ANÁLISIS CORRESPONDENCIAS SIMPLES Y MÚLTIPLES

Fac. Ciencias Económicas y Empresariales

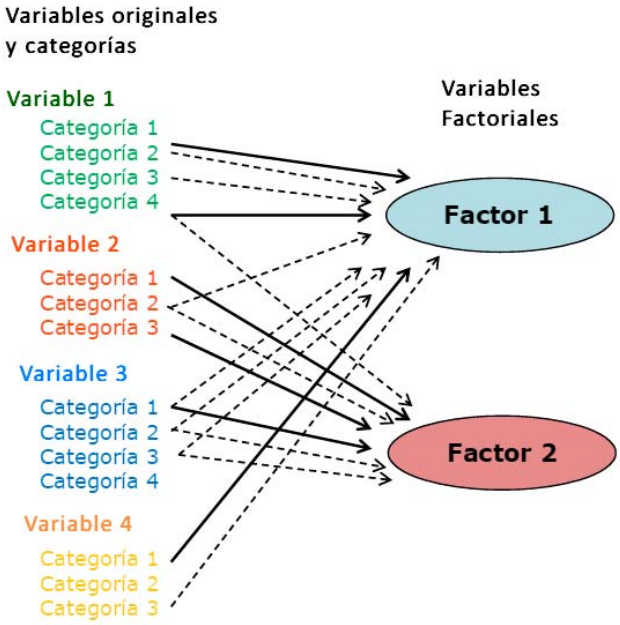
UNIVERSIDAD AUTONOMA DE MADRID

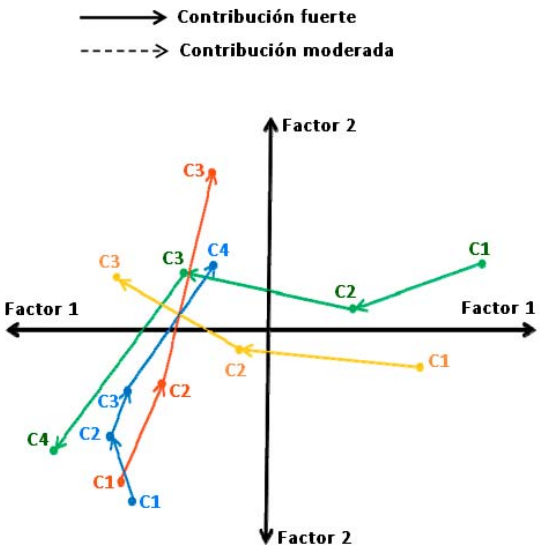
SIMPLES
MÚLTIPLES

ANÁLISIS DE CORRESPONDENCIAS

Es una técnica de reducción de dimensiones, una técnica para visualizar una nube de puntos multidimensional en dos dimensiones. Como las demás técnicas de reducción de dimensiones, en un procedimiento de traslado de una nube de puntos definida en un espacio de muchas dimensiones a un espacio de dos dimensiones donde poder visualizar la posición relativa de unos puntos.

Este traslado se hará respetando al máximo las posiciones relativas de los puntos en la nube de puntos original.





ANÁLISIS DE CORRESPONDENCIAS (AC)

Es una técnica descriptiva o exploratoria cuyo objetivo es resumir una gran cantidad de datos en un número reducido de dimensiones, con la menor pérdida de información posible.
En esta línea, su objetivo es similar al de los métodos factoriales, salvo que en el caso del análisis de correspondencias el método se aplica sobre variables categóricas u ordinales.

El análisis de "correspondencias simples" se utiliza a menudo en la representación de datos que se pueden presentar en forma de tablas de contingencia de dos variables nominales u ordinales. Otras utilizaciones implican el tratamiento de tablas de proximidad o distancia entre elementos, y tablas de preferencias.

Si se trata de una tabla de contingencia de dos variables cualitativas, una variable cuyas categorías aparecen en filas y la otra variable cuyas categorías son representadas en columnas, el análisis de correspondencias consiste en resumir la información presente en las filas y columnas de manera que pueda proyectarse sobre un subespacio reducido, y representarse simultáneamente los puntos fila y los puntos columna, pudiéndose obtener conclusiones sobre relaciones entre las dos variables nominales u ordinales de origen.

La extensión del análisis de correspondencias simples al caso de varias variables nominales (tablas de contingencia multidimensionales) se denomina "Análisis de Correspondencias Múltiples", y utiliza los mismos principios generales que la técnica anterior. En general se orienta a casos en los cuales una variable representa ítems o individuos y el resto son variables cualitativas u ordinales que representan cualidades.

Entre la utilización del Análisis de Correspondencias Simple y Múltiple, estudios:

- Preferencias de consumo en Investigación de Mercados.
- Posicionamiento de empresas a partir de las preferencias de consumidores.
- Búsqueda de tipologías de individuos respecto a variables cualitativas (patrones de enfermedades en medicina, perfiles psicológicos, comportamiento de especies en biología, etc.).

El Análisis de Correspondencias tiene dos objetivos básicos:

Asociación entre categorías de columnas o filas: Medir la asociación de solo una fila o columna, para ver, por ejemplo, si las modalidades de una variable pueden ser combinadas.

Asociación entre categorías de filas y columnas: Estudiar si existe relación entre categorías de las filas y columnas.

El análisis de correspondencias solo requiere que los datos representen las respuestas a una serie de preguntas y que estén organizadas en categorías. Dependiendo si existen dos o más variables el análisis será simple o múltiple.

Un análisis de correspondencias en SPSS: Analizar / Reducción de datos / Análisis de correspondencias.

PLANTEAMIENTO DEL PROBLEMA

El Análisis de Correspondencias es una técnica estadística que se utiliza para analizar, desde un punto de vista gráfico, las relaciones de dependencia e independencia de un conjunto de variables categóricas a partir de los datos de una tabla de contingencia.

Para ello asocia a cada una de las modalidades de la tabla, un punto en el espacio R^n (generalmente n = 2) de forma que las relaciones de "cercanía / lejanía "entre los puntos calculados reflejen las relaciones de dependencia y semejanza existentes entre ellas.

Se comienza analizando el problema bidimensional que es el que analiza el Análisis de Correspondencias propiamente dicho. Posteriormente se trata, brevemente, del problema n-dimensional ($n \geq 3$) que es el problema que analiza el Análisis de Correspondencias Múltiples.

Sea X e Y dos variables categóricas, respectivamente, con valores $\{x_1, \dots, x_k\}$ e $\{y_1, \dots, y_m\}$, observando dichas variables en N elementos de una población. La intersección entre una fila y una columna da lugar a una celda o casilla, cuya frecuencia observada es n_{ij}

X \ Y	y_1	y_2	.....	y_j	.....	y_m	
x_1	n_{11} (e_{11})	n_{12} (e_{12})	.....	n_{1j} (e_{1j})	.....	n_{1m} (e_{1m})	$n_{1\bullet}$
x_2	n_{21} (e_{21})	n_{22} (e_{22})	.....	n_{2j} (e_{2j})	.....	n_{2m} (e_{2m})	$n_{2\bullet}$
⋮	⋮	⋮	⋮		.....	⋮	⋮
x_i	n_{i1} (e_{i1})	n_{i2} (e_{i2})	.....	n_{ij} (e_{ij})	.....	n_{im} (e_{im})	$n_{i\bullet}$
⋮	⋮	⋮	.....		.....	⋮	⋮
x_k	n_{k1} (e_{k1})	n_{k2} (e_{k2})	.....	n_{kj} (e_{kj})	.....	n_{km} (e_{km})	$n_{k\bullet}$
	$n_{\bullet 1}$	$n_{\bullet 2}$		$n_{\bullet j}$		$n_{\bullet m}$	N

$$n_{i\bullet} = \sum_{j=1}^{m} n_{ij}$$

$$n_{\bullet j} = \sum_{i=1}^{k} n_{ij}$$

$$N = \sum_{i} n_{i\bullet} = \sum_{j} n_{\bullet j}$$

$$e_{ij} = \frac{n_{i\bullet} \times n_{\bullet j}}{N}$$

Los PERFILES MARGINALES describen la distribución marginal de las variables X e Y, respectivamente, se reflejan en las siguientes tablas:

$X = x_i$	x_1	x_2	.....	x_i	.....	x_k	Total
Frecuencias relativas marginales	$\frac{n_{1\bullet}}{N}$	$\frac{n_{2\bullet}}{N}$	.....	$\frac{n_{i\bullet}}{N}$	.....	$\frac{n_{k\bullet}}{N}$	1

$Y = y_j$	y_1	y_2	.....	y_j	.....	y_m	Total
Frecuencias relativas marginales	$\frac{n_{\bullet 1}}{N}$	$\frac{n_{\bullet 2}}{N}$	.....	$\frac{n_{\bullet j}}{N}$	.....	$\frac{n_{\bullet m}}{N}$	1

Los PERFILES CONDICIONALES describen las distribuciones condicionadas asociadas a la Tabla de Correspondencias.

- Los "perfiles fila" describen las distribuciones condicionadas de la variable Y por las distintas modalidades de la variable X. Se obtienen a partir de la tabla de correspondencias y el perfil marginal de X mediante las siguientes expresiones:

$f(Y / X = x_i)$	y_1	.....	y_j	.....	y_m	Total
x_1	$\frac{n_{11}}{n_{1\bullet}}$	.....	$\frac{n_{1j}}{n_{1\bullet}}$	.....	$\frac{n_{1m}}{n_{1\bullet}}$	1
x_2	$\frac{n_{21}}{n_{2\bullet}}$	.....	$\frac{n_{2j}}{n_{2\bullet}}$	.....	$\frac{n_{2m}}{n_{2\bullet}}$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	.....	$\vdots$	$\vdots$
x_i	$\frac{n_{i1}}{n_{i\bullet}}$	.....	$\frac{n_{ij}}{n_{i\bullet}}$	.....	$\frac{n_{im}}{n_{i\bullet}}$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	.....	$\vdots$	$\vdots$
x_j	$\frac{n_{j1}}{n_{j\bullet}}$	.....	$\frac{n_{jj}}{n_{j\bullet}}$	.....	$\frac{n_{jm}}{n_{j\bullet}}$	1
$\vdots$	$\vdots$	.....	$\vdots$	.....	$\vdots$	$\vdots$
x_k	$\frac{n_{k1}}{n_{k\bullet}}$	.....	$\frac{n_{kj}}{n_{k\bullet}}$	.....	$\frac{n_{km}}{n_{k\bullet}}$	1

Distancia Chi-cuadrado entre perfiles filas (i, j):

$$d_{ij} = \frac{1}{n_{\bullet 1}}\left[\frac{n_{i1}}{n_{i\bullet}} - \frac{n_{j1}}{n_{j\bullet}}\right]^2$$

Distancia Chi-cuadrado en la tabla entre perfiles filas:

$$d_{ij} = \sum_{h=1}^{m} \frac{1}{n_{\bullet h}}\left[\frac{n_{ih}}{n_{i\bullet}} - \frac{n_{jh}}{n_{j\bullet}}\right]^2$$

- Los "perfiles columna" describen las distribuciones condicionadas de la variable X por las distintas modalidades de la variable Y. Se obtienen a partir de la tabla de correspondencias y el perfil marginal de X mediante las siguientes expresiones:

$f(X / Y = y_j)$	y_1	.....	y_i	.....	y_j	.....	y_m
x_1	$\frac{n_{11}}{n_{\bullet 1}}$	.....	$\frac{n_{1i}}{n_{\bullet i}}$	.....	$\frac{n_{1j}}{n_{\bullet j}}$	.....	$\frac{n_{1m}}{n_{\bullet m}}$
x_2	$\frac{n_{21}}{n_{\bullet 1}}$	.....	$\frac{n_{2i}}{n_{\bullet i}}$	.....	$\frac{n_{2j}}{n_{\bullet j}}$	.....	$\frac{n_{2m}}{n_{\bullet m}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	$\frac{n_{i1}}{n_{\bullet 1}}$	.....	$\frac{n_{ii}}{n_{\bullet i}}$	.....	$\frac{n_{ij}}{n_{\bullet j}}$	.....	$\frac{n_{im}}{n_{\bullet m}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	$\frac{n_{k1}}{n_{\bullet 1}}$	.....	$\frac{n_{ki}}{n_{\bullet i}}$	.....	$\frac{n_{kj}}{n_{\bullet j}}$	.....	$\frac{n_{km}}{n_{\bullet m}}$
Total	1		1		1	1	1

Distancia Chi-cuadrado entre perfiles columnas (i, j):

$$d^c_{ij} = \frac{1}{n_{1\bullet}}\left[\frac{n_{1i}}{n_{\bullet i}} - \frac{n_{1j}}{n_{\bullet j}}\right]^2$$

Distancia Chi-cuadrado en la tabla entre perfiles columnas:

$$d^c_{ij} = \sum_{h=1}^{k} \frac{1}{n_{h\bullet}}\left[\frac{n_{hi}}{n_{\bullet i}} - \frac{n_{hj}}{n_{\bullet j}}\right]^2$$

Trabajar con perfiles facilita la interpretación, pero también puede producir una visión equivocada de la relación entre variables en la medida que todos los puntos tienen la misma importancia: los marginales de los perfiles y columna son iguales a 1.

Para evitar este problema el análisis de correspondencias deberá utilizar una distancia que no olvide las diferencias entre los efectivos de cada línea (o columna). La distancia Chi-cuadrado cumple la condición de ponderar cada perfil por un peso. Así cada fila (o columna) está afectada de un peso proporcional a su importancia en el conjunto, peso conocido como "masa".

Al considerar cada punto con una "masa proporcional" a su frecuencia se evita "privilegiar" las categorías con pocos efectivos. Se trata, de hecho, de una distancia euclídea ponderada por el inverso de la masa de las columnas cuando se mide la distancia entre filas, o por la masa de las filas para la distancia entre las columnas.

La distancia Chi-cuadrado cumple el principio de la equivalencia distribucional, que postula que si dos categorías tienen perfiles idénticos pueden ser sustituidas por una sola categoría que sea la suma de sus pesos, sin que con ello se modifique la distancia entre las filas o columnas. La importancia de esta propiedad estriba en que garantiza la estabilidad en los resultados con independencia de la codificación en las variables; de modo que es posible agrupar categorías que tienen perfiles coincidentes, tanto por filas como por columnas. Si el resultado se mantiene estable tras unir categorías, de igual modo estos resultados no mejoran al realizar más subdivisiones de categorías homogéneas.

DEPENDENCIA E INDEPENDENCIA EN TABLAS DE CORRESPONDENCIAS

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-cuadrado de Pearson.

Se contrasta la hipótesis nula que presupone la independencia entre ambas variables, mediante el estadístico χ^2 de Pearson.

H_0: Ambas variables son independientes $\qquad$ H_1: Existe una relación de dependencia

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

Se define el estadístico observado: $\displaystyle\sum_{i=1}^{k}\sum_{j=1}^{m}\frac{(n_{ij}-e_{ij})^2}{e_{ij}} = \chi^2_{(k-1)\,.\,(m-1)}$

siendo $e_{ij} = E\left[\, n_{ij} / H_0 \text{ es cierta} \,\right] = \dfrac{n_{i\bullet} \times n_{\bullet j}}{N}$, el estadístico observado se puede expresar también:

$$\sum_{i=1}^{k}\sum_{j=1}^{m}\frac{(n_{ij}-e_{ij})^2}{e_{ij}} = \sum_{i=1}^{k}\sum_{j=1}^{m}\frac{n_{i\bullet}\left[\dfrac{n_{ij}}{n_{i\bullet}}-\dfrac{n_{\bullet j}}{N}\right]^2}{\dfrac{n_{\bullet j}}{N}} = \sum_{i=1}^{k}\sum_{j=1}^{m}\frac{n_{\bullet j}\left[\dfrac{n_{ij}}{n_{\bullet j}}-\dfrac{n_{i\bullet}}{N}\right]^2}{\dfrac{n_{i\bullet}}{N}}$$

La región crítica para el contraste de independencia se determina: $P\left[\, \chi^2_{(k-1).(m-1)} \geq k / H_0 \,\right] = \alpha$

Así, pues, para un nivel de significación α:

$$\begin{cases} \chi^2_{(k-1).(m-1)} < \chi^2_{\alpha;\,(k-1).(m-1)} & \rightarrow \text{ X e Y son independientes al nivel } \alpha \\ \chi^2_{(k-1).(m-1)} \geq \chi^2_{\alpha;\,(k-1).(m-1)} & \rightarrow \text{ X e Y no son independientes al nivel } \alpha \end{cases}$$

- SPSS $\begin{cases} \text{Sig.asintótica}(p_\text{value}) \leq 0{,}05 & \Rightarrow \text{ Se rechaza } H_0 \\ \text{Sig.asintótica}(p_\text{value}) > 0{,}05 & \Rightarrow \text{ Se acepta } H_0 \end{cases}$

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 tipificados corregidos".

ANÁLISIS DE LOS RESIDUOS: Los residuos son las diferencias entre la frecuencia observada y la frecuencia esperada en cada casilla: $r_{ij} = n_{ij} - e_{ij}$. En el caso de que el contraste de χ^2 haya resultado significativo, estos residuos indicarán qué casillas contribuyen en mayor grado al valor del estadístico.

Cuanto mayor sea el valor de los residuos mayor es la probabilidad de que una determinada combinación de valores de las variables, esto es, una casilla, sea significativa.

Para que el análisis de los residuos resulte adecuado es necesario que previamente éstos hayan sido ajustados y estandarizados, para lo cual se suele aplicar la fórmula propuesta por Haberman (1978), que consiste en dividir el valor del residuo en cada casilla por su error típico.

Residuos tipificados: $r_{ij} = \dfrac{n_{ij} - e_{ij}}{\sqrt{e_{ij}}}$

Residuos tipificados corregidos: $r_{ij}^c = \dfrac{r_{ij}}{\sqrt{V(r_{ij})}} = \dfrac{(n_{ij} - e_{ij}) / \sqrt{e_{ij}}}{\sqrt{\left(1 - \dfrac{n_{i\bullet}}{N}\right)\left(1 - \dfrac{n_{\bullet j}}{N}\right)}} \approx N(0,1)$

ANÁLISIS DE CORRESPONDENCIAS CLASICO

Con los perfiles de filas y columnas descritos se elabora la matriz de coordenadas (distancias) utilizando la distancia Chi-cuadrado, que permitirá calibrar la magnitud de las diferencias entre la tabla de datos analizada y una tabla de datos sin relación entre las variables.

El método consiste en encontrar la descomposición en valores singulares de la matriz:

$$C = (r_{ij}) \quad \text{siendo} \quad r_{ij} = \frac{n_{ij} - e_{ij}}{\sqrt{e_{ij}}}$$

para construir un sistema de coordenadas (generalmente, bidimensional) asociado a las filas y columnas de la tabla de contingencia, que refleje las relaciones existentes entre dichas filas y columnas.

En la representación tienen un papel importante las llamadas distancias χ^2 entre perfiles, que son las que el análisis de correspondencias intenta reproducir en sus representaciones gráficas. Dichas distancias son distancias pitagóricas ponderadas entre perfiles que vienen dadas por las siguientes expresiones:

Distancia entre perfiles fila: $d_{ij} = \sum_{h=1}^{m} \dfrac{1}{n_{\bullet h}} \left[\dfrac{n_{ih}}{n_{i\bullet}} - \dfrac{n_{jh}}{n_{j\bullet}} \right]^2$

Distancia entre perfiles columnas: $d_{ij}^c = \sum_{h=1}^{k} \dfrac{1}{n_{h\bullet}} \left[\dfrac{n_{hi}}{n_{\bullet i}} - \dfrac{n_{hj}}{n_{\bullet j}} \right]^2$

Las distancias no se miden entre dos filas o dos columnas sino con relación al perfil medio de fila o columna, es decir, con relación al promedio de las coordenadas de esa fila (o columna) ponderada por su *masa* (peso proporcional a su importancia en el conjunto).

Este perfil medio aparecerá situado en el origen de coordenadas y es conocido como "centro de gravedad". La media de las distancias al cuadrado de cada punto de fila al centro de gravedad se conoce como "inercia de filas", o "inercia de columnas" cuando se trata de las columnas, e "inercia total de la nube de puntos" cuando se consideran todos los elementos de la tabla.

Una inercia baja significa que todos los productos están situados muy cerca del centro de gravedad y que en consecuencia son muy similares, mientras que altos valores de inercia en determinadas categorías implican grandes diferencias del perfil medio de las filas o las columnas.

Posteriormente se procede a diagonalizar la matriz C de varianza-covarianza con el fin de obtener los vectores y valores propios que definirán los nuevos ejes sobre los que será proyectada la nube de puntos.

Cuando la cantidad de inercia explicada con los primeros factores sea alta bastará con seleccionar un pequeño número de éstos (dos o tres) y representar la nube de puntos sobre gráficos de dos o tres dimensiones, obteniendo así una visión simplificada de las relaciones.

El análisis de correspondencias busca encontrar dos matrices (A, B) de coordenadas cartesianas:

$$A = \begin{pmatrix} a_1' \\ a_2' \\ \vdots \\ a_k' \end{pmatrix}$$ que represente a los puntos fila con $a_i = (a_{i1}, \cdots, a_{ih})'$

$$B = \begin{pmatrix} b_1' \\ b_2' \\ \vdots \\ b_m' \end{pmatrix}$$ que represente a los puntos columna con $b_j = (b_{j1}, \cdots, b_{jh})'$

(generalmente h = 2)

Hay diversas formas de calcular las matrices A y B, conocidas como normalizaciones. Una forma muy utilizada es la conocida como "normalización simétrica o canónica (ACC)", que busca satisfacer que el producto escalar $(a_i \,.\, b_j)$ sea proporcional a los residuos tipificados r_{ij}.

La normalización simétrica o canónica descompone la matriz $C = (r_{ij})$ en valores singulares calculando matrices U_{kxh}, D y V_{mxh}, con $H =$ mínimo $(k-1, m-1)$, tales que $C = U\,D\,V'$, siendo $U'U = V'V = I$, $D =$ diagonal $\{m_1, m_2, \ldots, m_h\}$ donde $\mu_i \equiv$ valores singulares $(i = 1, \ldots, H)$

Las matrices A y B se calculan a partir de las expresiones $A = D_k^{-1/2}\,U\,D$ y $B = D_m^{-1/2}\,V\,D$ con $D_k =$ diagonal $\{n_1, n_2, \ldots, n_k\}$, $D_m =$ diagonal $\{n_1, n_2, \ldots, n_m\}$

INTERPRETACIÓN BARICÉNTRICA:

$$\sum_{h=1}^{k} \frac{n_{h\bullet}}{N} a_{hj} = 0 \quad (j = 1, \cdots, h)$$

$$\sum_{h=1}^{k} \frac{n_{\bullet h}}{N} b_{hj} = 0 \quad (j = 1, \cdots, h)$$

Los puntos $(a_i')_{i=1,\cdots,k}$ y $(b_j')_{j=1,\cdots,m}$ tendrán una media baricéntrica igual al origen.

$$\mu_j \,.\, a_{ij} = \sum_{h=1}^{m} \frac{n_{ih}}{n_{i\bullet}} b_{hj} \quad (i = 1, \cdots, k) \;\; (j = 1, \cdots, H)$$

$$\mu_j \,.\, b_{ij} = \sum_{h=1}^{k} \frac{n_{hi}}{n_{i\bullet}} a_{hj} \quad (i = 1, \cdots, m) \;\; (j = 1, \cdots, H)$$

Las coordenadas de los puntos fila (columna) son medias ponderadas de las coordenadas de los puntos columna (fila) salvo un factor dado por los valores singulares.

Es decir, los puntos fila (columna) son, salvo un factor de dilatación $(1/\mu_j)$, el baricentro de los puntos columna (fila).

INTERPRETACIÓN RESULTADOS:

La "Inercia Total" (medida análoga a la variación total en el caso de las componentes principales) cuantifica el grado de dependencia entre las variables (X, Y):

$$IT = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{m}\frac{(n_{ij}-e_{ij})^2}{e_{ij}} = \sum_{h=1}^{H}\mu_h^2 = \sum_{h=1}^{H}\sum_{i=1}^{k} n_{i\bullet}a_{ih}^2 = \sum_{h=1}^{H}\sum_{j=1}^{m} n_{\bullet j}b_{jh}^2$$

A partir de la Inercia Total (IT) se calculan las proporciones de "inercia explicada" para cada una de las dimensiones $\left\{\frac{\mu_i^2}{IT}\right\}_{i=1,\cdots,H}$ que ayudan a ponderar la importancia de cada una de las dimensiones cuando se trate de explicar las dependencias observadas.

Las proporciones de inercia acumulada explicada por las i-ésimas primeras dimensiones $\left\{\sum_{h=1}^{i}\frac{\mu_i^2}{IT}\right\}_{i=1,\cdots,H}$ permiten decidir el número mínimo de dimensiones necesario para explicar dichas dependencias.

CONTRIBUCIONES TOTALES: Cuantifican la importancia de cada una de las modalidades de las variables analizadas en la construcción de los ejes factoriales construidos por el análisis de correspondencias (ACC):

$$\sum_{i=1}^{k}C_t(i) = \sum_{j=1}^{m}C_t(j) = 1 \quad \begin{cases} \text{contribución i-ésima fila: } C_t(i) = \dfrac{n_{i\bullet}a_{ih}^2}{\sum_{j=1}^{r} n_{j\bullet}a_{jh}^2} = \dfrac{n_{i\bullet}a_{ih}^2}{\mu_h^2} \\ \text{contribución i-ésima columna: } C_t(j) = \dfrac{n_{\bullet j}b_{jh}^2}{\sum_{i=1}^{m} n_{\bullet i}b_{ih}^2} = \dfrac{n_{\bullet j}b_{jh}^2}{\mu_h^2} \end{cases}$$

Se utilizan para interpretar el significado de los ejes utilizando, para cada uno de ellos, las modalidades con contribuciones más fuertes.

CONTRIBUCIONES RELATIVAS: Miden la importancia de cada factor para explicar la posición (en el diagrama cartesiano) de cada una de las modalidades de las variables analizadas, representando la parte de la distancia al origen de coordenadas, explicada por dicho factor. Vienen dadas por:

$$\sum_{i=1}^{k}C_t(i) = \sum_{j=1}^{m}C_t(j) = 1 \quad \begin{cases} \text{contribución i-ésima fila: } C_r(i) = \dfrac{a_{ih}^2}{\sum_{l=1}^{H} a_{il}^2} \\ \text{contribución i-ésima columna: } C_r(j) = \dfrac{b_{jh}^2}{\sum_{l=1}^{H} b_{jl}^2} \end{cases}$$

Se utilizan para analizar las proximidades entre los puntos, haciendo hincapié en aquellos factores cuyas contribuciones sean más elevadas cuando se desea explicar dichas proximidades.

ELEMENTOS SUPLEMENTARIOS: Son filas o columnas de la tabla de contingencia no utilizadas en el cálculo de los ejes factoriales pero que, una vez calculados éstos, se sitúan en el diagrama cartesiano con el fin de ayudar en la interpretación de los resultados obtenidos. Sus coordenadas se calculan utilizando las relaciones baricéntricas existentes entre los puntos fila y columna.

ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES

Se aplica a tablas de contingencias en donde por filas hay (n) individuos y por columnas (s) variables categóricas con $(p_i = 1, 2, \cdots, s)$ mutuamente excluyentes y exhaustivas.

La tabla de datos tiene la forma: $Z = \left[Z_1, Z_2, \cdots, Z_s\right]$, siendo Z_i una matriz $(n.\ p_i)$, de forma que:

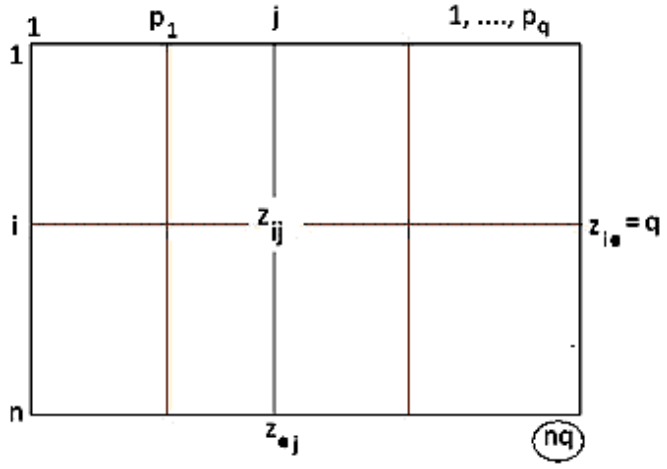
$$z_{ij} = \begin{cases} 1 & \text{si el individuo i-ésimo elige la modalidad j} \\ 0 & \text{si el individuo i-ésimo no elige la modalidad j} \end{cases}$$

El análisis de correspondencias múltiples se basa en realizar un análisis de correspondencias sobre la llamada matriz de Burt: $B = Z'Z$

La matriz de Burt se construye por superposición de cajas. En los bloques diagonales aparecen matrices diagonales conteniendo las frecuencias marginales de cada una de las variables analizadas. Fuera de la diagonal aparecen las tablas de frecuencias cruzadas correspondientes a todas las combinaciones 2 a 2 de las variables analizadas.

Se toman como dimensiones aquellas cuya distribución a la inercia supera (1/p).

ACM: ANÁLISIS FACTORIAL (AFC) DE UNA TABLA DISYUNTIVA



$n \equiv$ individuos

$q \equiv$ variables cualitativas

$p_k \equiv$ modalidades de la variable k

$$p = \sum_{k=1}^{q} p_k \equiv \text{total de columnas}$$

$$Z = (z_{ij}) \quad z_{ij} = \begin{cases} 1 \\ 0 \end{cases}$$

Márgenes $\begin{cases} z_{i\bullet} \equiv q \text{ pues hay q unos en la fila i-ésima} \\ z_{\bullet j} \equiv \text{individuos que tienen modalidad j-ésima} \end{cases}$

$$\sum_j \sum_i z_{ij} = nq$$

Matriz a diagonalizar: $S = \frac{1}{q} Z' Z D^{-1} = \frac{1}{q} B D^{-1}$ con $D = \text{diagonal}(z_{\bullet j})$

ACM: NUBE DE PUNTOS, PERFILES

- Perfiles- fila (individuos)

Puntos (coordenadas): $\frac{z_{ij}}{z_{i\bullet}} = \frac{z_{ij}}{q}$ Tabla : $\frac{1}{q} Z$

Peso: $\frac{z_{i\bullet}}{nq} = \frac{q}{nq} = \frac{1}{n}$

Métrica: diagonal $\left(\frac{nq}{z_{\bullet j}}\right)$

Distancia χ^2: $d^2(i, i') = \frac{n}{q} \sum_{j=1}^{p} (z_{ij} - z_{i'j})^2$

- Perfiles-columna (modalidades)

Puntos (coordenadas): $\frac{z_{ij}}{z_{\bullet j}}$

Peso: $\frac{z_{\bullet j}}{nq}$

Métrica: diagonal $\left(\frac{nq}{z_{i\bullet}}\right) = \text{diagonal}\left(\frac{nq}{q}\right) = \text{diagonal}\,(n)$

Distancia χ^2: $d^2(j, j') = n \sum_{j=1}^{p} \left[\frac{z_{ij}}{z_{\bullet j}} - \frac{z_{ij'}}{z_{\bullet j'}}\right]^2$

- Perfiles de fila: Más parecidos si tienen más modalidades en común.

