



Livret des Travaux Pratiques d'Analyse de Données

M1 WIC-MIASHS

Web, Informatique et connaissances

(Mathématiques et Informatique Appliquées aux Sciences de l'Homme et de la Société)

Intitulé du cours : Analyse des données 1

Chargé du cours

Professeur Mustapha RACHDI

Bureau : C008 du Bât. Michel Dubois

sur RDV

mustapha.rachdi@univ-grenoble-alpes.fr

Unité de Formation et de Recherche

Sciences de l'Homme et de la Société

Université Grenoble Alpes

Bât. Michel Dubois

UFR SHS, BP. 47

38040 Grenoble Cedex 09

Année universitaire

2020-2021

Table des matières

Table des matières	3
1 (re)Prise en main et algèbre	5
2 Mesure de la liaison entre une variable et un ensemble de variables	9
2.1 Exemple du cours	10
2.2 Etude du jeu de données Poids-Naissance	12
2.2.1 Régression linéaire simple	12
2.2.2 Régression linéaire multiple	16
2.3 Devoir	24
3 Analyse en Composantes Principales	33
3.1 Pratique de l'ACP	33
3.1.1 Notions de base	34
3.1.2 Calcul	35
3.1.3 Visualisation et interprétation	39
3.1.4 Biplot	60
3.2 Exemple sur les variétés d'eaux minérales	70
3.2.1 ACP avec d'autres fonctions de R	73
3.2.2 ACP "à la main"	73
3.2.3 Interprétation des résultats	74
3.3 Effet taille	74
3.4 Etude des données sur les pays l'OCDE	75
3.5 Devoir	76
4 Analyse Factorielle des Correspondances (AFC)	79
4.1 Pratique de l'AFC	79
4.1.1 Calcul	80
4.1.2 Visualisation et interprétation	84
4.1.3 Eléments supplémentaires	101
4.1.4 Filtrer les résultats	104
4.1.5 Outliers	105

4.1.6	Exportation des résultats	105
4.1.7	Résumé	106
4.1.8	Autres lectures	106
4.2	Exemple numérique	107
4.3	Etude des liens entre des catégories socio-professionnelles et le type d'hébergement	108
4.4	Devoir	109

Bibliographie **111**

5 Analyse Factorielle des Correspondances Multiples (AFCM ou ACM) **113**

5.1	Mise en oeuvre	113
5.1.1	Calcul	114
5.1.2	Visualisation et interprétation	117
5.1.3	Eléments supplémentaires	127
5.1.4	Filtrer des résultats	129
5.1.5	Exportation des résultats	131
5.1.6	Résumé	131
5.1.7	Autres lectures	132
5.2	Devoir	132

Bibliographie **134**

6 Classification Automatique

CAH et K-Means **137**

6.1	Objectifs de l'étude	137
6.1.1	Traitements réalisés	138
6.1.2	Fichier de données : Importation, statistiques descriptives et graphiques	138
6.2	Classification Ascendante Hiérarchique	138
6.2.1	La procédure <code>hclust()</code> de R (package <code>stats</code> - toujours chargée) . . .	138
6.2.2	Découpage en classes – Matérialisation des groupes	139
6.3	Méthode des centres mobiles	139
6.3.1	La procédure <code>kmeans()</code> de R (package <code>stats</code> également)	139
6.3.2	Aide à la détection du nombre adéquat de groupes	140
6.4	Analyses univariées et multivariées	140
6.4.1	Interprétation des classes : Statistiques comparatives	140
6.4.2	Analyse en composantes principales (ACP)	141
6.4.3	Compléter l'analyse à la lumière des résultats de l'ACP	142
6.5	Devoir	143

Chapitre 1

(re)Prise en main et algèbre

Les notions présentées dans ces annexes doivent être acquises. En effet, celles-ci devaient faire partie du programme de Licence. Pour les étudiants qui arrivent d'autres formations ou qui ne sont pas très familiers avec le calcul de ces notions sous R, sont invités à travailler sérieusement ces annexes. Bien entendu, les autres étudiants devraient trouver ceci très élémentaire, mais on ne sait jamais ! vraiment jamais !

