



# Análisis de correspondencias simple

Práctica 4  
30 noviembre 2009

# Objetivo

- Describir las relaciones entre 2 **variables categóricas** dispuestas en una tabla de contingencia. Variables  $I$  y  $J$ , con  $n$  y  $p$  categorías respectivamente

	1	2	...	p	Total
1	$k_{11}$	$k_{12}$	...	$k_{1p}$	$k_{1+}$
2	$k_{21}$	$k_{22}$	...	$k_{2p}$	$k_{2+}$
⋮	⋮	⋮	⋮	⋮	⋮
n	$k_{n1}$	$k_{n2}$	...	$k_{np}$	$k_{n+}$
Total	$k_{+1}$	$k_{+2}$	...	$k_{+p}$	$k_{++}$

$k_{ij}$  es el número de individuos en la combinación de las categorías " $I=i$ " e " $J=j$ "

# Frecuencias relativas

	1	2	...	p	Total
1	$f_{11}$	$f_{12}$	...	$f_{1p}$	$f_{1+}$
2	$f_{21}$	$f_{22}$	...	$f_{2p}$	$f_{2+}$
⋮	⋮	⋮	⋮	⋮	⋮
n	$f_{n1}$	$f_{n2}$	...	$f_{np}$	$f_{n+}$
Total	$f_{+1}$	$f_{+2}$	...	$f_{+p}$	1

donde

$$f_{ij} = \frac{k_{ij}}{k_{++}}, \quad f_{i+} = \frac{k_{i+}}{k_{++}}, \quad f_{+j} = \frac{k_{+j}}{k_{++}}.$$

# Test de independencia

- Contraste:

$H_0$  :  $I$  y  $J$  son independientes

vs

$H_1$  :  $I$  y  $J$  son dependientes

Estadístico: 
$$\chi^2 = k_{++} \sum_{i=1}^n \sum_{j=1}^p \frac{(f_{ij} - f_{i+} f_{+j})^2}{f_{i+} f_{+j}}$$

Rechazar a nivel  $\alpha$  si:  $\chi^2 > \chi^2_{(n-1)(p-1), \alpha}$

# Tabla de perfiles fila

- Dividiendo cada elemento entre el total de su fila

	1	2	...	p	Total
1	$f_{11}/f_{1+}$	$f_{12}/f_{1+}$	...	$f_{1p}/f_{1+}$	1
2	$f_{21}/f_{2+}$	$f_{22}/f_{2+}$	...	$f_{2p}/f_{2+}$	1
⋮	⋮	⋮	⋮	⋮	⋮
n	$f_{n1}/f_{n+}$	$f_{n2}/f_{n+}$	...	$f_{np}/f_{n+}$	1
	$f_{+1}$	$f_{+2}$	...	$f_{+p}$	1

$f_{1+}$  es el total de la fila 1, ...etc.

# Nubes de puntos fila

Las filas de la tabla de perfiles fila son  $n$  puntos en el espacio  $R^p$

	1	2	...	p	Total
1	$f_{11}/f_{1+}$	$f_{12}/f_{1+}$	...	$f_{1p}/f_{1+}$	1
2	$f_{21}/f_{2+}$	$f_{22}/f_{2+}$	...	$f_{2p}/f_{2+}$	1
⋮	⋮	⋮	⋮	⋮	⋮
n	$f_{n1}/f_{n+}$	$f_{n2}/f_{n+}$	...	$f_{np}/f_{n+}$	1
	$f_{+1}$	$f_{+2}$	...	$f_{+p}$	1

A estos les llamaremos “puntos fila”

$$r_i = (r_{i1}, r_{i2}, \dots, r_{ip}) = \left( \frac{f_{i1}}{f_{i+}}, \frac{f_{i2}}{f_{i+}}, \dots, \frac{f_{ip}}{f_{i+}} \right)$$



La *nube de puntos fila* es el conjunto de puntos fila.

