

Partie I : Généralités sur la statistique descriptive

Introduction : La statistique est la collecte de données, leur analyse, leur traitement, l'interprétation des résultats et leur présentation afin de les rendre compréhensibles par tous. C'est à la fois une science, une méthode et un ensemble de techniques.

I. Vocabulaire

Les statistiques consistent en diverses méthodes de classement de données telles que les tableaux et les graphiques permettant d'organiser un grand nombre de données. Les statistiques se sont développées dans la deuxième moitié du dix-neuvième siècle dans le domaine des sciences humaines (sociologie, économie, anthropologie, ...). Elles sont dotées d'un vocabulaire particulier :

1.1 Population statistique : Ensemble faisant l'objet d'une étude statistique. Ça peut être des personnes, des animaux, des objets, ... cet ensemble est noté Ω .

Exemple : On considère l'ensemble des étudiants de la section A (MI). On s'intéresse au nombre de frères et sœurs de chaque étudiants. Dans ce cas

La population statistique notée Ω = Ensemble des étudiants

1.2 Individu (unité statistique) : Une population est constitué d'éléments, chaque élément est appelé individu ou unité statistique. Autrement dit, un individu est un élément de la population Ω , il est noté ω

Exemple : Dans l'exemple précédent l'individu (unité statistique) = un étudiant

1.3 Caractère (variable statistique) : C'est ce qui est mesuré ou observé sur les individus. Le caractère est toute application

$$X: \Omega \rightarrow C$$

L'ensemble C est dit ensemble des valeurs du caractère X

Exemple : taille, température, couleur des yeux, ...

Remarque : Soit Ω un ensemble. On appelle et on note $\text{card } \Omega$ le nombre d'éléments de Ω . Ça représente en fait la taille de la population statistique.

1.4 Modalités d'un caractère: Différentes valeurs que peut prendre la variable statistique

Exemple : Variable est « situation familiale »

Modalités : célibataire, marié, divorcé, veuf

II. Types de caractères : Nous distinguons deux types de caractères : caractère qualitatif et caractère quantitatif

II.1 Caractère qualitatif : Les caractères qualitatifs sont des caractères dont les modalités ne sont pas mesurables (ordonnées)

Définition : Les éléments de C sont représentés par autre chose que des chiffres.

Exemple : l'état d'une maison : on peut considérer les modalités suivantes :

- Ancienne
- Dégradée
- Nouvelle
- Rénovée

II. Caractère quantitatif : Les caractères quantitatifs sont des caractères dont les modalités sont mesurables (ordonnées)

Exemple : la taille, l'âge, le salaire, ...

Définition : L'ensemble des valeurs est représenté par des chiffres. De même, il est partagé en deux sortes de caractères discret et continu

Exemple : le salaire des employés d'une usine

Modalités : 10 000 da, 20 000 da,

Type : discret

La taille des écoliers

Modalités : [70 – 110[

Type : continu

En général, la variable quantitative discrète est une variable ne prenant que des valeurs entières. Le nombre de valeurs distinctes d'une telle variable est habituellement assez faible.

Exemple : le nombre de maisons par quartier d'une ville

Une variable quantitative est dite continue lorsque les observations qui lui sont associés ne sont pas des valeurs précises, mais des intervalles. C'est le cas lorsque nous avons un grand nombre d'observations distinctes.

III. Organisation des données : On regroupe les données de la statistique dans un tableau indiquant la répartition des individus selon le caractère étudié. Le regroupement s'effectue par classe :

- Si le caractère est qualitatif, une classe contient tous les individus ayant la même modalité ou la même valeur du caractère.
- Si le caractère est continu, une classe est un intervalle. Dans ce cas, pour construire ses intervalles on doit appliquer les étapes suivantes :
 1. On détermine l'amplitude (longueur) des classes par la formule suivante :

$$a \geq \frac{e}{k} \quad \text{avec :} \quad e = X_{\max} - X_{\min} \text{ appelé étendue de la série et } k \approx \sqrt{N},$$

N : nombre d'observation

On pourra aussi utiliser d'autres formules à savoir : $n_c = 2.5\sqrt[4]{n}$ (formule de Yule) ou bien

$$a = \frac{e}{n_c} \text{ avec } n_c = 1 + 3.32\log_{10}(n) \text{ et } e : \text{l'étendue (formule de Sturge)}$$

2. Chaque fois que cela est possible, les amplitudes de classes sont égales

3. chaque classe contient sa borne inférieure mais pas sa borne supérieure

Remarque : dans les calculs, une classe est représentée par son centre qui est le milieu de l'intervalle soit $X_c = \frac{e_i + e_{i+1}}{2}$

Exemple : l'étude du poids de 33 étudiants a donné les résultats suivants :

61, 64, 64, 63, 65, 64, 65, 65, 65, 65, 63, 65, 58, 64, 61, 63, 60, 63, 63, 61, 55, 54, 58, 63, 54, 61, 58, 61, 58, 63, 64, 58, 58

Répartir ses données en classes

Calculons l'étendue de la série : $e = 65 - 54 = 11$ et $k \approx \sqrt{33} = 5.74 \approx 6$

$$a \geq \frac{e}{k} = \frac{11}{6} = 1.83 \approx 2$$

<i>Poids</i>	n_i
$[54 - 56[$	2
$[56 - 58[$	0
$[58 - 60[$	6
$[60 - 62[$	6
$[62 - 64[$	7
$[64 - 66[$	11
<i>Total</i>	33