

ANÁLISIS DE CORRESPONDENCIAS

Uno de los fines del análisis de correspondencias es describir las relaciones existentes entre dos variables nominales, recogidas en una tabla de correspondencias, sobre un espacio de pocas dimensiones, mientras que al mismo tiempo se describen las relaciones entre las categorías de cada variable. Para cada variable, las distancias sobre un gráfico entre los puntos de categorías reflejan las relaciones entre las categorías, con las categorías similares representadas próximas unas a otras. La proyección de los puntos de una variable sobre el vector desde el origen hasta un punto de categoría de la otra variable describe la relación entre ambas variables.

El análisis factorial es una técnica habitualmente utilizada para describir las relaciones existentes entre variables en un espacio de pocas dimensiones. Sin embargo, el análisis factorial requiere que estas variables sean de intervalo. Por su parte, el análisis de correspondencias asume que las variables son nominales y permite describir las relaciones entre las categorías (o modalidades) de cada variable, así como la relación entre las variables.

Para analizar la aplicación práctica del Análisis Factorial de Correspondencias, vamos a realizar un análisis de correspondencias sobre los datos del fichero **Resultados_electorales.sav**, que contiene los resultados que se produjeron en España en las elecciones generales del 14 de marzo de 2004. Las variables que se incluyen en este fichero son las siguientes:

- **NUTS1:** Clasificación estadística de unidades territoriales de nivel 1, definida por EUROSTAT. Esta variable presenta las siguientes categorías, que incluyen, en cada caso, las correspondientes Comunidades Autónomas:
 - **1:** Región NOROESTE (Galicia, Asturias y Cantabria).
 - **2:** Región NORESTE (País Vasco, Navarra, La Rioja y Aragón).
 - **3:** Región MADRID (Madrid).
 - **4:** Región CENTRO (Castilla y León, Castilla-La Mancha y Extremadura).
 - **5:** Región ESTE (Cataluña, Comunidad Valenciana e Illes Balears).
 - **6:** Región SUR (Andalucía, Murcia, Ceuta y Melilla).
 - **7:** Región CANARIAS (Canarias).
- **CCAA:** Denominación geográfica de cada Comunidad Autónoma. Esta variable presenta 19 modalidades diferentes, una para cada Comunidad Autónoma.
- **Partido:** Variable categórica que recoge, en cada Comunidad Autónoma, la denominación de los partidos políticos que han obtenido algún escaño por el territorio correspondiente.
- **Votos:** Número total de votos que ha obtenido cada partido político al que le ha sido asignado algún escaño en la circunscripción correspondiente.
- **Escaños:** Número de escaños obtenidos por cada partido político en cada circunscripción electoral.

Como la importancia relativa de la situación de cada partido en cada una de las entidades territoriales consideradas se mide en términos de los escaños que se consiguen en cada circunscripción, es necesario ponderar el conjunto de datos por la variable **Escaños**, mediante el procedimiento **ponderar casos...** del menú **Datos**.

Así, se comienza por seleccionar la opción de **Ponderar casos...**, bajo el menú de **Datos**:

