Universidad Nacional Agraria La Molina
Dpto. de Estadística e Informática
Curso: Técnicas Multivariadas

ANÁLISIS DE CORRESPONDENCIAS SIMPLES

I. Introducción[1]

El objetivo de las técnicas multivariadas consiste en determinar un menor número de dimensiones (factores o componentes), cuya mayor relevancia conceptual permite sintetizar la compleja información inicial. Cuando las variables son continuas, la técnica de Análisis de Componentes Principales o el modelo de Análisis Factorial son los procedimientos apropiados para analizar la interdependencia de un conjunto de indicadores. En cambio, cuando las variables estudiadas son cualitativas, es necesario acudir al Análisis de Correspondencias Simples o al Análisis de Correspondencias Múltiple para obtener estas dimensiones subyacentes que permitirán interpretar de forma rápida las relaciones de interdependencia del conjunto de variables originales.

Justificación

	CLASE A	CLASE B	CLASE C	TOTAL	marginal	
HOMBRES	100	200	200	500	0.25	
MUJERES	300	600	600	1500	0.75	(a)
	400	800	800	2000		
marginal	0.20	0.40	0.40			

	CLASE A	CLASE B	CLASE C		
HOMBRES	0.20	0.40	0.40	1.00	
MUJERES	0.20	0.40	0.40	1.00	(b)
	0.20	0.40	0.40	1.00	

	CLASE A	CLASE B	CLASE C		
HOMBRES	150	50	300	500	
MUJERES	250	750	500	1500	(c)
	400	800	800	2000	

	CLASE A	CLASE B	CLASE C		
HOMBRES	0.30	0.10	0.60	1.00	
MUJERES	0.17	0.50	0.33	1.00	(d)
	0.20	0.40	0.40	1.00	

[1] Basado en Renom Pinsach, Jordi. "Tratamiento Informatizado de Datos". Editorial Masson. 1998. España

II. Formulación del ACS[2]

Introducción

La matriz de datos está constituida de la siguiente manera:

Matriz de frecuencias absolutas (k)

	1	....	j	.....	p	Total fila
1						$k_{1.}$
.						
.						
i			k_{ij}			$k_{i.}$
.						
.						
n						$k_{n.}$
Total columna	$k_{.1}$		$k_{.j}$		$k_{.p}$	k

Donde:

n	Número de filas o de modalidades de la primera variable categórica
p	Número de columnas o de modalidades de la segunda variable categórica
k_{ij}	Número de individuos/casos que poseen a la vez la modalidad "i" de la primera variable y la modalidad "j" de la segunda variable
$k_{i.}$	Efectivo total de la fila "i"
$k_{.j}$	Efectivo total de la columna "j"
k	Efectivo total

$$k = \sum_i \sum_j k_{ij} \qquad k_{i.} = \sum_j k_{ij} \qquad k_{.j} = \sum_i k_{ij}$$

Geométricamente, la información de la matriz de contingencia puede ser analizada indistintamente, desde la perspectiva de las filas o de las columnas. Cualquiera de ellas nos proporciona la misma información.

En estadística, con la finalidad de eliminar la influencia del tamaño de las observaciones se trabajan con los perfiles (perfiles fila o perfiles columna) en vez de trabajar con las frecuencias observadas (k_{ij}).

El desarrollo se realizará trabajando con las frecuencias relativas.

[2] Basado en Pérez, César. "Técnicas de Análisis Multivariante de Datos: aplicaciones con SPSS". Pearson Prentice Hall. 2004. España.

Matriz de frecuencias relativas (f)

	1	……..	j	…..	p	Total fila
1						$f_{1.}$
.						
.						
i			f_{ij}			$f_{i.}$
.						
.						
n						$f_{n.}$
Total columna	$f_{.1}$		$f_{.j}$		$f_{.p}$	1

$$f_{ij} = \frac{k_{ij}}{k}$$

El peso de la fila "i" se determina con $\frac{f_{ij}}{f_{i.}}$ donde $f_{i.} = \frac{k_i}{k}$ y $f_{ij} = \frac{k_{ij}}{k}$

$$\sum_i f_{i.} = \sum_j f_{.j} = \sum_i \sum_j f_{ij} = 1$$

$$\sum_{i=1}^{n} \frac{f_{ij}}{f_{i.}} f_{i.} = f_{.j} \qquad (1)$$

$$\left(\frac{f_{ij}}{f_{i.}} - f_{.j} \right) \qquad j = 1, 2, \ldots p \qquad (2)$$

$$d^2(i, G) = \sum_{j=1}^{p} \left(\frac{f_{ij}}{f_{i.}} - f_{.j} \right)^2 \qquad (3)$$

$$d^2(i, i') = \sum_{j=1}^{p} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2 \qquad (4)$$

$$d_{\chi}^2(i, i') = \sum_{j=1}^{p} \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2 \qquad (5)$$

$$d_{\chi}^2(i, i') = d^2(z_i, z_{i'}) = \sum_{j=1}^{p} \left(\frac{f_{ij}}{f_{i.}\sqrt{f_{.j}}} - \frac{f_{i'j}}{f_{i'.}\sqrt{f_{.j}}} \right)^2 \qquad (6)$$

$$z_{ij} = \left(\frac{f_{ij}}{f_{i.}\sqrt{f_{.j}}} \right) - \sqrt{f_{.j}} = \left(\frac{f_{ij} - f_{i.} f_{.j}}{f_{i.}\sqrt{f_{.j}}} \right) \qquad (7)$$

Concepto de Inercia

$$\mathrm{I}(i) = \mathrm{d}^2(i, G)\, f_{i.}$$

$$\sum_{i=1}^{n} \mathrm{I}(i) = \mathrm{IT}$$

$$\mathrm{IT} = \sum_{\mathrm{i}} \mathrm{Inercia(i)} = \sum_{\mathrm{i}} \mathrm{f_i} \sum_{\mathrm{j}} \frac{1}{\mathrm{f_{.j}}} \left(\frac{\mathrm{f_{ij}}}{\mathrm{f_{i.}}} - \mathrm{f_{.j}}\right)^2 = \sum_{\mathrm{i}} \sum_{\mathrm{j}} \frac{(\mathrm{f_{ij}} - \mathrm{f_{i.}} \mathrm{f_{.j}})^2}{\mathrm{f_{i.}} \mathrm{f_{.j}}}$$

los resultados anteriores nos conducen a realizar un análisis de componentes principales, hallando los autovalores y autovectores usando la matriz X de término general:

$$x_{ij} = \frac{f_{ij} - f_{i.} f_{.j}}{\sqrt{f_{i.}}\sqrt{f_{.j}}}$$

En consecuencia, el análisis de correspondencias de la tabla de contingencia es lo mismo que realizar un análisis de componentes principales de la matriz X.

