



3 datos
tratamiento estadístico

Estudios de Mercado: el Análisis Factorial de Componentes Principales “vs” el Análisis de Correspondencias

Análisis Factorial de COMPONENTES PRINCIPALES

Análisis de CORRESPONDENCIAS Simple

- Son procedimientos de análisis estadístico: **MULTIVARIADOS** y de **INTERDEPENDENCIA** (estudian correlaciones entre variables)
- Posibilitan la identificación de **DIMENSIONES** o **FACTORES** teóricos complejos y subyacentes ("**constructos**") no observables directamente
- Extraen información simplificada, reduciendo el número de datos a interpretar con la menor pérdida de información posible.

Gran conjunto de datos empíricamente observados

se extraen

Dimensiones teóricas
(no observables)

que mantienen como mínimo 50% de la información inicial, aunque es deseable que esté por encima del 60% (este % de corte depende de la Ciencia donde aplica)

Ejemplo una encuesta con 20 preguntas = 20 ítems, genera 20 respuestas = 20 datos empíricos por persona

20 VARIABLES EMPÍRICAMENTE MEDIDAS que contienen el 100% de información de cada persona

Consecuencia: cumplen el objetivo de sintetizar la realidad observada mostrando la estructura de los datos

Por ejemplo determinar las dimensiones que identifican la imagen comercial de una marca

5	2	2	2	1	2	2	1	12	28	12	11	24	47
12	2	2	1	2	1	2	2	12	37	27	13	21	61
10	0	1	0	0	0	1	0	2	19	17	9	25	51
												74	
												94	
												70	
												33	
												39	
												62	
												71	
												92	
												32	
												20	
3	0	1	0	0	0	1	1	3	9	8	12	11	31
1	0	0	0	1	1	0	1	3	5	6	1	3	10
5	1	0	0	1	1	1	1	5	13	13	19	15	47
9	1	0	1	2	2	0	1	7	21	9	28	23	60
4	2	1	0	3	2	2	2	12	24	20	30	10	60
2	1	0	1	1	1	1	0	5	12	9	16	18	43
5	0	1	0	3	1	1	2	8	24	15	18	26	59

Análisis Factorial de COMPONENTES PRINCIPALES

Matriz de DATOS

Tabla de N individuos **x** P variables cuantitativas (ocasionalmente pueden ser dicotómicas)

	Var.1	Var.2	...	j	...	Var.P
Ind. 1				⋮		
Ind. 2				⋮		
Ind. 3				⋮		
...				⋮		
i				X_{ij}		
...				⋮		
Ind. N				⋮		

Resultado final

Una tabla con las variables y sus pesos (importancia) en las dimensiones extraídas

VARIABLES	Dimensiones	
	1	2
Renta per cápita	,79	,04
Población	,74	,12
Sistema de partidos políticos	,68	,15
Estabilidad política	,33	,89
Homogeneidad cultural	,11	,86
Libertad política	,02	-,84



Análisis de CORRESPONDENCIAS Simple

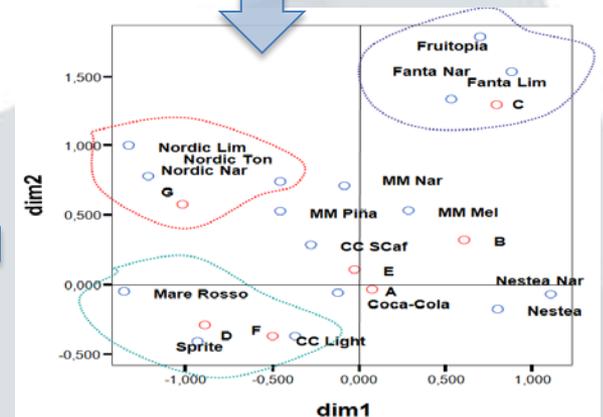
Matriz de DATOS

Conjunto de C columnas **x** F filas, en forma de tabla de contingencia (casillas = frecuencias)

	C ₁	C ₂	...	j	...	C _c
F ₁				⋮		
F ₂				⋮		
F ₃				⋮		
...				⋮		
i				f_{ij}		
...				⋮		
F _f				⋮		

Resultado final

Un gráfico con las variables posicionadas entre 2 ejes



Solo la puede dar el INVESTIGADOR

¿INTERPRETACIÓN TEÓRICA?

