



# Capítulo 3

*Análisis de Correspondencia*

---



# Capítulo 3

## Análisis de Correspondencia

---

### 1. Introducción

---

El **Análisis de Correspondencias** es una técnica que nos permite representar las categorías de dos o más variables cualitativas en un espacio de pequeñas dimensiones. Ello implica:

- Que la representación se hará agrupando las categorías en función de las similitudes que presenten en las variables relacionadas. Es una técnica, pues, para el estudio de las relaciones de dependencia entre variables categóricas presentadas en forma de tablas de contingencia. Guarda, en este sentido, cierta analogía con la prueba de chi-cuadrado. Sin embargo, y esta es la diferencia sustancial del análisis de correspondencias respecto a otros análisis, además de analizar la relación existente entre las variables, permite analizar como está **estructurada** esta asociación, describiendo "proximidades" que nos permiten, a su vez, identificar "categorías causa de asociación".
- Considerando la similitud entre categorías se produce una síntesis, reducción o resumen de las mismas en torno a pequeñas **dimensiones**. En este sentido, las dimensiones son sinónimo de **factores**. El análisis de correspondencias es un caso particular del análisis factorial (de hecho, la denominación original de la técnica es la de **análisis factorial de correspondencias**). Pese a ello, las diferencias son obvias: mientras que el análisis de correspondencias busca resumir la información aportada por variables cualitativas; el análisis factorial hace lo propio pero, en este caso, con variables cuantitativas.
- Al ser ésta una técnica factorial, sus resultados pueden ser presentados gráficamente sobre ejes de coordenadas, lo cual aporta una gran ayuda a la interpretación de los resultados.

En el análisis de correspondencias podemos diferenciar entre **análisis de correspondencia simple** y **análisis de correspondencias múltiple**. A ambos los separa el número de variables implicadas en el análisis: mientras que el primero trabaja con dos variables o con tablas de contingencia de dos en dos; el múltiple puede tener más de dos variables por lo que, en este caso, las tablas de contingencia son más complejas. Se denomina **tabla de Burt** a aquella que incluye a todas las variables. El paquete estadístico SPSS nos permite trabajar con el análisis simple.

En síntesis en esta lección vamos a:

1. En primer lugar, y como paso previo, valorar si los **datos se ajustan al modelo** de análisis propuesto.
2. En caso afirmativo, podremos proceder al análisis de correspondencias propiamente para el que es necesario reducir la información de las variables de partida. Para ello crearemos un **espacio factorial** en el que ubicar las variables y sujetos a partir del cual: analizar las semejanzas y diferencias entre las categorías de una misma variable; y analizar la relación de semejanzas y diferencias entre las categorías de las distintas variables incluidas en el análisis.

## 2. Bondad de ajuste al modelo: la tabla de contingencia y la prueba de $J_i^2$

Como previo paso al análisis de correspondencias propiamente debemos valorar si los datos se ajustan al modelo; o lo que es lo mismo, si entre las variables que pretendemos sintetizar existe, o no, relación. En el supuesto caso de no superar la prueba, carece de sentido aplicar el análisis que nos ocupa.

Para corroborar la bondad de ajuste de los datos al modelo, y al estar trabajando con dos variables cualitativas, se aplica la prueba de **Chi-cuadrado** a la **tabla de contingencia**. Pese a que el análisis de correspondencias guarde ciertas similitudes con el análisis factorial, al cambiar el nivel de medida no tiene sentido iniciar nuestro análisis con la matriz de correlaciones.

Con el estadístico Chi-cuadrado, para tablas de contingencia de doble entrada, contrastábamos la  $H_0$  de que las categorías de

una variable eran homogéneas entre sí respecto a las de la otra; o lo que es lo mismo, que existe independencia entre las variables. En la medida en que la distribución de las categorías de una variable se diferencian de la distribución de las categorías de la otra variable, estarán relacionadas. Si el *p-valor* asociado al estadístico es menor que el nivel de significación propuesto (normalmente, 0.05) se rechazará la hipótesis de independencia (ver punto en el que se expone el Cuadro de Diálogo de análisis de Tablas de Contingencia y más concretamente la lección que se dedica a este análisis).

