Table des matières

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_SSD_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_SSD_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 selectionner 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 TP3.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 :

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

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 $(A^{-1})^t = (A^t)^{-1}$, $(AB)^t = B^tA^t$ 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 $f : \mathbb{R} \longmapsto \mathbb{R}$ 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 traduisent 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 ??

1.1 Travail personnel

Exercice 13. Un échantillon de dossiers d'enfants a été saisi. Ce sont des enfants vus lors d'une visite en 1ère section de maternelle en $1996 - 1997$ dans des écoles de Bordeaux (Gironde, France). L'échantillon est constitué de 152 enfants âgés de 3 ou 4 ans.
Considérer le jeu de donnée Poids-Naissance (cf. fichier `Poids_naissance.txt`). II s'agit ici d'expliquer la variabilité du poids de naissance de l'enfant en fonction des caractéristiques de la mère, de ses antécédents et de son comportement pendant la grossesse. La variable à expliquer est le poids de naissance de l'enfant (variable quantitative `BWT`, exprimée en grammes) et les facteurs étudiés (variables explicatives) sont : Age de la mère, Poids de la mère lors du dernier cycle menstruel, "Race" de la mère, Tabagisme durant la grossesse, Nombre d'antécédents de prematurité, Antécédents d'hypertension, Présence d'irritabilité utérine, Nombre de visites à un médecin durant le premier trimestre de la grossesse, Poids de naissance et Poids de naissance inférieur ou égal à 2500 g.

1. Lire les données et donner la matrice des données.
2. Le poids de la mère étant exprimé en `livres`, nous commençons par effectuer une transformation du `data.frame` des données pour recoder cette variable en kilogrammes (1 livre = 0.45359237 kg). Convertir dans les donnés les poids des mamans en kg.

Description	Unité ou Codage	Variable
Age de la mère	Années	`AGE`
Poids de la mère lors du dernier cycle menstruel	Livres	`LWT`
Race de la mère	1=Blanche ; 2=Noire ; 3= Autre	`Race`
Tabagisme durant la grossesse	Oui=1 ; Non =0	`SMOKE`
Nombre d'antécédents de prématurité	0=Non ; 1 = Un ; 2 =Deux ; etc	`PTL`
Antécédents d'hypertension	Oui=1 ; Non =0	`HT`
Presence d'irritabilité utérine	Oui=1 ; Non =0	`UI`
Nombre de visites à un médecin durant le premier trimestre de la grossesse	O=Aucune ; 1=Une ; etc.	`FVT`
Poids de naissance	Grammes	`BWT`
Poids de naissance inférieur ou égal à 2500 g	Oui=1 ; Non =0	`LOW`

TABLE 1.1 – Variables et codage du jeu de données : Poids_naissance (txt, xls, ...)

3. Extraire les variables et en faire des tris à plat.

Exercice 14.

1. Créer une data frame "acteurs" qui renvoie :

	Mort.à	Années.de.carrière	Nombre.de.films	Prénom	Nom	Date.du.décès
1	93	66	211	Michel	Galabru	04-01-2016
2	53	25	58	André	Raimbourg	23-09-1970
3	72	48	98	Jean	Gabin	15-10-1976
4	68	37	140	Louis	De Funès	27-01-1983
5	68	31	74	Lino	Ventura	22-10-1987
6	53	32	81	Jacques	Villeret	28-01-2005

2. Changer le nom de la 1ère colonne par : `Age.du.décès`.
3. Extraire la colonne `Prénom`.
4. Ordonner la data frame par ordre croissant suivant l'âge de la mort.

Exercice 15. Le goût d'un fromage dépend de la concentration de plusieurs composés chimiques, dont :

— la concentration d'acide acétique (variable $X1$),
— la concentration d'hydrogène sulfuré (variable $X2$),
— la concentration d'acide lactique (variable $X3$).

Pour 30 types de `fromage` (cf. le jeu de données dans le fichier fromages.txt), on dispose du score moyen attribué par des goûteurs (caractère Y).

