

La práctica del análisis de correspondencias

MICHAEL GREENACRE

Catedrático de Estadística en la Universidad Pompeu Fabra

Separata del capítulo 16

Tablas de múltiples entradas

Primera edición: julio 2008

ISBN: 978-84-96515-71-0

Traducción: Jordi Comas Angelet
Revisión: Carles M. Cuadras Avellana

© **Michael Greenacre, 2008**
© **de la edición en español, Fundación BBVA, 2008**

www.fbbva.es

Tablas de múltiples entradas

Hasta ahora, en los mapas de AC, hemos representado las frecuencias de concurrencia de dos variables, dispuestas en tablas de contingencia de dos entradas. En este capítulo vamos a considerar situaciones en las que tenemos datos de más de dos variables y veremos cómo explorar gráficamente este tipo de datos. Una posibilidad es acomodar las tablas de múltiples entradas como si fueran tablas de dos entradas y a continuación llevar a cabo el AC habitual.

Contenido

Introducción de una tercera variable en los datos sobre valoración de la salud	165
Interacción entre variables	166
Codificación interactiva	166
AC de la tabulación codificada interactivamente	166
Conjunto de datos 9: opiniones sobre el trabajo de las mujeres	168
Mapa del AC de países por respuestas	169
Codificación interactiva de género con país	170
Codificación interactiva de país, género y grupo de edad	171
Configuración en arco (herradura) del mapa	172
RESUMEN: Tablas de múltiples entradas	173

Volvamos a los datos sobre la autopercepción de la salud (conjunto de datos 3) con el que hemos trabajado en los capítulos 6 y 7. Se trata de una muestra, representativa de 6371 españoles que hemos clasificado según edad y autopercepción de su salud (imagen 6.1). La encuesta incluye otras variables como género, educación, región de residencia, etc. Como ejemplo de cómo introducir una tercera variable en AC, vamos a trabajar con la variable más sencilla, el género, una variable que sólo consta de dos categorías. Con esta nueva variable, podemos construir dos tablas de contingencia adicionales: género por grupo de edad y género por salud. La primera tabla puede ser interesante desde un punto de vista demográfico, sin embargo, la segunda es más relevante con relación al tema que estamos tratando (imagen 16.1). Para ver la estructura que presen-

Introducción de una tercera variable en los datos sobre valoración de la salud

Imagen 16.1:

Cruce de género por autopercepción de la salud, que muestra los perfiles de las filas como porcentajes

GÉNERO	<i>Muy buena</i>	<i>Buena</i>	<i>Regular</i>	<i>Mala</i>	<i>Muy mala</i>	<i>Suma</i>
Hombre	448	1789	636	177	39	3089
%	14,5	57,9	20,6	5,7	1,3	
Mujer	369	1753	859	237	64	3282
%	11,2	53,4	26,2	7,2	2,0	
Suma	817	3542	1495	414	103	6371
%	12,8	55,6	23,5	6,5	1,6	

Fuente de datos: Encuesta Nacional de la Salud, 1997.

tan los datos, de la primera tabla, no hay necesidad de llevar a cabo un AC; es una tabla de 2×5 , de una sola dimensión, en la que podemos expresar todos los resultados en porcentajes. Así, vemos que, en general, los hombres tienen una opinión más optimista sobre su salud que las mujeres. En las categorías *muy buena* y *buena* los porcentajes de los hombres son mayores que los de las mujeres, mientras que las mujeres presentan mayores porcentajes en las categorías *regular*, *mala* y *muy mala*.

Interacción entre variables

Anteriormente, en el capítulo 6, vimos que la autopercepción de la salud empeoraba con la edad. En la tabla de la imagen 16.1 podemos ver el efecto del género. Vemos que en promedio los hombres son más optimistas sobre su salud que las mujeres. Queremos saber si el efecto del género se mantiene en todos los grupos de edad, o, si por el contrario, va variando. Puede ocurrir que para un determinado grupo de edad este efecto sea mayor, o incluso puede ocurrir que se dé el efecto contrario, es decir, existe una *interacción*; en este caso una interacción entre edad y género. La ausencia de interacción significaría que en todos los grupos de edad existe la misma disimilitud entre géneros.