La matriz de dispersión o de inercia tanto para las columnas como para las filas se expresa como sigue:

$V_F = X'X$ para el espacio que definen los puntos i en R^p

$V_C = XX'$ para el espacio que definen los puntos j en R^n

Obtenidas las matrices de inercia para las columnas y para las filas, el siguiente paso es calcular sus valores y vectores propios.

Para el espacio de p dimensiones que definen los puntos fila, es necesario diagonalizar la matriz V_F obteniendo sus valores y vectores propios que permitirán calcular las coordenadas de los puntos i en R^p.

Estas coordenadas tendrán la siguiente expresión:

$$F_\alpha(i) = \sum_{j=1} \left(\frac{f_{ij}}{f_{i.}\sqrt{f_{.j}}}\right) u_{\alpha j}$$

donde $u_{\alpha j}$ simboliza a los vectores propios de V_F

III. Interpretación de los resultados

Generalidades

- 🕮 En la práctica suelen ser suficientes dos o tres ejes para estudiar la relación entre líneas y columnas. Es posible obtener una visión global bastante buena si se representan simultáneamente los puntos fila y columna sobre el plano formado por los dos primeros ejes, que recoge la mayor cantidad de información.
- 🕮 Si dos filas (columnas) tienen una estructura semejante, su situación será próxima sobre el plano (no siempre es cierto lo inverso; dependerá de la calidad de representación de los puntos).
- 🕮 La situación cercana de un punto fila "i" y un punto columna "j" sólo se puede interpretar si están alejados del origen, en la periferia de la nube.
- 🕮 Cuando una línea tiene un perfil próximo al perfil medio, tiene un comportamiento medio, se encontrará próxima al origen.

Ayudas a la interpretación

Contribución Absoluta (CA)

Expresa la participación que tiene el elemento "i" en la inercia explicada por el factor α. $\lambda\alpha$. es la inercia explicada por el factor α.

$$CA_\alpha(i) = \frac{f_{i.} F_\alpha^2(i)}{\lambda_\alpha}$$

Contribución Relativa (CR)

Recoge la participación del factor α en la explicación del elemento "i". Mide la calidad de representación de "i" sobre el eje α

$$CR_\alpha(i) = \frac{F_\alpha^2(i)}{d^2(i,G)}, \qquad \sum_\alpha \mathrm{CR}_\alpha(i) = 1$$

Donde:

$$d^2(i,G) = \sum_{j=1}^{p} \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - f_{.j}\right)^2$$

Interpretación de los ejes

La interpretación de los ejes es fundamental, y para evitar errores es necesario seguir las siguientes indicaciones:

- Se buscan aquellos puntos i (j) de mayor CA
- Dentro de éstos se separan los puntos que se proyectan del lado positivo de los que intervienen del lado negativo, que estarán en oposición.
- Se estudia la calidad de representación CR de estos puntos. Si un punto tiene un CR pequeño es de suponer que tenga un papel importante sobre otro eje, y para su estudio sería necesario considerar el conjunto de los ejes.
- Se buscan aquellos puntos i (j) que si bien no contribuyen a la formación del factor, si se encuentran bien representados (CR alto). Estos puntos son ilustrativos de la significación del eje.

Luego de realizado este análisis en cada uno de los conjuntos de filas y columnas, se ponen estas interpretaciones en relación para obtener los resultados del análisis, teniendo cuidado de interpretar las proximidades entre puntos fila y columna únicamente cuando se encuentran en la periferia de la nube, pues es entonces cuando por las fórmulas de transición podemos afirmar que “i” desempeña un papel importante en “j” o recíprocamente. No ocurre así con los puntos próximos al origen de coordenadas o centros de gravedad de la nube.

IV. Ejemplos de Aplicación

Ejemplo de Aplicación Nº 1[3]

Salida con MINITAB

Se dispone de datos sobre el consumo de cuatro marcas de tres segmentos de consumidores.

Tabla de contingencia

	1	2	3	Total
A	30.000	30.000	155.000	215.000
B	30.000	130.000	30.000	190.000
C	80.000	30.000	30.000	140.000
D	80.000	30.000	5.000	115.000
Total	220.000	220.000	220.000	660.000

Perfiles de Filas

Perfiles de filas

	1	2	3	Total
A	0.140	0.140	0.721	0.326
B	0.158	0.684	0.158	0.288
C	0.571	0.214	0.214	0.212
D	0.696	0.261	0.043	0.174
Total	0.333	0.333	0.333	

0.696 = 80/220 **80/115** 80/660

1. 0.696 = 70% de los encuestados del segmento 1 prefieren la marca D

2. 0.696 = 70% de los encuestados de los que prefieren la marca D son del segmento 1

3. 0.696 = 70% de los encuestados prefieren la marca D y son del segmento 1

Perfiles de Columnas

Perfiles de columnas

	1	2	3	Total
A	0.136	0.136	0.705	0.326
B	0.136	0.591	0.136	0.288
C	0.364	0.136	0.136	0.212
D	0.364	0.136	0.023	0.174
Total	0.333	0.333	0.333	

0.705 = 155 / 220

70.5% de los encuestados del segmento 3 prefieren la marca A

Frecuencias esperadas

	1	2	3
A	71.667	71.667	71.667
B	63.333	63.333	63.333
C	46.667	46.667	46.667
D	38.333	38.333	38.333

Frecuencia esperada: 190 *220 / 660 = 63.333
Frecuencia observada: 30.0

[3] Basado en Pérez, César. "Técnicas de Análisis Multivariante de Datos: aplicaciones con SPSS". Pearson Prentice Hall. 2004. España.