Análisis Factorial de COMPONENTES PRINCIPALES

Abreviadamente: **AF de CP** o bien **ACP**

Fundamentación Teórica

- El **ACP** se basa en la existencia de muchas inter-correlaciones ($r > ,300$) estadísticamente significativas ($p < ,050$) entre las variables empíricas medidas en la muestra
- Las variables se agrupan en factores (componentes) en función de dichas correlaciones

El punto de partida de un A.F. de C.P. es una **matriz de correlaciones**

que es una tabla de doble entrada que contiene las correlaciones de cada variable con todas las demás

	Var.1	Var.2	Var.3	Var.4	Var.5
Var.1	1				
Var.2	0	1			
Var.3	0	0	1		
Var.4	0	0	0	1	
Var.5	0	0	0	0	1

La de al lado, es matriz donde las variables **NO** correlacionan entre sí. Por ello, **NO HAY POSIBILIDAD** de encontrar factores comunes a las variables.

Se llama: **matriz identidad** o **matriz unidad**

	Var.1	Var.2	Var.3	Var.4	Var.5
Var.1	1				
Var.2	0,787	1			
Var.3	0,315	0,316	1		
Var.4	0,325	0,321	0,773	1	
Var.5	0,873	0,855	0,317	0,281	1

El de al lado es un ejemplo ficticio de una matriz que contiene muchas correlaciones ($r > ,300$) entre todas las variables; por eso **SI** es posible extraer factores (dimensiones o componentes).

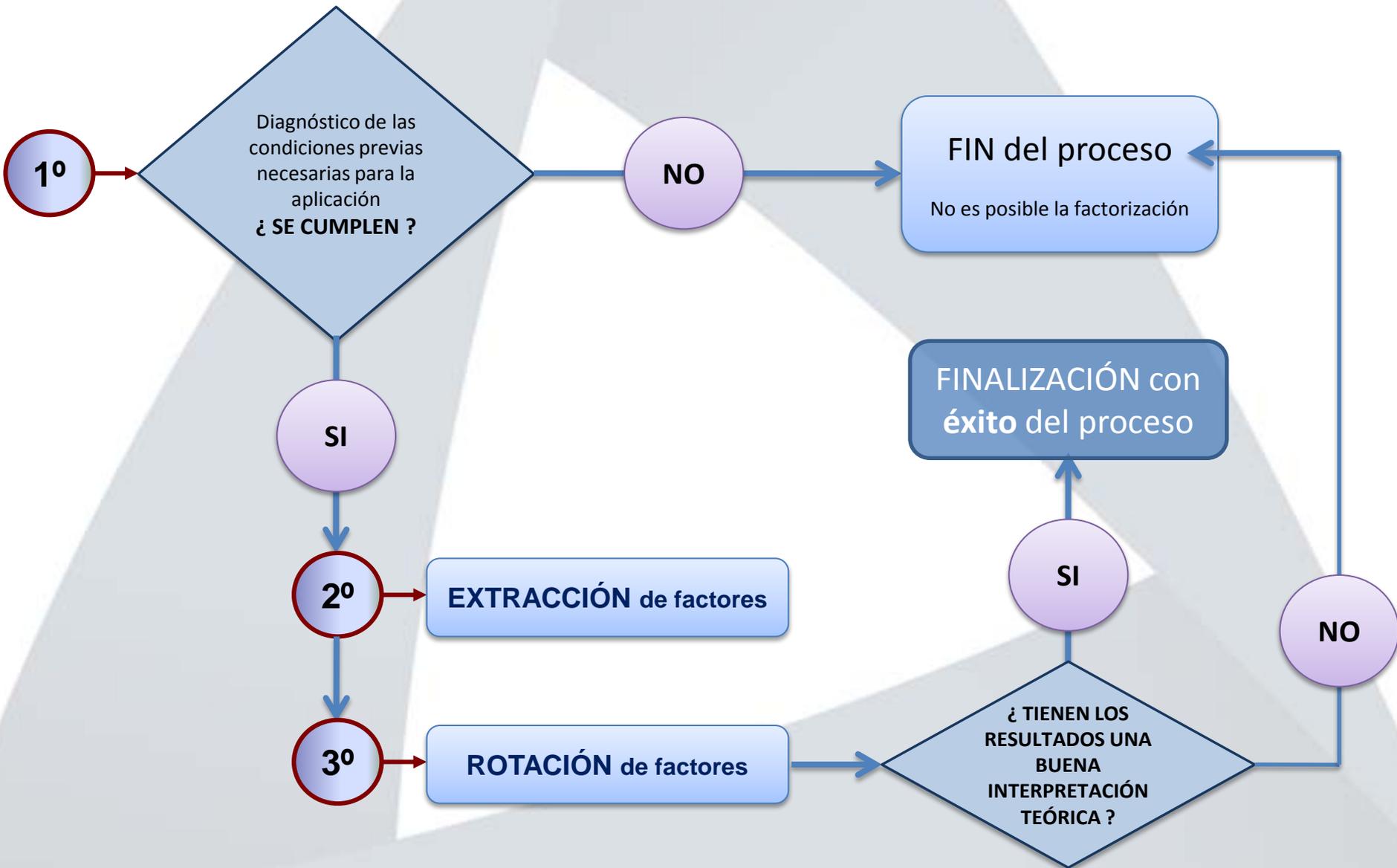
Las variables 1, 2 y 5 correlacionen entre sí con valores $> ,700$; por eso se definen como el **factor 1**. Las variables 3 y 4 correlacionan entre sí, pero no lo hacen tanto con la 1, 2 y 5; por ello son **otro factor** diferente (que sería la **dimensión 2**)