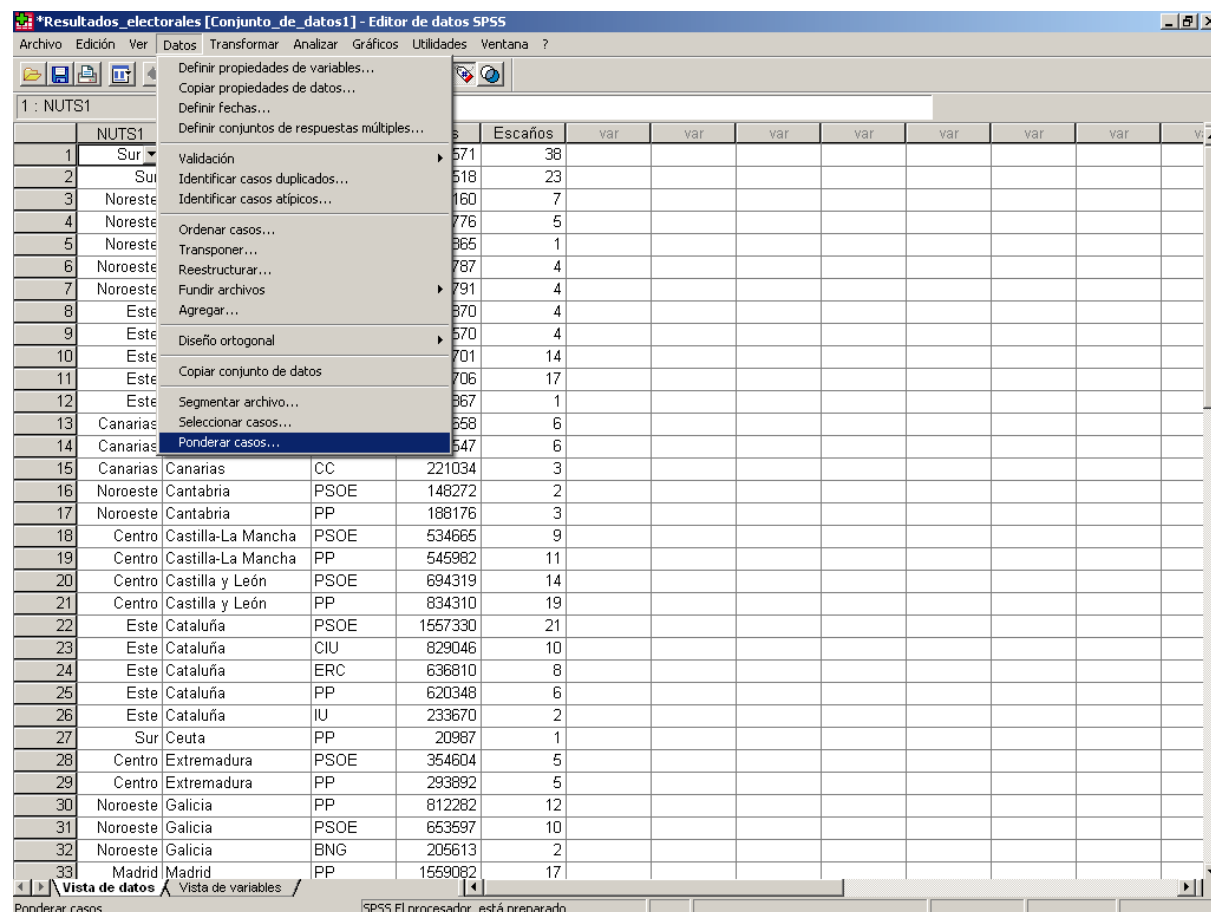


Figura 1: Selección del procedimiento de **Ponderar casos**.

Dentro del procedimiento de **Ponderar casos**, se selecciona la opción de **Ponderar casos mediante**, y se introduce como variable de ponderación el **Número de escaños**, como se muestra en la Figura 2.

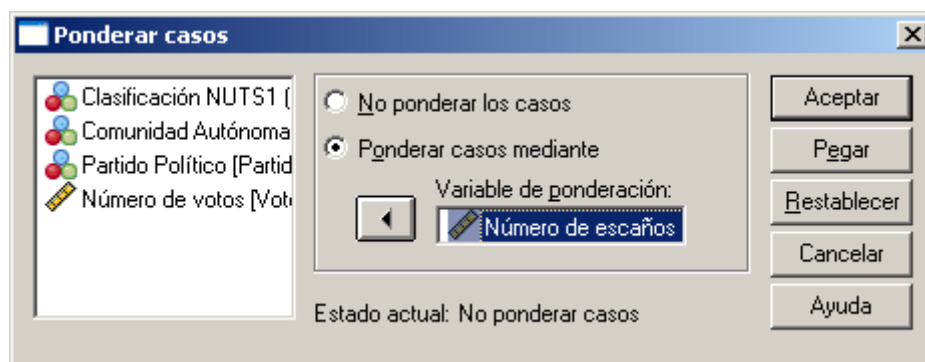


Figura 2: Selección de la variable de ponderación en el procedimiento de **Ponderar casos**.

Para realizar un análisis factorial de correspondencias simple (teniendo únicamente dos caracteres o variables cualitativas), se utiliza el procedimiento **Análisis de correspondencias** que se encuentra dentro del menú **Analizar**, en el submenú **Reducción de datos**, como se aprecia en la Figura 3:

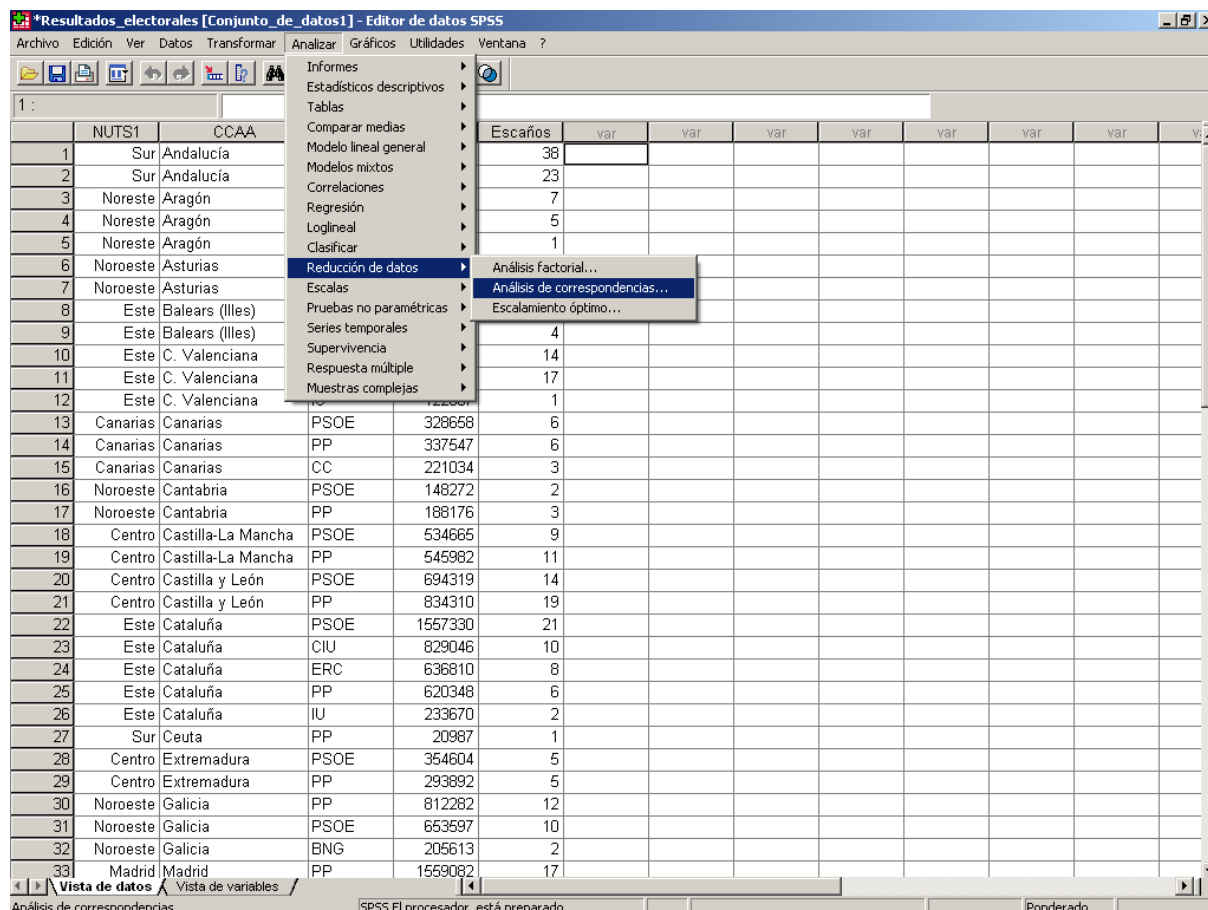


Figura 3: Selección del procedimiento **Análisis de correspondencias**.

El cuadro de diálogo del procedimiento de análisis de correspondencias sólo permite analizar dos variables en cada proceso, siendo éstas de tipo cualitativo, por lo que hay que definir el rango de valores que delimita las posibles modalidades de cada una de ellas.

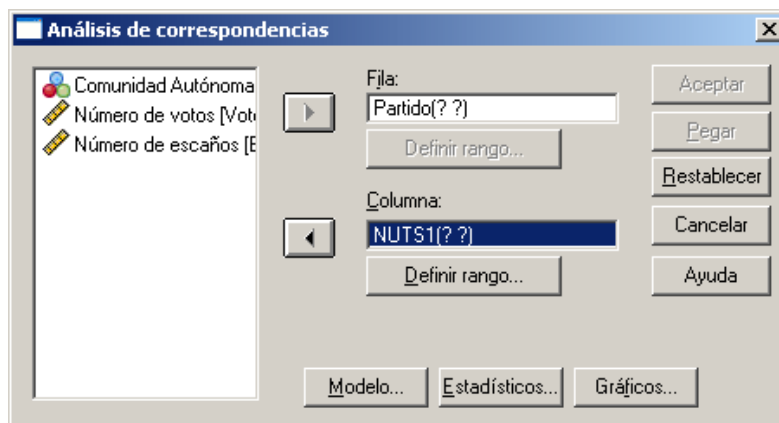


Figura 4: Cuadro de diálogo del procedimiento **Análisis de correspondencias**.