Ho: La preferencia por un producto es independiente del segmento
Ho: La preferencia por un producto es dependiente del segmento

Frecuencia esperada = (total fila)*(total columna) / (total de la tabla)

Frecuencias observadas - esperadas

	1	2	3
A	-41.667	-41.667	83.333
B	-33.333	66.667	-33.333
C	33.333	-16.667	-16.667
D	41.667	-8.333	-33.333

30-63.333=-33.333

Contribución al estadístico Chi-cuadrado

Distancias de chicuadrada

	1	2	3	Total
A	24.225	24.225	96.899	145.349
B	17.544	70.175	17.544	105.263
C	23.810	5.952	5.952	35.714
D	45.290	1.812	28.986	76.087
Total	110.868	102.164	149.381	362.413

X^2 calculado = (observado - esperado)^2 / esperado = (30-63.333)^2/63.3333 = 17.544

$$IT=\sum_i \text{Inercia}(i)=\sum_i f_{i.}\sum_j \frac{1}{f_{.j}}\left(\frac{f_{ij}}{f_{i.}}-f_{.j}\right)^2=\sum_i\sum_j \frac{(f_{ij}-f_{i.}f_{.j})^2}{f_{i.}f_{.j}}$$

Perfiles Filas

Marcas	Segmento 1	Segmento 2	Segmento 3	Total	$f_{i.}$
A	0.1395	0.1395	0.7209	1.000	0.3258
B	0.1579	0.6842	0.1579	1.000	0.2879
C	0.5714	0.2143	0.2143	1.000	0.2121
D	0.6957	0.2609	0.0435	1.000	0.1742
$f_{.j}$	0.3333	0.3333	0.3333		

Marcas	Segmento 1	Segmento 2	Segmento 3	$d^2(i,G)$
A	0.11267	0.11267	0.45069	0.6760
B	0.09234	0.36934	0.09234	0.5540
C	0.17007	0.04252	0.04252	0.2551
D	0.39382	0.01575	0.25205	0.6616

Inercias Absolutas

Marcas	Segmento 1	Segmento 2	Segmento 3	Inercia Filas	Inercia Relativa
A	0.0367	0.0367	0.1468	**0.2202**	0.4011
B	0.0266	0.1063	0.0266	**0.1595**	0.2905
C	0.0361	0.0090	0.0090	**0.0541**	0.0985
D	0.0686	0.0027	0.0439	**0.1153**	0.2099
Inercia Columnas	**0.1680**	**0.1548**	**0.2263**	0.5491	1.0000

Inercia Total = 0.5491

Inercia Total = Chi Cuadrado / (Total de la tabla)=362.413/660

La marca A esté más alejada del origen

El segmento 3 debería estar más alejado del origen

Ejemplo: Si la inercia de la marca B = 0. La marca B estará en el origen. Su perfil fila es similar a su perfil medio.

Inercias Relativas

0.0668 = 0.0367 / 0.5491

Marcas	Segmento 1	Segmento 2	Segmento 3	Total
A	0.0668	0.0668	0.2674	0.4011
B	0.0484	0.1936	0.0484	0.2905
C	0.0657	0.0164	0.0164	0.0985
D	0.1250	0.0050	0.0800	0.2099
Total	0.3059	0.2819	0.4122	1.0000

```
Inercias relativas

            1       2       3   Total
A       0.067   0.067   0.267   0.401
B       0.048   0.194   0.048   0.290
C       0.066   0.016   0.016   0.099
D       0.125   0.005   0.080   0.210
Total   0.306   0.282   0.412   1.000
```

Análisis de la Tabla de Contingencia

```
La tabla de análisis de contingencia

  Eje  Inercia  Proporción  Acumulada  Histograma
    1   0.3405      0.6201     0.6201  ******************************
    2   0.2086      0.3799     1.0000  ******************
Total   0.5491
```

$$x_{ij} = \frac{f_{ij} - f_{i.} f_{.j}}{\sqrt{f_{i.}}\sqrt{f_{.j}}}$$

$V_F = X'X$ para el espacio que definen los puntos i en R^p

$V_C = XX'$ para el espacio que definen los puntos j en R^n

Cálculo de la matriz de Inercia

Tabla de Datos

Marcas	Segmento 1	Segmento 2	Segmento 3	Total
A	30	30	155	215
B	30	130	30	190
C	80	30	30	140
D	80	30	5	115
Total	220	220	220	660

0.0455 = 30 / 660

Frecuencias Relativas

Marcas	Segmento 1	Segmento 2	Segmento 3	$f_{i.}$
A	0.0455	0.0455	0.2348	0.3258
B	0.0455	0.1970	0.0455	0.2879
C	0.1212	0.0455	0.0455	0.2121
D	0.1212	0.0455	0.0076	0.1742
$f_{.j}$	0.3333	0.3333	0.3333	660

Matriz X de 4 x 3

	Segmento 1	Segmento 2	Segmento 3
A	-0.191583537	-0.191583537	0.383167075
B	-0.163038663	0.326077325	-0.163038663
C	0.189934294	-0.094967147	-0.094967147
D	0.261956089	-0.052391218	-0.209564871

$$x_{ij} = \frac{f_{ij} - f_{i.} f_{.j}}{\sqrt{f_{i.}}\sqrt{f_{.j}}}$$

-0.19158 = (0.0455 - 0.3258 * 0.3333)/(raíz(0.3258)*raíz(0.3333))

Matriz X' de 3 x 4

	A	B	C	D
Segmento 1	-0.191583537	-0.163038663	0.189934294	0.261956089
Segmento 2	-0.191583537	0.326077325	-0.094967147	-0.052391218
Segmento 3	0.383167075	-0.163038663	-0.094967147	-0.209564871

$$V_F = X'X$$ para el espacio que definen los puntos i en R^p

$$V_C = XX'$$ para el espacio que definen los puntos j en R^n

Matriz de Inercia en R^4 (XX') 4x4

	Marca A	Marca B	Marca C	Marca D
Marca A	**0.220225511**	-0.093706571	-0.054582426	-0.12044754
Marca B	-0.093706571	**0.159489633**	-0.04644995	-0.02562538
Marca C	-0.054582426	-0.04644995	**0.054112554**	0.07463167
Marca D	-0.120447538	-0.025625382	0.074631667	**0.11528327**

Matriz de Inercia en R^3 (X'X) 3x3

	Segmento 1	Segmento 2	Segmento 3
Segmento 1	**0.167981886**	-0.048220676	-0.11976121
Segmento 2	-0.048220676	**0.154794273**	-0.106573597
Segmento 3	-0.11976121	-0.106573597	**0.226334807**

