TP n^o 2 : statistique descriptive univariée

La statistique est l'étude d'un phénomène par la collecte de données, leur analyse, leur traitement, l'interprétation des résultats et leur présentation afin de rendre les données compréhensibles par tous. Tout naturellement, cela commence par les organiser, les regarder, les représenter graphiquement . . . bref, faire appel aux outils et méthodes de la statistique descriptive.

Si les données ne sont relatives qu'à une seule variable, on parle de statistique descriptive univariée. Dans le cas où l'on s'intéresse à deux variables simultanément, on met en œuvre la statistique descriptive bivariée. Si l'ensemble de données provient de l'observation de plusieurs variables, on doit faire appel aux méthodes de la statistique descriptive multivariée.

Dans le TP1, nous avons présenté quelques outils graphiques de la statistique descriptive univariée. Ce TP est consacré à la présentation d'autres outils.

Le fichier "GPA.csv" contient des notes d'études secondaires et universitaires pour des diplômés en informatique dans une école publique locale.

1. Télécharger ce fichier puis importer-le dans R avec la commande

   ```
   data=read.csv("GPA.csv",header =TRUE,sep=";",dec =",")
   ```

2. Etudier les lignes de commande suivantes:

   ```
   dim(data)
   high=data$high_GPA
   univ=data$univ_GPA
   ```

3. Compléter la fonction suivante permettant de calculer la moyenne des entrées d'un vecteur z:

   ```
   moy <- function (z){
   n <- ...........
   zbar <- ......./n
   return (zbar)
    }
   ```

   Utiliser cette fonction pour calculer les moyennes des variables étudiées.
   Retrouver les résultats obtenus avec la fonction *mean*.

4. Ecrire une fonction pour calculer la variance des entrées d'un vecteur z.
   Calculez ensuite la variance puis l'écart-type des variables univ_GPA et high_GPA.

5. Utiliser la fonction *summary* pour donner un résumé statistique de chacune des deux variables étudiées.

6. Tracer dans le même graphique les boites à moustache de deux variables (*boxplot*).

7. Représenter, dans la même fenêtre graphique, les histogrammes (fonction *hist*) des deux variables univ_GPA et high_GPA.
   La commande par(mfrow=c(1,2)) permet de tracer 2 graphiques dans une même fenêtre.