Codificación interactiva

Para visualizar la posible interacción entre género y edad codificamos los datos de la siguiente manera: creamos una nueva variable a partir de todas las combinaciones posibles entre género y edad. En este caso tenemos dos géneros y siete grupos de edad, por tanto tenemos $2 \times 7 = 14$ combinaciones posibles: es lo que llamamos *codificación interactiva*. A continuación, cruzamos la variable codificada interactivamente con las categorías de salud, para, de esta manera, construir la tabla de contingencia que mostramos en la imagen 16.2.

AC de la tabulación codificada interactivamente

A pesar de que, como vimos en el capítulo 6, estos datos son más bien unidimensionales, en la imagen 16.3 mostramos el mapa simétrico bidimensional de la tabla de la imagen 16.2. En este mapa, los pares de puntos muestran las diferencias hombre-mujer de los distintos grupos de edad. Comparando los pares de

GÉNERO/EDAD	Muy buena	Buena	Regular	Mala	Muy mala	Suma
h16-24	145	402	84	5	3	639
h25-34	112	414	74	13	2	615
h35-44	80	331	82	24	4	521
h45-54	54	231	102	22	6	415
h55-64	30	219	119	53	12	433
h65-74	18	125	110	35	4	292
h75+	9	67	65	25	8	174
m16-24	98	387	83	13	3	584
m25-34	108	395	90	22	4	619
m35-44	67	327	99	17	4	514
m45-54	36	238	134	28	10	446
m55-64	23	195	187	53	18	476
m65-74	26	142	174	63	16	421
m75+	11	69	92	41	9	222

Imagen 16.2:
Cruce de la variable codificada interactivamente, género-edad, con la variable autopercepción de la salud (h = hombre, m = mujer, y siete grupos de edad como en la imagen 6.1). Hemos subdividido cada fila de la imagen 6.1 en dos filas, según su género

puntos de cada grupo de edad, vemos, de forma consistente, que el punto de las mujeres se halla a la izquierda del correspondiente de los hombres. Ello ilustra el efecto que observamos en la tabla de la imagen 16.1, es decir, que en general, las mujeres son menos optimistas sobre su salud que los hombres. No observamos el fenómeno contrario en ningún grupo de edad, sin embargo, podemos apreciar algunas diferencias en las distancias hombre-mujer. Así, en las edades más tempranas, hasta el grupo de edad 35-44, las distancias hombre-mujer son relativamente pequeñas. En cambio, en el grupo de edad 45-54, podemos observar cambios importantes en la autopercepción de la salud (capítulo 6), que van acompañadas de una mayor diferencia entre hombres y mujeres. Este cambio se mantiene para los grupos de más edad. Vemos que las mujeres del grupo de edad 55-64, incluso son más pesimistas que los hombres del grupo de edad, 65-74 años. De forma similar, las mujeres del grupo de edad 65-74, son más pesimistas que los hombres del grupo de más edad, 75+. Esta diferencia cambiante entre hombres y mujeres de los distintos grupos de edad constituye una evidencia de que existe una interacción género-edad en cuanto a la autopercepción de la salud.

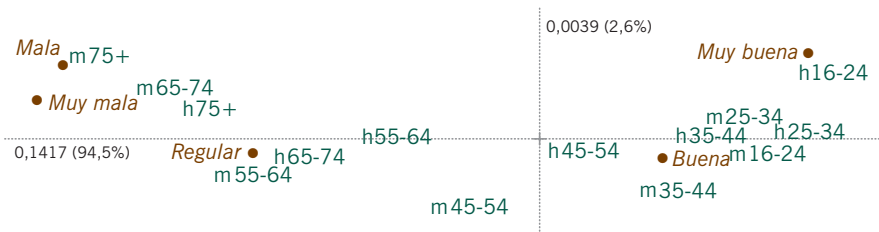


Imagen 16.3:
Mapa simétrico del AC correspondiente al cruce de género por edad, variable codificada interactivamente, con categoría de salud

