

UNIVERSIDAD DE LEON
Departamento de Dirección y Economía de la Empresa



Desarrollo de la metodología STATIS. Aplicación al estudio del sector asegurador español

Doctoranda: Dña. M. Eva VALLEJO PASCUAL

Directora: Dra. Dña. M. Jesús MURES QUINTANA
Catedrática de Escuelas Universitarias del
Área de Estadística e Investigación Operativa de la
Universidad de León

León, 2000

UNIVERSIDAD DE LEON
Departamento de Dirección y Economía de la Empresa



**Desarrollo de la metodología STATIS.
Aplicación al estudio del sector asegurador español**

M. Eva VALLEJO PASCUAL

León, 2000

Agradecimientos

Es de sobra conocido por todos lo largo, extenso, y en ocasiones cansino, que es el proceso de realización de una tesis doctoral. No han faltado voces que han afirmado que sin las ayudas que se reciben sería harto difícil concluir un trabajo de estas características. Ésta es la razón por la que, sin deseo de dejar a nadie en el tintero, hemos de agradecer a cuantas personas e instituciones nos han ayudado en alguno de los estadios del estudio. Saldada la deuda con todos, procede destacar el apoyo especial de algunos que nos han auxiliado, y entiéndase esta palabra con toda su carga semántica, en fondo, en forma, o en ambas cosas.

El Dr. Ventura Victoria, antiguo catedrático mío en los años de licenciatura ha seguido con su magisterio, aunque ahora sus tareas lo hallen en la Universidad de Oviedo. Sus sugerencias y observaciones han sido cuantiosas y valiosas, por lo que quiero dejar patente mi reconocimiento expreso.

A los miembros del Departamento de Dirección y Economía de la Empresa en general y en particular a los del Área de Estadística e Investigación Operativa debo darles las gracias. Unos me han pasado libros y trabajos que pudieran interesar para este estudio; otros, mis compañeros de Área, podría decirse que me han acompañado en el camino tortuoso: viviendo los propios altibajos de la investigación, ofreciendo la palabra de apoyo y, sobre todo, brindando sus conocimientos en las secciones que conocen y dominan mejor que yo.

Al Dr. Lanero, por sus indicaciones sobre normativa bibliográfica y por sus sabios consejos en materia lingüística, a D. Emilio Fernández, responsable de la enmaquetación y el mecanografiado. En la una y en el otro ha enterrado horas, paciencia y pulcritud. A él se deben presentación y distribución; a mí los errores que todavía hayan podido quedar por culpa de revisiones en las que se escapan erratas “invisibles” para el autor, pero que saltan a primera vista para el lector.

A la Directora de esta Tesis, Dra. Dña. María Jesús Mures Quintana, docente excelente, y excelencia de la que he podido disfrutar en mi época de estudiante, de postgraduada, de doctoranda y, si se me permite decirlo, de compañeras. El trabajo que aquí se presenta ha sido una travesía que hemos recorrido juntas. En ella siempre he tenido palabras de apoyo, de corrección, de adición de conocimiento y de acicate en los momentos en los que la voluntad flaquea. Por todo ello, muchas gracias.

A mis amigos, que concluidos sus estudios emprendieron trayectorias diversas, pero que, pese a todo, han continuado con su ayuda siempre valiosa. Quedan, cómo no, los que pertenecen al vínculo personal que, aún quedando en el anonimato, han sabido comprender, esperar y mostrar su afecto.

A todos, pues, debo mi agradecimiento y rindo mi particular homenaje, sin olvidar que el vocablo deriva, ante todo y sobre todo, de persona.

Los negocios deben estar claramente diferenciados de la vida. Aquéllos requieren seriedad y metodología; ésta debe llevarse de forma más libre

Goethe

Índice

INTRODUCCIÓN

Capítulo I: Técnicas descriptivas y predictivas para el análisis de la información: Contexto de la metodología STATIS	1
1. Técnicas Descriptivas Multidimensionales	5
1.1 Antecedentes históricos	6
1.2 Clasificación de técnicas	7
1.2.1 Técnicas para tablas de doble entrada	8
1.2.2 Técnicas para tablas de tres entradas	14
2. Técnicas de predicción	43
2.1 Antecedentes históricos	44
2.2 Clasificación de las técnicas de predicción	46
2.2.1 Basadas en información subjetiva	47
2.2.2 En función del análisis aislado de series	49
2.2.3 Según las relaciones entre variables	58
3. La Metodología STATIS como Técnica Descriptiva y de Predicción	67
3.1 Comparación con métodos simétricos	69
3.2 Comparación con el Análisis Factorial Múltiple	70
3.3 Comparación con métodos para datos temporales	73

Capítulo II: La metodología STATIS: <i>Statis</i> y <i>Statis dual</i>	77
1. <i>Análisis de Componentes Principales</i> (ACP)	81
1.1 Antecedentes históricos	81
1.2 Datos y objetivos	84
1.3 ACP <i>normado</i>	89
1.3.1 Descripción de las nubes (individuos y variables)	89
1.3.2 Desarrollo del ACP <i>normado</i>	99
1.3.3 Interpretación del ACP <i>normado</i>	133
1.4 ACP <i>no-normado</i> (<i>centrado</i>)	142
1.4.1 Análisis de la nube de individuos	142
1.4.2 Análisis de la nube de variables	146
1.4.3 Interpretación	149
2. Desarrollo de la metodología STATIS	153
2.1 Antecedentes históricos	153
2.2 Operadores asociados a una tabla de datos	157
2.2.1 Operador $W_t D$. Operador normado $\frac{W_t D}{\ W_t D\ _{HS}}$	161
2.2.2 Operador $V_t M$. Operador normado $\frac{V_t M}{\ V_t M\ _{HS}}$	171
2.3 Métodos <i>Statis</i> y <i>Statis dual</i>	176
2.3.1 Los datos y descripción general	177
2.3.2 Etapas de desarrollo	180
2.3.3 Interpretación	213
3. Posibilidades descriptivas y predictivas de STATIS	225
3.1 Descripción y STATIS	226
3.1.1 Introducción	226
3.1.2 La clasificación y el <i>Statis</i>	230
3.2 Predicción y STATIS	248
3.2.1 Conceptos básicos	248
3.2.2 <i>Análisis Factorial</i> de un proceso	252
3.2.3 <i>Statis</i> y procesos estocásticos	255
3.2.4 Modelos <i>ARIMA</i>	260
3.2.5 Predicción de las variables originales y validación de resultados	300

Capítulo III: Estudio del sector asegurador español	305
1. Análisis general del sector asegurador	311
1.1 Características de la actividad aseguradora	311
1.1.1 La función económica del seguro	314
1.1.2 El seguro dentro del sistema financiero	317
1.1.3 La empresa aseguradora	320
1.2 El sector asegurador español	328
1.2.1 Legislación	329
1.2.2 Indicadores económicos relevantes del sector	331
1.2.3 Perspectivas	342
1.3. Análisis estructural del sector asegurador español	346
1.3.1 El modelo de Porter para el análisis de sectores industriales	347
1.3.2 Aplicación al sector asegurador español	363
2. Análisis empírico	383
2.1 Introducción	383
2.2 Estudio descriptivo del sector asegurador español	386
2.2.1 Descripción de los datos	386
2.2.2 Estructura del sector asegurador español: <i>Statis</i> y <i>Statis dual</i>	395
2.2.3 Determinación de grupos estratégicos en el sector asegurador español: <i>Statis</i> y <i>Cluster</i>	446
2.3 Evolución del sector asegurador español	463
2.3.1 Descripción de los datos	464
2.3.2 Comportamiento del sector	468
2.3.3 Análisis temporal	488
2.3.4 Validación del modelo	506
2.3.5 Predicción e interpretación de resultados	509
CONCLUSIONES	513
BIBLIOGRAFÍA	545
ANEXO	565

Índice figuras

Capítulo I: Técnicas descriptivas y predictivas para el análisis de la información: Contexto de la metodología STATIS

Figura I.1	10
Clasificación de técnicas de Evrard y Le Maire	
Figura I.2	12
Clasificación de técnicas de Greenacre	
Figura I.3	17
Tablas de tres entradas “three-way o multiway data sets”	
Figura I.4	28
Tabla de tres entradas completa “three-way data array”	

Capítulo II: La metodología STATIS: *Statis* y *Statis dual*

Figura II.1	84
Tabla de datos (X)	
Figura II.2	86
Representación de los individuos en R^2	
Figura II.3	87
Cantidad de información	
Figura II.4	90
Nuevo origen: el centro de gravedad de la nube	
Figura II.5	95
Consecuencias de la transformación $x_{ij} - \bar{x}_j$ en N_J	
Figura II.6	105
Proyección del individuo e_1 sobre el eje u_1 (vector director a_1)	
Figura II.7	113
Imagen euclídea en el primer plano factorial de la nube de individuos N_I	
Figura II.8	123
Círculo de Correlaciones	
Figura II.9	124
Efecto talla	
Figura II.10	130
Relación entre los dos análisis en el eje k	

Figura II.11	134
Angulo entre e_i y su proyección sobre eje $k = 1$	
Figura II.12	157
Tabla de datos Y	
Figura II.13a	159
Esquema de dualidad	
Figura II.13b	160
Esquema de dualidad	
Figura II.14	177
Tabla total de tres entradas con individuos comunes	
Figura II.15	179
Tabla de tres entrada con variables comunes	
Figura II.16	216
Existe una estructura común al conjunto de tablas	
Figura II.17	216
El compromiso no representa bien todos los objetos	
Figura II.18	217
Objetos con normas muy diferentes	
Figura II.19	217
No hay una estructura común	
Figura II.20	218
Interestructura centrada	
Figura II.21	224
<i>Statis</i> y <i>Statis dual</i>	
Figura II.22	227
Tablas de datos con individuos y variables comunes	
Figura II.23	232
Tabla de n individuos p variables	
Figura II.24	247
El <i>Statis</i> y el Análisis Cluster	
Figura II.25	249
Tabla de datos temporales	
Figura II.26	255
Tabla de tres entradas con datos temporales	
Figura II.27	275
Fases de la metodología Box-Jenkins	
Figura II.28	304
Proceso predictivo del STATIS	

Capítulo III: Estudio del sector asegurador español

Figura III.1	347
Determinantes del nivel de competencia de un sector industrial	
Figura III.2	399
Primer plano de la interestructura del <i>Statis</i>	
Figura III.3	400
Primer plano de la interestructura del <i>Statis dual</i>	
Figura III.4a	405
Círculo de correlaciones de las variables de alcance (excepto PRIMEM). Factores 1 y 2	
Figura III.4b	406
Círculo de correlaciones de las variables de desarrollo y PRIMEM. Factores 1 y 2	
Figura III.5a	408
Círculo de correlaciones de las variables de alcance (excepto PRIMEM). Factores 2 y 3	
Figura III.5b	409
Círculo de correlaciones de las variables de desarrollo y PRIMEM. Factores 2 y 3	
Figura III.6a	410
Círculo de correlaciones de las variables de alcance, excepto PRIMEM. Factores 4 y 5	
Figura III.6b	411
Círculo de correlaciones de las variables de desarrollo y PRIMEM. Factores 4 y 5	
Figura III.7a	413
Individuos–compromiso de mayor inercia. Factores 1 y 2	
Figura III.7b	415
Otros individuos–compromiso. Factores 1 y 2	
Figura III.8	416
Individuos–compromiso de mayor inercia. Factores 2 y 3	
Figura III.9	418
Individuos–compromiso de mayor inercia. Factores 4 y 5	
Figura III.10	420
Círculo de correlaciones de las variables–compromiso. Factores 1 y 2	
Figura III.11	421
Círculo de correlaciones de las variables–compromiso. Factores 2 y 3	
Figura III.12	422
Círculo de correlaciones de las variables–compromiso. Factores 4 y 5	
Figura III.13a	426
Trayectorias de individuos. Factores 1 y 2	

Figura III.13b	427
Trayectorias de individuos. Factores 1 y 2	
Figura III.13c	428
Trayectorias de individuos. Factores 1 y 2	
Figura III.13d	428
Trayectorias de individuos. Factores 1 y 2	
Figura III.14a	430
Trayectorias de individuos. Factores 4 y 5	
Figura III.14b	430
Trayectorias de individuos. Factores 4 y 5	
Figura III.14c	431
Trayectorias de individuos. Factores 4 y 5	
Figura III.14d	431
Trayectorias de individuos. Factores 4 y 5	
Figura III.15	453
Clasificación de las entidades aseguradoras con cinco factores	
Figura III.16	459
Posicionamiento de los grupos obtenidos. Factores 1 y 2	
Figura III.17	461
Posicionamiento de los grupos obtenidos. Factores 2 y 3	
Figura III.18	462
Posicionamiento de los grupos obtenidos. Factores 4 y 5	
Figura III.19	471
Primer plano de la interestructura	
Figura III.20	475
Círculo de correlaciones. Factores 1 y 2	
Figura III.21	475
Círculo de correlaciones. Factores 1 y 3	
Figura III.22	476
Círculo de correlaciones. Factores 2 y 3	
Figura III.23	478
Los individuos–compromiso (años). Factores 1 y 2	
Figura III.24	479
Los individuos–compromiso (años). Factores 1 y 3	
Figura III.25	479
Los individuos–compromiso (años). Factores 2 y 3	
Figura III.26	490
Representación del primer factor	
Figura III.27	491
Correlograma muestral del primer factor	

Figura III.28	491
Correlograma parcial muestral del primer factor	
Figura III.29	494
Representación del segundo factor	
Figura III.30	495
Correlograma muestral del segundo factor	
Figura III.31	495
Correlograma muestral parcial del segundo factor	
Figura III.32	497
Representación del tercer factor	
Figura III.33	498
Correlograma muestral del tercer factor	
Figura III.34	499
Correlograma muestral parcial del tercer factor	
Figura III.35	502
Residuos del modelo $AR(1)$ sin término independiente	
Figura III.36	503
Función de Autocorrelación de los residuos del modelo $AR(1)$ sin término independiente	
Figura III.37	504
Función de Autocorrelación parcial de los residuos del modelo $AR(1)$ sin término independiente	

Índice tablas

Capítulo I: Técnicas descriptivas y predictivas para el análisis de la información: Contexto de la metodología STATIS

Tabla I.1	27
Métodos para tablas “three-way data sets”	
Tabla I.2	36
Métodos para tablas “three-way arrays”	
Tabla I.3	41
Métodos para tablas derivadas	
Tabla I.4	65
Técnicas de predicción	

Capítulo II: La metodología STATIS: *Statis* y *Statis dual*

Tabla II.1	125
Las dos nubes, ejes y factores respectivos	

Capítulo III: Estudio del sector asegurador español

Tabla III.1	333
Indicadores del crecimiento y evolución del sector asegurador	
Tabla III.2	335
Importancia relativa y evolución de los grandes ramos	
Tabla III.3	338
Porcentaje de gastos internos y externos sobre primas	
Tabla III.4	339
Siniestralidad, equilibrio y resultado técnico del sector	
Tabla III.5	370
Cuota de mercado de las empresas de seguros en 1995	
Tabla III.6	371
Cuota de mercado de las empresas de seguros en 1996	
Tabla III.7	371
Cuota de mercado de las empresas de seguros en 1997	
Tabla III.8	371
Cuota de mercado de las empresas de seguros en 1998	
Tabla III.9	372
Valores del índice de concentración de Gini	

Tabla III.10	379
Ramo de seguros de vida	
Tabla III.11	379
Ramo de seguros de automóviles	
Tabla III.12	379
Ramo de seguros de salud	
Tabla III.13	380
El sector asegurador	
Tabla III.14	388
Individuos (Entidades Aseguradoras)	
Tabla III.15	394
Variables analizadas para cada entidad aseguradora	
Tabla III.16	394
Años estudiados	
Tabla III.17a	398
Matriz S de productos escalares entre $W_t D$ normado	
Tabla III.17b	398
Matriz S de productos escalares entre $V_t M$ normado	
Tabla III.18a	401
Ponderaciones <i>Statis</i>	
Tabla III.18b	401
Ponderaciones <i>Statis dual</i>	
Tabla III.19	403
Primeros valores propios y porcentaje de inercia explicado por cada factor	
Tabla III.20	404
Correlaciones de las variables con los cinco primeros factores	
Tabla III.21	419
Valores propios y porcentajes de inercia explicado por cada factor	
Tabla III.22	419
Correlaciones de las variables con los cinco primeros factores	
Tabla III.23	467
Variables	
Tabla III.24	468
Ramos de seguros	
Tabla III.25	470
Matriz de coeficientes RV	
Tabla III.26	472
Ponderaciones del <i>Statis</i>	
Tabla III.27	473
Valores propios y porcentajes de inercia	

Tabla III.28	474
Correlaciones de las variables con los primeros factores	
Tabla III.29	489
Factores temporales	
Tabla III.30	505
Valores predichos para el primer factor	
Tabla III.31	506
Valores del primer factor para el 98 y el 99	
Tabla III.32	509
Variables relativas a los seguros de vida	
Tabla III.33	510
Variables relativas a los seguros de automóviles	
Tabla III.34	510
Variables relativas a los seguros de salud	
Tabla III.35	510
Variables relativas al resto de seguros	

INTRODUCCIÓN

Una de las transformaciones más destacables que deben afrontar las economías del tercer milenio es la desencadenada por las nuevas tecnologías de la información y de las comunicaciones, las que permiten adquirir, almacenar, procesar y transmitir gran cantidad de información a un ritmo hasta ahora desconocido y a bajo coste (Círculo de Empresarios, 1991: 1). Este hecho ha dado lugar a lo que se conoce con el nombre de la Sociedad de la Información. De tal manera, que cualquier empresa tiene acceso a gran cantidad de datos que, en su inmensa mayoría, están disponibles en la red y que son de utilidad extrema para el desarrollo de su actividad habitual.

No obstante, la mera generación de datos no estructurados no conduce automáticamente a la creación de información y no toda la información puede considerarse como fuente de conocimiento. Para llegar a éste es necesario clasificar, analizar y procesar la información: los datos (lo que en un proceso productivo serían los inputs) necesitan ser manipulados para que generen valor (outputs o conocimientos), de tal manera que la cuestión básica no es la disponibilidad de la información, sino su aprovechamiento por las empresas. Por esto, la próxima ventaja competitiva de las empresas dependerá de la capacidad de absorber, analizar y gestionar todo el volumen de información disponible de forma generalizada, transformándola en conocimiento (Círculo de Empresarios, 1999:33).

Las técnicas y procedimientos propios de la Estadística tienen un papel muy destacado en este proceso de transformación. Según manifiestan

Martín-Gúzman y Martín Pliego (1989: 1), la ciencia Estadística proporciona los métodos de análisis de la información recogida, y, en general, otros de investigación aplicables a todo el resto de ciencias. En opinión de Escuder Vallés (1987: 22), la Estadística puede considerarse como un método, que pueden utilizar la totalidad de las ciencias empíricas, incorporándose como un parte más del objeto formal de las mismas, dado que proporciona una metodología propia para el tratamiento de la información. Máxime en el caso de la Ciencia Económica, en la que la imposibilidad de establecer normas de comportamiento y leyes que regulen las relaciones económicas ha contribuido a una utilización, poco menos que imprescindible, de la Estadística como método de investigación (Mures Quintana, 1991: 2).

Resulta pertinente señalar que el Método Estadístico consta de una serie de estadios: el diseño del experimento, la recogida de la información, la descripción, el análisis de la misma, y la predicción. Para un desarrollo eficaz de todos estos estadios existen procedimientos adecuados a la naturaleza de la información que se va a analizar. En concreto, para las etapas de descripción y predicción, la Estadística aporta una serie de métodos para llevar a cabo las mencionadas tareas: las Técnicas Descriptivas y las Técnicas de Predicción. Debido a la gran cantidad de información de la que disponen las empresas y que deben procesar para que sea útil, el desarrollo de procedimientos para el tratamiento de grandes cantidades de datos es fundamental en la creación de conocimiento, y su empleo por parte de las empresas es una cuestión prioritaria, a fin de crear una ventaja competitiva frente a otras empresas.

En esta línea, las Técnicas de Descripción avanzan con la aplicación de algunos de los métodos básicos, ya conocidos, a las nuevas circunstancias, como es el caso de los artículos de Girard *et al.* (1998), que trata de

extender el *Análisis de Componentes Principales* al caso no lineal; y el de Tipping y Bishop (1999) que aportan un enfoque probabilístico del mismo. El desarrollo de nuevas técnicas para el tratamiento de las denominadas tablas de tres entradas, que, según veremos, se trata de un conjunto de tablas de doble entrada, es la característica común de trabajos como el de Casin (1996), que introduce el *Análisis de Componentes Principales Generalizado*, método que guarda cierta relación con técnicas ya conocidas para el análisis de este tipo de tablas, como el *Análisis Factorial Múltiple* y el STATIS. En la misma línea se encuentran los artículos de Chessel y Hanafi (1996), que definen lo que ellos denominan Co-Inercia de las K nubes de puntos que forma la tabla de tres entradas; y el *Análisis de la Concordancia* de Lafosse (1997), aplicado en primer lugar al tratamiento simultáneo de dos tablas de doble entrada, extendiendo posteriormente su desarrollo al caso de K tablas (Lafosse y Hanafi, 1997). Como ejemplos más recientes que hacen referencia al mismo tipo de tablas destacan los artículos de Boumaza (1998), que desarrolla un método con ciertas similitudes con el STATIS, y de Simier *et al.* (1999), que proponen otro procedimiento para el tratamiento de K nubes de puntos con el fin de combinar el punto de vista del STATIS con el del Análisis de Co-Inercia.

En cuanto a las Técnicas de Predicción, las nuevas tendencias tratan de obtener predicciones de conjuntos de series temporales, aprovechando la estructura de dependencia subyacente. Si bien, aunque existe una abundante literatura sobre la construcción de Modelos Multivariantes de series temporales, su utilización práctica es pequeña. Además, un problema pendiente de solución es el crecimiento exponencial del número de parámetros con la dimensión del vector de variables. Una posible solución es condensar la información en un número de factores comunes, lo que reduce el problema, en la línea marcada por los trabajos de Otter y Schuur (1982) y de Peña y Poncela (1996).

Las tendencias apuntadas, para ambos tipos de técnicas son las que subyacen en el artículo de Cabedo Semper y Moya Clemente (1999), que, a nuestro juicio, constituye un buen ejemplo de empleo simultáneo de Técnicas Factoriales y de Modelos de Predicción aplicados en el ámbito económico-financiero y en la línea señalada.

Teniendo en cuenta lo anterior, el objetivo de este trabajo es el desarrollo de la metodología STATIS (Structuration des Tableaux À Trois Indices de la Statistique) para el tratamiento de gran volumen de información, presentada en grupos y de carácter temporal, en el ámbito económico, con el fin de resaltar las ventajas que sus dos versiones¹, el *Statis* y el *Statis dual*, presentan al ser combinadas con otras técnicas cuando se trata de describir, clasificar o predecir.

Para lograr este objetivo, que parece estar en concordancia con las nuevas tendencias seguidas por los métodos estadísticos, nos ha parecido conveniente estructurar este trabajo en tres capítulos, encaminado a lograr un objetivo claro y preciso:

En el capítulo I hemos pretendido dar una visión global de las técnicas descriptivas y de predicción más frecuentemente empleadas en el ámbito económico, con el fin de conocer las circunstancias que motivan su aplicación. Asimismo se contextualiza la metodología STATIS, justificando su empleo frente a otras técnicas con objetivos y procedimientos similares.

¹ Cuando se haga referencia a la metodología de manera general se emplearán letras mayúsculas, mientras que cuando se mencione alguna de las versiones de manera individual se emplearán minúsculas.

Por tanto, el capítulo se estructura en tres partes: en la primera de ellas se realiza una revisión de las Técnicas Descriptivas Multidimensionales que se utilizan en el análisis de tablas de dos y de tres entradas. Se justifica el proceso en cuánto que el primer grupo de técnicas constituye la base para la comprensión de las segundas.

Las Técnicas de Predicción constituyen la segunda parte, considerándose en ella las más utilizadas en el ámbito económico, y realizando su clasificación según el tipo de materia de estudio que requiere su uso. Realizada esta revisión, en la tercera parte se aborda la metodología que se va a desarrollar en este estudio, relacionándola con otras técnicas y destacando su carácter de técnica descriptiva multidimensional, que se emplea en el análisis de la información de carácter temporal, lo que le imprime carácter de técnica de predicción.

El capítulo II pretende desarrollar desde una perspectiva teórica, y de forma amplia y detallada, las dos versiones de la metodología analizada, teniendo en cuenta que alguna de sus etapas son casos especiales del *Análisis de Componentes Principales* (ACP). Igualmente pretende poner de manifiesto las posibilidades descriptivas y predictivas de la misma, dependiendo de la manera en que se lleve a cabo su aplicación y de las técnicas con las que se combine.

Así, en este capítulo se hace, en primer lugar, una exposición exhaustiva del mencionado *Análisis de Componentes Principales*, puesto que los desarrollos teóricos de este método son fundamentales para la comprensión de las cuatro etapas (Interestructura, Compromiso, Intraestructura y Trayectorias) que componen el *Statis* y el *Statis dual* y que se desarrollan precisamente en la segunda parte del capítulo. En tercer lugar, se hace referencia a los procedimientos a seguir para aprovechar de la mejor manera posible las ventajas de la aplicación del STATIS, cuando el objeto del

estudio es la descripción, la clasificación o la predicción de un fenómeno a partir de un conjunto de datos, los cuales deben presentar ciertas características que justifiquen la aplicación de la metodología propuesta.

El objetivo propuesto con el desarrollo del Capítulo III consiste en el estudio del sector asegurador español, con el fin de obtener resultados y conclusiones relevantes sobre el mismo, tanto para las entidades que lo componen como para los ciudadanos y entidades públicas cuyas decisiones afectan al sector. El análisis empírico realizado sobre los datos del sector permite poner de manifiesto la utilidad de los desarrollos teóricos que sobre la metodología STATIS se llevaron a cabo en el capítulo anterior.

Para dicho propósito, se ha estructurado este último capítulo en dos partes: en la primera se analiza el sector asegurador español de una manera general, incidiendo en la importancia que la actividad aseguradora y el sector tienen en la economía y en el sistema financiero, incluyendo un análisis del mismo desde la óptica del Análisis Estratégico. Todo ello con el fin de que pueda ser tomado como referencia por cualquier entidad del sector en el proceso de formulación de su estrategia.

En la segunda parte se analiza, mediante la metodología STATIS, la información relativa al sector, con fines descriptivos y predictivos. De esta forma se logran los objetivos propuestos en términos de describir el sector (*Statis* y *Statis dual*), identificar los grupos estratégicos (*Statis* y *Análisis Cluster*), analizar la evolución del sector y predecir resultados (*Statis* y Modelos *ARIMA*).

Las conclusiones más generales obtenidas como consecuencia de los tres capítulos, con la justificación del objetivo alcanzado y las posibles vías para continuar la investigación, así como la bibliografía empleada y el

anexo, con las salidas de ordenador de los programas empleados, ponen fin a este trabajo, aunque sólo de manera temporal. Su conclusión definitiva dependerá, por un lado, de que otros estudios lo tomen en consideración, y, por otro, de que sus conclusiones rebasen el ámbito puramente académico, con el fin de resaltar, una vez más, la dependencia mutua entre Universidad y Empresa.

Capítulo I
Técnicas descriptivas y predictivas para el
análisis de la información:
Contexto de la metodología STATIS

El capítulo que vamos a desarrollar a continuación, como primera parte de nuestro trabajo, tiene una doble finalidad: en primer lugar, exponer la idea general de las técnicas multidimensionales y de predicción disponibles y las condiciones para su aplicación, facilitando la tarea de elegir la técnica adecuada para cada caso concreto. En segundo lugar, presentar el contexto en el que se desarrolla la metodología STATIS, relacionándolo con técnicas multidimensionales similares, con el fin de obtener una visión general de las posibilidades que ofrece su aplicación y los requisitos que debe cumplir la información que con ella se pretende analizar.

En los diferentes apartados del mismo, se recoge una relación amplia, aunque no exhaustiva, de las técnicas multidimensionales y de predicción más importantes que se han ido desarrollando a lo largo del tiempo. Aunque la clasificación de las diferentes técnicas no soluciona de manera definitiva el problema de elección de la técnica necesaria en cada caso, ni el de la elección entre variantes de una misma técnica, dicha clasificación proporciona una visión general de las mismas. Además, según manifiesta Pulido (1989b: 47) a propósito de las técnicas de predicción, pero también válido para todo tipo de métodos,

... no es recomendable considerar una sola técnica y se debe estar al tanto de los diversos métodos existentes. Conocer, a diferentes niveles, las diversas técnicas permite experimentar con las más simples, y emplear otras cuando se es consciente de las limitaciones de las primeras, justificándose así el empleo de una técnica más compleja.

Parece conveniente aclarar, que en muchos casos, una misma técnica o método puede ser aplicado de diferentes maneras, según los resultados que se persiga obtener con su aplicación. Precisamente, uno de los objetivos del presente estudio es poner de relieve las posibilidades tanto descriptivas como predictivas de STATIS que, aunque en principio es una Técnica Descriptiva Multidimensional, según el objetivo perseguido y la técnica con la que se combine, cumple una función descriptiva o predictiva. Por ese motivo, las clasificaciones que aquí se recogen y la consideración de ciertas técnicas como predictivas o descriptivas podrían realizarse según otros criterios.

Se consideran, en primer lugar, algunas de las clasificaciones propuestas para las Técnicas Multidimensionales y, a continuación, se hará lo mismo con las Técnicas de Predicción. Para las primeras, dada su diversidad, es difícil encontrar clasificaciones que permitan ordenar las técnicas existentes sin que su comprensión resulte difícil.

En cuanto a las Técnicas de Predicción, aunque también muy numerosas, las clasificaciones existentes parecen estar más estructuradas y, en principio, permiten catalogar más fácilmente las técnicas disponibles. No obstante, se han tratado de recopilar las clasificaciones existentes, con el fin de poner de manifiesto los criterios básicos para diferenciar las distintas técnicas.

Una vez clasificadas dichas técnicas, y con el fin de cumplir el segundo de los objetivos señalados, se hará especial referencia a la metodología STATIS y a su relación con técnicas de desarrollos y finalidades similares, estableciendo las diferencias y semejanzas oportunas e intentando justificar su aplicación en el caso que nos ocupa.

1. Técnicas Descriptivas Multidimensionales

No creemos necesario, por evidente, justificar la utilidad de estas técnicas en prácticamente la totalidad de las ciencias. Nos referiremos al ámbito que nos ocupa, el de las ciencias económicas y financieras, en el que existen muchos fenómenos (imagen de marca, calidad de los productos, estructura de un sector, etc.) que no pueden medirse directamente, sino que son el resultado de un conjunto de características. Para estos casos son necesarios métodos o técnicas que permitan sintetizar y resumir grandes conjuntos de datos, de cara a obtener informaciones válidas para la mejor comprensión del fenómeno objeto de estudio.

A partir de la recogida de datos, el estudio individual de cada una de esas características sería arduo, tedioso y al mismo tiempo incompleto, puesto que no se podría obtener una visión global del fenómeno estudiado. Así, el análisis de dos o más variables relativas a un fenómeno se realiza a través de los llamados Métodos Multidimensionales (Estadística Multivariante o Análisis de Datos)¹, mientras que el estudio de una única variable sería el objeto de la Estadística Univariante. A continuación se realiza una breve referencia histórica para, a continuación, centrarnos en la clasificación propiamente dicha.

¹ El término anglosajón es Multivariate Analysis que, según la disciplina estudiada, se ha traducido de diversas maneras: Análisis Multivariable, Análisis o Estadística Multivariante, entre otras. El término francés es Analyse des Données Multidimensionnelles, cuya traducción es Análisis de Datos Multidimensionales o, simplemente, Análisis de Datos. A lo largo del presente trabajo se emplearán indistintamente cualquiera de estas denominaciones.

1.1 Antecedentes históricos

Los pioneros de la Estadística moderna proceden de la Escuela Anglosajona y los orígenes del Análisis Multivariante están próximos a los desarrollos de los métodos de correlación. Este concepto, tal y como es empleado hoy estadísticamente, es debido a Galton, que fue el primero en proporcionar una medida del grado de asociación entre variables. Posteriormente, el mismo autor consideraría el problema más general de medir el grado en que una variable está afectada por la variación conjunta de otras variables. Este problema, que es el de la correlación múltiple, sería resuelto por Yule y Pearson proporcionando, además, otros coeficientes para medir la correlación parcial. Es precisamente en el trabajo de 1901 de este último autor, donde se establecen las primeras ideas del *Análisis de Componentes Principales (ACP)*.

Durante el primer tercio de siglo se producen importantes avances que posteriormente favorecerán el desarrollo de las técnicas multidimensionales, como los desarrollos de Spearman en 1904; los de Fisher, sobre el análisis de la varianza; y los de Neyman y Pearson, sobre la razón de verosimilitud, entre otros.

Las aportaciones de Hotelling, en 1931 y 1933; Wilks, en 1932 y 1935; Fisher, en 1935 y 1936; y Bartlett, en 1939, son fundamentales para el establecimiento definitivo de la mayoría de las técnicas multidimensionales.

Además de la Escuela Anglosajona, surgen nuevas escuelas de pensamiento dentro de la Estadística, algunas de las cuales se consideran muy relevantes: la Escuela India (con Mahalanobis, Rao, Krishnaiah, etc.); la Escuela Francesa (con Benzécri, Lebart, Morineau, Fénelon, etc.); y la Escuela Sueca (Jöreskog y Sörbom).

El empleo de los ordenadores en el proceso de análisis de datos constituyó un paso definitivo para la aplicación del Análisis Multivariante a disciplinas

como Psicología, Medicina y Economía. A partir de los años 80, cuando la informática empieza a ser de dominio público, cualquier investigación de carácter empírico en la mayoría de las disciplinas plantea la conveniencia de utilizar Análisis Multivariante, lo que ha contribuido de manera decisiva a la aparición de un gran número de técnicas. En la mayoría de los casos una nueva técnica surge cuando es preciso resolver un problema concreto en un ámbito de estudio particular, pero luego su aplicación se extiende a otras disciplinas. De la aplicación a estos nuevos ámbitos, pueden surgir aportaciones que mejoren la técnica o que den lugar a otra nueva, y así sucesivamente.

1.2 Clasificación de técnicas

La variedad de técnicas y diversidad de opiniones sobre su adecuación y validez, hacen difícil el ofrecer una visión equilibrada y de conjunto de las mismas. Aunque la tarea es ardua y difícil, diversos autores han realizado intentos de dar clasificaciones lo más amplias posibles. Cada clasificación varía según el criterio empleado, y algunas de ellas tratan de ser un compendio de las ya propuestas por otros. A continuación, se recoge una relación de técnicas cuya intención es clasificarlas según los criterios más convenientes en relación a los objetivos pretendidos.

Las técnicas que se van a considerar se agrupan, en primer lugar, según el tipo de tabla de datos para el que son adecuadas. Así, se distingue entre *técnicas para tablas de doble entrada* y *técnicas para tablas de tres entradas*. La elección de este criterio responde a uno de los objetivos del capítulo que ya hemos señalado: realizar una primera caracterización de la metodología STATIS. Precisamente la principal característica de este método es su adecuación a tablas de tres entradas. Por otra parte, aunque como ya se pondrá de manifiesto, una tabla de tres entradas puede considerarse

como una sucesión de tablas de dos entradas, los métodos adecuados para ellas no están indicados para el tratamiento de las de tres entradas. Su aplicación, de manera individual a cada una de las tablas que componen la tabla total de tres entradas, plantearía el problema de comparación de los resultados individuales obtenidos, además de dejar de lado las relaciones que pueden establecerse entre las diferentes tablas de doble entrada. Por otro lado, el tratamiento de una tabla de tres entradas como si fuese una "supertabla" de doble entrada por un método propio para ese tipo de tablas, no consideraría la existencia de las "subtablas" que la componen. Por ello, parece lógico buscar procedimientos específicos para tablas de tres índices o entradas.

Las *técnicas para tablas de doble entrada* constituyen el Análisis de Datos tradicional y en muchos casos, las *técnicas para tablas de tres entradas* son una generalización de las primeras, dando lugar a una técnica específica que, sin embargo, comparte alguno de los objetivos y características con las primeras. Por tanto, aunque las *técnicas para tablas de doble entrada* no son objeto de este trabajo, la clasificación de las mismas sirve para establecer criterios, como el tipo de datos que tratan o el tipo de resultados que obtienen, que serán empleados en la clasificación de las que nos ocupan. En definitiva, se hará una breve descripción de aquellas *técnicas para tablas de doble entrada* que sirven de base a alguna de las *técnicas para tablas de tres entradas*.

1.2.1 Técnicas para tablas de doble entrada

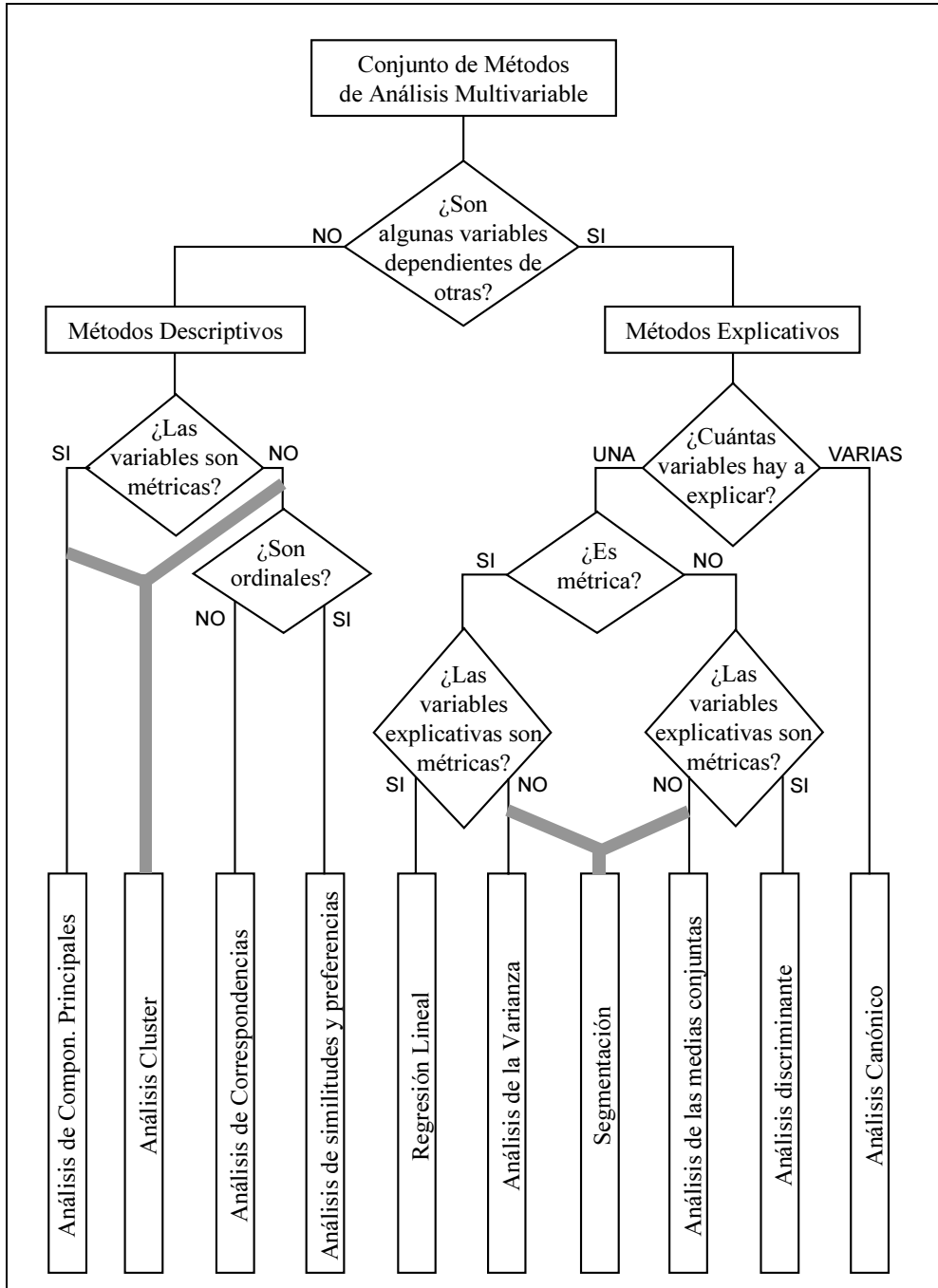
Estas tablas o matrices de datos representan los valores referidos a n individuos (objetos o unidades), respecto a p variables (o características), y por tanto sus elementos son de la forma x_{ij} ($i = 1, 2 \dots n; j = 1, 2 \dots p$). La terminología anglosajona las denomina "two-way data".

Una primera clasificación de los métodos o técnicas disponibles para este tipo de tablas se basa en los OBJETIVOS que se persiguen con el estudio. Así, encontramos las *técnicas descriptivas* (o *exploratorias*), con las que el investigador realiza una primera aproximación a los datos sin hipótesis de partida; y las *técnicas explicativas* (o *confirmatorias*), que parten de un marco teórico que justifica y fundamenta unas hipótesis que se intentan validar empíricamente. Entre las primeras pueden citarse el *Análisis de Componentes Principales* (ACP), el *Análisis de Correspondencias Simples* (ACS), las *Escalas Multidimensionales* (MS), y el *Análisis Cluster*. Entre las segundas se encuentran el *Análisis de Regresión*, el *Análisis de la Varianza* (ANOVA), el *Análisis de Correlaciones Canónicas* (ACC) y el *Análisis Discriminante*.

La siguiente clasificación se basa en el TIPO DE DATOS que se trata de analizar. Según que los datos sean cuantitativos, cualitativos o mixtos, se aplican unas técnicas u otras. Así, para datos cuantitativos tenemos el *Análisis de Componentes Principales*, *Análisis Cluster*, *Análisis de Regresión*, *Análisis de Correlaciones Canónicas* y *Escalas Multidimensionales* (métricas). Para datos cualitativos encontramos el *Análisis de Correspondencias* y las *Escalas Multidimensionales* (no métricas) y, en el caso de datos mixtos, son aplicables métodos como el *Análisis de la Varianza* y el *Análisis Discriminante*.

Esta clasificación puede ser ampliada en función del PAPEL SIMÉTRICO O NO DE LAS VARIABLES (en este caso, hay variables dependientes e independientes), dando lugar a la siguiente clasificación propuesta por Evrard y Le Maire en 1971:

Figura I.1
Clasificación de técnicas de Evrard y Le Maire

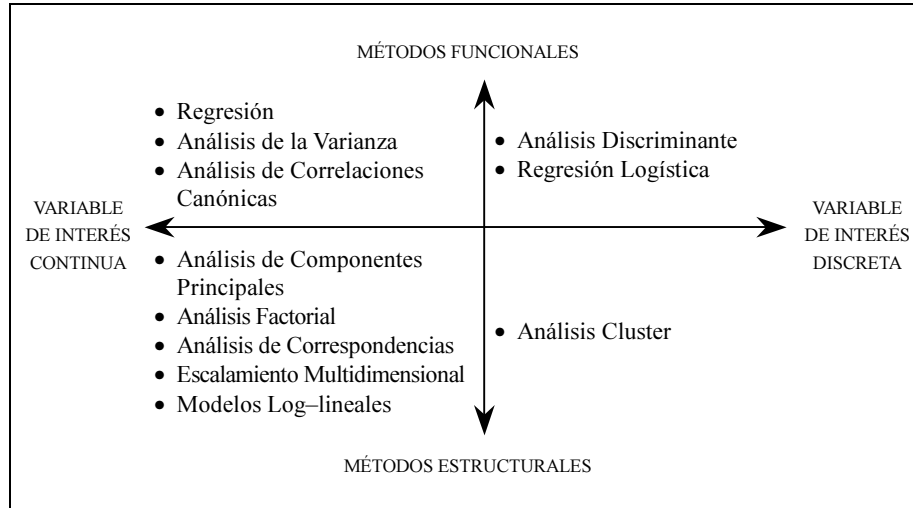


En función de los RESULTADOS que se persiguen, se distingue entre dos tipos de técnicas: *Técnicas Factoriales* y *Técnicas de Clasificación*. El objetivo de las primeras es encontrar nuevas variables (denominadas factores), en menor número que las variables iniciales, que permitan analizar la estructura de la tabla de datos, poniendo en evidencia las relaciones entre las variables, entre las observaciones o entre ambas. Las dos técnicas más importantes son el *Análisis de Componentes Principales* y el *Análisis de Correspondencias*. Dentro de este enfoque, el *Análisis Factorial (AF)* puede considerarse el marco general de las dos técnicas señaladas pero, en realidad, este método parte de un modelo que no se tiene en cuenta para la aplicación de los dos anteriores.

Las *Técnicas de Clasificación* se caracterizan porque proporcionan una o varias particiones del conjunto de observaciones o del conjunto de variables; se consideran, en muchos casos, como un complemento indispensable de las factoriales, y los factores obtenidos con éstas pueden emplearse como punto de partida de las técnicas de clasificación.

Una clasificación más actual es la que se puede encontrar en Greenacre (1999: 1-9). Este autor considera que el tipo de datos y el objeto del estudio determinan el método a aplicar y realiza una clasificación de las técnicas en lo que él denomina "las cuatro esquinas del Análisis Multivariante". Considera la existencia de dos tipos de métodos: *Métodos Funcionales* y *Métodos Estructurales*. Los primeros son aquéllos que consideran una variable a explicar y los *Métodos Estructurales* son los que consideran todas las variables por igual. Teniendo en cuenta si la variable de interés es discreta o continua se catalogan las diversas técnicas según muestra la figura siguiente:

Figura I.2
Clasificación de técnicas de Greenacre



Como puede comprobarse, esta clasificación, aunque más actual, no supone una gran diferencia respecto a las ya anteriormente citadas. Como ya hemos indicado, clasificaciones más complejas pueden ser dadas en un intento de recoger todas las técnicas existentes, si bien, para el objetivo propuesto, a nuestro juicio, no son necesarias.

De manera resumida, vamos a referirnos a algunas de las técnicas mencionadas más extensamente difundidas. El *Análisis de Componentes Principales* se desarrollará ampliamente en el siguiente capítulo en cuanto que proporciona los desarrollos teóricos que sirven de base a la metodología STATIS. En cuanto al *Análisis de Correspondencias*, es una técnica para el estudio de las relaciones entre dos variables nominales, que utiliza representaciones gráficas principalmente. Se trata de un doble análisis de componentes principales, por un lado, de las filas (que son las modalidades de una variable); y por otro lado, de las columnas (que son las modalidades de la otra), cumpliéndose ciertas propiedades especiales debido a la naturaleza de las mismas.

El *Escalamiento Multidimensional*, fue propuesto por Togerson, en primer lugar, y popularizado por Gower, Shepard y Kruskal, en los años sesenta. Dada la información acerca de las disimilaridades entre individuos, los

métodos de escalamiento tratan de encontrar coordenadas de los puntos, de modo que, cuanto más pequeña sea la disimilaridad entre dos puntos, más próximos estarán en el gráfico resultante.

La denominación *Análisis Cluster* se utiliza para definir una serie de técnicas estadísticas, fundamentalmente algoritmos, que permiten la obtención de grupos de elementos similares. Consiste en clasificar una serie de individuos, sobre los que se ha medido una serie de variables, en grupos lo más homogéneos posible en base a las variables observadas. Este análisis será desarrollado posteriormente de manera más amplia.

El *Análisis de Correlaciones Canónicas* (ACC) fue presentado por primera vez por Hotelling (1936: 129-149). En este método el número de grupos está limitado a dos y consiste en buscar, simultáneamente, una combinación lineal de las variables del primer grupo (variable canónica del primer grupo) y una combinación lineal de las variables del segundo grupo (variable canónica del segundo grupo) de forma que el coeficiente de correlación entre estas dos combinaciones sea máximo. Posteriormente, se busca una segunda pareja de combinaciones lineales, ortogonal a la primera (es decir, incorrelada con la primera), que satisfaga el mismo criterio, y así sucesivamente. Además de las propiedades teóricas que presenta, esta técnica constituye un marco general que unifica otra serie de técnicas, según recoge Bouroche y Saporta (1987).

Una referencia al *Análisis de Regresión* se hará al hablar de las técnicas de predicción, aunque también puede considerarse como una técnica multidimensional.

El *Análisis de la Varianza* es un procedimiento, creado por Fisher en 1925, para descomponer la variabilidad de un experimento en componentes independientes que pueden asignarse a causas distintas. Se dispone de n elementos que se diferencian en un factor. En cada uno se observa una característica continua que varía aleatoriamente de un elemento a otro y lo

que se desea conocer es si hay o no relación entre el valor medio esperado de la característica estudiada y el factor que los diferencia.

El *Análisis Discriminante* es una técnica que permite asignar o clasificar nuevos casos dentro de grupos previamente reconocidos o definidos. El punto de partida es una muestra de individuos sobre los que se ha medido un conjunto de variables. Dicha muestra ha sido dividida en dos o más grupos diferentes. Posteriormente, se considera un nuevo caso en el que son medidas las mismas variables y, según los valores de éstas, el nuevo individuo es asignado al grupo de máxima probabilidad, cuantificando, además, el valor de dicha probabilidad.

1.2.2 Técnicas para tablas de tres entradas

Las tablas de tres entradas pueden considerarse como una sucesión de tablas de doble entrada en la que cada una se refiere a un momento o situación diferentes. Se expresan mediante la siguiente notación $X^1, X^2 \dots X^T$, cuando las tablas se refieren a distintos momentos del tiempo (datos longitudinales), y por $X^1, X^2 \dots X^K$, cuando las diferentes tablas se refieren a situaciones no temporales. Por este motivo, para referirse al análisis de tablas de tres entradas, también se emplea la denominación de *Análisis Simultáneo de Varias Tablas de Datos*.

En estas tablas, los datos están clasificados según tres índices (modos o entradas) y los datos elementales son de la forma x_{ij}^t ($i = 1, 2 \dots n$; $j = 1, 2 \dots p$; y $t = 1, 2 \dots T$), considerando n individuos (objetos o unidades), p variables (características) y T situaciones (ocasiones o momentos del tiempo diferentes). La Escuela Francesa denomina a estas tablas, de forma general, "Tableaux à Trois Indices", mientras que la terminología anglosajona distingue entre las "three-way data arrays", para matrices completas o con los modos totalmente cruzados; y las "three-way o

multiway data sets”, para matrices que no tienen los tres modos totalmente cruzados, englobando ambas bajo la denominación común de “three-mode data”.

El objeto de estas distinciones es determinar la técnica adecuada, ya que varía según el tipo de tabla a estudiar. En algunos casos, la ambigüedad a la hora de definir el tipo de tabla tratada permite ser más flexible a la hora de aplicar un método u otro. Según Kiers (1988; 15), las técnicas para el análisis de tablas de tres entradas que no están totalmente cruzadas también pueden ser aplicadas para el análisis de matrices completas, si bien, en este caso, no se emplea toda la información disponible en la matriz completa. Por el contrario, las tablas no cruzadas totalmente no pueden ser analizadas por los métodos propuestos para tablas completas.

Antes de referirnos a los métodos o técnicas específicas para cada tipo de tablas, vamos a abordar algunas cuestiones generales. Según establece Coppi (1994: 7), la distinción tradicional entre *Técnicas de Factores* y *Técnicas de Escalamiento* que realizan Snyder y Law (1984: 4), es bastante restrictiva y no tiene en cuenta el creciente número de contribuciones al campo del Análisis Multivariable. Una distinción más general es la que se refiere al uso o no de modelos de probabilidad, enfoque que enlaza con la naturaleza muestreable o no de los datos. Si se emplea un Modelo de Probabilidad, el objeto principal es la obtención de los parámetros del mismo, o la realización de adecuadas reparametrizaciones.

Sin embargo, según Coppi (1994: 8), incluso en algunas estructuras de muestreo no es estrictamente necesario adoptar un enfoque probabilístico, puesto que, en algunos casos, el componente estocástico es un factor minoritario en la determinación de la variación de los datos. En estos casos, la parte paramétrica (estructural) del modelo explica la mayor parte de la variación. Autores como Kroonenberg (1994: 73-96) y Harshman y Lundy (1984: 122-215) plantean técnicas basadas en mínimos cuadrados que permiten la estimación de los parámetros del modelo estructural.

Cuando los modelos de probabilidad no juegan un papel fundamental en el análisis, esto es, desde una Perspectiva Descriptiva–Exploratoria, también es posible extraer información esencial de los datos y, si es necesario, tratar el componente estocástico adoptando procedimientos de remuestreo, sin necesidad de una formalización probabilística.

Es importante señalar que, durante las últimas décadas, se han desarrollado muchos métodos para analizar tablas de tres entradas según este enfoque. Algunos de éstos se han descrito en la literatura anglosajona y otros en la francesa. Sin embargo, muchos de los métodos desarrollados por la escuela francesa han sido ignorados por la literatura anglosajona y a la inversa, por lo que los métodos franceses se han desarrollado casi independientemente de los homólogos anglosajones.

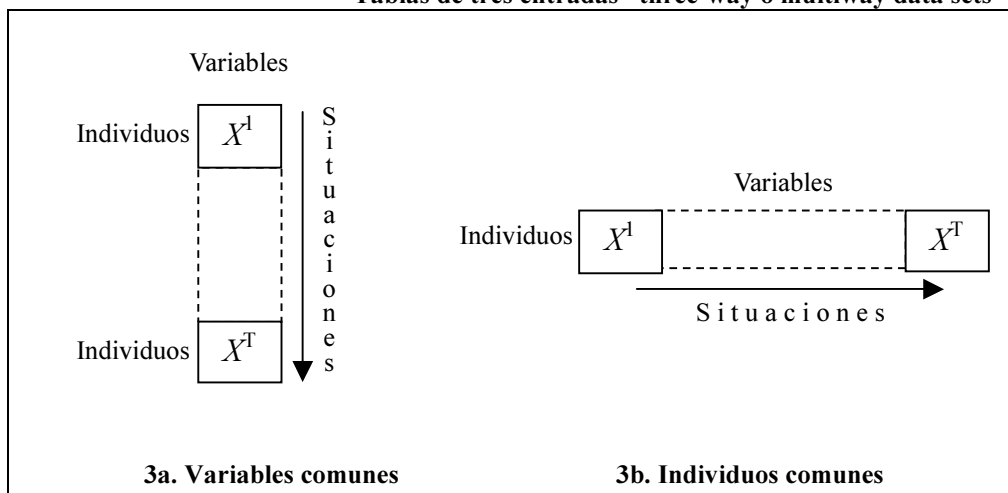
Ahora bien, a la hora de analizar ciertos datos se plantea el problema de elegir entre diferentes métodos que, ya sean franceses o anglosajones, sean adecuados al tipo de datos y a los objetivos de la investigación. Por este motivo, la clasificación propuesta, basada en la recogida en Bove y Di Ciaccio (1994: 15-37), no tiene en cuenta la Escuela de pensamiento a la que pertenece, sino los dos criterios que acabamos de mencionar, así como el tipo de tabla de tres entradas, poniendo especial énfasis en los métodos para datos cuantitativos y en los que consideran el aspecto temporal de la información.

a. Tablas “three-way o multiway data sets”

Cuando sobre un conjunto de individuos se observan distintos grupos de variables (agrupadas en varias tablas), o un conjunto de variables es medido sobre distintos grupos de individuos (agrupados en tablas diferentes), la tabla total resultante en ambos casos es una three-way o multiway data sets. Los modos o criterios de clasificación no están totalmente cruzados, puesto

que las variables o los individuos, según el caso, pueden ser diferentes de unas tablas a otras. La matriz resultante se denomina "multiway data set" tal y como señala Kiers (1988: 16). En la figura I.3 se representan las dos formas que puede presentar dicha matriz. Cuando existe un grupo de variables y varios grupos de individuos, la representación es la correspondiente a I.3a, y en I.3b se presenta la situación de un grupo de individuos y varios grupos de variables.

Figura I.3
Tablas de tres entradas "three-way o multiway data sets"



La clasificación de las técnicas idóneas para analizar este tipo de tablas se realiza considerando el criterio del tipo de datos, por lo que se distinguirán técnicas para datos cuantitativos, cualitativos y mixtos.

Datos cuantitativos

Las diferentes técnicas se van a especificar según la forma (individuos comunes o variables comunes) de la tabla de datos:

Individuos comunes

Se distinguen los siguientes métodos dependiendo del objeto del análisis:

- Si el objetivo es la cuantificación y descripción de las relaciones entre los diferentes grupos de variables (todos con un papel simétrico), la técnica adecuada es el *Análisis Canónico Generalizado* (ACG), desarrollado por la Escuela Anglosajona, que es una generalización del *Análisis de Correlaciones Canónicas* (ACC) ya comentado. Estas generalizaciones del ACC a más de dos grupos de variables, denominado también *Análisis Multicanónico*, han sido realizadas por varios autores. Destaca especialmente la propuesta efectuada por Carroll (1968: 227-228). En este caso, el ACG considera un grupo de variables canónicas para cada grupo de variables originales y un solo conjunto de variables generales para todos los grupos. Estas últimas marcan las tendencias generales de los grupos. Las variables generales son obtenidas haciendo máxima la suma de los cuadrados de los coeficientes de correlación múltiple entre ellas y los grupos, con restricción de ortogonalidad entre las variables generales. Las variables canónicas de cada grupo se obtienen maximizando la correlación entre la variable general y la combinación lineal de cada grupo más relacionado con dicha variable general. Así, las variables canónicas pueden interpretarse como la representación de la variable general en los grupos y la proyección de la variable general sobre el espacio engendrado por ese grupo.

Por otra parte, dado que el papel simétrico de las variables en la práctica casi nunca se verifica, versiones asimétricas del ACC han sido propuestas por Van den Wollenberg (1977: 207-219), bajo la denominación de *Análisis de la Redundancia*. Este autor propone un índice, denominado índice de redundancia, que explica el porcentaje de varianza de las variables de un grupo explicada por las del otro grupo. La redundancia, a diferencia de la correlación, no es simétrica: la redundancia asociada a un grupo de variables Y respecto un grupo

X puede ser diferente a la redundancia de las variables del grupo X respecto el grupo Y . El índice de correlación canónica, que es el que se hace máximo en un ACC, sólo mide parte de esa redundancia. Este análisis, a diferencia del ACC, tiene como objetivo maximizar la redundancia total recogida por el índice del mismo nombre.

- Cuando el objetivo principal es la definición de tipologías de objetos, o bien, la estrategia global requiere el análisis conjunto y separado de los tres modos de una manera explícita, nos encontramos ante las técnicas que siguen el enfoque Interestructura–Compromiso–Intraestructura (ICI), procedentes de la Escuela Francesa, tal y como se indica en Glaçon (1981: 22-26).

La interestructura, el compromiso y la intraestructura indican la denominación que reciben las tres etapas que siguen los métodos de dicho enfoque. La primera de ellas tiene como objeto el estudio de la relación entre grupos de variables desde un punto de vista global. La segunda y la tercera fase se basan en un ACP de la matriz obtenida ponderando y yuxtaponiendo las matrices de datos asociadas a cada grupo de variables. Así, el análisis del compromiso tiene como objeto el estudio de las proximidades entre individuos, las cuales son obtenidas como media ponderada de las proximidades en cada grupo de variables. La tercera fase, el análisis de la intraestructura, se refiere al estudio analítico de las relaciones entre variables y las proximidades entre individuos en los diferentes grupos empleando la configuración obtenida en la etapa anterior. Cuando los grupos de variables corresponden a diferentes puntos del tiempo ordenados, la unión de los puntos que representan los distintos objetos dan lugar a una trayectoria que facilita la interpretación de la configuración obtenida.

Los dos métodos más importantes según este enfoque son la metodología STATIS y el *Análisis Factorial Múltiple (AFM)*. El desarrollo de la primera se realizará en el siguiente capítulo, puesto que constituye el eje central de nuestro trabajo.

El AFM, desarrollado por Escofier y Pagès (1992: 117-176), procedentes de la Escuela Francesa, permite el tratamiento simultáneo de una serie de tablas con los mismos individuos y diferentes variables (una tabla de tres entradas con los modos sin cruzar totalmente) y constituye una alternativa al ACG. El AFM también puede ser empleado para analizar tablas de tres entradas formadas por diferentes individuos y un mismo grupo de variables, intercambiando los papeles de filas y columnas, aunque esta opción rara vez se considera en la práctica.

En primer lugar, el AFM normaliza las tablas de datos $X^1, X^2 \dots X^k$ de tal manera que si se realizase un ACP de cada tabla X^k , la primera componente principal obtenida en cada análisis explicaría la misma cantidad de inercia (uno en todos los casos). Esto se consigue multiplicando todas las variables de cada tabla de datos por el inverso del mayor valor propio obtenido en el ACP individual de dicha tabla. Por tanto, el elemento representativo de cada una de esas tablas es la matriz de productos escalares entre individuos, dividido entre el correspondiente valor propio. El compromiso será la suma de dichos elementos representativos de cada tabla.

En segundo lugar, se realiza un ACP sobre la tabla resultante de yuxtaponer las $X^1, X^2 \dots X^k$ que han sido ponderadas de la manera indicada. Los ejes son obtenidos de tal manera que la nube de individuos, definida para el conjunto de variables ponderadas, es de inercia máxima al proyectarse sobre esos ejes. Estos ejes se interpretan gracias a sus correlaciones con el conjunto de variables. Este ACP permite estudiar las relaciones entre las variables y los individuos-medios (etapa de intraestructura). Los ejes y factores obtenidos en esta segunda etapa son los que luego sirven de base para el análisis de la interestructura. En dicha etapa, se proyectan los elementos representativos de cada tabla sobre el cuadrado de los factores de la etapa de la intraestructura. Esta proyección se realiza haciendo máxima la suma de las proyecciones de todas las tablas. La

relación de los factores con las variables originales, dado que son los de la intraestructura, permite interpretar la posición de las tablas, y la coordenada de cada tabla coincide con la contribución de esa tabla a la inercia de la componente principal correspondiente.

Llegados a este punto, conviene hacer una distinción entre el enfoque ICI y el ACG, puesto que, en éste último, también se podría hacer referencia a la construcción de un compromiso. En el caso de ACG, rara vez se realizan representaciones gráficas similares a las obtenidas en la etapa de la interestructura en los otros métodos, por lo que, respecto a ese punto, no es necesario comparación alguna.

La aplicación del ACG está más justificada cuando el principal objeto de estudio es el análisis de las relaciones entre grupos de variables, y los objetos o individuos se analizan desde ese punto de vista. En el caso de los métodos ICI, los objetos y las variables tienen la misma importancia en el análisis. La determinación del compromiso según el ACG, está más influenciada por la organización de las variables en grupos que en los métodos del enfoque ICI. En el caso del ACG, los ejes del espacio (variables generales) están determinadas maximizando la correlación múltiple entre los conjuntos de variables canónicas, mientras que, en el otro caso, lo que se maximiza es la inercia explicada de las variables en la dirección de los componentes del compromiso.

- Otro enfoque diferente, procedente de la Escuela Anglosajona, es el que aporta el *Análisis Procrusteano Generalizado (APG)*. Propuesto por Gower (1975: 33-51) considera la construcción del compromiso de la intraestructura como un problema de comparación de configuraciones obtenidas separadamente. Para comparar las diferentes nubes de puntos-individuo, asociadas a las tablas de datos iniciales, el APG transforma estas matrices (busca desplazamientos de las nubes estudiadas) de forma que estén lo más próximas posible unas de otras. Esto es, conservando los puntos de una nube sus posiciones relativas, se intenta aproximar los individuos homólogos. Los desplazamientos

consisten en transformaciones ortogonales de las matrices originales. Previamente a estos desplazamientos, las matrices son centradas, coincidiendo todos los centros de gravedad, y normalizadas, teniendo todas ellas la misma inercia.

La solución se basa en el denominado teorema de la "Descomposición de Eckart-Young de una matriz real". El método de resolución de Gower es iterativo y puede ser enfocado como si de un *Análisis de la Varianza* se tratase. Una vez desplazadas las nubes iniciales y orientadas de manera análoga, éstas se representan en espacios de reducida dimensión a través de un ACP de todos los puntos correspondientes a todas las nubes.

Variables comunes

Cuando se trata de este tipo de tablas de tres entradas, los métodos antes considerados pueden también aplicarse realizando las precisiones oportunas. Además, muchas generalizaciones del *Análisis de Componentes Principales* se han propuesto para estos casos, por ejemplo, el *Análisis de Componentes Simultáneo*, propuesto por Millsap y Meredith en 1988 y denominado así por Kiers y Ten Berge en el mismo año, que es precisamente una sencilla generalización del ACP. Teniendo en cuenta que los individuos son distintos y, por tanto, también su ponderación, en este método la proporción de inercia es maximizada en todas las ocasiones simultáneamente, y las mismas ponderaciones son aplicadas a todas las variables en todas las situaciones, sin importar el grupo al que pertenecen, teniendo los componentes el mismo significado en todas ellas.

Otro método es el *Análisis Factorial Simultáneo* (AFS), propuesto por Levin en 1966. Se describe como un ACP de una supermatriz formada por la yuxtaposición de las diferentes tablas apiladas unas debajo de otras. Según Kiers (1991: 460), es equivalente a una de las etapas en el modelo de Tucker de 1966, que a su vez equivale a una de las etapas del método

de Jaffrenou para el análisis de tablas totalmente cruzadas. Estos dos autores puntualizaron que un ACP de dicha supermatriz no produce la misma solución que el ajuste del modelo de Tucker, pero puede considerarse una buena aproximación en el sentido de mínimos cuadrados.

Este método, denominado SUMPCAc (Kiers, 1991: 460), también puede considerarse como una aplicación a una suma de matrices de productos cruzados. El modelo desarrollado por Gower en 1966, también puede considerarse equivalente a este método cuando se consideran las matrices de productos cruzados como matrices de similitudes.

La diferencia entre el *Análisis de Componentes Simultáneo* y el *Análisis Factorial Simultáneo*, según señala Kiers (1988: 22, 28), es que el primero intenta buscar componentes para las matrices de datos originales que expliquen de manera óptima cada matriz, y el segundo busca componentes que expliquen de manera óptima las variables como si hubiesen sido obtenidas de una gran población, de tal manera que cada tabla representaría las distintas subpoblaciones. Es decir, el *Análisis de Componentes Simultáneo* permite considerar los diferentes grupos de individuos como observaciones separadas.

Datos cualitativos y mixtos

En el caso de las variables cualitativas, vamos a hacer una revisión de los métodos más utilizados en el tratamiento de tablas de tres entradas más breve que para datos cuantitativos, en cuanto que no son objeto del presente trabajo.

La presencia de una variable cualitativa permite estudiar la tabla de datos de distintas maneras. Vamos a ir distinguiendo las distintas maneras de organizar dichas tablas y las técnicas apropiadas para cada caso, teniendo en cuenta que la tabla original no tiene los modos (objetos, variables y ocasiones) totalmente cruzados, si bien, cualquiera que sea la estructura

que se analice, es necesario suponer que, bien los objetos, bien las variables, son comunes.

Tabla de individuos por variables

Las variables pueden ser sólo cualitativas, o cualitativas y cuantitativas (datos mixtos). Si las variables están agrupadas en varios grupos, se obtiene una tabla de tres entradas con los mismos individuos y diferentes variables en cada ocasión. También es posible definir una tabla de tres entradas en la que las variables son las mismas pero los individuos son diferentes en cada ocasión. Si se tiene una variable cualitativa (con K categorías) mientras que las otras son cuantitativas, podemos considerar las K modalidades de esta variable cualitativa para definir K ocasiones y aplicar los métodos ya descritos para variables cuantitativas. Si esto no es posible o hay más variables cualitativas, son necesarios métodos específicos, como los que se citan a continuación.

Además del *Análisis de Correlaciones Canónicas no lineal* (Van Der Burg, en 1988), el *Análisis de la Redundancia cualitativo y no lineal* (Van Der Burg y De Leeuw, en 1990), y el *Análisis de Correspondencias parcial y condicional* (Escofier, en 1988), los métodos de mayor aceptación en el tratamiento de este tipo de tablas, debido a la existencia de software apropiado desde hace tiempo, son el *Análisis Canónico no lineal*, con el programa OVERALS, y el *Análisis Factorial Múltiple* (AFM). Ambos métodos analizan una matriz de datos en la que las variables pueden variar de una tabla a otra pero no los individuos.

El método OVERALS se deriva del *Escalamiento Multidimensional*, de ahí que pueda considerarse como un *Análisis de Correlaciones Canónicas* en K -grupos con *Escalamiento Multidimensional*. El OVERALS analiza variables nominales (simples o múltiples), ordinales o cuantitativas. Si todas las variables son cualitativas en todos los grupos el AFM, puede considerarse

como una extensión del *Análisis de Correspondencias Múltiples (ACM)*, que es una generalización del ACS a más de dos variables nominales. En este caso, el AFM y el ACM son equivalentes y el programa OVERALS proporciona idénticos resultados.

En otros casos, AFM y OVERALS producen resultados diferentes. Si cada grupo tiene sólo cierto número de variables cualitativas, el AFM pondera cada uno de forma diferente y tampoco hay equivalencia con el ACM, interpretándose el AFM como un ACP ponderado de indicadores de variables categóricas. Además el AFM no puede considerar el aspecto ordinal de los datos.

Extensiones del STATIS para variables cualitativas fueron realizadas por Escoufier (1980: 59-76), pero no se conoce la existencia de software disponible.

Tabla de contingencia múltiple

Este tipo de tablas pueden obtenerse a partir de una o más tablas de objetos por variables cuando todas ellas son cualitativas. Hay que tener en cuenta que, en esta transformación a datos de frecuencia, se pierde la posibilidad de asignar etiqueta a cada objeto. Es complicado determinar si una tabla de contingencia múltiple es en realidad una tabla de tres entradas "multiway data set". En realidad, tres variables cualitativas son suficientes para obtener una tabla de contingencia a tres entradas con los modos totalmente cruzados, pero sólo si una de esas variables asume el papel de "ocasión" encontramos una "multiway data set".

