

Université A- mira de Bejaia
Faculté des sciences économiques, commerciales et des sciences de gestion
Département SEGC(LMD)
Module stat I
Enseignant : Dr. Mousli
Chapitre 1 : Vocabulaire de la statistique descriptive

Introduction

Pour avoir une idée de la place et de l'importance que revêt la statistique dans la vie de tous les jours, Il suffit d'allumer son ordinateur ou d'écouter les informations à la radio pour constater que les statistiques sont partout. Ceci révèle que le monde moderne est presque complètement tourné vers le quantitatif et le mesurable. D'où l'intérêt de la statistique, discipline relativement récente, mais qui correspond parfaitement à cette orientation du monde moderne.

1.1. La collecte de l'information

La collecte de l'information chiffrée remonte à l'antiquité. Le besoin d'information sur la population et sa richesse est devenu une nécessité pour la pérennité des États. Pour recueillir des informations sur une population statistique, l'on dispose de deux méthodes :

1.1.1. La méthode exhaustive (recensement)

La collecte statistique porte sur la totalité de la population, où chaque individu de la population est étudié selon le ou les caractères étudiés.

Exemple :

L'Algérie organise tous les 10 ans le recensement général de la population et de l'habitat (RGPH). Elle a réalisé depuis son indépendance six recensements : en 1966, 1977, 1987, 1998, 2008 et 2022. On y trouve des informations concernant la population totale (sa répartition spatiale par sexes, par âges, par emplois, par niveau d'instruction, la croissance démographique, l'espérance de vie etc.) et l'habitat (consistance du parc national de logements, le nombre d'occupants par logement, l'âge et le degré de vétusté des logements, l'habitat informel etc.)

Le RGPH a une utilité très importante, car il fournit des informations précieuses aux autorités publiques, mais aussi, aux universitaires, aux acteurs économiques, sociaux et culturels. Cependant, le recensement est une opération lourde nécessitant d'importants moyens humains, financiers et matériels pour sa réalisation.

En raison du manque d'argent ou du temps, il convient dans cette situation de faire porter l'étude sur un ensemble plus petit de la population.

1.2.2. La méthode des sondages (échantillonnage)

Le sondage est technique de collecte non exhaustive qui conduit à n'examiner qu'une fraction de la population (un échantillon). Cet échantillon tiré au hasard permet de connaître la population mère avec une certaine précision. On peut avoir recours à un échantillonnage pour plusieurs raisons :

- On ne peut pas observer qu'une partie de la population lorsque celle-ci est infinie
- Le choix d'un échantillon est de moindre coût qu'un recensement.
- Impossibilité de faire autrement (c'est le cas où observer un individu revient à le détruire).

Exemple :

On ne peut pas tester la durée de vie de toutes les ampoules, mais on peut tester un petit échantillon.

Pour choisir un échantillon dans une population, on distingue plusieurs méthodes :

1.2.2.1. Echantillonnage aléatoire

- **Méthodes élémentaires**

Dans une population de taille N où à chaque individu a une probabilité $\frac{1}{N}$ d'être choisi, on procède à un tirage au hasard de n individus (à l'aide d'une table de nombre au hasard par exemple). On distingue deux types de tirage : le tirage exhaustif (sans remise) et le tirage non exhaustif (avec remise).

- **Tirage systématique**

Un premier individu étant tiré au hasard, la méthode consiste à tirer un individu tous les k individus rencontrés. Par exemple, si on dispose d'une liste de 250 individus et qu'on veuille choisir au $(1/10)^{\text{ème}}$, on tire le premier individu parmi les 10 premiers. Supposons que c'est le 7^{ème}. On tire ensuite le 17, puis le 27^{ème}, etc.

- **Echantillonnage stratifié**

On tire les individus de l'échantillon dans des groupes homogènes de la population que nous appelons strates. On cherche le nombre optimal d'individus à tirer dans chaque strate. Cette méthode a pour but d'améliorer la précision de l'estimation.

- **Echantillonnage par grappes**

Un tirage au hasard ayant donné des grappes (groupes d'individus), on examine tous les individus de chaque grappe tirée. Cette méthode est utile si on ne dispose pas de liste de tous les individus de la population.

- **Echantillonnage par degrés**

C'est un tirage où l'échantillonnage n'a pas lieu directement parmi les individus de la population mais en plusieurs temps (plusieurs degrés) : c'est un tirage en cascade.

Exemple :

Si on s'intéresse à une population d'écoliers, on tire au hasard en premier lieu des villes, puis dans les villes tirées, on tire au hasard des écoles, puis enfin des écoliers.