Imagen 16.4:
Frecuencias de respuesta a la pregunta sobre el trabajo de las mujeres que tienen niños en edad escolar en casa en 24 países

PAÍSES	<i>T</i>	<i>t</i>	<i>C</i>	<i>?</i>	<i>Suma</i>	
AUS	Australia	256	1156	176	191	1779
AO	Alemania Occidental	101	1394	581	248	2324
AE	Alemania del Este	278	691	62	66	1097
GB	Gran Bretaña	161	646	70	107	984
IN	Irlanda del Norte	126	394	75	52	647
EU	Estados Unidos	482	686	107	172	1447
A	Austria	84	632	202	59	977
H	Hungría	285	736	447	32	1500
I	Italia	171	670	167	10	1018
IR	Irlanda	223	424	209	82	938
HO	Países Bajos	539	1205	143	81	1968
N	Noruega	487	1242	205	153	2087
S	Suecia	295	833	39	105	1272
CH	Checoslovaquia	228	585	198	13	1024
ES	Eslovenia	341	428	222	41	1032
P	Polonia	431	425	589	152	1597
B	Bulgaria	270	427	335	94	1126
RUS	Rusia	175	1154	550	119	1998
NZ	Nueva Zelanda	120	754	72	101	1047
CAN	Canadá	566	497	108	269	1440
F	Filipinas	243	448	484	25	1200
IS	Israel	468	664	92	63	1287
J	Japón	203	671	313	120	1307
E	España	738	1012	514	230	2494
<i>Suma</i>		<i>7271</i>	<i>17774</i>	<i>5960</i>	<i>2585</i>	<i>33590</i>
<i>%</i>		<i>21,6%</i>	<i>52,9%</i>	<i>17,7%</i>	<i>7,7%</i>	

Nota: Alemania del Este y Alemania Occidental todavía se tratan de forma separada, igual que Gran Bretaña e Irlanda del Norte). Expresamos los perfiles como porcentajes. Hemos utilizado las siguientes abreviaciones: *T*: tiempo completo, *t*: tiempo parcial, *C*: permanecer en casa, *?*: no sabe/no está seguro/no contesta.

Fuente: Encuesta ISSP sobre la familia y los cambios de rol de género, 1994.

Conjunto de datos 9:
opiniones sobre el trabajo de las mujeres

Como ilustración adicional de codificación interactiva, vamos a introducir un nuevo conjunto de datos que utilizaremos varias veces en este y en los siguientes capítulos. Son datos obtenidos de una encuesta de 1994 del Programa Internacional de Investigación (ISSP, *International Social Survey Programme*) sobre la familia y los cambios de rol de género. El total de la encuesta consta de 33.590 individuos. Se llevó a cabo en 24 países (en las encuestas del ISSP, la antigua Alemania del Este y Alemania Occidental se tratan de forma separada, igual que Gran Bretaña e Irlanda del Norte). Vamos a analizar la relación entre variables demográficas y las respuestas a la siguiente pregunta relacionada con la participación de las mujeres en el mercado de trabajo: «Una mujer con un niño en edad escolar en casa, ¿debe trabajar a tiempo completo, a tiempo parcial, o debe permanecer en casa?». Como en todas las encuestas de este tipo, existe una opción adicional de

