

Análisis multivariado de datos en R

CAMPO ELÍAS PARDO Y JORGE ORTIZ *

1. Introducción

Desde la conferencia del profesor Cabrera (2002), en el Simposio de Estadística, nos interesamos en el R para la enseñanza y la investigación de la estadística. Las notas de Correa & Salazar (2000), fueron de gran utilidad para “romperle el hielo” al R. Tuvimos algún contacto con el ADE4 (análisis de datos ecológicos y “environnementales” (ambientales) con procedimientos exploratorios euclidianos) (Thioulouse, Chessel, Dolédec & Olivier 1997), cuando buscamos software para ejecutar el método STATIS (Lavit 1988). El grupo de Bioestadística de Lyon implementó en ADE4 en R y tiene a disposición gran cantidad de información tanto didáctica como de investigación en la página: <http://pbil.univ-lyon1.fr/ADE-4/ADE-4.html>.

La parte práctica del cursillo tiene por objeto ayudar a los participantes para utilizar el R, para la enseñanza de la estadística y como herramienta para la ejecución de los métodos estadísticos y la implementación de nuevos métodos. Se utiliza el *ADE4* bajo R para la ejecución de los métodos factoriales descriptivos básicos: análisis en componentes principales (ACP), análisis de correspondencias simples (ACS) y análisis de correspondencia múltiples (ACM). Los interesados en ejecutar el análisis factorial múltiple (AFM) (Escofier & Pagès 1992) tienen una motivación para empezar a usar el ADE4 bajo R.

En la edición de textos matemáticos el Latex (<http://www.miktex.org/>) se ha constituido en una herramienta fundamental, ampliamente difundida en nuestro medio por De-Castro (2003). La Revista Colombiana de Estadística se edita en Latex, otras publicaciones estadísticas también exigen las contribuciones en Latex, lo mismo que algunos congresos (por ejemplo el congreso del 2004 de la Federación Internacional de Sociedades de Clasificación (ICFS): <http://www.classification-society.org/ifcs04/>). Estos hechos motivan la inclusión de algunas instrucciones que utilizamos para pasar tablas y gráficos de R a Latex.

2. Instalación y uso de R

El R (R Development Core Team 2004) se encuentra en la página <http://www.R-project.org>, de donde se puede instalar o bajar y luego instalar. Para lo que sigue se supone que el R y sus paquetes adicionales, se han bajado y se encuentran disponibles en el disco duro o en un CD. Estas notas hacen referencia a la versión 1.9.1, la última liberada al momento de escribirlas y al ambiente *RGui* bajo Windows.

*Profesores Asociados, Departamento de Estadística. Universidad Nacional de Colombia - Bogotá; e-mail: cepardot@unal.edu.co, jeortizp@unal.edu.co

2.1. Instalación del núcleo básico

El R (versión 1.9.1) bajo windows se instala haciendo doble click en el archivo **rw1091.exe**.

Para iniciar R se hace click en el acceso directo **R 1.9.1** o en *Inicio* → *Programas* → *R* → *R 1.9.1*. En la pantalla de *RGui* aparece la ventana *Consola de R* con lo siguiente:

```
R : Copyright 2004, The R Foundation for Statistical Computing
Version 1.9.1 (2004-06-21), ISBN 3-900051-00-3
```

```
R is free software and comes with ABSOLUTELY NO WARRANTY. You are
welcome to redistribute it under certain conditions. Type
'license()' or 'licence()' for distribution details.
```

```
R is a collaborative project with many contributors. Type
'contributors()' for more information and 'citation()' on how to
cite R in publications.
```

```
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for a HTML browser interface to help. Type 'q()' to
quit R.
>
```

Ahora R está listo para recibir comandos. Es importante tener en cuenta que R hace diferencia entre mayúsculas y minúsculas. La ayuda está disponible desde la barra de menú de *RGui* y se puede acceder a ella de diferentes maneras. Una buena opción es: *Help* → *Html help*, que utiliza el navegador en uso para acceder a la ayuda deseada. A esta ayuda se puede acceder sin necesidad de abrir *RGui* con *Inicio* → *Programas* → *R* → *R 1.9.1 Help*. Otra opción es utilizar el comando de R *help(nombre de la función)* o, el equivalente, *?nombre de la función*, comando que crea una ventana de ayuda cada vez. Estas ventanas se pueden intercambiar desde el menú *Windows* de la barra de control de *RGui*. En el manual de referencia (*Help* → *Manuals* → *R Reference Manual*) se accede al ítem de ayuda deseado con el índice del Acrobat. Para preguntas específicas se puede utilizar la lista *r-help*, cuya suscripción se puede hacer en la página: <https://www.stat.math.ethz.ch/mailman/listinfo/r-help>.

La ayuda de R es generosa en ejemplos con los cuales se pueden apreciar los comandos en acción copiándolos y pegándolos en la consola de R. Se recomienda leer totalmente *An Introduction to R* (*Help* → *Manuals* → *An Introduction to R*), disponible en español en una versión un poco más antigua (R Development Core Team 2000), en la página <http://cran.r-project.org/doc/contrib>. Tres lecturas complementarias en español, disponibles en la misma página son (Paradis 2002), (Díaz 2003) y para gráficos (Correa & González 2002). El tiempo invertido en estas lecturas se revertirá en ahorro durante la utilización de R.

Para quienes adopten el R como su programa básico para la enseñanza y ejecución de métodos estadísticos es conveniente además suscribirse a la lista *r-help* en

<https://www.stat.math.ethz.ch/mailman/listinfo/r-help>

2.2. Instalación y uso del ADE4

En el menú de R hacer click en *Packages*, luego en *Install package(s) from local zip files ...*, buscar el archivo *ade4.1.2-2.zip* y hacer click en *Abrir*.

En la consola de R aparece:

```
> install.packages(choose.files('',filters=Filters[c('zip','All')]),
.libPaths()[1], CRAN = NULL)
package 'ade4' successfully unpacked and MD5 sums checked updating
HTML package descriptions
>
```

Lo que indica que la instalación de ADE4 ha tenido éxito y se encuentra disponible permanentemente junto con la ayuda *Html*. Para hacer disponibles las funciones, la ayuda con los comandos *help()* o *?* y los datos de los ejemplos de la librería *ade4*, es necesario ejecutar la instrucción de R, en cada sesión:

```
library(ade4)
```

Ahora estamos listos para utilizar R y ADE4 para ejecutar los métodos básicos de análisis multivariado de datos. Sin embargo es más cómodo utilizar un editor de texto para almacenar los comandos de R en forma organizada, como programas fuente.

2.3. Un editor para R

Para la elaboración y modificación de los programas en R estamos utilizando el WinEdt 5.3, el mismo que utilizamos como editor de Latex. Se puede obtener una versión de prueba en Internet (<http://www.winedt.com/>) y un programa que lo configura para R (<http://cran.r-project.org/contrib/extra/winedt/>).

La opción *a)* de instalación, recomendada en el archivo *ReadMe.txt* es:

1. Instalar R
2. Instalar WinEdt 5 (V. 5.2 o superior)
3. Instalar *SWinRegistry* (disponible en <http://www.omegahat.org/SWinRegistry>): *Packages → Install package(s) from local zip files...*
4. Instalar *RWinEdt_1.6-1.zip*: *Packages → Install package(s) from local zip files...*
5. Activarlo desde *Rgui* mediante *library(RWinEdt)*, con lo cual se entra en WinEdt y aparece el menú R-WinEdt en la barra de control de *Rgui*

2.4. Salidas de R a Latex

En Latex es posible colocar texto sin alterar su formato colocándolo entre los comandos: *begin{verbatim}* y *end{verbatim}*.

Con esta herramienta se pasan tanto las instrucciones de R, desde el editor, como los resultados desde la consola de R o un archivo de salida. Así esta editado este documento.

Las gráficas se graban desde la consola *R Graphics*, usando el menú contextual (botón derecho del mouse), opción *Save as postscript*, con extensión **.eps* y quedan listas para incluir en Latex.

El paquete *xtable* permite convertir algunas salidas de R en tablas de Latex. Se carga de la misma manera que que los otros paquetes: *Packages → Install package(s) from local zip files...* y se hace disponible con *library(xtable)*.

2.5. Objetos de R

Es muy importante tener presente los tipos de objetos que se manejan en R, ya que las funciones de R operan según los tipos de objeto que admitan como argumento. En este documento por ejemplo la función `plot` produce diferentes resultados:

- la figura 2 se obtiene con `plot(cafe)`, puesto que el objeto `cafe` es de tipo `data.frame`;
- `plot(acp$eig)` produce el gráfico de la figura 3, debido a que `acp$eig` es un vector numérico.

Con la función `ls()` se obtiene un listado de los objetos activos en R en ese momento y con `class(nombre-objeto)`, el tipo de un objeto particular. Por ejemplo:

```
class(cafe)
[1] "data.frame"
```

Los ejemplos colocados en R y en sus distintos paquetes son generalmente de tipo `list`, que es una lista de objetos, que pueden ser de diferente tipo.

Obviamente en los manuales se encuentra información sobre los tipos de objetos que maneja R, por ejemplo en el manual de introducción: *Help* → *Html help* → *An Introduction to R* → *Objects*.

R tiene funciones para convertir objetos de un tipo a otro y además cada una de las otras funciones hace claridad sobre los tipos de objetos que recibe como argumento y los objetos que produce.

2.6. Funciones y objetos de ADE4 para los métodos factoriales básicos

Para el tema que nos ocupa, el objeto principal de ADE4 es `dudi`, el cual se obtiene con la función `as.dudi`, que es llamada por las funciones que realizan cualquier método factorial, por ejemplo `dudi.pca`, `dudi.coa`, `dudi.mca`.

La función interna `as.dudi(df,col.w,row.w)` realiza el $\text{ACP}(\mathbf{X}, \mathbf{M}, \mathbf{D})$ donde:

df: `data.frame` con n filas y p columnas (\mathbf{X})

col.w: un vector numérico con los pesos de las filas ($\text{col.w}[i] = D[i, i]$)

row.w: un vector numérico con los pesos de las columnas ($\text{row.w}[j] = M[j, j]$)

La salida de `as.dudi` es un objeto `dudi` que es un `list` de los objetos:

tab: \mathbf{X}

cw: pesos de las filas ($D[i, i]$)

lw: pesos de las columnas ($M[j, j]$)

eig: valores propios, un vector con $\min(n, p)$ componentes

nf: entero, número de ejes guardados

c1: ejes principales (vectores propios en \mathbb{R}^p), `data frame` con p filas y nf columnas

l1: componentes principales, `data.frame` con n filas y nf columnas

co: coordenadas de las columnas, *data.frame* con p filas y nf columnas

li: coordenadas de las filas, *data.frame* con n filas y nf columnas

call: llamado original de la función *as.dudi*

Para obtener las demás ayudas a la interpretación de los métodos factoriales básicos la librería *ade4* tiene la función *inertia.dudi(dudi, row.inertia = TRUE, col.inertia = TRUE)*, *dudi* es un objeto de tipo *dudi*. La salida de esta función es un objeto *list* de los siguientes objetos:

TOT: valores propios y sus porcentajes

row.abs: contribuciones absolutas de las filas a la inercia de los ejes y total

row.rel: contribuciones relativas para las filas (cosenos cuadrado = calidad de la representación)

row.cum: contribución relativa de las filas a la inercia total

col.abs: contribuciones absolutas de las columnas a la inercia de los ejes y total

col.rel: contribuciones relativas para las columnas (cosenos cuadrado = calidad de la representación)

col.cum: contribución relativa de las columnas a la inercia total

Luego de corrido el código hasta la sección ??, se tienen los siguiente objetos en R:

```
> ls()
[1] "acp"      "acpcotex" "acpI"     "cafe"     "cafecor"  "cafetex"
>
```

El objeto *acp* es la salida de la función *dudi.pca* de *ADE4* y su tipo y objetos son:

```
> class(acp)
[1] "pca" "dudi"
> names(acp)
[1] "tab" "cw"  "lw"  "eig" "rank" "nf"  "c1"  "l1"  "co"  "li"
[11] "call" "cent" "norm"
```

En comparación con *dudi* el subtipo de objeto *pca*, *dudi* tiene además los objetos:

rank: rango(\mathbf{X})

cent: vector con medias de *df* matriz de entrada al ACP

norm: desviaciones estándar de *df*

El ACP normado de \mathbf{Y} es el $\mathbf{ACP}(\mathbf{X}, \mathbf{I}_p, \frac{1}{n}\mathbf{I}_n)$ con \mathbf{X} de término general:

$$x_{ij} = \frac{y_{ij} - \bar{y}_j}{s_j}$$

que se puede escribir en términos de objetos:

$$tab[i, j] = \frac{df[i, j] - cent[j]}{norm[j]}$$

El objeto *acpI* es la salida de la función *inertia.dudi*, cuyo tipo y objetos son:

```
> class(acpI)
[1] "list"
> names(acpI)
[1] "TOT"      "row.abs" "row.rel" "row.cum" "col.abs" "col.rel"
"col.cum"
```

Las funciones para la proyección de elementos ilustrativos y para la construcción de gráficos se muestran en los ejemplos. La función para ejecutar el análisis factorial múltiple (AFM) es *mfa*, recibe un objeto de tipo *ktab* y entrega un objeto de tipo *mfa*, que son específicos de *ade4*.

