

MODELOS CAUSALES ENTRE ATRIBUTOS: UNA APLICACIÓN DEL ANÁLISIS DE CORRESPONDENCIAS SIMPLES

JOSÉ LUIS NARVAIZA SOLIS

Director General del EUSTAT

Palabras clave: Análisis de correspondencias, modelos causales, análisis de actividad de la población.
Nº de clasificación JEL: C1, C42

0. INTRODUCCIÓN

Cuando se trata de explicar la utilidad y funcionamiento de cualquier técnica estadística hay que empezar habitualmente por resaltar que esa técnica sirve para medir, o para explicar, el comportamiento *variable* de un fenómeno. En efecto, lo constante, lo invariable, no suscita ninguna curiosidad; cuando algo cambia, ya sea en el tiempo, o entre individuos, cuando no se da en todos en la misma cantidad o de la misma manera, es cuando provoca las preguntas ¿cuánto?, ¿cómo?, y, sobre todo, ¿por qué? Las dos primeras se responden con las técnicas de estadística descriptiva univariante, pero la última, y más ambiciosa, exige aceptar la existencia de causas que expliquen esos cambios, y, por lo tanto, la necesidad de técnicas multivariantes.

Si lo que tratamos de estudiar son fenómenos cuantificables, las diferentes alternativas están bien estudiadas: Un modelo causal con variables también cuantificables y directamente observables nos lleva a los modelos econométricos, con todos los problemas que plantean: analizar bien el carácter verdaderamente exógeno de las variables causa, tener en cuenta la falta o no de independencia entre estas variables exógenas, etc. Cuando los fenómenos a analizar son varios que se producen de forma simultánea, y sólo se pretende eliminar las redundancias, o rastrear,

a partir de esas redundancias, la existencia de fenómenos subyacentes, entonces las técnicas a utilizar son las diferentes variantes del Análisis Factorial.

Menos estudiadas están las alternativas que se pueden plantear cuando lo que analizamos es algo cualitativo, cuando lo que cambia es la *forma* de presentarse un fenómeno, y no la cantidad de fenómeno que se produce.

En este caso queda siempre perfectamente claro que la idea de relación o de asociación sólo significa que poseer una determinada modalidad de un atributo hace aumentar la probabilidad de poseer una determinada modalidad de otro atributo, aumentos de probabilidad que dependen de la intensidad de la relación y que no son necesariamente iguales para todas las modalidades.

La comprobación de la existencia de una asociación significativamente intensa entre dos atributos se había realizado, hasta la aparición del Análisis de Correspondencias, mediante las Tablas de Contingencia y el coeficiente χ^2 . Esta comprobación se hace de forma negativa: la falta de asociación no se puede mantener cuando la acumulación de frecuencias en las casillas no puede considerarse debida al azar. Cuando más radicalmente diferente de lo totalmente aleatorio, más indicio de una

relación. Todo, como se ve, muy impreciso, con afirmaciones globales que abarcan a todo el atributo cuando en realidad pueden deberse a unas pocas de sus modalidades.

El Análisis de Correspondencias Simples trasciende esta mera comprobación y profundiza el análisis mostrando los entresijos de la relación y las modalidades a las que afecta, aunque limitado a dos atributos. Cuando el número de atributos es mayor que dos, el enfoque usual es estudiar las interrelaciones de todos con todos por medio del Análisis de Correspondencias Múltiple.

En general, la postura ante la existencia de relaciones entre atributos suele ser extremadamente prudente y muy pocas veces se aventuran hipótesis de causalidad. Más bien se suele tender a un propósito mucho más modesto: poner de manifiesto las relaciones entre modalidades, en definitiva las redundancias, y definir colectivos de individuos caracterizados por esas modalidades redundantes.

Sin embargo, hay ocasiones en que interesa explicar el comportamiento de un atributo por la influencia de otros, de la misma forma que en un modelo econométrico. Habría un atributo endógeno y otros exógenos, y mayor o menor capacidad explicativa o, lo que es lo mismo, mayor o menor capacidad de predicción de la modalidad que tendrá un individuo del atributo endógeno, una vez que sabemos las modalidades que tiene de los exógenos.

Continuando con el paralelismo con econometría, los atributos exógenos pueden tener diferente grado de influencia y tener problemas de relaciones entre ellos (lo que en econometría se llama multicolinealidad).

En un caso como éste tendremos varios atributos, pero entre ellos se establece una diferencia radical: uno de ellos es el resultado y los demás la causa; fundamentalmente interesa la relación existente entre este segundo grupo y el atributo a explicar, mientras que las relaciones que pueda haber entre los atributos que explican interesan, no por sí mismas, sino por el efecto distorsionador que puedan producir. En esta situación no tiene mucho sentido utilizar una técnica

como el Análisis de Correspondencias Múltiple, que analiza de forma igual las interrelaciones de todos con todos a través del cuadro de Burt.

La utilización un poco especial del Análisis de Correspondencias Simples, a pesar de tratar con más de dos atributos, puede ser una manera de resolver el problema. Vamos a exponerlo utilizando un ejemplo práctico que permite desarrollar claramente todos los problemas que se pueden presentar.

1. ANÁLISIS DE LA SITUACIÓN DE LA POBLACIÓN RESPECTO A LA ACTIVIDAD: RESULTADOS DE LA ENCUESTA P.R.A.

Esta encuesta que realiza el Instituto Vasco de Estadística periódicamente mediante un panel de familias, clasifica a los individuos según su situación respecto a la actividad, pero no se limita a la clasificación tradicional de la O.I.T., activos, —que se dividen en ocupados y parados— e inactivos, sino que subdivide a estos grandes colectivos poniendo al descubierto subcolectivos fronterizos entre ocupación y paro, y distinguiendo entre diferentes tipos de inactividad. Las modalidades de este atributo que se van a tener en cuenta son: *Sobreocupados* (que han trabajado 45 o más horas la semana de referencia), *Ocupados Estrictos* (que trabajan entre 15 y 45 horas), *Subocupados* (menos de 15 horas en un empleo continuado), *Ocupados Marginales* (actividades laborales marginales tales como ayuda familiar, empleos estacionales, ocasionales, etc.), *Cuasiparados* (similar a los anteriores pero que declaran que buscan un empleo). Dentro de los que realizan actividades no laborales (tareas del hogar, estudios, servicio militar), se pueden distinguir entre los que se declaran disponibles para incorporarse a esa actividad laboral y están buscando empleo, y entonces se consideran *Parados con actividad no laboral*, los que buscan empleo pero no están disponibles, y son los que llamaremos *Actividad no laboral buscando empleo*, y, por fin, los que no buscan empleo y los consideraremos *Ocupados no laborales*. Por último, la población *Parada estricta* que no

realiza actualmente ninguna actividad, ni siquiera no laboral, pero busca activamente un empleo y podría ocuparlo, y los *Jubilados estrictos y otros*, que son los que por su edad o situación física no realizan ninguna actividad.

Los ocupados según la O.I.T. serían la suma de sobreocupados, ocupados, subocupados y ocupados marginales. Los parados según la O.I.T. serían los cuasiparados, los parados con actividad laboral y los parados estrictos, y, por último, los inactivos según la O.I.T. son los que tienen actividad no laboral y buscan empleo, los ocupados con actividad no laboral y los jubilados estrictos y otros.

La pregunta es si hay circunstancias personales, características sociológicas que, de forma general, provoquen un aumento de la probabilidad de estar en cada una de esas modalidades.

El Análisis de Correspondencias Simples parte, precisamente, de estas diferencias en la probabilidad (o en la frecuencia relativa). Su enfoque consiste en comparar lo que sucede en el colectivo completo con lo que sucede en los subcolectivos parciales que se definen a partir de un atributo determinado.

Por ejemplo, en el colectivo completo las dos modalidades del sexo se reparten, como es habitual, aproximadamente al 50 %. Esta frecuencia es la que se llama frecuencia marginal. Si el sexo no tuviera ninguna relación con la situación de actividad, ese mismo 50 % se debería dar entre los ocupados estrictos, entre los que están ocupados en actividades no laborales, etc., etc. Si nos encontramos que en los ocupados estrictos hay un 70 % de hombres y en los ocupados no laborales un 85 % de mujeres, parece que esas diferencias respecto a aquel 50 % no se pueden deber al azar, y que indican que hay una relación entre sexo y tipo de actividad, en el sentido de que el sexo que tiene una persona hace más probable que se encuentre en una situación que en otra. Estas frecuencias que se dan dentro de cada subcolectivo parcial se llaman frecuencias condicionales.

En este ejemplo, cada tipo de actividad

se encuentra definido por los dos valores de su frecuencia condicional, las dos modalidades del sexo, por lo que el conjunto de todas esas modalidades de la actividad forma una nube de puntos alrededor del centro de gravedad que es la frecuencia marginal: lo que sucede en el conjunto del colectivo. Si hay una relación, estos puntos se encuentran alejados del centro de gravedad, se producen unas distancias que ponderadas por el «peso» de esa modalidad —la importancia relativa que tiene dentro del conjunto del colectivo—, provoca lo que se llama una inercia. El Análisis de Correspondencias Simples trata de explicar el máximo de la inercia conjunta con el menor número de dimensiones que sea posible.

Naturalmente, también dentro de cada sexo la forma de repartirse los distintos tipos de actividad es diferente de lo que sucede en el colectivo total, de forma que también se puede considerar que hay dos puntos en un espacio definido por los diferentes tipos de actividad, con su distancia y su inercia. La técnica analiza estos dos espacios de forma simultánea y, mediante una serie de transformaciones, puede proyectarlos sobre un único espacio definido por lo que llama los «factores».