Cuando se observan las mismas variables cualitativas en diferentes grupos de objetos, tenemos una tabla de contingencia múltiple para cada grupo. En este caso podemos obtener una tabla de contingencia única uniendo los distintos grupos de objetos e introduciendo en cada uno de

ellos una variable "artificial" para distinguirlos. Se puede considerar que esta variable desempeña el papel de "ocasión".

La limitación general para el análisis de este tipo de tablas es que se pierde la identificación de los objetos. El análisis de una tabla de contingencia múltiple se ha desarrollado particularmente con el enfoque probabilístico y los métodos más frecuentemente empleados son los *Métodos Log-lineales y Logísticos* (Cox, 1972) y el *Análisis de Clase Latente* (Goodman, 1974). Un método propuesto recientemente es el *Modelo de Aproximación Log-no-lineal*, del mismo autor, en el que explícitamente se introduce una cuantificación de las variables categóricas. Sin embargo, estos métodos tienen fuertes limitaciones porque son esencialmente métodos para el tratamiento de tablas de dos entradas.

Un método de análisis de datos específico para este tipo de tablas a tres entradas es el *Análisis de la Estructura Simultánea Latente*, de Clogg y Goodman, que permite obtener correctamente la estructura latente para este tipo de datos. Este método puede emplearse para analizar las restricciones de homogeneidad en los parámetros de los modelos de cada tabla.

La ventaja de estos métodos es que tienen un enfoque probabilístico y permiten analizar la exactitud de los resultados obtenidos, algo que no permiten otras técnicas, a no ser mediante métodos de remuestreo (Bootstrap o Jackknife).

En la tabla siguiente se esquematizan los diferentes métodos descritos según los criterios considerados:

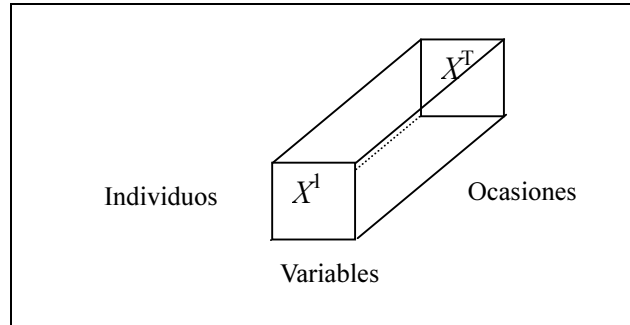
Tabla I.1
Métodos para tablas “three-way data sets”

TIPO DE DATOS	FORMA DE LA TABLA	TÉCNICA
DATOS CUANTITATIVOS	Individuos comunes	<i>Análisis Canónico Generalizado</i> <i>Análisis de la Redundancia</i> <i>Métodos ICI:</i> – <i>Statis</i> – <i>Análisis Factorial Múltiple</i> <i>Análisis Procrusteano Generalizado</i>
	Variables comunes	<i>Statis dual</i> <i>Análisis de Componentes Simultáneo</i> <i>Análisis Factorial Simultáneo</i>
DATOS CUALITATIVOS Y MIXTOS	Individuos por variables	<i>Análisis de Correlaciones Canónicas no lineal</i> <i>Análisis de la Redundancia cualitativo y no lineal</i> <i>Análisis de Correspondencias parcial y condicional</i> <i>Análisis Factorial Múltiple</i> <i>Análisis Canónico no lineal (OVERALS)</i> <i>Análisis de Correspondencias Múltiple</i>
	Tabla de contingencia múltiple	<i>Métodos Log-lineales y Logísticos</i> <i>Análisis de Clase Latente</i> <i>Modelo de Aproximación Log-no-lineal</i> <i>Análisis de la Estructura Simultánea Latente</i>

b. Tablas “three-way data arrays”

Cuando los modos están totalmente cruzados, es decir, cuando en cada tabla se han observado los mismos individuos respecto a las mismas variables, nos encontramos ante las matrices completas de tres entradas o “Three-way arrays”. Un ejemplo de este tipo de datos lo constituyen aquellos casos en los que se observa un grupo de variables sobre los mismos objetos en diversos periodos del tiempo. Los datos pueden presentarse por medio de una matriz de tres entradas completa o “array”, que presenta la siguiente forma, según se muestra en el artículo de Kiers (1988: 16):

Figura I.4
Tabla de tres entradas completa “three-way data array”



Las técnicas idóneas para este tipo de tablas se clasifican según el tipo de datos, por lo que se distinguirán las técnicas para datos cuantitativos, cualitativos y mixtos, siguiendo el mismo esquema que para las tablas anteriores.

Datos cuantitativos

Aunque este tipo de tablas pueden analizarse con los métodos para datos cuantitativos señalados para las tablas cuyos modos no están totalmente cruzados, según Bove y Di Ciaccio (1994: 21), algunos de dichos métodos pierden sus peculiaridades cuando se aplican a tablas totalmente cruzadas. Tal es el caso del ACG, en el que la relación entre grupos es menos interesante si los grupos están formados por las mismas variables.

El sistema de ponderaciones adoptado por los métodos ICI también es menos interesante cuando las variables son las mismas. Sin embargo, en este caso, una nueva característica de la infraestructura es que permite dibujar la trayectoria, no sólo de los individuos, sino también de las variables, cuando los periodos de tiempo están ordenados. Así, las evoluciones de los objetos y de las variables, y sus relaciones, pueden representarse en la configuración de la infraestructura. Como ya hemos indicado, el *STATIS* pertenece a este enfoque y, por tanto, la aplicación del mismo a este tipo de tablas es especialmente útil si el estudio quiere completarse con el análisis de la evolución de las variables (*Statis*) y de los individuos (*Statis*)

dua). Además, como se pondrá de manifiesto en este estudio, es posible la aplicación de las dos versiones para obtener resultados complementarios.

Los métodos propuestos para este tipo de tablas se dividen en dos grupos: los métodos basados en *Modelos de Componentes* y los métodos basados en *Modelos de Factores Comunes*. Los primeros son básicamente analíticos y exploratorios (los tres modos se consideran fijos), y los segundos pueden considerarse más probabilísticos y confirmatorios. Además, los parámetros de los segundos se suelen obtener ajustando las matrices de covarianzas o correlaciones más que los datos originales, por lo que serán expuestos al analizar los modelos para tablas derivadas.

Métodos de Componentes

Dentro estos métodos, Tucker (1966: 279-311) introdujo un nuevo enfoque factorial basado en la idea de que una estructura subyacente (factorial) estaba asociada a cada modo, y no hay un solo grupo de factores, sino tres. Los tres modos observados de los datos están asociados, cada uno, con otro modo (no observable) que puede considerarse como un conjunto de factores o categorías ideales. Así, los objetos, variables y ocasiones son considerados como combinaciones lineales de los objetos "idealizados", de las variables "latentes" y de las ocasiones "prototipo", respectivamente. Las relaciones entre los tres tipos de factores son consideradas en una tabla o matriz estimada por el modelo. Una nueva versión fue considerada por Tucker (1972: 3-27) para aquellos casos en los que uno de los tres modos no puede ser reducido (por ejemplo, en las tablas indizadas por el tiempo).

El modelo TUCKALS2 fue desarrollado por Kroonenberg y De Leeuw en 1980, sólo para tablas de tres entradas con los modos totalmente cruzados, sin posibilidad de aplicarlo a los otros tipos de tablas. El modelo subyacente al TUCKALS2 se deriva del modelo de Tucker de 1966. Asimismo, dichos autores desarrollaron un modelo más general, el TUCKALS3, que consiste en

ajustar el modelo original de Tucker en el sentido de mínimos cuadrados. Este método es totalmente diferente a los otros métodos considerados porque, no sólo reduce el número de objetos y el número de variables a un número reducido de objetos ideales y variables latentes, sino que reduce también el número de ocasiones a un número menor de ocasiones ideales.

No obstante, según Kiers (1989: 16) el TUCKALS3 puede considerarse como una variante restringida de la primera fase de la metodología STATIS (STATIS1, según Kiers, 1991: 450) en la que se realiza un ACP de las matrices que recogen cada grupo de variables. Por otra parte, el TUCKALS3 permite obtener una representación de los objetos que no posibilita el análisis de la interestructura del STATIS, y proporciona medidas que indican la relación entre objetos y variables, aunque el inconveniente que presentan es que dichas relaciones tienen una difícil interpretación. Los algoritmos desarrollados por Tucker pueden ser aplicados por el programa TUCKALS, elaborado por Kroonenberg (1994: 73-96).

Las dificultades para interpretar la tabla o matriz resultado del método de Tucker dieron lugar a nuevas aplicaciones considerando modelos basados en hipótesis más sencillas. Carroll y Chang, en 1970, y Harshman, en 1970, desarrollaron, aunque de una manera independiente, un modelo que descomponía las tablas de tres entradas con los modos totalmente cruzados de una manera muy simple. Carroll y Chang lo denominaron CANDECOMP (Canonical Decomposition) y Harshman lo denominó PARAFAC (Parallel Factor Analysis). El método de Carroll y Chang puede aplicarse usando el programa INDSCAL desarrollado por los mismos autores. Estos dos métodos están basados en el mismo modelo trilineal de factores, pero no tiene el problema de rotación. A diferencia de la concepción de Tucker, sólo un grupo de factores subyace a los datos observados y puede ser interpretado por los tres modos simultáneamente. El inconveniente es que los ejes sólo pueden ser interpretados en un único modo, dado que no permite rotaciones de los mismos.

Se puede considerar que este método es un caso especial del TUCKALS2 y que, mientras en éste los componentes se definen para las variables y los individuos, en CANDECOMP/PARAFAC los componentes se definen de modo simultáneo para ambos.

Otro método es el *Análisis Triadítico*, propuesto por Jaffrenou en 1978 para este tipo de tablas totalmente cruzadas. Consiste, además de otras etapas, en realizar un *Análisis de Componentes Principales* sin tener en cuenta los distintos momentos del tiempo. Cada matriz de individuos \times variables, que representa cada momento del tiempo, se expresa como un vector columna, por lo que el resultado es una matriz de un número de columnas igual al número de tablas (ocasiones) y un número de filas igual al producto de número de individuos y el de variables. Este ACP es lo que Kiers (1991: 451) denomina PCA-SUP (algunos autores lo denominan *Análisis Triadítico Parcial* o preSTATIS). Esta fase es equivalente al método usado por Tucker, en 1966, para encontrar una solución aproximada para cada ocasión en el modelo de análisis factorial de tres modos. Kroonenberg y De Leeuw lo han descrito como Tucker-1 puesto que uno de los modos es reducido. Como ya hemos indicado, esta etapa del método de Jaffrenou es equivalente al *Análisis Factorial Simultáneo*.

Según Kiers (1991: 456), también es posible realizar, sin más, un ACP de la suma de matrices $X^1, X^2 \dots X^T$ (que se puede denominar SUMPCA), y que considera que todas las tablas–ocasiones tienen los mismos pesos en cada componente. Es especialmente interesante si los datos son estables de unas ocasiones a otras.

El mismo autor (1991: 461) señala que la mayor parte de los métodos expuestos pueden considerarse una versión restringida del ACP de Jaffrenou ya descrito. Así, el SUMPCA es una versión restringida del CANDECOMP/PARAFAC que, a su vez, es una versión restringida del TUCKALS3, y éste, del PCA-SUP. En este caso, cuanto más alta sea la posición del método en la escala, mejor modelo proporciona. El inconveniente es que el número de parámetros a

estimar suele ser más elevado, y el modelo más complejo y con una más difícil interpretación. Por ello, es necesario establecer un equilibrio entre las ventajas y los inconvenientes a la hora de seleccionar el modelo adecuado.

Métodos para datos temporales o longitudinales

Cuando se trata de analizar datos longitudinales (tablas de datos referidas a distintos momentos del tiempo), se pueden aplicar otros métodos. Así, el *Doble Análisis de Componentes Principales* (DACP), desarrollado por Bouroche y Dussaix en 1975, tiene su aplicación en grupos de matrices compuestas por el mismo conjunto de individuos y de variables, medidas en diferentes momentos de tiempo. Estos autores suponen que, ante T matrices ($n \times p$) referidas a otros tantos momentos de tiempo, existen dos fenómenos distintos que hay que analizar:

- Un fenómeno de evolución global, esto es, de variación de los centros de gravedad de las nubes asociadas a cada tabla de datos: se trata de encontrar los ejes de evolución que diferencien lo mejor posible estas nubes. Se estudia la nube de los centros de gravedad y se busca una representación de éstos en un espacio de reducida dimensión, de forma que su dispersión esté representada lo mejor posible.
- Un fenómeno de deformación de las nubes de puntos de un instante a otro: se trata de comparar las dispersiones de las diferentes nubes. Para tal fin, se realiza, en una primera etapa, un ACP de cada una de las tablas centradas sobre su propio centro de gravedad. En una segunda etapa, se busca el sistema de ejes que resuma, de manera óptima, las diferencias y/o semejanzas entre los distintos ejes de inercia obtenidos anteriormente.

Una de las soluciones propuestas proporciona resultados similares al método *Statis dual* (GERI, 1996).

El *Análisis Factorial Evolutivo* (AFE) es un método desarrollado por Chevalier en 1977, que tiene su aplicación en datos cronológicos, cuyo objetivo es estudiar la evolución de los individuos. Partiendo de T matrices de datos referidas a distintos momentos de tiempo, X_t ($n \times p$), se construye una matriz global, de orden ($n \times pT$), yuxtaponiendo las T matrices dadas. Antes de proceder al análisis de esta tabla global, se ponderan los diferentes períodos de tiempo. Por ejemplo, los períodos alejados en el tiempo tienen un peso débil, mientras que los períodos actuales tienen un peso más fuerte. Se puede ponderar cada período t por la inercia al centro de gravedad de la nube asociada a dicho período, o por su inverso, en caso de privilegiar a los períodos de dispersión mínima. Para determinar los sujetos que evolucionan de manera parecida a lo largo del tiempo, se define una distancia de evolución ponderada. Así, la distancia de evolución entre dos individuos en un periodo es la distancia en ese momento multiplicada por el coeficiente de ponderación. Si se consideran todos los periodos, podemos decir que dos individuos presentan la misma evolución a lo largo del tiempo si están próximos en cada periodo t o lo están en media. Para que esa distancia sea interpretable, las métricas de todos los periodos deben ser del mismo tipo.

El AFE trata, por tanto, de realizar un ACP con distancia evolutiva de la matriz global. Los ejes factoriales obtenidos se interpretan como ejes de evolución global, y las proyecciones sobre los mismos indican claramente la evolución de cada individuo. Para completar el análisis, se pueden efectuar Análisis Factoriales Parciales de las tablas X_t , correspondientes a cada periodo de tiempo, y relacionarlos con el análisis global anterior. Para ello, se calcula la correlación entre las direcciones principales de cada periodo t y las direcciones principales de evolución global. Estas correlaciones permiten precisar los periodos determinantes para la evolución del conjunto de individuos, así como los periodos de estabilidad.

El *Análisis de Series Cronológicas Multidimensionales* (ASCM), de Tenenhaus y Priouret (1971), fue especialmente elaborado para el análisis de

datos temporales para tablas cruzadas formadas por los mismos individuos y por las mismas variables, y sigue un enfoque próximo al *Doble Análisis de Componentes Principales*. Así, a cada tabla X_t se le añade una columna (de tamaño n) en la que todas las componentes toman el valor t . En primer lugar, se trataría de estudiar la evolución simultánea de las p series cronológicas que corresponden a las p variables de cada tabla, poniendo en evidencia la evolución común a todas. Esta evolución común se denomina tendencia (h^t) y es de carácter polinomial. Este polinomio se obtiene realizando un ajuste, en el sentido de mínimos cuadrados, de los centros de gravedad de las tablas. La solución se denomina h^{t*} .

Posteriormente, en lo que podría llamarse análisis de la intraestructura, se realiza un ACP de las tablas de la forma $X_t - H^{t*}$ —siendo $H^{t*} = (h^{t*} \dots h^{t*})$ —, apiladas unas debajo de otras. Así, las posiciones de los individuos estarán medidas respecto a la tendencia y no respecto al centro de gravedad. Ahora bien, como la tendencia se ha obtenido como el polinomio más próximo a los centros de gravedad en el sentido de mínimos cuadrados, las soluciones obtenidas mediante este análisis no difieren fundamentalmente de los resultados obtenidos con el DACP, ni con los obtenidos por el *Statistical dual* cuando se emplean objetos normados. Por este motivo, se ha incluido como técnica descriptiva y no como técnica de predicción.

Recientemente y desarrollado por Pernin (1986), se ha propuesto otro nuevo método, el *método LONGI*, para el análisis de datos longitudinales que están recogidos en una tabla totalmente cruzada. Uno de los objetivos del método es encontrar combinaciones lineales de las variables (denominados índices de situación) que hacen máxima las diferencias entre objetos pero varían mínimamente entre los objetos de las diferentes ocasiones. Las matrices están ordenadas formando una supermatriz, de tal manera que cada vector columna corresponde a una variable, y contiene las puntuaciones de todas las combinaciones de individuos y ocasiones para dicha variable. Esa variable está centrada considerando la media para cada ocasión, pero se supone que esa transformación ya ha sido hecha. A continuación, se

realiza un Análisis Discriminante de tal manera que las funciones obtenidas maximizan la discriminación entre objetos. Esto lleva a realizar un Análisis de Correlaciones Canónicas sobre el conjunto de variables y sobre el conjunto de variables indicadores, señalando el objeto al que cada combinación de objetos y ocasiones se refiere.

Datos cualitativos y mixtos

Como ya indicamos, tres variables cualitativas son suficientes para obtener una tabla de tres entradas con los modos totalmente cruzados. Cuando se observan los mismos individuos para las mismas variables en T situaciones diferentes, se obtiene una tabla de variables \times objetos (con las variables agrupadas en T grupos), y se puede construir un grupo de tablas de contingencia múltiple. Se puede considerar como ejemplo aquel caso en el que los mismos datos sociológicos están observados, en distintos periodos de tiempo, sobre un grupo de individuos. Para estas situaciones, sería incorrecto transformar los datos en una única tabla de contingencia múltiple, porque cada individuo puede considerarse más de una vez en la tabla. Foucart (1978: 67-84) ha diseñado un método de análisis de un conjunto de tablas de contingencia indexadas en el tiempo, método disponible en el paquete estadístico ADE-4.

En este caso, la limitación general para el análisis de tablas de contingencia múltiple, es decir, la pérdida de la identificación de los objetos, es especialmente importante, puesto que un objetivo interesante podría ser la descripción del diferente comportamiento de los objetos en las T ocasiones. En este sentido, la transformación de la tabla de individuos \times variables original a una tabla de contingencia múltiple no es totalmente satisfactoria.

El análisis de un grupo de tablas de contingencia empleando un modelo log-bilineal, fue propuesto por Becker y Clogg en 1989. El empleo de modelos tri-lineales también ha sido propuesto por otros autores. El

enfoque probabilístico de estos métodos tiene la ventaja, ya señalada, de que permiten analizar la fiabilidad de los resultados.

Varias contribuciones se refieren al uso conjunto del enfoque log–lineal y Métodos Exploratorios: el *Análisis de Correspondencias Generalizado de Tablas de Contingencia Múltiple* (Van Der Heijden y Heijerink, 1989), o el *Análisis a tres entradas de los residuos del modelo log–lineal* (Kroonenberg, 1983) son algunos ejemplos. Extensiones del modelo PARAFAC también se han realizado para el tratamiento de datos cualitativos y mixtos.

La cuestión más importante con las tablas de contingencia múltiple es que, a menudo, hay muchos problemas para tratar más de cuatro o cinco variables al mismo tiempo. Además, las variables no deben tener muchas categorías ya que, en caso contrario, habrá muchos ceros o bajas frecuencias en las celdas de la tabla y los métodos no se podrán aplicar correctamente.

A modo de resumen, en la tabla siguiente se expresan los métodos descritos para este tipo de tablas de datos:

Tabla I.2
Métodos para tablas “three-way arrays”

TIPO DE DATOS	TÉCNICA
DATOS CUANTITATIVOS	<i>Modelos de factores comunes</i>
	<i>Modelos de componentes</i> – Método de Tucker – TUCKALS2, TUCKALS3 – PARAFAC/CANDECOMP – <i>Análisis Triadítico y Análisis Triadítico parcial</i> (PCA-SUP o preSTATIS) – SUMPCA
	<i>Modelos para datos temporales o longitudinales</i> – <i>Doble Análisis de Componentes Principales</i> – <i>Análisis Factorial Evolutivo</i> – <i>Análisis de Series Cronológicas Multidimensionales</i> – Método LONGI
DATOS CUALITATIVOS Y MIXTOS	<i>Análisis de Correspondencias Generalizado de Tablas de Contingencia Múltiple</i> <i>Análisis a tres entradas de los residuos del modelo log–lineal</i>

c. Tablas derivadas (matrices de proximidades)

También pueden considerarse como tablas de tres entradas aquellos casos en los que los datos son proximidades (similaridades, disimilaridades o distancias) entre pares de individuos agrupados en K tablas (por ejemplo, las puntuaciones otorgados por K jueces a una serie de individuos respecto a ciertas variables). En este caso, y por oposición al anterior, los tres modos están totalmente cruzados pero uno de ellos (en el ejemplo los individuos) se repite dos veces en cada tabla de datos. La terminología anglosajona denomina estas tablas con el nombre de "three-way, two-mode".

Como indica Bove y Di Ciaccio (1994: 27), este último tipo de tablas puede obtenerse de los dos tipos de tablas de tres entradas anteriores. Dicha tabla, llamada *matriz derivada*, depende del tipo de datos. Si tenemos varios grupos de individuos y uno solo de variables, la matriz derivada es la matriz de productos escalares entre variables (las covarianzas o las correlaciones); si hay más de un grupo de variables y uno solo de individuos, la matriz derivada es la matriz de productos escalares entre individuos. Cuando hay un solo grupo de objetos y uno solo de variables ambas matrices derivadas son posibles.

El análisis de este tipo de tablas es lo que algunos autores denominan *análisis de tablas derivadas* o *ajuste indirecto*, frente al *ajuste directo*, que consideraría la tabla de tres entradas de la que proviene directamente. De todos modos, es importante tener en cuenta que el ajuste directo, siempre y cuando sea posible, produce mejores resultados que el ajuste de tablas derivadas (Kiers, 1991: 466).

Siguiendo un procedimiento similar al de los dos tipos de tablas anteriores, haremos referencia a los métodos para datos cuantitativos y, posteriormente, a los métodos para datos cualitativos y mixtos.

Datos cuantitativos

El *análisis de tablas derivadas* o *ajuste indirecto* puede hacerse por medio de métodos derivados del ajuste directo. Las relaciones entre las estructuras factoriales de este tipo de tablas pueden estudiarse a través de métodos como TUCKALS3, CANDECOMP y PARAFAC, y el *Análisis Factorial Simultáneo* (denominado SUMPCAC, puesto que éste se aplica a la suma de tablas de productos cruzados). Según señala Kiers (1991: 458), el STATIS en su etapa de interestructura (STATIS1) también puede considerarse como un método aplicado a tablas derivadas. No obstante, muchos otros métodos pueden ser aplicados, como los desarrollados por McDonald (1984: 285-307), cuya puesta en práctica se realiza mediante programas como el LISREL (de Joreskog y Sorbom) o el COSAN (de McDonald), que son programas que permiten el análisis de estructura de la covarianza.

Por otra parte, los modelos especialmente propuestos para ese tipo de tablas adquieren especial relevancia cuando no existe una tabla de datos originales a partir de la que se ha obtenido la tabla derivada. Así, un conjunto de matrices simétricas de productos cruzados puede ser analizado por modelos espaciales y no espaciales para obtener la información referente a las diferencias entre las estructuras de similitudes y disimilitudes, así como su estabilidad. Los métodos de *Escalamiento Multidimensional* tratan este problema de las diferencias individuales. Históricamente, el primer enfoque para tratar con varios grupos de matrices de proximidades fue realizado por Tucker y Messick en 1963 (*Análisis de Puntos de Vista*). En éste se consideran dos etapas: en la primera, una matriz de correlaciones entre pares de matrices de datos es obtenida y analizada factorialmente para obtener una representación homogénea de los grupos de ocasiones. A continuación, en una segunda etapa, una media ponderada para cada grupo es obtenida y representada por un *Escalamiento Multidimensional* clásico. Cada una de las configuraciones obtenidas representa un punto de vista diferente. Es importante señalar que estas dos etapas están muy próximas a las fases de la interestructura y del compromiso del enfoque ICI.

Kroonenberg y De Leeuw puntualizan que la aplicación del método de Jaffrenou a la tabla de similaridades produce los mismos resultados que el *Análisis de Puntos de Vista* de Tucker y Messick (Tucker, 1972: 3-27), cuando éste se aplicada a tablas de disimilaridades.

Varios autores sucesivamente sugieren otras soluciones. Por ejemplo, la propuesta de Carroll y Chang, en 1970, representa una importante innovación como punto de partida para el tratamiento de, al menos, dos tablas de disimilaridades. Se trata de los modelos INDSCAL e IDIOSCAL. Estos modelos desarrollados también por Tucker en 1972, tienen como objetivo la comparación de los juicios proporcionados por K jueces sobre los mismos n individuos, es decir, comparar K matrices de similaridades o disimilaridades. Ambos consisten en postular un modelo y posteriormente, a partir de los datos, estimar los parámetros del mismo.

El modelo INDSCAL surge cuando el modelo CANDECOMP/PARAFAC se aplica a tablas derivadas. Además, el modelo INDSCAL constituye un caso particular del IDIOSCAL. Ambos métodos consideran las K matrices de similaridades, disimilaridades o distancias entre n individuos, provistos de los mismos pesos, recogidos en una matriz D . Para poder comparar las diferentes K visiones de los individuos se sustituyen estas matrices de distancias por K matrices W_k ($n \times n$) simétricas, semidefinidas positivas y centradas para D . Su construcción se basa en la existencia de s dimensiones (factores) comunes a los K análisis estudiados, asignándose, en cada análisis, diferentes pesos a cada uno de estos s factores.

El modelo INDSCAL, a partir de esta situación, supone que existe una matriz X ($n \times s$) y K matrices diagonales positivas $D_1, D_2 \dots D_k$ ($s \times s$) tales que, para todo $k = 1 \dots K$, la matriz de distancias es XD_kX_k . Por su parte, el modelo IDIOSCAL supone que existe una matriz X ($n \times s$) y K matrices simétricas definidas positivas, $M_1, M_2 \dots M_k$ ($s \times s$) tales que, para todo $k = 1, 2 \dots K$, la matriz de distancias es XM_kX_k .

Estos métodos, al ser aplicados a matrices de disimilaridades, poseen dualidad con el modelo procrusteano. Sin embargo, en éstos últimos, se supone que los individuos ocupan una posición dada y que cada juez evalúa las distancias entre éstos en función de su métrica, mientras que en el modelo procrusteano, la métrica es común a todos los estudios y los individuos tienen una posición diferente de una nube a otra. El modelo INDSCAL euclídeo ponderado posee la propiedad de unicidad, que también tienen los modelos CANDECOMP y PARAFAC.

Las tablas derivadas también pueden ser analizadas para modelos no espaciales de árboles, cuyo principal objetivo es agrupar los objetos observados. Se parte de la idea de que cada objeto está representado por un conjunto de características y que una función, basada en los factores comunes y específicos, permite cuantificar la similaridad entre pares de objetos. Es una opinión general, y así lo expresan Bove y Di Ciaccio (1994: 30), que, para este tipo de tablas, los análisis basados en métodos espaciales deben ir acompañados de otras técnicas basadas en datos no espaciales.

Datos cualitativos y mixtos

Para datos cualitativos, las tablas derivadas se refieren a un conjunto de tablas de proximidades (entre individuos–objetos, categorías, variables u ocasiones) obtenidas a partir de una tabla de contingencia múltiple o de una tabla de objetos \times variables, empleando un índice apropiado. Ejemplos de tablas derivadas son las de proximidad entre categorías u objetos, y las tablas de residuos con respecto a un modelo.

Dependiendo de la transformación realizada y del índice seleccionado, se obtienen un gran número de métodos. Otro aspecto a evaluar es el papel que se desea asignar a las variables en el análisis: éstas pueden jugar papeles simétricos (por ejemplo, en el enfoque log–lineal) o asimétricos (cuando se

quiere predecir un grupo de variables analizando otro, como en el *Análisis de la Redundancia Cualitativa*).

El enfoque buscado con tablas derivadas puede considerarse más general que otros enfoques ya que, empleando índices apropiados y aplicando técnicas de *Escalamiento Multidimensional* a tres entradas, se pueden realizar los mismos análisis que con los métodos descritos para datos cualitativos. No obstante, y como indica Bove y Di Ciaccio (1994: 28), el empleo de tablas derivadas puede considerarse menos "natural" que el análisis de las estructuras de las que provienen, debido a la transformación realizada y a la dificultad de interpretación de las mismas.

Tabla I.3
Métodos para tablas derivadas

TIPO DE DATOS	TÉCNICA
DATOS CUANTITATIVOS	<i>Métodos derivados del ajuste directo</i> – TUCKALS3 – CANDECOMP Y PARAFAC – <i>Análisis Factorial Simultáneo</i> (SUMPCA) – STATIS1 <i>Métodos de ajuste indirecto</i> – <i>Análisis de los Puntos de Vista</i> – INDESCAL – IDIOSCAL <i>Métodos no espaciales</i>
DATOS CUALITATIVOS Y MIXTOS	<i>Enfoque log-lineal</i> <i>Análisis de la Redundancia Cualitativa</i>

2. Técnicas de predicción

En relación al grupo de técnicas que permiten predecir resultados, es preciso comenzar señalando que cualquier proceso de toma de decisiones implica cierto grado de incertidumbre, dado que el decisor debe valorar ciertos sucesos situados generalmente en el futuro o que desconoce. En algunos casos, su intuición, basada en sus conocimientos previos sobre el entorno y en su experiencia, puede ser válida, si bien, en este caso, existe la posibilidad de que ignore relaciones económicas importantes o que las utilice inadecuadamente. En cualquier caso, se podrán lograr mejores resultados si, en alguna medida, se puede reducir la incertidumbre sobre dichos sucesos empleando alguna técnica de predicción que contraste las relaciones individuales de una u otra manera. Éste es precisamente el objetivo de toda técnica de predicción.

Según Aznar y Trávez (1993a: 8) una técnica de predicción es un proceso que incluye desde la recogida de la información considerada como relevante para predecir una variable, hasta la obtención de la predicción tras procesar dicha información. La predicción ideal debe ser informativa, exacta, barata, comprensible, simple y de fácil incorporación al proceso de toma de decisiones, debiendo estar disponible en el momento preciso y con la frecuencia requerida. Es importante recalcar, como señala Pulido (1989b: 23), que:

las empresas o Administraciones Públicas no suelen tener problemas directos de predicción y menos aún de aplicación de técnicas específicas. Los problemas se tienen en términos

genéricos de economía y gestión de empresas, aunque en su solución pueden intervenir planteamientos y métodos específicos de predicción.

Por todo ello, el conocimiento de las distintas técnicas de predicción es fundamental para elegir la técnica adecuada a la hora de resolver un problema concreto. En la selección de dicha técnica, debe tenerse en cuenta el tipo de variable que se va a predecir, el horizonte de predicción, los recursos y la información disponibles, que debe permitir aproximarse a lo que se denomina predicción ideal. Una clasificación de las distintas técnicas disponibles facilitará dicha tarea, al mismo tiempo que permite comprender el contexto actual de las técnicas de predicción y plantear nuevas posibilidades respecto a alguna de las existentes. Antes de proponer cualquier clasificación, comenzaremos por realizar una breve reseña histórica sobre las mismas.

2.1 Antecedentes históricos

Los antecedentes científicos más antiguos de la predicción pueden encontrarse en el *Análisis de Series Temporales*, a mediados del siglo XIX, respecto a datos demográficos, meteorológicos y astronómicos. Los trabajos de Verhulst en 1845, sobre la variación de la densidad de población en forma de curva logística, y los de Buys-Ballot en 1847, sobre los cambios periódicos de temperatura, son ejemplos significativos. Más tarde, en 1921, Sir William Beveridge elaborará el gráfico de evolución periódica del precio del trigo en Europa, con objeto de explicar la evolución de dicha variable.

El intento de explicar las fluctuaciones observables en las series económicas provoca nuevos desarrollos de gran trascendencia. Los trabajos de Yule, en 1927, y Slutsky, en 1937, plantean los primeros esquemas autorregresivos y de medias móviles, respectivamente, poco después integrados por el sueco Herman Wold.

A partir de la Segunda Guerra Mundial, empieza el desarrollo acelerado de estas técnicas: la descomposición de series en sus componentes (tendenciales, estacionales y cíclicos) por Shiskin, en 1957, considerando los desarrollos de Pearson, en 1919; el filtrado de series de Kalman, en 1960; *Análisis Espectral* por Turkey, en 1961, con los antecedentes teóricos de los resultados obtenidos por Kolgomorov, en 1941 y Wiener, en 1949; el alisado exponencial, con desarrollos ligados a los nombres de Holt, en 1957, de Brown, en 1959, y de Winters, en 1960. Posteriormente, serían desarrollados los *Modelos ARIMA* conjuntamente por Box y Jenkins, en 1970.

Paralelamente al desarrollo del *Análisis de Series Temporales*, se va cimentando el enfoque de los *Modelos Económicos*, es decir, del *Análisis Relacional o Causal* en Economía. Aunque los orígenes de este enfoque se han situado incluso en Le Tableau Économique de Quesnay, en 1758, antes de 1911 pueden encontrarse pocos estudios o desarrollos teóricos sobre esta línea. Los desarrollos de Moore, en 1911, sobre el mercado de trabajo, constituye un trabajo fundamental en el desarrollo de la Econometría. Los estudios de Cobb y Douglas, sobre la función de producción, aparecen en 1928, y en 1941, Joel Dean inicia sus trabajos sobre el estudio de las funciones de costes mediante *Análisis de Regresión*, cuyos orígenes se encuentran en los estudios de Galton.

Los primeros *Modelos Económicos Multiecuacionales* fueron propuestos por Jean Tinbergen en 1937. Además de esta línea de modelización econométrica, existen otras variantes tales como los *Modelos de Dinámica de Sistemas* de Forrester, en 1961, y los *Modelos ARIMA Multivariantes y Multiecuacionales* de Zellner y Palm, en 1974.

En cuanto a la predicción basada en información subjetiva, se puede considerar sus fundamentos remotos en la propia Historia de la Humanidad. De hecho, técnicas tan difundidas como el *Método Delphi*, fueron desarrolladas para aplicaciones militares y difundidas posteriormente. Diversos trabajos de Psicología Aplicada y desarrollos en Teoría de las Organizaciones han

puesto las bases de otras técnicas específicas de obtención subjetiva de información. La propia Teoría Estadística de Muestras en Poblaciones Finitas, con trabajos como el de Neyman, en 1934, señalan el inicio de múltiples encuestas de opinión, intenciones, actitudes o expectativas, que se realizan de manera sistemática a partir de los años cincuenta.

Como en el caso de las Técnicas Multidimensionales, el desarrollo de la informática ha sido fundamental en la aparición de nuevas técnicas, y en la aplicación de las existentes a otros campos de conocimiento, contribuyendo a su ampliación y mejora.

2.2 Clasificación de las técnicas de predicción

La literatura sobre el tema recoge múltiples clasificaciones sobre las técnicas de predicción disponibles, que varían según el punto de vista adoptado y según los autores. Wood y Files, en 1976, diferencian dos grandes categorías: *Métodos Causales* y *No Causales*. La diferencia entre ambos radica en si es, o no, necesario describir los factores que determinan el nivel de una variable a la hora de predecir sus valores futuros.

O'Donovan, en 1983, agrupa los métodos de predicción en dos categorías: *Métodos Cualitativos* y *Métodos Cuantitativos*. Los *Métodos Cualitativos* generalmente utilizan la opinión de expertos para predecir acontecimientos futuros. Los *Métodos Cuantitativos*, llevan a cabo el análisis de datos históricos para predecir los valores futuros de una variable de interés. Dentro de estos *Métodos Cuantitativos* pueden distinguirse dos grupos: *Series Temporales* y *Métodos Causales*.

Makridakis y Wheelright, en 1989, distinguen tres grandes categorías de métodos de predicción: *Métodos Subjetivos*, *Métodos Cuantitativos* y *Métodos Tecnológicos*. Dentro de los *Métodos Subjetivos*, se incluyen técnicas diferentes

que van desde aquellos procesos basados en opiniones individuales hasta otros basados en opiniones de grupo, incluyendo también los estudios de mercado. Dentro de los *Métodos Cuantitativos* se distinguen las *Series Temporales, Causales o Explicativas* y de *Cambio de Tendencia*. Dentro de los *Métodos Tecnológicos*, se distinguen dos grandes subcategorías: los basados en algún tipo de esquema extrapolativo y los basados en un panel de expertos. De forma separada, dentro de esta gran categoría se hace referencia a la *Técnica de la Matriz de Impacto Cruzado*.

Pulido (1989b: 33-41) distingue tres grandes grupos de métodos: *Métodos basados en Información Subjetiva, Métodos basados en el Análisis Aislado de Series* y *Métodos basados en Relaciones entre Variables*. En el primer grupo se incluyen las *Encuestas, el Consenso de Grupo en Interacción Abierta* y el *Panel de Expertos con Retroacción* controlada de la información. En el segundo grupo se incluyen *los Métodos Paramétricos y No Paramétricos Univariantes de Series Temporales*. Por último, en el tercer grupo, se incluyen el *Establecimiento de Analogías, el Análisis de Regularidades Estadísticas, los Modelos de Simulación Deterministas* y los *Modelos Econométricos*.

En líneas generales, la exposición de técnicas que realizan Aznar y Trávez (1993a: 9-15) sigue la misma línea que la clasificación anterior, distinguiendo este autor entre *Métodos Subjetivos, Estudios de Mercado* y *Métodos Objetivos*. Basándonos en la clasificación de Pulido, se recogen a continuación las técnicas de predicción que se aplican habitualmente en el ámbito económico, dedicando especial atención a alguna de ellas.

2.2.1 Basadas en información subjetiva

Este tipo de técnicas toman la opinión que tienen ciertas personas sobre el futuro de la cuestión en estudio. Los datos son suministrados por individuos, aisladamente o como componentes de un grupo, según sus experiencias,

opiniones y expectativas de futuro. Dentro de este grupo, se encuadran las siguientes:

Encuestas de intenciones o expectativas

La predicción se obtiene directamente a partir de las proporciones, medias y totales de una encuesta por muestreo, generalizando los resultados al conjunto de la población.

Encuestas de actitudes o sentimientos

Conocida a priori la conexión entre actitudes y comportamiento, se obtiene, de una encuesta por muestreo, un indicador de "sentimiento", según una escala preestablecida, que se transforma en previsiones en función de la historia del fenómeno en estudio.

Agregación de predicciones individuales

Las opiniones de determinados "expertos", que tienen un conocimiento privilegiado de la variable que se va a predecir, se combinan para obtener una previsión global, con posibles ponderaciones diferentes.

Consenso de grupo en interacción abierta

Considerando una discusión en grupo, se establecen predicciones con cierto grado de aceptación de sus componentes.

Panel de expertos con retroacción controlada de la información

Denominado también *Método Delphi*, se trata de un proceso en el que un grupo de expertos va emitiendo sus opiniones y modificándolas en cada

una de las etapas del proceso, según se van conociendo las opiniones emitidas por el resto de participantes. El proceso comienza pidiendo a cada experto una predicción sobre un determinado fenómeno. Obtenidas las primeras respuestas, se elabora un documento de síntesis en donde se destacan las coincidencias y las discrepancias, enviando a cada experto este documento de síntesis. El proceso sigue y se repite hasta que se logra un acuerdo aceptable por todos los participantes. Cada uno de éstos seguramente modificará sus propuestas predictivas a la vista de la opinión emitida por los restantes.

Otras técnicas exploratorias y normativas

Según este tipo de técnicas, el investigador examina los posibles condicionantes de la evolución de un fenómeno y, cuantificando o no sus probabilidades subjetivas de ocurrencia, establece la predicción. Se puede citar aquí la denominada *Técnica de Escenarios*, que consiste en una secuencia lógica de descripciones de la realidad que se intenta predecir, hasta llegar a una situación posible sobre la forma que tomará dicha realidad en el periodo para el que se hace la predicción. La secuencia de descripciones se efectúa a partir de una serie de supuestos que van modificando las alternativas plausibles según sea la evolución de los acontecimientos.

2.2.2 En función del análisis aislado de series

La característica clave de este enfoque es el estudio de un fenómeno en sí mismo, a través de su evolución temporal (series temporales o históricas), incluyendo en este grupo las siguientes:

Fórmulas "ad hoc" de predicción

Las predicciones se generan mediante un mecanismo automático establecido a priori y de cálculo recursivo.

Análisis de descomposición temporal de series

Las técnicas incluidas dentro de este grupo no parten de la estructura estocástica de la población de la que, hipotéticamente, se han extraído las observaciones. Todos estos métodos están basados en la idea de que, dentro de los valores de una serie observada, pueden distinguirse cuatro componentes: tendencia, variación estacional, variación cíclica y componente irregular, cuya descripción de manera precisa es:

- Tendencia, $T(t)$, también denominada tendencia secular o movimiento secular, se define como una función determinista del tiempo que describe la trayectoria de la serie a largo plazo. Para apreciar esta componente, es preciso que la serie se observe durante un período de tiempo amplio, de tal forma que se podrá comprobar si, a largo plazo, la variable adopta un movimiento que se ajuste a una ley determinada, ya sea de crecimiento, de decrecimiento, de fases sucesivas de crecimiento y decrecimiento, o de estabilidad. Es una componente que reviste gran interés, puesto que refleja la línea maestra de la evolución de un fenómeno económico.
- Variaciones Estacionales, $E(t)$, que se definen como los movimientos repetitivos que se producen sistemáticamente a lo largo de la trayectoria de la serie. Representan las fluctuaciones que se registran, de forma constante, en períodos de tiempo inferiores o iguales a un año, y la repetición de dichos movimientos se debe, generalmente, a las estaciones del año, de ahí su nombre. Su importancia radica en que el tiempo juega en ella un papel que tiene más sentido que en las otras componentes.

- Variación cíclica, $C(t)$, se define como una fluctuación con periodicidad superior a un año, que es debida, generalmente, a las propias fluctuaciones de la actividad económica y social. En Economía, se suelen diferenciar tres tipos de ciclos en cuanto a su periodicidad: los cortos o de Kintchine, que tienen lugar más o menos cada tres años; los medios o de Juglar, cuyo período oscila entre cinco y diez años; y los largos o de Kondratiev, que aparecen cada cuarenta o cincuenta años. El análisis de esta componente ha perdido importancia porque la duración de los ciclos no suele ser constante, razón por la cual, cualquier variación puede dar lugar a una predicción errónea.
- Variación irregular, u_t , componente que refleja las variaciones provocadas por factores esporádicos con repercusiones en la serie temporal. Sus fluctuaciones obedecen a la ocurrencia de sucesos raros, como guerras, huelgas, accidentes, etc.; o bien, a cambios de pequeña magnitud debidos a múltiples causas y no recogidos en otras componentes. La característica esencial de esta componente es que difícilmente se puede formalizar algún esquema de comportamiento para ella.

Considerando que el *Análisis Clásico de Series Temporales* consiste en suponer que la serie temporal se origina como resultado de estas cuatro componentes, nos planteamos, a continuación, el establecimiento de la relación entre ellas para originar la serie, es decir, plantear un modelo para la serie temporal, también denominado esquema de composición. Los considerados habitualmente son:

- Modelo aditivo, en el que se supone que las observaciones se generan como suma de las cuatro componentes, es decir:

$$y_t = T(t) + E(t) + C(t) + u_t \quad \text{siendo } y_t \text{ la serie considerada}$$

- Modelo multiplicativo, cuando las observaciones están generadas por el producto de las componentes según la forma: $y_t = T(t) E(t) C(t) u_t$

La mayoría de las técnicas incluidas dentro de este grupo tratan de predecir los dos componentes relevantes –tendencial y estacional– por separado y, posteriormente, elaborar una predicción conjunta de la variable, constituyendo el enfoque clásico del análisis de series temporales. Autores como Aznar y Trívez (1993a: 256-327) distinguen dos tipos de métodos dentro de este enfoque: de estructura fija y de estructura variable.

Métodos de estructura fija

Son métodos que consideran que la estructura de la tendencia y de la componente estacional es la misma a lo largo de todos los periodos de la muestra. Para una serie que no tiene componente estacional, la estructura tipo del modelo que sirve de punto de referencia para las técnicas de este grupo será:

$$y_t = \beta_0 + \beta_1 t + u_t \quad t = 1, 2 \dots T$$

El modelo se estima con las T observaciones y estas estimaciones son las mismas para todas ellas. En el caso en que la serie tuviera también componente estacional, el modelo tipo sería:

$$y_t = \beta_0 + \beta_1 t + \beta_2 D_{2t} + \dots + \beta_s D_{st} + u_t$$

considerando que hay s estaciones y D_{jt} es una variable ficticia que toma el valor uno, si la observación t -ésima corresponde al periodo estacional j -ésimo; y cero, en otro caso. También, en este caso, los parámetros $\beta_2 \dots \beta_s$ son los mismos para todos los períodos de la muestra.

Métodos de estructura variable

En estos métodos la aportación de la tendencia y de la componente estacional es diferente para los distintos periodos en los que se ha elegido la muestra. Todos ellos están basados en una media ponderada de los valores previos de la muestra, y la diferencia entre ellos radica en si ponderan

todos los valores previos de la misma o sólo unos cuantos, y en si la ponderación asignada a los elementos que se elijan es la misma o si es diferente. Se distinguen, a su vez, dos grandes subcategorías:

Métodos de medias móviles: Sólo ponderan un reducido grupo de observaciones muestrales previas y asignan la misma ponderación a cada una de las observaciones que entran en el promedio.

Métodos de alisado exponencial: Ponderan todas las observaciones y asignan una ponderación diferente según sea su grado de alejamiento desde el período actual. Dentro de este grupo, los métodos concretos varían según que la serie tenga o no tendencia, y/o componente estacional.

Análisis de series temporales: modelos *ARIMA*–univariantes

La obra de los estadísticos George Box y Gwilyn Jenkins (1976) ha concentrado su interés en las técnicas de tratamiento estocástico de las series temporales (denominados por algunos autores, como Aznar y Trívez (1993b), métodos paramétricos de tratamiento de series temporales, por oposición a los anteriormente descritos, que serían no paramétricos). La justificación de este interés radica en que este enfoque ha mostrado ser una herramienta potente para la predicción de series, proporcionando más información que en el caso determinista ya expuesto y, por tanto, mejores predicciones.

Este enfoque se basa en el supuesto de que la serie temporal que se trata de predecir, ha sido generada por un proceso estocástico, es decir, que cada valor de la serie se extrae aleatoriamente de una distribución de probabilidad. El concepto de proceso estocástico desempeña, en el análisis de series temporales, un papel análogo al de población en la Estadística clásica. El proceso, por tanto, se concibe como el conjunto generador de todas las series que podían haber sido observadas y, entre éstas, la serie

concreta que es objeto de estudio, interpretándose ésta como una muestra del proceso.

El enfoque de Box-Jenkins es de tipo iterativo, examinándose, en un primer paso, la serie sin procesar, para identificar uno de los muchos modelos disponibles que se pueden elegir como la mejor representación de la serie. Si la serie bruta no es estacionaria, hay que modificarla para que lo sea, como paso previo a su modelización. El análisis del comportamiento estocástico de la serie permite establecer un modelo estadístico de doble componente: autorregresiva (valores anteriores de la variable, *modelo AR*), y de medias móviles (componente aleatorio, *modelo MA*), con posible incorporación de estacionalidad y eliminación de tendencia. Si ambos componentes están presentes se obtiene un *modelo ARMA*. La expresión abreviada correspondiente es la siguiente: $AR(B) y_t = MA(B) \varepsilon_t$ donde y_t es la serie en cuestión, ε_t es la perturbación aleatoria o ruido blanco, $AR(B)$ es el polinomio en el operador B de la parte autorregresiva del modelo, y $MA(B)$ es el polinomio en el operador B de la parte de medias móviles del modelo.

El siguiente paso en el proceso, una vez identificado el modelo, consiste en la estimación real de los parámetros, cuyo procedimiento es similar al ajuste realizado en una regresión. A continuación, se realiza el diagnóstico para poder determinar si se ha elegido el modelo correcto, y el paso final consiste en predecir realmente, sirviéndose para ello del modelo elegido. Los errores que puedan observarse en la estimación se pueden ir corrigiendo a medida que se va disponiendo de más datos y, en algunos casos, pueden tomarse como una indicación de que el modelo entero debería cambiarse.

Esta metodología de Box-Jenkins es un procedimiento técnicamente sofisticado de predicción de una variable teniendo solamente en cuenta la pauta de la serie cronológica en el pasado. Así pues, se ignora cualquier tipo de información que pudiera incluirse en un modelo estructural de regresión; en su lugar, utiliza la observación más reciente como valor inicial

y analiza, a continuación, errores recientes de previsión, a fin de seleccionar el ajuste más apropiado para futuros periodos de tiempo. Por sus características, este proceso es más apto para la previsión a largo plazo que a corto plazo, aunque, en la práctica, se utiliza indistintamente en ambos casos. No obstante, ofrece muchas ventajas con respecto a otros métodos de series temporales, como la suavización exponencial, la descomposición de series y el análisis de la tendencia simple. Extrae mucha información de la serie, utilizando para ello un número reducido de parámetros, y permitiendo, asimismo, una mayor flexibilidad en la elección del modelo correcto, ya que incluye un proceso que permite examinar una gran variedad de modelos en la búsqueda del más adecuado.

Modelos derivados del *ARIMA*

Los supuestos del modelo de Box-Jenkins no siempre se cumplen, por lo que son necesarios otros modelos, por ejemplo, en aquellos casos en los que los residuos del *modelo ARMA* no tienen varianzas constantes en el tiempo (no hay homoscedasticidad) pero dichas varianzas pueden predecirse. Así han surgido los *modelos ARCH*, que fueron introducidos por Engle en 1982. En este caso el aspecto heterocedástico del modelo se introduce por medio de un componente autorregresivo sobre el cuadrado de los ruidos blancos y, en la práctica, esta formulación puede conducir a un número de retardos bastante grande y a la estimación de un número importante de parámetros. Por analogía con la manera de proceder de Box-Jenkins, Bollerslev en 1986 propone un componente autorregresivo de medias móviles, dando lugar a los *modelos GARCH*. Ambos modelos tienen su aplicación, sobre todo, en el campo de las finanzas y en dos ámbitos especialmente: a la hora de probar teorías económicas relativas a los diversos mercados (obligaciones, divisas, etc.), y a la hora de examinar las relaciones entre el corto y el largo plazo. El segundo caso se refiere a la forma de llevar a cabo ciertos cálculos financieros (determinación de carteras óptimas, de

carteras de cobertura, etc.), si bien, este último tipo de aplicación está, en la mayoría de los casos, sujeta a secreto bancario (Droesbeke, Fichet y Tassi, 1994).

Modelos probabilísticos

Utilizando las distribuciones teóricas de probabilidad, se establecen modelos de predicción sobre la base de los comportamientos reales observados.

Procesos estocásticos

Hasta épocas muy recientes, la previsión de variables económicas se realizaba, casi exclusivamente, mediante Regresión y Series Temporales. Los procesos estocásticos proporcionan ventajas importantes a la hora de perfeccionar la previsión. Además de lo indicado, de una manera intuitiva podemos decir que un proceso estocástico es un fenómeno aleatorio desarrollado en el tiempo y controlado mediante leyes probabilísticas. Todo fenómeno aleatorio está representado por una variable aleatoria que recoge los valores o resultados posibles del fenómeno en un contexto determinado, en un ambiente fijado de antemano, o en un momento fijo cualquiera. Este concepto sirve para estudiar los fenómenos aleatorios en un ambiente dinámico o cambiante, de forma que los valores de la variable aleatoria van a depender de un parámetro, que generalmente es el tiempo, lo que permitirá que los valores de la variable evolucionen.

Matemáticamente, un proceso estocástico se define como un conjunto de variables aleatorias $\{X(t), t \in T\}$ definidas sobre un espacio probabilístico (Ω, A, P) , siendo $t \in T$ un parámetro, y T el espacio paramétrico que origina los diferentes tipos de procesos: los de parámetro discreto y los de parámetro continuo. La caracterización de los procesos estocásticos viene dada por la especificación de la ley de probabilidad conjunta de las variables aleatorias

que la componen, diferenciándose los *Procesos Normales o de Gauss*, que son aquéllos formados por variables aleatorias que se distribuyen conjuntamente según una ley normal, y los *Procesos de Markov*, que tiene gran relevancia, a su vez, para el estudio de otros procesos. La característica de interés que presentan éstos, es la descripción de modelos matemáticos que se ajustan a situaciones en las que el estado presente determina la probabilidad de los estados futuros. Se caracterizan por las distribuciones condicionadas y se les llaman procesos sin memoria porque la función de distribución de una variable sólo depende de la anterior, pudiendo ser también de parámetro discreto o de parámetro continuo.

Los *Procesos de Difusión Unidimensionales* son aquellos procesos de Markov, de parámetro continuo, que verifican las ecuaciones atrasadas y adelantadas de Kolgomorov, siendo uno de los más significativos, en cuanto a las aplicaciones económicas se refiere, el *Proceso Logarítmico–Normal*. En él se introducen variables externas, que de alguna forma modifican los resultados, a las que se denominan factores exógenos y que vienen dadas por vectores cuyas componentes son sus valores en un instante o periodo de tiempo. El estudio de este tipo de procesos como modelos de predicción se completa cuando el proceso es multidimensional (Mures Quintana, 1991), dado que las situaciones reales están formadas por varias variables, especialmente en el ámbito económico. La metodología utilizada para su estudio es una generalización de la correspondiente al caso unidimensional, considerándose básico para el estudio del modelo con factores exógenos que completa el desarrollo de los procesos logarítmico–normales. Obviamente, este último proceso es el que mejor modeliza la situación real y por tanto el que mejores resultados puede proporcionar.

El *Análisis Armónico* se encarga del análisis de los procesos de segundo orden, basados en la generalización de las ecuaciones del Análisis Factorial a los procesos aleatorios, desarrollado por Deville (1978: 19-29). El *Análisis Armónico Cualitativo* es otra extensión hacia series temporales donde se

intenta estudiar la evolución de la variable nominal a lo largo de un periodo de tiempo finito.

Análisis espectral (Análisis frecuencial de procesos estocásticos)

Es bastante frecuente la observación de componentes cíclicas en una serie temporal. Sin embargo, ya hemos indicado que no es fácil detectar la presencia de dichas oscilaciones cíclicas ni detectar el periodo de los ciclos. El *Análisis Espectral* proporciona información sobre la estructura interna de una serie en relación con sus componentes cíclicas y, si se dispone de varias series, se tratará de relacionar las componentes cíclicas de una serie con las de las otras.

Las relaciones temporales de una serie estacionaria se representan a través de la función de autocorrelación, y el teorema de Wiener-Kintchine define la función de distribución espectral a través de la representación espectral de la función de covarianza. Las oscilaciones de una serie histórica de datos pueden descomponerse, según la frecuencia o periodicidad de estos movimientos, definiendo el espectro de la misma.

2.2.3 Según las relaciones entre variables

Para los métodos incluidos en este grupo, la información relevante a tener en cuenta no sólo está constituida por el pasado de la variable que se predice, sino también por los valores presentes y pasados de otras variables. La dependencia entre el comportamiento de diferentes fenómenos posibilita la introducción de predicciones condicionadas con diversos grados de formalización y un análisis de mera relación o de causalidad. Se diferencian las siguientes:

Establecimiento de analogías

La forma más simple de relación es la que se establece entre fenómenos similares en contextos distintos (temporal y espacialmente). En este sentido, la predicción se realiza por pura analogía de situaciones, y por tanto, algunos autores catalogan este método como de información subjetiva.

Análisis de regularidades estadísticas

La relación existente entre fenómenos puede establecerse por un mero análisis estadístico de regularidad en su comportamiento conjunto, incluso ante la inexistencia de una auténtica relación de causalidad, de forma que pueden realizarse predicciones condicionadas a los valores precedentes de ciertos indicadores o barómetros.

Modelos de simulación deterministas

Las relaciones históricamente observadas entre la actuación de un conjunto de factores y el comportamiento de un fenómeno sirven de base para el establecimiento de predicciones sobre el mismo, ante hipótesis de mantenimiento de la relación técnica preexistente.

Modelos econométricos

Un modelo econométrico especifica una o varias relaciones entre un grupo de variables endógenas y otro grupo de variables exógenas, incorporando una estructura estocástica que se concreta en las propiedades asignadas a una o varias perturbaciones aleatorias. Entre las múltiples posibilidades desarrolladas dentro de la econometría, la especificación más utilizada ha sido el *Modelo Lineal General o Regresión Múltiple* (MLG), cuya estructura tipo puede expresarse de la forma siguiente:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + \varepsilon_i$$

Dicha formulación del modelo se plantea como una generalización del *Modelo Lineal Simple* (MLS), considerando que la variable dependiente Y_i es función lineal de un conjunto de variables explicativas independientes $X_1 \dots X_k$ y un término de error ε_i . La representación del modelo se puede realizar matricialmente como $Y = X\beta + \varepsilon$, siendo Y el vector de las observaciones de la variable dependiente, X la matriz de las observaciones de las variables independientes, ε el vector de los errores, y β la matriz de los parámetros. Partiendo de los supuestos básicos del modelo,

- la relación entre Y y las variables $X_1 \dots X_k$ es aleatoria o estocástica y lineal;
- los elementos de X son fijos y su varianza es finita. Además, X es una matriz de rango k , siendo $k < n$ ($n =$ número de observaciones);
- se cumple que $E(\varepsilon) = 0$; $E(\varepsilon\varepsilon') = \sigma^2 I$; ε se distribuye según una ley normal,

la estimación del mismo se realiza por el método de los mínimos cuadrados y consiste en obtener los valores de los parámetros que minimizan la suma de los cuadrados de los errores. La expresión de los estimadores mínimo cuadráticos obtenida es la siguiente:

$$\hat{\beta} = (X'X)^{-1}(X'Y)$$

El contraste de hipótesis y la construcción de intervalos de confianza de los parámetros se realizan teniendo en cuenta la distribución del vector de parámetros $\hat{\beta}$, pudiéndose considerar contrastes basados en la distribución normal, cuando σ^2 sea conocida y en caso contrario, se aproxima σ^2 por S^2 y los contrastes se basan en el estadístico t de Student.

Para comprobar si el ajuste realizado es adecuado, se utiliza el Coeficiente de Determinación (R^2) y la prueba conjunta sobre los coeficientes de regresión. Cuando la media de la variable dependiente es distinta de cero, se modifica la definición de R^2 y se define el Coeficiente

de Determinación Corregido. Por último, se consideran pruebas relativas a la validez conjunta de los parámetros obtenidos, definiendo un estadístico que sigue una F de Snedecor.

El estudio del modelo lineal se completa con el análisis exhaustivo de aquellas situaciones en las que no se verifican algunos de los supuestos. Esto significa la determinación de procedimientos de estimación que mejoren el de mínimos cuadrados. Esta técnica es considerada como la técnica de análisis de datos cuantitativos que permite cuantificar las relaciones que pueden existir entre diversas variables, además de constituir un método idóneo para predecir. Dichos aspectos hacen que esta metodología estadística sea muy utilizada para aplicaciones económicas (Aznar y Trivez, 1993a: 87-225; Pulido, 1989a).

La utilización de variables exógenas en un modelo uniecuacional puede extenderse a la inclusión de variables artificiales dinámicas, es decir, variables temporales que son construidas y empleadas para tener en cuenta situaciones externas que no pueden ser explicadas mediante variables observadas. Surgen así los modelos de *Análisis de Intervención*.

Métodos no paramétricos: Indicadores y Modelo Input-Output

La *Técnica de los Indicadores* se basa en ciertas pautas regularmente observadas que establecen una asociación entre el comportamiento de algunas variables –llamados indicadores– y el comportamiento de la variable que se va a predecir. Normalmente, esa asociación tiene un retardo temporal, de forma que, cuando la variable indicador experimenta un cambio significativo en un periodo, la variable objeto de predicción experimenta también una variación, pero tras haber transcurrido algunos periodos, los indicadores tienen un carácter de señales y nos envían mensajes acerca de cómo será en el futuro el comportamiento de la variable objeto de estudio. Lo que más se destaca en este enfoque de los

indicadores, es ese carácter de anticipación que tienen las señales emanadas de las pautas seguidas por las variables–indicadores.

El *Modelo Input-Output* se basa en una formalización matemática de las relaciones intersectoriales existentes en una economía. Mediante unas hipótesis sobre la forma de la función de producción y el carácter exógeno de las demandas finales en cada uno de los sectores, se obtiene el nivel de producción de cada sector, en función de las demandas finales de todos los sectores de la economía. Conviene destacar que los parámetros del modelo no son estimados, sino calculados con los datos correspondientes a un solo período, razón por la que se considera como método causal y no paramétrico.

Análisis Multivariante de Series Temporales

Algunos autores consideran también este análisis dentro del enfoque causal, dado que se basa en las relaciones que pueden encontrarse entre varias series temporales. Se pueden seguir diversos enfoques: si se parte de un conocimiento apriorístico de la dirección de causalidad existente entre ellas, y enunciando una relación entre la variable output y la variable input (que aparece con desfases temporales), se configura el denominado *Modelo de Función de Transferencia*.

El estudio de los componentes a largo plazo de varias series y, particularmente, de las relaciones entre dichos componentes, se ve mejorado a través del concepto de *Cointegración*. Si consideramos en principio dos series, este concepto está relacionado con la idea de una tendencia común a ambas. Así, si entre dos series existe una relación lineal estable, la expresión del modelo es:

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t$$

donde la perturbación aleatoria ε_t es estacionaria, o sea $\varepsilon_t \in I(0)$, y las series Y_t e X_t estarán cointegradas, aunque sean estacionarias. Si las series son integradas de orden d , entonces $X_t, Y_t \in I(d)$, es decir, son del mismo

orden de integración. Las dos series se denominan cointegradas de orden d y 0, ó sea $CI(d,0)$.

En general dos series $X_t, Y_t \in I(d)$ son cointegradas $CI(d,b)$ si existe una constante α tal que $Y_t - \alpha X_t \in I(d,b)$.

Si se dispone de m series,

$$Z_t = \begin{bmatrix} z_{1t} \\ z_{2t} \\ \vdots \\ z_{mt} \end{bmatrix}$$

para las que $Z_t \in I(d)$, Z_t será cointegrada $CI(d,b)$ si existe un vector $\alpha = (\alpha_1 \dots \alpha_m)'$ tal que $\alpha' Z_t \in I(b)$.

El concepto de cointegración es importante en la elaboración de modelos econométricos dinámicos, ya que el método de mínimos cuadrados no es recomendable si las distintas variables tienen diferente grado de integración. Además, en el caso de que las series estén cointegradas, la predicción a corto plazo puede mejorar, pero en particular afecta a la calidad de la predicción a largo plazo, por lo que se emplea cuando ésta es el objeto del estudio.

Una generalización del *modelo ARMA univariante*, es el *modelo estocástico multivariante MARMA*, en el que se dispone de g series interrelacionadas. En este caso:

$$y_t = (y_{1t}, y_{2t} \dots y_{gt}) \quad \text{para } t = 1 \dots n$$

y se constituye un modelo dinámico: $H(B) = F(B) \varepsilon_t$ en el que $H(B)$ y $F(B)$ son matrices de dimensión $g \times g$ cuyos elementos son polinomios del operador B , y ε_t es un ruido blanco g -dimensional, con esperanza y covarianza nulas, y varianza constante. La formulación de este modelo aprovecha las interrelaciones entre las variables que componen el vector Y_t

multidimensional, para mejorar las predicciones que se obtendrían con un *modelo ARMA* uniecuacional.

Otro modelo multivariante que es útil para predecir varias variables $Y_1, Y_2 \dots Y_g$ que están correlacionadas entre sí es el modelo de tipo *VAR*, en el que las g variables endógenas anteriores se consideran función de unas variables exógenas $X_1, X_2 \dots X_k$, así como de las variables endógenas retardadas:

$$Y_t = A_{t-1} Y_{t-1} + \dots + A_{t-r} Y_{t-r} + \dots + B X_t + E_t$$

Las matrices $A_{t-1} \dots A_{t-r}$, B se estiman por mínimos cuadrados bajo la hipótesis de no existencia de autocorrelación (lo que no es muy restrictivo debido a la estructura de retardos), si bien las perturbaciones de una ecuación pueden estar relacionadas entre sí, aunque no con las variables exógenas ni con las endógenas retardadas.

Autores como Granger (1989: 251) consideran que las generalizaciones del enfoque Box-Jenkins para el *Análisis Multivariante de Series Temporales*, más que una técnica causal, emplea los datos para elegir el modelo apropiado. En este caso de múltiples series temporales, es posible simplificar la situación empleando un procedimiento, tal como el *Análisis de Componentes Principales* o cualquier método factorial, que permita reducir el número de series.

Una vez descritas las técnicas de predicción más significativas, se resumen en el siguiente esquema:

Tabla I.4
Técnicas de predicción

	TÉCNICAS
INFORMACIÓN SUBJETIVA	<ul style="list-style-type: none"> – Encuestas de intenciones o expectativas – Encuestas de actitudes o sentimientos – Agregación de predicciones individuales – Consenso de grupo en interacción abierta – Panel de expertos con retroalimentación controlada de la información – Otras técnicas exploratorias y normativas
ANÁLISIS AISLADO DE SERIES	<ul style="list-style-type: none"> – Fórmulas “ad hoc” de predicción – Análisis de descomposición temporal de series – Análisis de series temporales: Modelos ARIMA–Univariantes – Modelos derivados del ARIMA – Modelos probabilísticos – Procesos estocásticos – Análisis espectral
RELACIONES ENTRE VARIABLES	<ul style="list-style-type: none"> – Establecimiento de analogías – Análisis de regularidades estadísticas – Modelos de simulación deterministas – Modelos econométricos – Métodos no paramétricos – Análisis Multivariante de series temporales

3. La Metodología STATIS como Técnica Descriptiva y de Predicción

Para finalizar este capítulo, y como precedente para el siguiente, en este apartado queremos dejar patente el contexto en el que se incluye la metodología STATIS. Antes de comenzar con su desarrollo, y considerando la descripción realizada en los dos apartados anteriores de este capítulo, procederemos a posicionar dicha metodología, destacando su carácter descriptivo y predictivo.

Anteriormente ya se ha especificado que la metodología a la que nos estamos refiriendo, se incluía dentro de los métodos descriptivos que analizaban variables cuantitativas y cuya estrategia global requería análisis separados y conjuntos de los tres modos. Además de ser idónea para el análisis de matrices formadas por los mismos individuos y distintas variables (*Statis*), también permite el tratamiento de tablas en las que las variables son comunes y los individuos diferentes (*Statis dua*).

Además, el tratamiento de tablas de tres entradas totalmente cruzadas mediante este método, proporciona una nueva característica a la infraestructura, pues permite dibujar la trayectoria de las variables (en el *Statis*) y de los individuos (en el *Statis dua*) cuando las distintas tablas están ordenadas y representan periodos de tiempo distintos, como ya hemos especificado anteriormente.

Todo ello constituye la justificación de este apartado en el que también se van a establecer comparaciones con métodos que permiten el tratamiento de tablas en el tiempo, y con métodos de objetivos y procedimientos similares. Siguiendo a Kiers (1991: 450), entre dichos métodos se distinguen, por un lado, los que tratan de una manera simétrica los tres modos, y por otro, los que dan prioridad a uno de ellos, sobre cuya base se obtienen las representaciones de los otros dos modos. Los métodos con un enfoque simétrico permiten obtener representaciones mejores de todos los modos, mientras que los asimétricos permiten una mejor representación del modo prioritario. Entre los métodos simétricos encontramos el CANDECOMP/PARAFAC y el TUCKALS3, y entre los asimétricos se encuentra la metodología STATIS y el AFM.

Una circunstancia en la que el enfoque asimétrico tiene especial relevancia, es cuando uno de los modos es el tiempo. Respecto a los otros dos modos (individuos y variables) dado que AFM y STATIS pertenecen al enfoque ICI, ambos métodos conceden la misma importancia a individuos y variables. Sin embargo, mientras que el STATIS compara los datos entre los distintos momentos del tiempo (interestructura) y, posteriormente, a partir de dicha comparación, realiza un análisis detallado de las relaciones entre variables e individuos (intraestructura); el AFM pone especial énfasis en las relaciones de la intraestructura, condicionando a ésta el análisis de la interestructura.

Por este motivo, la metodología STATIS se va a comparar especialmente con el AFM y con aquellos métodos que ponen más énfasis en el modo temporal (*Doble Análisis de Componentes Principales, Análisis Evolutivo y método LONGI*). Asimismo, se van a poner de manifiesto su relación con ciertos métodos simétricos, cuyos desarrollos teóricos son similares.

3.1 Comparación con métodos simétricos

Tal y como recoge Kiers (1991: 450), el SUMPCA para productos cruzados es una versión restringida del método INDSCAL, y éste, a su vez, del que allí se denomina STATIS1, que es la etapa de la interestructura de STATIS. Esto nos permite asegurar que el análisis de la interestructura que realiza el STATIS permite un mejor ajuste de la matriz de productos escalares entre los objetos representativos de cada tabla que los otros métodos.

