

Chapitre 1

Notions de base et vocabulaire statistique

1.1 Concepts de base de la statistique

Introduction : La statistique descriptive est un ensemble des méthodes scientifiques qui permettent de décrire, de rassembler, de présenter et d'analyser les données d'une enquête, d'une étude et d'une expérience ... etc. L'objectif de la statistique descriptive est décrire les données de manière efficace, c'est-à-dire permettant de mieux visualiser, et comprendre les données.

1.1.1 Notions de population, d'échantillon, variables et modalités

La population désigne un ensemble d'unités statistiques. Les unités statistiques, aussi appelées individus, sont les entités abstraites qui représentent des personnes, des animaux ou des objets. La statistique sert à décrire l'ensemble des unités statistiques qui composent la population

Définition 1.1.1 *La population statistique est l'ensemble étudié, elle est notée Ω .*

les éléments de cette population sont appelés individus ou unités statistiques.

La taille de la population est le nombre de ses individus i.e., $\text{card}(\Omega)$

Un échantillon est un sous ensemble de la population considérée.

Exemple 1.1.2 *Les étudiants de première année Informatique de la faculté MI forment une population statistique, l'unité statistique est l'étudiant, un groupe de TD est un échantillon.*

Définition 1.1.3 *Une variable statistique(Caractère) est une caractéristique d'individu qui peut prendre différentes valeurs pour différents d'individus, elle est notée $X, Y \dots$ etc.*

Exemple 1.1.4 *Une étude sur les étudiants de première année Informatique dans la faculté MI peut porter sur les différentes variables : leur âges, leur sexes, leur nationalités, leur moyennes, ... etc.*

Définition 1.1.5 *Les modalités sont les différentes valeurs d'une variable peut prendre.*

Exemple 1.1.6 *Pour la variable «nationalité » les modalités sont par exemple algérienne, tunisienne, mauritanienne, ... etc.*

Pour la variable « note de TD » les modalités sont nombres dans $[0, 20]$.

Pour la variable "couleur des yeux" les modalités sont : noir, bleu, marron ou vert.

Remarque 1.1.7 *Pour définir une variable, il faut, après avoir indiqué sur quelle population on travaille, préciser l'ensemble des modalités de la variable.*

1.1.2 Différents types de variables statistiques

il existe deux types de variables statistiques

Variables quantitatives

Une variable statistique est dite quantitative lorsque ses modalités sont mesurables (numériques), elle s'exprime par des nombres réels, il ya deux types.

a) **Variable discrète** : les modalités sont dénombrables

Exemple 1.1.8 1) *Le nombre d'enfants par famille.*

Modalités : 0,1,2,3,4...

Type : discrète.

2) *Le salaire d'employés d'une usine.*

Modalités : 20000 DA , 25000 DA...

Type : discrète.

b) **Variable continue** : les modalités sont définies sur un intervalle continue.

Exemple 1.1.9 *Le poids, la taille. sont des variables statistiques continues.*

Définition 1.1.10 *Soit $[a_i, b_i[$ une classe d'une variable quantitative continue , L'amplitude de $[a_i, b_i[$ est $b_i - a_i$ et $c_i = \frac{a_i+b_i}{2}$ est le centre de la classe $[a_i, b_i[$.*

Variables qualitatives

Une variable statistique est dite qualitative lorsque ses modalités ne sont pas mesurables (non numériques), il ya deux types :