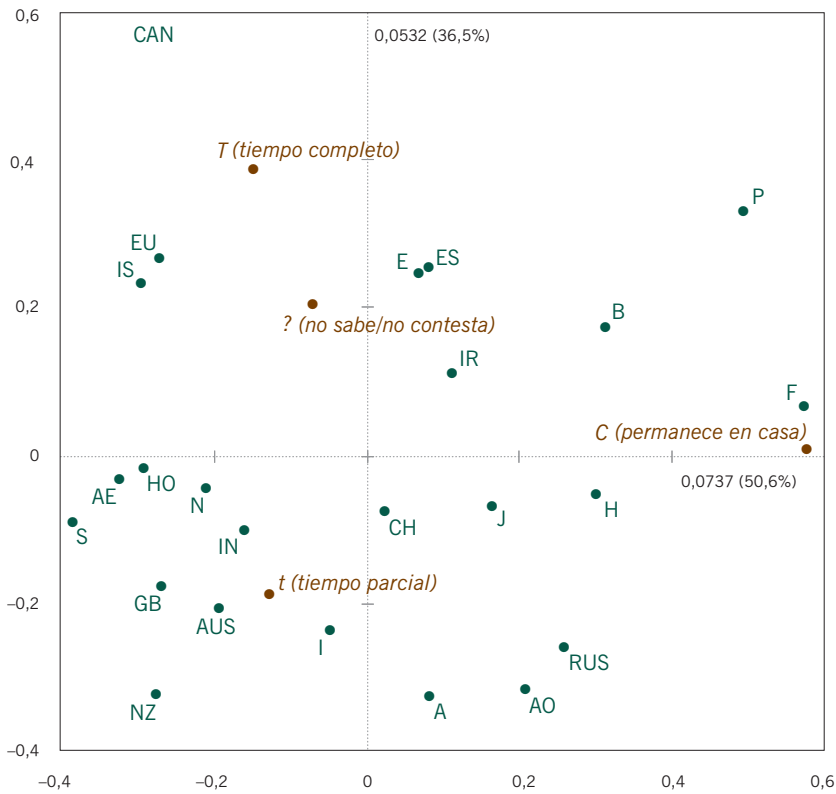


Imagen 16.5:
 Mapa simétrico del AC correspondiente a 24 países y a 4 categorías de respuesta (tabla de la imagen 16.4)

respuesta: «no está seguro/no sabe», a la que hemos añadido también algunas, pocas, encuestas sin respuesta (en el capítulo 21, veremos en detalle las no respuestas). Aparte de las respuestas a la pregunta anterior, tenemos datos sobre algunas variables demográficas de cada encuestado. Las tres siguientes tienen, para nosotros, un interés especial: género (dos categorías), edad (6 categorías) y país (24 categorías). En la tabla de la imagen 16.4 mostramos las frecuencias de respuesta de cada país.

En la imagen 16.5, mostramos el mapa de AC correspondiente a esta tabla (hemos cambiado el estilo de nuestros mapas de AC; además, al final del apéndice de cálculo (B), comentaremos algunas opciones de software para la creación de mapas). La interpretación de este mapa es bastante clara; de izquierda a derecha se produce un contraste entre las mujeres que trabajan (a la izquierda) y las que permanecen en casa (a la derecha), en vertical se produce un contraste entre las mujeres que trabajan a tiempo completo (arriba) *versus* las que trabajan a tiempo parcial (abajo). Con relación a este tema, los países más tradicionales son Filipinas y Polonia, mientras que países como Suecia, Alemania del Este, Israel, Nueva

[Mapa del AC de países por respuestas](#)

Imagen 16.6:

Frecuencias de respuesta a la pregunta sobre el trabajo de las mujeres cuando tienen un niño en edad escolar en casa, son los datos de los 24 países que aparecen en la imagen 16.4, que se han subdividido según el género del encuestado (cualquier pequeña diferencia en los subtotales de un país y los totales de la imagen 16.4 se deben a unos pocos valores perdidos sobre el género)

País	<i>T</i>	<i>t</i>	<i>C</i>	<i>?</i>	<i>Suma</i>
AUSh	117	596	114	82	909
AUSm	138	559	60	109	866
AOh	43	675	357	123	1198
AOm	58	719	224	125	1126
AEnh	146	316	29	37	528
AEm	132	375	33	29	569
...
...
ISh	220	275	57	29	581
ISm	247	387	35	34	703
Jh	85	279	171	57	592
Jm	118	392	142	63	715
Eh	347	445	294	111	1197
Em	390	566	218	118	1292

