

## Variables qualitatives

### Exercice 1

1- Charger les données contenues dans le fichier Classe.data dans R avec la commande `read.table` dans une table nommée `data`.

2- Définir les vecteurs `sexe`, `departement`, et `serie`.

3- Représenter les fréquences des variables `sexe`, `departement`, et `serie` en utilisant respectivement un digramme en bâtons, en digramme en barres, un diagramme par secteurs.

4- Déterminer les modes de ces séries.

### Exercice 2

On reprend les données du fichier Classe.

1- Définir le vecteur `difficulte`

2- Calculer les fréquences cumulées associées

3- Tracer la boîte à moustache associée.

4- Comparer les difficultés des étudiants provenant des filières “S” et “ES” au travers de leurs boîtes à moustache.

## Variables quantitatives discrètes

### Exercice 1

Lors d’une étude statistique portant sur le nombre d’enfants des ménages résidant dans une zone géographique particulière, on recueille les données suivantes:

Nombre d’enfants	Effectifs observés
0	43
1	58
2	49
3	25
4	13
5	8
6	4

1. Créer un vecteur correspondant à ces données.

2. Représenter la distribution par un diagramme en bâtons et déterminer la valeur du mode. Utiliser `range` pour retrouver les modalités maximales et minimales.

3. Donner le tableau des fréquences et des fréquences cumulées. Tracer la courbe des fréquences cumulées.

4. En déduire la valeur des quartiles. Utiliser la commande `boxplot.stats` pour contrôler vos résultats

5. Déterminer la moyenne, la variance, et l’écart-type (à la main puis avec R)

### Exercice 2

On étudie la fréquentation d'un guichet dans une grande banque au centre de Paris dans la but d'améliorer ce service. On fait une étude sur 2000 périodes de 2 minutes choisies sur la tranche horaire de 11h à 12h et on compte le nombre de personnes entrant dans l'agence pendant la période donnée. Les données recueillies se trouvent dans la librairie 'lycee' dans le fichier 'guichet'.

- 1- Charger ces données sur R. Créer un vecteur file contenant la colonne "queue" du fichier guichet. Déterminer la moyenne, le mode et la médiane de ces données.
- 2- Tracer la boîte à moustache associée à ces données, et déterminer les valeurs des quartiles, de la variance et de l'écart-type.
- 3- Construire le diagramme en bâton correspondant à la distribution des données. Représenter également la courbe des fréquences cumulées.
- 4- Proposer une modélisation de la distribution du caractère étudié. Représenter sur un même graphique la distribution empirique et la modélisation.

### Exercice 3

On observe le nombre d'enfant de 1000 ménages d'une région géographique particulière

Nombre d'enfants	Effectifs observés
1	510
2	248
3	132
4	55
5	29
6	15
7	2
8	8
10	1

1. Créer un vecteur correspondant à ces données.
2. Représenter la distribution par un diagramme en bâtons et déterminer la valeur du mode. Utiliser range pour retrouver les modalités maximales et minimales.
3. Donner le tableau des fréquences et des fréquences cumulées. Tracer la courbe des fréquences cumulées.
4. En déduire la valeur des quartiles. Utiliser la commande boxplot.stats pour contrôler vos résultats
5. Déterminer la moyenne, la variance, et l'écart-type (à la main puis avec R)
6. On voudrait voir si il est envisageable de modéliser la distribution du nombre d'enfants par une loi géométrique. Sachant que l'espérance d'une loi géométrique de paramètre p est donnée par  $(1-p)/p$ , on propose d'estimer p par  $1/(1+\text{mean}())$ . Comparer distribution empirique et modélisation.