

Chapitre 1

Concepts de base

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

L'analyse des données est utilisée pour d'écrire les phénomènes étudiés, faire des prévisions et prendre des décisions à leur sujet. En cela, la statistique est un outil essentiel pour la compréhension et la gestion des phénomènes complexes.

Les données étudiées peuvent être de toute nature, ce qui rend la statistique utile dans tous les champs disciplinaires et explique pourquoi elle est enseignée dans toutes les filières universitaires, de l'économie à la biologie en passant par la psychologie et bien sûr les sciences de l'ingénieur. La statistique est une discipline scientifique dont le but est :

- de planifier et recueillir des données pertinentes,
- d'extraire l'information contenue dans un ensemble de données,
- de fournir une analyse et une interprétation des données afin de pouvoir prendre des décisions.

1.1 Définition des concepts usuels de la statistique

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

1.1.1 Population

En statistique, on travaille sur des populations. Ce terme vient du fait que la démographie, étude des populations humaines, a occupé une place centrale aux

1.1.2 Individu

débuts de la statistique, notamment à travers des recensements de population. Mais, en statistique, le terme de population s'applique à tout objet statistique étudié, qu'il s'agisse d'étudiants (d'une université ou d'un pays), de ménages ou de n'importe quel autre ensemble sur lequel on fait des observations statistiques.

Définition 1.1.1 *On appelle population l'ensemble sur lequel porte notre étude statistique. Cet ensemble est noté Ω . On parlera de population d'élèves, population d'automobiles, population de bactéries, ...*

Exemple 1.1.1

- *On considère l'ensemble des étudiants de la filière MPCI. On s'intéresse au nombre de frères et sœurs de chaque étudiant. Dans ce cas la population est l'ensemble des étudiants.*
- *Si l'on s'intéresse maintenant à la circulation automobile dans une ville, la population est alors constituée de l'ensemble des véhicules susceptibles de circuler dans cette ville à une date donnée.*

1.1.2 Individu

Une population est composée d'individus. Les individus qui composent une population statistique sont appelés unités statistiques. Les individus d'une population peuvent être de nature très diverses : élèves dans une classe, voitures produites par une usine, ...

Définition 1.1.2 *On appelle individu ou unité statistique tout élément ω de la population Ω .*

Remarque 1.1.1

- *Les individus d'une population peuvent être de nature très diverses : élèves dans une classe, voitures produites par une usine, ...*
- *L'unité statistique est un objet pour lequel nous sommes intéressés à recueillir de l'information.*

1.1.3 Echantillon / Population mère

Lorsque la population est trop importante pour être connue entièrement, on prélève un échantillon. Un échantillon de taille n est un sous-ensemble formé de n individus de la population. La population initiale d'où est tiré l'échantillon est la population mère. Pour que les résultats observés lors d'une étude soient généralisables à la population mère, l'échantillon doit être représentatif de cette population, c'est-à-dire qu'il doit refléter fidèlement sa composition et sa complexité. Dans le cas contraire, l'échantillon est biaisé.

1.1.4 Caractère

La statistique « descriptive », comme son nom l'indique cherche à décrire une population donnée. Nous nous intéressons aux caractéristiques des unités qui peuvent prendre différentes valeurs.

Définition 1.1.3 :

On appelle caractère (ou variable statistique, dénotée $V.S$) toute application

$$X : \Omega \rightarrow C.$$

L'ensemble C est dit : ensemble des valeurs du caractère X (c'est la propriété qui est mesuré ou observé sur les individus)

Exemple 1.1.2 *Taille, température, nationalité, couleur des yeux, catégorie socio-professionnelle ...*

Remarque 1.1.2 *Soit Ω un ensemble. On appelle et on note $\text{Card}(\Omega)$, le nombre d'éléments de Ω .*

1.1.5 Modalités

Les modalités d'une variable statistique sont les différentes valeurs que peut prendre celle-ci. Ces modalités doivent être exhaustives (elles décrivent toutes les catégories ou valeurs possibles d'un caractère) et incompatibles (à chaque individu doit correspondre une seule catégorie ou valeur) de façon à créer une partition.

Pour une population étudiante universitaire, on peut s'intéresser par exemple aux caractères « mention du bac », « âge », « couleur des yeux », ...

- les modalités (catégories) du caractère « mention du bac » sont : passable, assez bien, bien, très bien, excellent.
- les modalités (valeurs) du caractère « âge » sont : 17, 18, 19, ...
- les modalités (catégories) du caractère « couleur des yeux » sont : marron, bleu, gris, vert, ...

1.1.6 Série statistique

Une série statistique est l'ensemble des résultats, sous forme de données brutes, d'une étude menée sur une certaine population, relatifs à un ou plusieurs caractères au sein de cette population. Si un seul caractère est étudié, on parlera de série statistique à une dimension et lorsque deux caractères sont étudiés on parlera de série statistique à deux dimensions. Les données brutes ainsi recueillies, présentées généralement sous forme de listes ou de tableaux, sont difficiles à analyser telles quelles. C'est pour cela que nous devons les trier, les classer, les exploiter et éventuellement les ordonner selon un ou plusieurs critères. Le résultat est un nouveau type de tableau appelé tableau statistique qui permet de présenter les données de façon concise dans le but d'interpréter son contenu. Ces données y sont généralement regroupées. Le regroupement peut être effectué par catégories ou classes de catégories pour un caractère qualitatif, et par valeurs ou classes de valeurs pour un caractère quantitatif; de ce fait, la série statistique est réorganisée en ce que l'on appelle une distribution statistique.

