



Trabajo final de grado  
GRADO DE  
MATEMÁTICAS  
Facultat de Matemàtiques  
Universitat de Barcelona

---

ANÁLISIS DE CORRESPONDENCIAS SIMPLE.  
Aplicación a un estudio de geografía electoral.

---

Roberto Gregorio Sanz

Director: Josep Vives  
Realizado en: Departament de Probabilitat, Lògica i Estadística

Barcelona, 21 de Junio de 2013

# Abstract

The analysis of correspondence, developed by Benzecri, comes within the frame of descriptive multidimensional statistics. Based on the principles of classic factorial analysis founded by Spearman and Thurstone, it has as its aim the extraction of the main axes from a large group of data which are difficult to perceive immediately.

However, the analysis of correspondence converts to brute data into relative frequencies, that permits of the treatment in general of information of a qualitative nature at the same time quantitative expressed in units of different sizes. In a more precise way, the analysis of correspondence sums up the information contained in a table of contingencies referring to two groups of large sizes and taking into account the probabilistic character of the data to remedy their heterogeneity.

After transforming the elements of the table, the analysis of correspondence gives a perfectly clear vision of two clusters of points when this is projected on to a sub-space of small dimensions (usually two) in such way as to maintain an important part of the information initially constituted. So, it deals with giving a coherent meaning to the axes on which the points have been projected and taking advantage of the certain help in interpretation which the method provides.

*Under the name of mathematical statistics, some authors have built a pompous science, rich in hypothesis that are never satisfied in practice"*

*(Benzecri, 1980)*

# Abstract

El análisis de correspondencias, desarrollado por Benzecri, se encuadra dentro de la estadística descriptiva multidimensional. Inspirado en los principios del análisis factorial clásico, fundado por Spearman y Thurstone, tiene por objeto extraer los principales ejes de un gran conjunto de datos de difícil percepción inmediata.

El análisis de correspondencias convierte los datos brutos en frecuencias relativas, que permite tratar en general informaciones de carácter cualitativo, al mismo tiempo que datos cuantitativos expresados en unidades de medida diferentes. Dicho de otro modo, el análisis de correspondencias resume la información contenida en una tabla de contingencia referida a dos conjuntos de grandes dimensiones, teniendo en cuenta el carácter probabilístico de los datos y así comprobar su heterogeneidad.

Después de transformar los elementos de la tabla, el análisis de correspondencias permite dar una visión fácilmente interpretable de dos nubes de puntos, al proyectar éstos sobre un subespacio de pocas dimensiones (generalmente dos), de forma que se conserve una parte importante de la información inicialmente constituida. Se trata entonces de dar un significado coherente a los ejes sobre los cuales se han proyectado los puntos, apoyándose para ello en ciertas ayudas de interpretación que proporciona la técnica.

*Bajo el nombre de estadística matemática, algunos autores han construido una ciencia pomposa, rica en hipótesis que nunca se satisfacen en la práctica.*

*(Benzecri, 1980)*

# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Motivación . . . . .	1
1.2. Objetivos . . . . .	1
<b>2. Teoría del análisis de correspondencias</b>	<b>2</b>
2.1. Perfiles, masas y centroides . . . . .	3
2.2. Distancia ji-cuadrado e inercia . . . . .	14
2.3. Reducción de la dimensión. . . . .	22
2.4. Descomposición en valores singulares (DVS) . . . . .	25
2.5. Simetría entre el análisis de filas y columnas . . . . .	27
2.6. Representaciones bidimensionales . . . . .	29
2.7. Cálculo de las coordenadas principales y las coordenadas estándares . . . . .	31
2.8. Análisis y diagnóstico de las contribuciones a la inercia . . . . .	33
2.9. Relaciones de transición . . . . .	38
<b>3. Aplicación a un estudio de geografía electoral</b>	<b>41</b>
3.1. Introducción . . . . .	41
3.2. Análisis de los resultados de las elecciones al Parlament de Catalunya 2012	43
3.2.1. Tablas preeliminares . . . . .	43
3.2.2. Análisis estadístico univariante de los resultados . . . . .	49
3.2.3. Análisis estadístico multivariante: análisis de correspondencias de los resultados . . . . .	51
3.3. Conclusiones del estudio de geografía electoral . . . . .	72
<b>A. Implementación del cálculo del AC con R</b>	<b>1</b>
<b>Bibliografía</b>	<b>4</b>

# Capítulo 1

## Introducción

### 1.1. Motivación

La motivación por este Trabajo Final de Grado, vino implicado por mi interés por la Estadística y en concreto por el análisis de datos, dentro del análisis estadístico multivariante.

Leer el libro "La práctica del análisis de correspondencias", de M-J Greenacre (2008), hizo que me decantaría por un tipo de análisis que desconocía y que fue clave para la decisión final de realizar este Trabajo Final de Grado.

### 1.2. Objetivos

El objeto de este trabajo final de grado es presentar los fundamentos teóricos del Análisis de Correspondencias Simple, y aplicarlo al estudio de comportamiento electoral de la comarcas de Cataluña, en las últimas elecciones autonómicas de 2012. El hecho de que haya escogido como aplicación un estudio de geografía electoral, es porque pienso que los resultados electorales constituyen un ámbito adecuado e idóneo para aplicar esta técnica.

## Capítulo 2

# Teoría del análisis de correspondencias

Los orígenes del análisis de correspondencias se remontan a la década de los 60 (J-M Benzécri, 1982). Se desarrolla como técnica estadística aplicada de naturaleza euclidiana (Foucart, 1984 y 1985) a la hora de analizar las tablas de contingencia, cuyo objeto es obtener una representación geométrica de los datos que ayude a visualizar y localizar las relaciones establecidas en grandes matrices de datos. Este nuevo enfoque del análisis de las tablas de contingencias, fuera del marco de la estadística clásica (Batista, 1984), ha derivado posteriormente a formulaciones nuevas. De este modo es plausible utilizarlo en tablas lógicas completas o datos ordinales, y en general a cualquier estructura (sujetos x variables) que contenga valores enteros positivos (Cornejo, 1988; Foucart, 1985) y también a formas multivariantes, de forma tal que es tomado en la actualidad como un caso particular del análisis factorial. Iniciado en los trabajos de principios de siglo de Pearson y Hottelling, y desarrollado sobre todo a partir de la década de los 60 gracias al uso de la informática (Batista y Martínez, 1989). Mientras la escuela anglosajona enfoca las aplicaciones sobre las variables (estructuras) con el desarrollo del análisis factorial, la escuela francesa se desvía hacia aplicaciones centradas en los individuos (análisis de correspondencias).

El nombre de análisis de correspondencias es una traducción del francés "Analyse des correspondances", el cual fue propuesto en los años 60 por el físico-matemático francés Benzécri, con el fin de definir, describir e interpretar el análisis a través de un gráfico geométrico. Esta técnica analiza los datos tal como fue diseñado por algunos precursores de la estadística entre los que destacan Pearson, Guttman, Fisher, los cuales, sin embargo no pudieron llevar a cabo los cálculos por la falta de herramientas que permitiesen cálculos matemáticos tan complejos como los que pueden hoy en día realizar los ordenadores y el software existente en el mercado (R, SPSS, Minitab, Xlstat, etc.).

## 2.1. Perfiles, masas y centroides

### Perfiles.

#### Tabla de contingencia.

Sean  $X$  e  $Y$  dos variables categóricas con  $I$  y  $J$  categorías respectivamente, y que han sido observadas cruzando las  $I$  categorías  $X$  con las  $J$  categorías  $Y$ , con valores  $\{x_1, \dots, x_I\}$  e  $\{y_1, \dots, y_J\}$ , obteniendo  $n = \sum_{ij} n_{ij}$  observaciones, donde  $n_{ij}$  es el número de veces en que aparece la intersección  $X_i \cap Y_j$  dando lugar a la tabla de contingencia  $N$  de  $I \times J$ :

**Tabla 2.1.** - Tabla de contingencia  $N$ .

		Y						Total fila
		1	2	j		J		
X	1	$n_{11}$	$n_{12}$	...	$n_{1j}$	...	$n_{1J}$	$n_{1+}$
	2	$n_{21}$	$n_{22}$	...	$n_{2j}$	...	$n_{2J}$	$n_{2+}$
	:	:	:	...	:	...	:	:
	i	$n_{i1}$	$n_{i2}$	...	$n_{ij}$	...	$n_{iJ}$	$n_{i+}$
	:	:	:	...	:	...	:	:
	I	$n_{I1}$	$n_{I2}$	...	$n_{Ij}$	...	$n_{IJ}$	$n_{I+}$
Total columna		$n_{+1}$	$n_{+2}$	...	$n_{+j}$	...	$n_{+J}$	$n_{++}$

Donde:

$n_{ij}$ , es el número de elementos de la muestra con  $X = x_i$ , y  $Y = y_j$ , situado en la  $i$ -ésima fila y en la  $j$ -ésima columna de la tabla de contingencia,

$n_{i+}$ , es el total de la  $i$ -ésima fila (el subíndice + indica suma de los elementos del correspondient índice):

$$n_{i+} = \sum_{j=1}^J n_{ij},$$

$n_{+j}$ , es el total de la  $j$ -ésima columna:

$$n_{+j} = \sum_{i=1}^I n_{ij},$$

$n_{++}$ , es el total de la tabla  $N$ :

$$n_{++} = \sum_i \sum_j n_{ij}.$$

De lo anterior obtenemos las relaciones:

$$n_{++} = \sum_i \sum_j n_{ij} = \sum_{i=1}^I n_{i+} = \sum_{j=1}^J n_{+j}.$$

**Tabla de perfiles fila.**

Definimos los perfiles fila como las frecuencias relativas de las filas de la tabla de contingencia  $N$ . Estos perfiles fila describen las distribuciones condicionadas de la variable  $Y$  por las distintas modalidades de la variable  $X$ , es decir, las probabilidades condicionadas  $P(y_1/x_i), \dots, P(y_J/x_i)$ .

**Tabla 2.2.** - Tabla de perfiles fila  $A$ .

		Y					Masas de las filas	
		1	2	j	J			
<b>X</b>	<b>1</b>	$\frac{n_{11}}{n_{1+}}$	$\frac{n_{12}}{n_{1+}}$	...	$\frac{n_{1j}}{n_{1+}}$	...	$\frac{n_{1J}}{n_{1+}}$	$\frac{n_{1+}}{n_{++}}$
	<b>2</b>	$\frac{n_{21}}{n_{2+}}$	$\frac{n_{22}}{n_{2+}}$	...	$\frac{n_{2j}}{n_{2+}}$	...	$\frac{n_{2J}}{n_{2+}}$	$\frac{n_{2+}}{n_{++}}$
		:	:				:	
		:	:				:	
	<b>i</b>	$\frac{n_{i1}}{n_{i+}}$	$\frac{n_{i2}}{n_{i+}}$	...	$\frac{n_{ij}}{n_{i+}}$	...	$\frac{n_{iJ}}{n_{i+}}$	$\frac{n_{i+}}{n_{++}}$
	:	:	...		...	:		
	:	:	...		...	:		
	<b>I</b>	$\frac{n_{I1}}{n_{I+}}$	$\frac{n_{I2}}{n_{I+}}$	...	$\frac{n_{Ij}}{n_{I+}}$	...	$\frac{n_{IJ}}{n_{I+}}$	$\frac{n_{I+}}{n_{++}}$
Perfil fila medio		$\frac{n_{+1}}{n_{++}}$	$\frac{n_{+2}}{n_{++}}$	...	$\frac{n_{+j}}{n_{++}}$	...	$\frac{n_{+J}}{n_{++}}$	$n_{++}$

Denotamos  $a_{ij}$  como el  $j$ -ésimo elemento del perfil de la fila  $i$ :

$$a_{ij} = \frac{n_{ij}}{n_{i+}},$$

y el perfil fila  $i$  por  $a_i$ :

$$a_i = (a_{i1}, a_{i2}, a_{i3}, \dots, a_{iJ}).$$

**Tabla de perfiles columna.**

Definimos los perfiles columna como las frecuencias relativas de las columnas de la tabla de contingencia  $N$ . Estos perfiles columna describen las distribuciones condicionadas de la variable  $X$  por las distintas modalidades de la variable  $Y$ , es decir, las probabilidades condicionadas  $P(x_1/y_j), \dots, P(x_I/y_j)$ .

**Tabla 2.3.** - Tabla de perfiles columna  $B$ .

		Y					Perfil columna medio	
		1	2	...	j	...		J
<b>X</b>	<b>1</b>	$\frac{n_{11}}{n_{+1}}$	$\frac{n_{12}}{n_{+2}}$	...	$\frac{n_{1j}}{n_{+j}}$	...	$\frac{n_{1J}}{n_{+J}}$	$\frac{n_{1+}}{n_{++}}$
	<b>2</b>	$\frac{n_{21}}{n_{+1}}$	$\frac{n_{22}}{n_{+2}}$	...	$\frac{n_{2j}}{n_{+j}}$	...	$\frac{n_{2J}}{n_{+J}}$	$\frac{n_{2+}}{n_{++}}$
		⋮	⋮				⋮	
	<b>i</b>	$\frac{n_{i1}}{n_{+1}}$	$\frac{n_{i2}}{n_{+2}}$	...	$\frac{n_{ij}}{n_{+j}}$	...	$\frac{n_{iJ}}{n_{+J}}$	$\frac{n_{i+}}{n_{++}}$
		⋮	⋮	...		...	⋮	
	<b>I</b>	$\frac{n_{I1}}{n_{+1}}$	$\frac{n_{I2}}{n_{+2}}$	...	$\frac{n_{Ij}}{n_{+j}}$	...	$\frac{n_{IJ}}{n_{+J}}$	$\frac{n_{I+}}{n_{++}}$
Masas de las columnas		$\frac{n_{+1}}{n_{++}}$	$\frac{n_{+2}}{n_{++}}$	...	$\frac{n_{+j}}{n_{++}}$	...	$\frac{n_{+J}}{n_{++}}$	$n_{++}$

Denotamos  $b_{ij}$  como el  $i$ -ésimo elemento del perfil de la columna  $j$ :

$$b_{ij} = \frac{n_{ij}}{n_{+j}},$$

y el perfil de la columna  $j$  por  $b_j$ :

$$b_j = (b_{1j}, b_{2j}, b_{3j}, \dots, b_{nI}).$$

### ***Perfil fila medio.***

Definimos el perfil fila medio como las frecuencias marginales de las filas de la tabla de contingencia  $N$ , con relación a sus totales  $n_{++}$  (Veáse última fila de la Tabla 2.2 de la Tabla de perfiles fila  $A$ ).

Denotamos  $c_j$  como el  $j$ -ésimo elemento del perfil fila medio:

$$c_j = \frac{n_{+j}}{n_{++}},$$

y el perfil fila medio por  $c$ :

$$c = (c_1, c_2, c_3, \dots, c_J).$$

### ***Perfil columna medio.***

Definimos el perfil columna medio como las frecuencias marginales de columnas de la tabla de contingencia  $N$ , con relación a sus totales  $n_{++}$ . (Veáse última columna de la Tabla 2.3 de la Tabla de perfiles columna  $B$ ).

Denotamos  $r_i$  como el  $i$ -ésimo elemento del perfil columna medio:

$$r_i = \frac{n_{i+}}{n_{++}},$$

y el perfil fila medio por  $r$ :

$$r = (r_1, r_2, r_3, \dots, r_I).$$

### ***Cociente de contingencia.***

El cociente de contingencia es el tratamiento simétrico de filas y columnas, es decir, comparando los perfiles fila con su perfil fila medio, llegamos al mismo resultado que comparando los perfiles columna con su perfil columna medio.

### ***Tratamiento simétrico de filas y columnas.***

El AC analiza de forma similar a las filas y columnas de una tabla de contingencia, decimos que trata filas y columnas de forma *simétrica*.

*Característica geométrica especial de los perfiles.*

Todos estos perfiles se pueden expresar en porcentajes (%), puesto que, tienen la característica geométrica especial de que la suma de sus elementos es 1.

Para un perfil fila  $i$ ,  $a_i = (a_{i1}, a_{i2}, a_{i3}, \dots, a_{iJ})$ :

$$\sum_{j=1}^J a_{ij} = \sum_{j=1}^J \frac{n_{ij}}{n_{i+}} = 1,$$

y para un perfil columna  $j$ ,  $b_j = (b_{1j}, b_{2j}, b_{3j}, \dots, b_{Ij})$ :

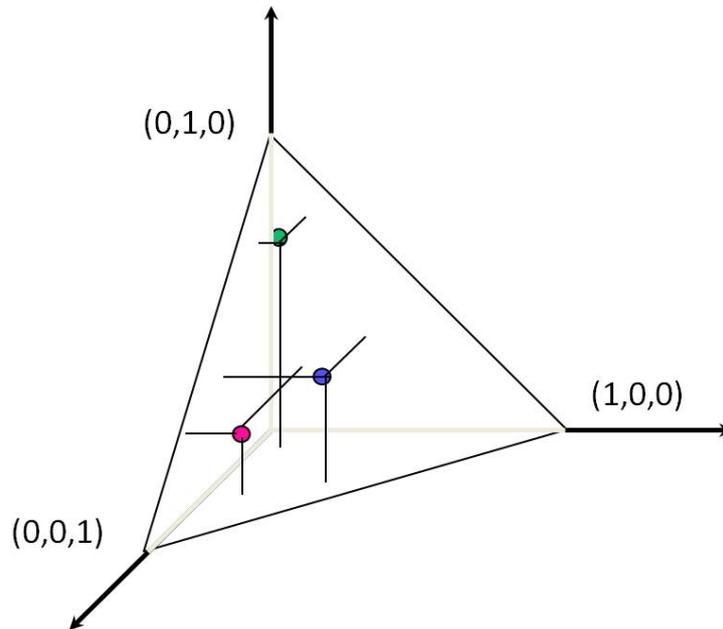
$$\sum_{i=1}^I b_{ij} = \sum_{i=1}^I \frac{n_{ij}}{n_{+j}} = 1.$$

## Geometría de los perfiles..

Representamos los perfiles con  $J$  componentes como puntos en un espacio  $J$ -dimensional. Sin embargo, debido a que la suma de estos  $J$  componentes es 1, los perfiles ocupan en realidad, una región restringida de este espacio de dimensión  $(J - 1)$ , que llamamos *sistema de coordenadas baricéntrico* ó ***símplex***.

Las líneas de unión de todos los pares de los  $J$  puntos unidad de los  $J$  ejes perpendiculares delimitan el *símplex*. A estos puntos unidad les llamamos *vértices* del *símplex* o del espacio de perfiles.

En el caso de tres dimensiones ( $J = 3$ ), a estas representaciones las llamamos *sistemas de coordenadas triangular* (o *ternario*).



Los vértices coinciden con los perfiles extremos, es decir perfiles totalmente concentrados en un solo tipo de categoría. En definitiva vemos que los vértices definen los extremos del espacio de perfiles.

Entonces los  $I$  perfiles fila se pueden representar como  $I$  puntos en el espacio de los perfiles columna, pero es suficiente con una dimensión menos, el  $\text{mím}\{I - 1, J - 1\}$ .

## Masas.

Podemos considerar los valores de los perfiles no sólo como unas coordenadas en un espacio multidimensional, sino también como los pesos asignados a los vértices de un *símplex*. Cada perfil es una media ponderada (centroide o baricentro) de los vértices.

El perfil medio es una media ponderada de los perfiles, a los que hemos asignado pesos de acuerdo con sus frecuencias marginales. Este perfil medio se halla en una posición central entre los perfiles, pero más cerca de los perfiles con mayor frecuencia.

### ***Masas.***

Definimos por *masas* a los pesos asignados a los perfiles. La masa asociada a cada perfil (categoría de la variable fila o de la variable columna) mide la importancia relativa de dicho perfil. La masa que corresponde a cada perfil será su total marginal dividido por el total de la tabla,  $(n_{++})$ .

### ***Masas de los perfiles.***

Por tanto, la masa de cada perfil fila, será el correspondiente elemento del perfil medio de columnas  $(r_i)$ , mientras que la masa de cada perfil columna, será el correspondiente elemento del perfil medio de las filas  $(c_j)$ .

Entonces:

$r_i$ , es la masa de la  $i$ -ésima fila, así  $r_i = \frac{n_{i+}}{n_{++}}$ .  
 $c_j$ , es la masa de la  $j$ -ésima columna, es decir  $c_j = \frac{n_{+j}}{n_{++}}$ .

### ***Relaciones entre perfiles fila y perfiles columna.***

Del  $i$ -ésimo elemento del perfil fila obtenemos:

$$a_{ij} = \frac{n_{ij}}{n_{i+}} \Rightarrow n_{ij} = a_{ij}n_{i+}$$

y del  $j$ -ésimo elemento del perfil columna obtenemos:

$$b_{ij} = \frac{n_{ij}}{n_{+j}} \Rightarrow n_{ij} = b_{ij}n_{+j}$$

entonces, tenemos la relación entre los elementos de los perfiles filas y columnas:

$$b_{ij} = \frac{a_{ij}r_i}{c_j} \qquad a_{ij} = \frac{b_{ij}c_j}{r_i}.$$

Mientras que el perfil que representa a una fila/columna es independiente del total marginal correspondiente, este total lo vamos a utilizar para medir la importancia del perfil en el análisis.

Las filas o columnas con perfiles iguales son *distribucionalmente equivalentes*, las podemos agregar en solo punto.

## *La matriz de correspondencias.*

Sea  $\mathbf{N}$  la tabla de contingencia de  $\mathbf{I} \times \mathbf{J}$  categorías, y denotamos  $n_{++}$  por:

$$n_{++} = \sum_i \sum_j n_{ij} = \mathbf{1}^T \mathbf{N} \mathbf{1},$$

donde utilizamos  $\mathbf{1}$  para simbolizar un vector de unos, el primer  $\mathbf{1}$  es  $\mathbf{I} \times \mathbf{1}$ , y el segundo es  $\mathbf{J} \times \mathbf{1}$  para coincidir con el número de filas y de columnas de  $N$ .

Entonces, la matriz  $P = \{p_{ij}\}$ , definida como:

$$P = \frac{1}{n_{++}} \mathbf{N}$$

es la transformación de la matriz de frecuencias absolutas  $N$ , en la matriz de frecuencias relativas  $P$ , llamada *matriz de correspondencias*, fundamental por su importancia y utilidad para el cálculo del AC.

Donde:

$p_{ij}$ , es  $n_{ij}$  dividido por el total de la tabla:

$$p_{ij} = \frac{n_{ij}}{n_{++}},$$

$p_{+j}$ , es la suma de frecuencias relativas de la  $j$ -ésima columna de  $P$ :

$$p_{+j} = \sum_{i=1}^I p_{ij} = \frac{n_{+j}}{n_{++}},$$

y  $p_{i+}$ , es la suma de frecuencias relativas de la  $i$ -ésima fila de  $P$ :

$$p_{i+} = \sum_{j=1}^J p_{ij} = \frac{n_{i+}}{n_{++}}.$$

Indicando por  $\mathbf{r}$  el vector  $I \times 1$  con los totales marginales de las filas de la matriz de correspondencias  $P$ , y por  $\mathbf{c}$  el vector  $J \times 1$  con los totales marginales de las columnas de la matriz de correspondencias de  $P$ , tenemos que las masas de las filas y columnas de los perfiles correspondientes las podemos representar por:

$$r_i = \sum_{j=1}^J p_{ij} \qquad c_j = \sum_{i=1}^I p_{ij},$$

matricialmente:

$$\mathbf{r} = \mathbf{P} \mathbf{1} \qquad y \qquad \mathbf{c} = \mathbf{P}^T \mathbf{1},$$

de donde:

$$p_{ij} = \frac{n_{ij}}{n_{++}}; \quad a_{ij} r_i = p_{ij}; \quad b_{ij} c_j = p_{ij}.$$

Comprobamos que todas estas definiciones y todos los resultados en términos de estos valores relativos  $P = \{p_{ij}\}$ ,  $r = \{r_i\}$  y  $c = \{c_j\}$ , la suma de sus elementos, en todos los casos suman 1.

Además multiplicando por  $n_{++}$  recuperamos los elementos de la matriz original  $N$ :

$$n_{++}p_{ij} = n_{ij},$$

$$n_{++}r_i = \text{suma de la } i\text{-ésima fila de } N,$$

$$\text{y } n_{++}c_j = \text{suma de la } j\text{-ésima columna de } N.$$

Denotemos las matrices que contienen los valores marginales de filas y columnas, como:

$$D_r = \text{diag}(r) \quad \text{y} \quad D_c = \text{diag}(c).$$

Entonces la matriz de perfiles fila es:

$$A = D_r^{-1}P,$$

y la matriz de perfiles columna es:

$$B = PD_c^{-1}.$$

## Centroide.

Podemos pensar en el *perfil medio de filas* como un punto de la hipernube de puntos de los perfiles fila que ocupa la posición central (**centro de gravedad, centroide o baricentro**) de dicha hipernube, pero que tiende a estar más cerca de aquellos perfiles que tienen masas más grandes.