$d^2(i, i') = \frac{n}{q} \sum_{j=1}^{p} (z_{ij} - z_{i'j})^2 = \frac{n}{q} \sum_{j \in M_{ii'}} \frac{1}{z_{\bullet j}}$ con $M_{ii'}$: modalidades que tiene solo un individuo i o i'.

- Perfiles de columna: Entre más objetos tengan sólo una de j o j' mayor es la distancia.

$$d^2(j, j') = n \sum_{j=1}^{p} \left[\frac{z_{ij}}{z_{\bullet j}} - \frac{z_{ij'}}{z_{\bullet j'}}\right]^2 = n \frac{\text{card}\left[\text{individuos } (j, \text{ no } j')\right] + \text{card}\left[\text{individuos } (j', \text{ no } j)\right]}{z_{\bullet j} \cdot z_{\bullet j'}}$$

INTERPRETACIÓN

- Dos modalidades escogidas por los mismos individuos coinciden
- Dos individuos son cercanos si escogen las mismas modalidades
- Modalidades con poco efectivo están alejadas del centro de gravedad

ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES (ACM): INERCIA

- Centro de gravedad de nube de modalidades: $G = \left(\frac{1}{n}, \cdots, \frac{1}{n}\right)$

$$d^2(j, G) = n \sum_{j=1}^{p} \left[\frac{z_{ij}}{z_{\bullet j}} - \frac{1}{n}\right]^2 = n \sum_{j=1}^{p} \left[\frac{z_{ij}}{z_{\bullet j}^2} - \frac{2\, z_{ij}}{n\, z_{\bullet j}} + \frac{1}{n^2}\right] = \frac{n}{z_{\bullet j}} - 1$$

La distancia es mayor si el efectivo es pequeño.

- Inercia de la modalidad j: $I(j) = \frac{z_{\bullet j}}{nq} d^2(j, G) = \frac{z_{\bullet j}}{nq}\left(\frac{n}{z_{\bullet j}} - 1\right) = \frac{1}{q}\left(1 - \frac{z_{\bullet j}}{n}\right)$

Hay mayor inercia si el efectivo es pequeño.

- Inercia de la variable k: $I(k) = \sum_{j=1}^{p_k} I(j) = \sum_{j=1}^{p_k} \frac{1}{q}\left(1 - \frac{z_{\bullet j}}{n}\right) = \frac{1}{q}(p_k - 1)$

La inercia crece con el número de modalidades. Si $p_k = 2 \rightarrow$ mínimo $\frac{1}{q}$

- Inercia total: $I = \sum_k I(k) = \sum_k \frac{1}{q}(p_k - 1) = \frac{1}{q}(p - q) = \frac{p}{q} - 1$

No tiene significado estadístico.

SOLUCIÓN DE ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES

- Diagonalización:

$$\text{En } \Re^p: \ \frac{1}{q} Z' Z D^{-1} u_\alpha = \lambda_\alpha u_\alpha$$

$$\text{factor } \phi_\alpha = D^{-1} u_\alpha \quad \Rightarrow \quad \frac{1}{q} D^{-1} Z' Z \phi_\alpha = \lambda_\alpha \phi_\alpha$$

$$\text{En } \Re^n: \ \frac{1}{q} Z D^{-1} Z' \psi_\alpha = \lambda_\alpha \psi_\alpha$$

- Relaciones de transición:

$$\begin{cases} \phi_\alpha = \dfrac{1}{\sqrt{\lambda_\alpha}} D^{-1} Z' \psi_\alpha \\ \psi_\alpha = \dfrac{1}{q\sqrt{\lambda_\alpha}} Z \phi_\alpha \end{cases}$$

- Relaciones baricéntricas:

$P(i) \equiv$ modalidades que tiene i
$I(j) \equiv$ modalidades que tiene j

$$\begin{cases} \psi_\alpha = \dfrac{1}{\sqrt{\lambda_\alpha}} \displaystyle\sum_{j=1}^{p} \frac{z_{ij}}{z_{i\bullet}} \phi_{\alpha j} = \frac{1}{q\sqrt{\lambda_\alpha}} \sum_{j \in P(i)} \phi_{\alpha j} \\ \phi_{\alpha j} = \dfrac{1}{\sqrt{\lambda_\alpha}} i = 1 \displaystyle\sum_{j=1}^{n} \frac{z_{ij}}{z_{\bullet j}} \psi_{\alpha i} = \frac{1}{z_{\bullet j}\sqrt{\lambda_\alpha}} \sum_{j \in I(j)} \psi_{\alpha i} \end{cases}$$

INTERPRETACIÓN ANÁLISIS CORRESPONDENCIAS MÚLTIPLES

- Proximidad entre individuos en términos de parecido: Dos individuos se parecen si tienen casi las mismas modalidades. Es decir, dos individuos están próximos se han elegido globalmente las mismas modalidades.
- Proximidad entre modalidades de variables diferentes en términos de asociación: Son cercanos puesto que globalmente están presentes en los mismos individuos. Es decir, dos modalidades están próximas si han sido elegidas globalmente por el mismo conjunto de individuos.
- Proximidad entre modalidades de una misma variable en términos de parecido: (a) Son excluyentes por construcción. (b) Si son cercanas es porque los individuos que las poseen presentan casi el mismo comportamiento en las otras variables.

Sea la tabla formada por 10 individuos de una empresa, que se distribuyen según el género, los años en la empresa y los ingresos obtenidos:

Individuos	Género	Años	Ingreso
1	Mujer	5	Medio
2	Mujer	3	Alto
3	Hombre	4	Bajo
4	Mujer	1	Bajo
5	Mujer	2	Medio
6	Hombre	5	Alto
7	Mujer	2	Medio
8	Hombre	3	Bajo
9	Hombre	1	Alto
10	Mujer	4	Medio

A partir de la tabla original se construye la tabla disyuntiva (matriz Z) con tantas columnas como categorías:

Género		Años					Ingresos		
Mujer	Hombre	1	2	3	4	5	Bajo	Medio	Alto
1	0	0	0	0	0	1	0	1	0
1	0	0	0	1	0	0	0	0	1
0	1	0	0	0	1	0	1	0	0
1	0	1	0	0	0	0	1	0	0
1	0	0	1	0	0	0	0	1	0
0	1	0	0	0	0	1	0	0	1
1	0	0	1	0	0	0	0	1	0
0	1	0	0	1	0	0	1	0	0
0	1	1	0	0	0	0	0	0	1
1	0	0	0	0	1	0	0	1	0

En la tabla disyuntiva completa (matriz Z), si hay alguna variable continua, debe transformarse en nominal, ordenándose en intervalos a los que se da un rango de valores.

- Las frecuencias marginales de las líneas de la tabla disyuntiva completa son iguales al número de preguntas (s), y las frecuencias marginales de las columnas corresponden al número de sujetos que han elegido la modalidad (j) de la pregunta (q), por lo que para cada subtabla, el número total de individuos es n.
- En consecuencia, si para n individuos se dispone de respuestas respecto a, por ejemplo, dos variables nominales que tienen respectivamente p1 y p2 modalidades, entonces es equivalente someter a un Análisis de Correspondencias Simples la tabla de contingencia (p1, p2) y analizar la tabla binaria de n líneas y (p1 + p2) columnas que describe las respuestas.

se tiene,

$$
Z = \begin{array}{c|c}
 & \begin{array}{cccccccccc} \text{Género} & & \text{Años} & & & & & \text{Ingresos} & & \\ M & H & 1 & 2 & 3 & 4 & 5 & B & M & A \end{array} \\
\begin{array}{c} \rightarrow \\ \text{Género} \\ \rightarrow \\ \\ \text{Años} \\ \\ \mapsto \\ \\ \text{Ingresos} \\ \mapsto \end{array} \begin{array}{c} M \\ H \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ B \\ M \\ A \end{array} &
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0
\end{bmatrix}
\end{array}
$$

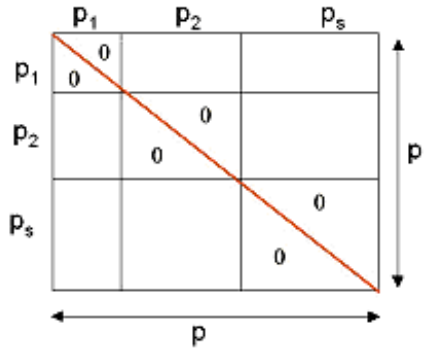
$$
Z' = \begin{array}{c|c}
 & \begin{array}{cccccccccc} \text{Género} & & \text{Años} & & & & & \text{Ingresos} & & \\ M & H & 1 & 2 & 3 & 4 & 5 & B & M & A \end{array} \\
\begin{array}{c} \rightarrow \\ \text{Género} \\ \rightarrow \\ \\ \text{Años} \\ \\ \mapsto \\ \\ \text{Ingresos} \\ \mapsto \end{array} \begin{array}{c} M \\ H \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ B \\ M \\ A \end{array} &
\begin{bmatrix}
1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0
\end{bmatrix}
\end{array}
$$

Relacionando cada variable con todas las demás la tabla disyuntiva se convierte a una "tabla de Burt" que contiene todas las tablas de contingencia simples entre las variables (cruzadas dos a dos).

A partir de la tabla disyuntiva completa se puede construir la tabla de contingencia de Burt (B), que es una tabla simétrica de orden (p, p): $B = Z'.Z$

B es una yuxtaposición de tablas de contingencia y está formada de s^2 bloques de la forma:

$B = Z'Z =$



Cada bloque es una submatriz formada por tablas de contingencia de las variables dos a dos, salvo los bloques que se están en la diagonal que son las tablas de contingencia de cada variable consigo misma.

con lo cual,

MATRIZ DE BURT

Z'.Z =		Género		Años					Ingresos		
		M	H	1	2	3	4	5	B	M	A
Género	M	6	0	1	2	1	1	1	1	4	1
	H	0	4	1	0	1	1	1	2	0	2
Años	1	1	1	2	0	0	0	0	1	0	1
	2	2	0	0	2	0	0	0	0	2	0
	3	1	1	0	0	2	0	0	1	0	1
	4	1	1	0	0	0	2	0	1	1	0
	5	1	1	0	0	0	0	2	0	1	1
Ingresos	B	1	2	1	0	1	1	0	3	0	0
	M	4	0	0	2	0	1	1	0	4	0
	A	1	2	1	0	1	0	1	0	0	3

La tabla disyuntiva completa es equivalente a la tabla de Burt y ambos producen los mismos factores.

Con la tabla de Burt se podrán obtener las puntuaciones (distancias al centro de gravedad), contribuciones absolutas de cada modalidad y variable a los ejes o factores obtenidos (contribución de cada modalidad o variable a la inercia de los nuevos ejes) y contribuciones relativas o correlaciones de cada modalidad con los nuevos ejes.

Como en la tabla de Burt las filas y las columnas representan las mismas modalidades, el estudio de ambas ofrece iguales resultados, por lo que sólo se representan los de las filas.

Para la resolución en el SPSS (Analizar / Reducción de datos / Escalonamiento óptimo), en primer lugar hay que introducir los datos de la misma forma que en el análisis de correspondencias simples, creando por tanto tres variables. Las dos primeras corresponden a los rangos de todas las modalidades y, en la tercera, se incluirán las frecuencias conjuntas de los pares de modalidades, las cuales funcionarán como ponderaciones.

EXAMEN DE LOS PUNTOS:

- Las distancias de las modalidades, mientras más alejadas se encuentren del origen, mejor representados estarán. Cuanto más alejadas estén las modalidades entre sí en el gráfico menor asociación existirá entre ellas y cuanto más cercanas, más asociación existirá entre ellas.
- La contribución de los puntos a la inercia de cada dimensión o contribución de cada una de las filas a la inercia o varianza explicada en cada uno de los ejes considerados
- La contribución de las dimensiones a la inercia de cada punto. Se refiere a la correlación existente entre cada uno de los caracteres y los nuevos ejes.

LÓGICA DEL ANÁLISIS DE CORRESPONDENCIAS (AC)

Antes de comenzar con una exposición detallada conviene precisar que el objetivo en estudio es la utilidad práctica e interpretativa que se deriva de su aplicación.

Partiendo de esta concepción se trata de mostrar la lógica del análisis de correspondencias utilizando una supuesta investigación comercial (productos-limpieza.sav).

Un fabricante de producto de limpieza desea conocer las características más importantes de cada uno de sus productos, para ello realiza una encuesta entre sus clientes.

Introducción de datos para mostrar las frecuencias de asociación de cada producto con las características analizadas.

El análisis de correspondencias sintetiza la información de partida (espacio original generado a partir de la representación o proyección de la nube de puntos de las categorías de las variables) en un número reducido de dimensiones o factores (subespacio dimensional que al presentarse como una simplificación de la nube de puntos original permite identificar las similitudes entre las categorías).

Si los factores sintetizan la información original, resulta obvio suponer que éstos estarán representados por aquellas categorías que originalmente tuvieron más protagonismo, más frecuencias o más masa (Perfiles de fila y columna): A mayor masa, mayor será la importancia relativa de la categoría correspondiente.

Cada masa resulta, pues, de una ponderación de manera que aquellas categorías que se presentan con mayor frecuencia, mayor peso o masa, influirán más en la dirección del eje.

Una vez representada la nube de puntos ponderados (a partir de su masa) hay que valorar si éstos se encuentran concentrados o dispersos.

Si las variables son muy dependientes, las distancias serán muy grandes (filas y columnas distintas). Cuando son independientes, los puntos de la nube aparecerán muy agrupados (filas y columnas parecidas).

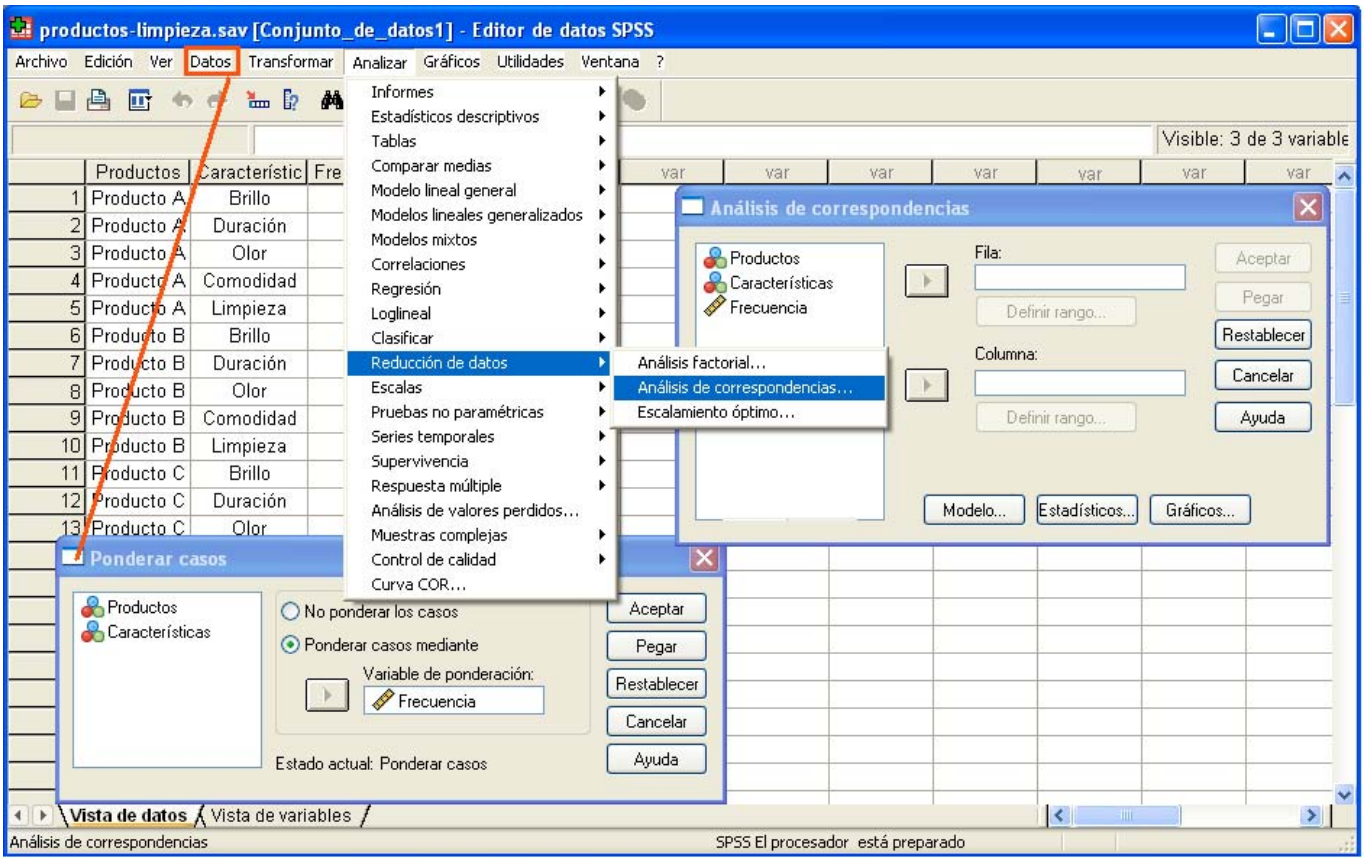
Cuando mayor sea la inercia total más dependientes serán las variables y más sentido tendrá identificar qué o cuáles categorías (una vez identificados los factores) son las que participan con mayor protagonismo en esa dependencia.

La representación gráfica de las correspondencias ayudan a detectar qué categorías se parecen (si el gráfico pertenece a una sola de las variables) o qué categorías están relacionadas entre sí (para gráficos con dos variables).

Cuanto mayor sea la distancia respecto al origen, mayor será la tendencia de la categoría en cuestión a concentrar su frecuencia en determinada celda (tanto más cuanto mayor sea su distancia al origen) y, en consecuencia muy poco en el resto.

Pese a que las representaciones gráficas facilitan identificar categorías similares, y dado que el tamaño del gráfico es siempre el mismo indistintamente de las puntuaciones factoriales, nada dicen acerca del grado de dicha similitud. Para superar este inconveniente se puede solicitar en el botón Estadísticos la opción Inspección de los puntos de fila y columna.

ANÁLISIS DE CORRESPONDENCIAS: Analizar / Reducción de datos / Análisis de correspondencias

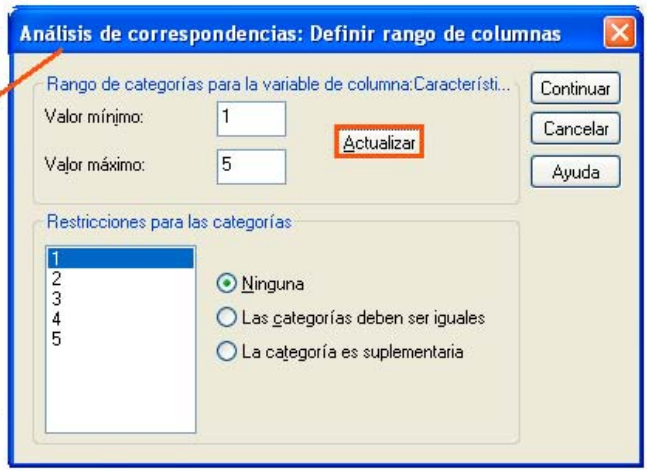


En la parte inferior del cuadro de diálogo aparecen una serie de Restricciones para las categorías que permiten unir las categorías o definir una categoría como suplementaria.

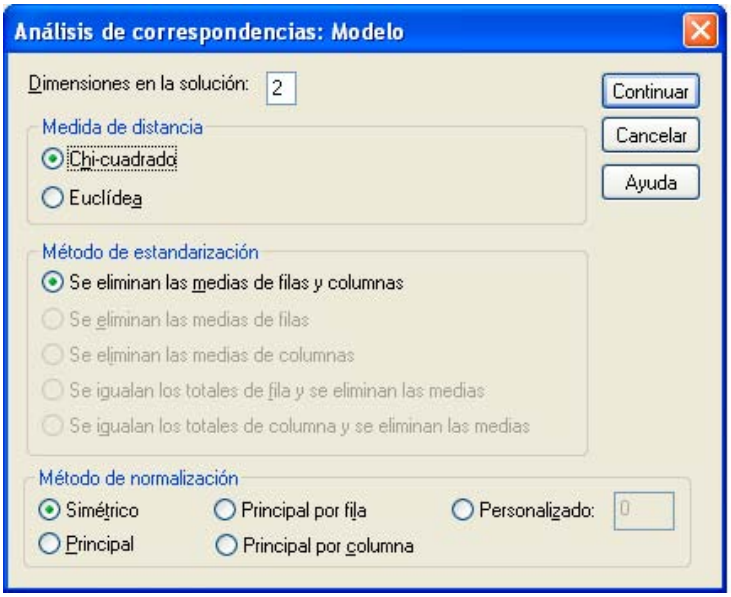
La reducción del número de categorías activas que se origina al unir categorías o definir determinadas categorías como suplementarias está sujeta a ciertas restricciones:

(a) El número máximo de categorías que pueden unirse equivale al total de categorías menos 1.

(b) El número máximo de categorías suplementarias es el total de categorías menos 2.



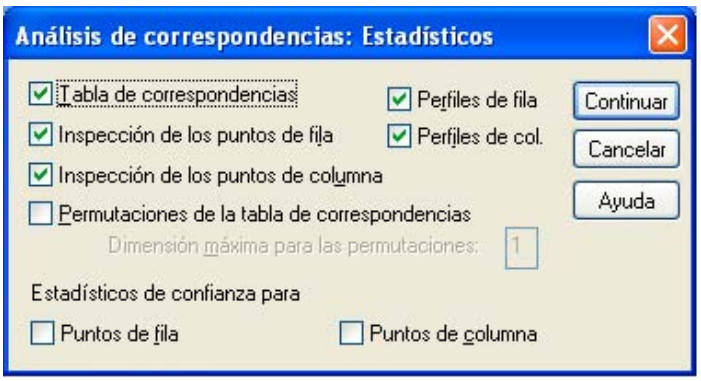
En este ejemplo, con dos variables, respectivamente, con 4 y 5 categorías cada variable, el número máximo será (4 - 1 = 3). Sin embargo, el investigador deberá especificar el menor número de factores para explicar el máximo de inercia, buscando siempre el equilibrio entre parsimonia e interpretabilidad.



El submenú Modelo se encuentra dividido en cuatro partes (número de dimensiones, medida de distancia, método de estandarización y normalización)

Dimensiones en la solución: Número de dimensiones necesarias para explicar la mayor parte de la variación.

El número máximo de factores es igual al número de filas menos 1 (o al de columnas menos 1), el que sea menor.



Botón Estadísticos.

La tabla adjunta muestra las frecuencias de asociación de cada producto con las características analizadas.

Tabla de correspondencias

Productos	Características					
	Brillo	Duración	Olor	Comodidad	Limpieza	Margen activo
Producto A	68	60	67	45	35	275
Producto B	12	13	13	15	12	65
Producto C	95	84	94	63	49	385
Producto D	30	32	32	36	28	158
Margen activo	205	189	206	159	124	883

Con el objetivo de conocer la relación entre los productos y las características el análisis de correspondencias convierte la tabla de contingencia en dos nubes de puntos de puntos fila y columna, para posteriormente realizar una representación de cada nube que permita detectar las relaciones entre las filas (productos), las columnas (características) y filas y columnas conjuntamente.

Para realizar esto el análisis de correspondencias no trabaja directamente con los datos de la tabla, puesto que su objetivo no es detectar las diferencias absolutas existentes entre las valoraciones de cada producto, sino que realiza unas transformaciones de los datos y los convierte en perfiles de filas y columnas.

Perfiles de fila

Productos	Características					
	Brillo	Duración	Olor	Comodidad	Limpieza	Margen activo
Producto A	,247	,218	,244	,164	,127	1,000
Producto B	,185	,200	,200	,231	,185	1,000
Producto C	,247	,218	,244	,164	,127	1,000
Producto D	,190	,203	,203	,228	,177	1,000
Masa	,232	,214	,233	,180	,140	

Los Perfiles de fila se calculan dividiendo el número de personas que eligen una determinada característica de un producto entre todos los que utilizan ese producto, obteniendo la distribución condicional de las características (columnas) dentro de cada fila (producto):

De las 275 personas que utilizaron el producto A, 68 consideran que deja la superficie brillante, el ratio (68/275 = 0,247) proporciona un perfil de 0,247.

El ratio (205/883 = 0,232) proporciona el perfil medio de brillo de 0,232.

Perfiles de columna

Productos	Características					
	Brillo	Duración	Olor	Comodidad	Limpieza	Masa
Producto A	,332	,317	,325	,283	,282	,311
Producto B	,059	,069	,063	,094	,097	,074
Producto C	,463	,444	,456	,396	,395	,436
Producto D	,146	,169	,155	,226	,226	,179
Margen activo	1,000	1,000	1,000	1,000	1,000	

Análogamente, los Perfiles de columna. En este caso, de las 205 elecciones recibidas de Brillo, 68 corresponden al producto A: El ratio (68/205 = 0,332) proporciona un perfil columna de 0,332

PROBLEMA DE LA DITANCIA: Cuando dos productos tienen varias características en la misma proporción, poseen el mismo perfil, se consideran similares (productos A y C). Lo mismo sucede con las características (columnas).

Frente a la sencillez de la lectura basada en los perfiles, los valores absolutos presentan mayor complicación.

La problemática de los números absolutos consiste en que los productos muy utilizados consiguen numerosas elecciones en todas las celdillas (producto C), mientras que los productos pocos utilizados tienen celdas con valores reducidos (producto B).

Al calcular las diferencias entre los productos utilizando los valores absolutos el resultado obtenido plasma la diferencia ente la distribución de las características, unida a la diferencia entre el número de elecciones de cada producto. Esto es, el mayor número de elecciones del producto C dificulta realizar una correcta comparación con el producto A.

En este sentido, el objetivo del análisis de correspondencias no es detectar las diferencias absolutas entre las valoraciones de los productos.

Habiendo detectado cómo trabajar con perfiles facilita la interpretación, hay que prevenir no producir una visión equivocada de la relación entre variables en la medida que todos los puntos tienen la misma importancia: los marginales de los perfiles de fila y columna son iguales a 1.

Para evitar este problema el análisis de correspondencias debe utilizar una distancia que no olvide las diferencias entre los efectivos de cada línea (o columna).

La distancia Chi-cuadrado cumple esta condición, al ponderar cada perfil por un peso. Así cada fila (o columna) está ponderada por un peso proporcional a su importancia real en el conjunto, peso conocido como "Masa".

Al considerar cada punto con una masa proporcional a su frecuencia se evita privilegiar las categorías con pocos efectivos.

De hecho, se trata de una distancia euclídea ponderada por el inverso de la masa de las columnas cuando se mide la distancia entre filas, o por la masa de las filas para la distancia entre las columnas.

La distancia Chi-cuadrado cumple también el principio de equivalencia distribucional, que postula que si dos categorías tienen perfiles idénticos pueden ser sustituidas por una sola categoría que sea la suma de sus pesos, sin que con ello se modifique la distancia entre las filas o columnas.

Esta propiedad garantiza la estabilidad de los resultados con independencia de la codificación de las variables, de modo que es posible agrupar categorías que tienen perfiles coincidentes tanto por filas como por columnas. Si el resultado se mantiene estable tras unir categorías, de igual forma estos resultados no mejoran al realizar más subdivisiones de categorías homogéneas.

Con los perfiles de fila y columna se elabora la matriz de coordenadas (distancias) utilizando la distancia Chi-cuadrado, que permitirá calibrar la magnitud de la diferencia entre la tabla de datos analizada y una tabla de datos sin relación entre las variables.

Las distancias no se miden entre dos filas o dos columnas sino con relación al perfil medio de filas o columnas o, dicho de otra manera, con relación al promedio de las coordenadas de esa fila (o columna) ponderada por su masa. Este perfil medio aparece situado en el origen de coordenadas y se le conoce como centro de gravedad.

La media de las distancias al cuadrado de cada punto de fila al centro de gravedad se conoce como "inercia de filas," o "inercia de columnas" cuando se trata de las columnas, e "inercia total " de la nube de *puntos* cuando se consideran todos los elementos de la tabla.

Una inercia baja significa que todos los puntos están situados muy cerca del centro de gravedad y que por tanto son muy similares; mientras que altos valores de la inercia en determinadas categorías implican grandes diferencias del perfil medio de las filas o las columnas.

Posteriormente se procede a diagonalizar la matriz de varianza-covarianza con el fin de extraer los vectores y valores propios que definirán los nuevos ejes sobre los que será proyectada la nube de puntos.

Cuando la cantidad de inercia explicada con los primeros factores sea alta bastará seleccionar un pequeño número de estos (tan sólo dos o tres) y representar la nube de puntos sobre gráficos de dos o tres dimensiones, obteniendo así una visión simplificada de las relaciones.

Trabajar con perfiles de fila (para comparar las distribuciones de las características en cada producto) y con perfiles de columna (para comparar cómo se distribuye cada característica en los productos) precisa de dos análisis diferentes: Uno sobre los perfiles de filas y otros sobre los perfiles de columnas, puesto que se consideran simétricos los papeles de las filas y de las columnas.

Al realizar estos análisis es posible utilizar unas fórmulas que relacionan a ambas dimensiones, conocidas como "fórmulas de transición", que permiten obtener las coordenadas factoriales del otro conjunto sin necesidad de una nueva diagonalización.

Además del ahorro de tiempo que esto supone al reducir los cálculos a una sola factorización, estas relaciones permiten representar sobre un mismo plano los puntos de fila y columna, permitiendo así interpretar la distancia de un punto a otro. De hecho, una de las grandes ventajas del análisis de correspondencias es la facilidad para sacar conclusiones basándose en la interpretación gráfica.

Debe considerarse que los factores o dimensiones extraídas no explican la totalidad de la inercia, ni tampoco cada una de ellas aportan lo mismo; de forma que será necesario extraer un número de factores que expliquen un porcentaje elevado de la variabilidad total, al tiempo que deberá interpretarse cada uno en relación con la cantidad de inercia explicada.

El análisis de las "contribuciones absolutas y relativas" indicará las modalidades más relacionadas con cada factor, mientras que el signo de las coordenadas situarán cada categoría en una parte del factor.

La "contribución absoluta" expresa la aportación de un elemento a la inercia explicada por el factor, mientras que la "contribución relativa " recoge la contribución de un factor a la explicación de una fila o columna.

PROCESO INTERPRETACIÓN DEL GRÁFICO:

1. Localización de las modalidades con mayores contribuciones absolutas, diferenciando mediante el signo de las coordenadas las que se sitúan en el lado positivo y en el lado negativo del factor.
2. Análisis de la calidad de representación (contribución relativa) del resto de modalidades. Cuando una modalidad tiene una baja contribución relativa es probable que esté muy relacionada con otro factor, de modo que para su estudio será conveniente considerar, sino la totalidad de los ejes, al menos un número elevado de éstos.
3. Búsqueda de aquellas modalidades que, aunque no contribuyen a la formación del factor, se encuentran bien representadas. Estas modalidades son ilustrativas de la significación de la dimensión.
4. Considerando todos estos elementos se procede con la denominación de cada factor, analizando por separado la variable fila y columna.
5. Se procede al análisis gráfico con el fin de detectar similitudes entre las modalidades de fila (o columna). Respecto a la situación de las modalidades en el gráfico hay que precisar que en el origen de coordenadas se encuentran las categorías similares a la media de las filas (o columnas), que son las que tienen menor tasa de inercia, y por tanto las que menos aportan en la definición de cada dimensión. Por otra parte, las modalidades más alejadas del origen se caracterizan por su gran contribución en la definición de cada factor. En el ejemplo de los productos de limpieza, si dos filas (productos) tienen características semejantes se situarán próximas una de la otra en el plano de coordenadas. La misma consideración se utiliza para interpretar las características (columnas) próximas. Resumiendo, mayor o menor proximidad entre las modalidades en el plano equivale a mayor o menor grado de relación o interdependencia entre las mismas.
6. Interpretación conjunta de ambas variables. Generalmente, pude decirse que dos modalidades de fila y columna con una situación cercana en el gráfico están indicando asociación entre ellas, mucho más cuando se encuentran lejos del centro de gravedad.