- **Echantillonnage à probabilité inégales**

On procède au tirage d'individu avec des probabilités proportionnelles à un caractère connu.

Exemple :

Si on s'intéresse à la taille des étudiants par sexe dans une université, et on sait a priori que le nombre de garçons est supérieur à celui de fille. Dans ce cas-là, le nombre de garçons tiré dans l'échantillon sera supérieur à celui des filles.

1.2.2.2. Echantillonnages empiriques

L'échantillonnage empirique est caractérisé par la construction de l'échantillon par un choix « raisonné ». Cet échantillon ressemble le plus possible à la population, mais cette ressemblance est due aux connaissances préalables qu'on a sur cette population. On ne peut calculer la précision de cet échantillonnage. Un exemple d'échantillonnage empirique est fourni par ce qu'on appelle la méthode des quotas. Cette méthode se base sur la construction d'un échantillon de taille n dans lequel les proportions des individus selon différents critères sont égales à celle de la population. Ces proportions dans la population sont connues (par un recensement par exemple). Une fois ces quotas déterminés, on est libre quant au choix de l'échantillon (à condition de respecter les quotas).

1.2.3. Les sources statistiques

La provenance ou l'origine de l'information statistique est appelée la source statistique. Elle est essentielle car elle permet à celui qui souhaite de vérifier l'exactitude des informations ou de les détailler selon ses préoccupations de recherche.

1.3. Différence entre la statistique et les statistiques

1.3.1. La statistique

La statistique est un outil (méthode) scientifique qui permet de recueillir, d'organiser, de résumer et de présenter des informations (données) qualitatives ou quantitatives en vue de les analyser pour en prendre des décisions pertinentes.

1.3.2. Les statistiques

Les statistiques (et on dit une statistique) représentent un ensemble de données ou de renseignements numériques provenant des recensements de population, des données de registres d'état civil et d'enquêtes appropriées. Les statistiques économiques sont par exemple, celles relatives aux prix, aux chiffres d'affaires d'une entreprise, aux taux de croissance de l'économie, etc.

1.4. Les concepts de base de la statistique descriptive

1.4.1. La population statistique

On appelle population statistique l'ensemble des individus sur lesquels porte l'étude statistique.

1.4.2. L'échantillon : un échantillon est une partie ou un sous ensemble d'individus représentatifs d'une population.

Exemple1 : Les étudiants de l'université de Bejaia.

La population statistique : l'ensemble des étudiants de l'université de Bejaia

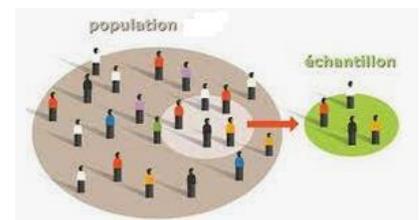
Exemple2 : Les voitures fabriquées par une usine.

La population statistique : l'ensemble des voitures.

1.4.3. L'individu (unité statistique)

On appelle individu chaque élément de la population ou de l'échantillon. On utilise également le terme **unité statistique** pour désigner un individu.

Exemple : Dans l'exemple précédent portant sur les étudiants de l'université de Bejaia, chaque **étudiant** de l'université représente un **individu** ou une **unité statistique**.



1.4.4. Le caractère ou la variable

Le caractère est l'aspect particulier que l'on veut étudier. Pour décrire les individus qui composent une population ou un échantillon, nous pouvons nous intéresser à leur âge, leur sexe, leur taille, leur religion, leur nationalité ou encore leur opinion sur un sujet donné. **Le caractère peut être quantitatif ou qualitatif**

A) Le caractère quantitatif

On dit d'un caractère qu'il est quantitatif si ses modalités sont mesurables c'est-à-dire elles s'expriment par un nombre.

En raison de la nature des valeurs que peut prendre un caractère quantitatif dans une série statistique, nous distinguons deux catégories : le caractère quantitatif discret (discontinu) et le caractère quantitatif continu.

- **Le caractère quantitatif discret (discontinu)**

Un caractère quantitatif est discontinu si l'ensemble des valeurs qu'il peut prendre est fini ou dénombrable. Lorsque ces valeurs sont entières, on dira que le caractère est discret.

Exemple :

Le nombre d'enfants par famille, le nombre de voiture, le nombre d'étudiants etc. sont des caractères quantitatifs discrets.

- **Le caractère quantitatif continu**

Un caractère quantitatif est continu s'il peut prendre n'importe quelle valeur dans un intervalle donné de nombres réels.

