



LABORATORIO R

Análisis Multivariado I



Andrés Castrillejo

Copyleft © 2005 Andrés Castrillejo.

Permission is granted to copy, distribute, and/or modify this document under the terms of the GNU Free Documentation License, Version 1.1 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the appendix entitled “GNU Free Documentation License.”

Índice general

1. Tipificar Datos	5
1.1. Uso	5
2. Análisis de Grupos	6
3. Análisis de Componentes Principales	7
3.1. Entrada	7
3.2. Salida	8
3.3. Uso	8
4. Análisis de Correspondencias Simples	9
4.1. Entrada	9
4.2. Salida	10
4.3. Uso	10
5. Análisis de Correspondencias Múltiples	11
5.1. Entrada	11
5.2. Salida	12
5.3. Uso	12
6. Análisis Discriminante	13

Capítulo 1

Tipificar Datos

Para tipificar (“*estandarizar*”) los datos en la forma habitual puede invocarse a la función **standard**.

`standard(X)`

Esta función toma la matriz de datos X y opera por columnas restando la media y dividiendo por el desvío: $\frac{x_{ij} - \bar{x}_j}{s_j}$.

Ademas de la operación se agregan dos propiedades útiles:

- Si se quiere tipificar datos que están en una estructura de “data.frame”, el resultado tipificado también pertenecerá a la misma *clase* y conservará los *atributos*.
- La forma particular de tratamiento de los valores faltantes que es: *'pairwise.complete.obs'* que se describe a continuación desde la ayuda del **R**.

`help(var)`

...

If 'use' is 'all.obs', then the presence of missing observations will produce an error. If 'use' is 'complete.obs' then missing values are handled by casewise deletion. **Finally, if 'use' has the value 'pairwise.complete.obs' then the correlation between each pair of variables is computed using all complete pairs of observations on those variables.** This can result in covariance or correlation matrices which are not positive semidefinite.

The denominator $n - 1$ is used which gives an unbiased estimator of the (co)variance for i.i.d. observations. These functions return 'NA' when there is only one observation ...

1.1. Uso

```
>DatosST <- standard(DATOS[,3:9])
```

Capítulo 3

Análisis de Componentes Principales

La función **acp** realiza un Análisis de Componentes Principales sobre un conjunto de datos numéricos y presenta la salida en la manera vista en el curso.

El R ya posee una función para realizar ACP (*princomp*) en la que se basa la que se presenta a continuación.

```
acp(DATOS,Fsup=NULL,Csup=NULL,Ncomps=2)
```

3.1. Entrada

Los parámetros que pueden pasarse son los siguientes:

- PARÁMETRO OBLIGATORIO

DATOS Puede ser un “data.frame” o una “matrix” con variables cuantitativas, también puede ser una parte de estos de la forma `x[,3:12]`.

- PARÁMETROS OPCIONALES (ya tienen un valor por defecto)

Fsup Se suministran las filas suplementarias de la forma `c(i:i+3,j,j+1 ...,z)`.
Valor por defecto: NULL.

Csup Se suministran las columnas suplementarias de la forma `c(i,j,...,z)`.
Valor por defecto: NULL.

Ncomps Número de componentes (y cosenos²) que se desean retener (impresos y guardados). Valor por defecto: 2.

3.2. Salida

- Se genera un archivo de salida (“ACP_salida.txt”) que queda en el *Directorio de Trabajo* actual.
- La salida impresa presenta
 - tabla con los autovalores, desvios, proporciones de inercia, e inercia acumulada.
 - la matriz de correlaciones.
 - vectores propios.
 - matriz de saturación.
- Además existe la posibilidad de acceder a los “Componentes Disponibles” a través de una lista que contiene:
 1. Matriz de Saturacion
 2. Componentes
 3. Norma
 4. Cosenos2
 5. DatosFsup
 6. DatosCsup
 7. Ptos. Suplementarios Fila
 8. Ptos. Suplementarios Columna

Esto permite operar, graficar y guardar cada uno de estos elementos de la lista como variables independientes. Recordar que el acceso a cada elemento de esta lista compuesta por diferentes objetos (matrices y vectores) es de la forma $x[[i]]$.

3.3. Uso

```
> ComPrin <- acp(XdatSTD,Fsup=c(3,8,70:78),Ncomps=4)
```

```
##Ejemplo de como acceder a los elementos de 'Componentes Disponibles'
```

```
> NORM <- CompPrin[[6]]
```

Capítulo 4

Análisis de Correspondencias Simples

La función **acs** permite realizar un Análisis de Correspondencias Simples sobre una *Tabla de Contingencia*.

```
acs(xtab,Fsup=NULL,Csup=NULL)
```

Un ancestro de esta función fue creado por Oscar Gutierrez y Andrés Castillejo en el año 2001 para S-Plus. En ese programa también se facilitaba la creación de la *Tabla de Contingencia* a partir de dos variables (cuantitativas o cualitativas).