7. Proyección de modalidades ilustrativas o suplementarias. La interpretación del análisis puede enriquecerse con la representación gráfica de las modalidades suplementarias, elementos que no participan en la definición de los ejes pero que se "proyectan" sobre el gráfico obtenido.

Hasta este momento el análisis se ha centrado en una tabla bidimensional, aunque el efecto práctico y eficacia del análisis de correspondencias aumenta cuando se trabaja con grandes tablas de datos. Señalar que el análisis de correspondencias, en su formato más simple, trata de representar dos variables cualitativas que forman parte de una tabla de contingencia, aunque existe una generalización del análisis de correspondencias para más de dos variables categóricas que recibe el nombre de análisis de correspondencias múltiples.

La lógica del análisis y el proceso de cálculo es similar en ambos, de modo que se anotan aquí los aspectos diferenciados del análisis de correspondencias múltiples:

- En el análisis de correspondencias múltiples los valores propios generan una idea pesimista de la variabilidad explicada, siendo conveniente medir la *tasa de inercia* realizando una modificación de éstos utilizando la corrección de Benzécri (1979):
    1. Calcular B=1/Q , siendo Q el número de variables.
    2. Seleccionar los valores propios (VP) iguales o superiores a B.
    3. Calcular los valores propios transformados (VPT): $VPT = (VP - B)^2$.
    4. Calcular el porcentaje de varianza explicada (VPE) con los valores propios transformados. Cada valor propio tiene una tasa de inercia sobre el total de varianza explicada por todos los Valores Propios transformados.
    5. Calcular el porcentaje acumulado de varianza explicada.
- La parte de inercia debida a una modalidad de respuesta aumenta cuanto menor sea el número de personas de esta modalidad, cuanto menor sea su masa. Por ello es conveniente no introducir en el análisis las modalidades con escaso número de respuestas, y de hecho algunos programas estadísticos eliminan las modalidades con un número de respuestas inferior al 2% de la muestra.
- La parte de inercia producida por una variable aumenta cuanto mayor sea el número de opciones de respuesta, de modo que debe procurarse que las variables tengan un número homogéneo de categorías.
- El punto de partida es una "tabla disyuntiva "completa donde las categorías de respuesta de una pregunta se excluyen mutuamente, y una única modalidad es escogida obligatoriamente por el encuestado (o unidad con la que se trabaje).

En esta tabla, las filas están formadas por los individuos encuestados, y las columnas por cada una de las categorías de las variables sometidas al análisis, de modo que cada celda está formada por un 1 cuando el individuo posee una característica, y un 0 cuando no es así.

TABLA ORIGINAL

	P1	P2	P3
n1	1	1	1
n2	2	2	1
n3	3	1	2
n4	2	3	1
n5	1	2	2
n6	3	2	2

En la tabla original presentada se ha dejado una única columna para las categorías de cada pregunta, mientras que la tabla disyuntiva tantas columnas como categorías.

Así las respuestas del segundo sujeto (2, 2, 1) son codificadas en la tabla disyuntiva como (0, 1, 0) en P1, (0, 1, 0) en P2, y (1, 0) en P3.

TABLA DISYUNTIVA COMPLETA

	P1			P2			P3		
Categorías	1	2	3	1	2	3	1	2	Total
n1	1	0	0	1	0	0	1	0	3
n2	0	1	0	0	1	0	1	0	3
n3	0	0	1	1	0	0	0	1	3
n4	0	1	0	0	0	1	1	0	3
n5	1	0	0	0	1	0	0	1	3
n6	0	0	1	0	1	0	0	1	3
Total	2	2	2	2	2	2	2	2	

Relacionando cada variable con todas las demás la tabla disyuntiva se convierte a una *tabla de Burt* que contiene todas las tablas de contingencia simples entre las variables (cruzadas dos a dos).

TABLA DE BURT

		P1			P2			P3	
	categorías	1	2	3	1	2	3	1	2
P1	n1	2	0	0	1	1	0	1	1
	n2	0	2	0	0	1	1	2	0
	n3	0	0	2	1	1	0	0	2
P2	n1	1	0	1	2	0	0	1	1
	n2	1	1	1	0	3	0	1	2
	n3	0	1	0	0	0	1	1	0
P3	n1	1	2	0	1	1	1	3	0
	n2	1	0	2	1	2	0	0	3

- El valor 2 de la esquina superior izquierda son las personas que han elegido el valor 1 en P1 (pregunta 1). En las dos primeras tablas estas personas han sido identificadas con n1 y n5.
- En la pregunta 2 (P2), el 1 de la izquierda representa a la persona que ha elegido 1 en P1 y el 1 en P2, se trata de la persona n1.

- En la pregunta 2, el 1 del centro es el 1 que eligió el 1 en P1 y el 2 en P2, persona n5.
- En la pregunta 2, el 1 de la segunda línea representa a la persona que ha elegido 2 en P1 y 3 en P2, identificada con n4.

O también:

TABLA DISYUNTIVA COMPLETA

	P1			P2			P3		
Categorías	1	2	3	1	2	3	1	2	Total
n1	1	0	0	1	0	0	1	0	3
n2	0	1	0	0	1	0	1	0	3
n3	0	0	1	1	0	0	0	1	3
n4	0	1	0	0	0	1	1	0	3
n5	1	0	0	0	1	0	0	1	3
n6	0	0	1	0	1	0	0	1	3
Total	2	2	2	2	2	2	2	2	

$$Z = \begin{array}{c} \begin{array}{ccc} \text{P1} & \text{P2} & \text{P3} \\ 1\ \ 2\ \ 3 & 1\ \ 2\ \ 3 & 1\ \ 2 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \end{array} \qquad Z' = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ & & & & & \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ & & & & & \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

$$Z' \,.\, Z = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ & & & & & \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ & & & & & \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} =$$

MATRIZ DE BURT

$$
= \begin{array}{cc}
 & \begin{array}{ccccccccc} & P1 & & & P2 & & & P3 & \\ & 1 & 2 & 3 & 1 & 2 & 3 & 1 & 2 \end{array} \\
\begin{array}{c} P1 \\ \\ \\ P2 \\ \\ \\ P3 \\ \\ \end{array} \begin{array}{c} n1 \\ n2 \\ n3 \\ n1 \\ n2 \\ n3 \\ n1 \\ n2 \end{array} &
\left[\begin{array}{ccc|ccc|cc}
2 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\
0 & 2 & 0 & 0 & 1 & 1 & 2 & 0 \\
0 & 0 & 2 & 1 & 1 & 0 & 0 & 2 \\
\hline
1 & 0 & 1 & 2 & 0 & 0 & 1 & 1 \\
1 & 1 & 1 & 0 & 3 & 0 & 1 & 2 \\
0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\
\hline
1 & 2 & 0 & 1 & 1 & 1 & 3 & 0 \\
1 & 0 & 2 & 1 & 2 & 0 & 0 & 3
\end{array}\right]
\end{array}
$$

Total categoría

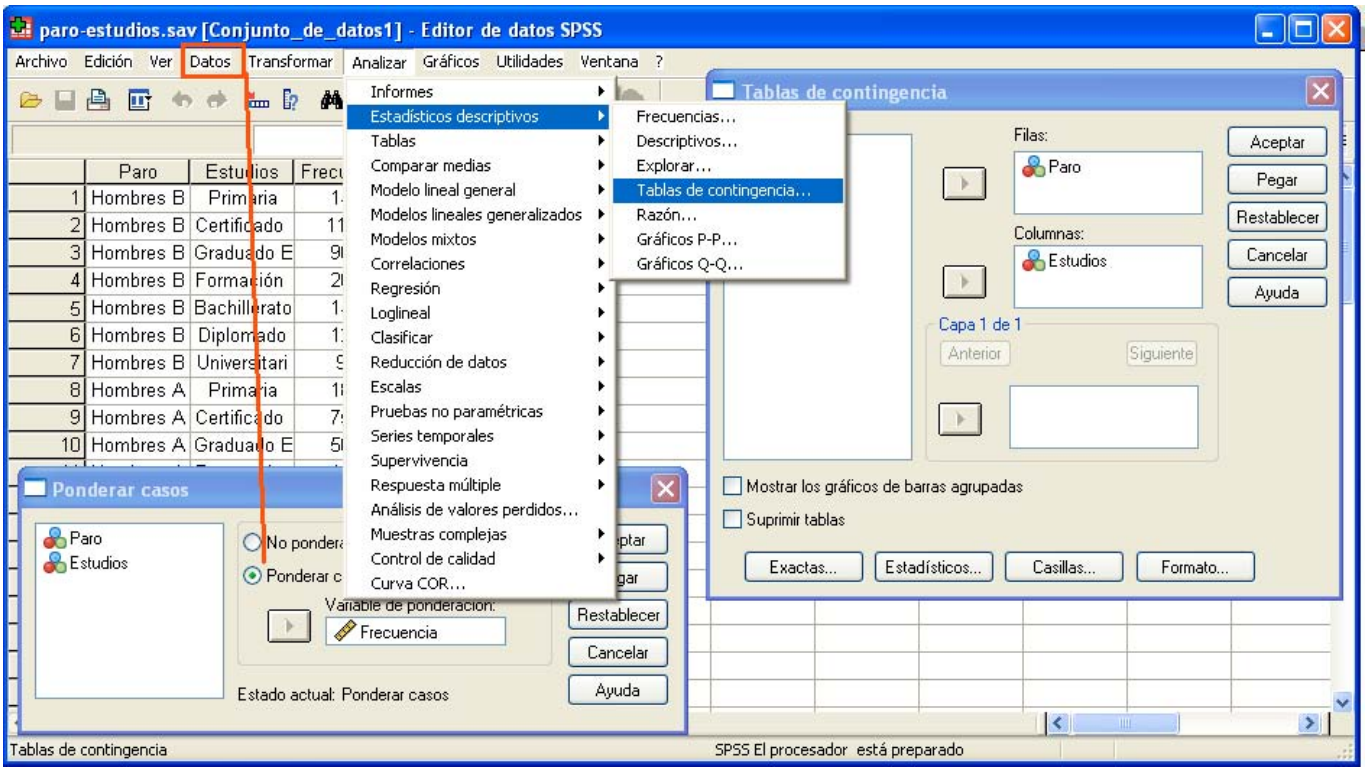
La tabla disyuntiva completa es equivalente a la tabla de Burt y ambas producen los mismos factores.

CORRESPONDENCIAS CON SPSS (paro-estudios.sav)

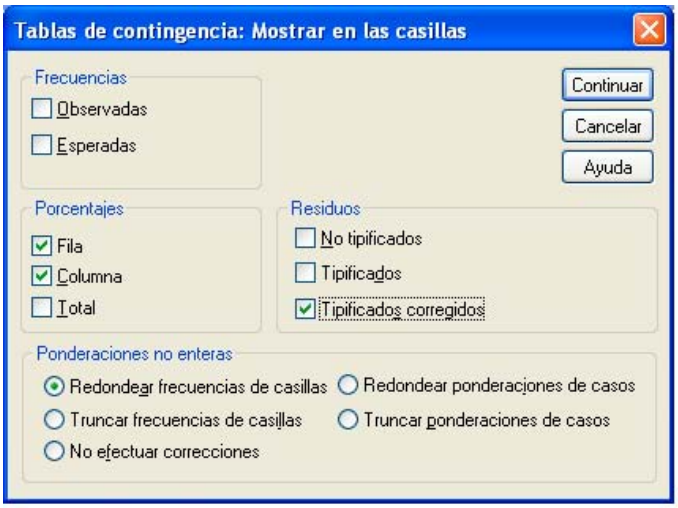
Los datos corresponden a la distribución del número de parados de tres provincias de Castilla-León en el año 2005 clasificados por Sexo, Provincia y Nivel de Estudios

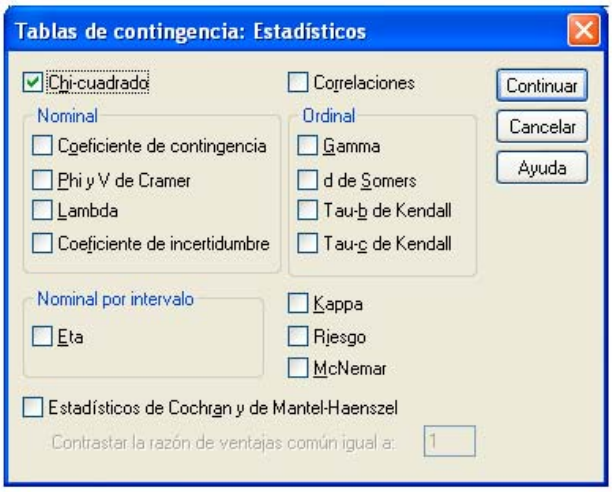
PRIMER ANÁLISIS DE LA INFORMACIÓN: Analizar / Estadísticos descriptivos / Tablas de contingencia

Se comienza con un análisis univariado de las variables consideradas, con la finalidad de identificar determinados patrones de comportamiento.



Se solicitan porcentajes de Fila y Columna, Residuos estandarizados corregidos y Pruebas de Chi-cuadrado.





% de Paro

Tabla de contingencia Paro * Estudios

		Estudios							Total
		Primaria	Certificado Escolaridad	Graduado Escolar	Formación Profesional	Bachillerato	Diplomado	Universitario	
Paro	Hombres Burgos	5,2%	39,8%	32,3%	9,5%	5,3%	4,5%	3,3%	100,0%
	Hombres Avila	9,8%	40,6%	30,5%	5,8%	7,5%	2,7%	3,1%	100,0%
	Hombres Soria	2,4%	37,2%	32,3%	11,3%	8,0%	3,8%	5,0%	100,0%
	Mujeres Burgos	1,7%	21,0%	38,3%	13,1%	9,7%	10,7%	5,5%	100,0%
	Mujeres Avila	2,1%	19,6%	41,4%	10,6%	12,2%	9,5%	4,7%	100,0%
	Mujeres Soria	,8%	23,0%	36,6%	12,6%	11,1%	8,5%	7,3%	100,0%
Total		2,0%	28,7%	35,2%	11,8%	9,7%	6,8%	6,0%	100,0%

Al comparar los porcentajes de fila se observa que un 9,8% de los Hombres de Avila parados tienen un nivel de estudios primario (cifra que contrasta a la correspondiente distribución marginal en la que únicamente un 2% de los parados poseen dicho nivel de estudios), mientras que en Mujeres de Soria paradas un 0,8% tienen estudios primarios.

% de Estudios

Tabla de contingencia Paro * Estudios

		Estudios							Total
		Primaria	Certificado Escolaridad	Graduado Escolar	Formación Profesional	Bachillerato	Diplomado	Universitario	
Paro	Hombres Burgos	13,6%	7,1%	4,7%	4,1%	2,8%	3,4%	2,9%	5,1%
	Hombres Avila	16,9%	4,8%	2,9%	1,7%	2,6%	1,3%	1,8%	3,4%
	Hombres Soria	38,5%	41,5%	29,4%	30,9%	26,6%	17,8%	26,7%	32,0%
	Mujeres Burgos	6,7%	5,7%	8,5%	8,7%	7,8%	12,2%	7,2%	7,8%
	Mujeres Avila	5,3%	3,4%	5,8%	4,4%	6,2%	6,9%	3,9%	4,9%
	Mujeres Soria	18,9%	37,6%	48,7%	50,2%	54,0%	58,3%	57,6%	46,8%
Total		100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Al comparar los porcentajes columna, por ejemplo, se observa que un 16,9% de los parados con un nivel de estudios primario son Hombres de Ávila (cifra mucho más elevada que la correspondiente a la distribución marginal en la que tan sólo un 3,4% son Hombres de Ávila).

Para interpretar el interior de la tabla, desde un punto de vista práctico, es más sencillo utilizar los Residuos estandarizados corregidos.

Residuos corregidos

Tabla de contingencia Paro * Estudios

		Estudios						
		Primaria	Certificado Escolaridad	Graduado Escolar	Formación Profesional	Bachillerato	Diplomado	Universitario
Paro	Hombres Burgos	12,9	13,4	-3,3	-3,8	-8,0	-5,0	-6,0
	Hombres Avila	24,9	11,5	-4,3	-8,0	-3,3	-7,2	-5,2
	Hombres Soria	4,7	30,3	-9,6	-2,1	-8,9	-19,3	-6,7
	Mujeres Burgos	-1,4	-11,6	4,5	2,8	,1	10,5	-1,3
	Mujeres Avila	,5	-10,7	7,0	-2,0	4,5	5,7	-2,9
	Mujeres Soria	-18,5	-27,5	6,6	5,8	11,0	14,7	12,8

En los residuos estandarizados corregidos la mayor parte son mayores que 2 en valor absoluto (a un nivel del 95% de confianza, residuos con un valor absoluto mayor que 2 se consideran como valores anormalmente altos).

Observando, además, el patrón de los signos:

Los residuos positivos para los hombres tienden a situarse en los niveles de estudios más bajos (estudios primarios y certificado escolar) y para las mujeres en los niveles de estudios superiores (graduado escolar, bachillerato y diplomados en todas las provincias; formación profesional en Burgos y Soria, y universitario en Soria).

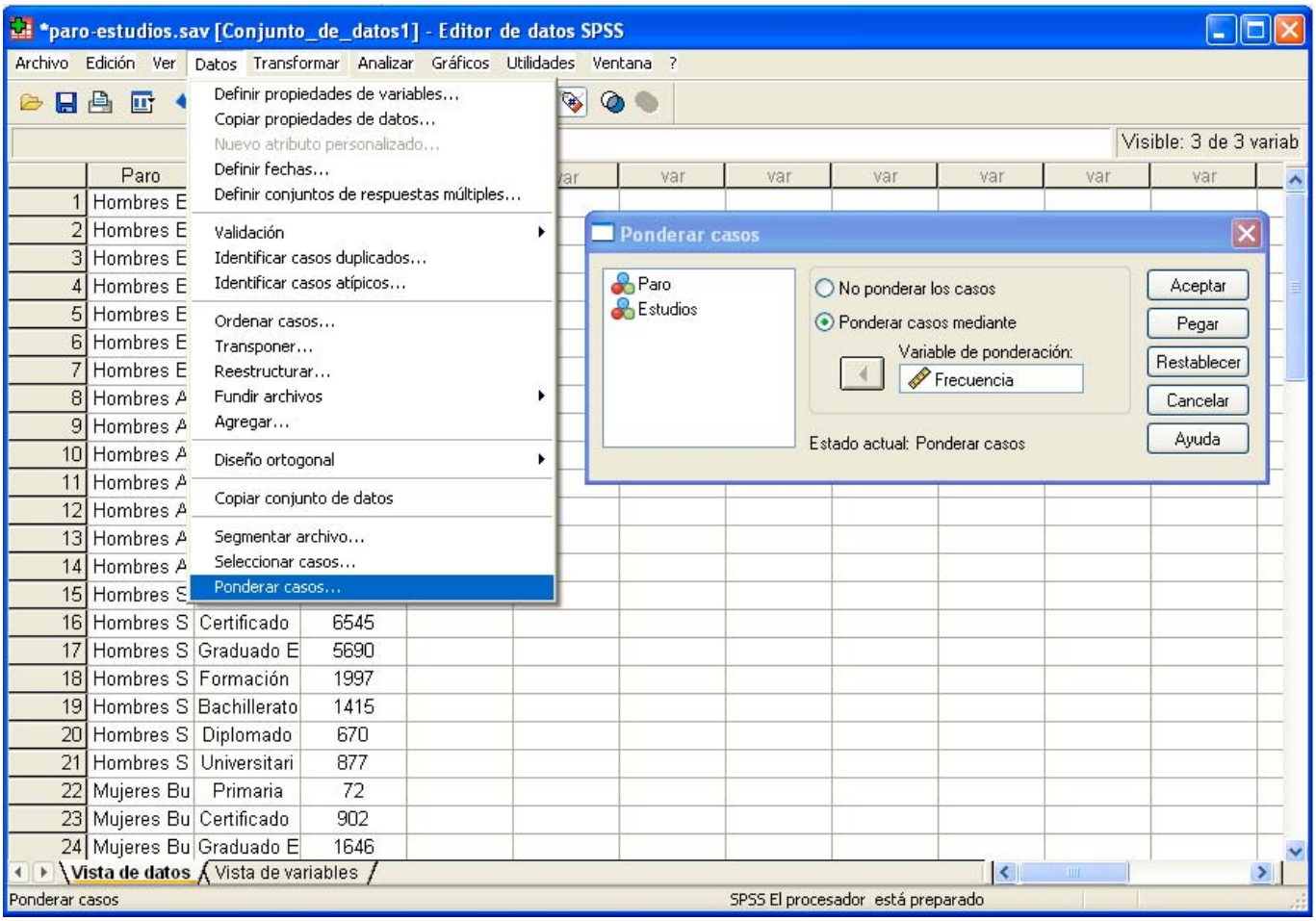
Pruebas de chi-cuadrado

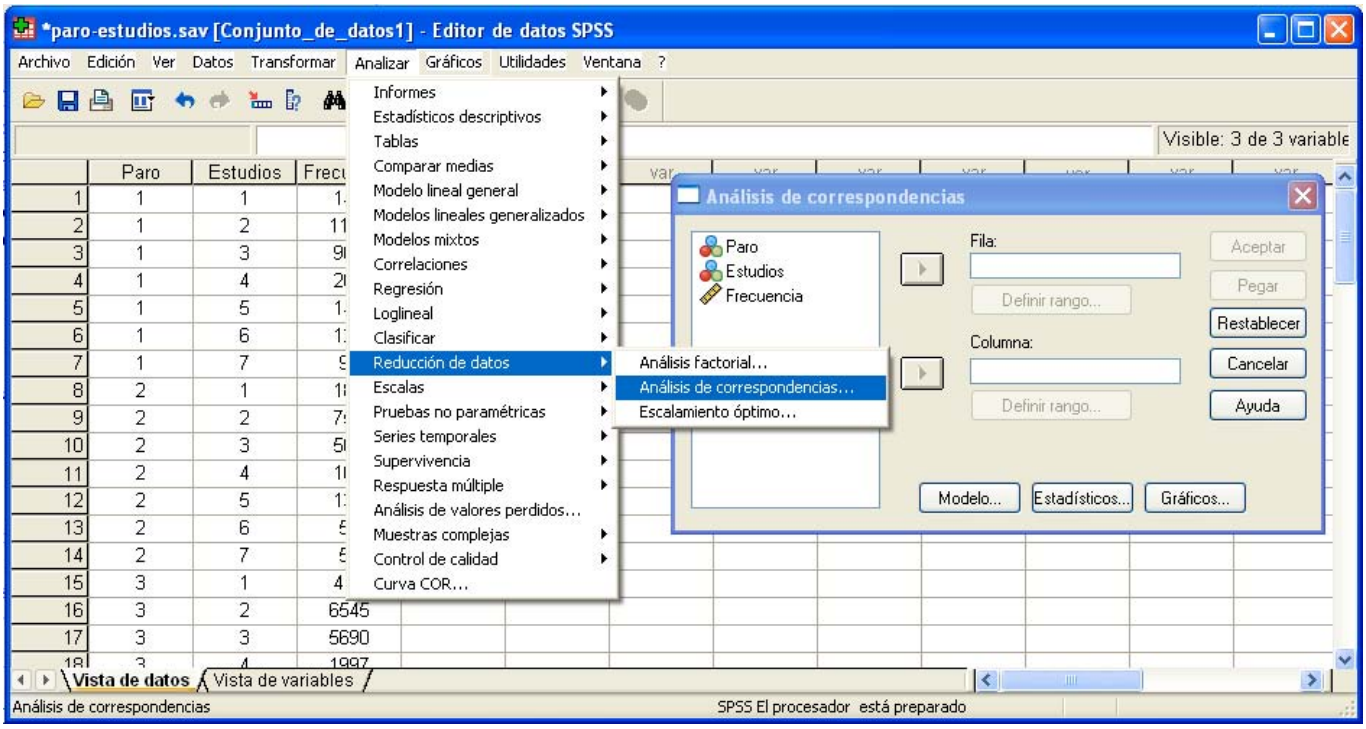
	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	3160,768[a]	30	,000
Razón de verosimilitudes	2919,072	30	,000
Asociación lineal por lineal	1572,065	1	,000
N de casos válidos	55069		

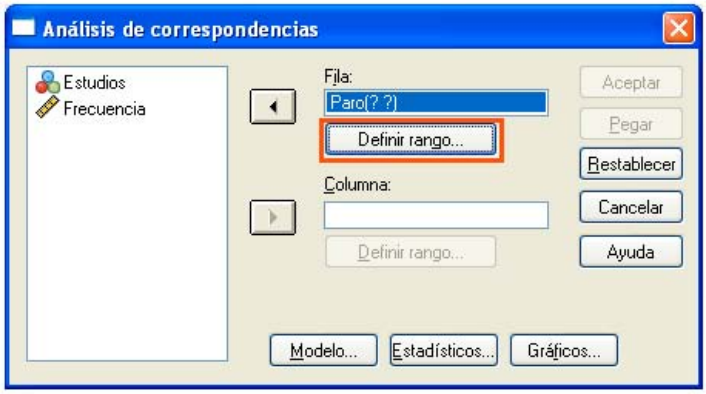
a. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 36,20.

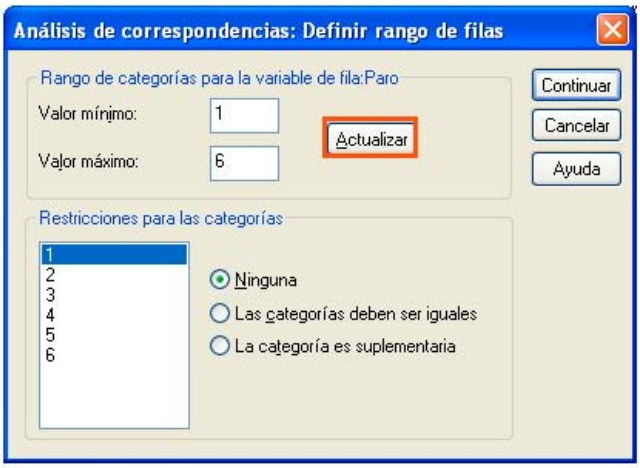
El $p_value = 0{,}000 < 0{,}05$ con lo que se rechaza la hipótesis nula, concluyendo que los patrones de paro son distintos según el sexo-provincia de residencia.

REDUCIR DATOS / ANÁLISIS DE CORRESPONDENCIAS: Analizar / Clasificar / Discriminante





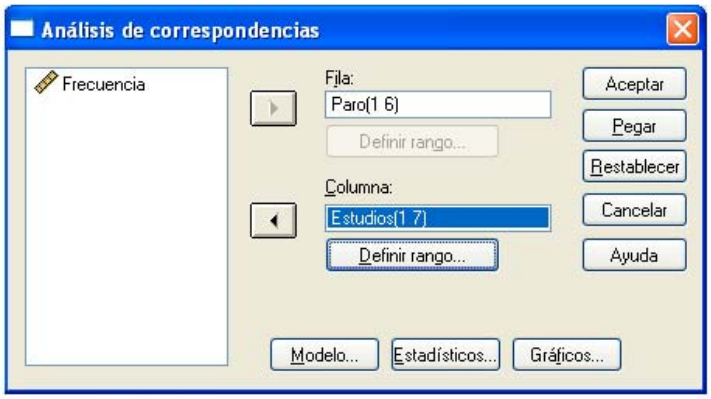


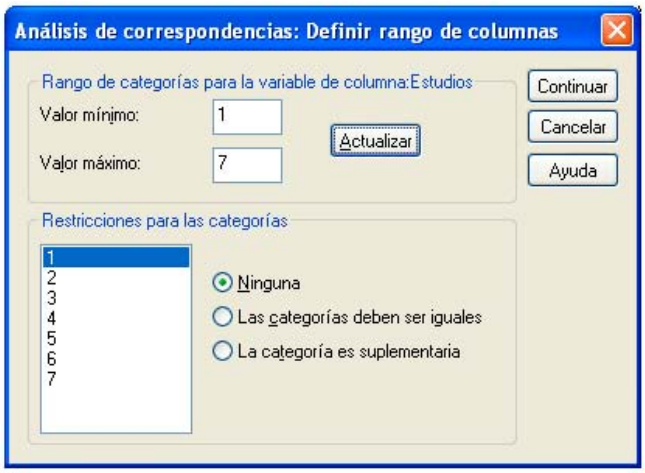


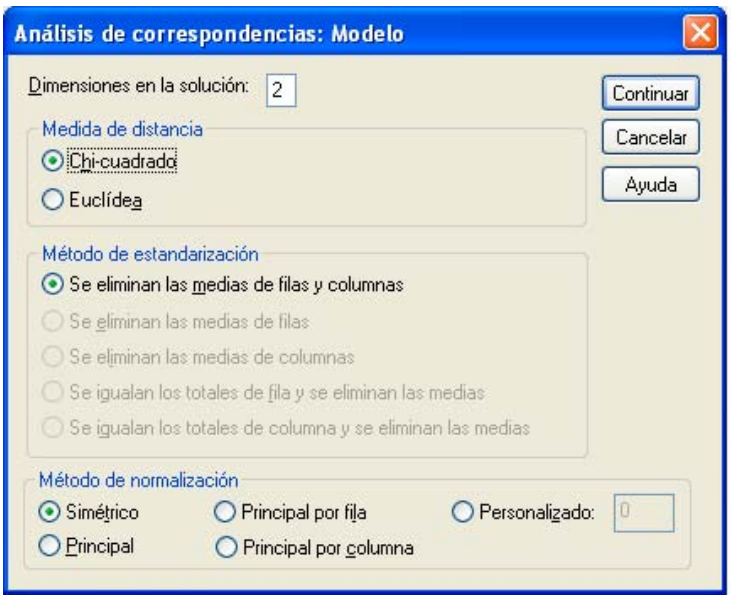
En la parte inferior del cuadro de diálogo aparecen una serie de Restricciones para las categorías que permiten unir las categorías o definir una categoría como suplementaria. La reducción del número de categorías activas que se origina al unir categorías o definir determinadas categorías como suplementarias está sujeta a ciertas restricciones:

(a) El número máximo de categorías que pueden unirse equivale al total de categorías menos 1.

(b) El número máximo de categorías suplementarias es el total de categorías menos 2.







El submenú Modelo se encuentra dividido en cuatro partes (número de dimensiones, medida de distancia, método de estandarización y normalización)

Dimensiones en la solución: Número de dimensiones necesarias para explicar la mayor parte de la variación. El número máximo de factores es igual al número de filas menos 1 (o al de columnas menos 1), el que sea menor.

En este ejemplo, con dos variables, respectivamente, con 6 y 7 categorías cada variable, el número máximo será (6 - 1 = 5). Sin embargo, el investigador deberá especificar el menor número de factores para explicar el máximo de inercia, buscando siempre el equilibrio entre parsimonia e interpretabilidad.

