Statistiques

Marie Delacre

###### Pagebreak

# Chapitre 1: qu’est-ce que la statistique, et à quoi sert-elle?

## Qu’est-ce que la statistique?

La statistique est l’ensemble des instruments de recherches mathématiques qui permettent de **récolter**, **traiter** et **interpréter** un ensemble de données (généralement vaste).

### La récolte des données: les concepts de population, individus et échantillons

La **population** est l’ensemble des éléments (individus ou objets) auxquels on s’intéresse. Les éléments qui constitue cette population sont appelés **individus** ou **unité statistique**. Admettons par exemple que l’on souhaite étudier les caractéristiques de l’ensemble des employés belges du secteur automobile. Chaque employé de nationalité belge travaillant dans ce secteur constituera une unité statistique, et leur ensemble constituera la population d’intérêt.

Il est bien souvent compliqué, voire impossible de recueillir des informations pour l’entièreté de la population tant celle-ci peut être vaste. Pour cette raison, on prélèvera généralement, aléatoirement, une partie de taille gérable de cette population que l’on appelle l’**échantillon**, et l’on mesurera, pour chaque individu de l’échantillon, un ensemble d’informations (des **variables**) qui nous intéresse, comme le genre, la situation géographique ou linguistique ou encore l’âge et le salaire moyen.

Les variables que l’on récolte au sein de l’échantillon peuvent être de diverses natures. Autrement dit, il existe plusieurs manières de les mesurer. Imaginez par exemple que je vous demande d’estimer la silhouette d’un individu. Vous pourriez prendre un mêtre et me dire que son tour de taille est de 72 cm. Vous pourriez également vous contenter de dire qu’il est de taille “Medium”. Décider de la manière dont on mesure les variables aura des implications importantes en termes de possibilités d’analyses et ne doit pas être laissé au hasard. Un chapitre ultérieur sera dédié aux variables et à leurs natures possibles (chapitre 2).

### Le traitement des données, appelée également la statistique descriptive

Après avoir récolté des données, on se retrouve avec une série de variables pour chaque individu. Ces données pourront être encodées dans une base de données tel qu’illustré à la Table 1.1. Cette table est une représentation fictive de données socio-démographiques récoltées auprès de 15 employés du secteur automobile.

Table 1.1  
*Données socio-démographiques récoltées auprès de 15 employés du secteur automobile* Table @ref(tab:Table1)