Análisis de correspondencias

utilizando el SPSS

Vamos a realizar el Análisis de correspondencias sobre el mismo ejemplo (ejemplo 6.1 de los apuntes), sobre el que ya se obtuvo el análisis mediante el R, pero ahora empleando el paquete estadístico SPSS.

Lo primero que conviene recordar es que la hoja de datos de SPSS mantiene la estructura de filas, que representan a los individuos, y columnas, que representan a las variables. En consecuencia, no se puede situar en la hoja de datos una tabla de contingencia donde las filas sean las categorías de una variable, y las columnas las categorías de la otra variable. La estructura de individuos por variables, obligaría a introducir tantas filas como individuos, pero eso se puede evitar haciendo uso de

Datos – Ponderar casos ...

🔛 epa04	99.sav [Co	onjunto_de_da	itos1] - Edito	or de datos SP	SS											[- 8 🗙
Archivo E	dición Ver	Datos Transform	mar Analizar	Gráficos Utilidad	des Vent	ana ?											
	🕒 📴 🔺	Definir propied	lades de variabl	es	🐼 🔕	•											
6.		Copiar propied	lades de datos.		P -	-									Vicible	3 do 3 variab	00
6 : Nuevo atributo personalizado Definir fechas Definir conjuntos de respuestas múltiples															VISIDIE.	J ue J valiaŭ	e5
1		Definir conjunt	tos de respuesta	as múltiples	var	var	Var	Var	Var	Var	Var	var	var	Var	Var	Var	Var 🔨
2		Identifican ene	aa dualaadaa														
3		Tuerruncar cas	us uupiicauus		-												
4		Ordenar casos			-												
5		Transponer															
6		Eundir archivo					1										
7	(Agregar			-												
8	Castilla	Diseño este es															
9		Diseno ortogor	nai														
10	Comunid	Copiar conjunt	to de datos														
11		Segmentar arc	hivo														=
12		Seleccionar casos															
13	Comuni	Ponderar caso	s														
14	Reg	gión de Murcia	ocupado	399200													
15	C. Fo	oral de Navarra	ocupado	206800													
16		La Rioja	ocupado	97000													
17		País Vasco	ocupado	784600													
18	(Ceuta y Melilla	ocupado	40000													
19		Andalucía	parado	771800													
20		Aragón	parado	39300													
21		Asturias	parado	71200													
22		Daleares	parado	24500													
23		Cantahria	parado	29400													
24	(Cantabria Castilla y León	parado	145700													
26	Castilla	a La Mancha	narado	97600													
27	oustine	Cataluña	parado	270300													
28	Comunid	ad Valenciana	parado	228600													
29		Extremadura	parado	103600													
30		Galicia	parado	182800													
31	Comuni	idad de Madrid	parado	275100													
32	Reg	gión de Murcia	parado	60600													
33	C. Fo	oral de Navarra	parado	15900													
34		La Rioja	parado	9500													
35		País Vasco	parado	123800													
36	(Ceuta y Melilla	parado	14100													
37		Andalucía	inactivo	2898000													
38		Aragón	inactivo	510500													
39		Asturias	inactivo	526000													
40		Baleares	inactivo	276300													
41		Canarias	inactivo	607200													
42		Cantabria	inactivo	248400													
43	Contille	Jastilla y León	inactivo	722100													
44	Castilla sta de datos	I La mancha Vista de varia	ables /	1 32 100		1											>
Ponderar ca	isos	The action					SP	SS El procesado	or está prepara	obe	1						

En el menú de diálogo que se abre tras efectuar la selección anterior, se escogería la variable que contiene las frecuencias de cada combinación de niveles de las variables categóricas.

Una vez situados los datos en la hoja, y efectuada la ponderación (lo cual deja una leyenda en el margen inferior, en la zona derecha), para obtener un Análisis de correspondencias se pincha en



Analizar – Reducción de datos – Análisis de correspondencias ...