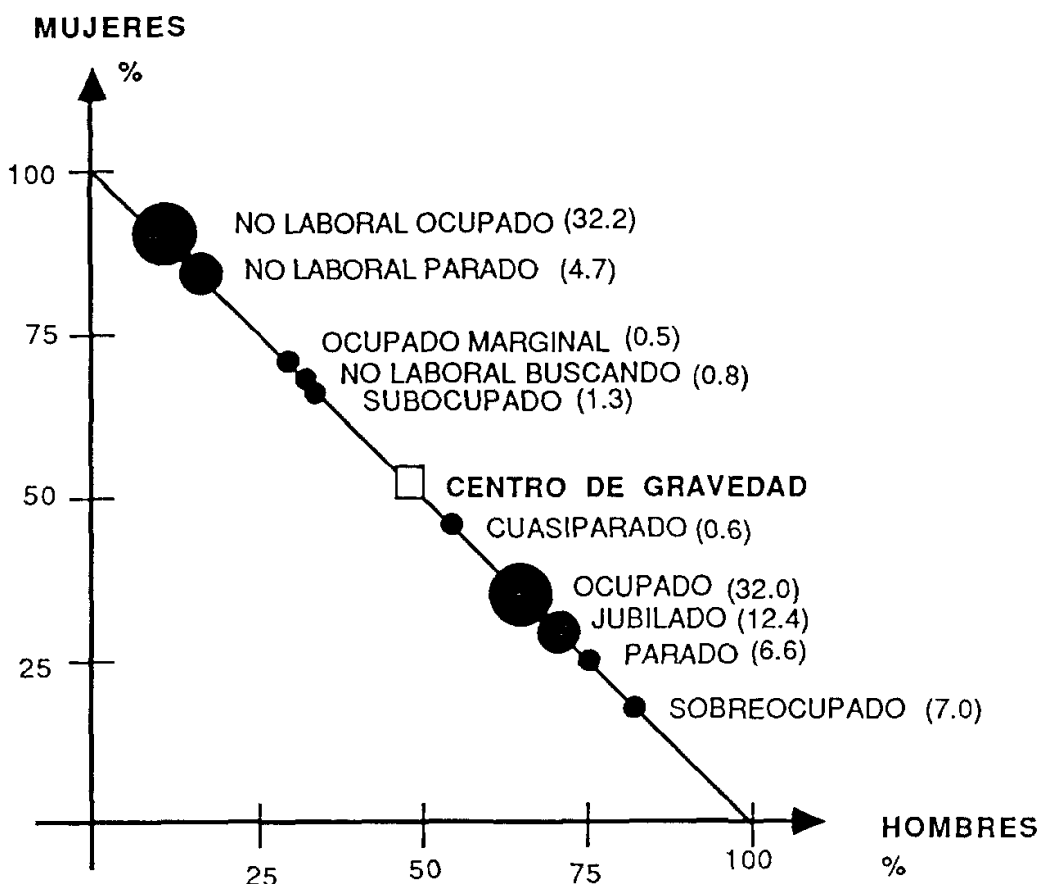
El gráfico n.º1 ilustra perfectamente la inercia de los tipos de actividad alrededor del centro de gravedad, y el eje que explica totalmente esa dispersión. Los datos que se han utilizado, al igual que en el resto del artículo, provienen de la Encuesta de Población en Relación con la Actividad (PRA) del primer trimestre de 1988, que publica el Instituto Vasco de Estadística (EUSTAT).

Este es, a grandes rasgos, el punto de partida básico de Correspondencias Simples, cuando analiza la relación entre dos atributos.

1.1. Problemas de la utilización de vanos atributos exógenos

Pero, en nuestro caso, está claro que el sexo no es la única característica que puede modificar la probabilidad de estar en cada modalidad de actividad, sino que puede haber otras. Supongamos que hacemos la hipótesis de que pueden

Gráfico n.º 1. Relación actividad-sexo



influir también la edad, el lugar de residencia, el estado civil, el nivel de instrucción y la situación familiar.

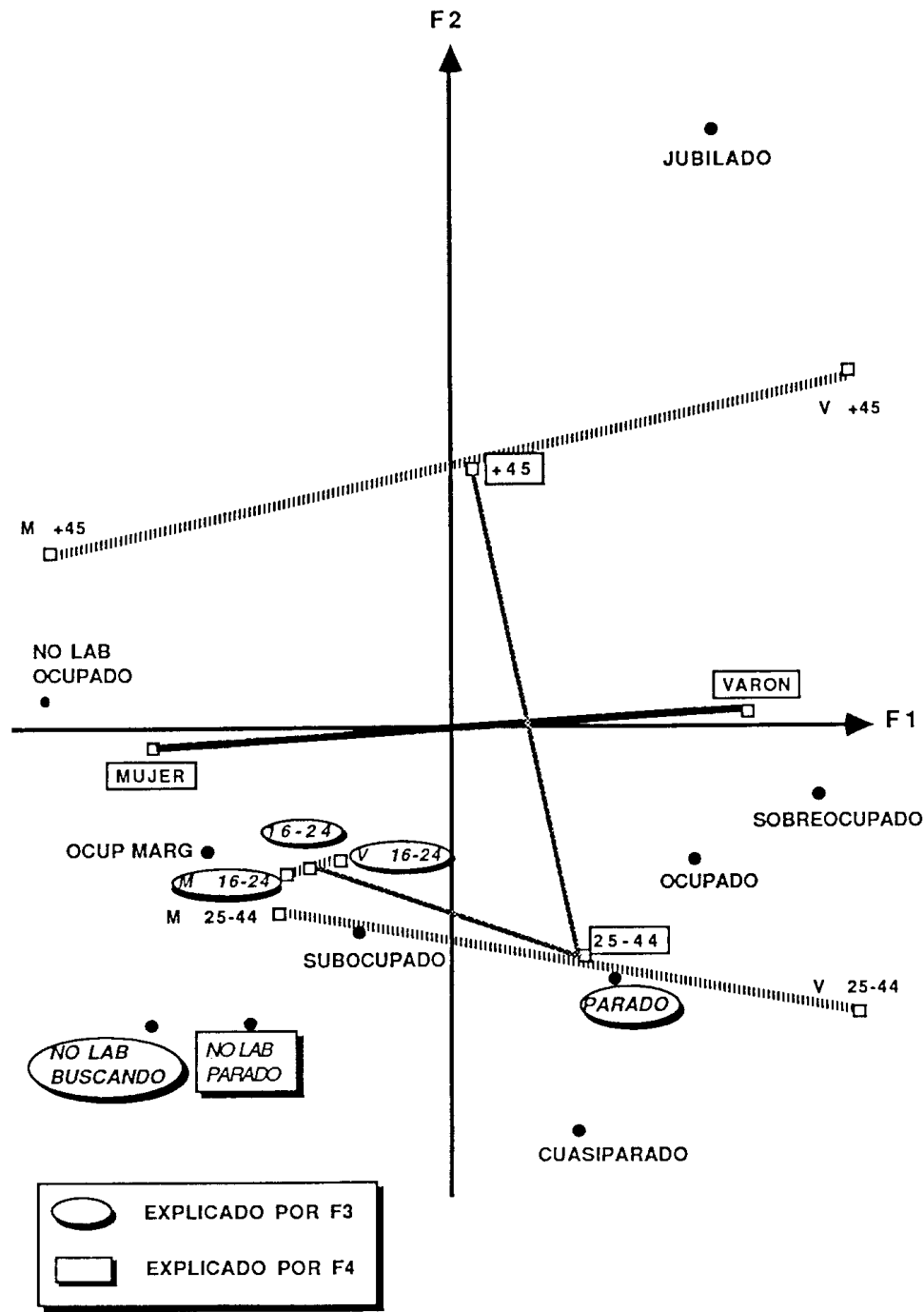
Al intentar comprobar esta hipótesis volvemos a encontrarnos con el paralelismo con econometría: habrá que analizar la influencia conjunta en lugar de la influencia aislada de cada atributo exógeno, pero, al igual que en econometría con el problema de la multicolinealidad, aquí nos encontramos con la duda de si las posibles interrelaciones entre los atributos exógenos perturbarán ese análisis de la influencia conjunta, y si lo harán de forma semejante. Otro interrogante es el procedimiento a utilizar para incluir la influencia conjunta de estos fenómenos exógenos.

En efecto, si tomamos dos atributos independientes entre ellos como son la

edad y el sexo, se produce una situación muy diferente según que se consideren separados o bien que se construya un atributo conjunto que combine las modalidades de los dos. En el gráfico n.º 2 tenemos el resultado en los dos casos. Evidentemente, el grupo global de varones es el centro de gravedad de los tres grupos de varones diferenciados por su edad, y lo mismo el grupo global de mujeres, o los grupos de edad.

Como se puede apreciar en la figura, estos grupos generados por cada atributo por separado son heterogéneos en el otro atributo y pueden dar una idea engañosa si esto no se tiene en cuenta. Es decir, no se puede hablar del comportamiento global de las mujeres en cuanto a la actividad si no se diferencia por edades, ni tampoco se puede hablar en general de cómo es la relación con la

Gráfico n.º 2. Actividad, edad y sexo



actividad de los mayores de 45 años si no se diferencia por el sexo.

Parece evidente, por tanto, que hay casos claros en los que es necesario construir un atributo que sea la combinación de las modalidades de los atributos que interesa introducir en el modelo, pero ¿hay que hacerlo siempre? La respuesta es que puede ser contraproducente porque cuando haya atributos que están relacionados entre ellos, nos encontraremos que bastantes de las combinaciones posibles se producen con una frecuencia muy escasa, con los grandes problemas de distorsión que esto produce en el Análisis de Correspondencias.