Como se aprecia, se debe definir el rango de modalidades que toman ambos atributos, en el caso del partido político, el rango es de 1 a 11, y en el caso de las unidades territoriales, de 1 a 7. Para ello, se pulsa en el botón **Definir rango...** de ambos atributos, seleccionándolos previamente, apareciendo los siguientes cuadros de opciones, en los que se reproducen los valores que hay que especificar:

The figure shows two side-by-side dialog boxes titled "Análisis de correspondencias: Definir rango de filas" and "Análisis de correspondencias: Definir rango de columnas".

The left dialog box, "Definir rango de filas", is for the variable "Partido". It has input fields for "Valor mínimo:" (1) and "Valor máximo:" (11), with an "Actualizar" button. Below, under "Restricciones para las categorías", there is a list of categories (1-9) and three radio buttons: "Ninguna" (selected), "Las categorías deben ser iguales", and "La categoría es suplementaria".

The right dialog box, "Definir rango de columnas", is for the variable "NUTS1". It has input fields for "Valor mínimo:" (1) and "Valor máximo:" (7), with an "Actualizar" button. Below, under "Restricciones para las categorías", there is a list of categories (1-7) and the same three radio buttons: "Ninguna" (selected), "Las categorías deben ser iguales", and "La categoría es suplementaria".

Figura 5: Definición del número de modalidades de las características, en las filas y en las columnas.

Si se pulsa el botón de **Modelo** del cuadro de diálogo del procedimiento de Análisis de correspondencias, se obtiene el siguiente conjunto de opciones:

The figure shows a dialog box titled "Análisis de correspondencias: Modelo".

At the top, "Dimensiones en la solución:" is set to 2.

Under "Medida de distancia", there are two radio buttons: "Chi-cuadrado" (selected) and "Euclídea".

Under "Método de estandarización", there are five radio buttons: "Se eliminan las medias de filas y columnas" (selected), "Se eliminan las medias de filas", "Se eliminan las medias de columnas", "Se igualan los totales de fila y se eliminan las medias", and "Se igualan los totales de columna y se eliminan las medias".

Under "Método de normalización", there are three radio buttons: "Simétrico", "Principal" (selected), and "Principal por columna". There is also a "Personalizado:" option with a value of 0.

Figura 6: Especificación del modelo en el procedimiento Análisis de correspondencias.

El cuadro de diálogo de la opción **Modelo** permite especificar el número de dimensiones, la medida de distancia, el método de estandarización y el método de normalización.

Para hacer el análisis de correspondencias típico, se debe elegir como medida de distancia la Chi-cuadrado, que utiliza una distancia ponderada entre los perfiles, donde la ponderación es

la masa de las filas o de las columnas. Por su parte, la distancia euclídea utiliza la raíz cuadrada de la suma de los cuadrados de las diferencias entre los pares de filas y entre los pares de columnas

Utilizaremos como método de normalización el Principal, pues sirve para examinar las diferencias entre las categorías dentro de una o de ambas variables.

Si se pulsa el botón de **Estadísticos** del cuadro de diálogo del procedimiento de Análisis de correspondencias, se obtiene el siguiente conjunto de opciones, de las cuales se señalan las que se muestran en la Figura 7:

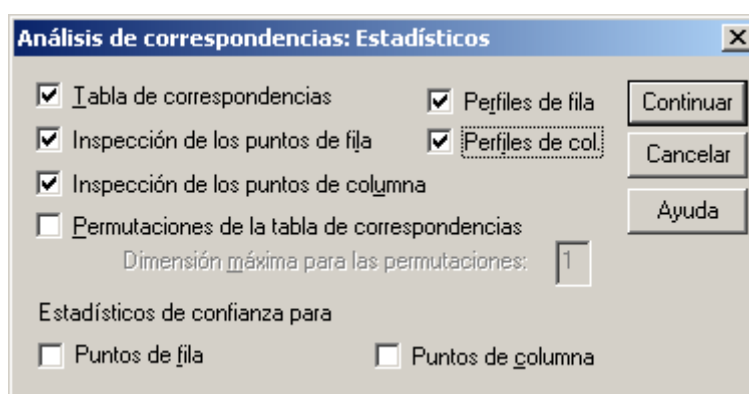


Figura 7: Opciones de **Estadísticos** del **Análisis de correspondencias**.