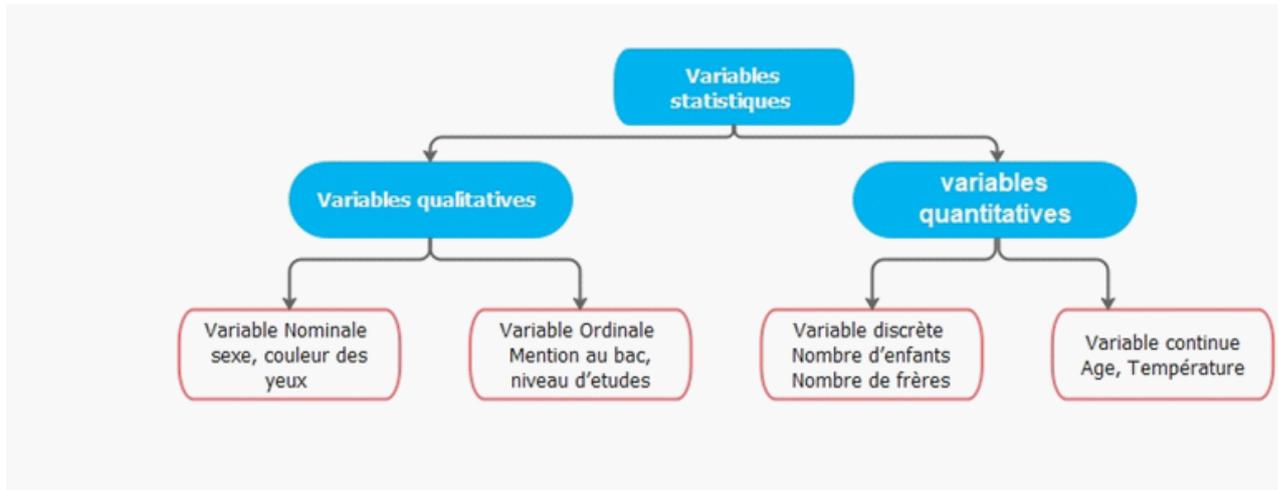
a) **Variable nominale** : Les modalités ne sont pas ordonnées.

b) **Variable ordinale** : Les modalités sont ordonnées.

Exemple 1.1.11 1) *Les activités sportives sont des variables nominales.*

2) *Qualité d'un métal, (très résistant, assez résistant et peu résistant), Mention au bac, (passable, assez bien, bien,...) sont des variables ordinales.*

Remarque 1.1.12 *On peut éclaircir les variables statistiques indiquées ci-dessus par le diagramme suivant :*



1.2 Série statistique d'une seule variable

On appelle série statistique de la variables X , la suite des valeurs prises par cette variable sur les individus. Les valeurs de la variable X sont notées x_1, \dots, x_k . Une série statistique se définit par une population et un caractère.

1.2.1 Effectif, Fréquence et Pourcentage

Définition 1.2.1 L'effectif total noté N est le nombre d'individus qui compose la population i.e., $\text{card}(\Omega) = N$.

Définition 1.2.2 L'effectif d'une valeur x_i noté n_i est le nombre d'individus associé à cette valeur.

Définition 1.2.3 La fréquence d'une valeur x_i notée f_i est définie par : $f_i = \frac{n_i}{N}$. Elle peut aussi s'exprimer en pourcentage d'un caractère $p_i = f_i \times 100$.

Remarque 1.2.4 1) La somme des effectifs est toujours égale au nombre d'individus de la population i.e., $\sum_{i=1}^n n_i = N$.

2) La somme des fréquences est toujours égale à 1. i.e., $\sum_{i=1}^n f_i = 1$, où $0 \leq f_i \leq 1$.

Exemple 1.2.5 Les résultats en Analyse 1 d'un groupe de 14 étudiants étaient :

9, 9, 6, 4, 6, 9, 4, 10, 12, 14, 12, 9, 14, 9

On ordonne cette série : 4, 4, 6, 6, 9, 9, 9, 9, 9, 10, 12, 12, 14, 14. On a $N = 14$ et par exemple : effectif (6) = 2, fréquence (6) = $\frac{2}{14} = \frac{1}{7} = 0,14$.

Pourcentage (6) = $0,14 \times 100 = 14\%$.

Effectifs (Fréquences) cumulés

Définition 1.2.6 L'effectif cumulé noté EC : Quand les valeurs d'un caractère sont ordonnées, on peut cumuler les effectifs de façon croissante ou décroissante.

1) L'effectif cumulé croissant noté ECC d'une valeur(ou d'une classe) est la somme des effectifs de cette valeur(ou de cette classe) et des effectifs(ou des classes) précédentes.

