

DOCUMENTOS DE TRABAJO

BILTOKI

D.T. 2003.11

Una propuesta para el análisis de tablas múltiples de variable cualitativas.

Juan Ignacio Modroño Herrán, Karmele Fdez. Aguirre y
M. Isabel Landaluce Calvo

eman ta zabal zazu



Universidad Euskal Herriko
del País Vasco Unibertsitatea

Facultad de Ciencias Económicas.
Avda. Lehendakari Aguirre, 83
48015 BILBAO.

Documento de Trabajo BILTOKI DT2003.11

Editado por los Departamentos de Economía Aplicada II (Hacienda), Economía Aplicada III (Econometría y Estadística), Fundamentos del Análisis Económico I, Fundamentos del Análisis Económico II e Instituto de Economía Pública de la Universidad del País Vasco.

23-7-2003

Depósito Legal No.: BI-2082-03

ISSN: 1134-8984

Una propuesta para el análisis de tablas múltiples de variables cualitativas. Aplicación a la cultura de investigación en una universidad*

J.I. Modroño, K. Fernández Aguirre, M.I. Landaluce

21 de julio de 2003

Resumen

El análisis de tablas múltiples ha seguido cuatro direcciones básicas: el Análisis Procusteano, la metodología STATIS, el Análisis Factorial Múltiple y el Análisis Canónico Generalizado. Estos métodos, en general, han sido diseñados para el análisis de un conjunto de matrices de naturaleza cuantitativa que recogen información sobre el mismo o distinto conjunto de variables medidas en el mismo conjunto de individuos. No obstante, cuando nuestro objetivo es realizar un estudio comparativo de encuestas de opinión las tablas a analizar recogen información de naturaleza, generalmente, cualitativa y el número de encuestados (filas) en cada tabla es diferente. En este trabajo, centrándonos en el Análisis Factorial Múltiple (AFM) y STATIS, vamos a mostrar una extensión de estos métodos que permite el análisis de tablas múltiples con estas últimas características. El trabajo se completa con un análisis empírico y una estrategia de simulación.

*Trabajo parcialmente financiado por el Programa Nacional de Socioeconomía del Programa de Fomento de la Investigación Técnica (PROFIT) FIT-080000-2001-0063, proyectos AD01UN02 y PE02UN05 del Gobierno Vasco, proyecto AE01/C04 de la UPV y el Gobierno Vasco y el grupo consolidado de la Universidad del País Vasco 9/UPV 00038.321-13631/2001.

1. Introducción

El análisis de tablas múltiples es una metodología factorial que está construida sobre el análisis factorial clásico basado en el teorema de Eckart y Young de descomposición en valores singulares de una tabla rectangular de datos (Eckart & Young 1936). La consideración de tablas rectangulares es un tanto restrictiva y pronto se pensó en ampliar el análisis con una tercera vía, mediante la consideración de *grupos* de tablas rectangulares cuya justificación se debe a la consideración, típicamente, de diferentes momentos de tiempo ó diferentes referentes espaciales; las primeras aproximaciones en este sentido pueden encontrarse en (Tucker 1964) y (Tucker 1966). Posteriormente el análisis de tablas múltiples ha seguido cuatro direcciones básicas: el Análisis Procrusteano (Tucker 1958) y Procrusteano Generalizado (Gower 1975), la metodología STATIS iniciada por el equipo de Y. Escoufier (L'Hermier des Plantes 1976), el Análisis Factorial Múltiple (Escoufier & Pagès 1992) y Análisis Canónico Generalizado (Carrol 1968).

Los métodos anteriormente citados normalmente analizan un conjunto ó grupo de matrices de datos, cuya característica más común es que contienen información sobre variables cuantitativas, generalmente no discretas y medidas sobre un mismo colectivo o sobre colectivos del mismo tamaño, para cada una de las tablas de las que se disponen. En muchos casos ésta es una restricción excesiva, en particular en el análisis comparado de encuestas de opinión (ó no) donde, por un lado, las variables medidas son cualitativas y, por otro, el número de encuestados en cada grupo es diferente, ya sea por razones de diferente tasa de no respuesta, diferente presupuesto para la encuesta, diferente tamaño de la población original a medir,...

En este trabajo, centrándonos en el Análisis Factorial Múltiple (AFM) y en STATIS, vamos a mostrar una nueva posibilidad de aplicación de estos métodos, que permite el análisis de tablas múltiples en este tipo de situaciones. Estas dos metodologías, cada una de ellas desde una perspectiva diferente, tienen en común que equilibran la influencia de cada tabla de cara a un análisis global. En la sección 2 se muestran los detalles del método propuesto, en la sección 4 se realiza un análisis de los datos descritos previamente en la sección 3 y en la sección 5 se efectúa un ejercicio de simulación del procedimiento propuesto. Finalmente, en la sección 6 se hace un comentario final sobre los resultados.

2. Metodología

Este método propuesto se lleva a cabo en dos etapas (Ver figura 1):

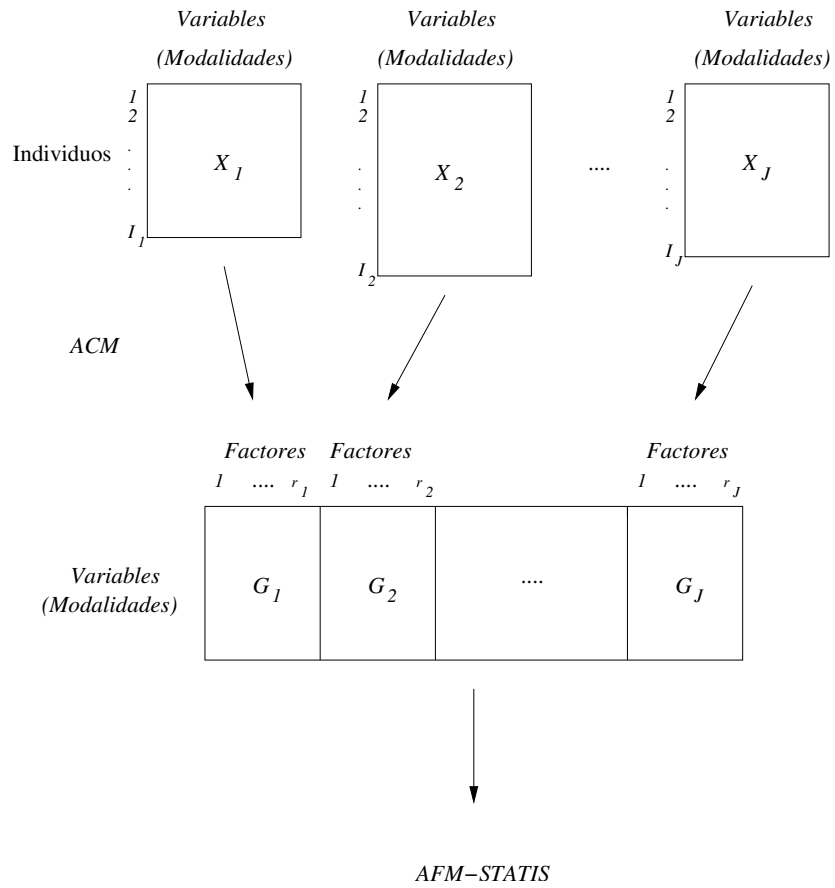


Figura 1: Esquema del método en dos etapas.

Se dispone de J tablas de datos de variables cualitativas, X_1, X_2, \dots, X_J cada una de ellas con, posiblemente diferente número de individuos, I_j , $j = 1, \dots, J$ pero el mismo número y tipo de variables indicadoras correspondientes a las variables categóricas originales $K_j = K_{j'} \quad \forall j, j' \in 1, \dots, J$.

En una primera etapa se lleva a cabo un Análisis de Correspondencias Múltiple sobre cada tabla ó matriz de datos X_j . El número de individuos no necesita ser el mismo en cada tabla, pero las tablas deben ser homogéneas en variables, en el sentido de que cada tabla debe incluir las mismas variables con las mismas categorías. A continuación, se seleccionan un número más reducido y no necesariamente igual de factores principales interpretables para cada tabla r_j , $j = 1, \dots, J$ y, a continuación, se forman unas nuevas tablas G_1, G_2, \dots, G_J con los mismos.

Posteriormente, en una segunda etapa, las tablas G_1, G_2, \dots, G_J obtenidas en el paso previo se yuxtaponen formando una gran tabla G . De esta forma, G contiene por filas las proyecciones de las categorías iniciales sobre cada uno de los factores obtenidos para cada tabla, mientras que por columnas contiene los factores r_j seleccionados para cada una de dichas tablas, que pueden ser distintos en número. Esta nueva tabla está compuesta por grupos de variables cuantitativas pero ahora con la misma dimensión en filas. Así, esta tabla, representante de las tablas iniciales, puede ser analizada, dentro del marco de tablas múltiples, tanto por medio del AFM como de STATIS y es posible comparar sus resultados, según la metodología de (Pagès 1996), (Landaluce, Fernández & Modroño 1999) y (Abascal, Fernández, Landaluce & Modroño 2001).

3. Datos

Los datos constituyen una pequeña parte de una encuesta¹ realizada a profesores de la Universidad del País Vasco, como parte de un proyecto titulado *Gestión del conocimiento en una universidad pública: El proceso de Investigación-Desarrollo-Transferencia del conocimiento científico técnico*. El objetivo de la encuesta es analizar la cultura organizacional de la universidad. El modelo que se adopta para este análisis es el *Modelo de Valores en Competencia* (Cameron & Quinn 1999) que se basa en dos dimensiones bipolares. La primera opone la posición organizacional hacia el interior frente al exterior y la segunda opone flexibilidad frente a control. Estos dos ejes forman cuatro cuadrantes que representan otros tantos tipos de cultura interpretados como Jerarquía, Mercado, Clan e Innovación, respectivamente (figura 2).

¹La encuesta ha sido diseñada por el grupo de trabajo dirigido por el Dr. Sabino Ayestarán e integrado además por K. Fernández, J. Landeta, J. Larrea, P. Mariel, A. Martín y F. Panera.

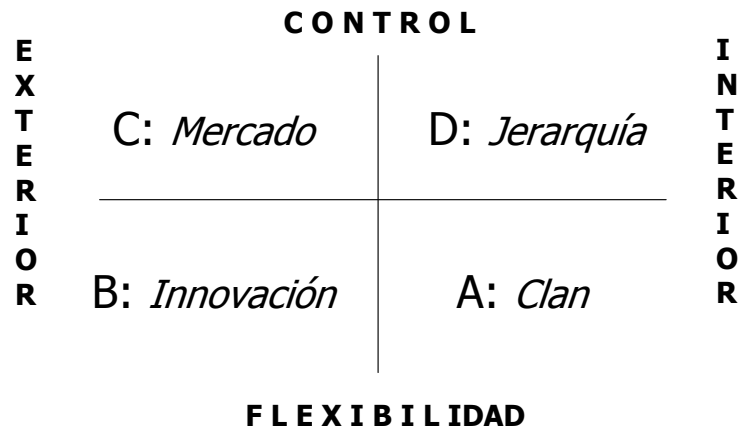


Figura 2: Tipos de cultura organizacional.

Jerarquía Cultura caracterizada por un espacio de trabajo formalizado y estructurado donde los procedimientos rigen lo que las personas hacen. Los líderes efectivos son buenos coordinadores y organizadores. Las metas de la organización a largo plazo son la estabilidad, la previsibilidad y la eficiencia. Las reglas formales y las políticas son las que hacen funcionar a la organización.

Mercado Se basa en la *dirección por objetivos* y la *transacción de costes*. Se orienta hacia el exterior más que a los asuntos internos, hacia transacciones con otros organismos externos, y funciona a través de mecanismos externos o *de mercado*.

Clan De forma similar a una familia, una organización así se caracteriza por equipos de trabajo, la implicación de los empleados en los programas y por el compromiso corporativo con los empleados. El éxito es definido en términos de clima interno. Se caracteriza por el trabajo en equipo, la participación y el consenso.

Innovación El objetivo último de la organización es la adaptabilidad, la flexibilidad y la creatividad. La tarea principal de la dirección es estimular el conocimiento, el riesgo y la creatividad con la finalidad de poseer los avances más recientes. Los equipos de mejora son parte fundamental de este tipo de corporaciones.

Según este modelo, una institución puede situarse en varias de estas posiciones en mayor o menor grado o en ninguna. Evidentemente, la cultura de *Innovación* es la que se trataría de impulsar en la universidad.

Con respecto a la parte de información disponible que se va a analizar para el caso de la Universidad del país Vasco, se tienen respuestas en una escala de 5 puntos (totalmente en desacuerdo a totalmente de acuerdo) para un total de 9 variables que reflejan opiniones sobre el tipo de cultura de investigación percibida en la universidad. La pregunta es la siguiente:

La cultura de la Universidad es:

Rígida	1	2	3	4	5	Flexible
Abierta hacia el exterior	1	2	3	4	5	Centrada en la propia Univ.
Individualista	1	2	3	4	5	Trabajo en equipo
Innovadora	1	2	3	4	5	Tradicional
Planificada	1	2	3	4	5	Imaginativa
Cooperación	1	2	3	4	5	Rivalidad
Resultados a largo plazo	1	2	3	4	5	Resultados a corto plazo
Valoración de la investig.	1	2	3	4	5	Indiferencia hacia la investig.
Investigación básica	1	2	3	4	5	Investigación Aplicada

Se dispone de información sobre estas variables para un total de 534 integrantes de la universidad encuadrados en 5 áreas: Ciencias exactas (exa-118 respuestas), Ingenierías (ing-103 respuestas), Ciencias médicas (med-79 respuestas), Ciencias sociales (soc-162 respuestas) y Humanidades (hum-72 respuestas). Además, se dispone de información de variables de caracterización como sexo, edad, categoría profesional (rango) y una binaria de participación ó no en proyectos de investigación.

Se pretende analizar la relación entre las variables (categóricas) por separado entre los grupos delimitados por las áreas de conocimiento para luego realizar un análisis global según el método propuesto en dos etapas.

