

L'objectif de ce TD est d'utiliser R pour réaliser l'ACP de l'exemple du TD précédent (Notes). R est un logiciel libre de calcul scientifique (voir <http://www.r-project.org/>). Voici tout d'abord les commandes de R qui peuvent vous être utiles.

Création d'un vecteur : avec la fonction `c`. Exemple :

```
> x=c(1,2,3, 5,1, 2,0.2 ,0.6, 7,2.5,3,4.2)
```

```
> y=1:5
```

Création d'une matrice avec la fonction `matrix`. Exemples:

```
> A=matrix(x,ncol=3,nrow=4)
```

Pour afficher la matrice A :

```
> A
```

```
> B=matrix(x,ncol=3,nrow=4,byrow=TRUE)
```

Quelques matrices particulières :

```
> matrix(0,ncol=3,nrow=3)
```

```
> diag(1,ncol=3,nrow=3)
```

Transposée d'une matrice

```
> C=t(B);
```

Multiplication de matrices

```
> E=C%*%A
```

Multiplication d'une matrice par un réel

```
> F=0.3*A;
```

Extraire des éléments d'une matrice

```
> A[1,2] // L'élément  $A_{12}$ 
```

```
> A[1,] // La première ligne
```

```
> A[,2] // La deuxième colonne
```

```
> D=A[1:3,1:3]
```

Somme de chaque colonne

```
> colSums(A)
```

Somme de chaque ligne

```
> rowSums(A)
```

Moyenne, variance et écart-type d'un vecteur (Pour les moyennes par colonne d'une matrice, utiliser `colMeans`, pour les variances et écarts-types par colonne d'une matrice, utiliser `sapply(matrice,sd)`)

```
> mean(x)
> var(x)
> sd(x)
```

Diagonalisation d'une matrice carrée

```
> vec=eigen(D)
```

la variable **vec** créée contient deux matrices **values** (qui contient les valeurs propres) et **vectors** qui contient des vecteurs propres unitaires associés.

```
> vec$values
> vec$vectors
```

Les Tableaux :

On peut créer un tableau à partir d'un fichier texte, chaque colonne étant séparée par une tabulation.

Créer un fichier texte contenant les notes des élèves.

On importe alors les données avec la commande :

```
> data=read.table(file.choose(),header=TRUE)
> data
```

data est un **data.frame** (tableau dont les colonnes ont des propriétés comme nom, type,...). On peut mélanger dans un **data.frame** les variables quantitatives, qualitatives et textuelles. Les variables des **data.frame** sont accessibles par leur nom :

```
> data$Maths
```

Les noms de colonne se trouve sur la première ligne. Pour créer les vecteurs correspondants aux colonnes du tableau, on utilise la fonction **attach** : **attach(data)** crée ici 5 vecteurs : **Maths**, **Science**, **Francais**, **Latin**, **DM**.

Pour sélectionner les 5 premiers élèves (5 premières lignes du tableau **data**)

```
> data[1:5,]
```

Pour sélectionner plusieurs colonnes (par exemple les colonnes 2, 3 et 4 du tableau **data**)

```
> data[,2:4]
> summary(data)
```

Tester les fonctions suivantes :

```
> plot(data)
> hist(data$Maths)
```

```
> boxplot(data)
> plot(Maths,Science)
```

Avec la fonction `read.table`, le tableau obtenu n'est pas une matrice. Pour le transformer en matrice il faut utiliser la fonction `as.matrix` :

```
> A=as.matrix(data)
```

Éléments de programmation :

Boucle for :

```
for (i in 1:n) {
  instruction
}
```

Boucle while :

```
while (condition) { }
```

Instruction conditionnelle :

```
if (condition) { }
```

Après avoir testé ces commandes, écrire (en utilisant l'éditeur) un programme qui réalise l'ACP du tableau de données du TD précédent (notes des élèves) :

1. Représenter les `boxplot` des variables. Calculer les moyennes, les variances et les écarts-types des variables.
2. Calculer la matrice X des données centrées (utiliser la fonction `scale`).
3. Calculer la matrice d'inertie, retrouver les écarts-types des cinq variables.
4. Quelle est l'inertie du problème (utiliser la fonction `diag`) ?
5. Calculer les valeurs propres et les vecteurs propres de la matrice d'inertie. Combien d'axes sélectionnez-vous ?
6. Expliquer comment on calcule les coordonnées de l'individu 1 sur les axes sélectionnés. Calculer les projections des individus sur les axes factoriels.
7. Expliquer comment on calcule les contributions et les cosinus carrés de l'individu 1. Calculer les contributions et les cosinus carrés des individus.
8. Représenter les différents résultats en utilisant la commande `plot` (voir l'aide associée).

9. Calculer les matrices $Y = {}^tX$ et $W = {}^tYY$.
10. Calculer les valeurs propres et vecteurs propres de W . Vérifier la relation entre les vecteurs propres de V et ceux de W .
11. En déduire les coordonnées des variables sur les axes factoriels et les représenter.
12. On veut reprendre l'étude avec les données sur les appartement loués par des étudiants dans le fichier **Appart.txt** : les variables étudiées sont les suivantes : le loyer mensuel en euros (**Loyer**), le nombre d'étudiants dans l'appartement (**NbPersonnes**), le nombre de chambres de l'appartement (**Chambres**), la distance de l'appartement avec l'université (**Distance**) et le loyer par personne (**LoyerParPersonne**).
 - (a) Tracer les diagrammes en boîte des 5 variables. Que constatez-vous ?
 - (b) Calculer la matrices des données centrées, la matrice d'inertie, les valeurs propres et les vecteurs propres associés. Combien d'axes sélectionnez-vous ? Représenter les individus et les variables sur les deux premiers axes factoriels.
 - (c) Calculer les cosinus carrés et les contributions. Interpréter les deux premiers axes factoriels. Conclusion ?
13. On souhaite maintenant que toutes les variables initiales aient exactement la même inertie égale à 1. Transformer les données pour que cela soit possible. Calculer la matrice Z des ces nouvelles données (utiliser la fonction **scale**).
14. Calculer la matrice d'inertie. Calculer les corrélations des données : quel est le lien avec la matrice d'inertie ?
15. Reprendre la question 12 pour ces nouvelles données. Quelles différences observe-t-on entre ces deux études ?