### 3. Extracción del espacio factorial

---

El análisis de correspondencias, como ya hemos apuntado, 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 nos 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** (ver tablas de 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 tal 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) debemos valorar si éstos se encuentran concentrados o dispersos puesto que si las variables son muy dependientes, las distancias serán muy grandes (filas y columnas distintas); pero si son independientes, los puntos de la nube aparecerán muy agrupados (filas y columnas parecidas). La **Inercia** es el estadístico que en el análisis de correspondencias mide la dispersión de la nube de puntos. 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. En el ejemplo que recogemos en la sección de resulta-

dos la inercia total es de 0.080, pudiendo concluir que ésta es muy elevada. Al ser grande la dispersión de la nube de puntos (lo que no anula la posibilidad de que algunas categorías de cualquiera de las dos variables sean parecidas entre sí), podemos concluir que son distintas. Al extraer el espacio factorial (análisis de correspondencias) identificaremos en qué consisten esas diferencias.

Por otra parte, el total de la inercia, en nuestro ejemplo (1.000), queda perfectamente representada (la nube de puntos no sufre ninguna deformación) por un espacio de tres dimensiones, de tal manera que: el primer factor, es el que mejor representa la dispersión de la muestra; el segundo factor es el segundo y así sucesivamente. Sin embargo, el SPSS de no especificarse lo contrario, aplica como criterio de selección, de las tres dimensiones las dos primeras. En nuestro ejemplo, la **proporción de la inercia explicada por la dimensión 1** es igual al 96,3% y la explicada por la segunda dimensión es del 2,3%. En consecuencia, los dos primeros factores (criterio de selección) explican el 98,6% de la inercia total (**proporción de inercia acumulada**). El 100% de la variabilidad total explicada solo se conseguirá considerando las tres dimensiones. Sin embargo, debemos recordar que la finalidad del análisis de correspondencias es reducir la información original sin que ello implique perder mucha información.

Con los datos arrojados podemos concluir que éstos están perfectamente representados (al 99%) a partir del espacio factorial generado con los dos primeros factores, pero no podemos concluir que la representación de cada categoría en particular coincida con este valor. Las **puntuaciones factoriales** o proyecciones de cada una de las categorías de las variables sobre cada uno de los dos primeros factores, nos permiten identificar similitudes entre categorías.

Dos categorías de una misma variable con puntuaciones factoriales parecidas en los factores, serán similares entre sí. Comparando las proyecciones de las puntuaciones de la variable PREOCUPACIÓN (ver tabla “Inspección de los puntos en la fila, 1ª variable”, columna Puntuaciones en la Dimensión 1 y Dimensión 2) se aprecia que las Preocupaciones trabajo, familia-salud y vida afectiva son bastante parecidas entre sí. Luego dentro de las categorías de la variable PREOCUPACIÓN pueden diferenciarse 3 subconjuntos de categorías. Esta reflexión debe

repetirse también con las categorías de la variable ESTADO CIVIL (ver tabla “Inspección de los puntos en la columna, 2ª variable”, columna Puntuaciones en las Dimensiones 1 y 2) . De su análisis se desprende que en la variable ESTADO CIVIL se pueden diferenciar, a su vez, cuatro grupos, a saber: solteros, casados/viven en pareja-viudos y separados y divorciados.

La **representación gráfica** de las correspondencias nos 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 consultar la segunda parte de la tabla “Inspección de los puntos en la fila”. En ella se recoge la **proporción de la inercia de cada uno de los dos ejes atribuible a cada una de las categorías de las variables** PREOCUPACIÓN y ESTADO CIVIL respectivamente. En nuestro caso, y para la primera variable, de la inercia total proyectada sobre el primer factor, el 71% es atribuible a la preocupación TRABAJO. La preocupación que suscita la situación laboral hace que la dispersión de la nube de puntos sea explicada, mayoritariamente, por esta categoría. Esta categoría es muy distinta al resto.

