



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

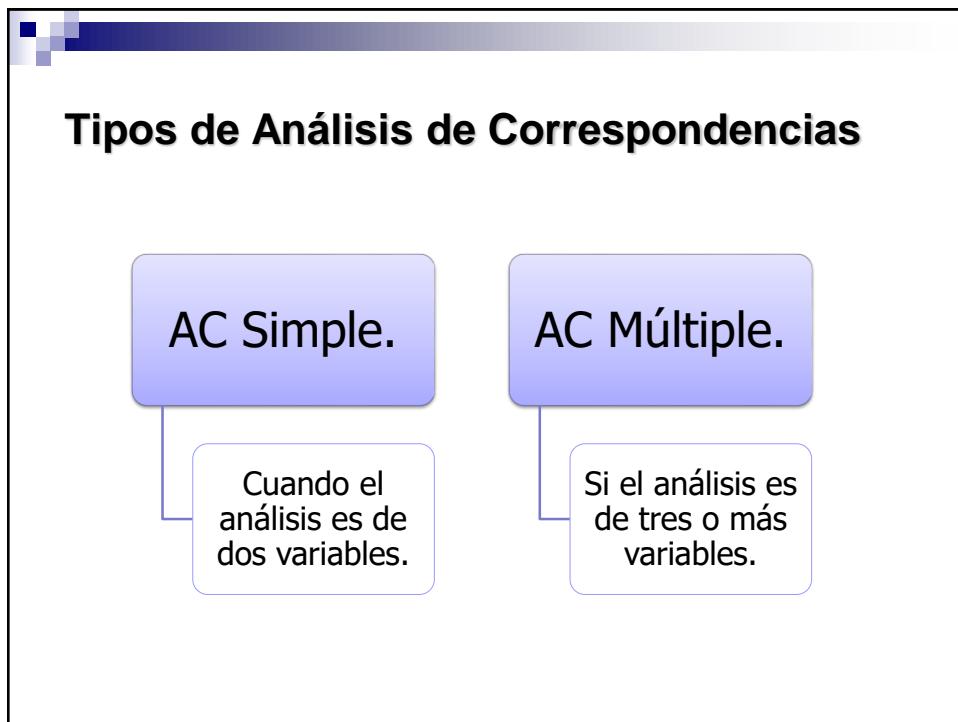
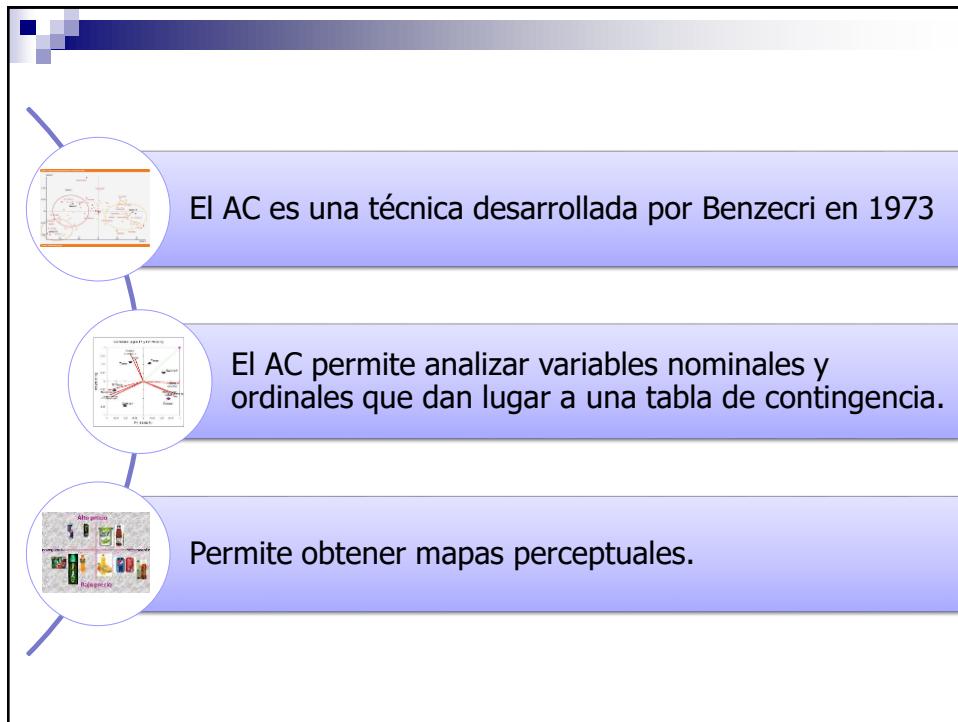
Unidad IV