El centro de gravedad será un punto con  $p$  coordenadas, tal que

$$\bar{r} = (\bar{r}_1, \dots, \bar{r}_p)$$

donde

$$\bar{r}_j = \sum_{i=1}^n f_{i+} \frac{f_{ij}}{f_{i+}} = \sum_{i=1}^n f_{ij} = f_{+j}, \quad j = 1, 2, \dots, p$$

# Tabla de perfiles columna

- Dividiendo cada elemento entre el total de su columna

	1	2	...	p	Total
1	$f_{11}/f_{+1}$	$f_{12}/f_{+2}$	...	$f_{1p}/f_{+p}$	$f_{1+}$
2	$f_{21}/f_{+1}$	$f_{22}/f_{+2}$	...	$f_{2p}/f_{+p}$	$f_{2+}$
⋮	⋮	⋮	⋮	⋮	⋮
n	$f_{n1}/f_{+1}$	$f_{n2}/f_{+2}$	...	$f_{np}/f_{+p}$	$f_{n+}$
	1	1	...	1	1

Con “perfil columna”:

$$c_j = (c_{1j}, c_{2j}, \dots, c_{nj}) = \left( \frac{f_{1j}}{f_{+j}}, \frac{f_{2j}}{f_{+j}}, \dots, \frac{f_{nj}}{f_{+j}} \right)$$

La *nube de puntos columna* es el conjunto de perfiles columna.

El centro de gravedad será un punto con  $p$  coordenadas, tal que

$$\bar{c} = (\bar{c}_1, \dots, \bar{c}_n)$$

donde

$$\bar{c}_i = \sum_{j=1}^p f_{+j} \frac{f_{ij}}{f_{+j}} = \sum_{j=1}^p f_{ij} = f_{i+}, \quad i = 1, 2, \dots, n,$$

# Inercia de las nubes de puntos filas y columna

- Distancia  $\chi^2$  entre 2 puntos fila:

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

- Distancia  $\chi^2$  entre 2 puntos columna:

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

Principio de equivalencia distribucional..

# Inercia

- Del punto fila  $r_i$

$$I(r_i) = f_{i+} d_{\mathcal{X}}(r_i, \bar{r})$$

- Del punto columna  $c_j$

$$I(c_j) = f_{+j} d_{\mathcal{X}}(c_j, \bar{c})$$

# Inercia de la nube de

- Puntos fila  $N(I)$

$$I(N(\mathcal{I})) = \sum_{i=1}^n I(r_i) :$$

Suma de la inercia de cada punto fila

- Puntos columna  $N(J)$ .

$$I(N(\mathcal{J})) = \sum_{j=1}^p I(c_j)$$

Además,  $I(N(\mathcal{I})) = I(N(\mathcal{J})) = \frac{\chi^2}{k_{++}}$

# Análisis de la nube de puntos fila

- Deseamos representar los puntos fila en un espacio de menor dimensión, con ejes ortogonales, y de manera que la inercia de los puntos proyectados en los nuevos ejes tengan la máxima inercia.

Esto es semejante al ACP reemplazando la varianza por la inercia de los puntos fila.

# Transformación

- Para utilizar la distancia euclídea al cuadrado  $d_e^2$  en lugar de la distancia  $d_x$

$$N^*(\mathcal{I}) = \{(r_i^*, f_{i+}), i = 1, 2, \dots, n\}$$

donde

$$r_i^* = (r_{i1}^*, \dots, r_{ip}^*) = \left( \frac{r_{i1}}{\sqrt{f_{+1}}}, \dots, \frac{r_{ip}}{\sqrt{f_{+p}}} \right) = \left( \frac{f_{i1}}{f_{i+}\sqrt{f_{+1}}}, \dots, \frac{f_{ip}}{f_{i+}\sqrt{f_{+p}}} \right)$$