En cuanto a la etapa de la intraestructura, STATIS debe ser comparado con métodos que consideren el *ajuste directo*, pues el objeto de esa etapa es analizar las tablas de datos originales. Lo mismo ocurre con el AFM, por lo que Kiers (1991: 462) considera ambos métodos como una versión ponderada del SUMPCA. Dado que partimos de las tablas de datos originales, el SUMPCA es, en realidad, un *Análisis Factorial Simultáneo*, que coincide con una de las fases del método de Jaffrenou (PCA-SUP), y que es el modelo más amplio de los considerados, ya que tal y como se ha expuesto, los modelos CANDECOMP/PARAFAC son una versión restringida del TUCKALS3, y éste, a su vez, de dicho PCA-SUP.

La diferencia entre el STATIS y el *Análisis de Componentes Simultáneo* es la misma que la señalada entre este último y el *Análisis Factorial Simultáneo*. Así, el *Análisis de Componentes Simultáneo* busca componentes para las matrices de datos que expliquen de manera óptima cada matriz, mientras que el STATIS trataría de explicar de manera óptima las variables (Kiers, 1988: 28).

3.2 Comparación con el Análisis Factorial Múltiple

La primera cuestión a tener en cuenta es que, mientras la metodología STATIS permite tratar solo variables cuantitativas, el AFM puede tratar datos mixtos, es decir, variables cualitativas y cuantitativas simultáneamente. Existen, además, otra serie de diferencias que se recogen en el artículo de Pagès (1996: 81-95), y que se pueden concretar en las siguientes:

Estudio de la interestructura

En el AFM, los elementos representativos de cada tabla de datos se representan en las direcciones obtenidas por las componentes principales del análisis de la intraestructura, siendo dicha representación óptima en el sentido de que los vectores ortogonales obtenidos, son los que más se relacionan con el conjunto de variables, y como tal, pueden ser interpretados. Sin embargo, según STATIS, los elementos representativos de cada tabla están proyectados sobre sus ejes principales de inercia máxima.

Comparando las soluciones de ambos métodos, el criterio de la inercia máxima empleado por STATIS se impone por razones técnicas. Aunque en el AFM cada eje puede interpretarse en función de los grupos de variables con los que está relacionado, la calidad global de la representación no es correcta en la mayoría de las ocasiones, aunque se conserven todos los ejes. Esto se debe a que dicha representación puede considerarse como un caso particular del modelo INDSCAL y, por tanto, sólo si los datos siguen exactamente ese modelo, se obtiene una representación perfecta. En el caso de STATIS, la calidad global de la representación es óptima, y el primer eje, combinación lineal de los elementos representativos de cada tabla, es interpretable en tanto que constituye el compromiso. Como inconveniente, cabe señalar que los ejes siguientes no son interpretables ya que no se pueden relacionar con las variables iniciales.

Búsqueda de un compromiso

Según la metodología a la que nos estamos refiriendo, el compromiso está formado por la suma ponderada de los elementos característicos de cada tabla de datos (con norma unitaria o no, dependiendo del caso), empleando como ponderación los elementos del primer vector propio del análisis de la interestructura. Así, cada variable resulta ponderada por la raíz cuadrada del elemento correspondiente del primer vector propio. Por su parte, en el AFM, el compromiso puede considerarse formado por la matriz de productos escalares entre individuos. Cada individuo estará caracterizado por las variables de las distintas tablas, ponderadas por el inverso del mayor valor propio de la matriz de productos escalares de la tabla a la que pertenecen. Geométricamente, esto equivale a hacer igual a uno la inercia axial máxima de las nubes de individuos y de variables para cada grupo o tabla.

La ponderación de STATIS es específica de este método y permite otorgar a priori más importancia a los grupos cuanto más grande sea su relación con el compromiso. Dado que todas las variables pertenecientes a un mismo grupo tienen una misma ponderación, todas las variables verán ampliado su peso original, aunque no participen en dicha relación. Asimismo, para los casos en los que los elementos representativos de cada tabla tienen norma unitaria, cada variable estará dividida por la raíz cuadrada de la norma, es decir, la raíz cuadrada de la suma de los valores propios de su matriz de productos escalares entre individuos. El objeto de esta ponderación es equilibrar la influencia de los grupos en el análisis global, que es una situación intermedia entre la del AFM y la que sería necesaria para igualar la inercia total de todos los grupos. Como consecuencia, con respecto al AFM, los primeros ejes obtenidos con STATIS están a priori más influenciados por los grupos unidimensionales.

Estudio de la infraestructura

Las representaciones obtenidas con ambos procedimientos, en lo que concierne a las posiciones globales y a las direcciones principales de inercia, conducen a conclusiones similares. Las mayores diferencias se producen respecto a las inercias, debido a las ponderaciones de STATIS. En cuanto a las direcciones de inercia menor, como resultado de tal ponderación, se refuerzan direcciones de inercia poco importantes de los grupos que están muy relacionados con el compromiso. En cambio, el AFM permite extraer de una manera más exacta las dimensiones de cada uno de los grupos.

Trayectorias de los individuos

En el AFM la imagen de cada individuo-compromiso o medio, es el isobaricentro de las imágenes del individuo de cada tabla. En el STATIS, la representación superpuesta se obtiene colocando, como individuos suplementarios en los ejes del compromiso, las tablas de datos iniciales. Debido al sistema de ponderaciones de este método, la representación del individuo correspondiente a cada tabla (la unión de los puntos que representan a cada individuo da lugar a las trayectorias) está fuertemente influenciada por dicha ponderación. Por ese motivo, la representación acentúa la proximidad del individuo-medio con las representaciones del individuo pertenecientes a los grupos más próximos a la estructura común, cualquiera que sea el eje. Dado que la suma de las ponderaciones no es uno, en una situación en la que las representaciones de un individuo para las distintas tablas fuesen coincidentes, el individuo-compromiso ocuparía una posición diferente. Por este motivo, la representación del AFM es más adecuada.

Resultados

Como consecuencia de todo lo indicado, el AFM permite extraer, una por una, las dimensiones comunes al conjunto de grupos, es decir, las más

relacionadas con el conjunto de los grupos, mientras que la metodología STATIS permite extraer la estructura común multidimensional y analizar esa estructura común, además de minimizar la importancia de las estructuras marginales que no participan en la estructura común. Las principales dimensiones comunes son puestas de relieve por ambos métodos, aunque la diferencia entre ambos radica en las estructuras marginales.

Por tanto, se empleará la metodología STATIS si el objeto del estudio es obtener una visión global de las semejanzas entre los grupos, mientras que se utilizará el AFM si lo que se busca es una visión analítica y parcial de las semejanzas entre grupos. En la práctica, si los datos poseen cierta estructura común, aunque sea débil, ambos métodos proporcionan resultados semejantes. No obstante, si el modo común a las tablas son las variables, la aplicación del AFM sólo es posible invirtiendo las tablas, mientras que la metodología STATIS se aplica directamente a través de la versión *Statis Dual*.

3.3 Comparación con métodos para datos temporales

El primer método con el que se hace la comparación es el *Doble Análisis en Componentes Principales*, que se aplica especialmente para datos cronológicos, es decir, grupos de matrices medidas para el mismo conjunto de individuos y de variables (cuantitativas) en diferentes momentos de tiempo, siendo el tiempo, un objetivo importante del estudio. Este método propone un análisis muy completo de los datos, ya que, por una parte, analiza de forma global todos los centros de gravedad; y por otra parte, analiza cada tabla centrada sobre su centro de gravedad, comparando, posteriormente, los diferentes análisis. Sin embargo, presenta el inconveniente de que en algunas etapas del mismo hay problemas para encontrar un óptimo global, y para relacionar los análisis correspondientes al espacio de los individuos y al de las variables.

En lo que se refiere al *Análisis de Series Cronológicas Multidimensionales*, si se compara con el *Statis dual*, ambos métodos proporcionan resultados similares. La ventaja que presenta el ASCM es que, debido a la estimación de la tendencia, es posible prever el centro de gravedad de una nueva tabla y hacer predicciones sobre los datos. Para conseguir dicho objetivo a través del *Statis dual*, será necesario apoyarse en otra técnica que, evidentemente, debe ser de predicción. La ventaja de ésta opción es que las tablas de datos no han de tener el mismo número de individuos ni el mismo número de variables en todas las tablas.

En cuanto al *Análisis Factorial Evolutivo*, es preciso señalar que tiene también su aplicación a datos temporales o cronológicos, pero posee la restricción frente al STATIS de que las variables deben ser las mismas en todas las tablas. El AFE introduce una ponderación para las diferentes tablas, pero no para equilibrar la influencia que éstas puedan ejercer en el análisis global, sino para dar, por ejemplo, un peso débil a los periodos alejados y un peso fuerte a los periodos actuales.

Respecto a las diferencias con el método LONGI, son prácticamente las mismas que entre el método LONGI y el *Análisis de Correlaciones Generalizado*. La diferencia entre éstos es que, en el primero, los grupos tienen que contener todos las mismas variables y, a partir de éstas, se obtienen las "variables canónicas", empleando los mismos pesos en cada tabla. Por esta razón, puede considerarse una versión del ACG con los pesos fijos.

Concluimos, por tanto, esquematizando las ventajas, frente a otros métodos, de la metodología STATIS, que desarrollaremos a continuación, y que, a nuestro juicio, puede considerarse como la técnica multivariante más apropiada para el análisis de datos en las siguientes circunstancias:

- El objeto global del estudio de la tabla de tres entradas debe ser el análisis de la estructura común de las tablas de datos de doble

entrada que la componen, a fin de obtener una visión global de las semejanzas entre los grupos, relacionando, además, individuos y variables (todas cuantitativas).

- Las tablas que originan la tabla de tres entradas analizada, representan momentos del tiempo diferentes (inicialmente fue desarrollado para este tipo de tablas, aunque también sea posible su aplicación en otros casos), con independencia del tipo de tablas que se traten:
 - * Tablas formadas por el mismo conjunto de individuos, pero distintos grupos de variables. En este caso, el análisis se centra en la evolución, si los periodos están ordenados, que sufren los individuos en las diferentes situaciones analizadas estudiando, asimismo, las relaciones entre los grupos considerados.
 - * Tablas totalmente cruzadas, que permitirán analizar también la evolución de las variables, si los periodos de tiempo están ordenados.
 - * Tablas formadas por distintos individuos sobre los que se han observado las mismas variables (*Statis dual*). En este caso, el objeto es analizar la evolución de las variables considerando, asimismo, las relaciones entre grupos.
- Su aplicación es adecuada, desde el punto de vista metodológico, puesto que proporciona mejores soluciones que otros métodos, al ser una versión más amplia de algunos de ellos, tanto en la etapa de la interestructura como de la intraestructura. La construcción del compromiso está menos influenciada que en otros métodos, como el ACG, por la organización de las variables en grupos, y otorga a individuos y variables un papel simétrico.
- Cuando se pretende equilibrar la influencia de cada uno de los grupos de variables sobre el análisis global de todas las matrices de datos, su aplicación es una solución adecuada.

Por último, STATIS es la única metodología de las expuestas que proporciona un análisis satisfactorio en las situaciones mencionadas. Si el objetivo es la previsión, combinado con alguna técnica de predicción apropiada, permite realizar predicciones para las distintas variables de una forma más sencilla que si éstas se analizasen de manera individual. La aplicación de las dos versiones de la técnica, cuando es posible, permite obtener resultados complementarios y conclusiones más precisas sobre los datos analizados. Además, la etapa de la intraestructura es, en realidad, un ACP ponderado y, por tanto, factorial, que puede complementarse con otras técnicas descriptivas, como el *Análisis Cluster*.

Capítulo II
La metodología STATIS:
Statis y Statis dual

Para justificar el desarrollo de este capítulo nos basamos en la conveniencia de la aplicación de la metodología STATIS como técnica multidimensional, cuando se trata de analizar de la estructura común de las tablas de datos que componen una tabla de tres entradas, con independencia de como sean dichas tablas y sin que unas tengan prioridad sobre otras. Además, al tratarse de una técnica factorial, es posible también el empleo de la misma para la reducción del número de variables en la aplicación de algún modelo de predicción y como alternativa a la predicción individual de cada variable.

Por estas razones, desarrollaremos aquí, desde el punto de vista teórico, la metodología STATIS, denominación empleada para referirnos a sus dos versiones¹: *método Statis* y *método Statis dual*. El *Statis* se utiliza para analizar tablas de tres entradas en las que los individuos observados son los mismos en todas las tablas (ocasiones o momentos del tiempo distintos) pero las variables son diferentes. Por el contrario, el *Statis dual* se aplicará cuando la tabla de triple entrada esté formada por las mismas variables, observadas sobre diferentes individuos en cada tabla. Para el caso en que los individuos y las variables sean los mismos, ambos procedimientos

¹ En algunos trabajos se emplea la denominación *Analyse Conjointe de Tableaux quantitatifs* (ACT) para referirse a las dos versiones del método. Podemos citar, entre otras, la obra de C. Lavit (1988), *Analyse conjointe de tableaux quantitatifs*, Paris: Ed. Masson; y C. Lavit, Y. Escoufier, R. Sabatier, P. Traissac (1994), "The ACT (STATIS method)", *Computational Statistics and Data Analysis*, 18, pp. 97-119.

pueden aplicarse y la elección de uno u otro dependerá de los objetivos perseguidos (GERI, 1996: 68).

El objetivo que pretendemos es sentar las bases teóricas con el fin de su posterior aplicación. Puesto que la base metodológica de ambas versiones es el *Análisis de Componentes Principales* (ACP), en el primer apartado se desarrolla en profundidad dicho método, para describir, a continuación, el *Statis* y *Statis dual*. Finalizaremos el capítulo poniendo de manifiesto las posibilidades descriptivas y predictivas de dichas técnicas, lo que va a permitir considerar esta metodología como una de las más completas para el análisis estadístico.

1. Análisis de Componentes Principales (ACP)

Con el fin de situar este método dentro de las técnicas multivariantes, comenzamos este apartado con una breve referencia histórica, lo que va a permitir aclarar ciertas cuestiones respecto a su empleo generalizado. A continuación se considerarán los objetivos más concretos del mismo y las características de los datos que pueden ser analizados con él. Se completa con el desarrollo teórico del método propiamente dicho: el ACP *normado*, cuando las variables están centradas y reducidas, que es el más utilizado y el ACP *no-normado (centrado)*, cuando las variables están centradas pero no reducidas.

1.1 Antecedentes históricos

Los orígenes del *Análisis de Componentes Principales (ACP)* se remontan a los trabajos de Karl Pearson en 1901 y a los desarrollos de Harold Hotelling en 1933, siendo en el trabajo de este último autor donde aparece por primera vez el método y la denominación actual. A pesar de ello, su empleo generalizado no se produce hasta épocas recientes, como consecuencia del desarrollo de la informática y de la simplificación de los cálculos matemáticos que ésta trae consigo.

Desde el punto de vista clásico, el ACP tal y como fue presentado por Hotelling, tiene por objeto la búsqueda de los ejes principales, que son los ejes del elipsoide descrito por una muestra de una distribución normal multidimensional. Este punto de vista es el que inspira también los trabajos de autores como Anderson, en 1958.

Como técnica factorial, se puede considerar como un caso particular del método de *Análisis Factorial* (AF), empleado en Psicometría por autores como Horst (1965) y Harman (1967). Aunque una de las posibles opciones para la extracción de factores en el *Análisis Factorial* es la de componentes principales (por ello su consideración como caso particular de dicho método), el ACP es una técnica que se emplea de manera independiente y que no presupone la existencia de un modelo subyacente a los datos como se supone en el AF. En Chatfield y Collins (1980: 82) se señala que en el AF los factores son estimados, mientras que en ACP sólo son transformaciones de las variables originales en un nuevo conjunto de combinaciones lineales; por tanto, a pesar de la similitud de métodos, la denominación empleada en algunos casos de AF para referirse al ACP no es adecuada.

Por otra parte, ambos métodos tienen en común ciertos aspectos polémicos, como la estimación de las comunalidades, el número de factores a extraer y los métodos de rotación de factores. Dado que no entra en nuestros objetivos detenernos en un análisis a fondo de estas problemáticas nos remitimos a las obras de Harman (1980) y Stevens (1986: 337-372) para abordar más detenidamente dichas cuestiones.

El planteamiento reciente se enfoca a considerar el ACP como una técnica de representación de datos óptima, debido a las propiedades geométricas y algebraicas que cumple, que se utiliza sin necesidad de hipótesis generales de naturaleza estadística, ni de un modelo particular. Este es el punto de vista adoptado en el trabajo de Pearson de 1901; la obra de Rao, en 1964, también realiza una presentación próxima a éste enfoque y así es considerado también por Lebart *et al.* (1995: 32). De

todos modos, y como precisa Chatfield (1980: 58), aunque no se requiere modelo estadístico subyacente para explicar el error de estructura, si los datos provienen de un modelo normal multivariante las componentes pueden dotarse de mayor significado. Este último planteamiento se corresponde con el que se va a seguir en el presente estudio, que es el que subyace en muchas de las aplicaciones del método en todos los ámbitos del conocimiento.

Su empleo generalizado ha llevado en muchos casos a conclusiones más allá de las posibilidades del mismo. Conviene precisar que, tal y como manifiesta Chatfield (1980: 57), el ACP no es siempre una herramienta tan eficaz como se pretende que sea. En algunos casos la identificación de componentes es extremadamente arbitraria y a veces resulta peligroso intentar atribuir un significado concreto a las componentes, pudiendo llevar a conclusiones erróneas. Sin embargo, sí proporciona un método adecuado para reducir la dimensionalidad de una serie de datos y para agrupar variables si los primeros componentes recogen la mayor parte de la variación de los datos originales. De aquí, el empleo de los resultados del ACP en análisis subsiguientes como el *Cluster* o *Discriminante* (Chatfield, 1980: 75), ó como en *Regresión Múltiple* (Stevens, 1986: 345), en el que el ACP se utiliza para reducir el número de predictores.

El Análisis de Componentes Principales presenta numerosas variantes según las transformaciones iniciales de la tabla de datos. La nube de individuos puede estar centrada o no, reducida o no. De todos ellos, el ACP *normado* (variables centradas y reducidas) es el más empleado y por tanto el que se desarrollará de manera amplia. Sin embargo, no sólo éste sino también el ACP *no-normado* (variables no reducidas pero centradas) son los métodos que sirven de referencia teórica para el *Statis* y el *Statis dual*, por lo que les dedicaremos especial atención.

1.2 Datos y objetivos

El *Análisis de Componentes Principales* (ACP) tiene por objeto la descripción de una tabla de individuos por variables, es decir, de doble entrada, mediante representaciones gráficas que contengan el máximo de información posible de la tabla de datos original. Dicha tabla (X), de dimensiones $n \times p$, representa en filas los n individuos (u objetos) y en columnas las p variables (o características), tal y como se expresa a continuación:

Figura II.1
Tabla de datos (X)

	1	2	...	j	...	p
1	x_{ij}					
2						
⋮						
i						
⋮						
n						

El elemento característico de dicha tabla es x_{ij} ($i = 1, 2 \dots n; j = 1, 2 \dots p$).

Cada una de las n filas de dicha tabla representa un individuo e_i , por tanto, cada individuo e_i está representado por p valores que constituyen un vector del espacio $E = R^p$, $e_i' = (x_{i1}, x_{i2} \dots x_{ip})$. El conjunto de e_i ($i = 1, 2 \dots n$) origina la nube de puntos-individuos N_I .

De la misma forma, cada una de las p columnas de dicha tabla, representa una variable, lo que significa que cada variable X_j vendrá determinada por n valores que constituyen un vector del espacio $F = R^n$:

$$X_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{bmatrix}$$

y para cada $j = 1, 2 \dots p$, el conjunto de X_j origina la nube de puntos-variables N_j .

Las medidas a tener en cuenta, en relación a la tabla de datos son:

- La media de X_j :

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

- La varianza:

$$s_j^2 = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$$

- La covarianza de dos variables X_j e $X_{j'}$:

$$Cov(X_j, X_{j'}) = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ij'} - \bar{x}_{j'})$$

- La matriz de varianzas-covarianzas de las variables X_j , que denominaremos V (de orden $p \times p$), que está formada por la covarianza de las diferentes variables.
- El coeficiente de correlación entre X_j y $X_{j'}$:

$$\rho_{jj'} = \frac{Cov(X_j, X_{j'})}{s_j s_{j'}}$$

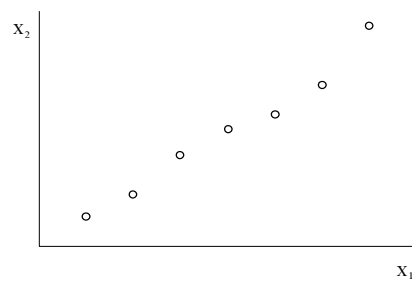
- La matriz de correlaciones, que denominaremos R (de orden $p \times p$), que viene dada por los coeficientes de correlación.

Los objetivos concretos de esta técnica se pueden resumir en los siguientes:

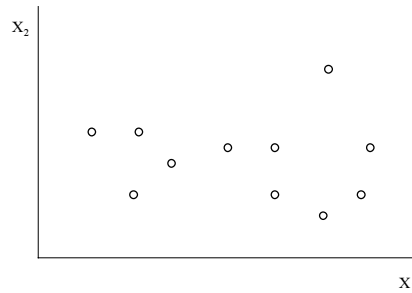
- Facilita el estudio de las relaciones existentes entre las variables X_j .
- Permite realizar el análisis de la dispersión de los individuos e_i poniendo en evidencia los individuos con características semejantes, detectando las variables que son responsables de dicha dispersión.

La posición de los puntos-individuos en la nube, viene dada por las distancias entre ellos y determina la forma de la misma. Esta forma es la que caracteriza la naturaleza e intensidad de las relaciones entre los individuos (filas) y las variables (columnas). En el caso de que solo fueran dos las variables consideradas, podríamos representar en el plano cada individuo e_i por un punto de dos coordenadas (x_{i1}, x_{i2}) . La relación entre esas dos variables se podría estudiar mediante las siguientes representaciones gráficas, al mismo tiempo que podríamos agrupar individuos con características semejantes:

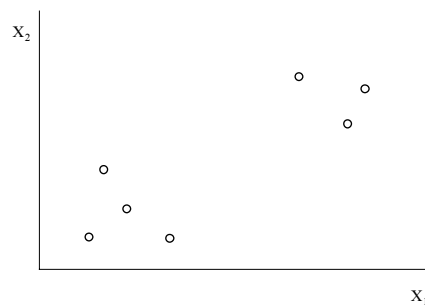
Figura II.2
Representación de los individuos en R^2



Fuerte relación entre las variables



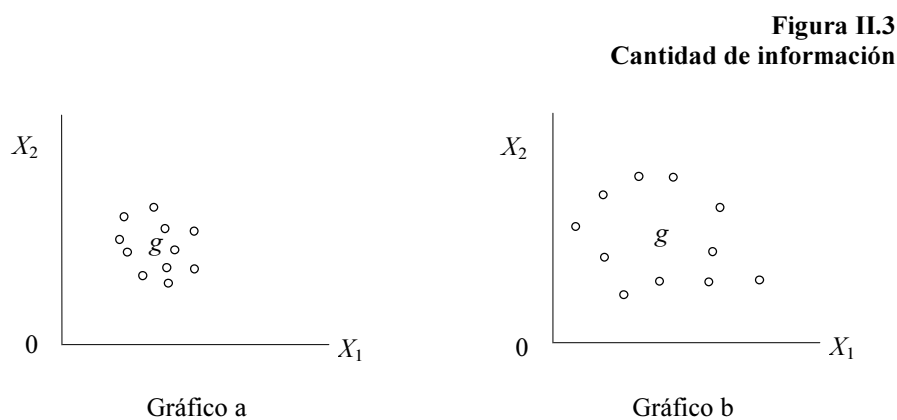
Ausencia de relación entre variables



Grupos homogéneos de individuos
(respecto a las dos variables)

Cuando las variables son tres la representación en el espacio de los individuos aún permitiría el estudio visual de los datos. Si bien, y este es el caso más frecuente, cuando el número de características es superior a tres, es difícil tener una representación gráfica del conjunto de individuos, por lo cual es necesario buscar una imagen de los mismos en un espacio de menor dimensión (dos o tres a lo sumo) que refleje, de forma adecuada, las distancias originales entre los individuos y que mantenga una cantidad importante de información de la tabla de datos original.

Existen diversas formas de medir la cantidad de información de una tabla de datos, aunque antes de tomar una de ellas conviene aclarar algunas cuestiones. En una tabla formada por dos variables y n individuos, podemos obtener diferentes representaciones gráficas según la cantidad de información que esas tablas presenten (Abascal y Grande, 1989: 30):



En el gráfico II.3a no existen muchas diferencias entre unos individuos y otros y ninguno de ellos aporta gran información al colectivo, por lo que el centro de gravedad g representa bastante bien a todos ellos; se dice en este caso que el gráfico tiene poca información. En el gráfico II.3b los individuos están más dispersos. Todos son distintos y las distancias entre ellos, entre éstos y el origen, y entre ellos y el centro de gravedad son elevadas, por tanto, este gráfico contiene mucha información.

Se puede deducir de todo ello, que una medida de la *cantidad de información* de una tabla de datos $n \times p$, viene dada por la suma de los cuadrados de las distancias de los individuos al origen (denominado punto 0), cada uno de ellos ponderado por su peso p_i , que se representa:

$$CI = \sum_{i=1}^n p_i d^2(e_i, 0)$$

Normalmente, mediante la transformación que luego indicaremos, el origen (0) de la nube coincide con su centro de gravedad (g), y así la cantidad de información coincide con lo que se denomina *ineria*, que es la medida de dispersión de los puntos en torno a su centro:

$$I = \sum_{i=1}^n p_i d^2(e_i, g)$$

Como consecuencia, la nueva imagen obtenida debe conservar, de la mejor manera posible, la *ineria* (si se mide respecto al centro de gravedad) o la *cantidad de información* (si se mide respecto al origen) de la nube de puntos original. Geométricamente, este proceso busca un sistema de ejes ortogonales que maximice la media de los cuadrados de las distancias entre las proyecciones de los individuos, a fin de conservar lo mejor posible las distancias iniciales. Es necesario aclarar que esta reducción sólo es posible si los p caracteres iniciales no son independientes y los coeficientes de correlación entre variables no son nulos. Por tanto, se debe analizar la matriz de correlaciones de las variables iniciales (a través del Test de esfericidad de Barlett²) y la inercia de los nuevos ejes. Tal y como se indicará en el epígrafe de interpretación, si los valores propios son todos similares, la nube es esférica y las componentes no explican mucha

² El test de esfericidad de Bartlett considera como hipótesis nula que las variables en la población están incorreladas, es decir, la matriz de correlaciones es la matriz identidad. Este test debe realizarse si el número de individuos es inferior a 100 y los elementos de la matriz de correlación son pequeños, tal y como se señala en J. Stevens (1986: 339).

más información que las variables originales. En ese caso un ACP no permite obtener conclusiones relevantes.

Puesto que la dimensión del nuevo espacio es inferior a la dimensión de la nube de puntos original, la distancia entre la proyección del individuo e_i y la proyección del individuo $e_{i'}$ en el nuevo espacio, es inferior a la distancia original entre e_i y $e_{i'}$. Analíticamente se puede expresar como $d^2(f_{si}, f_{si'}) < d^2(e_i, e_{i'})$, siendo f_{si} y $f_{si'}$ las proyecciones respectivas de e_i y $e_{i'}$ en el nuevo espacio respectivamente.

1.3 ACP *normado*

Este constituye el caso más general de ACP. Con el fin de comprender mejor dicho método, es conveniente explicar previamente las propiedades de los elementos que forman la tabla de datos, según esta se tome en filas (nube de individuos) o en columnas (nube de variables). El desarrollo del método propiamente dicho, es decir el análisis de las nubes indicadas, se expondrá a continuación junto con una serie de precisiones para la correcta interpretación de los datos.

1.3.1 Descripción de las nubes (individuos y variables)

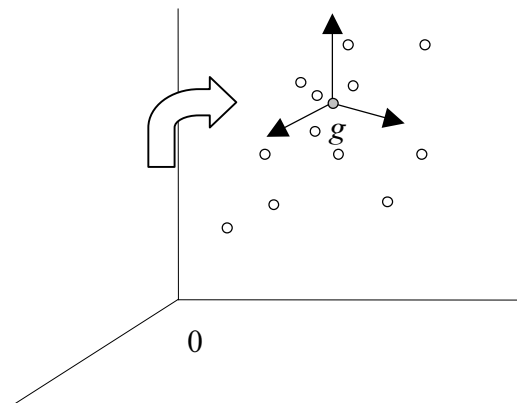
El proceso que permite describir la tabla de datos, utilizando las transformaciones que originan las nubes de los individuos y de las variables, incluye la definición de una métrica, el producto escalar, la distancia (adecuada a la naturaleza y al espacio en el que se presentan las nubes) y, como consecuencia, la cantidad de información de la nube correspondiente.

a. La nube de individuos

El conjunto de individuos e_i , ya descrito, origina la nube de puntos N_I , siendo el centro de gravedad de esa nube $g = (\bar{x}_1, \bar{x}_2 \dots \bar{x}_p)$, definido en el espacio $E = R^p$ y formado por la media de cada una de las variables de la tabla, que se puede interpretar como el individuo medio.

Para que el origen de dicha nube coincida con el centro de gravedad, es necesario realizar una transformación en la tabla de datos original. Cada elemento x_{ij} pasa a ser $y_{ij} = x_{ij} - \bar{x}_j$ de tal manera que cada individuo se representa por $e_i' = (y_{i1}, y_{i2} \dots y_{ip})$. Con esta transformación la cantidad de información de la tabla de datos original coincide con la inercia de dicha tabla. Gráficamente y en el caso de tener solo tres variables, esta transformación supone un desplazamiento del origen, tal y como muestra la siguiente figura:

Figura II.4
Nuevo origen: el centro de gravedad de la nube



Con la transformación indicada la media de la variable Y_j toma el valor:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_{ij} = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j) = 0$$

y la varianza de la variable X_j coincide con la varianza de la variable Y_j :

$$\text{Var}(Y_j) = \frac{1}{n} \sum_{i=1}^n (y_{ij} - \bar{y}_j)^2 = \frac{1}{n} \sum_{i=1}^n y_{ij}^2 = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 = s_j^2$$

Es necesario definir ahora las distancias entre individuos y las distancias al centro de gravedad teniendo en cuenta la transformación previa, por lo que el espacio E debe estar dotado de una estructura euclídea a fin de poder definir dichas distancias; por tanto nos referiremos previamente a la métrica y al producto escalar entre individuos.

— Denominaremos M a la matriz simétrica, definida no negativa, cuyos elementos indican el peso de cada variable Y_j y la perpendicularidad de los ejes. La matriz M , de orden $p \times p$, es la *métrica* del espacio E y tiene la siguiente expresión:

$$M = \begin{bmatrix} m_1 & 0 & \dots & \dots & 0 \\ 0 & m_2 & 0 & \dots & 0 \\ \vdots & & & & \vdots \\ 0 & \dots & \dots & 0 & m_p \end{bmatrix}$$

— El *producto escalar* entre e_i y e_i , se define:

$$\langle e_i, e_i \rangle = w_{ii} = e_i' M e_i = \sum_{j=1}^p m_j y_{ij} y_{ij}$$

La matriz de término general w_{ii} se denominará $W = YMY'$. El producto escalar del vector e_i por si mismo se denota:

$$\|e_i\|_M^2 = e_i' M e_i$$

y la norma de ese vector e_i que se denomina M -norma, es:

$$\|e_i\|_M = \sqrt{\|e_i\|_M^2}$$

— En el espacio R^p , la *distancia* entre dos individuos e_i y e_i , de la nube viene dada por la siguiente expresión:

$$d^2(e_i, e_i) = \sum_{j=1}^p m_j (y_{ij} - y_{i'j})^2 = (e_i - e_{i'})' M (e_i - e_{i'})$$

Si los ejes son perpendiculares, la matriz M suele ser la matriz identidad ($M = I$, de orden $p \times p$) y la distancia es la distancia euclídea usual. En este caso todas las variables tendrían el mismo peso (unitario) en la definición de las distancias entre los individuos. Si las variables están medidas en distintas unidades, o la dispersión de cada una de ellas es muy diferente, no todas las variables tendrán la misma influencia en la definición de la distancia entre dos individuos. Por tanto, en estos casos se toma como métrica la matriz diagonal M^* , formada por los inversos de las varianzas de las variables:

$$M^* = \begin{bmatrix} 1/s_1^2 & 0 & \dots & 0 \\ 0 & 1/s_2^2 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & 0 & & 1/s_p^2 \end{bmatrix}$$

Esta métrica permite otorgar a cada variable la misma importancia, cualquiera que sea su dispersión y cualquiera que sea la unidad de medida. Con esta transformación todas las variables van a tener varianza unitaria (también media nula) y por tanto la misma dispersión. Dado que M^* es simétrica y definida positiva, existirá una matriz T tal que $M^* = T^* T$ por lo que:

$$\langle e_i, e_{i'} \rangle = e_i' M e_{i'} = (T e_i)' (T e_{i'})$$

Por este motivo es posible utilizar la métrica $M = I_{(n \times n)}$ siempre y cuando hayamos transformado los datos originales de tal manera que cada individuo $e_i = (z_{i1}, z_{i2} \dots z_{ip})$, siendo:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

El producto escalar toma la expresión:

$$w_{ii'} = \sum_{j=1}^p m_j z_{ij} z_{i'j}$$

y la matriz de productos escalares entre individuos es $W = ZMZ'$.

Cuando se realiza esta transformación en los datos se realiza un *Análisis de Componentes Principales normado*, que es al que nos vamos a referir en este apartado. Así:

$$M = \begin{bmatrix} m_1 & 0 & \dots & 0 \\ 0 & m_2 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & \dots & 0 & m_p \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & \dots & 0 & 1 \end{bmatrix} = M^{-1} = I(p \times p)$$

— La *inerencia de la nube*, teniendo en cuenta que el centro de gravedad coincide con el origen, toma la siguiente expresión:

$$I = \sum_{i=1}^n p_i d^2(e_i, g) = \sum_{i=1}^n p_i d^2(e_i, 0)$$

Al ser:

$$d^2(e_i, 0) = \sum_{j=1}^p m_j (z_{ij} - 0)^2 = \sum_{j=1}^p m_j z_{ij}^2 = e_i' M e_i = \|e_i\|_M^2$$

se deduce que:

$$I = \sum_{i=1}^n p_i d^2(e_i, g) = \sum_{i=1}^n p_i \sum_{j=1}^p m_j z_{ij}^2 = \sum_{i=1}^n p_i e_i' M e_i = TR(WD)$$

siendo D , de orden $n \times n$, la matriz diagonal de pesos de los individuos.

Generalmente todos los individuos tienen el mismo peso e igual a $1/n$. Así:

$$D = \begin{bmatrix} p_1 & 0 & \dots & 0 \\ 0 & p_2 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & \dots & 0 & p_n \end{bmatrix} = \begin{bmatrix} 1/n & 0 & \dots & 0 \\ 0 & 1/n & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & \dots & 0 & 1/n \end{bmatrix}$$

Si la inercia de la nube se descompone según los individuos, el cociente:

$$\frac{p_i d^2(e_i, 0)}{Tr(WD)} = \frac{p_i e_i' M e_i}{Tr(WD)}$$

representa la *parte de la inercia* total que corresponde a cada individuo e_i .

b. La nube de variables

Siguiendo el mismo procedimiento que para el caso de los individuos, vamos a definir las métricas, los productos escalares, las distancias y la cantidad de información de la nube N_j teniendo en cuenta, previamente, las transformaciones realizadas.

La nube de variables se define en el espacio $R^n = F$ y supone tomar la tabla de datos como una yuxtaposición de columnas. Debido a las transformaciones ya indicadas, cuando se realiza un ACP *normado* (variables centradas y reducidas), cada variable se denomina Z_j :

$$Z_j = \begin{bmatrix} z_{1j} \\ z_{2j} \\ \vdots \\ z_{nj} \end{bmatrix}$$

siendo:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

La media de la nueva variable es cero y su varianza es unitaria. La matriz de varianzas-covarianzas de las matrices Z_j se denominará V_z . Por ello se tiene que la media de Z_j es:

$$\bar{z}_j = \frac{1}{n} \sum_{i=1}^n z_{ij} = \frac{1}{n} \sum_{i=1}^n \frac{x_{ij} - \bar{x}_j}{s_j} = \frac{1}{ns_j} \sum_{i=1}^n (x_{ij} - \bar{x}_j) = 0$$

y la varianza es:

$$\text{Var}(Z_j) = \frac{1}{n} \sum_{i=1}^n (z_{ij} - \bar{z}_j)^2 = \frac{1}{n} \sum_{i=1}^n z_{ij}^2 = \frac{1}{n} \sum_{i=1}^n \frac{(x_{ij} - \bar{x}_j)^2}{s_j^2} = \frac{s_j^2}{s_j^2} = 1$$

La covarianza entre Z_j y $Z_{j'}$ viene dada por la expresión:

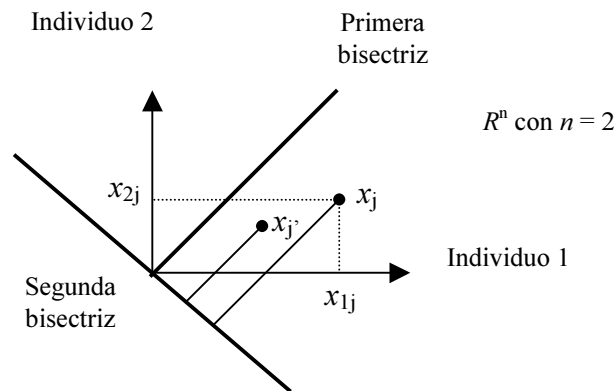
$$\text{Cov}(Z_j, Z_{j'}) = \frac{1}{n} \sum_{i=1}^n (z_{ij} - \bar{z}_j)(z_{ij'} - \bar{z}_{j'}) = \frac{1}{n} \sum_{i=1}^n z_{ij} z_{ij'}$$

La tabla completa está representada por la matriz Z y tiene la siguiente expresión:

$$Z = Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1j} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2j} & \cdots & z_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nj} & \cdots & z_{np} \end{bmatrix}$$

Las consecuencias de esta transformación en R^n son diferentes a las que se producían en el espacio de los individuos. En este caso no se produce un cambio de origen. De hecho, la nube de variables no va a estar centrada y su origen no coincide con el centro de gravedad. En R^n la transformación $x_{ij} - \bar{x}_j$ es una proyección paralela a la primera bisectriz de los ejes sobre el hiperplano que le es ortogonal, tal y como refleja la figura siguiente:

Figura II.5
 Consecuencias de la transformación $x_{ij} - \bar{x}_j$ en N_j



Para las variables definimos, una métrica y un producto escalar, con el fin de poder medir la distancia entre variables y la cantidad de información de la nube formada por ellas.

— La *métrica* en este caso es la matriz de pesos de los individuos D . La relación entre las métricas y los pesos de R^n y R^p tiene propiedades muy interesantes que se pondrán de manifiesto con las fórmulas de transición.

— El *producto escalar* entre dos variables Z_j y Z_j' viene dado por:

$$Z_j' D Z_j = \sum_{i=1}^n p_i z_{ij} z_{ij}$$

y la matriz³ (Lebart *et al.*, 1995: 41) que recoge dichos productos escalares es $V_z = Z' D Z$.

— La *distancia* entre dos variables vendrá dada por la siguiente expresión:

$$d^2(z_j, z_{j'}) = \sum_{i=1}^n p_i (z_{ij} - z_{ij'})^2 = \sum_{i=1}^n p_i z_{ij}^2 - \sum_{i=1}^n p_i z_{ij'}^2 - 2 \sum_{i=1}^n p_i z_{ij} z_{ij'}$$

Dado que $p_i = 1/n$ se cumple que:

$$\sum_{i=1}^n p_i z_{ij}^2 = \sum_{i=1}^n p_i z_{ij'}^2 = 1$$

Esta expresión coincide con las varianzas de Z_j y $Z_{j'}$ por lo que:

$$d^2(z_j, z_{j'}) = 2 - 2 \sum_{i=1}^n p_i z_{ij} z_{ij'} = 2(1 - \sum_{i=1}^n p_i z_{ij} z_{ij'}) = 2(1 - \rho(Z_j, Z_{j'}))$$

siendo $\rho(Z_j, Z_{j'})$ el coeficiente de correlación entre las dos variables Z_j y $Z_{j'}$ teniendo en cuenta que:

$$\rho(Z_j, Z_{j'}) = \frac{\frac{1}{n} \sum_{i=1}^n (z_{ij} - \bar{z}_j)(z_{ij'} - \bar{z}_{j'})}{\sqrt{\text{Var}(Z_j)} \sqrt{\text{Var}(Z_{j'})}}$$

Como Z_j y $Z_{j'}$ tienen media nula y varianza unitaria, el coeficiente de correlación es:

$$\rho(Z_j, Z_{j'}) = \frac{1}{n} \sum_{i=1}^n z_{ij} z_{ij'} = \sum_{i=1}^n p_i z_{ij} z_{ij'} = \text{Cov}(Z_j, Z_{j'})$$

Por otra parte este coeficiente de correlación coincide con el de las variables originales X_j y $X_{j'}$ puesto que:

³ Ese producto escalar $\sum_{i=1}^n p_i z_{ij} z_{ij'}$ coincide con $\text{Cov}(Z_j, Z_{j'})$ pues $p_i = 1/n$ para todo e_i .

$$\sum_{i=1}^n p_i Z_{ij} Z_{ij'} = \sum_{i=1}^n \frac{1}{n} \frac{(x_{ij} - \bar{x}_j)(x_{ij'} - \bar{x}_{j'})}{s_j s_{j'}} = \frac{\text{Cov}(X_j, X_{j'})}{s_j s_{j'}} = \rho(X_j, X_{j'})$$

En definitiva, la proximidad entre dos variables se expresará en términos de su coeficiente de correlación con las siguientes consecuencias:

- Dos variables fuertemente correlacionadas estarán muy próximas entre sí (coeficiente igual a uno y por tanto distancia igual a cero); o por el contrario, lo más alejadas posibles (coeficiente igual a uno y distancia igual a dos).
- Si las variables son ortogonales, y por lo tanto su coeficiente de correlación es 0, estarán a una distancia intermedia, cuyo valor es $\sqrt{2}$.
- Cada vector que representa a cada variable tiene norma unitaria, pues la raíz cuadrada de:

$$\|Z_j\|_D^2 = Z_j' D Z_j = \sum_{i=1}^n p_i z_{ij}^2 = 1 = \text{Var}(Z_j)$$

que es también uno.

- La distancia al origen de una variable Z_j es:

$$d^2(Z_j, 0) = \sum_{i=1}^n p_i (z_{ij} - 0)^2 = \sum_{i=1}^n p_i z_{ij}^2 = 1 = \|Z_j\|_D^2$$

Dado que se cumple la igualdad anterior para cualquier X_j , todas las variables estarán a la distancia unitaria del origen y por tanto los puntos-variables estarán situados sobre una esfera de radio uno y centro el origen de los ejes, formando lo que se denomina *esfera de correlaciones*. La representación de estos puntos en un espacio de dos dimensiones dará lugar a lo que se denomina *círculo de correlaciones*, en el interior del cual se encuentra la representación de las variables.

- Como el coeficiente de correlación que forman los vectores que representan a las dos variables X_j y $X_{j'}$, coincide con el coseno del

ángulo⁴ entre ambas y además están a la misma distancia del origen, la proximidad entre variables se interpreta en función del ángulo que forman. Por tanto, al estudiar el espacio de las variables, el objetivo son las direcciones determinadas por ellas, es decir, su tratamiento como vectores y no como puntos. La interpretación del coeficiente de correlación como un coseno es una propiedad del ACP *normado* muy importante porque le da apoyo geométrico y visual al método.

— La *cantidad de información* viene dada por la siguiente expresión:

$$\sum_{j=1}^p m_j d^2(Z_j, 0) = \sum_{j=1}^p m_j \text{Var}(Z_j)$$

teniendo en cuenta que $d^2(Z_j, 0) = \text{Var}(Z_j) = 1$ y $m_j = 1$ se tiene que:

$$\sum_{j=1}^p m_j \text{Var}(Z_j) = p$$

Al verificarse:

$$\sum_{j=1}^p m_j d^2(Z_j, 0) = \sum_{j=1}^p m_j \sum_{i=1}^n p_i z_{ij}^2 = \sum_{j=1}^p \sum_{i=1}^n m_j p_i z_{ij}^2 = \sum_{j=1}^p m_j Z_j' D Z_j$$

se cumple que:

$$\sum_{j=1}^p m_j d^2(Z_j, 0) = \text{Tr}(V_z M) = \text{Tr}(M V_z)$$

siendo V_z la matriz de varianzas-covarianzas de las matrices Z_j y Z_j' , que coincide con R , que es la matriz de correlaciones entre las variables originales X_j y X_j' , puesto que:

$$\text{Cov}(Z_j, Z_j') = \rho(Z_j, Z_j') = \rho(X_j, X_j')$$

Por tanto:

$$\text{Tr}(V_z M) = \text{Tr}(M V_z) = p = \text{Tr}(MR)$$

⁴ El coseno del ángulo entre dos vectores es igual al cociente entre el producto escalar de ambos y el producto de sus normas. Dado que en este caso las normas son unitarias, dicho coseno coincide con el producto escalar que es el coeficiente de correlación entre ambos.

Como ya habíamos demostrado que:

$$\sum_{j=1}^p m_j \sum_{i=1}^n p_i z_{ij}^2 = \sum_{i=1}^n \sum_{j=1}^p m_j p_i z_{ij}^2 = \sum_{i=1}^n p_i \sum_{j=1}^p m_j z_{ij}^2 = \sum_{i=1}^n p_i e_i' M e_i = \text{Tr}(WD)$$

la inercia de la nube de individuos coincide con la cantidad de información de la nube de variables y se tiene que:

$$\text{Tr}(WD) = \text{Tr}(RM) = p$$

Podemos, por tanto, descomponer la cantidad de información total y determinar que parte corresponde a cada variable. El cociente:

$$\frac{m_j d^2(Z_j, 0)}{\text{Tr}(V_z M)} = \frac{m_j Z_j' D Z_j}{\text{Tr}(V_z M)}$$

representa la parte que se debe a la variable Z_j . Dado que todas las variables tienen varianza unitaria, todas tienen la misma influencia ($1/p$) en la inercia total. Este valor tiene especial interés si las variables sólo están centradas, en cuyo caso tenemos la siguiente expresión:

$$\frac{m_j d^2(Y_j, 0)}{\text{Tr}(VM)} = \frac{m_j Y_j' D Y_j}{\text{Tr}(VM)}$$

Una vez caracterizadas ambas nubes estamos en condiciones de analizar detalladamente cada una de ellas a través del ACP *normado*.

1.3.2 Desarrollo del ACP *normado*

Es preciso recordar que, a partir de los datos originales no es posible obtener una imagen adecuada de las distancias entre individuos, razón por la cual es necesario buscar una representación de los mismos en un espacio de menor dimensión (dos o tres a lo sumo) que refleje de forma

adecuada las distancias originales entre los individuos y que mantenga una cantidad importante de información de la tabla de datos original.

Considerando las características diferenciales de las dos nubes, vamos a describir el procedimiento que permite obtener los ejes de cada una de ellas que resuman lo mejor posible la información original, así como las coordenadas de los individuos y de las variables en esos nuevos ejes. Asimismo, se establecerán las conexiones entre los ejes de ambas nubes, considerando las relaciones de transición, y la representación simultánea de individuos y variables.

a. Análisis de la nube de individuos N_I

Para los individuos, y sabiendo que se trata de maximizar la inercia explicada, vamos a ir obteniendo sucesivamente los nuevos ejes y las coordenadas de los individuos en ellos (las componentes principales). Comenzando por la primera de ellas, vamos a obtener posteriormente las siguientes, más exactamente los h primeras. Las componentes principales permiten reconstituir la tabla de datos original. También es posible calcular las coordenadas de otros individuos, denominados individuos suplementarios, que no han intervenido en la construcción de los ejes pero cuya posición puede tener una interesante interpretación.

LA PRIMERA COMPONENTE PRINCIPAL

Tal y como la define Cuadras (1980: 269) una componente (o factor) principal F_k es otra variable, obtenida a partir de los nuevos ejes, que es combinación lineal de las variables observables y que tiene varianza máxima. Para su obtención es necesario maximizar la inercia de las proyecciones de los individuos sobre el eje que la define, definiendo previamente dicho eje y el vector director correspondiente.

– Denominaremos por:

$$a_1 = \begin{bmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1p} \end{bmatrix}$$

el *vector director* de una recta que pasando por el centro de gravedad maximiza la inercia de la nube proyectada sobre esa recta. Tomando la matriz M como métrica del espacio R^p la proyección del individuo e_i sobre dicha recta es f_{1i} cuyo valor es:

$$f_{1i} = e_i' M a_1 = e_i' u_1 = \sum_{j=1}^p z_{ij} m_j a_{1j} = \sum_{j=1}^p z_{ij} u_{1j}$$

siendo a_1 un vector de norma unitaria $a_1' M a_1 = 1$.

Llamando $u_1 = M a_1$ se cumple que: $u_1' M^{-1} u_1 = 1$, siendo el *primer eje factorial*:

$$u_1 = \begin{bmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{1p} \end{bmatrix}$$

La proyección de todos los individuos sobre dicha recta da lugar a un vector F_1 :

$$F_1 = \begin{bmatrix} f_{11} \\ f_{12} \\ \vdots \\ f_{1n} \end{bmatrix}$$

perteneciente al espacio de variables R^n tal que $F_1 = Z M a_1 = Z u_1$

F_1 puede interpretarse como una nueva variable artificial, combinación lineal de las p variables, que recibe el nombre de *primera componente principal* o *primer factor principal*.

- Calculamos dicha componente comenzando por la obtención del vector a_1 . Dicho vector hace maximiza la inercia de N_1 proyectada sobre la recta de la que es vector director, inercia que vendrá dada como la suma ponderada de las distancias (al cuadrado) de las proyecciones de los puntos al origen (que en este caso es el centro de gravedad, pues la nube de individuos está centrada):

$$\begin{aligned} \sum_{i=1}^n p_i d^2(F_1, 0) &= \sum_{i=1}^n p_i f_{1i}^2 = F_1' D F_1 = (Z M a_1)' D (Z M a_1) = \\ &= a_1' M Z' D Z M a_1 \end{aligned}$$

dado que $M = M'$ por ser $M = I$.

A la matriz $S_1 = M Z' D Z$ se le denomina *matriz de inercia* de la nube de individuos y verifica las siguientes propiedades:

- Es una matriz simétrica y, en nuestro caso, $S_1 = Z' D Z = V_z = R$ coincide con la matriz de correlaciones entre las variables originales.
- La traza de la matriz de inercia $M V_z = R$ es la suma de las varianzas de las variables Z_j , es decir, la inercia de la nube N_1 . Ya hemos demostrado que puesto que son unitarias su traza es igual al número de variables que es p .

Se deduce, por tanto, que $a_1' M Z' D Z M a_1 = a_1' S_1 M a_1$, expresión que hay que maximizar considerando que $\|a_1\|_M^2 = 1$. Por tanto el programa de optimización se plantea de la forma siguiente:

$$\begin{cases} \text{Max : } & a_1' S_1 M a_1 \\ \text{sujeto a : } & a_1' M a_1 = 1 \end{cases}$$

Puesto que $u_1 = M a_1$ y que $u_1' M^{-1} u_1 = 1$, dicho programa se transforma en el siguiente:

$$\begin{cases} \text{Max : } & u_1' M^{-1} S_1 u_1 \\ \text{sujeto a : } & u_1' M^{-1} u_1 = 1 \end{cases}$$

La resolución de dicho programa se realiza mediante el método de los multiplicadores de Lagrange, lo que requiere la construcción de la función lagrangiana:

$$L = u_1' M^{-1} S_1 u_1 - \lambda_1 (u_1' M^{-1} u_1 - 1)$$

Derivando respecto a u_1 :

$$\frac{\partial L}{\partial u_1} = 2M^{-1}S_1u_1 - 2\lambda_1M^{-1}u_1 = 0$$

$$\Rightarrow M^{-1} S_1 u_1 = \lambda_1 M^{-1} u_1$$

$$\Rightarrow S_1 u_1 = \lambda_1 u_1$$

$$\Rightarrow MR u_1 = \lambda_1 u_1$$

Como: $S_1 = MZ' D Z$ se tiene que:

$$M^{-1} MZ' D Z u_1 = \lambda_1 M^{-1} u_1$$

$$\Rightarrow Z' D Z M a_1 = \lambda_1 a_1$$

$$\Rightarrow R M a_1 = \lambda_1 a_1$$

Así, la inercia de la nube proyectada sobre un eje u_1 es máxima cuando ese eje es colineal a los vectores propios asociados al mayor valor propio λ_1 de la matriz MR . El vector director de la recta, a_1 , es el vector asociado al primer valor propio de RM . Dado que $M = I$ se cumple que $a_1 = u_1$ que coincide con el vector propio asociado al primer valor propio de R , (matriz de correlaciones de las variables originales), que en este caso es la matriz de varianzas –covarianzas de las variables Z_j .

- Una vez especificado el proceso de determinación de la primera componente principal, se deducen las propiedades de la misma y del correspondiente eje.
 - La media de las proyecciones de los individuos sobre el primer eje factorial es nula:

$$\bar{F}_1 = \sum_{i=1}^n p_i f_{1i} = \sum_{i=1}^n p_i \sum_{j=1}^p z_{ij} m_j a_{1j} = \sum_{i=1}^n \sum_{j=1}^p p_i z_{ij} m_j a_{1j} =$$

$$= \sum_{j=1}^p m_j a_{1j} \sum_{i=1}^n p_j z_{ij} = 0$$

puesto que: $\sum_{i=1}^n p_j z_{ij} = 0$, por ser la media de la variable Z_j .

Es decir, la primera componente principal F_1 es una variable de media igual a cero.

- Puesto que $\bar{F}_1 = 0$, la inercia del eje obtenido toma la siguiente expresión:

$$\sum_{i=1}^n p_i d^2(F_1, 0) = \sum_{i=1}^n p_i f_{1i}^2 = \sum_{i=1}^n p_i (f_{1i} - \bar{F}_1)^2 = \text{Var}(F_1)$$

Es decir, la inercia del eje es en realidad la varianza de las proyecciones sobre dicho eje, y éste es el que maximiza la varianza de las proyecciones de los puntos de N_I , es decir hace máxima $F_1' D F_1 = a_1' M Z' D Z M a_1$.

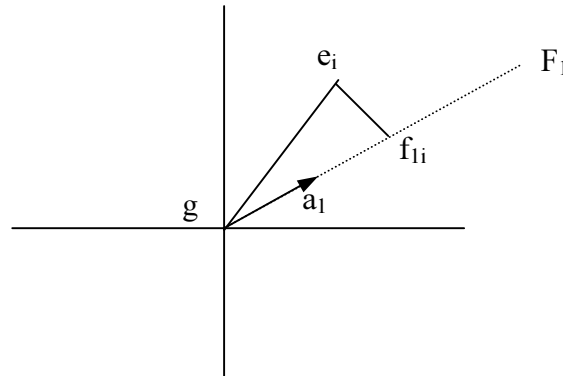
Teniendo en cuenta que $S_1 u_1 = \lambda_1 u_1$ y $u_1' M^{-1} u_1 = 1$, se deduce:

$$\begin{aligned} \text{Var}(F_1) &= \sum_{i=1}^n p_i f_{1i}^2 = F_1' D F_1 = a_1' M Z' D Z M a_1 = a_1' S_1 M a_1 = \\ &= u_1' M^{-1} S_1 u_1 = u_1' M^{-1} \lambda_1 u_1 = \lambda_1 \end{aligned}$$

- El primer eje factorial corresponde al eje que, pasando por el origen (ya que coincide con el centro de gravedad), mejor se ajusta a la nube de puntos N_I , en el sentido de mínimos cuadrados ordinarios ortogonales.

En la figura II.6, se observa que para el individuo e_i se cumple que $d^2(e_i, 0) = d^2(f_{1i}, 0) + d^2(f_{1i}, e_i)$.

Figura II.6
Proyección del individuo e_i sobre el eje u_1 (vector director a_1)



Como lo que se trata de maximizar son las distancias (al cuadrado) de las proyecciones de los puntos al origen, tenemos que:

$$d^2(f_{1i}, 0) = d^2(e_i, 0) - d^2(f_{1i}, e_i)$$

lo que implica que:

$$\text{Max: } d^2(f_{1i}, 0) = \text{Max: } d^2(e_i, 0) - d^2(f_{1i}, e_i)$$

Puesto que la distancia $d^2(e_i, 0)$ es fija, $d^2(f_{1i}, e_i)$ debe ser mínima.

Si consideramos los n individuos y las ponderaciones correspondientes, al aplicar el sumatorio se tiene:

$$\text{Max: } \sum_{i=1}^n p_i d^2(f_{1i}, 0) \Rightarrow \text{Min: } \sum_{i=1}^n p_i d^2(f_{1i}, e_i)$$

lo que supone minimizar la suma de los cuadrados de los errores (distancia entre la proyección y el punto) del conjunto de puntos de N_1 . De la misma expresión $d^2(f_{1i}, 0) = d^2(e_i, 0) - d^2(f_{1i}, e_i)$ se deduce también que $d^2(f_{1i}, 0) < d^2(e_i, 0)$, como ya habíamos señalado.

- El primer eje factorial maximiza la suma de los cuadrados de las distancias entre proyecciones de pares de puntos. Así, la representación de las proyecciones refleja lo mejor posible las

distancias entre pares de individuos de la nube original. Como la distancia entre las proyecciones de dos puntos es:

$$d^2(f_{1i}, f_{1i'}) = (f_{1i} - f_{1i'})^2 = f_{1i}^2 + f_{1i'}^2 - 2f_{1i}f_{1i'}$$

al tomar en consideración todos los pares de puntos, se tiene que:

$$\begin{aligned} \sum_{i=1}^n \sum_{i'=1}^n p_i p_{i'} d^2(f_{1i}, f_{1i'}) &= \sum_{i=1}^n \sum_{i'=1}^n p_i p_{i'} (f_{1i}^2 + f_{1i'}^2 - 2f_{1i}f_{1i'}) = \\ &= \sum_{i=1}^n \sum_{i'=1}^n p_i p_{i'} f_{1i}^2 + \sum_{i=1}^n \sum_{i'=1}^n p_i p_{i'} f_{1i'}^2 - 2 \sum_{i=1}^n \sum_{i'=1}^n p_i p_{i'} f_{1i} f_{1i'} = \\ &= \sum_{i=1}^n p_{i'} \sum_{i=1}^n p_i f_{1i}^2 + \sum_{i=1}^n p_i \sum_{i'=1}^n p_{i'} f_{1i'}^2 - 2 \sum_{i=1}^n p_i f_{1i} \sum_{i'=1}^n p_{i'} f_{1i'} = \\ &= \sum_{i=1}^n p_{i'} \sum_{i=1}^n p_i f_{1i}^2 + \sum_{i=1}^n p_i \sum_{i'=1}^n p_{i'} f_{1i'}^2 = \quad \downarrow \quad 0 \\ &= \sum_{i=1}^n p_{i'} \text{Var}(F_1) + \sum_{i=1}^n p_i \text{Var}(F_1) = \\ &= n \text{Var}(F_1) + n \text{Var}(F_1) = 2n \text{Var}(F_1) = 2n \lambda_1 \end{aligned}$$

Luego:

$$\sum_{i=1}^n \sum_{i'=1}^n p_i p_{i'} d^2(f_{1i}, f_{1i'}) = 2n \lambda_1$$

Como la $\text{Var}(F_1) = \lambda_1$ es máxima, hemos demostrado que también se hace máxima la suma de las distancias entre las proyecciones de pares de puntos. Así el eje obtenido, primer eje factorial, es el que menos deforma la nube de puntos.

LAS h COMPONENTES O FACTORES PRINCIPALES

Para su determinación, se va a seguir el mismo procedimiento que en el caso de la primera componente; es decir, se definen los conceptos necesarios, se plantea el problema de optimización, se resuelve y se extraen las conclusiones oportunas. Estas h componentes son la mejor aproximación, de orden h , de las variables originales como posteriormente veremos.

— Se denomina inercia de N_I , respecto al origen, de un conjunto de h ejes ($h \leq p$) ortogonales, asociados a los vectores unitarios $(a_1, a_2 \dots a_h)$ a la inercia, respecto al origen, de la proyección de la nube N_I en el subespacio engendrado por esos ejes de tal manera que:

$$I(N_I; a_1, a_2 \dots a_h) = \sum_{i=1}^n p_i d_h^2(e_i, 0)$$

siendo $I(N_I; a_1, a_2 \dots a_h)$ dicha inercia, y $d_h^2(e_i, 0)$ el cuadrado de la distancia al origen de la proyección en el subespacio R^h del individuo e_i .

Llamando:

$$F^i = \begin{bmatrix} f_{1i} \\ f_{2i} \\ \vdots \\ f_{hi} \end{bmatrix}$$

al vector cuyas componentes son las proyecciones del individuo e_i en cada uno de los h ejes, se cumple que:

$$d_h^2(e_i, 0) = \sum_{k=1}^h (f_{ki} - 0)^2 = \sum_{k=1}^h f_{ki}^2 = (F^i)' F^i$$

Por ello se tiene que:

$$I(N_I; a_1, a_2 \dots a_h) = \sum_{i=1}^n p_i \sum_{k=1}^h f_{ki}^2 = \sum_{k=1}^h \sum_{i=1}^n p_i f_{ki}^2 = \sum_{k=1}^h F_k' D F_k$$

Esta expresión es la suma de las varianzas de las proyecciones de los n individuos en los h ejes. Si de todos los posibles subespacios de h dimensiones, el subespacio engendrado por los vectores M -ortogonales $(a_1, a_2 \dots a_h)$ es aquél que maximiza la inercia de la nube N_I , se tiene que:

$$\begin{aligned} I(N_I; a_1, a_2 \dots a_h) &= \sum_{k=1}^h F_k' D F_k = \sum_{k=1}^h a_k' M Z' D Z M a_k = \\ &= \sum_{k=1}^h a_k' S_1 M a_k = \sum_{k=1}^h u_k' M^{-1} S_1 u_k \end{aligned}$$

ya que $u_k = M a_k$.

— Se deduce que el correspondiente programa de optimización se debe plantear en los términos siguientes:

$$\left\{ \begin{array}{l} \text{Max : } \sum_{k=1}^h u_k' M^{-1} S_I u_k \\ \text{sujeto a : } u_k' M^{-1} u_k = 1 \quad , , \quad k = 1, 2 \dots h \\ u_k' M^{-1} u_{k'} = 0 \quad , , \quad k \neq k' \quad (\text{condición de ortogonalidad}) \end{array} \right.$$

La resolución del mismo nos conduce a la obtención de los h ejes. Posteriormente demostraremos que la condición de ortogonalidad no es necesaria, en cuanto que los ejes obtenidos son los vectores propios de la matriz de inercia, y por tanto, ya son ortogonales dos a dos.

Determinando el Lagrangiano, como para el caso del primer eje, y sin considerar por tanto esas últimas condiciones, tendremos:

$$L = \sum_{k=1}^h u_k' M^{-1} S_I u_k - \sum_{k=1}^h \lambda_k (u_k' M^{-1} u_k - 1)$$

tomando derivadas respecto a u_k :

$$\frac{\partial L}{\partial u_k} = \begin{bmatrix} \frac{\partial L}{\partial u_{k1}} \\ \frac{\partial L}{\partial u_{k2}} \\ \vdots \\ \frac{\partial L}{\partial u_{kp}} \end{bmatrix} = 2 \sum_{k=1}^h M^{-1} S_I u_k - 2 \sum_{k=1}^h \lambda_k M^{-1} u_k = 0 \Rightarrow$$

$$\Rightarrow \sum_{k=1}^h M^{-1} S_I u_k = \sum_{k=1}^h \lambda_k M^{-1} u_k \Rightarrow$$

$$\Rightarrow \sum_{k=1}^h S_I u_k = \sum_{k=1}^h \lambda_k u_k \Rightarrow$$

$$\Rightarrow S_I \sum_{k=1}^h u_k = \sum_{k=1}^h \lambda_k u_k \Rightarrow$$

$$\Rightarrow MR \sum_{k=1}^h u_k = \sum_{k=1}^h \lambda_k u_k$$

Al sustituir S_I por $MZ'DZ$ y $u_k = Ma_k$ se tiene que:

$$\begin{aligned} \sum_{k=1}^h M^{-1} MZ'DZ u_k &= \sum_{k=1}^h \lambda_k M^{-1} Ma_k \Rightarrow \sum_{k=1}^h Z'DZ_I Ma_k = \sum_{k=1}^h \lambda_k a_k \Rightarrow \\ &\Rightarrow Z'DZM \sum_{k=1}^h a_k = \sum_{k=1}^h \lambda_k a_k \Rightarrow RM \sum_{k=1}^h a_k = \sum_{k=1}^h \lambda_k a_k \end{aligned}$$

Por tanto, la inercia de la nube proyectada sobre los h ejes $(u_1, u_2 \dots u_h)$ es máxima cuando dichos ejes son colineales a los vectores propios asociados a los valores propios $(\lambda_1, \lambda_2 \dots \lambda_k)$ de la matriz MR . De la misma forma que en el caso anterior se demuestra que dicha inercia vale precisamente:

$$\sum_{k=1}^h \lambda_k$$

Los vectores directores de las rectas $(a_1, a_2 \dots a_h)$ son los vectores asociados a los h primeros valores propios de RM . Como $M = I$, para todo k , se cumple que $a_k = u_k$, siendo éste el vector propio k de la matriz R , que es la matriz de correlaciones de las variables originales, que coincide con la matriz de varianzas-covarianzas de las variables Z_j .

— Una vez llevado a cabo todo este proceso, se pueden deducir las siguientes conclusiones:

- Estos h ejes $(u_1, u_2 \dots u_h)$ son los que hacen máxima la suma de las varianzas de las proyecciones de los puntos-observaciones sobre dichos ejes.
- El subespacio de dimensión h , según lo ya demostrado para F_1 , es el que mejor se ajusta a la nube de puntos en el sentido de los mínimos cuadrados ordinarios.
- Este subespacio de h dimensiones es, por tanto, el que menos deforma la nube de puntos original y el que mejor conserva la distancia entre pares de puntos cuando N_I se proyecta sobre R^h . En realidad la transformación realizada es una rotación ortogonal en R^p

de los ejes originales de tal manera que, con los nuevos ejes, las distancias entre individuos en dicho espacio se conservan.

- Como R es una matriz cuadrada, en tanto que es una matriz diagonalizable, su traza es igual a la suma de sus valores propios. R es de orden $p \times p$ y si su rango es p el número de valores propios no nulos será también p . Así se cumple que:

$$\begin{aligned} \sum_{k=1}^p \lambda_k &= Tr(R) = Tr(S_I) = Tr(MV_Z) = \\ &= \sum_{j=1}^p m_j Var(Z_j) = p = \text{Inercia total de la nube} \end{aligned}$$

Además, los vectores propios de una matriz cuadrada, asociados a los valores propios, son siempre ortogonales, motivo por el cual la condición de ortogonalidad es redundante⁵.

- Teniendo en cuenta lo demostrado para F_1 , se cumple que $F_k' D F_k = Var(F_k) = \lambda_k$, es decir, la varianza de la componente k es el valor propio correspondiente. Por tanto:

$$I(N_1; a_1, a_2 \dots a_h) = \sum_{k=1}^h F_k' D F_k = \sum_{k=1}^h \lambda_k$$

es decir, la suma de las varianzas de las proyecciones de los individuos sobre los h primeros ejes es igual a la suma de los valores propios.

Como consecuencia $F_1, F_2 \dots F_h$ son las *componentes o factores principales* y la suma de sus varianzas es la suma de los valores propios correspondientes, es decir:

$$\sum_{k=1}^h \lambda_k$$

Premultiplicando por Z la expresión $S_I u_k = \lambda_k u_k$ y tomando solamente el eje k se tiene que:

⁵ Para demostraciones algebraicas más detalladas nos remitimos a la obra de M. Queysanne (1975: 554 -60).

$$\begin{aligned} Z S_1 u_k &= \lambda_k Z u_k \Rightarrow Z M Z' D Z u_k = \lambda_k Z u_k \Rightarrow \\ &\Rightarrow Z M Z' D F_k = \lambda_k F_k \Rightarrow W D F_k = \lambda_k F_k \end{aligned}$$

Si consideramos el conjunto de las h componentes y le aplicamos sumatorios se tiene que:

$$\sum_{k=1}^h W D F_k = \sum_{k=1}^h \lambda_k F_k \Rightarrow W D \sum_{k=1}^h F_k = \sum_{k=1}^h \lambda_k F_k$$

Por lo que las h componentes principales son los vectores propios de la matriz WD asociados a los primeros h valores propios.

— Propiedades de las componentes principales:

- De la siguiente relación: $F_k = Z u_k$, se deduce que las componentes principales son combinaciones lineales de las variables iniciales. Dado que la nube está centrada, las posiciones de los individuos dependen de su comportamiento respecto al individuo-medio (centro de gravedad) y, por tanto, los ejes factoriales obtenidos pueden interpretarse como las direcciones de máximo alargamiento o variabilidad (Escofier, Pagès, 1992: 76) de la nube respecto de dicho centro de gravedad.
- Para F_1 la media de las proyecciones de los individuos en un eje k cualquiera es cero y la varianza de F_k es igual al vector propio correspondiente, λ_k .
- Las componentes principales son ortogonales y no están correlacionadas dos a dos. La característica de ortogonalidad se debe a que son los vectores propios, asociados a los h primeros valores propios, de la misma matriz⁶. Al ser su media cero, la covarianza de F_k y $F_{k'}$ toma la siguiente expresión:

$$\begin{aligned} \text{Cov}(F_k, F_{k'}) &= \sum_{i=1}^n p_i f_{ki} f_{k'i} = F_k' D F_{k'} = u_k' Z' D Z u_{k'} = \\ &= a_k' M Z' D Z M a_{k'} = a_k' S_1 M a_{k'} = 0 \end{aligned}$$

⁶ Véase nota anterior.

puesto que los vectores a_k y a_k' son M -ortogonales.