1. Créer le répertoire `TP_M1MIASHS_AD` sur le bureau ou ...
2. Lancer ensuite R et modifier le répertoire de travail en allant dans `Fichier -> Changer le Répertoire Courant` et en choisissant le répertoire `Bureau/TP_M1MIASHS_AD` qui a été créé.
3. Ouvrir une fenêtre d'éditeur `Fichier -> Nouveau Script`.
4. Sauver le fichier dans le répertoire courant sous le nom `TP0.R` : `Fichier -> Sauver sous`
5. Pour les différentes questions, on peut utiliser un "copier-coller" à partir de ce document. *Il est fortement recommandé de saisir toutes les commandes dans la fenêtre ouverte de l'éditeur.* Pour exécuter les commandes saisies, il suffit de les sélectionner avec la souris et d'appuyer simultanément sur les touches `Ctrl` et `R`.
6. Pour inclure des commentaires dans le programme, ce qui est fortement recommandé, utiliser le caractère `#`. Tout ce qui suit le caractère `#` sera négligé lors de l'exécution.
7. Penser à sauvegarder régulièrement le contenu du fichier `TP0.R` en appuyant sur les touches `Ctrl` et `S`.

Exercice 1. (Rappels des commandes de base)

On définit trois vecteurs x , y et z par les commandes R suivantes :

`x = c(1, 3, 5, 7, 9) ; y = c(2, 3, 5, 7, 11, 13) ; z = c(9, 3, 2, 5, 9, 2, 3, 9, 1)`

Reproduire et comprendre les résultats des commandes R suivantes :

```
x + 2; y * 3; length(x); x + y; sum(x > 5); sum(x[x > 5]); sum(x > 5 | x < 3);
y[3]; y[-3]; y[x]; (y > 7); y[y > 7]; sort(z); sort(z, dec = TRUE); rev(z);
order(z); unique(z); duplicated(z); table(z); rep(z, 3)
```

Exercice 2. Construire une matrice comportant 9 lignes et 9 colonnes avec des 0 sur la diagonale et des 1 partout ailleurs (on pourra utiliser la commande `diag`).

Exercice 3. Créer deux vecteurs de dimensions quelconques. Créer un vecteur en insérant le second vecteur entre les 2ème et le 3ème éléments du premier vecteur.

Exercice 4. On définit un vecteur x par les commandes R suivantes :

```
x = c (4.12, 1.84, 4.28, 4.23, 1.74, 2.06, 3.37, 3.83, 5.15, 3.76, 3.23, 4.87,
5.96, 2.29, 4.58)
```

1. Créer un vecteur égal à x sans ses 4 premiers éléments.
2. Créer un vecteur égal à x sans ses 1er et 15ème éléments.
3. Créer un vecteur contenant les éléments de x dont les valeurs sont strictement supérieures à 2.57 et strictement inférieures à 3.48.
4. Créer un vecteur contenant les éléments de x dont les valeurs sont strictement supérieures à 4.07 ou strictement inférieures à 1.48.
5. Déterminer la coordonnée de la plus petite valeur des éléments de x .

Exercice 5. On considère les matrices :

$$A = \begin{pmatrix} -2 & 1 & -3 & -2 \\ 1 & 2 & 1 & -1 \\ -2 & 1 & 1 & -1 \\ -1 & -3 & 1 & 1 \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} 2 & -1 & 3 & -4 \\ 2 & -2 & 4 & -5 \\ -2 & 1 & 3 & -1 \\ -1 & -3 & 1 & -1 \end{pmatrix}$$

1. Montrer que A et B sont inversibles, puis calculer leurs inverses.
2. Vérifier que $\det(A^t) = \det(A)$, $\det(A^{-1}) = (\det(A))^{-1}$ et $\det(AB) = \det(A)\det(B)$.
3. Vérifier que $(A^{-1})^t = (A^t)^{-1}$, $(AB)^t = B^t A^t$ et $(AB)^{-1} = B^{-1} A^{-1}$.