Zelanda, Gran Bretaña, y Canadá son los más liberales. A la izquierda, la dirección vertical separa países como Canadá, más favorable a que la mujer trabaje a tiempo completo, de países como, por ejemplo, Nueva Zelanda, más favorable al empleo a tiempo parcial. Recordemos que el origen del mapa representa el perfil medio mostrado en la última fila de la tabla de la imagen 16.4. Por tanto, todos los países situados a la izquierda son más liberales que la media. Si dos países se hallan en la misma posición en el eje horizontal (por ejemplo, Estados Unidos y Gran Bretaña), el país más positivo con relación al eje vertical estará más a favor que el valor de la media, de que las mujeres trabajen a tiempo completo.

Codificación interactiva de género con país

Con el objetivo de visualizar las diferencias hombre-mujer en los distintos países, codificamos género con país. En la tabla de la imagen 16.6 mostramos las primeras y las últimas filas de la tabla de contingencia de 48×4 . En el mapa de la imagen 16.7, podemos ver que no han cambiado mucho las posiciones de las categorías de respuesta. Sin embargo, es interesante comparar los pares de puntos de cada país. En casi todos los casos, el punto correspondiente a las mujeres queda más a la izquierda que el de su homólogo masculino (Bulgaria es la única excepción). Dentro de un mismo país, las respuestas son sorprendentemente homogéneas en comparación con las grandes diferencias entre países. Los países en los que existe la mayor distancia entre las opiniones de hombres y mujeres se hallan principalmente en el lado conservador del mapa: es el caso de Filipinas, Japón, Irlanda del Norte, Alemania Occidental y España. Sin embargo, en el lado izquierdo del mapa, Australia muestra una de las mayores diferencias hombre-mujer. En este análisis, la inercia es mayor que en el mapa de la imagen 16.5 debido a que la división de las muestras por géneros añade inercia. En realidad, la inercia total de este análisis es de 0,01546, mientras que la inercia del análisis anterior era