- Con centro de gravedad

$$\bar{r}^* = (\bar{r}_1^*, \dots, \bar{r}_p^*)$$

Donde 
$$\bar{r}^* = \left( \frac{f_{+1}}{\sqrt{f_{+1}}}, \dots, \frac{f_{+p}}{\sqrt{f_{+p}}} \right)$$

centrándola

$$N^{**}(\mathcal{I}) = \{(x_i, f_{i+}), i = 1, 2, \dots, n\}$$

# Determinación de los ejes principales

- La nube de puntos fila tiene  $n$  puntos con  $p$  coordenadas cada una (buscaremos  $p-1$  ejes principales)

- Definimos la matriz  $X^* = (x_{ij}^*)$  con

$$x_{ij}^* = \frac{f_{ij}}{\sqrt{f_{i+}} \sqrt{f_{+j}}}$$

- 
- Definimos  $T^* = (X^*)'X^*$

Calculamos los valores y vectores propios asociados a la matriz  $T^*$

# Análisis de la nube de puntos columna

Es semejante al anterior.

- Calculamos la matriz  $X^* = (x_{ij}^*)$

con 
$$x_{ij}^* = \frac{f_{ij}}{\sqrt{f_{i+}} \sqrt{f_{+j}}}$$

Por lo que calculamos los autovalores y autovectores de la matriz  $S^* = X^*(X^*)'$

- 
- Pero  $S^*$  y  $T^*$  tienen los mismos valores propios positivos.
  - Los ejes principales correspondientes a las filas y columnas se obtienen simultáneamente

# Análisis conjunto

- Podemos descomponer  $X^*$  como

$$X^* = B\Lambda^{1/2}A'$$

$\Lambda$ : matriz diagonal con los valores propios no nulos de  $S^*$  ó bien de  $T^*$

A: contiene los vectores propios normalizados de  $T^*$

B: contiene los vectores propios normalizados de  $S^*$

# Selección del número de ejes

- Cuando la proporción de inercia explicada sea  $\geq 75\%$

- Inercia total:  $I(N(\mathcal{I})) = \sum_{i=1}^r \lambda_i$

- Inercia explicada por los primeros  $r_1$  ejes:

$$\frac{\sum_{i=1}^{r_1} \lambda_i}{\sum_{i=1}^r \lambda_i}$$

# Contribución relativa

- Se define la contribución relativa del eje  $\alpha$  a la inercia del punto fila  $i$  como

$$cr_{\alpha}(i) = \frac{F_{i\alpha}^2}{d_e^2(F_i, O)} \text{ onde } d_e^2(F_i, O) = \sum_{\alpha=1}^r F_{i\alpha}^2.$$

- Idem, al punto columna  $j$

$$cr_{\alpha}(j) = \frac{G_{j\alpha}^2}{d_e^2(G_j, O)}$$

- Un punto es aceptablemente representado si:

$$\sum_{\alpha=1}^r cr(i) \geq 0,6$$

# Representaciones gráficas

- Nube de puntos fila respecto a sus ejes principales,  $F_i, i=1, \dots, n$
- Nube de puntos columna respecto a sus ejes principales,  $G_j, j=1, \dots, p$
- Representación simultánea de puntos fila y columna en un espacio conjunto



# Ejercicio 1 en SPSS

- Leer texto
- Abrir fichero DatosPractica4ACS.sav
- Ejecutar comandos

# Resultados

$k_{ij}$

Marca	Segmento de consumidores			Margen activo
	Segmento 1	Segmento 2	Segmento 3	
Marca A	30	30	<u>155</u>	215
Marca B	30	<u>130</u>	30	190
Marca C	<u>80</u>	30	30	140
Marca D	<u>80</u>	30	5	115
Margen activo	220	220	220	660

- El Segmento 3 consume la Marca A en mayor proporción
- El Segmento 2 la Marca B
- El Segmento 1 las marcas C y D

# Resultados

## Perfiles de fila

Marca	Segmento de consumidores			
	Segmento 1	Segmento 2	Segmento 3	Margen activo
Marca A	,140	,140	,721	1,000
Marca B	,158	,684	,158	1,000
Marca C	,571	,214	,214	1,000
Marca D	,696	,261	,043	1,000
Masa	,333	,333	,333	