4. Análisis de los datos

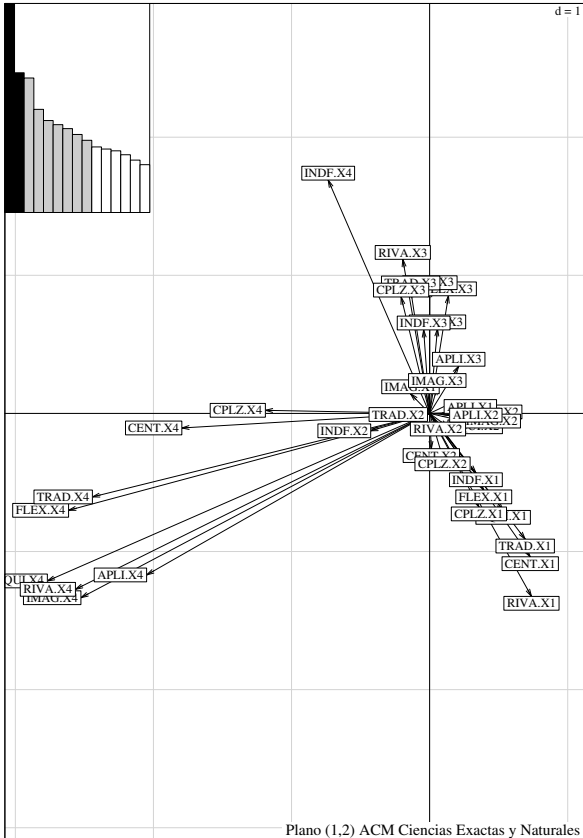
4.1. ACMs preliminares

Tras una inspección preliminar (ver cuadro 1), el número de efectivos resultantes de la categoría 1 (tot. en desacuerdo) es insignificante en relación al resto, incluso para algunas variables correspondientes al área de Humanidades no hay ninguno. El problema no es tan acusado en el caso de la otra categoría extrema (5) y se decide amalgamar las categorías 1 y 2 formando una nueva categoría 1, pasando las 3,4 y 5 a ser las 2, 3 y 4, respectivamente.

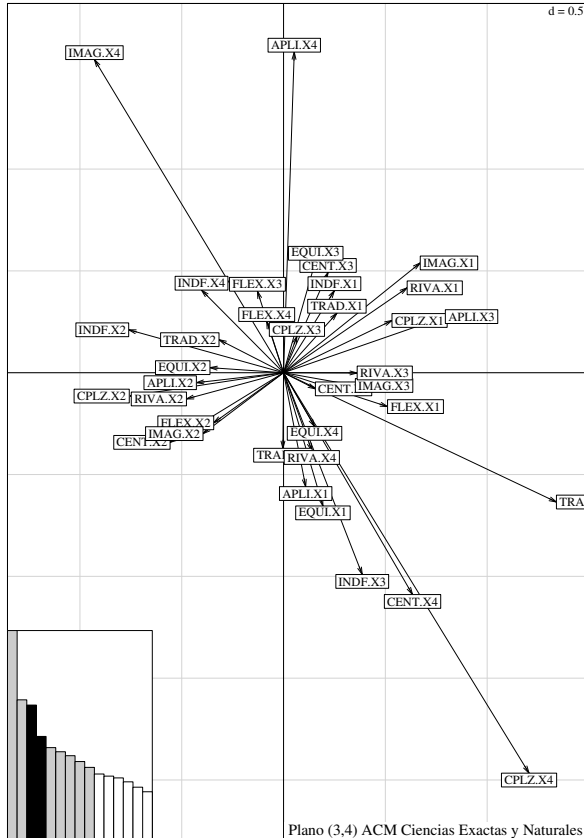
AREA	FLEX	CENT	EQUI	TRAD	IMAG	RIVA	CPLZ	INDF	APLI
exa	1: 2	1: 1	1: 5	1: 3	1: 1	1: 1	1: 7	1:18	1: 1
exa	2:34	2:19	2:27	2:26	2:20	2:21	2:27	2:35	2:29
exa	3:39	3:37	3:38	3:49	3:56	3:56	3:37	3:30	3:55
exa	4:32	4:51	4:40	4:34	4:36	4:35	4:41	4:30	4:25
exa	5:11	5:10	5: 8	5: 6	5: 5	5: 5	5: 6	5: 5	5: 8
ing	1: 2	1: 1	1: 2	1: 1	1: 1	1: 1	1: 2	1: 7	1: 1
ing	2:27	2:18	2:24	2:19	2:10	2:23	2:14	2:36	2:31
ing	3:30	3:28	3:33	3:43	3:43	3:38	3:35	3:36	3:43
ing	4:35	4:49	4:36	4:33	4:44	4:35	4:39	4:20	4:26
ing	5: 9	5: 7	5: 8	5: 7	5: 5	5: 6	5:13	5: 4	5: 2
med	1: 2	1: 1	1: 2	1: 1	1: 3	1: 2	1: 2	1: 4	1: 2
med	2:10	2: 9	2:13	2:13	2:13	2:12	2:15	2:17	2:16
med	3:23	3:25	3:22	3:27	3:45	3:32	3:25	3:31	3:34
med	4:27	4:35	4:27	4:25	4:15	4:22	4:22	4:18	4:20
med	5:17	5: 9	5:15	5:13	5: 3	5:11	5:15	5: 9	5: 7
soc	1: 4	1: 7	1:11	1: 8	1: 4	1: 8	1: 5	1:11	1: 8
soc	2:21	2:21	2:21	2:17	2:17	2:21	2:28	2:43	2:33
soc	3:52	3:54	3:39	3:61	3:65	3:54	3:57	3:71	3:69
soc	4:54	4:57	4:58	4:58	4:64	4:64	4:53	4:28	4:45
soc	5:31	5:23	5:33	5:18	5:12	5:15	5:19	5: 9	5: 7
hum	1: 0	1: 2	1: 0	1: 0	1: 0	1: 2	1: 5	1: 3	1: 3
hum	2:18	2:10	2:10	2:14	2: 2	2: 9	2:12	2:17	2:14
hum	3:21	3:30	3:21	3:30	3:42	3:34	3:26	3:25	3:39
hum	4:19	4:25	4:26	4:20	4:23	4:23	4:22	4:24	4:12
hum	5:14	5: 5	5:15	5: 8	5: 5	5: 4	5: 7	5: 3	5: 4

Cuadro 1: Número de efectivos original por variable y grupo.

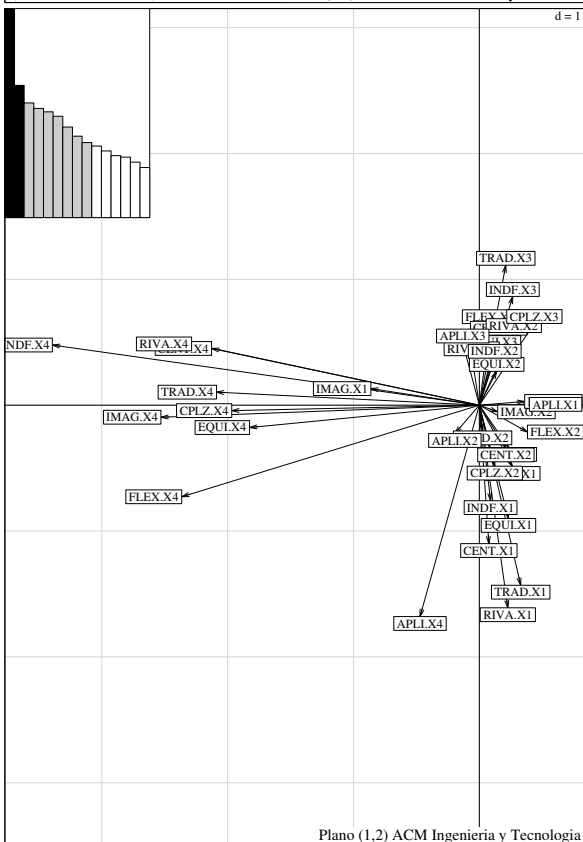
Como las variables son categóricas se examinan inicialmente las tablas individualmente mediante un ACM separado de cada una. A continuación se incluyen los planos (1,2) y (3,4) para las categorías o modalidades de las 9 cuestiones en las 5 áreas.



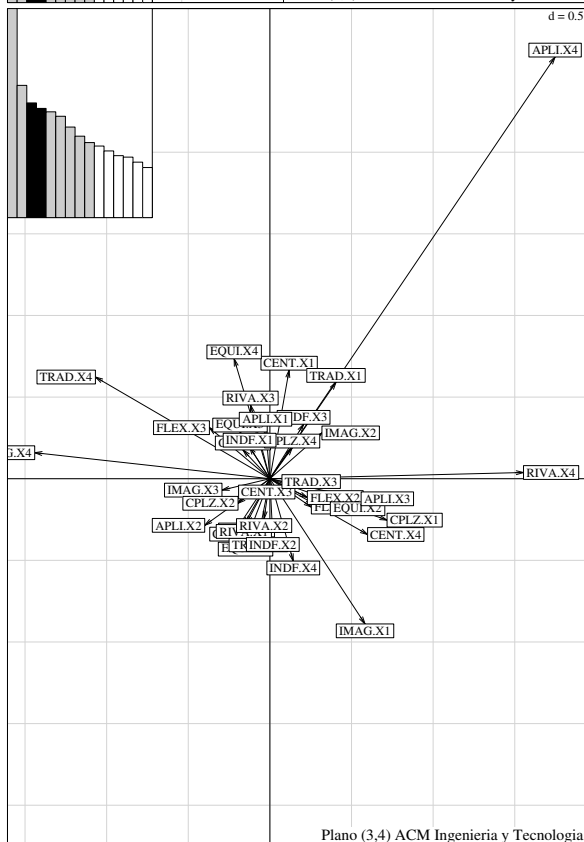
Plano (1,2) ACM Ciencias Exactas y Naturales



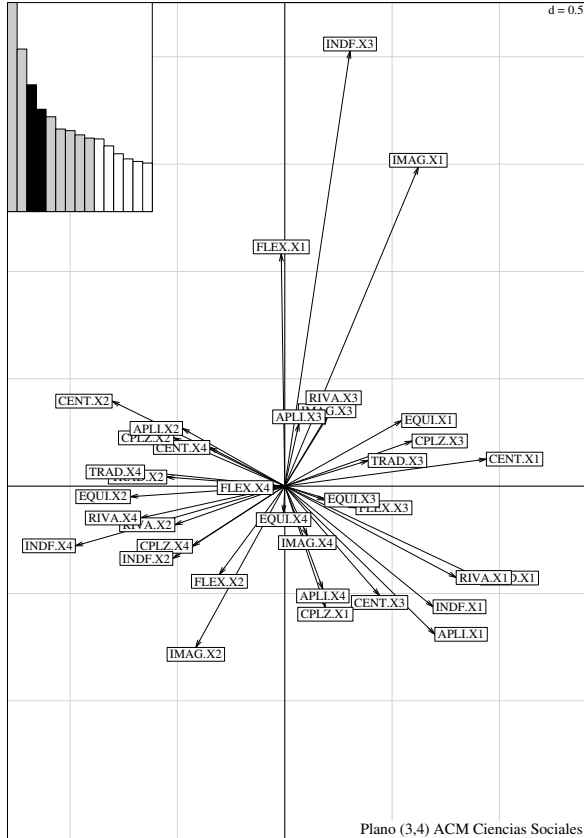
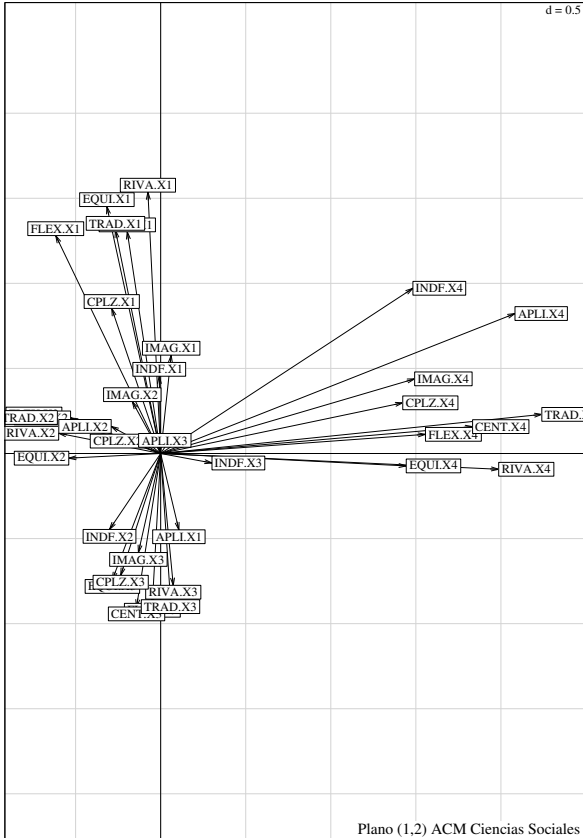
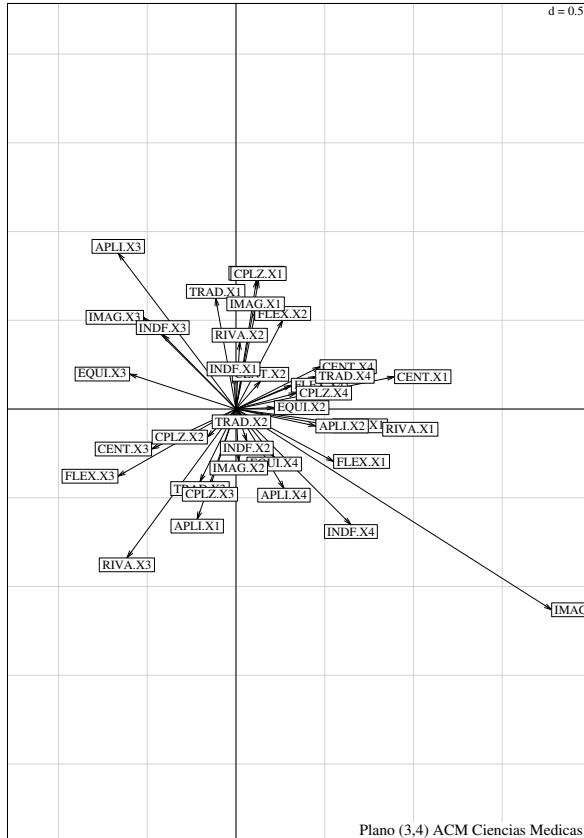
Plano (3,4) ACM Ciencias Exactas y Naturales

