Introduction

la statistique est une branche des mathématiques qui s’intéresse à l’étude des phénomènes aléatoires, en contre opposé  aux mathématiques déterministes plus familières aux étudiant(e)s, que sont le calcul différentiel, calcul intégral, géométrie, algèbre,…Le mot statistique dont le nom est dérivé de ``state’’ en référence à tout ce qui est étatique, est relativement nouveau, puisqu’il a été introduit en Allemagne au XVIIème siècle. Par contre la pratique de la statistique est plus ancienne, elle fut utile aux grands empires en Mésopotamie, dans l’Égypte ancienne, ainsi que chez les romains et les empires indiens et chinois. Il s’agissait de bien connaître la population pour administrer sa répartition sur les territoires, collecter les impôts et gérer les aspects militaires.

De nos jours, on ne peut trouver un domaine qui peut être compris, analysé sans les méthodes statistiques. Que ce soit dans le domaine des sciences sociales, sciences de la vie ou sciences de l’ingénieur, les méthodes statistiques sont omniprésentes pour mettre de l’ordre dans le protocole de travail, elles permettent quand on est devant un chaos apparent des données, de déterminer par où commencer et quelles sont les étapes à suivre selon le contexte pour analyser ces données.

La statistique grosso-modo est formée de trois  grandes classes : la statistique descriptive,  la statistique inférentielle et la nouvelle branche qu’est la statistique exploratrice. Ce chapitre est consacré à la statistique descriptive. La statistique descriptive  comme son nom l’indique, se propose de décrire les données, de les classer et de les présenter sous des formes claires et compréhensibles. Elle est à la base par exemple de toute organisation du système d’information d’une entreprise : statistiques de la production ou des ventes, statistiques financières, statistiques des ressources humaines…Elle est aussi une importante composante en sciences humaines de ce qu’on appelle les méthodes quantitatives. On va commencer par définir le lexique qu’on va utiliser tout le long de ce chapitre et même de ce livre.

**1.2   : Les concepts et le vocabulaire de base.**

Au début de tout travail statistique, il faut cerner avec précision sur quoi va porter l’étude. L’ensemble de tous les éléments sur lesquels porte l’étude s’appelle **population.**Une population peut être un ensemble d’êtres vivants (humains, oiseaux, poissons, bactéries,…) ou un ensemble de choses (maisons, voitures, rivières,…) ou un ensemble de faits (pannes, accidents, divorces,…). Chaque élément d’une population s’appelle **individu** ou **unité statistique**. Une population peut être finie (population d’un pays) ou presque infinie (population des insectes), on considère généralement les populations comme finies mêmes si elles sont très grandes. Le nombre d’unités statistiques dans une population s’appelle **taille de la population**et on le note par **N**.

Quand une étude porte sur toute la population, on dit qu’on fait un **recensement**. Mais pour des raisons techniques ou économiques, il n’est généralement pas possible de collecter des données sur tous les éléments d’une population. Alors on se contente d’extraire une partie de la population appelée **échantillon** et restreindre l’étude à cet échantillon. On verra dans le chapitre V, qu’il existe des méthodes spécifiques permettant de s’assurer que l’échantillon soit représentatif de la population, c'est-à-dire une réplique en miniature de ce qui se passe dans la population. Pour l’instant, on suppose qu’on dispose d’un échantillon sur lequel porte l’étude (sans savoir comment il a été extrait). Le nombre d’éléments dans l’échantillon s’appelle **taille de l’échantillon** et sera noté par **n**.

On appelle **variable** tout caractère observé ou mesuré sur chacun des éléments de l’échantillon. On va réserver les dernières lettres de l’alphabet pour noter les variables : X, Y, Z, U...

Les différentes valeurs que prend une variable s’appellent **modalités**. Afin que le classement d’une unité statistique soit toujours possible sans ambiguïté, les différentes modalités doivent être à la fois incompatibles (un individu ne peut avoir plusieurs modalités à la fois) et exhaustives (tous les cas doivent être prévus). Il existe deux types de variables : Les **variables qualitatives** et les **variables quantitatives**. Une variable est dite qualitative si elle ne peut être mesurée ou quantifiée, mais peut être classée en catégories comme le sexe, la race, l’espèce, le niveau scolaire,…. Une variable est de type quantitatif si elle peut être mesurée ou quantifiée, comme le poids, la hauteur, le revenu, le nombre d’enfants, le nombre de pannes.

Les variables qualitatives sont constituées de deux sous-classes :

* Les variables qualitatives **nominales**: ce sont celles dont les modalités ne peuvent qu’être constatées, nommées.

1. **:** Le sexe (masculin, féminin), la nationalité (Canadienne, Française, Marocaine,..), les cours suivis durant une session (mathématiques, anglais, philosophie,..) …

* Les variables qualitatives **ordinales**. ce sont les variables qualitatives dont les modalités appellent naturellement un ordre dans leur rangement. **Exemple :** Le niveau scolaire (primaire, secondaire, collégial, universitaire), le comportement lors d’une réception (incongru, correct, parfait,..), …

Les variables quantitatives sont elles aussi subdivisées en deux sous-classes :

* Les variables quantitatives **discrètes :**ce sont celles dont les modalités sont des valeurs isolées.

1. **:** Le nombre de pannes, le nombre d’accidents, le nombre d’enfants,…
2. **:** La taille, le poids, le revenu,…

* Les variables quantitatives **continues**, ce sont celles dont les modalités forment un continuum. Ce sont celles qui peuvent prendre n’importe quelle valeur dans un intervalle raisonnable.