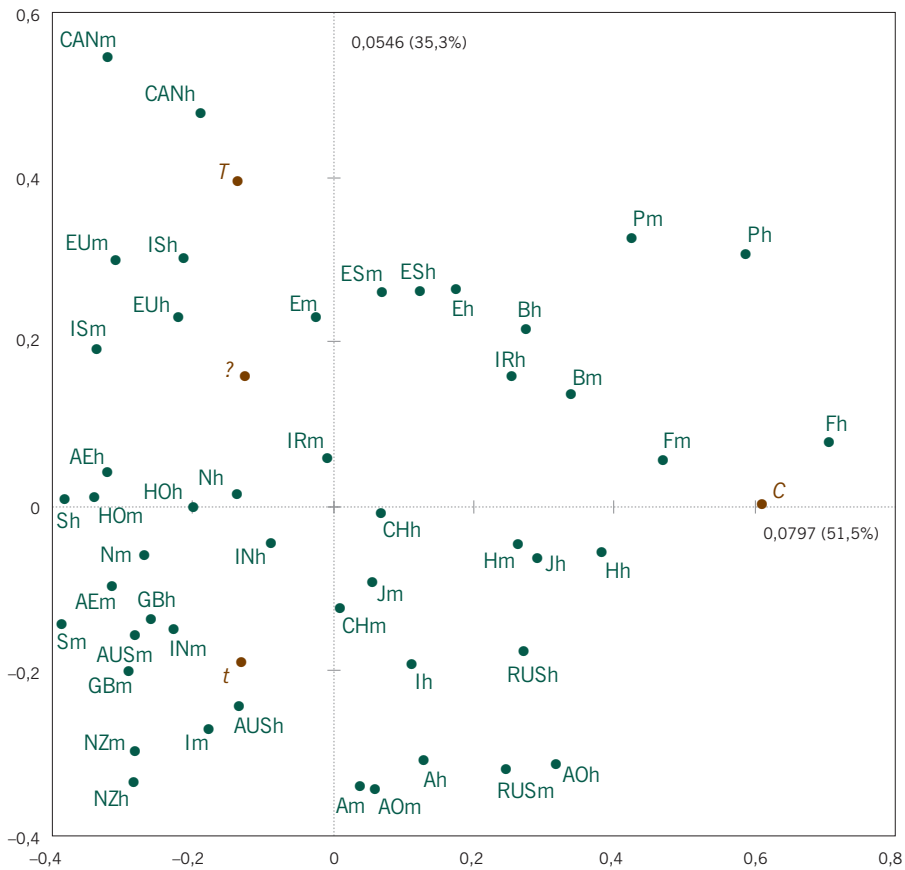


Imagen 16.7: Mapa simétrico de datos codificados interactivamente (imagen 16.6). Los puntos correspondientes a los hombres se halla, de forma consistente, más a la derecha de los de sus contrapartes femeninas, con la sola excepción de Bulgaria (B), país en el que las mujeres son más conservadoras que los hombres

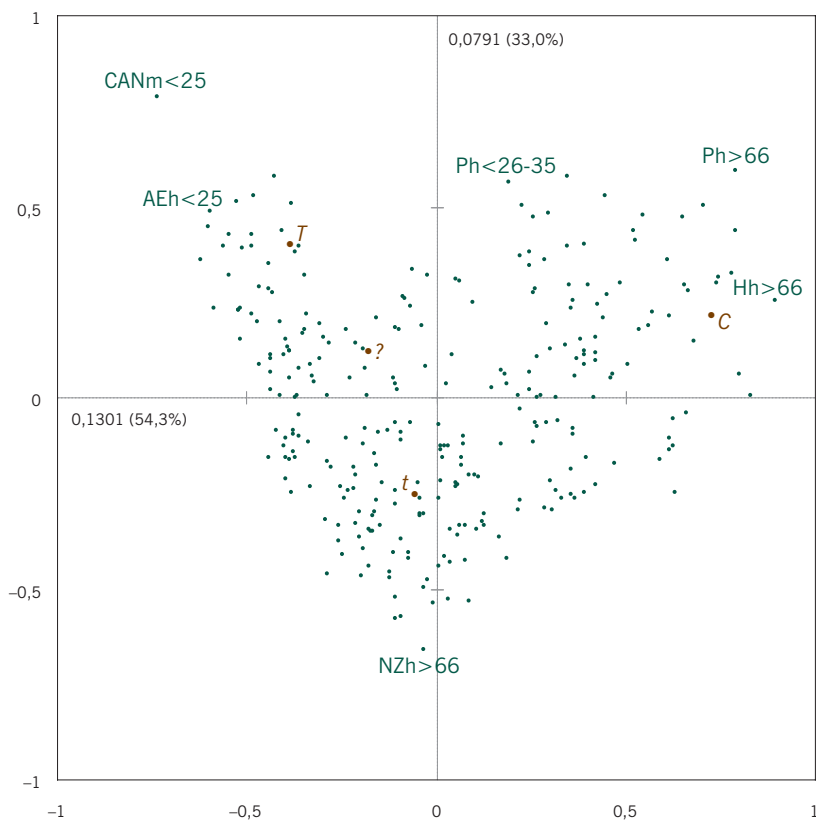
de 0,01456. Por tanto, en el presente análisis, la inercia atribuible a la diferencia entre géneros es de 0,00090, es decir el 5,8% de la inercia. Como podemos apreciar en el mapa, la mayor diferencia se produce entre países, y no entre géneros.