Indica que el fichero está ponderado

Tras ejecutar las orden anterior, se accede a un cuadro de diálogo en el que, como es costumbre en SPSS, las variables del fichero de datos figuran en un cuadro a la izquierda. Seleccionamos la variable que deseamos que figure en las filas de la tabla de contingencia (en este caso la comunidad autónoma), indicamos el rango de valores de la variables (de 1 a 18), y hacemos lo mismo con la variable de columna (la situación respecto del empleo).

Aparecen cinco botones a la derecha (Aceptar, Pegar, Restablecer, Cancelar, Ayuda), comunes a cualquier procedimiento de SPSS. En la parte inferior figuran tres botones específicos del procedimiento. En el caso del "Análisis de correspondencias" estos botones son Modelo, Estadísticos y Gráficos. Cada uno de ellos permite configurar cierto conjunto de características del método que se desea aplicar.

🔛 epa0-	199.sav [Conjunto_de_d	atos1] - Edit	or de datos	SPSS												- # X
Co II	Dictori ver Datos Marsion				aua 1											
1 : ccaa	9 11 9 9 10 19	#9 1 ≇ ⊞		* 🛞 🥨	•									Visible:	3 de 3 variabl	es
	ccaa	situacio	peso	var	var	var	var	var	var	var	var	var	var	var	var	var
1	Andalucía 🗸	ocupado	2107300	1.441												
2	Aragón	ocupado	446500							1						
3	Asturias	ocupado	319500													
4	Apálicie de	correction	opeiae								1					
5	Analisis de	correspond	eneias				4]]				
6	eso 🖉		Fila:			Aceptar										
7) ccaa	(1 18)										1		
8	Castill	1 A T	D	lefinir rango		regar										
9			Calum			Restablecer										
10	Comunic	ſ	Colum	rid.		Cancelar	<u> </u>			_	_					
11				30(13)			<u> </u>				_					
12	-			lefinir rango		Ayuda				_						
13	Comur						-		-	-	-	6 d		0 0		
14	Ke O. F						<u> </u>			-						
10	U. F		Modelo	Estadístico	os Gráfic	os	-			-	_					
10	Paic Vacco	ocupada	A 600		-					-	0					
18	Couta y Molilla	ocupado	10000							-				-		
10	Andalucía	narado	711800													
20	Aragón	parado	89300				Anál	isis de corre	espondencias:	Modelo			J			
21	Asturias	parado	71200				Dime		alusián:							
22	Baleares	parado	24500				Dillik					Continuar				
23	Canarias	parado	97900				Me	dida de distano	sia			Cancelar				
24	Cantabria	parado	29400				• •	Chi-cuadrado				Aunda				
25	Castilla y León	parado	145700				0	Euclídea				Ayuua				
26	Castilla La Mancha	parado	97600													
27	Cataluña	parado	270300					todo de estanc	Janzacion							
28	Comunidad Valenciana	parado	228600					Se ellimitari las	s medias de mas y	Columnas						
29	Extremadura	parado	103600					Se eliminan las	s medias de filas							
30	Galicia	parado	182800					Se eliminan las	s medias de colum	nas						
31	Comunidad de Madrid	parado	275100					Se igualan los	totales de fila y se	eliminan las me	edias					
32	Región de Murcia	parado	60600					Se igualan los	totales de column	a y se eliminan l	las medias					
33	C. Foral de Navarra	parado	15900				Mé	todo de normal	ización							
34	La Rioja	parado	9500					Simétrico	Principal pr	orfila	O Personalizad	o: 0				
35	País Vasco	parado	123800				- 0	Principal	O Principal p	or columna						
36	Ceuta y Melilla	parado	14100						C minoparp	or columna						
37	Andalucía	inactivo	2898000										-			
38	Aragón	inactivo	510500													
39	Asturias	inactivo	526000								_					
40	Baleares	inactivo	276300													
41	Canarias	inactivo	607200													
42	Cantabria Castilla y Lafa	inactivo	240400		-											
43	Castilla La Mancha	inactivo	732100		-		-				-			-		
44	sta de datos & Vista de vari	ables /	7 32 100		1	(1							> ×
	The set state of the de van					SF	SS El proc	esador está pr	eparado					Ponde	erado	
						-										

Al pinchar en **"Modelo..."** se abre un nuevo cuadro de diálogo, en el que vamos a dejar las opciones por defecto, que son:

- 1. Dos dimensiones en la solución
- 2. Se centra restando las medias de fila y columna
- 3. El método de normalización es simétrico, que es el que permite la representación simultánea y se corresponde con los apuntes.