Exemple :

Le poids, la taille et l'âge d'un groupe d'individu sont des caractères quantitatifs continus.

B) Le caractère qualitatif

Un caractère est dit qualitatif lorsque son observation ne peut être traduite par une mesure (n'est pas mesurable), ses diverses modalités sont simplement constatées, repérées par un mot traduisant un état.

Exemple :

La religion, le sexe et le lieu de résidence sont des caractères qualitatifs.

Deux types d'échelle se rapportent aux caractères qualitatifs : les échelles à variation **nominale** et **ordinaire** :

- **Le caractère qualitatif nominal**

Il décrit un nom ou une catégorie qui ne suit pas un ordre naturel.

Exemple :

Le mode de transport utilisé par les habitants d'une ville :

Voiture personnelle, Transport en commun, A pieds, Bicyclette, Autres moyens.

- **Le caractère qualitatif ordinal**

Un caractère qualitatif est ordinal si ses modalités peuvent être classées dans un ordre spécifique ou dans un ordre naturel quelconque.

Exemple :

Classement de l'état de santé à la suite d'une visite médicale :

Excellent, Très bon, Bon, Mauvais, Très mauvais.

1.4.5. Les modalités

Les modalités sont les différentes situations dans lesquelles les individus peuvent se trouver à l'égard du caractère considéré.

Exemple1 :

Le sexe est un caractère qui présente deux modalités : féminin ou masculin

Exemple2 :

Si nous voulons étudier la satisfaction tirée par les étudiants de l'université de Bejaia de leurs dernières vacances estivales, les modalités peuvent être, « peu », « modérément », « beaucoup » ou « énormément ».

Exemple3 :

Si le caractère retenu ou la variable étudiée est le nombre d'enfants par familles, les modalités de ce caractère peuvent être 0,1, 2, 3, etc.

1.4.6. La série statistique

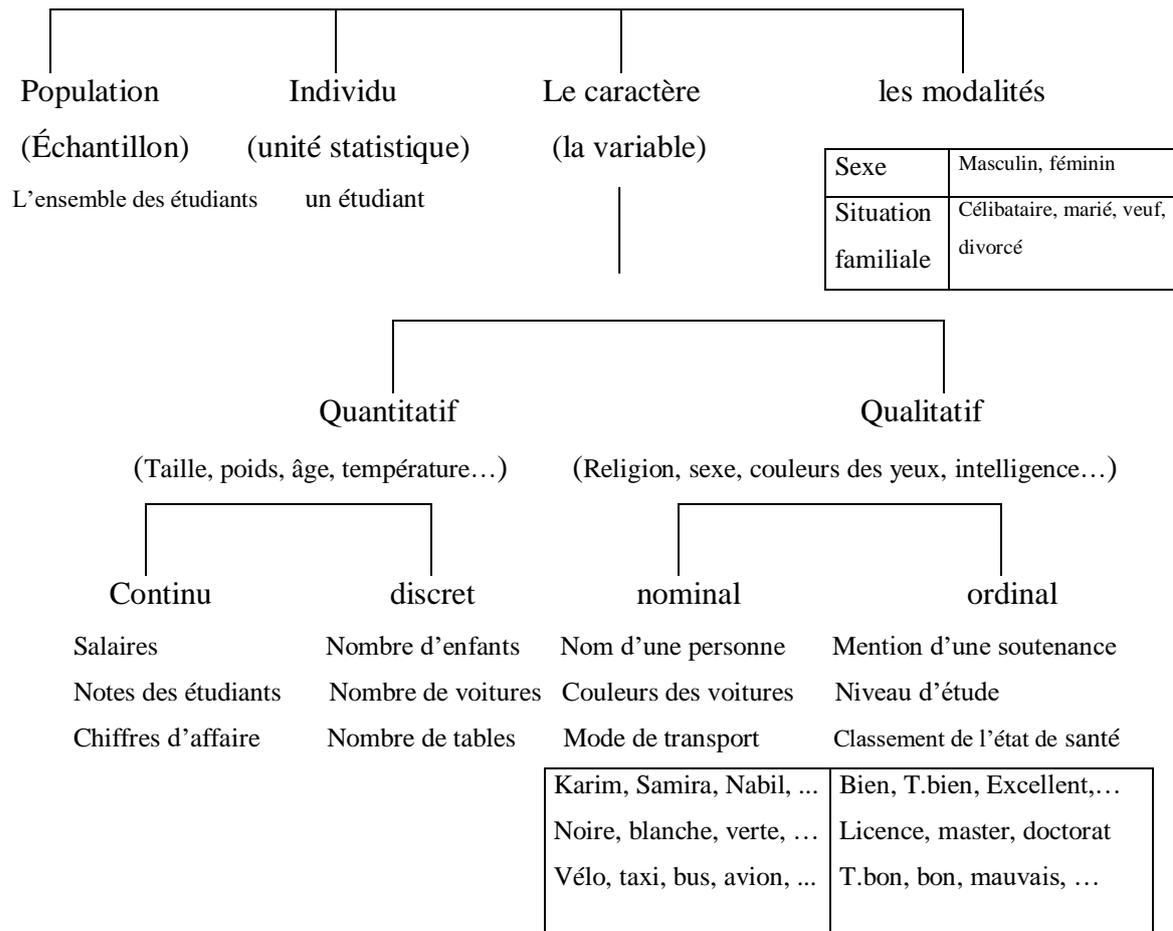
Une série statistique est une liste de mesures obtenues généralement lors d'une étude ou de relevés de mesures.

