
Projet MASTER I de SID :
Classification Non Supervisées

-- **Remarques** Vous devez rendre un document :

- Un programme source, écrit en Rmarkdown et qui rassemble le code commenté permettant de répondre aux questions ci-dessous.
La documentation sur Rmarkdown est disponible sur <https://rmarkdown.rstudio.com/>
- Pensez à mentionner noms et prénoms sur le script.
- **Le travail est par groupe (voir fichier groupe).**
- A rendre au plus tard le **03 Avril 2021** aux adresses emails suivants:
: **nourou01@yahoo.fr**

Chaque groupe propose les étapes suivantes :

1. Analyse descriptive des variables téléchargées
2. Analyse multivariées (ACP, AFC ou ACM) suivant la nature des données
3. Une méthode hiérarchique
4. Une méthode de k-means
5. Une méthode X (à choisir par vous)

Description des données

- Veuillez télécharger un jeu de données de votre choix dans l'un des plateformes suivant :
www.kaggle.com
zindi.africa
ou tout autre plateforme de téléchargement de données
- Décrire le contexte des données et le problématique posé
- Décrire les données

1 Question de cours

1. Comment mesure-t-on la ressemblance entre les individus ?
2. Le nuage est toujours centré. Pourquoi ? Quel est l'effet du centrage sur l'analyse du nuage des individus ? Quel type de liaisons évalue-t-on dans le nuage des variables ?

2 Travail à effectuer

- Faire le pre-processing (gestion des valeurs manquantes, des valeurs extremes, codage, re-codage, ...) nécessaire sur les données.
- Suivant la nature de vos données, proposer une méthode de classification et argumez
- Pour chaque méthode proposée, merci de répondre à ces questions.

2.1 Typologie des Individus

1. Analyser la matrice de distance entre les individus . Quelles sont les individus qui se ressemblent le plus ?
2. Etudier si les ressemblances ou les dissemblances des individus correspondent à des explications concernant la problématique
3. Estimer l'impact de chaque individus et interpréter.
4. Proposer une typologie des individus.
5. Visualiser les individus sur le plan 2d.
6. Faites une interprétation globale de vos résultats sur les individus.

2.2 Typologie des Variables

1. Analyser la matrice de corrélation entre les individus . Quelles sont les variables les plus proches ?
2. Etudier si les corrélations des variables correspondent à des explications concernant la problématique
3. Estimer l'impact de chaque variable et interpréter.
4. Proposer une typologie des variables.
5. Visualiser les variables sur le plan 2d.
6. Faites une interprétation globale de vos résultats sur les variables.
7. Faites une interprétation globale de vos données.

Faites une interprétation globale de vos données.