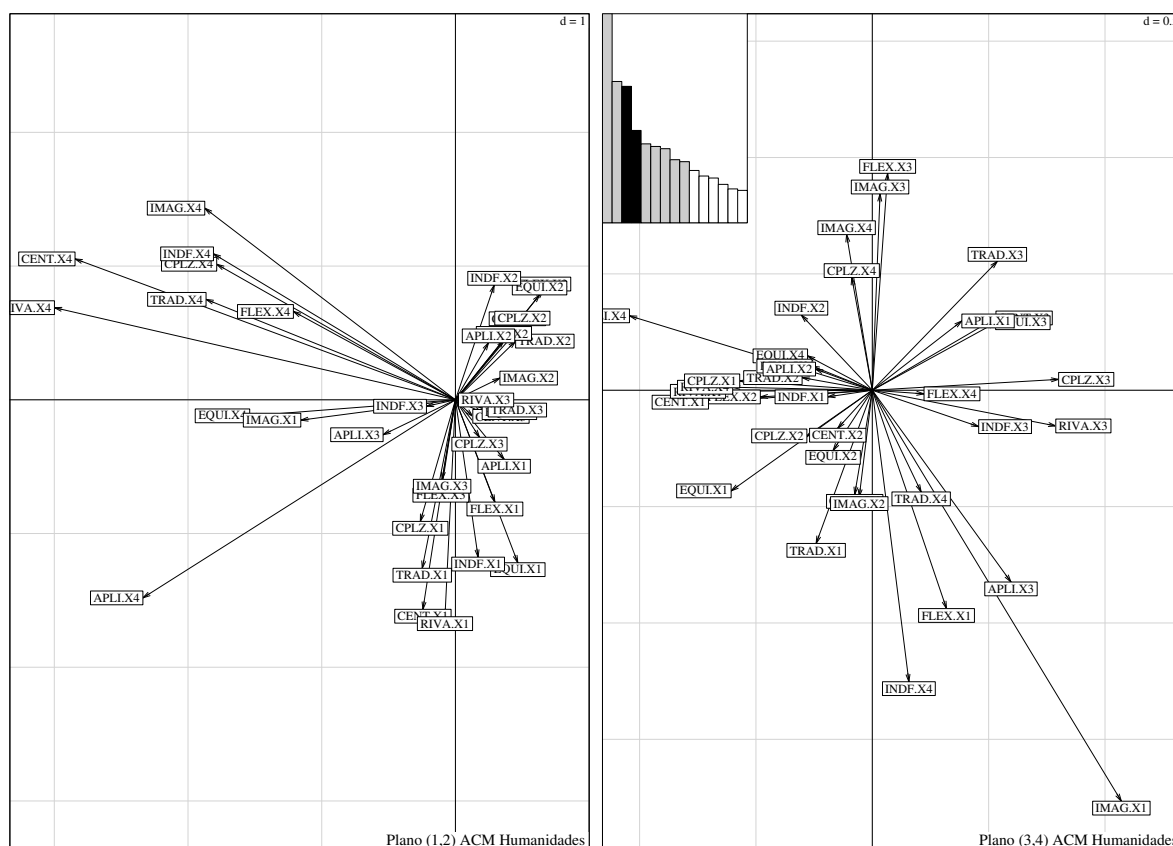


Plano (1,2) ACM Ingeniería y Tecnología



Plano (3,4) ACM Ingeniería y Tecnología





El primer factor (eje horizontal) de los 5 ACMs realizados es el mismo en todas ellos. Opone modalidad completamente de acuerdo (4) al resto de modalidades para todas las variables en casi todos los casos para las 5 áreas.

El segundo factor (eje vertical) opone, como característica general a todos los grupos excepto quizás el correspondiente a Ciencias médicas, la modalidad amalgamada completamente ó bastante en desacuerdo (1) a la modalidad bastante de acuerdo (3). En el caso del área de ciencias médicas (y en menor medida en el de Humanidades), el factor 3 tiene más que ver con esta interpretación que el 2. Si miramos a la matriz de correlaciones de los factores en el cuadro 2:

	1ex	2ex	3ex	1in	2in	3in	1me	2me	3me	1so	2so	3so	1hu	2hu	3hu
1ex	1.0	-0.3	-0.1	0.8	-0.1	-0.0	0.8	-0.3	-0.3	0.9	-0.1	-0.4	0.8	-0.5	0.1
2ex	-0.3	1.0	0.0	-0.1	-0.7	0.1	-0.1	-0.2	0.6	-0.3	-0.6	-0.1	-0.3	-0.2	-0.7
3ex	-0.1	0.0	1.0	-0.1	-0.0	0.2	-0.5	-0.1	-0.2	-0.3	-0.1	-0.3	-0.3	-0.3	0.3
1in	0.8	-0.1	-0.1	1.0	-0.1	0.1	0.8	-0.2	-0.5	0.9	0.1	-0.4	0.9	-0.5	0.0
2in	-0.1	-0.7	-0.0	-0.1	1.0	-0.1	-0.2	0.3	-0.4	0.0	0.7	0.3	0.0	0.5	0.6
3in	-0.0	0.1	0.2	0.1	-0.1	1.0	-0.1	0.1	-0.2	-0.1	-0.2	0.0	-0.3	-0.2	-0.2
1me	0.8	-0.1	-0.5	0.8	-0.2	-0.1	1.0	-0.1	-0.2	0.9	-0.1	-0.1	0.9	-0.2	-0.1
2me	-0.3	-0.2	-0.1	-0.2	0.3	0.1	-0.1	1.0	-0.3	-0.1	0.4	0.7	-0.1	0.6	0.0
3me	-0.3	0.6	-0.2	-0.5	-0.4	-0.2	-0.2	-0.3	1.0	-0.3	-0.7	0.1	-0.4	0.2	-0.5
1so	0.9	-0.3	-0.3	0.9	0.0	-0.1	0.9	-0.1	-0.3	1.0	0.0	-0.2	0.9	-0.3	0.1
2so	-0.1	-0.6	-0.1	0.1	0.7	-0.2	-0.1	0.4	-0.7	0.0	1.0	0.1	0.1	0.4	0.5
3so	-0.4	-0.1	-0.3	-0.4	0.3	0.0	-0.1	0.7	0.1	-0.2	0.1	1.0	-0.2	0.8	-0.2
1hu	0.8	-0.3	-0.3	0.9	0.0	-0.3	0.9	-0.1	-0.4	0.9	0.1	-0.2	1.0	-0.3	0.1
2hu	-0.5	-0.2	-0.3	-0.5	0.5	-0.2	-0.2	0.6	0.2	-0.3	0.4	0.8	-0.3	1.0	0.2
3hu	0.1	-0.7	0.3	0.0	0.6	-0.2	-0.1	0.0	-0.5	0.1	0.5	-0.2	0.1	0.2	1.0

Cuadro 2: Correlaciones entre 3 primeros factores ACMs individuales

puede verse que la correlación entre 2ex con 3me y 3hu es significativamente alta,

y mucho mayor que la existente con 2me y 2hu.

El tercer factor y sucesivos (no presentados aquí) presentan ya una correlación muy baja entre las distintas áreas. Esto indica que la estructura común de las 5 áreas de conocimiento de la universidad consideradas es únicamente bidimensional, siendo el resto específica de cada área. En cuanto a porcentajes de inercia que representa dicha estructura bidimensional, éste oscila entre el 19.6 % del área de ingenierías y el 25.2 % del área de humanidades (cuadro 3), que es un buen porcentaje para un ACM de 9 variables de 4 categorías cada una; es decir, 36 variables indicadoras.

Area	exa	ing	med	soc	hum
Pct.	20.5	19.6	22.3	21.4	25.2

Cuadro 3: Porcentajes de inercia recogidos por dos primeros factores de ACM por áreas.

4.2. AFM

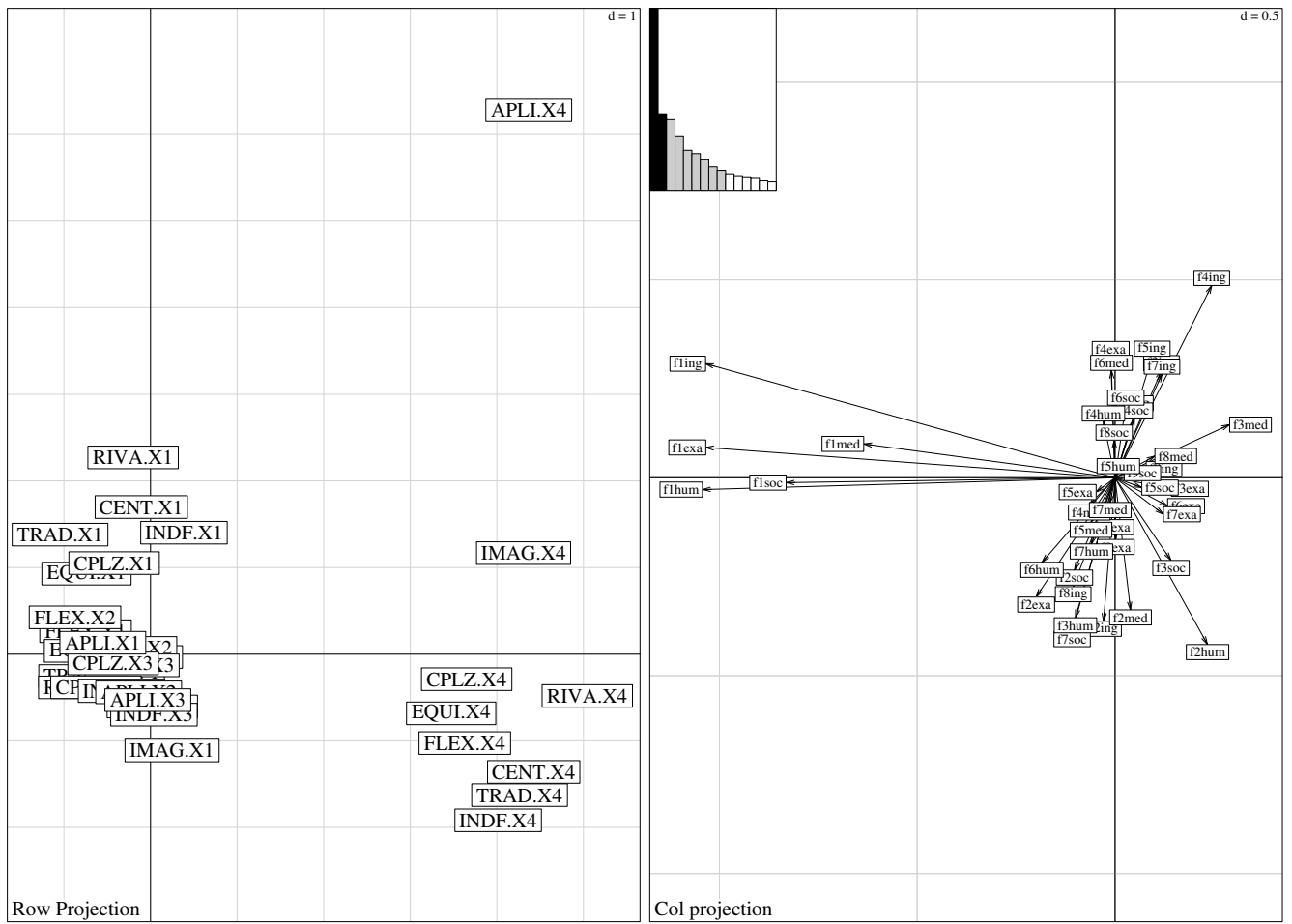
El Análisis Factorial Múltiple es una técnica que parte de un análisis factorial previo (ya sea componentes Principales o Correspondencias Múltiples) de las J tablas yuxtapuestas. Sin embargo, el AFM no es simplemente eso; tal análisis yuxtapuesto realmente puede llegar a perder la diferenciación entre los grupos desde el momento en que la mayor variabilidad en un sólo grupo podría suponer que los primeros factores globales reflejen principalmente el comportamiento de ese único grupo. Para evitar esto, el AFM realiza un Análisis de Componentes Principales yuxtapuesto, pero previa ponderación para cada tabla del inverso del primer valor singular correspondiente a un análisis parcial sobre ella. Con esto se evita que el análisis pueda estar determinado por una única tabla, al menos en cuanto al primer eje.

Para construir la tabla de análisis del AFM hemos seleccionado factores que proporcionan algo menos del 60 % de inercia de cada tabla original. Dichos factores recogen los siguientes porcentajes de inercia (cuadro 4):

Area	exa	ing	med	soc	hum
Factores	9	9	8	9	7
Inercia (%)	59	59	59	58	58

Cuadro 4: Porcentajes de inercia recogida por factores de ACM seleccionados en la primera etapa.

En las figuras 3 a 6 están los resultados correspondientes al AFM, añadiéndose en la primera las posiciones relativas de las 5 áreas en los dos primeros ejes globales del AFM de los factores principales considerados para cada área según la tabla 4 y un histograma de los valores propios globales.