2) L'effectif cumulé décroissant noté ECD d'une valeur (ou d'une classe) est la somme des effectifs de cette valeur (ou de cette classe) et des effectifs(ou des classes) suivantes

3) La fréquence cumulée croissante notée FCC d'une valeur (ou d'une classe) est la somme des fréquences de cette valeur (ou cette classe) et des fréquences (ou des classes) précédentes.

4) La fréquence cumulée décroissante notée FCD d'une valeur (ou d'une classe) est la somme des fréquences (ou des classes) suivantes.

1.3 Représentation des données

1.3.1 Tableaux statistiques

Un tableau statistique est un moyen de regrouper, organiser et de ranger par ordre croissant (ou décroissant) les données d'une série statistique pour les bien représenter.

Le choix du type de tableau dépend de la nature des variables.

Exemple 1.3.1 Répartition des enfants d'après leur l'âge de la façon suivante.

| Age | [3, 4[| [4, 5[| [5, 6[| [6, 7[| Total |
|-----------------|-----------------|-----------------|-----------------|----------------|-------|
| c_i | 3.5 | 4.5 | 5.5 | 6.5 | 20 |
| effectif n_i | 19 | 7 | 1 | 3 | 30 |
| ECC | 19 | 26 | 27 | 30 | |
| ECD | 30 | 11 | 4 | 3 | |
| Fréquence f_i | $\frac{19}{30}$ | $\frac{7}{30}$ | $\frac{1}{30}$ | $\frac{3}{30}$ | 1 |
| FCC | $\frac{19}{30}$ | $\frac{26}{30}$ | $\frac{27}{30}$ | 1 | |
| FCD | 1 | $\frac{11}{30}$ | $\frac{4}{30}$ | $\frac{3}{30}$ | |

Sachant que le centre d'une classe $[a_i, b_i[$ est $c_i = \frac{a_i+b_i}{2}$.

1) Quel est type de variable statistique étudiée

1) Compléter le tableau ci-dessus.

1.3.2 Représentations graphiques

Ces représentations permettent de résumer visuellement des séries statistiques. Le choix du type de graphe dépend de la nature des variables statistiques.

Cas d'une variable statistique qualitative

Dans ce cas, il ya deux types de présentation graphique

1) **Diagramme Circulaire (ou Camembert)**. Le principe du graphe dans ce cas consiste à diviser un disque circulaire en secteurs, où chaque secteur représente une modalité dont la surface est proportionnelle à la fréquence de cette modalité. Si x_i une modalité d'effectif n_i l'angle central α_i (degré de la modalité i) correspondant se calcule comme suit :

$$\begin{cases} N & \longrightarrow 360^\circ \\ n_i & \longrightarrow \alpha_i \end{cases} \implies \alpha_i = \frac{n_i \times 360}{N}.$$

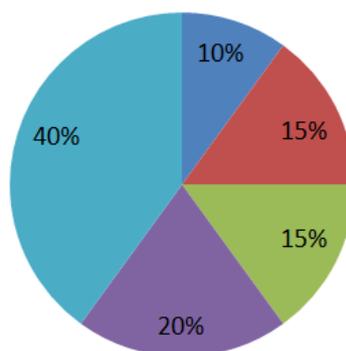
Exemple 1.3.2 Soit le tableau statistique suivant : Distribution de 200 employés selon la nationalité

| Modalité x_i | Allemande | Russe | Chinoise | Française | Alérienne | Total |
|------------------|------------|------------|------------|------------|-------------|-------------|
| effectif n_i | 20 | 30 | 30 | 40 | 80 | 200 |
| Fréquence f_i | 0.10 | 0.15 | 0.15 | 0.2 | 0.40 | 1 |
| Angle α_i | 36° | 54° | 54° | 72° | 144° | 360° |

- 1) Compléter le tableau suivant
- 2) Représenter cette série par un diagramme de (circulaire) camembert.