- Si la matriz R es de rango p , el número de valores propios no nulos es también p , como indicamos anteriormente. En este caso, para una representación completa de las observaciones mediante las nuevas variables (componentes principales), deberíamos tomar tantos ejes como valores propios distintos de cero existen, es decir p . Cuando tomamos h ejes, la parte de inercia de la nube N_1 explicada será:

$$\frac{\sum_{k=1}^h \lambda_k}{\sum_{k=1}^p \lambda_k}$$

y una medida de la desviación entre R^n y R^p es:

$$\sum_{k=k+1}^p \lambda_k$$

Puesto que $Tr(WD) = Tr(V_z M) = Tr(R) = p$, se deduce que la proporción de inercia explicada es:

$$\frac{\sum_{k=1}^h \lambda_k}{p}$$

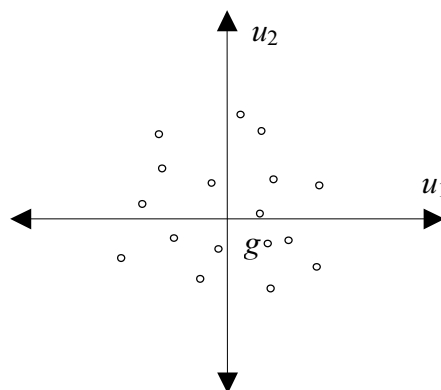
Si el rango de la matriz S_1 fuese $h < p$, toda la inercia estaría explicada por los h ejes y dicha nube estaría contenida en un subespacio de R^h dimensiones. En el caso de que $h > 3$ la representación gráfica de los individuos en los nuevos ejes no tiene una interpretación sencilla. Por ello, en la práctica se suelen tomar entre dos y tres ejes, los primeros, que deben explicar entre el 70% y 80% de la información de la tabla de datos inicial (Judez, 1989: 35). Otro criterio que puede combinarse con el anterior para determinar el número de ejes a retener, es el de considerar los ejes factoriales asociados a las componentes principales cuya varianza sea superior a uno (es decir, $\lambda_k > 1$). Las componentes asociadas a estos ejes explicarán más información que una variable cualquiera de la tabla, en

cuanto que su varianza es superior (la varianza de todas las variables es unitaria).

Cuando se analizan gráficamente las observaciones en el plano formado por los ejes primero y segundo, asociados a las componentes F_1 y F_2 , la representación resultante es la imagen en dos dimensiones de la nube de individuos que mejor mantiene la forma de la nube de puntos original. Si la inercia explicada por los dos primeros ejes fuese el 100% esas distancias en el plano serían las distancias originales entre dichos individuos. De no ser así, es necesario estudiar la calidad de representación de los individuos para saber si las posiciones de los mismos reflejan adecuadamente sus posiciones originales en la nube. Así, los individuos próximos en el plano, siempre y cuando estén bien representados, son los que poseen valores próximos para el conjunto de las variables.

La representación de los individuos en los dos primeros ejes se muestra en la figura siguiente:

Figura II.7
Imagen euclídea en el primer plano factorial de la nube de individuos N_1



A continuación vamos a considerar la manera de reconstituir la tabla de partida a partir de las h componentes principales obtenidas, que son la mejor aproximación de la nube de puntos original.

RECONSTITUCIÓN DE LOS DATOS**RELACIÓN DE LAS COMPONENTES PRINCIPALES CON LAS VARIABLES INICIALES**

La tabla original puede reconstruirse exactamente considerando las p componentes que se pueden obtener según el procedimiento expuesto. Es posible encontrar una aproximación óptima⁷ de los datos de partida, cuando se consideran solo h de las p componentes, a aproximación que podría emplearse con fines predictivos. También es posible establecer una relación entre variables originales y componentes principales a través del coeficiente de correlación entre ambas. A estos aspectos nos vamos a referir a continuación.

Partimos de la matriz F , formada por las p componentes principales en sus columnas:

$$F = \begin{bmatrix} f_{11} & f_{21} & \cdots & f_{h1} & \cdots & f_{p1} \\ f_{12} & f_{22} & \cdots & f_{h2} & \cdots & f_{p2} \\ \vdots & \vdots & & \vdots & & \vdots \\ f_{1n} & f_{2n} & \cdots & f_{hn} & \cdots & f_{pn} \end{bmatrix} = (F_1, F_2 \dots F_p)$$

y sea U la matriz cuyos vectores-columna son los p vectores propios que forman los ejes principales:

$$U = \begin{bmatrix} u_{11} & u_{21} & \cdots & u_{h1} & \cdots & u_{p1} \\ u_{12} & u_{22} & \cdots & u_{h2} & \cdots & u_{p2} \\ \vdots & \vdots & & \vdots & & \vdots \\ u_{1p} & u_{2p} & \cdots & u_{hp} & \cdots & u_{pp} \end{bmatrix} = (u_1, u_2 \dots u_p)$$

Evidentemente se verifica que $F_k = Z u_k$, cuya expresión matricial es:

$$F = Z U$$

Como $u_k' M^{-1} u_k = 1$ se cumple que $U' M^{-1} U = U M^{-1} U' = I_{(p \times p)}$.

⁷ Esta representación es óptima teniendo en cuenta la descomposición de matrices rectangulares de Eckart et Young de 1936 tal y como se menciona en Lebart *et al.* (1995: 23).

Post-multiplicando por $M^{-1} U'$ la expresión anterior ($F = Z U$) se tiene que:

$$FM^{-1}U' = ZUM^{-1}U' \Rightarrow FM^{-1}U' = Z$$

es decir:

$$\begin{bmatrix} f_{11} & f_{21} & \dots & f_{h1} & \dots & f_{p1} \\ f_{12} & f_{22} & \dots & f_{h2} & \dots & f_{p2} \\ \vdots & \vdots & & \vdots & & \vdots \\ f_{1n} & f_{2n} & \dots & f_{hn} & \dots & f_{pn} \end{bmatrix} \begin{bmatrix} m_1^{-1} & 0 & \dots & 0 \\ 0 & m_2^{-1} & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & \dots & 0 & m_p^{-1} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & \dots & u_{1h} & \dots & u_{1p} \\ u_{21} & u_{22} & \dots & u_{2h} & \dots & u_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ u_{p1} & u_{p2} & \dots & u_{ph} & \dots & u_{pp} \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1j} & \dots & z_{1p} \\ z_{21} & z_{22} & \dots & z_{2j} & \dots & z_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ z_{n1} & z_{n2} & \dots & z_{nj} & \dots & z_{np} \end{bmatrix}$$

cuya columna genérica es:

$$Z_j = \begin{bmatrix} z_{1j} \\ z_{2j} \\ \vdots \\ z_{nj} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^p f_{k1} m_k^{-1} u_{kj} \\ \sum_{k=1}^p f_{k2} m_k^{-1} u_{kj} \\ \vdots \\ \sum_{k=1}^p f_{kn} m_k^{-1} u_{kj} \end{bmatrix} = \sum_{k=1}^p F_k m_k^{-1} u_{kj} = (m_k = 1 = m_k^{-1}) = \sum_{k=1}^p F_k a_{kj}$$

Esta última expresión es la que se denomina *fórmula de reconstitución* de los datos, que permite calcular los valores z_{ij} en función de los factores.

Si solo consideramos los h primeros ejes, entonces:

$$Z_j = \sum_{k=1}^h F_k m_k^{-1} u_{kj} = \sum_{k=1}^h F_k a_{kj}$$

de tal forma que dicha aproximación es buena si $\lambda_1 + \lambda_2 + \dots + \lambda_h$ es una proporción importante de la inercia total. Además, la suma de esos h primeros términos constituye la mejor aproximación de la matriz Z por una matriz de rango h en el sentido de los mínimos cuadrados, según el Teorema de Eckart-Young (Saporta, 1990)⁸.

Se deduce, por tanto, que las variables originales son combinación lineal de las componentes principales. A partir de esta expresión se puede deducir el

⁸ Véase nota anterior.

coeficiente de correlación⁹ entre las variables originales y las componentes principales, que permite estudiar la relación entre ambas:

$$\rho(F_k, Z_j) = \frac{\text{Cov}(F_k, Z_j)}{\sqrt{\text{Var}(F_k)}\sqrt{\text{Var}(Z_j)}}$$

Demostrado que $\text{Var}(F_k) = \lambda_k$ y $\bar{F}_k = 0$; $\text{Var}(Z_j) = 1$ y $\bar{Z}_j = 0$, la covarianza de Z_j y F_k es el producto escalar entre esos dos puntos de R^p :

$$\text{Cov}(F_k, Z_j) = F_k' D Z_j = F_k' D \sum_{k=1}^p F_k m_k^{-1} u_{kj} \Rightarrow Z_j = \sum_{k=1}^p F_k m_k^{-1} u_{kj}$$

Al desarrollar la expresión de la covarianza, tendremos:

$$\begin{aligned} \text{Cov}(F_k, Z_j) &= F_k' D \sum_{k=1}^p F_k m_k^{-1} u_{kj} = F_k' D F_1 m_1^{-1} u_{1j} + F_k' D F_2 m_2^{-1} u_{2j} + \dots + \\ &F_k' D F_k m_k^{-1} u_{kj} + \dots + F_k' D F_p m_p^{-1} u_{pj} \end{aligned}$$

Como los F_k son incorrelacionados entre sí el único sumando distinto de cero es: $F_k' D F_k m_k^{-1} u_{kj}$, por lo que:

$$F_k' D \sum_{k=1}^p F_k m_k^{-1} u_{kj} = F_k' D F_k m_k^{-1} u_{kj} = \text{Var}(F_k) m_k^{-1} u_{kj} = \lambda_k m_k^{-1} u_{kj}$$

Por tanto,

$$\rho(F_k, Z_j) = \frac{\text{Cov}(F_k, Z_j)}{\sqrt{\text{Var}(F_k)}\sqrt{\text{Var}(Z_j)}} = \frac{\lambda_k m_k^{-1} u_{kj}}{\sqrt{\lambda_k}} = \sqrt{\lambda_k} m_k^{-1} u_{kj} = \sqrt{\lambda_k} a_{kj}$$

Es posible preguntarse si una misma variable Z_j puede tener un coeficiente de correlación próximo a uno, o menos uno, con más de un factor. Como se va a demostrar, la respuesta es no.

Considerando que:

$$Z_j = \sum_{k=1}^p F_k m_k^{-1} u_{kj} \quad \text{y que} \quad \text{Var}(Z_j) = Z_j' D Z_j = 1:$$

⁹ Algunos autores denominan cargas factoriales a estos valores (Stevens, 1986: 339).

$$\begin{aligned}
 1 &= \text{Var}(Z_j) = Z_j' D Z_j = Z_j' D \sum_{k=1}^p F_k m_k^{-1} u_{kj} = \\
 &= Z_j' D F_1 m_1^{-1} u_{1j} + Z_j' D F_2 m_2^{-1} u_{2j} + \dots + Z_j' D F_p m_p^{-1} u_{pj} = \\
 &= \lambda_1 m_1^{-1} u_{1j} m_1^{-1} u_{1j} + \lambda_2 m_2^{-1} u_{2j} m_2^{-1} u_{2j} + \dots + \lambda_p m_p^{-1} u_{pj} m_p^{-1} u_{pj}
 \end{aligned}$$

ya que $Z_j' D F_k = \lambda_k m_k^{-1} u_{kj}$ es la covarianza entre Z_j y F_k :

$$1 = \text{Var}(Z_j) = Z_j' D Z_j = 1 = \lambda_1 u_{1j}^2 + \lambda_2 u_{2j}^2 + \dots + \lambda_p u_{pj}^2$$

por ser: $m_k^{-1} = 1 \forall k$. Se verifica, por tanto, que:

$$1 = \rho^2(F_1, Z_j) + \rho^2(F_2, Z_j) + \dots + \rho^2(F_p, Z_j)$$

lo que significa que si una variable Z_j tiene un coeficiente de correlación con un factor próximo a uno o a menos uno, sus coeficientes de correlación con los otros factores estarán próximos a cero.

Además, la suma de los coeficiente de correlación, al cuadrado, de todas las variables con un factor es la inercia del eje correspondiente, tal y como se expresa en la siguiente expresión:

$$\sum_{j=1}^p \rho^2(F_k, Z_j) = \sum_{j=1}^p \lambda_k u_{kj}^2 = \lambda_k$$

Para el factor F_1 se cumplirá que:

$$\sum_{j=1}^p \rho^2(F_1, Z_j) = \lambda_1$$

motivo por el cual, F_1 , asociada al mayor valor propio λ_1 , es la variable más relacionada posible a las $Z_1, Z_2 \dots Z_p$ en el sentido de que la suma de los cuadrados de las correlaciones es máxima.

INDIVIDUOS SUPLEMENTARIOS

En algunas ocasiones es útil conocer la posición, sobre los nuevos ejes obtenidos, de ciertos individuos que, aunque no han intervenido en la

determinación de los ejes, tienen especial interés y facilitan la interpretación de los planos factoriales. Esto se consigue otorgándoles un peso nulo en el proceso de obtención de los ejes. Puede ser el caso de individuos que presentan caracteres especiales, o bien individuos que se sospecha están afectados por errores de medida. Estos individuos se denominan *individuos suplementarios* en oposición a los considerados inicialmente, denominados *individuos activos*.

Para situar los individuos suplementarios es necesario posicionarlos respecto al centro de gravedad de la nube, por lo que hay que dividir sus coordenadas por las desviaciones típicas de las variables y restarles las medias respectivas. Considerando como individuo suplementario:

$$e_{i+}' = (z_{i+1}, z_{i+2} \dots z_{i+p}) \quad ,, \quad z_{i+j} = \frac{x_{i+j} - \bar{x}_j}{s_j}$$

Sus coordenadas en el eje k , que se toman de la expresión $f_{i+k} = e_{i+}' u_k$, permiten su representación en los ejes obtenidos e interpretar su posición.

b. Análisis de la nube de variables N_J

Los resultados obtenidos pueden, en muchas situaciones, bastar para la aplicación del ACP *normado*, si bien en este apartado se realiza una nueva aproximación a dicho método para incidir en el estudio de las relaciones entre variables, estudiando la nube que éstas forman. Como se trata de un proceso similar al visto para los individuos, se van a considerar las h componentes a la vez, obteniendo los ejes de máxima inercia y las coordenadas de las variables en esos nuevos ejes. La representación de las variables, cuando están correlacionadas dos a dos (efecto talla), y el tratamiento de elementos suplementarios, serán considerados también en este caso.

LAS h COMPONENTES O FACTORES PRINCIPALES

Ya hemos indicado que la distancia entre las variables depende del coeficiente de correlación entre las mismas y que al estudiar dicho espacio el objetivo son las direcciones determinadas por las variables y por tanto, su tratamiento como vectores.

Por otra parte, la nube de variables no está centrada y la cantidad de información de las mismas se mide respecto al origen. Dicha inercia tomaba la siguiente expresión:

$$I = \sum_{j=1}^p m_j d^2(Z_j, 0) = \sum_{j=1}^p m_j \sum_{i=1}^n p_i z_{ij}^2 = \sum_{j=1}^p \sum_{i=1}^n m_j p_i z_{ij}^2 = \text{Tr}(V_z M) = \text{Tr}(M V_z)$$

siendo V_z la matriz de varianzas-covarianzas de las matrices Z_j y Z_j' , que coincide con R , matriz de correlaciones entre las variables originales X_j y X_j' .

En este caso, se trata de buscar un subespacio que conserve la mayor parte de información, o lo que es lo mismo, que represente de la mejor manera posible la *esfera de correlaciones*.

— Denominamos por t_1 el vector director de una recta que, pasando por el origen, maximiza la inercia de la nube N_J proyectada sobre esa recta. Considerando la métrica D del espacio R^n , la proyección de la variable Z_j sobre dicha recta es g_{1j} cuyo valor es:

$$g_{1j} = Z_j' D t_1 = Z_j' v_1 = \sum_{i=1}^n z_{ij} p_i t_{1i} = \sum_{i=1}^n z_{ij} v_{1i}$$

siendo:

$$t_1 = \begin{bmatrix} t_{11} \\ t_{12} \\ \vdots \\ t_{1n} \end{bmatrix} \text{ y } \begin{bmatrix} v_{11} \\ v_{12} \\ \vdots \\ v_{1n} \end{bmatrix} = D t_1 = v_1$$

el primer eje factorial de N_J .

El vector t_1 debe ser D -ortonormal, es decir: $t_1' D t_1 = 1$, por lo que v_1 es D^{-1} -ortonormal y se cumple que: $v_1' D^{-1} v_1 = 1$. El vector que recoge las proyecciones de las p variables en el primer eje se denominará G_1 y toma la siguiente expresión:

$$Z' D t_1 = \begin{bmatrix} g_{11} \\ g_{12} \\ \vdots \\ g_{1p} \end{bmatrix} = G_1$$

Se trata de un vector que pertenece a R^p (espacio de los individuos).

— La cantidad de información del primer eje, respecto al origen, viene dada por la siguiente expresión:

$$\sum_{j=1}^p m_j d^2(G_1, 0) = \sum_{j=1}^p m_j g_{1j}^2 = G_1' M G_1$$

Esta expresión es la que debe hacerse máxima y por ello el primer eje es el que hace máxima $G_1' M G_1 = t_1' D Z M Z' D t_1$ teniendo en cuenta que $D = D'$ por ser una matriz diagonal.

Si denominamos por S_J a la matriz de inercia en el caso de las variables tal que: $S_J = D Z M Z'$ se tiene que:

$$G_1' M G_1 = t_1' D Z M Z' D t_1 = t_1' S_J D t_1 = v_1' D^{-1} S_J v_1$$

Cuando se consideran h vectores correspondientes a otros tantos h ejes v_k ($k=1, 2 \dots h$) la expresión a maximizar es:

$$\sum_{k=1}^h v_k' D^{-1} S_J v_k$$

— El problema de optimización se plantea en los siguientes términos:

$$\left\{ \begin{array}{l} \text{Max : } \sum_{k=1}^h v_k' D^{-1} S_J v_k \\ \text{sujeto a : } v_k' D^{-1} v_k = 1 \text{ para } k = 1, 2 \dots h \\ \quad \quad \quad v_k' D^{-1} v_k = 0 \text{ para } k \neq k' \text{ (condición de ortogonalidad)} \end{array} \right.$$

Como en el caso de los individuos, la condición de ortogonalidad es redundante por el motivo ya indicado al analizar N_I . La función de Lagrange resultante es:

$$L = \sum_{k=1}^h v_k' D^{-1} S_J v_k - \lambda_k^* (v_k' D^{-1} v_k - 1)$$

Derivando respecto v_k :

$$\frac{\partial L}{\partial v_k} = \begin{bmatrix} \frac{\partial L}{\partial v_{k1}} \\ \frac{\partial L}{\partial v_{k2}} \\ \vdots \\ \frac{\partial L}{\partial v_{kp}} \end{bmatrix} = 2 \sum_{k=1}^h D^{-1} S_J v_k - 2 \sum_{k=1}^h \lambda_k^* D^{-1} v_k = 0$$

$$\Rightarrow \sum_{k=1}^h D^{-1} S_J v_k = \sum_{k=1}^h \lambda_k^* D^{-1} v_k \Rightarrow D Z M Z' \sum_{k=1}^h v_k = \sum_{k=1}^h \lambda_k^* v_k$$

$$\Rightarrow \sum_{k=1}^h S_J v_k = \sum_{k=1}^h \lambda_k^* v_k \Rightarrow S_J \sum_{k=1}^h v_k = \sum_{k=1}^h \lambda_k^* v_k$$

es decir, los h ejes son los vectores propios asociados a los h primeros valores propios de la matriz $S_J = D Z M Z' = DW$.

Al ser $v_k = D t_k$ tenemos que:

$$D Z M Z' \sum_{k=1}^h D t_k = \sum_{k=1}^h \lambda_k^* D t_k$$

y premultiplicando por D^{-1} esa expresión se tiene que:

$$D Z M Z' \sum_{k=1}^h D^{-1} D t_k = \sum_{k=1}^h \lambda_k^* D^{-1} D t_k \Rightarrow D Z M Z' \sum_{k=1}^h t_k = \sum_{k=1}^h \lambda_k^* t_k$$

$$\Rightarrow W D \sum_{k=1}^h t_k = \sum_{k=1}^h \lambda_k^* t_k$$

Se deduce, entonces, que los vectores propios de la matriz WD son los vectores directores de los ejes de mayor inercia, y como $WD = DW$, por ser D una matriz diagonal, t_k y v_k son los vectores propios de una misma matriz.

Se cumple, asimismo, que los vectores asociados a los valores propios distintos son todos ortogonales como ya habíamos indicado al analizar N_1 .

— De estos resultados se obtienen las siguientes conclusiones:

- Si consideramos el eje k , la inercia respecto al origen de dicho eje es:

$$G_k' M G_k = t_k' S_J D t_k = v_k' D^{-1} S_J v_k = v_k' D^{-1} \lambda_k^* v_k = \lambda_k^*$$

Considerando h ejes, tenemos que:

$$\sum_{k=1}^h v_k' D^{-1} S_J v_k = \sum_{k=1}^h v_k' D^{-1} \lambda_k^* v_k = \sum_{k=1}^h \lambda_k^*$$

- Dado que: $G_k = Z' D t_k = Z' v_k$, premultiplicando por Z' la expresión:

$$D Z M Z' \sum_{k=1}^h v_k = \sum_{k=1}^h \lambda_k^* v_k$$

tenemos que:

$$\begin{aligned} Z' D Z M Z' \sum_{k=1}^h v_k &= \sum_{k=1}^h \lambda_k^* Z' v_k \Rightarrow Z' D Z M \sum_{k=1}^h Z' v_k = \sum_{k=1}^h \lambda_k^* Z' v_k \\ &\Rightarrow \text{RM} \sum_{k=1}^h G_k = \sum_{k=1}^h \lambda_k^* G_k \end{aligned}$$

de donde se deduce que las G_k son los vectores propios, asociados a los h valores propios de la matriz RM .

Como la traza de esta matriz es la inercia total y es una matriz cuadrada se tiene que:

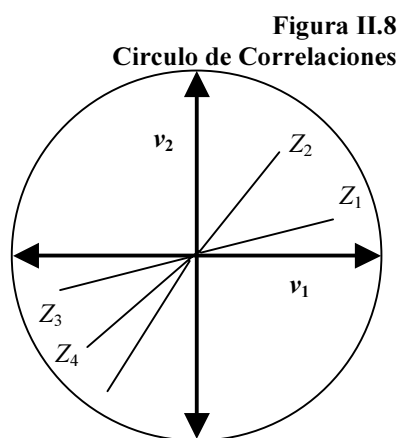
$$\text{Tr}(\text{RM}) = \sum_{k=1}^h \lambda_k^* = p$$

- La proporción de inercia, respecto al origen, explicada por los h ejes de la nube de variables será, por tanto:

$$\frac{\sum_{k=1}^h \lambda_k^*}{p}$$

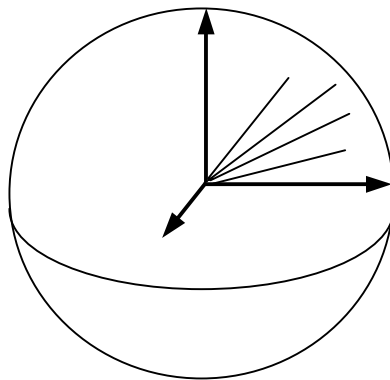
- Siguiendo un procedimiento similar al de los individuos, para obtener una representación de la nube de variables en un espacio de menor dimensión, a menudo se suele tomar un número de ejes (entre dos y tres generalmente) que explique entre el 70 y el 80% del total de la tabla. La representación obtenida en el primer plano factorial (ejes uno y dos) se denomina *círculo de correlaciones*, debido a que todas las variables están a una distancia uno del origen y sus proyecciones sobre los planos factoriales estarán en el interior de dicho círculo de radio uno y centro el origen. Las variables que se encuentren realmente sobre el plano estarán situadas en la periferia del círculo y los coeficientes de correlación entre estas variables podrán visualizarse teniendo en cuenta que la distancia entre las mismas está en función de dicho coeficiente. Las variables que estén mal representadas en el plano se encontrarán próximas al origen y las distancias con las otras variables, y por tanto su correlación, no tiene una interpretación adecuada. Como se verá posteriormente, las ayudas a la interpretación permiten extraer una serie de coeficientes para evaluar la calidad de representación de dichas variables.

La representación obtenida es la síntesis gráfica de la matriz de correlaciones y se recoge en la figura siguiente:



- Cuando todas las variables X_j están correlacionadas positivamente entre ellas dos a dos formando ángulos agudos, el centro de gravedad de la nube de variables queda lejos del origen. En este caso, algunos individuos tienen valores pequeños para el conjunto de variables, otros tienen valores grandes para el conjunto de las mismas existiendo, por último, otros que ocupan posiciones intermedias entre estos dos extremos. Esta estructura común al conjunto de variables es lo que refleja la primera componente principal y el primer eje factorial aparece entonces próximo a la dirección que une el origen con el centro de gravedad. La siguiente representación gráfica muestra dicha situación:

Figura II.9
Efecto talla



En este caso, todas las variables tienen la primera componente positiva. Este eje representa mal la forma de la nube N_j puesto que está indicando sobre todo la posición de la nube de variables respecto al origen y todas las proyecciones de las variables estarán próximas entre sí. Esta primera componente principal define un "factor de talla" o de tamaño. En este caso para que la segunda componente fuese un "factor de forma" las coordenadas de las variables sobre las mismas deberían ser unas positivas y otras negativas (Cuadras, 1980: 277).

VARIABLES SUPLEMENTARIAS

En algunas ocasiones interesa estudiar ciertas variables que no intervienen en la construcción de los factores pero cuya posición respecto a los mismos puede ser interesante para completar la interpretación de los mismos. A este tipo de variables se les denomina *variables suplementarias*, cuyo concepto es similar al de individuo suplementario.

Así, al considerar la variable suplementaria:

$$Z_{j+} = \frac{x_{ij+} - \bar{x}_{j+}}{s_{j+}}$$

La coordenada de dicha variable con el eje k toma la expresión: $g_{kj+} = Z_{j+}' v_k$ y coincide con su coeficiente de correlación con el factor k , lo que permite matizar la interpretación de dicho factor.

En el cuadro siguiente se esquematizan los desarrollos realizados sobre la nube de individuos y variables en un ACP *normado*:

Tabla II.1
Las dos nubes, ejes y factores respectivos

	Nube N_I	Nube N_J
<i>Espacio</i>	R^p	R^n
<i>Métrica</i>	M	D
<i>Coordenadas</i>	$e_i' \rightarrow Z$	$z_j \rightarrow Z$
<i>Peso</i>	D	M
<i>Vectores unitarios</i>	a_k	t_k
<i>Ejes Factoriales (nuevos ejes)</i>	u_k	v_k
<i>Ecuación</i>	$Z'DZMa_k = \lambda_k a_k \Rightarrow VMa_k = \lambda_k a_k$	$ZMZ'Dt_k = \lambda_k t_k \Rightarrow WDt_k = \lambda_k t_k$
<i>Norma</i>	$\ a_k\ _M = 1$ y $\ u_k\ _{M^{-1}} = 1$	$\ t_k\ _D = 1$ y $\ v_k\ _{D^{-1}} = 1$
<i>Factores</i>	$F_k = Z u_k$	$G_k = Z' v_k$
<i>Ecuaciones</i>	$WDF_k = \lambda_k F_k$	$V_z M G_k = \lambda_k G_k$
<i>Norma</i>	$\ F_k\ _D = \sqrt{\lambda_k}$	$\ G_k\ _M = \sqrt{\lambda_k}$
<i>Inercia del eje</i>	λ_k	λ_k

c. Relaciones de transición

Representación simultánea de individuos y variables

La nube de individuos y la nube de variables son dos representaciones de la misma tabla, según se tome ésta como una yuxtaposición de filas o de columnas, respectivamente. En este apartado vamos a profundizar precisamente en las relaciones entre ambas nubes a través de los dos aspectos siguientes: relaciones de transición y representación simultánea.

Así, las relaciones de transición van a permitir, precisamente, expresar los resultados de un análisis en función de los resultados del otro, razón por la cual, a veces sólo se lleva a cabo el análisis de una de las nubes. Además y con las consideraciones oportunas, es posible representar en un mismo plano individuos y variables con el fin de favorecer la interpretación de cada uno de los espacios.

RELACIONES DE TRANSICIÓN

En cuanto a las relaciones entre los dos espacios se pueden hacer las siguientes consideraciones:

— Del análisis de la nube de individuos se deduce que:

$$\text{WD} \sum_{k=1}^h F_k = \sum_{k=1}^h \lambda_k F_k$$

y del análisis de la nube de variables se obtiene:

$$\text{WD} \sum_{k=1}^h t_k = \sum_{k=1}^h \lambda_k^* t_k$$

Como los valores propios son únicos, se tiene que los $\lambda_k = \lambda_k^*$ y a partir de aquí los denominaremos λ_k para ambas nubes.

— Considerando el valor propio λ_k y las expresiones $W D F_k = \lambda_k F_k$ y $W D t_k = \lambda_k t_k$, los vectores F_k y t_k deben ser colineales, por lo que, en virtud de la igualdad $v_k = D t_k$, se cumple: $t_k = c F_k \Rightarrow v_k = c D Z u_k$.

Además sabemos que: $v_k' D^{-1} v_k = 1$, por lo que sustituyendo en la anterior expresión el valor de v_k , se tiene que:

$$\begin{aligned} c^2 u_k' Z' D D^{-1} D Z u_k = 1 &\Rightarrow c^2 u_k' Z' D Z u_k = 1 \Rightarrow \\ &\Rightarrow c^2 F_k' D F_k = 1 \Rightarrow c^2 \lambda_k = 1 \Rightarrow c = \frac{1}{\sqrt{\lambda_k}} \end{aligned}$$

Luego tenemos que:

$$t_k = \frac{1}{\sqrt{\lambda_k}} F_k \Rightarrow F_k = \sqrt{\lambda_k} t_k \Rightarrow f_{ki} = \sqrt{\lambda_k} t_{ki}$$

Teniendo en cuenta que: $v_k = D t_k$, se deduce que:

$$v_k = \frac{1}{\sqrt{\lambda_k}} D Z u_k \Rightarrow v_k = \frac{1}{\sqrt{\lambda_k}} D F_k \Rightarrow \sqrt{\lambda_k} v_k = D F_k \Rightarrow D^{-1} \sqrt{\lambda_k} v_k = F_k$$

por lo que el eje principal de la nube de variables se puede obtener a partir de las componentes principales de la nube de individuos. Esta es la razón por la que a veces sólo se realiza el análisis de esta última nube.

— Por otra parte, las coordenadas de las variables en el eje k se recogen en el vector $G_k = Z' v_k$. Sustituyendo v_k se tiene que:

$$G_k = \frac{1}{\sqrt{\lambda_k}} Z' D F_k$$

es decir, las coordenadas de las variables son combinaciones lineales de las componentes principales de los individuos.

— Puesto que con el análisis de la nube de individuos se cumple que:

$$RM \sum_{k=1}^h a_k = \sum_{k=1}^h \lambda_k a_k$$

y como consecuencia del análisis de la nube de variables tenemos que:

$$RM \sum_{k=1}^h G_k = \sum_{k=1}^h \lambda_k G_k$$

tomando un solo factor G_k se deduce que: $a_k = q G_k \Rightarrow u_k = q M Z' v_k$,
 dado que: $u_k' M^{-1} u_k = 1$.

Sustituyendo u_k por su valor:

$$q^2 v_k' Z M M^{-1} M Z' v_k = 1 \Rightarrow q^2 v_k' Z M Z' v_k = 1 \Rightarrow q^2 G_k' M G_k = 1$$

Por tanto:

$$q^2 \lambda_k = 1 \Rightarrow q = \frac{1}{\sqrt{\lambda_k}} \Rightarrow a_k = \frac{1}{\sqrt{\lambda_k}} G_k \Rightarrow G_k = \sqrt{\lambda_k} a_k$$

Tenemos que: $u_k = M a_k$, por lo que se cumple que:

$$u_k = \frac{1}{\sqrt{\lambda_k}} M Z' v_k \Rightarrow u_k = \frac{1}{\sqrt{\lambda_k}} M G_k \Rightarrow G_k = \sqrt{\lambda_k} M^{-1} u_k$$

Además, la proyección de un individuo sobre un eje es una combinación lineal de las proyecciones de todas las variables sobre dicho eje, es decir:

$$F_k = Z u_k = \frac{1}{\sqrt{\lambda_k}} Z M Z' v_k = \frac{1}{\sqrt{\lambda_k}} Z M G_k$$

Teniendo en cuenta que: $M = I = M^{-1}$, se concluye que:

$$G_k = \sqrt{\lambda_k} u_k \Rightarrow g_{kj} = \sqrt{\lambda_k} u_{kj} = \rho(F_k, Z_j)$$

De esta expresión se deduce que, para ACP *normado*, el coeficiente de correlación de una variable con una componente principal es la coordenada de la variable con el eje correspondiente, conclusión que constituye la base para proceder a la interpretación de las componentes F_k . En efecto, teniendo en cuenta que una variable consigo misma tiene un coeficiente de correlación igual a uno y con su opuesta igual a menos uno, la variable F_k se interpretará a partir de las variables Z_j cuyo $\rho(F_k, Z_j)$ sea próximo a uno o menos uno.

Teniendo en cuenta que:

$$G_k = \frac{1}{\sqrt{\lambda_k}} Z' D F_k$$

la proyección de una variable (Z_j) sobre otra se interpreta (F_k) como un mutuo coeficiente de correlación (G_k), y los factores principales (F_k) son las variables sintéticas más relacionadas con el conjunto de variables iniciales.

Para la primera componente, puesto que:

$$\sum_{j=1}^p \rho^2(F_k, Z_j) = \lambda_k$$

según se demostró, se tiene que:

$$\sum_{j=1}^p \rho^2(F_1, Z_j) = \lambda_1 = \sum_{j=1}^p g_{1j}^2$$

La suma de las correlaciones al cuadrado de las variables con la primera componente es la suma de las coordenadas al cuadrado de las variables con el primer eje, y esa expresión es máxima puesto que λ_1 es el mayor valor propio y la varianza de dicha componente.

— A la vista de los anteriores resultados, la fórmula de reconstitución de cada variable:

$$Z_j = \sum_{k=1}^p F_k m_k^{-1} u_{kj}$$

puede expresarse como:

$$\sum_{k=1}^p D^{-1} \sqrt{\lambda_k} v_k m_k^{-1} u_{kj}$$

Sustituyendo F_k por su valor respecto v_k tenemos que:

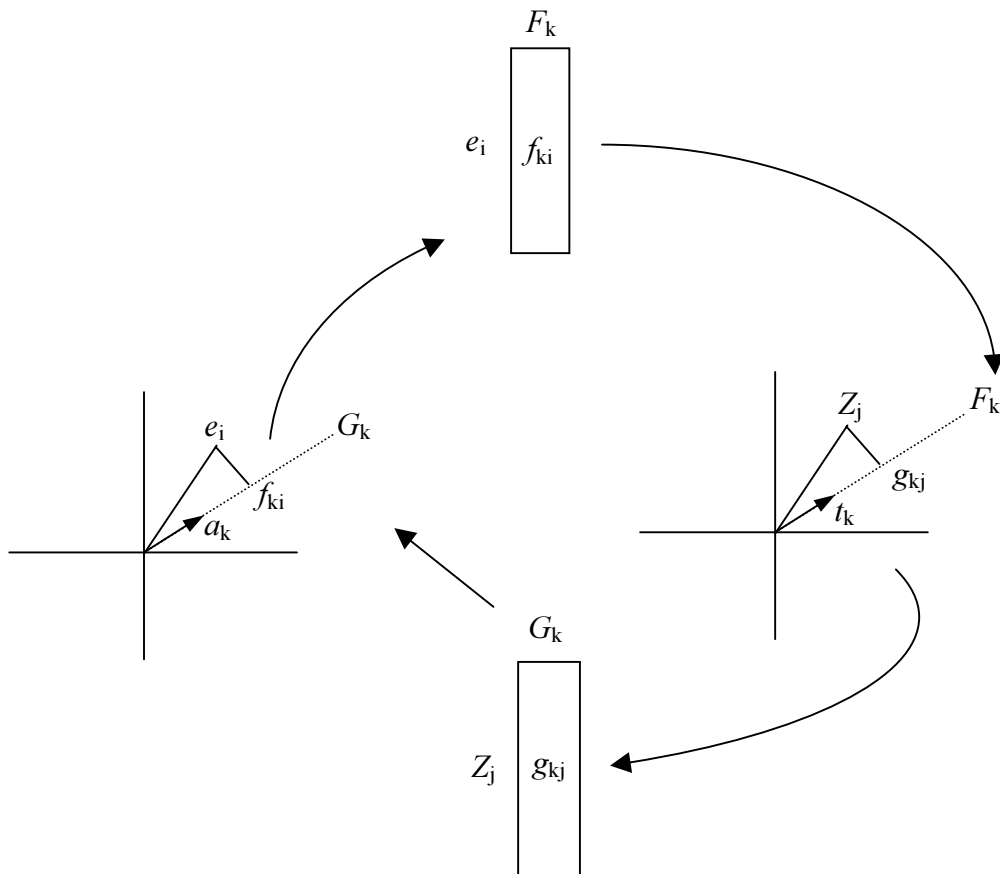
$$Z_j = \sum_{k=1}^p n \sqrt{\lambda_k} v_k u_{kj}$$

Expresando respecto t_k la misma expresión se tiene que:

$$Z_j = \sum_{k=1}^p \sqrt{\lambda_k} t_k a_{kj}$$

Gráficamente, se puede representar las relaciones de transición entre las dos nubes tal y como recoge la siguiente figura:

Figura II.10
Relación¹ entre los dos análisis en el eje k



Según dicho esquema, el conjunto de proyecciones de todos los puntos de la nube de individuos sobre a_k (que es F_k) es una nueva variable que se confunde con G_k , siendo ésta la proyección de la nube de variables sobre el eje k correspondiente. Asimismo, la proyección de las p variables sobre t_k (que es G_k) define unos valores que pueden considerarse como un nuevo individuo. Este individuo caracteriza el factor y se puede considerar como un individuo-tipo que está situado sobre el eje k de la nube de individuos. En el caso del ACP *normado*, la proyección de Z_j sobre F_k es su coeficiente de correlación.

¹ Escofier y Pagès (1992: 81).

REPRESENTACIÓN SIMULTÁNEA DE INDIVIDUOS Y VARIABLES

Esta representación se considera interesante en la medida que las variables ayudan a la interpretación de los individuos y viceversa. Ya se ha señalado que en la imagen de la nube de individuos las distancias entre ellos se interpretan en términos de similitud de comportamiento respecto al total de variables estudiadas, mientras que en la nube de variables las proximidades entre ellas se interpretan en términos de correlación. Por esta razón, en una representación simultánea no es posible interpretar directamente la distancia entre un punto-individuo y un punto-variable ya que dichas representaciones pertenecen a espacios y puntos de referencia diferentes. Además la representación simultánea de individuos y variables no puede hacerse superponiendo las representaciones gráficas obtenidas para unos y otras.

Tomando como referencia las relaciones de transición, que permiten relacionar los dos espacios y los resultados obtenidos para ambos, es posible plantearse una representación simultánea de individuos y variables, pero en otros términos. Si se consideran las variables no como puntos sino como vectores, es posible plantearse su representación simultánea junto con los individuos. En dicho espacio, después de una transformación de la tabla de datos, se dispone de dos sistemas de ejes unitarios:

- El sistema de referencia de la tabla original: $r_1, r_2 \dots r_p$ que corresponden a las p variables antes del análisis y que son los ejes de referencia de las coordenadas iniciales de los individuos, es decir:

$$e_i = \sum_{j=1}^p Z_{ij} r_j$$

- Los nuevos ejes factoriales $u_1, u_2 \dots u_p$ que se han obtenido al analizar la nube de individuos.

La proyección, como individuos suplementarios, de los ejes r_j originales sobre los nuevos ejes u_k permite la representación simultánea. La coordenada

de una variable viene dada por la proyección del r_j sobre u_k y será $r_j \cdot u_k = u_{kj}$. Así, es posible representar las direcciones de los datos para las variables originales sobre el plano de la nube de individuos, que se han expresado a través de los vectores unitarios. La representación obtenida es distinta de la representación de la nube de variables antes comentada. Por ello, la denominada representación simultánea es, en realidad, un aplastamiento de la referencia ortonormada de los ejes originales del espacio de los individuos sobre el nuevo plano factorial. Se dispone así de una perspectiva deformada del sistema de ejes original, a la vez que se tiene en cuenta las relaciones existentes entre las variables originales.

Considerando que en R^n la coordenada de la variable Y_j respecto al eje k es el coeficiente de correlación con F_k igual a $g_{kj} = \sqrt{\lambda_k} u_{kj}$, las dos nubes de variables no son coincidentes, siendo la diferencia entre una y otra $\sqrt{\lambda_k}$. Además, en la representación simultánea, que es una representación en R^p , no se interpreta la distancia entre dos variables en términos de correlación, pues los puntos representados son en realidad los extremos de dos vectores unitarios ortogonales (la distancia es igual a $\sqrt{2}$) en el espacio completo. La interpretación de las distancias entre variables (en términos de correlación) sólo se puede hacer en R^n y no en la representación simultánea. De todos modos, ambas dan una imagen similar, sobre todo si los valores propios son pequeños. En la representación simultánea no es posible tener variables cuantitativas suplementarias puesto que lo que se ha representado son los ejes del espacio de los individuos.

Para obtener una representación equilibrada de las dos nubes en el mismo gráfico es necesario tener en cuenta que las coordenadas de los individuos deben corregirse para que su posición en el plano sea compatible con el reparto de los puntos-variables y obtener una representación equilibrada de las dos nubes. La dispersión de ambas nubes debe ser la misma y por tanto el momento de orden dos, respecto al origen, debe valer lo mismo. Para la nube de individuos, dicho momento, para el eje k , vale:

$$\alpha_2(F_k) = \sum_{i=1}^n \frac{1}{n} f_{ki}^2 = F_k' D F_k = \text{Var}(F_k) = \lambda_k$$

Para la nube de variables y para el eje k tenemos que:

$$\alpha_2(G_k) = \sum_{j=1}^p \frac{1}{p} g_{kj}^2 = \frac{1}{p} G_k' G_k = \frac{1}{p} G_k' M G_k = \frac{\lambda_k}{p}$$

considerando que $M = I$.

Por tanto, para la representación simultánea, considerando:

$$\hat{f}_{ki} = \sqrt{1/p} f_{ki}$$

se cumple que:

$$\alpha_2(\bar{F}_\alpha) = \sum_{i=1}^n \frac{1}{n} \hat{f}_{ki}^2 = \frac{1}{p} F_k' D F_k = \frac{\lambda_k}{p} = \alpha_2(G_k)$$

y la dispersión respecto al origen es la misma en ambos casos.

1.3.3 Interpretación del ACP *normado*

Uno de los aspectos fundamentales de una técnica de Análisis de Datos es la interpretación de los resultados que se obtienen del desarrollo teórico de la misma. En este apartado nos ocuparemos del procedimiento que se debe seguir para una correcta interpretación de los resultados obtenidos en el ACP *normado*. En primer lugar, es aconsejable considerar una serie de indicadores numéricos que nos van a facilitar dicha tarea. Posteriormente se tendrán en cuenta los factores respecto a su inercia y su relación con los individuos y las variables. La representación simultánea de ambos enriquece la interpretación, siempre y cuando se tengan en cuenta sus limitaciones.

a. Ayudas a la interpretación

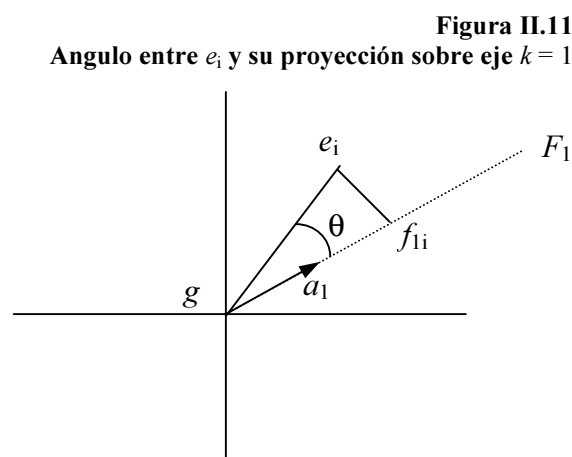
En este apartado se van a considerar una serie de indicadores numéricos que pretenden facilitar la tarea de interpretación de los resultados obtenidos en el análisis de las dos nubes. Ya hemos señalado que los ejes forman una imagen aproximada de una nube de puntos en cuanto que los mismos no explican el 100% de la inercia total de la nube. La calidad de representación total de las nubes por un eje ya se ha considerado precisamente al obtener la inercia del mismo. Ahora es necesario medir la calidad de la aproximación para cada punto, individuo o variable, por separado.

INDIVIDUOS

— La *calidad de representación* de un individuo e_i por un eje k viene dada por:

$$\text{Calidad}(e_i) = \frac{\text{Inercia de la proyección del individuo sobre } k}{\text{Inercia total de } e_i \text{ (respecto de todos los ejes)}}$$

Esta expresión no es más que el $\cos^2\theta$, siendo θ el ángulo formado entre la coordenada del individuo original y el eje k :



Por tanto la anterior expresión tiene la siguiente forma:

$$\frac{d^2(f_{ki}, 0)}{d^2(e_i, 0)} = \frac{f_{ki}^2}{\sum_{k=1}^p f_{ki}^2}$$

Si generalizamos esta definición para un plano, y dado que los ejes son ortogonales, la calidad de representación del individuo e_i en el plano de los ejes k y k' , es la suma de las calidades de representación de e_i para los dos ejes de manera separada, y además coincide con el coseno cuadrado del ángulo entre el individuo e_i y el plano de proyección:

$$\frac{d^2(f_{ki}, 0)}{d^2(e_i, 0)} + \frac{d^2(f_{k'i}, 0)}{d^2(e_i, 0)} = \frac{f_{ki}^2 + f_{k'i}^2}{\sum_{k=1}^p f_{ki}^2}$$

Si la calidad de un punto toma un valor cercano a uno ese punto está próximo al plano y por ello su distancia al centro de gravedad está bien representada en ese plano y su posición puede interpretarse correctamente. Si ese cociente es cero dicho individuo no está correctamente representado. Por otra parte, dos individuos cuya imagen está próxima en el plano solo pueden interpretarse como semejantes si la calidad de representación de ambos es buena. La calidad de representación de un punto también se denomina *contribución relativa*.

— La *contribución* de un individuo a la inercia de un eje k viene dada por la siguiente expresión:

$$\begin{aligned} \text{Contribución}(e_i) &= \frac{\text{Inercia de la proyección del individuo sobre } k}{\text{Inercia de la proyección de todos los individuos sobre } k} = \\ &= \frac{\text{Inercia de la proyección del individuo sobre } k}{\text{Var}(F_k)} = \frac{p_i f_{ki}^2}{\lambda_k} \end{aligned}$$

dado que:

$$\text{Var}(F_k) = \lambda_k = \sum_{i=1}^n p_i f_{ki}^2 \quad \text{para todos los individuos}$$

para el individuo e_i toma el valor $p_i f_{ki}^2$.

En este caso se trata de estudiar la inercia del eje k que es atribuida al individuo e_i . Esta relación se precisa para poner en evidencia los elementos que han contribuido principalmente a la construcción del eje y sobre la que se apoyará en primer lugar la interpretación. También se le denomina *contribución absoluta*.

VARIABLES

— La *calidad de representación* de una variable para el eje k viene dada por:

$$\text{Calidad } (Z_j) = \frac{\text{Inercia de la proyección de la variable sobre "k"}}{\text{Inercia total de } X_j} =$$

$$\frac{d^2(g_{kj}, 0)}{d^2(Z_j, 0)} = g_{kj}^2 = \lambda_k u_{kj}^2 = \rho^2(Z_j, F_k)$$

puesto que $d^2(Z_j, 0) = 1$ para todas las variables.

Las variables con ese valor próximo a uno estarán bien representadas en el plano correspondiente y se situarán en la periferia del círculo de correlaciones. Para más de un eje, la calidad de representación de una variable será la suma de las calidades de representación de cada una y que algunos autores denominan *comunalidad*.

— La *contribución* de una variable a la inercia de un eje k se calcula mediante la siguiente expresión:

$$\text{Contribución } (Z_j) = \frac{\text{Inercia de la proyección de la variable sobre "k"}}{\text{Inercia de la proyección de todas las variables sobre "k"}} =$$

$$= \frac{\lambda_k u_{kj}^2}{\lambda_k} = u_{kj}^2$$

puesto que la inercia de la proyección para todas las variables es:

$$\sum_{j=1}^p \lambda_k u_{kj}^2 = \lambda_k$$

para una variable Z_j se tiene que la inercia de su proyección es: $\lambda_k u_{kj}^2$.

En este caso se determinarán las variables que más influyen en la construcción del eje k . La calidad de representación es proporcional a la contribución dado que λ_k es común para todas las variables.

b. Estudio de la inercia de los factores

Sabemos que el primer valor propio (λ_1) tomará un valor entre uno y la inercia total de la nube (p en este caso): valdrá uno si las variables no están correlacionadas dos a dos y por tanto el primer factor no explica más que la varianza de una variable; λ_1 valdrá p cuando existe una relación lineal perfecta entre todas las variables, y en este caso la primera componente refleja esta relación.

Cuanto más grande sea dicho valor propio más información contendrá el factor al que está asociado sobre las relaciones entre las variables. Para el resto de valores propios el valor uno es un punto de referencia, es decir, un valor propio inferior a la unidad indica que la componente principal a la que está asociado sintetiza menos información que una variable aislada y por ello la interpretación de ese factor debe hacerse con cierta cautela.

Por otra parte, si dos factores están asociados a dos valores propios casi iguales significa que ambos representan la misma variabilidad y que no hay porque retener uno si y otro no. Recíprocamente, un fuerte descenso entre dos valores propios sucesivos lleva a retener, para la interpretación, los factores anteriores a este decrecimiento.

Generalmente los h primeros valores propios presentan un decrecimiento bastante regular, lo que indica que esos factores corresponden a características de la nube que deben ser interpretadas y sugiere que los factores siguientes solo representan el "ruido" que acompaña a toda

observación estadística. Como se indicó en su momento, un descenso regular desde el primer valor propio significa que la nube es poco esférica y que los factores sintetizan poca información e indica que las variables originales no están lo suficientemente correlacionadas para este análisis.

Por tanto, el número de factores que se debe retener dependerá del porcentaje de inercia extraída por el factor, puesto que es la medida de la importancia relativa de los factores que se extraen. Parece adecuado retener aquellos factores que en conjunto representan entre el 70 y el 80% de la inercia total, y que tengan además un valor propio superior a uno. Ahora bien, dicho porcentaje debe ser interpretado según los datos originales: un 10% sería un porcentaje débil si la tabla tiene diez variables y alto si la tabla tiene cien variables. También parece necesario establecer un criterio para los factores con valor propio superior a uno, ya que puede darse la situación de que dichos factores sean difícilmente interpretables mientras que otros, de inercia inferior, pongan de manifiesto informaciones difíciles de ver en la tabla de datos. Así, además del criterio de inercia indicado, se suele recomendar tomar para el análisis, factores que sean claramente interpretables al tiempo que será cuestionable considerar un factor que no sea claramente interpretable, aunque tenga más inercia.

c. Interpretación de las componentes o factores

Los factores deben retenerse teniendo en cuenta el criterio antes mencionado. Una vez que se ha decidido los factores a considerar, éstos pueden ser estudiados de forma separada o dos a dos, formando planos factoriales. Debido a la dualidad entre las nubes de individuos y variables, es necesario consultar alternativamente los resultados obtenidos para ambos, aunque puede seguirse el orden que se expresa a continuación:

CONTRIBUCIÓN DE LOS INDIVIDUOS

La contribución de éstos a los factores indicará la generalidad o no de un factor, en el sentido del número de individuos que participan en él. Se puede realizar una aproximación previa consultando la lista de contribuciones para ver si un individuo, o un grupo de ellos, tienen en el factor una contribución superior a la media.

En el caso de que un factor esté inducido por un solo individuo, se pueden hacer las siguientes consideraciones:

- Si el factor es uno de los primeros, el individuo en cuestión es necesariamente muy diferente de los otros, lo que dificultará el análisis del resto de la población. Se puede rehacer el análisis suprimiendo dicho individuo como elemento activo. Si la inercia del eje k del individuo suprimido es inferior a la diferencia entre λ_k y λ_{k+1} los nuevos factores obtenidos están correlacionados con los anteriores y la eliminación del individuo no perturba los resultados. De todas maneras, la eliminación del individuo del análisis puede afectar al objetivo y hay que tener en cuenta la situación general del estudio y no solo el criterio de inercia.
- Si dicho factor está después de los que expresan las tendencias generales, es natural que los factores, una vez extraídos los generales, reflejen fenómenos concretos y puntuales. Hay que señalar que en el ACP *normado* el problema de elementos excepcionales solo afecta a los individuos pues las variables poseen todas la misma varianza.

COORDENADAS DE LAS VARIABLES ACTIVAS

El estudio de las gráficas suele comenzar por el análisis de las variables: la interpretación de los factores debe hacerse según la correlación de las variables activas con los mismos, es decir, según su coordenada con el eje correspondiente. En este caso, la interpretación se apoya en la regla

siguiente: la coordenada de la variable X_j sobre el eje k es el coeficiente de correlación entre esta variable y el factor k . Además en el ACP *normado* las variables tienen el mismo peso entre sí y están equidistantes, por lo que el cuadrado de su coordenada es la calidad de representación de dicha variable y es proporcional a su contribución, como ya hemos indicado. Las variables que se encuentren realmente en el plano estarán situadas en la periferia del círculo de correlaciones. Se suelen producir las dos situaciones siguientes:

- Las variables más correlacionadas están situadas del mismo lado del eje y el factor es entonces una síntesis de esas variables que mide el efecto de "talla".
- Entre las variables muy correlacionadas con un factor, unas presentan una coordenada positiva y otras negativa. Es preciso buscar una relación entre las variables de un lado, en oposición a las variables del otro lado. Esta situación permitirá dotar de significado general al factor y en algunos casos atribuirle una denominación.

INTERPRETACIÓN DEL PLANO

La representación de las variables en el plano permite una visión más sugestiva para interpretar los resultados que en un solo eje factorial. Las variables próximas al círculo de correlaciones estarán bien representadas y los vectores que unen el origen a los puntos-variable permiten visualizar los ángulos entre variables y medir la relación entre ellas. Sólo las variables bien representadas serán tenidas en cuenta y la representación gráfica permite reagrupar aquellas que presentan una tipología común. Los factores que formarán el plano deben seguir las indicaciones ya apuntadas para su interpretación y además deben tener una importancia comparable para que el plano que determinan sea estable.

d. Interpretación simultánea de individuos y variables

Para una correcta interpretación de la representación simultánea, además de lo indicado, hay que precisar que:

- No es posible interpretar la posición de un individuo respecto a una sola variable (y viceversa) pues, como indican las fórmulas de transición, la posición del individuo depende de las coordenadas de todas las variables, y la proyección de un individuo sobre un eje es una combinación lineal de las proyecciones de todas las variables sobre dicho eje. Por su parte, el coeficiente de una variable (positivo o negativo) para un individuo depende del valor z_{ij} de esa variable para dicho individuo, de tal manera que si $z_{ij} > \bar{Z}_j$ entonces el individuo toma un valor superior a la media y el coeficiente es positivo; en caso contrario el coeficiente es negativo. Por ese motivo, sólo pueden interpretarse las posiciones respectivas de dos individuos en función del conjunto de variables y las posiciones de las variables en función del conjunto de individuos.
- Se debe considerar el distanciamiento de un individuo respecto al origen en la dirección de un conjunto de variables y no la proximidad de dicho individuo al conjunto de puntos que representan dichas variables. Es decir, interesa el movimiento de los individuos en la dirección de las variables y la dirección de una variable define las áreas de los individuos: dicho individuo estará al lado de las variables para las que tiene valores altos y estará al lado opuesto de las variables con valores bajos. Por otra parte, si una variable está muy correlacionada con un factor significa que los individuos que poseen una coordenada elevada positiva con él están caracterizados por valores de Z_j muy superiores a la media.
- Los individuos cuya representación está cerca del origen (que es el centro de gravedad) tienen una difícil interpretación. Éstos no presentan valores muy extremos para el conjunto de variables, sino

más bien en torno a los valores medios de todas ellas, y de ahí su proximidad al dicho centro de gravedad.

1.4 ACP *no-normado* (centrado)

Abordamos ahora el análisis de la nube de individuos y variables cuando las variables no están reducidas. En este caso la influencia de cada variable en el análisis dependerá de su varianza, y por ello se analiza tanto la forma como el tamaño de la dos nubes².

Las fases que vamos a considerar son básicamente las mismas que para el ACP *normado*: análisis de la nube de individuos, análisis de la nube de variables e interpretación del método, considerando solamente aquellos aspectos que difieren respecto al ACP *normado* e indicando cuando es aconsejable la elección de uno u otro.

Para su descripción se considera la nube de individuos centrada y no reducida, es decir, la matriz de datos Y de tal manera que $y_{ij} = x_{ij} - \bar{x}_j$. La matriz $M = I_{(p \times p)}$ recoge el peso de las distintas variables, y cada individuo tiene como peso p_i ($i = 1 \dots n$), pesos que se recogen en la matriz diagonal D y que con carácter general son $1/n$ para todos los individuos.

1.4.1 Análisis de la nube de individuos N_I

Partiendo del espacio R^p en el que cada individuo se representa como $e_i' = (y_{i1}, y_{i2} \dots y_{ip})$, la nube de individuos sigue tomando como origen el centro de gravedad.

² Para profundizar en los conceptos de forma y tamaño nos remitimos a Cuadras (1980: 275-6).

La media de Y_j y de $Y_{j'}$ es cero, y por ello, el valor de la covarianza entre dos variables en este espacio vendrá dada por:

$$Cov(Y_j, Y_{j'}) = Y_j' D Y_{j'} = \sum_{i=1}^n p_i y_{ij} y_{ij'}$$

Esta expresión coincide con la $Cov(X_j, X_{j'})$, dado que $p_i = 1/n$ para todo e_i :

$$Cov(X_j, X_{j'}) = \sum_{i=1}^n p_i y_{ij} y_{ij'} = \sum_{i=1}^n p_i (x_{ij} - \bar{x}_j)(x_{ij'} - \bar{x}_{j'})$$

En primer lugar vamos a considerar el producto escalar y la norma de los individuos para luego definir la inercia de la nube y describir el proceso para la obtención de las componente principales, con las consecuencias que de ello se derivan.

— El producto escalar entre individuos, en este caso, toma la siguiente expresión:

$$e_i' M e_{i'} = w_{ii'} = \sum_{j=1}^p m_j y_{ij} y_{ij'}$$

siendo $W = Y M Y'$ la matriz de productos escalares.

— La distancia de un individuo e_i al origen es su norma:

$$d^2(e_i, 0) = \sum_{j=1}^p m_j (y_{ij} - 0)^2 = \sum_{j=1}^p m_j y_{ij}^2 = e_i' M e_i = \|e_i\|_M^2$$

— La inercia de la nube será:

$$I = \sum_{i=1}^n p_i d^2(e_i, 0) = \sum_{i=1}^n p_i \sum_{j=1}^p m_j y_{ij}^2 = \sum_{i=1}^n p_i e_i' M e_i = Tr(WD)$$

que coincide con la $Tr(VM)$.

— Las coordenadas de los individuos en el eje k se obtienen como:

$$F_k = Y u_k$$

siendo:

$$f_{ki} = \sum_{j=1}^p y_{ij} u_{kj}$$

— La obtención de los ejes factoriales y las componentes principales se realiza teniendo en cuenta que los ejes deben hacer máxima la inercia sobre ellos proyectada. El procedimiento para su obtención coincide con el descrito para el ACP *normado*. La matriz de inercia es la matriz de varianzas-covarianzas entre las variables originales:

$$S_1 = M Y' D Y = M V_y = M V$$

Para el primer eje se cumple que:

$$S_1 u_1 = \lambda_1 u_1 \Rightarrow M V u_1 = \lambda_1 u_1$$

A partir de aquí se determina el primer eje factorial y se obtiene la primera componente. Dicha componente cumple que: $F_1' D F_1 = \lambda_1$, siendo λ_1 la inercia de ese eje y $\bar{F}_1 = 0$.

Los h ejes se obtienen como:

$$M V \sum_{k=1}^h u_k = \sum_{k=1}^h \lambda_k u_k$$

y la inercia de los h ejes es:

$$I(N_I; a_1, a_2, \dots, a_h) = \sum_{k=1}^h F_k' D F_k = \sum_{k=1}^h \lambda_k$$

— Una vez obtenidas dichas componentes podemos extraer la siguientes conclusiones:

- Si la matriz MV es de rango p la proporción de inercia explicada por los h ejes factoriales es:

$$\frac{\sum_{k=1}^h \lambda_k}{\sum_{k=1}^p \lambda_k} = \frac{\sum_{k=1}^h \lambda_k}{\sum_{k=1}^p m_j \text{Var}(Y_j)}$$

puesto que MV es cuadrada, y se tiene que:

$$\text{Tr}(MV) = \sum_{k=1}^p \lambda_k = \sum_{j=1}^p m_j \text{Var}(Y_j)$$

- Como para el caso de variables normadas, es posible demostrar que las componentes principales F_k son los vectores propios, asociados a los valores propios, de la matriz WD , de tal manera que:

$$WD \sum_{k=1}^h F_k = \sum_{k=1}^h \lambda_k F_k$$

También es posible demostrar que las componentes principales están incorrelacionadas, por lo que $\text{Cov}(F_k, F_{k'}) = 0$.

- La fórmula de reconstitución de los datos que se deduce en este caso es:

$$Y_j = \sum_{k=1}^p F_k m_k^{-1} u_{kj}$$

cuya aproximación de orden h es:

$$Y_j = \sum_{k=1}^h F_k m_k^{-1} u_{kj}$$

- En este caso la correlación entre F_k e Y_j toma la siguiente expresión:

$$\rho(F_k, Y_j) = \frac{\text{Cov}(F_k, Y_j)}{\sqrt{\text{Var}(F_k)} \sqrt{\text{Var}(Y_j)}}$$

Teniendo en cuenta que: $\text{Var}(F_k) = \lambda_k$; $\bar{F}_k = 0$ y la $\text{Var}(Y_j)$ no es unitaria, la covarianza de Y_j y F_k es el producto escalar entre esos dos puntos de R^n :

$$\text{Cov}(F_k, Y_j) = F_k' D Y_j = F_k' D \sum_{k=1}^p F_k m_k^{-1} u_{kj} = \lambda_k m_k^{-1} u_{kj}$$

Por ello:

$$\rho(F_k, Y_j) = \frac{\lambda_k m_k^{-1} u_{kj}}{\sqrt{\lambda_k} \sqrt{\text{Var}(Y_j)}} = \frac{\sqrt{\lambda_k} m_k^{-1} u_{kj}}{\sqrt{\text{Var}(Y_j)}}$$

1.4.2 Análisis de la nube de variables

Se parte de que la nube de variables no está centrada y el origen no coincide con el centro de gravedad. Como en el caso anterior, se van a ir definiendo el producto escalar, la distancia al origen, la inercia de la nube, la obtención de los ejes y las consecuencias que se deducen.

— El producto escalar entre dos variables Y_j e $Y_{j'}$ es:

$$Y_j' D Y_{j'} = \sum_{i=1}^n p_i y_{ij} y_{ij'} = \text{Cov}(Y_j, Y_{j'})$$

— La distancia (o inercia de esa variable) entre una variable Y_j y el origen no es uno; y por tanto, en este caso, la nube de variables no es la esfera de correlaciones sino que:

$$d^2(Y_j, 0) = \sum_{i=1}^n p_i y_{ij}^2 = \text{Var}(Y_j)$$

En el espacio R^n la distancia entre Y_j y $Y_{j'}$ es:

$$\begin{aligned} d^2(Y_j, Y_{j'}) &= \sum_{i=1}^n p_i (y_{ij} - y_{ij'})^2 = \sum_{i=1}^n p_i y_{ij}^2 + \sum_{i=1}^n p_i y_{ij'}^2 - 2 \sum_{i=1}^n p_i y_{ij} y_{ij'} = \\ &= \text{Var}(Y_j) + \text{Var}(Y_{j'}) - 2 \text{Cov}(Y_j, Y_{j'}) \end{aligned}$$

Como podemos observar, la distancia entre dos variables se expresa en términos de covarianza y aumenta con la varianza de las variables. Así, la

distancia disminuye si la relación entre las variables es positiva y aumenta en caso contrario.

— La inercia, respecto al origen, de la nube es:

$$\sum_{j=1}^p m_j d^2(Y_j, 0) = \sum_{j=1}^p m_j \sum_{i=1}^n p_i y_{ij}^2 = \sum_{j=1}^p m_j \text{Var}(Y_j) = \text{Tr}(V_y M) = \text{Tr}(VM)$$

siendo V_y la matriz de varianzas–covarianzas de las variables Y_j ($V_y = V$ es la matriz de varianzas–covarianzas de las variables iniciales). Además $\text{Tr}(VM) = \text{Tr}(WD)$.

— La coordenada de una variable se expresará entonces como:

$$G_k = Y' v_k = \sum_{i=1}^n y_{ij} v_{ki}$$

— Para obtener los ejes factoriales, se debe maximizar la inercia explicada por los ejes factoriales y la matriz de inercia que se obtiene es:

$$S_J = D Y M Y' = DW$$

de tal manera que los ejes son:

$$S_J \sum_{k=1}^h v_k = \sum_{k=1}^h \lambda_k v_k$$

siendo:

$$G_k' M G_k = \lambda_k$$

y la inercia explicada por los h ejes es:

$$I(N_I; t_1, t_2, \dots, t_h) = \sum_{k=1}^h G_k' M G_k = \sum_{k=1}^h \lambda_k$$

— De este proceso se pueden extraer las siguientes conclusiones:

- Es posible demostrar que las G_k son los vectores propios, asociados a los valores propios, de la matriz VM de tal manera que:

$$VM \sum_{k=1}^h G_k = \sum_{k=1}^h \lambda_k G_k$$

- Cuando el rango de la matriz VM sea p la proporción de inercia explicada por los h ejes obtenidos es:

$$\frac{\sum_{k=1}^h \lambda_k}{\sum_{k=1}^p \lambda_k} = \frac{\sum_{k=1}^h \lambda_k}{\sum_{j=1}^p m_j \text{Var}(Y_j)}$$

- Considerando el eje k tenemos que:

$$WDF_k = \lambda_k F_k$$

$$WDt_k = \lambda_k t_k$$

Igual que para el caso normado, se puede demostrar que:

$$t_k = c F_k \quad y \quad c = \frac{1}{\sqrt{\lambda_k}} \Rightarrow t_k = \frac{1}{\sqrt{\lambda_k}} F_k \Rightarrow$$

$$\Rightarrow F_k = \sqrt{\lambda_k} t_k \Rightarrow f_{ki} = \sqrt{\lambda_k} t_{ki}$$

$$a_k = q G_k \quad y \quad q = \frac{1}{\sqrt{\lambda_k}} \Rightarrow a_k = \frac{1}{\sqrt{\lambda_k}} G_k \Rightarrow$$

$$\Rightarrow G_k = \sqrt{\lambda_k} a_k \Rightarrow g_{kj} = \sqrt{\lambda_k} m_k^{-1} u_{kj}$$

Tenemos que:

$$g_{kj} = \sqrt{\lambda_k} m_k^{-1} u_{kj} = \frac{\text{Cov}(F_k, Y_j)}{\sqrt{\lambda_k}} \Rightarrow \text{Cov}(F_k, Y_j) = \sqrt{\lambda_k} g_{kj}$$

En este caso la coordenada de una variable con un eje es la covarianza de la variable con el eje correspondiente.

- Dado que el coeficiente de correlación entre la variable Y_j y F_k toma la siguiente expresión:

$$\rho(F_k, Y_j) = \frac{\lambda_k m_k^{-1} u_{kj}}{\sqrt{\lambda_k} \sqrt{\text{Var}(Y_j)}}$$

se cumple que:

$$\rho(F_k, Y_j) = \frac{\sqrt{\lambda_k} m_k^{-1} u_{kj}}{\sqrt{\text{Var}(Y_j)}} = \frac{g_{kj}}{\sqrt{\text{Var}(Y_j)}}$$

Si bien en este caso la coordenada de la variable con un eje no es el coeficiente de correlación de la variable con la componente correspondiente.

- En esta situación, la primera componente principal hace máximo el cuadrado de las covarianzas de las variables con dicha componente, ya que:

$$\sum_{j=1}^p g_{1j}^2 = \sum_j \frac{\text{Cov}^2(F_1, Y_j)}{\lambda_1} = \frac{1}{\lambda_1} \sum_j \text{Cov}^2(F_1, Y_j) = \lambda_1$$

1.4.3 Interpretación

En principio nos referiremos a las ayudas a la interpretación y posteriormente a las características peculiares de esta variante del método respecto al ACP *normado*.