Link

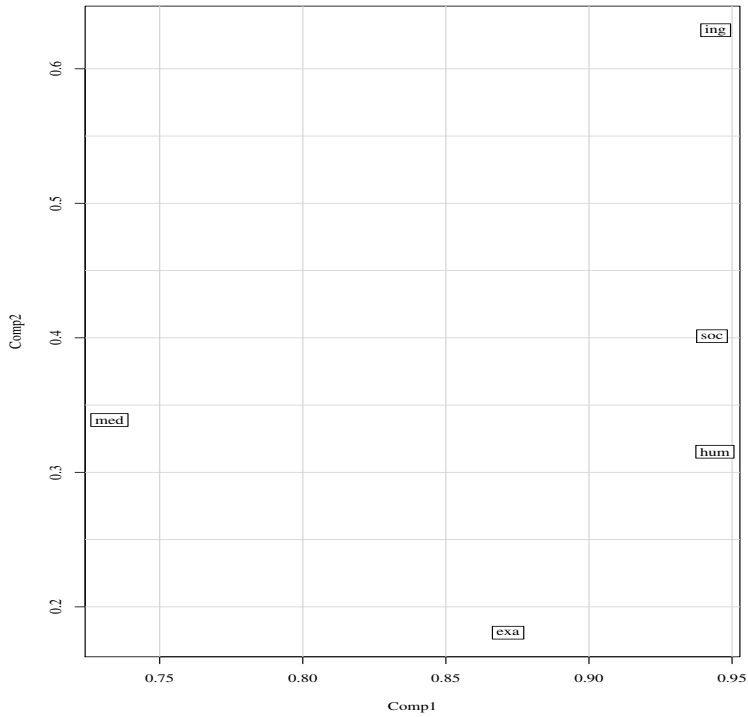
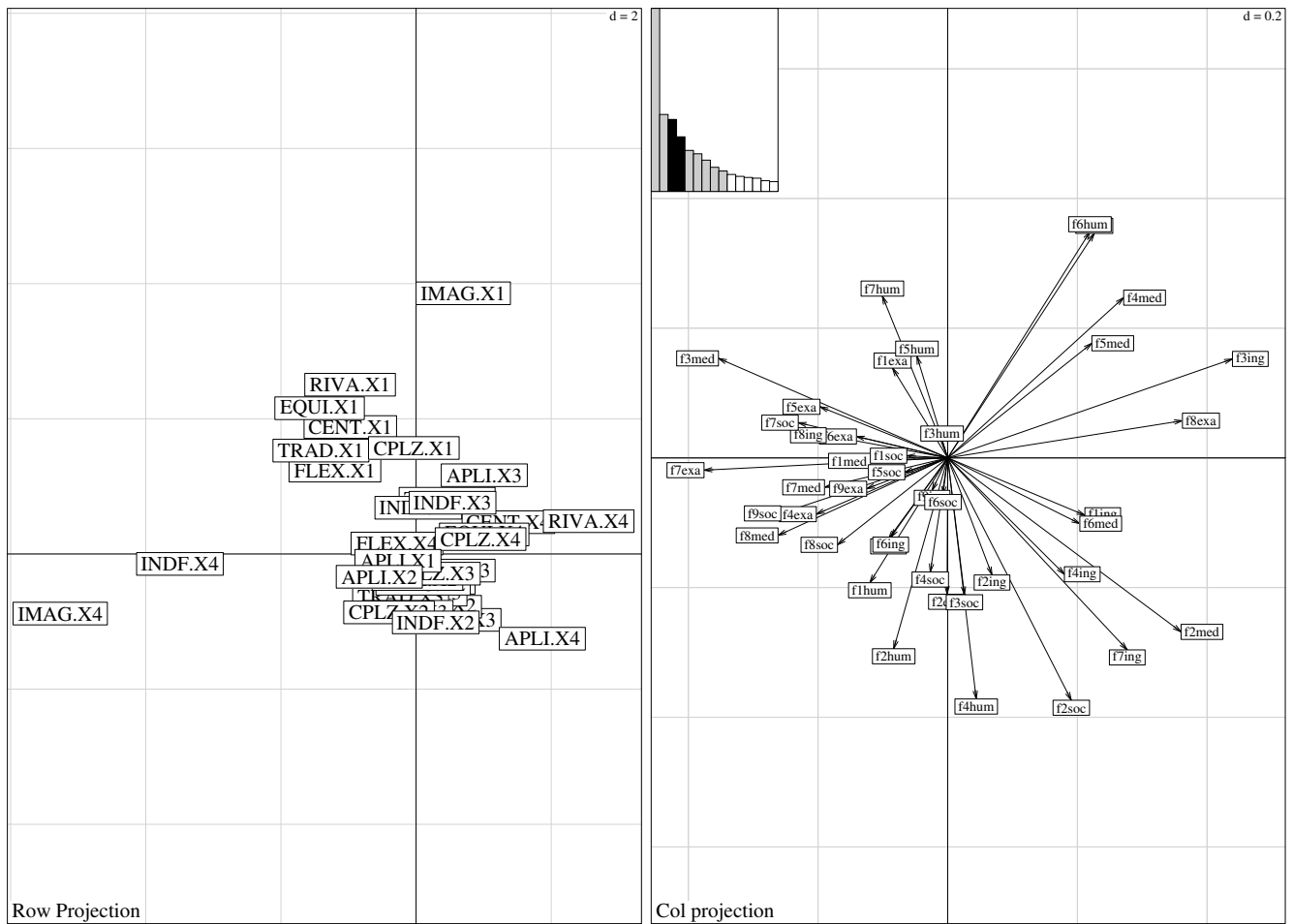


Figura 3: Plano (1,2) AFM
11



Link

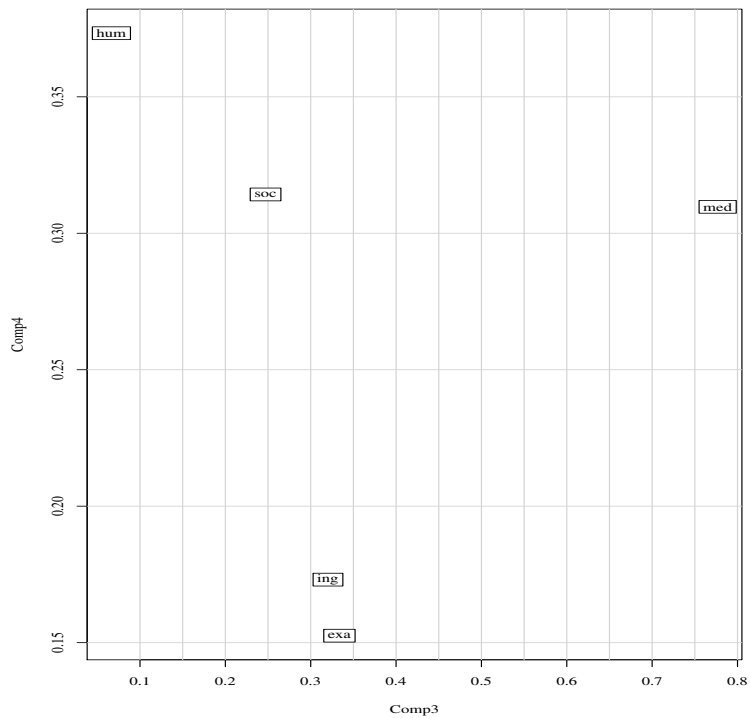


Figura 4: Plano (3,4) AFM

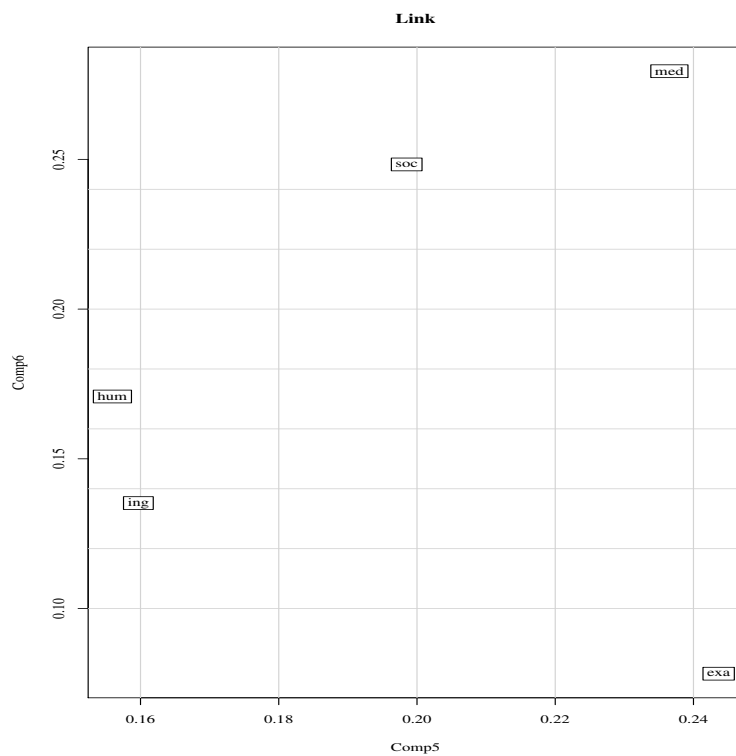
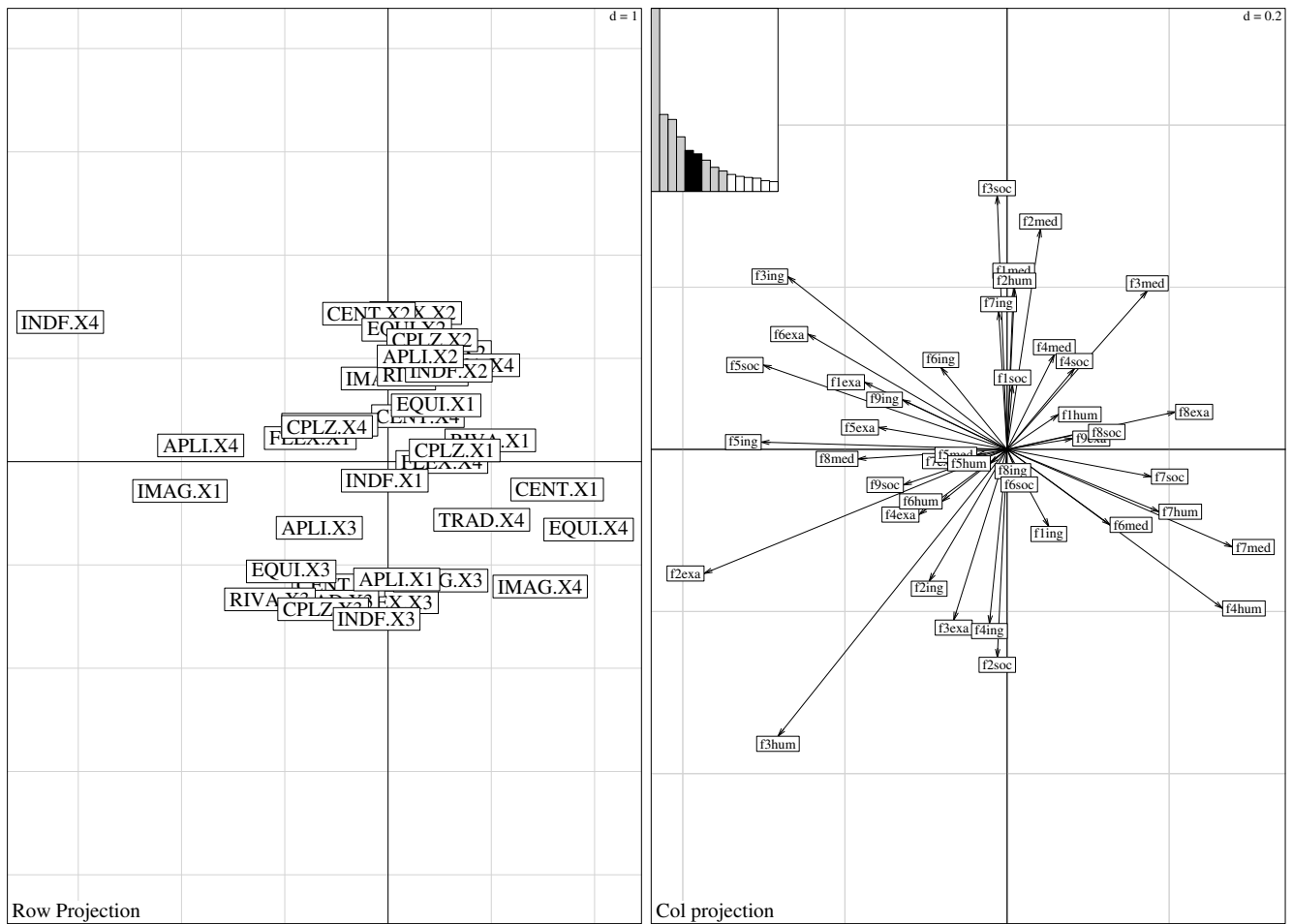


Figura 5: Plano (5,6) AFM
13

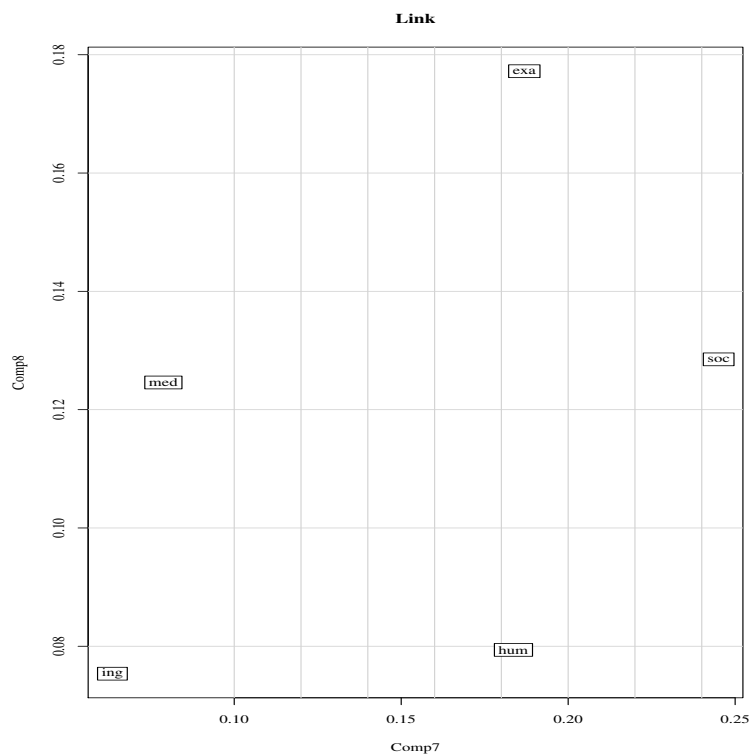
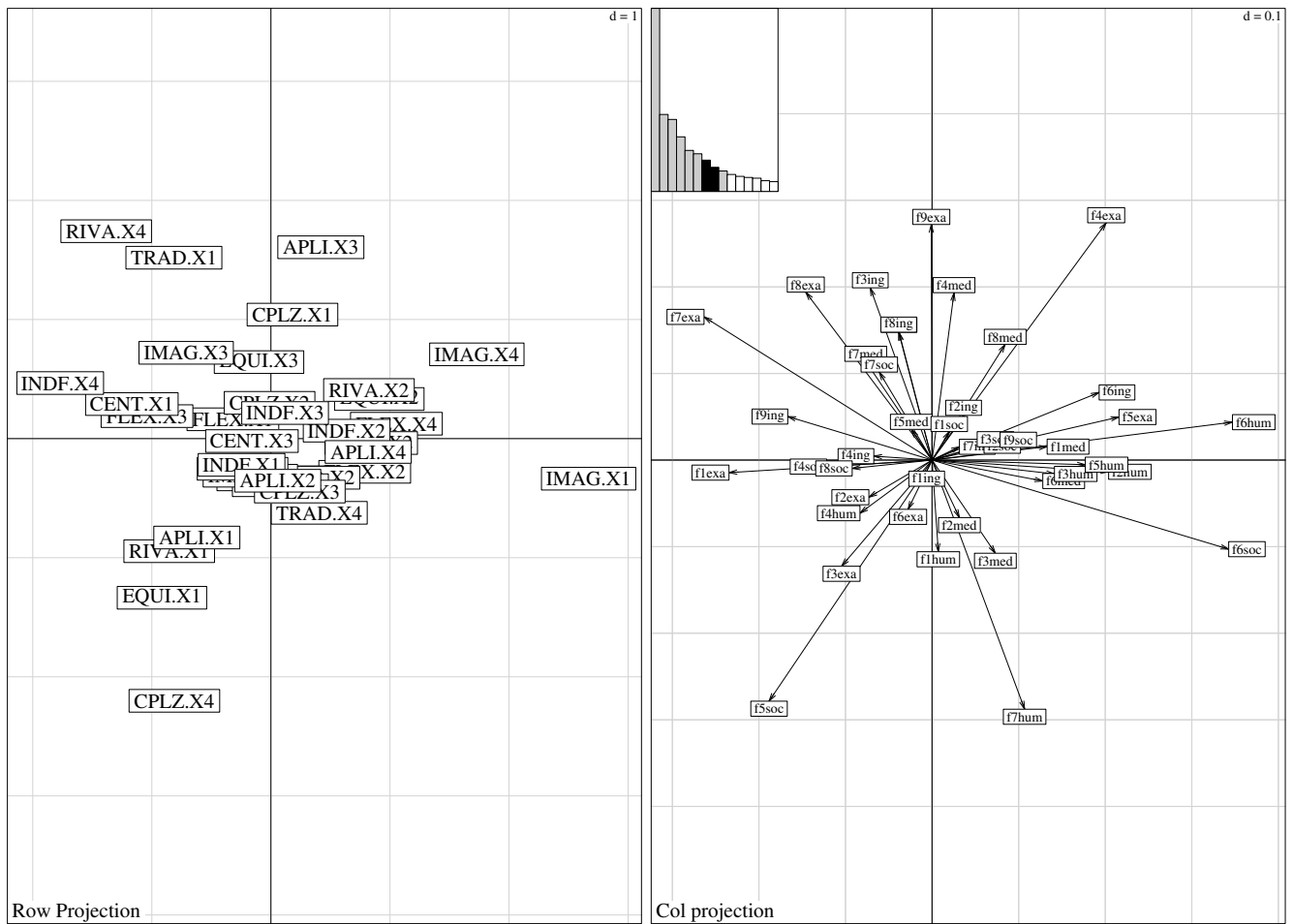


