

## TP3 : Indicateurs statistiques

### Exercice 1: fichier de données iris

On considère la variable longueur du pétale de ce jeu de données iris.

1. Taper la ligne de commande :

```
summary(iris$Petal.Length)
```

Quelle est l'information que vous obtenez?

2. Retrouvez ces valeurs en appliquant les formules vues en cours.

### Exercice 2: Données brutes ou regroupement par classes

On considère le jeu de données suivant, relatif au taux d'hémoglobine dans le sang mesuré en g/L chez des adultes supposés en bonne santé.

Femmes	105	110	112	112	118	119	120	120	125	126
	127	128	130	132	133	134	135	138	138	138
	138	142	145	148	148	150	151	154	154	158
Hommes	141	144	146	148	149	150	150	151	153	153
	153	154	155	156	156	160	160	160	163	164
	164	165	166	168	168	170	172	172	176	179

On considère les classes  $]104; 114]$ ,  $]114; 124]$ ,  $]124; 134]$ ,  $]134; 144]$ ,  $]144; 154]$ ,  $]154; 164]$ ,  $]164; 174]$ ,  $]174; 184]$ .

1. Pour chacune des deux séries hommes et femmes, déterminer les effectifs et les fréquences pour chacune des classes
2. Calculer les moyennes des trois distributions initiales : ensemble, femmes et hommes
3. Calculer les moyennes des trois distributions après le regroupement en classes : ensemble, femmes et hommes
4. Calculer les médianes des trois distributions initiales
5. Calculer l'écart interquartile des trois distributions initiales
6. Calculer les variances et écart-types des trois distributions initiales
7. Calculer les variances et écart-types des trois distributions après regroupement par classes
8. Pour la distribution des femmes, calculer les moments jusqu'à l'ordre 4, puis en déduire les moments centrés jusqu'à l'ordre 4 et le coefficient d'asymétrie et d'aplatissement.

9. Quelles sont les fonctions du logiciel qui permettent de faire cela?
10. Réaliser une représentation graphique permettant de visualiser de façon synthétique la dispersion des données.

### Exercice 3: retour sur les données iris

Précédemment, nous avons travaillé sur chaque variable pris séparément. A présent, nous allons regarder à l'aide de l'outil informatique comment croiser deux variables.

1. Taper, identifier le rôle de chaque argument de la fonction plot et commenter le résultat de la ligne de commande suivante :

```
plot(iris$Petal.Length,iris$Petal.Width,xlab="Longueur du pétale",ylab="Largeur du pétale",main="Nuage de points",pch=20)
```

2. Faire de même entre deux autres variables quantitatives de votre choix.
3. Que produit la ligne de commande suivante et commenter le résultat :

```
boxplot(iris$Petal.Length ~ iris$Species,col=grey(0.6))
```

Quel peut-être l'usage d'une telle commande?

Les travaux précédents semblent mettre en lumière le fait que les données morphologiques des iris semblent être liées à l'espèce.

Par conséquent, sur une même fenêtre graphique :

1. Tracer l'histogramme des longueurs de pétales pour l'ensemble des iris, les iris de catégorie setosa, de catégorie versicolor et de catégorie virginica
2. Tracer les nuages de points de la largeur du pétale en fonction de la longueur du pétale pour l'ensemble des iris, les iris de catégorie setosa, de catégorie versicolor et de catégorie virginica

Pour finir, taper et commenter :

```
pairs(iris[,1:4],main="Anderson's Iris Data",pch=21,  
bg=c("red","green3","blue")[unclass(iris$Species)])
```