Distribution des employés selon la nationalité

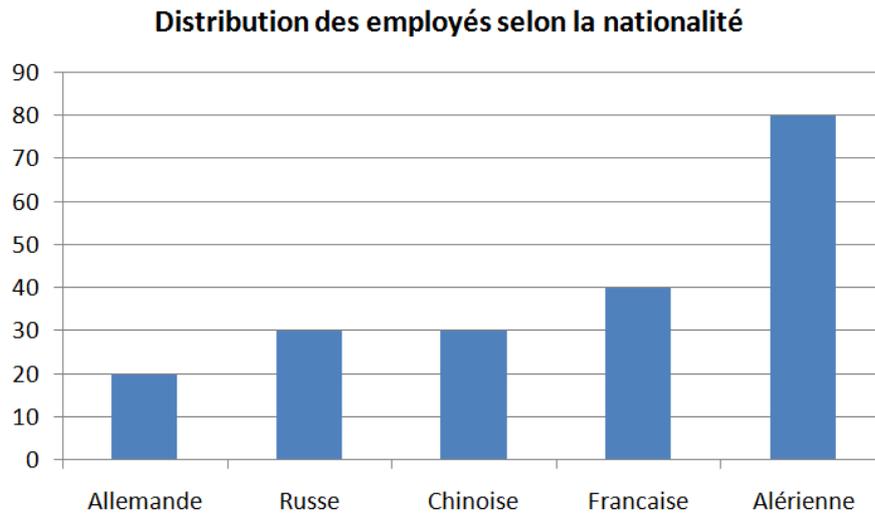
■ Allemande ■ Russe ■ Chinoise ■ Française ■ Alérienne



2) **Diagramme à bande (tuyaux d'orgue)** Dans ce cas le graphe se fait à l'aide des rectangles de même largeur, Les modalités sur l'axe horizontal, les effectifs sur l'axe vertical où à chaque modalité correspond une barre, les hauteurs des barres sont proportionnelles aux effectifs représentant les modalités.

Exemple 1.3.3 Refaire l'exemple précédent en utilisant le diagramme à bande.

| Modalité x_i | Allemande | Russe | Chinoise | Française | Alérienne | Total |
|----------------|-----------|-------|----------|-----------|-----------|-------|
| effectif n_i | 20 | 30 | 30 | 40 | 80 | 200 |



Cas d'une variable statistique quantitative Dans ce cas, il ya deux types de présentation graphique

1) Cas d'une variable quantitative discrète.