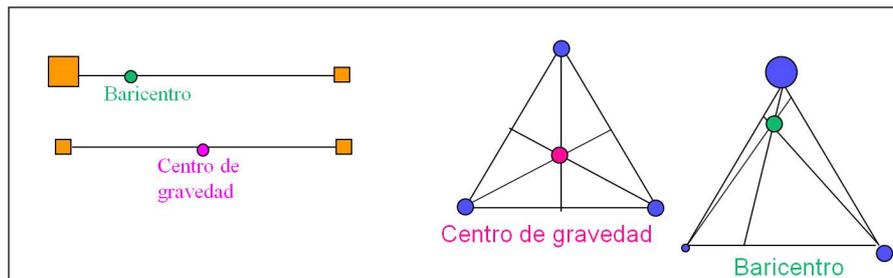
Entonces,  $c_j$  es la media ponderada o centro de gravedad de los perfiles fila, donde el elemento  $j$ -ésimo de cada perfil fila se pondera en relación a la masa del correspondiente perfil fila.

$$\sum_{i=1}^I a_{ij} r_i = \sum_{i=1}^I \frac{n_{ij}}{n_{i+}} \frac{n_{i+}}{n_{++}} = \sum_{i=1}^I \frac{n_{ij}}{n_{++}} = \frac{n_{+j}}{n_{++}} = c_j$$

Teniendo en cuenta el mismo razonamiento para las columnas, el *perfil medio de columnas* podemos verlo como un punto de la hipernube de puntos de los perfiles columna que ocupa la posición central de dicha hipernube, pero que tiende a estar más cerca de aquellos perfiles que tienen masas más pequeñas.

De la misma manera que para las filas,  $r_i$  es la media ponderada o centro de gravedad de los perfiles columna, donde el elemento  $i$ -ésimo de cada perfil columna se pondera en relación a la masa del correspondiente perfil columna.

$$\sum_{j=1}^J b_{ij} c_j = \sum_{j=1}^J \frac{n_{ij}}{n_{+j}} \frac{n_{+j}}{n_{++}} = \sum_{j=1}^J \frac{n_{ij}}{n_{++}} = \frac{n_{i+}}{n_{++}} = r_i$$



## ***Representación del espacio de perfiles.***

### ***Representación en el espacio $R^J$ (espacio de las filas).***

Cada fila de la matriz o tabla de contingencia de los perfiles fila  $A$ , representa la distribución de la variable en columnas condicionada al atributo que representa la fila (Véase Tabla 2.2.).

Entonces tenemos una nube de  $I$  puntos  $[N(I)]$ , que representan a las  $I$  filas y las  $J$  coordenadas del punto fila  $i$ -ésimo, los cuáles serán:  $a_{ij}$ . Esto nos lleva a que el vector que representa cada perfil fila  $i$  será:

$$a_i = (a_{i1}, a_{i2}, a_{i3}, \dots, a_{iJ}) = \left( \frac{p_{i1}}{r_i}, \frac{p_{i2}}{r_i}, \dots, \frac{p_{iJ}}{r_i} \right)$$

Además, cada punto de la nube está dotado de una masa  $r_i$ .

Por tanto el baricentro de la nube será el perfil medio de filas:  $c = (c_1, c_2, c_3, \dots, c_I)$

### ***Representación en el espacio $R^I$ (espacio de las columnas).***

Cada columna de la tabla de contingencia de los perfiles columna  $B$ , representa la distribución de la variable en filas condicionada al atributo que representa la columna (Véase Tabla 2.3.).

Entonces tenemos una nube de  $J$  puntos  $[N(J)]$ , que representan a las  $J$  columnas y las  $I$  coordenadas del punto columna  $i$ -ésimo, los cuáles serán  $b_{ij}$ . Esto nos lleva a que el vector que lo representa será:

$$b_j = (b_{1j}, b_{2j}, b_{3j}, \dots, b_{nI}) = \left( \frac{p_{1j}}{c_j}, \frac{p_{2j}}{c_j}, \dots, \frac{p_{Ij}}{c_j} \right)$$

Además, cada punto de la nube está dotado de una masa  $c_j$ .

Por tanto el baricentro de la nube será el perfil medio de columnas:  $r = (r_1, r_2, r_3, \dots, r_J)$

## 2.2. Distancia ji-cuadrado e inercia

En el AC las distancias las medimos mediante la distancia ji-cuadrado. Esta distancia es la clave de muchas de las propiedades interesantes del AC. Justificaremos la utilización de la distancia ji-cuadrado, técnicamente e intuitivamente.

***La hipótesis de independencia, o de homogeneidad, en tablas de contingencia.***

La existencia o no de algún tipo de relación entre las variables  $X$  e  $Y$  se analiza mediante contrastes de hipótesis sobre la independencia de dichas variables. El test de hipótesis habitualmente utilizado es el de la  $\chi^2$  de Pearson.

Así, dadas dos variables aleatorias  $X$  e  $Y$ , son independientes si

$$P(X = x_i, Y = y_j) = P(X = x_i) * P(Y = y_j) \quad \forall i, j$$

En el caso de una tabla de contingencia, si se aproxima la probabilidad de que sucedan  $x_i$  e  $y_j$  como la frecuencia relativa en un experimento con  $N$  tiradas totales (regla de Laplace), entonces:

$$p_{ij} = \frac{n_{ij}}{n_{++}}$$

$$p_{i+} = \frac{n_{i+}}{n_{++}}$$

$$p_{+j} = \frac{n_{+j}}{n_{++}}$$

Así, si

$$P(X = x_i, Y = y_j) = p_{ij} = p_{i+} \times p_{+j}$$

para todo  $i, j$ , las variables  $X$  e  $Y$  son independientes y la tabla es homogénea.

En dicho test la hipótesis nula es  $H_0$ :  $X$  e  $Y$  son independientes y la alternativa es  $H_1$ :  $X$  e  $Y$  son dependientes.

El test se basa en comparar los perfiles fila y columna con los perfiles marginales correspondientes, teniendo en cuenta que si  $H_0$  es cierta todos los perfiles fila (esp. columna) son iguales entre sí e iguales al perfil marginal de  $X$  (resp. de  $Y$ ).

El estadístico  $\chi^2$  del test viene dado por la expresión:

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - e_{ij})^2}{e_{ij}} = \sum_{i=1}^I \sum_{j=1}^J \frac{n_{i+} \left( \frac{n_{ij}}{n_{i+}} - \frac{n_{+j}}{n_{++}} \right)^2}{\frac{n_{+j}}{n_{++}}} = \sum_{i=1}^I \sum_{j=1}^J \frac{n_{+j} \left( \frac{n_{ij}}{n_{+j}} - \frac{n_{i+}}{n_{++}} \right)^2}{\frac{n_{i+}}{n_{++}}}$$

donde  $e_{ij} = E[n_{ij} | H_0 \text{ cierta}] = n_{++} p_{ij} = n_{++} p_{i+} p_{+j} = \frac{n_{i+} n_{+j}}{n_{++}}$

Intuitivamente, valores pequeños de  $\chi^2$  significan que los valores de  $n_{ij}$  y  $e_{ij}$  son cercanos y, por lo tanto, que  $H_0$  es cierta y, por el contrario, valores grandes de  $\chi^2$  darían evidencia de que  $H_0$  es falsa.

Bajo la hipótesis nula  $\chi^2$  se distribuye, asintóticamente, según una  $\chi^2_{(I-1)(J-1)}$  y el p-valor del test viene dado por:

$$P[\chi^2_{(I-1)(J-1)} \geq \chi^2_{obs}]$$

donde  $\chi^2_{obs}$  es el valor observado en la muestra del estadístico  $\chi^2$ . Para un nivel de significación  $0 < \alpha < 1$  la hipótesis  $H_0$  se rechaza si dicho p-valor es menor o igual que  $\alpha$ .

Si la hipótesis nula se rechaza, las variables  $X$  e  $Y$  son dependientes. En este caso conviene analizar los perfiles condicionales fila y columna así como los residuos del modelo para estudiar qué tipo de dependencia existe entre ellas. Los residuos más utilizados son los llamados residuos estandarizados que vienen dados por la expresión:

$$s_{ij} = \sqrt{r_i} \left( \frac{p_{ij}}{r_i c_j} - 1 \right) \sqrt{c_j}$$

y si toman valores grandes en valor absoluto será debido a que en la celda ( $X = x_i$ ,  $Y = y_j$ ) de la tabla de correspondencias los valores de  $n_{ij}$  y  $e_{ij}$  son muy diferentes y que, por lo tanto, existe un número anormalmente alto (si  $s_{ij} > 0$ ) o bajo (si  $s_{ij} < 0$ ) de casos.

Los residuos se distribuyen asintóticamente como una  $N(0, 1)$ . La hipótesis  $H_0$  y, a un nivel del 95.5 % de confianza, residuos con un valor absoluto mayor que dos se consideran como valores anormalmente altos.

### ***Distancia euclídea o pitagórica.***

La distancia euclídea o pitagórica ordinaria  $d$  entre los perfiles fila  $i$  e  $i'$  será:

$$d^2(a_i, a_{i'}) = \sum_{j=1}^J (a_{ij} - a_{i'j})^2.$$

### ***Distancia ji-cuadrado.***

#### ***Distancia ji-cuadrado entre dos perfiles fila $i$ e $i'$ .***

La expresión que vamos a considerar para calcular la distancia entre dos perfiles fila, será una distancia euclídea ponderada, ponderando cada dimensión del espacio, por el inverso del valor  $c_j$  (es decir, dividimos cada diferencia de cuadrados entre las coordenadas en la dimensión  $j$ -ésima del espacio de las columnas por el correspondiente elemento  $j$ -ésimo del perfil medio de filas).

Esta ponderación está en relación con la frecuencia marginal de la columna (dimensión) correspondiente, a mayor frecuencia marginal, mayor  $c_j$ , y por tanto, al dividir por algo mayor, la diferencia al cuadrado en el numerador, relativa a dicho valor, será menos importante que la misma diferencia con respecto a un  $c_j$  menor).

Esta es la distancia que se conoce con el nombre de distancia ji-cuadrado entre los perfiles fila  $i$  e  $i'$ :

$$d_{\chi^2}^2(a_i, a_{i'}) = \sum_{j=1}^J \frac{(a_{ij} - a_{i'j})^2}{c_j}.$$

#### ***Distancia ji-cuadrado entre dos perfiles columna $j$ e $j'$ .***

Para calcular la distancia ji-cuadrado entre dos perfiles columna, utilizaremos como ponderación de cada dimensión del espacio de las filas el inverso del valor  $r_i$  (es decir, dividimos cada diferencia de cuadrados en la dimensión  $i$ -ésima del espacio de las filas por el correspondiente elemento  $i$ -ésimo del perfil medio de columnas).

Esta ponderación está en relación con la frecuencia marginal de la fila (dimensión) correspondiente, a mayor frecuencia marginal, mayor  $r_i$ , y por tanto, al dividir por algo mayor, la diferencia al cuadrado en el numerador, relativa a dicho valor, será menos importante que la misma diferencia con respecto a un  $r_i$  menor).

Esta es la distancia que se conoce con el nombre de distancia ji-cuadrado entre los perfiles columna  $j$  e  $j'$ :

$$d_{\chi^2}^2(b_j, b_{j'}) = \sum_{i=1}^I \frac{(b_{ij} - b_{ij'})^2}{r_i}.$$

#### ***Una expresión alternativa del estadístico $\chi^2$ , en términos de perfiles y de masas.***

$$\chi^2 = \text{total de la fila} \times \frac{(\text{perfiles observados de la fila} - \text{perfiles esperados de la fila})^2}{\text{perfiles esperados de la fila}}.$$

### Relación de la distancia Ji-Cuadrado con el estadístico $\chi^2$ .

Nos interesa la habilidad del estadístico Ji-cuadrado para medir la heterogeneidad entre los perfiles.

$$\begin{aligned}
 \chi^2 &= \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - e_{ij})^2}{e_{ij}} = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - \frac{n_{i+}n_{+j}}{n_{++}})^2}{\frac{n_{i+}n_{+j}}{n_{++}}} = \sum_{i=1}^I \sum_{j=1}^J \frac{\frac{1}{n_{i+}^2} (n_{ij} - \frac{n_{i+}n_{+j}}{n_{++}})^2}{\frac{1}{n_{i+}^2} \frac{n_{i+}n_{+j}}{n_{++}}} = \\
 &= \sum_{i=1}^I \sum_{j=1}^J n_{i+} \frac{(\frac{n_{ij}}{n_{i+}} - \frac{n_{i+}n_{+j}}{n_{i+}n_{++}})^2}{\frac{n_{i+}n_{+j}}{n_{i+}n_{++}}} = \sum_{i=1}^I \sum_{j=1}^J n_{i+} \frac{(a_{ij} - c_j)^2}{c_j} = \sum_{i=1}^I n_{i+} \sum_{j=1}^J \frac{(a_{ij} - c_j)^2}{c_j} = \\
 &= \sum_{i=1}^I n_{i+} d_{\chi^2}^2(a_i, c) = \sum_{i=1}^I n_{i+} \frac{n_{++}}{n_{++}} d_{\chi^2}^2(a_i, c) = n_{++} \sum_{i=1}^I r_i d_{\chi^2}^2(a_i, c).
 \end{aligned}$$

En definitiva:

$$\boxed{\chi^2 = n \sum_{i=1}^I r_i d_{\chi^2}^2(a_i, c)}$$

Si rechazamos la  $H_o$ , significa que el valor del estadístico  $\chi^2$  es de magnitud considerable, pero esto puede deberse o bien a que todos los sumandos son grandes, o sólo algunos de ellos.

Un valor pequeño de este estadístico nos llevará a manifestar que no hay razones para pensar en una heterogeneidad entre los perfiles en relación a la otra variable (no tenemos razones para rechazar la  $H_o$  de homogeneidad), lo que supone que sus perfiles son similares, y por tanto, también similares a su perfil medio  $c$ .

Si los perfiles son similares, la distancia ji-cuadrado entre cada uno de ellos y su centroide será pequeña, lo que también es congruente con la hipótesis de homogeneidad.

### Inercia.

El estadístico  $\chi^2$  nos mide también cómo de lejos se encuentran los perfiles de su perfil medio.

$$\chi^2 = n \sum_{i=1}^I r_i d_{\chi^2}^2(a_i, c)$$

Definimos **la inercia (total)** como el estadístico  $\chi^2$  dividido por el tamaño de la muestra.

$$Inercia = \frac{\chi^2}{n}$$

Entonces, si la inercia es baja, habrá poca asociación entre la variable fila y la variable columna, y si la inercia es alta habrá una alta asociación entre ellas.

### ***Inercia de la nube de puntos.***

Con la definición anterior podemos decir que la inercia, es la media ponderada de los cuadrados de las distancias ji-cuadrado entre los perfiles y su perfil medio (tanto para filas como para columnas), ponderando dichas distancias al cuadrado por la masa de cada perfil.

Denotamos:

- La distancia ji-cuadrado entre el  $i$ -ésimo y el  $i'$ -ésimo perfil fila, por  $\|a_i - a_{i'}\|_c$ .
- La distancia ji-cuadrado entre el  $j$ -ésimo y el  $j'$ -ésimo perfil columna, por  $\|b_j - b_{j'}\|_r$ .
- La distancia ji-cuadrado entre el  $i$ -ésimo perfil fila  $a_i$  y el perfil fila medio  $c$  (el vector de las masas de las columnas), por  $\|a_i - c\|_c$ .
- La distancia ji-cuadrado entre el  $j$ -ésimo perfil columna  $b_j$  y el perfil columna medio  $r$  (el vector de las masas de las filas), por  $\|b_j - r\|_r$ .
- A la inercia, también se le conoce con el nombre de ***coeficiente medio cuadrático de contingencia***, dado que es una medida de la varianza total de la tabla. A su raíz cuadrada le denominamos  $\phi$ , por tanto expresamos la inercia como  $\phi^2$ .

Con la notación anterior, la fórmula de la inercia total, por filas, queda:

$$\phi^2 = \frac{\chi^2}{n} = \sum_{i=1}^I r_i d_{\chi^2}^2(a_i, c) = \sum_i r_i \|a_i - c\|_c^2 = \sum_i r_i \sum_j (p_{ij}/r_i - c_j)^2/c_j,$$

y por columnas:

$$\phi^2 = \frac{\chi^2}{n} = \sum_{j=1}^J c_j d_{\chi^2}^2(b_j, r) = \sum_j c_j \|b_j - r\|_r^2 = \sum_j c_j \sum_i (p_{ij}/c_j - r_i)^2/r_i.$$

### *Interpretación geométrica de la inercia.*

Como ya hemos visto anteriormente, la suma de las masas de los perfiles es 1, podemos decir que la inercia es la media ponderada de los cuadrados de las distancias ji-cuadrado entre los perfiles fila y su perfil media.

$$Inercia = \sum_i (i\text{-ésima masa}) \times (\text{distancia } \chi^2 \text{ de } i\text{-ésimo perfil media})^2.$$

La inercia será alta cuando los perfiles fila presenten grandes desviaciones en relación con su media (Gráfico B), y será baja cuando éstos se hallen cerca de la media (Gráfico A). Podemos pensar en la inercia como asociación, o correlación entre las filas y las columnas de una matriz, así cuando la inercia es baja, los perfiles fila presentan poca variación y se hallan cerca de su perfil medio, decimos que existe poca asociación, o correlación, entre las filas y las columnas, y cuánto mayor sea la inercia, más cerca se hallarán los perfiles fila de los vértices columna, es decir mayor será la asociación entre las filas y las columnas.

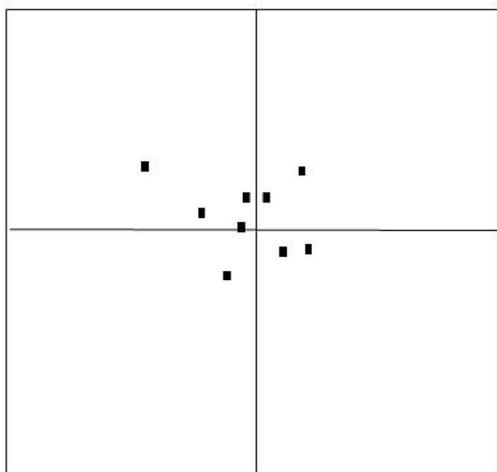


Gráfico A: Poca inercia

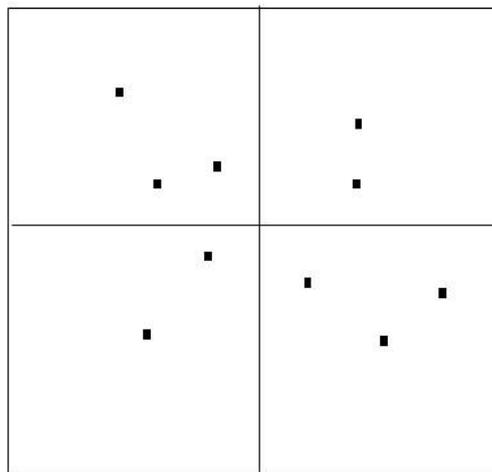


Gráfico B: Mucha inercia

### *Las inercias máxima y mínima.*

Si todos los perfiles fueran idénticos, y por tanto todos se hallaran en el mismo punto (media) de la nube de puntos, todas las distancias ji-cuadrado seran cero, y por lo tanto también lo será la inercia total, con lo cuál la inercia mínima es 0.

Por otro lado, se llegará a la inercia máxima cuando todos los perfiles estén exactamente en los vértices del espacio de perfiles. En tal caso, la inercia máxima sería igual a la dimensión del espacio,  $\min\{I-1, J-1\}$ .

### *La inercia de las filas es igual a la inercia de las columnas.*

La inercia total será la misma si lo calculamos a partir de los perfiles columna. Es decir, la inercia total de la tabla será igual a la media ponderada de los cuadrados de la distancia  $\chi^2$  entre los perfiles columna y su perfil media, ponderadas ahora con las masas de las columnas (Vease apartado *inercia de la nube de puntos*).

### ***El principio de equivalencia distribucional.***

Las distancias ji-cuadrado cumplen el *principio de equivalencia distribucional*, que garantiza la estabilidad de las distancias entre filas, cuando dividimos las columnas en componentes similares, o cuando unimos columnas similares.

Si dos filas de la tabla de contingencia tienen la misma estructura relativa (es decir, el mismo perfil) y las unimos en una nueva fila única, las distancias entre las columnas permanecen invariables. (Lo mismo respecto a las las columnas).

Garantiza que no perdemos nada al agregar categorías homogéneas, ni podemos ganar nada por desagregar una categoría en categorías homogéneas.

### ***Distancia euclídea ponderada.***

Llamamos distancia euclídea ponderada a:

$$d_{E,p}^2(x_j, y_j) = \sqrt{\sum_{j=1}^p w_j (x_j - y_j)^2}$$

dónde  $w_j$  son los valores positivos de los pesos y  $x_j$ , con  $j = 1, \dots, p$  e  $y_j$  con  $j = 1, \dots, p$  son dos puntos en un espacio  $p$ -dimensional.

### ***Métrica de las filas.***

Calcular la distancia ji-cuadrado entre dos perfiles columna, es un ejemplo de distancia euclídea ponderada, es decir, es equivalente a calcular la distancia euclídea ordinaria entre esos dos perfiles, pero dividiendo cada una de las componentes entre la raíz cuadrada del correspondiente elemento del perfil medio de columnas (raíz cuadrada de  $r_i$ ). Equivalentemente para la distancia ji-cuadrado entre dos perfiles fila.

$$d_{\chi^2}^2(b_j, b_{j'}) = \sum_{i=1}^I \frac{(b_{ij} - b_{ij'})^2}{r_i} = \sum_{i=1}^I \left( \frac{b_{ij}}{\sqrt{r_i}} - \frac{b_{ij'}}{\sqrt{r_i}} \right)^2 = d_E^2(\hat{b}_j, \hat{b}_{j'})$$

dónde:  $\hat{b}_j = \frac{b_{ij}}{\sqrt{r_i}}$ .

### ***Interpretación geométrica de la inercia y del estadístico $\chi^2$ .***

En el espacio de perfiles transformado por las distancias ji-cuadrado (que vimos en el apartado anterior), podemos trazar líneas de unión entre los perfiles y su media para mostrar, así, las distancias ji-cuadrado entre los perfiles y su media. Por la fórmula de la inercia, sabemos que la suma de las distancias de las filas a su media, ponderadas con sus respectivas masas, es igual a la inercia total de la tabla. Si en vez de ponderar con las masas, ponderamos con las frecuencias totales de las filas (la frecuencia total de una fila es  $n$  veces la masa de la fila, siendo  $n$  la suma total de la tabla), entonces la suma ponderada de los cuadrados de estas distancias es igual al estadístico  $\chi^2$ . Obtenemos resultados equivalentes con los perfiles columna y el perfil columna medio. Por tanto, *geométricamente*, podemos interpretar la inercia y el estadístico  $\chi^2$  como medidas de dispersión de los perfiles (de las filas o de las columnas) con relación a su media.

### ***Justificación teórica de la distancia ji-cuadrado.***

La distancia ji-cuadrado:

- 1.- Trabaja con perfiles y tiene en cuenta la ponderación de las dimensiones del espacio.
- 2.- Se relaciona con el estadístico ji-cuadrado de Pearson para contrastar la independencia en una tabla de contingencia.
- 3.- Verifica el *principio de equivalencia distribucional*.
- 4.- Las distancias ji-cuadrado hacen que las contribuciones de las categorías sean más parecidas.

Esto se basa en la constatación de que existen diferencias importantes en las varianzas de los valores de las frecuencias de las distintas categorías. Esta observación ilustra una regla general sobre los datos de frecuencias: los conjuntos de datos de frecuencias pequeñas presentan menor dispersión que los conjuntos de frecuencias grandes.

Este ejercicio ilustra el fenómeno general de que una categoría  $A$ , la categoría con más frecuencias más pequeñas, contribuye menos a la distancia euclídea, que por ejemplo otra  $B$ . Sin embargo, la contribución de  $A$  a la distancia ji-cuadrado se ve incrementada gracias a la división por las frecuencias medias.

- 5.- Hace que el análisis de filas y columnas sean simétricos.