Después de esta breve exposición en torno al análisis de correspondencias presentamos la secuencia de pasos a seguir para llevar a cabo este análisis. Recordemos que dos van a ser los cuadros de diálogos que vamos a utilizar: (1) el cuadro de diálogo del análisis de tablas de contingencia; y (2) el cuadro de diálogo del análisis de correspondencias propiamente.

#### 4. Cuadro de Diálogo del Análisis de Tablas de Contingencia

En primer lugar, debemos valorar si los datos se ajustan al modelo que se pretende aplicar; o lo que es lo mismo, si el objetivo último es analizar la similitud entre categorías de dos varia-

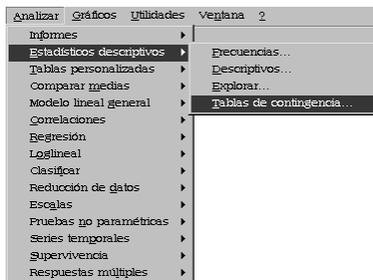


Figura 1

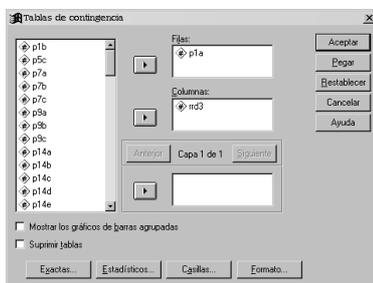


Figura 2

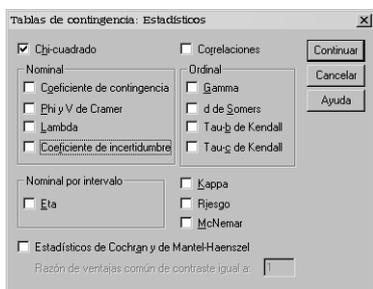


Figura 3

bles, lo primero que deberemos certificar es que entre las variables existe relación alguna. Puesto que las variables sometidas a análisis son cualitativas, tenemos que solicitar la tabla de contingencia con la prueba de la Chi-cuadrado. Para ello:

**1º paso:** Acceder al Cuadro de Diálogo de Tablas de doble entrada seleccionando **Estadísticos Descriptivos: Tablas de Contingencia** (figura 1) del Menú Principal **Analizar**.

**2º paso:** Indicar las dos variables que van a formar la tabla de doble entrada recordando que: en las filas situaremos a la dependiente y en las columnas la independiente (figura 2). Las variables seleccionadas son: la **P1a** MAS PREOCUPA y la **D3** ESTADO CIVIL. La variable estado civil ha sido recodificada (RRD3) en: 1 Solteros; 2 Casados/Pareja; 3 Separados/Divorciados; 4 Viudos; y 9 NS/NC. En ambas variables no se ha considerado la opción NS/NC en los análisis que se presentan.

**3º paso:** Aplicar la prueba de Chi-cuadrado para determinar si los datos se ajustan a un modelo. La prueba se selecciona clicando el botón de comando **Estadísticos** que figura en la parte inferior del cuadro de diálogo principal. En este subcuadro de diálogo, también podemos seleccionar las medidas de asociación más indicadas a nuestras variables. Sin embargo este análisis carece de sentido pues solo nos interesa saber si existe o no relación entre las variables seleccionadas (figura 3). Rechazamos la hipótesis de independencia si el *p-valor* asociado al estadístico es menor o igual a 0,05.

Una vez que demostrada la dependencia de las variables con la prueba de Chi-cuadrado podemos aplicar el análisis de correspondencias.