— Es necesario considerar, en principio, ciertos coeficientes que nos ayuden a la interpretación de los resultados:

Así, la *calidad de representación* de una variable para el eje k viene dada por:

$$\text{Calidad}(Z_j) = \frac{\text{Inercia de la proyección de la variable sobre "k"}}{\text{Inercia total de } X_j}$$

$$\text{Calidad}(Z_j) = \frac{d^2(g_{kj}, 0)}{d^2(Z_j, 0)} = \frac{g_{kj}^2}{\text{Var}(Y_j)} = \frac{\lambda_k (m_k^{-1})^2 u_{kj}^2}{\text{Var}(Y_j)} = \rho^2(F_k, Y_j)$$

es decir, el coeficiente de correlación al cuadrado entre variable y factor.

La *contribución* de una variable a la inercia de un eje k se calcula mediante la siguiente expresión:

$$\text{Contribución } (Z_j) = \frac{\text{Inercia de la proyección de la variable sobre "k"}}{\text{Inercia de la proyección de todas las variables sobre "k"}} = \frac{g_k^2}{\lambda_k} = u_{kj}^2$$

Las coordenadas de las variables no pueden interpretarse como el coeficiente de correlación con el factor correspondiente, ya que el cuadrado de ésta coordenada mide sólo la calidad de representación pero no es proporcional a su contribución.

— Para la correcta interpretación de este análisis es necesario tener en cuenta que:

- La inercia de cada variable no vale uno y los valores propios obtenidos no son comparables. Para interpretar la inercia de cada factor hay que analizar el porcentaje y no el valor total. Como el análisis se hace sobre la matriz de covarianzas, se considerarán las componentes cuya inercia sea superior a la mínima de las varianzas de las variables (Cuadras, 1981: 277).
- El peso de cada variable en la distancia entre individuos no es el mismo, sino que depende de la dispersión de la variable, por lo que el primer eje puede originarse a partir de una sola variable.
- La representación gráfica de las coordenadas de las variables con los ejes debe acompañarse del estudio de las correlaciones con los factores. La representación gráfica de estas últimas da lugar al círculo de correlaciones y permite la interpretación de éstos.

Conocidas las características de ambas variantes del método, estamos en condiciones de proponer cuando es recomendable una u otra. Generalmente cuando las unidades de las variables que intervienen en el estudio son

razonablemente comparables la covarianza entre ellas tiene mayor interés y la componente k del ACP *no-normado* permite extraer la k -ésima parte más grande del total de la varianza. Además, la maximización de la inercia a partir de las puntuaciones tipificadas es en cierto sentido artificial (Morrison, 1967: 272).

Si las medidas empleadas son, en cierta medida, arbitrarias (como por ejemplo en datos sociológicos) es aconsejable un ACP *normado*. En general las componentes obtenidas no son las mismas en uno y en otro análisis pero es posible pasar de unas a otras reescalando los coeficientes. Conviene señalar que si las varianzas de las variables originales son muy diferentes unas de otras, las componentes extraídas a partir de la matriz de covarianzas, van a reflejar en orden descendente las variables originales en orden descendente de varianza y en ese caso también es aconsejable tipificar las variables³.

En algunos casos también es posible trabajar con las variables sin centrar pero entonces no se trata de un ACP al no tener las mismas propiedades. El centrado de las variables permite interpretar los ejes factoriales en R^p como las direcciones de mayor variabilidad de la nube N_1 . Sin centrado previo, no sólo influye en los ejes la forma de la nube, sino también su posición respecto al origen. El centrado permite además interpretar el coseno del ángulo en R^n entre dos vectores columna como un coeficiente de correlación.

³ Así se demuestra en Chatfield y Collins (1980: 74).

2. Desarrollo de la metodología STATIS

Toda vez que se ha contextualizado la metodología STATIS, como técnica adecuada cuando las variables objeto de estudio son de tipo cuantitativo y la estrategia global requiere análisis separados y conjuntos de los tres modos, el paso siguiente es su desarrollo teórico. Éste es el objetivo del apartado que nos ocupa, para lo cual comenzaremos con los antecedentes históricos de dicha metodología y los conceptos generales necesarios para el tratamiento de más de una tabla de datos. A continuación se realiza el desarrollo teórico de los dos métodos (o versiones) de que consta, es decir, la versión *Statis*, cuando los individuos son los mismos en todas las tablas, y la versión *Statis dual*, cuando lo que es común a las tablas, son las variables.

Nos parece conveniente señalar la denominación a la que haremos referencia que se concreta indistintamente en metodología o método STATIS, entendiendo que es un proceso metodológico que se materializa en un método. Por esta razón, al hablar de *Statis* y *Statis dual*, éstas se entienden como versiones del método o bien, como métodos dentro de la metodología global.

2.1 Antecedentes históricos

La metodología STATIS (Structuration des Tableaux à Trois Indices de la Statistique) fue desarrollada inicialmente por el equipo del profesor Escoufier, en el ámbito del Laboratorio de Estadística de la Universidad de Ciencias y

Técnicas de Languedoc, así como en la unidad de Biometría, en la que dicho laboratorio colaboraba con la Cátedra de Matemáticas e Informática de la Escuela Nacional Superior de Agronomía de Montpellier, y con el Departamento de Biometría del Instituto Nacional de Investigación Agronómica.

La primera etapa del método (la interestructura) fue expuesta por Escoufier *et al.* (1976: 61-89) y Cailliez y Pagès (1976) y está basada en los coeficientes de asociación entre tablas, propuestos por Escoufier (1970: 1-47). Posteriormente, L'Hermier des Plantes desarrollaría el método completo (en su versión *Statis*) en su tesis doctoral en 1976. El artículo de Escoufier y L'Hermier des Plantes (1978: 477-483) desarrolla la versión dual del método al intentar comparar una serie de matrices de varianzas-covarianzas obtenidas en el tratamiento de ciertos datos meteorológicos. Posteriormente, Escoufier (1980: 59-76) presenta una versión del STATIS para variables cualitativas, pero sin software específico para ese caso; Foucart (1983: 61-75) se plantea una nueva aproximación, proponiendo una nueva forma de analizar la intraestructura que aplica al análisis de ratios de una muestra de empresas de la zona de Bretaña.

Dada la proximidad con otros métodos de tratamiento de tablas de tres entradas, las comparaciones con técnicas similares se recogen en varios trabajos, como los ya mencionados de Glaçon (1981), los artículos de Kiers (1988: 14-32) y la obra del GERI (1996). En éste último se compara, en particular, con otros métodos multidimensionales que consideran el aspecto temporal de la información. El artículo de Pagès (1996: 81-95) somete a una revisión exhaustiva al STATIS y el AFM, en cuanto que éste es el método más próximo en relación a los objetivos y al desarrollo.

Mientras que la interestructura nunca ha sido objeto de controversia, el cálculo de las trayectorias ha suscitado y propuesto diversas soluciones. La presentada por Lavit (1988) parece ser la que mejor refleja las desviaciones entre individuos. De todos modos, también tiene sus inconvenientes y algunos autores, como Morineau (CISIA, 1997: 27-53) en el programa SPAD-TM

versión 3, proponen otras soluciones. Los resultados obtenidos con el STATIS, han servido de punto de partida de otros métodos desarrollados, como el que propone el artículo de Carlier (1985: 27-53), sobre la clasificación de trayectorias de los individuos. Asimismo, los factores del compromiso también pueden tomarse como punto de partida de otros análisis, como luego demostraremos.

En cuanto a los programas informáticos desarrollados para su aplicación, hay que señalar que el primero fue el realizado por Place (1980) y Lavit y Roux en 1983, cuya implantación en la Biblioteca de Programas de Análisis de Datos (ANADO) del Centro Nacional Universitario Sud de Cálculo, permitió realizar las primeras aplicaciones del método a situaciones reales y perfeccionar la técnica. Posteriormente Lavit, en 1986, propondría otra versión del programa más perfeccionada. En 1989 se desarrolla el programa ACT (método STATIS), distribuido por CISIA, que permite la aplicación de los métodos *Statis* y *Statis dual* con el cálculo de las trayectorias ya mencionado y propuesto por Lavit. En 1993 el *Statis* y el *Statis dual* serán implementados en el paquete estadístico SPAD y en 1995 los autores Chessel y Dolédec lo incluyen en su paquete estadístico ADE-4. La tercera versión del SPAD (en concreto el módulo SPAD-TM de 1997) permite la aplicación del método en sus dos versiones, pero elimina el cálculo de las trayectorias por medio de la técnica de individuos suplementarios que había planteado Lavit; en él se plantea otra solución que consiste en el cálculo de las diferencias entre tablas, analizando las distancias entre los individuos o entre las variables, según el caso.

Un aspecto importante en relación a su proceso de evolución son las aplicaciones. Las más frecuentes provienen del ámbito de la Ecología y de la Biología; en este último campo destaca el artículo de Génard *et al.* (1994: 241-245) que aplica la metodología STATIS con el fin de obtener un patrón de correlación común entre los parámetros de calidad, obtenidos en diversas condiciones, en el cultivo del melocotón. En Ecología ha sido aplicado en diversos trabajos, entre los que destacan el de Amanieu *et al.*

(1981: 303-312) cuyo objetivo es determinar la tipología de diferentes grupos faunísticos en la caracterización espacio-temporal de un medio. Ecoutin (1989) la emplea para analizar las variaciones espacio-temporales de una serie de peces y Verneaux y Chessel (1994) con el fin de determinar la permanencia o no de las estructuras mesológicas en los diferentes ríos franceses. La tesis de Barán (1995) aplica el método para estudiar la dinámica espacio-temporal de los peces de los estuarios de Guinea y de sus relaciones con el medio abiótico.

Ya en el ámbito económico las aplicaciones quedan prácticamente restringidas al campo de la Economía Agraria. Así, destaca el artículo de L'Hermier des Plantes y Thiebaut (1977: 57-81) sobre la evolución de la pluviometría en Languedoc y el trabajo de Bernard *et al.* (1978) que analiza los resultados proporcionados por una de las primeras versiones del método. En la tesis doctoral de Pérez Hugalde¹³ de 1984, se emplea para estudiar la productividad y las estructuras agrarias de las provincias españolas. En el artículo de Pérez Hugalde *et al.* (1986: 317-351) se analizan los resultados de los sectores agrarios y no agrarios de las distintas provincias españolas con el mismo método. El estudio realizado por Pérez Pérez y Pérez Hugalde (1990: 95-108) sobre la evolución de la dimensión del tamaño de las explotaciones agrarias en las comarcas de Aragón en el período 1962-1982 es otro ejemplo de aplicación del método en este ámbito. La aplicación de Álvarez Esteban *et al.* (1996: 23-37) a la producción agrícola y a la producción cárnica de la Comunidad Autónoma de Castilla y León, y las de Landaluce Calvo y Valencia García (1998: 442-455) y Huerga Castro *et al.* (1998: 410-428) aplicadas, respectivamente, al Valor Añadido Bruto de las diferentes ramas industriales y a las exportaciones e importaciones castellano-leonesas, son otros dos ejemplos de aplicaciones del método en el campo económico.

¹³ La publicación de la misma es del año 1988 y la referencia bibliográfica es C. Pérez Hugalde (1988), "Evolución de las relaciones entre la productividad del trabajo en la agricultura, las estructuras agrarias y el desarrollo económico en España 1960-1979. Un análisis mediante métodos estadísticos multivariados". Madrid: Ed. Ministerio de Agricultura, Pesca y Alimentación. Colección Tesis Doctorales.

En el marco de las ciencias sociales encontramos el trabajo de Bernard y Lavit (1985: 183-196) sobre la evolución de la población activa de los cantones rurales de Languedoc–Roussillon en el período 1954-1975. En el ámbito financiero el trabajo ya mencionado de Foucart, aplicado al análisis de ratios de una muestra de empresas de la zona de Bretaña, constituye uno de los pocos ejemplos de aplicación encontrados en dicho campo.

Podemos señalar además que en la bibliografía española de Análisis de Datos, las referencias a la metodología STATIS son escasas. La obra de Judez Asensio (1989) constituye una de las pocas que la consideran, a pesar de ser muy numerosos los títulos que tratan de dar una visión general de las diferentes Técnicas Multidimensionales existentes.

2.2 Operadores asociados a una tabla de datos

Además de los aspectos teóricos ya considerados al desarrollar el *Análisis de Componentes Principales* para la adecuada comprensión de las etapas que conlleva la aplicación del STATIS, son necesarios los conceptos que van a desarrollarse en este apartado.

Se parte de una tabla de n individuos por p variables centradas, como la que se muestra a continuación:

Figura II.12
Tabla de datos Y

	1	2	...	j	...	p
1	y_{ij}					
2						
⋮						
i						
⋮						
n						

El elemento y_{ij} representa dicha tabla y se obtiene cuando X se centra para los pesos de los individuos $p_i = 1/n$ de tal manera que $y_{ij} = x_{ij} - \bar{x}_j$. El espacio de los individuos es el espacio $E = R^p$ y el espacio de las variables es $F = R^n$; como se indicó en su momento, el centro de gravedad de la nube de individuos es el origen, pero en el caso de la nube de variables esos dos puntos no son coincidentes.

En el espacio de los individuos $E = R^p$ el sistema de referencia original viene dado por la base canónica de dicho espacio, formada por p vectores M -ortogonales. La matriz $M = I_{(n \times n)}$ es la métrica del espacio y esos vectores unitarios serán los r_j ($j = 1, 2 \dots p$). Por ello el individuo $e'_i = (y_{i1}, y_{i2} \dots y_{ip})$ se expresa de la forma:

$$e_i = \sum_{j=1}^p y_{ij} r_j$$

De la misma forma, en el espacio de las variables $F = R^n$, la base canónica original de dicho espacio estará formada por n vectores D -ortogonales, siendo D la métrica de dicho espacio. Tales vectores l_i ($i = 1, 2 \dots n$), permiten expresar la variable Y_j de la forma:

$$Y_j = \sum_{i=1}^n y_{ij} l_i$$

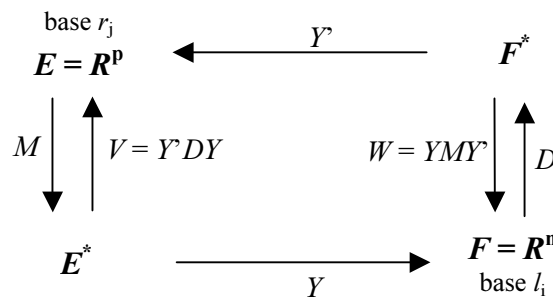
Asimismo, al individuo e_i se le asociará el vector l_i^* del espacio F^* , dual¹⁴ de F , siendo Y la matriz asociada a la transformación lineal que permite considerar e_i en el espacio E (de base r_j) y en el espacio F^* (de base l_i^*) cuando se pasa de F^* a E .

Recíprocamente, cuando se considera la variable Y_j se le asociará el vector r_j^* del espacio E^* , dual de E , siendo Y la matriz asociada a la transformación lineal que permite considerar Y_j en el espacio en el espacio F (de base l_i) y en el espacio E^* (de base r_j^*).

¹⁴ Los espacios duales están desarrollados más ampliamente en Queysanne (1975: 319-25).

Habiendo elegido para medir la semejanza entre individuos la métrica M sobre el espacio E , y la métrica D para medir la semejanza entre variables en el espacio F , el esquema de dualidad al que da lugar es el siguiente:

Figura II.13a
Esquema de dualidad

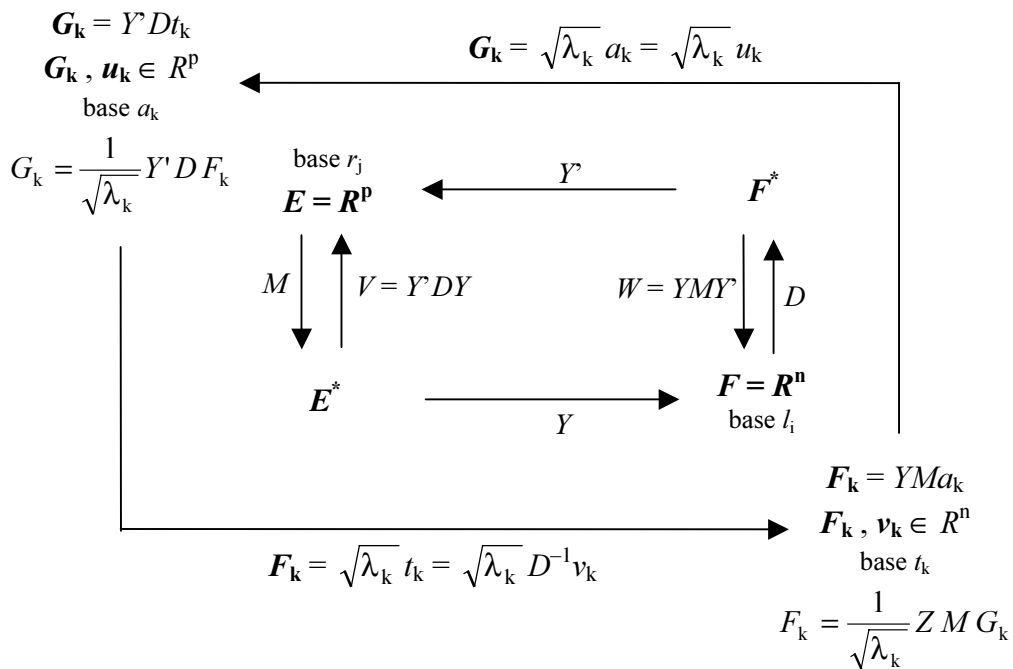


La métrica asociada al espacio de individuos, M , puede definirse como la aplicación de E en E^* , que es un isomorfismo de E en E^* (el espacio E dotado de la métrica M y el espacio E^* dotado de la métrica M^{-1}). La métrica D del espacio de las variables es una aplicación de F en F^* , que es otro isomorfismo de F en F^* (F está dotado de la métrica D y F^* de la métrica D^{-1}). Las matrices $Y M$ y $Y^* D$ definen las aplicaciones de F en E y de E en F , respectivamente.

La matriz: $W = Y M Y^*$, es la matriz de productos escalares entre individuos y la matriz: $V = Y^* D Y$, la de productos escalares entre variables (matriz de varianzas-covarianzas). W es la métrica euclídea sobre F^* si se desea que las distancias entre individuos sean las mismas en E y en F^* ; V es la métrica euclídea en E^* para que en este espacio y en F las distancias entre variables sean las mismas.

Cuando se buscan los nuevos vectores, las a_k en el espacio de los individuos, y los t_k en el espacio de las variables, las nuevas coordenadas de los individuos son $F_k = Y M a_k$ y las coordenadas de las variables $G_k = Y^* D t_k$. Por tanto, completando el esquema de dualidad anterior, se tiene que:

Figura II.13b
Esquema de dualidad



Es decir, si al vector a_k de E se le aplica sucesivamente $M Y D Y'$ se obtiene de nuevo a_k , excepto por $\sqrt{\lambda_k}$, es decir, se obtiene G_k . Lo mismo ocurre en el caso de las variables, estableciéndose así la relación entre los ejes factoriales de una nube y los factores de la otra.

Tomando como referencia los artículos de Escoufier (1976: 165-178) y Robert y Escoufier (1976: 257-265) definiremos el concepto de operador asociado a una tabla de datos. Según el esquema de dualidad, para una misma tabla pueden considerarse dos operadores diferentes según el objetivo perseguido: si se desea estudiar la nube de individuos definida por dicha tabla, o comparar dicha nube con las nubes de individuos de otras tablas de datos se tomará la matriz WD . Por el contrario, se considerará como operador representativo de la tabla la matriz VM si se desea estudiar la nube de variables, o comparar dicha nube con otras nubes de variables definidas por otras tablas de datos.

Vamos por tanto a profundizar en las características particulares de cada una de esas dos matrices. Cuando uno de los objetivos es establecer comparaciones entre diferentes $W_t D$ ($t = 1, 2 \dots T$), o entre diferentes $V_t M$ ($t = 1, 2 \dots T$) es necesario definir en el espacio de cada uno de esos operadores, una métrica, un producto escalar y una distancia. Ese es precisamente uno de los objetivos del STATIS y de ahí la necesidad de desarrollar estas cuestiones.

2.2.1 Operador $W_t D$. Operador normado $\frac{W_t D}{\|W_t D\|_{HS}}$

El operador $W_t D$ es un elemento de $L(F, F)$, conjunto de aplicaciones lineales D -simétricas positivas, del espacio de variables F . Además, según demuestra Lavit (1988: 68), puesto que las variables medidas sobre los individuos están centradas, la tabla de productos escalares entre individuos también estará centrada.

En este conjunto se puede definir un producto escalar que le dota de una estructura de espacio de Hilbert-Schmidt y permite calcular las distancias entre operadores derivados de distintas tablas de datos. Así, considerando solo dos tablas de datos X^1 y X^2 , (es decir, $t = 1, 2$), observadas sobre los mismos n individuos y las matrices centradas Y^1 y Y^2 derivadas, respectivamente, de las tablas anteriores, se definen:

- $W_1 D = Y^1 M_1 Y^1 D$, siendo M_1 la matriz de ponderaciones de las variables (Y_j^1) de Y^1 , cuyos elementos de la diagonal son los m_j^1 ; la matriz D recoge los pesos de los individuos (e_i^1) y $W_1 D$ es de orden $n \times n$, siendo p_1 el número de variables de Y^1 . El triplete (Y^1, M_1, D) define el estudio.
- $W_2 D = Y^2 M_2 Y^2 D$, siendo M_2 la matriz de ponderaciones de las variables (Y_j^2) de Y^2 , cuyos elementos de la diagonal son m_j^2 ; la

matriz D recoge los pesos de los individuos (e_i^2). Dado que en ambas éstos son los mismos, D es común a ambas y también los $p_i = 1/n$. En este caso tenemos (Y^2, M_2, D) . $W_2 D$ también es de orden $n \times n$ y el número de variables de Y^2 es p_2 .

En este caso interesa comparar las nubes de individuos que se deducen de cada tabla de datos (generalmente porque los individuos son los elementos comunes y las variables son las que varían de una tabla a otra, aunque también pueden ser las mismas). La matriz D es común a las dos tablas pues ambas están medidas sobre los mismos individuos, y por consiguiente, también el peso otorgado a cada uno de ellos es el mismo.

Producto escalar

El producto escalar, de Hilbert-Schmidt, entre esos dos operadores se define como:

$$\langle W_1 D, W_2 D \rangle_{\text{HS}} = \text{Tr}(W_1 D W_2 D)$$

que es la covarianza vectorial, $\text{Cov}(W_1 D, W_2 D)$, tal y como lo define Lavit (1988: 84).

Este producto escalar, así definido, es en realidad el producto escalar entre los operadores W_1 y W_2 de R^{n^2} tomando como métrica la matriz $D \otimes D$ (producto tensorial entre matrices D) y similar al espacio de las variables en un ACP *no-normado*. La matriz que recoge esos pesos es:

$$D \otimes D = \begin{bmatrix} p_1 p_1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ p_n p_1 & & & \\ p_1 p_2 & & & \\ \vdots & \vdots & \vdots & \vdots \\ p_n p_2 & & & \\ \vdots & \vdots & \vdots & \vdots \\ p_1 p_n & & & \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & p_n p_n \end{bmatrix}$$

Desarrollando el producto escalar se tiene que:

$$\begin{aligned} \langle W_1 D, W_2 D \rangle_{HS} &= \text{Tr} (W_1 D W_2 D) = \sum_{i=1}^n (W_1 D W_2 D)_{ii} = \\ &= \sum_{i=1}^n \sum_{i'=1}^n (W_1 D)_{ii'} (W_2 D)_{i'i} \end{aligned}$$

Como WD es la matriz de productos escalares entre individuos tenemos que:

$$(W_1 D)_{ii'} = p_i e_i^1{}' M_1 e_{i'}^1 \text{ y } (W_2 D)_{i'i} = p_i e_{i'}^2{}' M_2 e_i^2$$

y por tanto:

$$\begin{aligned} \langle W_1 D, W_2 D \rangle_{HS} &= \sum_{i=1}^n \sum_{i'=1}^n (W_1 D)_{ii'} (W_2 D)_{i'i} = \\ &= \sum_{i=1}^n \sum_{i'=1}^n p_i e_i^1{}' M_1 e_{i'}^1 p_i e_{i'}^2{}' M_2 e_i^2 = \sum_{i=1}^n \sum_{i'=1}^n p_i p_{i'} \langle e_i^1, e_{i'}^1 \rangle_{M_1} \langle e_{i'}^2, e_i^2 \rangle_{M_2} \end{aligned}$$

De esta expresión se deduce que el producto escalar entre operadores es la suma de los productos escalares de los individuos de una tabla por los productos escalares de los individuos de la otra.

Por otra parte, tenemos que:

$$W_1 = Y^1 M_1 Y^{1'} \Rightarrow w_{ii'} = (W_1)_{ii'} = \sum_{j=1}^{p_1} m_j^1 y_{ij}^1 y_{i'j}^1$$

donde se deduce que:

$$(W_1 D)_{ii'} = p_{i'} \sum_{j=1}^{p_1} m_j^1 y_{ij}^1 y_{i'j}^1 \quad (W_2 D)_{ii'} = p_i \sum_{j=1}^{p_2} m_j^2 y_{ij}^2 y_{i'j}^2$$

Por ello:

$$\begin{aligned} \sum_{i=1}^n \sum_{i'=1}^n (W_1 D)_{ii'} (W_2 D)_{ii'} &= \sum_{i=1}^n \sum_{i'=1}^n p_{i'} \sum_{j=1}^{p_1} m_j^1 y_{ij}^1 y_{i'j}^1 p_i \sum_{j'=1}^{p_2} m_{j'}^2 y_{ij'}^2 y_{i'j'}^2 = \\ &= \sum_{i=1}^n \sum_{i'=1}^n \sum_{j=1}^{p_1} \sum_{j'=1}^{p_2} p_i p_{i'} m_j^1 m_{j'}^2 y_{ij}^1 y_{i'j}^1 y_{ij'}^2 y_{i'j'}^2 \end{aligned}$$

Reordenando los sumatorios, se tiene que:

$$\begin{aligned} \langle W_1 D, W_2 D \rangle_{HS} &= \sum_{j=1}^{p_1} \sum_{j'=1}^{p_2} m_j^1 m_{j'}^2 \sum_{i=1}^n p_i y_{ij}^1 y_{ij'}^2 \sum_{i'=1}^n p_{i'} y_{i'j}^1 y_{i'j'}^2 = \\ &= \sum_{j=1}^{p_1} \sum_{j'=1}^{p_2} m_j^1 m_{j'}^2 \text{Cov}^2(Y_j^1, Y_{j'}^2) \end{aligned}$$

Esto es la suma del cuadrado de las covarianzas de todas las variables de una tabla con las de la otra. Teniendo en cuenta la relación entre las X_j y las Y_j en cualquier tabla se deduce que:

$$\langle W_1 D, W_2 D \rangle_{HS} = \sum_{j=1}^{p_1} \sum_{j'=1}^{p_2} m_j^1 m_{j'}^2 \text{Cov}^2(X_j^1, X_{j'}^2)$$

Si ese valor fuese cero indicaría que las X_j^1 tienen covarianza nula con la $X_{j'}^2$. En el caso de que las tablas estuviesen reducidas esa expresión sería:

$$\sum_{j=1}^{p_1} \sum_{j'=1}^{p_2} m_j^1 m_{j'}^2 \text{Cov}^2(Z_j^1, Z_{j'}^2) = \sum_{j=1}^{p_1} \sum_{j'=1}^{p_2} m_j^1 m_{j'}^2 \text{Cor}^2(X_j^1, X_{j'}^2)$$

que es la suma del cuadrado de las correlaciones de todas las variables de una tabla con las de la otra. Tal y como se recoge en Braun (1973: 81-106) otra expresión del producto escalar entre operadores viene dada por:

$$\langle W_1 D, W_2 D \rangle_{HS} = \sum_{k=1}^{p_1} \sum_{l=1}^{p_2} \lambda_k^1 \lambda_l^2 \text{Cov}^2(F_k^1, F_l^2)$$

considerando que F_k^1 y F_k^2 son, respectivamente, los vectores propios de $W_1 D$ y $W_2 D$ y λ_k^1 y λ_k^2 los valores propios de $W_1 D$ y $W_2 D$ de rangos p_1 y p_2 .

Norma

La norma de $W_1 D$ viene dada como la distancia al origen y es la raíz cuadrada de la siguiente expresión:

$$\begin{aligned} \|W_1 D\|_{HS}^2 &= \langle W_1 D, W_1 D \rangle_{HS} = \text{Tr} (W_1 D W_1 D) = \\ &= \text{Tr} ((W_1 D)^2) = d^2 (W_1 D, 0) \end{aligned}$$

Teniendo en cuenta la expresión de los productos escalares entre individuos, la norma viene dada por:

$$\begin{aligned} \|W_1 D\|_{HS}^2 &= \langle W_1 D, W_1 D \rangle_{HS} = \text{Tr} ((W_1 D)^2) = \sum_{i=1}^n \sum_{i'=1}^n (W_1 D)_{ii'}^2 = \\ &= \sum_{i=1}^n \sum_{i'=1}^n p_i p_{i'} \langle e_i^1, e_{i'}^1 \rangle_{M_1}^2 \end{aligned}$$

Por tanto, se obtiene como la suma de los cuadrados de los productos escalares entre individuos de la tabla a la que representa el operador; de aquí, realizando las correspondientes operaciones, se obtiene:

$$\begin{aligned} \|W_1 D\|_{HS}^2 &= \text{Tr} ((W_1 D)^2) = \sum_{i=1}^n \sum_{i'=1}^n (W_1 D W_1 D)_{ii'} = \sum_{i=1}^n \sum_{i'=1}^n (W_1 D)_{ii'} (W_1 D)_{i'i} = \\ &= \sum_{i=1}^n \sum_{i'=1}^n p_{i'} \sum_{j=1}^{p_1} m_j^1 y_{ij}^1 y_{i'j}^1 p_i \sum_{j=1}^{p_1} m_j^1 y_{ij}^1 y_{i'j}^1 = \sum_{i=1}^n \sum_{i'=1}^n \sum_{j=1}^{p_1} \sum_{j'=1}^{p_1} p_{i'} p_i m_j^1 m_{j'}^1 y_{ij}^1 y_{i'j}^1 y_{ij'}^1 y_{i'j'}^1 = \\ &= \sum_{j=1}^{p_1} m_j^1 \sum_{i=1}^n p_i y_{ij}^1 y_{i'j}^1 \sum_{j'=1}^{p_1} m_{j'}^1 \sum_{i=1}^n p_i y_{ij'}^1 y_{i'j'}^1 = \\ &= \sum_{j=1}^{p_1} \sum_{j'=1}^{p_1} m_j^1 m_{j'}^1 \text{Cov}^2 (Y_j^1, Y_{j'}^1) = \sum_{j=1}^{p_1} \sum_{j'=1}^{p_1} m_j^1 m_{j'}^1 \text{Cov}^2 (X_j^1, X_{j'}^1) \end{aligned}$$

Se puede observar que las normas de los operadores dependen de las covarianzas entre las variables que los componen; ahora bien, si las

variables estuviesen reducidas tendríamos que dichas normas dependen de las correlaciones:

$$\|W_1 D\|_{HS}^2 = \sum_{j=1}^{p_1} \sum_{j'=1}^{p_1} m_j^1 m_{j'}^1 \text{Cov}^2(Z_j^1, Z_{j'}^1) = \sum_{j=1}^{p_1} \sum_{j'=1}^{p_1} m_j^1 m_{j'}^1 \text{Cor}^2(X_j^1, X_{j'}^1)$$

Al analizar la nube de individuos, en el ACP, se demostró que los vectores propios, asociados a los mayores valores propios del operador $W_1 D$ son las componentes principales F_k^1 . Es decir, se cumple para el eje k que $W_1 D F_k^1 = \lambda_k^1 F_k^1$. Además, la varianza de cada F_k^1 es la inercia explicada por el eje k y es igual al valor propio correspondiente. La inercia de la nube de individuos era:

$$\text{Tr}(W_1 D) = \sum_{k=1}^{p_1} \lambda_k^1$$

siendo λ_k^1 los p_1 valores propios, suponiendo que el rango de $W_1 D$ es p_1 .

Por tanto, y considerando las propiedades de las trazas y de los valores propios (Lavit, 1988: 80) se obtiene:

$$\|W_1 D\|_{HS}^2 = \text{Tr}((W_1 D)^2) = \text{Tr}(W_1 D W_1 D) = \sum_{k=1}^{p_1} (\lambda_k^1)^2 = \|V_1 M_1\|_{HS}^2$$

Distancia

El concepto de distancia entre operadores se define a partir de los conceptos de producto escalar y norma, y viene dada por:

$$\begin{aligned} d^2(W_1 D, W_2 D) &= \|W_1 D - W_2 D\|_{HS}^2 = \\ &= \|W_1 D - W_2 D\|_{HS}^2 = \|W_1 D\|_{HS}^2 + \|W_2 D\|_{HS}^2 - 2\|W_1 D W_2 D\|_{HS} = \\ &= \sum_{i=1}^n \sum_{i'=1}^n p_i p_{i'} \langle e_i^1, e_{i'}^1 \rangle_{M_1}^2 + \sum_{i=1}^n \sum_{i'=1}^n p_i p_{i'} \langle e_i^2, e_{i'}^2 \rangle_{M_2}^2 + \\ &- 2 \sum_{i=1}^n \sum_{i'=1}^n p_i p_{i'} \langle e_i^1, e_{i'}^1 \rangle_{M_1} \langle e_i^2, e_{i'}^2 \rangle_{M_2} = \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=1}^n \sum_{i'=1}^n p_i p_{i'} [\langle e_i^1, e_{i'}^1 \rangle_{M_1} - \langle e_i^2, e_{i'}^2 \rangle_{M_2}]^2 = \\
 &= \sum_{i=1}^n p_i \sum_{i'=1}^n p_{i'} [\langle e_i^1, e_{i'}^1 \rangle_{M_1} - \langle e_i^2, e_{i'}^2 \rangle_{M_2}]^2 = \\
 &= \sum_{i=1}^n p_i \sum_{i'=1}^n p_{i'} [(W_1)_{ii'} - (W_2)_{ii'}]^2 = \sum_{i=1}^n p_i \|W_1(e_i) - W_2(e_i)\|_{HS}^2
 \end{aligned}$$

Para cada individuo e_i se considera la distancia, al cuadrado, del mismo con el resto, entre las dos tablas de los productos escalares; es decir, el cambio de posición de un individuo de una tabla a otra. Por tanto, la distancia entre dos tablas se obtiene teniendo en cuenta todos los individuos.

A partir de:

$$\|W_1 D\|_{HS}^2 = \sum_{k=1}^{p_1} (\lambda_k^1)^2$$

la distancia toma la siguiente expresión:

$$d^2(W_1 D, W_2 D) = \sum_{k=1}^{p_1} (\lambda_k^1)^2 + \sum_{l=1}^{p_2} (\lambda_l^2)^2 - 2 \langle W_1 D, W_2 D \rangle_{HS}$$

Por tanto, la distancia depende del producto escalar entre operadores y de la norma de cada uno, siendo λ_k^1 y λ_l^2 los valores propios de $W_1 D$ y $W_2 D$ respectivamente, y las matrices de rangos p_1 y p_2 . Considerando la expresión de producto escalar ya mencionada y expresada por Braun (Escoufier, 1976: 170) tenemos que:

$$d^2(W_1 D, W_2 D) = \sum_{k=1}^{p_1} (\lambda_k^1)^2 + \sum_{l=1}^{p_2} (\lambda_l^2)^2 - 2 \sum_{k=1}^{p_1} \sum_{l=1}^{p_2} \lambda_k^1 \lambda_l^2 \text{Cov}^2(F_k^1, F_l^2)$$

Si los operadores $W_1 D$ y $W_2 D$ tienen los mismos vectores propios esa distancia será cero. En particular se puede demostrar que el operador que se deduce de la matriz de componentes principales de un estudio y el operador que se deduce de la matriz de datos original de ese estudio se confunden y su distancia es nula. Inversamente, si dos operadores están a distancia nula sus vectores propios son los mismos en la medida que todos los valores propios son diferentes para cada estudio. Si éste no es el caso,

los vectores propios asociados a los valores propios iguales solo difieren porque están rotados. Por ese motivo, y con la métrica definida, dos operadores asociados a diversos estudios se confunden si las distancias entre los individuos que los componen son las mismas en ambos estudios.

En definitiva, este producto escalar y esta norma permiten definir las distancias entre operadores y determinar si las nubes a las que representan son o no equivalentes. Así (Y^t, M_t, D) y $(Y^{t'}, M_{t'}, D)$ son equivalentes si los operadores asociados $W_t D$ y $W_{t'} D$ son idénticos o si son homotéticos, es decir, $W_t D = c W_{t'} D$, siendo c un número real positivo. Luego, dos estudios son equivalentes si las imágenes de los individuos, asociadas a las tablas W_t y $W_{t'}$ son también equivalentes y la imagen de uno se deduce de la imagen del otro.

Operador Normado $\frac{W_t D}{\|W_t D\|_{HS}}$

Si la norma de los operadores varía mucho de unos a otros, es posible tomar como elementos representativos de las tablas Y^1 y Y^2 los operadores normados, con el fin de eliminar la influencia de la misma en el estudio. Así, tomamos:

$$\frac{W_1 D}{\|W_1 D\|_{HS}} \quad \frac{W_2 D}{\|W_2 D\|_{HS}}$$

Seguiremos, para su descripción, el mismo esquema que para el operador $W_t D$.

Producto escalar

En este caso el producto escalar entre operadores tiene la siguiente expresión:

$$\begin{aligned} \left\langle \frac{W_1 D}{\|W_1 D\|_{HS}}, \frac{W_2 D}{\|W_2 D\|_{HS}} \right\rangle_{HS} &= \text{Tr} \left(\frac{W_1 D}{\|W_1 D\|_{HS}} \frac{W_2 D}{\|W_2 D\|_{HS}} \right) = \\ &= \frac{\text{Tr}(W_1 D W_2 D)}{\|W_1 D\|_{HS} \|W_2 D\|_{HS}} = \frac{\text{Tr}(W_1 D W_2 D)}{\sqrt{\text{Tr}(W_1 D)^2} \sqrt{\text{Tr}(W_2 D)^2}} \end{aligned}$$

Dicha expresión el nombre de coeficiente *RV* de Escoufier y es el coseno del ángulo entre esos dos vectores, siendo *O* el origen y $D \otimes D$ la métrica del espacio R^{n^2} :

$$\left\langle \frac{W_1 D}{\|W_1 D\|_{HS}}, \frac{W_2 D}{\|W_2 D\|_{HS}} \right\rangle_{HS} = \cos(O\vec{W}_1, O\vec{W}_2) = RV(W_1 D, W_2 D)$$

Norma

La norma de: $\frac{W_1 D}{\|W_1 D\|_{HS}}$

viene dada por la raíz cuadrada de: $\left\| \frac{W_1 D}{\|W_1 D\|_{HS}} \right\|_{HS}^2 = \frac{1}{\|W_1 D\|_{HS}^2} \|W_1 D\|_{HS}^2 = 1$

que es la distancia, al origen de ese operador, elevado al cuadrado.

Distancia

La distancia al cuadrado entre: $\frac{W_1 D}{\|W_1 D\|_{HS}}$ y $\frac{W_2 D}{\|W_2 D\|_{HS}}$ será:

$$\begin{aligned} &\left\| \frac{W_1 D}{\|W_1 D\|_{HS}} - \frac{W_2 D}{\|W_2 D\|_{HS}} \right\|_{HS}^2 = \\ &= \left\| \frac{W_1 D}{\|W_1 D\|_{HS}} \right\|_{HS}^2 + \left\| \frac{W_2 D}{\|W_2 D\|_{HS}} \right\|_{HS}^2 - 2 \left\langle \frac{W_1 D}{\|W_1 D\|_{HS}}, \frac{W_2 D}{\|W_2 D\|_{HS}} \right\rangle_{HS} = \\ &= 1 + 1 - 2 RV(W_1 D, W_2 D) = 2[1 - RV(W_1 D, W_2 D)] \end{aligned}$$

Con los operadores así transformados la distancia entre ellos depende del coseno del ángulo que forman, que es el coeficiente RV . Este coeficiente puede interpretarse como un coeficiente de correlación: si el coeficiente RV es uno la distancia será cero y, en este caso, existe una relación de equivalencia entre esos dos estudios, tal que la imagen euclídea de los individuos de la primera tabla (asociada a $W_1 D$) es proporcional a la imagen de los individuos de la segunda (asociado a $W_2 D$). La razón es:

$$\frac{\|W_1 D\|_{HS}}{\|W_2 D\|_{HS}}$$

Si el coeficiente¹⁵ RV es cero, la distancia es $\sqrt{2}$, que será la distancia máxima ya que:

$$\frac{\text{Tr}(W_1 D W_2 D)}{\|W_1 D\|_{HS} \|W_2 D\|_{HS}} = 0 \Rightarrow \text{Tr}(W_1 D W_2 D) = \langle W_1 D, W_2 D \rangle_{HS} = 0$$

En este caso, las variables de la tabla X^2 tienen covarianza nula (o correlación si están centradas y reducidas) con las de la tabla X^1 .

Dado que cada operador: $\frac{W_t D}{\|W_t D\|_{HS}}$

está a una distancia unitaria del origen (norma unitaria), el espacio R^{n^2} de los operadores:

$$\frac{W_t D}{\|W_t D\|_{HS}}$$

se asemeja al espacio de variables cuando se realiza un ACP *normado*. La imagen de cada operador se situará sobre una esfera de radio uno y centro el origen y la imagen de los operadores de las tablas en un espacio de dos dimensiones se interpreta de manera análoga al círculo de correlaciones.

¹⁵ El coeficiente RV es un producto escalar entre operadores semi-positivos y el resultado siempre es positivo o nulo. Por ello el valor varía entre uno y cero. Dado que es el coseno del ángulo entre operadores, esos ángulos sólo pueden ser agudos (Lavit, 1988: 86).

2.2.2 Operador $V_t M$. Operador normado $\frac{V_t M}{\|V_t M\|_{HS}}$

El operador $V_t M$ es un elemento del conjunto de aplicaciones lineales M -simétricas positivas del espacio E , y que se expresa $L(E, E)$. Dicho operador no está centrado, ya que es el resultado del producto escalar entre variables. En este conjunto de aplicaciones se puede definir un producto escalar que le atribuye una estructura de espacio de Hilbert-Schmidt. Considerando dos tablas de datos X^1 y X^2 (es decir, $t = 1, 2$), ambas observadas sobre las mismas p variables pero distintos individuos en cada una de ellas, y las matrices centradas Y^1 y Y^2 derivadas, respectivamente, de las tablas anteriores, se definen:

- $V_1 M = Y^1 D_1 Y^1 M$, siendo D_1 la matriz de ponderaciones de los distintos individuos (e_i^1) de Y^1 (siendo los elementos de la diagonal p_i^1) y M la matriz de pesos de las p variables (Y_j^1) que es común a las dos tablas. $V_1 M$ es de orden $p \times p$ siendo n_1 el número de individuos de Y^1 . El triplete (Y^1, M, D_1) define el estudio.
- $V_2 M = Y^2 D_2 Y^2 M$, siendo D_2 la matriz de ponderaciones de los individuos (e_i^2) de Y^2 (siendo los elementos de la diagonal p_i^2) y M la matriz de pesos de las variables (Y_j^2), puesto que todas las variables tienen el mismo peso $m_j = 1$ para todo p . En este caso tenemos (Y^2, M, D_2) , siendo $V_2 M$ de orden $p \times p$ y n_2 el número de individuos de Y^2 .

Ambos conceptos permiten comparar las nubes de variables que se deducen de las tablas de datos originales (generalmente porque los variables son los elementos comunes y los individuos son los que varían de una tabla a otra, aunque también pueden ser los mismos). La matriz M es común a las dos tablas pues las variables son comunes a ambas y por consiguiente el peso otorgado a cada una de ellas, es el mismo.

Producto escalar

El producto escalar de Hilbert-Schmidt entre esos dos operadores se define como:

$$\langle V_1M, V_2M \rangle_{HS} = \text{Tr} (V_1MV_2M)$$

y es la covarianza vectorial $\text{Cov}(V_1M, V_2M)$. El producto escalar definido es un producto escalar entre los operadores V_1 y V_2 de R^{p^2} tomando como métrica la matriz $M \otimes M$. Se puede demostrar que:

$$\begin{aligned} \langle V_1M, V_2M \rangle_{HS} &= \text{Tr} (V_1MV_2M) = \sum_{j=1}^p (V_1MV_2M)_{jj} = \\ &= \sum_{j=1}^p \sum_{i=1}^p (V_1M)_{ji} (V_2M)_{ij} \end{aligned}$$

Dado que:

$$(V_1M)_{ji} = m_j \sum_{i=1}^{n_1} p_i^1 y_{ij}^1 y_{ij}^1 = Y_j^1 D^1 Y_j^1 = \text{Cov}(Y_j^1, Y_j^1) = \text{Cov}(X_j^1, X_j^1)$$

$$(V_2M)_{ij} = m_j \sum_{i=1}^{n_2} p_i^2 y_{ij}^2 y_{ij}^2 = Y_j^2 D^2 Y_j^2 = \text{Cov}(Y_j^2, Y_j^2) = \text{Cov}(X_j^2, X_j^2)$$

se cumple que:

$$\begin{aligned} \langle V_1M, V_2M \rangle_{HS} &= \sum_{j=1}^p \sum_{j'=1}^p m_j \sum_{i=1}^{n_1} p_i^1 y_{ij}^1 y_{ij}^1 m_{j'} \sum_{i'=1}^{n_2} p_{i'}^2 y_{i'j'}^2 y_{i'j'}^2 = \\ &= \sum_{j=1}^p \sum_{j'=1}^p m_j Y_j^1 D^1 Y_j^1 m_{j'} Y_{j'}^2 D^2 Y_{j'}^2 = \sum_{j=1}^p \sum_{j'=1}^p m_j m_{j'} \langle Y_j^1, Y_{j'}^1 \rangle_{D_1} \langle Y_{j'}^2, Y_j^2 \rangle_{D_2} \end{aligned}$$

Es decir, el producto escalar entre los operadores asociados a las nubes de variables, es igual a la suma de las covarianzas de una tabla por las covarianzas de la otra tabla.

Haciendo transformaciones similares a las realizadas para los individuos es posible demostrar, tal y como se deduce de la siguiente expresión, que el producto escalar entre dos operadores es la suma de los productos de las covarianzas (correlaciones si las variables están reducidas y centradas)

de dos variables de una tabla por las covarianzas de las mismas variables en la otra tabla:

$$\langle V_1 M, V_2 M \rangle_{HS} = \sum_{j=1}^p \sum_{j'=1}^p m_j m_{j'} \text{Cov}(Y_j^1, Y_{j'}^1) \text{Cov}(Y_{j'}^2, Y_j^2)$$

Norma

La norma de $V_1 M$ será igual a:

$$\|V_1 M\|_{HS}^2 = \sum_{j=1}^p \sum_{j'=1}^p m_j m_{j'} \langle Y_j^1, Y_{j'}^1 \rangle^2 = \sum_{j=1}^p \sum_{j'=1}^p m_j m_{j'} \text{Cov}^2(Y_j^1, Y_{j'}^1)$$

expresión que viene dada por la suma de las covarianzas al cuadrado de las variables de la tabla y verifica:

$$\|V_1 M\|_{HS}^2 = \langle V_1 M, V_1 M \rangle_{HS} = \text{Tr}((V_1 M)^2) = d^2(\|V_1 M\|_{HS}^2, 0)$$

siendo ésta la distancia al cuadrado del operador $V_1 M$ al origen.

En el ACP cuando se analiza la nube de variables se demostró que los vectores propios, asociados a los mayores valores propios, del operador $V_1 M$ son las componentes principales G_k . Es decir, se cumple para el eje k que: $V_1 M G_k^1 = \lambda_k^1 G_k^1$ y la cantidad de información explicada por el eje k es igual al valor propio correspondiente. Por ello, para toda la nube de variables, se tiene que:

$$\text{Tr}(V_1 M) = \sum_{k=1}^{p_1} \lambda_k^1$$

siendo λ_k^1 los p_1 valores propios, dado que el rango de $V_1 M$ es p_1 . Teniendo en cuenta las propiedades de las trazas y de los valores propios, como se indicó para los operadores $W_t D$, se tiene que:

$$\|V_1 M\|^2 = \text{Tr}((V_1 M)^2) = \text{Tr}(V_1 M V_1 M) = \sum_{k=1}^{p_1} (\lambda_k^1)^2$$

Distancia

La distancia entre los operadores V_1M y V_2M se define como:

$$\begin{aligned}
 d^2(V_1M, V_2M) &= \|V_1M\|_{HS}^2 + \|V_2M\|_{HS}^2 - 2\langle V_1M, V_2M \rangle_{HS} = \\
 &= \sum_{j=1}^p \sum_{j'=1}^p m_j m_{j'} \langle Y_j^1, Y_{j'}^1 \rangle_{D_1}^2 + \sum_{j=1}^p \sum_{j'=1}^p m_j m_{j'} \langle Y_j^2, Y_{j'}^2 \rangle_{D_2}^2 + \\
 &\quad - 2 \sum_{j=1}^p \sum_{j'=1}^p m_j m_{j'} \langle Y_j^1, Y_{j'}^1 \rangle_{D_1} \langle Y_j^2, Y_{j'}^2 \rangle_{D_2} = \\
 &= \sum_{j=1}^p \sum_{j'=1}^p m_j m_{j'} [\langle Y_j^1, Y_{j'}^1 \rangle_{D_1} - \langle Y_j^2, Y_{j'}^2 \rangle_{D_2}]^2 = \\
 &= \sum_{j=1}^p m_j \sum_{j'=1}^p m_{j'} [(V_1)_{jj'} - (V_2)_{jj'}]^2 = \sum_{j=1}^p m_j \|V_1(Y_j) - V_2(Y_j)\|_{HS}^2
 \end{aligned}$$

que es, para cada variable, la suma de las diferencias (al cuadrado) de una tabla a otra, de sus covarianzas con las otras variables de la tabla correspondiente. En el caso de variables centradas y reducidas, Z_j , las covarianzas entre las Y_j y las $Y_{j'}$ se transformaría en covarianzas entre las Z_j y las $Z_{j'}$. En ambos casos, la distancia refleja, para cada variable, el cambio de posición de una tabla a la otra.

Estas distancias entre operadores permiten determinar si las nubes a las que representan son o no equivalentes, es decir, en este caso se trata de comparar V_1M y V_2M . Cuando: $V_1M = c V_2M$, siendo c un número real positivo, entonces se da dicha equivalencia, y una de las nubes se deduce de la otra porque es la misma o es proporcional.

Operador Normado $\frac{V_t M}{\|V_t M\|_{HS}}$

De manera muy breve, en cuanto que es similar al caso de los operadores normados $W_t D$ vamos a definir el producto escalar (coeficientes RV), la norma y la distancia entre ellos.

Producto escalar

El producto escalar entre operadores normados:

$$\begin{aligned} \left\langle \frac{V_1M}{\|V_1M\|_{HS}}, \frac{V_2M}{\|V_2M\|_{HS}} \right\rangle &= \text{Tr} \left(\frac{V_1M}{\|V_1M\|_{HS}} \frac{V_2M}{\|V_2M\|_{HS}} \right) = \frac{\text{Tr}(V_1MV_2M)}{\|V_1M\|_{HS} \|V_2M\|_{HS}} = \\ &= \frac{\text{Tr}(V_1MV_2M)}{\sqrt{\text{Tr}(V_1M)^2} \sqrt{\text{Tr}(V_2M)^2}} = \cos(O\vec{V}_1, O\vec{V}_2) = RV(V_1M, V_2M) \end{aligned}$$

siendo O el origen y $M \otimes M = I$ la métrica del espacio R^{p^2} .

Norma

La norma de los operadores así definidos es unitaria tal y como se demuestra a continuación:

$$d^2 \left(\frac{V_1M}{\|V_1M\|_{HS}}, 0 \right) = \left\| \frac{V_1M}{\|V_1M\|_{HS}} \right\|_{HS}^2 = \frac{1}{\|V_1M\|_{HS}^2} \|V_1M\|_{HS}^2 = 1$$

Distancia

La distancia entre los operadores normados se define como:

$$\begin{aligned} d^2 \left(\frac{V_1M}{\|V_1M\|_{HS}}, \frac{V_2M}{\|V_2M\|_{HS}} \right) &= \left\| \frac{V_1M}{\|V_1M\|_{HS}} \right\|_{HS}^2 + \left\| \frac{V_2M}{\|V_2M\|_{HS}} \right\|_{HS}^2 + \\ &- 2 \left\langle \frac{V_1M}{\|V_1M\|_{HS}}, \frac{V_2M}{\|V_2M\|_{HS}} \right\rangle_{HS} = 1+1 - 2 RV(V_1M, V_2M) = \\ &= 2 [1 - RV(V_1M, V_2M)] \end{aligned}$$

Con los operadores así transformados la distancia entre ellos depende del coseno del ángulo que forman, que es el coeficiente RV , y que puede interpretarse como un coeficiente de correlación. Cuando los estudios son equivalentes, las matrices de covarianzas son proporcionales y el coeficiente

RV entre operadores será uno. Si dicho coeficiente es cero la suma de los productos de las covarianzas (correlaciones si las variables están reducidas y centradas) de las variables de una tabla con las dos mismas de la otra será nula y la distancia entre las tablas es máxima e igual a $\sqrt{2}$. Además como cada operador:

$$\frac{V_t M}{\|V_t M\|}$$

está a una distancia unitaria del origen, el espacio R^p de esos operadores se asemeja al espacio de variables obtenido cuando se realiza un ACP *normado* y que ya ha sido desarrollado ampliamente.

Una vez abordadas estas cuestiones generales estamos en condiciones de desarrollar los métodos *Statis* y *Statis dual*.

2.3 Métodos *Statis* y *Statis dual*

El desarrollo teórico de las dos versiones que componen la metodología STATIS se va realizar de manera simultánea, dada la similitud del procedimiento entre ambas, haciendo referencia en cada uno de los apartados, en primer lugar, al *Statis* y, en segundo lugar, al *Statis dual*, evitando de esta manera la repetición de algunos desarrollos comunes.

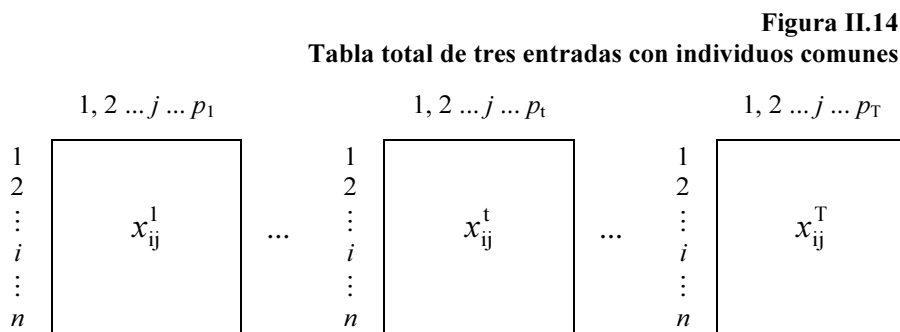
Para una adecuada exposición de dichos métodos, es necesario comenzar abordando ciertas cuestiones relativas al tipo de datos, y a las tablas, que son objeto de estudio, así como los objetivos más generales perseguidos en cada caso. Seguidamente se exponen los métodos, estructurados en cuatro fases o etapas, cada una de las cuales tiene un objetivo preciso. Por último, se hacen una serie de consideraciones que facilitan la interpretación de los resultados obtenidos.

2.3.1 Los datos y descripción general

Recordemos que el STATIS, según lo expuesto en el capítulo I, en sus dos versiones, es una técnica exploratoria de Análisis de Datos cuantitativos para tablas de tres entradas (o lo que es lo mismo para una sucesión de tablas de dos entradas), cuyo objetivo fundamental es la búsqueda de la estructura común al conjunto de tablas que componen la tabla total, así como su descripción. Las tablas a las que se aplica corresponden a los siguientes casos:

- ❶ – La observación de un conjunto de n individuos en T situaciones (ocasiones o momentos de tiempo distintos).
- En cada situación se miden p_t variables o características sobre cada uno de los n individuos.
- Las p_t variables pueden ser las mismas o diferentes en cada situación, pero los individuos tienen que ser los mismos.

Ello da lugar a una supertabla o tabla total de tres entradas, que se presenta en la figura II.14, formada por n filas (los individuos) y $\sum_{t=1}^T p_t$ columnas (las variables) agrupadas en t tablas ($t = 1, 2 \dots T$) de n filas y p_t columnas cada una.



El elemento característico de la tabla a analizar es x_{ij}^t :

$$i = 1, 2 \dots n$$

$$j = 1, 2 \dots p_t$$

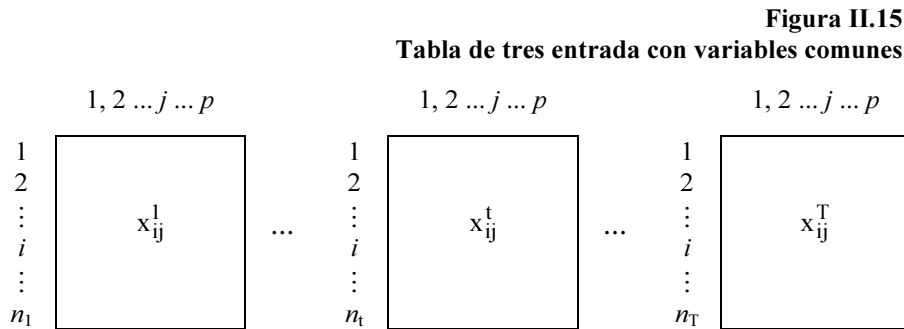
$$t = 1, 2 \dots T$$

y cada una de las tablas se representa en el estudio por (X^t, M_t, D) , siendo X^t la matriz de datos t , M_t la matriz diagonal que recoge el peso de las variables de cada tabla t , y D la matriz diagonal de pesos de los individuos, que es común para todas las tablas.

El análisis de esta tabla de tres entradas se realiza mediante la versión *Statis*, puesto que el objetivo es estudiar la nube de individuos que se deduce de cada una de las tablas, determinando si éstos cambian de posición o bien mantienen las distancias de unas tablas a otras.

- ② – La observación de un conjunto de p variables en T situaciones (ocasiones ó momentos distintos).
 - Para cada situación se han observado las p variables sobre n_t individuos.
 - Los individuos pueden ser los mismos o diferentes, pero las variables son las mismas.

La tabulación de esta información da lugar a la tabla de tres entradas formada por $\sum_{t=1}^T n_t$ filas y p columnas agrupadas en t tablas ($t = 1, 2 \dots T$) cada una de n_t filas y p columnas. Su representación se expresa en la figura II.15:



De la misma forma que en el caso anterior, el elemento característico es x_{ij}^t con:

$$i = 1, 2 \dots n_t$$

$$j = 1, 2 \dots p$$

$$t = 1, 2 \dots T$$

y la tabla se expresa por (X^t, M, D_t) , donde X^t es la matriz de datos t , M es la matriz diagonal de pesos de las variables, común a todas las tablas, y D_t la matriz que recoge los pesos de los individuos, una para cada tabla.

En este caso se trata de estudiar la tabla descrita, lo cual se realiza a través del *Statis dual*; ello va a permitir el análisis de las nubes de variables que se deducen de las tablas, y las correlaciones entre las mismas (o covarianzas según el caso), comprobando si varían de unas tablas a otras.

Remitiéndonos a lo ya indicado (GERI, 1996: 68, 86), cuando todas las tablas están formadas por los mismos individuos y por las mismas variables, pueden aplicarse los dos métodos, lo que enriquece las conclusiones obtenidas.

Aunque ya se ha hecho referencia a ello, conviene volver a señalar que el tratamiento de cualquiera de las dos "supertablas" antes consideradas por un método propio para tablas de doble entrada no sería óptimo. Por ejemplo, un ACP de cualquiera de las mismas no tendría en cuenta la pertenencia de las variables a distintos grupos y seguramente presentaría problemas de correlación entre variables (efecto talla). Por otra parte, los ACP de las tablas individuales que forman cada "supertabla" nos llevaría a multitud de

representaciones gráficas difíciles de comparar ya que los ejes obtenidos a partir de cada una de ellas son diferentes. El STATIS con sus dos versiones resuelve de manera satisfactoria esos dos problemas.

2.3.2 Etapas de desarrollo

Los métodos *Statis* y *Statis dual* están estructurados en una serie de etapas que van a ser descritas y desarrolladas ampliamente a continuación, especificando en cada una las características diferenciales de cada versión.

a. Estudio de la interestructura

Esta etapa tiene como objeto el estudio de las tablas de doble entrada que componen la tabla total desde un punto de vista global, estableciendo comparaciones entre las mismas. Para ello es necesario definir el objeto representativo (operador) de cada una de ellas, un producto escalar y una distancia entre ellas. La imagen euclídea de los objetos representativos de dichas tablas, en la que cada uno está representado por un punto de dicha imagen, permitirá determinar si dos tablas son semejantes o no. La proximidad, tal y como luego explicaremos que debe ser entendida, se interpreta como semejanza entre tablas e indicará la existencia de un comportamiento común de los individuos en las mismas si se trata de la interestructura del *Statis*, o un comportamiento común de las variables, si es el caso del *Statis dual*. Si las tablas están alejadas, los individuos o las variables, en el *Statis* o en el *Statis dual* respectivamente, se comportan de manera diferente en cada tabla y no hay una estructura común de distancias entre individuos en el primer caso, o de correlaciones (o covarianzas) entre variables en el segundo.

Comenzamos el desarrollo de esta etapa definiendo los objetos representativos de cada tabla, que son el producto escalar y las distancias.

— En el caso del *Statis*, el objeto representativo de cada tabla es la matriz de productos escalares entre individuos, considerando el peso de cada uno en la matriz D . En este caso, el operador asociado a cada tabla es $W_t D$; si las variables sólo están centradas (para que el origen de cada operador coincida con su centro de gravedad y la imagen euclídea de los individuos refleje las direcciones de máximo alargamiento de la nube) el operador toma la forma tal que $W_t D = Y^t M Y^{t'}$. Si las variables están medidas en distintas unidades, o la varianza entre ellas es muy distinta, debemos considerarlas centradas y reducidas, siendo el operador $W_t D = Z^t M Z^{t'}$. En el caso de que las normas de los operadores sean muy diferentes, el operador representativo de cada tabla será:

$$\frac{W_t D}{\|W_t D\|_{HS}}$$

con el fin de equilibrar la influencia de todas las tablas en el análisis y que cada una intervenga según su forma y no según su tamaño.

— Cuando las variables son las mismas, que es el caso del *Statis dual*, el operador representativo es $V_t M$ que es la matriz de varianzas–covarianzas entre las variables (si estas están centradas) o la de correlaciones (si estas están centradas y reducidas) y por ello $V_t M$ no está centrado. Si las normas de los operadores son muy diferentes es aconsejable tomar:

$$\frac{V_t M}{\|V_t M\|_{HS}}$$

para que la construcción del compromiso no esté influenciada por los operadores de mayor norma (que será la tabla de datos con las variables de mayor covarianza, o correlación, según el caso).

Dado que ciertas propiedades difieren según se consideren los operadores normados o no, esta distinción es la que se va a tener en cuenta para el desarrollo de esta etapa y de todo el método.

Operadores normados

El producto escalar y la norma entre operadores varían según el objeto representativo de la tabla de datos que corresponda.

— El método *Statis* considera los operadores: $\frac{W_t D}{\|W_t D\|_{HS}}$

y el producto escalar se define como:

$$\left\langle \frac{W_t D}{\|W_t D\|_{HS}}, \frac{W_t D}{\|W_t D\|_{HS}} \right\rangle_{HS} = \cos(O\vec{W}_t, O\vec{W}_t) = RV(W_t D, W_t D)$$

siendo O el origen.

Su norma, que es la distancia al origen, es la raíz de la expresión:

$$d^2\left(\frac{W_t D}{\|W_t D\|_{HS}}, 0\right) = \left\| \frac{W_t D}{\|W_t D\|_{HS}} \right\|_{HS}^2 = 1$$

Cuando se trata de los operadores: $\frac{V_t M}{\|V_t M\|_{HS}}$

en el *Statis dual*, el producto escalar toma el valor:

$$\left\langle \frac{V_t M}{\|V_t M\|_{HS}}, \frac{V_t M}{\|V_t M\|_{HS}} \right\rangle = \cos(O\vec{V}_t, O\vec{V}_t) = RV(V_t M, V_t M)$$

y su norma, al cuadrado: $d^2\left(\frac{V_t M}{\|V_t M\|_{HS}}, 0\right) = \left\| \frac{V_t M}{\|V_t M\|_{HS}} \right\|_{HS}^2 = 1$

— La distancia entre operadores normados depende de su producto escalar, que es el coeficiente RV y coincide con el coseno del ángulo que forman, de tal manera que, para el *Statis*, será:

$$d^2\left(\frac{W_t D}{\|W_t D\|_{HS}}, \frac{W_t D}{\|W_t D\|_{HS}}\right) = 2[1 - RV(W_t D, W_t D)]$$

y para el *Statis dual*:

$$d^2\left(\frac{V_t M}{\|V_t M\|_{HS}}, \frac{V_{t'} M}{\|V_{t'} M\|_{HS}}\right) = 2 [1 - RV(V_t M, V_{t'} M)]$$

En este caso la distancia entre operadores normados depende de la dimensión y forma de la nube, y no del tamaño, puesto que éste ha sido eliminado al emplear operadores normados.

Con relación a estos conceptos son necesarias algunas precisiones:

- La dimensión se refiere al número de valores propios no nulos de la matriz de productos escalares. Si dos nubes son de diferente dimensión significa que el número de factores que resumen esas nubes es distinto y el rango de las matrices también es diferente.
- La forma de la nube se refiere a la repartición de los valores propios de la matriz de productos escalares. Así, si la dimensión es la misma, la distancia entre nubes puede deberse a que tienen distinta forma. Una nube redonda será aquella cuyos valores propios son del mismo orden de grandeza, mientras que una nube alargada será la que presente un primer valor propio mucho mayor que el resto y los demás son pequeños con relación al primero.
- Si ambas tienen una forma similar, la repartición de los individuos (o de las variables) en el interior de la nube es lo que varía de una nube a otra. Estas son las nubes que deben compararse mediante el método *Statis* o el método *Statis dual*, según el caso. Por ello el paso previo para su aplicación es el análisis de los operadores asociados a cada tabla de doble entrada (generalmente W_t ó V_t) mediante la descomposición en sus valores propios. Al analizar cada una de ellos, si todas las variables están centradas, los porcentajes de inercia explicada coinciden con los porcentajes de las matrices de varianzas-covarianzas; si las variables están reducidas, con los de las de correlación.

— Una vez abordadas estas cuestiones sobre los operadores, se va a desarrollar el procedimiento propiamente dicho.

Partimos de la matriz S , que es la formada por los productos escalares entre operadores normados. Es una matriz simétrica, semi-definida positiva y de orden $T \times T$ que viene dada por los siguientes términos:

$$* \textit{Statis} \Rightarrow s_{tt'} = \text{RV}(W_t D, W_{t'} D) = \frac{\text{Tr}(W_t D W_{t'} D)}{\|W_t D\|_{\text{HS}} \|W_{t'} D\|_{\text{HS}}}$$

expresión que se puede considerar como el producto escalar, en R^{n^2} , entre operadores W_t y $W_{t'}$, donde la métrica del espacio (el peso de los n^2 elementos de las filas) viene dada por la matriz $D \otimes D$.

$$* \textit{Statis dual} \Rightarrow s_{tt'} = \text{RV}(V_t M, V_{t'} M) = \frac{\text{Tr}(V_t M V_{t'} M)}{\|V_t M\|_{\text{HS}} \|V_{t'} M\|_{\text{HS}}}$$

que es el producto escalar en R^{p^2} entre operadores V_t y $V_{t'}$, la métrica del espacio (el peso de los p^2 elementos de las filas) viene dada por la matriz $M \otimes M$.

Por tanto, la matriz S viene dada por la siguiente expresión:

$$S = \begin{bmatrix} 1 & s_{12} & \cdots & s_{1T} \\ s_{21} & 1 & \cdots & s_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ s_{T1} & \cdots & \cdots & 1 \end{bmatrix}$$

La matriz S tiene características similares a la matriz V en el análisis de la nube N_I (tabla II.1: 125), por lo que para obtener los ejes sobre los que se va a obtener la imagen euclídea de las tablas es necesario considerar una matriz que recoja la importancia de cada una de ellas. Así vamos a denominar π_t al peso de cada tabla y Δ a la matriz que recoge dichos pesos, tal que:

$$\Delta = \begin{bmatrix} \pi_1 & 0 & \cdots & 0 \\ & \pi_2 & & \\ & & \ddots & \\ & & & \pi_t \\ & & & & \ddots \\ 0 & \cdots & 0 & & \pi_T \end{bmatrix}$$

Ambas versiones suponen que todas las tablas tienen la misma importancia en el análisis, por lo que los π_t son todos iguales a uno, y por tanto, $\Delta = I$ es la matriz identidad de orden $T \times T$. En este caso la inercia total, respecto al origen, viene dada como:

$$\sum_{t=1}^T \pi_t d^2\left(\frac{W_t D}{\|W_t D\|_{HS}}, 0\right) = \sum_{t=1}^T \pi_t \frac{\text{Tr}((W_t D)^2)}{\|W_t D\|_{HS}^2} = \text{Tr}(S\Delta) = T \text{ (Statis)}$$

$$\sum_{t=1}^T \pi_t d^2\left(\frac{V_t M}{\|V_t M\|_{HS}}, 0\right) = \sum_{t=1}^T \pi_t \frac{\text{Tr}((V_t M)^2)}{\|V_t M\|_{HS}^2} = \text{Tr}(S\Delta) = T \text{ (Statis dual)}$$

Podrían tomarse distintos pesos y así ponderar más las tablas relativas a los momentos del tiempo más recientes, aunque estas cuestiones requieren un estudio más profundo.