Figura 6: Plano (7,8) AFM

El primer eje global, eje horizontal en las gráficas, refleja casi miméticamente los primeros ejes de los ACMs, fundamentalmente las categorías tot. de acuerdo (*.X4) de todas las variables originales, en oposición al resto. El segundo panel de la figura 3 refleja que en todas las áreas hay pocas respuestas extremas de este tipo, y que donde proporcionalmente más hay es en Ciencias médicas (y después en Ciencias sociales), que tiene una componente negativa menor. Por la posición (en el 2º cuadrante) que ocupan las respuestas poco o nada de acuerdo (modalidad *.X1), el área que más respuestas tiene de este estilo es Ingenierías (ver paneles 1 y 2). En todo caso, predominan las respuestas medias y medias-altas a todas las preguntas, que es hacia donde se proyectan los factores primeros de los ACMs y las modalidades 2 y 3.

El segundo factor, eje vertical en las gráficas, tiene, sin embargo, como principal causante a la oposición entre encuestados que opinan que la cultura de la universidad está orientada a la investigación aplicada (ver APLI.X4 en la parte superior del primer panel de la figura 3) frente a los que opinan que lo está a la indiferencia hacia la investigación (ver INDF.X4 en la parte inferior del mismo panel). En la dirección asociada a la cultura orientada a inv. aplicada, están también, con menor representatividad, opiniones hacia cooperación o escasa rivalidad (RIVA.X1), resultados a L/P (CPLZ.X1), abierta (CENT.X1), innovadora o poco tradicional (TRAD.X1) y valoración de la investigación (INDF.X1). En la dirección asociada a indiferencia de investigación están también opiniones de tradicional (TRAD.X4) y, curiosamente, flexible (FLEX.X4). Este factor parece ser claramente un factor de buena consideración de la investigación dentro de la universidad, en particular de la investigación aplicada. Hay que recalcar que APLI.X4, investigación aplicada, es una característica general de todas las áreas, pero de forma particularmente predominante en el área de ingenierías (al que corresponde el punto más alto de todo el plano). El punto arriba y a la derecha del baricentro corresponde a Ciencias sociales, mientras que los de debajo del baricentro son, de mayor a menor, Ciencias exactas, médicas y humanidades, que es la que menos respuestas tiene en favor de la investigación aplicada. El detalle sobre las coordenadas de las variables (que en este caso son los individuos parciales, en la terminología del AFM) están en el apéndice A.

El tercer factor (en el eje horizontal de la figura 4) se oponen cultura imaginativa (IMAG.X4) y la indiferencia hacia la investigación (INDF.X4) en el lado izquierdo a la rivalidad (RIVA.X4) y a investigación aplicada (APLI.X4) en el lado derecho. Hay que destacar que (en el primer panel) la posición de IMAG.X4 se debe sobre todo al punto parcial correspondiente al área de ciencias médicas, que también es el más influyente en la posición de RIVA.X4. Es, por tanto, el área de Ciencias Médicas quien más contribuye en la formación de este factor.

El cuarto factor (en el eje vertical de la figura 4) opone principalmente cultura planificada (IMAG.X1) en la parte superior a investigación aplicada (APLI.X4) en la parte inferior, principalmente. La posición de IMAG.X1 se debe, fundamentalmente, al área de Humanidades, que es el principal contribuyente a la formación de este factor, y donde más claramente se manifiesta la cultura planificada. Por su parte, APLI.X4 se sitúa en la parte negativa debido al área de Ingenierías, donde más claramente se está a favor de la investigación aplicada (nuevamente) y menos a la cultura planificada.

A partir de aquí, los factores del ACM empiezan a ser repetitivos o a carecer de sentido. El factor quinto (eje horizontal, figura 5) opone trabajo en equipo (EQUI.X4), cultura abierta (CENT.X1) e imaginativa (IMAG.X4) en la parte de-

recha a indiferencia a la investigación (INDF.X4), cultura planificada (IMAG.X1) e investigación aplicada (APLI.X4), asociación de esta última que no parece tener ya mucho sentido, en el lado izquierdo. En el derecho, los puntos parciales más importantes corresponden a Ciencias exactas (RIVA.X4 y EQUI.X4) y a Ciencias médicas (CENT.X1) lo que no es ya un factor de opinión general, sino que parece responder a ciertos grupos concretos, puesto que los baricentros de esas variables no están en esa misma zona. El factor sexto opone modalidades intermedias (.X2) en la parte positiva a un grupo de modalidades de escasa caracterización, y donde nuevamente son determinantes algunos puntos parciales que ni siquiera corresponden a dichas modalidades (por ejemplo, INDF.X4 de Ciencias Médicas, abajo e IMAG.X1 de Ciencias Sociales, arriba). Se incluyen también el plano de los factores 7 y 8 en la figura 6.

4.2.1. AFM: Comentarios

El análisis en dos etapas culminado en un AFM de factores revela un primer factor *tamaño*, muy frecuente en análisis factoriales sobre datos de encuestas. Este es común a todas las áreas igual que el segundo, un factor de valoración *positiva* de la investigación, como algo fundamentalmente aplicado y, a mayor distancia, cooperativo, en búsqueda de resultados a L/P, en base a una cultura abierta al exterior, innovador y menos en negativo como indiferencia hacia ella e investigación tradicional, siendo esto más fuerte en las áreas de ingeniería y ciencias sociales y menos en ciencias exactas, médicas y humanidades. Esto configura una estructura en la que predominan este tipo de opiniones sobre la investigación en la universidad, a nivel general. Es una opinión que encuadra en el tipo de cultura organizativa de **innovación**, fundamentalmente, y un poco menos de mercado y de clan, siguiendo el modelo de valores y competencia de Cameron y Quinn.

El resto de factores, representan opiniones concretas de unas áreas u otras. El tercero opone algo que no tiene demasiado sentido como que es imaginativa (principalmente debido a opiniones de profesores de ciencias médicas) frente a una cultura aplicada (debido a las ingenierías) pero de rivalidad, de difícil clasificación en el modelo. El cuarto, sin embargo, es más interpretable, oponiendo la cultura de investigación *planificada*, opiniones principalmente procedentes del sector de humanidades, acompañada a más distancia de cultura rígida, individualista, pero cooperativa, innovadora, a largo plazo y abierta, por lo que parece encajar con bastante claridad en la cultura **jerárquica** del modelo. El quinto parece ser una mezcla entre mercado y clan, debido a opiniones en las áreas de ciencias médicas y sociales, mientras que el sexto ya es una mezcla poco clara de opiniones.

4.3. STATIS

Los factores obtenidos a través de los ACMs de cada grupo pueden ser utilizados, extrayendo las coordenadas de las modalidades sobre esos factores como base para el análisis mediante el procedimiento Statis desarrollado en (Dazy & Le Barzic 1996) ó (Lavit 1988). Fundamentalmente, el método Statis es una técnica factorial múltiple que tiene tres modalidades:

- a) Análisis Triádico de Jaffrenou. Requiere que el número de individuos y de variables contenidos en cada una de las matrices a analizar sea el mismo $I_1 =$

$\dots = I_J \quad K_1 = \dots = K_J$. Es demasiado restrictivo, por ello poco utilizado y aquí inaplicable.

- b) Statis. Calcula las matrices de productos escalares entre individuos para las J tablas, y posteriormente calcula una matriz compromiso que resulta ser la que maximiza la suma de las correlaciones² al cuadrado entre esa matriz compromiso y las matrices de productos escalares.
- c) Statis Dual. Idéntico al anterior en cuanto a la búsqueda de una matriz compromiso, pero, en este caso, las matrices objeto son matrices de productos escalares *entre variables*, típicamente matrices de correlación, y por tanto, adecuado al caso en que el número de variables es idéntico entre las tablas, pudiendo diferir en cuanto al número de individuos.

Dado que, en todo caso, el número de modalidades total ha de ser el mismo y éste se ha convertido en el número de individuos, el método elegido es Statis, aunque en el caso de que se hayan extraído inicialmente el mismo número de factores de las variables, también podría haberse usado Statis Dual.

En primer lugar, en el cuadro 5 están los coeficientes de asociación RV (Robert & Escoufier 1976), calculados a partir de los productos escalares entre las tablas, entendidos como una medida de asociación entre ellas similar a un coeficiente de correlación. Puede apreciarse como los coeficientes están entre 0.53 y 0.69, con lo

	exa	ing	med	soc	hum
exa	1.00	0.55	0.54	0.65	0.58
ing	0.55	1.00	0.53	0.66	0.66
med	0.54	0.53	1.00	0.58	0.53
soc	0.65	0.66	0.58	1.00	0.69
hum	0.58	0.66	0.53	0.69	1.00

Cuadro 5: Coeficientes RV entre matrices de factores

que las 5 áreas de conocimiento son en buena parte parecidas, especialmente ciencias exactas y sociales por un lado, e ingenierías, ciencias sociales y ciencias humanas por otro, teniendo entre ellas coeficientes mayores que 0.60.

Esto queda corroborado por el gráfico 7 de la interestructura (de una interpretación parecida a la del gráfico 3 del AFM) en el que las áreas se proyectan sobre los dos primeros ejes de un ACP de la matriz de productos escalares entre las 5 áreas, donde se ve más claramente la relación de ciencias humanas e ingeniería y de éstas con ciencias sociales, por un lado, y de ciencias sociales por otro lado con ciencias exactas. Ciencias médicas es la más diferente de todas, y lo es, en particular a través del segundo factor, de ciencias humanas e ingeniería. Por tanto, la matriz compromiso no es del todo un mal reflejo de la estructura común de las tablas.

²Medidas a través del coeficiente RV de Escoufier (Robert & Escoufier 1976).

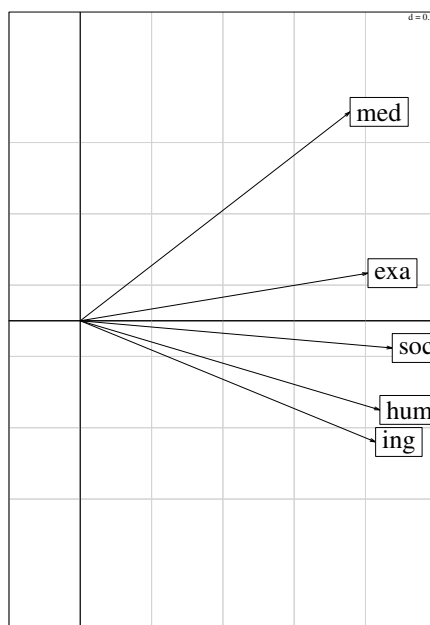


Figura 7: Interestructura Statis: ACP de la matriz 5×5 de productos escalares

Los pesos de las 5 áreas (de sus matrices de productos escalares) vienen dados en el cuadro 6 los cuales reflejan, por su menor peso, la relativa lejanía del área de

exa	ing	med	soc	hum
0.44	0.45	0.42	0.48	0.46

Cuadro 6: Pesos de las matrices de productos escalares entre individuos

ciencias médicas y permiten el cálculo de la matriz compromiso.

La figura 8 contiene las proyecciones de las modalidades de las variables originales (lo que se conoce en la terminología Statis como el estudio de las trayectorias de los individuos) sobre los planos factoriales (1,2), (3,4), (5,6) y (7,8) de la matriz compromiso entendida como la matriz de mayor correlación con las matrices originales. La representación es muy similar a la obtenida en el AFM por lo que no añadimos nada a la interpretación allí hecha.

En el apéndice B están las figuras 13-16 que contienen las proyecciones de los factores principales de los ACMs iniciales en los planos principales extraídos a partir de la matriz compromiso. Nuevamente estos factores ocupan posiciones parecidas que aquéllas del AFM.