1.2 Types des caractères

Nous distinguons deux catégories de caractères : les caractères qualitatifs et les caractères quantitatifs.

1.2.1 Caractères qualitatifs

Un caractère est dit qualitatif si ses modalités ne sont pas mesurables. Dans ce cas, les modalités sont aussi appelées catégories. Si les modalités du caractère qualitatif ne sont pas naturellement ordonnées on dira qu'il est nominal, par contre si les modalités sont organisées selon un ordre hiérarchique on dira qu'il est ordinal. Le caractère « couleur des yeux » est nominal alors que le caractère « mention du bac » est ordinal.

1.2.2 Caractères quantitatifs

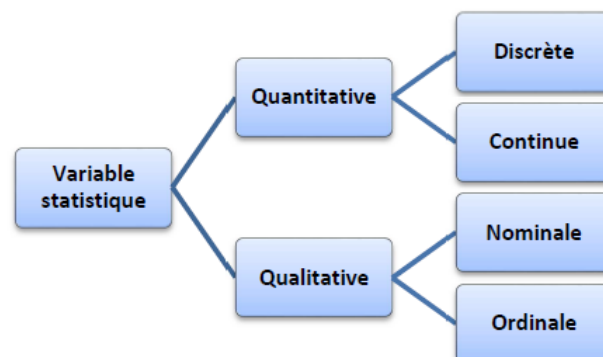
Un caractère est dit quantitatif si ses modalités sont mesurables. Dans ce cas, on lui donne le nom de variable statistique, et bien entendu les différentes modalités constituent l'ensemble des valeurs numériques du caractère quantitatif. Il existe deux types de variables statistiques : les variables discrètes et les variables continues.

a) Variables discrètes

Une variable statistique est dite discrète si l'ensemble de ses modalités est un ensemble fini ou dénombrable. En général, elles sont représentées par des nombres entiers. Elles résultent le plus souvent d'un comptage : nombre d'enfants par famille, nombre de places dans une automobile, ...

b) Variables continues

Une variable statistique est dite continue si l'ensemble de ses modalités n'est pas dénombrable. Ainsi, une variable continue peut prendre toutes les valeurs d'un intervalle. En général, elles sont représentées par des nombres en écriture décimale. Elles résultent le plus souvent d'une mesure : poids d'une personne, longueur d'un saut...



1.3 Elaboration de statistiques

L'étude statistique des phénomènes suppose d'abord une collecte des données de base. Cette collecte se fait à partir d'enquêtes (collecte auprès de personnes morales ou physiques), de résultats d'expériences ou d'exploitation de fichiers administratifs. L'observation des faits peut se faire de façon instantanée (enquêtes par sondages et recensements) ou de façon continue (enregistrement des naissances à l'état civil, comptabilité d'une entreprise).

1.3.1 Recensement

C'est une méthode exhaustive, c'est-à-dire que toute la population fait l'objet d'observation suivant le ou les caractères étudiés.

Exemple 1.3.1 : *Recensement de la population du Burkina Faso en décembre 2006 suivant des caractères démographiques (âge, sexe, etc.), économiques (activités économiques), sociaux (niveau d'éducation, alphabétisation, etc.), géographiques (lieu de résidence).*

1.3.2 Enquête par sondage

Elle porte sur un échantillon.

Exemple 1.3.2 :

- *Enquête sur les conditions de vie des ménages*
- *Enquête démographique et de santé*
- *Sondages d'opinion*
- *Etudes du temps passé sur le net par les étudiants d'une filière en une journée*

1.3.3 Les grandes étapes d'une enquête statistique

Le déroulement d'une enquête statistique peut être résumé en quatre (4) grandes étapes :

- a) La conception : Elle consiste à définir les objectifs de l'étude, définir l'ensemble de l'étude ainsi que les critères à étudier, à concevoir les outils nécessaires à la collecte des informations (questionnaires, guide d'entretien, manuels des agents, etc.). Elle doit également définir les résultats attendus, notamment les indicateurs essentiels à calculer.
- b) La phase de collecte : Elle comprend la formation des acteurs, la sensibilisation des personnes cibles, l'observation et l'enregistrement de l'information à l'aide de questionnaires. La collecte peut se faire par interview directe, par courrier (poste, e-mail), par téléphone, etc.
- c) Traitement des données : pour manipuler plus facilement les données brutes recueillies, il faut passer :
 - au regroupement de ces données et à leur présentation en tableaux ;

1.3.3 Les grandes étapes d'une enquête statistique

- à la représentation graphique de ces données pour obtenir un résumé visuel ;
- aux calculs permettant de caractériser et de résumer ces données à l'aide d'indicateurs ;

L'ensemble des étapes 1, 2 et 3 constituent la Statistique Descriptive.

- d) Interprétation et Analyse des données : les données recueillies à l'issue d'une étude statistique ne constituent en fait qu'un échantillon d'une réalité souvent bien plus vaste. En s'appuyant sur le Calcul des Probabilités, on essaye de voir si les résultats observés sur l'échantillon peuvent se généraliser à la population toute entière pour ainsi prendre des décisions ou faire des prévisions. On fabrique alors un estimateur dont il faut contrôler l'incertitude. C'est le rôle de la Statistique Inférentielle.