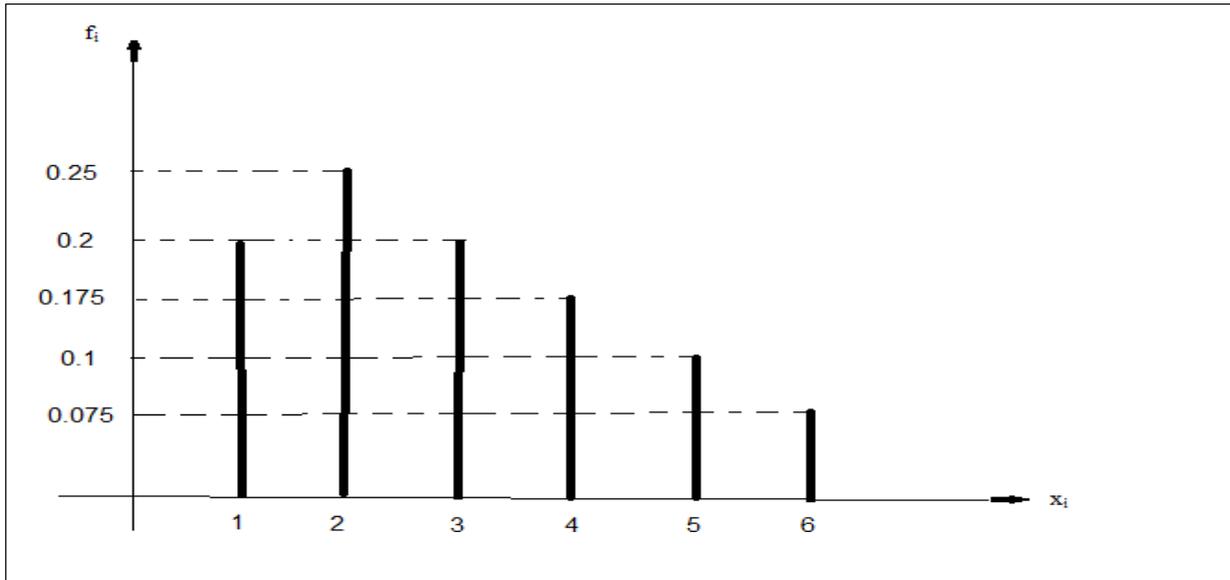
i) **Diagramme en bâton (en barre) :**

Dans ce cas l'axe vertical représente les effectifs (les fréquences) et l'axe horizontal représente les valeurs de la variable. A chaque valeur on associé un segment (bâton) dont sa hauteur est proportionnelle à l'effectif (ou la fréquence) de cette modalité.

Exemple 1.3.4 Soit la tableau statistique donnant le nombre de pièces par logement :

| | | | | | | | |
|------------------------------------|-----|------|-----|-------|-----|-------|----------|
| <i>Modalité x_i</i> | 1 | 2 | 3 | 4 | 5 | 6 | |
| <i>effectif n_i</i> | 8 | 10 | 8 | 7 | 4 | 3 | $N = 40$ |
| <i>fréquences f_i</i> | 0.2 | 0.25 | 0.2 | 0.175 | 0.1 | 0.075 | 1 |

Représenter ces données par un diagramme en bâton



ii) **Courbe des effectifs cumulés croissants (ou décroissants) :**

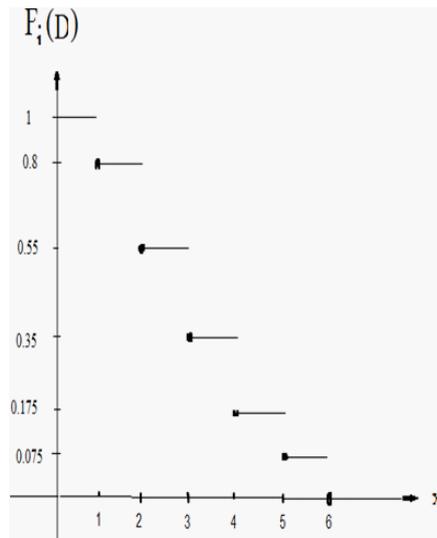
Pour Tracer cette courbe : On complète le tableau en calculant ECC (ECD) ou f_iCC (f_iCD). On trace dans un repère le point dont l'abscisse est la modalité et l'ordonnée est l'effectif cumulé qui lui correspond. On trace les segments qui relient ces points.

Exemple 1.3.5 Compléter le tableau ci-dessus

| | | | | | | | |
|--------------------|-----|------|------|-------|-------|-------|----------|
| Modalité x_i | 1 | 2 | 3 | 4 | 5 | 6 | |
| effectif n_i | 8 | 10 | 8 | 7 | 4 | 3 | $N = 40$ |
| fréquences f_i | 0.2 | 0.25 | 0.2 | 0.175 | 0.1 | 0.075 | 1 |
| fréquences f_iCC | 0.2 | 0.45 | 0.65 | 0.825 | 0.925 | 1 | |
| fréquences f_iCD | 1 | 0.8 | 0.55 | 0.35 | 0.175 | 0.075 | |

Tracer la courbe de fréquences f_iCD (le diagramme en escalier).