3. Análisis en componentes principales

En R existen varias posibilidades para ejecutar el ACP, aquí solo se ilustra la del paquete ADE4, seleccionada por ser la más cercana al programa SPAD (CISIA-CERESTA 2000), que es el que hemos venido utilizando para la enseñanza y utilización de los métodos estadísticos multivariados descriptivos.

3.1. Ejemplo café

Este ejemplo académico se utilizó en el documento (Cabarcas & Pardo 2001). Los datos conforman una tabla, de 10 preparaciones de café, contaminadas o no con 20 o 40 % de maíz o cebada. A las tazas de café se les miden varias características mediante un panel de degustación. Adicionalmente se tienen tazas preparadas con dos cafés comerciales. Los datos activos del análisis se tienen en el archivo *cafe.txt*:

```
IDEN;IntensidadAroma;Aroma;Cuerpo;AcidezTasa;Amargo;Astringencia
ExcelsoClaro;7.72;7;6.84;5.02;5.04;5.36
Claro40Maiz;6.02;5.42;6.22;4.34;4.6;4.78
Claro40Cebad;6.48;5.98;6.44;4.58;4.82;4.8
Claro20Maiz;6.82;6.44;6.7;4.62;4.38;4.8
Claro20Cebad;7.08;6.2;6.72;4.78;4.94;4.9
ExcelsoOscuro;7.66;7.42;6.98;5.12;5.18;5.22
Oscuro40Maiz;6.18;5.82;6.26;4;4.46;4.96
Oscuro40Ceba;6.84;6.56;6.82;4.3;4.96;4.84
Oscuro20Maiz;6.66;7.06;6.7;4.64;5;4.9
Oscuro20Ceba;7;6.7;7.04;4.6;4.88;5.18
```

El programa en R que encadena los procedimientos para realizar el ACP se encuentra en el archivo *cafe.R*. A continuación se presenta paso a paso, como una inducción a los nuevos usuarios de R. El símbolo *#* indica comienzo de comentario, el cual termina con el salto de línea. Para cada secuencia de instrucciones R se supone que se han ejecutado todas las anteriores.

3.1.1. Lectura de los datos

```
#-----
# Simposio04 - Ejemplo ACP-cafe
#-----
# activación de ade4
library(ade4)
```

```
# lectura de datos
cafe <- read.table("cafe.txt", header = TRUE, sep = ";",row.names=1)
# se leen los datos del archivo cafe.txt
# la primera fila contiene los nombres de las columnas
# los datos están separados por ;
# la primera columna contiene los nombres de las filas
#-----
```

3.2. Tabla de datos a Latex

Esta sección es para quienes escriben documentos en Latex. La tabla 1 son los datos del archivo *cafe.txt* obtenidos mediante las siguientes instrucciones de R. La salida (*cafetex*) se escribe en el archivo *cafe.tex*, de donde se copia para este documento. Es también posible copiar la tabla desde la consola de R, escribiendo *cafetex*.

```
#-----
# construcción de tabla de datos en latex
library(xtable) # activa paquete para convertir a Latex
cafetex <- xtable(cafe) #convierte cafe a Latex
caption(cafetex) <- "Datos del ejemplo cafe" # título para la tabla
label(cafetex) <- "t:cafe-datos"
# escribe tabla en archivo cafe.tex con título arriba
print.xtable(cafetex, type="latex", file="cafe.tex", append=FALSE,
  caption.placement="top")
#-----
```

Las líneas para el fuente de Latex obtenidas son:

```
% latex table generated in R 1.9.1 by xtable 1.2-3 package
% Mon Jun 28 18:57:41 2004
\begin{table}[ht]
\begin{center}
\caption{Datos del ejemplo cafe} \label{t:cafe-datos}
\begin{tabular}{rrrrrrr}
\hline
& IntensidadAroma & Aroma & Cuerpo & AcidezTasa & Amargo & Astringencia \\
\hline
ExcelsoClaro & 7.72 & 7.00 & 6.84 & 5.02 & 5.04 & 5.36 \\
Claro40Maiz & 6.02 & 5.42 & 6.22 & 4.34 & 4.60 & 4.78 \\
Claro40Cebad & 6.48 & 5.98 & 6.44 & 4.58 & 4.82 & 4.80 \\
Claro20Maiz & 6.82 & 6.44 & 6.70 & 4.62 & 4.38 & 4.80 \\
Claro20Cebad & 7.08 & 6.20 & 6.72 & 4.78 & 4.94 & 4.90 \\
ExcelsoOscur & 7.66 & 7.42 & 6.98 & 5.12 & 5.18 & 5.22 \\
Oscuro40Maiz & 6.18 & 5.82 & 6.26 & 4.00 & 4.46 & 4.96 \\
Oscuro40Ceba & 6.84 & 6.56 & 6.82 & 4.30 & 4.96 & 4.84 \\
Oscuro20Maiz & 6.66 & 7.06 & 6.70 & 4.64 & 5.00 & 4.90 \\
Oscuro20Ceba & 7.00 & 6.70 & 7.04 & 4.60 & 4.88 & 5.18 \\
\hline
\end{tabular}
\end{center}
\end{table}
```

3.2.1. Características distribucionales de las variables

Con *summary* se obtienen las estadísticas básicas que se muestran y se utiliza *apply* para calcular la desviación estándar. Los resultados se copian directamente de la consola de R (tabla 2).

Tabla 1: Datos del ejemplo cafe

	IntensidadAroma	Aroma	Cuerpo	AcidezTasa	Amargo	Astringencia
ExcelsoClaro	7.72	7.00	6.84	5.02	5.04	5.36
Claro40Maiz	6.02	5.42	6.22	4.34	4.60	4.78
Claro40Cebad	6.48	5.98	6.44	4.58	4.82	4.80
Claro20Maiz	6.82	6.44	6.70	4.62	4.38	4.80
Claro20Cebad	7.08	6.20	6.72	4.78	4.94	4.90
ExcelsoOscur	7.66	7.42	6.98	5.12	5.18	5.22
Oscuro40Maiz	6.18	5.82	6.26	4.00	4.46	4.96
Oscuro40Ceba	6.84	6.56	6.82	4.30	4.96	4.84
Oscuro20Maiz	6.66	7.06	6.70	4.64	5.00	4.90
Oscuro20Ceba	7.00	6.70	7.04	4.60	4.88	5.18

Con la función *boxplot* se obtiene la figura 1, donde se pueden comparar las distribuciones de las seis variables. Se observan notas más altas en las tres primeras variables con respecto a las tres últimas, mostrando una diferencia esos dos grupos de variables y justificando la normalización de los datos antes de realizar el ACP (ACP normado).

```
#-----
# impresión de estadísticas básica y boxplot
# la función cat titula el objeto que se imprime luego con print
# \n indica salto de linea
#-----
cat("\n Estadísticas básicas \n")
print(summary(cafe))
cat("\n Desviaciones estándar \n")
print(apply(cafe,2,sd))
#-----
par(ask=TRUE) # para pausa antes de la gráfica
boxplot(cafe,cex.axis=0.75)

# cex.axis modifica tamaño etiqueta ejes
#-----
```

Tabla 2: Café - estadísticas básicas

Estadísticas básicas						
IntensidadAroma	Aroma	Cuerpo	AcidezTasa	Amargo	Astringencia	
Min. :6.020	Min. :5.420	Min. :6.220	Min. :4.000	Min. :4.380	Min. :4.780	
1st Qu.:6.525	1st Qu.:6.035	1st Qu.:6.505	1st Qu.:4.400	1st Qu.:4.655	1st Qu.:4.810	
Median :6.830	Median :6.500	Median :6.710	Median :4.610	Median :4.910	Median :4.900	
Mean :6.846	Mean :6.460	Mean :6.672	Mean :4.600	Mean :4.826	Mean :4.974	
3rd Qu.:7.060	3rd Qu.:6.925	3rd Qu.:6.835	3rd Qu.:4.745	3rd Qu.:4.990	3rd Qu.:5.125	
Max. :7.720	Max. :7.420	Max. :7.040	Max. :5.120	Max. :5.180	Max. :5.360	
Desviaciones estándar						
IntensidadAroma	Aroma	Cuerpo	AcidezTasa	Amargo	Astringencia	
0.5577773	0.6173780	0.2812986	0.3332000	0.2623060	0.2052749	

3.3. Correlaciones y gráficas de dispersión

El espacio de las variables del ACP normado es una imagen de la matriz de correlaciones, que se obtiene en R con la función *cor*, cuando se aplica sobre un objeto de tipo *data.frame* o *matrix* de datos numéricos. La función *plot* sobre el mismo *data.frame*, *cafe* produce la figura 2.