1. Construire une data frame `w` constituée du jeu de données `fromages` avec les noms des colonnes.
2. Associer (attacher) les noms aux colonnes respectives. Taper `X1` pour voir si cela a marché.
3. Afficher les caractéristiques de `w`.
4. Donner les paramètres statistiques élémentaires pour les variables Y, $X1$, $X2$ et $X3$.
5. Faite les commandes : `pairs(w)`. Que cela renvoie t'il ?
6. Construire une nouvelle data frame `ww` des individus vérifiant $X1 > 5.1$ et $X3 < 1.77$.
7. Afficher les caractéristiques de `ww`.
8. A partir de `ww`, donner les paramètres statistiques élémentaires pour les variables Y, $X1$, $X2$ et $X3$.

Exercice 16. On travaille avec le jeu de données `airquality`, disponible dans `R`.

1. Charger les données et comprendre d'où elles émanent.
2. Afficher les noms des variables considérées.
3. Afficher le nombre de lignes et de colonnes.
4. Calculer les paramètres statistiques de base à l'aide de la commande `summary`.
5. Représenter la boîte à moustaches de la variable Ozone pour chaque mois avec la commande `plot`.
6. Créer une variable qualitative `saison` qui vaut `printemps` quand le mois est 5, `été` quand les mois sont 6, 7 et 8, et `automne` quand le mois est 9.
7. Proposer des commandes `R` permettant d'obtenir le graphique suivant :

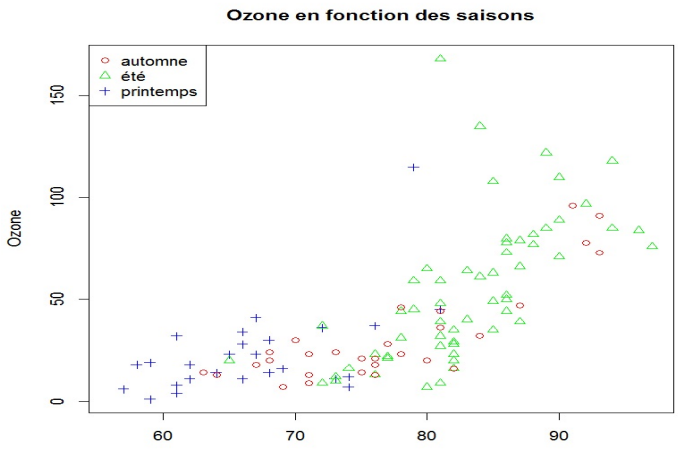

FIGURE 1.1 – Ozone en fonction des saisons

Exercice 17.

1. Simuler 100 valeurs $e_1, \ldots, e_{100}$ d'une *var* suivant la loi normale $\mathcal{N}(0, 5^2)$.
2. Pour tout $i \in \{1, \ldots, 100\}$, on pose $y_i = 1.7 + 2.1\, i + e_i$.
   (a) Représenter le nuage de points $\{(i, y_i)$ pour $i \in \{1, \ldots, 100\}\}$.
   (b) Sur ce même graphique, tracer en rouge la droite qui ajuste au mieux ce nuage de points.

Exercice 18. On considère un tableau de contingence obtenu en ventilant 592 femmes suivant la couleur de leurs yeux et la couleur de leurs cheveux.

	brun	chatin	roux	blond
marron	68	119	26	7
noisette	15	54	14	10
vert	5	29	14	16
bleu	20	84	17	94

1. Saisir les données du tableau ci-dessus.
2. Calculer la matrice des fréquences (arrondir au 100ème près).
3. Donner les lois marginales (nommer `c` pour le vecteur colonne et `r` pour le vecteur ligne).
4. Utiliser la commande `sweep` pour donner la matrice des profils lignes `L` (distributions conditionnelles en ligne).
5. Utiliser la commande `sweep` pour donner la matrice des profils colonnes `C` (distributions conditionnelles en colonne).
6. Calculer la distance de chi-deux entre les profils lignes.
7. Donner la matrice des taux de liaison (arrondir au 100ème près).
8. Faire un test permettant de juger de la liaison entre la couleur des yeux et la couleur des cheveux.