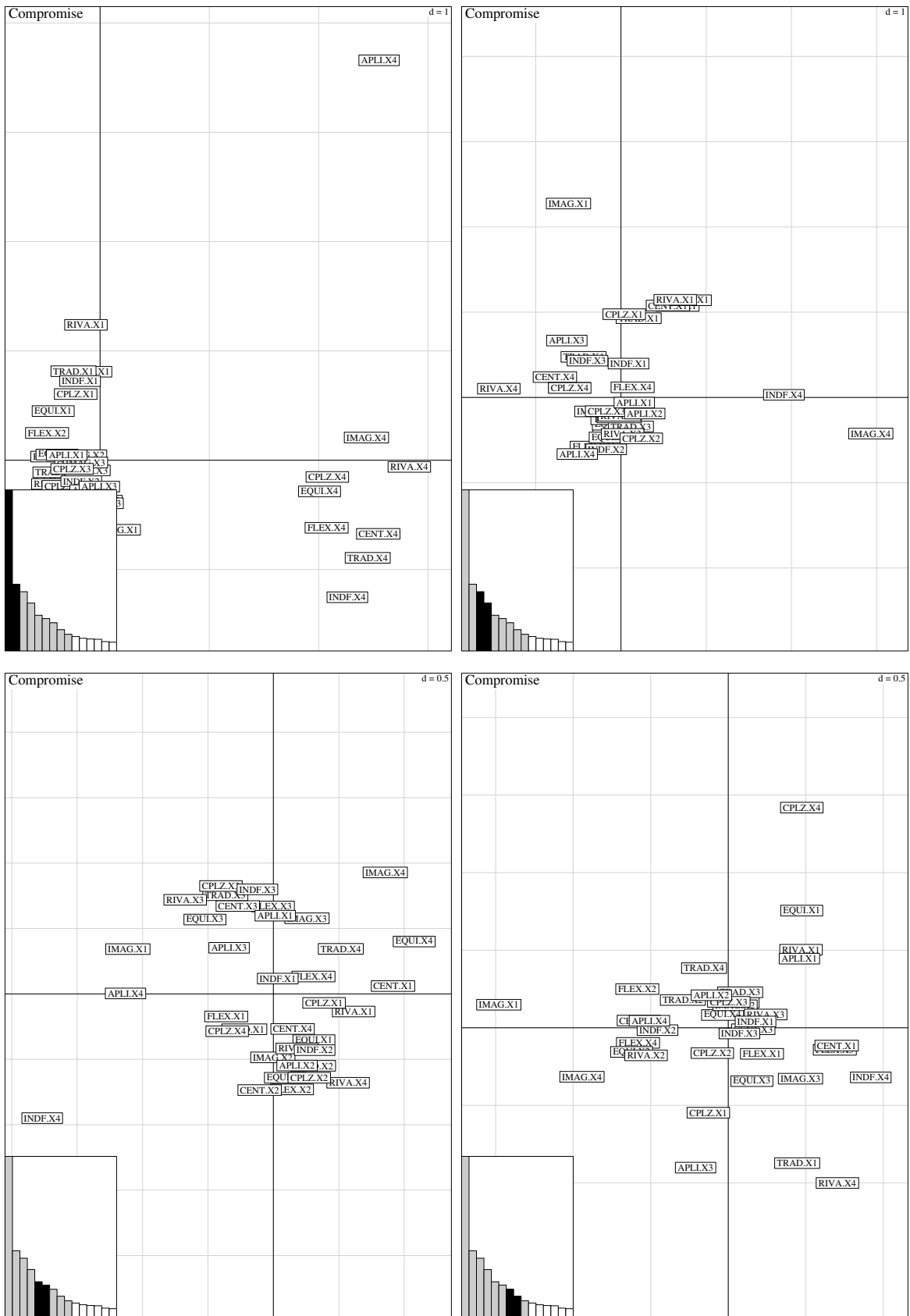


Figura 8: Stasis: Proyecciones modalidades en planos principales (1,2) (3,4) (5,6) y (7,8) de izquierda a derecha y de arriba a abajo.

4.4. Comentarios: STATIS

STATIS pone de manifiesto la misma relación entre las variables originales que teníamos ya en AFM. Proporciona mayor énfasis en la relación global entre las tablas, a partir de medidas fácilmente interpretables como productos escalares o coeficientes RV. Por lo demás, basta con tener en cuenta la tendencia, por la propia característica del método, a dar un menor peso a aquellas tablas que estén más alejadas del conjunto, a la inversa que el AFM, lo que, sin embargo, no hace que los resultados difieran sustancialmente y, pueden considerarse, por ello, dos métodos complementarios.

5. Estrategia de simulación

La validación de los resultados puede hacerse mediante un procedimiento de simulación. Una estrategia posible es el uso de bootstrap siguiendo, por ejemplo, a (Holmes 1989). La estrategia que se va a exponer aquí no sigue esta metodología, sino que es un estudio de Monte Carlo que recoge y modifica una idea de (Markus 1994).

Se construyen artificialmente, al igual que en el análisis realizado sobre la cultura de la universidad, $K = 5$ tablas de tamaños $N_i \quad i = 1, \dots, K$ iguales a los disponibles en las tablas de datos correspondientes a las áreas de conocimiento, es decir, $N_1 = 118, N_2 = 103, N_3 = 79, N_4 = 162, N_5 = 72$ y conteniendo cada una de ellas $K_i = 9 \forall i$ variables categóricas medidas en una escala de 4 puntos (hasta formar un total de 36 modalidades). El proceso es como sigue:

1. Se supone que la dimensionalidad subyacente es de orden dos ($D=2$). Así, se generan $K \sum_i (K_i + D) = 55$ muestras independientes de tamaño $\max_i N_i = 162$ con distribución $N(0, 1)$. Estas muestras se denotan $\nu_1^{(i)}, \dots, \nu_{K_i}^{(i)}, f_1^{(i)}, f_{2(=D)}^{(i)}$ $i = 1, \dots, K (= 5)$.
2. Si $F_1 = (f_1^{(1)} \dots f_1^{(K)})$ es la matriz que contiene futuros factores de primer orden subyacentes para cada una de las K tablas y $F_2 = (f_2^{(1)} \dots f_2^{(K)})$ una matriz similar pero de factores de segundo orden, y Σ_1, Σ_2 son dos matrices simétricas $K \times K$ de elementos $\sigma_{ll}^{(d)} = 1$ y $-1 < \sigma_{kl}^{(d)} < 1, k \neq l, d = 1, 2$, entonces los productos de matrices $F_1 \Sigma_1^{1/2}$ y $F_2 \Sigma_2^{1/2}$ hacen que $f_d^{(l)}$ y $f_d^{(k)}$ estén correlacionados con coeficiente igual a $\sigma_{kl}^{(d)}$. Esto permite la existencia de correlación entre las K tablas.
3. Se construyen las variables

$$z_j^{(i)} = \alpha \nu_j^{(i)} + \beta_{1j} f_1^{(i)} + \beta_{2j} f_2^{(i)} \quad j = 1, \dots, K_i, i = 1, \dots, K \quad (1)$$

con la restricción, impuesta en (Markus 1994), de que $\alpha^2 + \sqrt{\beta_{1j}^2 + \beta_{2j}^2} = 1$.

4. Se discretizan las variables construidas $z_j^{(i)}$ en 4 clases, tratando de preservar la igualdad de tamaño entre clases y se eliminan aleatoriamente para cada tabla i el número de observaciones sobrante para que las K tengan finalmente el número de filas que les corresponde, K_i .

El proceso requiere la elección de los parámetros $\alpha, \beta_{1j}, \beta_{2j} (j = 1, \dots, K_i)$ y $\sigma_{kl}^{(d)}$. La propuesta es escoger los valores de $\sigma_{kl}^{(d)}$ a partir de la tabla 2, donde aparecen las correlaciones obtenidas entre los factores de los ACMs iniciales, el valor de α y β_{1j} y β_{2j} a partir de la inercia residual obtenible del cuadro (3) y de las coordenadas de las variables en los ACMs (figuras de la sección 4.1). Posteriormente puede replicarse un número arbitrario de veces el procedimiento y observar la estabilidad de los ejes, coeficientes *RV* o tasas de inercia, como por ejemplo en (Modroño, Fernández & Landaluce 2001), donde un procedimiento similar es aplicado a otro conjunto de datos con resultados de simulación conciliadores.

Sin embargo, en este caso, dicho procedimiento no parece resultar adecuado. La elección de β_{2j} resulta inapropiada por la no linealidad de las variables en el segundo factor, visible ya en los ACMs (donde las modalidades de la mayoría de las variables no se representan sobre el eje de forma ordenada, sino en el orden 1,2,4,3, con lo cual no hay una relación lineal entre las variables y el segundo eje). Parece mejor modificar la ecuación (1) de forma que la relación entre $z_j^{(i)}$ y $f_2^{(i)}$ sea, por ejemplo, cuadrática (*e.g.*, $f_2^{(i)} = [f_1^{(i)}]^2$).

6. Comentarios Finales

Este artículo muestra cómo puede analizarse un conjunto de variables de datos categóricos divididos en grupos de diferente tamaño mediante técnicas factoriales de tablas múltiples, utilizando para ello toda la información disponible en la generación de los ejes de inercia máxima. La ventaja de este método es que permite diferenciar claramente la relación común entre los grupos de otras relaciones internas de los grupos considerados de una manera equilibrada, según las características del Análisis Factorial Múltiple y de Statis. La simulación propuesta creemos que confirmará la idoneidad del análisis como hace con otros tipos de datos más *lineales* y será objeto de análisis posteriores.

En cuanto al conjunto de datos analizado, conviene señalar que hay una opinión generalizada, entre el conjunto de docentes de la universidad analizado, de que existe una cultura de investigación principalmente aplicada y también innovadora, abierta, cooperativa y a largo plazo que encaja en los tipos de cultura innovadora, principalmente, y también de mercado y clan definidas en el modelo propuesto. Otras percepciones (ó percepciones con peso distinto entre esos 3 tipos de cultura), no existen de forma importante a nivel general, aunque sí aparecen, de forma secundaria, a nivel de áreas particulares como es el caso de la cultura jerárquica en el área de humanidades.

