k}{N}x_k$$

Donc si $f_i$ est la fréquence de la variable $x_i$ alors :

$$\bar{x} = \sum_{i=1}^{k} f_ix_i$$

**c) Propriétés**

Soit une série statistique sur une population et une partition de cette population en deux sous-populations d'effectifs respectifs $n_1$ et $n_2$. Si $m_1$ et $m_2$ sont les moyennes respectifs des deux sous-populations, alors la moyenne de la population est $\bar{x} = \dfrac{n_1m_1+n_2m_2}{n_1+n_2}$

Si $\bar{x}$ est la moyenne de la série $(x_i, n_i)$,

alors la moyenne de la série $(x_i - a, n_i)$ est $\bar{x} - a$

la moyenne de la série $(hx_i, n_i)$ est $h.\bar{x}$

Donc , si on pose $y_i = ax_i + b$ alors la moyenne de la série $(y_i, n_i)$ est $\bar{y} = a.\bar{x} + b$

### 3.1.2 Caractéristique de dispersion.

Une caractéristique de dispersion est utilisée pour évaluer la dispersion d'une série. On utilise le plus souvent la variance et l'écart type .

**a) Variance. Écart type**

($\iota$)–Définition

La variance d'une série est la moyenne des carrés des écarts à la moyenne.

$$V = n_1(x - x_1)^2 + n_2(x - x_2)^2 + \ldots + n_k(x - x_k)^2 = \frac{\sum_{i=1}^{k} n_i (x - x_i)^2}{\sum_{i=1}^{k} n_i}$$

L'écart type d'une série est la moyenne quadratique des écarts à la moyenne. C'est la racine carrée de la variance

$$\sigma = \sqrt{V}$$

(ii)-Remarques

$R_1$. Plus la variance est grande, plus la série est dispersée.

$R_2$. Plus la variance est petite (voisin de 0), pus la série est resserrée autour de la moyenne.

$R_3$. La variance est une quantité positive ou nulle.

(iii)-Méthode de calcul

Même avec des valeurs observées $x_i$ très simples, il arrive souvent que la moyenne $\overline{x}$ soit un nombre décimal. Dans ce cas, le calcul de la variance V nécessite des calculs fastidieux.

La formule de Koenig : $V = \frac{\sum_{i=1}^{k} n_i x_i^2}{N} - (\overline{x})^2$ permet de simplifier les calculs.