Matriz de componentes

	Dimensión	
	1	2
Var.5	,804	
Var.2	,737	
Var.1	,670	
Var.4		,677
Var.3		,641

Solución final de este ACP ficticio

Esquema metodológico de un A.F. de C.P.



1º

DIAGNÓSTICO de la condiciones

1

La muestra debe tener al menos 200 sujetos y/o a ser posible en una ratio de 10 casos/variable (mínimo: 5/1). Se utiliza la Prueba de **adecuación muestral** de **Kaiser-Meier-Olkin**. Es un valor descriptivo (en una escala de 0 a 1). Se precisa un **KMO >,500** y cuanto más cerca de 1, mejor será la adecuación muestral.

2

Comprobación de la existencia de inter-correlaciones en la matriz:

- Un índice descriptivo es el **Determinante** de la matriz; se precisa que sea: **D≈0**
- Una prueba inferencial llamada **Test de Esfericidad de Bartlett**, cuya hipótesis nula es que la matriz de correlaciones es una matriz identidad o unidad (ver anterior). Para que sea factible realizar el ACP se precisa que en este test la **p < ,050** para **R** la H₀ (significación).



Ejemplo con datos reales, resuelto con IBM-SPSS.19

Matriz de correlaciones ^a

a. Determinante = 8,15E-006

INTERPRETACIÓN:

El valor de **KMO=,691** > ,500 es bueno.

KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,691
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	2142,370
	gl	300
	Sig.	,000

El **determinante** es casi **0**
Y la prueba de **Bartlett** tiene una **p=,000** < ,050 que permite rechazar la hipótesis de matriz identidad.

Luego, el diagnóstico es positivo, es decir que se cumplen satisfactoriamente las condiciones para utilizar el ACP

2º

EXTRACCIÓN de factores

- Existen varios procedimientos: máxima verosimilitud, componentes principales, mínimos cuadrados, etc...
- El **ACP** utiliza lógicamente el de **Componentes Principales**, que le da nombre, y que prioriza obtener la máxima cantidad de información inicial (varianza total) en el menor número de factores (dimensiones) posible = Solución más fácil

¿Cuántos factores hay que extraer?