### ***Implicaciones de trabajar con perfiles.***

1.- Todas las filas no tienen la misma importancia, ya que el perfil de cada fila está referido a un total marginal diferente. De forma que al representar los perfiles deberemos dar más importancia a aquéllos que tengan mayor masa ( $r_i$ ). El mismo análisis, pero para las columnas.

2.- La distancia euclídea ordinaria entre dos puntos perfil fila no es una buena medida de su proximidad y deberemos modificarla: deberemos utilizar la distancia ji-cuadrado, ponderando cada dimensión del espacio de las columnas. Razonamiento análogo para las columnas.

Queremos proyectar estos puntos en un espacio de dimensión menor de manera que las filas que tengan la misma estructura estén próximas y las que tengan una estructura muy diferente, alejadas.

## 2.3. Reducción de la dimensión.

La mayoría de tablas de contingencia tienen más de tres filas y columnas, y por lo tanto se sitúan en un espacio de representación de mayor dimensión. Dado que no podemos ni observar, ni imaginar fácilmente puntos en un espacio de más dimensiones, es necesario reducir la dimensión de los puntos.

La reducción de la dimensión es un aspecto analítico fundamental del AC, por lo que llevarlo a cabo implica una cierta pérdida de información. Debemos restringir en lo posible esta pérdida y así conservar la máxima información.

### *Identificación de subespacios de baja dimensionalidad.*

El AC nos permite la visualización que nos proporciona más agudeza en el análisis de datos. Pero si la dimensión es mayor que tres, la visualización directa es imposible. Por tanto, es interesante poder visualizar los perfiles, aunque fuera de forma aproximada en un espacio de pocas dimensiones. Así podríamos visualizar los perfiles de forma aproximada en un subespacio de una, dos o tres dimensiones.

Esta es precisamente la esencia del AC: la identificación de subespacios de pocas dimensiones que contengan los perfiles, aunque sea de forma aproximada. Dicho de otro modo, indentifica dimensiones para las cuáles existe muy poca dispersión de los perfiles, y elimina las direcciones de dispersión que aportan poca información.

### *Proyección de perfiles en subespacios.*

Podemos imaginar los perfiles formando una nube de puntos en el espacio de perfiles. Si identificamos la recta, plano o ternario, más próxima a la nube de puntos, podemos proyectar los puntos perpendicularmente sobre la recta, plano o ternario, sacarla del espacio multidimensional y representar las proyecciones en este espacio unidimensional, bidimensional o ternario. En general, los perfiles están situados en espacios de altas dimensiones de modo que no se pueden observar directamente. Se pueden determinar subespacios de dimensión menor al número mínimo entre filas y columnas menos uno, donde se puede aproximar la posición original de los perfiles.

### *Determinación de la calidad de la representación.*

Como las proyecciones de los perfiles en subespacios de pocas dimensiones no son sus verdaderas posiciones, deberíamos conocer cuál la magnitud de la discrepancia entre las posiciones exactas y las aproximadas. Para eso utilizaremos la inercia total de los perfiles como una medida de variabilidad total; es decir, como una medida de la dispersión geométrica de los puntos en sus verdaderas posiciones.

Expresaremos tanto la calidad de la representación, como su contrapartida, la pérdida de calidad, o error de representación, como porcentajes de la inercia total; y por tanto, su

suma debe ser 100 %.

Entonces cuánto menor sea la pérdida de inercia, mayor será la calidad, y cuánto mayor su pérdida, menor la calidad.

Por último, medimos la calidad de representación en subespacios de dimensión menor, en porcentajes de inercia con respecto a la total.

### ***Una aproximación de la distancia entre los perfiles.***

En el espacio multidimensional de las columnas podemos representar los perfiles fila como puntos, y la distancia entre dichos perfiles es la distancia ji-cuadrado, una distancia abstracta que no podemos ver.

Las distancias entre las proyecciones de los perfiles fila son aproximaciones de las verdaderas distancias ji-cuadrado en el espacio real completo. Queremos encontrar una representación de esos perfiles como puntos en un subespacio de baja dimensión donde la distancia euclídea que los separe coincida tanto como sea posible con la distancia ji-cuadrado original.

Hemos visto cómo si transformamos las coordenadas de un punto fila, dividiendo cada una de ellas por la raíz cuadrada del peso de la dimensión correspondiente, la distancia euclídea entre los puntos, utilizando las coordenadas transformadas, coincide con la distancia ji-cuadrado original. Por tanto, los puntos perfil fila con coordenadas transformadas vienen separados por la distancia euclídea ordinaria.

### ***Representación de las proyecciones de los vértices.***

Como ya vimos anteriormente, los vértices representan el perfil fictio concentrado en una categoría. Estos vértices los podemos proyectar en el espacio de perfiles sobre la dimensión que representemos gráficamente. También se pueden proyectar perfiles fila y perfiles columna de modo equivalente en el subespacio extraído.

### ***Criterio de proximidad en el AC.***

Tenemos un espacio con  $I$  perfiles y que  $S$  es un candidato a subespacio de pocas dimensiones en el espacio original. Denotamos como  $d_{\chi^2}(a_i, S)$  la distancia ji-cuadrado entre el  $i$ -ésimo perfil de masa  $r_i$  y  $S$ . Calculamos la proximidad de este perfil al subespacio  $S$  como  $r_i[d_{\chi^2}(a_i, S)]^2$ , es decir, el cuadrado de la distancia ponderada con la masa.

Calculamos la proximidad de todos los perfiles a  $S$ , como la suma de estos valores:

$$\text{Proximidad a } S = \sum_i r_i [d_{\chi^2}(a_i, S)]^2.$$

El objetivo del AC es identificar el subespacio  $S$  que minimize el criterio anterior.

El cálculo matemático de los subespacios se basa en minimizar las sumas de las distancias entre los perfiles y el subespacio, ponderadas por las masas de los puntos. Es decir, se calcula por *el método de los mínimos cuadrados ponderados*.

### Reducción de la dimensión.

Para reducir la dimensión, buscamos las coordenadas en el plano (subespacio de dimensión  $K$ , en general) de forma que se minimicen las distancias al cuadrado entre cada punto y su proyección:  $d_i^2 = \hat{d}_i^2 + e_i^2$

Utilizando el criterio de ajuste de *suma mínimo-cuadrática*, es decir, minimizar  $e_i^2$  es lo mismo que maximizar  $\hat{d}_i^2$ . Entonces:

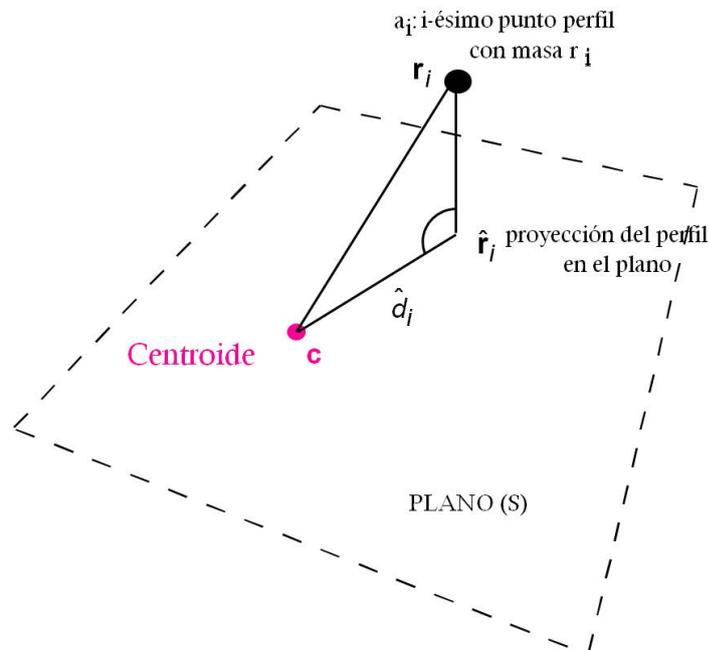
$$\sum_{I=1}^I r_i d_i^2 = \sum_{I=1}^I r_i \hat{d}_i^2 + \sum_{I=1}^I r_i e_i^2$$

donde:

$\sum_{I=1}^I r_i d_i^2$  es la Inercia Total,

$\sum_{I=1}^I r_i \hat{d}_i^2$  es la Inercia Explicada,

y  $\sum_{I=1}^I r_i e_i^2$  es la Inercia Residual.



La inercia residual es una medida de cuánta inercia ha sido perdida a la hora de reducir los perfiles a un formato bidimensional (o de menor dimensión). El procedimiento busca el plano para el cual la pérdida es mínima.

El método más interesante para calcular la minimización anterior, es la *descomposición en valores singulares (DVS)*.

## 2.4. Descomposición en valores singulares (DVS)

El análisis de correspondencias se basa en resultados directos de la teoría de matrices. Utiliza especialmente la descomposición de una matriz en valores singulares (DVS), el fundamento de muchos métodos multivariantes.

**Teorema.** Sea  $A$  una matriz de orden  $m \times n$ . Entonces  $A$  se puede factorizar como  $A = U\Sigma V^T$ , donde  $U$  es una matriz unitaria (ortogonal) de orden  $m$ ,  $V$  es una matriz unitaria (ortogonal) de orden  $n$  y  $\Sigma$  es una matriz de orden  $m \times n$  de la forma  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_p)$ ,  $p = \min(m, n)$ , y  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$ . Los valores  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p$  se denominan valores singulares de la matriz  $A$ .

*Demostración:*

Consideremos la matriz  $A^T A$ , de orden  $n$ . Es hermitiana (simétrica) y sus valores propios son reales mayores o iguales que cero. Los ordenamos en forma decreciente, y supongamos que  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$  son los positivos ( $r = \text{rango}(A)$ ). Sea  $\{v_1, \dots, v_n\}$  una base ortonormal de valores propios, y llamemos:

$$\sigma_i = \sqrt{\lambda_i}, i = 1, \dots, p = \min\{m, n\}, \mathbf{U}_i = \frac{1}{\sigma_i} A \mathbf{v}_i \in K^m, i = 1, \dots, r.$$

Observemos que los vectores  $u_i$  están bien definidos, pues  $r \geq p$ . Tenemos que:

$$\mathbf{U}_i^T \bullet \mathbf{U}_i = \frac{1}{\lambda_i} \mathbf{v}_i^T A^T A \mathbf{v}_i = 1, i = 1, \dots, r.$$

y si  $i \neq j$  entonces:

$$\mathbf{U}_i \bullet \mathbf{U}_j = \frac{1}{\sigma_i \sigma_j} \mathbf{v}_j^T A^T A \mathbf{v}_i = \frac{\lambda_i}{\sigma_i \sigma_j} v_i * v_j = 0$$

Por tanto, los vectores  $u_1, \dots, u_r$  son unitarios y ortogonales entre sí. Completamos, mediante Gram-Schmidt, a una base ortonormal del espacio  $u_1, \dots, u_r, u_{r+1}, \dots, u_m$ . Sean:

$$U = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m), V = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$$

Entonces las entradas de  $U^T A V$  son los números  $\mathbf{u}_j^T A \mathbf{v}_i$ . Si  $i > r$ , entonces son cero, porque  $A \mathbf{v}_i = 0$  en estos casos. Si  $i \leq r$ , entonces  $\mathbf{u}_j^T A \mathbf{v}_i = \sigma_i \mathbf{u}_j^T \mathbf{v}_i$ , que vale  $\sigma_i$  cuando  $i = j$  y cero en los restantes casos. Por tanto,  $U^T A V$  es una matriz de orden  $m \times n$  con los valores  $\sigma_i$  en la diagonal principal y cero en el resto, esto es, la matriz  $\Sigma$ . Entonces  $A = U\Sigma V^T$ .

**Proposición.** Los valores singulares no nulos de  $A$  son las raíces cuadradas de los valores propios de  $A^T A$  o de  $AA^T$  (ambas matrices tienen los mismos valores propios).

*Demostración:*

De

$$A^T A = (V \Sigma^T U^T)(U \Sigma V^T) = V(\Sigma^T \Sigma)V^T$$

tenemos que  $A^T A$  y  $\Sigma^T \Sigma$  son semejantes, por lo que tienen los mismos valores propios. Los de  $\Sigma^T \Sigma$ , que es una matriz diagonal, son  $\sigma_1^2, \dots, \sigma_p^2$  y  $n - p$  valores propios nulos adicionales si  $n > p$ .

**Teorema de Eckart-Young.** Sea  $A = U \Sigma V^T$  descomposición en valores singulares de  $A$ . Para cada  $0 \leq k \leq r = \text{rango}(A)$  definimos:

$$A_k = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

Si  $k = p = \min\{m, n\}$  definimos  $\sigma_{k+1} = 0$ . Entonces:

$$\min_{\text{rango}(B) \leq k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1}.$$

*Demostración:*

Tenemos que  $A - A_k = \sum_{i=k+1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ , por lo que  $A - A_k = U_1 \Sigma_1 V_1^T$  donde  $U_1, V_1$  son unitarias y  $\Sigma_1 = \text{diag}(\sigma_{k+1}, \dots, \sigma_r, 0, \dots, 0)$ . Entonces  $\|A - A_k\|_2 = \sigma_{k+1}$ . Supongamos que  $B$  es una matriz con  $\text{rango}(B) \leq k$  y  $\|A - B\|_2 < \|A - A_k\|_2 = \sigma_{k+1}$ . Entonces  $W = \ker(B)$  es un subespacio de dimensión, al menos,  $n - k$ . Si  $w \in W$ , entonces:

$$\|Aw\|_2 = \|(A - B)w\|_2 \leq \|A - B\|_2 \|w\|_2 < \sigma_{k+1} \|w\|_2$$

Sea  $L = \langle v_1, \dots, v_{k+1} \rangle$ . Es una variedad vectorial de dimensión  $k + 1$ , y  $Av_i = \sigma_i u_i$ . Si  $w \in L$ , con  $w = \sum_{j=1}^{k+1} \alpha_j v_j$ , entonces:

$$\|Aw\|_2 = \left\| \sum_{j=1}^{k+1} \alpha_j \sigma_j v_j \right\|_2 = \left( \sum_{j=1}^{k+1} \alpha_j \sigma_j \right)^{\frac{1}{2}} \geq \sigma_{k+1} \|w\|_2$$

Como la suma de las dimensiones de  $W$  y  $L$  es mayor que  $n$ , tiene que haber un vector común, lo que es contradictorio.

El Teorema de Eckart-Young viene a decir que si aproximamos  $A$  por las primeras  $k$  componentes de la SVD, perdemos una aportación del orden del valor singular  $\sigma_{k+1}$ .

## 2.5. Simetría entre el análisis de filas y columnas

En el análisis de filas de una tabla, tomamos las columnas como puntos de referencia (vértices de las filas) para visualizar e interpretar los datos, y en el análisis de columnas tomamos las filas (vértices de las columnas) como puntos de referencia. Veremos que los valores de los perfiles tienen una interpretación simétrica.

### *Interpretación simétrica de los valores de los perfiles.*

1. *Cociente de contingencia idéntico.*

Si comparamos los perfiles fila con su perfil medio y comparamos los perfiles columna con su perfil medio, obtenemos el mismo resultado. Es decir, el cociente entre un elemento del perfil fila y su masa es el mismo que entre un elemento del perfil columna y su respectiva masa.

2. *La misma inercia total.*

Si hacemos el cálculo por filas:

$$\phi^2 = \frac{\chi^2}{n} = \sum_i r_i \sum_j (p_{ij}/r_i - c_j)^2 / c_j$$

o por columnas:

$$\phi^2 = \frac{\chi^2}{n} = \sum_j c_j \sum_i (p_{ij}/c_j - r_i)^2 / r_i$$

llegamos al mismo resultado.

3. *Igual dimensión.*

Aunque los perfiles fila y los perfiles columnas se encuentran en espacios distintos, la dimensión de la nube de puntos de los perfiles es la misma. Esta es *la primera coincidencia geométrica* entre el análisis de los perfiles fila y el de los perfiles columna.

4. *La misma aproximación para reducir la dimensión.*

*La segunda coincidencia geométrica* entre el análisis de los perfiles fila y el de los perfiles columna es, la calidad y el porcentaje de representación son los mismos.

5. *Los mismos valores de coordenadas, pero redimensionados.*

*La tercera coincidencia geométrica* entre el análisis de los perfiles fila y el de los perfiles columna, es decir en el análisis de columnas, es que las posiciones de los vértices fila, son una expansión o contracción de las posiciones de los perfiles fila del análisis de filas.

### ***Ejes e inercias principales.***

Un *eje principal*, es la dispersión de puntos de un espacio multidimensional que optimiza la inercia o, de forma equivalente, el eje que mejor se ajusta a los puntos en el sentido de distancias mínimo-cuadráticas ponderadas y una *inercia principal* es la inercia explicada por ese eje, también llamado *valor propio*.

*El primer eje principal* de perfiles es la recta, o dimensión, que mejor se ajusta y *la primera inercia principal* es la inercia explicada por esa dimensión.

### ***El factor de escala.***

El factor de escala entre las coordenadas de un perfil y las coordenadas de un vértice vienen dadas por la raíz cuadrada de la inercia, es decir:

$$\text{Coordenada del perfil} = \text{coordenadas del vértice} \times \sqrt{\text{inercia principal}}.$$

Con este factor escala, lo que pone de manifiesto es que «las posiciones de los vértices de un análisis, son las posiciones de los perfiles del otro análisis».

Por otro lado, el factor escala nos indica la dispersión de los perfiles *internos* en comparación con los *vértices* exteriores.

Como las inercias principales (o su raíz cuadrada), miden la dispersión para los ejes principales (de forma individual), y tenemos que la inercia principal  $\leq 1$ , entonces los perfiles deben hallarse en el interior de sus correspondientes vértices. A mayor inercia principal, mayor dispersión de los perfiles en el interior de sus correspondientes vértices.

### ***Coordenadas principales y coordenadas estándares.***

Denominamos *coordenadas principales* a las posiciones de las coordenadas de los perfiles en un eje principal, y *coordenadas estándares* a las posiciones de coordenadas de los vértices en un eje principal.

## 2.6. Representaciones bidimensionales

En el AC proyectamos una nube de perfiles sobre un espacio reducido que mejor se ajusta. Sin embargo en la práctica la mayor parte de representaciones, no son en un solo eje principal, la recta que mejor se ajusta. Sino en representaciones bidimensionales. Normalmente, representaremos el primer eje principal horizontalmente (eje  $x$ ) y el segundo eje principal verticalmente (eje  $y$ ).

### *Anidado de los ejes principales.*

La representación bidimensional también contiene la mejor representación unidimensional. Si proyectamos los puntos de la representación bidimensional verticalmente sobre el eje horizontal, esta representación unidimensional será al misma, que si de entrada, hubiesemos realizado la representación unidimensional. Por eso decimos que los ejes están *anidados*. Dicho de otro modo, la representación óptima de una determinada dimensionalidad contiene todas las representaciones óptimas de menor dimensionalidad. Es decir, el eje principal de una representación unidimensional es idéntico al primer eje principal de una solución bidimensional, y así sucesivamente.

### *Mapa asimétrico del AC.*

El mapa asimétrico, o con escalas asimétricas es una representación conjunta de perfiles y vértices, en el que representamos las filas en coordenadas principales, y las columnas en coordenadas estándares o viceversa (Véase el gráfico 3.4. con las filas en coordenadas principales y las columnas en coordenadas estándares, pág. 64, y el gráfico 3.5. con las filas en coordenadas estándares y la columnas en coordenada principales, pág. 65).

En este caso los factores de contracción (es decir, las raíces de las inercias principales), de los dos ejes no son los mismos.

En un mapa asimétrico, en el que representamos las filas en coordenadas principales, las distancias entre las filas, son aproximadamente distancias ji-cuadrado. Las distancias entre las filas y un vértice columna son, en general, inversamente proporcionales a los valores de los elementos del perfil de esa columna. Aparte de estos factores de escala, las posiciones relativas de los perfiles y los vértices son las mismas.

Los mapas asimétricos van bien cuando la inercia es alta, pero resultan problemáticos cuando la inercia total es pequeña, porque las coordenadas principales se hallan demasiado cerca del origen, lo que nos complica el etiquetado.

### *Mapas en filas principales y mapas en columnas principales.*

Llamamos mapa en *filas principales*, a un mapa asimétrico que expresa las filas en coordenadas principales, y mapa en *columnas principales* cuando expresamos las columnas en coordenadas principales.

### ***Mapa simétrico del AC.***

En los mapas simétricos solapamos en la misma representación, los perfiles fila y los perfiles columna, aunque, en un sentido más estricto, las configuraciones de filas y columnas salen de espacios distintos. Es decir, en los mapas simétricos representamos tanto las filas como las columnas en coordenadas principales.

En los mapas simétricos, las distancias entre las filas y las distancias entre las columnas son aproximadamente distancias ji-cuadrado de sus respectivos perfiles. De todas maneras, en un mapa simétrico no existe una interpretación específica para las distancias entre las filas y las columnas.

Una regla de oro para la interpretación de esta tipología de mapas es que podemos interpretar la distancias entre puntos siempre que se hallen en el mismo espacio, como, por ejemplo, los perfiles fila y los vértices columna en el espacio de perfiles fila.

Justificamos este hecho por la estrecha relación existente entre el análisis de filas y de columnas, es decir, por la existencia de un solo factor de escala entre las filas y las columnas. A la hora de interpretar un mapa simétrico, tenemos que tener claro que no es más que un *solapamiento de dos mapas distintos* (Véase el gráfico 3.1. con las coordenadas principales de las filas, pág. 61, el gráfico 3.2. con las coordenadas principales de las columnas, pág. 62, y el gráfico 3.3. con las filas y columnas en coordenadas principales, pág. 63).

### ***La interpretación dimensional de los mapas.***

Tenemos que tener en cuenta que la interpretación *dimensional* de los mapas es siempre igual, tanto si recurrimos a mapas simétricos como a mapas asimétricos. Es decir, tenemos que interpretar los ejes uno a uno. Basamos la interpretación en asignar nombres descriptivos a los ejes principales a partir de las posiciones relativas de los puntos de uno de los dos conjuntos de coordenadas. A continuación, interpretamos las posiciones relativas del otro conjunto de coordenadas con relación a las dimensiones que previamente hemos asignado nombres descriptivos.

### ***La importancia de la razón de escalas del mapa.***

Es importante que las utilidades de representación gráfica mantengan la razón de escalas de los mapas, es decir el cociente entre una unidad de longitud horizontal y una unidad en el eje vertical debería ser igual a 1. Una unidad en el eje horizontal debe aproximarse tanto como sea posible a una unidad en el eje vertical. Cuando las escalas son distintas, las distancias se distorsionan.

### ***La calidad global de un mapa bidimensional.***

La calidad global de un mapa bidimensional del AC, viene determinado por el valor de la inercia explicada por los dos primeros ejes principales.

## 2.7. Cálculo de las coordenadas principales y las coordenadas estándares

El algoritmo de cálculo para obtener las coordenadas de los perfiles fila y de los perfiles columna en relación con los ejes principales, (utilizando el teorema de la descomposición en valores singulares (DVS) y el teorema de Eckart-Young) , es el siguiente:

*Paso 1.- Cálculo de la matriz  $\mathbf{S}$  de residuos estandarizados:*

$$\mathbf{S} = \mathbf{D}_r^{-\frac{1}{2}}(\mathbf{P} - \mathbf{r}\mathbf{c}^T)\mathbf{D}_c^{-\frac{1}{2}}$$

algebraicamente:

$$s_{ij} = \frac{p_{ij} - r_i c_j}{\sqrt{r_i c_j}}$$

*Paso 2.- Aplicando el teorema, realizamos el cálculo de la DVS de  $\mathbf{S}$ :*

$$\mathbf{S} = \mathbf{U}\mathbf{D}_\alpha\mathbf{V}^T$$

donde

$$\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V}$$

y

$$\mathbf{D}_\alpha$$

es la matriz diagonal de valores singulares (positivos) en orden descendiente

$$\alpha_1 \geq \alpha_2 \geq \dots$$

Las columnas de las matrices  $\mathbf{U}$  y  $\mathbf{V}$  son, respectivamente, los *vectores singulares izquierdo y derecho*.

*Paso 3.- Cálculo de las coordenadas estándares de las filas  $\Phi$ :*

$$\Phi = \mathbf{D}_r^{-\frac{1}{2}}\mathbf{U}$$

Denotamos por  $\phi_{ik}$  la coordenada estándar de la fila  $i$  en el eje  $k$ .

*Paso 4.- Cálculo de las coordenadas estándares de las columnas  $\Gamma$ :*

$$\Gamma = \mathbf{D}_c^{-\frac{1}{2}}\mathbf{V}$$

Denotamos por  $\gamma_{jk}$  la coordenada estándar de la columna  $j$  en el eje  $k$ .