Aunque por defecto el programa selecciona 2 dimensiones, cuando se realiza el primer análisis de un conjunto de datos es conveniente solicitar un número elevado de dimensiones, con el fin de explicar un alto porcentaje de inercia, y disponer así de la máxima información para decidir la dimensionalidad adecuada.

Medidas de distancia: Entre las filas y las columnas, se pueden medir entre dos medidas:

- Distancia Chi-cuadrado: Las modalidades se ponderan en función de la masa de las filas o columnas. Utilizada para el análisis de correspondencias estándar.
- Distancia Euclídea: Raíz cuadrada de la suma cuadrática de las diferencias entre pares de filas y entre pares de columnas. Joaristi y Lizasoain desaconsejan la utilización de esta diferencia porque considera a todos los elementos con el mismo peso, llegando a desvirtuar el análisis de correspondencias.

Método de estandarización: Cuando se realiza la distancia Chi-cuadrado el programa centra las filas y las columnas al presentar únicamente la opción Se eliminan las medias de filas y columnas. Con la distancia euclídea se activan todas las opciones disponibles.

Método de normalización: La normalización se utiliza para distribuir la inercia de la tabla por filas y/o columnas, de modo que el método elegido únicamente afectará a las puntuaciones y a las varianzas de las filas y columnas. El resto de resultados no cambian (autovalores, inercia explicada por cada factor y el porcentaje de inercia explicada). El SPSS contempla cinco métodos de distribución de inercia:

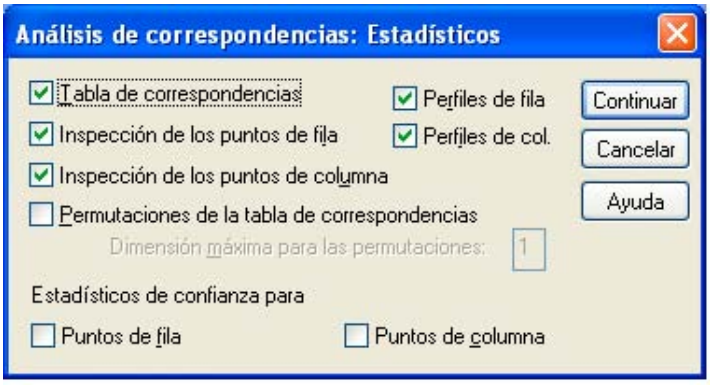
- Simétrico (canónico): En cada dimensión las puntuaciones de fila son el promedio ponderado de las puntuaciones de columna divididas por el autovalor correspondiente, mientras que las puntuaciones de columna son el promedio ponderado de las puntuaciones de fila divididas por el autovalor. Se aconseja utilizar este método cuando se desean examinar las diferencias (o similitudes) entre las dos variables. Es el método elegido para el ejemplo.

- Principal por fila: Las puntuaciones de la fila son la media ponderada de la puntuación de las columnas, maximizando así las distancias entre las categorías de la variable fila. Método aconsejable cuando el objetivo es analizar las diferencias entre las categorías de la variable situada en las filas.

- Principal por columna: Cuando el objetivo es analizar las diferencias o similitudes entre las categorías de la variable columna se utiliza este método, que considera las puntuaciones de las columnas como la media ponderada de la puntuación de las filas, llegando así a maximizar las distancias entre las columnas.

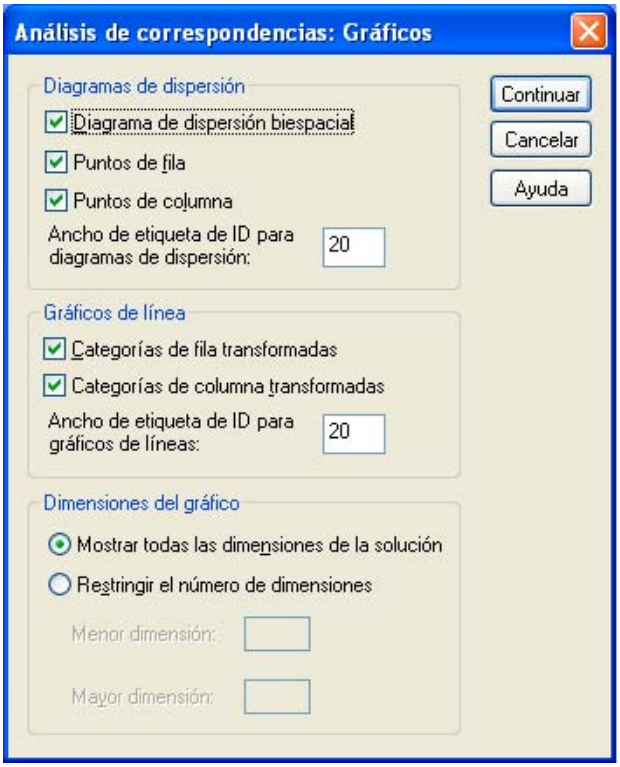
- Principal: Las distancias obtenidas con este método representan la distancia existente entre cada fila (o columna) a la distancia promedio del perfil fila (o columnas). Este método se utiliza cuando se desean examinar las diferencias entre las categorías de la variable fila y las diferencias entre las categorías de la variable columna, pero no las diferencias entre variables. Con este método no es posible representar el Diagrama de dispersión biespacial (submenú Estadísticos).

- Personalizado: Introduciendo un valor entre –1 y 1 en la ventana situada a la derecha de esta opción. El valor –1 realiza un análisis principal por columna, el 1 un análisis principal por fila, y el 0 un análisis simétrico. Dentro de estos limites, el valor elegido dispersará la inercia sobre las puntuaciones de fila o columna en grados diversos. Una de las ventajas de este método es que permite crear diagramas de dispersión biespaciales a medida.



El submenú Estadísticos se encuentran varias opciones.

- Tabla de correspondencias: Muestra una tabla de contingencia de las variables de entrada (fichero de datos), incluyendo el número de casos en cada celda y los totales marginales de fila y columna.
- Inspección de los puntos de la fila: Para cada categoría de la variable fila se muestran las masas, puntuaciones, inercia, contribución absoluta (contribución de la dimensión a la inercia) y relativa (contribución de la dimensión a la inercia del punto).
- Inspección de los puntos de columna: Masas, puntuaciones, inercia, contribución de la dimensión a la inercia (contribución absoluta) y contribución de la dimensión a la inercia del punto (contribución relativa) para cada categoría de la variable situada en la columna.
- Permutaciones de la tabla de correspondencias: Realiza una ordenación de la tabla de correspondencias situando las filas y las columnas en orden ascendente en función de las puntuaciones de la primera dimensión. Es posible representar el resto de dimensiones colocando un número en la ventana Dimensión máxima para las permutaciones, al realizarlo se genera una tabla permutada para cada dimensión, desde la primera hasta el número especificado.
- Perfiles de fila: Proporción de cada categoría con relación al marginal de fila.
- Perfiles de columna: Proporción de cada categoría con relación al marginal de columna.
- Estadísticos de confianza para puntos de fila: Desviación típica y correlaciones para los puntos de fila activos.
- Estadísticos de confianza para puntos de columna: Desviación típica y correlaciones para los puntos de columna activos.



El submenú Gráficos:

- Diagrama de dispersión biespacial: Gráfico de dispersión con las puntuaciones de fila y columna. Señalar que no es posible representar este gráfico cuando se utiliza el método de normalización principal.
- Puntos de fila: Gráfico de dispersión con las puntuaciones de la fila.
- Puntos de columna: Gráfico de dispersión con las puntuaciones de la columna.

Gráficos de línea: Generan un gráfico para cada dimensión de la variable seleccionada.

- Categoría de fila transformadas: Representación gráfica de las coordenadas de la variable fila.
- Categoría de columna transformadas: Representa los valores de la variable colocada en la columna.

Todas las soluciones gráficas permiten delimitar el número de caracteres de las etiquetas de valor que serán utilizadas en la representación gráfica (Ancho de etiqueta de ID para diagramas de dispersión o para gráficos de líneas). Aunque el programa permite utilizar hasta 20 caracteres, se aconseja utilizar cuatro o cinco para facilitar la claridad e interpretación del gráfico.

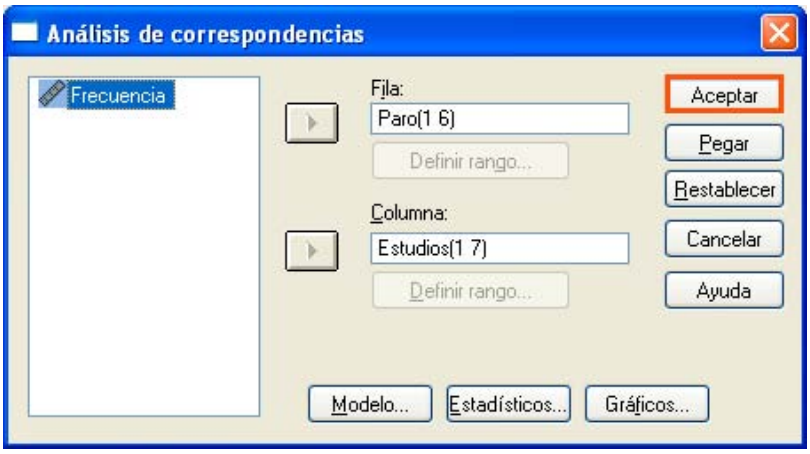


Tabla de correspondencias

Paro	Estudios							
	Primaria	Certificado Escolaridad	Graduado Escolar	Formación Profesional	Bachillerato	Diplomado	Universitario	Margen activo
Hombres Burgos	147	1120	908	268	149	127	94	2813
Hombres Avila	182	751	564	108	138	50	58	1851
Hombres Soria	415	6545	5690	1997	1415	670	877	17609
Mujeres Burgos	72	902	1646	561	417	461	236	4295
Mujeres Avila	57	534	1127	288	331	260	127	2724
Mujeres Soria	204	5931	9434	3250	2872	2196	1890	25777
Margen activo	1077	15783	19369	6472	5322	3764	3282	55069

Tabla de correspondencias: Tabla de doble entrada de resumen de los datos.

Perfiles de fila

Paro	Estudios							
	Primaria	Certificado Escolaridad	Graduado Escolar	Formación Profesional	Bachillerato	Diplomado	Universitario	Margen activo
Hombres Burgos	,052	,398	,323	,095	,053	,045	,033	1,000
Hombres Avila	,098	,406	,305	,058	,075	,027	,031	1,000
Hombres Soria	,024	,372	,323	,113	,080	,038	,050	1,000
Mujeres Burgos	,017	,210	,383	,131	,097	,107	,055	1,000
Mujeres Avila	,021	,196	,414	,106	,122	,095	,047	1,000
Mujeres Soria	,008	,230	,366	,126	,111	,085	,073	1,000
Masa	,020	,287	,352	,118	,097	,068	,060	

Perfiles de columna

Paro	Estudios							
	Primaria	Certificado Escolaridad	Graduado Escolar	Formación Profesional	Bachillerato	Diplomado	Universitario	Masa
Hombres Burgos	,136	,071	,047	,041	,028	,034	,029	,051
Hombres Avila	,169	,048	,029	,017	,026	,013	,018	,034
Hombres Soria	,385	,415	,294	,309	,266	,178	,267	,320
Mujeres Burgos	,067	,057	,085	,087	,078	,122	,072	,078
Mujeres Avila	,053	,034	,058	,044	,062	,069	,039	,049
Mujeres Soria	,189	,376	,487	,502	,540	,583	,576	,468
Margen activo	1,000	1,000	1,000	1,000	1,000	1,000	1,000	

Perfiles por fila y columna: son las distribuciones marginales. Los perfiles por fila son el resultado de dividir cada valor por el total de su fila mientras que los perfiles por columna es el valor de cada casilla entre el total su columna.

Resumen

Dimensión	Valor propio	Inercia	Chi-cuadrado	Sig.	Proporción de inercia		Confianza para el Valor propio	
					Explicada	Acumulada	Desviación típica	Correlación 2
1	,218	,047			,825	,825	,004	,244
2	,091	,008			,144	,969	,006	
3	,035	,001			,021	,989		
4	,024	,001			,010	,999		
5	,006	,000			,001	1,000		
Total		,057	3160,768	,000[a]	1,000	1,000		

a. 30 grados de libertad

La primera dimensión contribuye con 0,825 (82,5%) a dicha inercia, y las dos primeras dimensiones contribuyen con un 96,9%, mientras que las tres primeras dimensiones contribuyen a un 98,9%, por lo que se concluye que las dependencias observadas en la tabla vienen adecuadamente capturadas por las 2 primeras dimensiones.

Chi-cuadrado: Con este valor se obtiene el contraste de independencia entre las variables, presenta un p-valor < 0,05 rechazando la hipótesis nula de independencia entre variables. Señalar que si las variables fueran independientes no tendría lógica el análisis de correspondencias.

Examen de los puntos de fila[a]

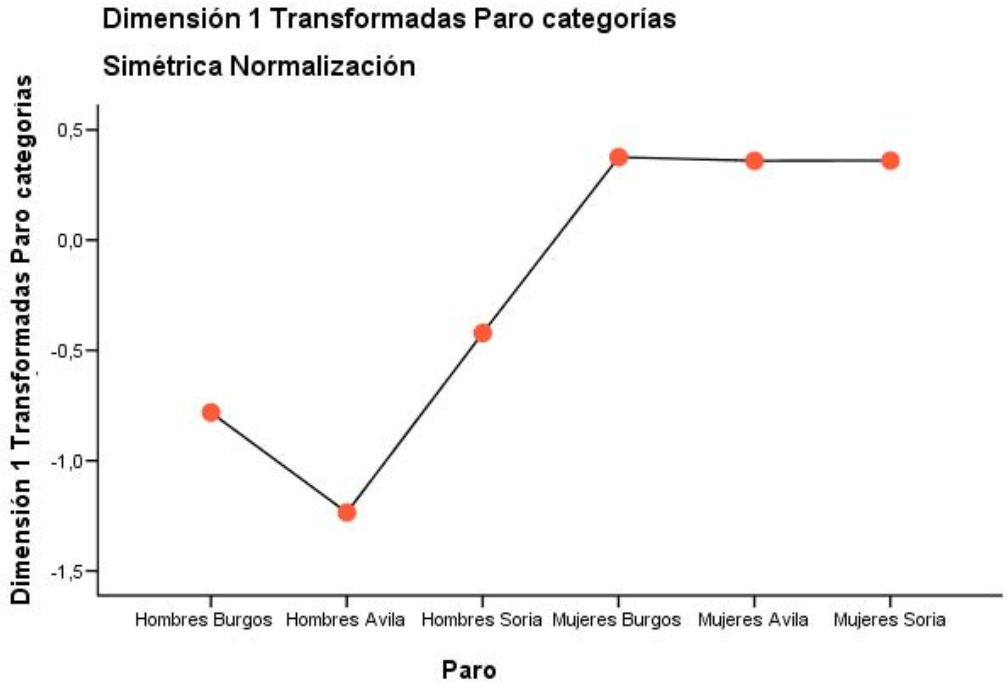
Paro	Masa	Puntuación en la dimensión		Inercia	Contribución				
					De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto		
		1	2		1	2	1	2	Total
Hombres Burgos	,051	-,781	,253	,007	,143	,036	,924	,040	,964
Hombres Avila	,034	-1,235	1,082	,015	,236	,433	,743	,238	,982
Hombres Soria	,320	-,421	-,290	,015	,261	,295	,833	,164	,997
Mujeres Burgos	,078	,376	,330	,004	,051	,094	,645	,207	,852
Mujeres Avila	,049	,360	,508	,003	,029	,141	,463	,386	,849
Mujeres Soria	,468	,361	-,016	,014	,280	,001	,983	,001	,983
Total activo	1,000			,057	1,000	1,000			

a. Normalización Simétrica

El análisis de las contribuciones relativa del examen de los puntos de fila, refleja una elevada representatividad, con valores cercanos a la unidad en la última columna, de las categorías en los factores.

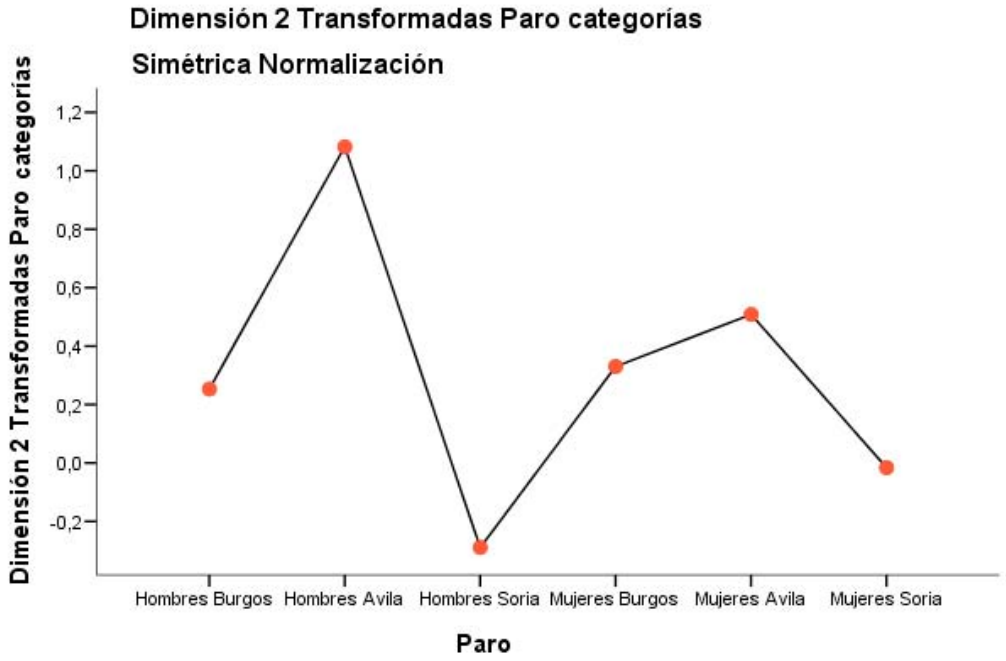
En la tabla de examen de los puntos fila (primer factor): Hombres de Burgos tiene una puntuación de (- 0,781) y Hombres de Avila (- 1,235), respectivamente, contribuyen a su inercia con una puntuación de (0,143) y (0,236).

En la parte positiva del primer factor se encuentran las mujeres, es decir, el primer factor discrimina por sexos.

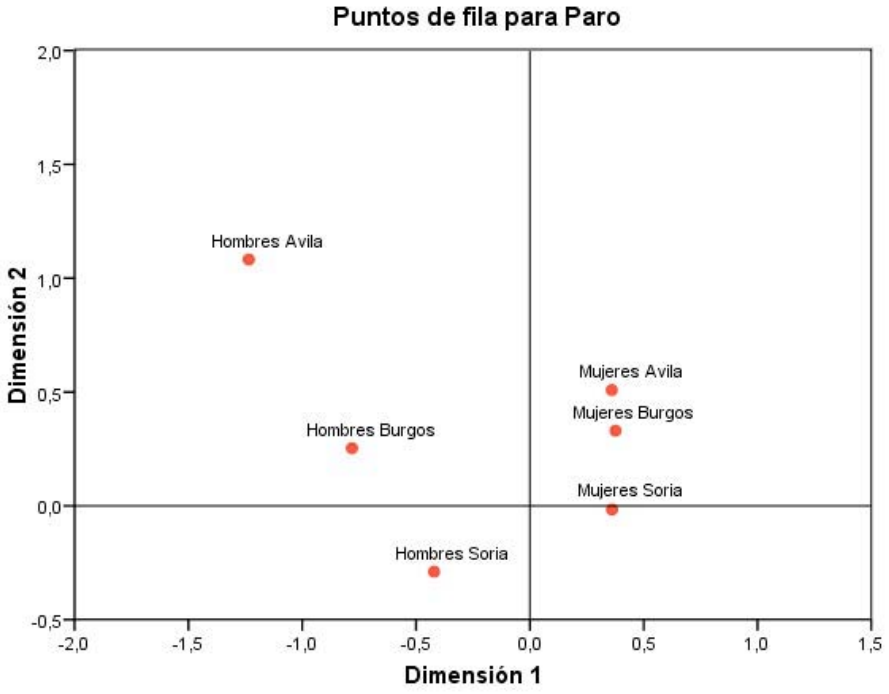


En el examen de los puntos fila (segundo factor): discrimina separando, esencialmente a Hombres de Avila con una puntuación de (1,082) y a Hombres de Soria (-0,290).

Los puntos que más contribuyen a su inercia son, consecuentemente, los puntos fila de Hombres Avila (0,433) y Hombres Soria (0,295).



Por otra parte, el segundo factor tiene una contribución relativa a la inercia nada despreciable de los puntos fila de Avila (Hombres y Mujeres), Mujeres de Burgos y Hombres de Soria.



Se observa que la primera dimensión discrimina por sexos.

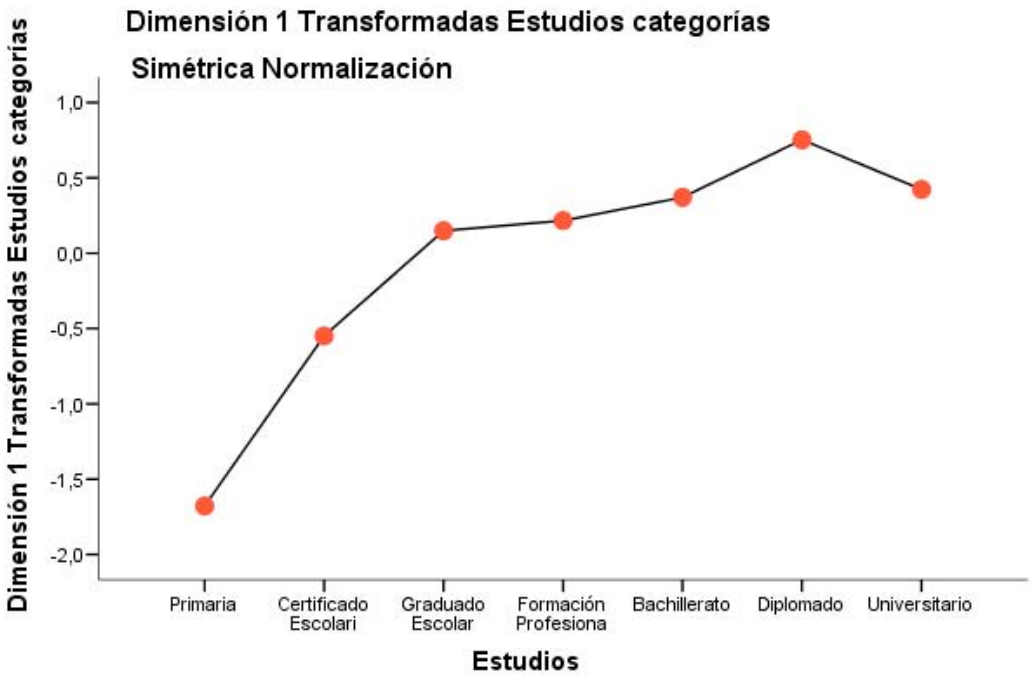
La segunda dimensión discrimina por provincias separando, especialmente a Ávila de Soria.

Examen de los puntos columna[a]

Estudios	Masa	Puntuación en la dimensión		Inercia	Contribución				
					De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto		
		1	2		1	2	1	2	Total
Primaria	,020	-1,678	1,669	,017	,253	,600	,705	,291	,996
Certificado Escolaridad	,287	-,549	-,228	,020	,398	,164	,932	,067	,999
Graduado Escolar	,352	,149	,088	,002	,036	,030	,805	,119	,924
Formación Profesional	,118	,216	-,195	,002	,025	,049	,690	,235	,925
Bachillerato	,097	,371	,076	,003	,061	,006	,845	,015	,860
Diplomado	,068	,753	,412	,010	,178	,128	,864	,108	,973
Universitario	,060	,423	-,186	,003	,049	,023	,726	,059	,785
Total activo	1,000			,057	1,000	1,000			

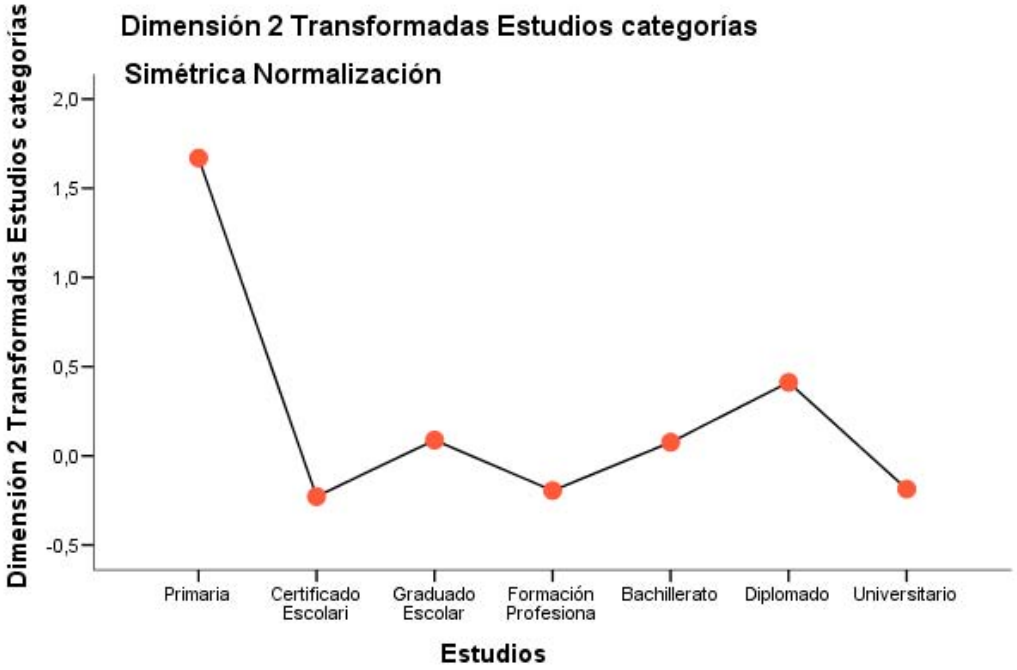
a. Normalización Simétrica

En la tabla se muestra la masa (peso proporcional a su importancia en el conjunto) de cada columna, las coordenadas de las columnas en los factores, la contribución de cada columna a la inercia total, las contribuciones absolutas y relativas.

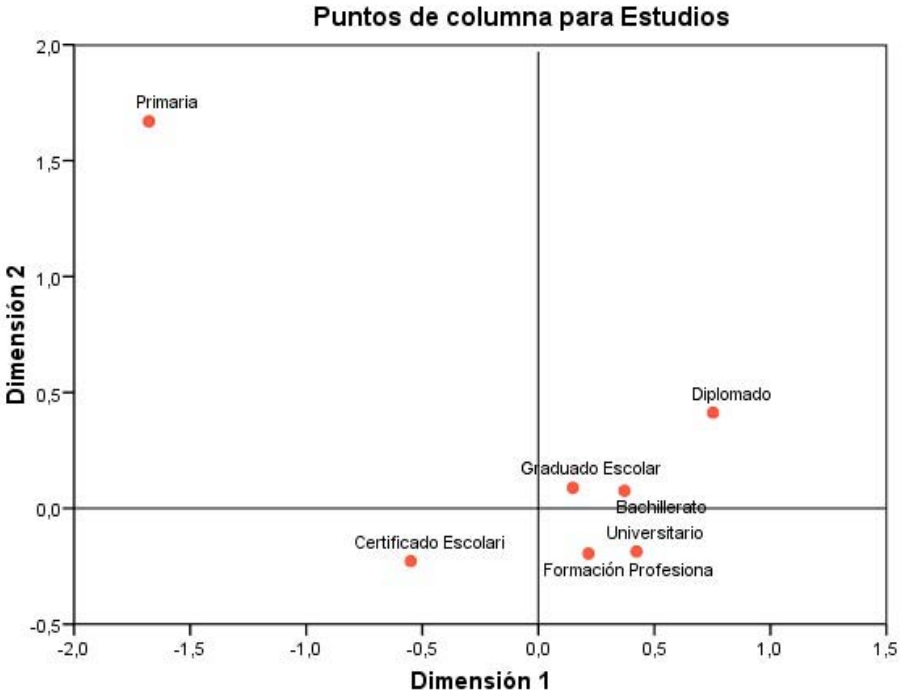


El primer factor discrimina los niveles de estudio más bajos - Primaria (0,253) y Certificado Escolar (0,398) - frente al resto, siendo éstas modalidades junto a la de Diplomados (0,178) las que más contribuyen a su inercia.

De otra parte, el primer factor es el que más contribuye a la inercia de todos los perfiles de columna

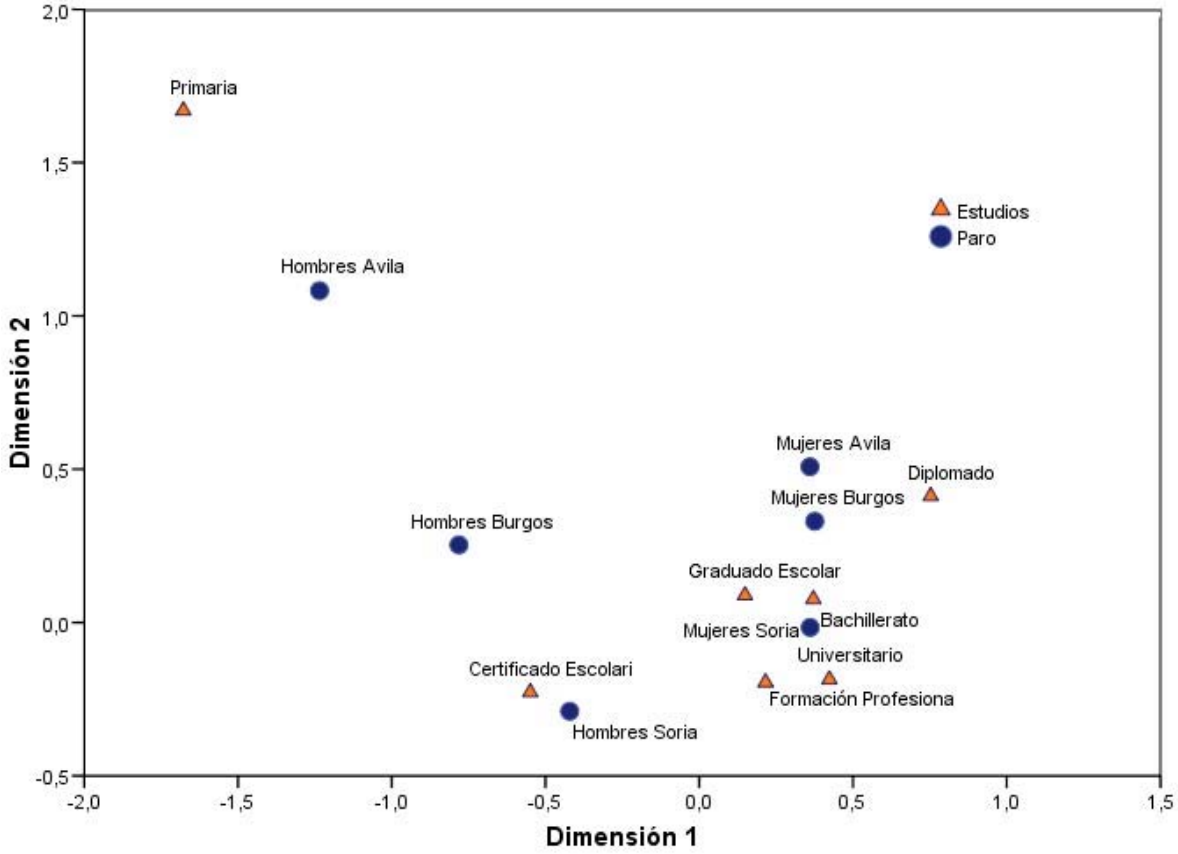


El segundo factor separa el perfil correspondiente al nivel de estudios de Primaria (0,600) del resto de los niveles. Destacar que el segundo factor tiene una contribución relativa nada despreciable a la inercia de los que tienen Certificado Escolaridad (0,164) y Formación Profesional (0,049).