Análisis de Correspondencias

Mg. Jesús Salinas Flores

jsalinas@lamolina.edu.pe

Introducción



Ejemplo - Cervezas

Se quiere determinar el posicionamiento del mercado cervecero en los estudiantes de la UNA La Molina entre las edades de 18 a 28 años.

Materiales y métodos

MARCAS	ATRIBUTOS				
	Sabor	Precio	Presentación	Publicidad	Calidad
Cristal					
Pilsen					
Cusqueña					
Brahma					
Barena					
Dorada					

Tabla de Datos

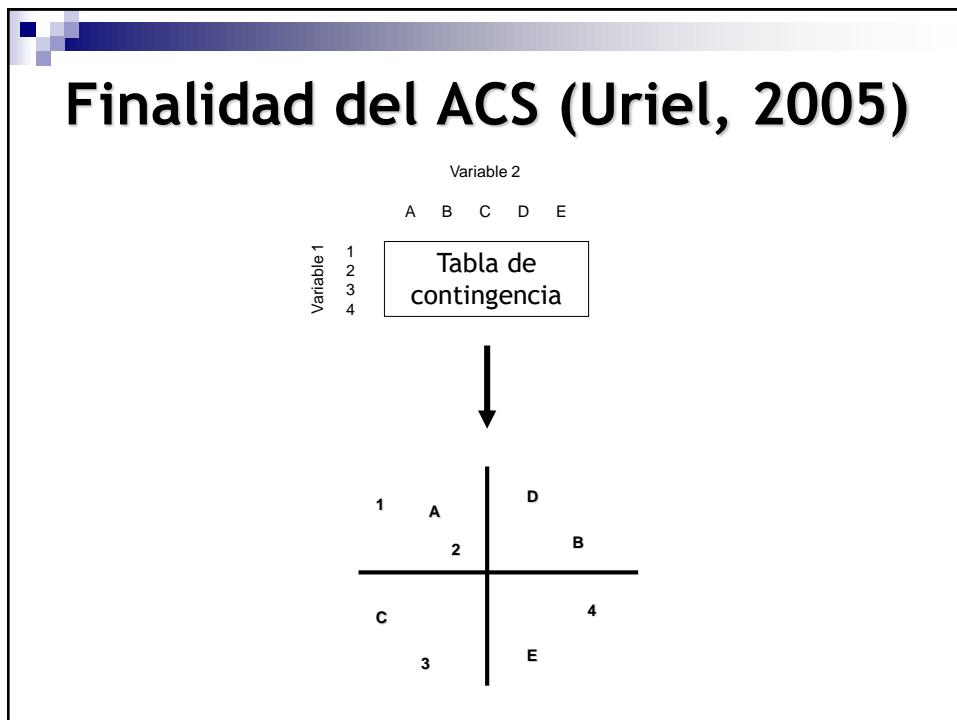
Marca	Atributo					
	Sabor	Precio	Presentación	Publicidad	Calidad	Total
Cristal	32	5	14	49	23	123
Pilsen	22	28	25	31	17	123
Cusqueña	51	4	24	10	34	123
Brahma	9	61	12	33	8	123
Barena	15	29	34	40	5	123
Dorada	10	91	11	11	0	123
Total	139	218	120	174	87	738

Prueba de Independencia

H_0 : El atributo preferido es independiente de la marca de cerveza

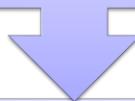
H_1 : El atributo preferido es dependiente de la marca de cerveza

Chi Square	Sig.
332.527	.000 ^a



Definición del ACS

Es un modo de representación gráfica de las tablas de contingencia.