Como en un ACP, para obtener la imagen euclídea de las distintas tablas el primer paso es la obtención de los ejes que hacen máxima la inercia sobre ellos proyectados, descomponiendo en valores y vectores propios la matriz $S\Delta$.

Estableciendo una equivalencia con dicho análisis, las coordenadas de las tablas (variables) en los nuevos ejes vienen dadas por su coeficiente de correlación con el factor correspondiente. Así denominando μ_k al valor propio asociado al vector propio γ_k de la matriz $S\Delta$ se tiene que:

$$S\Delta \gamma_k = \mu_k \gamma_k$$

γ_k es el eje correspondiente a la componente principal k , que denominaremos δ_k , de tal manera que:

$$\gamma_k = \begin{bmatrix} \gamma_{k1} \\ \gamma_{k2} \\ \vdots \\ \gamma_{kT} \end{bmatrix}$$

Por tanto, la coordenada de la tabla t con el eje k , será el coeficiente de correlación del operador representativo de dicha tabla con la componente δ_k , es decir:

$$\xi_{kt} = \sqrt{\mu_k} \gamma_{kt}$$

siendo:

$$\xi_k = \begin{bmatrix} \xi_{k1} \\ \xi_{k2} \\ \vdots \\ \xi_{kT} \end{bmatrix}$$

el vector que recoge las coordenadas para todas las T tablas en el eje k . La inercia de ξ_k y de δ_k será:

$$d^2(\xi_k, 0) = \xi_k' \Delta \xi_k = \mu_k = d^2(\delta_k, 0)$$

ya que μ_k es el vector propio correspondiente a la componente δ_k teniendo en cuenta la relación entre δ_k y ξ_k demostrada al considerar las relaciones de transición en un ACP *normado*. Al desarrollar dicho análisis también se demostró que la primera componente, en este caso δ_1 , es la que hace máxima la suma de los coeficientes de correlación al cuadrado de cada una de las tablas.

— La imagen euclídea que permite visualizar mejor las proximidades entre los operadores representativos de las tablas de datos es la que se obtiene considerando únicamente dos dimensiones. Por tanto, la inercia explicada por este primer plano factorial es $\mu_1 + \mu_2$ y la proporción de inercia que representa será:

$$\frac{\mu_1 + \mu_2}{T}$$

La representación así obtenida tiene una interpretación similar al círculo de correlaciones, por lo que el origen no es el centro de gravedad. Los puntos próximos a los extremos del círculo estarán bien representados (y por tanto las tablas a las que se refieren), y los ángulos entre ellos darán una visión correcta del ángulo real. Los puntos con normas próximas al origen estarán mal representados, siendo su estructura diferente a la del conjunto de las tablas. Esta representación es óptima en el sentido de que es la mejor aproximación posible de la distancia de Hilbert-Schmidt entre los objetos representados (GERI, 1996: 74).

Dado que $S\Delta$ es simétrica y definida positiva, admite un primer valor con todas las componentes del mismo signo (Teorema de Frobenius (GERI, 1996: 75)). Así, es posible tomar un primer vector propio con todas las componentes positivas, y las diferencias entre tablas se ponen de manifiesto respecto al segundo eje, de tal forma que los ángulos entre los operadores que representan dichas tablas son agudos y la representación en dos dimensiones obtenida presenta un aspecto similar al efecto talla. Por este motivo puede ser interesante el análisis de la tabla de coeficientes RV pero centrada previamente. Este análisis se presenta a continuación.

La Interestructura centrada

En este caso se considera la matriz de productos escalares centrada \tilde{S} (en filas y en columnas) con el objeto de que los autovalores obtenidos, además de la inercia, representen la varianza de las componentes. El elemento característico es: $\tilde{s}_{tt'} = s_{tt'} - s_{t.} - s_{.t'} + s_{..}$, tal que:

$$s_{t.} = \sum_{t'=1}^T \pi_{t'} s_{tt'} \quad s_{.t'} = \sum_{t=1}^T \pi_t s_{tt'}$$

siendo:

$$s_{..} = \sum_{t=1}^T \sum_{t'=1}^T \pi_t \pi_{t'} s_{tt'}$$

De esta forma, el origen coincide con el centro de gravedad de la nube y la representación obtenida permite interpretar los puntos en términos de distancia al mismo (que coincide con el operador asociado al compromiso WD en el *Statis* y VM en el *Statis dual*, estableciendo distancias y proximidades entre ellos. Por otra parte, es posible también obtener una partición del conjunto de tablas mediante un método de clasificación basado en las distancias que se deducen de los productos escalares entre objetos. También podrían explicarse las posiciones de las tablas sobre los ejes mediante el cálculo de las correlaciones de dicho eje con las variables cuantitativas que forman las diferentes tablas.

Como en la mayoría de las ocasiones el análisis de la interestructura se restringe al análisis de una imagen en dos dimensiones, el análisis centrado de ésta es una ayuda suplementaria de gran valor para la interpretación de la posición de los puntos relativos a cada operador.

Operadores no normados

Cuando las normas de los operadores son muy diferentes no es aconsejable tomar este tipo de operadores, ya que los que representen a las nubes de mayor tamaño tendrán más influencia en el análisis. De todos modos, aunque las normas sean similares la situación más frecuente es considerar operadores normados siguiendo el proceso anteriormente descrito.

Las características de estos operadores son:

— El producto escalar y la norma. Al considerar los operadores $W_t D$ se tiene que:

$$\langle W_1 D, W_2 D \rangle_{HS} = \sum_{i=1}^n \sum_{i'=1}^n p_i p_{i'} \langle e_i^1, e_{i'}^1 \rangle_{M_1} \langle e_{i'}^2, e_i^2 \rangle_{M_2}$$

y la norma:

$$\|W_t D\|_{HS}^2 = \sum_{j=1}^{p_1} \sum_{j'=1}^{p_1} m_j^t m_{j'}^t \text{Cov}^2(X_j^t, X_{j'}^t) \text{ para variables centradas}$$

$$\|W_t D\|_{HS}^2 = \sum_{j=1}^{p_1} \sum_{j'=1}^{p_1} m_j^t m_{j'}^t \text{Cor}^2(X_j^t, X_{j'}^t) \text{ para variables centradas y reducidas}$$

Por otra parte, para los operadores $V_t M$ tendremos, con variables centradas:

$$\langle V_t M, V_{t'} M \rangle_{HS} = \sum_{j=1}^p \sum_{j'=1}^p m_j m_{j'} \text{Cov}(X_j^t, X_{j'}^t) \text{Cov}(X_{j'}^{t'}, X_j^{t'})$$

y con variables centradas y reducidas:

$$\langle V_t M, V_{t'} M \rangle_{HS} = \sum_{j=1}^p \sum_{j'=1}^p m_j m_{j'} \text{Cor}(X_j^t, X_{j'}^t) \text{Cor}(X_{j'}^{t'}, X_j^{t'})$$

La norma de dicho operador viene dada, respectivamente, por:

$$\|V_t M\|_{HS}^2 = \sum_{j=1}^p \sum_{j'=1}^p m_j m_{j'} \text{Cov}^2(X_j^t, X_{j'}^t)$$

y:

$$\|V_t M\|_{HS}^2 = \sum_{j=1}^p \sum_{j'=1}^p m_j m_{j'} \text{Cor}^2(X_j^t, X_{j'}^t)$$

De las expresiones anteriores se deduce que la norma de un operador depende de las varianzas, o de las correlaciones, de las variables que intervienen en el estudio: normas muy diferentes entre dos de ellos indican que la relación global entre las variables que los forman es muy distinta en uno y en otro.

— La distancia entre operadores vendrá dada como:

$$d^2(W_t D, W_{t'} D) = \sum_{i=1}^n p_i \|W_t(e_i) - W_{t'}(e_i)\|_{HS}^2$$

en el caso del *Statis*, lo que implica que para cada individuo e_i consideramos la distancia, al cuadrado, entre las dos tablas de los productos escalares de

dicho individuo con el resto, es decir, el cambio de posición de un individuo de una tabla a otra.

Para el *Statis dual*:

$$d^2(V_t M, V_{t'} M) = \sum_{j=1}^p m_j \|V_t(Y_j) - V_{t'}(Y_j)\|_{HS}^2$$

que es, para cada variable, el cuadrado de la diferencia, de una tabla a otra, de sus covarianzas con las otras variables. En el caso de variables centradas y reducidas, Z_j , las covarianzas entre las Y_j y las $Y_{j'}$ se transformarían en covarianzas entre las Z_j y las $Z_{j'}$.

Como los productos escalares también son positivos, los vectores representativos de los diferentes estudios o tablas también forman ángulos agudos entre sí. La diferencia es que ahora, el coseno del ángulo entre dos tablas no coincide con el producto escalar entre los objetos representativos de dichas tablas, sino que éste es la covarianza, y la distancia entre nubes depende del tamaño y de la forma de cada una de ellas. Si dos estudios son similares en forma y tamaño, el ángulo entre ellos es pequeño y las normas son similares.

— Al considerar operadores no normados, el análisis de la interestructura se asemeja al ACP *centrado (no-normado)*. Como en el caso anterior, se define la matriz S , que viene dada por la siguiente expresión:

$$S = \begin{bmatrix} 1 & s_{12} & \cdots & s_{1T} \\ s_{21} & 1 & \cdots & s_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ s_{T1} & \cdots & \cdots & 1 \end{bmatrix}$$

Es una matriz simétrica, semi-definida positiva y de orden $T \times T$ cuyos términos representativos son:

* *Statis* $\Rightarrow s_{tt'} = \text{Cov}(W_t D, W_{t'} D) = \text{Tr}(W_t D W_{t'} D)$

* *Statis dual* $\Rightarrow s_{tt'} = \text{Cov}(V_t M, V_{t'} M) = \text{Tr}(V_t M V_{t'} M)$

Denominando π_t al peso de cada tabla y Δ a la matriz que recoge dichos pesos, se tiene:

$$\Delta = \begin{bmatrix} \pi_1 & 0 & \cdots & 0 \\ & \pi_2 & & \\ & & \ddots & \\ & & & \pi_t \\ & & & & \ddots \\ 0 & \cdots & 0 & & & \pi_T \end{bmatrix}$$

Dado que todas las tablas tienen la misma importancia en el análisis los π_t son todos iguales a uno, y por ello, $\Delta = I$ es la matriz identidad de orden $T \times T$. En este caso, la inercia total, respecto al origen, viene dada por las siguientes expresiones:

$$\textit{Statis} \Rightarrow \sum_{t=1}^T \pi_t d^2(W_t D, 0) = \sum_{t=1}^T \pi_t \text{Tr}((W_t D)^2)$$

$$\textit{Statis dual} \Rightarrow \sum_{t=1}^T \pi_t d^2(V_t M, 0) = \sum_{t=1}^T \pi_t \text{Tr}((V_t M)^2)$$

Como en un ACP, para obtener la imagen euclídea de las distintas tablas el primer paso es la obtención de los ejes, que hacen máxima la inercia sobre ellos proyectados, descomponiendo en valores y vectores propios la matriz $S\Delta$. Estableciendo una equivalencia con dicho análisis, la coordenada de la tabla t con el eje k será:

$$\xi_{kt} = \sqrt{\mu_k} \gamma_{kt}$$

siendo: $\xi_k = \begin{bmatrix} \xi_{k1} \\ \xi_{k2} \\ \vdots \\ \xi_{kT} \end{bmatrix}$

el vector que recoge las coordenadas de las T tablas en el eje k , ya que μ_k es el valor propio asociado al vector propio γ_k de la matriz $S\Delta$ tal que:

$$S\Delta \gamma_k = \mu_k \gamma_k$$

y γ_k es el eje correspondiente a la componente principal k , que denominaremos δ_k , de tal manera que:

$$\gamma_k = \begin{bmatrix} \gamma_{k1} \\ \gamma_{k2} \\ \vdots \\ \gamma_{kT} \end{bmatrix}$$

La inercia de δ_k será: $d^2(\delta_k, 0) = \delta_k' \Delta \delta_k = \mu_k = d^2(\xi_k, 0)$

— La imagen euclídea en dos dimensiones de los operadores representativos de las tablas de datos representa la parte de inercia dada por:

$$\frac{\mu_1 + \mu_2}{\sum_{t=1}^T \pi_t \text{Tr}((W_t D)^2)} \text{ en el caso del } \textit{Statis}$$

$$\frac{\mu_1 + \mu_2}{\sum_{t=1}^T \pi_t \text{Tr}((V_t M)^2)} \text{ para el } \textit{Statis dual}$$

En este caso el coeficiente de correlación de una tabla t con la componente k será, respectivamente:

$$\langle W_t D, \delta_k \rangle = \frac{\sqrt{\mu_k} \gamma_{kt}}{\sqrt{(\text{Tr}(W_t D)^2)}} = \frac{\xi_{kt}}{\sqrt{(\text{Tr}(W_t D)^2)}}$$

$$\langle V_t M, \delta_k \rangle = \frac{\sqrt{\mu_k} \gamma_{kt}}{\sqrt{(\text{Tr}(V_t M)^2)}} = \frac{\xi_{kt}}{\sqrt{(\text{Tr}(V_t M)^2)}}$$

que no coincide con la coordenada correspondiente. También es posible mostrar una imagen euclídea centrada.

A la vista de lo expuesto cabe indicar que si los individuos y las variables son los mismos en todas las matrices D y M también serían constantes, y tendríamos que:

$$\|V_t M\|_{HS}^2 = \sum_{j=1}^p \sum_{j'=1}^p m_j m_{j'} \text{Cov}^2(Y_j^t, Y_{j'}^t) = \|W_t D\|_{HS}^2$$

$$\left\| \frac{V_t M}{\|V_t M\|_{HS}} \right\|_{HS}^2 = \left\| \frac{W_t D}{\|W_t D\|_{HS}} \right\|_{HS}^2 = 1$$

Conviene precisar que, a pesar de estas igualdades, la aplicación simultánea del *Statis* y del *Statis dual* no permite obtener la misma imagen euclídea de la interestructura puesto que, aunque la norma es la misma las matrices S en uno y en otro caso reflejan realidades diferentes: mientras los elementos que forman la matriz S resultado de la interestructura del *Statis* recoge los productos escalares entre los $W_t D$ (nubes de individuos), los elementos de la matriz S resultado de la interestructura del *Statis dual* indican los productos escalares entre los $V_t M$ (nubes de variables) para cada par de tablas.

b. Búsqueda de un compromiso

El objetivo de esta etapa es buscar la estructura común a las tablas, es decir, una nube "media", de individuos en el *Statis*, o de variables en el *Statis dual*, que represente lo mejor posible las T tablas de datos, y que va a venir caracterizada por un operador de la misma naturaleza que los de las tablas originales. Cuando las diferencias entre tablas son muy acusadas, el comportamiento de los individuos, o de las variables, es muy diferente de unas tablas a otras y la búsqueda de esa estructura común o compromiso carece de sentido.

La solución más sencilla consiste en definir el compromiso como una media ponderada de los operadores representativos de cada tabla, que debe ser

de la misma naturaleza que las tablas de partida, y que denominaremos *WD* en el caso del *Statis* y *VM* en el caso del *Statis dual*. Desarrollamos ahora esta segunda etapa del método distinguiendo como en el caso anterior entre objetos normados y no normados, haciendo referencia a las dos versiones del *STATIS* cuando surjan diferencias entre ellas.

Operadores normados

Se trata de determinar el valor de las ponderaciones para la obtención de *WD* ó de *VM* teniendo en cuenta que el tipo de operador que representa cada tabla es:

$$\frac{W_t D}{\|W_t D\|_{HS}} \quad \text{ó} \quad \frac{V_t M}{\|V_t M\|_{HS}}$$

para el *Statis* ó para el *Statis dual*, respectivamente.

De esta manera el operador compromiso toma las expresiones:

$$WD = \sum_{t=1}^T \alpha_t \frac{W_t D}{\|W_t D\|_{HS}}$$

$$VM = \sum_{t=1}^T \alpha_t \frac{V_t M}{\|V_t M\|_{HS}}$$

Para la determinación de los α_t se establece el criterio según el cual el operador compromiso es el objeto más correlacionado, en el sentido de producto escalar de Hilbert-Schmidt, con los otros objetos. Es decir, su distancia con el resto de objetos debe ser mínima, por lo que hay que minimizar las expresiones:

$$\sum_{t=1}^T \pi_t \left\| WD - \sum_{t=1}^T \alpha_t \frac{W_t D}{\|W_t D\|_{HS}} \right\|_{HS}$$

para que:

$$\sum_{t=1}^T \pi_t < \frac{W_t D}{\|W_t D\|_{HS}}, WD >^2$$

sea máxima en *Statis*, y:

$$\sum_{t=1}^T \pi_t \left\| VM - \sum_{t=1}^T \alpha_t \frac{V_t M}{\|V_t M\|_{HS}} \right\|_{HS} \Rightarrow \sum_{t=1}^T \pi_t < \frac{V_t M}{\|V_t M\|_{HS}}, VM >^2$$

sea máxima en el *Statis dual*.

Además, dado que deben ser de la misma naturaleza, la matriz WD es una matriz *normada* (norma unitaria) media de productos escalares entre individuos; y la matriz VM será la matriz *normada* media de varianzas-covarianzas si las variables están centradas, y una matriz *normada* media de correlaciones si las variables están centradas y reducidas.

Teniendo en cuenta la etapa de la interestructura, y dada la semejanza de dicho análisis con el de un ACP *normado* el primer vector propio:

$$\gamma_1 = \begin{bmatrix} \gamma_{11} \\ \gamma_{12} \\ \vdots \\ \gamma_{1T} \end{bmatrix}$$

asociado al primer valor propio μ_1 de la matriz de $S\Delta$ define la primera componente principal δ_1 que hace máxima su correlación con el conjunto de las tablas. En este caso, la primera componente principal se obtendría como:

$$\delta_1 = \gamma_{11} \frac{W_1 D}{\|W_1 D\|_{HS}} + \gamma_{12} \frac{W_2 D}{\|W_2 D\|_{HS}} + \dots + \gamma_{1t} \frac{W_t D}{\|W_t D\|_{HS}} + \dots + \gamma_{1T} \frac{W_T D}{\|W_T D\|_{HS}}$$

(*Statis*)

$$\delta_1 = \gamma_{11} \frac{V_1 M}{\|V_1 M\|_{HS}} + \gamma_{12} \frac{V_2 M}{\|V_2 M\|_{HS}} + \dots + \gamma_{1t} \frac{V_t M}{\|V_t M\|_{HS}} + \dots + \gamma_{1T} \frac{V_T M}{\|V_T M\|_{HS}}$$

(*Statis dual*)

y se deduce que:

$$\sum_{t=1}^T \pi_t < \frac{W_t D}{\|W_t D\|_{HS}}, \delta_1 >^2 = \mu_1 \quad (\textit{Statis})$$

$$\sum_{t=1}^T \pi_t < \frac{V_t M}{\|V_t M\|_{HS}}, \delta_1 >^2 = \mu_1 \quad (\textit{Statis dual})$$

siendo μ_1 máximo. Por tanto, si consideramos que WD ó VM es la primera componente de la interestructura, dicho compromiso cumple el criterio de maximización deseado y se tiene que $\alpha_t = \gamma_{1t} \forall t$.

Además, ya se demostró que el coeficiente de correlación de un operador con el primer factor es su coordenada con el primer eje. Por tanto, para todo t se tiene que:

$$\left\langle \frac{W_t D}{\|W_t D\|_{HS}}, \delta_1 \right\rangle = \sqrt{\mu_1} \gamma_{1t} = \xi_{1t} \quad \text{en el caso del } \textit{Statis}$$

$$\left\langle \frac{V_t M}{\|V_t M\|_{HS}}, \delta_1 \right\rangle = \sqrt{\mu_1} \gamma_{1t} = \xi_{1t} \quad \text{en el caso del } \textit{Statis dual}$$

De esta manera, el compromiso se sitúa sobre el primer eje principal a una distancia del origen $\sqrt{\mu_1}$ siendo μ_1 el valor propio al que está asociado, lo que facilita la interpretación de la imagen euclídea de la interestructura.

Para que la norma del operador WD y de VM sea uno (dado que si no $\|WD\|_{HS}^2 = \mu_1$ ó $\|VM\|_{HS}^2 = \mu_1$) será necesario dividir por la raíz cuadrada del valor propio correspondiente, y la ponderación α_t para cada t tabla será:

$$\alpha_t = \frac{\gamma_{1t}}{\sqrt{\mu_1}}$$

multiplicando por $\sqrt{\mu_1}$ el numerador y el denominador, tenemos que:

$$\alpha_t = \frac{\sqrt{\mu_1} \gamma_{1t}}{\mu_1} = \frac{\xi_{1t}}{\mu_1}$$

Es decir, la ponderación de cada operador normado es su coordenada con el primer eje dividida por el valor propio asociado a dicho eje. Considerando ese α_t cada tabla t tendrá más importancia en el análisis cuanto mayor sea la correlación de dicha tabla con el primer factor de la interestructura. Puesto que todas las variables de una tabla tienen la misma ponderación, todas las de una tabla que esté relacionada con el compromiso tiene su peso aumentado, incluso para las variables que no participen en esta relación.

Es posible demostrar, para el *Statis*, que con esta ponderación la matriz de productos escalares entre individuos WD es la misma que se obtendría al realizar un ACP *centrado* de las tablas yuxtapuestas, ponderadas cada una de ellas por $\sqrt{\alpha_t \beta_t}$ siendo:

$$\beta_t = \frac{1}{\|W_t D\|_{HS}}$$

de la siguiente manera, para variables centradas:

$$\sqrt{\alpha_1 \beta_1} Y^1, \sqrt{\alpha_2 \beta_2} Y^2, \dots, \sqrt{\alpha_t \beta_t} Y^t, \dots, \sqrt{\alpha_T \beta_T} Y^T = \Psi$$

y con variables centradas y reducidas:

$$\sqrt{\alpha_1 \beta_1} Z^1, \sqrt{\alpha_2 \beta_2} Z^2, \dots, \sqrt{\alpha_t \beta_t} Z^t, \dots, \sqrt{\alpha_T \beta_T} Z^T = \Pi$$

Tanto Ψ como Π tienen n filas y $\sum p_t$ columnas. Es decir, n individuos-compromiso e_i^c representados por $\sum p_t$ valores cada uno; y $\sum p_t$ variables (Y_j^t ó Z_j^t) representadas por n valores cada una.

Esta definición del compromiso permite demostrar que la distancia entre dos puntos e_i^c y $e_{i'}^c$ representa la distancia media ponderada entre dichos individuos para el conjunto de las tablas (o distancia media ponderada en el periodo si las tablas corresponden a diferentes momentos del tiempo). Así, por ejemplo en el caso de variables centradas y operadores normados:

$$\begin{aligned} d^2(e_i^c, e_{i'}^c) &= \\ &= \sum_{j=1}^{\sum p_t} m_j (\sqrt{\alpha_t \beta_t} y_{ij} - \sqrt{\alpha_t \beta_t} y_{i'j})^2 = \sum_{t=1}^T \sum_{j=1}^{p_t} m_j^t (\sqrt{\alpha_t \beta_t} y_{ij}^t - \sqrt{\alpha_t \beta_t} y_{i'j}^t)^2 = \\ &= \sum_{t=1}^T \alpha_t \beta_t \sum_{j=1}^{p_t} m_j^t (y_{ij}^t - y_{i'j}^t)^2 = \sum_{t=1}^T \alpha_t \beta_t d^2(e_i^t, e_{i'}^t) = \sum_{t=1}^T \alpha_t \beta_t \|e_i^t - e_{i'}^t\|_{M_t}^2 \end{aligned}$$

y la distancia entre individuos compromiso está influida por la ponderación. Si las variables estuviesen reducidas la demostración es similar, cambiando Y_{ij} por Z_{ij} . En el caso de objetos no normados la demostración es equivalente, con la ponderación $\sqrt{\alpha_t}$.

Para el *Statis dual* la matriz VM es la tabla de productos escalares entre variables resultado de un ACP *centrado* de las tablas de datos apiladas y debidamente ponderadas. Como ya se ha indicado VM tiene que ser una matriz de varianzas-covarianzas (si las variables están centradas) o de correlaciones (si las variables están centradas y reducidas) según el caso, pero con norma unitaria ambas.

$$\text{Denominando: } \beta_t = \frac{1}{\|V_t M\|_{\text{HS}}}$$

$$\text{se tiene: } \left\{ \begin{array}{c} \sqrt{\alpha_1 \beta_1} Y^1 \\ \vdots \\ \sqrt{\alpha_t \beta_t} Y^t \\ \vdots \\ \sqrt{\alpha_T \beta_T} Y^T \end{array} \right\} = \Psi \quad (\text{variables centradas})$$

$$\left\{ \begin{array}{c} \sqrt{\alpha_1 \beta_1} Z^1 \\ \vdots \\ \sqrt{\alpha_t \beta_t} Z^t \\ \vdots \\ \sqrt{\alpha_T \beta_T} Z^T \end{array} \right\} = \Pi \quad (\text{variables centradas y reducidas})$$

Tanto Ψ como Π tienen Σn_t filas y p columnas, es decir, p variables-compromiso (Y_j^c ó Z_j^c) representadas por Σn_t valores cada una, y Σn_t individuos e_i^t representados por p valores cada uno de ellos.

Conviene señalar que el resultado no sería el mismo si se realizara un ACP *normado* puesto que este análisis supondría que todas las variables de las tablas tienen la misma varianza. Debido a la ponderación, este hecho no se cumple ya que cada variable tiene una varianza diferente según esa ponderación, excepto para el caso de que la matriz VM fuese una matriz

de correlaciones (es decir, el caso de variables centradas y reducidas del *Statis dual*). Para una variable de la tabla t se tiene que:

$$\text{Var} (\sqrt{\alpha_t \beta_t} Y_j^t) = \alpha_t \beta_t \text{Var} (Y_j^t)$$

ó:

$$\text{Var} (\sqrt{\alpha_t \beta_t} Z_j^t) = \alpha_t \beta_t \text{Var} (Z_j^t) = \alpha_t \beta_t$$

según se considere Y_j^t ó Z_j^t . Con operadores no normados la situación es la misma, sustituyendo $\sqrt{\alpha_t \beta_t}$ por $\sqrt{\alpha'_t}$

Operadores no normados

Cuando los operadores asociados a las tablas son $W_t D$ y $V_t M$, el operador compromiso, se obtiene según las siguientes expresiones respectivamente:

$$WD = \sum_{t=1}^T \alpha'_t W_t D$$

$$VM = \sum_{t=1}^T \alpha'_t V_t M$$

Puesto que el compromiso tiene que ser de la misma naturaleza, la norma de dicho operador debe cumplir que:

$$\|WD\|_{HS} = \sum_{t=1}^T \pi_t \|W_t D\|_{HS}$$

$$\|VM\|_{HS} = \sum_{t=1}^T \pi_t \|V_t M\|_{HS}$$

siendo estas expresiones las que nos dan las distancias del compromiso al origen. Como ya se ha indicado, la primera componente δ_1 se obtiene por combinación lineal de las tablas (siendo μ_1 el primer vector propio asociado al primer valor propio γ_1) de tal manera que:

$$\delta_1 = \sum_{t=1}^T \gamma_{1t} W_t D$$

$$\delta_1 = \sum_{t=1}^T \gamma_{1t} V_t M$$

Dado que en este caso el análisis de la interestructura es similar a un ACP *centrado*, se tiene que el primer vector propio, del que se deduce la primera componente principal, hace máxima las siguientes expresiones:

$$\sum_{t=1}^T \pi_t \langle W_t D, \delta_1 \rangle^2 \quad \text{siendo} \quad \langle W_t D, \delta_1 \rangle = \frac{\sqrt{\mu_1} \gamma_{1t}}{\sqrt{(\text{Tr}(W_t D)^2)}} \quad \text{en } \textit{Statis}$$

$$\sum_{t=1}^T \pi_t \langle V_t M, \delta_1 \rangle^2 \quad \text{siendo} \quad \langle V_t M, \delta_1 \rangle = \frac{\sqrt{\mu_1} \gamma_{1t}}{\sqrt{(\text{Tr}(V_t M)^2)}} \quad \text{en } \textit{Statis dual}$$

que son los coeficientes de correlación correspondientes entre los operadores y el primer factor.

La primera componente así definida tendría norma $\sqrt{\mu_1}$, y dado que la norma debe tomar la expresión indicada, consideramos:

$$\alpha_t' = \frac{\gamma_{1t}}{\sqrt{\mu_1}} \sum_{t=1}^T \pi_t \|W_t D\|_{HS}$$

para la construcción de WD y:

$$\alpha_t' = \frac{\gamma_{1t}}{\sqrt{\mu_1}} \sum_{t=1}^T \pi_t \|V_t M\|_{HS}$$

para la construcción de VM .

Los $W_t D$ y los $V_t M$ no están normados por lo que en la construcción del compromiso influye tanto la dimensión como el tamaño de las nubes. La matriz WD es la que se deduciría al aplicar un ACP *centrado* a la yuxtaposición de las tablas de manera que para las variables centradas:

$$\sqrt{\alpha_1'} Y^1, \sqrt{\alpha_2'} Y^2, \dots, \sqrt{\alpha_t'} Y^t, \dots, \sqrt{\alpha_T'} Y^T = \Psi$$

y con variables centradas y reducidas:

$$\sqrt{\alpha_1'} Z^1, \sqrt{\alpha_2'} Z^2, \dots, \sqrt{\alpha_t'} Z^t, \dots, \sqrt{\alpha_T'} Z^T = \Pi$$

En ambos casos la tabla total tendrá n filas (individuos-compromiso e_i^c) y $\sum p_t$ columnas-variables (Y_j^t ó Z_j^t), como para el caso ya comentado.

La matriz VM (según el caso matriz de varianzas–covarianzas o de correlaciones pero sin normar) es la que se deduciría al aplicar un ACP *centrado* de las tablas apiladas y debidamente ponderadas. Para variables centradas se aplica a:

$$\left\{ \begin{array}{c} \sqrt{\alpha'_1} Y^1 \\ \vdots \\ \sqrt{\alpha'_t} Y^t \\ \vdots \\ \sqrt{\alpha'_T} Y^T \end{array} \right\} = \Psi$$

y para variables centradas y reducidas:

$$\left\{ \begin{array}{c} \sqrt{\alpha'_1} Z^1 \\ \vdots \\ \sqrt{\alpha'_t} Z^t \\ \vdots \\ \sqrt{\alpha'_T} Z^T \end{array} \right\} = \Pi$$

Tanto ψ como Π tienen Σn_t filas y p columnas, es decir, p variables–compromiso (Y_j^c ó Z_j^c) representadas por Σn_t valores cada una, y Σn_t individuos e_i^t representados por p valores cada uno de ellos.

c. Estudio de la intraestructura

La imagen euclídea de la intraestructura permite obtener las posiciones–compromiso de los individuos (o posiciones medias para el conjunto de las tablas) en el caso del *Statis*, o las posiciones–compromiso (covarianzas o correlaciones medias para el conjunto de las tablas) de las variables. Si las distancias entre tablas son débiles entonces la estructura común o compromiso es la que se deduce de las distancias entre los individuos–

compromiso en el primer caso, o de las covarianzas (o correlaciones) de las variables en el segundo. En el caso del *Statis*, los factores obtenidos a partir de las posiciones-compromiso, que definen la imagen euclídea, serán interpretados, como en un ACP, a través de las correlaciones de las variables de todas las tablas de datos, analizando de esa manera la estructura obtenida. En el caso del *Statis dual*, las correlaciones de las variables-compromiso con los factores permiten definir la estructura común obtenida, y la proyección de los individuos sobre ésta estructura permite la caracterización de éstos, de acuerdo con la significación otorgada a los factores.

La descomposición en valores y vectores propios de WD y VM permitirá obtener las puntuaciones de los individuos-compromiso, en el primero, y las correlaciones de las variables-compromiso, en el segundo, alcanzado así los objetivos planteados. Considerando las diferencias entre ambos operadores, en el caso del compromiso WD su descomposición, por ser una matriz de productos escalares entre individuos de orden $n \times n$, es similar a un ACP *centrado* de la nube de individuos. Por este motivo, los vectores propios de WD son las coordenadas de los n individuos en los ejes de inercia (que dan lugar a los factores), y las coordenadas de las Σp_i variables y sus correlaciones con esos factores se obtienen a través de las relaciones de transición. Para el compromiso VM su descomposición, por ser una matriz de covarianzas o de correlaciones de orden $p \times p$, es equivalente a un ACP *centrado* de la nube de las p variables, cuyas coordenadas son los vectores propios de VM , obteniéndose las coordenadas de los Σn_i individuos mediante las fórmulas de transición. A continuación se analizan WD y VM teniendo en cuenta estas circunstancias.

— **El operador WD define el espacio de los individuos**, que es: $R^{\Sigma p_i}$

Denominando F_k^c a la componente k , que representa las coordenadas de los n individuos-compromiso (de los e_i^c para $i = 1, 2 \dots n$) en el eje k y λ_k^c el valor propio asociado a dicha componente, se tiene que:

$$WD F_k^c = \lambda_k^c F_k^c$$

de tal manera que:

$$F_k^c = \begin{bmatrix} f_{k1}^c \\ f_{k2}^c \\ \vdots \\ f_{ki}^c \\ \vdots \\ f_{kn}^c \end{bmatrix}$$

siendo f_{ki}^c la coordenada del individuo e_i^c en el eje k del compromiso y la componente F_k^c pertenece a R^n .

— Al eje asociado del espacio de las variables lo denominaremos t_k^c , y teniendo en cuenta el esquema de dualidad de la figura II.13b (p. 160), se cumple que:

$$t_k^c = \frac{F_k^c}{\sqrt{\lambda_k^c}}$$

por lo que sustituyendo este valor en la expresión anterior se tiene que:

$$WD \sqrt{\lambda_k^c} t_k^c = \lambda_k^c \sqrt{\lambda_k^c} t_k^c \Rightarrow WD \frac{t_k^c}{\sqrt{\lambda_k^c}} = \sqrt{\lambda_k^c} t_k^c = F_k^c$$

Considerando que f_{ki}^c es la coordenada del individuo e_i^c en los ejes del compromiso, se cumple que:

$$f_{ki}^c = \sum_{i'=1}^n (WD)_{ii'} \frac{t_{ki'}^c}{\sqrt{\lambda_k^c}}$$

— La cantidad de información explicada por esta la componente k es su valor propio, por lo que el porcentaje de inercia explicada es:

$$\frac{\lambda_k^c}{\sum_{k=1}^p \lambda_k^c}$$

considerando que p es el número de vectores propios no nulos de WD .

— Para interpretar los ejes del compromiso es necesario obtener las coordenadas de las $\sum p_t$ variables con los ejes, y sus correlaciones con los factores. Dichas coordenadas, teniendo en cuenta el eje de la nube de variables t_k^c , toman la siguiente expresión:

$$G_k = \Psi' D t_k^c = \begin{bmatrix} g_{k1}^1 \\ \vdots \\ g_{kp_1}^1 \\ g_{k1}^2 \\ \vdots \\ g_{kp_2}^2 \\ \vdots \\ g_{k1}^T \\ \vdots \\ g_{kp_T}^T \end{bmatrix}$$

considerando que Ψ es la matriz formada por las tablas yuxtapuestas. Las coordenadas de la variable Y_j^t (variable Y_j de la tabla t) son¹⁶:

$$g_{kj}^t = (\sqrt{\alpha_t \beta_t} Y_j^t)' D t_k^c$$

Cuando las variables están centradas, el coeficiente de correlación de una variable con un eje k es igual a:

$$\begin{aligned} \rho (F_k^c, \sqrt{\alpha_t \beta_t} Y_j^t) &= \frac{g_{kj}^t}{\sqrt{\text{Var}(\sqrt{\alpha_t \beta_t} Y_j^t)}} = \\ &= \frac{(\sqrt{\alpha_t \beta_t} Y_j^t)' D t_k^c}{\sqrt{\alpha_t \beta_t} \sqrt{\text{Var}(Y_j^t)}} = \frac{(Y_j^t)' D t_k^c}{\sqrt{\text{Var}(Y_j^t)}} \end{aligned}$$

expresión en la que las ponderaciones se eliminan.

Si las variables están reducidas se verifica que $G_k = \Pi' D t_k^c$, siendo Π la matriz formada por la yuxtaposición de las tablas. La coordenada de una variable Z_j^t con el eje k toma la siguiente expresión:

¹⁶ Dado que lo más habitual es el empleo de operadores normados, las ponderaciones consideradas son las relativas a dicho caso. En el caso de operadores sin normar las ponderaciones serían $\sqrt{\alpha_t}$ pero el proceso sería el mismo.

$$g_{kj}^t = (\sqrt{\alpha_t \beta_t} Z_j^t)' D t_k^c$$

Puesto que se trata de un ACP *centrado* el coeficiente de correlación de una variable con la componente k se expresa como:

$$\begin{aligned} \rho(F_k^c, \sqrt{\alpha_t \beta_t} Z_j^t) &= \frac{g_{kj}^t}{\sqrt{\text{Var}(\sqrt{\alpha_t \beta_t} Z_j^t)}} = \\ &= \frac{(\sqrt{\alpha_t \beta_t} Z_j^t)' D t_k^c}{\sqrt{\alpha_t \beta_t} \sqrt{\text{Var}(Z_j^t)}} = (Z_j^t)' D t_k^c \end{aligned}$$

Los coeficientes de correlación de las variables con los factores permiten caracterizar éstos e interpretar las posiciones de los individuos—compromiso como en cualquier ACP *centrado*, siendo necesario estudiar también su calidad de representación. Si las variables son las mismas en todas las tablas, es decir p , podemos estudiar la trayectoria de una variable teniendo en cuenta las T correlaciones relativas a dicha variable, con el fin de completar la interpretación otorgada a los factores.

Teniendo en cuenta el esquema de dualidad de la figura II.13b (p. 160) se tiene que:

$$G_k = \sqrt{\lambda_k^c} a_k \Rightarrow g_{kj}^t = \sqrt{\lambda_k^c} a_{kj}^t \Rightarrow a_{kj}^t = \frac{g_{kj}^t}{\sqrt{\lambda_k^c}}$$

siendo a_{kj}^t el elemento representativo del vector unitario a_k^t asociado al espacio de los individuos: $R^{\sum p_t}$

De esta manera es posible reconstituir las tablas de partida Ψ ó Π según el caso:

$$\sqrt{\alpha_t \beta_t} Z_j^t = \sum_{k=1}^h F_k^c a_{kj}^t \quad \text{ó} \quad \sqrt{\alpha_t \beta_t} Y_j^t = \sum_{k=1}^h F_k^c a_{kj}^t$$

— Para obtener la imagen euclídea de los individuos en los factores, caracterizados a través de los coeficientes de correlación con las variables,

podemos considerar los dos primeros obteniendo una representación en dos dimensiones del conjunto de n individuos–compromiso e_i^c , imagen que puede considerarse, para cada individuo e_i como su posición–media para el conjunto de las tablas. La cantidad de información explicada por ese plano será la suma de sus valores propios $\lambda_1^c + \lambda_2^c$ por lo que el porcentaje explicado por el mismo será:

$$\frac{\lambda_1^c + \lambda_2^c}{\sum_{k=1}^p \lambda_k^c}$$

Si consideramos la imagen euclídea en los h primeros factores, la cantidad de información será:

$$\frac{\sum_{k=1}^h \lambda_k^c}{\sum_{k=1}^p \lambda_k^c}$$

— **El operador VM , por tratarse de una matriz de covarianzas o de correlaciones**, no está centrado y el espacio que se considera es: $R^{\sum p_t}$

Denominando G_k^c al vector que representa las coordenadas de las p variables–compromiso (Y_j^c ó Z_j^c , para $j = 1, 2 \dots p$) en el eje k y λ_k^c el valor propio asociado a dicha componente, se tiene que:

$$VM G_k^c = \lambda_k^c G_k^c$$

de tal manera que:

$$G_k^c = \begin{bmatrix} g_{k1}^c \\ g_{k2}^c \\ \vdots \\ g_{kj}^c \\ \vdots \\ g_{kp}^c \end{bmatrix}$$

siendo g_{kj}^c la coordenada de la variable Y_j^c ó Z_j^c en los ejes del compromiso.

— El eje asociado al espacio de los individuos, que pertenece a R^p , es u_k^c (tiene el vector unitario a_k^c asociado), por lo que considerando las fórmulas de transición se tiene que:

$$u_k^c = \frac{G_k^c}{\sqrt{\lambda_k^c}}$$

Sustituyendo en la expresión anterior se deduce:

$$VM \sqrt{\lambda_k^c} u_k^c = \lambda_k^c \sqrt{\lambda_k^c} u_k^c \Rightarrow VM \frac{u_k^c}{\sqrt{\lambda_k^c}} = \sqrt{\lambda_k^c} u_k^c = G_k^c$$

por lo que si g_{kj}^c es la coordenada de la variable Y_j^c ó Z_j^c en los ejes del compromiso, se tiene:

$$g_{kj}^c = \sum_{j'=1}^p (VM)_{jj'} \frac{u_{kj'}^c}{\sqrt{\lambda_k^c}}$$

Si tomamos los dos primeros ejes, obtenemos una representación en dos dimensiones del conjunto de p variables–compromiso (Y_j^c ó Z_j^c) y esa imagen puede considerarse para cada variable, como su posición–media para el conjunto de las tablas.

— Es necesario considerar también las correlaciones de las variables con los factores a fin de caracterizar éstos y representar posteriormente los individuos. Si la matriz VM es una matriz de correlaciones entonces g_{kj}^c es también el coeficiente de correlación de la variable (Y_j^c ó Z_j^c) con la componente k . En el caso de que sea una matriz de varianzas–covarianzas el coeficiente de correlación tomará la expresión:

$$\frac{\sqrt{\lambda_k^c} u_{kj}^c}{\sqrt{\text{Var}(\sqrt{\alpha_t} \beta_t Y_j^c)}}$$

coeficientes que permiten dotar de significado a las componentes.

— La proyección de los $\sum n_t$ individuos en esos ejes se puede interpretar según la posición de cada uno de ellos con respecto a los factores, dotados

de significado a través de las correlaciones de las variables. Las posiciones de los individuos en los ejes del compromiso vienen dadas a través de sus coordenadas con los ejes u_k^c y tomarán la siguiente expresión:

$$F_k = \Psi u_k^c = \begin{bmatrix} f_{k1}^1 \\ \vdots \\ f_{kn_1}^1 \\ f_{k1}^2 \\ \vdots \\ f_{kn_2}^2 \\ \vdots \\ f_{k1}^T \\ \vdots \\ f_{kn_T}^T \end{bmatrix}$$

considerando Ψ la matriz formada por las tablas apiladas si las variables están centradas. Las coordenadas del individuo e_i^t (individuo e_i de la tabla t) en el eje k son $f_{ki}^t = (e_i^t)' u_{ki}^c$, siendo:

$$(e_i^t)' = (y_{i1}^1, y_{i2}^1, \dots, y_{ip}^1, y_{i1}^2, y_{i2}^2, \dots, y_{ip}^2, \dots, y_{i1}^T, y_{i2}^T, \dots, y_{ip}^T)$$

Si las variables estuviesen centradas y reducidas tenemos $F_k = \Pi u_{ki}^c$, siendo Π la matriz formada por las tablas de datos apiladas tal y como se indicó, siendo los valores de:

$$(e_i^t)' = (z_{i1}^1, z_{i2}^1, \dots, z_{ip}^1, z_{i1}^2, z_{i2}^2, \dots, z_{ip}^2, \dots, z_{i1}^T, z_{i2}^T, \dots, z_{ip}^T)$$

— Para obtener la imagen euclídea de los individuos en los factores, caracterizados a través de los coeficientes de correlación con las variables, podemos considerar los dos primeros obteniendo una representación en dos dimensiones del conjunto de individuos-compromiso e_i^t . La cantidad de información explicada por ese plano viene dada por la suma de los valores propios de los factores correspondientes, que son los del compromiso, por lo que dicha inercia es $\lambda_1^c + \lambda_2^c$ y el porcentaje explicado por ese plano será:

$$\frac{\lambda_1^c + \lambda_2^c}{\sum_{k=1}^p \lambda_k^c}$$

Si consideramos la imagen euclídea en los h primeros factores, la cantidad de información será:

$$\frac{\sum_{k=1}^h \lambda_k^c}{\sum_{k=1}^p \lambda_k^c}$$

Cuando los individuos son los mismos en todas las tablas, es decir n , podemos estudiar la trayectoria, como individuos activos, de cada uno teniendo en cuenta los T puntos representativos correspondientes a los mismos.

d. Trayectorias

En la imagen euclídea del compromiso es posible representar la posición de cada individuo para cada tabla en el *Statis*, o de cada variable en el *Statis dual*, mediante la técnica de individuos suplementarios. Si las tablas representan momentos del tiempo diferentes y están ordenadas, la unión de los T puntos correspondientes a cada individuo, o a cada variable, define su trayectoria. Dichas trayectorias ponen en evidencia los individuos, o las variables, responsables de las diferencias entre las tablas, que son aquéllos cuya trayectoria difiere del comportamiento medio. Analizamos a continuación las trayectorias de los individuos que se obtienen como resultado del *Statis*, y las trayectorias de las variables obtenidas como resultado del *Statis dual*.

Trayectorias de los individuos

En el caso de la intraestructura del *Statis*, cada individuo está representado por un punto que refleja el comportamiento medio del individuo para el conjunto de las tablas, es decir la representación del individuo-compromiso. Sin embargo, para determinar los individuos responsables de las diferencias entre tablas, es necesario representar las posiciones de cada individuo para

cada una de las ellas, respecto a los ejes del compromiso. Para un e_i concreto hay T puntos, e_i^t ($t = 1, 2 \dots T$) cuya representación en el compromiso permite configurar la trayectoria del individuo, siempre y cuando los puntos están ordenados cronológicamente.

Para llevar a cabo dicha representación se consideran cada una de las T tablas de n individuos, representadas por el operador:

$$\frac{W_t D}{\|W_t D\|_{HS}}$$

o el $W_t D$ correspondiente, como tablas suplementarias del operador WD compromiso.

Denominado F_k^t al vector representativo de las coordenadas de los n individuos de la tabla t en el eje k se tiene, en el caso de que los objetos sean normados:

$$F_k^t = \frac{W_t D}{\|W_t D\|_{HS}} \frac{t_k^c}{\sqrt{\lambda_k^c}}$$

siendo $F_k^t = \begin{bmatrix} f_{k1}^t \\ f_{k2}^t \\ \vdots \\ f_{ki}^t \\ \vdots \\ f_{kn}^t \end{bmatrix}$

$$y \quad f_{ki}^t = \frac{W_t D}{\|W_t D\|_{HS}} \frac{t_{ki}^c}{\sqrt{\lambda_k^c}}$$

expresión que refleja las proyecciones en t_k^c del individuo, pero sometidas a la transformación:

$$\frac{1}{\sqrt{\lambda_k^c}}$$

Cuando se trabaja con objetos no normados se tiene que: $F_k^t = W_t D \frac{t_k^c}{\sqrt{\lambda_k^c}}$

siendo: $F_k^t = \begin{bmatrix} f_{k1}^t \\ f_{k2}^t \\ \vdots \\ f_{ki}^t \\ \vdots \\ f_{kn}^t \end{bmatrix}$

El vector F_k^t representa las coordenadas de todos los individuos de la tabla t para el eje k . Para el individuo e_i los puntos $f_{ki}^1, f_{ki}^2 \dots f_{ki}^T$ representan la trayectoria de dicho individuo respecto al eje k . Además se puede demostrar que el punto f_{ki}^c es el centro de gravedad de los $f_{ki}^1, f_{ki}^2 \dots f_{ki}^T$, debidamente ponderados, según el caso, de forma que:

$$\begin{aligned} F_{ki}^c &= WD \frac{t_k^c}{\sqrt{\lambda_k^c}} = \\ &= \left[\alpha_1 \frac{W_1 D}{\|W_1 D\|_{HS}} + \alpha_2 \frac{W_2 D}{\|W_2 D\|_{HS}} + \dots + \right. \\ &\quad \left. + \dots + \alpha_t \frac{W_t D}{\|W_t D\|_{HS}} + \dots + \alpha_T \frac{W_T D}{\|W_T D\|_{HS}} \right] \frac{t_k^c}{\sqrt{\lambda_k^c}} = \\ &= \alpha_1 \frac{W_1 D}{\|W_1 D\|_{HS}} \frac{t_k^c}{\sqrt{\lambda_k^c}} + \alpha_2 \frac{W_2 D}{\|W_2 D\|_{HS}} \frac{t_k^c}{\sqrt{\lambda_k^c}} + \dots + \\ &\quad + \dots + \alpha_t \frac{W_t D}{\|W_t D\|_{HS}} \frac{t_k^c}{\sqrt{\lambda_k^c}} + \dots + \alpha_T \frac{W_T D}{\|W_T D\|_{HS}} \frac{t_k^c}{\sqrt{\lambda_k^c}} = \\ &= \alpha_1 F_k^1 + \alpha_2 F_k^2 + \dots + \alpha_t F_k^t + \dots + \alpha_T F_k^T = \\ &= \sum_{t=1}^T \alpha_t F_k^t \quad \Rightarrow \quad f_{ki}^c = \sum_{t=1}^T \alpha_t f_{ki}^t \end{aligned}$$

De manera semejante se haría la demostración si el compromiso hubiese sido obtenido a través de los operadores no normados. Por este motivo, se puede decir que los coeficientes α_t se visualizan en las representaciones gráficas y que el método acentúa la proximidad de los individuos-compromiso con los individuos de las tablas más relacionadas con la estructura común, cualquiera que sea el eje factorial. Por otra parte, dado que:

$$\sum_{t=1}^T \alpha_t \neq 1$$

aunque los T puntos representativos de un individuo fuesen coincidentes, su posición no coincidiría con la posición compromiso.

Las trayectorias permiten determinar los individuos responsables de las desviaciones entre las tablas t y t' sobre cualquier eje k , teniendo en cuenta el coeficiente:

$$\frac{1}{\sqrt{\lambda_k^c}}$$

las distancias entre f_{ki}^t y $f_{ki}^{t'}$ están influenciadas por las distancias entre $W_t D$ y $W_{t'} D$, de tal forma que:

$$\begin{aligned} \left\| f_{ki}^t - f_{ki}^{t'} \right\|^2 &= \left\| \sum_{i'=1}^n (W_t D)_{i i'} \frac{t_{k i'}^c}{\sqrt{\lambda_k^c}} - \sum_{i'=1}^n (W_{t'} D)_{i i'} \frac{t_{k i'}^c}{\sqrt{\lambda_k^c}} \right\|_{HS}^2 = \\ &= \frac{1}{\lambda_k^c} \left\| \sum_{i'=1}^n (W_t D)_{i i'} t_{k i'}^c - \sum_{i'=1}^n (W_{t'} D)_{i i'} t_{k i'}^c \right\|_{HS}^2 = \frac{1}{\lambda_k^c} \left\| \sum_{i'=1}^n p_{i'} t_{k i'}^c [W_t(e_{i'}) - W_{t'}(e_{i'})] \right\|_{HS}^2 = \\ &= \frac{1}{\lambda_k^c} \left\| [W_t(e_i) - W_{t'}(e_i)] D t_k^c \right\|_{HS}^2 \end{aligned}$$

Análogamente se demostraría si los objetos fuesen normados. Como consecuencia del anterior desarrollo, el estudio de las trayectorias se limita al examen de las posiciones de los individuos sobre los dos primeros ejes debido a que los coeficientes λ_k^c son cada vez más pequeños y distorsionan cada vez más las distancias en los ejes sucesivos. No obstante, si no hay una gran diferencia entre λ_2^c y λ_3^c se considerarán también las trayectorias sobre el tercer eje.

Trayectorias de las variables

Para determinar las variables responsables de las diferencias entre tablas es necesario representar las posiciones de cada variable para cada tabla,

considerando que cada variable (Y_j ó Z_j) está representada por T puntos (Y_j^t ó Z_j^t con $t = 1, 2 \dots T$); la representación de éstas sobre el compromiso refleja la posición de cada variable en cada tabla respecto a los ejes de éste. En este caso, la técnica para esta representación se basa también en la de elementos suplementarios, de forma que:

$$G_k^t = V_t M \frac{u_k^c}{\sqrt{\lambda_k^c}}$$

siendo: $G_k^t = \begin{bmatrix} g_{k1}^t \\ g_{k2}^t \\ \vdots \\ g_{kj}^t \\ \vdots \\ g_{kp}^t \end{bmatrix}$

El vector G_k^t representa las coordenadas de todas las variables de la tabla t para el eje k . Para la variable (Y_j ó Z_j) los puntos $g_{kj}^1, g_{kj}^2 \dots g_{kj}^T$ representan la trayectoria de dicha variable respecto al eje k y son las coordenadas de ($Y_j^1, Y_j^2 \dots Y_j^T$ ó $Z_j^1, Z_j^2 \dots Z_j^T$). También es posible demostrar que el punto g_{kj}^c es el centro de gravedad de los $g_{kj}^1, g_{kj}^2 \dots g_{kj}^T$ debidamente ponderados por los α_t . De esta manera, las trayectorias de las variables así consideradas están influenciadas por las ponderaciones, como las trayectorias de los individuos anteriormente descritas.

2.3.3 Interpretación

Para una correcta interpretación de los resultados obtenidos son necesarias una serie de puntualizaciones sobre cada una de las etapas indicadas.

a. Interestructura y compromiso

Remitiéndonos a lo ya comentado, como fase previa a la aplicación de cualquiera de las dos versiones de la metodología STATIS es conveniente realizar un análisis individual de los operadores asociados a las nubes (como indicamos W_t ó V_t), para determinar si son comparables o no. Según se trabaje con los operadores normados o no, se obtienen unos valores propios u otros, pero los porcentajes de inercia son iguales para los operadores W_t y V_t (los correspondientes a la matriz de varianzas, si las variables están centradas, o a la de correlaciones si las variables están centradas y reducidas). Respecto al ACP, mientras éste analiza el $W_t D$ ó $V_t M$ (normado o no), el *Statis* analiza el W_t (normado o no) y el *Statis dual* el V_t (normado o no). Por ello los porcentajes de inercia serán los mismos, pero no los valores propios.

El análisis de la interestructura permite obtener la imagen euclídea de las tablas de datos que forman la tabla total, y comparar globalmente dichas tablas. Dado que el compromiso es colineal al primer eje, está situado a una distancia al origen igual a uno (si se trata de objetos normados), o a una distancia de:

$$\sum_{t=1}^T \pi_t \|W_t D\|_{HS} \quad (\text{para los } WD)$$

$$\sum_{t=1}^T \pi_t \|V_t M\|_{HS} \quad (\text{para los } VM)$$

en el caso de objetos no normados. Por tanto, dicho eje permite comparar las tablas entre sí, y éstas con el compromiso o estructura común.

En el caso de operadores normados, tanto de $W_t D$ como $V_t M$, es necesario interpretar la norma y el ángulo que forman; la norma indica la relación global que existe entre las variables que componen la tabla a la que se refiere: si es elevada significa que dicha tabla está bien representada sobre el primer eje y por tanto respecto al compromiso. Si la norma es pequeña, y por tanto próxima al origen, indicaría una mala representación, como sucede en un círculo de correlaciones.

El ángulo entre los vectores correspondientes a los distintos operadores depende de su coeficiente RV , que puede interpretarse como el coseno del ángulo que forman. Si estos coeficientes RV fuesen pequeños los ángulos serían casi rectos y no existiría una estructura común a todas las nubes; en este caso el compromiso sólo sería una media ponderada de operadores, pero con poco sentido, siendo además los ángulos formados por ellos y el primer eje bastante grandes. Si los RV son elevados, los ángulos entre operadores, y de éstos con el compromiso, son pequeños.

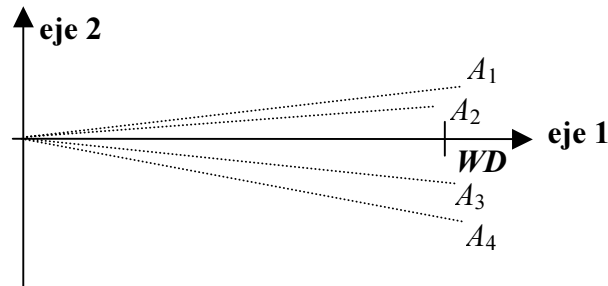
Si se emplean operadores no normados la interpretación es análoga, si bien todos ellos deben tener una norma similar puesto que en caso contrario el compromiso estaría más influenciado por los de norma más elevada. Además, los ángulos no reflejan los coeficientes de correlación entre los estudios sino las covarianzas, aunque siguen siendo agudos.

Una vez abordadas estas cuestiones generales de interpretación se van a considerar cuatro casos típicos en el estudio de la interestructura que sirven de modelo para la interpretación de los demás y que son los considerados en Lavit (1988: 97) y en GERI (1996, 78-9). En todas las representaciones es posible sustituir el operador compromiso WD , que corresponde al *Statis*, por el operador VM del *Statis dual*, cuando esta versión es la aplicada, y los criterios para su interpretación siguen siendo válidos.

— El primer caso considerado, que se expresa en la figura II.16, indica que las tablas están representadas por objetos de normas muy similares, elevadas y coeficientes RV altos. Además, todos los puntos están próximos al círculo de correlaciones, y por tanto bien representados, lo que indica que la relación global entre las variables que forman cada tabla es similar. En todos los casos el ángulo con el compromiso es pequeño, así como el que forman los objetos entre sí, lo que indicaría la existencia de una estructura común que es precisamente dicho compromiso. Dado que las coordenadas de las

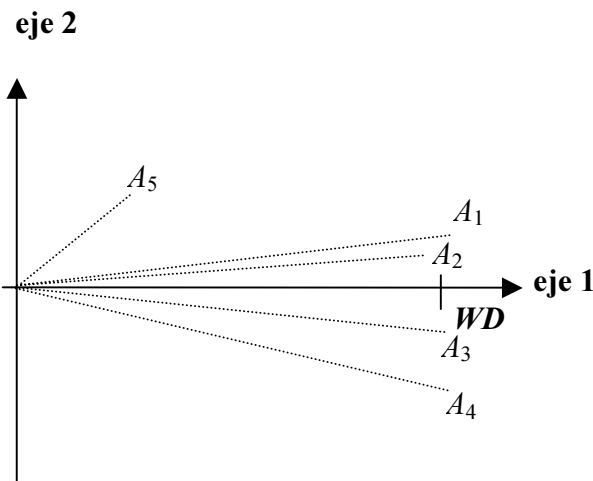
tablas con el primer eje son similares para todas, las ponderaciones de los diferentes objetos en la construcción del compromiso también lo serán.

Figura II.16
Existe una estructura común al conjunto de tablas



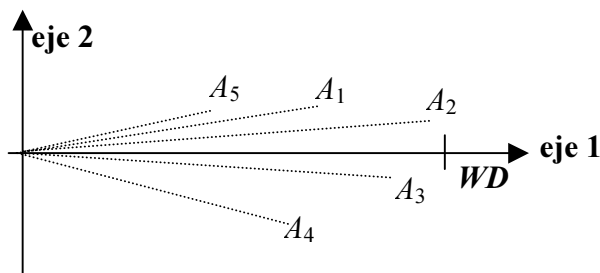
— En el segundo caso (figura II.17) la tabla a la que corresponde el punto A_5 no está bien representada (su norma es pequeña y próxima al origen). Además, el ángulo que forma con el compromiso y con el resto de las tablas es grande, por lo que su estructura es diferente a la de las demás. El compromiso está construido mayoritariamente por éstas últimas, y la ponderación de la A_5 en el estudio será baja, como corresponde a su correlación con el primer eje. En este caso el compromiso se puede interpretar como un “resumen mayoritario” del conjunto de objetos.

Figura II.17
El compromiso no representa bien todos los objetos



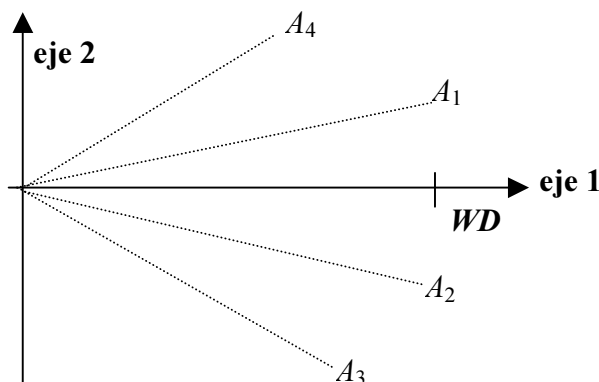
— En el tercer caso (figura II.18) el compromiso no constituye un buen resumen en cuanto que los objetos tienen normas muy diferentes, y por tanto sólo aquéllos con normas elevadas contribuyen a la construcción del mismo (serán las nubes de mayor tamaño y en las que las variables están más relacionadas globalmente). En este caso, deben emplearse objetos normados para equilibrar la influencia de los estudios en la construcción de la estructura común. Si los objetos normados están próximos unos de los otros (RV grandes) se obtiene una imagen similar a la primera.

Figura II.18
Objetos con normas muy diferentes



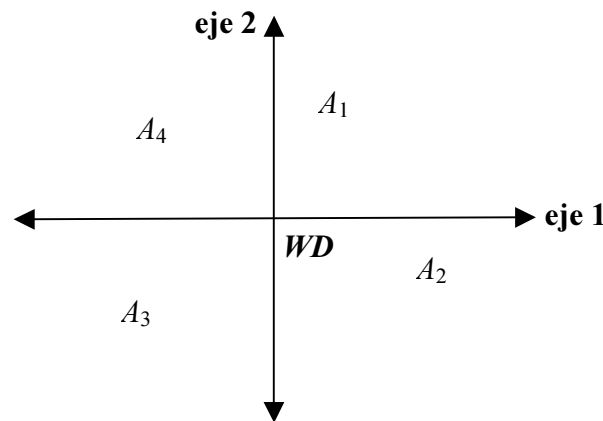
— Cuando los operadores presentan RV pequeños se obtiene una representación similar a la figura II.19, que representa el último caso típico. Dado que los ángulos entre el compromiso y los objetos son grandes, lo mismo que los ángulos que éstos forman entre sí, el compromiso solo es una media ponderada de los mismos y no existe una estructura común.

Figura II.19
No hay una estructura común



Cuando se analiza la *interestructura centrada* la representación obtenida muestra una imagen similar a la de la figura II.20, interpretándose las diferencias entre estudios respecto al origen, estableciendo distancias y proximidades entre ellos. Esta imagen complementa la obtenida con la interestructura no centrada.

Figura II.20
Interestructura centrada



Es necesario hacer una serie de precisiones respecto a los operadores: el espacio de operadores de $V_t M$ (ó $V_t M$ normado) tiene propiedades similares al espacio de operadores $W_t D$ (ó $W_t D$ normado). Además, cuando tenemos los mismos individuos y las mismas variables en todas las tablas, se deduce que:

$$\left\| \frac{V_t M}{\|V_t M\|_{HS}} \right\|_{HS}^2 = \left\| \frac{W_t D}{\|W_t D\|_{HS}} \right\|_{HS}^2 = 1 \quad \text{ó} \quad \|V_t M\|_{HS}^2 = \|W_t D\|_{HS}^2$$

Si bien, es preciso insistir en que el compromiso obtenido es de diferente naturaleza. Así WD es una matriz de productos escalares entre individuos (si se ha obtenido a partir de operadores normados también su norma será uno), mientras que VM es una matriz de covarianzas (o correlaciones si las variables están centradas y reducidas) en la que su norma será uno si se ha obtenido a partir de los $V_t M$ normados. La diferente naturaleza de esas matrices hace que su descomposición en valores singulares, en la etapa de

la intraestructura proporcione resultados diferentes aunque los individuos y las variables sean las mismas.

b. Intraestructura

Dado que el estudio de la intraestructura supone realizar un ACP *centrado* de una tabla, de tamaño $n \times \sum_t p_t$ formada por las tablas yuxtapuestas y ponderadas en el *Statis*, o de tamaño $\sum_t n_t \times p$ formada por las tablas apiladas y también ponderadas, la interpretación de las posiciones de los individuos (n individuos–compromiso en el *Statis* y $\sum_t n_t$ individuos en el *Statis dual*), y las posiciones de las variables ($\sum_t p_t$ variables en el *Statis* y p variables–compromiso en el *Statis dual*), se realiza teniendo en cuenta las normas de interpretación ya comentadas en dicho caso (epígrafe 1.4). A diferencia de este último, la representación simultánea de individuos y variables no suele realizarse, interpretándose cada una de ellas de manera individual: las variables, a través de las correlaciones con las componentes, dan sentido a los factores; los individuos se alejan o no del centro de gravedad según su comportamiento difiera o no del comportamiento medio. La característica peculiar de este análisis es que las posiciones de los individuos, en el *Statis*, resumen el comportamiento de los mismos para el conjunto de las tablas (o para el periodo considerado si las tablas se refieren a momentos del tiempo distintos), y en el caso del *Statis dual* son las correlaciones de las variables las que presentan la estructura común.

En el *Statis*, los factores tienen una interpretación más sencilla cuando los puntos representativos de una misma variable se agrupan respecto a esos factores, siendo más fácil dotarlos de un significado en función de todas las variables, facilitando también la interpretación del recorrido de las trayectorias a lo largo de dichos factores. Cuando las correlaciones entre las variables en el interior de una misma tabla son fuertes, los puntos representativos de éstas para una tabla se agruparán en torno a un factor, dificultando su interpretación, así como el de las trayectorias.

Por su parte, en el *Statis dual*, dado que las coordenadas de los individuos correspondientes a las tablas no se obtienen como elementos suplementarios sino como coordenadas de individuos activos resultado de un ACP, la representación de éstos mejora y las distancias entre ellos se interpretan en función del significado de los factores, que se ha otorgado en función de las correlaciones de las variables–compromiso.

c. Trayectorias

Para la correcta interpretación de las trayectorias de los individuos y de las variables, obtenidas al ser tratados como elementos suplementarios en el *Statis* y en el *Statis dual* respectivamente, son necesarias las siguientes recomendaciones.

Conviene recalcar que el estudio de las trayectorias así obtenidas se suele limitar al examen de las mismas sobre los dos primeros ejes, debido a que los coeficientes λ_k^c son cada vez más pequeños y las distancias se distorsionan cada vez más en los ejes sucesivos. Teniendo en cuenta que la posición–compromiso es el centro de gravedad de los puntos que constituyen la trayectoria ponderados por los α_t correspondientes, las trayectorias están influenciadas por esos coeficientes acentuándose la proximidad de los individuos–compromiso (o de las variables) con los individuos (o las variables) de las tablas más relacionadas con la estructura común, cualquiera que sea el eje factorial. Como también se ha señalado ya, dado que:

$$\sum_{t=1}^T \alpha_t \neq 1$$

los T puntos representativos de un individuo (o de una variable) pueden coincidir y sin embargo no coincidirían con la posición compromiso.

En el caso de los individuos, puesto que se trata de elementos suplementarios, las trayectorias se interpretan con respecto a la evolución media, es decir con respecto a la evolución de un individuo ficticio formado por los valores medios de las variables de cada tabla. Al estar las variables centradas, la trayectoria del individuo medio es el origen del plano del compromiso.

Existen dos formas típicas de trayectorias, según la organización de los puntos que las forman en torno a la posición *compromiso*:

- Una trayectoria poco extensa y que gira en torno al individuo-compromiso representa un individuo que sigue la evolución media, es decir, para todas las variables, la desviación entre la media y el valor de la variable es regular de unos años a otros.
- Una trayectoria de gran amplitud refleja un cambio en la posición del individuo a lo largo de las tablas (o periodos considerados si es el caso) diferente a la evolución media. Si los factores están dotados de significado, a través de sus correlaciones con las variables iniciales, la modificación del comportamiento del individuo se interpretará en función del factor respecto al que se desplace.

Por este motivo, algunos autores como Morineau (CISIA, 1997), aconsejan sustituir el análisis de las trayectorias por la descomposición de las distancias entre nubes en contribuciones de los individuos. Esta descomposición algebraica permite detectar los individuos con una trayectoria anormal, que serán los que tienen un porcentaje de inercia elevada. Para determinar en qué sentido evoluciona dicha trayectoria se podría realizar un *Análisis de Componentes Principales* de la tabla formada por las diferentes tablas yuxtapuestas (sin ponderar) puesto que dicho plano es óptimo para representar el conjunto de puntos que forman dicha trayectoria. Según este planteamiento, el cambio de posición de un individuo e_i de la tabla t a la tabla t' se representa:

$$\sum_{i'=1}^n p_{i'} [(W_t D)_{i'i'} - (W_{t'} D)_{i'i}]^2$$

Para todos los individuos se tiene que:

$$\sum_{i=1}^n p_i \sum_{i'=1}^n p_{i'} [(W_t D)_{i'i'} - (W_{t'} D)_{i'i}]^2 = \|W_t D - W_{t'} D\|_{HS}^2$$

que es la distancia al cuadrado entre tablas, por lo que la contribución del individuo e_i a la distancia entre tablas es:

$$\frac{\sum_{i'=1}^n p_{i'} [(W_t D)_{i'i'} - (W_{t'} D)_{i'i}]^2}{\sum_{i=1}^n p_i \sum_{i'=1}^n p_{i'} [(W_t D)_{i'i'} - (W_{t'} D)_{i'i}]^2}$$

Si se consideran todas las nubes tenemos que para el individuo e_i la expresión:

$$\sum_{t=1}^T \sum_{t'=1}^T \sum_{i'=1}^n p_{i'} [(W_t D)_{i'i'} - (W_{t'} D)_{i'i}]^2$$

representa la dispersión, en la configuración de la interestructura que es debida a dicho individuo e_i .