*Paso 5.- Cálculo de las coordenadas principales de las filas F:*

$$\mathbf{F} = \mathbf{D}_r^{-\frac{1}{2}} \mathbf{U} \mathbf{D}_\alpha = \Phi \mathbf{D}_\alpha$$

Denotamos por  $f_{ik}$  la coordenada principal de la fila  $i$  en el eje  $k$ .

*Paso 6.- Cálculo de las coordenadas principales de las columnas G:*

$$\mathbf{G} = \mathbf{D}_c^{-\frac{1}{2}} \mathbf{V} \mathbf{D}_\alpha = \Gamma \mathbf{D}_\alpha$$

Denotamos por  $g_{jk}$  la coordenada principal de la columna  $j$  en el eje  $k$ .

*Paso 7.- Inercias principales  $\lambda_k$ :*

Por la proposición tenemos que los valores propios de:

$$\mathbf{S} \mathbf{S}^T = \mathbf{U} \mathbf{D}_\alpha^2 \mathbf{U}^T \quad \text{o} \quad \text{de} \quad \mathbf{S}^T \mathbf{S} = \mathbf{V} \mathbf{D}_\alpha^2 \mathbf{V}^T$$

representados por

$$\lambda_k, \quad k = 1, 2, \dots, \mathbf{K}, \quad \mathbf{K} = \min[I - 1, J - 1]$$

son la raíz cuadrada de los valores singulares de  $\mathbf{S}$ :

$$\lambda_k = \alpha_k^2$$

Tengamos en cuenta el cálculo de las escalas de las coordenadas principales:

$$\mathbf{F} \mathbf{D}_r \mathbf{F}^T = \mathbf{G} \mathbf{D}_c \mathbf{G}^T = \mathbf{D}_\lambda$$

Es decir, la suma de cuadrados ponderada de las coordenadas principales en la  $k$ -ésima dimensión (es decir, su inercia en la dirección de esta dimensión) es igual a la inercia principal (o valor propio)  $\lambda_k = \alpha_k^2$ , el cuadrado del  $k$ -ésimo valor singular, como nos indica la proposición.

También, fijémonos en el cálculo de las coordenadas estándares:

$$\Phi \mathbf{D}_r \Phi^T = \Gamma \mathbf{D}_c \Gamma^T = \mathbf{I}$$

Es decir, la suma de cuadrados ponderada de las coordenadas estándares es igual a 1. Todas las matrices de coordenadas tienen columnas ortogonales, en las que siempre utilizamos las masas para el cálculo de los productos escalares (ponderados).

## 2.8. Análisis y diagnóstico de las contribuciones a la inercia

*La inercia total mide la variación total de los perfiles.*

Anteriormente vimos que la inercia total, geoméricamente es una media ponderada de las distancias ji-cuadrado entre los perfiles y el perfil medio:

$$\text{Inercia total} = \sum_i (\text{i-ésimo masa}) \times (\text{distancia ji-cuadrado de i-ésimo perfil medio})^2.$$

Si existen pequeñas diferencias entre los perfiles y su media, la inercia total toma un valor próximo a cero, es decir existe poca variación. En cambio, si los perfiles se concentran en unas pocas categorías, y además en categorías distintas en los diferentes perfiles, la inercia es grande. En definitiva, la inercia es una medida de la dispersión de los perfiles en el espacio de perfiles.

*Inercia de filas e inercia de columnas.*

La inercia de las filas, es la contribución de cada fila a la inercia.

$$\text{Inercia de la fila i-ésima} = r_i \sqrt{\sum_j (a_{ij} - a_{i'j})^2 / c_j}$$

La inercia de las columnas, es la contribución de cada columna a la inercia.

$$\text{Inercia de la columna j-ésima} = c_j \sqrt{\sum_i (b_{ij} - b_{ij'})^2 / r_i}$$

Para facilitar la interpretación, podemos expresar los componentes de la inercia con relación a la inercia total, como porcentajes o más habitualmente en tantos por mil.

*Contribuciones grandes y contribuciones pequeñas a la inercia.*

Para decidir que contribuciones son pequeñas o grandes, como regla general, utilizaremos como valor umbral la media de las contribuciones a la inercia.

Por tanto consideramos contribuyentes principales, las filas o columnas con contribuciones mayores que este umbral.

*Contribuciones de las celdas a la inercia.*

Denominamos *contribuciones ji-cuadrado* a las contribuciones de las celdas, ya que sus valores son idénticos a las contribuciones relativas de cada celda al estadístico  $\chi^2$

### ***Descomposición de la inercia en ejes principales.***

Un diagrama de descomposición ("Scree plot", en inglés), es un diagrama de barras, donde se muestran todos los valores de todas las inercias principales. Dichas inercias principales se pueden interpretar por ellas mismas, como el cuadrado de correlaciones canónicas (Véase la tabla 3.12., pág. 54).

### ***Componentes de cada inercia principal.***

Como vimos anteriormente una inercia principal, es una inercia, calculada a partir de las proyecciones de los perfiles fila (o columna) sobre los ejes principales. También vimos que, la suma ponderada de los cuadrados de las distancias de los perfiles fila (o columna) a su centroide es igual a la inercia total. Dijimos que el primer eje principal es la recta mínimo-cuadrática más próxima a los perfiles, y que este eje pasa por el centroide de las filas (o columnas) que se halla en el origen, o punto cero de representación. Entonces, si proyectamos todos los perfiles fila (o columna) sobre este eje, tendremos que la primera inercia principal es la suma ponderada de los cuadrados de las distancias de estas proyecciones al centroide.

### ***Descomposición completa de la inercia de los perfiles en los ejes principales.***

Podemos repetir el mismo proceso para todos los ejes principales. El examen de las contribuciones de las filas (o de las columnas) a la inercia de los ejes principales nos proporciona un respaldo numérico a la interpretación de los mapas.

### ***Componentes de la inercia de cada perfil.***

Podemos descomponer las inercias principales con relación a las filas (o columnas). Para ello tenemos dos posibilidades para expresar la contribución del k-ésimo eje a la inercia de las filas (o columnas):

a) *con relación a la inercia principal del eje:*

Esta posibilidad nos permite diagnosticar que filas (o columnas) han tenido un mayor papel en la determinación de la orientación de los ejes principales.

b) *con relación a la inercia de la fila o columna:*

Esta posibilidad nos permite diagnosticar la posición de los perfiles con relación a los ejes y si estos están bien representados en el mapa. Si están bien representados los podemos interpretar con seguridad. Si están pobremente representados, debemos interpretar sus posiciones con más cautela.

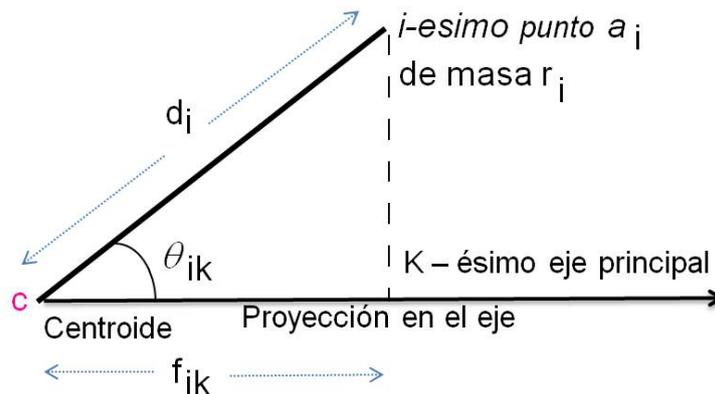
Estos valores de inercia los podemos interpretar como correlaciones al cuadrado, pues son los cuadrados de los cosenos de los ángulos formados por los perfiles y los ejes principales.

## Geometría de la inercia y álgebra de descomposición de la inercia.

En este apartado, tomaremos como referencia el análisis de las filas, pero analógicamente se realiza para el análisis de columnas.

Representamos gráficamente la descomposición de la inercia.

		Ejes				
		Eje 1	Eje 2	..	Eje p	
<b>1</b>	$r_1 f_{11}^2$	$r_1 f_{12}^2$	..	$r_1 f_{1p}^2$	$r_1 d_1^2$	
<b>2</b>	$r_2 f_{21}^2$	$r_2 f_{22}^2$	..	$r_2 f_{2p}^2$	$r_2 d_2^2$	
<b>3</b>	$r_3 f_{31}^2$	$r_3 f_{32}^2$	..	$r_3 f_{3p}^2$	$r_3 d_3^2$	
:	:	:	:	:	:	
:	:	:	:	:	:	
:	:	:	:	:	:	
:	:	:	:	:	:	
<b>n</b>	$r_n f_{n1}^2$	$r_n f_{n2}^2$	..	$r_n f_{np}^2$	$r_n d_n^2$	
	$\lambda_1$	$\lambda_2$	..	$\lambda_p$		



Recordemos que, por  $a_i$  (Tabla de perfiles, sección 2.1., pág. 4), denotamos el  $i$ -ésimo perfil, de masa  $r_i$ , a una distancia  $d_i$  del perfil fila medio  $c$ . También, que la coordenada principal de  $a_i$ , en el eje principal  $k$ , es  $f_{ik}$  (Paso 5, sección 2.7., pág. 32).

Entonces, la contribución absoluta de un punto  $i$  a la inercia principal del eje  $k$  es:

$$r_i f_{ik}^2$$

Por la ecuación de la inercia total:

$$\phi^2 = \frac{\chi^2}{n} = \sum_i r_i \|a_i - c\|_c^2$$

Tenemos que la inercia total es:

$$\phi^2 = \sum_i r_i d_i^2$$

Por tanto, la contribución absoluta de la inercia en este eje (i.e. la inercia principal  $k$ -ésima), es:

$$\sum_i r_i f_{ik}^2 = \lambda_k$$

Con lo obtenemos que, la contribución relativa de un punto  $i$  a la inercia principal del eje  $k$  es:

$$\frac{r_i f_{ik}^2}{\lambda_k}.$$

Por el teorema de Pitágoras, sabemos que

$$d_i^2 = \sum_k f_{ik}^2$$

Entonces, tenemos que la inercia total es:

$$\phi^2 = \sum_i r_i d_i^2 = \sum_i r_i \sum_k f_{ik}^2 = \sum_k \lambda_k$$

y, en consecuencia, la contribución relativa de de la inercia principal  $k$ -ésima, es:

$$\frac{\lambda_k}{\sum_k \lambda_k}.$$

Por tanto, la contribución de los ejes a la inercia de las filas es:

$$\sum_k r_i f_{ik}^2 = r_i d_i^2$$

La contribución relativa del eje  $k$  a la inercia del punto  $i$  es (i.e. la proporción de inercia del punto  $i$  explicada por el eje  $k$ ):

$$\frac{r_i f_{ik}^2}{r_i d_i^2} = \left( \frac{f_{ik}}{d_i} \right)^2$$

***Las contribuciones relativas como cuadrados de los cosenos de los ángulos.***

Sea  $\theta_{ik}$ , el ángulo formado por el punto  $i$  el eje  $k$ .

Entonces:

la contribución relativa del eje  $k$  a la inercia del punto  $i$  es  $\cos^2(\theta_{ik})$

Es decir:

$$\left( \frac{f_{ik}}{d_i} \right)^2 = \cos^2(\theta_{ik})$$

Esto nos explica si un perfil se halla cerca del eje, o si no se halla cerca de ese eje, y más bien se encuentra en otras dimensiones del espacio.

### *Las contribuciones relativas como correlaciones al cuadrado.*

Podemos interpretar los cosenos de los ángulos entre vectores como coeficientes de correlación. Por tanto, las contribuciones relativas son correlaciones al cuadrado. Si la correlación es 1, el perfil se halla sobre el eje principal, mientras que si la correlación es cero, el perfil es perpendicular al eje principal (ángulo de  $90^\circ$ ).

### *Calidad de la representación de un espacio.*

Por el teorema de Pitágoras, podemos sumar los cuadrados de los cosenos de los ángulos formados por un perfil y cada uno de los ejes, y así obtener una suma de cosenos al cuadrado que relaciona el perfil con el subespacio definido por estos ejes.

Así, de esta manera, podemos calcular el ángulo entre un perfil fila y el plano principal a partir de la suma de las contribuciones relativas de los dos ejes principales.

Interpretamos estas sumas como una medida de la **calidad de la representación** de los perfiles en los mapas bidimensionales (mapa asimétrico y mapa simétrico), de la misma manera que la suma de los dos primeros porcentajes de inercia nos da una medida de la **calidad global (o media) de la representación**. Además, podemos ver que perfiles están bien representados y cuáles no.

Es decir, la suma de los cuadrados de las correlaciones de un perfil con las dimensiones de un determinado subespacio nos proporciona una medida de la calidad de la representación del perfil en dicho subespacio.

## 2.9. Relaciones de transición

En esta sección veremos las ecuaciones de transición, es decir las relaciones matemáticas existentes entre el análisis de filas y columnas.

### *La relación entre perfiles y vértices.*

La posición de los perfiles fila resulta de calcular las medias ponderadas de los vértices de las columnas, siendo estos los pesos de los valores de los perfiles fila.

Para los perfiles columna y los vértices fila se cumple la misma relación.

Estas relaciones, basadas en el cálculo de medias ponderadas, también se cumplen para las proyecciones de filas y columnas sobre cualquier subespacio. En concreto, para las proyecciones de las coordenadas en los ejes principales. Es decir, en un eje principal  $k$ , las posiciones de las coordenadas principales de las filas son medias ponderadas de las coordenadas estándares de las columnas y viceversa.

### *Ecuaciones de transición entre filas y columnas.*

Llamamos *ecuaciones de transición*, a aquellas que expresan la relación entre filas y columnas basada en el cálculo de medias ponderadas (o medias condicionales).

Entonces, *las ecuaciones de transición entre las coordenadas principales y las coordenadas estándares* son:

para las coordenadas principales de una fila a partir de las coordenadas estándares de todas las columnas:

$$f_{ik} = \sum_j \left( \frac{p_{ij}}{r_i} \right) \gamma_{jk},$$

matricialmente:

$$\mathbf{F} = \mathbf{D}_r^{-1} \mathbf{P} \mathbf{\Gamma},$$

y para las coordenadas principales de una columna a partir de las coordenadas estándares de todas las filas:

$$g_{jk} = \sum_i \left( \frac{p_{ij}}{c_j} \right) \phi_{ik},$$

matricialmente:

$$\mathbf{G} = \mathbf{D}_c^{-1} \mathbf{P} \mathbf{\Phi}.$$

Entre paréntesis hemos mostrado los pesos, que son los perfiles de las filas  $\left( \frac{p_{ij}}{r_i} \right)$ , y los de las columnas  $\left( \frac{p_{ij}}{c_j} \right)$ , respectivamente.

Las relaciones existentes entre las coordenadas principales y las coordenadas estándar son las que siguen a continuación:

para la coordenada principal de una fila a partir de la coordenada estándar de esta fila:

$$f_{ik} = \sqrt{\lambda_k} \phi_{ik},$$

matricialmente:

$$\mathbf{F} = \mathbf{D}_r^{-\frac{1}{2}} \mathbf{U} \mathbf{D}_\alpha = \Phi \mathbf{D}_\alpha,$$

y para la coordenada principal de una columna a partir de la coordenada estándar de esta columna:

$$g_{jk} = \sqrt{\lambda_k} \gamma_{jk},$$

dónde  $\lambda_k$  es la inercia principal del  $k$ -ésimo eje, y en notación matricial:

$$\mathbf{G} = \mathbf{D}_c^{-\frac{1}{2}} \mathbf{V} \mathbf{D}_\alpha = \Gamma \mathbf{D}_\alpha.$$

Por otro lado, las ecuaciones de transición entre las coordenadas principales de las filas y las coordenadas principales de las columnas, las expresamos de la siguiente manera:

para las coordenadas principales de las filas a partir de las coordenadas principales de las columnas:

$$f_{ik} = \frac{1}{\sqrt{\lambda_k}} \sum_j \left( \frac{p_{ij}}{r_i} \right) g_{jk},$$

matricialmente:

$$\mathbf{F} = \mathbf{D}_r^{-1} \mathbf{P} \mathbf{G} \mathbf{D}_\lambda^{-\frac{1}{2}},$$

y para la coordenada principal de una columna a partir de las coordenadas principales de las filas:

$$g_{jk} = \frac{1}{\sqrt{\lambda_k}} \sum_i \left( \frac{p_{ij}}{c_j} \right) f_{ik},$$

matricialmente:

$$\mathbf{G} = \mathbf{D}_c^{-1} \mathbf{D}^T \mathbf{F} \mathbf{D}_\lambda^{-\frac{1}{2}}.$$

De manera similar, expresamos *las ecuaciones de transición entre las coordenadas estándares de las filas y las coordenadas estándares de columnas*:

para la coordenada estándar de una fila a partir de las coordenadas estándares de las columnas:

$$\phi_{ik} = \frac{1}{\sqrt{\lambda_k}} \sum_j \left( \frac{p_{ij}}{r_i} \right) \gamma_{jk},$$

matricialmente:

$$\Phi = \mathbf{D}_r^{-1} \mathbf{P} \mathbf{\Gamma} \mathbf{D}_\lambda^{-\frac{1}{2}},$$

y para la coordenada estándar de una columna a partir de las coordenadas estándares de las filas:

$$\gamma_{jk} = \frac{1}{\sqrt{\lambda_k}} \sum_i \left( \frac{p_{ij}}{c_j} \right) \phi_{ik},$$

matricialmente:

$$\mathbf{\Gamma} = \mathbf{D}_c^{-1} \mathbf{P} \mathbf{\Phi} \mathbf{D}_\lambda^{-\frac{1}{2}}.$$

### *Cálculo recíproco de medias.*

Las ecuaciones de transición «de la obtención de las coordenadas principales a partir de las coordenadas estándares», son la base del conocido algoritmo para hallar la solución del AC, llamado *cálculo recíproco de medias*.

Empezamos el algoritmo con unas coordenadas estándares de las columnas que hemos centrado y normalizado con medias y sumas de cuadrados ponderados-. Aplicando la fórmula de *la primera ecuación de transición* de cálculo de medias ponderadas, calculmos valores para las coordenadas de las filas, a continuación aplicando la fórmula de *la segunda ecuación de transición* a los valores anteriores a las filas, calculamos nuevos valores y repetimos el proceso desde el inicio hasta la convergencia de los resultados, es decir, hasta obtener las coordenadas principales en el primer eje principal (es necesario estandarizar los valores de las coordenadas de las columnas que obtenemos en cada iteración, en caso contrario en los sucesivos cálculos de medias llegaríamos al valor cero). Se halla el segundo conjunto de coordenadas, de la misma manera, asegurando la ortogonalidad con las primeras coordenadas.

# Capítulo 3

## Aplicación a un estudio de geografía electoral

### 3.1. Introducción

Normalmente en los resultados electorales se utilizan técnicas estadísticas de tipo univariante, i.e., las estadísticas se hacen por separado para cada variable.

En geografía electoral, es poco frecuente utilizar técnicas estadísticas multivariantes. Pero vemos que los resultados son de tipo multivariante, es decir, pensemos por ejemplo, en lo que es el perfil electoral de una comarca, es decir un vector con tantas componentes como opciones electorales (en nuestro caso 10), en la que cada componente recoge la frecuencia relativa de voto de aquella opción electoral.

Estos perfiles se pueden interpretar geoméricamente como puntos de un cierto espacio multidimensional, y por tanto pueden ser analizadas con métodos estadísticos multivariantes.

El objetivo de esta aplicación a un estudio electoral es doble. Desde el punto de vista metodológico de la teoría del AC, es introducirlo dentro de un estudio de geografía electoral, ya que el AC es una técnica estadística multivariante, del tipo descriptivo, muy adecuada para el tratamiento de frecuencias y que permite representar de una manera óptima los perfiles electorales en un plano bidimensional.

Desde el punto de vista de geografía electoral, hacer una descripción en profundidad transversalmente de las diferencias en el comportamiento electoral de las comarcas de Cataluña, en las elecciones al Parlament 2012.

Los datos de nuestro análisis son los resultados de las elecciones al Parlament de Catalunya 2012, agrupados por comarcas. Es decir, las entidades territoriales consideradas son las 41 comarcas de Cataluña.

Evidentemente, tomar las comarcas como entidades territoriales representa una pérdida de información, pero nos simplifica el análisis y se trata de una unidad tradicional en Cataluña.

Como diferentes opciones de voto, hemos considerado 10 categorías: CiU, ERC-Cat Sí, PSC, P.P., ICV-EUiA, Ciutadans (C's), CUP-Alternativa d'Esquerres, Otros, Nulos + Blancos, y Abstención. Por un lado los 7 partidos con representación parlamentaria, la categoría de Abstención, Otros, que recoge los votos de los partidos extra-parlamentarios, y los votos Nulos y Blancos.

Los datos utilizados son los datos oficiales publicados por la Generalitat de Catalunya. Vemos que para cada uno de los recuentos electorales, se pueden recoger en una tabla de doble entrada con 41 filas, que representan la comarcas de Cataluña, y 10 columnas que representan las opciones de voto que hemos considerado.

El punto de vista adecuado para nuestro análisis es considerar que tenemos 41 vectores de frecuencias, de dimensión 10.

## 3.2. Análisis de los resultados de las elecciones al Parlament de Catalunya 2012

### 3.2.1. Tablas preeliminares

Analizamos a fondo en esta sección los resultados de las elecciones al Parlament de Catalunya realizadas el 25 de Noviembre de 2012.

**Tabla 3.1.** - Resultados electorales (frecuencias absolutas).

	CiU	ERC	PSC	P.P.	ICV-EUiA	C's	CUP-A.E.	Otros	Nulos+Blancos	Abstención
Alt Camp	8.721	4.086	2.303	2.158	1.218	973	1.044	1166	523	9.856
Alt Empordà	22.995	9.219	5.653	7.474	2.847	2.270	2.384	2904	1141	26.647
Alt Penedès	20.933	9.468	6.170	4.613	4.452	2.178	3.099	2981	1251	20.343
Alt Urgell	4.978	1.732	1.004	1.015	629	248	461	339	303	4.944
Alta Ribagorça	743	348	321	155	133	58	84	88	80	988
Anoia	20.744	9.018	7.854	6.603	4.534	3.255	1.965	4410	1566	26.646
Bages	37.850	16.672	10.887	7.793	6.444	3.654	4.014	5417	2151	40.322
Baix Camp	28.578	10.801	10.626	13.348	5.013	6.497	3.789	4014	2124	42.050
Baix Ebre	11.766	7.446	4.003	3.830	2.542	878	1.126	1718	1566	19.918
Baix Empordà	25.195	10.548	7.120	5.758	3.647	2.016	2.037	2768	1227	26.528
Baix Llobregat	78.137	42.146	79.272	56.333	51.739	42.615	9.978	27790	10219	180.344
Baix Penedès	11.829	4.930	7.351	6.899	3.597	3.800	1.361	2774	1260	26.179
Barcelonès	277.791	127.738	163.027	166.920	132.317	94.883	37.810	61536	21959	481.132
Berguedà	11.373	5.050	1.751	1.252	1.050	380	1.277	1010	501	8.374
Cerdanya	4.812	1.450	593	962	409	284	226	426	201	3.820
Conca de Barberà	4.972	2.990	698	655	587	226	420	451	287	4.158
Garraf	19.020	8.396	10.092	8.718	6.553	5.524	3.989	3640	1569	33.148
Garrigues	5.347	2.406	838	769	388	136	353	331	290	4.102
Garrotxa	15.351	6.065	1.943	1.508	1.272	419	1.390	1124	555	10.012
Gironès	35.444	15.541	8.317	7.421	5.173	3.330	4.683	4164	1692	31.588
Maresme	78.295	31.268	26.018	27.322	19.148	14.828	6.988	13623	4522	90.273
Montsià	9.918	6.286	3.810	3.712	1.716	618	1.041	1409	1593	16.917
Noguera	9.709	3.836	1.594	1.857	869	357	530	656	543	7.858
Osona	42.302	17.618	4.883	4.241	3.499	1.747	2.980	5200	1333	24.626
Pallars Jussà	3.470	1.452	655	467	344	95	271	200	183	3.238
Pallars Sobirà	1.879	905	298	191	209	60	251	90	118	1.392
Pla de l'Estany	8.771	3.784	654	820	633	199	971	590	255	4.788
Pla d'Urgell	8.961	3.434	1.387	1.463	741	351	475	663	645	6.616
Priorat	2.533	1.449	422	302	317	56	506	164	119	1.552
Ribera d'Ebre	4.809	2.860	1.311	947	685	208	358	523	423	4.819
Ripollès	7.729	2.928	1.268	956	797	308	563	643	383	4.972
Segarra	4.953	2.002	635	919	406	209	387	434	326	3.946
Segrià	33.802	13.843	12.179	13.427	6.013	4.720	2.334	4195	3052	46.185
Selva	28.382	11.951	9.265	8.292	5.677	3.616	2.274	3630	1678	35.948
Solsonès	3.452	1.758	407	442	266	79	309	256	159	2.608
Tarragonès	27.537	11.707	17.395	20.726	8.483	12.656	2.931	5809	2761	55.214
Terra Alta	2.765	1.396	604	920	318	100	249	203	295	2.565
Urgell	8.769	3.384	1.254	1.517	765	267	680	739	502	7.195
Val d'Aran	1.297	262	825	911	235	235	56	189	241	2.440
Vallès Occidental	115.588	51.044	80.416	53.685	51.706	43.697	14.297	28600	10245	189.869
Vallès Oriental	61.105	27.249	28.434	23.458	21.489	16.622	6.257	12862	5106	85.459

Para comparar los comportamientos electorales de las comarcas, consideraremos los vectores de frecuencias relativas de voto o "perfiles electorales". Los componentes del perfil electoral de una comarca nos indicarán el peso relativo de cada partido en la comarca.