A. Individuos parciales de las 5 áreas en los ejes del AFM

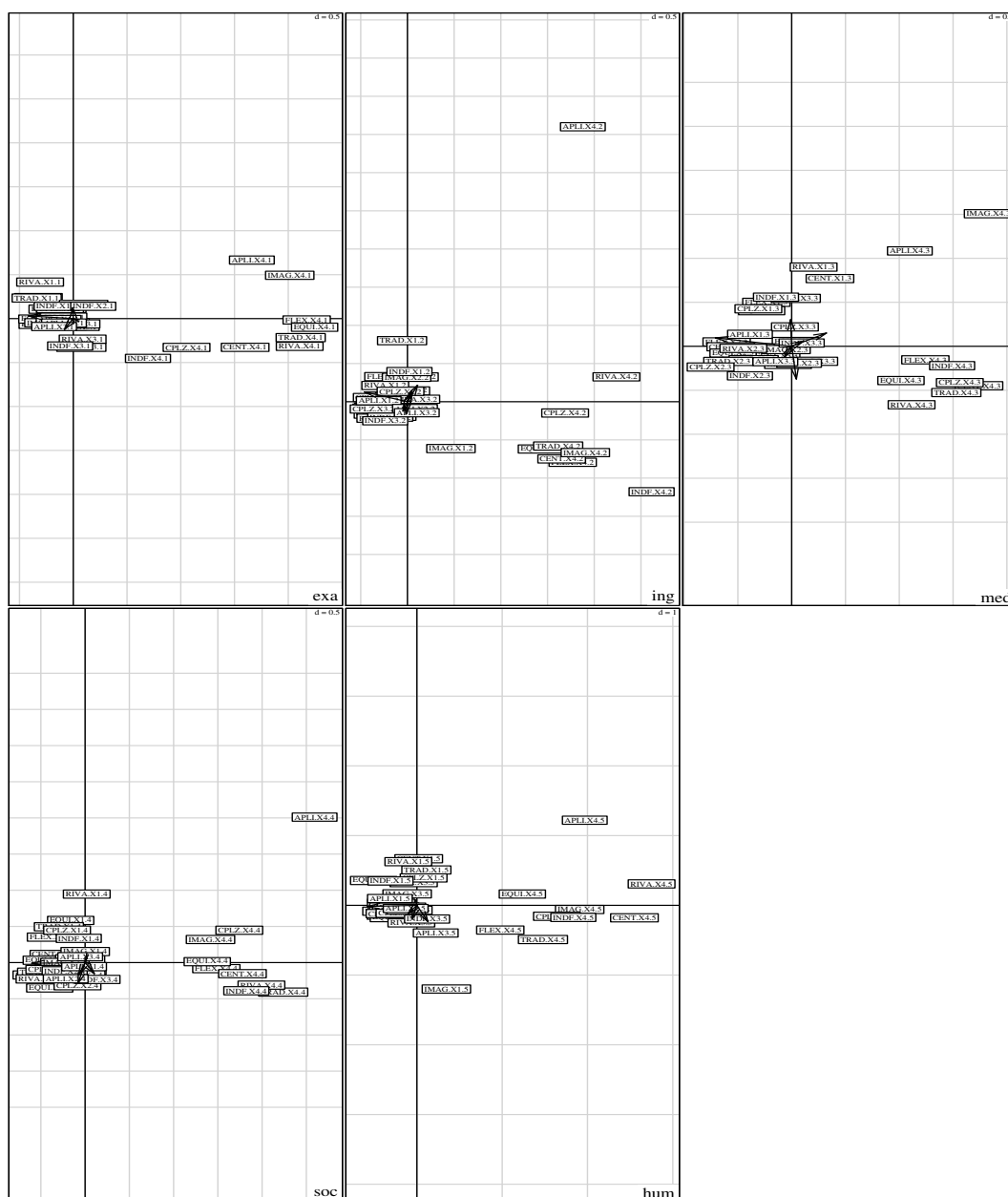


Figura 9: Individuos parciales en el plano (1,2) AFM

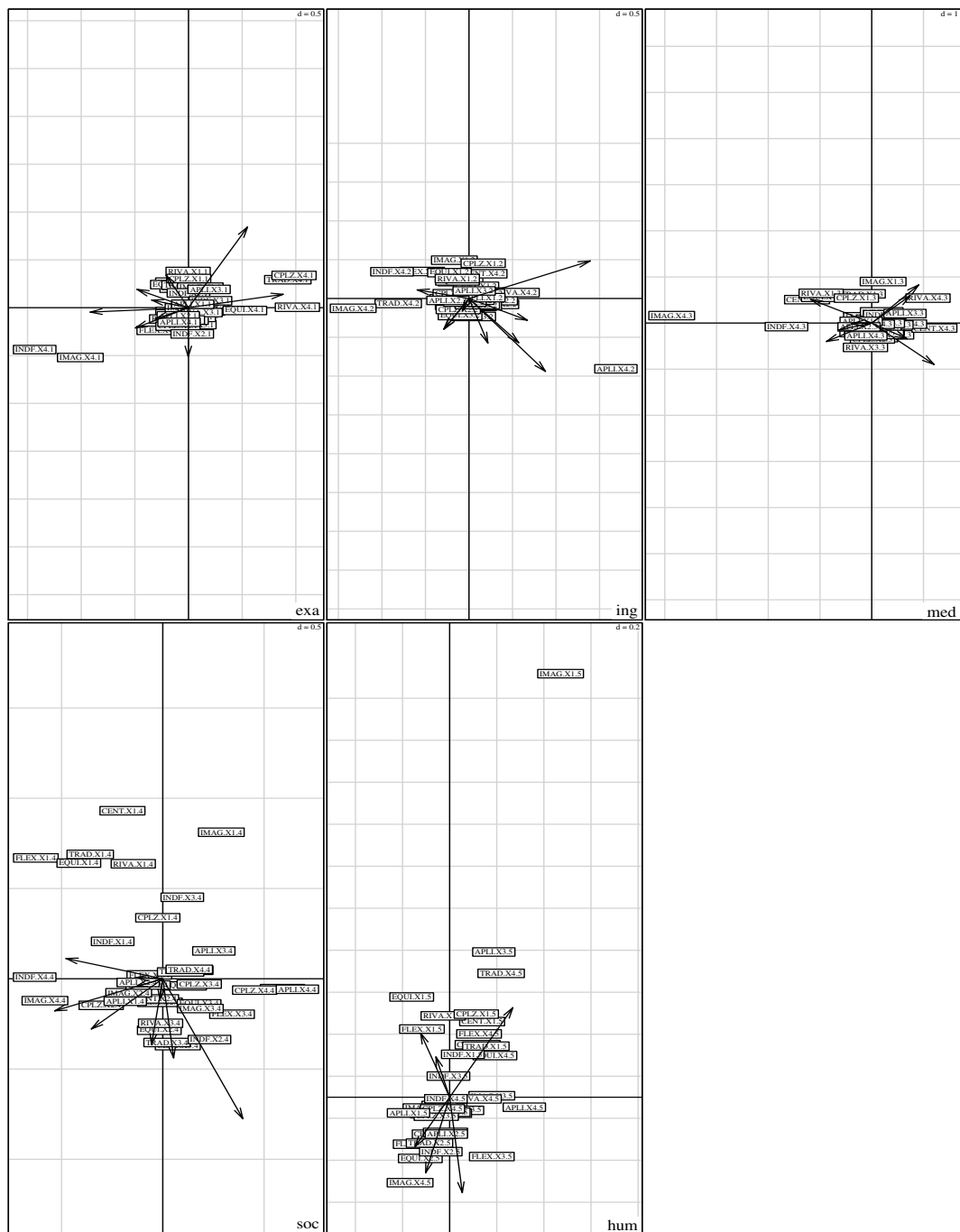


Figura 10: Individuos parciales en el plano (3,4) AFM

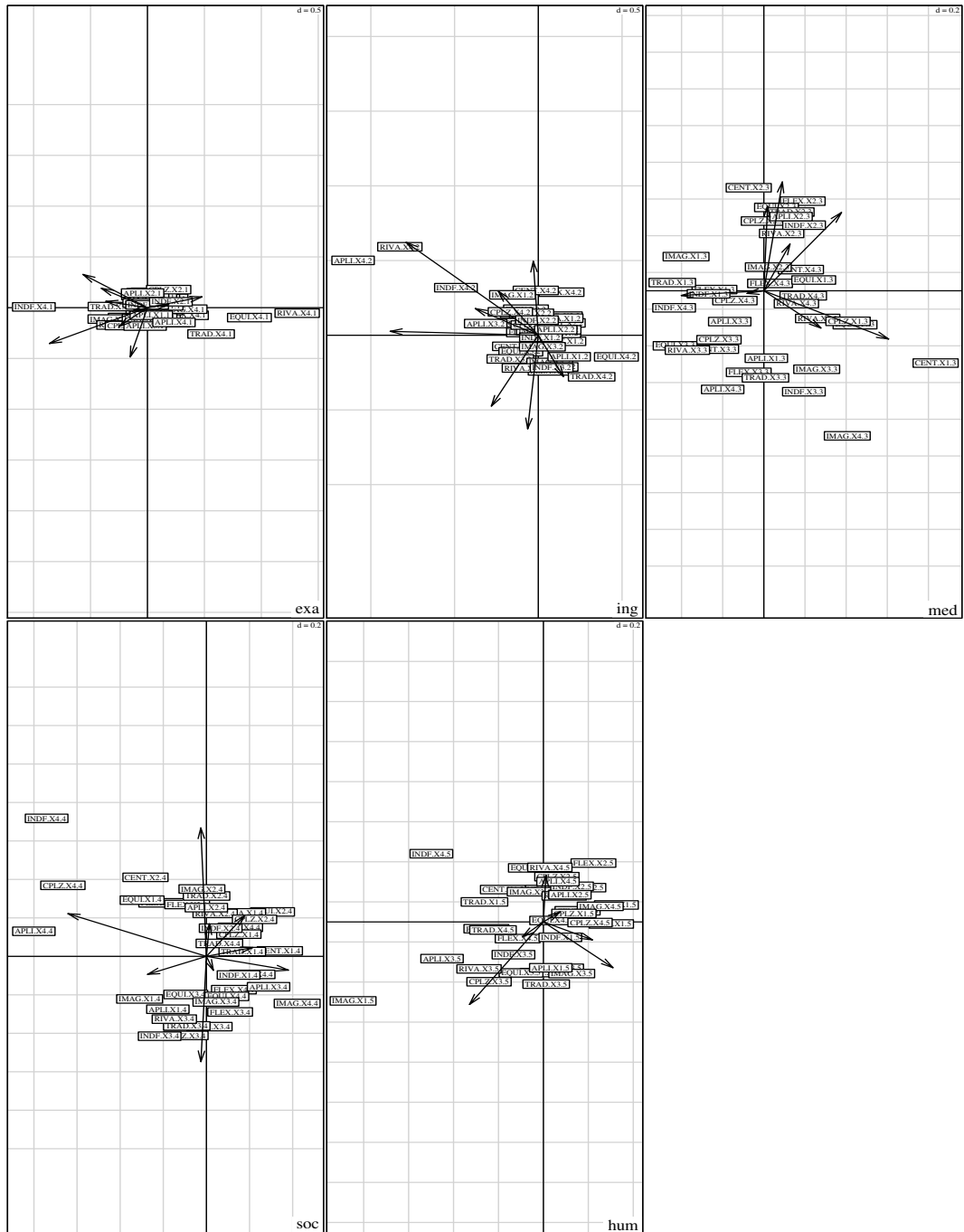


Figura 11: Individuos parciales en el plano (5,6) AFM

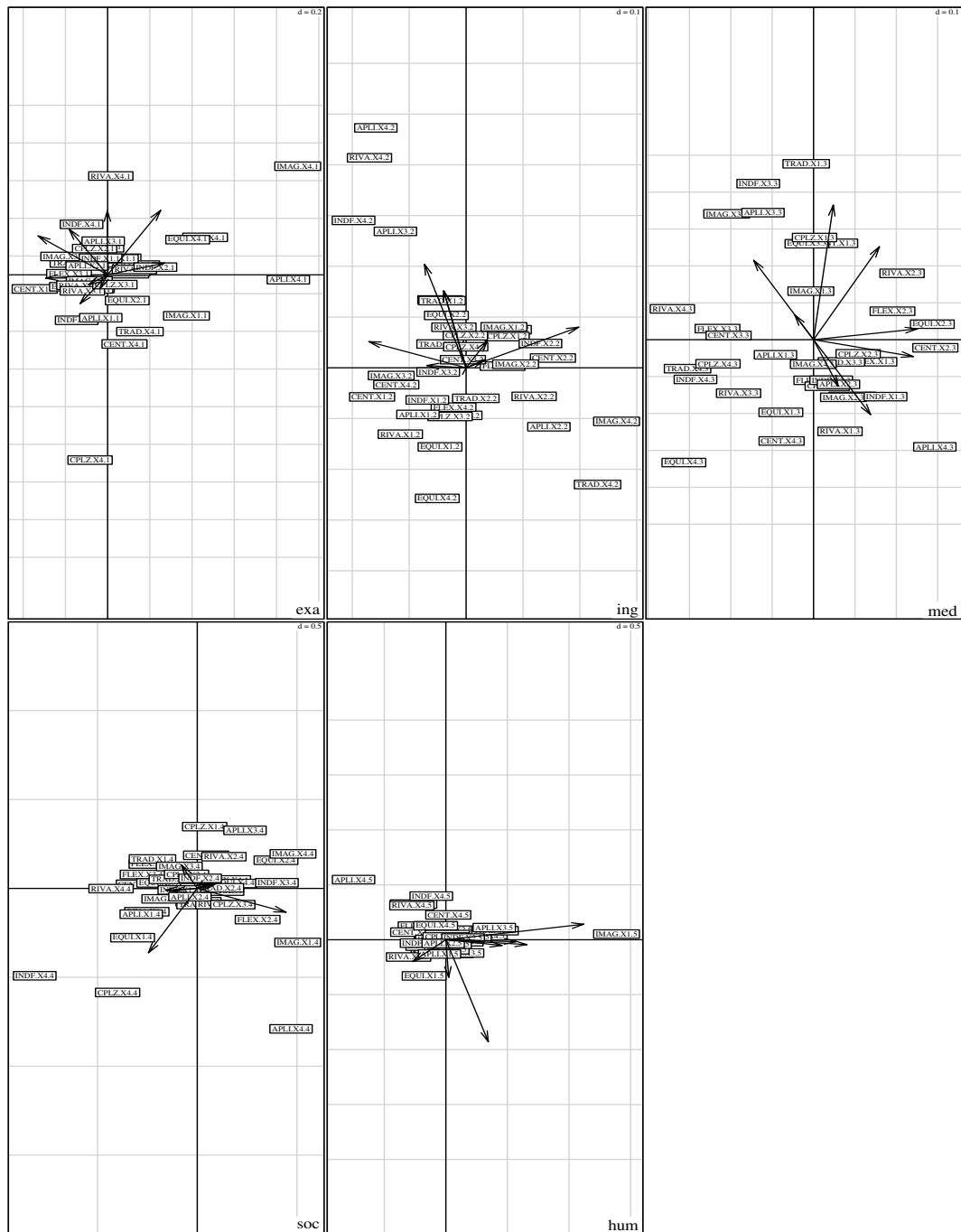


Figura 12: Individuos parciales en el plano (7,8) AFM

B. Proyecciones de Factores de los ACMs en los ejes de STATIS

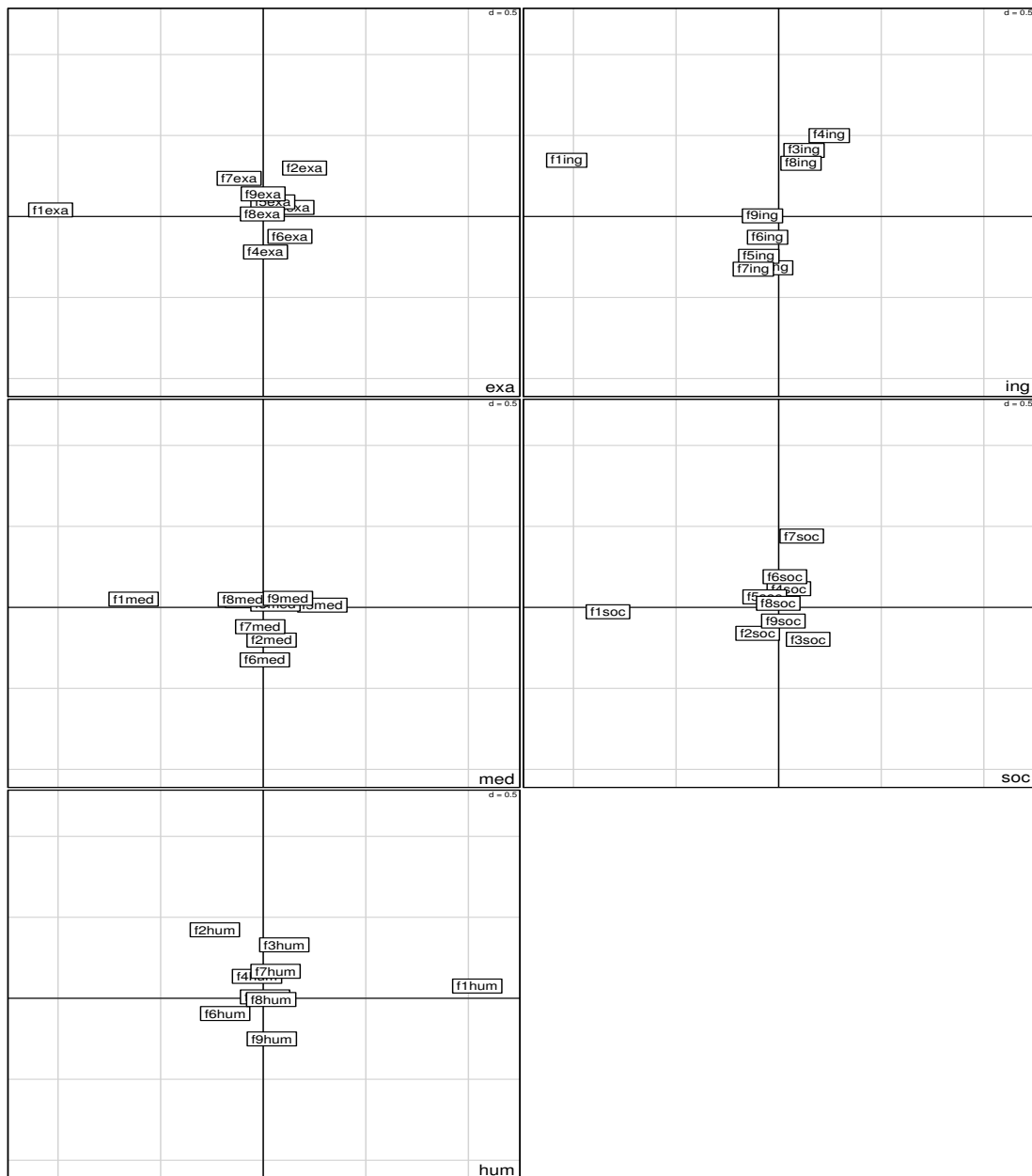


Figura 13: Plano (1,2) proyecciones factores principales ACMs

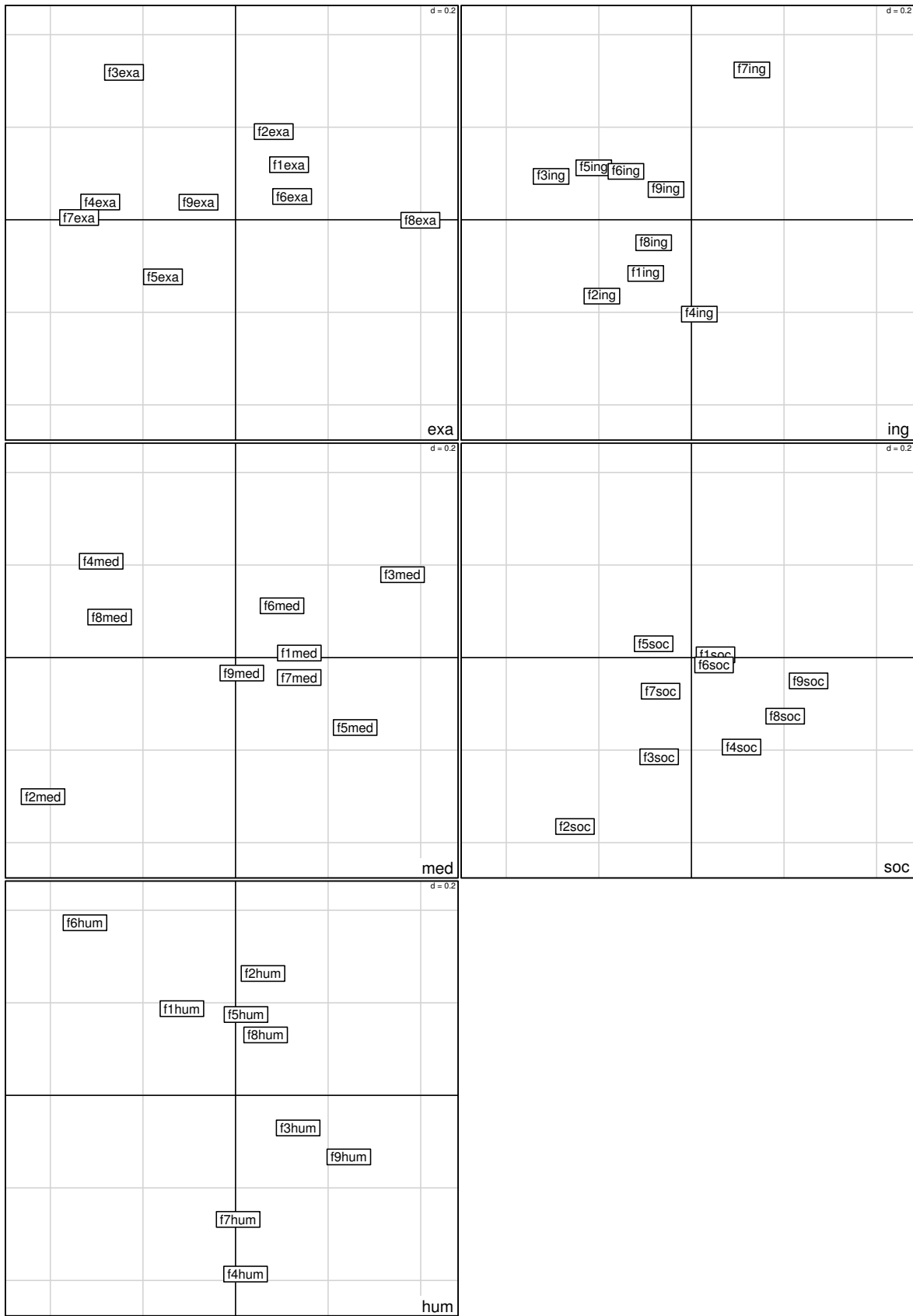


Figura 14: Plano (3,4) proyecciones factores principales ACMs

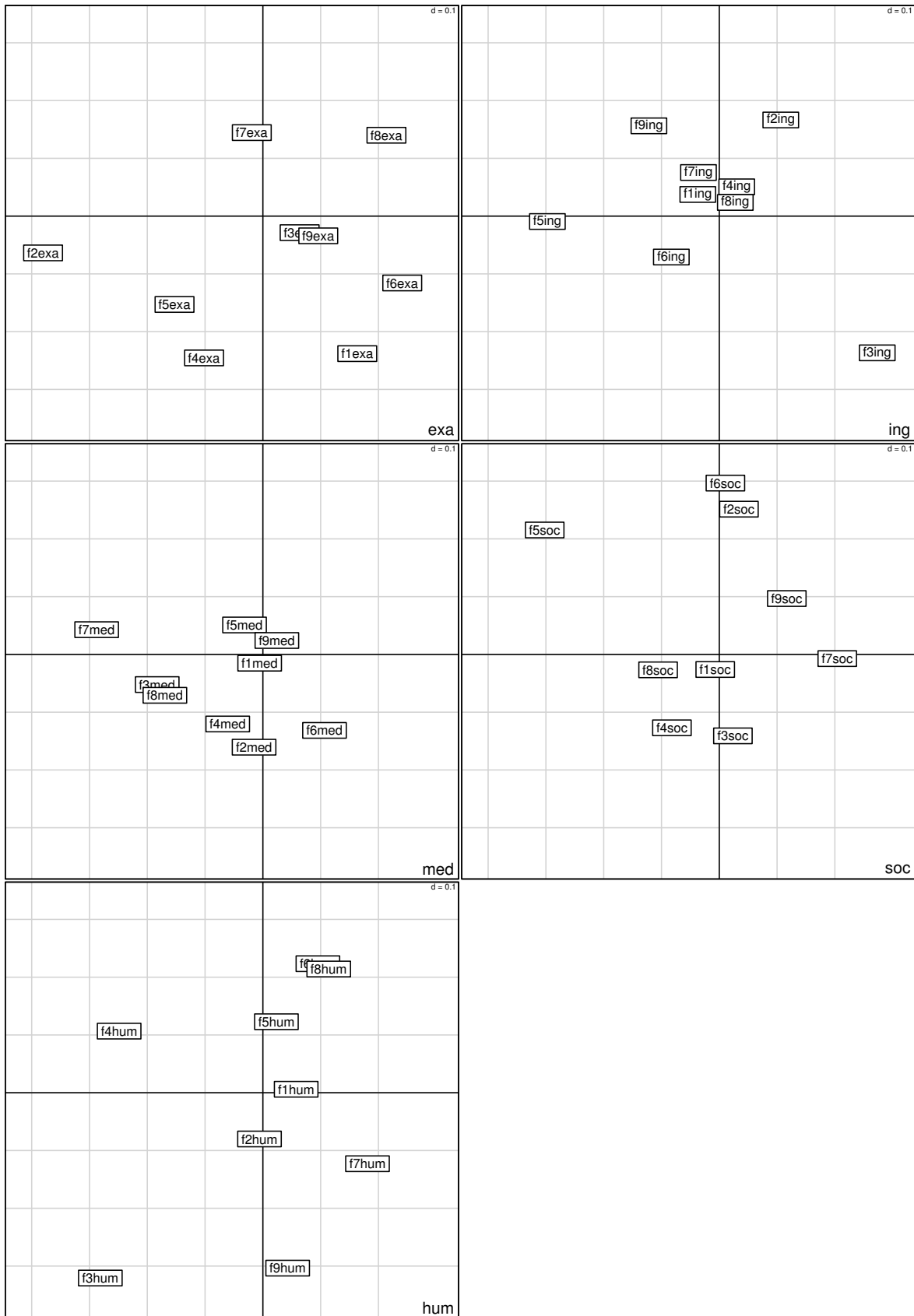


Figura 15: Plano (5,6) proyecciones factores principales ACMs

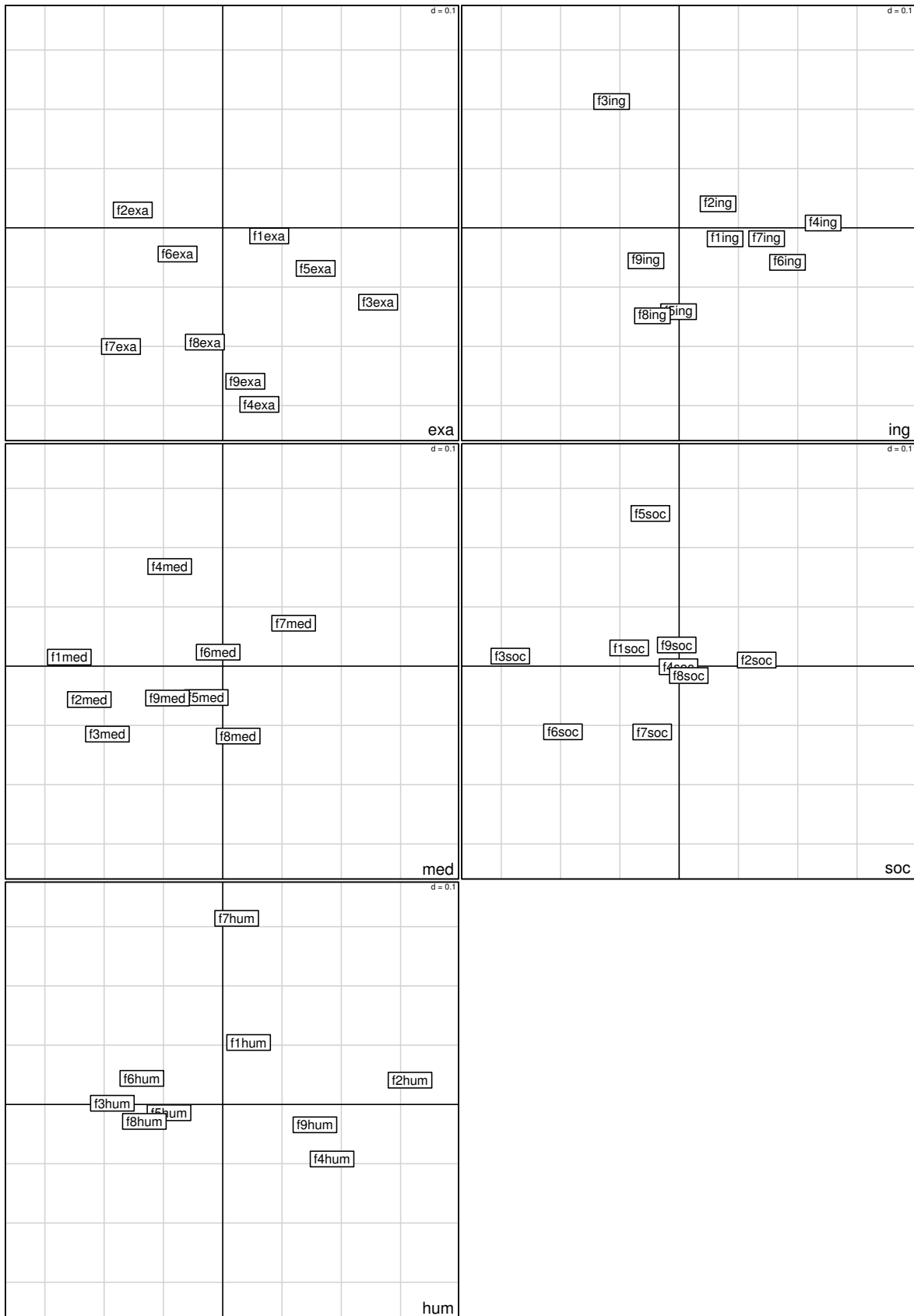


Figura 16: Plano (7,8) proyecciones factores principales ACMs

Referencias

- Abascal, E., Fernández, K., Landaluce, M. & Modroño, J. (2001), 'Diferentes aplicaciones de las técnicas factoriales de análisis de tablas múltiples en las investigaciones mediante encuestas', *Metodología de Encuestas* **3**(2), 251–279.
- Cameron, K. S. & Quinn, R. E. (1999), *Diagnosing and Changing Organizational Culture*, Addison Wesley Longman, Reading, MA.
- Carrol, J. (1968), Generalization of canonical correlation to three or more sets of variable, in 'Proceedings of the American Psychological Association', pp. 227–228.
- Dazy, F. & Le Barzic, J. (1996), *L'Analyse des Données Évolutives. Méthodes et Applications.*, Technip, Paris.
- Eckart, C. & Young, G. (1936), 'The approximation of one matrix by another of lower rank', *Psychometrika* **1**, 211–218.
- Escofier, B. & Pagès, J. (1992), *Análisis Factoriales Simples y Múltiples*, Servicio Editorial de la Universidad del País Vasco, Bilbao.
