EFREI 2019/2020

L3APP-LSI

Analyse de Données

Travail pratique

**Analyse de Données**

**Réalisation d’une ACP**

A l’aide du logiciel R et RStudio

**Réalisation du travail**

Par équipes de 4 étudiants.

Composition des équipes à fournir en début de séance « TP ».

**Données brutes et analyse**

Un fichier global vous est fourni.

Vous devez décider à partir de ce fichier les données ce que vous souhaitez utiliser pour votre (vos) analyse(s). C’est-à-dire que vous devez décider des « variables » que vous allez utiliser et des « individus », et identifier en quelques mots l’objectif de votre analyse (ce que vous attendez comme type de résultat).

Attention : cette étape peut vous amener à manipuler le fichier, éventuellement en effectuant certaines sélections de données et/ou certains calculs de nouvelles données, afin de constituer le tableau ‘X’ des données initiales.

**Script R, analyse**

Vous devez développer un script R exécutable sur n’importe quelle machine disposant d’au moins la console R.

Ceci implique que l’utilisation des menus de RStudio est interdite. Seules des fonctions du langage R peuvent être utilisées.

Votre script doit effectuer les tâches suivantes (voir le modèle de rapport attendu) :

* Import du fichier ‘X’
* Calcul des statistiques de base sur ‘X’ (moyenne, variance, covariance, corrélation) et les afficher
* Calcul des données centrées-réduites « Z » et des indices de corrélation sur ces données
* Calcul et affichage des valeurs et vecteurs propres
* Tracé de l’ébouli/coude des valeurs propres pour identifier le nombre de vecteurs propres à utiliser pour l’analyse
* Calcul des composantes principales
* Tracé du cercle de corrélation pour l’interprétation des axes
* Tracé des schémas finaux

Si le script est bien pensé et réalisé, il doit pouvoir s’appliquer à n’importe quel tableau de données initiales ‘X’. Une fois ce script réalisé pour une première étude, il doit pouvoir être utilisé rapidement pour d’autres études.

Vous pourrez par exemple (hors travail demandé mais utile pour tester…) l’utiliser avec les données du bilan financier des groupes pétroliers en entrée.

Important :

Afin de pouvoir constituer le rapport d’analyse, il est nettement préférable que les tableaux et graphiques soient sauvegardés dans des fichiers qui pourront ensuite simplement être insérés dans un rapport MSWord.

**Résultat à fournir**

1. Le script R utilisé, dans un fichier « .R »
2. Pour chaque analyse effectuée : le fichier « .xls » ou « .csv » des données initiales
3. Le rapport d’analyse

Chaque fichier rendu doit être préfixé par le numéro de l’équipe, par exemple :

« 5-Rapport.pdf », « 10-Script.R », …