Dicho análisis trata de evidenciar en uno o varios gráficos la mayor cantidad posible de información contenida en una tabla, fijándose no en los valores absolutos, sino en las correspondencias entre las características, es decir, en los valores relativos

Metodología

Metodología del Análisis de Correspondencias (Pedret, 2000)

DISEÑO DEL ANÁLISIS

¿Cómo deben ser las variables?

¿En qué escala deben medirse?



Ejemplo - Niveles de Inglés (Renom, 1997)

	Clase A	Clase B	Clase C	Total	Marginal
Hombres	100	200	200	500	0.25
Mujeres	300	600	600	1500	0.75
Total	400	800	800	2000	
Marginal	0.20	0.40	0.40		

	Clase A	Clase B	Clase C	Total
Hombres	0.20	0.40	0.40	1
Mujeres	0.20	0.40	0.40	1
Total	0.20	0.40	0.40	1

	Clase A	Clase B	Clase C	Total
Hombres	0.25	0.25	0.25	0.25
Mujeres	0.75	0.75	0.75	0.75
Total	1	1	1	1

Ejemplo - Niveles de inglés (Renom, 1997)

	Clase A	Clase B	Clase C	Total	Marginal
Hombres	150	50	300	500	0.25
Mujeres	250	750	500	1500	0.75
Total	400	800	800	2000	
Marginal	0.20	0.40	0.40		

	Clase A	Clase B	Clase C	Total
Hombres	0.30	0.10	0.60	1
Mujeres	0.17	0.50	0.33	1
Total	0.20	0.40	0.40	1

Tabla de Datos

Parte de una tabla de contingencia que contiene dos variables, con sus diversas modalidades, y cada casilla recoge la frecuencia en que se presentan.

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

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

Consideraciones generales...

El AC trata de determinar la posición que tienen los distintos niveles de las variables y relacionarlos entre sí en unas coordenadas a las que denominamos dimensiones.



Éstas dimensiones tienen mucha similaridad con lo que en el análisis factorial se denominaban factores.

...Consideraciones generales

Lo que en AF se denomina autovalores en ACS se denomina inercia.

La inercia va a determinar la cantidad de información o variancia que conseguimos explicar de las relaciones existentes entre las variables.

Distancia de Benzecri e Inercia

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

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

$$IT = \frac{\chi^2_{\text{calculado}}}{\text{Total de la tabla}}$$

...Consideraciones generales

La inercia total va a ser igual al valor que en las chi-cuadrado de las tablas de contingencia dividido por el número total de individuos.

Esta inercia total va a ser posteriormente repartida entre las distintas dimensiones.

Si en el AF extraía factores, en ACS se extraen dimensiones o componentes. El número máximo de éstas será igual al mínimo del número de filas o del número de columnas menos uno.



Aplicaciones

Ejemplo - Renta/Opinión (Luque, 2000)

Se pretende buscar la relación entre el poder económico de la persona y la opinión que tiene sobre el sistema sanitario público.

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.

Tabla de datos

		Opinión sobre el sistema sanitario público			
Nivel de renta		Bueno	Malo	Regular	Total
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

Pertinencia del AC

Se debe realizar previamente la prueba chi-cuadrado de independencia

Tabla de contingencia Nivel de Renta * Opinión sobre el sistema sanitario público

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

Prueba chi-cuadrado de independencia

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	40.049 ^a	6	.000
Razón de verosimilitudes	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 5.
La frecuencia mínima esperada es 25.60.

Datos – Ponderar casos...

	Renta	Opinion	frecuencia	var
1	1	1	75	
2	1	2	40	
3	1	3	35	
4	2	1	60	
5	2	2	50	
6	2	3	70	
7	3	1	20	
8	3	2	40	
9	3	3	30	
10	4	1	15	
11	4	2	40	
12	4	3	25	

Ponderar casos

No ponderar casos
 Ponderar casos mediante
 Variable de frecuencia:

Estado actual: No ponderar casos

Aceptar Pegar Restablecer Cancelar Ayuda

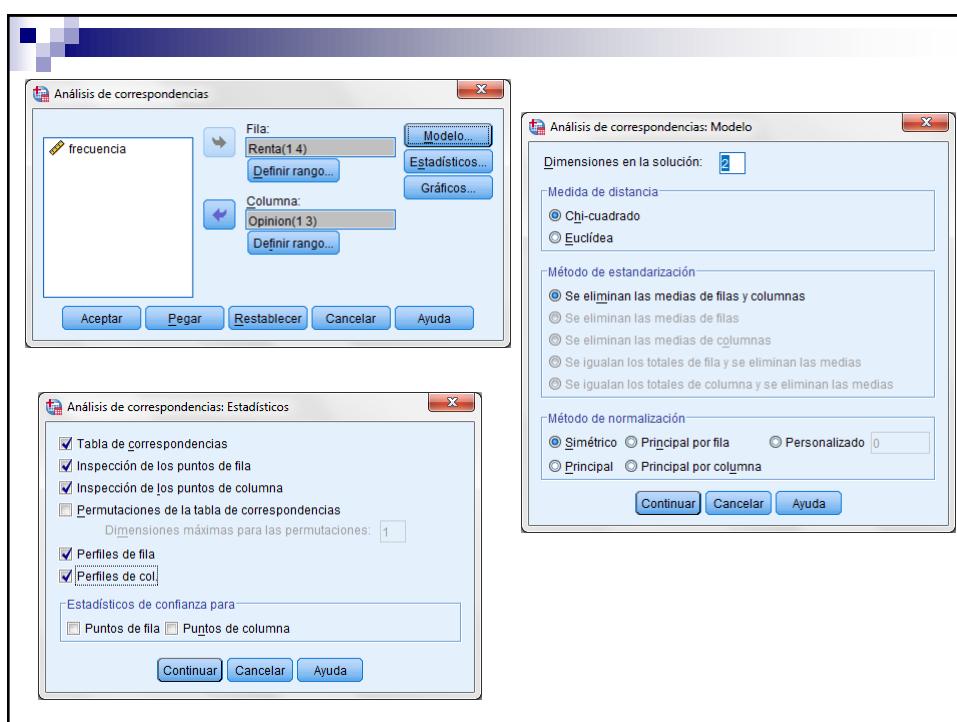
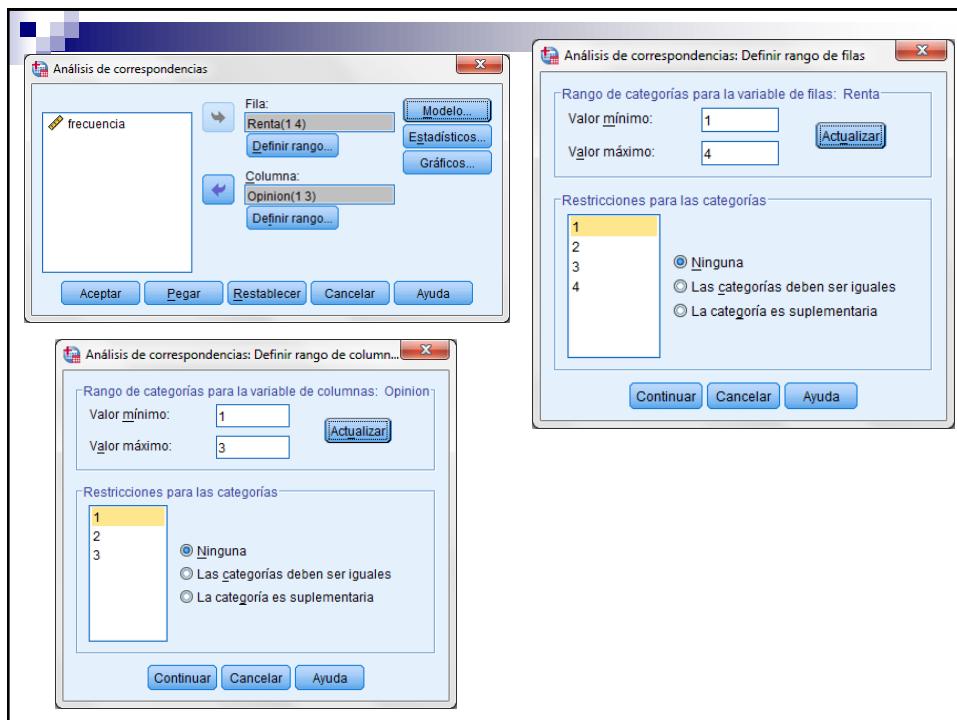


Tabla de Datos

Tabla de correspondencias

Nivel de Renta	Opinión sobre el sistema sanitario público			
	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 Fila

Perfiles de fila

Nivel de Renta	Opinión sobre el sistema sanitario público			
	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 Columna

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		Confianza para el Valor propio	
					Explicada	Acumulada	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

Interpretación de los resultados...