Al solicitar la **Tabla de correspondencias**, aparecerá en los resultados la tabla de contingencia de las variables de entrada con los totales marginales de fila y columna. En **Perfiles de fila** y **Perfiles de columna**, la distribución a través de las categorías de la variable de columna y de fila, respectivamente. La **Inspección de los puntos de fila y de columna** proporciona, para cada categoría de fila y de columna, respectivamente, las puntuaciones, la masa, la inercia, la contribución a la inercia de la dimensión y la contribución de la dimensión a la inercia del punto. Las **Permutaciones de la tabla de correspondencias** presentan la tabla de correspondencias reorganizada de tal manera que las filas y las columnas estén en orden ascendente de acuerdo con las puntuaciones en la primera dimensión. Por último, los **Estadísticos de confianza para puntos de fila y de columna** incluyen la desviación típica y las correlaciones para todos los puntos de fila y de columna no suplementarios, respectivamente.

Si se pulsa el botón de **Gráficos** del cuadro de diálogo del procedimiento de Análisis de correspondencias, se obtiene el conjunto de opciones que aparecen en la Figura 8, en la que se solicita la realización de todos los gráficos posibles (el diagrama de dispersión espacial no se puede realizar en este caso).

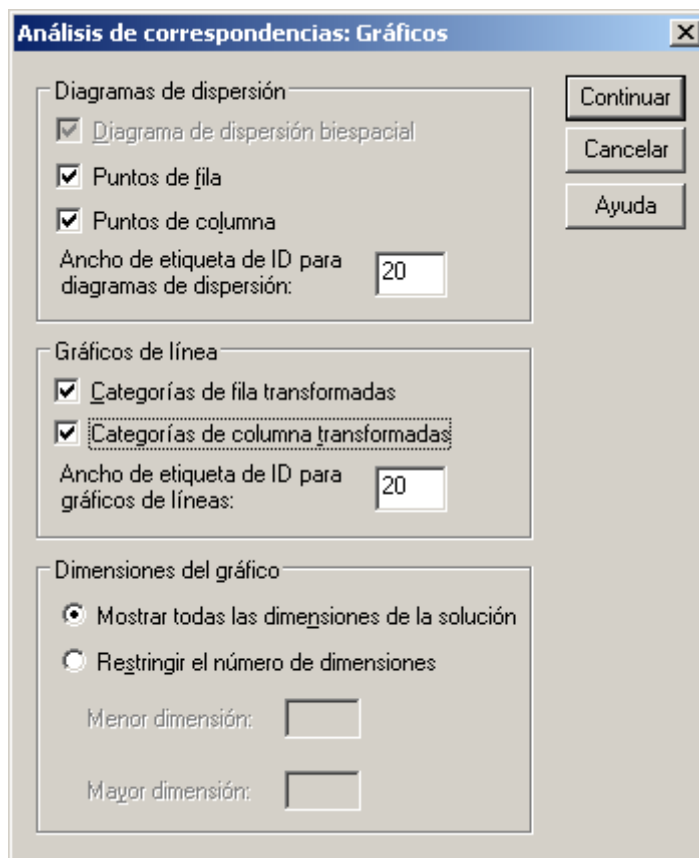


Figura 8: Opciones de **Gráficos** del **Análisis de correspondencias**.

Una vez seleccionadas las variables y definidas sus categorías, y marcadas todas las opciones necesarias para este procedimiento, los resultados que se obtienen son:

Correspondencias

Inicialmente, se muestra un cuadro de créditos en el que aparecen los responsables del procedimiento de Análisis de correspondencias (es un procedimiento incorporado a SPSS en su versión 10, y actualizado en la versión 14, y está realizado por un grupo de investigación holandés).