Hay, por lo tanto, una seria limitación a la construcción de atributos combinados: el gran número de modalidades que resultan y el alto riesgo de que muchas de ellas estén prácticamente vacías, riesgo que se agrava si hay atributos con altas relaciones entre ellos, y más todavía si se utilizan estimaciones muestrales que puedan tener altísimos errores de muestreo para colectivos tan pequeños.

Pero este problema que pueden crear los atributos que son redundantes, desaparece si no se combinan y sólo se yuxtaponen en el análisis, con la ventaja adicional de que podemos estudiar si esa relación se produce en todas las modalidades o sólo en algunas, y, además, nos evitamos el riesgo de eliminar a priori variables importantes.

En definitiva, parece que puede ser conveniente combinar los atributos influyentes e independientes entre ellos con el fin de no mezclar colectivos diferenciados mientras que los atributos que previsiblemente estarán bastante relacionados entre ellos, es preferible yuxtaponerlos, para no multiplicar el número de modalidades prácticamente vacías.

1.2. La actividad relacionada con edad, sexo, estado civil, nivel de instrucción y situación familiar

Esto es lo que se ha hecho en nuestro ejemplo con la P.R.A. en el que la edad y el sexo se han combinado con el estado civil, con el nivel de instrucción y con la situación familiar, pero estos tres

últimos atributos no se han combinado entre ellos, sino que se han yuxtapuesto las tres tablas que relacionan la actividad con cada uno de estos tres grandes atributos. La tabla que vamos a analizar mediante Análisis de Correspondencias Simples está, por lo tanto, compuesta por tres tablas que tienen en común el atributo que define las columnas: la actividad, de tal manera que cada individuo interviene tres veces —según su estado civil, su nivel de instrucción y su situación familiar—.

¿Afecta esto a lo que tratamos de estudiar, las relaciones entre atributos? En realidad parece que no: las frecuencias condicionales de las filas no cambian al añadir otra vez el mismo colectivo clasificado por medio de otro atributo, por ejemplo, la forma en que los varones entre 16 y 24 años solteros se reparten en las distintas formas de actividad no cambia al añadir otras filas que dependen del nivel de instrucción. Lo único que sucede es que estamos superponiendo varias nubes de puntos, cada una definida por un atributo, alrededor del mismo centro de gravedad. Además, la proximidad de las filas, cada una perteneciente a un atributo distinto, nos indica la semejanza de comportamiento, por ejemplo, que los varones jóvenes casados tienen un comportamiento parecido a los varones jóvenes que son persona principal (por lo menos en cuanto a la actividad), aunque esto no nos debe llevar a la conclusión de que son los mismos individuos.

En cuanto a las frecuencias condicionales de las columnas, al añadir nuevos atributos cambia la situación pero solamente porque vamos añadiendo más elementos que definen a cada modalidad y que pueden acentuar o debilitar el parecido que podría tener con otra modalidad. Es decir, si con un atributo dos columnas tienen perfiles semejantes esto no cambia al añadir otro atributo nuevo, pero sí es cierto que añadimos más elementos que caracterizan a ambas columnas, y en las cuales se pueden parecer o no, modificándose, por lo tanto, la semejanza global.

En definitiva, no parece que la yuxtaposición de tablas que tienen en común el atributo que define las columnas sea un obstáculo para poder utilizar el Análisis de Correspondencias

Cuadro n.º 1. Columnas y filas activas

COLUMNAS ACTIVAS

	NOMBRE	EXPLICACION
1	SOBREOCU	SOBREOCUPADA
2	OCUP EST	OCUPADA Estricta
3	SUBOCUPA	SUBOCUPADA
4	OCUP MRG	OCUPADA MARGINAL
5	CUASIPAR	CUASIPARADA
6	PAR.N.LA	PARADA CON ACTIVIDAD NO LABORAL
7	N.LA.BUS	NO LABORAL BUSCANDO EMPLEO
8	N.LA.OCU	OCUPADA EN ACTIVIDADES NO LABORALES
9	PARO EST	PARADA Estricta
10	JUBILADA	JUBILADA Estricta Y OTRAS

FILAS ACTIVAS

	NOMBRE	EXPLICACION
1	V1 SOLT.	VARON 16-24 SOLTERO
2	V1 NO SO	VARON 16-24 NO SOLTERO
3	V1 EPRIM	VARON 16-24 ESTUDIOS PRIMARIOS
4	V1 ESECU	VARON 16-24 ESTUDIOS SECUNDARIOS
5	V1 EUNIV	VARON 16-24 ESTUDIOS UNIVERSITARIOS
6	V1 PPRIN	VARON 16-24 PERSONA PRINCIPAL
7	V1 CGRUP	VARON 16-24 CABEZA DE GRUPO
8	V1 OTROS	VARON 16-24 OTROS
9	V2 SOLT.	VARON 25-44 SOLTERO
10	V2 NO SO	VARON 25-44 NO SOLTERO
11	V2 EPRIM	VARON 25-44 ESTUDIOS PRIMARIOS
12	V2 ESECU	VARON 25-44 ESTUDIOS SECUNDARIOS
13	V2 EUNIV	VARON 25-44 ESTUDIOS UNIVERSITARIOS
14	V2 PPRIN	VARON 25-44 PERSONA PRINCIPAL
15	V2 CGRUP	VARON 25-44 CABEZA DE GRUPO
16	V2 OTROS	VARON 25-44 OTROS
17	V3 SOLT.	VARON 45- SOLTERO
18	V3 NO SO	VARON 45- NO SOLTERO
19	V3 EPRIM	VARON 45- ESTUDIOS PRIMARIOS
20	V3 ESECU	VARON 45- ESTUDIOS SECUNDARIOS
21	V3 EUNIV	VARON 45- ESTUDIOS UNIVERSITARIOS
22	V3 PPRIN	VARON 45- PERSONA PRINCIPAL
23	V3 CGRUP	VARON 45- CABEZA DE GRUPO
24	V3 OTROS	VARON 45- OTROS
25	M1 SOLT.	MUJER 16-24 SOLTERO
26	M1 NO SO	MUJER 16-24 NO SOLTERO
27	M1 EPRIM	MUJER 16-24 ESTUDIOS PRIMARIOS
28	M1 ESECU	MUJER 16-24 ESTUDIOS SECUNDARIOS
29	M1 EUNIV	MUJER 16-24 ESTUDIOS UNIVERSITARIOS
30	M1 PPRIN	MUJER 16-24 PERSONA PRINCIPAL
31	M1 CGRUP	MUJER 16-24 CABEZA DE GRUPO
32	M1 OTROS	MUJER 16-24 OTROS
33	M2 SOLT.	MUJER 25-44 SOLTERO
34	M2 NO SO	MUJER 25-44 NO SOLTERO
35	M2 EPRIM	MUJER 25-44 ESTUDIOS PRIMARIOS
36	M2 ESECU	MUJER 25-44 ESTUDIOS SECUNDARIOS
37	M2 EUNIV	MUJER 25-44 ESTUDIOS UNIVERSITARIOS
38	M2 PPRIN	MUJER 25-44 PERSONA PRINCIPAL
39	M2 CGRUP	MUJER 25-44 CABEZA DE GRUPO
40	M2 OTROS	MUJER 25-44 OTROS
41	M3 SOLT.	MUJER 45- SOLTERO
42	M3 NO SO	MUJER 45- NO SOLTERO
43	M3 EPRIM	MUJER 45- ESTUDIOS PRIMARIOS
44	M3 ESECU	MUJER 45- ESTUDIOS SECUNDARIOS
45	M3 EUNIV	MUJER 45- ESTUDIOS UNIVERSITARIOS
46	M3 PPRIN	MUJER 45- PERSONA PRINCIPAL
47	M3 CGRUP	MUJER 45- CABEZA DE GRUPO
48	M3 OTROS	MUJER 45- OTROS

Simples, siempre que se sea consciente de los cambios en la interpretación que se producen respecto al caso habitual.

En nuestro ejemplo, las columnas y filas son las que aparecen en el cuadro n.º 1, en donde van acompañadas por la abreviaturas que utiliza el ordenador.

Utilizando el programa de Análisis de Correspondencias Simples incluido dentro del paquete ESTIN, que ha elaborado el Centro de Cálculo de la Facultad de Ciencias Económicas de la Universidad de Deusto, se obtienen los resultados que comentamos a continuación.

Hay dos factores, o dos dimensiones, que explican el 82 % del total de inercia a explicar, y si consideramos las cuatro primeras dimensiones se consigue explicar el 97 %, tal como aparece

en el cuadro n.º 2. Representando gráficamente la situación de las filas y columnas en el espacio de las dos primeras dimensiones pero señalando claramente los casos que están explicados fundamentalmente por las otras dos, tenemos el gráfico n.º 3, que nos puede ayudar, junto con la tabla anterior, a hacer las observaciones siguientes:

Este plano está definido fundamentalmente por la oposición entre la ocupación en el sentido habitual (ocupados y sobreocupados), y la ocupación marginal y, sobre todo, la no laboral, oposición que define el primer eje, mientras que el segundo está definido por la modalidad «jubilados». Las otras dos dimensiones las definen los otros colectivos: «parados» y los «no laborales que buscan empleo» provocan el tercer eje, y los «parados no laborales» el cuarto.