Matriz de Inercia en R^4 (XX') 4x4

Análisis de Componentes Principales

0.220225511	-0.093706571	-0.054582426	-0.120447538
-0.093706571	**0.159489633**	-0.04644995	-0.025625382
-0.054582426	-0.04644995	**0.054112554**	0.074631667
-0.120447538	-0.025625382	0.074631667	**0.115283267**

$$V_C = XX'$$

Análisis de Componentes Principales:

Valores propios:

	F1	F2
Valor propio	0.3405	0.2086
Variabilidad (%)	62.0091	37.9909
% acumulado	62.0091	100.0000

Vectores propios:

	F1	F2
A	0.8012	-0.0893
B	-0.2894	0.7924
C	-0.2286	-0.4173
D	-0.4713	-0.4360

0.167981886	-0.04822068	-0.11976121
-0.048220676	**0.15479427**	-0.1065736
-0.11976121	-0.1065736	**0.22633481**

$$V_f = X'X$$

Análisis de Componentes Principales:

Valores propios:

	F1	F2
Valor propio	**0.3405**	**0.2086**
Variabilidad (%)	62.0091	37.9909
% acumulado	62.0091	100.0000

Vectores propios:

	F1	F2
Segmento 1	-0.4682	0.6689
Segmento 2	-0.3452	-0.7399
Segmento 3	0.8134	0.0710

Tabla de Datos

Marcas	Segmento 1	Segmento 2	Segmento 3	Total
A	30	30	155	215
B	30	130	30	190
C	80	30	30	140
D	80	30	5	115
	220	220	220	660

Frecuencias Relativas

Marcas	Segmento 1	Segmento 2	Segmento 3	$f_{i.}$
A	0.045454545	0.045454545	0.234848485	0.32575758
B	0.045454545	0.196969697	0.045454545	0.28787879
C	0.121212121	0.045454545	0.045454545	0.21212121
D	0.121212121	0.045454545	0.007575758	0.17424242
$f_{.j}$	0.333333333	0.333333333	0.333333333	1

Vectores propios:

	F1	F2
A	0.8012	-0.0893
B	-0.2894	0.7924
C	-0.2286	-0.4173
D	-0.4713	-0.4360

0.819 = 0.0454 / 0.3257 * raíz(0.3333)0.8012

$$F_\alpha(i) = \sum_{j=1} \left(\frac{f_{ij}}{f_{i.}\sqrt{f_{.j}}}\right) u_{\alpha j}$$

Coordenadas de las Marcas (Filas)

Marcas	F1	F2
A	0.819	0.071
B	-0.315	-0.675
C	-0.290	0.414
D	-0.659	0.477

Contribuciones Absolutas y Relativas

Tabla de Datos

Marcas	Segmento 1	Segmento 2	Segmento 3	Total
A	30	30	155	215
B	30	130	30	190
C	80	30	30	140
D	80	30	5	115
Total	220	220	220	660

Frecuencias Relativas

Marcas	Segmento 1	Segmento 2	Segmento 3	$f_{i.}$
A	0.0455	0.0455	0.2348	0.3258
B	0.0455	0.1970	0.0455	0.2879
C	0.1212	0.0455	0.0455	0.2121
D	0.1212	0.0455	0.0076	0.1742
$f_{.j}$	0.3333	0.3333	0.3333	660

Matriz X

	Segmento 1	Segmento 2	Segmento 3
A	-0.1916	-0.1916	0.3832
B	-0.1630	0.3261	-0.1630
C	0.1899	-0.0950	-0.0950
D	0.2620	-0.0524	-0.2096

Matriz de Inercia en R^4 (XX')

	Marca A	Marca B	Marca C	Marca D
Marca A	**0.2202**	-0.0937	-0.0546	-0.1204
Marca B	-0.0937	**0.1595**	-0.0464	-0.0256
Marca C	-0.0546	-0.0464	**0.0541**	0.0746
Marca D	-0.1204	-0.0256	0.0746	**0.1153**

Coordenadas de las Marcas

	Coordenada 1	Coordenada 2
Marca A	0.8191	0.0715
Marca B	-0.3147	-0.6745
Marca C	-0.2896	0.4138
Marca D	-0.6588	0.4770

Autovalores	0.340499	0.208612

Autovectores	F1	F2
Segmento 1	-0.4682	0.6689
Segmento 2	-0.3452	-0.7399
Segmento 3	0.8134	0.0710

Contribuciones Absolutas (Contr)

	Componente 1	Componente 2
Marca A	0.6419	0.0080
Marca B	0.0837	0.6278
Marca C	0.0522	0.1741
Marca D	0.2221	0.1901
Total	1.0000	1.0000

Contribuciones Relativas (Corr)

	Componente 1	Componente 2	Total
Marca A	0.9924	0.0076	1.0000
Marca B	0.1788	0.8212	1.0000
Marca C	0.3288	0.6712	1.0000
Marca D	0.6561	0.3439	1.0000

```
Contribuciones de las filas

                                             Componente  1          Componente  2
ID  Nombre     Cal  Total  Inercia   Coord   Corr  Contr   Coord   Corr  Contr
 1  A        1.000  0.326    0.401   0.819  0.992  0.642   0.071  0.008  0.008
 2  B        1.000  0.288    0.290  -0.315  0.179  0.084  -0.675  0.821  0.628
 3  C        1.000  0.212    0.099  -0.290  0.329  0.052   0.414  0.671  0.174
 4  D        1.000  0.174    0.210  -0.659  0.656  0.222   0.477  0.344  0.190

Contribuciones de columnas

                                             Componente  1          Componente  2
ID  Nombre     Cal  Total  Inercia   Coord   Corr  Contr   Coord   Corr  Contr
 1  Segm 1   1.000  0.333    0.306  -0.473  0.444  0.219   0.529  0.556  0.447
 2  Segm 2   1.000  0.333    0.282  -0.349  0.262  0.119  -0.585  0.738  0.547
 3  Segm 3   1.000  0.333    0.412   0.822  0.995  0.662   0.056  0.005  0.005
```

Cal: Proporción de la inercia de la fila (columna) representada por los dos componentes.