Exemple :

La série suivante provient d'une enquête auprès de 20 étudiants pour connaître leur taille.

169	170	165	169	167	168	180	170	179	180
180	160	158	175	173	172	167	165	167	177

Récapitulatif des concepts de base de la statistique descriptive



1.5. Exercices corrigés

Exercice 1 :

Pour chacune des variables suivantes, préciser si elle est qualitative nominale, qualitative ordinale, quantitative discrète ou quantitative continue.

- | | |
|---------------------|--------------------------|
| a) Sexe | f) Poids |
| b) État matrimonial | g) Tension artérielle |
| c) Âge | h) Taux de cholestérol |
| d) Taille | i) Mention du Bac |
| e) Couleur des yeux | j) Nombre de téléviseurs |

Corrigé exercice 1 :

- Les variables (a, b, e) sont des variables qualitatives nominales ;
- La variable (i) est une variable qualitative ordinale ;
- Les variables (c, d, f, g, h) sont des variables quantitatives continues ;
- La variable (j) est une variable quantitative discrète

Exercice 2 :

On dispose pour toutes les grandes surfaces d'une entreprise de distribution de leurs chiffres d'affaires aux mois de décembre 2020 et décembre 2021

Supermarché I

Décembre 2020 :458500 DA

Décembre 2021 :700000 DA

Supermarché II

Décembre 2020 :356500 DA

Décembre 2021 :131250 DA

Supermarché III

Décembre 2020 :750000 DA

Décembre 2021 :760000 DA

- 1) Quelle est la population étudiée ?
- 2) De quelle(s) variable(s) dispose-t-on ?
- 3) La variable « chiffre d'affaires en 2020 » est-elle quantitative ? oui ? Non ?
- 4) Discrète ? continue ?

Corrigé exercice 2 :

- 1) La population étudiée est les supermarchés, car c'est sur chacun des supermarchés qu'on a effectué des observations.
- 2) Les variables étudiées sont les chiffres d'affaires en 2020 et 2021 car elles se sont des variables mesurées.
- 3) Oui, la variable « chiffre d'affaires en 2020 » est quantitative (on peut calculer le chiffre d'affaires moyen).
- 4) Elle est continue (on peut avoir n'importe quelle valeur de chiffre d'affaires dans un intervalle).

Exercices 3 :

Pour les sujets d'étude qui suivent, spécifier : la population statistique, l'individu (unité statistique), le caractère et sa nature.

- Étude de la durée de vie des lampes électriques ;
- Étude de l'absentéisme des ouvriers, en jours, dans une usine ;
- Répartition des étudiants selon la mention obtenue sur leur diplôme ;
- Répartition de 100 familles selon le nombre de filles ;
- Répartition d'un groupe de 30 étudiants selon leur âge ;
- Répartition des jeunes selon leur sport préféré.

Corrigé exercice 3 :

La population statistique	L'individu	Le caractère	Sa nature
L'ensemble des lampes	Une lampe	La durée de vie d'une lampe	Quant. continu
L'ensemble des ouvriers	Un ouvrier	Le nombre de jours d'absences	Quant. discret
L'ensemble des étudiants	Un étudiant	La mention du diplôme	Qual. ordinal
Les 100 familles	Une famille	Le nombre de filles par famille	Quant. discret
Les 30 étudiants	Un étudiant	L'âge d'un étudiant	Quant. continu
L'ensemble des jeunes	Un jeune	Le sport préféré	Qual. nominal