Cuadro n.º 2. Contribuciones de los factores a las columnas y filas activas

INERCI A TOTAL = 0.77504742 G|**2 (423) = *****

VALOR PROPIO PORCENTAJE % ACUMULADO

1	0.42543360	54.89	54.89	1*****
2	0.21320999	27.51	82.40	1*****
3	0.07304722	9.42	91.83	1*****
4	0.04386495	5.66	97.49	1*****

CONTRIBUCIONES DE LOS FACTORES A LAS COLUMNAS ACTIVAS

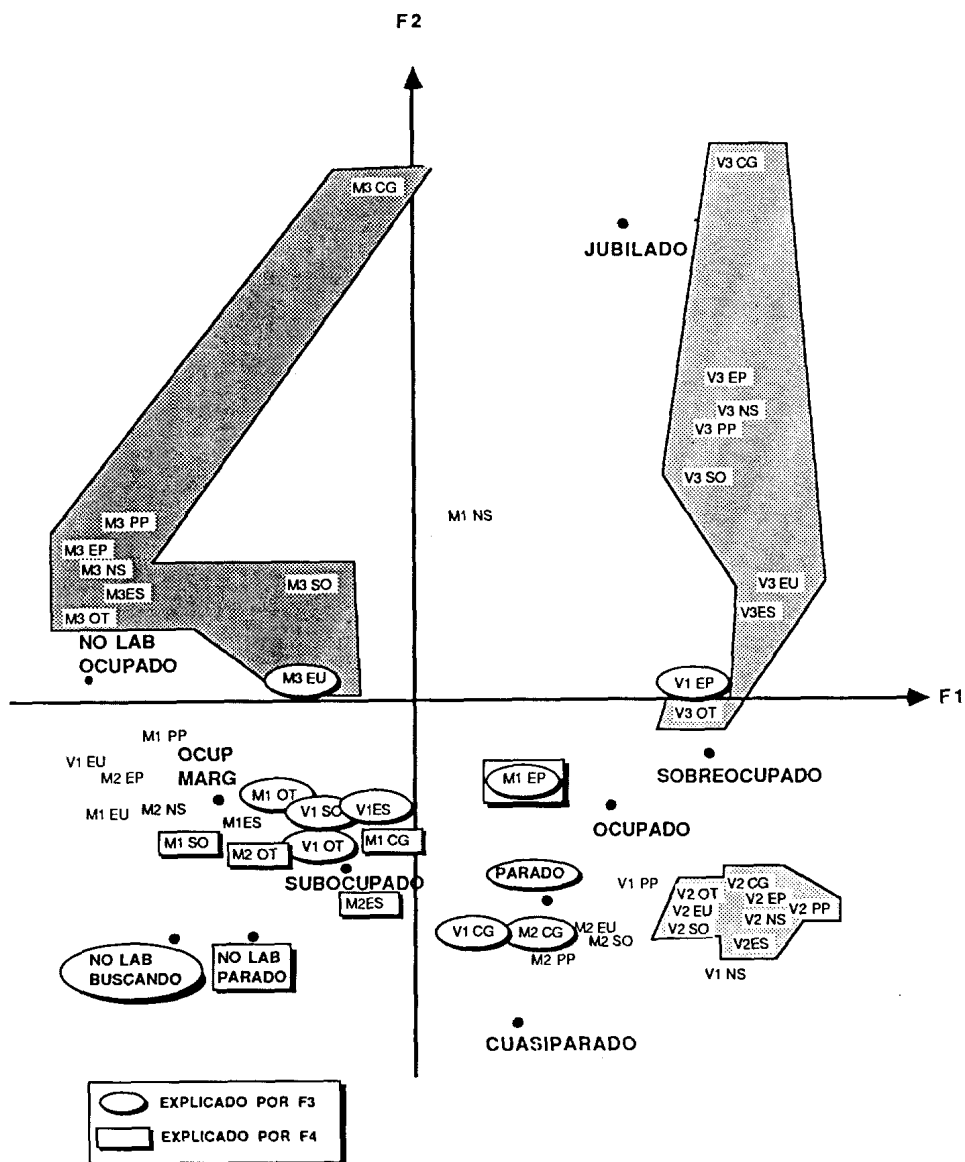
		COORDENADAS						CALIDAD
		MASA	DISTANCIA	F1	F2	F3	F4	EXPLICAC
1	SOBREOCU	0.0703	0.6747	0.7263	-0.1270	-0.1882	-0.0753	86.6740
2	OCUP EST	0.3192	0.4164	0.5595	-0.2558	-0.1834	0.0048	98.9680
3	SUBOCUPA	0.0128	0.3957	-0.1850	-0.3859	0.1054	0.2489	64.7403
4	OCUP MRG	0.0050	0.5455	-0.4465	-0.2294	0.1585	0.0568	51.3820
5	CUASIPAR	0.0061	1.2288	0.2119	-0.7556	0.3724	-0.1375	62.9502
6	PAR.N.LA	0.0467	1.2073	-0.3722	-0.5390	0.3468	0.8048	99.1486
7	N.LA.BUS	0.0076	1.9535	-0.5506	-0.5520	0.7642	0.2751	64.8871
8	N.LA.OCU	0.3419	0.7219	-0.8393	0.0679	-0.0859	-0.0732	99.9899
9	PARO EST	0.0664	1.0825	0.3055	-0.4632	0.7983	-0.3671	99.7759
10	JUBILADA	0.1241	1.5280	0.4989	1.1136	0.1753	0.0881	99.9745

ACUMULADOS

Cuadro n.º 2. Contribuciones de los factores a las columnas y filas activas

			COORDENADAS						CALIDAD
			MASA	DISTANCIA	F1	F2	F3	F4	EXPLICAC
1	V1	SOLT.	0.0338	0.4473	-0.2296	-0.2519	0.4938	-0.2759	97.4924
2	V1	NO SO	0.0007	1.6392	0.7541	-0.6557	0.4731	-0.5123	90.5837
3	V1	EPRIM	0.0010	1.9068	0.6957	0.0109	0.9186	-0.4962	82.5541
4	V1	ESECU	0.0288	0.5296	-0.1344	-0.2886	0.5503	-0.3382	97.9135
5	V1	EUNIV	0.0048	0.9419	-0.8451	-0.1473	0.0634	0.1082	79.8024
6	V1	PPRIN	0.0006	0.9823	0.5345	-0.4187	-0.1854	-0.3572	63.4195
7	V1	CGRUP	0.0007	1.5571	0.1487	-0.5389	0.7931	-0.6631	88.7026
8	V1	OTROS	0.0332	0.4505	-0.2310	-0.2511	0.4994	-0.2708	97.4864
9	V2	SOLT.	0.0229	0.8796	0.6549	-0.5366	0.2204	-0.2535	94.3275
10	V2	NO SO	0.0392	1.2191	0.8513	-0.5053	-0.4698	-0.0796	99.0186
11	V2	EPRIM	0.0256	0.9807	0.8134	-0.4686	-0.2820	-0.1150	99.3046
12	V2	ESECU	0.0267	1.0261	0.7931	-0.5677	-0.1266	-0.2135	98.7170
13	V2	EUNIV	0.0097	0.7641	0.6492	-0.5040	-0.2837	-0.0272	99.0438
14	V2	PPRIN	0.0381	1.2194	0.8511	-0.5080	-0.4716	-0.0848	99.3887
15	V2	CGRUP	0.0032	1.1763	0.7981	-0.5092	-0.2297	-0.1190	81.8806
16	V2	OTROS	0.0208	0.8822	0.6440	-0.5341	0.2554	-0.2551	94.1277
17	V3	SOLT.	0.0056	0.7990	0.6847	0.5113	-0.0325	0.1077	92.9848
18	V3	NO SO	0.0612	1.0147	0.7319	0.6845	0.0732	0.0660	99.9345
19	V3	EPRIM	0.0557	1.1063	0.7176	0.7572	0.1029	0.0823	99.9384
20	V3	ESECU	0.0072	0.6056	0.7407	0.2088	-0.0700	0.0093	98.6150
21	V3	EUNIV	0.0039	1.0803	0.8536	0.2746	-0.2404	-0.0023	79.7788
22	V3	PPRIN	0.0605	0.9635	0.7261	0.6540	0.0626	0.0616	99.9034
23	V3	CGRUP	0.0043	2.2251	0.7884	1.2169	0.1459	0.1764	96.8423
24	V3	OTROS	0.0020	0.5103	0.6561	-0.0174	-0.0581	0.0798	86.3227
25	M1	SOLT.	0.0310	0.3104	-0.3525	-0.2566	0.2939	0.1717	98.5923
26	M1	NO SO	0.0022	1.8303	-0.4510	-0.3703	0.3125	1.0779	87.4224
27	M1	EPRIM	0.0017	1.4719	0.2448	-0.1895	0.6613	0.5466	56.5185
28	M1	ESECU	0.0251	0.3362	-0.3255	-0.2641	0.2890	0.2281	92.5850
29	M1	EUNIV	0.0064	0.8402	-0.6463	-0.2836	0.2247	0.1630	68.4693
30	M1	PPRIN	0.0001	0.8807	-0.5961	-0.1010	-0.2862	0.0290	50.9013
31	M1	CGRUP	0.0016	0.5726	-0.0527	-0.2718	0.0959	0.3112	31.9144
32	M1	OTROS	0.0314	0.3597	-0.3739	-0.2645	0.3082	0.2282	99.1898
33	M2	SOLT.	0.0131	0.8097	0.4662	-0.5608	0.1949	0.3607	86.4384
34	M2	NO SO	0.0474	0.4501	-0.5419	-0.2534	-0.1194	0.2634	98.0760
35	M2	EPRIM	0.0306	0.5364	-0.6569	-0.1859	-0.1021	0.2238	98.1823
36	M2	ESECU	0.0222	0.3519	-0.1199	-0.4210	-0.0058	0.3897	97.6165
37	M2	EUNIV	0.0078	0.6459	0.4031	-0.5578	0.0182	0.2220	81.0158
38	M2	PPRIN	0.0018	1.2338	0.3223	-0.5708	-0.0921	0.6997	75.2037
39	M2	CGRUP	0.0016	1.2554	0.3704	-0.5616	0.6729	0.1816	74.7495
40	M2	OTROS	0.0572	0.3089	-0.3626	-0.3056	-0.0700	0.2744	98.7488
41	M3	SOLT.	0.0065	0.1669	-0.2243	0.2853	-0.0901	0.1050	90.3516
42	M3	NO SO	0.0696	0.8422	-0.8159	0.3248	-0.2021	-0.1669	99.7224
43	M3	EPRIM	0.0688	0.8006	-0.7893	0.3401	-0.1901	-0.1549	99.7819
44	M3	ESECU	0.0050	0.5676	-0.6857	0.1900	-0.1661	-0.0326	94.2470
45	M3	EUNIV	0.0023	0.3261	-0.2138	0.0476	-0.3243	-0.0455	47.6024
46	M3	PPRIN	0.0147	0.6624	-0.6565	0.4233	-0.1624	-0.1148	98.0914
47	M3	CGRUP	0.0085	1.6210	-0.0307	1.2479	0.2060	0.1014	99.3836
48	M3	OTROS	0.0529	0.9635	-0.9135	0.1443	-0.2650	-0.1910	99.8568