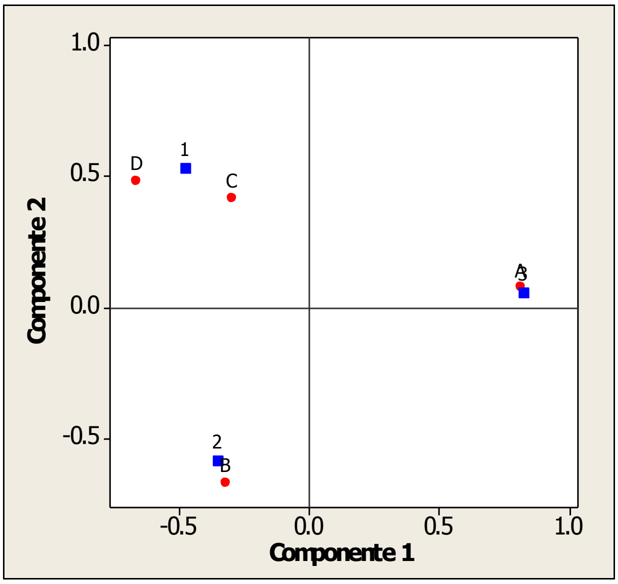
Total: Proporción, frecuencia de la clase en el total de datos

Inercia: Proporción con que contribuye cada fila (columna) a la inercia total

Coord: Coordenadas principales de las filas (columnas)

Corr: Contribución del componente a la inercia de la fila (columna). Contribución Relativa

Contr: Contribución de cada fila (columna) a la inercia del eje. Contribución Absoluta



Salida con SPSS

Tabla de correspondencias

Marca	Segmento 1	2	3	Margen activo
A	30	30	155	215
B	30	130	30	190
C	80	30	30	140
D	80	30	5	115
Margen activo	220	220	220	660

Resumen

Dimensión	Valor propio	Inercia	Chi-cuadrado	Sig.	Proporción de inercia		Confianza para el Valor propio	
					Explicada	Acumulada	Desviación típica	Correlación 2
1	.584	.340			.620	.620	.032	.155
2	.457	.209			.380	1.000	.038	
Total		.549	362.413	.000[a]	1.000	1.000		

a. 6 grados de libertad

Examen de los puntos de fila[a]

Marca	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
A	.326	-1.072	.106	.220	.642	.008	.992	.008	1.000
B	.288	.412	-.998	.159	.084	.628	.179	.821	1.000
C	.212	.379	.612	.054	.052	.174	.329	.671	1.000
D	.174	.862	.706	.115	.222	.190	.656	.344	1.000
Total activo	1.000			.549	1.000	1.000			

a. Normalización Simétrica

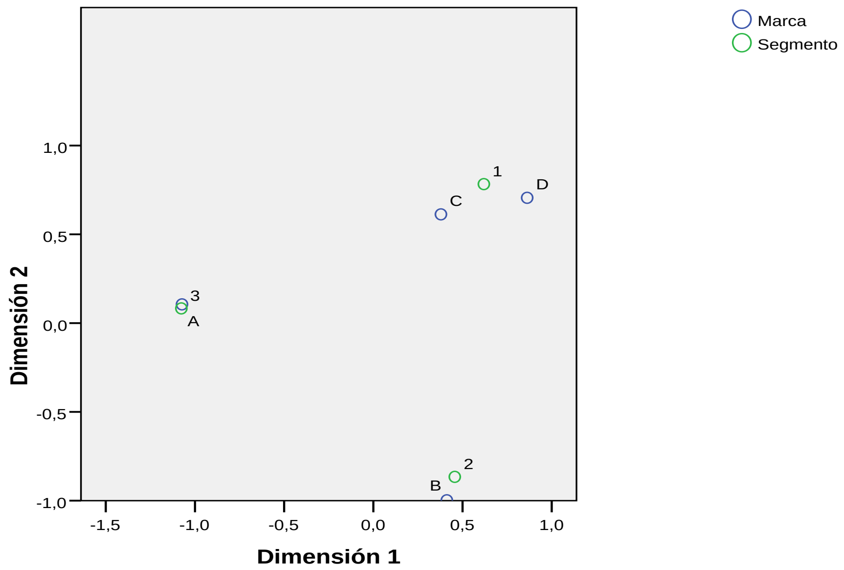
Examen de los puntos columna[a]

Segmento	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
1	.333	.619	.783	.168	.219	.447	.444	.556	1.000
2	.333	.457	-.866	.155	.119	.547	.262	.738	1.000
3	.333	-1.076	.083	.226	.662	.005	.995	.005	1.000
Total activo	1.000			.549	1.000	1.000			

a. Normalización Simétrica

Puntos de columna y de fila

Simétrica Normalización



Ejemplo de Aplicación Nº 2[4]

Se pretende buscar la relación entre el poder económico de la persona y la opinión que tiene sobre el sistema sanitario público. Para ello, se recogen para ello opiniones de 500 personas. En cuanto al nivel de renta se han dividido a las personas entrevistadas en cuatro niveles según distintos tramos de renta.

Nivel de renta	***Opinión sobre el sistema sanitario público***			***Total***
	Bueno	***Malo***	***Regular***	
Bajo	75	40	35	150
Medio	60	50	70	180
Alto	20	40	30	90
Muy Alto	15	40	25	80
Total	170	170	160	500

[4] Basado en Pérez, César. "Técnicas de Análisis Multivariante de Datos: aplicaciones con SPSS". Pearson Prentice Hall. 2004. España.

abla de contingencia Nivel de Renta * Opinión sobre el sistema sanitario públic

			Opinión sobre el sistema sanitario público			Total
			Bueno	Malo	Regular	
Nivel de Renta	Bajo	Recuento	75	40	35	150
		Frecuencia esperad	51.0	51.0	48.0	150.0
	Medio	Recuento	60	50	70	180
		Frecuencia esperad	61.2	61.2	57.6	180.0
	Alto	Recuento	20	40	30	90
		Frecuencia esperad	30.6	30.6	28.8	90.0
	Muy Alto	Recuento	15	40	25	80
		Frecuencia esperad	27.2	27.2	25.6	80.0
Total		Recuento	170	170	160	500
		Frecuencia esperad	170.0	170.0	160.0	500.0

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pears	40.049[a]	6	.000
Razón de verosimilitude	39.693	6	.000
Asociación lineal por lineal	14.525	1	.000
N de casos válidos	500		

a. 0 casillas (.0%) tienen una frecuencia esperada inferior a
La frecuencia mínima esperada es 25.60.

