

Laboratorio 6: Paquete ade4

Setiembre 2023

Instrucciones: Trabaje las siguientes instrucciones en R, creando uno o varios scripts. Entregue en la plataforma Multiversa los scripts o el archivo Markdown, y genere un archivo Word o L^AT_EX con las respuestas.

Debe tener instalado el paquete **ade4**. El paquete fue creado por investigadores de Biometría y Biología Evolutiva de la Universidad de Lyon 1, Francia. El nombre de este paquete es un acrónimo de *Analyse de Données destinée d'abord à la manipulation des données Écologiques et Environnementales avec des procédures Exploratoires d'essence Euclidienne*, es decir, Análisis de Datos destinados en primer lugar a la manipulación de datos Ecológicos y Ambientales con procedimientos Exploratorios esencialmente Euclídeos. Ciertamente, el paquete está muy orientado a los análisis ecológicos, pero los métodos básicos son muy generales.

Para el Análisis Multivariado de Datos, se basan en el concepto de *Diagrama de Dualidad*, es decir, esquema de dualidad según nuestro texto; por ello los métodos factoriales de base usan el prefijo **dudi** para referirse al diagrama de dualidad (*duality diagram*).

Recuérdese que un diagrama de dualidad está descrito por el triplete $(\mathbf{X}, \mathbf{M}, \mathbf{D})$, donde \mathbf{X} es una matriz $n \times p$ de n individuos descritos como vectores de \mathbb{R}^p . Los autores de **ade4** suponen que la métrica \mathbf{M} es diagonal y es guardada como un vector de p dimensiones (en **ade4** usan la notación \mathbf{Q} para la métrica). Finalmente, \mathbf{D} es la métrica diagonal de pesos de los individuos.

El paquete puede funcionar a través de comandos de R o R-Studio, o bien a través de ventanas tipo GUI, cargando la librería **ade4TkGUI**. Veremos el primer caso primero.

Los métodos vistos en el curso, presentes en **ade4**, son:

- **dudi.pca**: realiza el Análisis en Componentes Principales.
- **dudi.coa**: lleva a cabo el Análisis de Correspondencias Simples.
- **dudi.mca**: calcular el Análisis de Correspondencias Múltiples.

Hacer lo siguiente, habiendo cargado el paquete `library(ade4)`

1. Análisis en Componentes Principales.

- (a) Cargar la tabla de datos: `data(USArrests)`
- (b) Hacer el ACP:

```
ACP.res <- dudi.pca(USArrests, scannf = FALSE, nf = 3)
```

De esta forma se hace un ACP con 3 componentes principales. La instrucción `verb+scannf = 3+` indica que no se le pregunte al usuario por el número de factores, sino que él lo debe indicar en los parámetros con la instrucción `nf = 3`. En caso de que el usuario no haga esta indicación, por ejemplo con la instrucción

```
ACP.res <- dudi.pca(USArrests)
```

entonces se le abrirá una pregunta para que indique el número de ejes. Por defecto, `nf = 2`; además, el ACP es centrado y estandarizado. El usuario puede escoger un ACP no centrado o no normado (con las opciones `cent = FALSE`, `norm = FALSE`), además puede cambiar los pesos de las variables o de los objetos.

El resultado, `ACP.res`, es del tipo **dudi**, es decir diagrama de dualidad, creado por los autores. Esto se puede chequear con el comando `is.dudi(ACP.res)`.

- (c) Rango de la matriz: `ACP.res$rank`

- (d) Número de factores: `ACP.res$nf`
- (e) Valores propios: `ACP.res$eig barplot(ACP.res$eig)`
- (f) Tabla de datos analizada (centrada y estandarizada en el caso por defecto): `ACP.res$tab`
- (g) Pesos de las columnas: `ACP.res$cw`
- (h) Pesos de las filas: `ACP.res$lw`
- (i) Coordenadas de las filas: `ACP.res$li`
- (j) Coordenadas estandarizadas de las filas: `ACP.res$l1`
- (k) Coordenadas de las columnas: `ACP.res$co`
- (l) Coordenadas estandarizadas de las columnas: `ACP.res$c1`
- (m) Se puede hacer una serie de gráficos interesantes entre cada variable y sus coordenadas en los ejes principales, con la instrucción `score(ACP.res)`. Se puede ver para cada variable el efecto de minimización de la suma de cuadrados de los errores, al desplegar rectas de regresión para cada variable. En este caso, se ve el gráfico con la primera componente principal.
- (n) Graficación del círculo de correlaciones: `s.corcircle(ACP.res$co)`
- (o) Graficación del plano principal: `s.label(ACP.res$li)`
- (p) Gráfico simultáneo de filas y columnas tipo biplot, incluyendo el gráfico de barras de los valores propios: `scatter(ACP.res)`
- (q) El usuario puede pedir a `ade4` que lea una talba usual del R con el parámetro `df`, es decir, *data frame*. El usuario puede cambiar los pesos de las filas con el parámetro `row.w` o cambiar el peso de las columnas con el parámetro `col.w`