Gráfico n.º 3. Actividad según edad, sexo, estado civil, nivel de instrucción y situación familiar



En cuanto a las filas se puede apreciar inmediatamente el comportamiento homogéneo y compacto de los varones en edad intermedia, entre 25 y 44 (V2), que se encuentran caracterizados por la ocupación, sin que los otros atributos influyan en absoluto, ni el nivel de estudios, ni el estado civil, ni la situación familiar. Los varones de más edad, mayores de 45 años (V3), forman un abanico que oscila entre la ocupación y la jubilación, en donde los estudios superiores acercan a la ocupación; (posiblemente por la mayor abundancia de profesiones liberales), y la situación familiar como cabeza de grupo acerca a la jubilación. Esta segunda circunstancia seguramente enmascara la influencia de la edad en un colectivo heterogéneo en ese terreno (+ 45 años), de forma que en los cabeza de grupo abundarán los padres mayores que viven con hijos casados, mientras que en Persona Principal y Otros, abundarán más los jóvenes y, por tanto, todavía en activo. Este atributo pone de manifiesto que la escala de edad no está bien definida, por lo menos entre los varones, y que el colectivo de más de 45 años no es realmente homogéneo.

Entre los varones más jóvenes la situación es mucho más confusa y es en donde los otros atributos modifican mucho las relaciones: el estado civil y la situación familiar por ejemplo, son determinantes: los jóvenes casados y/o personas principales están próximos a situaciones más parecidas a los varones de edad intermedia, mientras que los solteros y/o los que no tienen responsabilidades familiares quedan fuera de este plazo y están más caracterizados por la situación de parados. Aunque esto último tampoco se puede afirmar totalmente porque los estudios introducen otra modificación, ya que los que tienen estudios universitarios se encuentran próximos a la ocupación no laboral.

Otro aspecto que puede observarse al analizar el comportamiento de los varones es la proximidad que se da en las modalidades estado civil y situación familiar, que hace sospechar una relación entre ellos, aunque esta relación no es tan clara en los niveles altos de edad, por el cambio que se produce en el comportamiento de los «Cabeza de Grupo».

Entre las mujeres la situación es mucho más confusa, y sólo las mayores de 45 años (M3), tienen un comportamiento bien explicado por este plano —salvo las universitarias, aunque es un grupo muy pequeño—: en efecto, al igual que los varones, forman un arco pero que en lugar de ir desde la ocupación a la jubilación, va desde la ocupación no laboral también hasta la jubilación, y diferenciándose mucho según la situación familiar que, de nuevo, parece poner en evidencia un fallo en la escala de edad.

Las mujeres de edad intermedia (M2) tienen grandes diferencias entre ellas: las solteras y/o de estudios universitarios se parecen a los varones de su misma edad, mientras que las no solteras y/o de estudios primarios se acercan más a sus mayores, es decir, a la ocupación no laboral. Las mujeres más jóvenes (M1) también son, como los varones, las que se aproximan más a las modalidades peor definidas por este plano, aunque al revés que los varones, no tanto a los parados estrictos sino a los parados no laborales.

De todos modos, en este análisis apresurado se pone de manifiesto la dificultad de analizar bien estos colectivos más reducidos y fronterizos entre ocupación, inactividad y paro, que la PRA define, y que parecen estar fundamentalmente asociados a las edades más jóvenes.

Si se pretende estudiar más detenidamente estos colectivos puede haber la tentación de utilizarlos de forma exclusiva eliminando del análisis a todos los demás, pero así se deformaría totalmente la realidad, y tomarían una importancia desmesurada colectivos que siguen siendo pequeños. Por eso pensamos que es más razonable conservar el resto de las modalidades pero agrupadas, de forma que sigan produciendo el necesario contrapunto.

1.3. Análisis del colectivo no laboral

A modo de ejemplo vamos a analizar ese colectivo que la mayoría de las encuestas sobre actividad incluyen entre los inactivos o entre los parados, y que la P.R.A. diferencia porque aunque no realizan una actividad económicamente

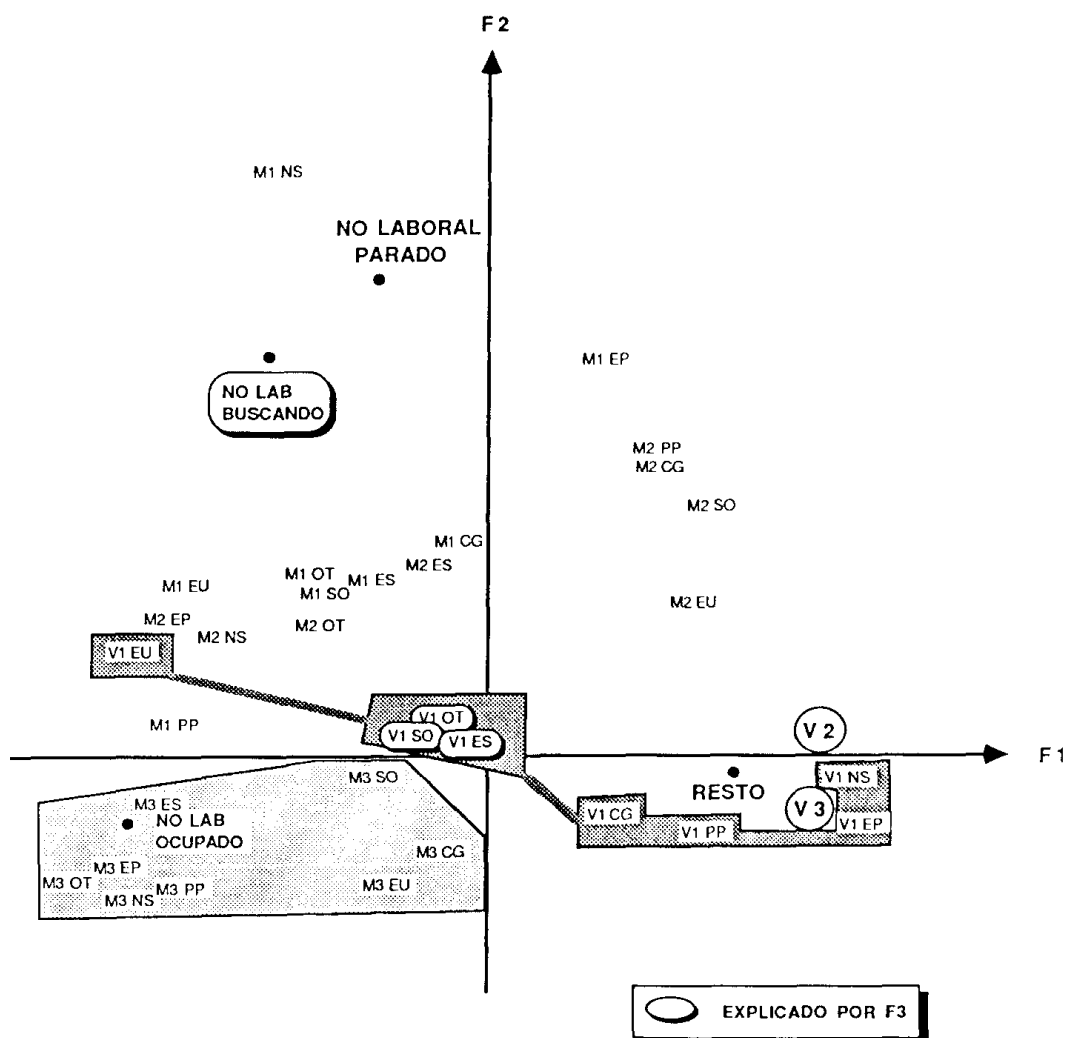
retribuida, tampoco se pueden considerar en realidad inactivos. Nos referimos a los que realizan actividades no laborales — es decir, tareas de hogar, estudios, servicio militar—, y que, a su vez, se diferencian según el deseo de encontrar un empleo y su disponibilidad actual para poder ocuparlo. De esta forma, los NO LABORALES PARADOS son los que están en esa situación pero quieren y podrían ocupar un empleo, los NO LABORALES BUSCANDO, quisieran pero no podrían, y, por fin, los NO LABORALES OCUPADOS son los que no buscan empleo remunerado. En lo

sucesivo les llamaremos PARADOS N.L., BUSCANDO N.L. y OCUPADOS N.L.