- Hay varios criterios en el AF, pero en el AF de CP el más usado es el del "autovalor"
- ACP**: Los que tengan un **autovalor** > 1. O bien: aquellos que logren explicar al menos el 60% de la variabilidad inicial
- También es habitual usar el criterio que fija el número de dimensiones a extraer y en este caso por similitud con el ACS, el ACP suele fijar ese número en 2 = espacio de 2 ejes



Ejemplo con datos reales, resuelto con IBM-SPSS.19

Utilizando el convenio de extraer los factores con **autovalor** >1, se han encontrado 8 dimensiones

Inicialmente hay tantos factores como variables (100% de información = 100% de varianza)

Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	5,820	23,279	23,279	5,820	23,279	23,279	4,446	17,784	17,784
2	3,109	12,436	35,715	3,109	12,436	35,715	2,528	10,110	27,895
3	2,080	8,321	44,036	2,080	8,321	44,036	1,943	7,774	35,668
4	1,503	6,013	50,049	1,503	6,013	50,049	1,913	7,650	43,319
5	1,445	5,781	55,830	1,445	5,781	55,830	1,871	7,484	50,803
6	1,332	5,328	61,158	1,332	5,328	61,158	1,787	7,150	57,952
7	1,223	4,890	66,048	1,223	4,890	66,048	1,706	6,826	64,778
8	1,020	4,079	70,126	1,020	4,079	70,126	1,337	5,348	70,126
9	,835	3,342	73,468						
10	,827	3,308	76,776						
11	,703	2,811	79,587						
12	,690	2,758	82,346						
13	,579	2,316	84,661						
14	,522	2,089	86,750						
15	,492	1,969	88,719						
16	,446	1,784	90,503						
17	,403	1,611	92,114						
18	,388	1,554	93,668						
19	,352	1,407	95,075						
20	,298	1,190	96,265						
21	,270	1,078	97,343						
22	,263	1,051	98,394						
23	,161	,642	99,036						
24	,123	,490	99,527						
25	,118	,473	100,000						

% de variabilidad explicada por los factores extraídos

3º

ROTACIÓN de factores

Existen 2 bloques de métodos

En muchos casos, hay que emplear 2 ó más métodos de rotación y comparar las soluciones buscando la más adecuada

- **Rotación ORTOGONAL:** que se emplea cuando los factores NO están correlacionados entre sí, o lo hacen de una forma muy débil. De entre ellos, el más utilizado es el método **Varimax**.
- **Rotación OBLÍCUA:** que se emplea cuando los factores SÍ están correlacionados entre sí. El más habitual en este caso es el método **Oblimín** (aunque también se utiliza bastante el **Promax**).

Solución FINAL

Es una matriz que nos indica qué variables, y cuánto, **saturan** (*correlacionan*) en cada uno de los factores encontrados. La saturación representa el *peso*, la *importancia*, de la variable dentro del componente (o dimensión).



Ejemplo con datos reales, resuelto con IBM-SPSS.19

Ejemplo: El factor nº 1 está formado por las variables:
dolores interfieren actividades + tengo dolores + me siento enfermo + estoy en cama + no disfruto cosas + no soy capaz de... + no duermo bien + me canso con facilidad

Ejemplo: El factor nº 2 está formado por las variables:
me preocupa que el trasplante no funcione + me preocupa la muerte + estoy preocupado por empeorar