Como las muestras son grandes, podemos seguir dividiéndolas sin problemas. Vamos a hacerlo por la edad, es decir, subdividimos cada grupo país-género en los seis grupos de edad siguientes: hasta 25 años, 26-35, 36-45, 46-55, 56-65, y 66 o más años. Ahora vamos a codificar interactivamente las tres variables en una nueva variable, con $24 \times 2 \times 6 = 288$ categorías en total. En la imagen 16.8 mostramos el mapa de AC de la tabla resultante de 288×4 . De nuevo vemos que el mapa no cambia demasiado con relación a la posición de las respuestas. Dado que es imposible etiquetar las 228 filas, las hemos representado mediante puntos. Solamente hemos etiquetado algunas observaciones atípicas. Por ejemplo, el grupo más liberal que se halla, lejos, arriba a la izquierda. Es el grupo más joven de mujeres canadienses, el grupo de hasta 25 años. En esta submuestra de 168 mujeres, 101

[Codificación interactiva de país, género y grupo de edad](#)

Imagen 16.8:

Mapa simétrico del AC correspondiente a una codificación de tres entradas. Vemos que los puntos del mapa, que representan los grupos país-género-edad, forman un patrón curvado que surge con frecuencia en los mapas de AC cuando los perfiles forman un gradiente de un extremo (T) a otro (C).



(60,1%) están a favor de que las mujeres con un niño en edad escolar en casa trabajen a tiempo completo, 32 (19,0%) a favor de que lo hagan a tiempo parcial, 3 (1,8%) dicen que las mujeres deben permanecer en casa, y 32 (19,0%) o no saben o no responden (en el capítulo 21 veremos que en la muestra de Canadá hay muchos «no saben»). El grupo de hombres más liberal es el grupo de hombres más joven de Alemania del Este. En el otro extremo, a la derecha, tenemos el grupo de más edad de hombres húngaros y polacos; así, de los 76 hombres polacos de 66 o más años, 16 (21,1%) responden que a tiempo completo, 13 (17,1%) que a tiempo parcial, 41 (53,9%) dicen que las mujeres deben permanecer en casa, las preguntas sin respuesta son 6 (7,9%). En la parte inferior tenemos el grupo de más edad de hombres neozelandeses —son los que están más a favor del trabajo de las mujeres a tiempo parcial.

Configuración en arco (herradura) del mapa

Finalmente, fijémonos en que la nube de puntos del mapa de la imagen 16.8 forma una curva, es lo que en AC llamamos *efecto arco* o *de herradura*. Este fenómeno, habitual, se debe a que el espacio de perfiles es un símplex, en el caso que nos

ocupa un tetraedro de tres dimensiones, ya que tenemos cuatro columnas. Cualquier gradiente de cambio de una esquina extrema del espacio T (trabajo a tiempo completo) a la otra C (permanecer en casa) seguirá, en este espacio limitado, más una trayectoria curva que recta. Los puntos que se hallan en el interior del arco, como por ejemplo el grupo de hombres polacos de 26-35 años, tienden a estar polarizados en el sentido de que presentan valores elevados en las dos respuestas extremas. Así, de los 141 individuos de este grupo, 45 (31,9%), responden que a tiempo completo, 31 (22,0%) tiempo parcial, 45 (31,9%) permanecer en casa y 20 no responden (14,2%) —en este grupo se observan respuestas por encima de la media en los dos extremos.

1. Podemos *codificar interactivamente* dos o más variables categóricas en una nueva variable formada por todas las combinaciones de categorías. Por ejemplo, podemos codificar dos variables con J_1 y J_2 categorías en una nueva variable con $J_1 J_2$ categorías.
2. Cruzamos la variable codificada interactivamente con otra variable y aplicamos el AC. El mapa resultante muestra la estructura de la interacción entre las variables que hemos codificado interactivamente.
3. En general, la codificación interactiva de tablas de múltiples entradas, no va más allá de tres variables, ya que aumenta mucho el número de categorías y por tanto la complejidad del mapa. El nivel de interacción que podemos investigar depende del tamaño de la muestra, ya que el codificado interactivo fragmenta la muestra en submuestras que no deberían ser demasiado pequeñas.

RESUMEN:
Tablas de múltiples
entradas