Se han mantenido estas tres modalidades y el resto de las que tenía el atributo actividad se suman en una columna que, lógicamente, llamaremos RESTO. Esta modalidad en realidad no es tan heterogénea, ya que incluye a todos los que tienen una actividad laboral, la han tenido, o la quieren tener.

El resultado del análisis es la obtención de tres dimensiones que explican la totalidad de la inercia (lo cual es lógico puesto que había sólo cuatro columnas),

Gráfico n.º 4. Análisis del Colectivo No Laboral



aunque sólo el primer eje ya explica el 85 %, y entre los dos primeros el 98 %.

El gráfico n.º 4 muestra la posición de las filas y las columnas en ese plano que explica casi la totalidad de la dispersión, el primer eje definido por la oposición de los OCUPADOS N.L. con el RESTO, mientras que el segundo lo definen los PARADOS N.L. y queda un resto de capacidad explicativa que lo completa en parte los que están BUSCANDO N.L.

En cuanto a las filas, los varones por encima de los 25 años (V2 y V3) tienen un comportamiento tan compacto alrededor del RESTO (es decir, de ocupación, paro y jubilación estrictas), que no merece la pena destacar las pequeñas variaciones que se pueden dar por su estado civil, etc. Algo parecido pasa con las mujeres mayores que se agrupan alrededor de OCUPADOS N.L. aunque al haber unos subgrupos definidos por ser solteras y/o tener estudios universitarios que se alejan de esa situación, hemos preferido señalar todas las modalidades.

Los varones menores de 25 años tienen un comportamiento muy distinto según sus otras características: los no solteros, los que tienen estudios primarios y/o los que tienen responsabilidades familiares no se diferencian de los que son mayores que ellos y se aproximan a las características laborales —paro u ocupación estrictas—. Por el contrario, los que tienen estudios secundarios son solteros y/o no tienen responsabilidades familiares están definidos sobre todo por ese tercer eje que se relaciona con querer un trabajo pero no poder ocuparlo de momento, que hemos llamado BUSCANDO N.L. Quedan al margen los jóvenes varones universitarios que no parecen querer un trabajo de momento y que se aproximan a OCUPADOS N.L.

Hemos dejado para el final el análisis de los dos colectivos más complejos: las mujeres menores de 44 años (M1 y M2), que se sitúan a lo largo de una serie de ejes que oscilan entre la actividad no laboral, ya sea sin buscar empleo (OCUPADOS N.L.), o buscándolo (PARADOS N.L.), y la actividad estricta (paro u ocupación estricta), ejes que dependen de las otras características y que, además, parecen de dirección contraria según la edad. Para que quede

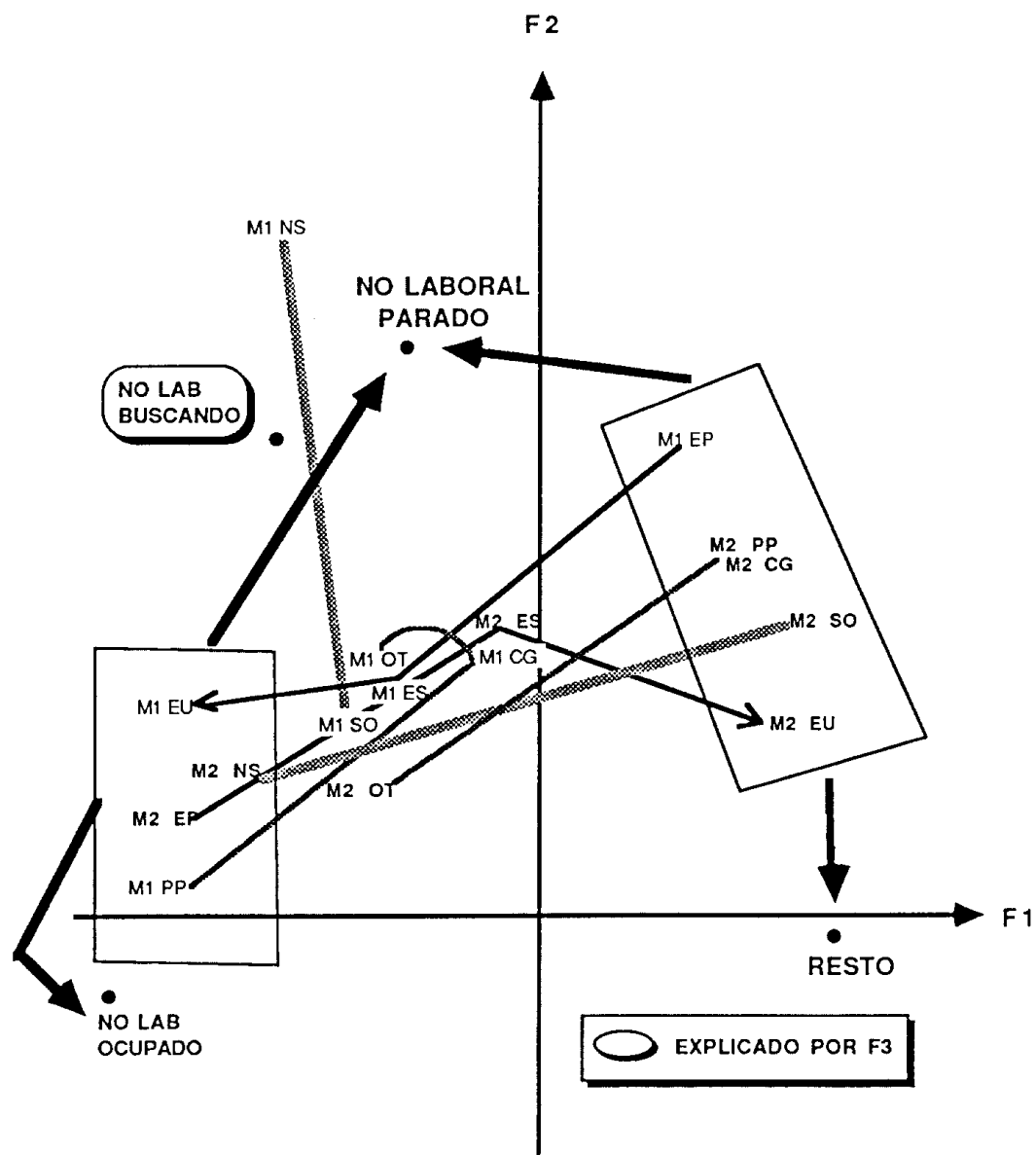
más claro, puede interesar realizar otra figura que destaque estas relaciones y en donde todas las demás modalidades de sexo y edad se eliminan (ver gráfico n.º 5).

El estado civil influye de forma totalmente distinta según la edad. Entre las más jóvenes produce que al casarse tiendan a querer un empleo, de tal forma que este grupo es el que queda más próximo al de PARADOS N.L. Entre las de edad intermedia provoca, por el contrario, la conformidad de las casadas con su ocupación no laboral, mientras que las solteras se encuentran entre el RESTO (actividad en sentido estricto), y el deseo de buscar empleo.

También el nivel de instrucción se comporta de forma distinta: entre las más jóvenes, al aumentar el nivel educativo se desplaza de esa zona que oscila entre la actividad estricta y la búsqueda de empleo y se acerca a esa otra más próxima a la ocupación no laboral; entre las de edad intermedia sucede justamente lo contrario. Es evidente que entre las más jóvenes los estudios universitarios están todavía en curso, lo que provoca el que no se busque un empleo inmediato, y su clasificación en ocupadas no laborales, mientras que al pasar de los 25 años son universitarias que ya han terminado la carrera y están trabajando o en disposición de hacerlo en un porcentaje importante, lo que las sitúa en la zona de actividad estricta (ocupación o paro).

En todo lo anterior se puede apreciar que la relación con la actividad está indudablemente relacionada con otras características de las personas, y entre ellas de forma fundamental, por la edad y el sexo, aunque otras características como las responsabilidades familiares—que se deben al estado civil y a la situación familiar—, y el nivel de instrucción, pueden también influir, sobre todo en algunos segmentos de edad. Precisamente esta fuerte influencia de la edad es la que hace más evidente las deficiencias de la escala de edad que se ha elegido en el ejemplo anterior, que, por ser excesivamente limitada en el número de estratos, puede provocar una mezcla de colectivos dentro de cada modalidad, sobre todo al comienzo y al final de la escala, que es cuando la relación con la actividad está más sujeta a cambios.

Gráfico n.º 5. **Análisis del Colectivo de Mujeres menores de 45 años**



Por ello, hemos considerado que se podría ampliar esa escala de edad hasta 6 estratos, y mantener su combinación con el sexo, eliminando todas las demás características.

1.4. Actividad y edad-sexo con escala de edad ampliada

Este nuevo análisis que no yuxtapone ya atributos distintos sino que trata la actividad en su relación con un solo

atributo, la combinación de edad y sexo, da los resultados siguientes:

Las dos primeras dimensiones explican el 86 % de la dispersión, que llega al 95 % cuando se añade una tercera dimensión. Los gráficos n.^{os} 6, 7 y 8 contienen la representación gráfica de los resultados en los tres planos que resultan.

En estas figuras se han añadido los centros de gravedad por niveles de edad, por sexo y por provincia.