2. Análisis de Correspondencias Simples.

- (a) Cargar el paquete: `library(ade4)` y cargar la tabla de datos sobre tareas en el hogar:
`data(housetasks):`


```
library(ade4)
data(housetasks)
housetasks
```
- (b) Test de chi cuadrado: `chisq.test(housetasks)`
- (c) Los perfiles fila:


```
round(housetasks/apply(housetasks,1,sum),2)
```

y los perfiles columna:


```
round(t((housetasks)/apply(t(housetasks),1,sum)),2)
```
- (d) Pesos de las filas:


```
round(apply(housetasks,1,sum)/sum(housetasks),2)
```

y pesos de las columnas


```
round(apply(housetasks,2,sum)/sum(housetasks),2)
```
- (e) Para el Análisis de Correspondencias Simples se usa la instrucción `dudi.coa`


```
z<-dudi.coa(df = housetasks, scannf = F, nf = 3)
```
- (f) Valores propios


```
round(z$eig,2)
round(z$eig/sum(z$eig)*100,2)
```

y gráfico de Valores Propios

```

inercia<-z$eig/sum(z$eig)*100
barplot(inercia,ylab="% de inercia",names.arg=round(inercia,2))
title("Valores Propios en %")

```

- (g) Plano factorial

```

scatter.coa(z, method = 1, sub = "Tareas en el Hogar", posieig = "none")

Otra representación del plano factorial

plot(z$li[,1],z$li[,2],type="n",xlab="Axe 1",ylab="Axe 2",xlim=c(-1.2,1.6))
text(z$li[,1], z$li[,2], label=row.names(housetasks))
text(z$co[,1], z$co[,2], label= colnames(housetasks),col="red")
title("Distribución de Tareas en el Hogar")
abline(h=0,v=0)

```

- (h) Coordenadas de los perfiles fila y columna:

```

round(z$li,2)
round(z$co,2)

```

- (i) Contribuciones absolutas de las columnas a la inercia

```

inertia.dudi(z,col.inertia = T)$col.abs

```

- (j) Contribución absoluta de las filas

```

inertia.dudi(z,row.inertia = T)$row.abs

```

- (k) Resolver los ejercicios 3 y 4 del capítulo 4 del libro con ade4.

3. Análisis de Correspondencias Múltiples.

- (a) Cargar tabla de datos: `data(ours)`

- (b) La tabla se ve con: `ours`

- (c) Ver las variables, modalidades y sus frecuencias: `summary(ours)`

- (d) Cajas de dispersión por variable y modalidad: `boxplot(dudi.acm(ours, scan = FALSE))`

- (e) Código disyuntivo completo: `ours.disyuntivo <- acm.disjonctif(ours)`

- (f) Tabla de Burt: `acm.burt(ours, ours)`

- (g) Resultados del ACM: `ours.acm <- dudi.acm(ours, scann = FALSE, nf = 3)`

- (h) Valores propios `apply(ours.acm$cr, 2, mean)` ó bien `ours.acm$eig[1:ours.acm$nf]`

- (i) Gráfico de valores propios:

```

inercia<-ours.acm$eig/sum(ours.acm$eig)*100
barplot(inercia,ylab="% de inercia",names.arg=round(inercia,2))
title("Valores Propios en %")

```

- (j) Representación del plano factorial

```

plot(ours.acm$li[,1], ours.acm$li[,2],type="n",xlab="Axe 1",ylab="Axe 2",xlim=c(-1.2,1.6))
text(ours.acm$li[,1], ours.acm$li[,2], label=row.names(ours))
text(ours.acm$co[,1], ours.acm$co[,2], label= colnames(ours.disyuntivo),col="red")
title("Distribución de 38 Usos")
abline(h=0,v=0)

```

- (k) Plano principal con intensidades: `s.value(ours.acm$li, ours.acm$li[,3])`

- (l) Otros aspectos: Relación entre los ACM de la tabla disyuntiva y la tabla de Burt:

```

bb <- acm.burt(ours, ours)
bbcoa <- dudi.coa(bb, scann = FALSE)
plot(ours.acm$c1[,1], bbcoa$c1[,1])
bd <- acm.disjonctif(ours)
bdcoa <- dudi.coa(bd, scann = FALSE)
plot(ours.acm$li[,1], bdcoa$li[,1])
plot(ours.acm$co[,1], dudi.coa(bd, scann = FALSE)$co[,1])

```

- (m) Resolver el ejercicio 6 del capítulo 5 del libro con ade4.
- (n) Considerar los datos del Comedor estudiantil del I.T.C.R. Hacer un A.C.M. con las primeras 9 variables como activas, y usando como suplementarias cualitativas: la razón, sexo y ocupación, y como suplementaria cuantitativa la edad.