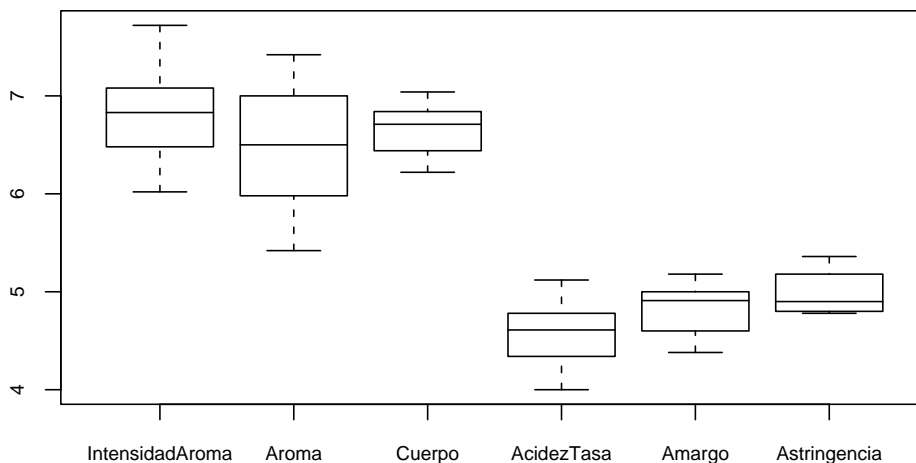


Figura 1: Café - boxplot

```

#-----
# matriz de correlaciones y matriz de gráficas de dispersión
#-----
cat("\n Matriz de correlaciones \n")
print(cor(caffe),2)
plot(caffe)
#-----

```

Las siguientes instrucciones son para mostrar la matriz de correlaciones en la tabla 3.

```

#-----
# Matriz de correlaciones en formato Latex
#-----
cafecor <- xtable(cor(caffe))#convierte cor(caffe) a Latex
caption(cafecor) <- "Café - matrix de correlaciones"
label(cafetex) <- "t:cafe-cor" # label de la tabla
# escribe tabla en archivo cafe.tex con título arriba
print.xtable(cafecor, type="latex", file="cafe.tex", append=TRUE,
  caption.placement="top")
#-----

```

3.4. ACP normado de *cafe*

Se realiza el ACP normado del ejemplo café con todas las variables como activas. Las funciones utilizados de *ade4* son *dudi.pca*, que realiza el ACP; *inertia.dudi*, que calcula las contribuciones

Tabla 3: Café - matrix de correlaciones

	IntensidadAroma	Aroma	Cuerpo	AcidezTasa	Amargo	Astringencia
IntensidadAroma	1.00	0.83	0.84	0.87	0.70	0.78
Aroma	0.83	1.00	0.86	0.72	0.71	0.66
Cuerpo	0.84	0.86	1.00	0.67	0.66	0.62
AcidezTasa	0.87	0.72	0.67	1.00	0.67	0.61
Amargo	0.70	0.71	0.66	0.67	1.00	0.56
Astringencia	0.78	0.66	0.62	0.61	0.56	1.00

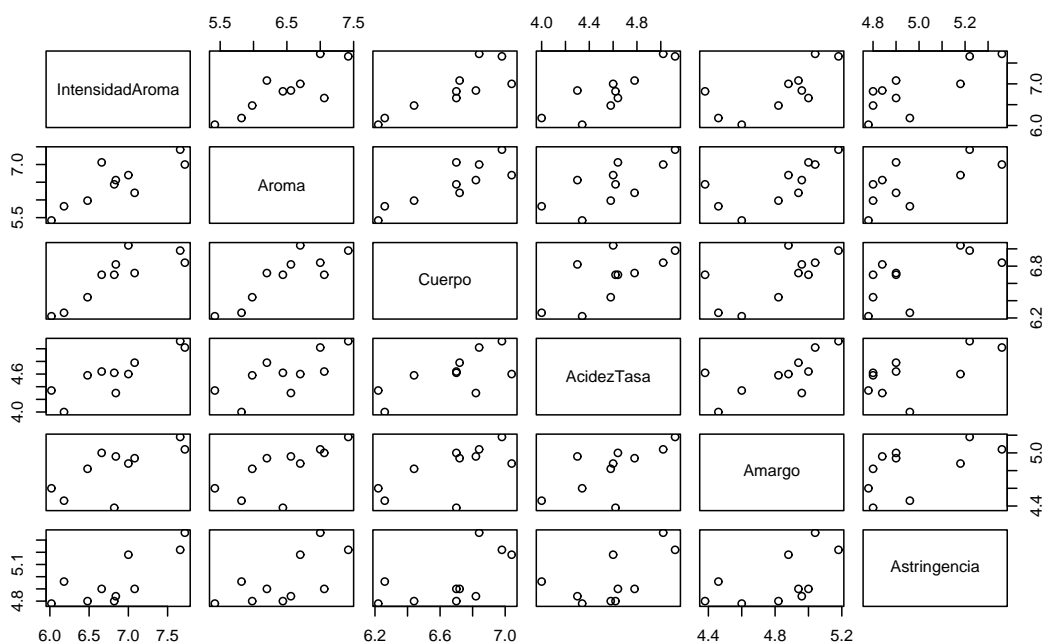


Figura 2: Café - gráficas de dispersión

absolutas y relativas a la inercia. Ver por ejemplo Lebart, Morineau & Warwick (1984). Las siguientes instrucciones producen las salidas numéricas del ACP, las cuales se han copiado desde la consola de R, a este texto y se han separado en tablas para facilitar su referencia:

```
#-----
# ACP normado del ejemplo café
#-----
acp <- dudi.pca(cafe,scannf=F,nf=2) # realiza el ACP
# acpI contiene las ayudas a la interpretación del ACP
acpI <- inertia.dudi(acp,row.inertia=T,col.inertia=T)
# impresión de objetos de acp y de acpI con títulos
cat("\n Valores propios \n")
print(acpI$TOT,2)
plot(acp$eig)

cat("\n Vectores propios \n")
print(acp$c1)
```

```

cat("\n Coordenadas de las columnas \n")
print(acp$co)
cat("\n Contribuciones de las columnas a los ejes \n")
print(acpI$col.abs/100)
cat("\n Calidad de representación de las columnas \n")
print(acpI$col.rel/100)
cat("\n Calidad de representación de las columnas en el plano \n")
print(acpI$col.cum/100)
cat("\n Coordenadas de las filas \n")
print(acp$li)
cat("\n Contribuciones de las filas a los ejes \n")
print(acpI$row.abs/100)
cat("\n Calidad de representación de las filas en los ejes \n")
print(acpI$row.rel/100)
cat("\n Calidad de representación de las filas en el plano \n")
print(acpI$row.cum/100)
#-----

```

La instrucción `plot(acp$eig)` produce la gráfica 3 de los valores propios, que sirve como complemento a la tabla 3. En este ejemplo las correlaciones (tabla 3) son positivas relativamente altas (entre 0.56 y 0.86), lo que explica que el resultado del ACP normado sea un primer de tamaño (4.6, 76.7% de la inercia). El primer componente principal se puede entonces tomar como un índice de la calidad del café, los coeficientes para calcularlo están en la columna CS1 de la tabla 5. El primer plano factorial conserva el 84.5% de la inercia total (6 = número de variables). El segundo eje separa un poco los cafés por *astringencia*, arriba y *amargo*, abajo.

Tabla 4: Valores propios

	inertia	cum	ratio
1	4.60147661	4.601477	0.7669128
2	0.46937103	5.070848	0.8451413
3	0.38451212	5.455360	0.9092266
4	0.34461612	5.799976	0.9666626
5	0.14487978	5.944856	0.9908093
6	0.05514434	6.000000	1.0000000

El ACP se llevó a cabo con `dudi.pca(cafe,scannf=F,nf=2)` si se usa `dudi.pca(cafe)` aparece en la pantalla una ventana con el “histograma” de valores propios (figura 3, derecha) y la consola de R espera un número que indica el número de ejes a retener. El usuario debe seleccionar tal número basado en el gráfico.

3.5. Gráficos del ACP

La fortaleza de los métodos descriptivos multivariados son las gráficas. Algunas de las salidas numéricas son insumo para las gráficas (coordenadas) y las otras son ayudas para su interpretación. En este ejemplo se hace un montaje de 4 gráficas (figura 4), arriba se presenta el círculo de correlaciones, construido con la función `s.corcircle` y el primer plano factorial de los cafés, construido con `s.label`. En la parte inferior está la posición de dos cafés comerciales proyectados como filas suplementarias y de nuevo el primer plano factorial de los cafés pero mostrando su agrupación por los tipos de contaminante de los cafés, utilizando la función `s.class`.

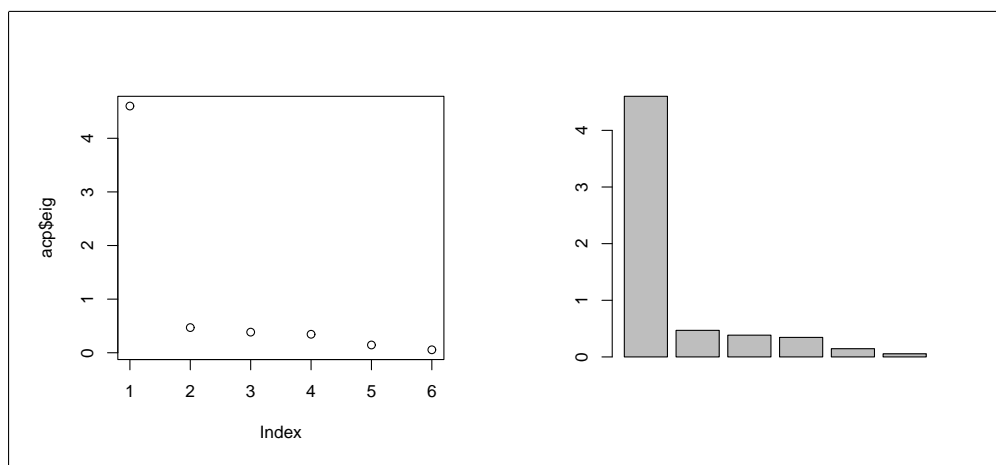


Figura 3: Café - valores propios

Tabla 5: Vectores propios

	CS1	CS2
IntensidadAroma	-0.4472587	0.1528586
Aroma	-0.4263091	-0.1445889
Cuerpo	-0.4144891	-0.1442163
AcidezTasa	-0.4040392	-0.1031639
Amargo	-0.3800191	-0.5342186
Astringencia	-0.3725008	0.7993107

A continuación se presentan las instrucciones de R que producen los dos gráficos de la parte superior de la figura 4:

```
#-----
# gráficas del acp
#-----
par(mfrow=c(2,2)) # para 4 gráficas simultáneas
s.corcircle(acp$co,sub="Cafe - Círculo de correlaciones",possub= "bottomright")
# se define iden para tener etiquetas más cortas en las gráficas
iden <- c("EC","C4M","C4C","C2M","C2C","EO","O4M","O4C","O2M","O2C")
# s.label es de ade4 y coloca puntos con etiquetas en planos factoriales
s.label(acp$li,label=iden,sub="Preparaciones de café",possub= "bottomright")
#-----
```

El círculo de correlaciones muestra el efecto “tamaño” del primer factor, con coordenadas negativas para los cafés mejor calificados, mostrando además a las variables *amargo* y *astringencia* como las más influyentes en el segundo eje. Este círculo es la clave para la lectura del primer plano factorial de los cafés: los mejores son los excelsior (EC, EO) y los peores los más contaminados con maíz (O4M, C4M). Los cafés más astringentes son el excelsior (EC) y el oscuro contaminado con 40% de maíz (O4M). Los café más amargo es el oscuro contaminado con el 20% de maíz (O2M).

Tabla 6: Coordenadas y ayudas a la interpretación de las columnas

Coordenadas de las columnas			
	Comp1	Comp2	
IntensidadAroma	-0.9594169	0.10472444	
Aroma	-0.9144779	-0.09905878	
Cuerpo	-0.8891227	-0.09880354	
AcidezTasa	-0.8667067	-0.07067830	
Amargo	-0.8151809	-0.36599671	
Astringencia	-0.7990534	0.54761303	

Contribuciones de las columnas a los ejes			
	Comp1	Comp2	
IntensidadAroma	20.00	2.34	
Aroma	18.17	2.09	
Cuerpo	17.18	2.08	
AcidezTasa	16.32	1.06	
Amargo	14.44	28.54	
Astringencia	13.88	63.89	

Calidad de representación de las columnas			
	Comp1	Comp2	con.tra
IntensidadAroma	-92.05	1.10	16.67
Aroma	-83.63	-0.98	16.67
Cuerpo	-79.05	-0.98	16.67
AcidezTasa	-75.12	-0.50	16.67
Amargo	-66.45	-13.40	16.67
Astringencia	-63.85	29.99	16.67

Calidad de representación de las columnas en el plano			
	Comp1	Comp2	remain
IntensidadAroma	92.05	93.14	6.86
Aroma	83.63	84.61	15.39
Cuerpo	79.05	80.03	19.97
AcidezTasa	75.12	75.62	24.38
Amargo	66.45	79.85	20.15
Astringencia	63.85	93.84	6.16

3.6. Elementos suplementarios

Las funciones *suprow.pca* y *supcol*, permiten el cálculo de filas y columnas suplementarias en ACP. En la gráfica se colocan con la misma función *s.label* con la opción *add.plot = T*. Para las variables nominales existe la función *s.clas*, que se utiliza para producir la gráfica inferior derecha de la figura 4.