**Tabla 3.2.** - Perfiles electorales del año 2012 en %.

	CiU	ERC	PSC	P.P.	ICV-EUiA	C's	CUP-A.E.	Otros	Nulos+Blancos	Abstención
Alt Camp	27,21	12,75	7,19	6,73	3,80	3,04	3,26	3,64	1,63	30,75
Alt Empordà	27,53	11,04	6,77	8,95	3,41	2,72	2,85	3,48	1,37	31,90
Alt Penedès	27,73	12,54	8,17	6,11	5,90	2,89	4,11	3,95	1,66	26,95
Alt Urgell	31,80	11,06	6,41	6,48	4,02	1,58	2,95	2,17	1,94	31,58
Alta Ribagorça	24,78	11,61	10,71	5,17	4,44	1,93	2,80	2,94	2,67	32,96
Anoia	23,96	10,41	9,07	7,63	5,24	3,76	2,27	5,09	1,81	30,77
Bages	27,99	12,33	8,05	5,76	4,77	2,70	2,97	4,01	1,59	29,82
Baix Camp	22,53	8,52	8,38	10,52	3,95	5,12	2,99	3,16	1,67	33,15
Baix Ebre	21,47	13,59	7,31	6,99	4,64	1,60	2,06	3,14	2,86	36,35
Baix Empordà	29,01	12,15	8,20	6,63	4,20	2,32	2,35	3,19	1,41	30,55
Baix Llobregat	13,51	7,28	13,70	9,74	8,94	7,37	1,72	4,80	1,77	31,17
Baix Penedès	16,90	7,04	10,50	9,86	5,14	5,43	1,94	3,96	1,80	37,41
Barcelonès	17,75	8,16	10,42	10,67	8,45	6,06	2,42	3,93	1,40	30,74
Berguedà	35,52	15,77	5,47	3,91	3,28	1,19	3,99	3,15	1,56	26,15
Cerdanya	36,50	11,00	4,50	7,30	3,10	2,15	1,71	3,23	1,52	28,98
Conca de Barberà	32,19	19,36	4,52	4,24	3,80	1,46	2,72	2,92	1,86	26,92
Garraf	18,90	8,34	10,03	8,66	6,51	5,49	3,96	3,62	1,56	32,93
Garrigues	35,74	16,08	5,60	5,14	2,59	0,91	2,36	2,21	1,94	27,42
Garrotxa	38,73	15,30	4,90	3,80	3,21	1,06	3,51	2,84	1,40	25,26
Gironès	30,20	13,24	7,09	6,32	4,41	2,84	3,99	3,55	1,44	26,92
Maresme	25,07	10,01	8,33	8,75	6,13	4,75	2,24	4,36	1,45	28,91
Montsià	21,09	13,37	8,10	7,89	3,65	1,31	2,21	3,00	3,39	35,98
Noguera	34,91	13,79	5,73	6,68	3,12	1,28	1,91	2,36	1,95	28,26
Osona	39,01	16,25	4,50	3,91	3,23	1,61	2,75	4,80	1,23	22,71
Pallars Jussà	33,45	14,00	6,31	4,50	3,32	0,92	2,61	1,93	1,76	31,21
Pallars Sobirà	34,84	16,78	5,53	3,54	3,88	1,11	4,65	1,67	2,19	25,81
Pla de l'Estany	40,86	17,63	3,05	3,82	2,95	0,93	4,52	2,75	1,19	22,31
Pla d'Urgell	36,23	13,88	5,61	5,91	3,00	1,42	1,92	2,68	2,61	26,75
Priorat	34,14	19,53	5,69	4,07	4,27	0,75	6,82	2,21	1,60	20,92
Ribera d'Ebre	28,38	16,88	7,74	5,59	4,04	1,23	2,11	3,09	2,50	28,44
Ripollès	37,62	14,25	6,17	4,65	3,88	1,50	2,74	3,13	1,86	24,20
Segarra	34,84	14,08	4,47	6,46	2,86	1,47	2,72	3,05	2,29	27,76
Segrià	24,19	9,91	8,71	9,61	4,30	3,38	1,67	3,00	2,18	33,05
Selva	25,64	10,79	8,37	7,49	5,13	3,27	2,05	3,28	1,52	32,47
Solsonès	35,46	18,06	4,18	4,54	2,73	0,81	3,17	2,63	1,63	26,79
Tarragonès	16,67	7,09	10,53	12,54	5,13	7,66	1,77	3,52	1,67	33,42
Terra Alta	29,37	14,83	6,42	9,77	3,38	1,06	2,64	2,16	3,13	27,24
Urgell	34,98	13,50	5,00	6,05	3,05	1,06	2,71	2,95	2,00	28,70
Val d'Aran	19,38	3,92	12,33	13,62	3,51	3,51	0,84	2,82	3,60	36,47
Vallès Occidental	18,08	7,99	12,58	8,40	8,09	6,84	2,24	4,47	1,60	29,71
Vallès Oriental	21,21	9,46	9,87	8,14	7,46	5,77	2,17	4,47	1,77	29,67

Evidentemente, estos datos en bruto son difíciles de interpretar. Si nos interesa conocer el peso de un determinado partido en una comarca, el dato que necesitamos conocer es la frecuencia relativa de los votos del partido en la comarca respecto el número de votantes de la comarca (el total de la fila) y el mismo, si nos interesa conocer el peso de la comarca sobre el partido. Pero este último dato, es mucho menos interesante ya que evidentemente las comarcas más pobladas tienen más peso en todos los partidos.

Así pues, lo primero que tenemos que hacer es obtener las tablas de frecuencias relativas (pueden ser frecuencias porcentuales) o perfiles.

Tabla 3.3. - Perfiles fila, vectores de frecuencias o perfiles electorales, que indican el peso relativo de los partidos en cada comarca.

	CiU	ERC	PSC	PP.	ICV-EUiA	C's	CUP-A.E.	Otros	Nulos+Blancos	Abstención	MASAS DE LAS FILAS
Alt Camp	0.272123	0.127496	0.071861	0.067336	0.038005	0.030361	0.032576	0.036383	0.016319	0.307539	0.006096
Alt Empordà	0.275277	0.110362	0.067673	0.089473	0.034082	0.027175	0.028539	0.034764	0.013659	0.318996	0.015889
Alt Penedès	0.273302	0.125424	0.081735	0.061109	0.038976	0.028852	0.041053	0.039490	0.016572	0.269487	0.014359
Alt Urgell	0.318022	0.110650	0.064141	0.064844	0.040184	0.015844	0.029451	0.021657	0.019357	0.315850	0.002977
Alta Ribagorça	0.247832	0.116077	0.107071	0.051701	0.044363	0.019346	0.028019	0.029353	0.026684	0.329553	0.000570
Anoia	0.239552	0.104140	0.090698	0.076252	0.032359	0.037589	0.022692	0.030027	0.018084	0.307708	0.016471
Bages	0.279947	0.123310	0.080523	0.057639	0.047661	0.027026	0.029688	0.040065	0.015009	0.298231	0.023717
Baix Camp	0.225307	0.085155	0.083775	0.102355	0.039522	0.031222	0.029872	0.031646	0.016746	0.331520	0.024126
Baix Ebre	0.214735	0.135893	0.073057	0.069899	0.046393	0.016024	0.020550	0.031354	0.028580	0.363514	0.010422
Baix Empordà	0.290118	0.121459	0.081986	0.066303	0.041995	0.023214	0.023456	0.031873	0.014129	0.305467	0.016519
Baix Llobregat	0.135051	0.072845	0.137013	0.097365	0.089425	0.073655	0.017246	0.048032	0.017662	0.311705	0.110051
Baix Penedès	0.169034	0.070449	0.105044	0.098585	0.051400	0.054301	0.019448	0.039640	0.018005	0.374093	0.013311
Barcelona	0.177489	0.081616	0.104163	0.106650	0.084541	0.060624	0.024158	0.039317	0.014030	0.307410	0.297701
Berguedà	0.355206	0.157724	0.054688	0.039103	0.032794	0.011868	0.039884	0.031545	0.015647	0.261540	0.006090
Cerdanya	0.365016	0.109990	0.044982	0.072973	0.031025	0.021543	0.017143	0.032314	0.015247	0.289767	0.002508
Conca de Barberà	0.321937	0.193603	0.045196	0.042411	0.038908	0.014634	0.027195	0.029202	0.018583	0.269231	0.002938
Garraf	0.188974	0.083419	0.100269	0.086618	0.065107	0.054884	0.039633	0.036165	0.015589	0.329343	0.019145
Garrigues	0.357420	0.160829	0.056016	0.051404	0.025936	0.009091	0.025596	0.022126	0.019385	0.274198	0.002846
Garrotxa	0.387270	0.153006	0.049017	0.038043	0.032090	0.010570	0.035066	0.028356	0.014001	0.252580	0.007540
Gironès	0.302029	0.132430	0.070872	0.063237	0.044081	0.028376	0.039905	0.035483	0.014418	0.269171	0.022322
Maresme	0.250716	0.100126	0.083315	0.087491	0.061316	0.047482	0.022377	0.043624	0.014480	0.289072	0.059400
Montsià	0.210932	0.133688	0.081029	0.078945	0.036495	0.013143	0.022240	0.029966	0.033879	0.339783	0.008944
Noguera	0.349132	0.137941	0.057320	0.066777	0.031249	0.012838	0.019059	0.023589	0.019526	0.282570	0.005290
Osona	0.390135	0.162484	0.045034	0.039113	0.032270	0.016112	0.027483	0.047958	0.012294	0.227116	0.020624
Pallars Jussà	0.334458	0.139952	0.063133	0.045012	0.033157	0.009157	0.026120	0.019277	0.017639	0.312096	0.001973
Pallars Sobirà	0.348415	0.167810	0.055257	0.035416	0.038754	0.011126	0.046542	0.016688	0.021880	0.258112	0.001026
Pla de l'Estany	0.408619	0.176287	0.030468	0.038202	0.029490	0.009271	0.045236	0.027487	0.011880	0.223061	0.004083
Pla d'Urgell	0.362266	0.138826	0.056072	0.059145	0.029956	0.014190	0.019203	0.026803	0.026075	0.267464	0.004705
Priorat	0.341375	0.195283	0.056873	0.040701	0.042722	0.007547	0.068194	0.022102	0.016038	0.209164	0.001411
Ribera d'Ebre	0.283834	0.168801	0.077377	0.055893	0.040430	0.012276	0.021130	0.030868	0.024966	0.284424	0.003223
Ripollès	0.376162	0.142303	0.061712	0.046527	0.038789	0.014990	0.027401	0.031294	0.018640	0.241982	0.003908
Segarra	0.348386	0.140817	0.044665	0.064641	0.028557	0.014701	0.027221	0.030527	0.022930	0.277555	0.002704
Segrià	0.241875	0.090055	0.087148	0.096079	0.043027	0.033775	0.016701	0.030018	0.021839	0.339483	0.026582
Selva	0.256357	0.107946	0.083685	0.074896	0.051277	0.032661	0.020540	0.032787	0.015156	0.324695	0.021059
Solsonès	0.354560	0.180567	0.041804	0.045399	0.027321	0.008114	0.031738	0.026294	0.016331	0.267872	0.001852
Tarragonès	0.166670	0.070857	0.105285	0.125446	0.051344	0.076601	0.017740	0.035159	0.016711	0.334187	0.031426
Terra Alta	0.293680	0.148274	0.064153	0.097716	0.033776	0.010621	0.029447	0.021561	0.031333	0.272488	0.001791
Urgell	0.349753	0.134971	0.050016	0.060506	0.030512	0.010649	0.027122	0.029475	0.020022	0.286974	0.004769
Val d'Aran	0.193842	0.039157	0.123300	0.136153	0.035122	0.035122	0.008369	0.028247	0.036019	0.364669	0.001273
Valles Occidental	0.180847	0.079863	0.125818	0.083995	0.080898	0.068368	0.022369	0.044747	0.016029	0.297066	0.121573
Valles Oriental	0.212140	0.094960	0.098715	0.081440	0.074604	0.057707	0.021233	0.044653	0.017727	0.296690	0.054788
PERFIL FILA MEDIO	0.211629	0.094433	0.099582	0.089543	0.068259	0.052242	0.024404	0.039893	0.016158	0.304257	

Tabla 3.4. -Perfiles columna, vectores de frecuencias o perfiles electorales, que indican el peso relativo de cada comarca sobre los votantes del partido.

	CIU	ERC	PSC	P.P.	ICV-EUIA	C's	CUP-A.E.	Otros	Nulos+Blancos	Abstención	PERFIL COLUMNA MEDIO
Alt Camp	0.007838	0.008230	0.004399	0.004584	0.003394	0.003543	0.008273	0.005560	0.006157	0.006162	0.006096
Alt Empordà	0.020668	0.018569	0.010798	0.015876	0.007933	0.008265	0.018891	0.013846	0.013432	0.016659	0.015889
Alt Penedès	0.018814	0.019071	0.011785	0.009799	0.012406	0.007930	0.024557	0.014214	0.014727	0.012718	0.014359
Alt Urgell	0.004474	0.003489	0.001918	0.002156	0.001753	0.000903	0.003653	0.004616	0.003567	0.003091	0.002977
Alta Ribagorça	0.000668	0.000701	0.000613	0.000329	0.000371	0.000211	0.000666	0.000420	0.000942	0.000618	0.000570
Anoia	0.018645	0.018164	0.015002	0.014026	0.012634	0.011851	0.015371	0.021027	0.018435	0.016658	0.016471
Bages	0.034019	0.033581	0.020795	0.016554	0.017957	0.013304	0.031807	0.025829	0.025322	0.025208	0.025717
Baix Camp	0.025686	0.021756	0.020297	0.028354	0.013969	0.023655	0.030024	0.019139	0.025004	0.026288	0.024126
Baix Ebre	0.010575	0.014998	0.007646	0.008136	0.007084	0.003197	0.008922	0.008192	0.018435	0.012452	0.010422
Baix Empordà	0.022645	0.021246	0.013600	0.012231	0.010163	0.007340	0.016141	0.013198	0.014444	0.016584	0.016519
Baix Llobregat	0.070229	0.084892	0.151416	0.119064	0.144176	0.155160	0.079066	0.132504	0.120209	0.112745	0.110051
Baix Penedès	0.010632	0.009930	0.014041	0.014655	0.010023	0.013836	0.010785	0.013227	0.014833	0.016366	0.013311
Barcelona	0.249676	0.257295	0.311395	0.354576	0.368715	0.345466	0.299609	0.263407	0.258502	0.300787	0.297701
Berguedà	0.010222	0.010172	0.003345	0.002660	0.002926	0.001384	0.010119	0.004816	0.005898	0.005235	0.006090
Cerdanya	0.004325	0.002921	0.001133	0.002044	0.001140	0.000823	0.001791	0.002031	0.002366	0.002388	0.002508
Conca de Barberà	0.004469	0.006023	0.001333	0.001391	0.001636	0.000823	0.003379	0.002150	0.003379	0.002599	0.002638
Garraf	0.017095	0.016912	0.019277	0.018519	0.018261	0.020113	0.031609	0.017356	0.018470	0.020723	0.019145
Garrigues	0.004806	0.004846	0.001601	0.001634	0.001081	0.000495	0.002797	0.001578	0.003414	0.002564	0.002846
Garrotxa	0.013797	0.012216	0.003711	0.003203	0.003545	0.001526	0.011014	0.005359	0.006533	0.006259	0.007540
Gironès	0.031857	0.031303	0.015886	0.015764	0.014415	0.012124	0.037108	0.019854	0.019918	0.019748	0.022322
Maresme	0.070371	0.062981	0.049697	0.058038	0.053358	0.053988	0.055373	0.064955	0.053253	0.056435	0.059400
Montsià	0.008914	0.012661	0.007277	0.007885	0.004782	0.002250	0.008249	0.006718	0.018753	0.010576	0.008944
Noguera	0.008726	0.007727	0.003045	0.003945	0.002422	0.001300	0.004200	0.003128	0.006392	0.004913	0.005290
Osona	0.038021	0.035487	0.009327	0.009009	0.009750	0.006361	0.023614	0.024794	0.015692	0.015395	0.020624
Pallars Jussà	0.003119	0.002925	0.001251	0.000692	0.000959	0.000346	0.002147	0.000954	0.002154	0.002024	0.001973
Pallars Sobirà	0.001689	0.001823	0.000569	0.000406	0.000582	0.000218	0.001989	0.000429	0.001389	0.000870	0.001026
Pla de l'Estany	0.007883	0.007622	0.001249	0.001742	0.001764	0.000725	0.007694	0.002813	0.003002	0.002993	0.004083
Pla d'Urgell	0.008054	0.006917	0.002649	0.003108	0.002065	0.001278	0.003764	0.003161	0.007503	0.004136	0.004705
Priorat	0.002277	0.002919	0.000806	0.000642	0.000883	0.000294	0.004010	0.000782	0.001401	0.000970	0.001411
Ribera d'Ebre	0.004322	0.005761	0.002504	0.002012	0.001909	0.000757	0.002837	0.002494	0.004980	0.003013	0.003223
Ripollès	0.006947	0.005898	0.002422	0.002031	0.002221	0.001121	0.004461	0.003066	0.004509	0.003108	0.003908
Segarra	0.004452	0.004033	0.001213	0.001952	0.001131	0.000761	0.003067	0.002069	0.003838	0.002467	0.002704
Segrà	0.030381	0.027883	0.023263	0.028522	0.016756	0.017185	0.018495	0.020002	0.035928	0.028873	0.026582
Selva	0.025510	0.024072	0.017697	0.017614	0.015820	0.013166	0.018019	0.017308	0.019753	0.022473	0.021059
Solsonès	0.003103	0.003541	0.000777	0.000939	0.000741	0.000288	0.002449	0.001221	0.001872	0.001630	0.001852
Tarragonès	0.024750	0.023581	0.003226	0.004027	0.023639	0.046080	0.023225	0.027698	0.032503	0.034518	0.031426
Terra Alta	0.002485	0.002812	0.001154	0.001054	0.000886	0.000364	0.001973	0.000968	0.003473	0.001604	0.001791
Urgell	0.007882	0.006816	0.002395	0.003222	0.002132	0.000972	0.005388	0.003524	0.005910	0.004498	0.004769
Val d'Aran	0.001166	0.000528	0.001576	0.001935	0.000655	0.000856	0.000444	0.000901	0.002837	0.001525	0.001273
Vallès Occidental	0.103890	0.102815	0.153601	0.153601	0.144084	0.159100	0.113290	0.113636	0.120605	0.118699	0.121573
Vallès Oriental	0.054921	0.054886	0.054311	0.049830	0.059881	0.060520	0.049581	0.061327	0.060108	0.053426	0.064788
MASAS DE LAS COLUMNAS	0.211629	0.094433	0.099582	0.089543	0.068259	0.052242	0.024004	0.039893	0.016158	0.304257	

Sumando cada una de las filas de la tabla 3.1., y dividiendo por el resultado total de los electores de Cataluña, obtenemos los pesos demográficos de cada una de las comarcas.

**Tabla 3.5.- Masas de las comarcas (Masas de las filas).**

COMARCAS	MASA FILA	Peso de la comarca (%)
Alt Camp	0,006096	0,61 %
Alt Empordà	0,015889	1,59 %
Alt Penedès	0,014359	1,44 %
Alt Urgell	0,002977	0,30 %
Alta Ribagorça	0,000570	0,06 %
Anoia	0,016471	1,65 %
Bages	0,025717	2,57 %
Baix Camp	0,024126	2,41 %
Baix Ebre	0,010422	1,04 %
Baix Empordà	0,016519	1,65 %
Baix Llobregat	0,110051	11,01 %
Baix Penedès	0,013311	1,33 %
Barcelonès	0,297701	29,77 %
Berguedà	0,006090	0,61 %
Cerdanya	0,002508	0,25 %
Conca de Barberà	0,002938	0,29 %
Garraf	0,019145	1,91 %
Garrigues	0,002846	0,28 %
Garrotxa	0,007540	0,75 %
Gironès	0,022322	2,23 %
Maresme	0,059400	5,94 %
Montsià	0,008944	0,89 %
Noguera	0,005290	0,53 %
Osona	0,020624	2,06 %
Pallars Jussà	0,001973	0,20 %
Pallars Sobirà	0,001026	0,10 %
Pla de l'Estany	0,004083	0,41 %
Pla d'Urgell	0,004705	0,47 %
Priorat	0,001411	0,14 %
Ribera d'Ebre	0,003223	0,32 %
Ripollès	0,003908	0,39 %
Segarra	0,002704	0,27 %
Segrià	0,026582	2,66 %
Selva	0,021059	2,11 %
Solsonès	0,001852	0,19 %
Tarragonès	0,031426	3,14 %
Terra Alta	0,001791	0,18 %
Urgell	0,004769	0,48 %
Val d'Aran	0,001273	0,13 %
Vallès Occidental	0,121573	12,16 %
Vallès Oriental	0,054788	5,48 %

Sumando cada una de las columnas de la tabla 3.1., y dividiendo por el resultado total de los electores de Cataluña, obtenemos los pesos demográficos de cada uno de los partidos.

**Tabla 3.6.** - Masas de los partidos (Masas de las columnas).

<b>PARTIDOS</b>	<b>MASA COLUMNA</b>
<b>CiU</b>	0,211629
<b>ERC-Cat Sí</b>	0,094433
<b>PSC</b>	0,099582
<b>P.P.</b>	0,089543
<b>ICV-EUiA</b>	0,068259
<b>C's</b>	0,052242
<b>CUP-Alternativa d'Esquerres</b>	0,024004
<b>Otros</b>	0,039893
<b>Nulos+Blancos</b>	0,016158
<b>Abstención</b>	0,304257

### 3.2.2. Análisis estadístico univariante de los resultados

Haré un estudio univariante para cada variable, en este caso para cada partido político y su frecuencia de voto, utilizando la tabla 3.2.

Este estudio es el más habitual para cualquier estudio electoral.

Haré una síntesis de los hechos más destacados.

Echando un vistazo a la tabla 3.6. vemos que la proporción global de voto de Catalunya es igual a la masa de las columnas, es decir, la suma de cada una de las columnas de la tabla 3.1., dividido por el número de electores total de Cataluña, que nos da el perfil global de Cataluña.

**Tabla 3.7.** - Masas de los partidos (Masas de las columnas) en %.

<b>PARTIDO</b>	<b>Peso del partido (%)</b>
<b>CiU</b>	21,16 %
<b>ERC-Cat Sí</b>	9,44 %
<b>PSC</b>	9,96 %
<b>P.P.</b>	8,95 %
<b>ICV-EUiA</b>	6,83 %
<b>C's</b>	5,22 %
<b>CUP-Alternativa d'Esquerres</b>	2,40 %
<b>Otros</b>	3,99 %
<b>Votos nulos + Votos en blanco</b>	1,62 %
<b>Abstención</b>	30,43 %

#### **CIU**

El voto de CiU tiene una media global de 21,16 %, y fluctua entre el 13,51 del Baix Llobregat y el 40,86 de Pla d'Estany.

#### **ERC-Cat Sí**

ERC-Cat Sí en su caso tiene una proporción global de toda Cataluña de 9,44 %, y fluctua entre 3,92 de la Vall d'Aran y el 19,53 del Priorat.

ERC tiene una cierta implantación en todas las comarcas.

Coinciden bastante con las comarcas dónde CiU obtiene los mejores resultados.