Tabla de correspondencias

	Opinión sobre el sistema sanitario público			
Nivel de Renta	Bueno	Malo	Regular	Margen activo
Bajo	75	40	35	150
Medio	60	50	70	180
Alto	20	40	30	90
Muy Alto	15	40	25	80
Margen activo	170	170	160	500

Perfiles de fila

	Opinión sobre el sistema sanitario público			
Nivel de Renta	Bueno	Malo	Regular	Margen activo
Bajo	.500	.267	.233	1.000
Medio	.333	.278	.389	1.000
Alto	.222	.444	.333	1.000
Muy Alto	.188	.500	.313	1.000
Masa	.340	.340	.320	

Perfiles de columna

Nivel de Renta	Opinión sobre el sistema sanitario público: Bueno	Malo	Regular	Masa
Bajo	.441	.235	.219	.300
Medio	.353	.294	.438	.360
Alto	.118	.235	.188	.180
Muy Alto	.088	.235	.156	.160
Margen activo	1.000	1.000	1.000	

Resumen

Dimensión	Valor propio	Inercia	Chi-cuadrado	Sig.	Proporción de inercia: Explicada	Acumulada	Confianza para el Valor propio: Desviación típica	Correlación 2
1	.255	.065			.813	.813	.043	.078
2	.122	.015			.187	1.000	.045	
Total		.080	40.049	.000[a]	1.000	1.000		

a. 6 grados de libertad

Examen de los puntos de fila[a]

Nivel de Renta	Masa	Puntuación en la dimensión: 1	2	Inercia	Contribución: De los puntos a la inercia de la dimensión: 1	2	De la dimensión a la inercia del punto: 1	2	Total
Bajo	.300	-.637	-.302	.034	.477	.223	.903	.097	1.000
Medio	.360	-.053	.458	.009	.004	.616	.027	.973	1.000
Alto	.180	.530	-.122	.013	.198	.022	.975	.025	1.000
Muy Alto	.160	.716	-.326	.023	.322	.139	.909	.091	1.000
Total activo	1.000			.080	1.000	1.000			

a. Normalización Simétrica

Contribución Absoluta indica el peso de una fila(columna) en la inercia de un componente (dimensión)

0.477 = Que el 47.7% de la inercia de la dimensión o del componente 1 es explicada por el nivel de renta Bajo

0.223 = Que el 22.3% de la inercia de la dimensión o del componente 2 es explicada por el nivel de renta Bajo

Contribución Relativa indica el peso de una componente o dimensión en la inercia de una fila(columna)

0.903 = Que el 90.3% de la inercia del nivel de renta Bajo es explicada por la dimensión o componente 1

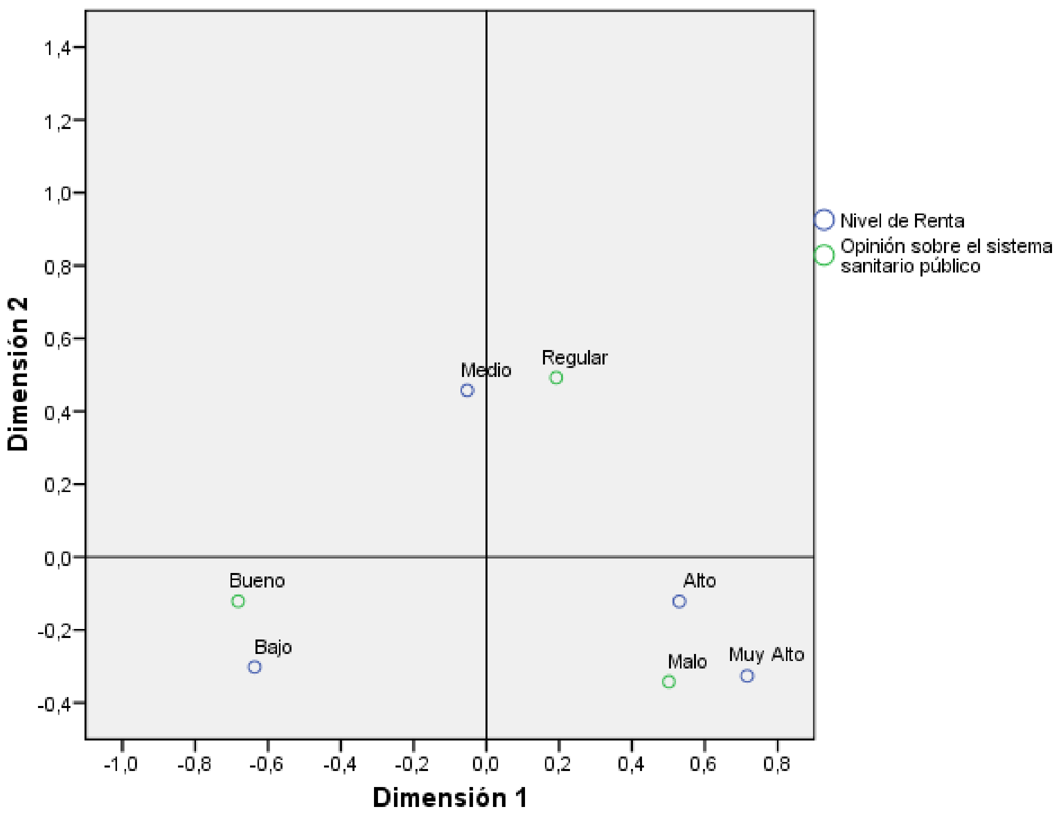
0.097 = Que el 9.7% de la inercia del nivel de renta Bajo es explicada por la dimensión o componente 2

Examen de los puntos columna[a]

Opinión sobre el sistema sanitario público	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
Bueno	.340	-.682	-.121	.041	.619	.041	.985	.015	1.000
Malo	.340	.501	-.342	.027	.335	.325	.817	.183	1.000
Regular	.320	.192	.493	.013	.046	.634	.240	.760	1.000
Total activo	1.000			.080	1.000	1.000			

a. Normalización Simétrica

Categorías	Componente 1	Componente 2
Bajo	CA (-)	
Medio		CA (+)
Alto	CR (+)	
Muy Alto	CA (+)	
Bueno	CA (-)	
Malo	CA (+)	
Regular		CA (+)