Créditos

CORRESPONDENCE
Version 1.1
by
Data Theory Scaling System Group (DTSS)
Faculty of Social and Behavioral Sciences
Leiden University, The Netherlands

La tabla de correspondencias (de contingencia) entre las variables en estudio es:

Tabla de correspondencias

	Clasificación NUTS1 (UE)							
Partido Político	Noroeste	Noreste	Madrid	Centro	Este	Sur	Canarias	Margen activo
PSOE	16	18	16	28	39	41	6	164
PP	19	13	17	35	27	31	6	148
CIU	0	0	0	0	10	0	0	10
ERC	0	0	0	0	8	0	0	8
EAJ-PNV	0	7	0	0	0	0	0	7
IU	0	0	2	0	3	0	0	5
CC	0	0	0	0	0	0	3	3
BNG	2	0	0	0	0	0	0	2
CHA	0	1	0	0	0	0	0	1
EA	0	1	0	0	0	0	0	1
NA-BAI	0	1	0	0	0	0	0	1
Margen activo	37	41	35	63	87	72	15	350

Los perfiles de fila y columna, que detallan la estructura de los datos para cada característica, son los siguientes:

Perfiles de fila

Partido Político	Clasificación NUTS1 (UE)							
	Noroeste	Noreste	Madrid	Centro	Este	Sur	Canarias	Margen activo
PSOE	.098	.110	.098	.171	.238	.250	.037	1.000
PP	.128	.088	.115	.236	.182	.209	.041	1.000
CIU	.000	.000	.000	.000	1.000	.000	.000	1.000
ERC	.000	.000	.000	.000	1.000	.000	.000	1.000
EAJ-PNV	.000	1.000	.000	.000	.000	.000	.000	1.000
IU	.000	.000	.400	.000	.600	.000	.000	1.000
CC	.000	.000	.000	.000	.000	.000	1.000	1.000
BNG	1.000	.000	.000	.000	.000	.000	.000	1.000
CHA	.000	1.000	.000	.000	.000	.000	.000	1.000
EA	.000	1.000	.000	.000	.000	.000	.000	1.000
NA-BAI	.000	1.000	.000	.000	.000	.000	.000	1.000
Masa	.106	.117	.100	.180	.249	.206	.043	

Perfiles de columna

[illegible]

El resumen del procedimiento, en el que se muestra la proporción de inercia que explica cada valor propio, y su dimensión asociada, es el siguiente:

Resumen

Dimensión	Valor propio	Inercia	Chi-cuadrado	Sig.	Proporción de inercia		Confianza para el Valor propio	
					Explicada	Acumulada	Desviación típica	Correlación
							2	
1	.490	.240			.360	.360	.054	.051
2	.454	.206			.309	.669	.089	
3	.391	.153			.229	.898		
4	.210	.044			.066	.964		
5	.133	.018			.026	.991		
6	.079	.006			.009	1.000		
Total		.667	233.386	.000 ^a	1.000	1.000		

a. 60 grados de libertad

El examen de los puntos de fila y de columna es:

Examen de los puntos de fila

Partido Político	Masa	Puntuación en la dimensión		Inercia	Contribución				
		1	2		De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto		
					1	2	1	2	Total
PSOE	.469	.008	.012	.006	.000	.000	.005	.012	.016
PP	.423	.019	-.075	.021	.001	.012	.007	.114	.121
CIU	.029	.959	.837	.086	.109	.097	.304	.232	.536
ERC	.023	.959	.837	.069	.087	.078	.304	.232	.536
EAJ-PNV	.020	-2.551	.489	.151	.542	.023	.864	.032	.896
IU	.014	.659	.497	.029	.026	.017	.212	.121	.333
CC	.009	.263	-4.277	.191	.002	.761	.003	.819	.822
BNG	.006	.070	-.220	.048	.000	.001	.001	.006	.006
CHA	.003	-2.551	.489	.022	.077	.003	.864	.032	.896
EA	.003	-2.551	.489	.022	.077	.003	.864	.032	.896
NA-BAI	.003	-2.551	.489	.022	.077	.003	.864	.032	.896
Total activo	1.000			.667	1.000	1.000			