La versión actual es una variante basada originalmente en código de F. Murtagh profesor de Queen's University Belfast y del O.A. de la Université Louis Pasteur. Las modificaciones son principalmente en la forma en que se produce la salida.

4.1. Entrada

Los parámetros de entrada, son los siguientes:

- PARÁMETRO OBLIGATORIO
 - **xtab** “data.frame” conteniendo la Tabla de Contingencia §
- PARÁMETROS OPCIONALES (ya tienen un valor por defecto)
 - **Fsup** Se suministran las modalidades-filas suplementarias si las hay, en la forma: $c(i,j,z)$
 - **Csup** Se suministran las modalidades-columnas suplementarias si las hay, en la forma: $c(i,j,z)$

4.2. Salida

- Se genera un archivo de salida (“ACS_salida.txt”) que queda en el *Directorio de Trabajo* actual.
- La salida impresa presenta
 1. TABLA DE CONTINGENCIA
 2. PERFILES FILA
 3. PERFILES COLUMNA
 4. Test “Pearson’s Chi-squared”
 5. Tabla de “Inercia y Descomposicion Chi^2 ”
 6. Coordenadas de las FILAS
 7. Contribuciones parciales a la Inercia: FILAS
 8. Cosenos cuadrados: FILAS
 9. Coordenadas de las COLUMNAS
 10. Contribuciones parciales a la Inercia: COLUMNAS
 11. Cosenos cuadrados: COLUMNAS
 12. Tabla de Resumen para las modalidades de FILAS y COLUMNAS
- Se grafica la proyección de las *modalidades filas y columnas* en el plano principal.

4.3. Uso

```
>acorresp <- acs(datACS1,Fsup=c(3,5))
```

§Una manera de crear una Tabla de Contingencia

```
# hace TABLA DE CONTINGENCIA con las columnas 4 y 5 de base
a <- table(base[,4], base[,5])

# convierte la tabla a data frame de 3x4
# porque es la cantidad de modalidades de cada una
b <- data.frame(matrix(a,3,4))

# le pone los nombres de modalidades a las filas y las columnas
attr(b,'names') <- c('A','B','C')
attr(b,'row.names') <- c('mod1','mod2','mod3','mod4')
```


Capítulo 5

Análisis de Correspondencias Múltiples

La función **acm** permite realizar un Análisis de Correspondencias Múltiples sobre un conjunto de variables de tipo cualitativo *factor*.

$$resultado <- -acm(DATOS, NF = 2, Csup = NULL, ByG = F)$$

La versión actual de la presente función y la forma de realizar el Análisis de Correspondencias Múltiples provienen de la *biblioteca ade4*¹

Esta función se basa en la función *dudi.acm* de la biblioteca *ade4* y calcula y presenta además las Contribuciones y Calidades de representación.

5.1. Entrada

Los parámetros de entrada, son los siguientes:

- PARÁMETRO OBLIGATORIO
 - **datos** “data.frame” con variables de tipo “factor”
- PARÁMETROS OPCIONALES (ya tienen un valor por defecto)
 - **NF** Se elige un numero de factores a retener (imprimir)[valor por defecto = 2]
 - **Csup** Se suministran las modalidades-columnas suplementarias si las hay, en la forma: $c(i, j, z)$ [valor por defecto = NULL]

¹J. Thioulouse, A.B Dufour, D.Chessel *Analysis of Environmental Data: Exploratory and Euclidean methods in Environmental sciences*,2004

- **ByG** Valor logico T/F. Presenta los indices de Benzecri y Greenacre como elementos de la salida [valor por defecto = FALSE]

5.2. Salida

- Se genera un archivo de salida (“ACM_salida.txt”) que queda en el *Directorio de Trabajo* actual.
- La salida disponible (accesible de la manera habitual *probando[[i]]*)
 1. INERCIA e INERCIA ACUMULADA
 2. Masa
 3. Inercia de las Modalidades
 4. Contribuciones Parciales de las Modalidades
 5. Calidad de Representacion(cos2) de las Modalidades
 6. Coordenadas Modalidades
 7. Coordenadas Individuos
 8. Indice de Benzecri y Greenacre

5.3. Uso

```
## Ejecuta la funcion acm y guarda en probando una lista con resultados
## genera un grafico de las modalidades en el plano principal
> probando <- acm(ejemploACM,NF=6, Csup=c(2:4), ByG=T)
```