ANÁLISIS DE CORRESPONDENCIAS USANDO R

```
#Librerias
library(MASS)
library(ca)
#                    Opinión
#Renta      Bueno Malo   Regular
#Bajo       75       40       35
#Medio      60       50      70
#Alto       20       40       30
#Muy Alto 15       40       25


datosacs<-matrix(c(75,40,35,
                   60,50,70,
                   20,40,30,
                   15,40,25),nrow=4,byrow=T)
datosacs
```

```
          col.opinion
col.renta  Bueno Malo Regular
  Bajo        75   40      35
  Medio       60   50      70
  Alto        20   40      30
  Muy Alto    15   40      25
```

```
#Nombres de las filas y las columnas de la tabla.

dimnames(datosacs)<-list(col.renta=c("Bajo", "Medio", "Alto", "Muy Alto")
                        ,col.opinion=c("Bueno","Malo","Regular"))
datosacs
```

```
         col.opinion
col.renta  Bueno Malo Regular
  Bajo        75   40      35
  Medio       60   50      70
  Alto        20   40      30
  Muy Alto    15   40      25
```

```
# Prueba de Independencia Chi-Cuadrado

prueba=chisq.test(datosacs)
prueba
```

```
        Pearson's Chi-squared test
data:  datosacs
X-squared = 40.049, df = 6, p-value = 4.455e-07
```

```
prueba$observed
```

```
          col.opinion
col.renta  Bueno Malo Regular
  Bajo        75   40      35
  Medio       60   50      70
  Alto        20   40      30
  Muy Alto    15   40      25
```

```
prueba$expected
```

```
          col.opinion
col.renta  Bueno Malo Regular
  Bajo      51.0 51.0    48.0
  Medio     61.2 61.2    57.6
  Alto      30.6 30.6    28.8
  Muy Alto  27.2 27.2    25.6
```

```
library(ca)

# Perfiles Fila
prop.table(datosacs, 1)
```

```
          col.opinion
col.renta      Bueno      Malo   Regular
  Bajo     0.5000000 0.2666667 0.2333333
  Medio    0.3333333 0.2777778 0.3888889
  Alto     0.2222222 0.4444444 0.3333333
  Muy Alto 0.1875000 0.5000000 0.3125000
```

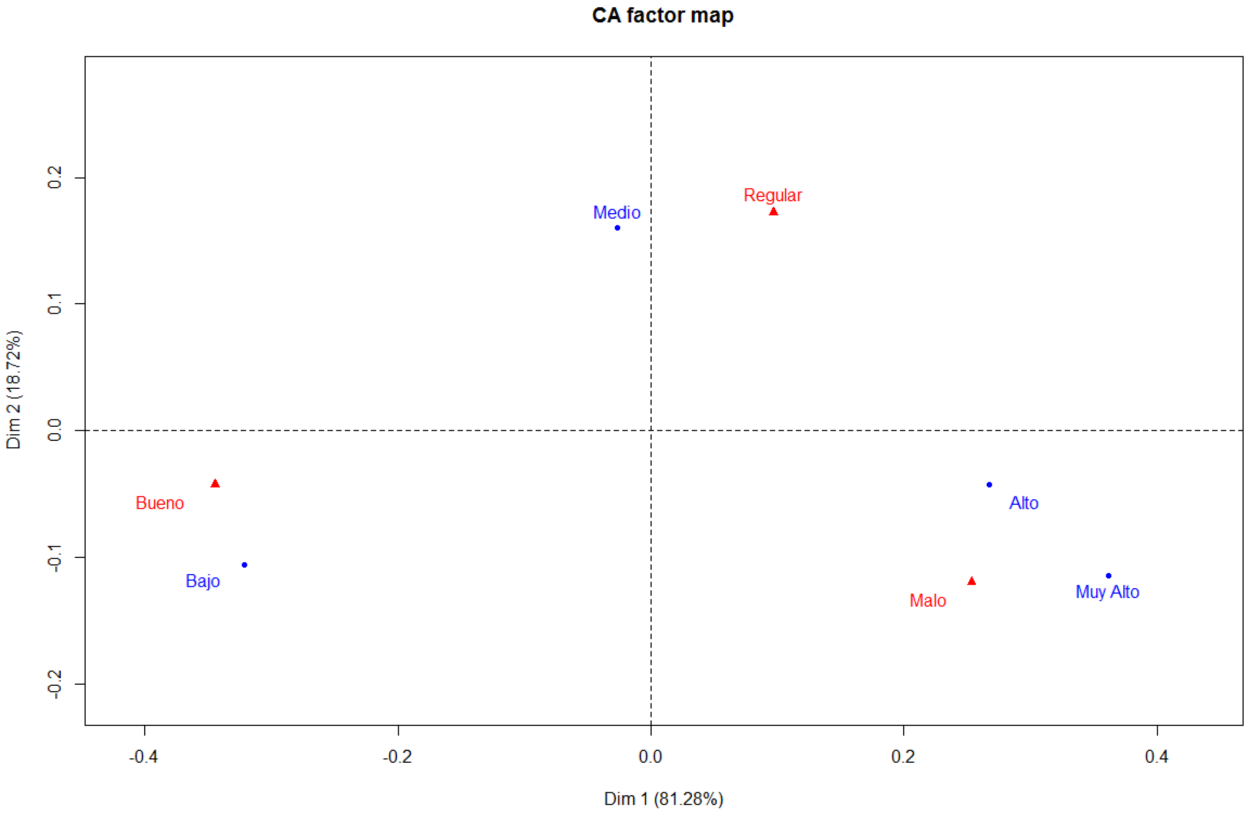
```
# Perfiles Columna
prop.table(datosacs, 2)
```

```
          col.opinion
col.renta       Bueno      Malo Regular
  Bajo     0.44117647 0.2352941 0.21875
  Medio    0.35294118 0.2941176 0.43750
  Alto     0.11764706 0.2352941 0.18750
  Muy Alto 0.08823529 0.2352941 0.15625
```

```
# Análisis de Correspondencia
fit = CA(datosacs,ncp=2,graph=TRUE)
print(fit)
summary(fit,nb.dec = 3, ncp = 2)
```

```
Call:
CA(X = datosacs, ncp = 2, graph = TRUE)

The chi square of independence between the two variables is equal to 40.04927
(p-value =  4.454704e-07 ).

Eigenvalues
                       Dim.1   Dim.2
Variance               0.065   0.015
% of var.             81.279  18.721
Cumulative % of var.  81.279 100.000

Rows
           Iner*1000    Dim.1    ctr   cos2    Dim.2    ctr   cos2
Bajo     |    34.375 | -0.322 47.655  0.903 | -0.106 22.341  0.097 |
Medio    |     9.485 | -0.027  0.391  0.027 |  0.160 61.558  0.973 |
Alto     |    13.219 |  0.268 19.803  0.975 | -0.043  2.178  0.025 |
Muy Alto |    23.019 |  0.362 32.151  0.909 | -0.114 13.923  0.091 |

Columns
           Iner*1000    Dim.1    ctr   cos2    Dim.2    ctr   cos2
Bueno    |    40.923 | -0.344 61.919  0.985 | -0.042  4.081  0.015 |
Malo     |    26.667 |  0.253 33.467  0.817 | -0.120 32.533  0.183 |
Regular  |    12.509 |  0.097  4.614  0.240 |  0.172 63.386  0.760 |

# Resultados extendidos, tomar las coordenadas, contribuciones absolutas y
relativas
# En summary se tienen dos indicadores importantes: la contribución absoluta
(ctr) y la contribución relativa (cos2)

# Interpretación
# Por ejemplo: para la fila BAJO y la dimensión 1 se tiene una ctr = 47.655
# El 47,65% de la inercia de la dimensión 1 es explicada por la fila BAJO
#
# Por ejemplo: para la fila BAJO y la dimensión 1 se tiene una cos2 = 0.903
# El 90,3% de la inercia de la fila BAJO es explicada por la dimensión 1
```

```
# biplot
plot(fit) # Mapa Simétrico
```