## 5. Cuadro de Diálogo del Análisis de Correspondencias

**4º paso:** En el Menú principal seleccionamos **Analizar: Reducción de datos: Análisis de correspondencias** (figura 4).

**5º paso:** Seleccionamos, nuevamente, las variables con las que queremos realizar el análisis de correspondencias colocándolas: una (**P1a**) en las filas con un rango de 1 a 4; y la otra (**rrd3**) en las columnas con un rango de 1 a 4 (figura 5). Al respecto cabe

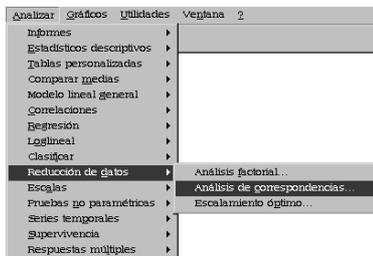


Figura 4

recodar que: (1) la ubicación de las variables debe coincidir con la posición establecida en la tabla de contingencia (la variable que hemos colocado en la fila seguirá estando en las filas y lo mismo con la segunda); y (2), el rango define la amplitud de categorías de la variable que hemos seleccionado (figura 6).

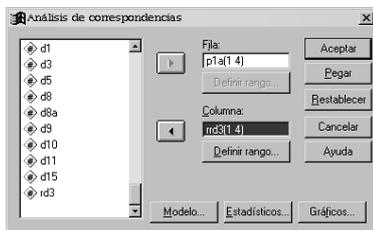


Figura 5

**6° paso:** El siguiente paso, consistirá en seleccionar el método con el cual se obtendrán las dimensiones y/o factores. Para acceder a esta selección debemos clicar el botón de comando **Modelo** que figura en la parte inferior del cuadro de diálogo principal (figura 7). Una vez en el subcuadro de diálogo Modelo seleccionamos el método de **Chi-cuadrado**. De esta forma obtendremos: el **estadístico de inercia** (que mide la dispersión de la nube de puntos) y el **valor propio** (o variabilidad total explicada). Cuanto mayor sea ésta, mayor será la relación entre las variables y más diferentes serán las categorías de dichas variables.

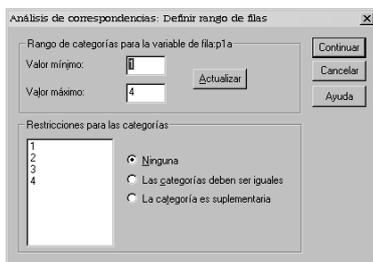


Figura 6

El subcuadro de Modelo se inicia seleccionando el número de **Dimensiones en la solución** en torno a los que nos gustaría sintetizar las categorías. A este respecto cabe recordar que: si seleccionamos pocas dimensiones (normalmente 2), tendremos una representación simplificada fácil de interpretar; si nos quedamos con muchas dimensiones, el resultado factorial será más representativo pero complejo a la hora de representar e interpretar. No obstante, este criterio podemos modificarlo si una vez aplicado el análisis con dos dimensiones (es el que viene por defecto), los resultados obtenidos no satisfacen nuestros propósitos.

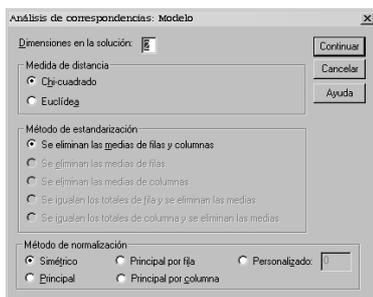


Figura 7

**7° paso:** Una vez seleccionado el criterio de extracción de factores, solicitaremos la batería de estadísticos que nos permiten identificar y comentar las similitudes entre las categorías. Para ello cliquemos sobre el botón de comando **Estadísticos**. El subcuadro de diálogo de Estadísticos (figura 8) hace referencia a una serie de tablas que el programa calcula y que nos van a permitir determinar: el número de factores o dimensiones; relación o similitud entre las categorías; y la asignación de cada una de las categorías en las dimensión.