- Se obtenía al dividir cada elemento entre el total de su fila

$$\frac{f_{ij}}{f_{i+}} = \frac{\frac{k_{ij}}{k_{++}}}{\frac{k_{i+}}{k_{++}}} = \frac{k_{ij}}{k_{i+}}$$

# Resultados

## Perfiles de columna

Marca	Segmento de consumidores			
	Segmento 1	Segmento 2	Segmento 3	Masa
Marca A	,136	,136	,705	,326
Marca B	,136	,591	,136	,288
Marca C	,364	,136	,136	,212
Marca D	,364	,136	,023	,174
Margen activo	1,000	1,000	1,000	

- Se obtenía al dividir cada elemento entre el total por columna

# Resultados

Resumen

6

7

1 Dimensión	2 Valor propio	3 Inercia	4 Chi-cuadrado	5 Sig.	6 Proporción de inercia		7 Confianza para el Valor propio	
					Explicada	Acumulada	Desviación típica	Correlación
1	,584	,340			,620	,620	,032	,155
2	,457	,209			,380	1,000	,038	
Total		,549	362,413	,000 <sup>a</sup>	1,000	1,000		

a. 6 grados de libertad

1  $\alpha$

2  $(\lambda_{\alpha})^{-1/2}$

3 Inercia indiv.:  $\lambda_{\alpha}$

Inercia total:  $\lambda_1 + \lambda_2 = \frac{\chi^2}{k_{++}}$

4  $\chi^2$

5 p valor

6  $\lambda_{\alpha} / (\lambda_1 + \lambda_2)$

7 a)  $\lambda_1 / (\lambda_1 + \lambda_2)$   
b)  $(\lambda_1 + \lambda_1) / (\lambda_1 + \lambda_2)$

# Resultados

Examen de los puntos de fila<sup>a</sup>

1	2	Puntuación en la dimensión		Inercia	4 Contribución 5				
		3			De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto		
		1	2		1	2	1	2	Total
Marca A	,326	-1,072	,106	,220	,642	,008	,992	,008	1,000
Marca B	,288	,412	-,998	,159	,084	,628	,179	,821	1,000
Marca C	,212	,379	,612	,054	,052	,174	,329	,671	1,000
Marca D	,174	,862	,706	,115	,222	,190	,656	,344	1,000
Total activo	1,000			,549	1,000	1,000			

a. Normalización Simétrica

1 Punto fila "i":

$$r_i = (r_{i1}, r_{i2}, \dots, r_{ip}) = \left( \frac{f_{i1}}{f_{i+}}, \frac{f_{i2}}{f_{i+}}, \dots, \frac{f_{ip}}{f_{i+}} \right)$$

2  $f_{i+}$

3 Transformadas  
 $F_{\alpha}(i^*)$

4 Contribución absoluta: sirve para interpretar los ejes

$$ca_{\alpha}(i) = \frac{f_{i+} F_{i\alpha}^2}{\lambda_{\alpha}}$$

5 Contribución relativa.

$$cr_{\alpha}(i) = \frac{F_{i\alpha}^2}{d_e^2(F_i, O)}$$

nota: en la tabla "1" refiere al eje 1, y "2" refiere al eje 2.

# Resultados

Examen de los puntos columna <sup>a</sup>									
1	2	Puntuación en la dimensión		Inercia	4 Contribución 5				
		3			De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto		
		1	2		1	2	1	2	Total
Segmento de consumidores	Masa								
Segmento 1	,333	,619	<u>,783</u>	,168	,219	<u>,447</u>	,444	<u>,556</u>	1,000
Segmento 2	,333	,457	<u>-,866</u>	,155	,119	<u>,547</u>	,262	<u>,738</u>	1,000
Segmento 3	,333	<u>-1,076</u>	,083	,226	<u>,662</u>	,005	<u>,995</u>	,005	1,000
Total activo	1,000			,549	1,000	1,000			

a. Normalización Simétrica

1 Punto columna "j"

$$c_j = (c_{1j}, c_{2j}, \dots, c_{nj}) = \left( \frac{f_{1j}}{f_{+j}}, \frac{f_{2j}}{f_{+j}}, \dots, \frac{f_{nj}}{f_{+j}} \right)$$