Archivo 8	Edición Ver Datos Transfor	mar Analizar	Gráficos Uti	idades Vent	ana ?											
1 : ccaa	H W 777 60 67	#99 " ⊞ ⊞	- -	\$ 💌 W	•									Visible: 3	de 3 variab	es
	ссаа	situacio	peso	var	var	var	var	var	var	var	var	var	var	var	var	var 👗
1	Andalucía	ocupado	2107300													
2	2 Aragón		446500													
3	Asturias	ocupado	319500													
4	🗖 Análisis de	correspond	lencias		12	×										
6	A Date		Fila:			Acentar										
7	♦ pcao	ſ	Ccaa	(1 18)		//ooptai										
8	Castill			lefinir rango		Pegar										
9						Restablecer										
10	Comunic		Colum	na:		Cancelar										
11		L		30(1.5)								_				
12	Comun			lefinir rango		Ayuda										
13	Do															
14	C F										-	-				
16			Modelo	Estadístico	Gráfic	xos										
17	Pais Vasco	ocupado	/84600	-	1	1	-								5	
18	Ceuta y Melilla	ocupado	40000													
19	Andalucía	parado	771800													
20	Aragón	parado	39300				Análisi	s de correst	ondencias:	Estadísticos						
21	Asturias	parado	71200													
22	Baleares	parado	24500				✓ Tab	a de correspon	dencias	Perfiles d	e fila 🛛 Co	intinuar				
23	Canarias	parado	97900				🗹 İnsp	ección de los p	untos de fila	Perfiles d	e col.	ancelar				
24	Cantabria	parado	29400				🗹 Insp	ección de los p	untos de colum	na						
20	Castilla La Mancha	parado	97600				Per	nutaciones de la	tabla de corre	spondencias		Ayuda				
20	Casula La Malicia Cataluña	parado	270300				-	Dimensión má	xima para las p	ermutaciones				-		
28	Comunidad Valenciana	parado	228600				Estadís	ticos de confiar	iza para							
29	Extremadura	parado	103600				Pun	tos de fila		Puntos de colu	mna					
30	Galicia	parado	182800								1					
31	Comunidad de Madrid	parado	275100													
32	Región de Murcia	parado	60600													
33	C. Foral de Navarra	parado	15900													
34	La Rioja	parado	9500								_					
35	Pais Vasco	parado	123800	1								_				
36	Ceuta y Melilla	parado	14100	-												
37	Anualucia	inactivo	510500		-	4				-			1			
39	Asturias	inactivo	526000		-											
40	Baleares	inactivo	276300													
41	Canarias	inactivo	607200													
42	Cantabria	inactivo	248400													
43	Castilla y León	inactivo	1114600													
44	Castilla La Mancha	inactivo	732100													~
4 > \Vi	sta de datos 🔏 Vista de varia	ables /				<								1.5.2.4		>
						SP	55 El procesad	or esta prepar	ado	ali al				Ponde	rado	

En el submenú de "Estadísticos..." solicitamos:

