

# TD3: Variables quantitatives

Statistiques descriptives avec R

Mars 2023

## Exercice 1: Comment grouper des données ? (Notes de cours Myriam et Frédéric Bertrand)

*Vous allez vous intéresser ici à la variable **masse** du jeu de données **Mesures**. Dans les rappels de cours, vous avez vu comment grouper ces données automatiquement à l'aide de l'une des trois règles dues à Sturges, Scott et Freedman-Diaconis. Vous allez voir comment grouper des données suivant différents critères.*

1. Groupez les données en 5 classes à l'aide de l'option `breaks=5` de la fonction `hist`. Que se passe-t-il si vous cherchez à en obtenir seulement 4 ?
2. Groupez les données en utilisant les classes suivantes `[0; 5]`, `]5; 10]`, `]10; 15]`, `]15; 20]` et `]20; 50]` à l'aide de l'option `breaks=c(0,5,10,15,20,50)` de la fonction `hist`.
3. Comparez le résultat obtenu avec :

```
brk <- c(0, 5, 10, 15, 20, 50)
table(cut(Mesures$masse, brk))
data.frame(table(cut(Mesures$masse, brk)))
```

4. Si vous cherchez à créer des groupes dont les effectifs sont équilibrés, vous pouvez par exemple utiliser la fonction `cut2` de la bibliothèque `Hmisc`. Après avoir téléchargé et installé cette bibliothèque, commentez les lignes de code suivantes et en particulier le rôle des options `g` et `m`.

```
library(Hmisc)
brk <- c(0, 5, 10, 15, 20, 50)
res <- cut2(Mesures$masse, brk)
table(res)
table(cut2(Mesures$masse, g = 10))
table(cut2(Mesures$masse, m = 50))
```

## Exercice 2. Jeu de données Europe (Notes de cours Myriam et Frédéric Bertrand)

*Le but de cet exercice est de calculer des résumés numériques et de tracer une boîte à moustaches.*

1. Affichez les six premières lignes de ce jeu de données qui est disponible dans la bibliothèque `BioStatR`
2. De quoi est constitué ce jeu de données ? C'est-à-dire : combien de variables composent ce jeu de données ? quelle est la nature de ces variables ? combien

d'unités statistiques sont présentes dans ce jeu de données ?

3. Quelles sont la classe et la taille de ce jeu de données ?
4. Donnez la moyenne, la valeur minimale, la valeur maximale, la médiane et la ou les classes modales de la variable **Duree**.
5. Donnez l'écart-type corrigé, le coefficient de variation et l'étendue de la variable **Duree**.
6. Tracez la boîte à moustaches de la variable **Duree** en mettant un label pour l'axe vertical qui est "Durée en heures". Représentez sur cette même boîte la moyenne. Tracez la boîte à moustaches de la variable **Duree** en mettant un label pour l'axe vertical qui est "Durée en heures". Représentez sur cette même boîte la moyenne.
7. Sauvegardez la boîte à moustaches au format **.pdf** et au format **.png** en utilisant les fonctions **pdf** et **png**.

**Exercice 3 Coefficient de Gini.** (*Notes de cours Christophe Ambroise*)

Le coefficient de Gini permet de mesurer l'inégalité des revenus dans une population. Si tous les individus gagnent le même salaire le coefficient de Gini vaut 0 (situation égalitaire), alors que si un seul individu gagne tous le revenu disponible et les autres rien l'index de gini vaut 1. Les Etats-unis ont par exemple un coefficient de Gini de 0.47.

1. Charger le jeu de données **gini.Rdata**.
2. Sélectionner les lignes du tableau correspondant à l'année 2007.
3. Tracer l'histogramme des coefficients.
4. Tracer le boxplot des coefficients.
5. Tracer un diagramme des fréquences cumulées des coefficients.
6. Ecrire une fonction R qui rende les pays de coefficient Gini d'index maximum et minimum.
7. Classer les pays par leur coefficient de Gini.
8. Calculer la moyenne, la variance, des coefficients de gini.
9. Combien de pays sont plus égalitaires que la France en Europe.

**Exercice 4 Statistiques de tendance centrale et de dispersion**

*Il s'agit ici de calculer les principales caractéristiques évoquées dans le cours pour le cas d'une variable statistique discrète. Notons que les valeurs sont ponctuelles et ne nécessitent pas de calculs préalables.*

Lors d'un contrôle de connaissances, on fait subir à huit étudiants un test noté sur 50 points. La série des notes obtenues est la suivante :

25, 43, 35, 20, 32, 30, 35, 24.

Calculer et expliquer les résumés statistiques suivants :

1. Mode, médiane. Que devient la médiane lorsque nous ajoutons la note 45 à la série ?
2. Moyennes arithmétique, quadratique, géométrique, harmonique.
3. Variance et écart type. Donner le coefficient de variation.