En las siguientes instrucciones de R se introducen los valores de las variables para dos cafés comerciales, se calculan sus coordenadas como variables suplementarias y se producen las gráficas inferiores de la figura 4.

```
#-----
# posición de elementos suplementarios
#-----
```

Tabla 7: Coordenadas de las filas

	Axis1	Axis2
ExcelsoClaro	-2.9947077	1.01622318
Claro40Maiz	3.2057321	0.03615753
Claro40Cebad	1.3866259	-0.55662815
Claro20Maiz	0.9814076	0.21905587
Claro20Cebad	-0.3457128	-0.50136938
ExcelsoOscur	-3.5409698	-0.08826572
Oscuro40Maiz	3.0213005	1.11233544
Oscuro40Ceba	0.1375515	-0.84616260
Oscuro20Maiz	-0.4982973	-0.90730984
Oscuro20Ceba	-1.3529300	0.51596367

```
# variables para los dos cafés comerciales
co1 <- c(6.48,5.46,7.06,4.60,5.26,5.16)
co2 <- c(6.66,6.26,7.36,4.44,5.64,5.24)
com <- rbind(co1,co2) # se unen los dos en una matriz
colnames(com) <- names(cafe) # asigna nombres de columnas a com
corcom <- suprow.pca(acp,com) # coordenadas de columnas suplementarias
cat("\n Coordenadas de los cafés comerciales \n")
print(corcom)
# se repite gráfica de filas para posicionar cafes comerciales
s.label(acp$li,label=iden,clabel=1,sub="Posición de cafés comerciales",
        possub= "bottomright")
s.label(corcom,add.plot=T) # cafes comerciales sobre gráfica anterior
# gráfica de una variable nominal
# tipo de contaminante (E)xcelsior (sin), (M)añz, (C)ebada
tipo <- as.factor(c("E","M","C","M","C","E","M","C","M","C"))
# s.class muestra las subnubes según la variable tipo
s.class(acp$li,tipo,sub="Cafés según contaminación",possub= "bottomright")
#-----
```

En la gráfica inferior izquierda de la figura 4, se puede apreciar la calidad de los cafés comerciales (*co1*, *co2*) con respecto a los 10 que dieron origen al ACP. La gráfica inferior derecha de la figura 4 equivale a la proyección de la variable nominal *tipo*, con categorías E (excelsior, sin contaminación), C (cebada) y M (maíz). Se observa la mejor calidad de los cafés excelsior, seguidos de cebada y finalmente maíz. El lector puede producir otra gráfica clasificando los cafés por su porcentaje de contaminación (0,20 y 40).

Para completar el panorama fundamental del ACP falta proyectar al menos una variable suplementaria sobre el círculo de correlaciones, eso se consigue con las siguientes instrucciones de R: (figura 5).

```
#-----
# Impresión global como variable suplementaria
#-----
impresion <- c(7.46,6.24,6.12,6.04,6.22,7.40,5.90,6.94,6.90,7.16)
impresion <- as.data.frame(impresion) rownames(impresion) <-
rownames(cafe) corimp <- supcol(acp,scalewt(impresion))
par(mfrow=c(1,1)) # para volver a una gráfica
```

Tabla 8: Ayudas a la interpretación de las filas

Contribuciones de las filas a los ejes

	Axis1	Axis2
ExcelsoClaro	19.49	22.00
Claro40Maiz	22.33	0.03
Claro40Cebad	4.18	6.60
Claro20Maiz	2.09	1.02
Claro20Cebad	0.26	5.36
ExcelsoOscur	27.25	0.17
Oscuro40Maiz	19.84	26.36
Oscuro40Ceba	0.04	15.25
Oscuro20Maiz	0.54	17.54
Oscuro20Ceba	3.98	5.67

Calidad de representación de las filas en los ejes

	Axis1	Axis2	con.tra
ExcelsoClaro	-86.17	9.92	17.35
Claro40Maiz	93.83	0.01	18.25
Claro40Cebad	70.98	-11.44	4.51
Claro20Maiz	23.90	1.19	6.72
Claro20Cebad	-10.83	-22.78	1.84
ExcelsoOscur	-98.65	-0.06	21.18
Oscuro40Maiz	83.49	11.32	18.22
Oscuro40Ceba	0.95	-35.78	3.34
Oscuro20Maiz	-13.54	-44.90	3.06
Oscuro20Ceba	-55.13	8.02	5.53

Calidad de representación de las filas en el plano

	Axis1	Axis2	remain
ExcelsoClaro	86.17	96.09	3.91
Claro40Maiz	93.83	93.84	6.16
Claro40Cebad	70.98	82.42	17.58
Claro20Maiz	23.90	25.10	74.90
Claro20Cebad	10.83	33.61	66.39
ExcelsoOscur	98.65	98.71	1.29
Oscuro40Maiz	83.49	94.81	5.19
Oscuro40Ceba	0.95	36.73	63.27
Oscuro20Maiz	13.54	58.45	41.55
Oscuro20Ceba	55.13	63.15	36.85

```
s.corcircle(acp$co,sub="Cafe - Círculo de correlaciones",possub=
"bottomright",
           clabel=0.5)
s.arrow(corimp$cosup,add.plot=T)
#-----
```

En la figura 5 se observa una alta correlación entre el primer eje y la *impresión* global dada por el panel de degustadores.

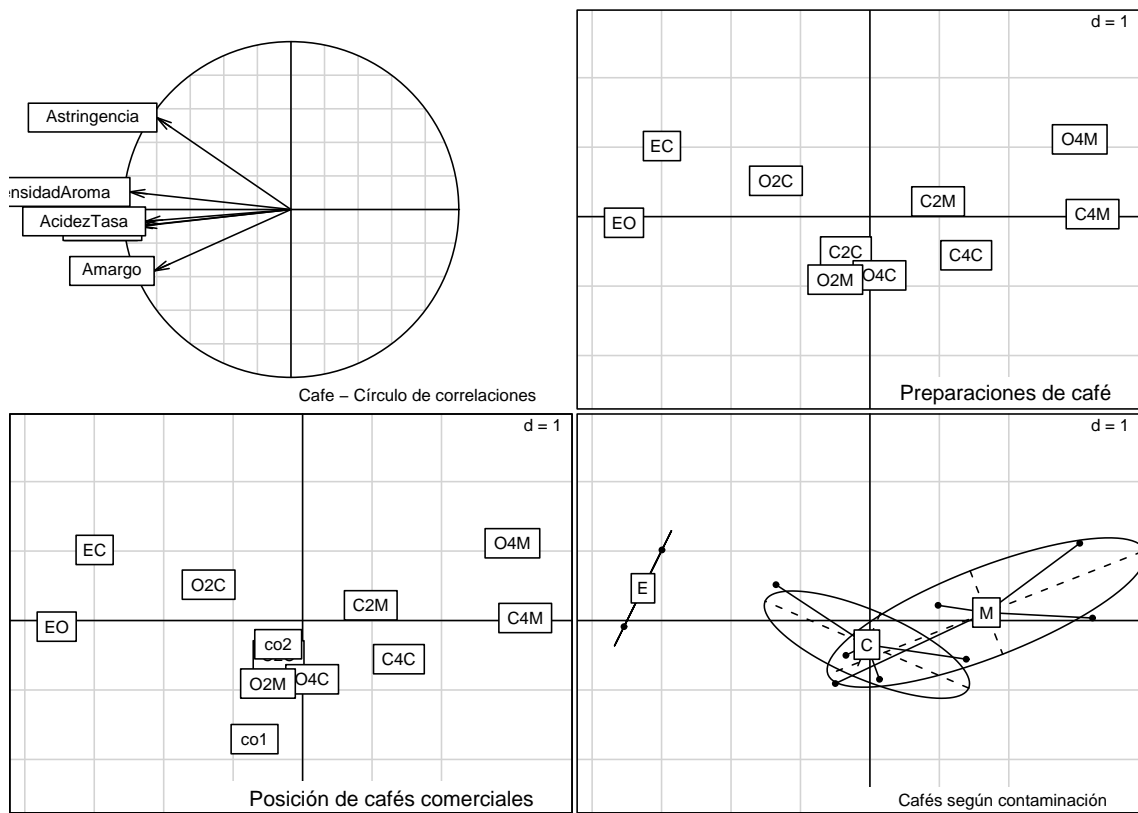


Figura 4: ACP de café

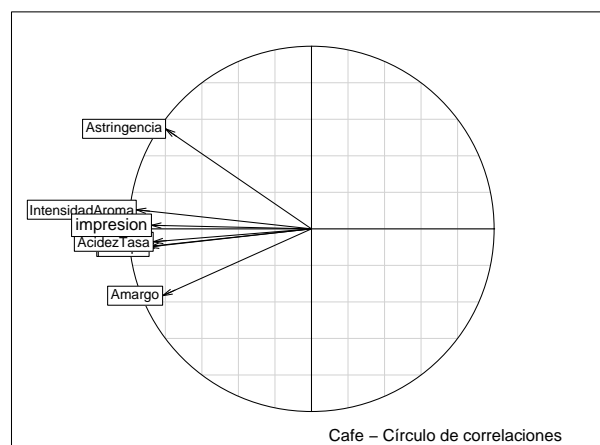


Figura 5: Cafe - proyección de impresión global

4. Análisis de correspondencias simples (ACS)

4.1. Ejemplo “Bogotá”: TC localidades x estrato

Para ilustrar el ACS se escogió una tabla de contingencia que clasifica las manzanas de Bogotá según la localidad (alcaldía menor) y el estrato a que pertenecen, ejemplo empleado en (Pardo 1999). Los datos están en el archivo *bogota.txt*. Las instrucciones de R están en el archivo *bogota.R*.

4.2. Lectura de los datos

Las siguientes instrucciones sirven para leer los datos y ponerlos en formato Latex para la tabla 9.

```
#-----
# Simposio04 - Ejemplo ACS-Estratificación de Bogotá
#-----
library(ade4)      # activación de ade4
# lectura de datos
bogota <- read.table("bogota.txt", header = TRUE, sep = ";", row.names=1)
# datos en formato Latex
library(xtable)   # activa paquete para convertir a Latex
# convierte bogota a Latex con formato entero
bogotatex <- xtable(bogota, display=c("s", rep("d", dim(bogota)[2])))
caption(bogotatex) <- "Datos del ejemplo Bogotá" # título para la tabla
label(bogotatex) <- "t:bogota-datos" # label de la tabla
# escribe tabla en archivo bogota.tex con título arriba
print.xtable(bogotatex, type="latex", file="bogota.tex", append=FALSE,
             caption.placement="top")
#-----
```

Tabla 9: Datos del ejemplo Bogotá

	SinEstrato	Estrato1	Estrato2	Estrato3	Estrato4	Estrato5	Estrato6
Usaquén	327	109	383	487	303	308	367
Chapinero	113	97	121	57	197	125	306
Santafé	91	40	439	145	25	1	0
SanCristóbal	150	225	1969	288	0	0	0
Usme	244	1593	1029	0	0	0	0
Tunjuelito	50	17	535	221	0	0	0
Bosa	219	263	2306	78	0	0	0
Kennedy	471	86	2077	1830	25	0	0
Fontibón	357	0	296	715	215	0	0
Engativá	334	67	857	2335	122	0	0
Suba	453	27	1982	955	373	482	112
BarrUnidos	98	1	1	675	387	15	0
Teusaquillo	148	0	2	156	625	38	0
LosMártires	79	0	32	584	4	0	0
AntonioNariñ	61	0	25	459	0	0	0
PteAranda	319	1	10	1836	0	0	0
LaCandelaria	42	0	73	39	0	0	0
RafaelUribe	182	239	1087	835	0	0	0
CiuBolívar	350	2946	1058	91	0	0	0