2  $f_{+j}$

3 Transformadas  
 $G_\alpha(j^*)$

4 Contribución absoluta

$$ca_\alpha(j) = \frac{f_{+j} G_{j\alpha}^2}{\lambda_\alpha}$$

5 Contribución relativa.

$$cr_\alpha(j) = \frac{G_{j\alpha}^2}{d_e^2(G_j, O)}$$

nota: en la tabla "1" refiere al eje 1,  
y "2" refiere al eje 2.

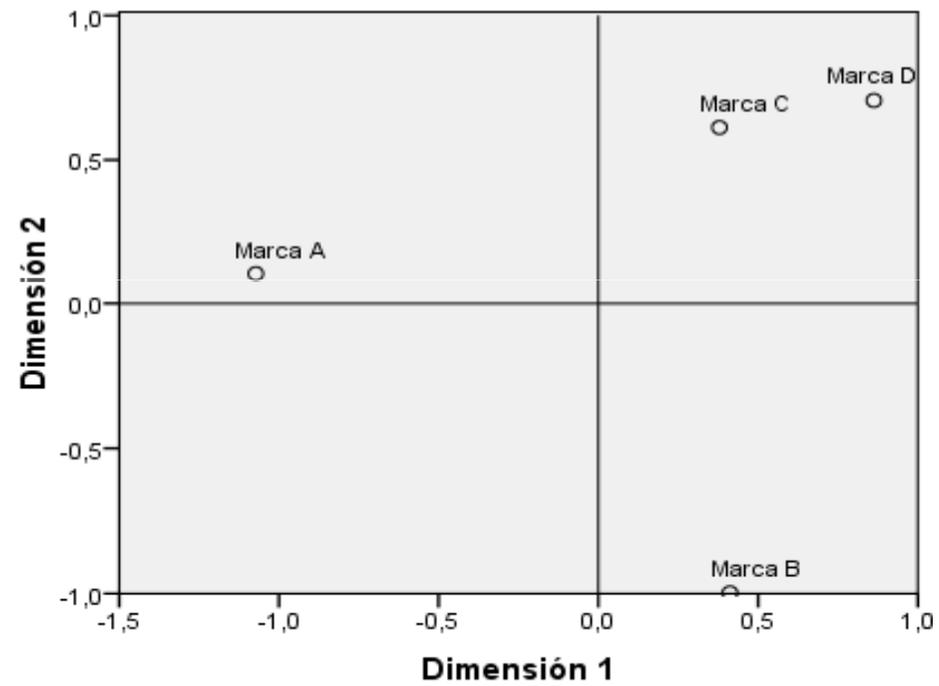
# Resultados

- El eje 1: Determinando por Marca A, Marca D y Segmento 3
- El eje 2: Determinado por Marca B, Marca C, Segmento 1 y 2.