#### **PSC**

La tercera fuerza más votada fue PSC con una proporción de votos del 9,96 %, y fluctua entre 3,05 % de la Pla d'Estany y el 13,70 % del Baix Llobregat.

#### **P.P.**

La cuarta fuerza más votada por orden de votos fue P.P., y fluctua entre 3,54 % de la Pallars Sobirà y el 13,62 % del Vall d'Aran.

Su proporción global fue del 8,95 %.

### **IC-EUiA**

IC-EUiA tiene una proporción global de votos del 6,83 %, y fluctua entre 2,59 % de la Garrigues y el 8,94 % del Baix Llobregat.

### **C's**

C's tiene una proporción de votos del 5,22 %, y fluctua entre 0,75 % del Priorat y el 7,66 % del Tarragonés.

### **CUP-Alternativa d'Esquerres**

El voto de CUP-Alternativa d'Esquerres tiene una media 2,40 %, y fluctua entre 0,84 % de la Vall d'Aran y el 6,82 % del Priorat.

### **Otros**

El voto a Otros partidos varia entre el 1,67 % de Pallars Sobirà y el 5,09 % de l'Anoia, con una media de votos del 3,99 %.

### **Votos nulos y votos en blanco**

Votos nulos y votos en blanco tiene una proporción de votos del 1,62 %, y fluctua entre 1,19 % del Plà d'Estany y el 3,60 % d'Urgell.

### **Abstención**

**La abstención tiene una media global de 30,43 %** y fluctua entre el 20,92 % del Priorat hasta el 37,41 % del Baix Penedès.

Como norma general **la abstención es más alta en comarcas pobladas e industrializadas**, algunas de ellas con mucha población sobretodo de origen inmigrante, todo y que no es evidente cuál es la causa de la abstención.

Considerando la tabla 3.4. de los perfiles columna, y la tabla 3.5. de las masas de las comarcas, vemos que este caso está claro que la mayoría de votantes de todos los partidos se encuentran en las comarcas muy pobladas. Pero es muy interesante comparar el peso relativo de estas comarcas pobladas sobre cada uno de los partidos. Por ejemplo el Barcelonès representa entre el 24,97 % y el 29,96 % de los votantes de partidos como CiU, ERP o CUP-Alternativa d'Esquerres, y en cambio representa el 36,87 % de votos de ICV-EUiA y el 35,46 % de votos del PP.

El Barcelonès, el Baix Llobregat y el Vallès Occidental - juntas tienen el 52,93 de la población.

El peso de esta macro-comarca sobre los partidos es muy diferente. Por un lado CiU 42,38 % ,ERC 44,50 % y CUP-Alternativa d'Esquerres están subrepresentadas. El PSC esta equilibradamente representado con un 61,64 %, junto con el PP, con un 58,83 %. ICV-EUiA y C's están superrepresentados con un 65,70 % y el 65,97 % respectivamente.

### 3.2.3. Análisis estadístico multivariante: análisis de correspondencias de los resultados

Después de haber visto los perfiles de los partidos y de las comarcas, ahora vamos a reflejarlo en mapas bidimensionales, utilizando la distancia ji-cuadrado, y el cálculo de las coordenadas principales y las coordenadas estándares.

**Tabla 3.8.-** Valores esperados del estadístico  $\chi^2$ .

	CiU	ERC	PSC	P.P.	ICV-EUiA	C's	CUP-A.E.	Otros	Nulos+ Blancos	Abstención
Alt Camp	6782,292813	3026,391023	3191,412255	2869,684553	2187,563061	1674,242238	769,286304	1278,480224	517,825668	9750,821860
Alt Empordà	17678,234144	7888,372057	8318,504473	7479,912303	5701,943731	4363,958785	2005,166068	3332,394128	1349,726952	25415,787359
Alt Penedès	15975,465548	7128,563577	7517,265612	6759,446691	5152,732161	3943,622008	1812,028349	3011,417721	1219,721170	22967,737162
Alt Urgell	3312,631971	1478,160842	1558,761109	1401,621702	1068,457457	817,739446	375,737597	624,439932	252,918285	4762,531658
Alta Ribagorça	634,464361	283,110343	298,547614	268,450895	204,640354	156,620639	71,964564	119,598219	48,441131	912,161880
Anoia	18326,031146	8177,431684	8623,325770	7754,004428	5910,884399	4523,870652	2078,642895	3454,505585	1399,186025	26347,117417
Bages	28613,115249	12767,728776	13463,919826	12106,616024	9228,884050	7063,287806	3245,462602	5393,648283	2184,601275	41136,736108
Baix Camp	26843,048565	11977,890580	12631,013807	11357,675635	8657,966132	6626,338165	3044,691551	5059,986008	2049,457306	38591,932250
Baix Ebre	11595,799117	5174,271196	5456,410750	4906,347533	3740,113042	2862,479873	1315,261622	2185,838957	885,335180	16671,142729
Baix Empordà	18378,726818	8200,945518	8648,121752	7776,300716	5927,880880	4536,878837	2084,619939	3464,438859	1403,209321	26422,877360
Baix Llobregat	122442,944951	54636,424524	57615,606681	51807,351506	39492,789647	30225,641370	13888,176637	23080,820598	9348,475735	176034,768352
Baix Penedès	14809,811878	6608,426228	6968,766526	6266,242044	4776,761825	3655,874683	1679,813267	2791,688906	1130,723922	21291,890719
Barcelonès	331223,622431	147798,425258	155857,489061	140145,425620	106832,981288	81764,175379	37569,270948	62436,533279	25288,806948	476196,269786
Berguedà	6775,943937	3023,558031	3188,424788	2866,998253	2185,515292	1672,674987	768,566178	1277,283444	517,340934	9741,694156
Cerdanya	2789,907831	1244,911160	1312,792928	1180,449684	899,857833	688,702429	316,447246	525,905042	213,008481	4011,017367
Conca de Barberà	3268,401467	1458,424331	1537,948417	1382,907147	1054,191335	806,820929	370,720721	616,102364	249,541301	4698,941987
Garraf	21300,268015	9504,594048	10022,854846	9012,446349	6870,195784	5258,075607	2415,997792	4015,157141	1626,267963	30623,142455
Garrigues	3165,972931	1412,718725	1489,750604	1339,568176	1021,154000	781,535943	359,102693	596,794313	241,720919	4551,681697
Garrotxa	8388,770120	3743,232456	3947,341188	3549,407951	2705,716805	2070,809039	951,502116	1581,305467	640,479645	12060,435213
Gironès	24835,322282	11082,004047	11686,276910	10508,178089	8010,393405	6130,721087	2816,963793	4681,524267	1896,168112	35705,448008
Maresme	66088,626784	29490,031219	31098,045938	27963,037966	21316,248451	16314,301593	7496,148699	12457,881824	5045,843385	95014,834143
Montsià	9950,805294	4440,242945	4682,357846	4210,327250	3209,536168	2456,405082	1128,677048	1875,753249	759,740481	14306,154639
Noguera	5885,196606	2626,089240	2769,283057	2490,110389	1898,213341	1452,789641	667,532540	1109,375204	449,332689	8461,077294
Osona	22946,743243	10239,283331	10797,604826	9709,093436	7401,250471	5664,516091	2602,750395	4325,522098	1751,975767	32990,260343
Pallars Jussà	2195,653018	979,743100	1033,165943	929,012021	708,186681	542,007714	249,043479	413,886430	167,637335	3156,664279
Pallars Sobirà	1141,316311	509,277546	537,047126	482,907164	368,120556	281,739528	129,454601	215,141158	87,139099	1640,856911
Pla de l'Estany	4542,620920	2027,005844	2137,533202	1922,047520	1465,178516	1121,368249	515,249954	856,296118	346,827508	6530,872168
Pla d'Urgell	5234,860054	2335,896416	2463,266773	2214,943744	1688,453565	1292,250874	593,767662	986,785035	399,679722	7526,096154
Priorat	1570,288713	700,693378	738,900366	664,411489	506,481462	387,633469	178,111095	296,003597	119,890990	2257,585441
Ribera d'Ebre	3585,633647	1599,979503	1687,222224	1517,132594	1156,511512	885,131248	406,703003	675,901474	273,761866	5155,022927
Ripollès	4348,345308	1940,316286	2046,116688	1839,846746	1402,516794	1073,410362	493,214107	819,674649	331,994633	6251,564427
Segarra	3008,732430	1342,554954	1415,760985	1273,037484	970,437595	742,720354	341,267579	567,154061	229,715662	4325,618896
Segrià	29575,181623	13197,021512	13916,620778	12513,679990	9539,189269	7300,778608	3354,585653	5575,000352	2258,054703	42519,887511
Selva	23430,104350	10454,968463	11025,050707	9913,610395	7557,153929	5783,836150	2657,575967	4416,636951	1788,880217	33685,182870
Solsonès	2060,421955	919,400368	969,532879	871,793835	664,569207	508,625263	233,704801	388,395016	157,312491	2962,244185
Tarragonès	34965,165885	15602,137368	16452,881348	14794,258993	11277,676650	8631,322653	3965,948386	6591,026713	2669,578104	50269,003898
Terra Alta	1992,488980	889,087353	937,566974	843,050427	642,658082	491,855675	225,999456	375,589469	152,125832	2864,577753
Urgell	5305,967468	2367,625927	2496,726413	2245,030302	1711,388575	1309,804088	601,833070	1000,188972	405,108748	7628,326438
Val d'Aran	1416,011021	631,851670	666,304899	599,134403	456,720683	349,549264	160,612040	266,921842	108,111943	2035,782235
Vallès Occidental	135262,172512	60356,613298	63647,702473	57231,349014	43627,507721	33390,130554	15342,206485	25497,279069	10327,219245	194464,819630
Vallès Oriental	60957,888299	27200,595874	28683,773636	25792,149539	19661,378292	15047,753639	6914,191099	11490,723981	4654,113308	87638,432334

Tabla 3.9. - Contribuciones al total del estadístico  $\chi^2$ .

	CiU	ERC	PSC	P.P.	ICV-EUiA	C's	CUP-A.E.	Otros	Nulos+Blancos	Abstención
Alt Camp	554,176244	370,993430	247,312560	176,498460	429,725911	293,709396	98,100817	9,895969	0,051704	1,134514
Alt Empordà	1599,028440	224,453247	854,109548	0,004673	1429,460565	1004,744456	71,572700	55,071976	32,278336	59,643423
Alt Penedès	1538,430775	767,751135	241,460755	681,599191	95,294214	790,496926	914,056356	0,307243	0,802122	299,953153
Alt Urgell	837,234772	43,590871	197,438778	106,645281	180,749224	396,951667	19,347751	130,478451	9,916951	6,914549
Alt a Ribagorça	18,566819	14,872885	1,688540	47,945847	25,079806	62,099290	2,012820	8,348347	20,560260	6,305263
Anoia	319,031073	86,403057	68,635020	170,855099	320,732148	355,897163	6,213048	264,283717	19,887922	3,390534
Bages	2981,850776	1193,895489	493,208210	1536,951628	840,359368	1645,585408	181,992463	0,110101	0,516820	16,136305
Baix Camp	112,135418	115,635673	318,270602	348,785368	1534,514908	2,524526	181,954414	216,223272	2,711261	309,863535
Baix Ebre	2,498175	997,387180	387,141457	236,127588	383,805207	1375,786221	27,234096	100,132395	523,309825	632,355101
Baix Empordà	2528,008635	671,710930	270,018873	523,840053	877,618443	1400,705273	1,087804	140,001629	22,127650	0,418227
Baix Llobregat	16032,093632	2855,434011	8140,144635	395,339536	3797,393634	5078,344091	1100,899112	960,813787	81,062680	105,487556
Baix Penedès	599,956267	426,291299	20,965321	63,895175	291,376881	5,681843	60,507856	0,112082	14,780181	1121,733971
Barcelonès	8619,690585	2722,766909	329,800560	5115,242471	6078,976753	2104,877333	1,542497	12,988552	438,439596	51,158386
Berguedà	3118,816307	1358,157182	648,028465	909,738733	589,972984	999,003773	336,347029	55,931547	0,516151	192,018685
Cerdanya	1465,588467	33,786694	394,656193	40,425496	267,754976	237,815418	25,851716	18,978745	0,676985	9,096853
Conca de Barberà	887,971686	1608,396117	458,736675	383,141280	1005,044538	418,126178	6,550611	44,243931	5,622933	62,273217
Garraf	244,110648	129,303867	0,477015	9,619880	14,644876	13,448986	1024,146610	35,052895	2,016654	208,172810
Garrigues	1502,501500	698,375178	285,134202	243,024618	392,579364	533,202161	0,103711	118,376827	9,642813	44,426136
Garrotxa	5778,277888	1440,093447	1017,744200	1174,096210	759,703999	1317,587982	202,080890	132,250406	11,408278	347,921675
Gironès	4531,611937	1794,138030	971,398075	906,976306	1005,044538	1279,464275	1236,114973	57,210291	21,983609	474,812082
Maresme	2254,480905	107,194630	829,854930	14,695459	220,550129	135,408336	34,446368	108,967189	54,383751	236,647164
Montsià	0,108151	767,259619	162,526709	58,981175	695,007056	1375,885952	6,810863	116,144592	913,892894	476,474194
Noguera	2484,449267	557,438805	498,789844	160,968271	558,040594	826,516726	28,335996	185,283640	19,525767	42,985333
Osona	16325,888174	5317,311565	3239,843540	3079,591933	2057,430538	2709,310394	54,679567	176,790590	100,195846	2120,651682
Pallars Jussà	739,625166	227,637816	138,418694	229,765711	187,283865	368,658769	1,935762	110,531300	1,407869	2,095725
Pallars Sobirà	476,797904	307,487071	106,403192	176,451705	68,780053	174,517288	114,119422	72,790858	10,929597	37,742330
Pla de l'Estany	3935,875338	1522,949957	1029,631147	631,882784	472,653042	758,683142	403,121054	82,814369	24,312637	465,114508
Pla d'Urgell	2652,242610	516,217874	470,249581	255,274833	531,651137	685,589173	23,756359	106,240716	150,575662	110,053737
Priorat	590,218228	799,155262	135,912562	197,681842	70,887539	283,723586	603,618398	58,867357	0,006622	220,523576
Ribera d'Ebre	417,394910	992,294996	83,891239	214,253636	192,235964	518,009875	5,832223	34,589155	81,355452	21,903182
Ripollès	2628,316138	502,762939	295,909605	424,592468	261,423313	545,786628	9,874152	38,080880	7,836113	261,900064
Segarra	1256,401648	323,910591	430,572478	98,459426	328,294988	383,532530	6,128488	31,261354	40,357169	33,315581
Segrià	604,087367	31,619878	216,958270	66,659323	1303,466197	912,288754	310,498876	341,596565	279,155830	315,923920
Selva	1046,571119	214,071460	280,976348	265,253542	467,765885	812,525364	55,362678	140,106081	6,872692	152,005746
Solsonès	939,850912	764,900001	326,387322	211,888101	239,038179	362,895593	24,258667	45,130446	0,018102	42,362795
Tarragonès	1578,074836	972,436965	53,947241	2378,324686	692,537818	1876,656498	270,078694	92,787635	3,130818	486,442630
Terra Alta	299,511456	289,015957	118,676243	7,023586	164,010806	312,186842	2,340824	79,307668	134,185152	31,329863
Urgell	2260,208791	436,308896	618,557535	236,089517	523,347735	830,231311	10,152431	68,206790	23,173814	24,615072
Val d'Aran	10,002481	216,491092	37,796713	162,334445	107,637037	37,538440	68,137351	22,747533	163,342136	80,260059
Vallès Occidental	2861,650503	1436,872642	4417,689736	219,750042	1495,891948	3181,525679	71,205964	377,564882	0,654581	108,613775
Vallès Oriental	0,355030	0,086136	2,174988	211,236914	169,886417	164,692463	62,465751	163,644860	43,875507	54,199113

Se tiene que el valor del estadístico ji-cuadrado es 273640,8844, y por tanto la inercia total es 0,0520493825.

Veamos la inercia relativa de las comarcas (inercia de las filas) y de los partidos (inercia de las columnas).

Es interesante considerar las inercias relativas de cada comarca, es decir, su peso en la inercia total. Es obvio que las comarcas con perfil más parecido al perfil global de Cataluña contribuirán poco a la inercia, y en cambio, las comarcas con perfil muy diferente al global de Cataluña contribuirán mucho. El peso de la comarca, en número de censados, también influye positivamente.

**Tabla 3.10.** - Inercias relativas de las comarcas (Inercias de las filas).

COMARCA	INERCIA
Alt Camp	0,000415
Alt Empordà	0,001014
Alt Penedès	0,001014
Alt Urgell	0,000367
Alta Ribagorça	0,000039
Anoia	0,000307
Bages	0,001691
Baix Camp	0,000598
Baix Ebre	0,000887
Baix Empordà	0,001224
Baix Llobregat	0,007332
Baix Penedès	0,000496
Barcelonès	0,004846
Berguedà	0,001561
Cerdanya	0,000475
Conca de Barberà	0,000776
Garraf	0,000320
Garrigues	0,000728
Garrotxa	0,002317
Gironès	0,002336
Maresme	0,000760
Montsià	0,000870
Noguera	0,001020
Osona	0,006692
Pallars Jussà	0,000382
Pallars Sobirà	0,000294
Pla de l'Estany	0,001774
Pla d'Urgell	0,001047
Priorat	0,000563
Ribera d'Ebre	0,000487
Ripollès	0,000947
Segarra	0,000558
Segrià	0,000834
Selva	0,000655
Solsonès	0,000562
Tarragonès	0,001599
Terra Alta	0,000273
Urgell	0,000957
Val d'Aran	0,000172
Vallès Occidental	0,002696
Vallès Oriental	0,000166

**Tabla 3.11.** - Inercias relativas de los partidos (Inercias de las columnas).

<b>PARTIDO</b>	<b>INERCIA</b>
<b>CiU</b>	0,018381
<b>ERC-Cat Sí</b>	0,006440
<b>PSC</b>	0,005486
<b>P.P.</b>	0,004255
<b>ICV-EUiA</b>	0,005769
<b>C's</b>	0,006956
<b>CUP-Alternativa d'Esquerres</b>	0,001458
<b>Otros</b>	0,000916
<b>Votos nulos + Votos en blanco</b>	0,000623
<b>Abstención</b>	0,001765

**Tabla 3.12.** - Diagrama de descomposición (Scree plot) de la inercia en ejes principales.

Principal inertias (eigenvalues):				
dim	value	%	cum %	scree plot
1	0,043423	83,4	83,4	*****
2	0,003678	7,1	90,5	**
3	0,001897	3,6	94,1	*
4	0,001213	2,3	96,5	*
5	0,000835	1,6	98,1	
6	0,000382	0,7	98,8	
7	0,000327	0,6	99,4	
8	0,000172	0,3	99,8	
9	0,000124	0,2	100,0	
Total:	0,052049	100,0		

Con estos datos, se obtiene un plano bidimensional que recoge el 90,5% de la inercia total, es decir la inercia de la nube de puntos proyectada es del 90,5% de la inercia de la nube de puntos iniciales. En particular el primer eje recoge un 83,4% de la inercia y el segundo, por sí solo, un 7,1%. Por tanto podemos hablar de una buena calidad de la representación gráfica bidimensional de la nube de puntos de las comarcas.

**Tabla 3.13.** - Coordenadas principales de las comarcas en el plano bidimensional.

COMARCA	EJE 1	EJE 2
Alt Camp	-0,252146	-0,029046
Alt Empordà	-0,227103	-0,095257
Alt Penedès	-0,240243	0,074046
Alt Urgell	-0,330403	-0,069968
Alta Ribagorça	-0,188382	-0,042228
Anoia	-0,109811	-0,006291
Bages	-0,249555	0,019900
Baix Camp	-0,049687	-0,123676
Baix Ebre	-0,177299	-0,134385
Baix Empordà	-0,263508	-0,025073
Baix Llobregat	0,249309	0,035901
Baix Penedès	0,100536	-0,139541
Barcelonès	0,117010	-0,002637
Berguedà	-0,499295	0,060618
Cerdanya	-0,402642	-0,055100
Conca de Barberà	-0,483677	0,041268
Garraf	0,037121	-0,028129
Garrigues	-0,498462	-0,014491
Garrotxa	-0,548035	0,066416
Gironès	-0,312413	0,043065
Maresme	-0,096972	0,011200
Montsià	-0,172926	-0,169501
Noguera	-0,427062	-0,046665
Osona	-0,549453	0,123279
Pallars Jussà	-0,425730	-0,041148
Pallars Sobirà	-0,509722	0,057161
Pla de l'Estany	-0,642668	0,103751
Pla d'Urgell	-0,456077	-0,022110
Priorat	-0,548101	0,147532
Ribera d'Ebre	-0,349596	-0,003387
Ripollès	-0,480223	0,072979
Segarra	-0,448580	-0,037889
Segrià	-0,111771	-0,124218
Selva	-0,158051	-0,049960
Solsonès	-0,542609	0,017326
Tarragonès	0,155842	-0,126361
Terra Alta	-0,337510	-0,101805
Urgell	-0,444048	-0,037074
Val d'Aran	0,101787	-0,277471
Vallès Occidental	0,127262	0,061461
Vallès Oriental	0,009793	0,038668

**Tabla 3.14.** - Coordenadas principales de los partidos en el plano bidimensional.

<b>PARTIDOS</b>	<b>EJE 1</b>	<b>EJE 2</b>
CiU	-0,291085	0,021027
ERC-Cat Sí	-0,249920	0,029718
PSC	0,218858	0,045766
P.P.	0,179086	-0,088620
ICV-EUiA	0,254571	0,112261
C's	0,348947	0,055837
CUP-Alternativa d'Esquerres	-0,132251	0,050186
Otros	0,080548	0,091263
Nulos+Blancos	-0,022106	-0,084602
Abstención	0,039718	-0,058953

**Tabla 3.15.** - Coordenadas estandares de las comarcas en el plano bidimensional.

COMARCA	EJE 1	EJE 2
Alt Camp	-1,210023	-0,478957
Alt Empordà	-1,089844	-1,570753
Alt Penedès	-1,152900	1,220997
Alt Urgell	-1,585569	-1,153753
Alta Ribagorça	-0,904026	-0,696329
Anoia	-0,526973	-0,103744
Bages	-1,197588	0,328145
Baix Camp	-0,238443	-2,039374
Baix Ebre	-0,850838	-2,215963
Baix Empordà	-1,264547	-0,413447
Baix Llobregat	1,196410	0,591993
Baix Penedès	0,482460	-2,300990
Barcelonès	0,561517	-0,043489
Berguedà	-2,396068	0,999578
Cerdanya	-1,932240	-0,908581
Conca de Barberà	-2,321118	0,680491
Garraf	0,178141	-0,463838
Garrigues	-2,392068	-0,238949
Garrotxa	-2,629966	1,095176
Gironès	-1,499238	0,710131
Maresme	-0,465357	0,184688
Montsià	-0,829856	-2,795016
Noguera	-2,049425	-0,769493
Osona	-2,636767	2,032840
Pallars Jussà	-2,043034	-0,678517
Pallars Sobirà	-2,446102	0,942571
Pla de l'Estany	-3,084100	1,710822
Pla d'Urgell	-2,188668	-0,364581
Priorat	-2,630279	2,432752
Ribera d'Ebre	-1,677678	-0,055854
Ripollès	-2,304540	1,203404
Segarra	-2,152690	-0,624772
Segrià	-0,536378	-2,048314
Selva	-0,758471	-0,823821
Solsonès	-2,603926	0,285704
Tarragonès	0,747869	-2,083652
Terra Alta	-1,619674	-1,678739
Urgell	-2,130942	-0,611334
Val d'Aran	0,488463	-4,575409
Vallès Occidental	0,610718	1,013476
Vallès Oriental	0,046996	0,637617

**Tabla 3.16.** - Coordenadas estandares de los partidos en el plano bidimensional.

<b>PARTIDOS</b>	<b>EJE 1</b>	<b>EJE 2</b>
CiU	-1,396886	0,346734
ERC-Cat Sí	-1,199340	0,490037
PSC	1,050277	0,754666
P.P.	0,859418	-1,461317
ICV-EUiA	1,221659	1,851154
C's	1,674562	0,920737
CUP-Alternativa d'Esquerres	-0,634660	0,827554
Otros	0,386540	1,504902
Nulos+Blancos	-0,106086	-1,395064
Abstención	0,190604	-0,972112

Para interpretar las gráficas utilizaremos la siguiente tabla de códigos para las comarcas. Esto nos proporcionará una mejor visualización de las etiquetas de los puntos en las representaciones bidimensionales.