- Al seleccionar **Tablas de correspondencias** el programa genera: la tabla de correspondencias propiamente en la que figuran el cómputo de encuestados por cada uno de los cruces de categorías y una tabla resumen final de las

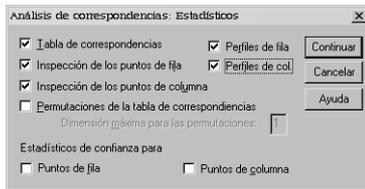


Figura 8

dimensiones y las inercias obtenidas una vez aplicado el análisis de correspondencias con estas variables.

- Al seleccionar las **Tablas de perfiles de filas** y de **columnas** obtenemos los tantos por uno por filas y los tantos por uno por columnas calculados en base a los totales que hemos obtenido en la “tabla de correspondencias”. Estas tablas nos ofrecen una primera aproximación de la similitud de las categorías. si las filas o columnas mostrasen las mismas frecuencias, las categorías serían iguales y no habría dependencia entre las categorías.
- Y por último, las **Tablas de Inspección de los puntos de fila** y de **columna**, nos ofrecen una doble información: en la primera parte de cada una de estas tablas podemos identificar las similitudes entre categorías a partir de las puntuaciones factoriales o proyecciones de cada una de las categorías de las variables (fila o columna respectivamente) sobre cada uno de dos primeros factores; en la segunda parte se aprecia la proporción de inercia de cada uno de los dos ejes atribuible a cada una de las categorías de las respectivas variables.

A su vez, la puntuación de cada categoría en cada dimensión nos permitirá representar el gráfico de correspondencias. Las puntuaciones que se obtienen cuanto más altas sean, o sea, cuanto más alejada esté la categoría del cruce de las dimensiones, querrá decir que esa categoría tendrá un mayor peso a la hora de explicar una relación y será mas fácil que entre en esa relación.

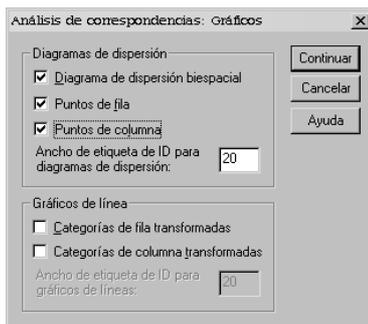


Figura 9

**8º paso:** Para terminar, realizamos la representación gráfica de las dimensiones en base a los resultados obtenidos en el paso anterior y en función a las puntuaciones de las tablas de inspección de los puntos por filas y por columnas.

- Dentro del subcuadro de **Gráficos** (figura 9), seleccionamos el **Diagrama de dispersión biespacial**, gráfico que representa de forma conjunta la dispersión de las categorías de la primera variable junto con las de la segunda variable. Es el gráfico donde se observa de manera más fácil el grado de relación y el peso de cada una de las categorías en cada una de las dimensiones.

- También seleccionamos, el **Diagrama de Puntos de fila** y el de **Puntos de columna**. Son dos gráficos que representan por separado la inercia de las categorías para cada una de las variables seleccionadas.

## 6. Resultados

Los resultados que se recogen en el anexo de este capítulo hacen referencia a una serie de tablas y gráficos dirigidos a:

- En primer lugar, la identificación de relación de dependencia entre las variables estado civil y elementos de preocupación. **Tabla de contingencia** y tabla de **bondad de ajuste**.
- En segundo lugar, y una vez ratificada la bondad de ajuste de los datos al modelo, se presentan una serie de tablas cuyo propósito no es otro que el de identificar relaciones entre las categorías de las variables. **Tabla de correspondencia**, tabla de **perfiles de filas**, tabla de **perfiles de columna**, tabla resumen de **extracción de factores**, tabla de **inspección de los puntos en la fila** y tabla de **inspección de los puntos en la columna**.
- Y por último, y con el propósito de simplificar las relaciones obtenidas y/o dimensiones subyacentes, se concluye el anexo de resultados con la exposición de los respectivos gráficos de correspondencias y **diagramas de dispersión de puntos**. Diagramas de dispersión de los puntos en la fila, de los puntos en la columna y diagrama de dispersión biespacial.