El segundo factor separa el perfil correspondiente al nivel de Estudios de primaria (0,600) del resto de los niveles

Gráfico conjunto de los puntos de columna y de fila



Se observa que la primera dimensión discrimina entre sexos debido a la tendencia de haber más parados varones en los niveles de estudios más bajos (estudios primarios y certificado escolaridad) y más parados mujeres en el resto de los niveles.

Este hecho se manifiesta analizando las relaciones de proximidad y alejamiento de los puntos fila y columna.

Así, por ejemplo, la cercanía entre los puntos fila Hombres de Ávila y columna Estudios Primarios es debida a la tendencia en ambos perfiles a tener mayor número de parados de la modalidad representada por el otro perfil.

También se observa que la segunda dimensión pone de manifiesto la asociación positiva existente entre las categorías (Mujer de Burgos y Mujer de Ávila y Diplomado) y las de (Mujer de Soria y Universitario) mostrando una especialización en el tipo de paro existente en las mujeres de Castilla-León.

En Mujeres de Burgos y Ávila tiende a haber mayores niveles de paro relativo en Diplomado, mientras que en Mujeres de Soria hay mayor nivel de paro relativo en Universitario

ANÁLISIS DE CORRESPONDENCIAS (AC)

En la tabla adjunta se presenta la Encuesta de Presupuestos Familiares durante el período 1990-1991, realizado en la distintas Comunidades Autónomas españolas. En las filas se recogen cada Comunidad Autónoma, mientras que las columnas recogen las cantidades gastadas en cada concepto según los nueve grandes grupos de gasto formulados por el INE:

Grupo 1: Alimentos, bebidas y tabaco
Grupo 2: Vestido y calzado
Grupo 3: Vivienda, calefacción y alumbrado
Grupo 4: Muebles, enseres y servicios del hogar
Grupo 5: Servicios médicos y gastos sanitarios
Grupo 6: Transportes y comunicaciones
Grupo 7: Esparcimiento, enseñanza y cultura
Grupo 8: Otros bienes y servicios
Grupo 9: Otros gastos no mencionados

De este modo, el cruce de una fila y una columna representan las pesetas que los habitantes de una Comunidad dedican a un determinado concepto.

		Gastos									
		Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7	Grupo 8	Grupo 9	Total
CCAA	Andalucía	604906	222072	183773	120660	50308	255114	114433	281153	83027	1915446
	Aragón	547436	255127	201639	125630	50268	246867	110719	262864	81832	1882382
	Asturias	586561	280510	232693	130525	57682	336226	150503	310705	120725	2206130
	Baleares	549567	227004	205880	143722	87284	357122	151111	333944	131271	2186905
	Canarias	572256	185538	179574	134070	74322	305326	159185	279855	98012	1988138
	Cantabria	587826	288598	261429	118398	63395	302423	116816	276978	105707	2121570
	Castilla, La Mancha	546618	217780	191246	119423	41204	251981	109661	259545	107719	1845177
	Castilla, León	543406	220646	210043	125781	53318	243869	96367	252823	101476	1847729
	Cataluña	685925	262378	282701	162944	92869	360872	227525	362571	107370	2545155
	Comunidad Valenciana	541577	217826	177228	132318	61803	280136	122874	281020	99310	1914092
	Extremadura	469635	211088	137556	100379	43829	207527	86593	222698	71039	1550344
	Galicia	615209	247674	198830	131660	51307	291052	128064	255713	104412	2023921
	Madrid	673620	253666	252591	145747	86853	369559	215523	432997	129751	2560307
	Murcia	603683	210393	189394	127725	46726	318856	103650	310354	113789	2024570
	Navarra	643258	324500	251319	220787	80939	406681	185658	408892	154978	2677012
	País Vasco	636443	267178	232181	157854	64959	342386	174320	394733	121880	2391934
	Rioja	602255	209670	196262	127146	54824	262629	126054	313107	121721	2013668
	Ceuta y Melilla	683373	193283	133987	81436	27191	141950	84439	234961	64365	1644985
Total		10693554	4294931	3718326	2406205	1089081	5280576	2463495	5474913	1918384	4E+007

PLANTEAMIENTO DEL PROBLEMA: El objetivo principal de las Encuestas de Presupuestos Familiares es actualizar las ponderaciones de los distintos bienes y servicios que integran la cesta de la compra para la elaboración del IPC (Índice de Precios al Consumo). Aunque en su origen estas encuestas analizaban únicamente el gasto de las unidades familiares, en la actualidad recogen gran cantidad de información sobre las familias entrevistadas (número de miembros, origen de los ingresos, etc.), información que proporciona una visión detallada de los modos de vida al permitir establecer conexiones entre los patronos de gastos, los ingresos de la unidad familiar, el equipamiento del hogar, las condiciones de la vivienda, etc.

En este ejercicio se utiliza como criterio clasificador la distribución de los gastos familiares. En este sentido, la pregunta esencial de la investigación cuestiona la existencia de relación o dependencia entre Comunidad Autónoma y tipo de gasto, es decir, si hay Comunidades que pueden caracterizarse por una serie de gastos diferenciados, o si más bien Comunidad y tipo de gasto son independientes.

Señalar que cuando se postula que las Comunidades Autónomas tienen patronos diferenciales de gasto se asume la existencia de una estructura de independencia entre tipos de gastos y Comunidades Autónomas.

En esta línea, surgen varias preguntas: ¿Cuáles son las características de esa relación?, ¿Qué Comunidades Autónomas tienen pautas similares (diferenciadas) de gasto?, ¿Qué grupos de gasto tienen una distribución semejante en las Comunidades?.

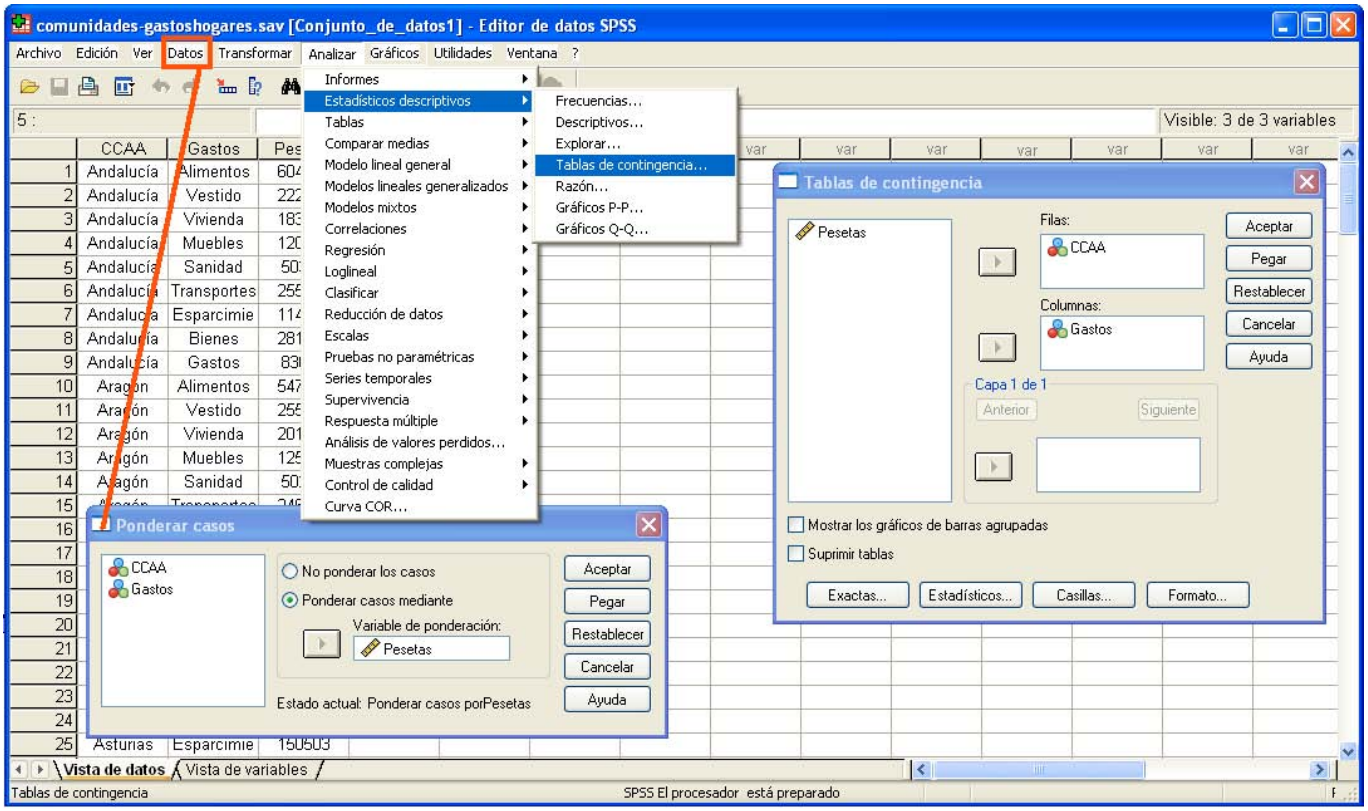
El Análisis de correspondencias permite plantear también cómo las categorías de una variable explican las diferencias (similitudes) observadas en la otra, es decir, ¿Qué gastos explican las diferencias (o similitudes) entre las Comunidades?, ¿Qué Comunidades explican la similitud (o diferencia) en los patrones de gasto?.

La gran heterogeneidad de las regiones españolas es adoptada como punto de partida para analizar la distribución de los gastos en bienes y servicios por parte de las familias de cada Comunidad Autónoma, partiendo de la hipótesis que en función del desarrollo de cada Comunidad sus habitantes manifestarán diferentes comportamientos en relación con la distribución de sus gastos.

PRIMER ANÁLISIS DE LA INFORMACIÓN: comunidades-gastohogares.sav

Se comienza con un análisis univariado de las variables consideradas, con la finalidad de identificar determinados patrones de comportamiento.

Analizar / Estadísticos descriptivos / Tablas de contingencia



Se solicitan porcentajes de Fila y Columna, Residuos estandarizados corregidos y Pruebas de Chi-cuadrado.

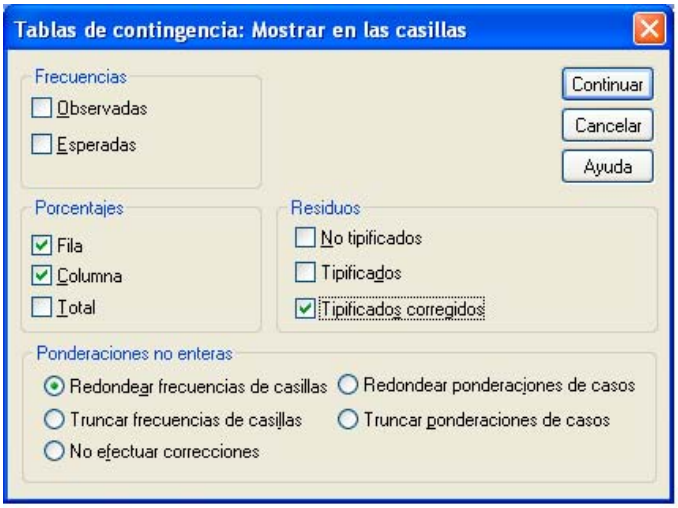


Tabla de contingencia CCAA * Gastos

			Gastos									
			Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7	Grupo 8	Grupo 9	Total
CCAA	Andalucía	Recuento	604906	222072	183773	120660	50308	255114	114433	281153	83027	1915446
		% de CCAA	31,6%	11,6%	9,6%	6,3%	2,6%	13,3%	6,0%	14,7%	4,3%	100,0%
		% de Gastos	5,7%	5,2%	4,9%	5,0%	4,6%	4,8%	4,6%	5,1%	4,3%	5,1%
		% del total	1,6%	,6%	,5%	,3%	,1%	,7%	,3%	,8%	,2%	5,1%
	Aragón	Recuento	547436	255127	201639	125630	50268	246867	110719	262864	81832	1882382
		% de CCAA	29,1%	13,6%	10,7%	6,7%	2,7%	13,1%	5,9%	14,0%	4,3%	100,0%
		% de Gastos	5,1%	5,9%	5,4%	5,2%	4,6%	4,7%	4,5%	4,8%	4,3%	5,0%
		% del total	1,5%	,7%	,5%	,3%	,1%	,7%	,3%	,7%	,2%	5,0%
	Asturias	Recuento	586561	280510	232693	130525	57682	336226	150503	310705	120725	2206130
		% de CCAA	26,6%	12,7%	10,5%	5,9%	2,6%	15,2%	6,8%	14,1%	5,5%	100,0%
		% de Gastos	5,5%	6,5%	6,3%	5,4%	5,3%	6,4%	6,1%	5,7%	6,3%	5,9%
		% del total	1,6%	,8%	,6%	,3%	,2%	,9%	,4%	,8%	,3%	5,9%
	Baleares	Recuento	549567	227004	205880	143722	87284	357122	151111	333944	131271	2186905
		% de CCAA	25,1%	10,4%	9,4%	6,6%	4,0%	16,3%	6,9%	15,3%	6,0%	100,0%
		% de Gastos	5,1%	5,3%	5,5%	6,0%	8,0%	6,8%	6,1%	6,1%	6,8%	5,9%
		% del total	1,5%	,6%	,6%	,4%	,2%	1,0%	,4%	,9%	,4%	5,9%
	Canarias	Recuento	572256	185538	179574	134070	74322	305326	159185	279855	98012	1988138
		% de CCAA	28,8%	9,3%	9,0%	6,7%	3,7%	15,4%	8,0%	14,1%	4,9%	100,0%
		% de Gastos	5,4%	4,3%	4,8%	5,6%	6,8%	5,8%	6,5%	5,1%	5,1%	5,3%
		% del total	1,5%	,5%	,5%	,4%	,2%	,8%	,4%	,7%	,3%	5,3%
	Cantabria	Recuento	587826	288598	261429	118398	63395	302423	116816	276978	105707	2121570
		% de CCAA	27,7%	13,6%	12,3%	5,6%	3,0%	14,3%	5,5%	13,1%	5,0%	100,0%
		% de Gastos	5,5%	6,7%	7,0%	4,9%	5,8%	5,7%	4,7%	5,1%	5,5%	5,7%
		% del total	1,6%	,8%	,7%	,3%	,2%	,8%	,3%	,7%	,3%	5,7%
	Castilla, La Mancha	Recuento	546618	217780	191246	119423	41204	251981	109661	259545	107719	1845177
		% de CCAA	29,6%	11,8%	10,4%	6,5%	2,2%	13,7%	5,9%	14,1%	5,8%	100,0%
		% de Gastos	5,1%	5,1%	5,1%	5,0%	3,8%	4,8%	4,5%	4,7%	5,6%	4,9%
		% del total	1,5%	,6%	,5%	,3%	,1%	,7%	,3%	,7%	,3%	4,9%
	Castilla, León	Recuento	543406	220646	210043	125781	53318	243869	96367	252823	101476	1847729
		% de CCAA	29,4%	11,9%	11,4%	6,8%	2,9%	13,2%	5,2%	13,7%	5,5%	100,0%
		% de Gastos	5,1%	5,1%	5,6%	5,2%	4,9%	4,6%	3,9%	4,6%	5,3%	4,9%
		% del total	1,5%	,6%	,6%	,3%	,1%	,7%	,3%	,7%	,3%	4,9%
	Cataluña	Recuento	685925	262378	282701	162944	92869	360872	227525	362571	107370	2545155
		% de CCAA	27,0%	10,3%	11,1%	6,4%	3,6%	14,2%	8,9%	14,2%	4,2%	100,0%
		% de Gastos	6,4%	6,1%	7,6%	6,8%	8,5%	6,8%	9,2%	6,6%	5,6%	6,8%
		% del total	1,8%	,7%	,8%	,4%	,2%	1,0%	,6%	1,0%	,3%	6,8%
	Comunidad Valenciana	Recuento	541577	217826	177228	132318	61803	280136	122874	281020	99310	1914092
		% de CCAA	28,3%	11,4%	9,3%	6,9%	3,2%	14,6%	6,4%	14,7%	5,2%	100,0%
		% de Gastos	5,1%	5,1%	4,8%	5,5%	5,7%	5,3%	5,0%	5,1%	5,2%	5,1%
		% del total	1,5%	,6%	,5%	,4%	,2%	,8%	,3%	,8%	,3%	5,1%
	Extremadura	Recuento	469635	211088	137556	100379	43829	207527	86593	222698	71039	1550344
		% de CCAA	30,3%	13,6%	8,9%	6,5%	2,8%	13,4%	5,6%	14,4%	4,6%	100,0%
		% de Gastos	4,4%	4,9%	3,7%	4,2%	4,0%	3,9%	3,5%	4,1%	3,7%	4,2%
		% del total	1,3%	,6%	,4%	,3%	,1%	,6%	,2%	,6%	,2%	4,2%
	Galicia	Recuento	615209	247674	198830	131660	51307	291052	128064	255713	104412	2023921
		% de CCAA	30,4%	12,2%	9,8%	6,5%	2,5%	14,4%	6,3%	12,6%	5,2%	100,0%
		% de Gastos	5,8%	5,8%	5,3%	5,5%	4,7%	5,5%	5,2%	4,7%	5,4%	5,4%
		% del total	1,6%	,7%	,5%	,4%	,1%	,8%	,3%	,7%	,3%	5,4%
	Madrid	Recuento	673620	253666	252591	145747	86853	369559	215523	432997	129751	2560307
		% de CCAA	26,3%	9,9%	9,9%	5,7%	3,4%	14,4%	8,4%	16,9%	5,1%	100,0%
		% de Gastos	6,3%	5,9%	6,8%	6,1%	8,0%	7,0%	8,7%	7,9%	6,8%	6,9%
		% del total	1,8%	,7%	,7%	,4%	,2%	1,0%	,6%	1,2%	,3%	6,9%
	Murcia	Recuento	603683	210393	189394	127725	46726	318856	103650	310354	113789	2024570
		% de CCAA	29,8%	10,4%	9,4%	6,3%	2,3%	15,7%	5,1%	15,3%	5,6%	100,0%
		% de Gastos	5,6%	4,9%	5,1%	5,3%	4,3%	6,0%	4,2%	5,7%	5,9%	5,4%
		% del total	1,6%	,6%	,5%	,3%	,1%	,9%	,3%	,8%	,3%	5,4%
	Navarra	Recuento	643258	324500	251319	220787	80939	406681	185658	408892	154978	2677012
		% de CCAA	24,0%	12,1%	9,4%	8,2%	3,0%	15,2%	6,9%	15,3%	5,8%	100,0%
		% de Gastos	6,0%	7,6%	6,8%	9,2%	7,4%	7,7%	7,5%	7,5%	8,1%	7,2%
		% del total	1,7%	,9%	,7%	,6%	,2%	1,1%	,5%	1,1%	,4%	7,2%
	País Vasco	Recuento	636443	267178	232181	157854	64959	342386	174320	394733	121880	2391934
		% de CCAA	26,6%	11,2%	9,7%	6,6%	2,7%	14,3%	7,3%	16,5%	5,1%	100,0%
		% de Gastos	6,0%	6,2%	6,2%	6,6%	6,0%	6,5%	7,1%	7,2%	6,4%	6,4%
		% del total	1,7%	,7%	,6%	,4%	,2%	,9%	,5%	1,1%	,3%	6,4%
	Ríoja	Recuento	602255	209670	196262	127146	54824	262629	126054	313107	121721	2013668
		% de CCAA	29,9%	10,4%	9,7%	6,3%	2,7%	13,0%	6,3%	15,5%	6,0%	100,0%
		% de Gastos	5,6%	4,9%	5,3%	5,3%	5,0%	5,0%	5,1%	5,7%	6,3%	5,4%
		% del total	1,6%	,6%	,5%	,3%	,1%	,7%	,3%	,8%	,3%	5,4%
	Ceuta y Melilla	Recuento	683373	193283	133987	81436	27191	141950	84439	234961	64365	1644985
		% de CCAA	41,5%	11,7%	8,1%	5,0%	1,7%	8,6%	5,1%	14,3%	3,9%	100,0%
		% de Gastos	6,4%	4,5%	3,6%	3,4%	2,5%	2,7%	3,4%	4,3%	3,4%	4,4%
		% del total	1,8%	,5%	,4%	,2%	,1%	,4%	,2%	,6%	,2%	4,4%
Total		Recuento	10693554	4294931	3718326	2406205	1089081	5280576	2463495	5474913	1918384	37339465
		% de CCAA	28,6%	11,5%	10,0%	6,4%	2,9%	14,1%	6,6%	14,7%	5,1%	100,0%
		% de Gastos	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
		% del total	28,6%	11,5%	10,0%	6,4%	2,9%	14,1%	6,6%	14,7%	5,1%	100,0%

% de CCAA

Tabla de contingencia CCAA * Gastos

		Gastos									Total
		Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7	Grupo 8	Grupo 9	
CCAA	Andalucía	31,6%	11,6%	9,6%	6,3%	2,6%	13,3%	6,0%	14,7%	4,3%	100,0%
	Aragón	29,1%	13,6%	10,7%	6,7%	2,7%	13,1%	5,9%	14,0%	4,3%	100,0%
	Asturias	26,6%	12,7%	10,5%	5,9%	2,6%	15,2%	6,8%	14,1%	5,5%	100,0%
	Baleares	25,1%	10,4%	9,4%	6,6%	4,0%	16,3%	6,9%	15,3%	6,0%	100,0%
	Canarias	28,8%	9,3%	9,0%	6,7%	3,7%	15,4%	8,0%	14,1%	4,9%	100,0%
	Cantabria	27,7%	13,6%	12,3%	5,6%	3,0%	14,3%	5,5%	13,1%	5,0%	100,0%
	Castilla, La Mancha	29,6%	11,8%	10,4%	6,5%	2,2%	13,7%	5,9%	14,1%	5,8%	100,0%
	Castilla, León	29,4%	11,9%	11,4%	6,8%	2,9%	13,2%	5,2%	13,7%	5,5%	100,0%
	Cataluña	27,0%	10,3%	11,1%	6,4%	3,6%	14,2%	8,9%	14,2%	4,2%	100,0%
	Comunidad Valenciana	28,3%	11,4%	9,3%	6,9%	3,2%	14,6%	6,4%	14,7%	5,2%	100,0%
	Extremadura	30,3%	13,6%	8,9%	6,5%	2,8%	13,4%	5,6%	14,4%	4,6%	100,0%
	Galicia	30,4%	12,2%	9,8%	6,5%	2,5%	14,4%	6,3%	12,6%	5,2%	100,0%
	Madrid	26,3%	9,9%	9,9%	5,7%	3,4%	14,4%	8,4%	16,9%	5,1%	100,0%
	Murcia	29,8%	10,4%	9,4%	6,3%	2,3%	15,7%	5,1%	15,3%	5,6%	100,0%
	Navarra	24,0%	12,1%	9,4%	8,2%	3,0%	15,2%	6,9%	15,3%	5,8%	100,0%
	País Vasco	26,6%	11,2%	9,7%	6,6%	2,7%	14,3%	7,3%	16,5%	5,1%	100,0%
	Rioja	29,9%	10,4%	9,7%	6,3%	2,7%	13,0%	6,3%	15,5%	6,0%	100,0%
	Ceuta y Melilla	41,5%	11,7%	8,1%	5,0%	1,7%	8,6%	5,1%	14,3%	3,9%	100,0%
Total		28,6%	11,5%	10,0%	6,4%	2,9%	14,1%	6,6%	14,7%	5,1%	100,0%

Se comparan los porcentajes de cada Comunidad Autónoma con la distribución de gastos para todo el país (Porcentajes por fila), permitiendo detectar las Comunidades que emplean más proporción del gasto en determinados conceptos: En Andalucía emplean en el Grupo 1 el 31,6% de los gastos, mientras que en Navarra en ese mismo Grupo 1 se emplean el 24% de los gastos

De forma análoga, se tiene la distribución de porcentajes de Columna:

% de Gastos

Tabla de contingencia CCAA * Gastos

		Gastos									Total
		Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7	Grupo 8	Grupo 9	
CCAA	Andalucía	5,7%	5,2%	4,9%	5,0%	4,6%	4,8%	4,6%	5,1%	4,3%	5,1%
	Aragón	5,1%	5,9%	5,4%	5,2%	4,6%	4,7%	4,5%	4,8%	4,3%	5,0%
	Asturias	5,5%	6,5%	6,3%	5,4%	5,3%	6,4%	6,1%	5,7%	6,3%	5,9%
	Baleares	5,1%	5,3%	5,5%	6,0%	8,0%	6,8%	6,1%	6,1%	6,8%	5,9%
	Canarias	5,4%	4,3%	4,8%	5,6%	6,8%	5,8%	6,5%	5,1%	5,1%	5,3%
	Cantabria	5,5%	6,7%	7,0%	4,9%	5,8%	5,7%	4,7%	5,1%	5,5%	5,7%
	Castilla, La Mancha	5,1%	5,1%	5,1%	5,0%	3,8%	4,8%	4,5%	4,7%	5,6%	4,9%
	Castilla, León	5,1%	5,1%	5,6%	5,2%	4,9%	4,6%	3,9%	4,6%	5,3%	4,9%
	Cataluña	6,4%	6,1%	7,6%	6,8%	8,5%	6,8%	9,2%	6,6%	5,6%	6,8%
	Comunidad Valenciana	5,1%	5,1%	4,8%	5,5%	5,7%	5,3%	5,0%	5,1%	5,2%	5,1%
	Extremadura	4,4%	4,9%	3,7%	4,2%	4,0%	3,9%	3,5%	4,1%	3,7%	4,2%
	Galicia	5,8%	5,8%	5,3%	5,5%	4,7%	5,5%	5,2%	4,7%	5,4%	5,4%
	Madrid	6,3%	5,9%	6,8%	6,1%	8,0%	7,0%	8,7%	7,9%	6,8%	6,9%
	Murcia	5,6%	4,9%	5,1%	5,3%	4,3%	6,0%	4,2%	5,7%	5,9%	5,4%
	Navarra	6,0%	7,6%	6,8%	9,2%	7,4%	7,7%	7,5%	7,5%	8,1%	7,2%
	País Vasco	6,0%	6,2%	6,2%	6,6%	6,0%	6,5%	7,1%	7,2%	6,4%	6,4%
	Rioja	5,6%	4,9%	5,3%	5,3%	5,0%	5,0%	5,1%	5,7%	6,3%	5,4%
	Ceuta y Melilla	6,4%	4,5%	3,6%	3,4%	2,5%	2,7%	3,4%	4,3%	3,4%	4,4%
Total		100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

El gasto de Andalucía en el Grupo 1 es el 5,7% del gasto Nacional, mientras que en Navarra el gasto en el Grupo 1 es del 6% del gasto Nacional.

Para interpretar el interior de la tabla, desde un punto de vista práctico, es más sencillo utilizar los residuos estandarizados corregidos, que se muestran en la siguiente tabla:

Tabla de contingencia CCAA * Gastos

			Gastos									Total
			Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7	Grupo 8	Grupo 9	
CCAA	Andalucía	% de Gastos	5,7%	5,2%	4,9%	5,0%	4,6%	4,8%	4,6%	5,1%	4,3%	5,1%
		Residuos corregidos	92,5	4,1	-17,3	-8,4	-24,5	-33,6	-35,7	,6	-51,7	
	Aragón	% de Gastos	5,1%	5,9%	5,4%	5,2%	4,6%	4,7%	4,5%	4,8%	4,3%	5,0%
		Residuos corregidos	13,8	90,5	35,4	13,2	-20,6	-41,5	-40,6	-27,8	-50,4	
	Asturias	% de Gastos	5,5%	6,5%	6,3%	5,4%	5,3%	6,4%	6,1%	5,7%	6,3%	5,9%
		Residuos corregidos	-69,5	58,2	30,1	-32,9	-27,5	48,3	13,8	-25,1	23,2	
	Baleares	% de Gastos	5,1%	5,3%	5,5%	6,0%	8,0%	6,8%	6,1%	6,1%	6,8%	5,9%
		Residuos corregidos	-118,3	-53,6	-27,7	7,9	97,3	95,7	19,2	26,2	59,7	
	Canarias	% de Gastos	5,4%	4,3%	4,8%	5,6%	6,8%	5,8%	6,5%	5,1%	5,1%	5,3%
		Residuos corregidos	4,6	-98,6	-44,8	17,7	70,8	50,5	82,3	-24,0	-13,6	
	Cantabria	% de Gastos	5,5%	6,7%	7,0%	4,9%	5,8%	5,7%	4,7%	5,1%	5,5%	5,7%
		Residuos corregidos	-30,9	98,7	118,4	-52,7	6,4	4,8	-65,9	-68,1	-10,5	
	Castilla, La Mancha	% de Gastos	5,1%	5,1%	5,1%	5,0%	3,8%	4,8%	4,5%	4,7%	5,6%	4,9%
		Residuos corregidos	30,4	13,1	18,9	1,6	-56,6	-19,4	-36,7	-23,5	44,2	
	Castilla, León	% de Gastos	5,1%	5,1%	5,6%	5,2%	4,9%	4,6%	3,9%	4,6%	5,3%	4,9%
		Residuos corregidos	23,8	19,2	65,6	20,6	-2,6	-37,8	-77,6	-38,6	22,4	
	Cataluña	% de Gastos	6,4%	6,1%	7,6%	6,8%	8,5%	6,8%	9,2%	6,6%	5,6%	6,8%
		Residuos corregidos	-61,7	-61,8	63,4	-2,8	71,9	1,7	155,9	-19,5	-68,8	
	Comunidad Valenciana	% de Gastos	5,1%	5,1%	4,8%	5,5%	5,7%	5,3%	5,0%	5,1%	5,2%	5,1%
		Residuos corregidos	-10,8	-5,4	-33,2	27,1	26,3	20,1	-10,2	,8	3,3	
	Extremadura	% de Gastos	4,4%	4,9%	3,7%	4,2%	4,0%	3,9%	3,5%	4,1%	3,7%	4,2%
		Residuos corregidos	46,5	84,2	-46,1	1,6	-6,8	-27,6	-51,9	-10,7	-32,0	
	Galicia	% de Gastos	5,8%	5,8%	5,3%	5,5%	4,7%	5,5%	5,2%	4,7%	5,4%	5,4%
		Residuos corregidos	56,9	33,7	-6,6	3,6	-33,2	10,0	-15,9	-83,9	1,4	
	Madrid	% de Gastos	6,3%	5,9%	6,8%	6,1%	8,0%	7,0%	8,7%	7,9%	6,8%	6,9%
		Residuos corregidos	-85,4	-82,9	-5,1	-50,7	46,9	13,9	121,6	105,4	-5,2	
	Murcia	% de Gastos	5,6%	4,9%	5,1%	5,3%	4,3%	6,0%	4,2%	5,7%	5,9%	5,4%
		Residuos corregidos	38,2	-50,9	-29,5	-8,1	-52,9	67,5	-87,1	27,6	32,0	
	Navarra	% de Gastos	6,0%	7,6%	6,8%	9,2%	7,4%	7,7%	7,5%	7,5%	8,1%	7,2%
		Residuos corregidos	-173,2	33,0	-32,3	124,7	10,8	51,1	23,1	29,4	50,1	
	País Vasco	% de Gastos	6,0%	6,2%	6,2%	6,6%	6,0%	6,5%	7,1%	7,2%	6,4%	6,4%
		Residuos corregidos	-71,8	-16,7	-13,4	10,1	-19,1	7,9	44,5	83,2	-3,1	
	Rioja	% de Gastos	5,6%	4,9%	5,3%	5,3%	5,0%	5,0%	5,1%	5,7%	6,3%	5,4%
		Residuos corregidos	41,0	-49,8	-10,3	-7,7	-16,8	-46,0	-19,8	36,6	59,9	
	Ceuta y Melilla	% de Gastos	6,4%	4,5%	3,6%	3,4%	2,5%	2,7%	3,4%	4,3%	3,4%	4,4%
		Residuos corregidos	374,4	10,2	-79,4	-79,8	-98,5	-207,5	-77,4	-14,1	-72,8	
Total		% de Gastos	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

En cada casilla se recogen los valores absolutos y los residuos estandarizados. Un análisis de éstas permitirá detectar los grupos de gasto que predominan en cada Comunidad Autónoma:

Grupo 1 (alimentos, bebida y tabaco): Las Comunidades con un mayor residuo positivo son Ceuta y Melilla con Andalucía, este elevado valor detecta la existencia de una relación elevada entre estos elementos, en otras palabras, estas Comunidades destacan por dedicar un mayor porcentaje de ingresos a los gastos en alimentación, bebida y tabaco. Por otra parte, los elevados coeficientes negativos de Navarra, Baleares y Madrid reflejan que el porcentaje de ingresos dedicados a este gasto es deficiente con relación al resto de grupos.