De todos modos, y según recoge Lavit *et al.* (1994: 107), a pesar de que el análisis del gráfico de las trayectorias no es óptimo, el vector de la trayectoria sí refleja de manera adecuada la dirección y la importancia del cambio de posición del individuo de un estudio a otro, y ésta es la interpretación que debe hacerse al observar el gráfico.

En el caso de las variables, al tratarse de elementos suplementarios, no han intervenido en la construcción de los factores, pero es posible obtener su coeficiente de correlación con los mismos y por tanto interpretar su posición. Las variables que presenten trayectorias cuyo comportamiento difiere del comportamiento medio serán las responsables de las diferencias entre tablas.

Este análisis puede ser completado con la descomposición de las distancias entre nubes en contribuciones de las variables, detectando las variables

con una trayectoria anormal, que serán las que tienen un porcentaje de inercia elevada. Según este planteamiento, el cambio de posición de una variable Y_j de la tabla t a la tabla t' se representa por:

$$\sum_{j'=1}^p m_{j'} [(V_t M)_{jj'} - (V_{t'} M)_{jj'}]^2$$

Para las p variables se tiene que:

$$\sum_{j=1}^p m_j \sum_{j'=1}^p m_{j'} [(V_t M)_{jj'} - (V_{t'} M)_{jj'}]^2 = \|V_t M - V_{t'} M\|_{HS}^2$$

que es la distancia al cuadrado entre tablas, por lo que la contribución de la variable Y_j a la distancia entre dos tablas es:

$$\frac{\sum_{j'=1}^p m_{j'} [(V_t M)_{jj'} - (V_{t'} M)_{jj'}]^2}{\sum_{j=1}^p m_j \sum_{j'=1}^p m_{j'} [(V_t M)_{jj'} - (V_{t'} M)_{jj'}]^2}$$

Si se consideran todas las nubes, tenemos que para la variable Y_j la expresión:

$$\sum_{t=1}^T \sum_{t'=1}^T \sum_{j'=1}^p m_{j'} [(V_t M)_{jj'} - (V_{t'} M)_{jj'}]^2$$

representa la dispersión, en la configuración de la interestructura que es debida a la variable Y_j .

Es conveniente señalar que el estudio de las trayectorias para las variables no aporta conclusiones significativas si los ejes ya han sido dotados de significado convenientemente a través del estudio de las correlaciones.

Remitiéndonos al análisis de la intraestructura, conviene señalar que algunos de los inconvenientes de las trayectorias pueden ser paliados en aquellos casos en los que los mismos individuos y las mismas variables son considerados en todas las tablas, siendo posible la aplicación de las dos versiones *Statis* y *Statis dual*.

Para terminar este desarrollo teórico y a modo de resumen, se esquematizan a continuación las fases de que constan y los resultados fundamentales de las mismas.

Figura II.21
Statis y Statis dual

	<i>Statis</i>	<i>Statis dual</i>
TIPO DE DATOS	mismos individuos × variables (mismas o diferentes variables)	individuos × mismas variables (mismos o diferentes individuos)
OBJETOS REPRESENTATIVOS DE LAS TABLAS	$\frac{W_t D}{\ W_t D\ _{HS}}$ ó $W_t D$	$\frac{V_t M}{\ V_t M\ _{HS}}$ ó $V_t M$
INTERESTRUCTURA	<ul style="list-style-type: none"> • Comparación global de las nubes de individuos • Descomposición de matriz S: $s_{tt'} = RV(W_t D, W_{t'} D) = \frac{Tr(W_t D W_{t'} D)}{\ W_t D\ _{HS} \ W_{t'} D\ _{HS}}$ o de la matriz S: $s_{tt'} = Cov(W_t D, W_{t'} D) = Tr(W_t D W_{t'} D)$ 	<ul style="list-style-type: none"> • Comparación global de las nubes de variables • Descomposición de matriz S: $s_{tt'} = RV(V_t M, V_{t'} M) = \frac{Tr(V_t M V_{t'} M)}{\ V_t M\ _{HS} \ V_{t'} M\ _{HS}}$ o de la matriz S: $s_{tt'} = Cov(V_t M, V_{t'} M) = Tr(V_t M V_{t'} M)$
COMPROMISO	Nube media de individuos WD $WD = \sum_{t=1}^T \alpha_t \frac{W_t D}{\ W_t D\ _{HS}}$ $WD = \sum_{t=1}^T \alpha_t W_t D$	Nube media de variables VM $VM = \sum_{t=1}^T \alpha_t \frac{V_t M}{\ V_t M\ _{HS}}$ $VM = \sum_{t=1}^T \alpha_t V_t M$
INTRAESTRUCTURA	<ul style="list-style-type: none"> • Coordenadas de n individuos–compromiso $WD F_k^c = \lambda_k^c F_k^c$ • Correlaciones de $\sum_t p_t$ variables a partir de las G_k: $G_k = \Pi' D t_k^c \text{ ó } G_k = \Psi' D t_k^c$ 	<ul style="list-style-type: none"> • Correlaciones de p variables–compromiso $VM G_k^c = \lambda_k^c G_k^c$ • Coordenadas de $\sum_t n_t$ individuos $F_k = \Pi' u_{ki}^c \text{ ó } F_k = \Psi u_k^c$
TRAYECTORIAS Elementos suplementarios	Coordenadas de $n \times T$ individuos $F_k^t = \frac{W_t D}{\ W_t D\ _{HS}} \frac{t_k^c}{\sqrt{\lambda_k^c}} \text{ ó } F_k^t = W_t D \frac{t_k^c}{\sqrt{\lambda_k^c}}$	Coordenadas de $p \times T$ variables $G_k^t = \frac{V_t M}{\ V_t M\ _{HS}} \frac{u_k^c}{\sqrt{\lambda_k^c}} \text{ ó } G_k^t = V_t M \frac{u_k^c}{\sqrt{\lambda_k^c}}$

3. Posibilidades descriptivas y predictivas de STATIS

Realizado el desarrollo teórico de los métodos que engloba la metodología STATIS, se procede a analizar su significado para el tratamiento de la información. De este forma, se completa este capítulo de nuestro trabajo con este apartado en el que trataremos de especificar las posibilidades que tiene la metodología desarrollada para describir y predecir. Con ello se pretende mejorar el conocimiento y obtener conclusiones más relevantes.

Primeramente nos referiremos a su utilidad como técnica descriptiva, tomando en consideración que una de sus etapas es un caso especial del *Análisis de Componentes Principales*, como quedó puesto de manifiesto en los apartados anteriores. Completaremos esta parte con una breve descripción del *Análisis Cluster* como técnica que permite complementar los resultados descriptivos del *Statis*.

Con relación al carácter predictivo, se destaca la conexión entre la metodología que nos ocupa y la teoría de los procesos estocásticos. Es necesario, por tanto, una breve descripción de tales procesos, especificando de forma más detallada los modelos ARIMA por su importancia en la predicción.

Ambos aspectos conducen a resultados fundamentales, poniendo de manifiesto la utilidad de la metodología STATIS para el análisis de datos, sin olvidar la importancia que tiene para cualquier estudio estadístico la predicción de resultados, razón por la cual se enfatizará en mayor medida en las posibilidades predictivas.

3.1 Descripción y STATIS

3.1.1 Introducción

La capacidad descriptiva del STATIS va a ser abordada desde dos puntos de vista diferentes. En primer lugar, se van a considerar las ventajas de aplicar las dos versiones de la metodología cuando los individuos y las variables son las mismas para todas las ocasiones, lo que induce a que los resultados obtenidos sean más interesantes, si bien es necesario hacer algunas precisiones para su correcta interpretación.

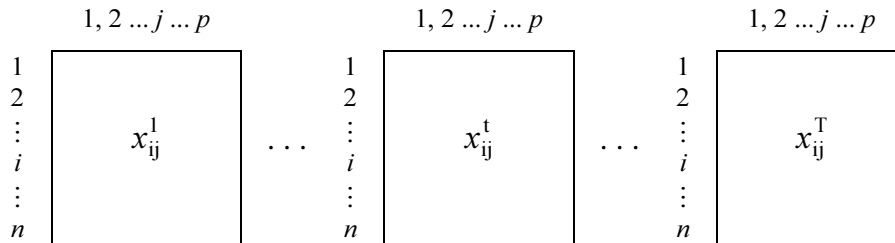
Ya se ha indicado que se trata de un caso especial de *Análisis de Componentes Principales*, por lo que los resultados del mismo pueden emplearse para posteriores estudios con *Análisis Cluster* o *Discriminante*, habiendo reducido la dimensionalidad de los datos con el primero. Por ello, en segundo lugar, se analizará la aplicación del STATIS como paso previo cuando el objetivo es la clasificación de una serie de individuos, contribuyendo a la descripción de los resultados obtenidos.

Cuando los individuos y las variables de las tablas de datos que componen la tabla total son los mismos, la aplicación de las dos versiones del STATIS permite obtener resultados complementarios a los ya especificados.

Para ello se considera la tabla de datos formada por n individuos y Tp variables: T tablas de doble entrada, con n filas y p columnas cada una, según recoge la figura siguiente:

Figura II.22

Tablas de datos con individuos y variables comunes



Cada tabla $X^1, X^2 \dots X^T$ representa un momento del tiempo diferente, en cada uno de los cuales se han observado las mismas variables sobre los mismos individuos. Si las variables se centran y se reducen, tal que:

$$Z_j^t = \frac{X_j^t - \bar{X}_j^t}{S_{x_j^t}}$$

tenemos las tablas de datos $Z^1, Z^2 \dots Z^T$. Por otra parte, los pesos de los individuos y los pesos de las variables serán comunes a todas las tablas de datos y, por tanto, también las matrices M y D , que representan las métricas.

Hay que tener en cuenta que si se consideran las nubes de individuos, cada tabla estará representada por $W_t D$ (ó $W_t D$ normado según el caso) y el método apropiado para su comparación es el *Statis*; al tratar con las nubes de variables tenemos $V_t M$ (ó $V_t M$ normado) y el método a aplicar es el *Statis dual*.

Como ya se demostró, cuando se consideran operadores normados, las distancias, entre las nubes de individuos o entre las nubes de variables, dependen del coseno del ángulo que forman los operadores (es decir, de los coeficientes RV), que coincide con el producto escalar entre esos operadores normados. Aunque se cumple que:

$$1 = \left\| \frac{W_t D}{\|W_t D\|_{HS}} \right\|_{HS}^2 = \left\| \frac{V_t M}{\|V_t M\|_{HS}} \right\|_{HS}^2 \quad \text{y} \quad 1 = \left\| \frac{W_t D}{\|W_t D\|_{HS}} \right\|_{HS}^2 = \left\| \frac{V_t M}{\|V_t M\|_{HS}} \right\|_{HS}^2$$

y, debido a que los individuos y las variables son comunes en todas las tablas, se tiene que:

$$\|W_t D\|_{HS}^2 = \|V_t M\|_{HS}^2 \quad \text{y} \quad \|W_{t'} D\|_{HS}^2 = \|V_{t'} M\|_{HS}^2$$

el coeficiente RV entre dos nubes de individuos t y t' no coincide con el coeficiente RV entre las nubes de variables t y t' . Esto se debe a que el numerador de dicho coeficiente, es decir, el producto escalar entre los operadores no normados, es diferente en uno y otro caso y, por tanto, también las distancias y los resultados de la interestructura. Las diferencias se deben a que en el caso de las nubes de individuos, el producto escalar es la suma de los productos escalares de los individuos en una tabla por los productos escalares de los individuos en la otra, mientras que en el caso de las nubes de variables ese producto escalar se interpreta como la suma de los productos de las correlaciones de las variables (por estar centradas y reducidas) de una tabla por las de la otra. Por el mismo motivo, la interestructura con operadores no normados tampoco permite obtener los mismos resultados para las nubes de individuos y de variables.

Dado que su significado es diferente, la matriz S analizada por el *Statis* (con operadores normados o no) es diferente de la matriz S analizada por el *Statis dual*. La descomposición en valores propios de esas matrices tampoco proporciona los mismos resultados, aunque como se pondrá de manifiesto en el capítulo práctico, permite extraer conclusiones similares: si el comportamiento de los individuos varía de una tabla a otra y, por tanto, dan lugar a nubes distintas, se debe a que éstos presentan valores diferentes para el conjunto de variables, y ese hecho debe reflejarse al comparar las nubes de variables.

Como consecuencia de las diferencias apuntadas, la ponderación de la nube de individuos t en el compromiso obtenido con el *Statis* (normado o no) no coincide con la ponderación que la nube de variables t presenta en el compromiso del *Statis dual*. Además, en el primer caso, el compromiso viene dado por una nube de individuos cuyo objeto representativo es WD ,

mientras que en el caso del *Statis dual* se obtiene una nube de variables representada por VM .

Parece conveniente volver a señalar que, debido a esa naturaleza diferente, el análisis del operador WD en valores y vectores propios es similar al análisis de la nube de individuos en un ACP *centrado*, obteniendo las coordenadas de los n individuos—compromiso y luego las correlaciones de las $p \times T$ variables; en el caso del operador VM su análisis es similar al ACP *centrado* o *normado* (según el caso, como ya se indicó) de la nube de variables, obteniendo las correlaciones de las p variables—compromiso y luego las de los $n \times T$ individuos.

Teniendo en cuenta que todas las tablas tienen las mismas p variables, cuando las T correlaciones de una de ellas se agrupan en torno a un factor del *Statis*, (es decir, presentan todas un valor elevado y del mismo signo) la correlación de la variable correspondiente presenta un valor elevado (y del mismo signo) con el respectivo factor del *Statis dual*, por lo que los factores obtenidos en uno y otro caso tienen un significado similar. Además, los T puntos correspondientes a cada una de las variables en el *Statis* permiten definir la trayectoria de cada variable en un espacio óptimo, matizando el significado otorgado a los factores, lo que es una alternativa a las trayectorias obtenidas con el *Statis dual*. Por el mismo motivo, la representación gráfica de los $n \times T$ individuos obtenida con el *Statis dual*, en la que a cada individuo n le corresponden T puntos, permite definir las trayectorias de los mismos en un espacio óptimo, interpretándose como elementos activos, en lugar de considerar las obtenidas como suplementarios con el *Statis*, eliminando también la distorsión debida a los coeficientes λ_k^c .

No obstante, la descomposición de distancias, por individuos o por variables, a la que se hizo referencia al considerar las trayectorias en uno y otro caso, sigue siendo un complemento imprescindible para su correcta interpretación.

3.1.2 La Clasificación y el STATIS

Cuando el objetivo final de un estudio es la clasificación de una serie de individuos, una técnica factorial, y por tanto el STATIS, puede ser empleada para reducir la dimensionalidad de los datos, siendo necesaria posteriormente una técnica de clasificación como puede ser el *Análisis Cluster*. La necesidad de clasificación, y no solo de descripción, se plantea en muchas situaciones desde el punto de vista estadístico. Lebart *et al.* (1995: 145) justifican dicha necesidad en aquellos casos en los que resulta interesante no solo una visualización general de las asociaciones estadísticas, sino que hay un interés manifiesto en poner en evidencia clases o grupos de individuos o de caracteres.

En este sentido, es un hecho probado que la clasificación es fundamental en la mayoría de las ramas de la ciencia. En Biología, por ejemplo la clasificación de organismos ha sido una preocupación desde las primeras investigaciones. Así, la mayoría de las técnicas numéricas para obtener clasificaciones se originaron en este ámbito de las ciencias naturales. Si bien, desde la segunda mitad del siglo XX, la necesidad de clasificación se ha hecho patente en muchos ámbitos del saber, como por ejemplo en la Arqueología, en Psiquiatría, en Marketing o en Economía de la Empresa, lo que ha dado lugar a una pluralidad de métodos. En concreto, en este último caso, la necesidad de establecer grupos de empresas con estrategias similares para analizar un sector industrial (Ventura Victoria, 1994: 163), ha dado lugar al empleo de técnicas sofisticadas de clasificación que permiten determinar las variables relevantes para caracterizar las distintas empresas, y clasificarlas según su comportamiento estratégico.

La elección de la Técnica de Clasificación apropiada depende del tipo de datos. En concreto, la aplicación del *Análisis Cluster* está especialmente indicada cuando de todo lo que se dispone es de una colección de observaciones, y hay poca o ninguna información sobre los posibles grupos.

Como se expondrá más adelante, esta denominación recoge una amplia variedad de procedimientos que pueden ser usados para crear cierta clasificación. Así, un *Método o Técnica Cluster* es un procedimiento estadístico multivariante que comienza sobre una muestra de entidades e intenta reorganizarlas en grupos relativamente homogéneos a los que se denomina *clusters*.

Tal y como manifiesta Lebart *et al.* (1995: 147), estos métodos de clasificación no pueden mostrar siempre la importancia de ciertas tendencias existentes, y el empleo previo de métodos factoriales, permite el posterior posicionamiento de las clases o grupos sobre los ejes factoriales, y la caracterización de esas tendencias o fenómenos latentes.

Los Métodos de Clasificación también resuelven ciertas limitaciones de los Métodos Factoriales, en aquellos casos en los que hay muchos puntos sobre las gráficas resultantes convirtiéndose casi en ilegibles (Lebart *et al.*, 1995: 185-6), ya que la organización de individuos en grupos homogéneos, facilita su interpretación. Si la clasificación se hace sobre el subespacio definido por los factores más significativos, que recogen la mayor parte de la información, los grupos obtenidos consideran la dimensión real de la nube de puntos, y corrigen ciertas deformaciones debidas a la proyección en planos de como mucho tres dimensiones. De esa manera ambos tipos de métodos se validan mutuamente.

A continuación se expone brevemente el procedimiento básico del *Análisis Cluster*.

a. Análisis Cluster

El *Análisis Cluster* define una serie de técnicas, fundamentalmente algoritmos, que permiten la determinación de grupos similares. El objetivo global consiste en clasificar una serie de individuos, en grupos lo más

homogéneos posibles según las variables observadas, pudiéndose obtener también un conjunto de variables similares, a partir de una serie de individuos.

Las técnicas que se incluyen dentro del *Análisis Cluster* han sido desarrolladas a lo largo de este siglo, aunque la mayoría de la literatura sobre este tema se ha escrito durante las tres décadas pasadas. El libro pionero fue *Principios de Taxonomía Numérica* publicado en 1963 por los biólogos Sokal y Sneath, y su gran desarrollo comienza a partir de este momento debido al avance de los ordenadores y a la importancia de la clasificación como procedimiento científico.

El punto de partida es una matriz de datos formada por los valores de las variables para cada uno de los individuos que constituyen el objeto del estudio, de la forma:

Figura II.23
Tabla de n individuos p variables

	1	2	...	j	...	p
1	x_{ij}					
2						
⋮						
i						
⋮						
n						

Una vez que se tiene la información expresada en la matriz de datos, se procede a la aplicación del método. Esquemáticamente, las etapas a seguir cuando se quiere obtener una clasificación de individuos son las siguientes:

1. SELECCIÓN DE VARIABLES

La primera etapa consiste en elegir las características o variables relevantes para identificar los grupos o clusters. En este sentido es importante tener en cuenta que sólo hay limitadas directrices matemáticas y estadísticas; de aquí que esta etapa esté, en parte, condicionada por la opinión del investigador acerca de su propósito de clasificación.

Una vez seleccionadas las variables hay que tener presente los siguientes problemas: número de las mismas, unidades de medida y la relación entre ellas.

Número de variables

En muchas aplicaciones es probable que el investigador se equivoque tomando demasiadas variables, lo cual puede dar origen a diversos problemas de carácter computacional o bien porque tantas variables enmascaren la estructura de los grupos.

En este caso la aplicación de una técnica factorial, como hemos indicado, resuelve el problema de exceso de dimensionalidad.

Unidades de medida

En la mayoría de las aplicaciones las variables consideradas no suelen estar medidas en las mismas unidades. La solución para variables cuantitativas consiste en tipificar las variables previamente, calculando las desviaciones típicas a partir de todos los individuos.

Cuando las variables son de tipos diferentes se suelen transformar todas en binarias, lo que resuelve el problema, aunque a consta de perder bastante información. Una alternativa es usar lo que se denomina coeficiente de similitud, que puede incorporar información de diferentes tipos de variables de una forma sensible. Asimismo, para variables mixtas existe la posibilidad de hacer un análisis por separado e intentar sintetizar los resultados a partir de los diferentes estudios.

Relación entre variables

La aplicación de una técnica factorial (como el STATIS) en este caso resuelve el problema y permite determinar variables no correlacionadas entre sí (los factores).

2. ELECCIÓN DE LA MEDIDA DE ASOCIACIÓN

La determinación de los grupos (tanto de individuos como de variables) lleva implícita una medida que permite analizar la proximidad óptima de los objetos en estudio. Cuando se agrupan individuos, la proximidad se suele expresar en términos de distancia, mientras que en el caso de las variables se trata de medidas del tipo de coeficientes de correlación.

Medidas de asociación entre variables

Son aquellas que nos permiten agrupar las variables consideradas, pero teniendo en cuenta que cada medida refleja asociación en un sentido particular, lo que significa que es necesario elegir una medida apropiada para el problema concreto que se esté tratando. Las más usuales son:

- *Coseno del ángulo de vectores*

Se trata de una medida de similaridad entre variables que toma valores entre -1 y 1 en virtud de la desigualdad de Schwarz. Se define mediante la siguiente expresión:

$$\text{Cos}(\beta) = \frac{x_j' x_j}{\|x_j'\| \|x_j\|} = \frac{\sum_{k=1}^n x_{kj}' x_{kj}}{\left(\sum_{k=1}^n x_{kj}'^2 \sum_{k=1}^n x_{kj}^2 \right)^{1/2}}$$

siendo $x_j = (x_{1j} \dots x_{nj})'$ $x_j' = (x_{1j}' \dots x_{nj}')$

los vectores cuyas k -ésimas componentes indican el valor de la variable correspondiente al k -ésimo individuo. Además, corresponden a las variables X_j y X_j' que se han muestreado sobre n individuos.

Se considera esta medida como la mejor para establecer paralelismos entre dos vectores, ya que dos vectores son paralelos cuando el coseno del ángulo que forman es cero.

- *Coefficiente de correlación*

Considerando dos variables genéricas X_j y $X_{j'}$, medidas sobre los n individuos objeto de estudio y siendo x_j y $x_{j'}$ los vectores asociados a ellas, se define la correlación muestral entre ambos, como:

$$r = \frac{Cov(x_j, x_{j'})}{[\text{Var}(x_j) \text{Var}(x_{j'})]^{1/2}} = \frac{\sum_{k=1}^n (x_{kj} - \bar{x}_j)(x_{kj'} - \bar{x}_{j'})}{\left[\sum_{k=1}^n (x_{kj} - \bar{x}_j)^2 \sum_{k=1}^n (x_{kj'} - \bar{x}_{j'})^2 \right]^{1/2}}$$

Se puede observar que dicha expresión corresponde al coseno del ángulo entre los vectores centrados \hat{x}_j y $\hat{x}_{j'}$:

$$\hat{x}_j = (x_{1j} - \bar{x}_j \dots x_{nj} - \bar{x}_j) \quad \hat{x}_{j'} = (x_{1j'} - \bar{x}_{j'} \dots x_{nj'} - \bar{x}_{j'})'$$

Con relación a las dos medidas descritas se deduce que el coeficiente de correlación posee una varianza más restrictiva que el coseno a la hora de establecer diferencias entre dos variables. No obstante, la diferencia esencial radica en que el coseno se basa en datos originales y el coeficiente de correlación en datos centrados respecto a la media. De aquí se deduce que si el origen está bien establecido es más apropiado el coseno y en caso contrario el coeficiente de correlación.

En el caso de disponer de variables dicotómicas, existen también medidas adecuadas pero dado que no será nuestro caso, no van a ser consideradas.

Medidas de asociación entre individuos

De una forma análoga a como se definieron las medidas entre variables se realiza para individuos. En este caso se consideran dos vectores fila que corresponden a los individuos:

$$e_i = (x_{i1} \dots x_{ip}) \quad e_{i'} = (x_{i'1} \dots x_{i'p})'$$

Las medidas más utilizadas son las siguientes:

- *Distancia euclídea*

$$d(e_i, e_{i'}) = \sqrt{(x_i - x_{i'})'(x_i - x_{i'})} = \sqrt{\sum_{j=1}^p (x_{ij} - x_{i'j})^2}$$

- *Distancia de Minkowski*

$$d(e_i, e_{i'}) = \left(\sum_{j=1}^p |x_{ij} - x_{i'j}|^p \right)^{1/p}$$

- *Distancia de Mahalanobis*

$$D_V(e_i, e_{i'}) = \sqrt{(x_i - x_{i'})' V^{-1} (x_i - x_{i'})}$$

donde V es la matriz de varianzas-covarianzas, que viene dada por la siguiente expresión:

$$V = \frac{1}{n} Y' Y$$

$$y_{ij} = x_{ij} - \bar{x}_j \quad \text{con} \quad i = 1 \dots n ; j = 1 \dots p$$

Esta distancia presenta la desventaja de que el cálculo de V está basado en todos los individuos de forma conjunta y no trata separadamente los objetos de cada cluster. No se suele emplear porque su cálculo es más laborioso que el de otras métricas.

También es posible definir una medida de *correlación entre individuos* y otras *medidas derivadas de la χ^2 y medidas no métricas*.

3. ELECCIÓN DE LA TÉCNICA CLUSTER

Una vez que se ha especificado la medida de agrupamiento, el siguiente paso corresponde a la selección de la técnica cluster más adecuada para el estudio. Aunque existen diversas técnicas clusters, todas ellas se incluyen bien, en los *métodos jerárquicos*, o en los *no jerárquicos*. En general, la

elección de la técnica depende de la naturaleza de los datos y de los objetivos perseguidos. Sin embargo, lo que es importante y conveniente en la práctica es considerar más de una técnica con el fin de contrastar resultados.

Métodos jerárquicos

Su objetivo consiste en agrupar clusters para formar uno nuevo o bien separar alguno ya existente para dar origen a otros dos, de forma que, si sucesivamente se va efectuando este proceso de aglomeración o división, se minimice alguna distancia o se maximice alguna medida de similitud.

Estos métodos permiten la construcción de un árbol de clasificación, denominado *DENDOGRAMA*, en el que se observan los grupos que se van uniendo, en qué nivel y el valor de la medida de asociación.

Se subdividen en *aglomerativos* y *disociativos*, de forma que cada uno de ellos presenta distintas variantes:

- ***Métodos aglomerativos***

Comienzan el análisis con tantos grupos como individuos haya. Se van formando grupos, de forma ascendente, hasta que al final todos los casos están englobados en un mismo conglomerado. Cabe destacar los siguientes:

- *De la distancia mínima o similitud máxima*

En este método se considera que la distancia o similitud entre dos clusters viene dada, respectivamente, por la mínima distancia (o máxima verosimilitud) entre sus componentes.

La distancia entre dos cluster viene dada por la siguiente expresión:

$$d(C_i, C_j) = \min\{d(e_r, e_s)\} \quad \begin{array}{ll} e_r \in C_i & r = 1, 2 \dots n_i \\ e_s \in C_j & s = 1, 2 \dots n_j \end{array}$$

siendo C_i un cluster con n_i elementos y C_j con n_j elementos.

La similitud, por su parte, se expresa de la forma siguiente:

$$S(C_i, C_j) = \max\{s(e_r, e_s)\} \quad \begin{array}{l} e_r \in C_i \quad r = 1, 2 \dots n_i \\ e_s \in C_j \quad s = 1, 2 \dots n_j \end{array}$$

– *De la distancia máxima o similitud mínima*

En este método se considera que la distancia o similitud entre dos clusters se mide atendiendo a sus elementos más dispares, es decir, mediante la máxima distancia o mínima similitud.

De análoga forma al caso anterior, la distancia entre dos clusters será:

$$d(C_i, C_j) = \max\{d(e_r, e_s)\} \quad \begin{array}{l} e_r \in C_i \quad r = 1, 2 \dots n_i \\ e_s \in C_j \quad s = 1, 2 \dots n_j \end{array}$$

y la similitud se calcula mediante la siguiente expresión:

$$S(C_i, C_j) = \min\{s(e_r, e_s)\} \quad \begin{array}{l} e_r \in C_i \quad r = 1, 2 \dots n_i \\ e_s \in C_j \quad s = 1, 2 \dots n_j \end{array}$$

– *De la distancia, o similitud, promedio no ponderada*

Este método determina la distancia (o similitud) entre dos clusters como la media aritmética entre la distancia, o similitud, de las componentes de dichos clusters.

Esto significa que si el cluster C_i (con n_i elementos) está formado por dos clusters C_{i1} y C_{i2} (con n_{i1} y n_{i2} elementos respectivamente) y el cluster C_j por n_j elementos, la distancia o similitud entre ellos se calcula:

$$d(C_i, C_j) = \frac{d(C_{i1}, C_j) + d(C_{i2}, C_j)}{2}$$

lo que indica que este método no tiene en cuenta el tamaño de los clusters involucrados.

– *De la distancia, o similitud, promedio ponderada*

Para este método, la distancia o similitud entre dos clusters se define como el promedio ponderado de las distancias o similitudes de los componentes de un cluster respecto a los del otro. Tenemos que:

$$\begin{aligned} e_r &\in C_i & r &= 1, 2 \dots n_i \\ e_s &\in C_j & s &= 1, 2 \dots n_j \end{aligned}$$

Si se consideran dos clusters C_i y C_j de manera que C_i está formado a su vez por C_{i1} y C_{i2} , con n_{i1} y n_{i2} elementos respectivamente, y $n_i = n_{i1} + n_{i2}$ entonces la distancia promedio ponderada será:

$$\begin{aligned} d(C_i, C_j) &= \frac{1}{(n_{i1} + n_{i2})n_j} \sum_{r=1}^{n_{i1}+n_{i2}} \sum_{s=1}^{n_j} d(e_r, e_s) = \\ &= \frac{1}{(n_{i1} + n_{i2})n_j} \sum_{r_1=1}^{n_{i1}} \sum_{s=1}^{n_j} d(e_{r_1}, e_s) + \\ &\quad + \frac{1}{(n_{i1} + n_{i2})n_j} \sum_{r_2=1}^{n_{i2}} \sum_{s=1}^{n_j} d(e_{r_2}, e_s) = \\ &= \frac{n_{i1}}{(n_{i1} + n_{i2})n_{i1}n_j} \sum_{r_1=1}^{n_{i1}} \sum_{s=1}^{n_j} d(e_{r_1}, e_s) + \\ &\quad + \frac{n_{i2}}{(n_{i1} + n_{i2})n_{i2}n_j} \sum_{r_2=1}^{n_{i2}} \sum_{s=1}^{n_j} d(e_{r_2}, e_s) = \\ &= \frac{n_{i1}}{n_{i1} + n_{i2}} d(C_{i1}, C_j) + \frac{n_{i2}}{n_{i1} + n_{i2}} d(C_{i2}, C_j) = \\ &= \frac{n_{i1} d(C_{i1}, C_j) + n_{i2} d(C_{i2}, C_j)}{n_{i1} + n_{i2}} \end{aligned}$$

de lo que se deduce que $d(C_i, C_j)$ es el promedio ponderado de las distancias de cada uno de los dos clusters C_{i1} y C_{i2} con respecto al cluster C_j .

Existen otros métodos como son los basados en el *centroide*, que se caracterizan porque la semejanza entre dos clusters viene dada por la

semejanza entre sus centroides, es decir, los vectores de medias de las variables medidas sobre los individuos del cluster. Se destacan entre ellos, el *Método del centroide ponderado* y el *Método de Ward*. Éste último, conocido también por la denominación "pérdida de inercia mínima", calcula la media de todas las variables de cada cluster, luego calcula la distancia euclídea al cuadrado entre cada individuo y la media de su grupo, y se suman las distancias de todos los casos; en cada paso los clusters que se forman son aquellos que resultan con el menor incremento en la suma total de las distancias al cuadrado intra-cluster.

- *Métodos disociativos*

Se caracterizan por seguir el proceso inverso a los aglomerativos en el sentido de que comienzan con un conglomerado que engloba a todos los casos y, a través de sucesivas divisiones, se forman grupos cada vez más pequeños.

La clasificación de estos métodos se hace siguiendo un procedimiento similar a los aglomerativos en cuanto a la forma de cálculo de la distancia entre grupos. Con relación a este aspecto, se maximizan distancias o se minimizan similitudes con el fin de buscar los individuos menos similares para separarlos del resto del conglomerado.

Esta clase de métodos es esencialmente de dos tipos: monotéticos, que dividen los datos sobre la base de un solo atributo y se suelen emplear cuando los datos son binarios; politéticos, cuyas divisiones se basan en los valores tomados por todas las variables.

Métodos no jerárquicos

También denominados partitivos o de optimización, su objetivo consiste en realizar una sola partición de los individuos en K grupos, lo que significa que el investigador debe especificar a priori los grupos que deben ser

formados. Esta es una de las diferencias con respecto a los métodos jerárquicos; otra diferencia importante reside en que la matriz de datos con la que se trabaja en los no jerárquicos es la original.

Estos métodos se utilizan para clasificar individuos (no siendo válidos para variables) en K clusters, donde K se especifica a priori. La idea general es similar a la señalada para los algoritmos descendentes en más de un paso utilizados en la optimización sin restricciones en programación no lineal.

No desarrollaremos los métodos no jerárquicos porque son menos utilizados que los jerárquicos, aunque reseñaremos una de las clasificaciones existentes:

- *Métodos de reasignación*

Se caracterizan porque permiten que un individuo asignado a un grupo en un determinado paso del proceso, sea reasignado a otro grupo en un paso posterior si ello optimiza el criterio de selección. El proceso termina cuando no quedan individuos cuya reasignación permita optimizar el resultado que se ha conseguido. Dentro de estos métodos están: *Método K -Medias*, *Quick-Cluster Analysis*, *Método de Forgy* y el *Método de las nubes dinámicas*.

- *Métodos de búsqueda de la densidad*

Se dividen en dos tipos: los que proporcionan una aproximación tipológica y los que proporcionan una aproximación probabilística.

En el primer tipo, los grupos se forman buscando las zonas en las cuales se da una mayor concentración de individuos. Se destacan entre ellos el *Análisis modal de Wishart*, el *Método Taxmap* y el *Método de Fortin*.

En el segundo tipo se parte de que las variables siguen una ley de probabilidad según la cual los parámetros varían de un grupo a otro. Se trata de encontrar los individuos que pertenecen a la misma

distribución. Entre los métodos de este tipo destaca el de las *combinaciones de Wolf*.

- *Métodos directos*

Se caracterizan porque permiten clasificar simultáneamente a los individuos y a las variables. El algoritmo más conocido es el *Block-Clustering*.

- *Métodos de reducción de dimensión*

Consisten en buscar factores en el espacio de individuos de manera que cada factor corresponde a un grupo. Se les conoce como *Análisis Factorial tipo Q*.

4. VALIDACIÓN E INTERPRETACIÓN DE RESULTADOS

Ambos aspectos constituyen la última fase de un estudio realizado mediante *Análisis Cluster*. Para validar los resultados obtenidos, es decir, para comprobar en qué medida la estructura final obtenida es válida, se utilizan diferentes métodos, que dependen de los que se hayan aplicado en las fases anteriores.

Métodos jerárquicos

Los métodos jerárquicos imponen una estructura sobre los datos y es necesario, con frecuencia, considerar si es aceptable o si se introducen distorsiones inaceptables en las relaciones originales. El método más usado para verificar este hecho, o sea, para ver la relación entre el dendograma y la matriz de proximidades original, es el Coeficiente de Correlación Cofenético, el cual es simplemente la correlación entre los $n(n-1)/2$ elementos de la parte superior de la matriz de proximidades observada y los correspondientes en la llamada matriz cofenética, C , cuyos elementos,

$c_{ii'}$, se definen como aquéllos que determinan la proximidad entre los elementos i y i' cuando éstos se unen en el mismo cluster.

Así, si tras el empleo de varios procedimientos cluster distintos, éstos conducen a soluciones parecidas surge la pregunta de qué método elegiremos como definitivo. La respuesta la da dicho coeficiente cofenético, ya que aquel método que tenga dicho coeficiente más elevado será el que presente una menor distorsión en las relaciones originales existentes entre los elementos en estudio.

Por otra parte, cuando se utilizan técnicas clusters jerárquicas uno de los objetivos es obtener un subconjunto de particiones y, por tanto, determinar el número de clusters adecuado para el problema en estudio. Esta cuestión ha motivado la aparición de múltiples reglas, de las que se pueden destacar las siguientes:

- Cortar el dendograma de una forma subjetiva una vez que se ha visualizado. Es obvio que este procedimiento no es adecuado ya que está sesgado por la opinión que el investigador tiene sobre sus datos.
- Otro método, también heurístico, se basa en representar en una gráfica el número de clusters que se observan en los distintos niveles del dendograma, frente a los niveles de fusión a los que los clusters se unen en cada nivel. La presencia de una pendiente poco pronunciada sugiere que la siguiente unión de clusters no aporta apenas información adicional sobre la aportada en el nivel anterior. Se basa, por tanto, en la existencia de saltos o discontinuidades en los niveles de fusión. Tal y como hemos indicado, se suele utilizar el Coeficiente de Correlación Cofenético ya que se ha utilizado un método jerárquico.

Métodos no jerárquicos

En el caso de los no jerárquicos algunos autores proponen utilizar el *Análisis Multivariante de la Varianza* o bien desarrollar *Análisis de la Varianza*

sobre cada variable en cada cluster. Otra técnica empleada normalmente es la de tomar varias submuestras de la muestra original y repetir el análisis sobre cada una.

En cuanto a la interpretación de resultados hay que señalar que ninguna de las técnicas proporciona una solución óptima para todos los problemas, ya que es posible llegar a distintos resultados según la técnica elegida. Por esta razón, el criterio del investigador, el conocimiento del problema y la experiencia deben sugerir el método adecuado y la solución correcta. Cuando previamente se ha aplicado una técnica factorial para reducir la dimensionalidad, la fase de interpretación se simplifica con el posicionamiento de los grupos obtenidos en los factores, interpretando la posición de cada grupo según el significado de dichos factores.

b. El *Statis* y el *Análisis Cluster*

Lebart *et al.* (1995: 187) consideran necesarias una serie de fases a seguir cuando el objetivo es describir un conjunto de datos de gran tamaño mediante la aplicación de Técnicas Factoriales y de Clasificación, que son las siguientes:

1ª etapa: Aplicación del *Método Factorial*

La aplicación de un *Método Factorial*, como etapa previa a la clasificación, es aconsejable en cuanto permite describir de una manera más precisa la información de partida, además de filtrar los datos y considerar solamente aquellos con significado relevante, que son los factores, menos numerosos que las variables de partida e incorrelados entre sí.

La elección del *Método Factorial* depende de las características y naturaleza de los datos originales. Si las variables de las que se dispone para clasificar los individuos son cuantitativas, el *Análisis de Componentes Principales* es

el método más frecuentemente empleado, cualquiera que sea el campo de aplicación, pudiéndose citar como ejemplo el trabajo de Pinillos Costa y Martín Peña (1998) para la determinación de grupos estratégicos en el ámbito de la Economía de la Empresa.

Ahora bien, cuando las variables cuantitativas de las que se dispone están agrupadas de cierta manera, porque corresponden a momentos de tiempo o situaciones diferentes, que dan lugar a una tabla de tres entradas, debe aplicarse un *Método Factorial* adecuado para este tipo de tablas como puede ser el STATIS. Otra posible opción sería la aplicación de un *Análisis de Componentes Principales* a cada grupo de variables por separado, con el inconveniente adicional de comparar las clasificaciones obtenidas a partir de cada grupo. También se podría aplicar un ACP a la tabla global, con el inconveniente de que las variables podrían estar muy correlacionadas y la clasificación obtenida podría estar distorsionada por esta circunstancia.

El empleo del STATIS como *Método Factorial* de manera previa a una clasificación se realiza en trabajos como el de Pérez Hugalde (1988), completándose la tipología obtenida con las ventajas descriptivas de la Técnica Multidimensional. De esta manera, la aplicación del STATIS, que proporciona mejores soluciones que otros métodos para tablas de tres entradas por ser una versión más amplia de algunos de ellos, permite obtener una clasificación equilibrando la influencia de todos los grupos en el análisis. Además permite clasificar el mismo conjunto de individuos pero a partir de grupos formados por distintas variables, siempre y cuando esto tuviese sentido en el contexto correspondiente.

2ª etapa: Clasificación a partir de factores

El procedimiento es el mismo que cuando el punto de partida son las variables. La única diferencia es que las distancias entre individuos se calculan en el espacio definido por los primeros h factores, obtenidos con la Técnica

Factorial, y empleando generalmente la distancia euclídea usual. La clasificación que toma como base esos h primeros factores presenta la ventaja de eliminar las fluctuaciones aleatorias que están recogidas en los datos y que se supone son variaciones no sistemáticas de los mismos.

Cuando el *Método Factorial* empleado es el *Statis*, los factores que se consideran son los obtenidos de la etapa de la intraestructura, es decir, los que representan la estructura común de las nubes de individuos. Las clases obtenidas se basan en dicha estructura, y por tanto, son más homogéneas que si la clasificación se realizase directamente a partir de las variables originales.

3ª etapa: Descripción y posicionamiento de los grupos en los planos factoriales

La descripción de los grupos obtenidos se puede hacer calculando, para todas las variables, las diferencias entre los valores medios de las mismas en cada grupo y los valores medios globales. Si bien, teniendo en cuenta el significado de los factores obtenidos como resultado de un ACP, y por tanto también con el *Statis*, la interpretación de la posición de un individuo respecto a un factor depende de que presente valores superiores o inferiores a la media, respectivamente, para las variables que se correlacionan positiva o negativamente con dicho factor. Así, la descripción de los grupos se puede hacer mediante el posicionamiento de las clases obtenidas en los planos factoriales: el alejamiento de un grupo en la dirección de un factor permite su caracterización en función de las variables que definen el factor, y son las que destacan por algún motivo en el grupo.

El posicionamiento de un grupo o clase sobre un factor se lleva a cabo otorgando a cada individuo un número correspondiente a la clase a la que pertenece, lo que también permite poner de manifiesto el comportamiento de ciertos individuos, que forman parte de un grupo, pero que presentan

un comportamiento un tanto diferente, revelando cierta discontinuidad en la clase obtenida.

A modo de resumen proponemos el siguiente cuadro que servirá de guía en la aplicación práctica del capítulo posterior y que permitirá poner de manifiesto las posibilidades descriptivas del STATIS.

Figura II.24
El *Statis* y el Análisis Cluster

<i>Statis y Análisis Cluster</i>	<p>1. Método factorial: el <i>Statis</i></p> <ul style="list-style-type: none"> • <i>Interestructura</i>: Imagen euclídea de los $W_1 D$ • <i>Compromiso</i>: Operador-compromiso $W D$ • <i>Intraestructura</i> <ul style="list-style-type: none"> – Círculo de correlaciones: pT puntos-variables – Individuos-compromiso: n puntos-individuos
	<p>2. Análisis Cluster</p> <ul style="list-style-type: none"> • <i>Las variables</i>: coordenadas de los n individuos-compromiso en los h primeros ejes (los h primeros factores obtenidos). • <i>Medida de asociación</i>: distancia euclídea usual entre las coordenadas • <i>Técnica de asociación</i>: Agrupación de los n individuos-compromiso, según Método Jerárquico en este caso el resultado se recoge en el dendograma • <i>Validación e interpretación de resultados</i>
	<p>3. Descripción y posicionamiento de los grupos en los planos factoriales</p>

3.2 Predicción y STATIS

Para abordar la capacidad predictiva del STATIS es necesario considerar ciertos aspectos relacionados con la teoría de procesos estocásticos y su conexión con el *Análisis de Componentes Principales*.

Comenzamos definiendo una serie de conceptos básicos para comprender la naturaleza de los procesos estocásticos, sus propiedades, y la consideración de una serie temporal como un proceso estocástico de ciertas características. A continuación se realiza una breve aproximación al *Análisis Factorial de un proceso*, que permite descomponer el mismo en una serie de variables aleatorias incorreladas que pueden considerarse como los factores obtenidos de un ACP. Por ese motivo, los factores obtenidos en el *Statis*, como un ACP que es, tienen un carácter estocástico y pueden emplearse en la predicción tratándolos mediante un modelo ARIMA. Las fases de este enfoque para el tratamiento de series temporales se desarrollarán de manera adecuada con el fin de ser empleadas en la modelización de dichos factores. Para finalizar se propone un modelo de predicción de las variables de la tabla original, basado en la modelización ARIMA de esos factores y en las fórmulas de reconstitución del método *Statis*.

3.2.1 Conceptos básicos

Partimos de una tabla de datos formada por p variables o características cuantitativas (las columnas de la tabla), observadas en T periodos de tiempo (las filas) sobre un conjunto de individuos. Los individuos, que no aparecen especificados, serán los mismos en todos los periodos del tiempo y cada valor de la variable constituye el total agregado para esos individuos en el momento del tiempo correspondiente. Esta tabla se representa de la forma siguiente:

Figura II.25
Tabla de datos temporales

	1	2	...	<i>j</i>	...	<i>p</i>
1	x_{tj}					
2						
⋮						
<i>t</i>						
⋮						
<i>T</i>						

y el elemento característico de la misma es x_{tj} ($t = 1, 2 \dots T ; j = 1, 2 \dots p$).

Cada variable (en columna) vendrá dada por el vector:

$$X_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{tj} \\ \vdots \\ x_{Tj} \end{bmatrix}$$

y a diferencia de lo expuesto hasta ahora, se va a considerar que los valores de X_j definen una serie temporal y tiene un carácter estocástico, por lo que definiremos previamente estos dos conceptos.

Un *proceso estocástico* se define como un conjunto de variables aleatorias dependientes del tiempo, es decir: $\{X_j(t) \dots t \in T\}$, siendo T el espacio paramétrico que designa el conjunto de momentos del tiempo.

Los valores de la *serie temporal* $x_{1j}, x_{2j} \dots x_{Tj}$ se consideran como realizaciones muestrales de las variables $X_{1j}, X_{2j} \dots X_{Tj}$ con unas probabilidades de ocurrencia $f(x_{1j}, x_{2j} \dots x_{Tj})$, deducidas de la función de distribución conjunta, pudiéndose afirmar que $x_{1j}, x_{2j} \dots x_{Tj}$ es una muestra de tamaño uno de cada una de las variables aleatorias tomadas en sucesivos periodos de tiempo.

Si se pudiera especificar numéricamente la función de distribución correspondiente a la serie se podría determinar la probabilidad de un determinado resultado futuro. Sin embargo, esto es casi imposible y

generalmente se construye un modelo simplificado que explique la aleatoriedad de la serie de forma que resulte útil para pronosticar.

Si se supone que el proceso estocástico verifica el supuesto de normalidad, o lo que es lo mismo, su distribución conjunta es normal multivariante quedará determinado por sus momentos:

$$\text{Media: } E(X_{tj}) = \mu_{tj}$$

$$\text{Varianza: } V(X_{tj}) = E[X_{tj} - \mu_{tj}]^2 = \sigma_{tj}^2$$

$$\text{Autocovarianza: } Cov(X_{tj}, X_{t'j}) = E[(X_{tj} - \mu_{tj})(X_{t'j} - \mu_{t'j})]$$

No obstante, a partir de una serie con pocas observaciones es imposible obtener estas características ya que tendríamos que calcular T medias, T varianzas y $T(T-1)/2$ autocovarianzas. Esto justifica la necesidad de introducir el supuesto de invariabilidad en el tiempo, es decir, la estacionariedad. Así, un proceso estocástico se dice que es *estrictamente estacionario* cuando su distribución de probabilidad conjunta no varía al realizar un cambio de origen, es decir:

$$f(X_{1j}, X_{2j}, \dots, X_{Tj}) = f(X_{(1+m)j}, X_{(2+m)j}, \dots, X_{(T+m)j})$$

para cualquier valor entero arbitrario m . Un proceso es *débilmente estacionario* cuando su media y su varianza son constantes para todos los valores de t y además la covarianza entre valores del proceso sólo depende de la distancia temporal entre los mismos y no del momento del tiempo a que están referidos, es decir:

$$\mu_j = \mu_{tj} = E(X_{tj})$$

$$\sigma_j^2 = \sigma_{tj}^2 = V(X_{tj})$$

$$Cov(X_{tj}, X_{(t+m)j}) = \sigma_{mj}$$

La propiedad de estacionariedad es una característica importante de los procesos estocásticos para los que vamos a construir modelos ya que, si el proceso no es estacionario, resulta difícil representar la serie temporal para intervalos de tiempo pasados y futuros mediante un modelo algebraico

sencillo. No obstante, como la propiedad de la covarianza no siempre se verifica a partir de una sola serie, es necesario introducir el concepto de ergodicidad que desarrollaremos posteriormente.

Los estimadores de los parámetros del proceso se definen a partir de la serie temporal $x_{1j}, x_{2j} \dots x_{Tj}$. El estimador que se puede considerar adecuado de la media de un proceso estacionario es la media muestral:

$$\bar{X}_j = \frac{1}{T} \sum_{t=1}^T x_{tj}$$

que es insesgado y consistente. La estimación de la autocovarianza poblacional de orden m , es decir, la $Cov(X_{tj}, X_{(t+m)j})$, se efectúa utilizando el estimador C_m :

$$C_m = T^{-1} \sum_{t=m+1}^T (x_{tj} - \bar{X}_j)(x_{(t-m)j} - \bar{X}_j)$$

siendo C_m la autocovarianza muestral de orden m . También es posible usar otro estimador:

$$C_m^1 = (T - m)^{-1} \sum_{t=m+1}^T (x_{tj} - \bar{X}_j)(x_{(t-m)j} - \bar{X}_j)$$

Al analizar las propiedades de estos estimadores se observa que los dos son sesgados. Si bien, se puede comprobar bajo la hipótesis de normalidad (según Barlett, 1946), que la varianza de C_m es inferior a la de C_m^1 , por esta razón suele considerarse C_m como mejor estimador. Se recomienda que cuando el sesgo de este estimador pueda ser importante solo se calcule alrededor de $T/4$ autocovarianzas muestrales.

Hay que señalar que el estimador de la varianza del proceso será la varianza muestral de orden cero (C_0), que se define como:

$$S_j^2 = C_0 = \frac{1}{T} \sum_{t=1}^T (x_{tj} - \bar{X}_j)^2$$

Para establecer la conexión entre los procesos estocásticos y la notación empleada en los métodos de *Análisis de Datos Multidimensionales* considerados se van a expresar las X_j como desviaciones respecto a su media y en unidades estándar. Así, una vez definidas la media y la varianza muestral de X_j , pasamos a la variable Z_j de tal manera que:

$$Z_j = \frac{X_j - \bar{X}_j}{S_j}$$

3.2.2 *Análisis Factorial de un proceso*

La teoría abstracta del *Análisis Factorial de un proceso* fue desarrollada en primer lugar por Escoufier (1970: 1-47) y por Dauxois y Pousse en 1976 y ampliamente desarrollada en Saporta (1981). Basándose en ella, tal y como manifiestan Deville (1974: 10) y Saporta (1985: 111-2), cualquier proceso estocástico Z_j que varía en un intervalo finito de tiempo T puede ser representado bajo la forma de una suma de funciones dependientes del tiempo ponderadas por una serie de funciones aleatorias no correladas. Como se recoge en dichos artículos, Z_j se expresa por:

$$Z_j = \sum_{k=1}^{\infty} \zeta_k f_k(t)$$

donde las ζ_k son combinaciones lineales de la variable aleatoria Z_j :

$$\zeta_k = \int_0^T Z_j f_k(t) dt$$

y $f_k(t)$ son funciones dependientes del tiempo.

El operador WD propuesto por Escoufier (1970: 20) permite transformar una variable aleatoria en otra. Así, la descomposición de dicho operador WD en vectores propios permite obtener la única reconstitución de la variable original en la que ζ_k y $f_k(t)$ son variables aleatorias ortogonales (la denominada descomposición de Karhunen-Loeve). En ella las ζ_k serán las

componentes principales, es decir, los vectores propios de WD , las $f_k(t)$ serán los ejes factoriales y la varianza de las ζ_k será el vector propio correspondiente, siendo su esperanza nula.

Por tanto, la mejor aproximación de la variable Z_j , en el sentido de mínimos cuadrados, por una suma de h términos vendrá dada¹⁷ por:

$$Z_j = \sum_{k=1}^h \zeta_k f_k(t)$$

Las relaciones entre la descomposición del proceso de Karhunen-Loeve y el *Análisis de Componentes Principales* presenta una serie de ventajas que Deville (1974: 10) considera y que son las siguientes:

- Permite el empleo del vocabulario geométrico habitual del *Análisis Multivariante* para generalizar los métodos en el espacio de Hilbert.
- Simplifica las notaciones y hace más intuitivos los desarrollos que en otro caso serían bastante abstractos.
- Facilita el uso de desarrollos matemáticos relativos a los espacios de Hilbert y los procesos estocásticos.

Cuando no tenemos una sola variable aleatoria sino p variables, un problema que se plantea es la reducción de la dimensionalidad tal y como lo manifiestan Peña y Box (1984: 494; 1987: 836). Estos autores consideran que si es posible encontrar un nuevo conjunto de variables construidas a partir de las originales que explican las características principales de las series originales, será posible obtener una representación más sencilla del conjunto original. Además, esta solución puede ser una alternativa mucho más interesante a la modelización univariante de cada serie, en cuanto permite establecer relaciones entre variables. Según consideran Peña y Poncela (1996: 1), la extracción de factores comunes constituye también una

¹⁷ Para desarrollos más amplios, y además de los artículos ya citados, conviene revisar el artículo de Pardoux (1989: 41-60).

alternativa a la modelización multivariante (modelos *VARMA*) en la que la estimación de un gran número de parámetros altamente correlacionados entre sí complica en exceso el modelo.

El modelo del *Análisis Factorial* sirve de marco de referencia para la extracción de factores en el artículo de Anderson (1963). Los artículos de Box y Tiao (1977), de Flury (1984) y los ya citados de Peña y Box y Peña y Poncela siguen el mismo enfoque y suponen la existencia de una serie de factores comunes que son los responsables de las distintas series.

Otros autores como Otter y Schuur (1982) emplean el modelo de *Análisis de Componentes Principales* como método para la obtención de factores comunes a las series y permiten reconstituir las variables originales con el fin de la previsión. Además, y dado que es posible establecer la relación entre las variables originales y las nuevas variables ortogonales, se puede analizar la evolución de cierto fenómeno de una manera más sencilla.

Al considerar la fórmula de reconstitución de los datos para el *Análisis de Componentes Principales*, se tiene que:

$$Z_j = \sum_{k=1}^h F_k m_k^{-1} u_{kj} = \sum_{k=1}^h F_k a_{kj}$$

expresión que constituye la mejor aproximación, de orden h en el sentido de mínimos cuadrados, de la variable Z_j . En esta expresión las F_k son las componentes principales y los a_{kj} son las coordenadas del vector del espacio de los individuos. Poniendo en relación dicha expresión con la descomposición de Karhunen-Loeve se deduce que F_k es la variable aleatoria que permite reconstituir las Z_j originales. Ese carácter estocástico se tendrá en cuenta para su modelización con objeto de predecir.

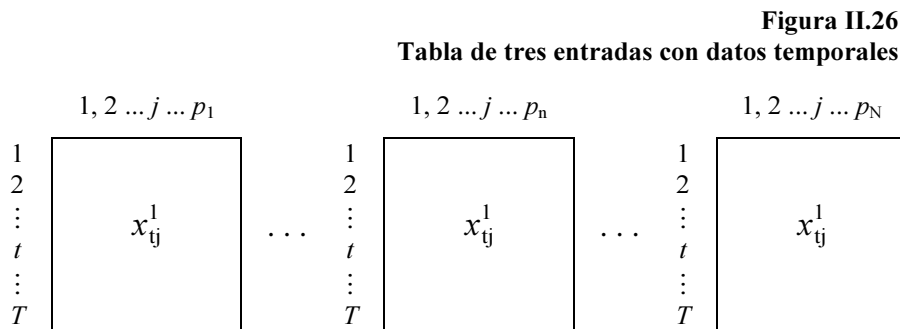
Como ya hemos indicado, tanto el método *Statis* como el *Statis dual* son casos especiales de ACP y, por tanto, para ambos son aplicables las consideraciones que acabamos de hacer como vamos a demostrar a continuación.

3.2.3 *Statis* y procesos estocásticos

Para relacionar ambos procedimientos, se considera una tabla de T filas (cada una de ellas representa un momento del tiempo diferente) y:

$$\sum_{n=1}^N p_n$$

columnas, que recogen las variables de N tablas diferentes de datos. Cada una de esas tablas estará formada por T filas y p_n columnas, tal y como se muestra a continuación:



El elemento representativo de la tabla será x_{tj}^n que representa el valor de la variable X_j de la tabla n en el instante t . Los momentos del tiempo son los mismos para todas las tablas pero las variables de cada tabla pueden ser las mismas o diferentes.

Dicha "supertabla" (o conjunto de tablas $X^1, X^2 \dots X^N$) constituye el objeto del método *Statis*, tal y como hemos expuesto; la peculiaridad de esta tabla es que ahora los distintos momentos del tiempo son las filas (los individuos antes), y cada una de esas tablas representa una situación espacial (y no temporal) diferente.

Si las variables se centran y se reducen pasaremos a tener las tablas $Z^1, Z^2 \dots Z^N$ y cada tabla estará representada por el operador $W_n D = Z^n M Z^n' D$. Considerando dichos operadores normados, que es el caso más habitual, tendremos los siguientes:

$$\frac{W_n D}{\|W_n D\|_{HS}}$$

La aplicación del *Statis* en estas condiciones permite extraer las siguientes conclusiones para las diferentes etapas:

1. Interestructura

Esta etapa permite obtener el posicionamiento de las N tablas de datos, unas respecto a las otras, pudiendo establecer comparaciones entre las mismas. La representación gráfica del primer plano factorial es la que consideraremos en este caso y se obtiene a partir de la descomposición en valores y vectores propios de la matriz de coeficientes RV .

2. Compromiso

Se trata de construir un operador WD de la misma naturaleza que los que representan las tablas de datos y cuyo análisis se realiza en la siguiente etapa. Las ponderaciones para la obtención de dicho compromiso son las obtenidas, de la manera indicada, al desarrollar la fase del compromiso del *Statis* con los operadores normados (pp. 194-5).

3. Intraestructura

Remitiéndonos a la etapa del mismo nombre al desarrollar el *Statis*, esta fase equivale a realizar un ACP *centrado* de las siguientes tablas de datos:

$$\sqrt{\alpha_1 \beta_1} Z^1, \sqrt{\alpha_2 \beta_2} Z^2 \dots \sqrt{\alpha_n \beta_n} Z^n \dots \sqrt{\alpha_N \beta_N} Z^N = \Pi$$

El operador representativo de esa tabla Π es el operador WD de la etapa del compromiso.

- La obtención de los valores y vectores propios permitirá obtener las posiciones compromiso de las filas (en este caso momentos del tiempo) de la tabla que darán lugar a las componentes principales:

$$F_k^c = \begin{bmatrix} f_{k1}^c \\ f_{k2}^c \\ \vdots \\ f_{kt}^c \\ \vdots \\ f_{kT}^c \end{bmatrix}$$

tal que: $WD F_k^c = \lambda_k^c F_k^c$

siendo t_k^c el eje asociado al espacio de las variables y verificándose:

$$F_k^c = \sqrt{\lambda_k^c} t_k^c$$

- Considerando que p es el rango de WD , la cantidad de información explicada por el eje k es su valor propio y el porcentaje de inercia es:

$$\frac{\lambda_k^c}{\sum_{k=1}^p \lambda_k^c}$$

- La interpretación de las componentes principales del compromiso es posible a través de las correlaciones de las variables originales con dichos ejes. Las coordenadas de todas las variables Z_j^n con el eje k toman la siguiente expresión:

$$G_k = \Pi' D t_k^c = \begin{bmatrix} g_{k1}^1 \\ \vdots \\ g_{kp_1}^1 \\ g_{k1}^2 \\ \vdots \\ g_{kp_2}^2 \\ \vdots \\ g_{k1}^N \\ \vdots \\ g_{kp_N}^N \end{bmatrix}$$

teniendo en cuenta que: $g_{kj}^n = (\sqrt{\alpha_t \beta_t} Z_j^n)' D t_k^c$

Por tanto, el coeficiente de correlación de una variable con un eje vendrá dado por:

$$\rho(F_k^c, Z_j^n) = \frac{g_{kj}^n}{\sqrt{\alpha_n \beta_n}} = (Z_j^n)' D t_k^c$$

- Como para cualquier ACP podemos reconstituir las variables que han intervenido en la tabla a través de las componentes obtenidas. Así, considerando la ponderación y la fórmula de reconstitución de los datos, se tiene que:

$$\sqrt{\alpha_n \beta_n} Z_j^n = \sum_{k=1}^h F_k^c m_k^{-1} u_{kj}^n = \sum_{k=1}^h F_k^c a_{kj}^n$$

que puede considerarse la mejor aproximación de orden h de la variable Z_j de la tabla n . Teniendo en cuenta la relación que existe entre los a_{kj} y los g_{kj} de cualquier ACP y según los desarrollos del *Statis*, tenemos que:

$$g_{kj}^t = \sqrt{\lambda_k^c} a_{kj}^t$$

por lo que en este caso:

$$g_{kj}^n = \sqrt{\lambda_k^c} a_{kj}^n \Rightarrow a_{kj}^n = \frac{g_{kj}^n}{\sqrt{\lambda_k^c}}$$

y la fórmula toma la expresión:

$$\sqrt{\alpha_n \beta_n} Z_j^n = \sum_{k=1}^h F_k^c \frac{g_{kj}^n}{\sqrt{\lambda_k^c}} \Rightarrow Z_j^n = \frac{\sum_{k=1}^h F_k^c \frac{g_{kj}^n}{\sqrt{\lambda_k^c}}}{\sqrt{\alpha_n \beta_n}}$$

Como:

$$Z_j^n = \frac{X_j^n - \bar{X}_j^n}{S_{x_j^n}}$$

tenemos que la reconstitución, de orden h , de la variable original de tabla X_j^n toma la siguiente expresión:

$$X_j^n = \frac{\sum_{k=1}^h F_k^c \frac{g_{kj}^n}{\sqrt{\lambda_k^c}}}{\sqrt{\alpha_n \beta_n}} S_{x_j^n} + \bar{X}_j^n \Rightarrow x_{tj}^n = \frac{\sum_{k=1}^h f_{kt}^c \frac{g_{kj}^n}{\sqrt{\lambda_k^c}}}{\sqrt{\alpha_n \beta_n}} S_{x_j^n} + \bar{X}_j^n$$

siendo \bar{X}_j^n su media y $S_{x_j^n}$ su desviación típica.

Si las X_j^n son variables aleatorias, ya hemos demostrado que las F_k^c tienen un comportamiento estocástico y, por tanto:

$$F_k^c = \begin{bmatrix} f_{k1}^c \\ f_{k2}^c \\ \vdots \\ f_{kt}^c \\ \vdots \\ f_{kT}^c \end{bmatrix}$$

puede considerarse como una serie temporal. La modelización de F_k^c permitirá obtener un modelo que explique dicho factor y además hacer predicciones para el periodo $T+1$ y para el periodo $T+2$ y, por tanto, obtener los valores de las variables de las tablas originales para $T+1$ y $T+2$ sobre la base del modelo obtenido y de la fórmula de reconstitución indicada.

Como resumen, esta etapa de la infraestructura permitirá resumir la evolución de un fenómeno a través de la posición de los distintos momentos del tiempo (individuos-compromiso) respecto a unos ejes, caracterizados a través de las correlaciones con las variables empleadas en el estudio. Además, la modelización de las componentes F_k^c permitirá predecir a corto plazo.

El carácter estocástico de los F_k^c nos lleva a considerar como modelo adecuado para el tratamiento de dicha serie temporal la modelización ARIMA, aunque el número de datos no sea muy elevado. Por otra parte, este modelo ofrece algunas ventajas con respecto a otros métodos de análisis de series temporales, como la suavización exponencial, la descomposición de series y el análisis de la tendencia simple. Extrae mucha información de

la serie y la consigue utilizando para ello un número reducido de parámetros. Asimismo, ofrece una mayor flexibilidad en la elección del modelo correcto, incluyendo un proceso que permite examinar una gran variedad de modelos en la búsqueda del más adecuado.

4. Trayectorias de los individuos

Dado el carácter especial de la tabla de datos, la trayectoria de cada individuo–momento vendría dada por N puntos (cada uno representando un conjunto de variables) con una interpretación difícil de establecer. Por ese motivo esta etapa no va a ser objeto de atención en este caso, si bien puede ser objeto de investigación en estudios posteriores.

3.2.4 Modelos ARIMA

Con objeto de modelizar los factores obtenidos como resultado del *Statis*, se presenta a continuación de manera breve el enfoque moderno de tratamiento de series temporales univariantes recogido en la obra de los estadísticos Box y Jenkins (1976). Dicha obra se centra en un tipo de modelos estocásticos, especialmente útiles para la predicción: los denominados modelos *ARIMA*. En primer lugar, se van a considerar unos conceptos introductorios sobre dicho enfoque y, a continuación, las características de los modelos básicos. Finalmente, se desarrollarán las fases de aplicación de la metodología Box-Jenkins para modelos lineales univariantes que es el caso que nos ocupa. Es importante señalar que las series temporales, que en nuestro caso vamos a modelizar, se representarán por $f_{k1}^c, f_{k2}^c \dots f_{kT}^c$ que es el equivalente a los $x_1, x_2 \dots x_T$ de la siguiente exposición teórica.

a. Conceptos básicos

— Los conceptos de *proceso estocástico* y de *serie temporal*, así como los elementos para su descripción, ya han sido definidos por lo que sólo vamos a recordarlos de manera breve con la modificación de notación correspondiente. Un *proceso estocástico* se define como un conjunto de variables aleatorias dependientes del tiempo, es decir: $\{X(t) \dots t \in T\}$, siendo T el espacio paramétrico que designa el conjunto de momentos del tiempo.

— Los valores de la *serie temporal* $x_1, x_2 \dots x_T$ se consideran como realizaciones muestrales de las variables $X_1, X_2 \dots X_T$ con unas probabilidades de ocurrencia deducidas de la función de distribución conjunta.

También se ha indicado que la estacionariedad es una característica importante de los procesos estocásticos ya que si el proceso no es estacionario resulta difícil representar la serie temporal para intervalos de tiempo pasados y futuros mediante un modelo algebraico sencillo. Como la propiedad de la covarianza no siempre se verifica a partir de una sola serie, es necesario introducir el concepto de ergodicidad para lo que previamente definiremos la *Función de Autocorrelación (FAC)*.

— Se define la *Función de Autocorrelación* de la forma siguiente:

$$\rho_m = \frac{Cov(X_t, X_{(t+m)})}{\sigma_t \sigma_{(t+m)}}$$

A la sucesión de valores que toma ρ_m para los diferentes valores de m (entre 0 y $T+m$) se le denomina *Correlograma*. Esta función resulta muy útil para obtener una descripción del proceso pues proporciona una medida cuantitativa de la correlación que existe entre observaciones contiguas de la serie. Así, si al representar el correlograma, la dependencia entre observaciones tiende a cero a medida que aumenta el retardo m , se dice que el proceso es *ergódico*. La ergodicidad es necesaria para poder estimar los momentos del proceso a partir de una única realización. Si el proceso no

fuese ergódico al aumentar el tamaño muestral no se dispondría de información adicional, por ser las observaciones dependientes entre sí, y no sería posible obtener estimadores consistentes de los parámetros del modelo. Por tanto, se consideran solo procesos estacionarios ergódicos.

— La *Función de Autocorrelación Parcial (FAP)* responde a la misma idea que la *FAC* y trata de medir la correlación que existe entre valores de la serie temporal en distintos instantes de tiempo. A diferencia de la *FAC*, la *FAP* mide la correlación ajustada por el efecto de los retardos inmediatos.

Cuando un proceso estocástico es estacionario y ergódico, los parámetros que caracterizan el mismo (la media, la varianza, autocovarianzas, autocorrelaciones y autocorrelaciones parciales), pueden estimarse a partir de una sola realización del proceso. Considerando la serie temporal de T observaciones vamos a definir los distintos estimadores de las funciones *FAC* y *FAP* que son respectivamente la *Función de Autocorrelación Muestral* y la *Función de Autocorrelación Parcial Muestral*.

— La *Función de Autocorrelación Muestral (FACM)* se estimará utilizando la expresión de la autocorrelación muestral de orden m , que denotaremos mediante r_m y que definiremos como el cociente entre la autocovarianza muestral de orden m (C_m) y la varianza muestral (C_0), esto es:

$$r_m = \frac{C_m}{C_0}$$

El conjunto de las autocorrelaciones muestrales obtenidas se denomina *correlograma muestral*.

— La *Función de Autocorrelación Parcial Muestral (FAPM)* se obtendrá reemplazando las autocorrelaciones ρ_m por sus estimaciones r_m , en la expresión que toma la *FAP*.

A pesar de las condiciones de estacionariedad y ergodicidad de los procesos considerados, el número de parámetros a estimar para caracterizar los

procesos sigue siendo elevado, por ese motivo solo se consideran una serie de modelos básicos en los que debido a sus características especiales la estimación de dichos parámetros resulta más sencilla. Vamos a ir analizando brevemente las peculiaridades de cada uno de ellos.

b. Características de los Modelos Básicos Univariantes

Comenzaremos por el modelo más sencillo, que es el de ruido blanco y a partir de éste se deducen los demás.