Cuadro n.º 3. Contribuciones de los factores a las columnas y filas activas

CONTRIBUCIONES DE LOS FACTORES A LAS COLUMNAS ACTIVAS

		COORDENADAS						CALIDAD EXPLICAC
		MASA	DISTANCIA	F1	F2	F3	F4	
1	SOBREOCU	0.0703	0.7579	0.6338	-0.5423	0.1811	0.0844	97.0671
2	OCUP EST	0.3191	0.5330	0.6111	-0.3653	0.1500	-0.0511	99.8101
3	SUBOCUPA	0.0128	0.3978	0.3051	0.3210	-0.3051	-0.2192	84.7861
4	OCUP MRG	0.0051	0.5036	-0.1000	0.4298	-0.3436	-0.0219	62.1983
5	CUASIPAR	0.0061	1.3146	0.5521	0.0515	-0.6411	0.0898	55.2737
6	PAR C AC	0.0466	1.2249	0.2377	0.4868	-0.6409	-0.7185	99.6299
7	BUSCA EM	0.0075	1.5303	0.1505	0.5749	-0.9773	-0.0481	85.6360
8	OCU E AC	0.3419	0.6832	-0.3329	0.7441	0.1246	0.0555	99.9951
9	PARO EST	0.0665	1.0005	0.4583	-0.1239	-0.7353	0.4831	99.8935
10	JUBILADA	0.1241	3.0161	-1.4125	-1.0080	-0.0608	-0.0363	99.9993
ACUMULADOS		*****						

CONTRIBUCIONES DE LOS FACTORES A LAS FILAS ACTIVAS

		COORDENADAS						CALIDAD EXPLICAC
		MASA	DISTANCIA	F1	F2	F3	F4	
1	HO 16-24	0.1036	0.4438	0.0954	0.2649	-0.4582	0.3848	98.5312
2	MU 16-24	0.0996	0.3468	0.0442	0.4050	-0.3815	-0.1617	97.3527
3	HO 25-34	0.0954	0.8680	0.7612	-0.4642	-0.1004	0.1745	96.2553
4	MU 25-34	0.0939	0.3662	0.3108	0.2320	-0.2613	-0.3747	98.0732
5	HO 35-44	0.0908	1.1832	0.8323	-0.6063	0.3400	-0.0447	99.5468
6	MU 35-44	0.0877	0.4295	0.0055	0.5950	0.1166	-0.2258	97.4461
7	HO 45-54	0.0738	0.9952	0.7394	-0.6070	0.2611	0.0068	98.8102
8	MU 45-54	0.0708	0.6709	-0.1833	0.7507	0.2609	0.0223	99.2363
9	HO 55-64	0.0696	0.9807	-0.3590	-0.9198	-0.0056	0.0315	99.5217
10	MU 55-64	0.0739	0.9096	-0.3523	0.7863	0.3793	0.1437	99.6929
11	HO 65-	0.0571	5.0863	-1.8206	-1.3166	-0.1305	-0.1313	99.9209
12	MU 65-	0.0836	1.0994	-0.9407	0.3628	0.2456	0.1343	99.5966
ACUMULADOS		*****						