Con las siguientes líneas de R se convierte el *data.frame bogota* en un objeto *list* de una

tabla de las columnas activas *tab*, un vector con una columna ilustrativa *sinest*, un vector para las etiquetas de las columnas *lest* y un vector para las etiquetas de las filas *lloc*. Con esto queda estructurada la información para el análisis de correspondencias.

```
#-----
# conversión a lista tab = columnas activas, sinest = columna ilustrativa
#-----
sinest <- bogota[,1]
sinest <- as.data.frame(sinest)
rownames(sinest) <- rownames(bogota)
lest <- c("E1","E2","E3","E4","E5","E6")
lloc <- c("Usaq","Chap","StFe","SCri","Usme","Tunj","Bosa","Kenn","Font","Enga",
        "Suba","BUni","Teus","Mart","AnNa","PtAr","Cand","RaUr","CBol")
bogota <- list(tab=bogota[,2:7],sinest=sinest,lest=lest,lloc=lloc)
#-----
```

4.3. Análisis de correspondencias

Los cálculos del ACS se realizan mediante la función *dudi.coa* de *ade4*. Para ver los resultados se ha creado la función *afg.tex*, que llama la función *inertia.dudi* e imprime los resultados en consola y crea un archivo con la salida en varias tablas Latex. La función se encuentra en el archivo *afg.tex.R* y en la tabla 10 Con las siguientes instrucciones de R se obtienen las salidas numéricas:

```
#-----
# análisis de correspondencias
#-----
ac <- dudi.coa(bogota$tab,scannf=F,nf=2)
# impresión de resultados
source("afg.tex.R") # función para salida de ACS a Latex
# crea archivo Bogota.tex con tablas Latex
# e imprime en consola
afg.tex(ac,2,"Bogota")
#-----
```

En la edición se han agrupado las tablas para ahorrar espacio. Recuérdese que en el ACS de una tabla de contingencia (TC) se representan los perfiles fila y columna (distribuciones condicionales) en dos espacios que se visualizan en proyecciones simultáneas. En la tabla 11 se puede observar que el primer plano factorial retiene el 67% de la inercia, de modo que esta proyección es suficiente para observar las principales semejanzas y diferencias entre las distribuciones condicionales de las localidades según estratos, de los estratos según localidades y las correspondencias entre ellas. Las tablas que se presentan sirven para hacer las lecturas apropiadas en los gráficos.

4.4. Gráficas del ACS

Como resultado principal se hace un montaje de 4 gráficas (figura 6) y luego se hace un montaje de 20 gráficas una para cada localidad, adicionando al final la gráfica global para referencia (figura 7). Las instrucciones de R son las siguientes:

```
#-----
```

Tabla 10: Función *afg.tex*

```

#-----
# afg.tex - función para crear tablas en Latex de análisis factoriales
# C.E. Pardo Julio/04
# requiere paquete xtable
# entra afc de tipo "coa" "dudi", scannf=número de ejes,
# job = título del trabajo para las tablas
# se imprimen los resultados en consola
# se crea el archivo job.tex si existe se sobrescribe
#-----
library(xtable)
#-----
# función para una tabla
latex <- function(obj,job="latex",tit="",lab="",append=T){
  latex <- xtable(obj)
  caption(latex) <- paste(job,"-",tit)
  label(latex) <- paste("t:",job,"-",lab,sep="")
  print.xtable(latex, type="latex", file=paste(job,".tex",sep=""),
    append,caption.placement="top")
  # impresion en consola
  cat("\n",tit,"\n"); print(obj)
}
#-----
afg.tex <- function(afg,scannf=2,job=""){
  afgI <- inertia.dudi(afg,row.inertia=T,col.inertia=T)
  porvp <- diff(afgI$TOT$ratio,1)*100
  porvp <- c(afgI$TOT$ratio[1]*100,porvp)
  tvalp <-cbind(afgI$TOT,porvp)
  latex(tvalp,job,"Valores propios","valores-propios",F)
  latex(afg$c1,job,"Vectores propios","vectores-propios")
  latex(afg$co,job,"Coordenadas de las columnas",
    "coordenadas-columnas")
  latex(afgI$col.abs/100,job,"Contribuciones de las columnas",
    "cont-abs-columnas")
  latex(afgI$col.rel/100,job,"Calidad de representación de las columnas",
    "calidad-columnas")
  latex(afgI$col.cum/100,job,"Calidad de representación acumulada de las columnas",
    "cal-acum-columnas")
  latex(afg$li,job,"Coordenadas de las filas",
    "coordenadas-filas")
  latex(afgI$row.abs/100,job,"Contribuciones de las filas",
    "cont-abs-filas")
  latex(afgI$row.rel/100,job,"Calidad de representación de las filas",
    "calidad-filas")
  latex(afgI$row.cum/100,job,"Calidad de representación acumulada de las filas",
    "cal-acum-filas")
}
#-----

# gráficas del ACS
#-----
par(mfrow=c(2,2)) scatter.coa(ac,posieig = "none",clab.row =
0,clab.col = 0) cat("\n")

```

Tabla 11: Bogota - valores propios

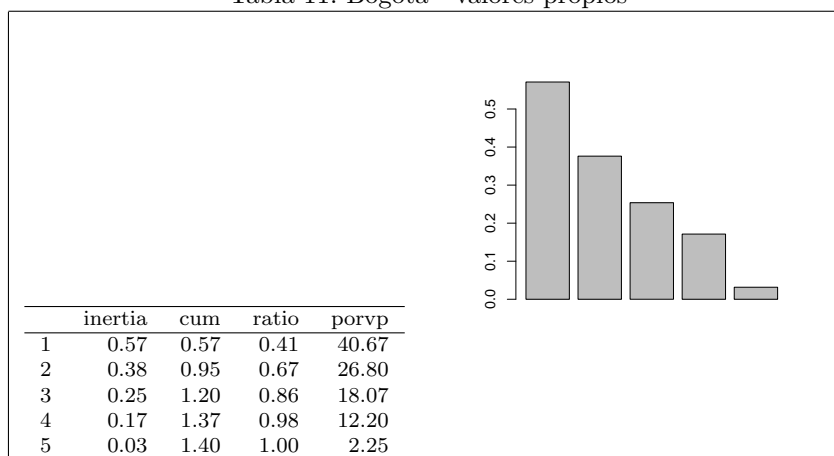


Tabla 12: Bogota - Coordenadas y ayudas a la interpretación de las columnas

Coordenadas				Contribuciones		
	Comp1	Comp2		Comp1	Comp2	
Estrato1	1.43	0.32		56.98	4.27	
Estrato2	0.21	-0.20		3.06	4.21	
Estrato3	-0.65	-0.44		24.07	16.98	
Estrato4	-1.05	1.40		12.21	33.20	
Estrato5	-0.62	1.34		1.85	12.98	
Estrato6	-0.69	2.21		1.83	28.36	

Calidad de representación				Calidad acumulada		
	Comp1	Comp2	con.tra	Comp1	Comp2	remain
Estrato1	81.76	4.04	28.34	81.76	85.81	14.19
Estrato2	11.01	-9.96	11.32	11.01	20.97	79.03
Estrato3	-52.72	-24.50	18.57	52.72	77.22	22.78
Estrato4	-24.84	44.52	19.98	24.84	69.36	30.64
Estrato5	-10.17	47.00	7.40	10.17	57.17	42.83
Estrato6	-5.17	52.84	14.38	5.17	58.01	41.99

```
s.label(ac$li,label=lloc,add.p=T,clab=0.8) scatter.coa(ac,posieig
= "none",clab.row = 0,clab.col = 0) cat("\n")
s.label(ac$co,label=lest,add.p=T,clab=0.8) scatter.coa(ac,posieig
= "none",clab.row = 0,clab.col = 0) cat("\n")
s.label(ac$li,label=lloc,add.p=T,clab=0.8)
s.label(ac$co,label=lest,add.p=T,clab=0.9) scatter.coa(ac,posieig
= "none",clab.row = 0,clab.col = 0) cat("\n")
s.label(ac$co,label=lest,add.p=T,clab=0.8)
s.label(supcol(ac,bogota$sinest)$cosup,label="SE",add.p=T,clab=1)
#-----

#-----
# una gráfica por cada localidad
#-----
```

Tabla 13: Bogota - Coordenadas y ayudas a la interpretación de las filas

Coordenadas			Contribuciones		
	Axis1	Axis2		Axis1	Axis2
Usaquén	-0.57	1.16	Usaquén	3.10	19.54
Chapinero	-0.54	1.99	Chapinero	1.29	26.48
Santafé	0.06	-0.26	Santafé	0.01	0.32
SanCristóbal	0.29	-0.29	SanCristóbal	1.03	1.59
Usme	1.26	0.19	Usme	20.27	0.68
Tunjuelito	-0.01	-0.42	Tunjuelito	0.00	1.01
Bosa	0.40	-0.25	Bosa	2.11	1.25
Kennedy	-0.21	-0.47	Kennedy	0.90	6.58
Fontibón	-0.67	-0.10	Fontibón	2.73	0.08
Engativá	-0.53	-0.49	Engativá	4.70	5.92
Suba	-0.31	0.25	Suba	1.90	1.87
BarrUnidos	-1.04	0.40	BarrUnidos	5.73	1.29
Teusaquillo	-1.26	1.70	Teusaquillo	6.33	17.70
LosMártires	-0.80	-0.68	LosMártires	1.94	2.12
AntonioNariñ	-0.80	-0.70	AntonioNariñ	1.50	1.75
PteAranda	-0.85	-0.72	PteAranda	6.49	7.02
LaCandelaria	-0.12	-0.46	LaCandelaria	0.01	0.18
RafaelUribe	0.02	-0.38	RafaelUribe	0.00	2.36
CiuBolívar	1.41	0.27	CiuBolívar	39.96	2.26

Calidad de representación				Calidad acumulada			
	Axis1	Axis2	con.tra		Axis1	Axis2	remain
Usaquén	-14.73	61.11	8.57	Usaquén	14.73	75.84	24.16
Chapinero	-5.01	67.81	10.46	Chapinero	5.01	72.82	27.18
Santafé	0.99	-19.23	0.44	Santafé	0.99	20.22	79.78
SanCristóbal	12.72	-12.90	3.31	SanCristóbal	12.72	25.61	74.39
Usme	92.93	2.06	8.87	Usme	92.93	94.98	5.02
Tunjuelito	-0.03	-38.76	0.70	Tunjuelito	0.03	38.79	61.21
Bosa	16.87	-6.60	5.09	Bosa	16.87	23.47	76.53
Kennedy	-15.14	-72.65	2.43	Kennedy	15.14	87.79	12.21
Fontibón	-68.66	-1.40	1.62	Fontibón	68.66	70.06	29.94
Engativá	-44.86	-37.25	4.26	Engativá	44.86	82.11	17.89
Suba	-17.88	11.63	4.31	Suba	17.88	29.51	70.49
BarrUnidos	-48.86	7.24	4.77	BarrUnidos	48.86	56.11	43.89
Teusaquillo	-18.96	34.95	13.57	Teusaquillo	18.96	53.92	46.08
LosMártires	-37.60	-27.04	2.10	LosMártires	37.60	64.64	35.36
AntonioNariñ	-36.50	-28.01	1.67	AntonioNariñ	36.50	64.51	35.49
PteAranda	-35.88	-25.57	7.36	PteAranda	35.88	61.45	38.55
LaCandelaria	-3.17	-49.19	0.10	LaCandelaria	3.17	52.35	47.65
RafaelUribe	0.20	-89.36	0.71	RafaelUribe	0.20	89.56	10.44
CiuBolívar	82.63	3.08	19.67	CiuBolívar	82.63	85.71	14.29

```

s.label(ac$li,label=lloc)
windows() # abre otra ventana de gráficos
par(mfrow=c(5,4))
for (i in 1:19) {
#   s.label(ac$li,clab=0)
#   s.label(ac$co)
  scatter.coa(ac,posieig = "none",clab.row = 0,
clab.col = 0)
  s.label(ac$co,label=lest,add.p=T)
  s.label(ac$li[i,],label=lloc[i],add.p=T,clab=1.5)
}
scatter.coa(ac,posieig = "none",clab.row = 0,clab.col = 0.8)
s.label(supcol(ac,bogota$sinest)$cosup,label="SE",add.p=T,clab=1.5)