V. Puntos Suplementarios[5]

Hay ocasiones en las cuales existen columnas y filas adicionales que no forman parte de los datos iniciales pero que son útiles para interpretar aspectos encubiertos de los mismos. Cualquier fila (o columna) adicional de una matriz de datos puede posicionarse sobre un mapa existente, mientras que el perfil de esta fila (o columna) sea comparable a los perfiles de las filas (o las columnas) que han determinado el mapa.

Por ahora, todas las filas y todas las columnas de una tabla de datos se han usado para determinar los ejes y de aquí el mapa perceptual. Sin embargo, existen situaciones en las que interesa eliminar puntos y los ejes se calculan sin que hayan intervenido estos puntos en la formación de los gráficos finales. Habría que pensar que tales puntos tienen una posición pero no tienen masa. Por tanto, su contribución a la inercia sería cero. Tales puntos se llaman puntos suplementarios (o puntos pasivos), a diferencia de los puntos activos usuales que tienen una masa positiva.

El uso de puntos suplementarios es útil para:

- Representar algunas filas o columnas, en especial cuando su número es demasiado grande
- Mostrar distintos grupos de individuos o variables de naturales diferente a las analizadas, por ejemplo, al considerar consumidores habituales y no habituales, o la suma de dos filas o dos columnas, etc.
- Visualizar elementos que perturbaban un análisis anterior y no permitían una interpretación clara de los ejes por ser muy diferentes o tener un gran peso (outliers)
- Enriquecer el análisis con nuevas variables que faciliten la interpretación.

Para proyectar filas suplementarias basta aplicar la siguiente expresión:

$$F_\alpha(i)=\frac{1}{\sqrt{\lambda_\alpha}}\sum_j \frac{k_{ij}}{k_{i.}} G_\alpha(j)$$

Y para posicionar columnas suplementarias:

$$G_\alpha(j)=\frac{1}{\sqrt{\lambda_\alpha}}\sum_i \frac{k_{ij}}{k_{.j}} F_\alpha(i)$$

No hay que olvidar que aunque los puntos suplementarios tienen una posición en el mapa, no contribuyen nada a la formación de éste. Entonces es lógico que su contribución absoluta sea cero. Sin embargo, es posible calcular la calidad de la representación de la fila o columna suplementaria, así como su contribución relativa.

[5] Basado en Pérez, César. "Técnicas de Análisis Multivariante de Datos: aplicaciones con SPSS". Pearson Prentice Hall. 2004. España.

VI. Consideraciones Finales[6]

Basándonos en el hecho de que, para la aplicación del ACS es suficiente con disponer de una matriz de números no negativos, es evidente que sus posibilidades de aplicación en investigación comercial son numerosísimas.

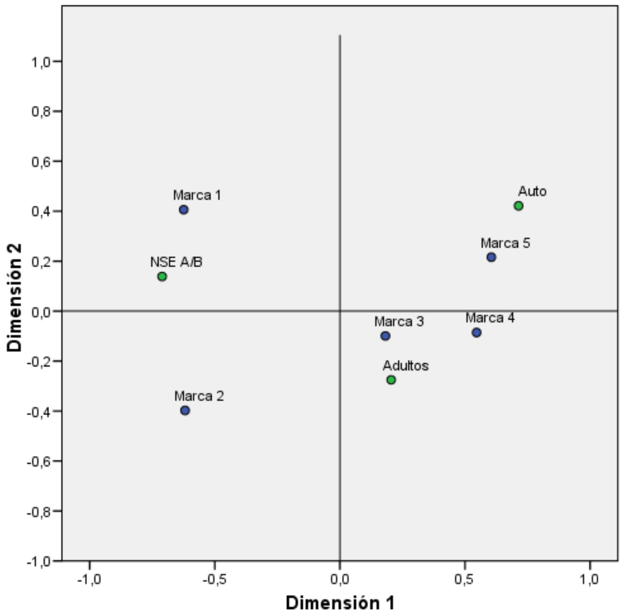
El método se aplica al análisis de matrices de valores absolutos (obtenidas por simple conteo). Por lo tanto, se hace imprescindible tener siempre en cuenta las siguientes precauciones:

- Es necesario que los márgenes de filas y columnas tengan un sentido; es decir, la suma de cada fila y columna debe poderse interpretar.
- Dado que en el cálculo de las distancias entre dos elementos intervienen, como hemos visto, la suma de las columnas; es preciso, para que el cálculo sea lícito, que esta suma tenga algún sentido.

En caso de no cumplirse estas dos condiciones anteriores, el análisis puede también realizarse, pero se corre el riesgo de que los resultados que se obtengan no reflejen la estructura del fenómeno que se está analizando, sino que más bien estén reflejando el método que hayamos utilizado para recoger los datos.

Ejemplo de Aplicación Nº 3

	Adultos	*NSE A/B*	*Auto*	*Total*
Marca 1	25	35	10	70
Marca 2	30	30	5	65
Marca 3	35	20	15	70
Marca 4	40	15	20	75
Marca 5	25	10	15	50
Total	155	110	65	330



[6] Basado en Pedret, Ramón & otros. "Herramientas para segmentar mercados y posicionar productos: análisis de información cuantitativa en investigación comercial". Ediciones Deusto. 2000. España.