Gráfico n.º 6. Actividad y Edad-Sexo. Escala ampliada. Plano F1-F2

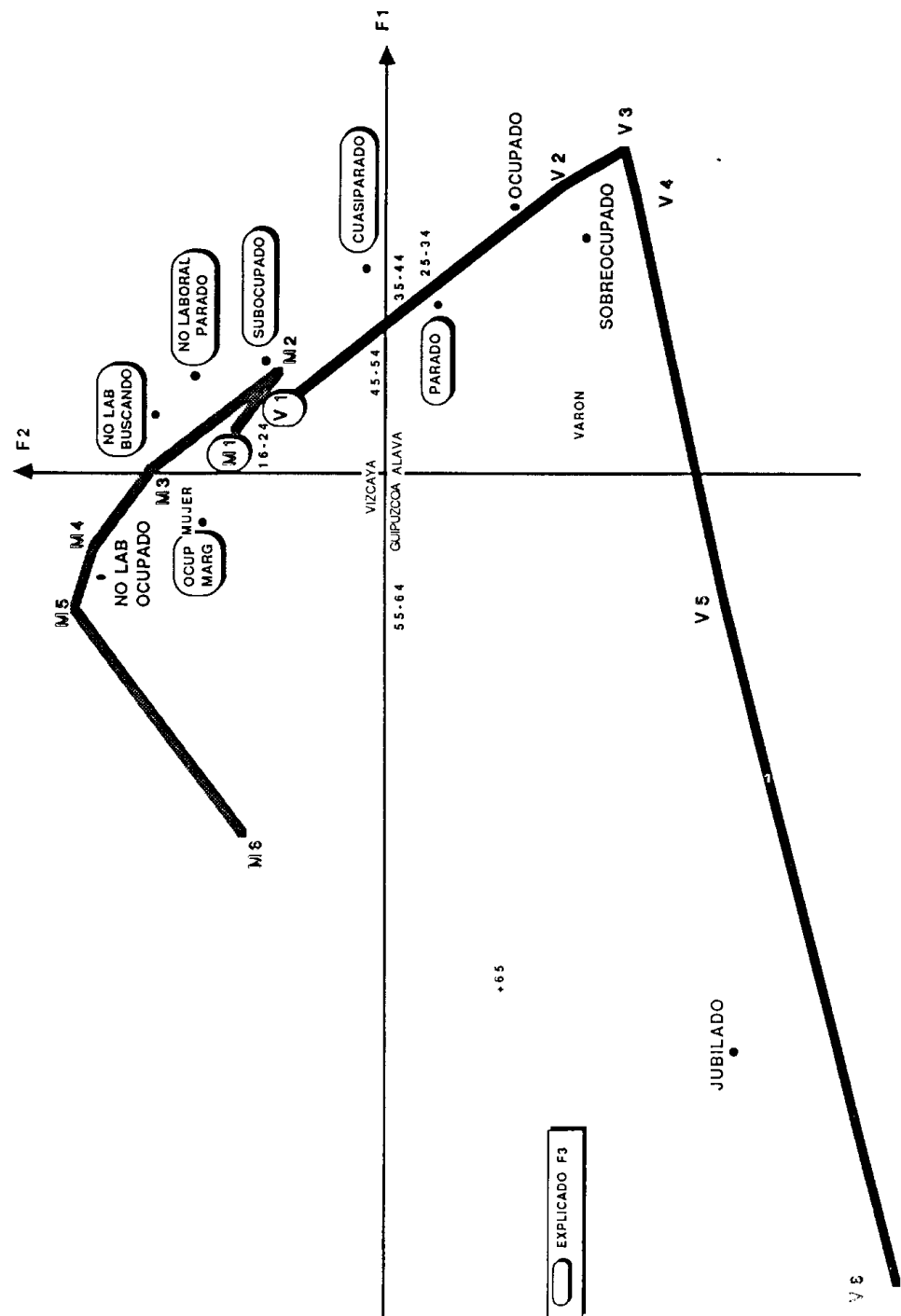


Gráfico n.º 7. Actividad y Edad-Sexo. Escala ampliada. Plano F1-F3

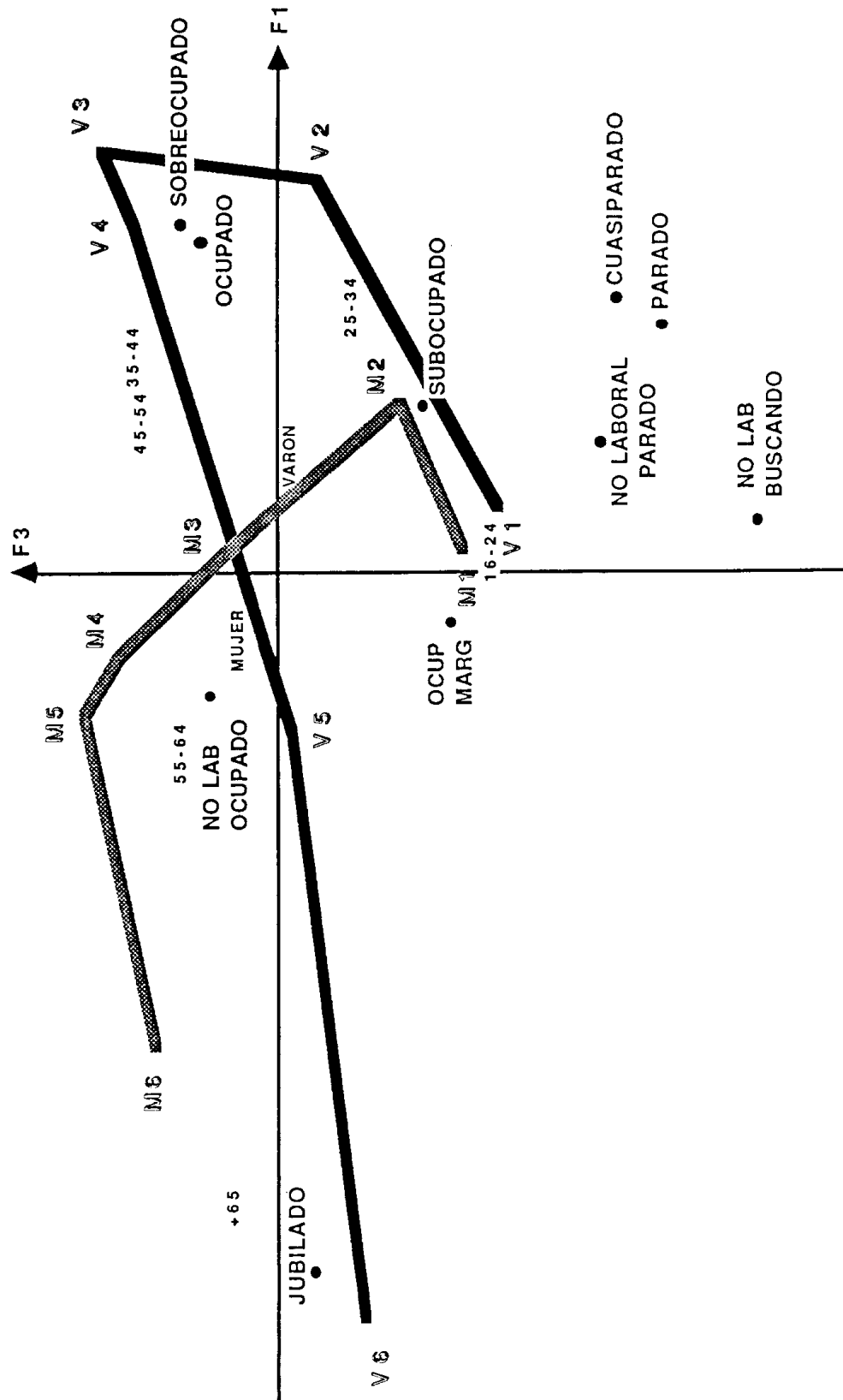
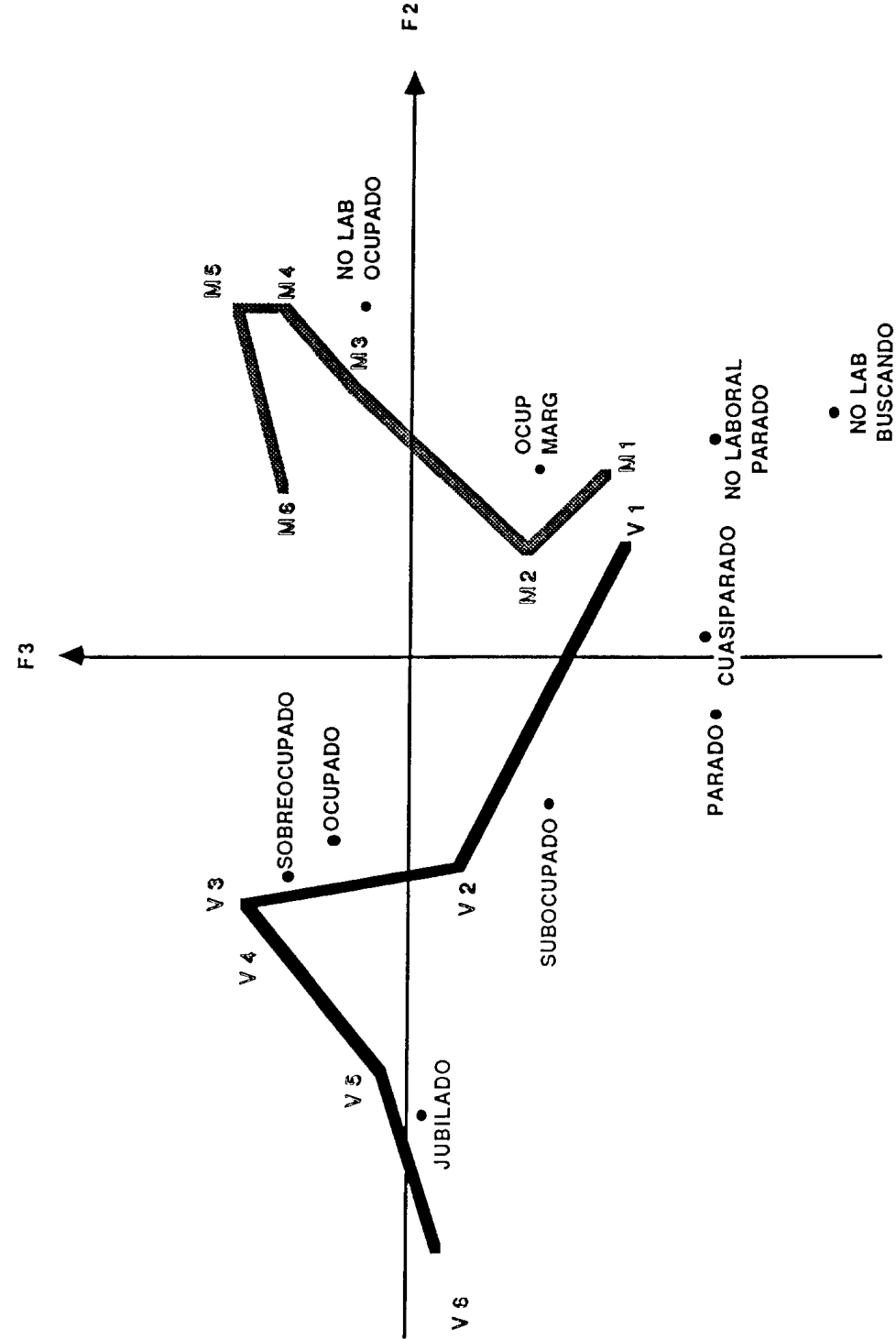


Gráfico n.º 8. Actividad y Edad-Sexo. Escala ampliada. Plano F2-F3



En el gráfico n.º 6 vemos de nuevo que el plano que mayor cantidad de inercia explica está definido por la ocupación tanto laboral como no laboral, y por la jubilación, mientras que el tercer eje queda definido por los parados y por estos otros colectivos marginales. Los diferentes colectivos definidos por el sexo y la edad se sitúan formando dos arcos muy claros que tienen un origen común: los más jóvenes no se diferencian por el sexo y se hunden ambos hacia el tercer factor definido por el paro. A partir de ahí las mujeres se sitúan próximas a actividad no laboral, primero buscando empleo, pero en los últimos estratos de edad ya sin hacerlo, con una cierta flexión hacia la jubilación al tener más de 65 años. Entre los varones el arco se dirige primero hacia la ocupación laboral y al final tiende hacia la jubilación.

La relación del tercer eje con el paro y las demás modalidades más marginales de la actividad se pueden percibir en el gráfico n.º 7, en donde la evolución de hombres y mujeres es bastante similar y depende fundamentalmente de la edad: al principio paro o subocupación, ocupación —laboral o no—, en las edades intermedias y, por fin, jubilación.

Por fin, en el gráfico n.º 8 vemos el plano que queda cuando quitamos el eje más explicativo, el F1, que diferencia ocupación y jubilación, y nos queda la diferencia clara entre los distintos tipos de ocupación: ocupación no laboral a la derecha y la ocupación laboral y jubilación a la izquierda, lo primero asociado a las mujeres de edades intermedias y lo segundo a los varones de esas mismas edades.

2. CONCLUSIONES

Todos estos ejemplos dan resultados que eran bastante previsibles, aunque esto es algo que se pretendía para que la utilidad de la técnica se pusiera claramente de manifiesto: el Análisis de Correspondencias Simples puede utilizarse para analizar las relaciones entre dos o más atributos a partir de modelos bastante próximos a los modelos causales econométricos, es decir, cuando haya uno de ellos cuyo comportamiento estamos

fundamentalmente interesados en describir a partir de la influencia de los otros.

Las recomendaciones para abordar este tipo de análisis tal como parecen deducirse de nuestros ejemplos serían las siguientes:

- Cuando dos atributos de los que consideramos exógenos no tengan excesiva relación entre ellos, no hay que yuxtaponerlos, no hay que mantenerlos separados, sino que hay que crear un nuevo atributo que combine las modalidades de ambos.
- Cuando sea previsible que dos atributos exógenos estarán bastante relacionados entre ellos, se puede optar por quitar uno de ellos y evitar la redundancia, pero, al contrario de lo que sucede en econometría, no hay excesivo problema en mantener los dos yuxtapuestos, ya que cada uno no afecta a la relación del otro. En el análisis quedará claro si la hipótesis de la relación entre ellos era cierta y en qué modalidades se produce cuando se ponen en relación con el atributo endógeno.
- Puede ser frecuente que alguna de las modalidades del atributo endógeno tenga pocos efectivos y su análisis resulte difícil. Para facilitararlo se puede concentrar el estudio en esas modalidades poco frecuentes, pero no eliminando los individuos caracterizados por los demás, sino reuniéndolos en una sola modalidad, de forma que su peso conjunto siga apareciendo. De otro modo se podría falsear totalmente el análisis.
- Es fundamental una definición adecuada de las distintas modalidades de los atributos, y esto es especialmente importante en el caso en que esas modalidades no sean naturales, sino escalas artificialmente colocadas en un fenómeno continuo como sucede con la edad, la renta, etc.
- La mayoría de las salidas de ordenador presentan dos tipos de resultados, las coordenadas, que dependen de la distancia original y de cómo queda esta distancia explicada en el espacio que se conserva, y, además, lo que se

suele llamar correlaciones y que más propiamente son esas mismas coordenadas tipificadas, es decir, haciendo que la distancia original sea uno. Esta segunda información es algo engañosa porque puede potenciar la importancia de modalidades poco importantes, pero la primera también puede serlo si no se tiene en cuenta la capacidad de

explicación que se ha conseguido. Lo preferible es utilizar las dos — olvidando el impropio nombre de «correlación» que se suele dar al segundo—, y tener siempre en cuenta los tres elementos que hay que considerar: distancia original, masa, y distancia que se consigue explicar por el espacio concreto que se analiza en cada caso y en cada figura.