```

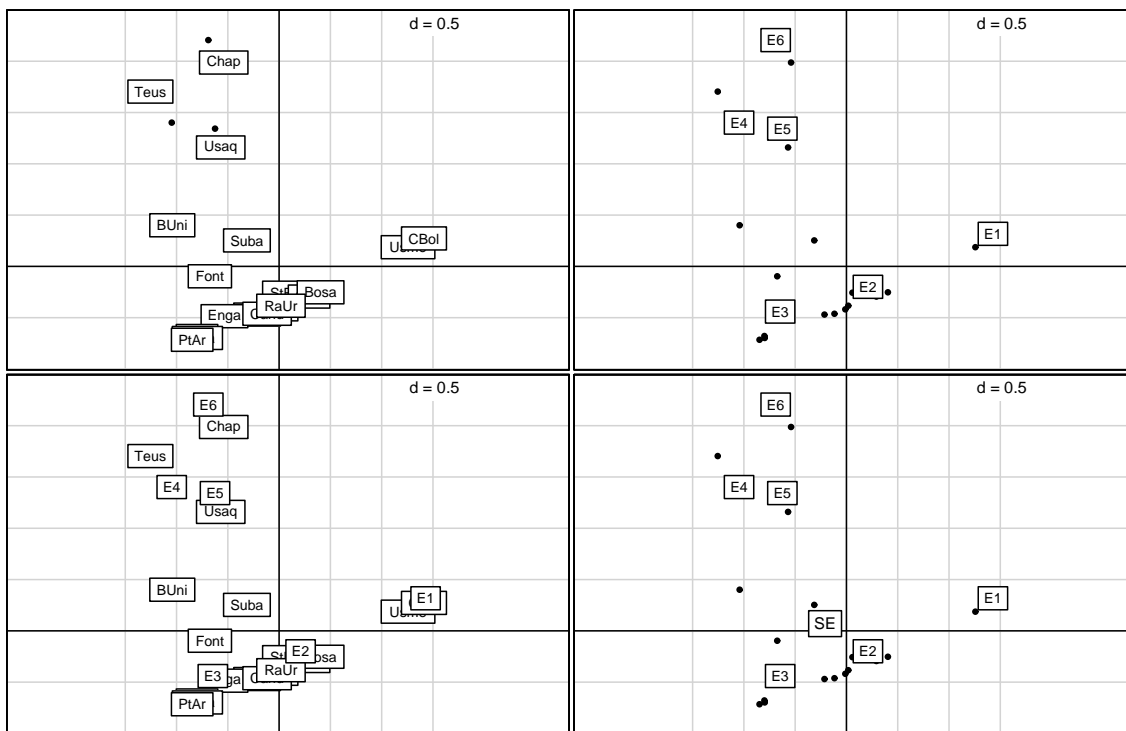


Figura 6: Bogotá - gráficas del ACS

#-----

Recuérdese que en el centro de las gráficas se encuentra, por un lado, la distribución de las manzanas de Bogotá según los estratos y por el otro según las localidades. Un perfil de categoría (localidad o estrato) cercano al centro, si está bien representado (tabla 13), está indicando una distribución parecida a la de la ciudad, por ejemplo *Suba* y el *Estrato2*. Alejarse del centro significa aumentar la proporción en algunas categorías y disminuir en otras. Por ejemplo *Usme* y *CiudadBolivar* tienen más proporción de estrato 1, su posición cercana indica perfiles muy parecidos. En el perfil de los estratos 4, 5 y 6 hay más proporción de las localidades Chapinero, Teusaquillo y Usaquén, en comparación con la distribución de las manzanas según localidad dentro de la ciudad. En realidad es suficiente la gráfica situada en la parte inferior izquierda de la figura 6, pues se trata de las dos gráficas superiores superpuestas. Sobre esta gráfica es más fácil la lectura simultánea que esta justificada por las relaciones de transición del ACS, que intuitivamente se puede entender como un "jalonamiento" mutuo. Por ejemplo *E1* está jalonando a *CBol* y *Usme* y recíprocamente *E1* está siendo jalonado por *CBol* y *Usme*. En la gráfica inferior derecha de la figura 6 se proyecta el perfil de la columna *Sin estrato* de la tabla de datos (tabla 9). Su posición cercana al centro está indicando una distribución muy parecida a las manzanas estratificadas de la ciudad.

La figura 7 se ha incluido para ubicar cada una de las localidades de la ciudad, aunque los puntos tapados de las localidades de la figura 6, se pueden dilucidar con las coordenadas de la tabla 13.

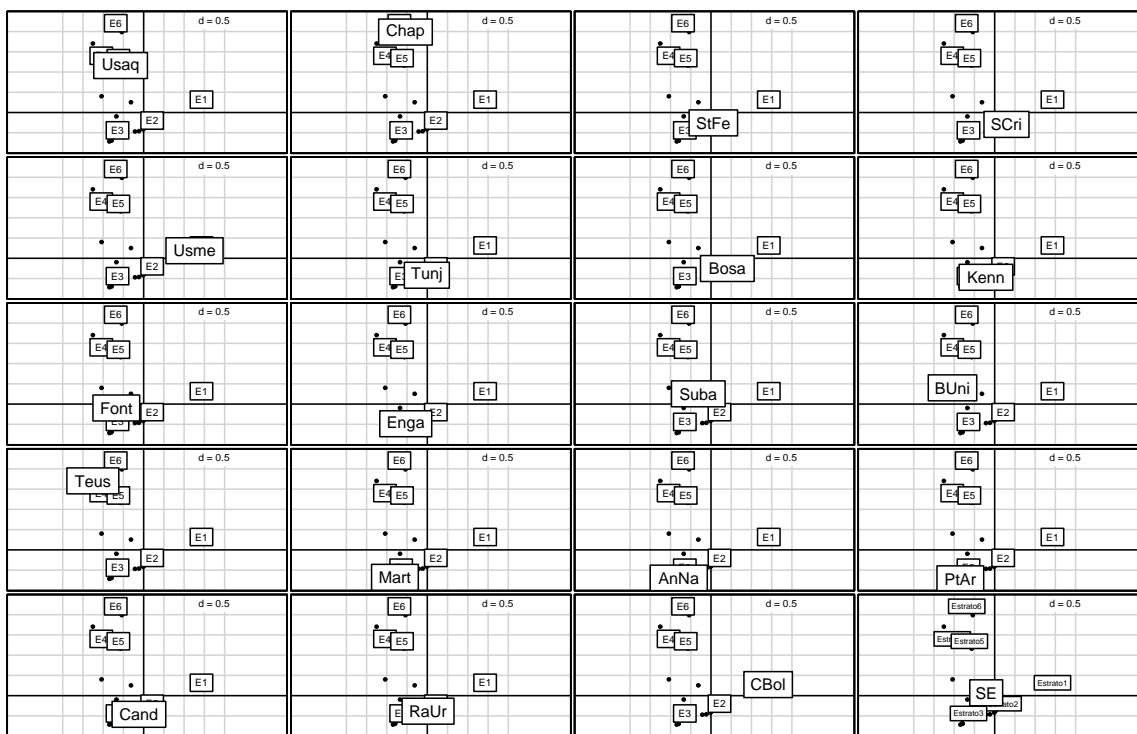


Figura 7: Bogotá - gráficas del ACS - localidades

5. Análisis de correspondencias múltiples (ACM)

5.1. Ejemplo razas de perros

Este ejemplo ha sido utilizado por Presta (<http://www.ulb.ac.be/assoc/presta/>), para la enseñanza del método. Se trata de 27 razas de perros que se han caracterizado por 6 variables físicas (tamaño, peso, velocidad) y psíquicas (inteligencia, afectividad, agresividad) y una variable adicional que clasifica a las razas de perros según su función (Fine 1996).

A continuación se muestra las instrucciones de R, que leen los datos del archivo *perros.txt*, la escribe en consola como una tabla Latex, escribe la TDC en formato Latex y finalmente convierte el objeto *perros* en una lista de *tab* y *funcion*. El ACM se realiza sobre *perros\$tab* y *perros\$funcion* se proyecta como ilustrativa.

```
#-----
# Simposio04 - Ejemplo ACM-Razas de perros
#-----
# activación de ade4
library(ade4)
# lectura de datos
perros <- read.table("perros.txt", header = TRUE,row.names=1)
library(xtable) # para convertir a Latex
print.xtable(xtable(perros)) # tabla latex en consola
# perros como list de tab (activas) y funcion (ilustrativa)
perros <- list(tab=perros[,1:6],funcion=perros[,7])
#-----
```

Las siguientes instrucciones de R ponen en el archivo *Perros.tex* la tabla disyuntiva completa (TDC, tabla 15) y la tabla de Burt (tabla 16). Los ACS sobre la TDC y sobre la tabla de Burt producen las mismas imágenes. En la tabla de Burt se pierde la información de los individuos. Las funciones *acm.disjonctif* y *acm.burt* son de la librería *ade4*.

```
#-----
# salida a Latex de TDC y Burt a archivo Perros.tex
#-----
# TDC en Latex print(xtable(acm.disjonctif(perros$tab),
  display=c("s",rep("d",dim(acm.disjonctif(perros$tab))[2])),
  file="Perros.tex")
# tabla de Burt print(xtable(acm.burt(perros$tab,perros$tab),
  display=c("s",rep("d",dim(acm.burt(perros$tab,perros$tab))[2])),
  file="Perros.tex",append=T)
#-----
```

5.2. ACM del ejemplo razas de perros

Las salidas numéricas (tablas 17 a 19) se obtienen con la función de ADE4 *dudi.mca* y se imprimen con la función *afg.tex* de la tabla 10. A este documento se copian las tablas y se agrupan para ahorrar espacio. Esto se consigue con las dos primeras instrucciones del siguiente código en R, las demás son para las gráficas:

Tabla 14: Razas de perros - datos

	TAM	PES	VEL	INT	AFE	AGR	FUN
bass	peq	liv	len	baj	baj	alt	caz
beau	gra	med	alt	med	alt	alt	uti
boxe	med	med	med	med	alt	alt	com
buld	peq	liv	len	med	alt	baj	com
bulm	gra	pes	len	alt	baj	alt	uti
cani	peq	liv	med	alt	alt	baj	com
chih	peq	liv	len	baj	alt	baj	com
cock	med	liv	len	med	alt	alt	com
coll	gra	med	alt	med	alt	baj	com
dalm	med	med	med	med	alt	baj	com
dobe	gra	med	alt	alt	baj	alt	uti
dogo	gra	pes	alt	baj	baj	alt	uti
foxx	gra	med	alt	baj	baj	alt	caz
foxt	peq	liv	med	med	alt	alt	com
galg	gra	med	alt	baj	baj	baj	caz
gasc	gra	med	med	baj	baj	alt	caz
labr	med	med	med	med	alt	baj	caz
masa	gra	med	alt	alt	alt	alt	uti
mast	gra	pes	len	baj	baj	alt	uti
peki	peq	liv	len	baj	alt	baj	com
podb	med	med	med	alt	alt	baj	caz
podf	gra	med	med	med	baj	baj	caz
poin	gra	med	alt	alt	baj	baj	caz
sett	gra	med	alt	med	baj	baj	caz
stbe	gra	pes	len	med	baj	alt	uti
teck	peq	liv	len	med	alt	baj	com
tern	gra	pes	len	med	baj	baj	uti