# Resultados

Puntos de fila para Marca

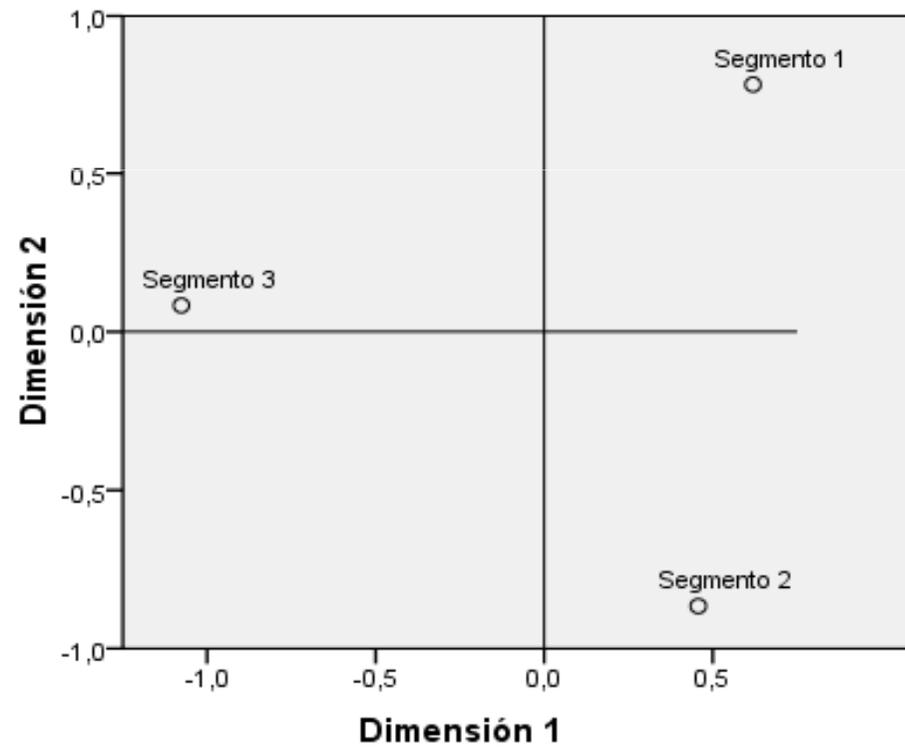
Simétrica Normalización



# Resultados

Puntos de columna para Segmento de consumidores

Simétrica Normalización



# Ejercicio 2 en SPSS

- Leer texto
- Abrir fichero DatosPractica5ACS.sav
  - >Tenemos 25 atributos y 9 marcas
- Ejecutar comandos

# Resultados

Tabla de correspondencias

atributo	marca									
	Ecua	JackDa	Marlbo	Empres	Absor	Imati	Necesi	Coloca	Novocla	Margen activo
cara	22	3	1	19	0	91	1	12	4	153
masivo	19	30	9	43	5	60	8	3	9	186
pocavar	29	20	9	38	10	57	12	5	9	189
tv	17	13	3	18	3	122	2	8	9	195
perfecto	21	25	6	27	4	74	11	2	9	179
descono	23	23	8	54	21	47	10	2	5	193
ahorratie	18	25	8	24	12	53	14	2	8	164
dura	20	11	6	73	12	57	11	2	3	195
Innovador	39	21	7	42	12	61	10	5	14	211
Exclus	22	6	6	38	13	90	5	17	14	211
facilenco	11	27	6	21	1	86	9	12	9	182
imitaci	30	22	10	56	5	52	13	3	6	197
economi	19	11	6	30	7	79	5	9	14	180
antigua	20	29	8	50	5	48	9	4	7	180
artesan	20	10	6	50	21	50	10	3	6	176
moda	15	22	6	42	6	78	13	14	4	200
lujos	13	8	2	11	31	58	13	10	17	
diseño	13	14	3	24	2	107	7	11	6	187
rendim	15	7	7	19	15	63	12	13	16	167
barata	21	17	8	53	16	42	8	3	4	172
clasealta	28	22	7	51	5	55	6	5	11	190
servicio	37	12	10	44	9	52	10	7	7	188
orgullo	20	10	5	36	6	94	14	7	6	198
regalos	25	27	9	40	7	40	11	3	8	170
calidad	9	11	3	20	1	120	6	12	7	189
Margen activo	513	418	157	912	198	1678	217	164	195	4452

a. Fila suplementaria

Tabla de contingencia ( $kij$ )

# Resultados

- Después, aparecen las tablas de:
- Perfiles fila: nos permitirá comparar los “atributos” para cada “marca”.
- Perfiles columna: nos permitirá comparar las “marcas” para cada “atributo”.