Ítems de una escala de Calidad de Vida en enfermos con trasplante de médula

Matriz de componentes rotados

	Componente							
	1	2	3	4	5	6	7	8
Los dolores interfieren en mis actividades	,776							
Tengo dolores	,773							
Me siento enfermo	,731							
Estoy forzado a permanecer en cama	,677							
Disfruto de las cosas que me gustaban	-,616							
Soy capaz de desenvolverme por mi mismo	-,601							
Duermo bien	-,601							
Me canso con facilidad	,598							
Me preocupa que el trasplante no funcione		,806						
Me preocupa la muerte		,777						
Estoy preocupado de empeorar		,768						
He perdido la esperanza			,835					
Respiración entrecortada			,635					
Me siento triste			,527					
Me siento nervioso			,508					
Necesito descansar				,841				
Tengo nauseas				,664				
Me siento alejado de la gente					,789			
Me siento apartado de amigos					,709			
Tengo molestias bucales					-,533			
Hablamos poco de enfermedad en familia						-,786		
Me gusta mi aspecto físico						,731		
He aceptado mi enfermedad							,783	
Tengo molestias oculares							-,721	
Me siento unido a mi pareja								,889

Habitualmente sólo se indican las saturaciones **>,500** que indican una asociación elevada

Análisis de CORRESPONDENCIAS Simple

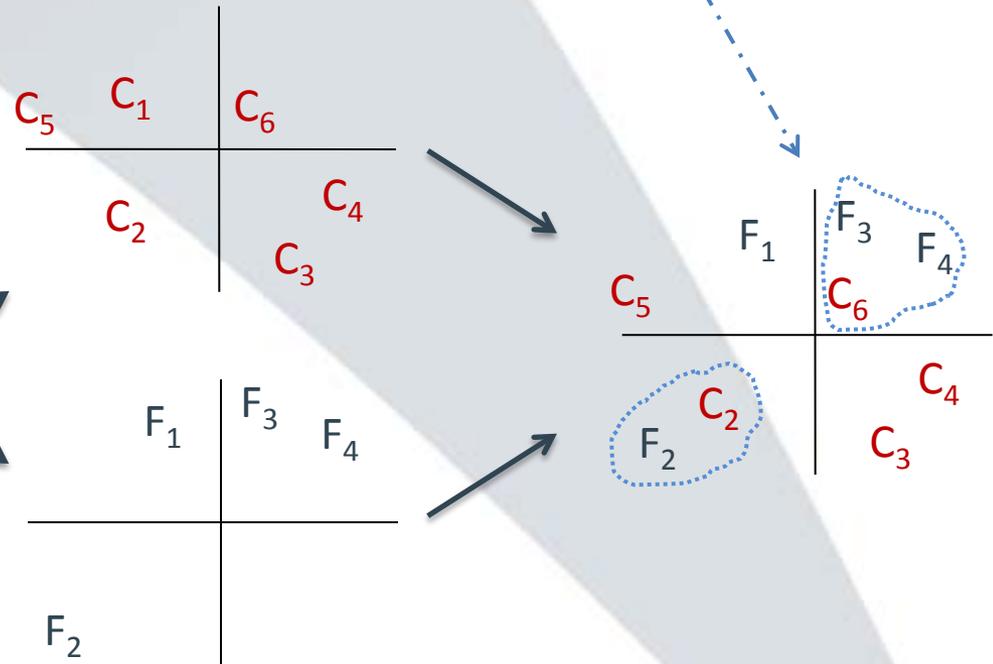
Abreviadamente: **ACS**

Fundamentación Teórica

- El **ACS** es un método estadístico que nos permite describir una tabla de contingencia en la que un conjunto de N individuos homogéneos ha sido clasificado en base a 2 criterios o variables categóricas.
- Las categorías de las filas y de las columnas quedarán representadas como puntos de un espacio vectorial bidimensional (la solución más simple) de forma que los correspondientes espacios se superponen para dar lugar a una representación gráfica conjunta: *Mapa Perceptual*, en el que se definen 4 cuadrantes resultado de la interpretación bipolar de ambos ejes.