### 6.1 Resumen del procedimiento

	Casos					
	Válidos		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
<b>PIA * RRD3</b>	1153	96,1%	47	3,9%	1200	100,0%

6.2. Tabla de contingencia

Tabla de contingencia P1A \* RRD3

			RRD3				Total
			Solteros	Casados/Pareja	Separa/Divor	Viudos	
P1A	Vida afectiva	Recuento	25	46	3	5	79
		Frecuencia esperada	20,4	51,1	1,4	6,1	79,0
		% de RRD3	8,4%	6,2%	15,0%	5,6%	6,9%
		% del total	2,2%	4,0%	,3%	,4%	6,9%
		Residual	4,6	-5,1	1,6	-1,1	
		Residuos corregidos	1,2	-1,2	1,5	-,5	
	Trabajo	Recuento	113	119	4	3	239
		Frecuencia esperada	61,8	154,6	4,1	18,4	239,0
		% de RRD3	37,9%	16,0%	20,0%	3,4%	20,7%
		% del total	9,8%	10,3%	,3%	,3%	20,7%
		Residual	51,2	-35,6	-,1	-15,4	
		Residuos corregidos	8,5	-5,4	-,1	-4,2	
	Familia	Recuento	65	217	5	24	311
		Frecuencia esperada	80,4	201,2	5,4	24,0	311,0
		% de RRD3	21,8%	29,1%	25,0%	27,0%	27,0%
		% del total	5,6%	18,8%	,4%	2,1%	27,0%
		Residual	-15,4	15,8	-,4	,0	
		Residuos corregidos	-2,3	2,2	-,2	,0	
Salud	Recuento	95	364	8	57	524	
	Frecuencia esperada	135,4	339,0	9,1	40,4	524,0	
	% de RRD3	31,9%	48,8%	40,0%	64,0%	45,4%	
	% del total	8,2%	31,6%	,7%	4,9%	45,4%	
	Residual	-40,4	25,0	-1,1	16,6		
	Residuos corregidos	-5,5	3,1	-,5	3,7		
Total	Recuento	298	746	20	89	1153	
	Frecuencia esperada	298,0	746,0	20,0	89,0	1153,0	
	% de RRD3	100,0%	100,0%	100,0%	100,0%	100,0%	
	% del total	25,8%	64,7%	1,7%	7,7%	100,0%	

### 6.3. Bondad de Ajuste

#### Pruebas de chi-cuadrado

	Valor	gl	Sig. asint. (bilateral)
<b>Chi-cuadrado de Pearson</b>	92,337 <sup>a</sup>	9	,000
<b>Razón de verosimilitud</b>	91,919	9	,000
<b>Asociación lineal por lineal</b>	44,270	1	,000
<b>N de casos válidos</b>	1153		

<sup>a</sup>. 2 casillas (12,5%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 1,37.

### 6.4. Tablas de Correspondencia

PIA	RRD3				
	Solteros	Casados/Pareja	Separa/Divor	Viudos	Margen activo
<b>Vida afectiva</b>	25	46	3	5	79
<b>Trabajo</b>	113	119	4	3	239
<b>Familia</b>	65	217	5	24	311
<b>Salud</b>	95	364	8	57	524
<b>Margen activo</b>	298	746	20	89	1153

#### Perfiles de fila

PIA	RRD3				
	Solteros	Casados/Pareja	Separa/Divor	Viudos	Margen activo
<b>Vida afectiva</b>	,316	,582	,038	,063	1,000
<b>Trabajo</b>	,473	,498	,017	,013	1,000
<b>Familia</b>	,209	,698	,016	,077	1,000
<b>Salud</b>	,181	,695	,015	,109	1,000
<b>Masa</b>	,258	,647	,017	,077	