Grupo 2 (vestido y calzado): En estos conceptos destacan Cantabria, Aragón y Extremadura; por el contrario, Canarias y Madrid presentan un elevado coeficiente negativo.

Grupo 3 (vivienda, calefacción y alumbrado): Hay gastos elevados en Cantabria, Castilla y León y Cataluña. Gastos reducidos en Ceuta y Melilla, Extramadura y Canarias.

Grupo 4 (muebles, enseres y servicios del hogar): Navarra y la Comunidad Valencia presenta un alto coeficiente, reflejando que dedican un alto porcentaje de ingresos a estos conceptos. Por el contrario, el porcentaje de ingresos dedicados a este concepto es deficiente en Ceuta y Melilla, Cantabria y Madrid.

Grupo 5 (servicios médicos y gastos sanitarios): Gastos elevados en Baleares, Canarias y Cataluña. Gastos reducidos en Ceuta y Melilla, Castilla - La Mancha y Murcia.

Grupo 6 (transportes y comunicaciones): Gastos elevados en Baleares, Canarias y Murcia. Gastos reducidos en Ceuta y Melilla, y la Rioja.

Grupo 7 (esparcimiento, enseñanza y cultura): Gastos elevados en Cataluña, Madrid y Canarias. Gastos reducidos en Murcia, Castilla y León, Cantabria, Ceuta y Melilla.

Grupo 8 (otros bienes y servicios): Gastos elevados en Madrid y País Vasco. Gastos reducidos en Galicia y Cantabria.

El análisis clásico de las tablas de contingencia permite resolver el objetivo propuesto, aunque presenta dificultad para revelar las estructuras de dependencia entre las categorías de las variables.
Por otra parte, a medida que la tabla aumenta de tamaño se va incrementando la complejidad para una adecuada interpretación. Subrayar que la utilización de tablas de contingencia no permite representar gráficamente las relaciones entre variables.

Pruebas de chi-cuadrado

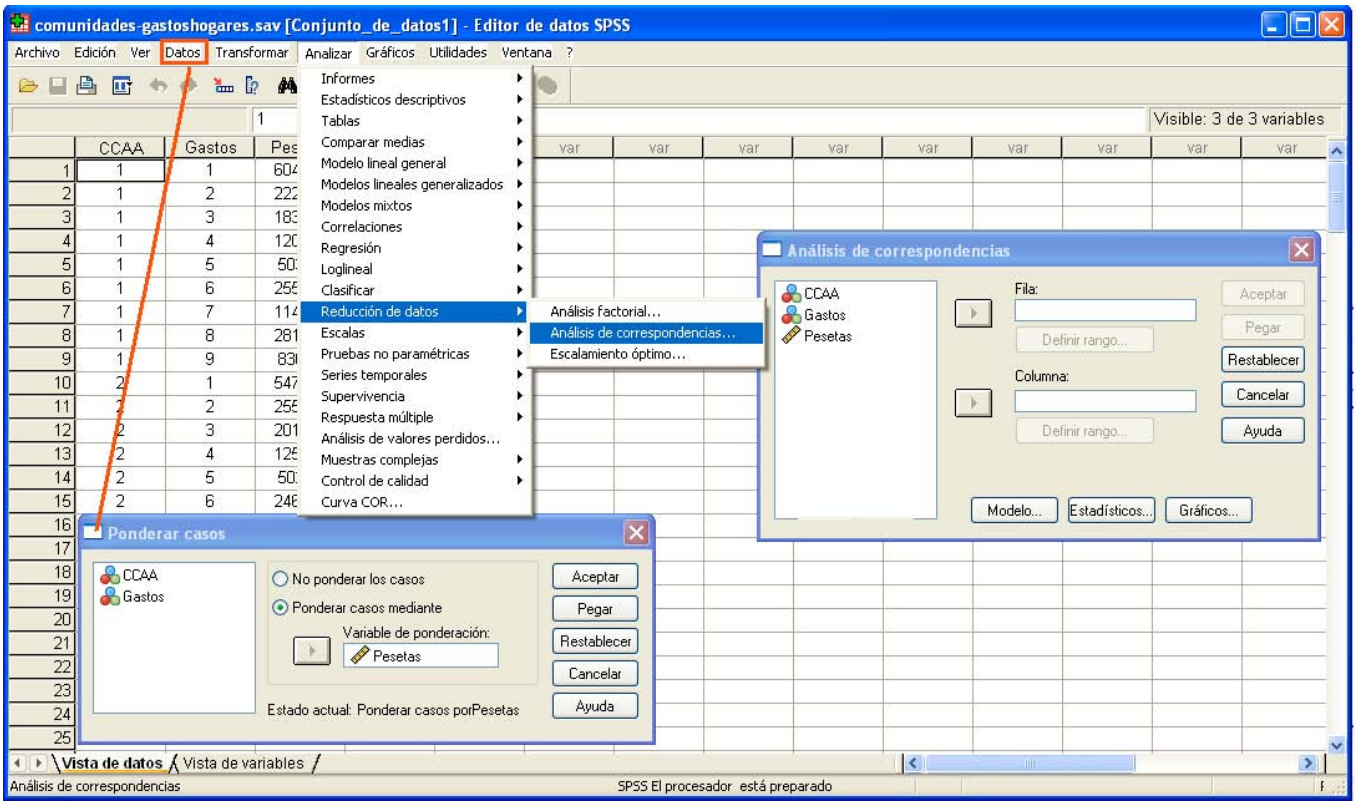
	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	489287,3[a]	136	,000
Razón de verosimilitudes	482338,7	136	,000
Asociación lineal por lineal	138,407	1	,000
N de casos válidos	37339465		

a. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 45218,92.

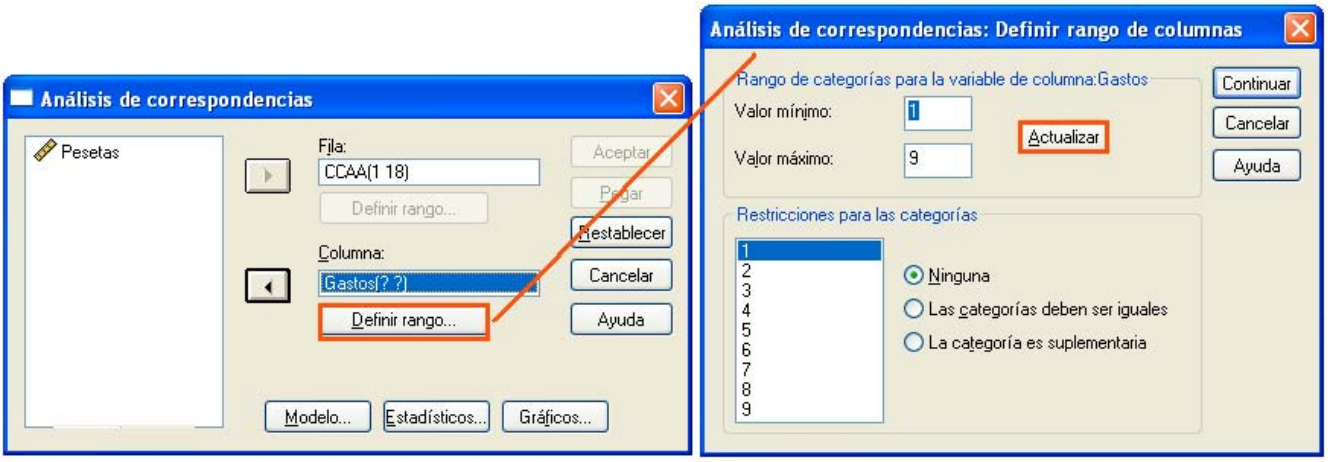
El p-value = 0 < 0,05, se rechaza la hipótesis nula, y se concluye que los patrones de gasto son distintos según la Comunidad Autónoma de residencia.

El Análisis factorial de Correspondencias se utiliza en la medida que su objetivo es profundizar en el conocimiento de las relaciones que se establecen entre dos variables cualitativas observadas en una misma población.

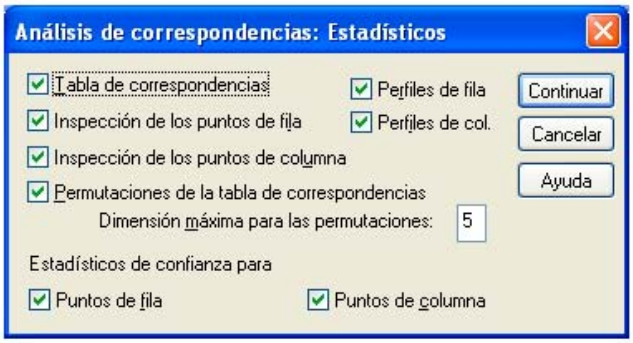
Analizar / Reducción de datos / Análisis de correspondencias (comunidades-gastoshogares.sav)



Análisis de correspondencias
Gastos
Pesetas
Fila:
CCAA(? ?)
Definir rango...
Columna:
Definir rango...
Aceptar
Pegar
Restablecer
Cancelar
Ayuda
Modelo...
Estadísticos...
Gráficos...
Análisis de correspondencias: Definir rango de filas
Rango de categorías para la variable de fila:CCAA
Valor mínimo:
1
Valor máximo:
18
Actualizar
Continuar
Cancelar
Ayuda
Restricciones para las categorías
1
2
3
4
5
6
7
8
9
Ninguna
Las categorías deben ser iguales
La categoría es suplementaria


Análisis de correspondencias
Pesetas
Fila:
CCAA(1 18)
Definir rango...
Columna:
Gastos(? ?)
Definir rango...
Aceptar
Pegar
Restablecer
Cancelar
Ayuda
Modelo...
Estadísticos...
Gráficos...
Análisis de correspondencias: Definir rango de columnas
Rango de categorías para la variable de columna:Gastos
Valor mínimo:
1
Valor máximo:
9
Actualizar
Continuar
Cancelar
Ayuda
Restricciones para las categorías
1
2
3
4
5
6
7
8
9
Ninguna
Las categorías deben ser iguales
La categoría es suplementaria

Análisis de correspondencias
Pesetas
Fila:
CCAA(1 18)
Definir rango...
Columna:
Gastos(1 9)
Definir rango...
Aceptar
Pegar
Restablecer
Cancelar
Ayuda
Modelo...
Estadísticos...
Gráficos...
Análisis de correspondencias: Modelo
Dimensiones en la solución:
5
Continuar
Cancelar
Ayuda
Medida de distancia
Chi-cuadrado
Euclídea
Método de estandarización
Se eliminan las medias de filas y columnas
Se eliminan las medias de filas
Se eliminan las medias de columnas
Se igualan los totales de fila y se eliminan las medias
Se igualan los totales de columna y se eliminan las medias
Método de normalización
Simétrico
Principal por fila
Personalizado:
0
Principal
Principal por columna



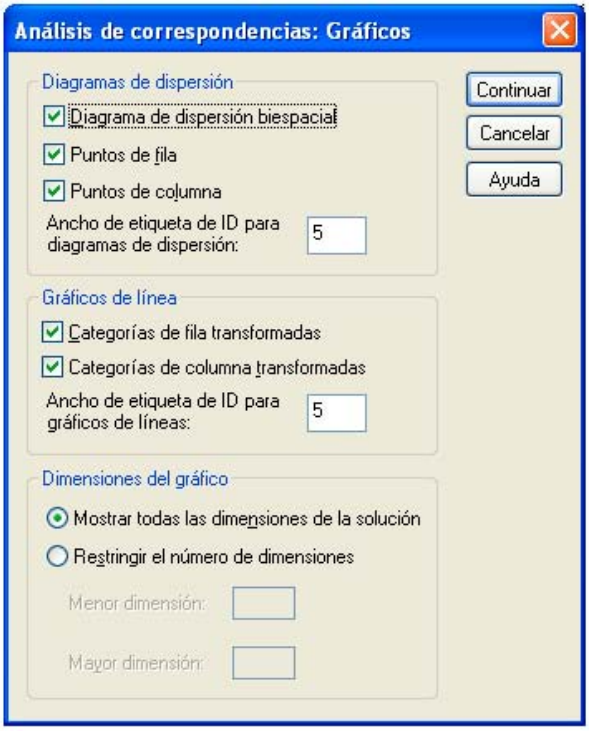


Tabla de correspondencias

CCAA	Gastos									
	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7	Grupo 8	Grupo 9	Margen activo
Andalucía	604906	222072	183773	120660	50308	255114	114433	281153	83027	1915446
Aragón	547436	255127	201639	125630	50268	246867	110719	262864	81832	1882382
Asturias	586561	280510	232693	130525	57682	336226	150503	310705	120725	2206130
Baleares	549567	227004	205880	143722	87284	357122	151111	333944	131271	2186905
Canarias	572256	185538	179574	134070	74322	305326	159185	279855	98012	1988138
Cantabria	587826	288598	261429	118398	63395	302423	116816	276978	105707	2121570
Castilla, La Mancha	546618	217780	191246	119423	41204	251981	109661	259545	107719	1845177
Castilla, León	543406	220646	210043	125781	53318	243869	96367	252823	101476	1847729
Cataluña	685925	262378	282701	162944	92869	360872	227525	362571	107370	2545155
Comunidad Valenciana	541577	217826	177228	132318	61803	280136	122874	281020	99310	1914092
Extremadura	469635	211088	137556	100379	43829	207527	86593	222698	71039	1550344
Galicia	615209	247674	198830	131660	51307	291052	128064	255713	104412	2023921
Madrid	673620	253666	252591	145747	86853	369559	215523	432997	129751	2560307
Murcia	603683	210393	189394	127725	46726	318856	103650	310354	113789	2024570
Navarra	643258	324500	251319	220787	80939	406681	185658	408892	154978	2677012
País Vasco	636443	267178	232181	157854	64959	342386	174320	394733	121880	2391934
Ríoja	602255	209670	196262	127146	54824	262629	126054	313107	121721	2013668
Ceuta y Melilla	683373	193283	133987	81436	27191	141950	84439	234961	64365	1644985
Margen activo	10693554	4294931	3718326	2406205	1089081	5280576	2463495	5474913	1918384	37339465

En la tabla de correspondencias se recogen las cantidades medias gastadas por hogar en cada Comunidad Autónoma, que oscilan entre 2.677.012 pesetas de Navarra y 1.644.985 pesetas de Ceuta y Melilla.

En las columnas aparecen las cantidades gastadas en cada concepto según los nueve grupos de gasto, se observa que en 1991 los hogares españoles en el Grupo 1 gastaron 10.693.554 pesetas, y considerando que el gasto total de las Comunidades fue 37.339.465 pesetas, el gasto en el Grupo 1 representa 0,286 (10.693.554 /37.339.465) de los gastos efectuados.

El gasto del Grupo 2 representa (masa) 2 el 0,115 (4.294.931 /37.339.465) de los gastos efectuados.

El gasto del Grupo 5 representa (masa) 2 el 0,029 (1.089.081 /37.339.465) de los gastos efectuados.

Perfiles de fila

CCAA	Gastos									
	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7	Grupo 8	Grupo 9	Margen activo
Andalucía	,316	,116	,096	,063	,026	,133	,060	,147	,043	1,000
Aragón	,291	,136	,107	,067	,027	,131	,059	,140	,043	1,000
Asturias	,266	,127	,105	,059	,026	,152	,068	,141	,055	1,000
Baleares	,251	,104	,094	,066	,040	,163	,069	,153	,060	1,000
Canarias	,288	,093	,090	,067	,037	,154	,080	,141	,049	1,000
Cantabria	,277	,136	,123	,056	,030	,143	,055	,131	,050	1,000
Castilla, La Mancha	,296	,118	,104	,065	,022	,137	,059	,141	,058	1,000
Castilla, León	,294	,119	,114	,068	,029	,132	,052	,137	,055	1,000
Cataluña	,270	,103	,111	,064	,036	,142	,089	,142	,042	1,000
Comunidad Valenciana	,283	,114	,093	,069	,032	,146	,064	,147	,052	1,000
Extremadura	,303	,136	,089	,065	,028	,134	,056	,144	,046	1,000
Galicia	,304	,122	,098	,065	,025	,144	,063	,126	,052	1,000
Madrid	,263	,099	,099	,057	,034	,144	,084	,169	,051	1,000
Murcia	,298	,104	,094	,063	,023	,157	,051	,153	,056	1,000
Navarra	,240	,121	,094	,082	,030	,152	,069	,153	,058	1,000
País Vasco	,266	,112	,097	,066	,027	,143	,073	,165	,051	1,000
Ríoja	,299	,104	,097	,063	,027	,130	,063	,155	,060	1,000
Ceuta y Melilla	,415	,117	,081	,050	,017	,086	,051	,143	,039	1,000
Masa	,286	,115	,100	,064	,029	,141	,066	,147	,051	

En los Perfiles de fila aparecen las masas descritas para los nueve grupos.

El dato de 0,286 corresponde a la masa (peso proporcional a su importancia en el conjunto) del perfil de la primera fila. La masa indica el promedio total de gasto, esto es, una Comunidad Autónoma ideal cuyo patrón de gasto coincide con el conjunto nacional.

Perfiles de columna

CCAA	Gastos									
	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7	Grupo 8	Grupo 9	Masa
Andalucía	,057	,052	,049	,050	,046	,048	,046	,051	,043	,051
Aragón	,051	,059	,054	,052	,046	,047	,045	,048	,043	,050
Asturias	,055	,065	,063	,054	,053	,064	,061	,057	,063	,059
Baleares	,051	,053	,055	,060	,080	,068	,061	,061	,068	,059
Canarias	,054	,043	,048	,056	,068	,058	,065	,051	,051	,053
Cantabria	,055	,067	,070	,049	,058	,057	,047	,051	,055	,057
Castilla, La Mancha	,051	,051	,051	,050	,038	,048	,045	,047	,056	,049
Castilla, León	,051	,051	,056	,052	,049	,046	,039	,046	,053	,049
Cataluña	,064	,061	,076	,068	,085	,068	,092	,066	,056	,068
Comunidad Valenciana	,051	,051	,048	,055	,057	,053	,050	,051	,052	,051
Extremadura	,044	,049	,037	,042	,040	,039	,035	,041	,037	,042
Galicia	,058	,058	,053	,055	,047	,055	,052	,047	,054	,054
Madrid	,063	,059	,068	,061	,080	,070	,087	,079	,068	,069
Murcia	,056	,049	,051	,053	,043	,060	,042	,057	,059	,054
Navarra	,060	,076	,068	,092	,074	,077	,075	,075	,081	,072
País Vasco	,060	,062	,062	,066	,060	,065	,071	,072	,064	,064
Ríoja	,056	,049	,053	,053	,050	,050	,051	,057	,063	,054
Ceuta y Melilla	,064	,045	,036	,034	,025	,027	,034	,043	,034	,044
Margen activo	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	

En la tabla Perfiles de columna para los distintos grupos, que se interpretan como la participación de cada Comunidad Autónoma en los grupos de gasto: Andalucía contribuye con un 5,7% (604.906 / 1.069.3554) del gasto total del Grupo 1.

En línea con lo expuesto, los perfiles son las proporciones de cada celdilla en función de la distribución marginal de la fila y la columna correspondiente. La importancia de esta información radica en que el análisis de correspondencias determina la distancia entre las modalidades considerando las distribuciones marginales y la masa de cada celda.

Este hecho obliga a considerar las diferencias entre los elementos en términos de distancias, de manera que cuanto mayor sea la diferencia entre perfiles con masa similar, mayor será la distancia entre las Comunidades.

ANÁLISIS PERFILES DE FILA:

Perfiles de fila

	Gastos									
CCAA	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7	Grupo 8	Grupo 9	Margen activo
Canarias	,288	,093	,090	,067	,037	,154	,080	,141	,049	1,000
Madrid	,263	,099	,099	,057	,034	,144	,084	,169	,051	1,000
Masa	,286	,115	,100	,064	,029	,141	,066	,147	,051	

Un análisis detallado muestra grandes similitudes en la distribución de gastos de Canarias y Madrid, concluyendo que ambas Comunidades tienen unos patrones de gastos similares.
Si se consideran las diferencias entre los elementos en función de las distancias, estas Comunidades tendrán escasas distancias. De forma análoga, cuanto mayor sea la diferencia entre los perfiles mayor será la distancia entre las Comunidades.

Resaltar que al establecer las diferencias entre las modalidades no hay que fijarse únicamente en las distancias entre los perfiles, también hay que prestar atención a las distancias ponderadas en función de su masa.

La masa (peso proporcional a su importancia en el conjunto) indica la influencia de un objeto basándose en su frecuencia marginal.

La masa afecta al centroide (baricentro), que es el perfil medio de la columna o fila. Las filas (o columnas) con una masa elevada: Por ejemplo, Navarra (0,072), Madrid (0,069) y Cataluña (0,068), influyen en la inercia aunque estén cerca del centroide.
Mientras que las filas (o columnas) con una masa pequeña: Extremadura (0,042), Ceuta y Melilla (0,044) influyen sobre la inercia únicamente cuando están lejos del centroide.

Tras los perfiles de filas y columnas, en una tabla-resumen donde aparecen el número de dimensiones calculadas, los valores propios o autovalores de cada dimensión, la inercia, el valor de Chi-cuadrado con su significación, la proporción de inercia explicada para cada dimensión, y la proporción explicada acumulada.

Resumen

					Proporción de inercia		Confianza para el Valor propio				
								Correlación			
Dimensión	Valor propio	Inercia	Chi-cuadrado	Sig.	Explicada	Acumulada	Desviación típica	2	3	4	5
1	,085	,007			,547	,547	,000	,033	,011	-,012	,007
2	,050	,002			,189	,737	,000		,035	-,006	,011
3	,037	,001			,106	,843	,000			,000	,000
4	,026	,001			,052	,894	,000				,012
5	,025	,001			,047	,941	,000				
6	,018	,000			,026	,967					
7	,016	,000			,020	,987					
8	,013	,000			,013	1,000					
Total		,013	489287,344	,000[a]	1,000	1,000					

a. 136 grados de libertad

- El número máximo de dimensiones es un número inferior al número de categorías de la variable con menos categorías (9 - 1 = 8). Al igual que el análisis de componentes principales, la primera dimensión explica la mayor cantidad de inercia, la segunda una cantidad menor, y así hasta la última.

Aunque el objetivo del análisis de correspondencias es representar la relación entre filas y columnas con el menor número de dimensiones, conviene comenzar el análisis solicitando un número elevado de dimensiones para observar la contribución relativa de cada una.

Se han solicitado cinco dimensiones, aunque SPSS proporciona información sobre el valor propio y el porcentaje de varianza explicada de todas las dimensiones. No ocurre lo mismo con la desviación típica y la correlación entre dimensiones, que únicamente presenta los valores de las dimensiones especificadas.

- En la segunda columna aparecen los valores propios o autovalores, es decir, la correlación entre las puntuaciones de filas y de columnas. La inercia de cada dimensión es el cuadrado de cada autovalor, que es una medida relativa de la importancia de cada dimensión.
- En la tercera columna el estadístico Chi-cuadrado, suma ponderada de todas las distancias al cuadrado entre los perfiles de fila y el perfil medio de la fila, y entre los perfiles de columna y el perfil medio de columna.
  Su significación 0,000 < 0,05 hace que se rechace la hipótesis nula, indicando una relación significativa entre las variables utilizadas. Cuanto mayor sea esta distancia menor será su significación, lo que conducirá a una elevada interrelación entre las variables, y una mayor diferencia de los perfiles respecto al perfil medio.
- La "inercia total" es la suma ponderada de todas las distancias al centroide (baricentro) , que coincide con la distancia Chi-cuadrado, dividida entre el total de casos:

  Chi-cuadrado / N = 489287,344 / 37339465 = 0,0131.

  Una vez que se conoce la inercia explicada por todas las dimensiones es sencillo valorar la aportación individual de cada una: el cociente de la inercia de cada dimensión entre el total (0,00717 / 0,013 1 = 0,547) informa que la primera dimensión explica el 54,7% de la inercia, mientras que la segunda dimensión sólo explica el 18,9% (0,00248 / 0,0131 = 0,189).

  En la siguiente columna se muestra la proporción de inercia acumulada, que indica la explicación conseguida cuando se considera un modelo con dos, tres o más dimensiones. En este sentido, el modelo con dos dimensiones explica el 73,7%, llegando a una explicación del 84,3% al seleccionar las tres primeras dimensiones.

Las siguientes casillas proporcionan información para realizar el intervalo de confianza de la población de que la se han extraído los datos, presentando la desviación típica para cada dimensión y la correlación entre las dimensiones solicitadas. Cuando las desviaciones típicas son pequeñas es probable que la solución presentada esté cerca del valor poblacional.

En la tabla aparecen desviaciones típicas muy pequeñas, indicando que los resultados serían muy similares aun cuando se realizase este análisis en otras muestras de igual tamaño, extraídas de la misma población y con un procedimiento idéntico.

Aparecen las tablas Examen de los puntos fila y Examen de los puntos columna:

Examen de los puntos de fila[a]

CCAA	Masa	Puntuación en la dimensión					Inercia	Contribución										
								De los puntos a la inercia de la dimensión					De la dimensión a la inercia del punto					
		1	2	3	4	5		1	2	3	4	5	1	2	3	4	5	Total
Andalucía	,051	-,240	,061	,009	,041	-,052	,000	,035	,004	,000	,003	,01	,840	,032	,000	,008	,012	,892
Aragón	,050	-,153	-,245	,190	,200	-,029	,000	,014	,061	,049	,078	,00	,260	,392	,176	,136	,003	,967
Asturias	,059	,105	-,205	,081	-,030	,114	,000	,008	,050	,011	,002	,03	,178	,403	,047	,005	,062	,696
Baleares	,059	,367	,048	-,171	-,189	-,079	,001	,093	,003	,046	,080	,01	,763	,008	,073	,062	,010	,916
Canarias	,053	,175	,335	,018	-,173	-,334	,001	,019	,120	,000	,062	,24	,220	,476	,001	,066	,236	,998
Cantabria	,057	-,052	-,422	,315	-,193	,135	,001	,002	,203	,151	,082	,04	,016	,610	,254	,067	,031	,978
Castilla, La Mancha	,049	-,127	-,140	-,096	-,069	,095	,000	,009	,019	,012	,009	,02	,327	,231	,081	,029	,054	,723
Castilla, León	,049	-,128	-,240	,025	-,138	,019	,000	,010	,057	,001	,036	,00	,198	,411	,003	,071	,001	,684
Cataluña	,068	,235	,256	,433	-,009	-,070	,001	,045	,090	,344	,000	,01	,303	,211	,452	,000	,008	,974
Comunidad Valenciana	,051	,053	,008	-,104	,023	-,155	,000	,002	,000	,015	,001	,05	,162	,002	,280	,009	,416	,869
Extremadura	,042	-,215	-,150	-,025	,231	-,150	,000	,023	,019	,001	,085	,04	,452	,130	,003	,160	,065	,810
Galicia	,054	-,142	-,124	,036	-,092	-,219	,000	,013	,017	,002	,018	,10	,340	,153	,010	,044	,236	,783
Madrid	,069	,264	,347	,074	,044	,278	,001	,057	,166	,010	,005	,21	,413	,420	,014	,004	,134	,986
Murcia	,054	-,072	-,031	-,385	-,243	,075	,000	,003	,001	,216	,123	,01	,049	,005	,607	,168	,015	,844
Navarra	,072	,323	-,171	-,201	,319	-,121	,001	,088	,042	,078	,281	,04	,582	,096	,099	,174	,024	,975
País Vasco	,064	,148	,097	-,069	,210	,202	,000	,017	,012	,008	,109	,10	,382	,096	,036	,238	,210	,963
Ríoja	,054	-,085	,114	-,198	-,077	,181	,000	,005	,014	,057	,012	,07	,124	,132	,295	,031	,163	,745
Ceuta y Melilla	,044	-1,037	,371	,000	,088	,003	,004	,559	,122	,000	,013	,00	,926	,070	,000	,002	,000	,998
Total activo	1,000						,013	1,00	1,00	1,000	1,000	1,0						

a. Normalización Simétrica

Examen de los puntos de fila[a]

CCAA	Contribución										
	De los puntos a la inercia de la dimensión					De la dimensión a la inercia del punto					
	1	2	3	4	5	1	2	3	4	5	Total
Andalucía	,035	,004	,000	,003	,01	,840	,032	,000	,008	,012	,892
Aragón	,014	,061	,049	,078	,00	,260	,392	,176	,136	,003	,967
Asturias	,008	,050	,011	,002	,03	,178	,403	,047	,005	,062	,696
Baleares	,093	,003	,046	,080	,01	,763	,008	,073	,062	,010	,916
Canarias	,019	,120	,000	,062	,24	,220	,476	,001	,066	,236	,998
Cantabria	,002	,203	,151	,082	,04	,016	,610	,254	,067	,031	,978
Castilla, La Mancha	,009	,019	,012	,009	,02	,327	,231	,081	,029	,054	,723
Castilla, León	,010	,057	,001	,036	,00	,198	,411	,003	,071	,001	,684
Cataluña	,045	,090	,344	,000	,01	,303	,211	,452	,000	,008	,974
Comunidad Valenciana	,002	,000	,015	,001	,05	,162	,002	,280	,009	,416	,869
Extremadura	,023	,019	,001	,085	,04	,452	,130	,003	,160	,065	,810
Galicia	,013	,017	,002	,018	,10	,340	,153	,010	,044	,236	,783
Madrid	,057	,166	,010	,005	,21	,413	,420	,014	,004	,134	,986
Murcia	,003	,001	,216	,123	,01	,049	,005	,607	,168	,015	,844
Navarra	,088	,042	,078	,281	,04	,582	,096	,099	,174	,024	,975
País Vasco	,017	,012	,008	,109	,10	,382	,096	,036	,238	,210	,963
Ríoja	,005	,014	,057	,012	,07	,124	,132	,295	,031	,163	,745
Ceuta y Melilla	,559	,122	,000	,013	,00	,926	,070	,000	,002	,000	,998
Total activo	1,00	1,00	1,000	1,000	1,0						

a. Normalización Simétrica

Examen de los puntos columna[a]

Gastos	Masa	Puntuación en la dimensión					Inercia	Contribución										
								De los puntos a la inercia de la dimensión					De la dimensión a la inercia del punto					
		1	2	3	4	5		1	2	3	4	5	1	2	3	4	5	Total
Grupo 1	,286	-,395	,127	-,010	-,066	-,060	,004	,529	,093	,001	,048	,041	,929	,056	,000	,008	,006	1,000
Grupo 2	,115	-,132	-,396	,129	,222	-,013	,001	,024	,362	,052	,217	,001	,126	,669	,053	,109	,000	,958
Grupo 3	,100	,071	-,224	,320	-,166	,194	,001	,006	,100	,274	,105	,150	,047	,275	,420	,078	,102	,922
Grupo 4	,064	,185	-,125	-,193	,262	-,312	,001	,026	,020	,065	,170	,252	,263	,070	,126	,161	,218	,838
Grupo 5	,029	,507	,212	,243	-,188	-,318	,001	,088	,026	,046	,040	,119	,614	,063	,062	,026	,071	,837
Grupo 6	,141	,318	-,091	-,112	-,167	-,089	,002	,169	,023	,048	,151	,045	,761	,037	,042	,064	,018	,922
Grupo 7	,066	,394	,445	,327	,151	-,025	,002	,121	,262	,190	,058	,002	,458	,343	,139	,021	,001	,962
Grupo 8	,147	,095	,171	-,162	,153	,244	,001	,016	,086	,103	,132	,350	,139	,262	,174	,109	,265	,949
Grupo 9	,051	,189	-,161	-,402	-,199	,140	,001	,022	,027	,223	,078	,041	,213	,092	,426	,073	,035	,837
Total activo	1,000						,013	1,000	1,000	1,000	1,000	1,000						

a. Normalización Simétrica

Examen de los puntos columna[a]

Gastos	Inercia	Contribución										
		De los puntos a la inercia de la dimensión					De la dimensión a la inercia del punto					
		1	2	3	4	5	1	2	3	4	5	Total
Grupo 1	,004	,529	,093	,001	,048	,041	,929	,056	,000	,008	,006	1,000
Grupo 2	,001	,024	,362	,052	,217	,001	,126	,669	,053	,109	,000	,958
Grupo 3	,001	,006	,100	,274	,105	,150	,047	,275	,420	,078	,102	,922
Grupo 4	,001	,026	,020	,065	,170	,252	,263	,070	,126	,161	,218	,838
Grupo 5	,001	,088	,026	,046	,040	,119	,614	,063	,062	,026	,071	,837
Grupo 6	,002	,169	,023	,048	,151	,045	,761	,037	,042	,064	,018	,922
Grupo 7	,002	,121	,262	,190	,058	,002	,458	,343	,139	,021	,001	,962
Grupo 8	,001	,016	,086	,103	,132	,350	,139	,262	,174	,109	,265	,949
Grupo 9	,001	,022	,027	,223	,078	,041	,213	,092	,426	,073	,035	,837
Total activo	,013	1,000	1,000	1,000	1,000	1,000						

a. Normalización Simétrica

La tabla muestra la masa (peso proporcional a su importancia en el conjunto) de cada columna, las coordenadas de las columnas en los factores, la contribución de cada columna a la inercia total, las contribuciones absolutas y relativas.