**Tabla 3.17.** - Tabla de los códigos de las comarcas.

COMARCAS	CÓDIGO
Alt Camp	AC
Alt Empordà	AE
Alt Penedès	AP
Alt Urgell	AU
Alta Ribagorça	AR
Anoia	A
Bages	BA
Baix Camp	BC
Baix Ebre	BEB
Baix Empordà	BEM
Baix Llobregat	BL
Baix Penedès	BP
Barcelonès	B
Berguedà	BE
Cerdanya	C
Conca de Barberà	CB
Garraf	GF
Garrigues	GG
Garrotxa	GR
Gironès	GI
Maresme	M
Montsià	MO
Noguera	N
Osona	O
Pallars Jussà	PJ
Pallars Sobirà	PS
Pla de l'Estany	PE
Pla d'Urgell	PU
Priorat	P
Ribera d'Ebre	RE
Ripollès	R
Segarra	SE
Segrià	S
Selva	SL
Solsonès	SO
Tarragonès	T
Terra Alta	TA
Urgell	U
Val d'Aran	VA
Vallès Occidental	VOC
Vallès Oriental	VOR

Para los partidos utilizaremos la siguiente tabla de codificación.

**Tabla 3.18.** - Tabla de los códigos de los partidos.

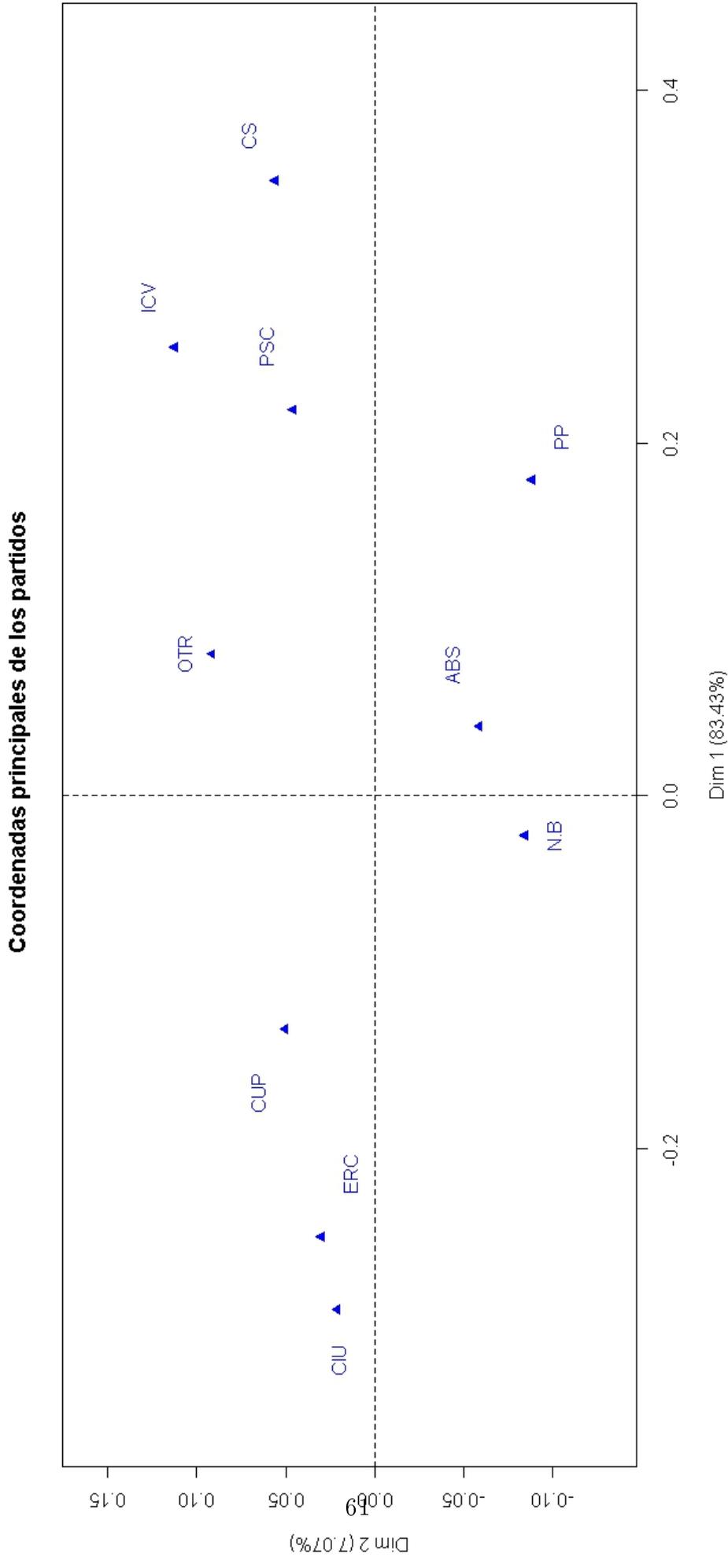
PARTIDOS	CÓDIGO
CiU	CIU
ERC-Cat Sí	ERC
PSC	PSC
P.P.	PP
ICV-EUiA	ICV
C's	CS
CUP-Alternativa d'Esquerres	CUP
Otros	OTR
Nulos+Blancos	N+B
Abstención	ABS

Presentamos un mapa geográfico de las comarcas de Cataluña, para poder localizar la ubicación correcta de las comarcas.

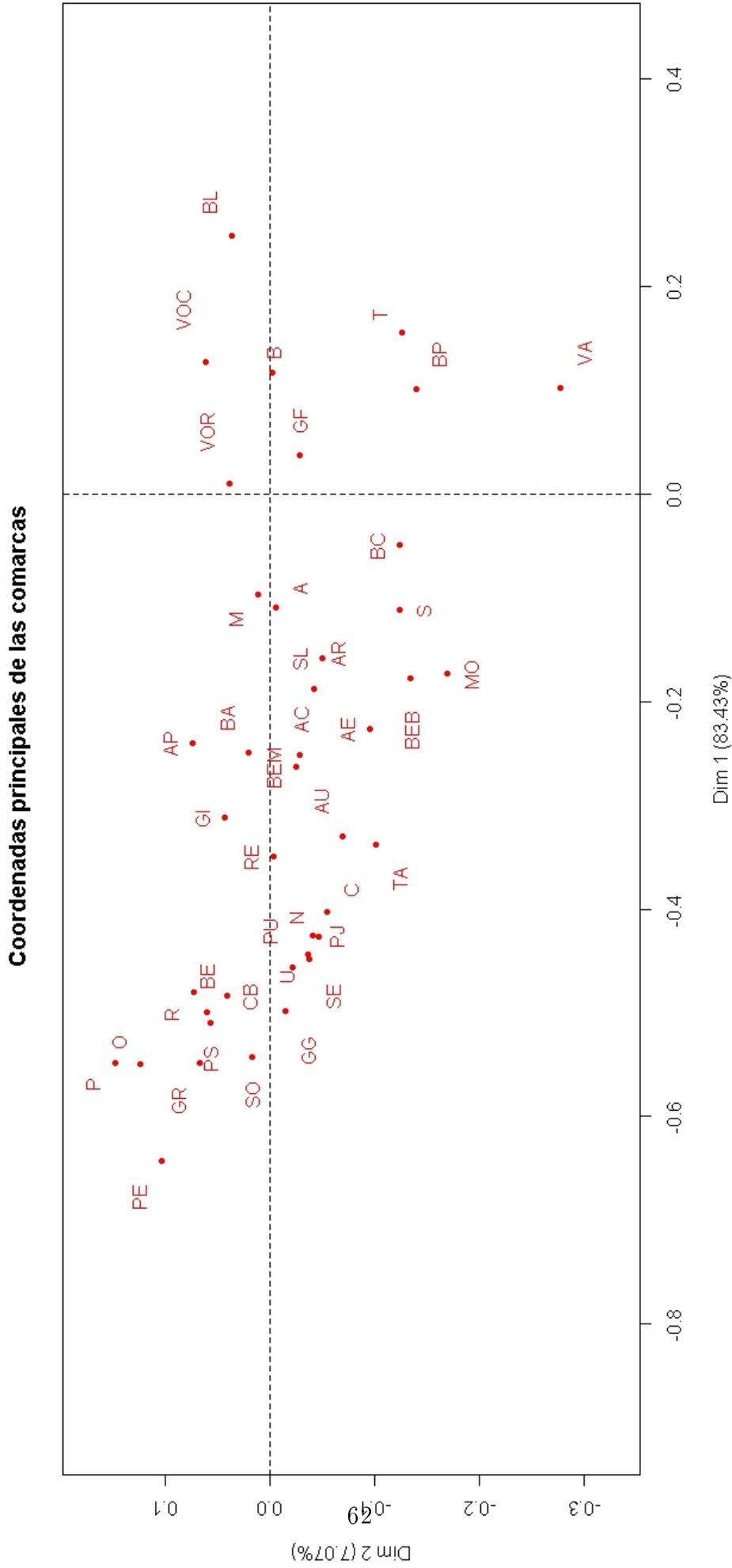


Las gráficas que representamos son vistas desde el punto de vista del análisis de filas, es decir, el análisis electoral de las comarcas, que es lo que nos interesa.

**Gráfico 3.1.** La siguiente gráfica es la representación bidimensional en la que representamos los perfiles de los partidos como coordenadas principales. Esta gráfica, sirve básicamente, para interpretar los ejes principales.

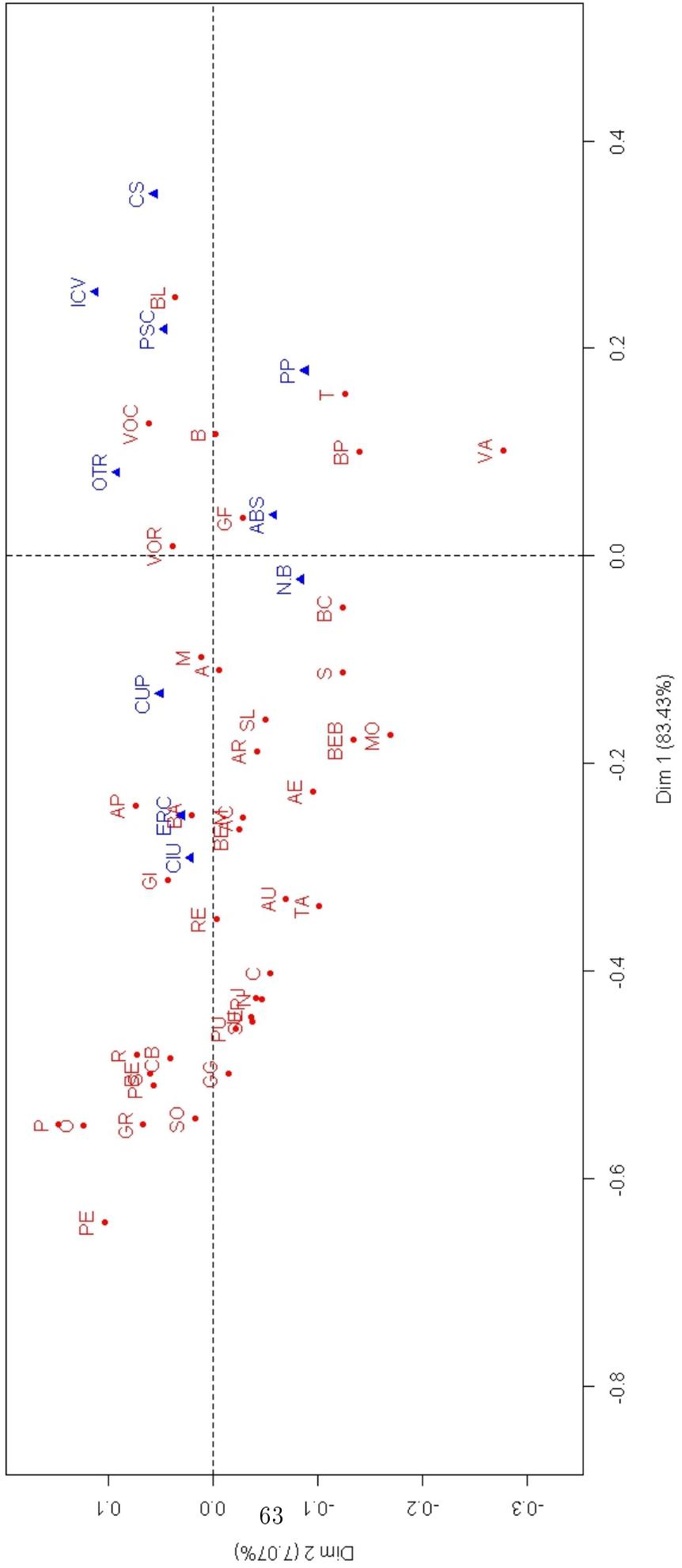


**Gráfica 3.2.** La siguiente gráfica es la representación bidimensional en la que representamos los perfiles de las comarcas como coordenadas principales. Esta gráfica permite analizar en detalle el comportamiento de cada comarca.

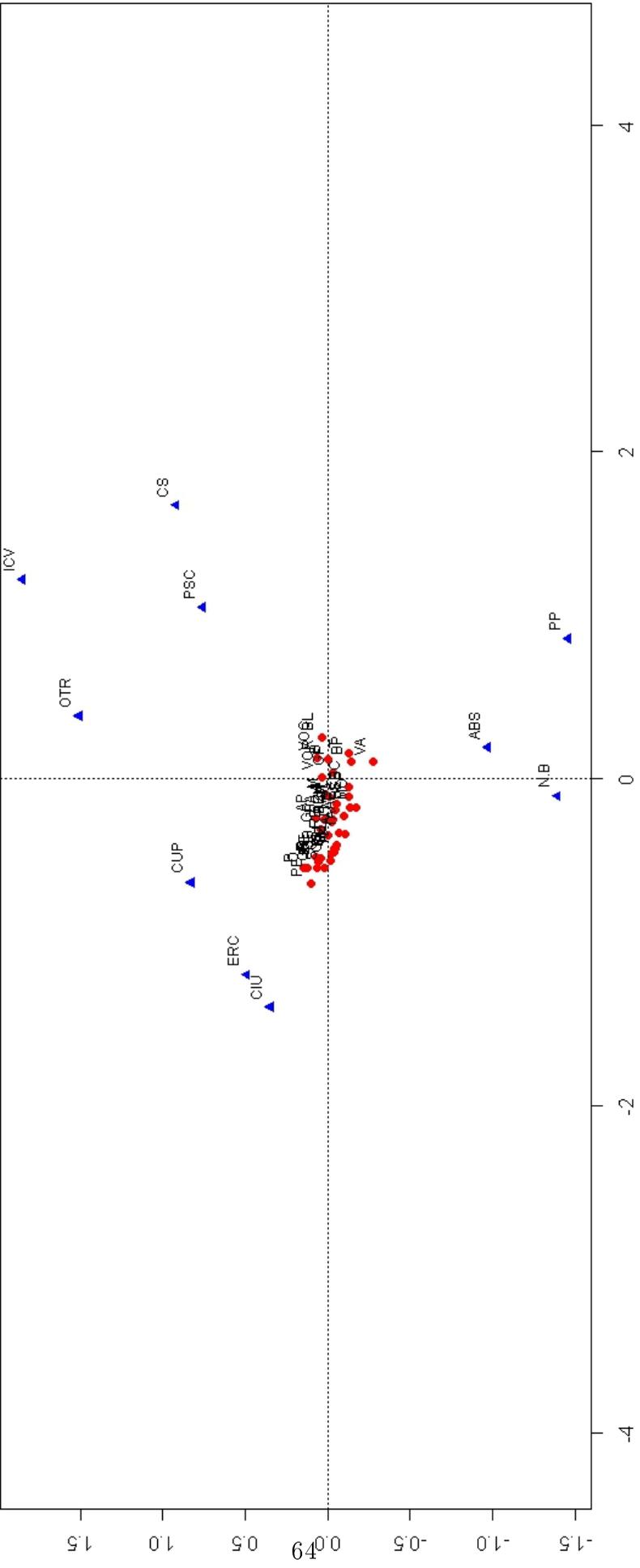


**Gráfico 3.3.** La siguiente gráfica es la representación bidimensional en la que representamos los perfiles conjuntos de las comarcas y de los partidos, ambos con coordenadas principales. Esta gráfica es el solapamiento de las dos anteriores representaciones, de la gráfica 3.1. y la gráfica 3.2.

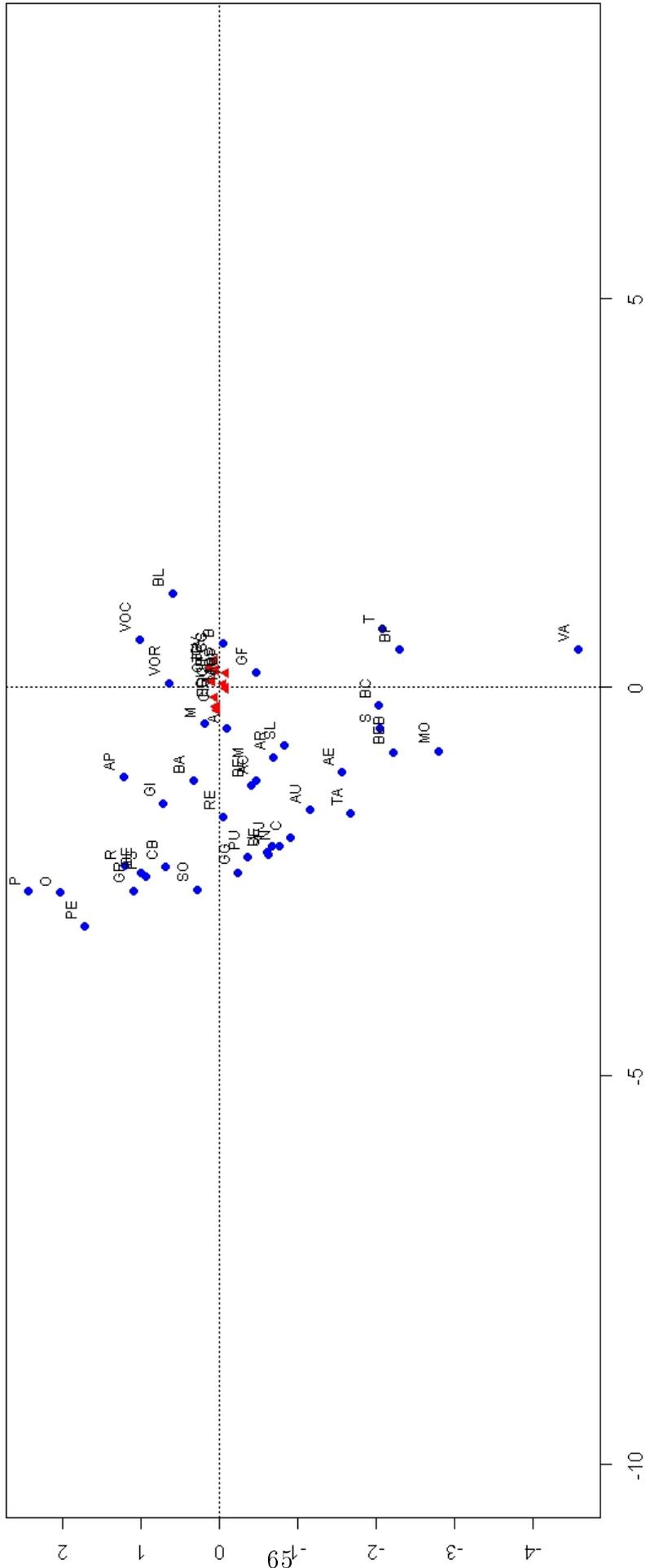
**Mapa conjunto de comarcas y partidos**



**Gráfico 3.4.** La siguiente gráfica es la representación bidimensional en la que representamos los perfiles conjuntos de las comarcas y de los partidos, en coordenadas principales y en coordenadas estándares respectivamente.



**Gráfica 3.5.** La siguiente gráfica es la representación bidimensional, la única vista desde el punto del análisis de columnas, en la que representamos los perfiles conjuntos de los partidos y de las comarcas, en coordenadas principales y en coordenadas estándares respectivamente.



**Tabla 3.19.** - Calidad de la representación de las comarcas.

COMARCA	CALIDAD	MASAS
Alt Camp	0,946358	0,006096
Alt Empordà	0,950461	0,015889
Alt Penedès	0,895057	0,014359
Alt Urgell	0,925431	0,002977
Alta Ribagorça	0,538550	0,000570
Anoia	0,648558	0,016471
Bages	0,953110	0,025717
Baix Camp	0,716996	0,024126
Baix Ebre	0,581238	0,010422
Baix Empordà	0,945488	0,016519
Baix Llobregat	0,952265	0,110051
Baix Penedès	0,794513	0,013311
Barcelonès	0,841564	0,297701
Berguedà	0,986730	0,006090
Cerdanya	0,872779	0,002508
Conca de Barberà	0,891531	0,002938
Garraf	0,129882	0,019145
Garrigues	0,971992	0,002846
Garrotxa	0,991706	0,007540
Gironès	0,950546	0,022322
Maresme	0,744565	0,059400
Montsià	0,602869	0,008944
Noguera	0,957122	0,005290
Osona	0,977281	0,020624
Pallars Jussà	0,945516	0,001973
Pallars Sobirà	0,917718	0,001026
Pla de l'Estany	0,975292	0,004083
Pla d'Urgell	0,937382	0,004705
Priorat	0,807464	0,001411
Ribera d'Ebre	0,808401	0,003223
Ripollès	0,974153	0,003908
Segarra	0,982600	0,002704
Segrià	0,890458	0,026582
Selva	0,883904	0,021059
Solsonès	0,970478	0,001852
Tarragonès	0,791332	0,031426
Terra Alta	0,813911	0,001791
Urgell	0,989512	0,004769
Val d'Aran	0,644897	0,001273
Vallès Occidental	0,900810	0,121573
Vallès Oriental	0,525200	0,054788

Tabla 3.20. - Contribución absoluta de las comarcas.

CONTRIBUCIÓN ABSOLUTA		
COMARCA	EJE 1	EJE 2
Alt Camp	0,000388	0,000005
Alt Empordà	0,000819	0,000144
Alt Penedès	0,000829	0,000079
Alt Urgell	0,000325	0,000015
Alta Ribagorça	0,000020	0,000001
Anoia	0,000199	0,000001
Bages	0,001602	0,000010
Baix Camp	0,000060	0,000369
Baix Ebre	0,000328	0,000188
Baix Empordà	0,001147	0,000010
Baix Llobregat	0,006840	0,000142
Baix Penedès	0,000135	0,000259
Barcelonès	0,004076	0,000002
Berguedà	0,001518	0,000022
Cerdanya	0,000407	0,000008
Conca de Barberà	0,000687	0,000005
Garraf	0,000026	0,000015
Garrigues	0,000707	0,000001
Garrotxa	0,002265	0,000033
Gironès	0,002179	0,000041
Maresme	0,000559	0,000007
Montsià	0,000267	0,000257
Noguera	0,000965	0,000012
Osona	0,006226	0,000313
Pallars Jussà	0,000358	0,000003
Pallars Sobirà	0,000267	0,000003
Pla de l'Estany	0,001686	0,000044
Pla d'Urgell	0,000979	0,000002
Priorat	0,000424	0,000031
Ribera d'Ebre	0,000394	0,000000
Ripollès	0,000901	0,000021
Segarra	0,000544	0,000004
Segrià	0,000332	0,000410
Selva	0,000526	0,000053
Solsonès	0,000545	0,000001
Tarragonès	0,000763	0,000502
Terra Alta	0,000204	0,000019
Urgell	0,000940	0,000007
Val d'Aran	0,000013	0,000098
Vallès Occidental	0,001969	0,000459
Vallès Oriental	0,000005	0,000082

Tabla 3.21. - Contribución relativa de las comarcas.