	C_1	C_2	C_3	C_4	C_5	C_6
F_1	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}
F_2	f_{21}	f_{22}	f_{23}	f_{24}	f_{25}	f_{26}
F_3	f_{31}	f_{32}	f_{33}	f_{34}	f_{35}	f_{36}
F_4	f_{41}	f_{42}	f_{43}	f_{44}	f_{45}	f_{46}

Las categorías de F o de C pueden ser otras variables. Sólo serían agrupables como categorías, si ha empleado igual forma de medir a todas ellas (ej.: opinión sobre los productos que ofrece una empresa; C sería cada producto. Y en F podría estar la variable zona de venta)



Desarrollo

- ✓ Preparar tablas de contingencia: Perfiles filas y columnas (perfiles = proporciones)
- ✓ Distancias chi-cuadrado entre filas y columnas
- ✓ Valores propios e inercia de valores propios → Significación de la asociación
- ✓ Contribuciones absolutas y relativas
- ✓ Coordenadas de filas y columnas
- ✓ Representación factorial de filas y columnas → Mapa perceptual

Tabla resumen

VALOR PROPIO

Otras denominaciones: Autovalor, Lambda λ , o "Eiguenvalue"
Indica la cantidad total de la varianza que está explicada por una dimensión

INERCIA

El cuadrado del valor propio

Inercia EXPLICADA

Proporción de la cantidad de inercia de cada eje con respecto a la inercia total
Indica la importancia de las dimensiones representadas en cada uno de los ejes del mapa

Chi-Cuadrado

Estadístico que contrasta la hipótesis de independencia entre las variables de fila y columna



Ejemplo con datos reales,
resuelto con IBM-SPSS.19

Resumen

Dimensión	Valor propio	Inercia	Chi-cuadrado	Sig.	Proporción de inercia		Confianza para el Valor propio	
					Explicada	Acumulada	Desviación típica	Correlación
								2
1	,170	,029			,560	,560	,030	,127
2	,150	,023			,440	1,000	,027	
Total		,051	56,541	,000 ^a	1,000	1,000		

a. 4 grados de libertad

INTERPRETACIÓN:

Se han extraído 2 dimensiones. La primera con una inercia de 0,029 ($\lambda=0,17$) acumula el 56% de la inercia total.

La segunda, con una inercia de 0,023 ($\lambda=0,15$) explica el 44% restante.

La asociación entre las variables es altamente significativa con $p<,05$ ($p=,000$).

Por tanto el ACS nos permitirá reducir dimensiones en las variables estudiadas y elaborar un mapa perceptual.

Tablas de puntos de fila y de columna

MASA

La MASA (peso) asociada a cada categoría de fila F_f o de columna C_c expresa la importancia relativa de dicha categoría dentro de la variable que entra en fila o en columna. Para cada perfil, se calcula en función de su total marginal y del gran total (N).

PUNTUACIONES

La puntuación en cada dimensión de cada categorías (bien de fila o bien de columna) Corresponde con las coordenadas de la categoría en el mapa perceptual

CONTRIBUIÓN

De los puntos a la dimensión: importancia de la categoría para definir/explicar el eje
Del eje a los puntos: porcentaje de la categoría explicado desde cada dimensión



Ejemplo con datos reales, resuelto con IBM-SPSS.19

Examen de los puntos columna^a

Ideología Política	Masa	Puntuación en la dimensión		Inercia	Contribución				
		1	2		De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto		
					1	2	1	2	Total
Izqda / Ext.Izqda	,504	-,240	,311	,012	,171	,325	,401	,599	1,000
Centro	,381	,504	-,138	,018	,571	,048	,938	,062	1,000
Dcha / Extr.DCha	,115	-,616	-,904	,022	,258	,627	,344	,656	1,000
Total activo	1,000			,051	1,000	1,000			

