



SAE – ECHANTILLONNAGE ET ESTIMATION

SHAABAN Wissam
ROUSSELIN Quentin

D'une part, ce projet nous amène à la production, sur le logiciel R ; d'une estimation de la population totale d'une région de France. Celle-ci se fait à partir d'un échantillon de 100 communes parmi les nombreuses que comptent les régions françaises. Pour obtenir cet échantillon, deux méthodes d'échantillonnage sont utilisées (relatives au sondage) : l'échantillonnage aléatoire simple, et l'échantillonnage aléatoire stratifié. Pour réaliser ces échantillonnages, nous avons à notre disposition un fichier au format csv contenant (en outre), la population de toutes les communes de France, avec la région d'indiquée ; daté du 1^{er} janvier 2023. **La finalité est d'appréhender, à la fois l'incertitude et la précision de l'estimation d'une grandeur mesurable dans une population.**

D'autre part, ce projet nous conduit aussi à réaliser, sur le logiciel SAS ; un test du khi-deux d'indépendance sur des notes dans trois matières : mathématiques, informatique, statistiques. Pour cela, nous avons à notre disposition deux fichiers Excel : Notes.xls (contenant les notes) et Liste.xls (contenant les caractéristiques des élèves).

PARTIE 1 – Echantillonnage aléatoire simple

Pour commencer l'échantillonnage, il nous a d'abord fallu importer le fichier csv intitulé 'population_française_communes', avant de créer un second data frame contenant uniquement les variables utiles à la conception d'un échantillon (le nombre d'habitants, le nom de la commune), uniquement sur les communes de la région étudiée, à savoir la Bretagne.

```
table <- read.csv2("population_francaise_communes.csv", sep=";", dec=",", header=TRUE)
donnees <- table[table$Nom.de.la.region=="Bretagne", c("Code.département", "Commune", "Population.totale")]
head(donnees)
```