courbe cumulative fréquences décroissants f_i^CD

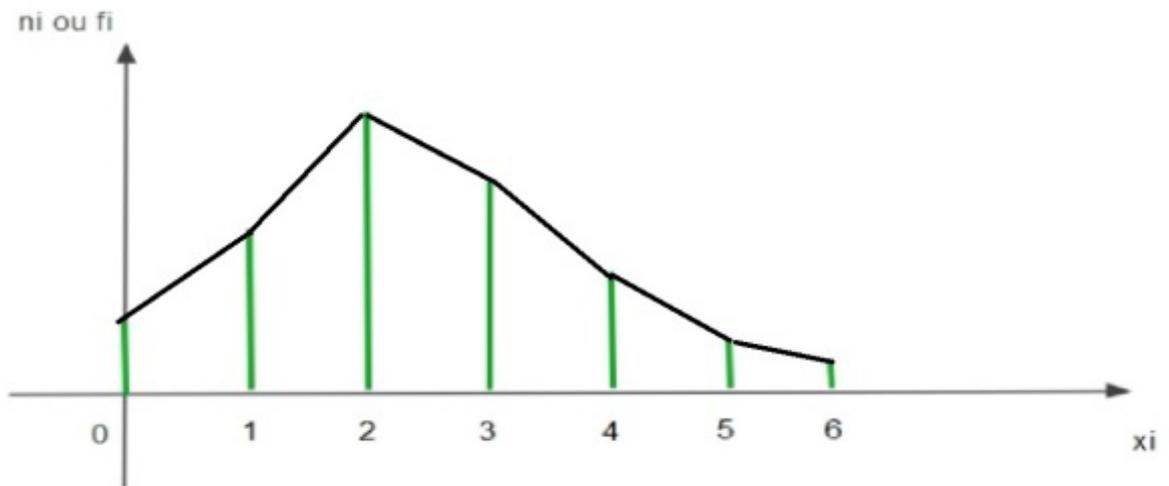


3i) **Polygone des effectifs(ou des fréquences)** : Le polygone des effectifs(ou des fréquences) est obtenu en reliant les extrémités des bâton.

Exemple 1.3.6 Une enquête réalisée dans un village porte sur le nombre d'enfants acharge par famille.

| Modalité x_i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | totale |
|----------------|----|----|----|----|----|---|---|--------|
| effectif n_i | 18 | 32 | 66 | 41 | 32 | 9 | 2 | 200 |

- 1) Représenter ces données par un diagramme en bâton
- 2) Représenter ces données par un polygone des effectifs



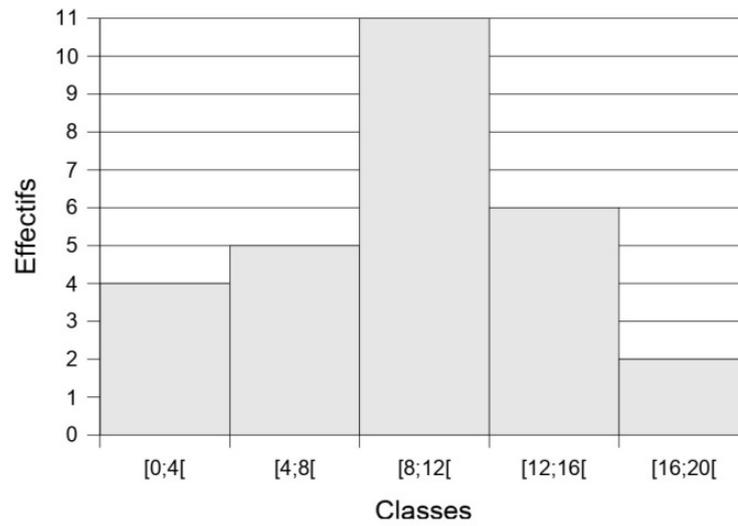
2) Cas d'une variable quantitative continue.

i) **Histogrammes** : Dans ce cas, on représente les données par des rectangles juxtaposés où la largeur de chaque rectangle correspond à l'intervalle de la classe correspondante, sa hauteur est telle que l'aire du rectangle est proportionnelle à l'effectif de la classe.

Exemple 1.3.7 Soit le tableau suivant

| | | | | | |
|----------------|--------|--------|---------|----------|----------|
| Age | [0, 4[| [4, 8[| [8, 12[| [12, 16[| [16, 20[|
| effectif n_i | 4 | 5 | 11 | 6 | 2 |

Repésenter ces données par un histogramme.



ii) **Polygone de fréquence** : Dans ce cas ,on joigne les milieux des cotés supérieurs des rectangles dans un histogramme, on trouve une ligne brisée qui est le polygone de fréquence.