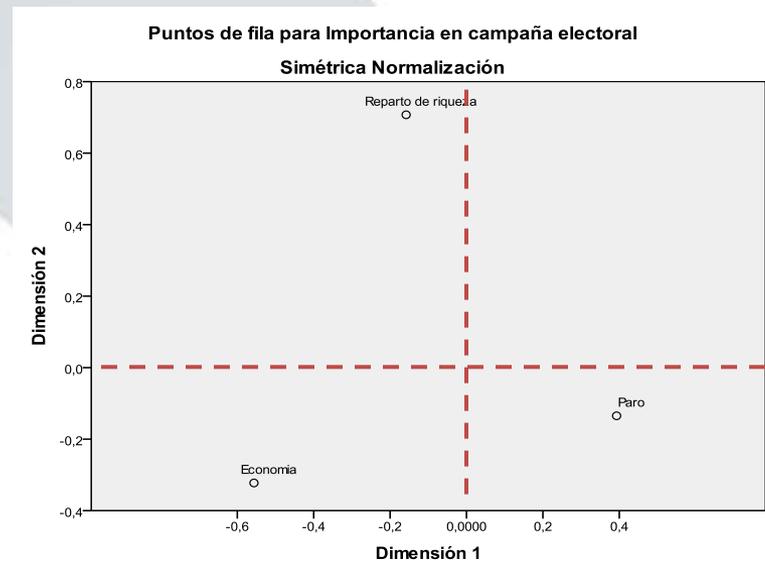
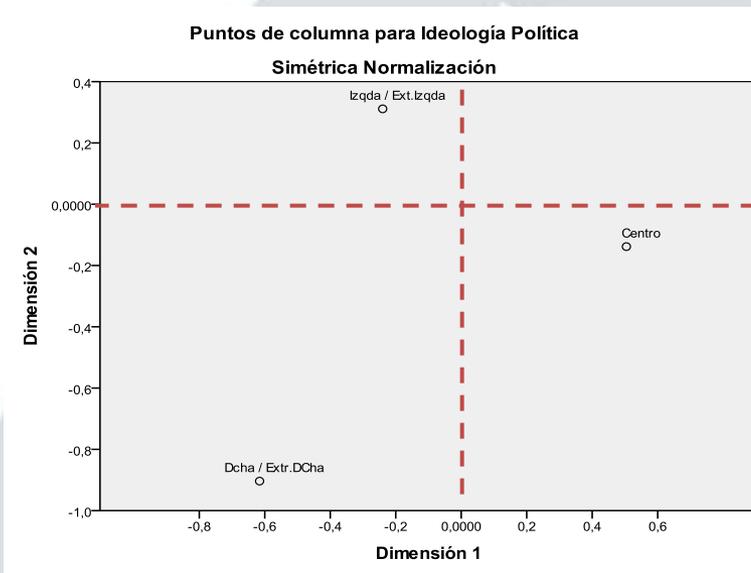
a. Normalización Simétrica

INTERPRETACIÓN:

La variable analizada en columnas es ideología política (con 3 categorías.)
La masa nos indica que la categoría más representada es Izqda./Extr.Izqda. (0,504). Su inercia es 0,012 (sobre el mismo total 0,051 anterior). Su situación en los ejes es: -0,240 en eje X (dimensión 1), es decir a la izquierda del origen) y 0,311 en el eje Y, es decir por encima del origen. Esto se verá mejor en el gráfico.
Tiene una mayor contribución, importancia, en la definición de la dimensión representada en el eje 2 (0,325 vs 0,171). Y dicha dimensión explica el 59,9% de la categoría ideología de izquierdas.

Puntos, coordenadas, de fila y de columna

Representación gráfica, por variables, a partir de las puntuaciones en cada dimensión



En el eje **X** siempre estará representada la Dimensión 1 del ACS y en el eje **Y** la Dimensión 2 del mismo.

INTERPRETACIÓN:

Según Ideología política:

El eje 1 sitúa en el lado positivo a los sujetos de Centro frente a los de izquierda; mientras que en el eje 2 se oponen los de ideología de izquierda (en el lado positivo, superior) frente a los de derecha