En la práctica suelen ser suficientes dos o tres componentes 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 componentes, que recoge la mayor cantidad de información.

...Interpretación de los resultados

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.

Cuando una línea tiene un perfil próximo al perfil medio, tiene un comportamiento medio y 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 componente α .
- λ_α . es la inercia explicada por el componente α .

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

Contribución Relativa (CR)

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

$$CR_\alpha(i) = \frac{F_\alpha^2(i)}{d^2(i, G)}, \quad \sum_\alpha CR_\alpha(i) = 1$$

Interpretación de los componentes...

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.

...Interpretación de los componentes

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.

Examen de los puntos de fila									
Nivel de Renta	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
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

Examen de los puntos columna									
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
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

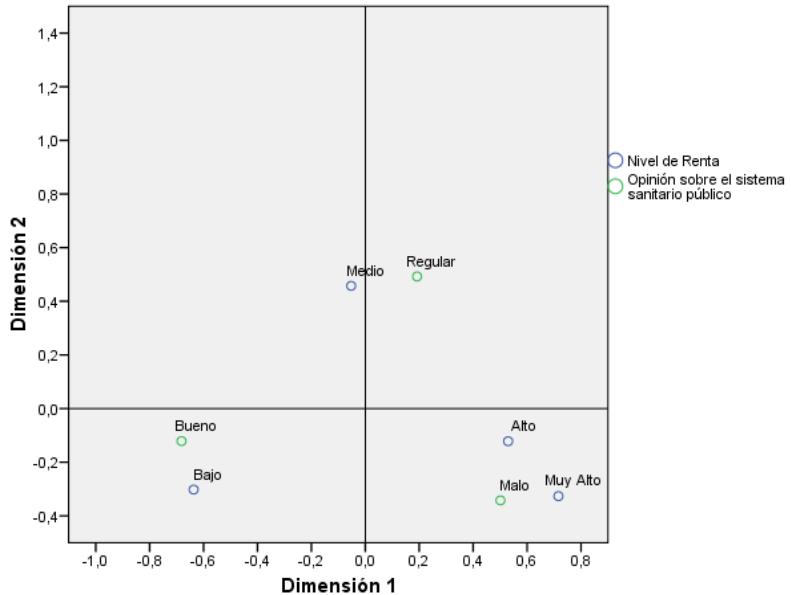
Notación

Masa	Proporción, frecuencia de la clase en el total de datos
Puntuación en la Dimensión	Coordenadas principales de las filas (columnas)
Contribución de los puntos a la inercia de la dimensión	Contribución de cada fila (columna) a la inercia del componente. Contribución Absoluta
Contribución de la dimensión a la inercia del punto	Contribución del componente a la inercia de la fila (columna). Contribución Relativa

Interpretación

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

Mapa Perceptual



Ejemplo - Supermercado

Se les pide a los encuestados que marque con una cruz el atributo que más identifica a cada supermercado

Tiendas	Atributos							
	Confiabilidad	Sofisticación	Amabilidad	Competitividad	Eficiencia	Sinceridad	Acogedor	Tradición
Jumbo								
Líder								
Tiendas por dpto								
Home Center								
Easy								
Home Depot								
Grandes Ferreterías								

Tabla de datos

Tabla de correspondencias

Tiendas	Atributos								
	Confabilidad	Sofisticación	Amabilidad	Competitividad	Eficiencia	Sinceridad	Acogedor	Tradición	Margen activo
Jumbo	121	78	69	65	84	35	71	87	610
Líder	49	29	43	92	58	17	29	14	331
Tiendas por dpto	90	81	48	82	57	28	61	87	534
Home Center	119	61	78	92	108	50	59	81	648
Easy	45	20	42	57	39	19	14	10	246
Home Depot	51	26	27	80	36	22	25	10	277
Grandes Ferreterías	39	17	31	22	32	28	22	50	241
Margen activo	514	312	338	490	414	199	281	339	2887

