

## **Analyse de Données**

### **TP4 – Langage R**

1. Spécifier le répertoire de travail
2. Charger le fichier de données « pHconductiviteRiviere.csv »
3. Faire les transformations nécessaires pour convertir les attributs à des valeurs numériques (variables quantitatives).
4. Afficher les statistiques descriptives pour chaque variable. Est-il nécessaire d'effectuer l'opération de centrage-réduction ?  
Si votre réponse est oui, centrer réduire.

On veut étudier si l'on peut effectuer une analyse par composantes principales (ACP).

5. Afficher la matrice de corrélation entre toutes les variables. Que peut-on dire ?
6. Calculer le test de sphéricité de Bartlett. Conclusion ?
7. Calculer l'indice de KMO tout en affichant les  $MSA_i$ . Que peut-on conclure ?  
 $\Rightarrow$  Les variables pH, DB, Cs, O2, Ox, P4 et N4 contribuent fortement à la construction des facteurs.
8. Extraire les facteurs à l'aide d'un ACP.
9. Afficher la variance totale expliquée par les facteurs extraits. Quels sont les pourcentages d'inertie expliquée par l'ACP si l'on extrait 3 facteurs ? 4 facteurs ? 5 facteurs ?
10. Tracer le graphique des valeurs propres. Quel est le nombre de facteurs à garder ?