# Resultados

Resumen

Dimensión					Proporción de inercia		Confianza para el Valor propio	
	Valor propio	Inercia	Chi-cuadrado	Sig.	Explicada	Acumulada	Desviación típica	Correlación
1	,300	,090			,577	,577	,014	,004
2	,157	,024			,157	,735	,015	
3	,125	,016			,101	,835		
4	,106	,011			,072	,908		
5	,081	,006			,042	,949		
6	,068	,005			,030	,979		
7	,051	,003			,017	,996		
8	,025	,001			,004	1,000		
Total		,156	693,478	,000 <sup>a</sup>	1,000	1,000		

a 184 grados de libertad

- Interpretación igual que el ejercicio anterior

# Resultados

- Después, aparecen las tablas de:
- Examen de los puntos fila
- Examen de los puntos columna

La interpretación de las tablas es semejante que en el ejercicio 1.

Cara, calidad,  
Diseño y tv son  
parecidas

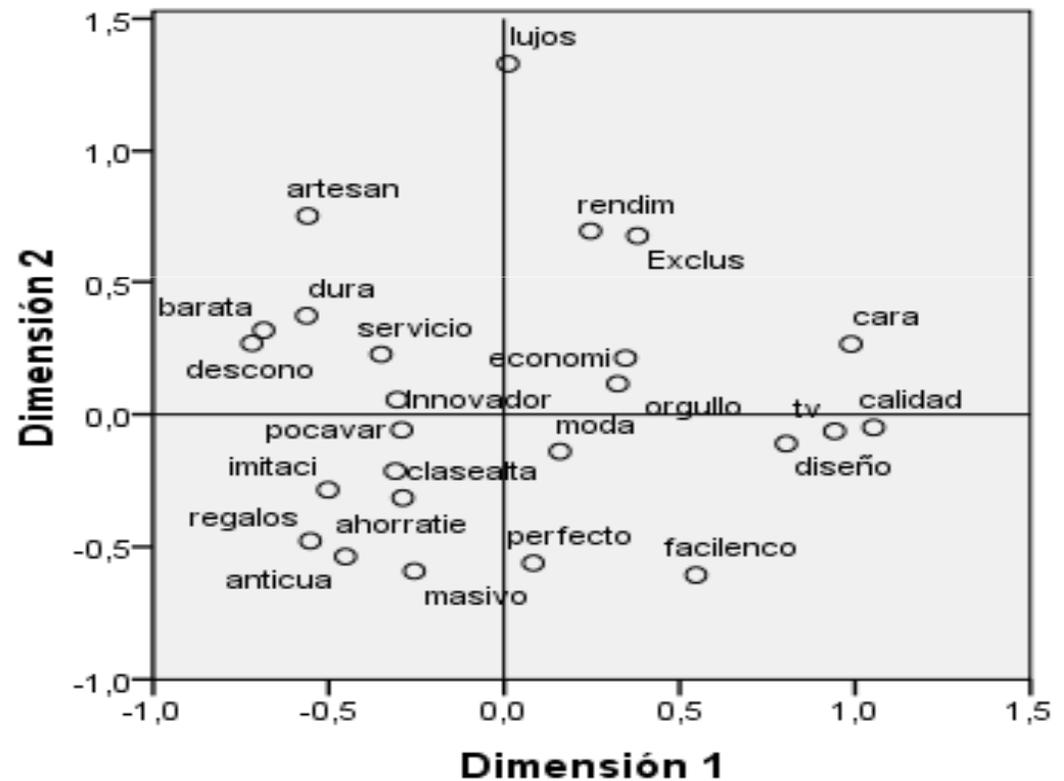
Anticuada, masivo,  
perfecto  
etc se parecen

Exclusivo se parece a  
Rendimiento

etc..

## Puntos de fila para atributo

### Simétrica Normalización



Eje 1: Coloca e Imati se oponen a Empres, Absor, Necesi y Marlboro

Eje 2: JackDa se opone a Absor

## Puntos de columna para marca

### Simétrica Normalización