#### Perfiles de columna

PIA	RRD3				
	Solteros	Casados/Pareja	Separa/Divor	Viudos	Masa
<b>Vida afectiva</b>	,084	,062	,150	,056	,069
<b>Trabajo</b>	,379	,160	,200	,034	,207
<b>Familia</b>	,218	,291	,250	,270	,270
<b>Salud</b>	,319	,488	,400	,640	,454
<b>Margen activo</b>	1,000	1,000	1,000	1,000	

### 6.5. Extracción de Factores o Dimensiones

#### Resumen

Dimensión					Proporción de inercia		Valor propio de confianza	
	Valor propio	Inercia	Chi-cuadrado	Sig.	Explicada por	Acumulado	Desviación típica	Correlación
<b>1</b>	,278	,077			,963	,963	,029	-,006
<b>2</b>	,043	,002			,023	,986	,039	
<b>3</b>	,034	,001			,014	1,000		
<b>Total</b>		,080	92,337	,000 <sup>a</sup>	1,000	1,000		

<sup>a</sup> 9 grados de libertad

### 6.6. Tabla de inspección de los puntos en la fila (1ª variable)

P1A	Puntuación en la dimensión				Contribución				
	Masa	1	2	Inercia	De puntos para la inercia de la dimensión		De dimensión para la inercia del punto		Total
					1	2	1	2	
<b>Vida afectiva</b>	,069	-,285	-,734	,003	,020	,860	,484	,497	,982
<b>Trabajo</b>	,207	-,978	,097	,055	,714	,045	,998	,002	,999
<b>Familia</b>	,270	,193	,123	,004	,036	,094	,766	,048	,814
<b>Salud</b>	,454	,374	-,006	,018	,229	,000	,980	,000	,980
<b>Total activo</b>	1,000			,080	1,000	1,000			

<sup>a</sup> Symmetrical normalization

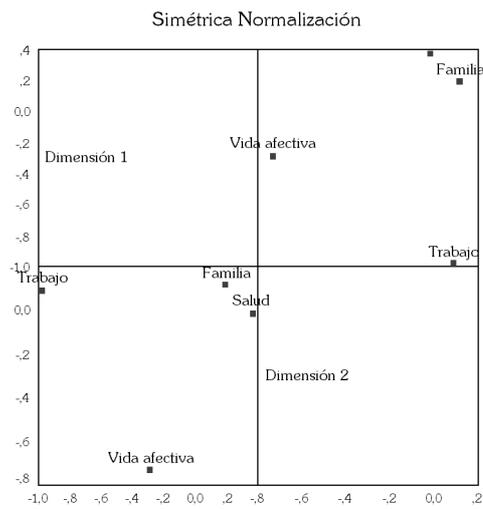
### 6.7. Tabla de inspección de los puntos en la columna (2ª variable)

RRD3	Puntuación en la dimensión				Contribución				
	Masa	1	2	Inercia	De puntos para la inercia de la dimensión		De dimensión para la inercia del punto		Total
					1	2	1	2	
<b>Solteros</b>	,258	-,840	-,002	,051	,657	,000	,998	,000	,998
<b>Casados/Pareja</b>	,647	,235	,065	,010	,129	,063	,971	,011	,982
<b>Separa/Divor</b>	,017	-,145	-,1458	,002	,001	,858	,056	,868	,924
<b>Viudos</b>	,077	,875	-,208	,017	,213	,078	,950	,008	,958
<b>Total activo</b>	1,000			,080	1,000	1,000			

<sup>a</sup> Symmetrical normalization

6.8. Diagrama de dispersión (puntos de fila y columna respectivamente).

Puntos de fila para Más Preocupa



Puntos de columna para Estado Civil Recodificada 2