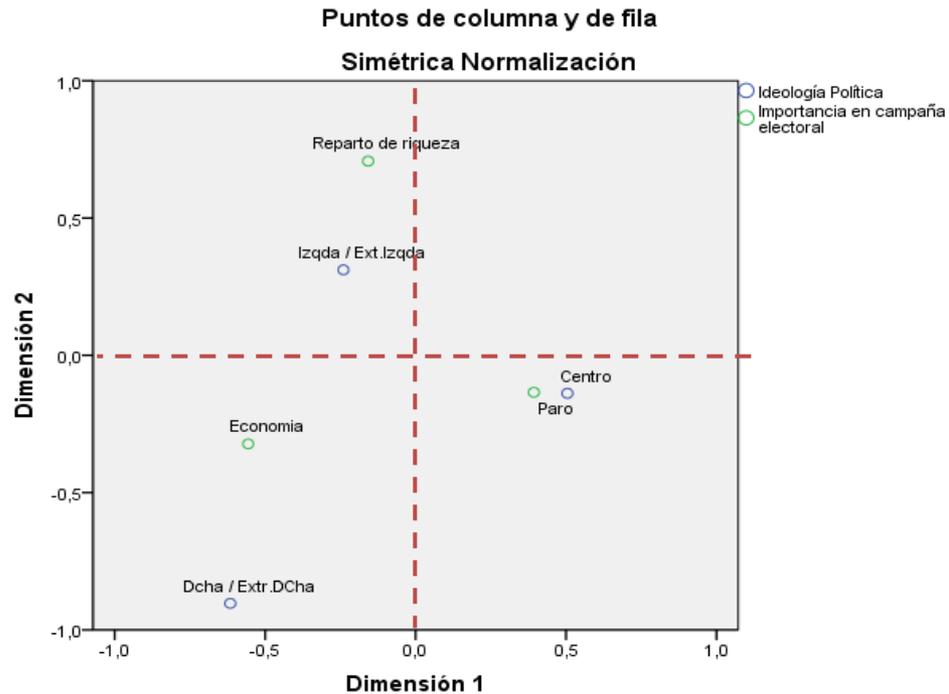
INTERPRETACIÓN:

Según importancia en la campaña electoral:

El eje 1 sitúa en el lado positivo al tema del paro frente a la economía; mientras que en el eje 2 se oponen el reparto de riqueza (en el lado positivo, superior) frente a la economía

Mapa perceptual

Representación gráfica, conjunta de todas las categorías de ambas variables, en los mismo ejes dimensionales



INTERPRETACIÓN:

Por cuadrantes:

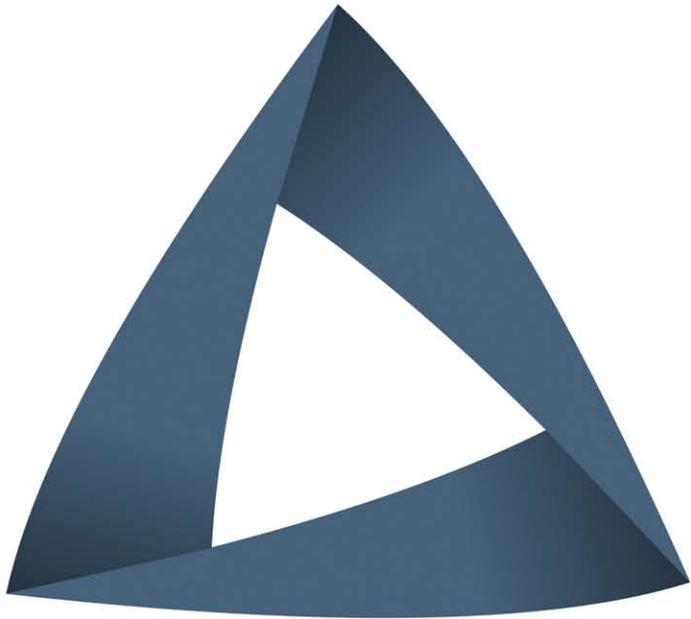
El cuadrante I (superior, derecha: ++) está vacío.

El cuadrante II (superior, izquierda: - +) asocia a los casos con ideología de izquierdas con el reparto de riqueza

El cuadrante III (inferior, izquierda: - -) asocia a las personas de derecha con la importancia de la economía

Y en el cuadrante IV (inferior, derecha: + -) asocia estrechamente a los sujetos de centro con la importancia del paro

Muchas gracias por su atención



3 datos

tratamiento estadístico

www.3datos.es