- 1. Tabla de correspondencias
- 2. Inspección de los puntos de fila
- 3. Inspección de los puntos de columna
- 4. Perfiles de fila
- 5. Perfiles de columna

		er un ui			• 💌 🗢														
: ccaa			1		Visible var var var var var var var var var var														
	CC	caa	situacio	peso	var	var	var	var	var	var	var	var	var	var	var	var	var		
1		Andalucia	ocupado	2107300		-			-										
3	3 Asturias		ocupado	319500															
4	F																		
5	-	Analisis de	correspond	encias			<u>×</u>		0										
6		peso		Fila:			Aceptar												
7				CC33	(1 18)		Recor												
8	Castill			D	efinir rango		regar												
9	Comunie			Colum	na:		Restablecer	-	-										
11	Comuni		1	∢ stua	cio(1 3)		Cancelar		-										
12					efinir rango		Avuda									1			
13	Comur																		
14	Re																		
15	C. F			Modelo	Estadístico	s Gráfic	105												
16																			
17	0	Pais Vasco	ocupado	/84600		T		-											
10	C	Andalusía	ocupado	771900					Análisis de co	rresponden	cias: Gráfic	DS							
20	Andalucia pai Aragón pai		parado	39300					Diagramas de dispersión Continuar										
21	Asturias parac		parado	71200					Diagrama de dispersión biespacial										
22	Baleares parad		parado	24500					Puntos de fila										
23		Canarias	parado	97900			Puntos de columna Ayuda												
24		Cantabria	parado	29400		Ancho de etiqueta de ID para 20													
25	C	astilla y León	parado	145700				diagramas de dispersión:											
26	Castilla	La Mancha	parado	97600		1			Gráficos de lír	ea									
21	Comunida	d Valanciana	parado	270300		1			Categorías	de fila transfor	madas								
20	Comunida	Extremadura	narado	103600		1			Categorías de columna transformadas										
30		Galicia	parado	182800		1			Ancho de etiqueta de ID para 20										
31	Comunid	ad de Madrid	parado	275100							20								
32	Regi	ión de Murcia	parado	60600					- Dimensiones o	el aráfico									
33	C. For	al de Navarra	parado	15900					Mostrarto	tas las dimensi									
34		La Rioja	parado	9500					Restores	a núman de d	mensiones								
35	~	País Vasco	parado	123800															
36	C	euta y Melilla	parado	14100					Menor dim	ensión:			_						
38		Andalucia	inactivo	2090000	-+			<u> </u>		un aidea	_			1					
39		Asturias	inactivo	526000	- 1-	-				maon:									
40		Baleares	inactivo	276300	1														
41		Canarias	inactivo	607200															
42		Cantabria	inactivo	248400															
43	C	astilla y León	inactivo	1114600															
44	Castilla	La Mancha	inactivo	732100	1														
» (Vi	sta de datos	A vista de varia	bies /		1		<	SS El process	idor está precor	do			_		Pondor	oher	>		
					1		58	oo ci procesa	iuur esta prepara	100					Ponder	auu			

En el submenú de **"Gráficos..."** solicitamos las opciones por defecto, que son:

- 1. El diagrama de dispersión biespacial
- 2. Que muestre todas las dimensiones de la solución

Una vez realizadas las selecciones que hemos indicado, se puede ejecutar el procedimiento. Se proporcionarán los resultados del análisis en la hoja de resultados de SPSS.

La mayor parte de los resultados proporcionados por SPSS se entienden a la luz de los contenidos expuestos en los apuntes, y coinciden con los que se obtienen con el programa R.

La principal diferencia consiste en que el SPSS aplica un enfoque basado en el análisis de correlación canónica, lo cual da lugar a un valor propio, que coincide con la raíz cuadrada del valor propio de los apuntes, que SPSS denomina Inercia (igual que en los apuntes).

Las puntuaciones en la dimensión se corresponden con las matrices A y B, en cada caso, con la diferencia de que mientras las puntuaciones de las matrices A y B están estandarizadas de modo que la suma ponderada de cuadrados sea el autovalor (en términos de inercia), las puntuaciones están estandarizadas para que dicha suma sea el autovalor (de la correlación canónica).

Las tablas de contribuciones consisten en proporciones de inercia por filas, por columnas o por componentes, donde la inercia bruta coincide con lo que figura en los apuntes.

Nótese que el enfoque de correlación canónica también es aplicado por el comando "corresp" de R, que por este motivo, da los mismos resultados que SPSS.