Tabla 15: Razas de perros - tabla disyuntiva completa

Raza	Tamaño			Peso			Velocidad			Inteligencia			Afectividad		Agresividad	
	gra	med	peq	liv	med	pes	alt	len	med	alt	baj	med	alt	baj	alt	baj
bass	0	0	1	1	0	0	0	1	0	0	1	0	0	1	1	0
beau	1	0	0	0	1	0	1	0	0	0	0	1	1	0	1	0
boxe	0	1	0	0	1	0	0	0	1	0	0	1	1	0	1	0
buld	0	0	1	1	0	0	0	1	0	0	0	1	1	0	0	1
bulm	1	0	0	0	0	1	0	1	0	1	0	0	0	1	1	0
cani	0	0	1	1	0	0	0	0	1	1	0	0	1	0	0	1
chih	0	0	1	1	0	0	0	1	0	0	1	0	1	0	0	1
cock	0	1	0	1	0	0	0	1	0	0	0	1	1	0	1	0
coll	1	0	0	0	1	0	1	0	0	0	0	1	1	0	0	1
dalm	0	1	0	0	1	0	0	0	1	0	0	1	1	0	0	1
dobe	1	0	0	0	1	0	1	0	0	1	0	0	0	1	1	0
dogo	1	0	0	0	0	1	1	0	0	0	1	0	0	1	1	0
foxx	1	0	0	0	1	0	1	0	0	0	1	0	0	1	1	0
foxt	0	0	1	1	0	0	0	0	1	0	0	1	1	0	1	0
galg	1	0	0	0	1	0	1	0	0	0	1	0	0	1	0	1
gasc	1	0	0	0	1	0	0	0	1	0	1	0	0	1	1	0
labr	0	1	0	0	1	0	0	0	1	0	0	1	1	0	0	1
masa	1	0	0	0	1	0	1	0	0	1	0	0	1	0	1	0
mast	1	0	0	0	0	1	0	1	0	0	1	0	0	1	1	0
peki	0	0	1	1	0	0	0	1	0	0	1	0	1	0	0	1
podb	0	1	0	0	1	0	0	0	1	1	0	0	1	0	0	1
podf	1	0	0	0	1	0	0	0	1	0	0	1	0	1	0	1
poin	1	0	0	0	1	0	1	0	0	1	0	0	0	1	0	1
sett	1	0	0	0	1	0	1	0	0	0	0	1	0	1	0	1
stbe	1	0	0	0	0	1	0	1	0	0	0	1	0	1	1	0
teck	0	0	1	1	0	0	0	1	0	0	0	1	1	0	0	1
tern	1	0	0	0	0	1	0	1	0	0	0	1	0	1	0	1

Tabla 16: Razas de perros - tabla de Burt

	Tamaño			Peso			Velocidad			Inteligencia			Afectividad		Agresividad	
	gra	med	peq	liv	med	pes	alt	len	med	alt	baj	med	alt	baj	alt	baj
TAM.gra	15	0	0	0	10	5	9	4	2	4	5	6	3	12	9	6
TAM.med	0	5	0	1	4	0	0	1	4	1	0	4	5	0	2	3
TAM.peq	0	0	7	7	0	0	0	5	2	1	3	3	6	1	2	5
PES.liv	0	1	7	8	0	0	0	6	2	1	3	4	7	1	3	5
PES.med	10	4	0	0	14	0	8	0	6	4	3	7	7	7	6	8
PES.pes	5	0	0	0	0	5	1	4	0	1	2	2	0	5	4	1
VEL.alt	9	0	0	0	8	1	9	0	0	3	3	3	3	6	5	4
VEL.len	4	1	5	6	0	4	0	10	0	1	4	5	5	5	5	5
VEL.med	2	4	2	2	6	0	0	0	8	2	1	5	6	2	3	5
INT.alt	4	1	1	1	4	1	3	1	2	6	0	0	3	3	3	3
INT.baj	5	0	3	3	3	2	3	4	1	0	8	0	2	6	5	3
INT.med	6	4	3	4	7	2	3	5	5	0	0	13	9	4	5	8
AFE.alt	3	5	6	7	7	0	3	5	6	3	2	9	14	0	5	9
AFE.baj	12	0	1	1	7	5	6	5	2	3	6	4	0	13	8	5
AGR.alt	9	2	2	3	6	4	5	5	3	3	5	5	5	8	13	0
AGR.baj	6	3	5	5	8	1	4	5	5	3	3	8	9	5	0	14

```

#-----
# ACM
#-----
acm <- dudi.acm(perros$tab,scannf=F,nf=2)
afg.tex(acm,2,"Perros")
#-----
# gráficas
par(ask=T,mfrow=c(1,1)) # para pausa antes de graficar
boxplot(acm) # boxplot del primer eje
scatter.acm(acm) # subnubes en el primer plano
# montaje de cuatro graficas
par(mfrow=c(2,2))
cat("\n")
s.label(acm$co) #1-columnas en primer plano
cat("\n")
s.label(acm$li,clab=1) #2-filas en primer plano
cat("\n")
s.label(acm$co) #3-filas y columnas en primer plano
s.label(acm$li,add.p=T,clab=1) # ""
cat("\n")
s.label(acm$co) #4-proyección de funcion como ...
s.class(acm$li,perros$fun,cellipse = 0,add.p=T) # ... ilustrativa
par(mfrow=c(1,1)) # vuelve a una gráfica
#-----

```

En la figura 1 muestra las posiciones de las categorías de cada variable en el primer eje factorial mediante boxplot. Se observa, sobretodo, una oposición de las razas por tamaño grandes (+) versus medianas y pequeñas (-); un ordenamiento por peso liviano (-) a pesado (+) y una oposición de las razas por agresividad alta (-) contra baja (+).

La figura 9 muestra las subnubes de las categorías por cada una de las variables, sobre el primer plano factorial. Se observa una muy buena separación de las categorías de tamaño y de peso y una buena separación para las de velocidad y afectividad.

En la figura 10 se presenta sobre el primer plano factorial del ACM: las categorías de todas las variables activas (arriba-izquierda), las 27 razas de perros (arriba-derecha), las categorías y

las razas simultáneamente (abajo-izquierda) y la proyección de la variable ilustrativa (función), superpuesta con las categorías activas. Esta es la síntesis gráfica del ACM. Se pueden leer las principales asociaciones entre las categorías de las variables activas y su relación con la función que cumplen las razas de perros.

Tabla 17: Perros - Valores propios

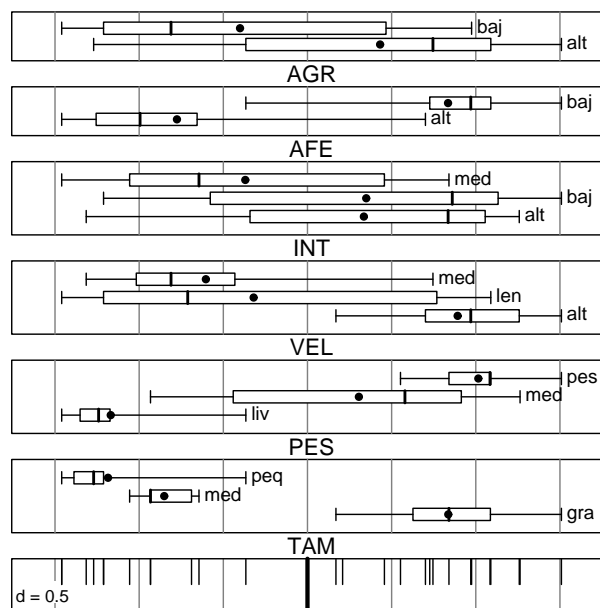
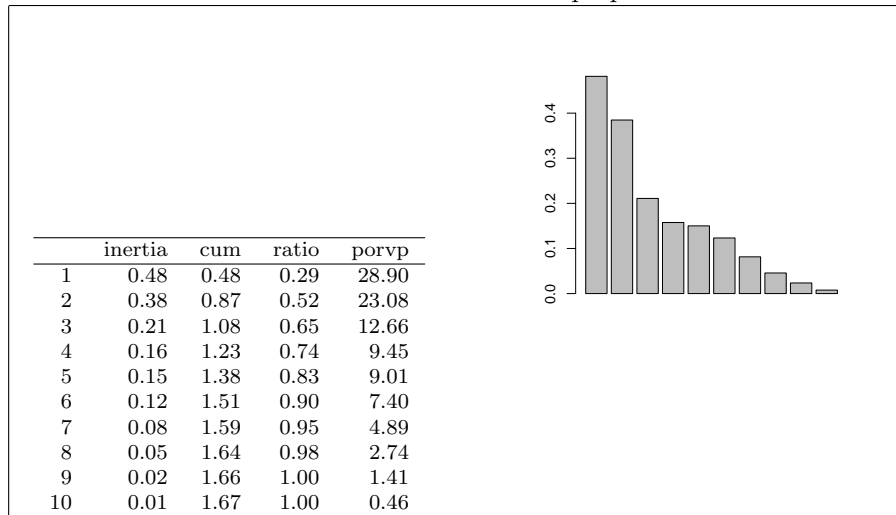


Figura 8: Perros - boxplot de categorías por variable

Tabla 18: Perros - Coordenadas y ayudas a la interpretación de las columnas

Coordenadas				Contribuciones			
	Comp1	Comp2		Comp1	Comp2		
TAM.gra	0.84	-0.02		TAM.gra	13.46	0.01	
TAM.med	-0.85	-1.23		TAM.med	4.64	12.17	
TAM.peq	-1.18	0.92		TAM.peq	12.60	9.59	
PES.liv	-1.17	0.82		PES.liv	14.01	8.72	
PES.med	0.31	-0.82		PES.med	1.67	15.06	
PES.pes	1.02	0.97		PES.pes	6.60	7.61	
VEL.alt	0.89	-0.37		VEL.alt	9.18	2.00	
VEL.len	-0.32	1.04		VEL.len	1.31	17.52	
VEL.med	-0.60	-0.89		VEL.med	3.74	10.12	
INT.alt	0.34	-0.46		INT.alt	0.86	2.03	
INT.baj	0.35	0.81		INT.baj	1.25	8.39	
INT.med	-0.37	-0.29		INT.med	2.27	1.70	
AFE.alt	-0.78	-0.27		AFE.alt	10.79	1.60	
AFE.baj	0.84	0.29		AFE.baj	11.62	1.72	
AGR.alt	0.43	0.21		AGR.alt	3.10	0.91	
AGR.baj	-0.40	-0.19		AGR.baj	2.88	0.85	

Calidad				Calidad acumulada			
	Comp1	Comp2	con.tra	Comp1	Comp2	remain	
TAM.gra	87.50	-0.05	4.44	TAM.gra	87.50	87.56	12.44
TAM.med	-16.46	-34.48	8.15	TAM.med	16.46	50.94	49.06
TAM.peq	-49.14	29.88	7.41	TAM.peq	49.14	79.02	20.98
PES.liv	-57.53	28.61	7.04	PES.liv	57.53	86.14	13.86
PES.med	10.04	-72.21	4.81	PES.med	10.04	82.26	17.74
PES.pes	23.42	21.56	8.15	PES.pes	23.42	44.98	55.02
VEL.alt	39.79	-6.91	6.67	VEL.alt	39.79	46.71	53.29
VEL.len	-6.02	64.22	6.30	VEL.len	6.02	70.25	29.75
VEL.med	-15.34	-33.19	7.04	VEL.med	15.34	48.53	51.47
INT.alt	3.21	-6.03	7.78	INT.alt	3.21	9.24	90.76
INT.baj	5.13	27.53	7.04	INT.baj	5.13	32.66	67.34
INT.med	-12.67	-7.57	5.19	INT.med	12.67	20.24	79.76
AFE.alt	-64.77	-7.67	4.81	AFE.alt	64.77	72.44	27.56
AFE.baj	64.77	7.67	5.19	AFE.baj	64.77	72.44	27.56
AGR.alt	17.29	4.06	5.19	AGR.alt	17.29	21.36	78.64
AGR.baj	-17.29	-4.06	4.81	AGR.baj	17.29	21.36	78.64

