

1 INTRODUCTION ET DÉFINITIONS

1.1 Introduction

Dans les conversations, les reportages, ... de tous les jours, apparaissent des phrases telles que : « *On a pu évaluer...* », « *Tel sondage attribue x % des intentions de vote en faveur de...* », « *On a prouvé statistiquement que...* », qui à elles seules montrent l'importance de la statistique.

La notion de **statistique** est née de l'envie de dénombrement, d'inventaire, de recensement (de biens, de personnes, ...), de la nécessité de s'informer pour connaître une situation, pour élaborer une stratégie, pour prendre une décision. La statistique contribue à gérer l'information dans un grand nombre de domaines comme celui de la politique (orientation de la politique générale, étude de la composition de la population, etc.), de l'économie (contrôle de la qualité de production, études de marché, ...), des assurances (établissement des primes d'assurance vie, ...), de la physique (théorie cinétique des gaz), de la biologie (lois de *Mendel*), de la psychologie, etc.

1.2 Les statistiques et la statistique

Les statistiques reprennent l'ensemble des données souvent numériques obtenues lors d'enquêtes (relevés de mesures, de ventes, de sondages, ...) qu'il est ensuite nécessaire d'organiser et synthétiser.

Il faut savoir comment recueillir les informations, auprès de qui et sous quelle forme, comment organiser ces informations et comment en décrire les principales caractéristiques.

La statistique reprend l'ensemble des méthodes ayant notamment pour but de traiter les données ci-dessus, elle permet donc l'amélioration de la connaissance, de la compréhension des situations et des phénomènes. Les méthodes de traitement et d'analyse sont fonctions des caractéristiques des observations et de la manière dont elles ont été recueillies.

La statistique descriptive rassemble, ordonne et représente les données ; elle utilise des opérations simples (rangement, représentations graphiques, ...) pour obtenir une idée générale sur la manière dont ces données se situent les unes par rapport aux autres. Elle se complète par une analyse exploratoire qui décrit les données sous forme de tableaux, de graphiques, et construit des paramètres représentatifs des caractéristiques.

La statistique mathématique se base sur les lois du calcul des probabilités pour tirer des conclusions.

1.3 Vocabulaire

La **population** est l'ensemble des éléments (personnes ou choses) soumis à une étude statistique, par exemple, les habitants d'une commune, des catégories sociales, des pièces usinées, des marchandises.

Un **échantillon** représente une partie plus ou moins restreinte d'une population, il doit néanmoins être suffisamment important et représentatif de la population entière. Par exemple, les habitants d'un quartier par rapport aux habitants d'une région, 100 lampes électriques parmi toutes celles produites par une usine.

Le **caractère** est le trait commun à tous les éléments d'une population et qui fait l'objet de l'étude statistique. Il peut prendre différentes *valeurs* ou *modalités*.

- Le caractère peut être **quantitatif** si ses modalités s'expriment par des nombres (il est mesurable grâce à une unité de mesure). Par exemple, le nombre d'enfants par famille, l'âge, la taille, le prix, la puissance d'un moteur.

Si les nombres traduisant la mesure sont entiers, le caractère est dit **discret** (le nombre d'enfants par famille).

Si les valeurs ne sont pas obligatoirement entières et qu'elles peuvent varier d'aussi peu que souhaitable suivant la précision de l'appareil de mesure (elles peuvent prendre toutes les valeurs d'un intervalle fini ou non), le caractère est qualifié de **continu** (taille d'une personne).

Il est à remarquer que dans le cas d'un caractère quantitatif, l'application qui à tout membre de la population associe la valeur numérique liée à ce caractère porte le nom de *variable statistique* (qualifiée de discrète ou continue).

- Le caractère peut être **qualitatif** s'il n'est pas mesurable directement (profession, état civil, appréciation des qualités et défauts, ...).

Néanmoins, le recours à l'informatique pour faciliter le traitement des données a répandu l'usage des codes qui consiste à attribuer à des informations qualitatives des étiquettes numériques plus faciles à manipuler mais avec un caractère numérique purement conventionnel (modalité homme & femme, codes 1 & 2).

Une série statistique peut être à une dimension (ou univariée) si elle reprend les données relatives à une population ou un échantillon pour lequel un seul caractère a été étudié (par exemple, l'ensemble des véhicules de la marque **X** pour lesquelles le prix est relevé). Elle peut être à plusieurs dimensions si elle reprend les données relatives à une population ou un échantillon pour lequel plusieurs caractères sont mesurés (l'ensemble des véhicules de la marque **X** pour lesquels le prix, la puissance et le poids sont relevés).