Exercice 6. On considère la matrice :

$$A = \begin{pmatrix} 1 & 1 & 3 \\ 5 & 2 & 6 \\ -2 & -1 & -3 \end{pmatrix}$$

1. Montrer que A est nilpotente, i.e. il existe un entier n tel que A^n est la matrice nulle.
2. Remplacer la 3ème ligne de A par la somme des deux premières.

Exercice 7. On considère la fonction $f : \mathbb{R} \mapsto \mathbb{R}$ définie par $f(x) = ax^2 + bx + c$, où a , b et c sont trois réels inconnus tels que : $f(0.5) = 7$, $f(1) = 4$ et $f(1.5) = 5$.

1. Créer dans **R** la matrice X telle que les informations dont on dispose sur f se tra-

duisent sous la forme matricielle : $X\beta = r$, avec $\beta = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ et $r = \begin{pmatrix} 1 \\ 4 \\ 5 \end{pmatrix}$.

2. Montrer que X est inversible et calculer X^{-1} .
3. Déterminer a , b et c .
4. Calculer les valeurs propres et représenter les vecteurs de X .

Exercice 8. On considère la matrice B décrite par les commandes suivantes :

```
A = matrix(0, nrow = 5, ncol = 5)
```

```
B = abs(col(A) - row(A)) + 1
```

1. Montrer que B est inversible et calculer B^{-1} .
2. On considère le système linéaire à 5 réels inconnus : a , b , c , d et e , défini par :

$$\begin{cases} a + 2b + 3c + 4d + 5e = 1 \\ 2a + b + 2c + 3d + 4e = 2 \\ 3a + 2b + c + 2d + 3e = 2 \\ 4a + 3b + 2c + d + 2e = 3 \\ 5a + 4b + 3c + 2d + e = 2 \end{cases}$$

Résoudre ce système en utilisant la matrice B .

Exercice 9. On définit deux vecteurs x et y par les commandes **R** suivantes :

```
x = 1:6 ; y = 5:10
```

1. Remplacer les éléments de $x + y$ dont les valeurs sont supérieures à 11 par 1.
2. Calculer le produit scalaire de x et y .
3. On définit la matrice M par les commandes **R** suivantes : `M = matrix(1:36, nrow = 6)` Calculer Mx , xM , M^t et MM^t .

Exercice 10. Créer la matrice à 16 lignes (et 3 colonnes) :

$$A = \begin{pmatrix} 3 & 2 & 1 \\ 3 & 2 & 1 \\ \vdots & \vdots & \vdots \\ 3 & 2 & 1 \end{pmatrix}$$

Exercice 11. Proposer des commandes **R** renvoyant la matrice :

	John	Lilly	Stef	Bob	Anna	Marik	Boris
Poids	95	68	85	72	55	86	115
Taille	189	169	179	167	171	178	179

Exercice 12. On considère les matrices :

$$A = \frac{1}{4} \begin{pmatrix} 3 & 1 & \sqrt{6} \\ 1 & 3 & -\sqrt{6} \\ -\sqrt{6} & \sqrt{6} & 2 \end{pmatrix}, \quad B = -\frac{1}{3} \begin{pmatrix} -2 & -1 & 2 \\ 2 & -2 & 1 \\ 1 & 2 & 2 \end{pmatrix}$$

1. Montrer que A est orthogonale, i.e. AA^t est égale à la matrice identité.
2. Vérifier que $A^{-1} = A^t$.
3. Montrer que B est orthogonale.
4. Est-ce que A et B commutent, i.e. $AB = BA$?
5. Calculer $\det(A)$.
6. Calculer et représenter les valeurs et les vecteurs propres de A et B ??
7. Créer une nouvelle matrice C construite en remplaçant la 3ème ligne de A par la somme des deux premières. Cette matrice est-elle inversible??