Tabla 19: Perros - Coordenadas y ayudas a la interpretación de las filas

	Coordenadas		Contribuciones		
	Axis1	Axis2	Axis1	Axis2	
bass	-0.25	1.10	bass	0.50	11.67
beau	0.32	-0.42	beau	0.77	1.68
boxe	-0.45	-0.88	boxe	1.54	7.48
buld	-1.01	0.55	buld	7.90	2.91
bulm	0.75	0.55	bulm	4.36	2.88
cani	-0.91	-0.02	cani	6.40	0.00
chih	-0.84	0.84	chih	5.44	6.85
cock	-0.73	0.08	cock	4.14	0.06
coll	0.12	-0.53	coll	0.11	2.66
dalm	-0.65	-0.99	dalm	3.22	9.44
dobe	0.87	-0.32	dobe	5.86	0.96
dogo	1.05	0.51	dogo	8.43	2.47
foxh	0.88	0.03	foxh	5.91	0.01
foxt	-0.88	0.14	foxt	5.98	0.19
galg	0.68	-0.08	galg	3.52	0.07
gasc	0.52	-0.11	gasc	2.06	0.12
labr	-0.65	-0.99	labr	3.22	9.44
masa	0.49	-0.46	masa	1.82	2.08
mast	0.76	0.89	mast	4.39	7.58
peki	-0.84	0.84	peki	5.44	6.85
podb	-0.48	-1.04	podb	1.76	10.35
podf	0.14	-0.52	podf	0.16	2.56
poin	0.67	-0.42	poin	3.49	1.73
sett	0.50	-0.38	sett	1.95	1.37
stbe	0.58	0.59	stbe	2.62	3.39
teck	-1.01	0.55	teck	7.90	2.91
tern	0.38	0.49	tern	1.13	2.27

	Calidad			Calidad acumulada			
	Axis1	Axis2	con.tra	Axis1	Axis2	remain	
bass	-3.38	63.49	4.24	bass	3.38	66.87	33.13
beau	8.86	-15.37	2.52	beau	8.86	24.23	75.77
boxe	-11.13	-43.25	3.99	boxe	11.13	54.39	45.61
buld	-62.45	18.39	3.65	buld	62.45	80.84	19.16
bulm	27.07	14.30	4.65	bulm	27.07	41.36	58.64
cani	-38.52	-0.01	4.80	cani	38.52	38.53	61.47
chih	-37.99	38.27	4.13	chih	37.99	76.26	23.74
cock	-27.92	0.32	4.28	cock	27.92	28.24	71.76
coll	1.24	-24.93	2.47	coll	1.24	26.17	73.83
dalm	-23.63	-55.30	3.94	dalm	23.63	78.93	21.07
dobe	48.76	-6.36	3.47	dobe	48.76	55.13	44.87
dogo	56.08	13.15	4.34	dogo	56.08	69.23	30.77
foxh	55.83	0.05	3.06	foxh	55.83	55.88	44.12
foxt	-43.63	1.08	3.96	foxt	43.63	44.71	55.29
galg	33.88	-0.51	3.00	galg	33.88	34.39	65.61
gasc	18.60	-0.89	3.20	gasc	18.60	19.50	80.50
labr	-23.63	-55.30	3.94	labr	23.63	78.93	21.07
masa	15.37	-14.02	3.42	masa	15.37	29.39	70.61
mast	30.00	41.36	4.23	mast	30.00	71.36	28.64
peki	-37.99	38.27	4.13	peki	37.99	76.26	23.74
podb	-10.50	-49.40	4.84	podb	10.50	59.89	40.11
podf	1.75	-22.21	2.66	podf	1.75	23.97	76.03
poin	29.46	-11.68	3.42	poin	29.46	41.13	58.87
sett	22.39	-12.53	2.52	sett	22.39	34.92	65.08
stbe	20.16	20.87	3.75	stbe	20.16	41.03	58.97
teck	-62.45	18.39	3.65	teck	62.45	80.84	19.16
tern	8.84	14.15	3.70	tern	8.84	22.99	77.01

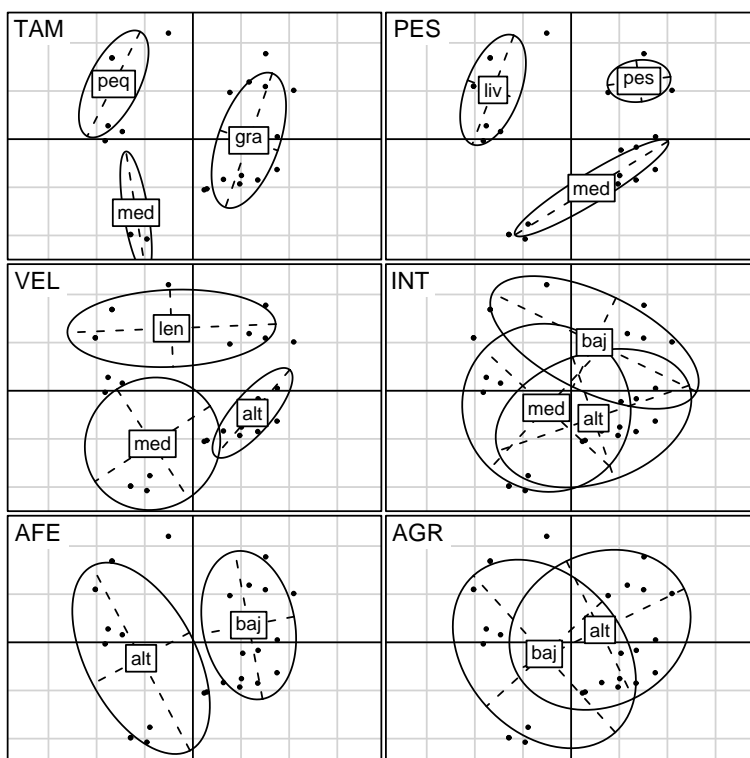


Figura 9: Perros - subnubes de categorías por variable

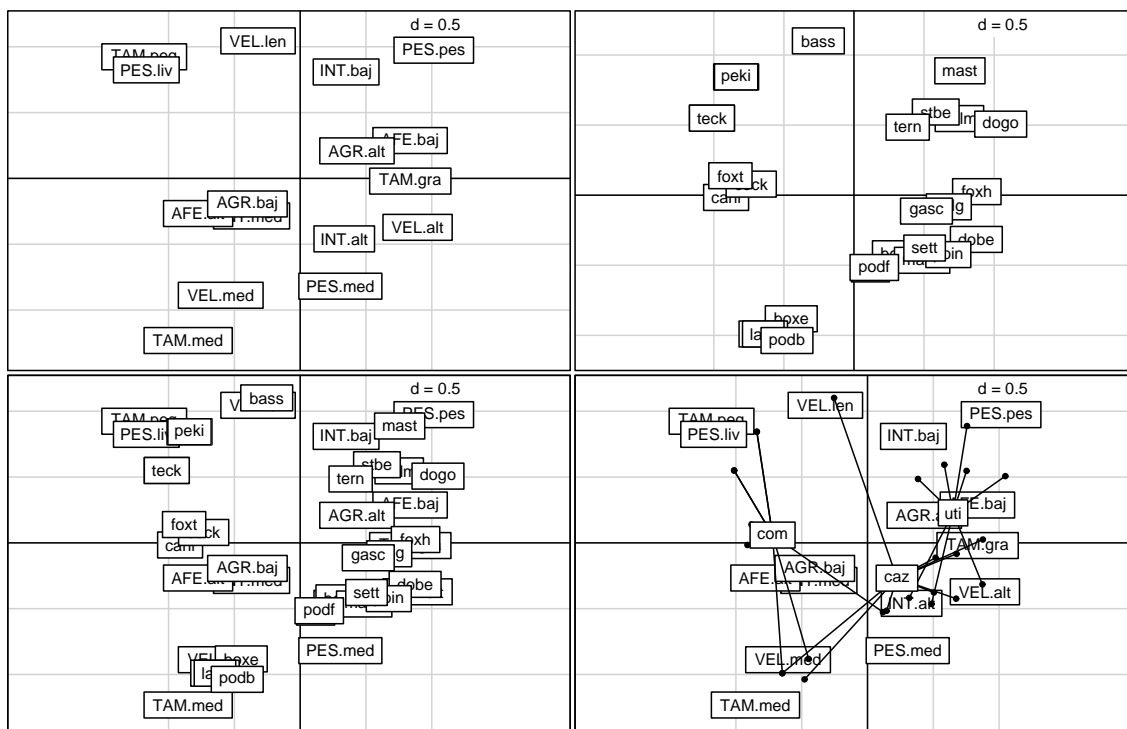


Figura 10: Perros - primer plano del ACM

Referencias

- Cabarcas, G. & Pardo, C.-E. (2001), 'Métodos estadísticos multivariados en investigación social', *Simposio de Estadística*.
- Cabrera, K. R. (2002), Aplicaciones en ciencias ambientales y del suelo utilizando el lenguaje estadístico R, *in* 'Memorias Simposio de Estadística 2002: Estadística Aplicada a las Ciencias Ambientales', Universidad Nacional de Colombia - Bogotá. Facultad de Ciencias. Departamento de Estadística.
- CISIA-CERESTA (2000), *SPAD. Versión 4.5. Manuel de prise en main*, Montreuil.
- Correa, J. C. & González, N. (2002), *Gráficos estadísticos con R*, Universidad Nacional de Colombia - Sede Medellín. Facultad de Ciencias Departamento de Matemáticas. Posgrado en Estadística.
*<http://cran.r-project.org/doc/contrib/grafi3.pdf>
- Correa, J. C. & Salazar, J. C. (2000), *R: Un Lenguaje Estadístico*, Universidad Nacional de Colombia - Sede Medellín. Facultad de Ciencias Departamento de Matemáticas. Posgrado en Estadística.
- De-Castro, R. (2003), *El Universo Latex*, 2 edn, Universidad Nacional de Colombia - Bogotá. Facultad de Ciencias. Departamento de Matemáticas, Bogotá.
- Díaz, R. (2003), Introducción al uso y programación del sistema estadístico R, Unidad de Bioinformática. Centro Nacional de Investigaciones Oncológicas (CNIO).
*<http://cran.r-project.org/doc/contrib/curso-R.Diaz-Uriarte.pdf>
- Escofier, B. & Pagès, J. (1992), *Análisis factoriales simples y múltiples. Objetivos, métodos e interpretación*, Universidad del País Vasco, Bilbao.
- Fine, J. (1996), Iniciación a los análisis de datos multidimensionales a partir de ejemplos, Folleto, PRESTA: Programme de recherche et a'enseignement en statistique appliquée, Sao Carlos.
- Lavit, C. (1988), *Analyse conjointe de tableaux quantitatives*, Masson, Paris.
- Lebart, L., Morineau, A. & Warwick (1984), *Multivariate Descriptive Statistical Analysis*, Wiley, New York.
- Paradis, E. (2002), *R para Principiantes*, Institut des Sciences de l'Evolution Universit Montpellier II, Montpellier. Traduciod por: Jorge A. Ahumada University of Hawaii & National Wildlife Health Center.
*http://cran.r-project.org/doc/contrib/rdebuts_es.pdf
- Pardo, C. E. (1999), Métodos estadísticos multivariados en investigación social, *in* 'Memorias IV Jornadas de Reflexión Investigativa', Universidad de Medellín, Medellín, pp. 65–90.
- R Development Core Team (2000), *Introducción a R*, R Foundation for Statistical Computing, Vienna, Austria.
*<http://www.R-project.org>
- R Development Core Team (2004), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3.
*<http://www.R-project.org>
- Thioulouse, J., Chessel, D., Dolédec, S. & Olivier, J. (1997), 'Ade-4: a multivariate analysis and graphical display software', *Stat. Comp.* **7**, 75–83.
*<http://pbil.univ-lyon1.fr/ADE-4/ADE-4F.html>