La inercia de una dimensión es la suma ponderada de las distancias al cuadrado entre cada punto y el centro de coordenadas. Las "contribuciones absolutas" aparecen bajo el título "Contribución de los puntos a la inercia de la dimensión", y aparece la contribución de cada punto a la definición de la dimensión o, expresado de otra forma, "la proporción de la inercia explicada por un factor debida a cada categoría". La suma de las contribuciones absolutas de todas las modalidades en un factor será igual a la unidad.

"La contribución de cada punto" depende de la distancia de cada categoría al centro de gravedad y de su masa, siendo el producto de la coordenada al cuadrado por la masa de la categoría, dividido por el valor propio del factor:

La contribución 0,529 de los gastos del Grupo 1 se obtiene de la expresión:

$[(-0{,}395)^2 \cdot 0{,}286] / 0{,}085 = 0{,}529.$

Con este enfoque, se puede afirmar que las contribuciones absolutas "ponderan" la distancia de cada modalidad en función de su masa, de modo que no basta con observar las coordenadas (como en el análisis de componentes principales) sino que es necesario considerar el número de elementos incluidos en cada modalidad, su masa.

Esto explica que puntos lejanos del centroide (baricentro) tengan contribuciones absolutas pequeñas, y viceversa, que Comunidades con coordenadas bajas consigan elevadas contribuciones en la explicación del factor.

- Las modalidades de gastos dominantes del primer factor son los gastos en el Grupo 1 (con una contribución de 0,529 a la inercia del eje), seguidos de los gastos en el Grupo 6 (con una contribución de 0,169 a la inercia del eje), y en el Grupo 7 (con una contribución de 0,121). Con estas tres modalidades (Grupo 1, Grupo 6, y Grupo 7) se explica el 82% de la inercia de este primer factor.
  El primer factor explica el 54,7% de la inercia total (tabla-resumen).
- En el segundo factor los gastos en Grupo 2 (con una contribución de 0,362 a la inercia del eje) y Grupo 7 (con una contribución de 0,262) llegan a explicar el 62,4% (0,362 + 0,262 = 0,624) de la inercia del segundo factor. El segundo factor explica el 18,9% de la inercia total (tabla-resumen).
- El tercer factor está explicado fundamentalmente por los gastos en el Grupo 3 (con una contribución de 0,274 a la inercia del eje), gastos en el Grupo 8 (contribución de 0,103) y gastos en el Grupo 9 (contribución de 0,223), llegando a explicar el 60% (0,274 + 0,103 + 0,223 = 0,6) de la inercia del tercer factor. El tercer factor sólo explica el 10,6% de la inercia total (tabla-resumen).

- En el cuarto factor destaca la contribución en el Grupo 2 (0,217), Grupo 6 (0,151), y Grupo 8 (0,132). En definitiva puede decirse que no tiene ninguna contribución absoluta elevada.
  El cuarto factor sólo explica el 5,2% de la inercia total (tabla resumen).
- En el quinto factor destacan las aportaciones de los gastos del Grupo 4 (0,252), gastos del Grupo 5 (0,119) y Grupo 8 (0,350), llegando a explicar el 72,1% de la inercia del quinto factor. El quinto factor sólo explica el 4,7% de la inercia total (tabla-resumen).

Las contribuciones relativas expresan la contribución de cada factor a la inercia de cada categoría, es decir, "la calidad de la representación de un punto sobre una dimensión".

La contribución relativa se calcula dividiendo la coordenada al cuadrado entre la distancia de la categoría al centro de gravedad, y la suma de las contribuciones relativas de una modalidad en todos los factores es igual a la unidad.

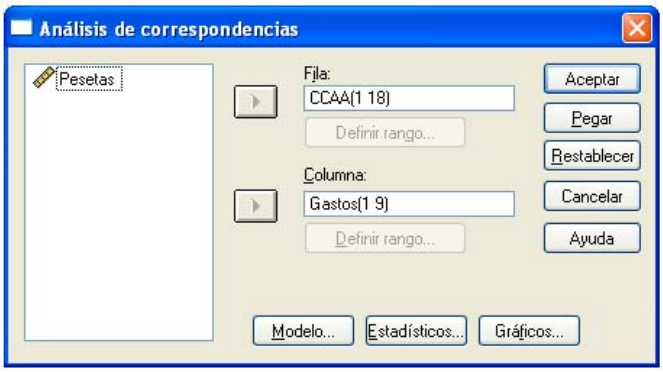
La última columna de las contribuciones relativas, con valores cercanos a la unidad, refleja la elevada calidad de todos los puntos sobre las cinco dimensiones; si bien las tres primeras dimensiones explican la mayor parte de los grupos de gasto, a excepción del gasto en Grupo 4 y Grupo 9.

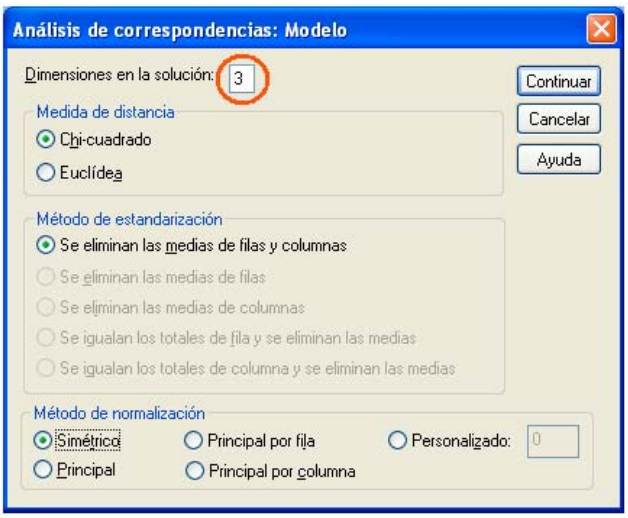
INFORME PUNTUAL DE LA INVESTIGACIÓN: El objetivo del Análisis de correspondencias es reducir una serie de variables observadas a un conjunto menor de variables latentes llamadas factores, con la menor pérdida de información.

En un principio se hacia referencia a la conveniencia de extraer un número elevado de dimensiones para analizar la contribución de cada una, a la vez que se ha mencionado en la conveniencia de buscar el equilibrio entre parsimonia e interpretabilidad.

En la tabla-resumen se observa que con cinco factores se explica el 94% de la inercia, mientras que con tres factores se explica el 84,3% de la inercia total. En esta línea, es aconsejable realizar el análisis con los "tres primeros factores", decisión que queda avalada por la ausencia de elevadas contribuciones absolutas a partir de la cuarta dimensión.

Para construir el nuevo modelo se modifica la línea /DIMENSION=5 con el editor de sintaxis, o se cambia el número de dimensiones en el cuadro de diálogo de la figura adjunta





Los resultados analizados hasta el momento no cambian, a excepción de las contribuciones relativas:

Resumen

Dimensión	Valor propio	Inercia	Chi-cuadrado	Sig.	Proporción de inercia: Explicada	Proporción de inercia: Acumulada	Confianza para el Valor propio: Desviación típica	Correlación 2	Correlación 3
1	,085	,007			,547	,547	,000	,033	,011
2	,050	,002			,189	,737	,000		,035
3	,037	,001			,106	,843	,000		
4	,026	,001			,052	,894			
5	,025	,001			,047	,941			
6	,018	,000			,026	,967			
7	,016	,000			,020	,987			
8	,013	,000			,013	1,000			
Total		,013	489287,344	,000[a]	1,000	1,000			

a. 136 grados de libertad

Se obtienen dos factores que conjuntamente explican el 73,7% de la inercia total.

Examen de los puntos de fila[a]

CCAA	Masa	Puntuación en la dimensión 1	Puntuación en la dimensión 2	Puntuación en la dimensión 3	Inercia	Contribución: De los puntos a la inercia de la dimensión 1	Contribución: De los puntos a la inercia de la dimensión 2	Contribución: De los puntos a la inercia de la dimensión 3	Contribución: De la dimensión a la inercia del punto 1	Contribución: De la dimensión a la inercia del punto 2	Contribución: De la dimensión a la inercia del punto 3	Contribución: De la dimensión a la inercia del punto Total
Andalucía	,051	-,240	,061	,009	,000	**,035**	,004	,000	,840	,032	,000	**,873**
Aragón	,050	-,153	-,245	,190	,000	,014	,061	,049	,260	,392	,176	**,828**
Asturias	,059	,105	-,205	,081	,000	,008	,050	,011	,178	,403	,047	**,629**
Baleares	,059	,367	,048	-,171	,001	**,093**	,003	,046	,763	,008	,073	**,844**
Canarias	,053	,175	,335	,018	,001	,019	**,120**	,000	,220	,476	,001	**,696**
Cantabria	,057	-,052	-,422	,315	,001	,002	**,203**	**,151**	,016	,610	,254	**,880**
Castilla, La Mancha	,049	-,127	-,140	-,096	,000	,009	,019	,012	,327	,231	,081	**,640**
Castilla, León	,049	-,128	-,240	,025	,000	,010	,057	,001	,198	,411	,003	**,612**
Cataluña	,068	,235	,256	,433	,001	**,045**	,090	**,344**	,303	,211	,452	**,966**
Comunidad Valenciana	,051	,053	,008	-,104	,000	,002	,000	,015	,162	,002	,280	**,444**
Extremadura	,042	-,215	-,150	-,025	,000	,023	,019	,001	,452	,130	,003	**,585**
Galicia	,054	-,142	-,124	,036	,000	,013	,017	,002	,340	,153	,010	**,503**
Madrid	,069	,264	,347	,074	,001	**,057**	**,166**	,010	,413	,420	,014	**,848**
Murcia	,054	-,072	-,031	-,385	,000	,003	,001	**,216**	,049	,005	,607	**,661**
Navarra	,072	,323	-,171	-,201	,001	**,088**	**,042**	**,078**	,582	,096	,099	**,777**
País Vasco	,064	,148	,097	-,069	,000	,017	,012	,008	,382	,096	,036	**,515**
Ríoja	,054	-,085	,114	-,198	,000	,005	,014	,057	,124	,132	,295	**,551**
Ceuta y Melilla	,044	-1,037	,371	,000	,004	**,559**	**,122**	,000	,926	,070	,000	**,996**
Total activo	1,000				,013	1,000	1,000	1,000				

a. Normalización Simétrica

Examen de los puntos columna[a]

Gastos	Masa	Puntuación en la dimensión 1	Puntuación en la dimensión 2	Puntuación en la dimensión 3	Inercia	Contribución: De los puntos a la inercia de la dimensión 1	Contribución: De los puntos a la inercia de la dimensión 2	Contribución: De los puntos a la inercia de la dimensión 3	Contribución: De la dimensión a la inercia del punto 1	Contribución: De la dimensión a la inercia del punto 2	Contribución: De la dimensión a la inercia del punto 3	Contribución: De la dimensión a la inercia del punto Total
Grupo 1	,286	-,395	,127	-,010	,004	**,529**	,093	,001	,929	,056	,000	**,986**
Grupo 2	,115	-,132	-,396	,129	,001	,024	**,362**	,052	,126	,669	,053	**,848**
Grupo 3	,100	,071	-,224	,320	,001	,006	,100	**,274**	,047	,275	,420	**,741**
Grupo 4	,064	,185	-,125	-,193	,001	,026	,020	,065	,263	,070	,126	**,459**
Grupo 5	,029	,507	,212	,243	,001	,088	,026	,046	,614	,063	,062	**,740**
Grupo 6	,141	,318	-,091	-,112	,002	**,169**	,023	,048	,761	,037	,042	**,840**
Grupo 7	,066	,394	,445	,327	,002	**,121**	**,262**	,190	,458	,343	,139	**,941**
Grupo 8	,147	,095	,171	-,162	,001	,016	,086	**,103**	,139	,262	,174	**,575**
Grupo 9	,051	,189	-,161	-,402	,001	,022	,027	**,223**	,213	,092	,426	**,730**
Total activo	1,000				,013	1,000	1,000	1,000				

a. Normalización Simétrica

Una primera observación de las contribuciones absolutas del primer factor refleja la gran aportación de dos grupos de gasto: Grupo 1 (Alimentación, bebidas y tabaco) con una contribución absoluta de 0,529, Grupo 6 (Transportes y comunicaciones) con una contribución absoluta de 0,169. Estas dos modalidades explican el 69,8% de la inercia del factor.

Las contribuciones relativas mejoran la interpretación al desvelar que el Grupo 5 (Servicios médicos y gastos sanitarios), con una contribución relativa de 0,614 se encuentran fuertemente explicados por este primer factor.

El análisis de coordenadas indica que los gastos en el Grupo 1 (Alimentación, bebidas y tabaco) se encuentran situados en la parte izquierda del factor (- 0,395) (Tabla examen de los puntos columna con tres dimensiones, Gráfico de línea de las columnas) con una contribución absoluta de 0,529, mientras que a la derecha del origen de coordenadas se encuentran los gastos del Grupo 6 (Transportes y comunicaciones) con una representación de 0,318, contribución absoluta de 0,169 y Grupo 5 (servicios médicos y gastos sanitarios) con una representación de 0,507, contribución absoluta 0,088.

Aunque en la parte derecha de este primer factor aparece también el Grupo 4 (Muebles, enseres y servicios del hogar), el análisis de contribución absoluta (0,026) desvela que este grupo de gasto tiene escasa influencia en la definición del factor. La baja contribución relativa de (0,263) está indicando que estos gastos están muy poco definidos por el modelo de los dos factores.

Concluyendo, el primer factor que explica el 54,7% de la inercia total (tabla resumen: proporción inercia explicada), está señalando una oposición entre los gastos del Grupo 5 (Servicios médicos y gastos sanitarios) y Grupo 6 (Transportes y comunicaciones), frente al gasto en el Grupo I (Alimentos, bebidas y tabaco).
Los tipos de gastos definitorios del primer factor, junto con las Comunidades Autónomas en los que predominan, aparecen representados en los extremos derecho e izquierda del gráfico puntos de columna para gastos.

El análisis de las contribuciones relativas de las dos tablas (examen de los puntos fila y columna con tres dimensiones), con valores cercanos a la unidad en la última columna, sigue reflejando una elevada calidad de la representación de las categorías en los factores, aunque determinados grupos de gasto (Grupo 4 y Grupo 8), así como algunas Comunidades Autónomas (Valencia, Galicia, País Vasco, etc.) disminuyan ligeramente su representatividad.

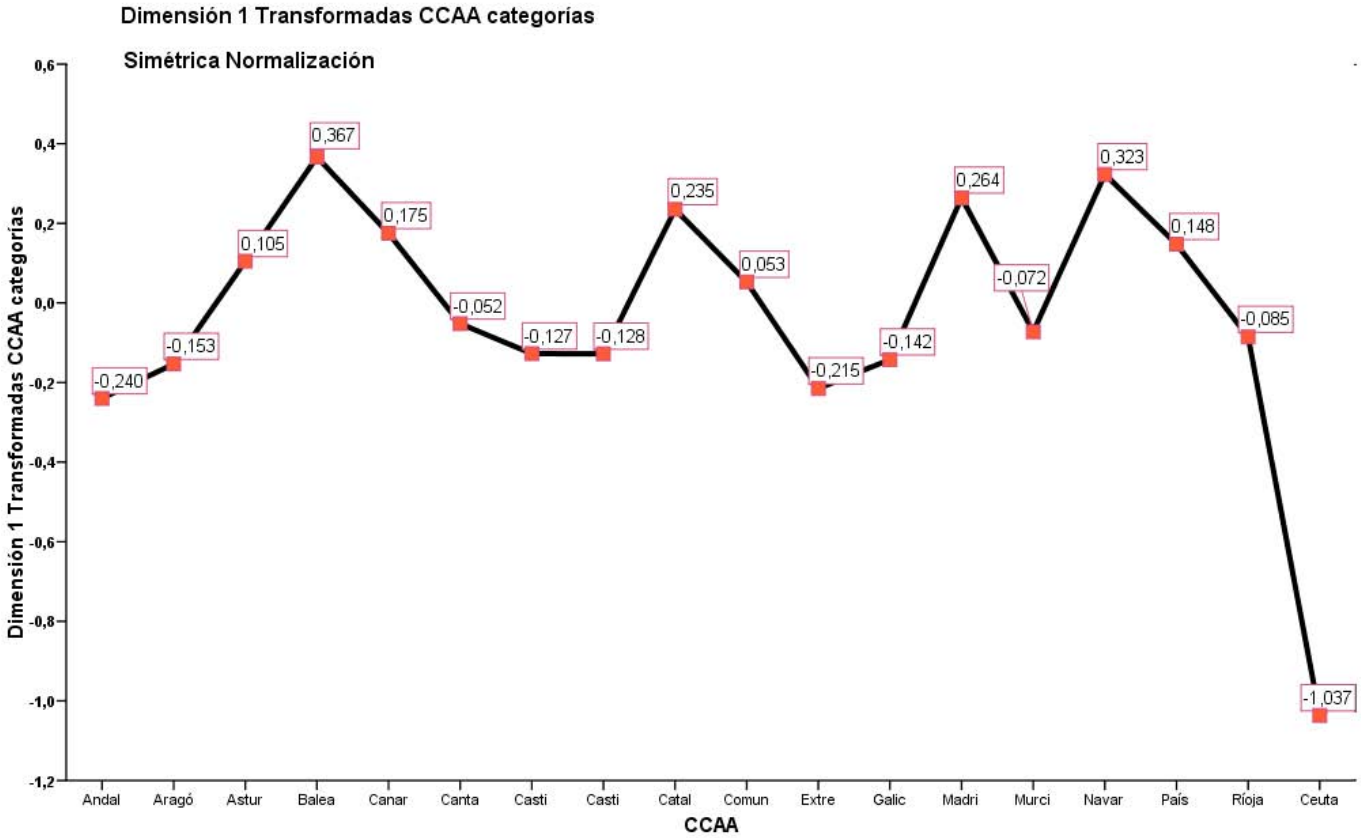
ANÁLISIS COMPONENTES GRÁFICOS:

En las tablas de examen de los puntos fila y columna con tres dimensiones se recogen las puntuaciones de columnas y filas que serán utilizadas como coordenadas en las representaciones gráficas.

Se comienzan analizando los Gráficos de Líneas en la medida que implican el acercamiento más sencillo a la realidad estudiada.

Estos gráficos no son más que una representación de las puntuaciones de las categorías de cada dimensión, siendo muy ilustrativos para conocer la distancia de cada categoría al punto de coordenadas.

Al estar representadas las coordenadas de cada categoría estos gráficos permiten una rápida comparación entre las distancias de las categorías respecto al origen de coordenadas, conociendo así los elementos que definen cada factor.



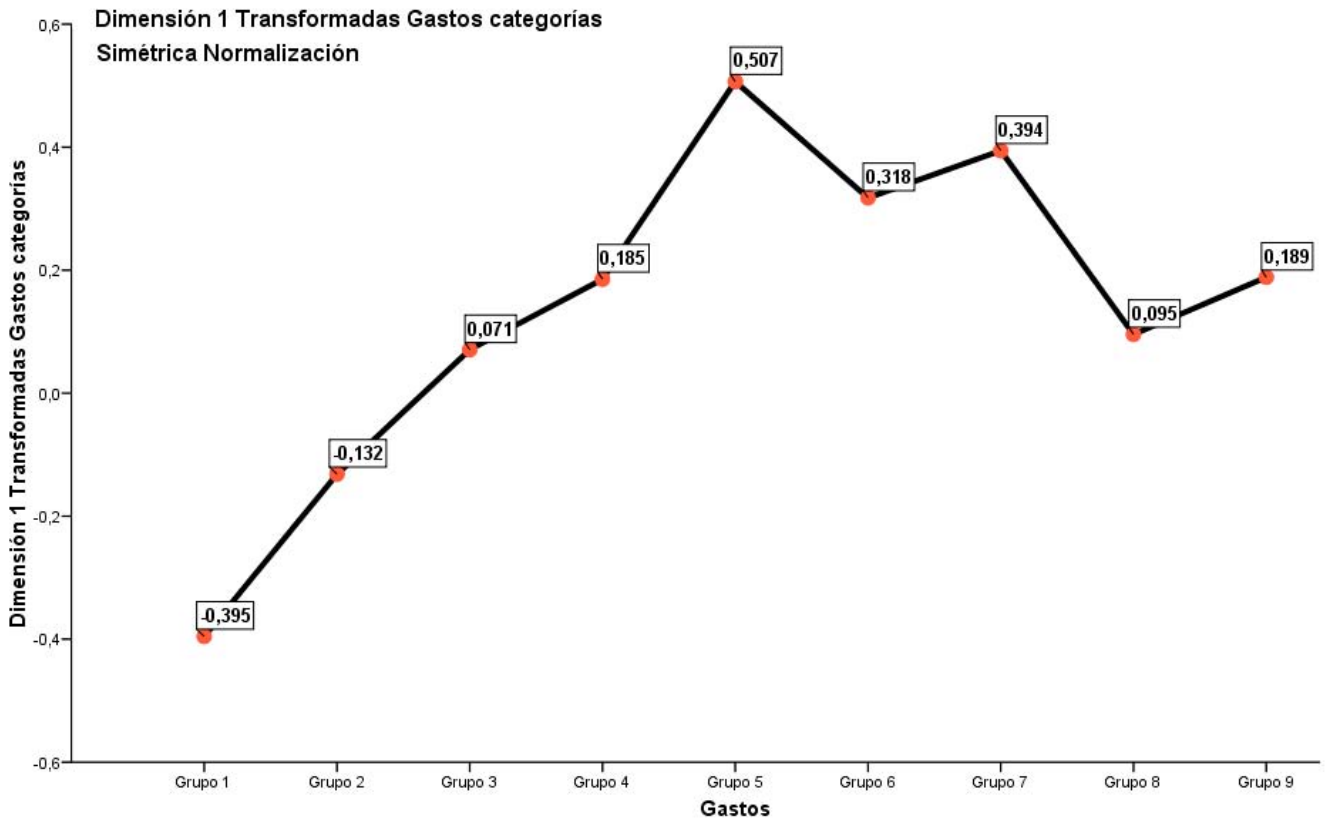
En la tabla de Examen de los puntos de Fila (primer factor): Andalucía tiene una puntuación de (– 0,240) en la primera dimensión, razón por la que esta Comunidad Autónoma aparece situada algo más abajo del valor (- 0,20). En la parte derecha, Ceuta y Melilla, con una puntuación de (- 1,037), es el valor más bajo de todos los valores considerados.

Alta presencia de gastos en Andalucía, Baleares, Cataluña, Madrid y Navarra.

En el gráfico de Examen de Líneas de Fila (primer factor) aparecen en la parte negativa las Comunidades de Ceuta y Melilla, Andalucía, Extremadura y Galicia.

Se podía pensar que en casos como las modalidades de Extremadura (-0,215) y Galicia (-0,142), al aparecer juntas tienen perfiles semejantes. Cuando la calidad de representación es baja se debe matizar esta semejanza: Estos tres factores explican el 50,3% de Galicia, es decir, la semejanza se produce considerando únicamente la mitad de esta modalidad, pudiendo suceder que estas Comunidades sean muy diferentes en cualquier otro aspecto no recogido por estos tres factores. En cualquier otro aspecto incluido en el 50% restante que queda sin especificar.

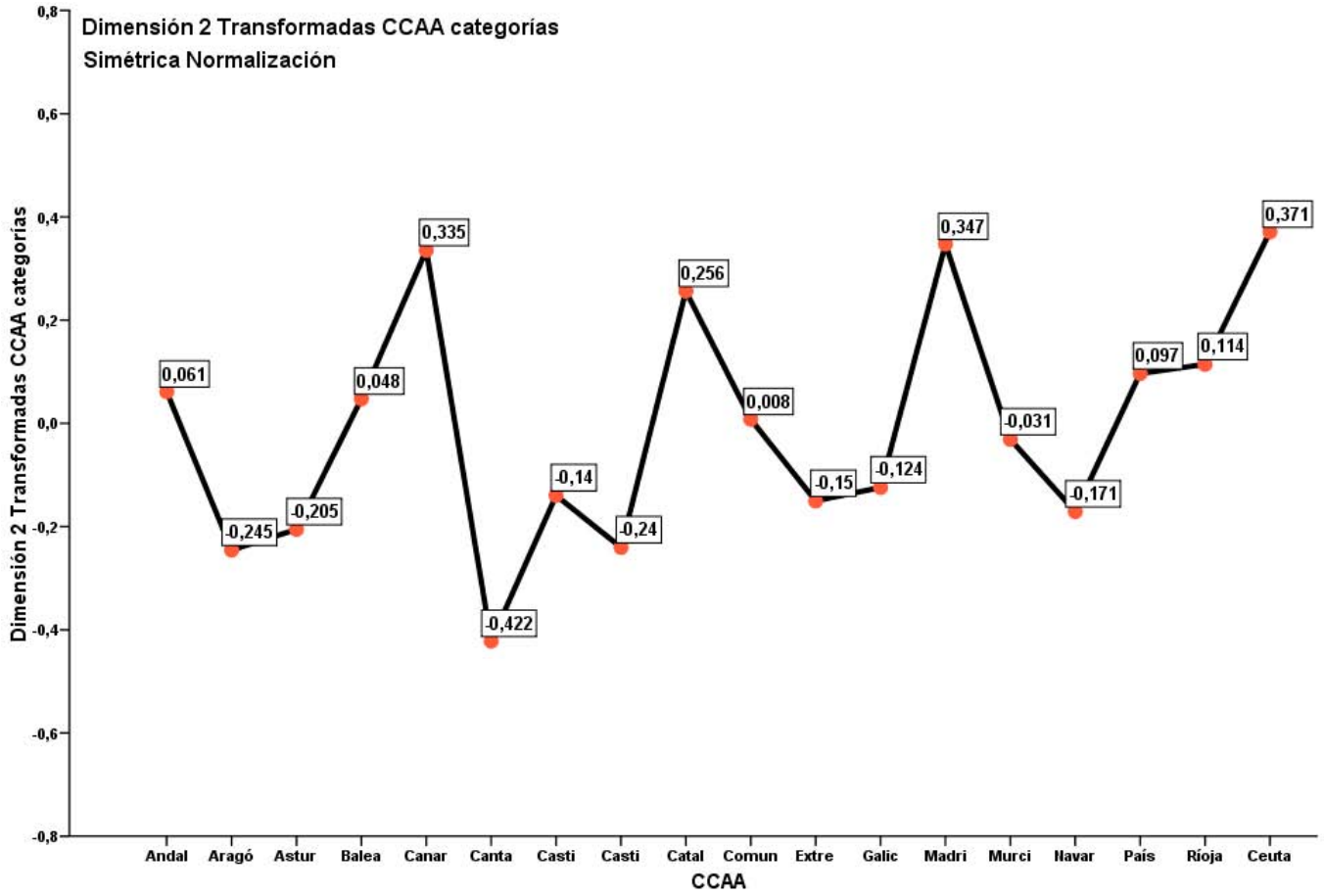
Para analizar el origen de este problema se considera que Galicia se diferencia del perfil medio, y de otras Comunidades, en grupos de gastos distintos a los destacados por estos tres factores. De hecho, al observar la tabla Examen de los puntos de Fila (cinco dimensiones) se aprecia que los cinco primeros factores explican el 78,3% de esta Comunidad, destacando la elevada contribución de Galicia en el quinto factor.



En el gráfico de Examen de Líneas de Columna (primer factor) se desvela que los gastos del Grupo 9, Grupo 8 y Grupo 1 están muy lejos del origen de coordenadas, respectivamente, los dos primeros situados en la parte positiva y el tercero en la parte negativa.