a. Normalización Principal

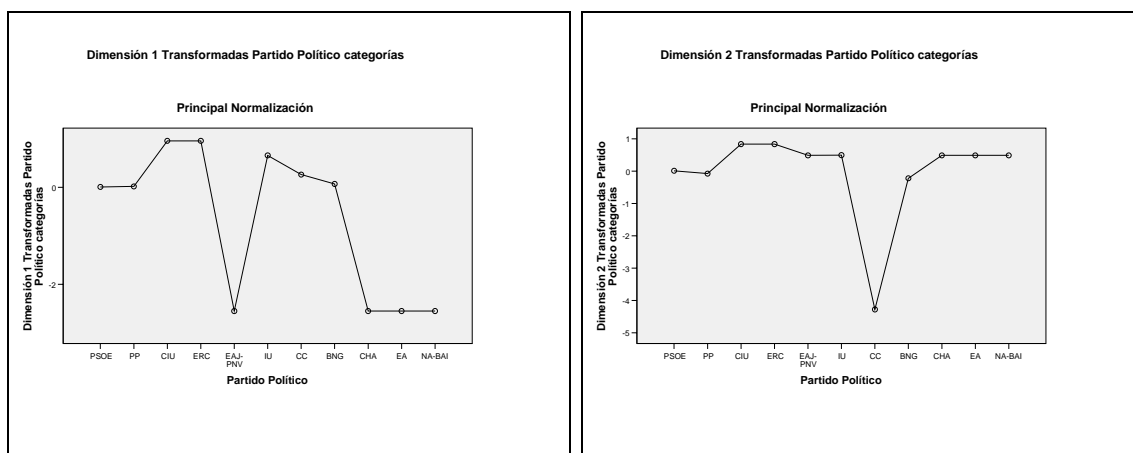
Examen de los puntos columna

Clasificación NUTS1 (UE)	Masa	Puntuación en la dimensión		Inercia	Contribución				
		1	2		De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto		
					1	2	1	2	Total
Noroeste	.106	.034	-.100	.056	.001	.005	.002	.019	.021
Noreste	.117	-1.251	.222	.203	.763	.028	.903	.028	.932
Madrid	.100	.103	-.006	.023	.004	.000	.045	.000	.045
Centro	.180	.028	-.080	.027	.001	.006	.005	.043	.048
Este	.249	.470	.380	.142	.228	.174	.386	.252	.638
Sur	.206	.025	-.056	.027	.001	.003	.005	.024	.029
Canarias	.043	.129	-1.941	.188	.003	.784	.004	.859	.862
Total activo	1.000			.667	1.000	1.000			

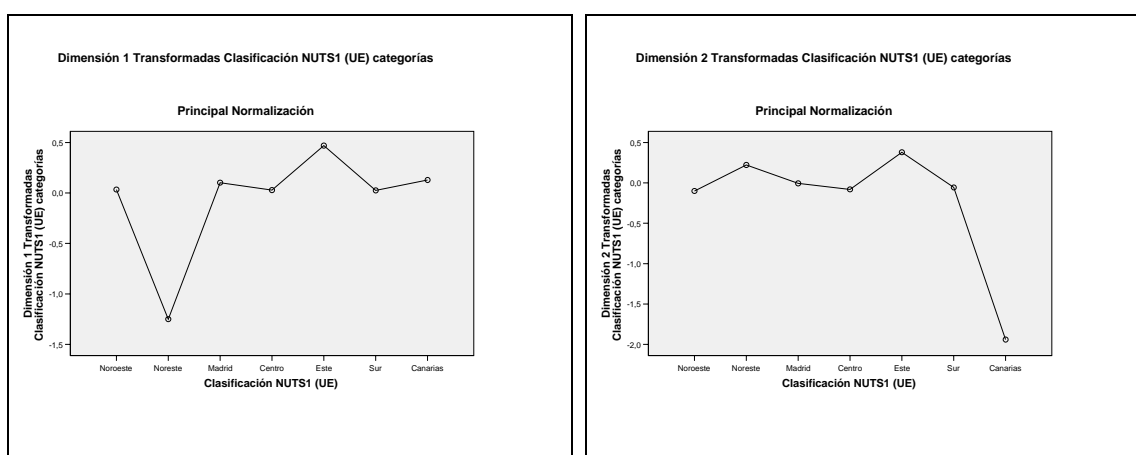
a. Normalización Principal

Las representaciones gráficas de las modalidades con respecto a las dos dimensiones son:

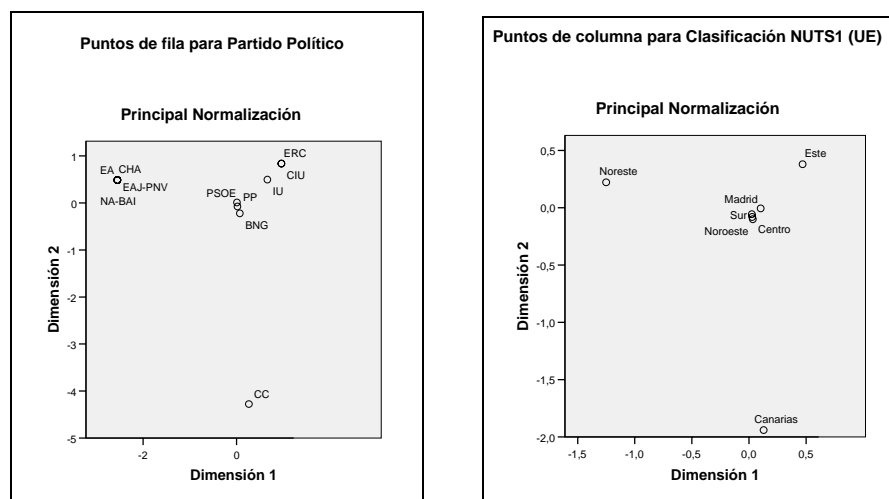
Para las modalidades de la variable Partido político:



Para las modalidades de la variable NUTS1:



Las representaciones de las modalidades de cada característica con respecto a las dos dimensiones resultan:



Para representar en el mismo espacio las coordenadas de las modalidades de ambas variables, se recuperan los datos del fichero **Puntuaciones_modalidades.sav**, en el que se han almacenado las puntuaciones en ambas dimensiones de las distintas modalidades que presentan ambas características (véase la relación de estas puntuaciones en las tablas de examen de los puntos fila y columna). Una vez abierto el fichero, se realiza una representación gráfica en un diagrama de dispersión, utilizando las siguientes opciones:

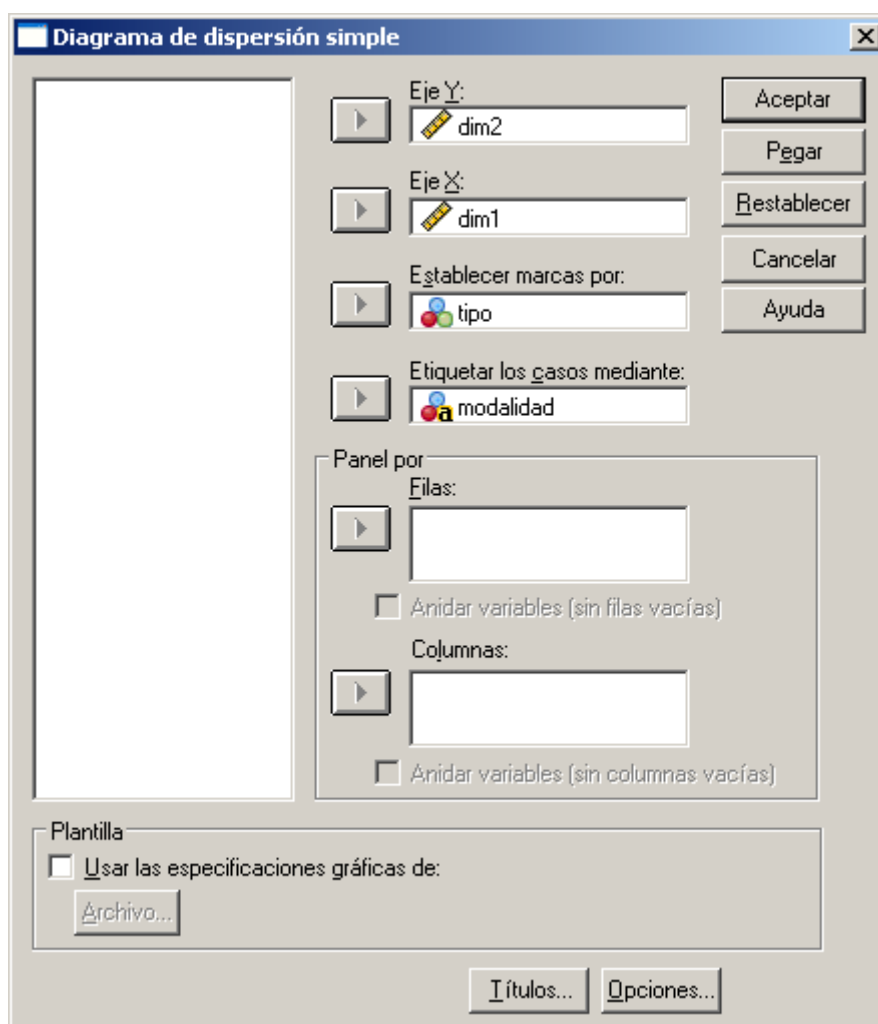


Figura 9: Representación gráfica del diagrama de estímulos combinados.

El resultado obtenido es el siguiente gráfico, en el que se pueden analizar las relaciones de atracción-repulsión entre las modalidades de los dos atributos considerados:

Gráfico