- Gower, J. (1975), 'Generalized procrustes analysis', *Psychometrika* **40**(1), 33–51.
- Holmes, S. (1989), Using the bootstrap and the RV coefficient in the multivariate context, in E. Diday, ed., 'Data Analysis, Learning Symbolic and Numeric Knowledge', Nova Science, pp. 119–131.
- Landaluce, M., Fernández, K. & Modroño, J. (1999), 'Reflexiones sobre el uso comparativo del análisis factorial múltiple y de la metodología statis para el análisis factorial múltiple', *Methodologica* **7**, 37–65.
- Lavit, C. (1988), *Analyse Conjointe de Tableaux Quantitatifs*, Masson, Paris.
- L'Hermier des Plantes, H. (1976), STATIS, Structuration des tableaux à trois indices de la statistique, PhD thesis, Université de Montpellier.
- Markus, M. (1994), *Bootstrap Confidence Regions in Nonlinear Multivariate Analysis*, DWSO Press, Leiden University.
- Modroño, J., Fernández, K. & Landaluce, M. (2001), A multivariate two step method for tables of categorical variables, in V. Esposito, C. Lauro, A. Morineau & M. Tenenhaus, eds, 'PLS and Related Methods. Proceedings of the PLS'01 International Symposium', CISIA ceresta, pp. 211–223.
- Pagès, J. (1996), 'Eléments de comparaison entre l'analyse factorielle multiple et la méthode statis', *Rev. Statistique Appliquée* **XLIV**(4), 81–95.
- Robert, P. & Escoufier, Y. (1976), 'A unifying tool for linear multivariate methods: the RV-coefficient', *Journal of Applied Statistics* **25**(3), 257–265.
- Tucker, L. (1958), 'An inter-battery method of factor analysis', *Psychometrika* **23**(2).
- Tucker, L. (1964), The extension of factor analysis to three-dimensional matrices, in C. Harris, ed., 'Contribution to mathematical Psychology', Univ. of Wisconsin Press, Madison, pp. 109–127.
- Tucker, L. (1966), 'Some mathematical notes on three-mode factor analysis', *Psychometrika* **31**, 279–311.