CONTRIBUCIÓN RELATIVA		
COMARCA	EJE 1	EJE 2
Alt Camp	0,933964	0,012393
Alt Empordà	0,808262	0,142199
Alt Penedès	0,817407	0,077650
Alt Urgell	0,885711	0,039720
Alta Ribagorça	0,512783	0,025767
Anoia	0,646436	0,002122
Bages	0,947088	0,006022
Baix Camp	0,099644	0,617352
Baix Ebre	0,369158	0,212081
Baix Empordà	0,937004	0,008483
Baix Llobregat	0,932920	0,019345
Baix Penedès	0,271491	0,523022
Barcelonès	0,841136	0,000427
Berguedà	0,972397	0,014333
Cerdanya	0,856736	0,016044
Conca de Barberà	0,885088	0,006443
Garraf	0,082506	0,047375
Garrigues	0,971171	0,000821
Garrotxa	0,977352	0,014354
Gironès	0,932821	0,017725
Maresme	0,734763	0,009802
Montsià	0,307465	0,295404
Noguera	0,945828	0,011293
Osona	0,930442	0,046839
Pallars Jussà	0,936765	0,008751
Pallars Sobirà	0,906320	0,011398
Pla de l'Estany	0,950519	0,024773
Pla d'Urgell	0,935184	0,002198
Priorat	0,752914	0,054550
Ribera d'Ebre	0,808325	0,000076
Ripollès	0,952163	0,021990
Segarra	0,975640	0,006960
Segrià	0,398394	0,492064
Selva	0,803608	0,080295
Solsonès	0,969489	0,000988
Tarragonès	0,477442	0,313890
Terra Alta	0,746033	0,067878
Urgell	0,982662	0,006850
Val d'Aran	0,076490	0,568407
Vallès Occidental	0,730442	0,170368
Vallès Oriental	0,031657	0,493543

Tabla 3.22. - Contribución relativa de las comarcas en ‰.

CONTRIBUCIÓN RELATIVA EN ‰		
COMARCA	EJE 1	EJE 2
Alt Camp	934	12
Alt Empordà	808	142
Alt Penedès	817	78
Alt Urgell	886	40
Alta Ribagorça	513	26
Anoia	646	2
Bages	947	6
Baix Camp	100	617
Baix Ebre	369	212
Baix Empordà	937	8
Baix Llobregat	933	19
Baix Penedès	271	523
Barcelonès	841	0
Berguedà	972	14
Cerdanya	857	16
Conca de Barberà	885	6
Garraf	83	47
Garrigues	971	1
Garrotxa	977	14
Gironès	933	18
Maresme	735	10
Montsià	307	295
Noguera	946	11
Osona	930	47
Pallars Jussà	937	9
Pallars Sobirà	906	11
Pla de l'Estany	951	25
Pla d'Urgell	935	2
Priorat	753	55
Ribera d'Ebre	808	0
Ripollès	952	22
Segarra	976	7
Segrià	398	492
Selva	804	80
Solsonès	969	1
Tarragonès	477	314
Terra Alta	746	68
Urgell	983	7
Val d'Aran	76	568
Vallès Occidental	730	170
Vallès Oriental	32	494

**Tabla 3.23.** - Calidad de la representación de los partidos.

PARTIDOS	CALIDAD	MASAS
CiU	0,980645	0,211629
ERC-Cat Sí	0,928843	0,094433
PSC	0,907488	0,099582
P.P.	0,840128	0,089543
ICV-EUiA	0,915901	0,068259
C's	0,937845	0,052242
CUP-Alternativa d'Esquerres	0,329438	0,024004
Otros	0,645485	0,039893
Nulos+Blancos	0,198176	0,016158
Abstención	0,871124	0,304257

**Tabla 3.24.** - Contribución absoluta de los partidos.

CONTRIBUCIÓN ABSOLUTA		
PARTIDOS	EJE 1	EJE 2
CiU	0,017931	0,000094
ERC-Cat Sí	0,005898	0,000083
PSC	0,004770	0,000209
P.P.	0,002872	0,000703
ICV-EUiA	0,004424	0,000860
C's	0,006361	0,000163
CUP-Alternativa d'Esquerres	0,000420	0,000060
Otros	0,000259	0,000332
Nulos+Blancos	0,000008	0,000116
Abstención	0,000480	1,057420

**Tabla 3.25.** - Contribución relativa de los partidos.

CONTRIBUCIÓN RELATIVA		
PARTIDOS	EJE 1	EJE 2
CiU	0,975554	0,005091
ERC-Cat Sí	0,915893	0,012950
PSC	0,869468	0,038020
P.P.	0,674871	0,165257
ICV-EUiA	0,766787	0,149114
C's	0,914431	0,023414
CUP-Alternativa d'Esquerres	0,287970	0,041468
Otros	0,282642	0,362844
Nulos+Blancos	0,012666	0,185510
Abstención	0,271966	0,599158

**Tabla 3.26.** - Contribución relativa de los partidos en ‰.

CONTRIBUCIÓN RELATIVA EN ‰		
PARTIDOS	EJE 1	EJE 2
CiU	976	5
ERC-Cat Sí	916	13
PSC	869	38
P.P.	675	165
ICV-EUiA	767	149
C's	914	23
CUP-Alternativa d'Esquerres	288	41
Otros	283	363
Nulos+Blancos	13	186
Abstención	272	599

### 3.3. Conclusiones del estudio de geografía electoral

Un análisis parecido para datos electorales de los años 80 y 90, se encuentra en Puig-Vives.

En la práctica son suficiente dos o tres ejes para estudiar la relación entre filas y columnas. En nuestro caso hemos cogido dos ejes, ya que estos representan más del 90% de la inercia total.

#### *Criterios de interpretación.*

Para poder interpretar bien un gráfico tendremos que tener en cuenta algunos puntos:

- 1.— Si dos filas o columnas, tienen una estructura parecida, su situación será próxima sobre el plano (no siempre será cierto a la inversa, depende de la calidad de representación de los puntos).
- 2.— La situación próxima de un punto fila  $i$  y un punto columna  $j$  solo se puede interpretar si están lejos del origen o centro de gravedad, en la periferia de las nubes.
- 3.— Cuando una fila o columna tiene un perfil próximo al perfil medio, tendrá un comportamiento medio, y se situará cerca del origen.

Para interpretar bien el gráfico de la nube de puntos, debemos conocer bien la significación de los ejes a donde hemos proyectado la nube y estudiar su forma. La visualización directa del gráfico nos puede conducir a errores. Para evitarlo utilizamos dos indicadores:

Contribución Absoluta ( $CA_{eje_k, ele_i}$ )

Expresa la participación que tiene el elemento  $i$  en la inercia explicada por el eje  $k$ .

Contribución Relativa ( $CR_{eje_k, ele_i}$ )

Expresa la calidad de representación del elemento  $i$  sobre el eje  $k$ .

Pautas a seguir:

- 1.— Se buscan aquellos puntos tanto filas como columnas con el  $CA_{eje_k, ele_i}$  más alto.
- 2.— De estos se separan los puntos que se proyectan al lado positivo por una banda y los que se proyectan al lado negativo por la otra.
- 3.— Se estudia la calidad de los puntos a partir de su  $CR_{eje_k, ele_i}$ . Si se obtiene un  $CR_{eje_k, ele_i}$  bajo no se tiene en cuenta, ya que tendrá su importancia en otro eje.
- 4.— Se buscan aquellos puntos, tanto filas como columnas, que estén bien representados, esto quiere decir que tengan un  $CR_{eje_k, ele_i}$  alto. Estos puntos son ilustrativos de la significación del eje.

De esta manera ya tenemos los elementos necesarios para la interpretación de los ejes. Estos nuevos ejes son nuevas dimensiones que condensan y resumen los criterios interrelacionados. A veces se puede dar un significado a estos ejes. Esto pasa cuando hay elementos muy bien representados y opuestos entre sí.

## **Resultados.**

La **gráfica 3.1.** es la representación gráfica de los partidos respecto las dos primeras coordenadas principales. La manera más interesante de interpretar esta gráfica es considerar que estos puntos que representan a los partidos, son los extremos de vectores que salen del origen. El coseno del ángulo entre dos vectores nos indica la proximidad que representan.

En este caso, ERC y CiU tienen perfiles muy parecidos. Pasa lo mismo con PSC y Ciutadans, y por otro lado con P.P. y Abstención. ICV-EUiA y CUP-Alternativa d'Esquerres son variables que no están muy bien explicadas.

La **gráfica 3.2.** es la representación gráfica de las 41 comarcas respecto las dos primeras coordenadas principales.

La **gráfica 3.3.** es la representación gráfica de las 41 comarcas y los 10 partidos, respecto las dos primeras coordenadas principales.

La interpretación si nos fijamos en las tres gráficas. A la izquierda del primer eje, aparecen comarcas con un gran peso de ERC y CiU, que son comarcas más del interior. Por ejemplo las más extremas: Pla d'Estany, Osona, Priorat, Solsonès, Garrigues, Noguera, Berguerà, Garrotxa, Ripollès y Pallars Sobirà.

Al otro extremo, con una importancia relativa de PSC, Ciutadans, Abstención e ICV-EUiA, se encuentran las comarcas más pobladas de Cataluña, urbanas y con mucha inmigración: Baix Llobregat, Vallès Occidental, Barcelonès y Tarragonés.

Cerca del centro se encuentran comarcas con un perfil parecido al perfil de Cataluña, en este caso son Garraf, Vallès Oriental, Baix Penedès, Maresme, Baix Camp y Anoia.

A partir de la gráfica 3.3., podemos imaginar grupos de comarcas según su comportamiento político.

Por ejemplo, Baix Llobregat, Barcelonès, y el Vallés Occidental son comarcas con un peso importante del PSC, ICV-EUiA y Abstención.

Tarragonés y Baix Penedès tienen un peso importante del P.P.

Maresme, Anoia, Garraf, Baix Camp y Vallés Oriental, forman un grupo de comarcas equilibradas, en el sentido que recogen unas proporciones muy parecidas a las globales de Cataluña.

Otro grupo está formado por comarcas con voto Nacionalista (ERC y CiU) y un voto más de derechas: Priorat, Pla d'Estany, Osona, Garrotxa y Ripollès.

Otros presentan altos votos nacionalistas y un voto más bien de izquierdas: Pallars Jussà, Alta Ribagorça, Selva, Alt Camp y Baix Empordà.

Vamos a continuar con el esquema descrito en *criterios de interpretación*.

La inercia de la nube de puntos es 0,0520493825.

En la Tabla 3.12, en el diagrama de descomposición, describimos los porcentajes de las inercias principales, de donde obtenemos lo siguiente.

La inercia del primer eje = 0,043423 = 83,4 % de la inercia total.

La inercia del segundo eje = 0,003678 = 7,1 % de la inercia total.

Sumando los dos ejes tenemos casi la representación exacta de todos los puntos, tanto los puntos de los perfiles fila, como los puntos de los perfiles columna = 90,5 %.

Además vemos que el primer eje coge un 83,4 % de la información de la tabla. Esto nos indica que podemos interpretar en un solo eje, es decir, proyectar estos puntos sobre una recta, sean del espacio  $\mathbb{R}^{10}$ , como  $\mathbb{R}^1$ .

### ***Espacio de los partidos.***

Vamos a estudiar el comportamiento de los partidos. Miramos la Tabla 3.23. - Calidad de la representación de los partidos, columna calidad, para comprobar si están bien representados en el plano bidimensional.

<b>PARTIDOS</b>	<b>CALIDAD</b>
CiU	0,980645
ERC-Cat Sí	0,928843
PSC	0,907488
P.P.	0,840128
ICV-EUiA	0,915901
C's	0,937845
CUP-Alternativa d'Esquerres	0,329438
Otros	0,645485
Nulos+Blancos	0,198176
Abstención	0,871124

Todos los partidos tienen una buena representación en el plano bidimensional, excepto los votos Nulos y Blancos, y el partido CUP-Alternativa d'Esquerres.

Para interpretar los ejes seguiremos las pautas antes definidas.

A.- Cogemos las comarcas con el  $CA_{eje_k, ele_i}$ ,  $k = 1, 2$ ,  $j = 1, \dots, 10$ , (contribución absoluta, es decir la importancia que tiene el punto proyectado en el eje) más alto de cada eje.

En nuestro caso hemos encontrado:

EJE 1		EJE 2	
CiU	0,017931	Abstención	1,057420
C's	0,006361	ICV-EUiA	0,000860
ERC-Cat Sí	0,005898	P.P.	0,000703
PSC	0,004770		

B.- De estos puntos separamos los que se proyectan al lado positivo y al lado negativo:

EJE 1	SIGNO	EJE 2	SIGNO
CiU	(-)	Abstención	(+)
C's	(+)	ICV-EUiA	(+)
ERC-Cat Sí	(-)	P.P.	(-)
PSC	(+)		

C.- Se estudia la calidad de cada partido en cada eje a partir del  $CR_{eje_k, ele_j}$ ,  $k = 1, 2$ ,  $j = 1, \dots, 10$ , (contribución relativa), es decir la inercia proyectada del punto al eje dividido por la inercia del punto en la nube de puntos. Si un punto está bien correlacionado con el eje, el  $CA_{eje_k, ele_j}$  será muy elevado.

EJE 1		EJE 2	
CiU	0,975554	Abstención	0,599158
ERC-Cat Sí	0,915893	Otros	0,362844
C's	0,914431	Nulos+Blancos	0,185510
PSC	0,869468	P.P.	0,165257
		ICV-EUiA	0,149114

De aquí podemos ver que en el primer eje principal (eje de la abcisas), todos los puntos partidos están bien representados. Entonces la interpretación de la nube de puntos de los partidos, viene dada por:

*ERC y CiU contra PSC y Ciudadans.*

El nombre de este eje sería alto voto de ERC y CiU contra alto voto de PSC y Ciudadans.

Dicho de otro modo,

**el primer eje principal opone *soberanismo* a *no-soberanismo*.**

El segundo eje principal (eje de las ordenadas), no es un eje tan importante, pero si a tener en cuenta, pues tiene el 7,1%, e iría más ligado a un alto voto más partidario del P.P. y no P.P., como PSC y Ciutadans, y después ICV-EUiA, es decir, esencialmente opone comarcas con un alto voto relativo de ICV-EUiA a comarcas con un alto voto relativo del P.P. Por tanto, podríamos decir:

**el segundo eje principal va de *derechas* a *izquierdas*.**

***Espacio de las comarcas.***

Primero miramos las comarcas que tienen más baja cualificación (columna Calidad de Tabla 18.- Calidad de la representación), es decir su representación en el plano bidimensional.

COMARCA	CALIDAD
Garraf	0,129882

Entonces, Garraf, estaría mejor representada en otros ejes. A la hora de hacer el estudio no la tendremos en cuenta.

Para interpretar los ejes seguiremos las pautas antes definidas.

A.- Cogemos las comarcas con el  $CA_{eje_k, ele_i}$ ,  $k = 1, 2$ ,  $i = 1, \dots, 41$ , (contribución absoluta, Tabla 3.21. - Contribución relativa de las comarcas ), más alto de cada eje:

EJE 1	
Baix Llobregat	0,006840
Osona	0,006226
Barcelonès	0,004076

EJE 2	
Tarragonès	0,000502
Vallès Occidental	0,000459
Segrià	0,000410
Baix Camp	0,000369
Osona	0,000313

B.- De estos puntos separamos los que se proyectan al lado positivo y al lado negativo:

EJE 1	SIGNO
Baix Llobregat	(+)
Osona	(-)
Barcelonès	(+)

EJE 2	SIGNO
Tarragonès	(-)
Vallès Occidental	(+)
Segrià	(-)
Baix Camp	(-)
Osona	(+)

C.- Se estudia la calidad de cada comarca en cada eje a partir del  $CR_{eje_k, ele_i}$ ,  $k = 1, 2$ ,  $i = 1, \dots, 41$ , (Contribución Relativa):

EJE 1		EJE 2	
Baix Llobregat	0,932920	Tarragonès	0,313890
Osona	0,930442	Vallès Occidental	0,170368
Barcelonès	0,841136	Segrià	0,492064
		Baix Camp	0,617352
		Osona	0,046839

Tendremos claro que *el primer eje principal* estará compuesto por: Baix Llobregat, Barcelonès y Vallès Occidental, por un lado y por el otro lado, como más importantes Osona, Garrotxa y Gironès.

En cambio para *el segundo eje principal* tendremos como más importantes: Vallès Occidental, al lado positivo, y Baix Camp y Baix Penedès, al lado negativo.

Otros puntos importantes de cada eje, es decir están bien representados, son:

EJE 1	CALIDAD	SIGNO	EJE 2	CALIDAD	SIGNO
Garrotxa	0,991706	(-)	Baix Penedès	0,794513	(-)
Gironès	0,950546	(-)	Alt Empordà	0,950461	(-)
Pla de l'Estany	0,975292	(-)			
Bages	0,953110	(-)			
Berguedà	0,986730	(-)			
Baix Empordà	0,945488	(-)			

### *Interpretación conjunta.*

A partir de estos datos, y de las gráficas podemos decir que el primer eje principal, que es el más importante, tiene al lado negativo, un alto voto de CIU y ERC y se agrupan comarcas como Garrigues (característica de este eje), Osona, Priorat, Berguerà, Garrotxa, Pla d'Estany, etc.

Por el otro lado, el positivo, está el PSC como más importante, junto con Ciutadans. En esta parte del eje tenemos como más destacado a Barcelonès y Baix Llobregat.

Observando las comarcas de este eje, podríamos pensar que está formado por comarcas de interior contra comarcas del cinturón de Barcelona.

En el segundo eje principal, no tan importante como el primero, en el lado negativo hay un alto voto del P.P. donde se agrupan comarcas como el Tarragonès y el Baix Penedès, y en el lado positivo hay un alto voto de ICV-EUiA, del PSC y de Ciutadans, con comarcas características como las formadas por el cinturón de Barcelona (Barcelonès, Baix Llobregat y Vallès Occidental).

# Apéndice A

## Implementación del cálculo del AC con R

En este apéndice veremos el cálculo de los resultados del AC utilizando el lenguaje de programación R. Para información sobre el uso del programa R en estadística veáse Peter Dalgaard.

**# Se captura los datos de la tabla  $N$ , y la guardamos en el data frame de R llamado `eleccions`.**

```
eleccions <- read.table("clipboard")
```

**# Calculamos el total de la muestra.**

```
sum(eleccions)
```

**# Calculamos el estadístico ji-cuadrado.**

```
eleccions.row <- apply(eleccions, 1, sum)
```

```
eleccions.col <- apply(eleccions, 2, sum)
```

```
eleccions.exp <- eleccions.row %o %eleccions.col / sum(eleccions)
```

```
eleccions.chi^2 <- sum((eleccions - eleccions.exp)^2 / eleccions.exp)
```

**# Calculamos la inercia total.**

```
eleccions.chi^2 / sum(eleccions)
```

**# CÁLCULO DE LAS COORDENADAS PRINCIPALES Y LAS COORDENADAS ESTÁNDARES.**

**# Obtenemos la matriz  $P$  de correspondencias de la tabla.**

```
eleccions.P <- eleccions / sum(eleccions)
```

**# Calculamos la masa de las filas.**

```
eleccions.r <- apply(eleccions.P, 1, sum)
```

**# Calculamos la masa de las columnas.**

```
eleccions.c <- apply(eleccions.P, 2, sum)
```

```

eleccions.Dr <- diag(eleccions.r)

eleccions.Dc <- diag(eleccions.c)

eleccions.Drmh <- diag(1/sqrt(eleccions.r))

eleccions.Dcmh <- diag(1/sqrt(eleccions.c))

eleccions.P <- as.matrix(eleccions.P)

# Cálculo de la matriz de residuos estandarizados S:
eleccions.S <- eleccions.Drmh %*% (eleccions.P - eleccions.r %o% eleccions.c) %*%
eleccions.Dcmh

# Realizamos los cálculos de la DVS:

eleccions.svd <- svd(eleccions.S)

# Calculamos las coordenadas principales y estándares (pc y sc).
eleccions.rsc <- eleccions.Drmh %*% eleccions.svd$u

eleccions.csc <- eleccions.Dcmh %*% eleccions.svd$v

eleccions.rpc <- eleccions.rsc %*% diag(eleccions.svd$d)

eleccions.cpc <- eleccions.csc %*% diag(eleccions.svd$d)

# Cálculo de las coordenadas principales de los dos primeros ejes.
eleccions.rpc[,1:2]
eleccions.cpc[,1:2]

# Cálculo de las coordenadas estándares de los dos primeros ejes.
eleccions.rsc[,1:2]
eleccions.csc[,1:2]

# CÁLCULO DE LAS GRÁFICAS.

# Cálculo de las gráficas en coordenadas estándares y principales de los dos primeros ejes.
plot(ca(eleccions), what = c("all", "none"), map = "rowprincipal", arrows = c(FALSE,FALSE))

plot(ca(eleccions), map = "rowprincipal", col = c("red", "blue"),arrows = c(FALSE,FALSE))

plot(ca(eleccions), map = "symmetric", col = c("red", "blue"),arrows = c(FALSE,FALSE))

plot(ca(eleccions), map = "rowprincipal", what = c("all", "none"), col = c("red",

```

```
"blue"),arrows = c(FALSE,FALSE))
```

```
plot(ca(eleccions), map = "rowprincipal", what = c("none", "all"), col = c("red",  
"blue"),arrows = c(FALSE,FALSE))
```

```
# CÁLCULO DE LAS CONTRIBUCIONES A LA INERCIA.
```

```
# Cálculo de las inercias principales.
```

```
eleccions.svd$d^2
```

```
# Cálculo de las inercias relativas de las comarcas.
```

```
round(apply(eleccions.S^2, 1, sum), 5)
```

```
# Cálculo de las inercias relativas de los partidos.
```

```
round(apply(eleccions.S^2, 2, sum), 5)
```

```
# Cálculo de las inercias relativas de las comarcas, en tantos por mil.
```

```
round(1000*eleccions.S^2/sum(eleccions.S^2), 0)
```

```
# Cálculo de las inercias principales de la tabla.
```

```
eleccions.svd <- svd(eleccions.S)
```

```
eleccions.svd$d^2
```

```
# (Contribución absoluta) Componentes individuales de la inercia de las co-  
marcas (filas) en los 10 ejes.
```

```
eleccions.F <- eleccions.Drmh %* %eleccions.svd$u %* %diag(eleccions.svd$d)
```

```
eleccions.rowi <- diag(eleccions.r) %* %eleccions.F^2
```

```
eleccions.rowi [,1:10]
```

```
# (Contribución absoluta) Componentes individuales de la inercia de los par-  
tidos (columnas) en los 10 ejes.
```

```
eleccions.G <- eleccions.Dcmh %* %eleccions.svd$v %* %diag(eleccions.svd$d)
```

```
eleccions.G
```

```
eleccions.colj <- diag(eleccions.c) %* %eleccions.G^2
```

```
eleccions.colj [,1:10]
```

```
# (Contribucion relativa) Contribución de las comarcas a la inercia.
```

```
(eleccions.rowi/apply(eleccions.rowi,1,sum))[,1:10]
```

```
(eleccions.rowi/apply(eleccions.rowi,1,sum))[,1:2]
```

```
# (Contribucion relativa) Contribución de los partidos a la inercia.
```

```
(eleccions.colj/apply(eleccions.colj,1,sum))[,1:10]
```

```
(eleccions.colj/apply(eleccions.colj,1,sum))[,1:2]
```

**# RESUMEN GENERAL Y COMPLETO DEL AC.**

**# Cálculo del AC simple: inercias principales; y para filas y columnas, las masas, las distancias ji-cuadrado al centroide, las inercias y las coordenadas estándares en las dos primeras dimensiones.**

ca(eleccions)

**# Resultados completos del AC simple.**

summary(ca(eleccions))

# Bibliografía

- [1] C. M. Cuadras: *Métodos de análisis multivariante*. PPU, 2ª edición, 1991.
- [2] J.M. Cornejo: *Técnicas de Investigación Social: El Análisis de Correspondencias*. Barcelona, PPU, 2ª edición, 1988.
- [3] J.M. Batista y M.R. Martínez: *Análisis Multivariado. Análisis en Componentes Principales*. Barcelona, Hispano Europea, 1989.
- [4] J.P. Benzécri: *Correspondence Analysis Handbook*. Marcel Dekker. New York·Basel·Hong Kong, Inc, 1ª edición, 1992.
- [5] J.P. & F. Benzécri; et collaborateurs: *Pratique de l'Analyse des Données. 1.- Analyse Des Correspondences: Exposé Élémentaire*. Bordas, Paris, 1ª edición, 1980.
- [6] J.P. Benzécri: *Histoire et Préhistoire de l'Analyse des Données*. Dunod, Paris, 1982.
- [7] Michael Greenacre: *La práctica del análisis de correspondencias*. Fundación BBVA, 1ª edición, 2008.
- [8] Peter Dalgaard: *Introductory Statistics with R*. Springer - Verlag. New York, Inc, 1ª edición, 2002.
- [9] T. Foucart: *Analyse Factorielle de Tableaux Multiples*. Masson, Paris, 1984.
- [10] T. Foucart: *Analyse Factorielle. Programmation Sur Micro-ordinateurs*. Masson, Paris, 2ème édition, 1985.
- [11] Ramón Puig y Josep Vives: *Anàlisi de correspondències i geografia electoral: les eleccions al Parlament de Catalunya (1984-1999)*. Documento no publicado.