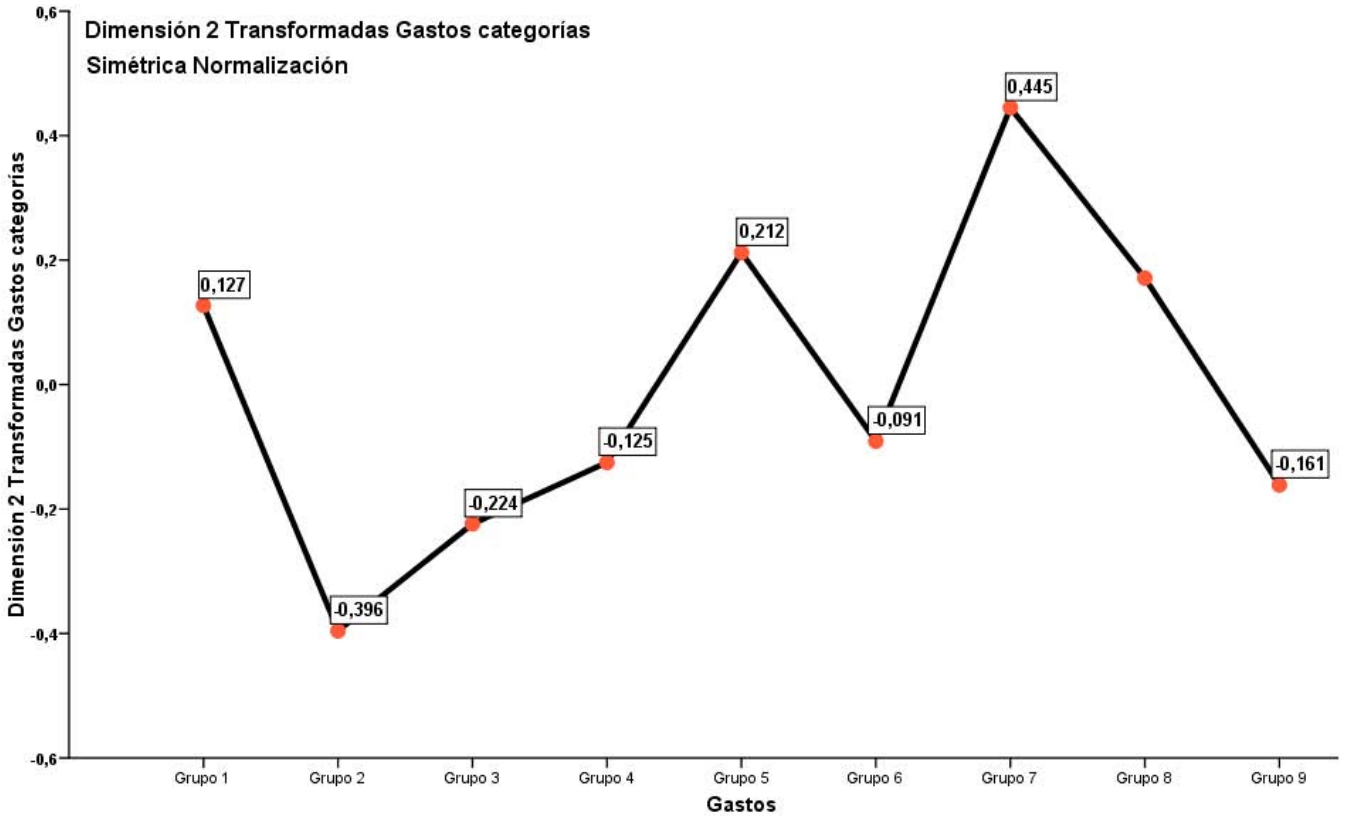
Los gastos del Grupo 8 y Grupo 3 se encuentran situados muy cerca del origen de coordenadas, de forma que tienen un escaso poder discriminante en la primera dimensión.

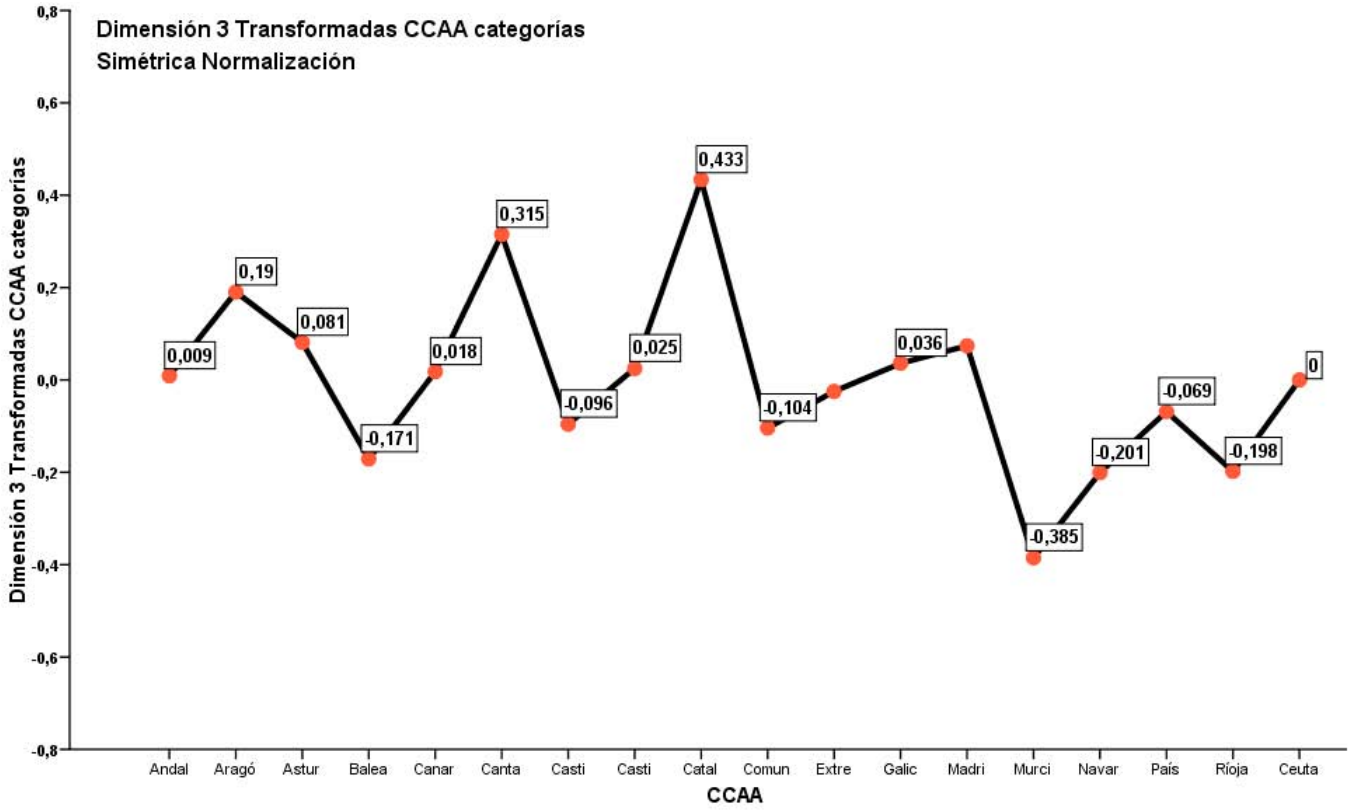
Alta presencia de gastos en el Grupo 1, Grupo 6 y Grupo 7 (extremo positivo del primer factor): Baleares, Navarra, Madrid y Cataluña

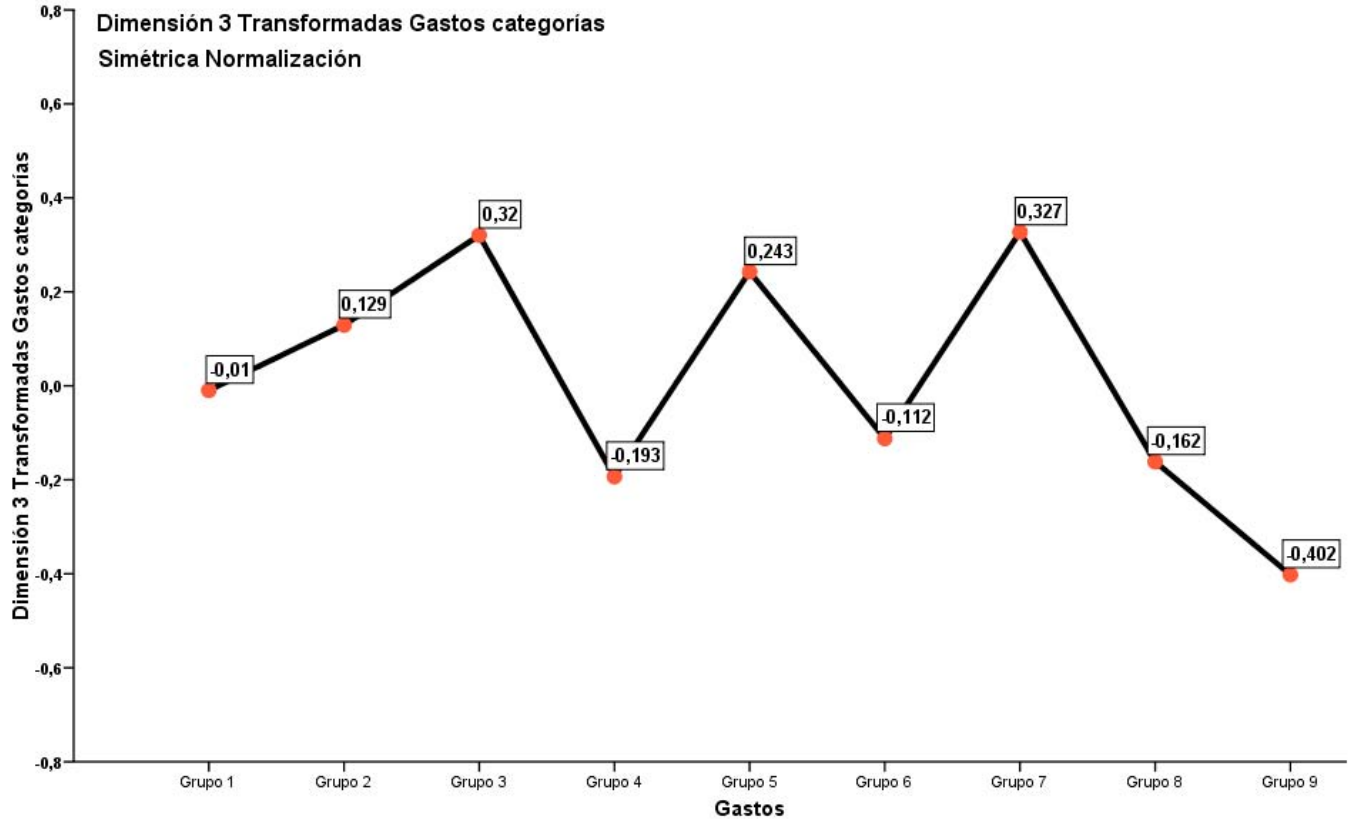


En la segunda dimensión se aprecian las puntuaciones extremas de los gastos del Grupo 2 y Grupo 7, el primero con signo negativo y el segundo positivo, que serán elementos definitorios de la segunda dimensión. Presencia elevada de gastos en el Grupo 7 (extremo positivo del segundo factor): Canarias, Cantabria, Madrid, Ceuta y Melilla

Predominio de Grupo 2 (parte negativa del segundo factor): Cantabria y Aragón.







En la tercera dimensión, se recogen puntuaciones extremas en el Grupo 7 y Grupo 3 (signo positivo) y en Grupo 9 (parte negativa).

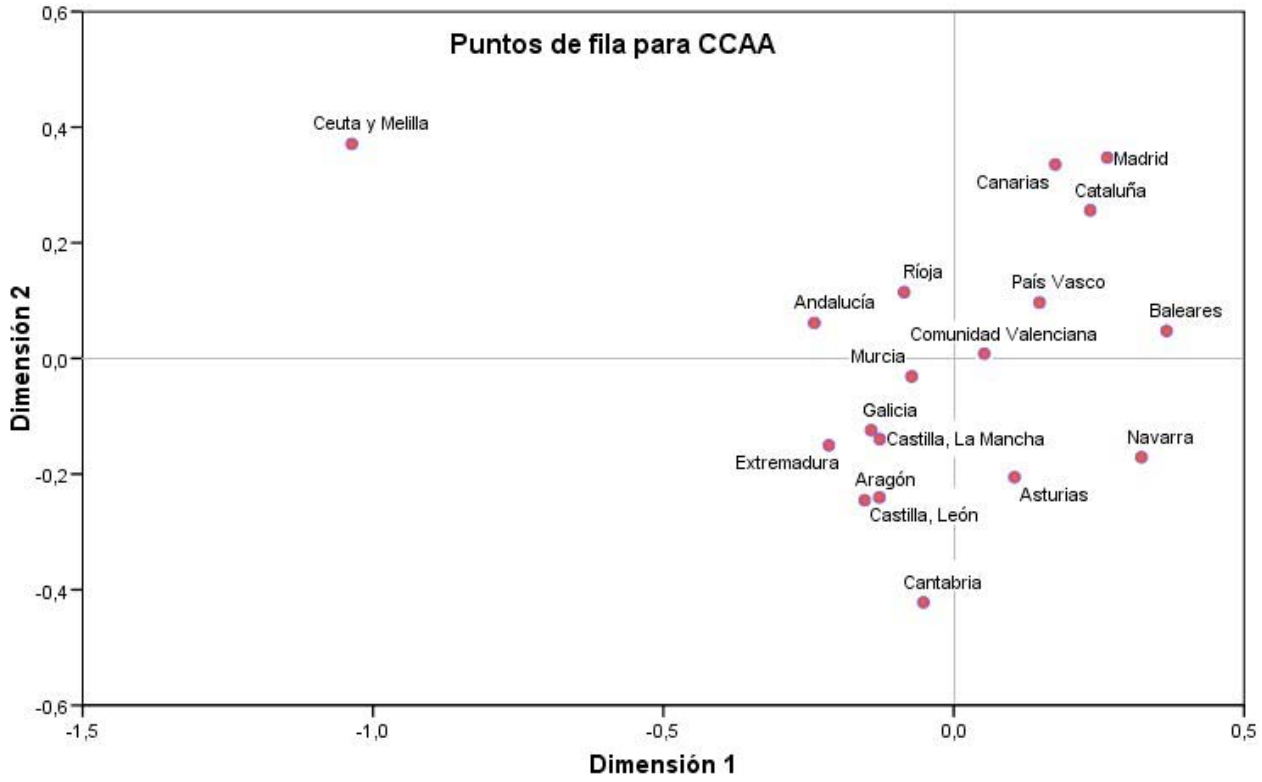
Elevados gastos en Grupo 9 (extremo negativo del tercer factor): Murcia y Navarra.

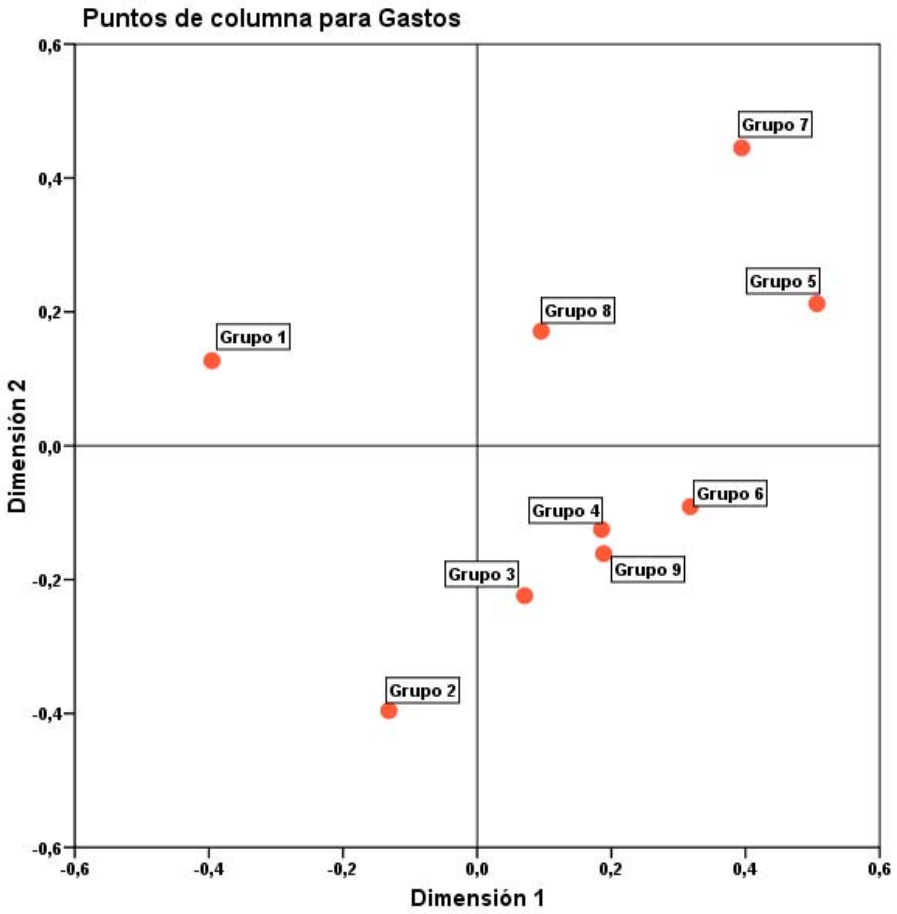
Predomino de gastos en Grupo 7 y Grupo 3 (parte positiva del tercer factor): Cataluña y Cantabria.

Las representaciones gráficas anteriores, con una gran facilidad de interpretación, dejan aspectos sin explicar, razón por la que es necesario analizar los gráficos de dispersión que presentan las puntuaciones de fila (o columna) en los diversos factores.

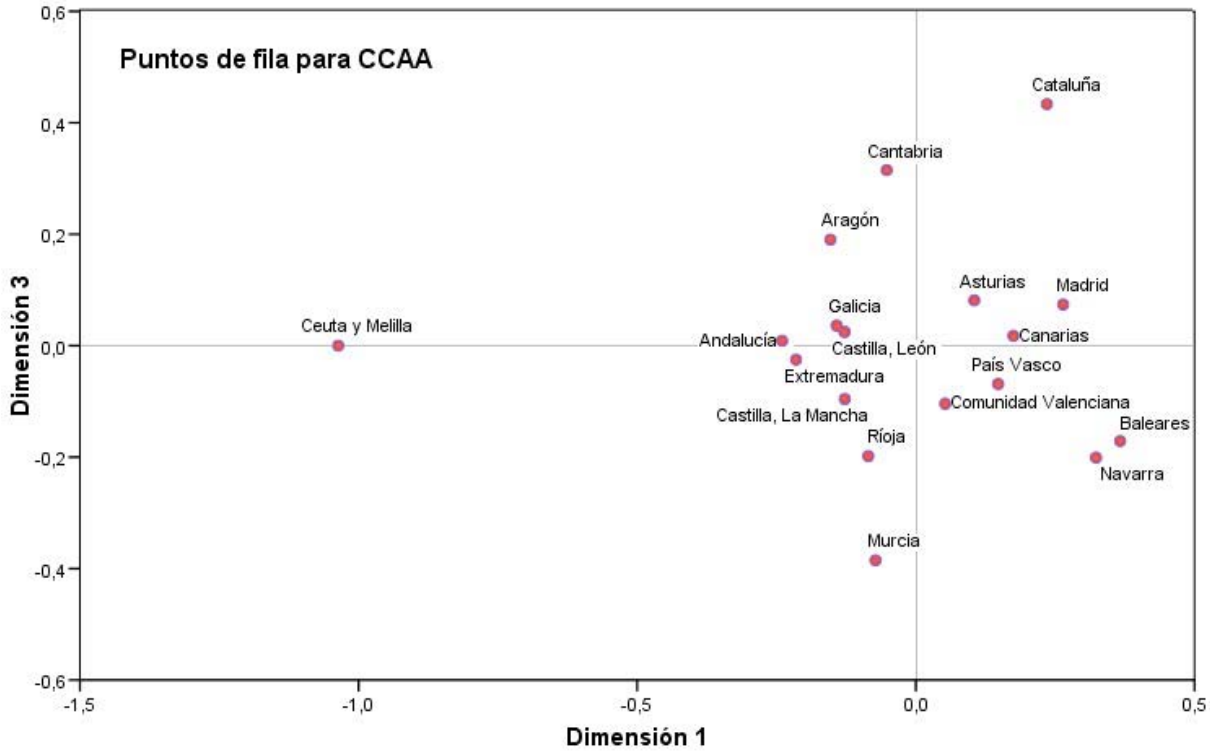
Mientras que los gráficos de línea únicamente consideraban la puntuación de cada categoría en un factor, los gráficos de dispersión realizan una representación en varias dimensiones reflejando una mejor caracterización de la realidad.

Las puntuaciones de cada Comunidad (filas) en las tres dimensiones seleccionadas se combinan para realizar una representación de los dos primeros factores, o del primero con el tercer factor.





Los tipos de gastos definitorios del primer factor, junto con las Comunidades Autónomas en los que predominan, aparecen representados en los extremos derecho e izquierda del gráfico.


Puntos de fila para CCAA
Cataluña
Cantabria
Aragón
Asturias
Madrid
Galicia
Ceuta y Melilla
Andalucía
Castilla, León
Canarias
País Vasco
Extremadura
Comunidad Valenciana
Castilla, La Mancha
Baleares
Ríoja
Navarra
Murcia
Dimensión 3
Dimensión 1
0,6
0,4
0,2
0,0
-0,2
-0,4
-0,6
-1,5
-1,0
-0,5
0,0
0,5

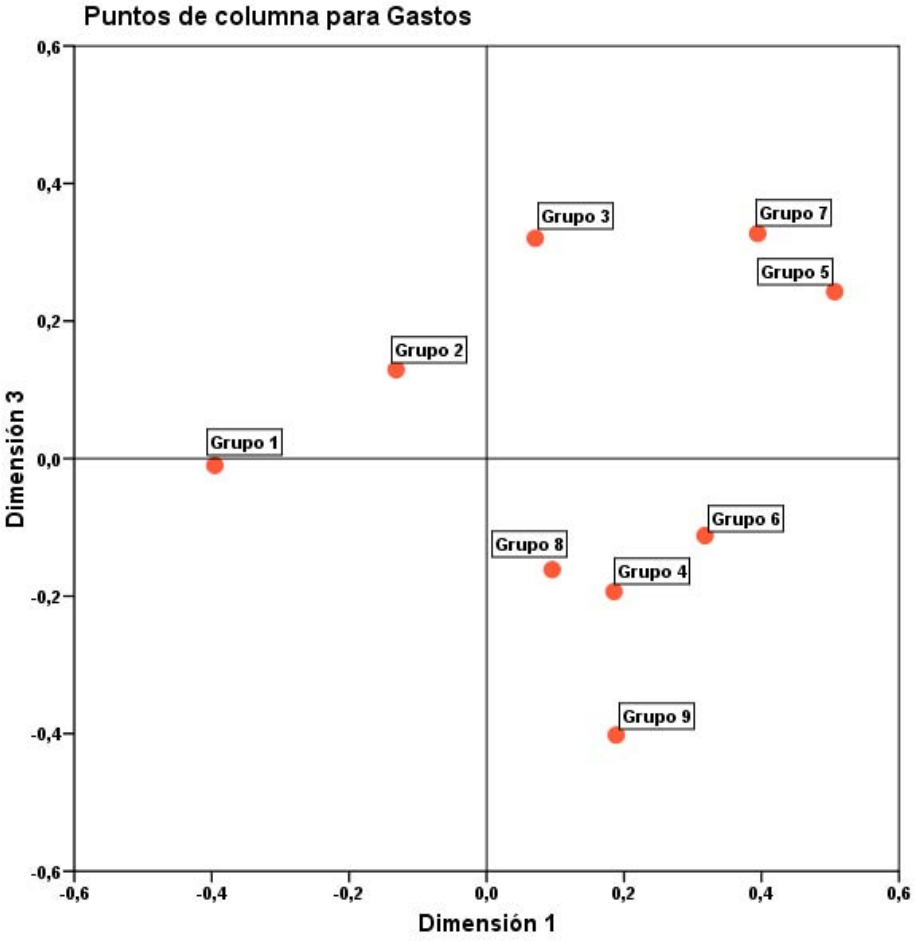

Puntos de columna para Gastos
Grupo 3
Grupo 7
Grupo 5
Grupo 2
Grupo 1
Grupo 6
Grupo 8
Grupo 4
Grupo 9
Dimensión 3
Dimensión 1
0,6
0,4
0,2
0,0
-0,2
-0,4
-0,6
-0,6
-0,4
-0,2
0,0
0,2
0,4
0,6

Tabla de correspondencias permutada de acuerdo con la dimensión 1

	Gastos									
CCAA	Grupo 1	Grupo 2	Grupo 3	Grupo 8	Grupo 4	Grupo 9	Grupo 6	Grupo 7	Grupo 5	Margen activo
Ceuta y Melilla	683373	193283	133987	234961	81436	64365	141950	84439	27191	1644985
Andalucía	604906	222072	183773	281153	120660	83027	255114	114433	50308	1915446
Extremadura	469635	211088	137556	222698	100379	71039	207527	86593	43829	1550344
Aragón	547436	255127	201639	262864	125630	81832	246867	110719	50268	1882382
Galicia	615209	247674	198830	255713	131660	104412	291052	128064	51307	2023921
Castilla, León	543406	220646	210043	252823	125781	101476	243869	96367	53318	1847729
Castilla, La Mancha	546618	217780	191246	259545	119423	107719	251981	109661	41204	1845177
Ríoja	602255	209670	196262	313107	127146	121721	262629	126054	54824	2013668
Murcia	603683	210393	189394	310354	127725	113789	318856	103650	46726	2024570
Cantabria	587826	288598	261429	276978	118398	105707	302423	116816	63395	2121570
Comunidad Valenciana	541577	217826	177228	281020	132318	99310	280136	122874	61803	1914092
Asturias	586561	280510	232693	310705	130525	120725	336226	150503	57682	2206130
País Vasco	636443	267178	232181	394733	157854	121880	342386	174320	64959	2391934
Canarias	572256	185538	179574	279855	134070	98012	305326	159185	74322	1988138
Cataluña	685925	262378	282701	362571	162944	107370	360872	227525	92869	2545155
Madrid	673620	253666	252591	432997	145747	129751	369559	215523	86853	2560307
Navarra	643258	324500	251319	408892	220787	154978	406681	185658	80939	2677012
Baleares	549567	227004	205880	333944	143722	131271	357122	151111	87284	2186905
Margen activo	10693554	4294931	3718326	5474913	2406205	1918384	5280576	2463495	1089081	37339465

Tabla de correspondencias permutada de acuerdo con la dimensión 2

	Gastos									
CCAA	Grupo 2	Grupo 3	Grupo 9	Grupo 4	Grupo 6	Grupo 1	Grupo 8	Grupo 5	Grupo 7	Margen activo
Cantabria	288598	261429	105707	118398	302423	587826	276978	63395	116816	2121570
Aragón	255127	201639	81832	125630	246867	547436	262864	50268	110719	1882382
Castilla, León	220646	210043	101476	125781	243869	543406	252823	53318	96367	1847729
Asturias	280510	232693	120725	130525	336226	586561	310705	57682	150503	2206130
Navarra	324500	251319	154978	220787	406681	643258	408892	80939	185658	2677012
Extremadura	211088	137556	71039	100379	207527	469635	222698	43829	86593	1550344
Castilla, La Mancha	217780	191246	107719	119423	251981	546618	259545	41204	109661	1845177
Galicia	247674	198830	104412	131660	291052	615209	255713	51307	128064	2023921
Murcia	210393	189394	113789	127725	318856	603683	310354	46726	103650	2024570
Comunidad Valenciana	217826	177228	99310	132318	280136	541577	281020	61803	122874	1914092
Baleares	227004	205880	131271	143722	357122	549567	333944	87284	151111	2186905
Andalucía	222072	183773	83027	120660	255114	604906	281153	50308	114433	1915446
País Vasco	267178	232181	121880	157854	342386	636443	394733	64959	174320	2391934
Ríoja	209670	196262	121721	127146	262629	602255	313107	54824	126054	2013668
Cataluña	262378	282701	107370	162944	360872	685925	362571	92869	227525	2545155
Canarias	185538	179574	98012	134070	305326	572256	279855	74322	159185	1988138
Madrid	253666	252591	129751	145747	369559	673620	432997	86853	215523	2560307
Ceuta y Melilla	193283	133987	64365	81436	141950	683373	234961	27191	84439	1644985
Margen activo	4294931	3718326	1918384	2406205	5280576	10693554	5474913	1089081	2463495	37339465

Tabla de correspondencias permutada de acuerdo con la dimensión 3

	Gastos									
CCAA	Grupo 9	Grupo 4	Grupo 8	Grupo 6	Grupo 1	Grupo 2	Grupo 5	Grupo 3	Grupo 7	Margen activo
Murcia	113789	127725	310354	318856	603683	210393	46726	189394	103650	2024570
Navarra	154978	220787	408892	406681	643258	324500	80939	251319	185658	2677012
Ríoja	121721	127146	313107	262629	602255	209670	54824	196262	126054	2013668
Baleares	131271	143722	333944	357122	549567	227004	87284	205880	151111	2186905
Comunidad Valenciana	99310	132318	281020	280136	541577	217826	61803	177228	122874	1914092
Castilla, La Mancha	107719	119423	259545	251981	546618	217780	41204	191246	109661	1845177
País Vasco	121880	157854	394733	342386	636443	267178	64959	232181	174320	2391934
Extremadura	71039	100379	222698	207527	469635	211088	43829	137556	86593	1550344
Ceuta y Melilla	64365	81436	234961	141950	683373	193283	27191	133987	84439	1644985
Andalucía	83027	120660	281153	255114	604906	222072	50308	183773	114433	1915446
Canarias	98012	134070	279855	305326	572256	185538	74322	179574	159185	1988138
Castilla, León	101476	125781	252823	243869	543406	220646	53318	210043	96367	1847729
Galicia	104412	131660	255713	291052	615209	247674	51307	198830	128064	2023921
Madrid	129751	145747	432997	369559	673620	253666	86853	252591	215523	2560307
Asturias	120725	130525	310705	336226	586561	280510	57682	232693	150503	2206130
Aragón	81832	125630	262864	246867	547436	255127	50268	201639	110719	1882382
Cantabria	105707	118398	276978	302423	587826	288598	63395	261429	116816	2121570
Cataluña	107370	162944	362571	360872	685925	262378	92869	282701	227525	2545155
Margen activo	1918384	2406205	5474913	5280576	10693554	4294931	1089081	3718326	2E+006	37339465

Panorámica Nacional (tabla de correspondencias permutadas de acuerdo con la dimensión 1): Ceuta y Melilla, Andalucía, Extremadura, y Galicia tienen un gasto superior en el Grupo 1 (Alimentos, bebidas y tabaco), respectivamente, (683373 / 1644985 = 41,5%), (604906 / 1915446 = 31,6%), (469635 / 1550344 = 30,3%), y (615209 / 2023921 = 30,4%).

Instrumentos Estadísticos Avanzados
Facultad Ciencias Económicas y Empresariales
Departamento de Economía Aplicada
Profesor: Santiago de la Fuente Fernández