Perfiles Fila

Perfiles de fila

Tiendas	Atributos								
	Confabilidad	Sofisticación	Amabilidad	Competitividad	Eficiencia	Sinceridad	Acogedor	Tradición	Margen activo
Jumbo	.198	.128	.113	.107	.138	.057	.116	.143	1.000
Líder	.148	.088	.130	.278	.175	.051	.088	.042	1.000
Tiendas por dpto	.169	.152	.090	.154	.107	.052	.114	.163	1.000
Home Center	.184	.094	.120	.142	.167	.077	.091	.125	1.000
Easy	.183	.081	.171	.232	.159	.077	.057	.041	1.000
Home Depot	.184	.094	.097	.289	.130	.079	.090	.036	1.000
Grandes Ferreterías	.162	.071	.129	.091	.133	.116	.091	.207	1.000
Masa	.178	.108	.117	.170	.143	.069	.097	.117	

Perfiles Columna

Perfiles de columna

Tiendas	Atributos								
	Confiabilidad	Sofisticación	Amabilidad	Competitividad	Eficiencia	Sinceridad	Acogedor	Tradición	Masa
Jumbo	.235	.250	.204	.133	.203	.176	.253	.257	.211
Líder	.095	.093	.127	.188	.140	.085	.103	.041	.115
Tiendas por dpto	.175	.260	.142	.167	.138	.141	.217	.257	.185
Home Center	.232	.196	.231	.188	.261	.251	.210	.239	.224
Easy	.088	.064	.124	.116	.094	.095	.050	.029	.085
Home Depot	.099	.083	.080	.163	.087	.111	.089	.029	.096
Grandes Ferreterías	.076	.054	.092	.045	.077	.141	.078	.147	.083
Margen activo	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	

Resumen

Dimensión	Valor propio	Inercia	Chi-cuadrado	Sig.	Proporción de inercia		Confianza para el Valor propio	
					Explicada	Acumulada	Desviación típica	Correlación
								2
1	.233	.054			.711	.711	.017	.004
2	.117	.014			.179	.891	.018	
3	.069	.005			.063	.953		
4	.043	.002			.024	.978		
5	.038	.001			.019	.997		
6	.016	.000			.003	1.000		
Total		.076	219.953	.000 ^a	1.000	1.000		

a. 42 grados de libertad

Examen de los puntos de fila										
Tiendas	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
Jumbo	.211	-.369	.114	.009	.124	.024	.775	.037	.812	
Líder	.115	.750	.082	.016	.277	.007	.935	.006	.940	
Tiendas por dpto	.185	-.343	.495	.011	.093	.388	.457	.479	.936	
Home Center	.224	-.089	-.259	.003	.008	.129	.149	.639	.788	
Easy	.085	.638	-.365	.011	.149	.097	.767	.126	.893	
Home Depot	.096	.735	.208	.014	.223	.036	.847	.034	.881	
Grandes Ferreterías	.083	-.594	-.670	.013	.127	.321	.532	.340	.872	
Total activo	1.000			.076	1.000	1.000				

a. Normalización Simétrica

Examen de los puntos columna										
Atributos	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
Confiabilidad	.178	-.052	-.005	.001	.002	.000	.084	.000	.084	
Sofisticación	.108	-.254	.612	.007	.030	.347	.242	.707	.949	
Amabilidad	.117	.148	-.392	.004	.011	.154	.154	.541	.694	
Competitividad	.170	.797	.225	.027	.463	.074	.923	.037	.960	
Eficiencia	.143	.163	-.280	.004	.016	.096	.236	.349	.585	
Sinceridad	.069	-.055	-.636	.005	.001	.238	.010	.677	.687	
Acogedor	.097	-.250	.328	.003	.026	.090	.465	.402	.868	
Tradición	.117	-.945	-.047	.025	.451	.002	.959	.001	.960	
Total activo	1.000			.076	1.000	1.000				

a. Normalización Simétrica