MODELO DE RUIDO BLANCO

En el análisis de series temporales el proceso más elemental es el de *ruido blanco*, también denominado *proceso puramente aleatorio* o *proceso sin correlación* ya que viene caracterizado por una serie de variables incorreladas. Se define dicho proceso como una sucesión de variables aleatorias con distribución normal, con esperanza cero, igual varianza e incorreladas entre sí, es decir:

$$E(X_t) = 0 \quad E(X_t^2) = \sigma^2 \quad Cov(X_t, X_{t+m}) = 0 \quad ,, \quad m = \pm 1, \pm 2 \dots$$

Existen autores que definen este proceso sin exigir que las variables tengan distribución normal, suponiendo que son independientes en vez de incorreladas. Por tanto, un *ruido blanco* es un proceso estocástico en el que las distribuciones de probabilidad correspondientes a dos instantes cualesquiera de tiempo son independientes y todas ellas con esperanza nula o igual varianza. Es inmediato que todo ruido blanco es un proceso estacionario con una función de autocorrelación que toma valores nulos excepto para $\rho_0 = 1$.

La característica fundamental de este proceso es que, a partir de él, va a ser posible la obtención de todo tipo de modelos probabilísticos. La predicción

para un proceso ruido blanco es la media del mismo (Aznar y Trivez, 1993b: 268, 272).

MODELO DE CAMINO ALEATORIO

Este modelo recoge un comportamiento especial en cuanto que la diferencia entre dos valores consecutivos de la variable es un ruido blanco:

$$X_t - X_{t-1} = \varepsilon_t \text{ (ruido blanco)}$$

Por lo tanto, al introducir una relación de dependencia entre X_t y X_{t-1} nos encontramos con el primer ejemplo de generación de un modelo de series de tiempo a partir del ruido blanco. A partir del concepto de *camino aleatorio* se puede deducir que un procedimiento para alcanzar la estacionariedad de una serie observada que no lo es, consiste en aplicar diferencias sucesivas.

Se puede demostrar que este proceso es estacionario en media pero no en varianza. La autocovarianza es dependiente del tiempo y la función de autocorrelación está formada por una serie de coeficientes próximos a uno lo que da lugar a una serie bastante regular a pesar de sus características no estacionarias. La predicción para este proceso viene dada por:

$$X_{T+1} = E[X_{T+1} / X_T \dots X_1]$$

donde: $X_{T+1} = X_T + \varepsilon_{T+1}$ y, ε_{T+1} es independiente de $X_T, X_{T-1} \dots X_1$, de lo que se deduce que:

$$X_{T+1} = X_T + E(\varepsilon_{T+1}) = X_T$$

MODELOS DE MEDIAS MÓVILES (MA)

Examinaremos a continuación algunos procesos sencillos de media móvil, indicando para cada uno de ellos la media, la varianza, las covarianzas y la función de autocorrelación. Estos estadísticos son importantes porque proporcionan información que ayuda a caracterizar e identificar el proceso.

Proceso de medias móviles de orden 1, $MA(1)$

— Se formula a partir de un proceso de ruido blanco, siendo su expresión:

$$X_t = \mu + \varepsilon_t - b_1 \varepsilon_{t-1}$$

siendo b_1 una constante, ε_t un proceso de ruido blanco y μ la media. A veces el modelo se formula excluyendo μ , ya que ésta no afecta a la función de autocorrelación.

— Este proceso tiene las siguientes características:

- $E(X_t) = \mu$

Cuando el proceso se expresa por: $X_t = \varepsilon_t - b_1 \varepsilon_{t-1}$,, $E(X_t) = 0$

- $Var(X_t) = E(X_t - \mu)^2 = \sigma^2(1 + b_1^2)$

- Función de autocovarianza:

para $m = 1$ será: $Cov(X_t, X_{t+1}) = -b_1 \sigma^2$

y para $m = 2$: $Cov(X_t, X_{t+2}) = 0$

De donde se puede deducir que para valores de $m > 1$, la covarianza se anula, puesto que en todos los casos hay expresiones de la forma:

$$E(\varepsilon_t \varepsilon_{t'}) = 0 \quad ,, \quad t \neq t'$$

- Función de autocorrelación:

$$\rho_m = \begin{cases} \frac{-b_1}{1+b_1^2} & m = 1 \\ 0 & m > 1 \end{cases}$$

De estas condiciones se deduce que el modelo $MA(1)$ tiene memoria para un sólo período, lo que significa que cualquier valor de X_t está relacionado con X_{t-1} y X_{t+1} , pero con ningún otro valor de la serie. En definitiva, una consecuencia inmediata de este resultado es que, a partir de un modelo $MA(1)$, la predicción sólo se puede obtener para el momento inmediatamente posterior al último valor de la serie de tiempo.

- Es un proceso estacionario, lo que se deduce de las características anteriores.

Hay que tener en cuenta una situación que puede ocurrir en la práctica: el hecho de que a una misma función de autocorrelación le correspondan dos procesos de medias móviles de orden uno distintos; la solución a este problema viene dada por las condiciones de invertibilidad. Sin entrar en más detalles decir que el proceso $MA(1)$ es invertible si y solo $|b_1| < 1$ y por tanto el modelo debe ser invertible.

Proceso de medias móviles de orden q , $MA(q)$

— Se define de la forma siguiente:

$$X_t = \varepsilon_t - b_1 \varepsilon_{t-1} - b_2 \varepsilon_{t-2} - \dots - b_q \varepsilon_{t-q}$$

— Como características del modelo se tienen las siguientes:

- $E(X_t) = 0$
- $Var(X_t) = 1 + b_1^2 + b_2^2 + \dots + b_q^2$
- Función de autocorrelación:

$$\rho_m = \begin{cases} \frac{-b_m + b_1 b_{m+1} + b_{q-m} b_q}{1 + b_1^2 + b_2^2 + \dots + b_q^2} & m = 1, \dots, q \\ 0 & m > q \end{cases}$$

- La condición de invertibilidad exige que las raíces de:

$$(1 - b_1 B - \dots - b_q B^q)$$

estén fuera del círculo unidad.

MODELOS AUTORREGRESIVOS (AR)

La expresión general de un modelo autorregresivo viene dada de la forma siguiente:

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + \delta + \varepsilon_t$$

expresión que permite interpretar este modelo como un Modelo de Regresión Lineal Múltiple donde las variables explicativas son los valores de la propia variable desplazados en el tiempo. Además, mientras que el modelo de medias móviles resulta de una media ponderada sobre una serie de ruido blanco, el autorregresivo se obtiene como una media ponderada de los valores de la propia variable desplazados en el tiempo.

Seguiremos el mismo procedimiento que en el modelo de medias móviles, definiendo y caracterizando el de orden 1 y el de orden p .

Proceso autorregresivo de orden 1, $AR(1)$

— Viene expresado de la forma:

$$X_t = a_1 X_{t-1} + \delta + \varepsilon_t \quad , \quad a_1, \delta \rightarrow \text{constante}$$

siendo su expresión abreviada en función del operador de retardos:

$$\varepsilon_t = (1 - a_1 B)X_t - \delta$$

De la misma definición se deduce que el modelo de camino aleatorio es un caso particular de este proceso autorregresivo cuando $a_1 = 1$.

— Las características de este proceso son:

- $E(X_t) = \delta + a_1 E(X_{t-1})$
- $Var(X_t) = a_1^2 Var(X_{t-1}) + Var(\varepsilon_t)$

Cuando el proceso sea estacionario se verifica que la esperanza y la varianza son constantes en el tiempo.

• Autocovarianza:

$$\text{Para } m = 1: \quad Cov(X_t, X_{t+1}) = \frac{a_1 \sigma_\varepsilon^2}{1 - a_1^2}$$

$$\text{y para } m = 2: \quad Cov(X_t, X_{t+2}) = \frac{a_1^2 \sigma_\varepsilon^2}{1 - a_1^2}$$

- Autocorrelación: $\rho_0 = 1$

$$\rho_1 = a_1$$

$$\rho_m = a_1^m$$

Se observa que este proceso tiene una *memoria infinita*; es decir, el valor actual del proceso depende de todos los valores pasados, aunque el grado de dependencia disminuye con el tiempo.

- Una propiedad importante de este proceso es que cuando es estacionario ($|a_1| < 1$) es equivalente a un proceso de media móvil de orden infinito. Además, para cualquier proceso autorregresivo, de cualquier orden, existe un proceso de media móvil de orden infinito equivalente, de forma que el proceso autorregresivo es invertible en un proceso de media móvil.

Proceso autorregresivo de orden p , $AR(p)$

— La expresión de este proceso es:

$$X_t = a_1 X_{t-1} + \dots + a_p X_{t-p} + \delta + \varepsilon_t$$

expresión equivalente a: $a(B)X_t = \varepsilon_t - \delta$

siendo $a(B)$ el operador autorregresivo:

$$a(B) = 1 - a_1 B - \dots - a_p B^p$$

— Los coeficientes de autocorrelación:

$$\rho_1 = a_1 + a_2 \rho_1 + \dots + a_p \rho_{p-1}$$

$$\vdots$$

$$\rho_p = a_1 \rho_{p-1} + a_2 \rho_{p-2} + \dots + a_p$$

son las ecuaciones de Yule-Walker y permiten obtener estimaciones de los parámetros $a_1 \dots a_p$ para $p \geq 2$. Además:

$$\rho_m = a_1 \rho_{m-1} + a_2 \rho_{m-2} + \dots + a_p \rho_{m-p} \quad \text{,,} \quad m > p$$

MODELOS *ARMA*

Existen muchos procesos aleatorios para los que no es posible construir un modelo únicamente de media móvil o un modelo puramente autorregresivo, puesto que presentan características de ambos tipos de procesos. La generalización lógica de los modelos presentados en las dos secciones anteriores es el proceso mixto, autorregresivo de media móvil, de orden (p, q) , que se simboliza por *ARMA*(p, q) y se representa mediante la ecuación:

$$X_t = a_1 X_{t-1} + \dots + a_p X_{t-p} + \delta + \varepsilon_t - b_1 \varepsilon_{t-1} - \dots - b_q \varepsilon_{t-q}$$

Proceso *ARMA*(1, 1)

— La ecuación del proceso será:

$$X_t = a_1 X_{t-1} + \delta + \varepsilon_t - b_1 \varepsilon_{t-1}$$

— Las características del proceso son las siguientes:

- La varianza y las covarianzas de este proceso se determinan a partir de la expresión general, multiplicando por X_{t+m} y aplicando esperanzas se tiene que:

para $m = 0$:
$$E(X_t^2) = \text{Var}(X_t) = \frac{1 + b_1^2 - 2a_1b_1}{1 - a_1^2} = \sigma_\varepsilon^2$$

para $m = 1$:
$$\text{Cov}(X_t, X_{t+1}) = a_1 \text{Var}(X_t) - b_1 \sigma_\varepsilon^2$$

para $m = 2$:
$$\text{Cov}(X_t, X_{t+2}) = a_1 \text{Cov}(X_t, X_{t+1})$$

$$\text{Cov}(X_t, X_{t+m}) = a_1 \text{Cov}(X_t, X_{t+m-1}) \quad m \geq 2$$

- La función de autocorrelación viene dada por:

$$\rho_1 = \frac{(1 - a_1b_1)(a_1b_1)}{1 + b_1^2 - 2a_1b_1}$$

y para un desplazamiento $m > 1$, $\rho_m = a_1 \rho_{m-1}$

Proceso ARMA(p, q)

— La ecuación que representa este proceso es:

$$X_t = a_1 X_{t-1} + \dots + a_p X_{t-p} + \delta + \varepsilon_t - b_1 \varepsilon_{t-1} - \dots - b_q \varepsilon_{t-q}$$

expresión que puede presentarse también en función del operador de retardos B :

$$B \varepsilon_t = \varepsilon_{t-1}, \quad B^2 \varepsilon_t = \varepsilon_{t-2} \quad \dots \quad B^n \varepsilon_t = \varepsilon_{t-n}$$

tal que:

$$(1 - a_1 B - a_2 B^2 - \dots - a_p B^p) X_t = \delta + (1 - b_1 B - b_2 B^2 - \dots - b_q B^q) \varepsilon_t$$

cuya expresión abreviada es:

$$AR(B) X_t = MA(B) \varepsilon_t + \delta \quad \text{ó} \quad A(B) X_t = B(B) \varepsilon_t + \delta$$

— Como características del proceso se tienen:

- Si el proceso es estacionario, la media permanece constante a lo largo del tiempo y viene dada por la expresión:

$$\mu = a_1 \mu + \dots + a_p \mu + \delta$$

$$\mu = \frac{\delta}{1 - a_1 - \dots - a_p}$$

de lo que se deduce que una condición necesaria para que el proceso sea estacionario es que $(a_1 + \dots + a_p) < 1$.

- La varianza, las covarianzas y la función de autocorrelación son soluciones de ecuaciones en diferencias que generalmente no pueden resolverse directamente.
- Las covarianzas vienen dadas por:

$$Cov(X_t, X_{t+m}) = a_1 Cov(X_t, X_{t+m-1}) + a_2 Cov(X_t, X_{t+m-2}) + \dots +$$

$$+ \dots + a_p Cov(X_t, X_{t+m-p}) \quad m \geq q + 1$$

- La función de autocorrelación:

$$\rho_m = a_1 \rho_{m-1} + a_2 \rho_{m-2} + \dots + a_p \rho_{m-p} \quad , \quad m \geq q + 1$$

Teniendo en cuenta que q es la memoria de la parte del proceso correspondiente a la media móvil, para $m > q + 1$ la función de autocorrelación y las covarianzas tienen las propiedades de un proceso puramente autorregresivo.

MODELOS *ARIMA*

Al describir los modelos de medias móviles se vio que eran estacionarios; en los modelos autorregresivos y en los *ARMA* se suponía la verificación de las condiciones necesarias para garantizar que el proceso fuese estacionario. Sin embargo, existen muchas series temporales que no son estacionarias, lo que significa que las características del proceso estocástico subyacente varían con el transcurso del tiempo. La mayoría de las series económicas son no estacionarias, ya que la simple observación de una tendencia en el tiempo, o de unas fluctuaciones que crecen en tamaño con el paso del tiempo bastan para rechazar el supuesto de estacionariedad.

Cuando se plantea esta situación es frecuente que, al tomar primeras o segundas diferencias de la serie, respectivamente:

$$\Delta X_t = X_t - X_{t-1} \quad \text{ó} \quad \Delta^2 X_t = \Delta X_t - \Delta X_{t-1}$$

se obtengan series transformadas que son estacionarias. En estos casos se trabaja con las series en primeras o segundas diferencias, especificando y estimando un modelo para ellas. Los procesos no estacionarios que pueden transformarse en estacionarios mediante sus diferencias de orden d se les denomina *procesos estocásticos no estacionarios, homogéneos de orden d* .

Se dice, por tanto, que la serie X_t es no estacionaria homogénea de orden d si la serie $W_t = \Delta^d X_t$ es estacionaria. De esta expresión se deduce que $X_t = \Sigma^d W_t$ donde Σ es el operador suma, que es el inverso del operador en diferencias Δ .

Si $W_t = \Delta^d X_t$ y W_t es un proceso $ARMA(p, q)$, se dice que X_t es un proceso autorregresivo integrado de media móvil de orden (p, d, q) ó, simplemente, un proceso $ARIMA(p, d, q)$.

La ecuación correspondiente al proceso $ARIMA(p, d, q)$ se puede escribir de la forma siguiente:

$$AR(B) \Delta^d X_t = \delta + MA(B) \varepsilon_t$$

$$AR(B) = 1 - a_1 B - a_2 B^2 - \dots - a_p B^p \rightarrow \text{operador autorregresivo}$$

$$MA(B) = 1 - b_1 B - b_2 B^2 - \dots - b_q B^q \rightarrow \text{operador de media móvil}$$

- Si W_t es $AR(p)$ se dice que X_t es un proceso autorregresivo integrado de orden (p, d) y se representa por $ARI(p, d, 0)$.
- Si W_t es $MA(q)$ decimos que X_t es un proceso integrado de media móvil de orden (d, q) y se representa por $IMA(0, d, q)$.
- Si $p = q = 0$ entonces X_t sigue un modelo $ARIMA(0, 1, 0)$ que es el denominado *paseo aleatorio*.
- Si $p = q = d = 0$ entonces X_t sigue un modelo $ARIMA(0, 0, 0)$ que es el denominado *ruido blanco*.

Al analizar la evolución de la mayoría de las series temporales económicas suele observarse que éstas presentan una tendencia creciente o decreciente. Como ya hemos indicado, la eliminación de esta tendencia (es decir, la eliminación de la no-estacionariedad en media) de la serie suele conseguirse mediante diferenciaciones implícitas en los modelos $ARIMA$. Ahora bien, en ocasiones se observa también que existe una tendencia en la varianza, esto es, la dispersión de las observaciones no es constante a lo largo del tiempo, la cual no se elimina mediante esas diferenciaciones. Cuando se presenta este hecho la transformación adecuada puede consistir en tomar logaritmos neperianos.

Esta posibilidad de transformar la serie puede concretarse de forma más general mediante la transformación de Box-Cox. De esta forma, el modelo *ARIMA* más general puede enunciarse como:

$$AR(B) \Delta^d X_t^{(\lambda)} = \delta + MA(B) \varepsilon_t$$

o bien:

$$AR(B) \Delta^d (X_t^{(\lambda)} - \mu) = MA(B) \varepsilon_t$$

donde μ es la media de $X_t^{(\lambda)}$, siendo:

$$X_t^{(\lambda)} = \begin{cases} \frac{X_t^\lambda - 1}{\lambda} & \text{para } \lambda \neq 0 \\ \ln X_t & \text{para } \lambda = 0 \end{cases}$$

Restringiéndose a los casos $\lambda = 1$ y $\lambda = 0$ tendremos respectivamente la serie sin transformar (serie original) o la serie transformada aplicando logaritmos neperianos.

Con el fin de comprender mejor las características de las funciones de autocorrelación y autocorrelación parcial descritas para los distintos modelos, se recoge en el epígrafe A.1.1 (p. 567) la representación gráfica de dichas funciones teóricas, que son de gran utilidad en el proceso de identificación de los modelos apropiados para modelizar series reales.

MODELOS ESTACIONALES

Hay que señalar que cuando se trabaja con series temporales referidas a períodos inferiores al año, es frecuente que los datos de dichas series temporales presenten patrones estacionales. Por ejemplo, al trabajar con datos trimestrales o mensuales suele ocurrir que en determinados trimestres o meses se produzcan bruscos aumentos (o disminuciones) en la serie de manera sistemática a lo largo de los años. Para tales casos será

conveniente tener en cuenta el factor estacional en la elaboración de modelos estocásticos de series temporales, analizando las correlaciones entre los mismos meses o trimestres de años sucesivos, así como las existentes entre las sucesiones de trimestres o meses.

c. Aplicación de la metodología Box-Jenkins

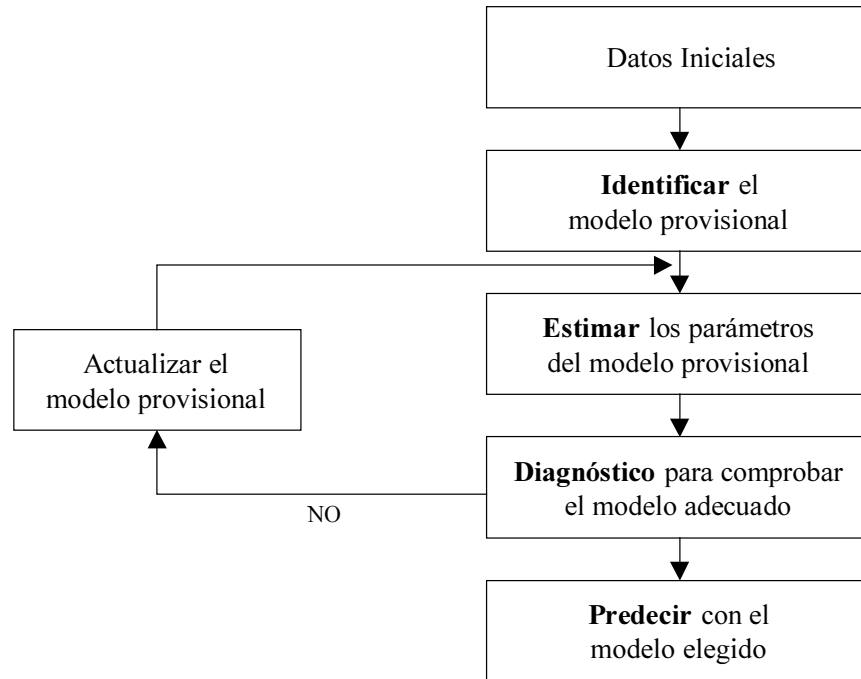
— El enfoque Box-Jenkins es de tipo iterativo y la aplicación del mismo supone siempre realizar una serie de etapas. En un primer paso se examinan las series sin procesar para identificar uno de los muchos modelos disponibles que se pueden elegir como la mejor representación de la serie (fase de *Identificación*). Si la serie bruta no es estacionaria, hay que modificarla para producir una serie estacionaria que hay que modelizar.

— El segundo paso en el proceso, una vez identificado el modelo, consiste en la estimación real de los parámetros (fase de *Estimación*). Este procedimiento es similar al ajuste de una regresión a un conjunto de datos. A continuación se realiza el diagnóstico para poder determinar si se ha elegido el modelo correcto (fase de *Validación*).

— El paso final consiste en predecir realmente, sirviéndose para ello del modelo elegido (fase de *Predicción*). Si se dispone de nuevas observaciones de la serie cronológica, debería estimarse de nuevo el modelo, comprobándose otra vez mediante el proceso de Box-Jenkins. Es muy probable que tengan que calcularse de nuevo los parámetros del modelo, o que tal vez tenga que seleccionarse otro modelo como la mejor representación de la serie. Los errores consistentes que puedan observarse en la estimación, a medida que se vaya disponiendo de más datos, pueden tomarse como una indicación de que el modelo entero podría requerir un cambio.

Esquemáticamente, la filosofía de Box-Jenkins queda reflejada en el siguiente cuadro, procediendo a continuación a su desarrollo.

Figura II.27
Fases de la metodología Box-Jenkins



IDENTIFICACIÓN

Partiendo de que los modelos expuestos se refieren a procesos estacionarios, el primer requisito que debe tenerse en cuenta para la identificación del modelo es que la serie objeto de estudio sea *estacionaria* tanto en media como en varianza. En caso de que la serie no sea estacionaria en media, debemos transformar la misma diferenciándola y determinando así el valor apropiado para el parámetro d . Respecto a la estacionariedad en varianza, se debe decidir si la serie se modifica o no mediante la transformación de Box-Cox determinando así el valor apropiado para el correspondiente parámetro λ .

Identificación de d

Una serie temporal será *estacionaria en media* cuando pueda mantenerse el supuesto de que existe una única media para toda la serie, esto es,

cuando ésta fluctúe en torno a un único valor medio. Así, será no estacionaria cuando presente cierta tendencia (creciente o decreciente), o cuando, sin un patrón tendencial claro, pueda aceptarse que hay varias medias en la serie temporal. Los instrumentos que permitirán responder a la cuestión de si la serie es o no estacionaria en media son:

- *El gráfico de la serie temporal*

Éste dará la primera idea de si la serie es estacionaria o no. Este gráfico es el que nos permite ver si la serie oscila en torno a algún valor sin alejarse del mismo, o si por el contrario, la serie presenta una tendencia creciente o decreciente, o bien existen varios tramos de la serie con valores por encima o por debajo de la misma.

- *El correlograma muestral de la serie temporal*

Basándonos en las consideraciones teóricas de los apartados anteriores, diremos que en los procesos estacionarios estudiados, la *FAC* tiende rápidamente a cero a medida que aumenta el desfase temporal considerado, es decir, a medida que es mayor el orden de la autocorrelación. Así, si X_t sigue un proceso $MA(q)$, la *FAC* se anula a partir del orden del proceso, y si sigue un $AR(p)$ o un $ARMA(p, q)$, el correlograma tiene un comportamiento amortiguado rápidamente hacia cero. Si la serie no es estacionaria, su correlograma no tenderá lentamente a cero, sino que las autocorrelaciones muestrales decrecerán lentamente.

Es conveniente seguir las siguientes reglas básicas:

- Si la serie temporal original es estacionaria en media no deberemos diferenciarla, es decir, asignaremos el valor $d = 0$.
- En el caso de que al analizar el gráfico de la serie temporal y su correlograma muestral se deduzca, siguiendo las indicaciones anteriores, que la serie no es estacionaria, se procederá a diferenciar una vez la serie temporal.

- Si el gráfico de esta serie temporal diferenciada así como su correlograma, indican que la serie diferenciada es estacionaria, el valor del parámetro d será igual a uno. Generalmente se da esta situación cuando existen segmentos de la serie temporal que difieren en el nivel, pero tienen la pendiente constante.
- Cuando la nueva gráfica y el nuevo correlograma, presentan a su vez las mismas características que los primeros, se procederá a practicar una segunda diferenciación (suele darse este caso cuando los distintos segmentos de la serie temporal difieran tanto en nivel como en pendiente).
- El proceso se repetirá hasta que se obtenga el valor de d que permita concluir que la serie $(1-B)^d X_t$ es estacionaria en media. Si bien los valores que suelen presentarse para d son cero, uno y dos. Además, mientras que una diferenciación se requiere con frecuencia, la segunda diferenciación sólo ocasionalmente resulta necesaria.

También suele ser útil analizar la *FAPM*. Así, si es preciso diferenciar la serie una vez, la representación gráfica de dicha función, tendrá el primer coeficiente de autocorrelación parcial muestral muy próximo a la unidad. Alternativamente el valor de d puede obtenerse a través del contraste de Dickey-Fuller (DF), o el de Dickey-Fuller Aumentado (DFA) según se indica en Aznar y Trávez (1993: 156).

Los instrumentos descritos pueden, no obstante, no ofrecer una clara evidencia de cuál es el grado apropiado de diferenciación a realizar en la serie temporal. Por ello un problema que se nos puede presentar es el de la *sobrediferenciación*, es decir, tomar más diferencias de las que son adecuadas. En este caso, el proceso seguirá siendo estacionario pero dejará de ser invertible. Si al estimar un proceso se obtiene que alguna (o algunas) de las raíces de la parte de medias móviles están próximas a la unidad, esto puede ser indicativo de que se ha efectuado una *sobrediferenciación*. Por

otra parte, en un proceso *sobrediferenciado* la varianza aumenta. Así, algunos autores proponen también como criterio para determinar el valor de d el que consiste en calcular la varianza tanto de la serie temporal como de la serie transformada mediante diferenciaciones, y tomar como diferenciación adecuada aquella para la cual la varianza es mínima. Si bien este criterio debe tomarse con cautela, pues no siempre se produce este hecho y debe ser acompañado de los otros métodos.

Identificación de λ

Se dice que una serie es *estacionaria en varianza* cuando la dispersión de la serie permanece invariable a lo largo del tiempo. Así pues, cuando observamos que una serie temporal fluctúa más fuertemente (la dispersión es mayor), o más débilmente (la dispersión es menor) en determinados períodos temporales, tendremos cierta evidencia de que dicha serie no es estacionaria en varianza. Para determinar cuando una serie es estacionaria en varianza tenemos dos instrumentos:

- *El gráfico de la serie temporal*

Si la dispersión de la misma es constante a lo largo del tiempo, la serie será estacionaria en varianza y por tanto asignaremos el valor $\lambda = 1$, ó lo que es lo mismo, no transformaremos la serie original. Si por el contrario, la dispersión no permanece constante, no habrá estacionariedad en varianza y transformaremos la serie original mediante logaritmos neperianos ($\lambda = 0$).

- *El gráfico rango-media*

Consiste en representar cada par de valores de rango y media en ordenadas y abscisas, respectivamente. Para calcular dichos valores se procede de la siguiente manera:

- Dada una serie temporal se divide la misma en grupos de observaciones, calculando para cada uno de ellos su valor medio y el recorrido (diferencia entre el valor mayor y el menor de las

observaciones incluidas en cada grupo). El tamaño del grupo es irrelevante, si bien el criterio general, es considerar dicho tamaño según la periodicidad de los datos: si son cuatrimestrales se tomarán tres observaciones (correspondientes al mismo año); si son trimestrales se tomarán cuatro observaciones; y así sucesivamente.

- Si los puntos resultantes de la representación están más o menos alineados en torno a una línea recta con pendiente positiva, será indicativo de que la dispersión aumenta a medida que lo hacen los valores medios o, lo que es lo mismo, que no hay estacionariedad en varianza ($\lambda = 0$). En este caso, por lo tanto, transformaremos la serie temporal original mediante logaritmos neperianos.
- Si el gráfico no muestra un esquema claro, o bien los puntos se encuentran alineados en torno a una línea paralela al eje de abscisas, indicará que la dispersión de la serie temporal se mantiene constante, y por lo tanto la serie es estacionaria en varianza y $\lambda = 1$, dejando la serie sin transformar.

En la práctica hay que determinar el valor de λ con anterioridad al de d , y esto es debido a que si diferenciamos previamente la serie temporal original podrían existir valores negativos a los que no podríamos posteriormente aplicar la transformación logarítmica.

Transformación de la serie

Una vez determinados los valores d y λ , y efectuadas las correspondientes transformaciones de la serie temporal según dichos parámetros, la serie temporal será estacionaria en media y en varianza (W_t) y se podrá escribir:

$$W_t = \Delta^d X_t^{(\lambda)}$$

Al hacer d diferenciaciones regulares en la serie temporal, el número de observaciones disponibles de la serie transformada es $T - d$, donde T es el número de observaciones de la serie temporal original.

Determinación del proceso generador de la serie

A continuación, siguiendo la metodología Box-Jenkins, se tratará de identificar los procesos *ARMA* susceptibles de generar la serie temporal estacionaria W_t . Una vez que ya sea estacionaria, se trata de averiguar, mediante las funciones *FAC* y *FAP* (*Función de Autocorrelación* y *Función de Autocorrelación Parcial*, respectivamente) los posibles valores de la parte regular del modelo (autorregresiva, p , y medias móviles, q). Dado que se van a obtener a partir de la serie temporal original y no del proceso estocástico que la genera, la identificación se realiza a través de la *Función de Autocorrelación Muestral (FACM)* y de la *Función de Autocorrelación Parcial Muestral (FAPM)*.

Por ello, es necesario algún procedimiento que nos permita comparar el comportamiento de las *FACM* y *FAPM* con el comportamiento teórico de las *FAC* y *FAP* para cada uno de los diferentes modelos. Para realizar adecuadamente esta comparación hay que tener en cuenta la distribución muestral de estos estimadores. Para los coeficientes de autocorrelación muestrales y por tanto para la *FACM*, Barlett demostró que para los procesos estocásticos estacionarios gaussianos (es decir, bajo el supuesto de normalidad) y para $m > 0$, los r_m se distribuyen asintóticamente como una normal. Además, en procesos en los que la *FAC* se anula a partir de un determinado desfase temporal (por ejemplo en los *MA* y en el *ruido blanco*) dicha normal tiene media cero para $m > q$ y su varianza se simplifica. Siguiendo este planteamiento se establecen al nivel de significación $\alpha = 0,05$ las correspondientes bandas de significatividad de las autocorrelaciones muestrales que casi todos los programas de ordenador incorporan a los correlogramas.

Puesto que es frecuente la existencia de más de un modelo que presumiblemente haya podido generar la serie temporal, se trata de catalogar todos los posibles, discriminando en las etapas posteriores de la metodología, bien porque algunos de ellos no superen la etapa de

Validación, o bien porque alguno de ellos no resulte el más apropiado atendiendo a diversos criterios de selección de modelos.

Identificación del término independiente

Además de determinar los posibles procesos tipo $ARMA(p, q)$ que pueden generar la serie temporal W_t , también debe determinarse si en el modelo hay que incluir término constante o no. En el caso de procesos que no contienen parte autorregresiva se denota con μ pues en este tipo de procesos la esperanza matemática coincide con dicho parámetro. En el resto se denota por δ .

Se incluirá término independiente si la serie W_t tiene media significativamente distinta de cero. Para realizar este contraste, tomando como hipótesis nula $\mu = 0$, se utiliza el siguiente estadístico:

$$t = \frac{\bar{W}}{\hat{\sigma}_{\bar{w}}}$$

siendo \bar{W} la media muestral de la serie W_t y aproximando $\hat{\sigma}_{\bar{w}}$, mediante la expresión:

$$\hat{\sigma}_{\bar{w}} = (C_0 / T)^{1/2}$$

Considerando el nivel de significación habitual se realiza el contraste, teniendo en cuenta que el estadístico indicado se distribuye según una t de Student con $T - 1$ grados de libertad. También puede optarse por incluir el término independiente en el modelo y analizar, una vez estimado el mismo, si la estimación de dicho parámetro es significativa o no, pudiendo optar en este último caso por su eliminación.

ESTIMACIÓN

Una vez identificados los posibles procesos estocásticos generadores de la serie temporal objeto de estudio, se trata de estimar los parámetros de

dichos modelos. Hay dos enfoques: condicional y no condicional (exacto). La diferencia entre uno y otro radica en la forma de afrontar el problema de los valores iniciales del modelo. En caso de que se les asigne a priori algún valor, y se proceda a efectuar la estimación de los parámetros considerando estos valores permanentemente asignados, estaremos aplicando el enfoque de estimación condicional. Por el contrario, si se estiman los parámetros del modelo junto con dichos valores iniciales, estaremos aplicando el enfoque no condicional (o exacto).

Ahora bien, cualquiera que sea el enfoque que adoptemos y como consecuencia de la no linealidad, será necesario aplicar algoritmos de estimación no lineales. En algunos casos para lograr la convergencia del proceso iterativo se usan algoritmos, como por ejemplo el de Marquardt. Generalmente los paquetes estadísticos establecen una serie de criterios en cuanto a la variación de los parámetros, la función de verosimilitud y el número de iteraciones completadas, para dar por finalizado el proceso. En nuestro caso y para el SPSS se indicarán al realizar el análisis empírico.

VALIDACIÓN

La fase de Validación o evaluación va dirigida a establecer si existe o no adecuación entre los datos y el modelo estimado. Después de especificar un modelo de series temporales, y estimar sus parámetros, se debe realizar una comprobación para verificar si la especificación inicial era o no correcta. Como señalan Box y Jenkins, este aspecto se corresponde sólo en parte con el problema clásico del contraste de la bondad del ajuste. La validación busca no sólo el contraste del modelo identificado y estimado sino además obtener información sobre la dirección a seguir en las modificaciones subsiguientes.

En virtud de lo expuesto, el modelo más adecuado para representar el comportamiento de una serie temporal es aquél que verifica:

- Los coeficientes son estadísticamente significativos y están poco correlacionados entre sí.
- Los residuos se aproximan al comportamiento de un ruido blanco.
- Los coeficientes del modelo son suficientes para representar la serie.
- La bondad del ajuste es elevada en comparación a la de otros modelos alternativos.
- El modelo es estable.

Por tanto, la finalidad de esta fase es analizar en qué medida se cumplen los anteriores requisitos. En este sentido, conviene señalar que es esencial que se cumpla el primero de ellos, ya que en caso contrario el modelo debe ser rechazado, pues esto sería indicativo de que los residuos contienen información relevante para la predicción.

Análisis de los coeficientes estimados

Con este análisis se pretende estudiar si los coeficientes son significativos, así como si el modelo es estacionario e invertible. La significatividad de los parámetros se estudia teniendo en cuenta la distribución muestral de los estadísticos que se emplean como estimadores. El vector de coeficientes estimados, para el modelo general *ARIMA* se expresa como:

$$\hat{\beta}' = (\hat{\delta}, \hat{a}_1 \dots \hat{a}_p, \hat{b}_1 \dots \hat{b}_q)$$

Bajo el supuesto de normalidad adoptado se distribuirá asintóticamente como una normal multivariante:

$$\hat{\beta} \approx N(\beta, V(\hat{\beta}))$$

siendo $V(\hat{\beta})$ la matriz de varianzas y covarianzas de los estimadores de los coeficientes del modelo, cuya estimación puede expresarse como:

$$\hat{V}(\hat{\beta}) = \hat{\sigma}_\varepsilon^2 (X_f' X_f)^{-1}$$

donde X_f es el valor de la matriz X obtenida en la última iteración del proceso de estimación y $\tilde{\sigma}_\varepsilon^2$ la varianza estimada.

Por tanto, dentro de un contexto asintótico, será posible contrastar la hipótesis nula del tipo: $H_0 : \beta_i = \beta_{i0}$ mediante el estadístico:

$$t = \frac{\hat{\beta}_i - \beta_{i0}}{\hat{\sigma}_{\hat{\beta}_i}}$$

con los grados de libertad iguales al tamaño de la serie utilizada en la estimación menos el número de parámetros estimados en el modelo; es decir, $T - p - q - 1$.

A partir de la expresión anterior puede efectuarse el contraste de significatividad individual de cada coeficiente, adoptando como hipótesis nula:

$$H_0 : \beta_{i0} = 0, \quad i = 1, 2 \dots (p + q + 1)$$

y como alternativa:

$$H_1 : \beta_{i0} \neq 0, \quad i = 1, 2 \dots (p + q + 1)$$

El estadístico de contraste que resulta en este caso es:

$$t = \frac{\hat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}}$$

y fijando un nivel de significación α , concluiremos que el coeficiente β_i es significativo cuando el valor absoluto del estadístico $|t|$ sea superior al valor en tablas $t_{\alpha/2}$ (con $T - p - q - 1$ grados de libertad). Cuando $\alpha = 0,05$ el valor de $t_{0,05}$ (con $T - p - q - 1$ grados de libertad) es, para tamaños muestrales suficientemente elevados, aproximadamente igual a dos.

Una cuestión importante en el análisis de los coeficientes estimados es comprobar si cumplen las condiciones de estacionariedad e invertibilidad. Así, por ejemplo, si hemos estimado un modelo $ARMA(1, 1)$, bastaría con comprobar si se cumple que los coeficientes estimados son en valor absoluto inferiores a uno.

Hay que señalar que, dado que se consideran estimaciones de los parámetros en lugar de valores conocidos de los mismos, no basta con que se cumplan las condiciones sin más, sino que debe ser de forma significativa. Así, en modelos simples, en los que la parte autorregresiva y de medias móviles no supere el orden uno, será necesario comprobar no sólo que los valores absolutos de los coeficientes estimados son distintos de uno, sino que además, en los intervalos de confianza relativos a cada coeficiente, se incluyen sólo valores comprendidos entre -1 y 1 .

Se puede completar el análisis estudiando la matriz de correlación entre los coeficientes estimados con el objeto de comprobar si existe multicolinealidad. Si la correlación entre dos coeficientes estimados está próxima a uno existe un problema de multicolinealidad; esto significa que dichos coeficientes pueden cambiar mucho de una realización a otra. Entonces es conveniente eliminar algún parámetro del modelo con el objeto de conseguir estimadores más estables aún a costa de que el grado de ajuste sea menor.

Análisis de los residuos

Sabemos que si los residuos constituyen un proceso de ruido blanco se distribuyen normalmente y son independientes entre sí; si el modelo se ha especificado correctamente, los residuos $\hat{\varepsilon}_t$ tienden a tener las mismas propiedades que ε_t .

- *Media Cero*

Los residuos de un modelo *ARIMA* no están, en principio, sujetos a la restricción de media cero, salvo en el caso en que se aplique el enfoque de estimación mínimo-cuadrática condicional a un proceso *AR* puro, puesto que el procedimiento de estimación que se utiliza es no lineal.

Si bien, es frecuente analizar si esto se cumple o no empleando un contraste de significatividad de la media de los residuos, calculando el cociente entre la media muestral de los residuos ($\bar{\hat{\varepsilon}}$) y su desviación

típica ($\hat{\sigma}_{\hat{\varepsilon}}$), de manera que si el valor del mismo no excede de dos, suponiendo que adoptamos el nivel de significación habitual, se puede aceptar que la media no difiere significativamente de cero.

- *Varianza Constante*

La constancia de la varianza suele comprobarse examinando el gráfico de los residuos. Si se observa que la dispersión de los mismos no es constante a lo largo del tiempo concluiremos que no puede aceptarse la hipótesis de homocedasticidad, lo cual nos sugiere que efectuar una transformación (habitualmente, la logarítmica) en la serie original puede ser recomendable. Así pues, desde el análisis gráfico de los residuos puede llegar a cuestionarse la transformación de la serie original, debiendo en este caso volver a identificar la nueva serie transformada.

- *Distribución Normal*

La comprobación de si puede aceptarse que los residuos se distribuyen como una normal se efectuará calculando los coeficientes de asimetría y curtosis, respectivamente:

$$g_1 = \frac{m_3}{m_2^{3/2}} \quad ; \quad g_2 = \frac{m_4}{m_2^2} - 3$$

En el caso de que a partir de la aplicación de los contrastes referidos exista evidencia de que la hipótesis de normalidad pueda no ser cierta, suele optarse como solución efectuar una transformación Box-Cox a los datos (es decir, transformando la serie original en logaritmos neperianos), dado que muy a menudo este tipo de transformación elimina coeficientes significativos de asimetría y curtosis.

Para analizar de forma más exacta la normalidad de los residuos podemos usar el contraste de Kolgomorov-Smirnov, cuya hipótesis nula es:

$$H_0 : F(x) = F_0(x)$$

siendo $F_0(x)$ la distribución de probabilidad de la muestra objeto de análisis (los residuos en este caso) y $F(x)$ la distribución normal. Considerando el estadístico de prueba, cuya distribución es independiente del modelo propuesto, y al nivel de significación α habitual, rechazaremos la hipótesis nula de normalidad cuando el valor del estadístico se encuentre dentro de la región crítica.

Se puede señalar que el análisis de los residuos es el principal criterio a tener en cuenta de cara a evaluar la adecuación de un determinado modelo. Los dos instrumentos a los que en la práctica se ha prestado atención preferente son el gráfico de los residuos a lo largo del tiempo, ya comentado, y su correlograma muestral.

- *Residuos incorrelados*

Por tanto, para analizar la independencia de los residuos se utiliza la *Función de Autocorrelación Estimada (FACE)* y la *Función de Autocorrelación Parcial Estimada (FACPE)*, de forma que $\hat{\epsilon}_t$ será de ruido blanco cuando los coeficientes de ambas no sean significativamente distintos de cero. Este estudio se realiza mediante un contraste individual y mediante un contraste global de los coeficientes.

- Para el análisis de los coeficientes individuales de la *FACE* y la *FACPE* se construyen intervalos de confianza utilizando la distribución de una variable de *ruido blanco* cuya varianza es $1/T$. Sin embargo, este valor no constituye una buena aproximación ya que la distribución de $\rho_m \sim N(0, 1/T)$ no es aplicable a los coeficientes de autocorrelación de una serie de residuos, especialmente en los retardos de orden bajo.
- Para hacer un contraste global de los coeficientes se emplea el estadístico:

$$Q = T \sum_{m=1}^M \hat{\rho}_m^2$$

siendo $\hat{\rho}_m$:

$$\hat{\rho}_m = \frac{\sum \hat{\epsilon}_t \hat{\epsilon}_{t+m}}{\sum \hat{\epsilon}_t^2}$$

propuesto por Box y Pierce en 1970, que se distribuye como una χ^2_m con m grados de libertad bajo la hipótesis nula de que:

$$H_0 : \rho_1 = \dots = \rho_m = 0$$

(El valor de m es arbitrario; al tomar m elevado se tiene la ventaja de que puede captar los ρ_m significativos correspondientes a retardos elevados pero a medida que aumenta m disminuye la potencia del contraste).

Box y Pierce han demostrado que utilizando m razonablemente elevado, el estadístico Q construido a partir de los residuos de un modelo *ARMA* es apropiado si se supone una distribución χ^2 con $m - p - q$ grados de libertad bajo la hipótesis nula de que los residuos son independientes. De esta forma, al reducir los grados de libertad en razón al número de parámetros ajustados, la probabilidad de rechazar la hipótesis nula aumenta.

El estadístico Q de Box-Pierce ha sido refinado posteriormente con el fin de disminuir el sesgo en pequeñas muestras. Realizado por Ljung y Box en 1978, estos autores han propuesto el estadístico:

$$Q^* = T(T+2) \sum_{m=1}^M (T-m)^{-1} \hat{\rho}_m^2$$

que se distribuye como una χ^2 con $m - p - q$ grados de libertad, bajo la hipótesis nula de que los residuos son independientes.

En definitiva, se puede llevar a cabo una prueba estadística de hipótesis que verifique la exactitud del modelo comparando el valor observado de Q con los valores teóricos de χ^2 hallados en las tablas, de forma que:

$$Q^* < \chi^2_{m-p-q}(\alpha) \rightarrow \text{acepta } H_0$$

$$Q^* > \chi^2_{m-p-q}(\alpha) \rightarrow \text{rechaza } H_0$$

La determinación del nivel de significación se trata, en principio, de una decisión arbitraria ($\alpha = 5\%$ y $\alpha = 1\%$ son los más utilizados), sin embargo, con la utilización de los ordenadores, el contraste de hipótesis se puede realizar desde otra perspectiva, es decir, se determina el nivel de significación crítico: α' (*NSC* ó *P-value*) que se calcula de la forma:

$$P[\chi^2 > Q^*] = \alpha'$$

La operación de determinar α' es la inversa de buscar el valor de las tablas para un nivel de significación dado. Determinado α' se rechaza la hipótesis nula para todo nivel de significación $\alpha: \alpha > \alpha'$ y se acepta cuando $\alpha < \alpha'$; por tanto el *NSC* es un indicador del nivel de admisibilidad de H_0 : cuanto mayor sea el *NSC*, mayor confianza se puede depositar en la hipótesis nula. En definitiva, la utilización del *NSC* implica dar la vuelta al problema del contraste de hipótesis: en lugar de fijar a priori un nivel de significación se calcula un valor α' que permite determinar a posteriori para qué niveles de significación se puede rechazar la hipótesis nula.

Algunos autores como Vandaele completan el análisis de los residuos tomando diferencias de primer orden y calculando la *FACE* y la *FACPE* de la serie diferenciada.

Conviene señalar que puesto que los verdaderos valores de los parámetros van a desconocerse siempre, resulta imposible calcular la verdadera serie ε_t y en definitiva, los coeficientes de autocorrelación muestrales. Por ello para analizar si los residuos son o no un ruido blanco, debemos comprobar además de si están o no incorrelados como ya hemos indicado, si tienen media cero, varianza constante, y distribución normal.

Introducción de parámetros adicionales en el modelo

Una vez estimados los parámetros es aconsejable examinar si se puede mejorar la adecuación entre datos y modelo introduciendo parámetros adicionales. Este aspecto se justifica cuando en el proceso de estimación los coeficientes nuevos resultaran estadísticamente significativos. No obstante, conviene no introducir parámetros de forma simultánea, en la parte autorregresiva y en la de medias móviles, con el fin de evitar posibles problemas de redundancia. Si los nuevos parámetros no son significativos, los coeficientes existentes son suficientes para representar la serie.

Bondad del ajuste

Si en la fase de identificación se especifican varios modelos alternativos que al estimar satisfacen los contrastes, hay que seleccionar mediante los criterios de bondad de ajuste cuál es el más adecuado.

Se suelen considerar para ello dos estadísticos:

$$R^2 = 1 - \frac{\hat{\sigma}_\varepsilon^2}{\hat{\sigma}_x^2} \quad \text{y} \quad \bar{R}^2 = 1 - \frac{\frac{\hat{\sigma}_\varepsilon^2}{T - p - q}}{\frac{\hat{\sigma}_x^2}{T - 1}}$$

que tienen estas expresiones cuando se trata de un proceso $ARMA(p, q)$ siendo:

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^N \hat{\varepsilon}_i^2}{T} \quad \text{y} \quad \hat{\sigma}_x^2 = \frac{\sum_{t=1}^N (x_t - \bar{x})^2}{T}$$

donde x_t es una serie estacionaria (es más aconsejable utilizar \bar{R}^2 porque tiende a evitar la sobreparametrización).

En los modelos $ARIMA$ estas expresiones presentan el inconveniente de que $\hat{\sigma}_x^2$ depende del número de diferencias. Luego estos estadísticos sólo tienen validez cuando se comparen modelos con el mismo número de

diferencias. Obviamente cuando R^2 ó \bar{R}^2 son aproximadamente igual a uno el modelo sería el adecuado.

Análisis de estabilidad

Diremos que un modelo no es estable, lo cual conllevará problemas de cara a obtener buenas predicciones en el futuro, cuando aun pudiendo aceptarse la permanencia de la especificación del mismo para todo el período muestral considerado, los valores de los parámetros cambian de manera significativa.

Para analizar dicha estabilidad se pueden tomar dos períodos muestrales distintos, y estimar sus parámetros, pudiendo concluir que el modelo es estable cuando las estimaciones de los parámetros no sean muy diferentes. La forma de dividir la muestra es totalmente aleatoria y nada impide hacer cualquier división. Si bien resulta recomendable, con el fin de que las submuestras no queden con un número de observaciones muy pequeño, eliminar un número de ellas de los datos disponibles originalmente (por ejemplo ocho si los datos son cuatrimestrales) y volver a estimar los parámetros del modelo. Si, por ejemplo, ningún coeficiente estimado para cada una de estas dos muestras difiere en $\pm 0,1$ podríamos aceptar la estabilidad del modelo.

Para esta finalidad también se puede aplicar como aproximación, el test de Chow, que considera el estadístico:

$$F_{M,T-2M} = \frac{\sum_{t=1}^T \hat{\epsilon}_t^2 - \left(\sum_{t=1}^{N_1} \hat{\epsilon}_{1t}^2 + \sum_{t=1}^{N_2} \hat{\epsilon}_{2t}^2 \right)}{\frac{M}{\frac{\sum_{t=1}^{N_1} \hat{\epsilon}_{1t}^2 + \sum_{t=1}^{N_2} \hat{\epsilon}_{2t}^2}{T - 2M}}}$$

$N = N_1 + N_2$, , $M = p + q + \delta$

N_1 : Tamaño de la primera submuestra

N_2 : Tamaño de la segunda submuestra

$\hat{\varepsilon}_t \rightarrow$ residuo del modelo utilizando en la estimación todo el período muestral N .

$\hat{\varepsilon}_{1t} \rightarrow$ residuo del modelo utilizando en la estimación los N_1 primeros datos.

$\hat{\varepsilon}_{2t} \rightarrow$ residuo del modelo utilizando en la estimación los N_2 últimos datos.

Si $F < F(\text{teórico}) \rightarrow$ acepta hipótesis estabilidad

Si $F > F(\text{teórico}) \rightarrow$ rechaza hipótesis estabilidad

Cuando el modelo analizado cumple adecuadamente los requisitos que hemos señalado en esta etapa estará bien especificado y no omitirá parámetros relevantes.

Reformulación del modelo

Si después de comprobar todos los análisis y contrastes correspondientes a la fase de validación se llega a que el modelo seleccionado no es el adecuado, se debe reformular. La *FACE* y la *FACPE* de los residuos del modelo estimado son instrumentos valiosos a la hora de reformular el modelo. La forma en que se realiza para un modelo *AR*(1) es la siguiente:

Se considera el modelo estimado: $(1 - \hat{a}_1 B)X_t = \hat{\varepsilon}_t$

Examinadas ambas funciones se llega a que $\hat{\varepsilon}_t$ sigue un modelo *MA*(1), es decir:

$$\hat{\varepsilon}_t = (1 - b_1 B)\varepsilon_t$$

de lo que se deduce sustituyendo, que:

$$(1 - \hat{a}_1 B)X_t = (1 - b_1 B)\varepsilon_t$$

luego X_t sigue un modelo *ARMA*(1, 1).

Si, por el contrario, los $\hat{\varepsilon}_t$ siguieran un $AR(1)$ en vez de un $MA(1)$ se tendría:

$$\begin{aligned}(1 - a_1^* B)\hat{\varepsilon}_t &= \varepsilon_t \Rightarrow \\(1 - \hat{a}_1 B)X_t &= (1 - a_1^* B)^{-1}\varepsilon_t \Leftrightarrow \\(1 - \hat{a}_1 B)(1 - a_1^* B)X_t &= \varepsilon_t\end{aligned}$$

lo que llevaría a la reformulación de un modelo $AR(2)$.

PREDICCIÓN

En la fase de Predicción se realizan los pronósticos sobre valores futuros de la variable en términos probabilísticos. En el caso de existir discrepancias de carácter sistemático entre los valores pronosticados y los observados se deberá cuestionar la validez del modelo identificado.

El objetivo fundamental de la predicción consiste en determinar o pronosticar los valores futuros de una serie temporal de forma que el error que se cometa sea lo más pequeño posible. Por esta razón, se considera como predicción "óptima" aquella que tenga el "mínimo error cuadrático medio". Ahora bien, como el error de predicción es una variable aleatoria se minimiza el "valor esperado", siendo el error medio cuadrático de la predicción la varianza del error de predicción. En definitiva, la medida de la bondad de la predicción viene dada por la varianza, de forma que la predicción será mejor cuanto menor sea ésta.

Designando por $\hat{x}_T(h)$ la predicción tomando como base T para el período futuro h , se pretende que dicha predicción sea tal que minimice la expresión:

$$E[e_T^2(h)] = E\{[x_{t+h} - \hat{x}_t(h)]^2\}$$

Sin entrar en más consideraciones vamos a realizar el cálculo de las predicciones para los distintos modelos.

Predicciones puntuales

- *Modelos autorregresivos*

A partir de la expresión general del modelo $AR(1)$:

$$x_t = a_1 x_{t-1} + \varepsilon_t + \delta$$

la predicción para h períodos sucesivos será:

$$\hat{x}_{T+h} = a_1^h x_T + (1 - a_1^h) E(x_T)$$

ya que: $E(x_t) = a_1 E(x_{t-1}) + E(\varepsilon_t) + \delta$

$$\delta = E(x_t) - a_1 E(x_{t-1}) = E(x_t) (1 - a_1)$$

De esta igualdad se deduce que la predicción óptima, h períodos hacia el futuro, es una combinación lineal convexa de la última observación de la variable (x_T) y de su esperanza matemática.

Por tanto, al avanzar la última observación, se tiene una ponderación menor y la esperanza recibe un peso mayor; esto significa que, en un proceso estacionario, cuanto más lejos hacia el futuro se quiera predecir, mayor será la incertidumbre bajo la que se obtiene la predicción. Ante una total incertidumbre, la predicción óptima del valor de una variable aleatoria es igual a su esperanza matemática; por el contrario, cuando se predice a horizontes cortos, la información muestral permite mejorar la predicción que se haría si se utilizase únicamente la esperanza matemática de la variable aleatoria.

La propiedad característica del proceso $AR(1)$ es que toda la información muestral relevante para la predicción queda resumida en el último valor observado de la variable. En consecuencia, dicho valor es todo lo que se precisa, junto con la esperanza matemática del proceso, para elaborar predicciones.

Siguiendo un razonamiento análogo en los procesos $AR(p)$, la información relevante para la predicción se resume en las p últimas observaciones de la variable a predecir.

- *Modelos de medias móviles*

Para un modelo $MA(1)$, dado por: $x_t = \varepsilon_t - b_1 \varepsilon_{t-1} + \delta$

se puede expresar el valor: $x_{T+1} = \varepsilon_{T+1} - b_1 \varepsilon_T + \delta$

y el valor de la predicción será:

$$\hat{x}_{T+h} = \delta$$

Esto significa que para $h > 1$ la predicción es la constante, resultado coherente con la afirmación ya establecida de que este modelo tiene memoria para un sólo periodo.

De la misma forma se puede comprobar que en un modelo $MA(q)$:

$$\hat{x}_{T+h} = \delta \quad \text{para } h \geq q + 1$$

- *Modelos ARMA*

Las expresiones analíticas correspondientes a las predicciones de valores futuros del modelo $ARMA(1, 1)$:

$$x_t = a_1 x_{t-1} + \delta + \varepsilon_t - b_1 \varepsilon_{t-1}$$

son:

$$\hat{x}_{T+h} = a_1^h x_T + (1 - a_1^h) E(x_T) - a_1^{h-1} b_1 \hat{\varepsilon}_T$$

- *Modelos ARIMA*

Dado el modelo $ARIMA(p, d, q)$:

$$w_t = a_1 w_{t-1} + \dots + a_p w_{t-p} + \varepsilon_t - b_1 \varepsilon_{t-1} - \dots - b_q \varepsilon_{t-q} + \delta$$

siendo: $x_t = \sum^d w_t$

Para el periodo $T + 1$ tendremos:

$$w_{T+1} = a_1 w_T + \dots + a_p w_{T-p+1} + \varepsilon_{T+1} - b_1 \varepsilon_T - \dots - b_q \varepsilon_{T-q+1} + \delta$$

La predicción $\hat{w}_T(h)$ se obtiene a partir de la esperanza condicionada.

Si $h > p$ y $h > q$, la predicción será:

$$\hat{w}_T(h) = a_1 \hat{w}_T(h-1) + \dots + a_p \hat{w}_T(h-p)$$

donde ha desaparecido el componente de medias móviles.

Una vez efectuada la predicción de la serie w_t , se puede obtener la predicción de la serie original x_t , aplicando el operador suma a w_t , es decir, sumando d veces w_t , ya que $w_t = x_t - x_{t-1}$.

– Si $d = 1$, la predicción será:

$$\hat{x}_{T+h} = x_T + \hat{w}_{T+1} + \dots + \hat{w}_{T+h}$$

– Si $d = 2$:

$$\hat{x}_{T+h} = x_T + h\Delta x_T + h\hat{w}_{T+1} + (h-1)\hat{w}_{T+2} + \dots + \hat{w}_{T+h}$$

Se seguirá un procedimiento análogo para $d > 2$.

Error de predicción

Se define el *error de predicción* como la diferencia entre la realización de la variable aleatoria y la predicción hecha para dicho valor; llamando $e_T(h)$ al error de predicción h periodos hacia adelante, se tiene:

$$e_T(h) = x_{T+h} - \hat{x}_{T+h}$$

En todos los modelos considerados, la esperanza del error de predicción es cero, por ser el predictor óptimo la esperanza condicionada de la variable a predecir, ya que:

$$\begin{aligned} E[e_T(h)] &= E(x_{T+h} - \hat{x}_{T+h}) = \\ &= E(x_{T+h}) - E(\hat{x}_{T+h}) = E(x_{T+h}) - E(x_{T+h}) = 0 \end{aligned}$$

El error de predicción es una variable aleatoria con esperanza cero, siendo su varianza la que nos da una medida de su importancia, por lo que es la "varianza del error de predicción" una de las variables de más interés en dicho análisis.

La raíz cuadrada de dicha varianza es el valor que se utiliza para construir los intervalos de confianza de las predicciones puntuales obtenidas

en el apartado anterior. Los errores y sus varianzas para los modelos considerados son:

- *Modelo AR(1)*

$$e_T(h) = \varepsilon_{T+h} + a_1 \varepsilon_{T+h-1} + a_1^2 x_{T+h-2} + \dots + a_1^{h-1} x_{T+1}$$

La varianza del error de predicción será:

$$Var[e_T(h)] = \sigma_\varepsilon^2 (1 + a_1^2 + a_1^4 + \dots + a_1^{2(h-1)}) = \sigma_\varepsilon^2 \frac{1 - a_1^{2h}}{1 - a_1^2}$$

De esta expresión se deduce que la varianza del error de predicción aumenta, en forma no lineal, a medida que crece h .

- *Modelo MA(1)*

$$e_T(h) = \varepsilon_{T+h} - b_1 \varepsilon_{T+h-1} \quad , \quad h > 1$$

$$Var[e_T(h)] = (1 + b_1^2) s_\varepsilon^2 \quad , \quad h > 1$$

- *Modelo ARIMA*

Expresando el modelo *ARIMA* como un proceso de media móvil de orden ∞ , el error de predicción para h periodos posteriores será:

$$e_T(h) = x_{T+h} - \hat{x}_{T+h} = \psi_0 \varepsilon_{T+h} + \psi_1 \varepsilon_{T+h-1} + \dots + \psi_{h-1} \varepsilon_{T+1}$$

siendo:

$$\psi(B) = \alpha^{-1}(B) (1 - B)^{-d} \beta(B)$$

Supuesto que los parámetros $a_1 \dots a_p, b_1 \dots b_q$ se conocen exactamente, la varianza del error de predicción será:

$$E[e_T^2(h)] = (\psi_0^2 + \psi_1^2 + \dots + \psi_{h-1}^2) \sigma_\varepsilon^2$$

de donde se deduce que la forma algebraica de la varianza del error de predicción depende de la especificación concreta del modelo. No obstante, por la definición de $\psi(B)$ se sabe, que cualquiera que sea la especificación se verifica que $e_T(1) = \varepsilon_{T+1}$. Además, como los parámetros no se conocen con certeza puesto que son estimaciones, la varianza real del error será mayor que la que hemos calculado.

Intervalos de confianza para la predicción

Para determinar los intervalos de confianza de los pronósticos obtenidos, es necesario conocer la estimación de la varianza de la perturbación σ_ε^2 , que se basará en la suma de los cuadrados de los residuos, obtenida después de estimar los parámetros del modelo:

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{t=1}^T \hat{\varepsilon}_t^2}{T - p - q}$$

Teniendo en cuenta que: $E[e_T^2(h)] = (\psi_0^2 + \dots + \psi_{h-1}^2) \sigma_\varepsilon^2$

y que $\psi_0 = 1$, un intervalo de confianza de n desviaciones será:

$$\hat{x}_t(h) \pm n \left(1 + \sum_{j=1}^{h-1} \psi_j^2 \right)^{1/2} \hat{\sigma}_\varepsilon$$

Por otro lado, sabiendo que $e_T(1) = \varepsilon_{T+1}$ cualquiera que sea el modelo de la serie temporal en estudio, se tiene que el error es la componente del valor futuro de la serie que no puede predecirse sobre la base de la información muestral; esto se conoce en la literatura de series temporales como *innovación de la serie*. En este sentido, si la innovación ε_T sigue una distribución normal, el intervalo de confianza, para el nivel de confianza α será:

$$\hat{x}_t(h) \pm \lambda_\alpha \hat{\sigma}_{\varepsilon_T(h)}$$

Indicadores de validez de las predicciones

La validación o chequeo de los modelos identificados no se agota con la consideración de los criterios de diagnosis enunciados en la denominada etapa de Validación descrita anteriormente, sino que la evaluación de la calidad de las predicciones de valores futuros de la serie objeto de estudio se convierte en un instrumento sumamente importante de validación del modelo.

Éste es el motivo por el que puede observarse que, tras la obtención de predicciones, debemos analizar el hecho de si el modelo predice de forma satisfactoria, debiendo volver, en caso negativo, a reiniciar las etapas del método de Box-Jenkins. Adicionalmente, si varios modelos han cumplido los requisitos enunciados en la etapa de chequeo, y en consecuencia han sido utilizados para predecir, parece lógico proponer que el modelo que seleccionemos es aquel que logre predicciones lo más satisfactorias posibles.

Existe en el contexto de los modelos econométricos, numerosos criterios para la selección de modelos alternativos. Si bien, todos ellos se pueden agrupar según dos líneas básicas de actuación: verificacionista y preferencialista, distinguiendo dentro de estos últimos los criterios de información y de corroboración.

En cuanto al campo que nos ocupa los dos criterios más utilizados son el Criterio de Información de Akaike (*AIC*) que se define como:

$$AIC(j) = \ln \tilde{\sigma}_\varepsilon^2(j) + \frac{2k_j}{T}$$

donde $\tilde{\sigma}_\varepsilon^2(j)$ es el estimador máximo-verosímil del parámetro σ_ε^2 del modelo j -ésimo, k_j es el número de parámetros del modelo j -ésimo y T el número de observaciones disponibles. Está especialmente indicado para modelos autorregresivos y se tomaría aquel modelo cuyo valor para dicho indicador fuese el menor.

Un criterio más general sería el desarrollado por Schwartz, denominado *SBC* (Schwartz's Bayesian Criterion):

$$SBC(j) = T \ln \tilde{\sigma}_\varepsilon^2(j) + (k_j) \ln T$$

Este criterio tiene un carácter más general (Wei, 1990: 153) y el modelo seleccionado sería el que presentase un menor valor. En caso de discrepancia entre los dos indicadores, es el investigador el que debe elegir en última instancia.

Éste es el procedimiento que se puede aplicar a los h factores, obtenidos de la etapa de la infraestructura del *Statis*, y que constituyen la mejor aproximación de orden h de las variables originales. Una vez identificados, estimados y validados los modelos para dichos factores se obtienen los valores predichos de los mismos para uno y dos periodos hacia adelante ($T + 1$ y $T + 2$). Los valores de dichos factores, del momento 1 a T , permiten estimar los valores de las variables originales y comparar con los valores reales con el fin de estudiar el error cometido. Los valores de los factores para $T + 1$ y $T + 2$ permiten predecir, es decir, obtener los posibles valores de las variables analizadas para dos periodos más, fuera del horizonte temporal considerado, tal y como se explica a continuación.

3.2.5 Predicción de las variables originales y validación de resultados

Una vez modelizados los h factores y obtenido un modelo *ARIMA* para cada uno de ellos, se pueden determinar los valores predichos de cada factor, como consecuencia del modelo estimado, y que se expresan mediante un vector de la forma:

$$\hat{F}_k^c = \begin{Bmatrix} \hat{f}_{k1}^c \\ \hat{f}_{k2}^c \\ \vdots \\ \hat{f}_{kT}^c \end{Bmatrix} \quad \text{para } k = 1, 2 \dots h$$

La consideración de estos valores, en lugar de los reales, según la fórmula de reconstitución expresada en la página 115, nos va a permitir obtener los valores estimados de las variables originales desde el momento 1 al momento T . Así tendremos que para la variable X_j de la tabla n el valor predicho toma la siguiente expresión:

$$\hat{X}_j^n = \frac{\sum_{k=1}^h \hat{F}_k^c \frac{g_{kj}^n}{\sqrt{\lambda_k^c}}}{\sqrt{\alpha_n \beta_n}} S_{x_j^n} + \bar{X}_j^n$$

Cada valor de esa variable se expresa como:

$$\hat{x}_{ij}^n = \frac{\sum_{k=1}^h \hat{f}_{kt}^c \frac{g_{kj}^n}{\sqrt{\lambda_k^c}}}{\sqrt{\alpha_n \beta_n}} S_{x_j^n} + \bar{X}_j^n$$

siendo \bar{X}_j^n la media de la variable original y $S_{x_j^n}$ su desviación típica. La comparación de estos valores con los realmente observados, es decir, la comparación de \hat{x}_{ij}^n y x_{ij}^n permite obtener los errores cometidos al emplear los factores predichos y no los originales para la obtención de X_j^n . Si denominamos e_{ij}^n a la diferencia entre esos dos valores, y E_j^n al vector que recoge dichas diferencias para la variable X_j^n , se tiene que:

$$E_j^n = \begin{Bmatrix} e_{1j}^n \\ e_{2j}^n \\ \vdots \\ e_{Tj}^n \end{Bmatrix} = \begin{Bmatrix} x_{1j}^n - \hat{x}_{1j}^n \\ x_{2j}^n - \hat{x}_{2j}^n \\ \vdots \\ x_{Tj}^n - \hat{x}_{Tj}^n \end{Bmatrix} \quad j = 1 \dots p_n ; n = 1 \dots N$$

Puesto que los E_j^n pueden considerarse como una nueva variable aleatoria, para que el modelo de previsión sea adecuado, su comportamiento debería asemejarse a un ruido blanco. El análisis de su media, funciones de autocorrelación y autocorrelación parcial muestrales, así como su normalidad, de la forma indicada para los modelos vistos, nos permitirá llegar o no a esa conclusión.

También interesa analizar si el modelo propuesto es estable o no, o lo que es lo mismo, si la introducción o eliminación de un año-individuo le afecta de manera considerable, modificando los parámetros de los modelos identificados para los factores que se han considerado. Una manera sencilla de valorar dicha estabilidad es la eliminación de filas (es decir, años-

individuos) de la tabla original y la aplicación de nuevo del *Statis*. El estudio de los nuevos factores obtenidos nos permitirá saber si éstos tienen un comportamiento similar a los obtenidos en el primer análisis, y por tanto, determinar si son estables o no. Tanto los errores como la estabilidad serán dos de las cuestiones más ampliamente consideradas en la parte empírica del presente estudio.

El procedimiento finaliza con la predicción de las variables para periodos sucesivos. Como la predicción a largo plazo entraña muchos errores nos limitaremos a formular las expresiones que predicen para uno y dos periodos hacia delante ($T + 1$) y ($T + 2$):

- para un periodo hacia delante:

$$\hat{x}_{(T+1)j}^n = \frac{\sum_{k=1}^h \hat{f}_{k(T+1)}^c \frac{g_{kj}^n}{\sqrt{\lambda_k^c}}}{\sqrt{\alpha_n \beta_n}} S_{x_j^n} + \bar{X}_j^n$$

$$j = 1, 2 \dots p_n ; n = 1, 2 \dots N$$

- para dos periodos hacia delante:

$$\hat{x}_{(T+2)j}^n = \frac{\sum_{k=1}^h \hat{f}_{k(T+2)}^c \frac{g_{kj}^n}{\sqrt{\lambda_k^c}}}{\sqrt{\alpha_n \beta_n}} S_{x_j^n} + \bar{X}_j^n$$

$$j = 1, 2 \dots p_n ; n = 1, 2 \dots N$$

Cuando se disponga de los datos reales, la comparación con los valores previstos permitirá valorar el modelo de nuevo e incorporar dicha información al mismo, reformándolo si es necesario, con el fin de predecir a más largo plazo. Entre las ventajas del procedimiento propuesto destacamos las siguientes:

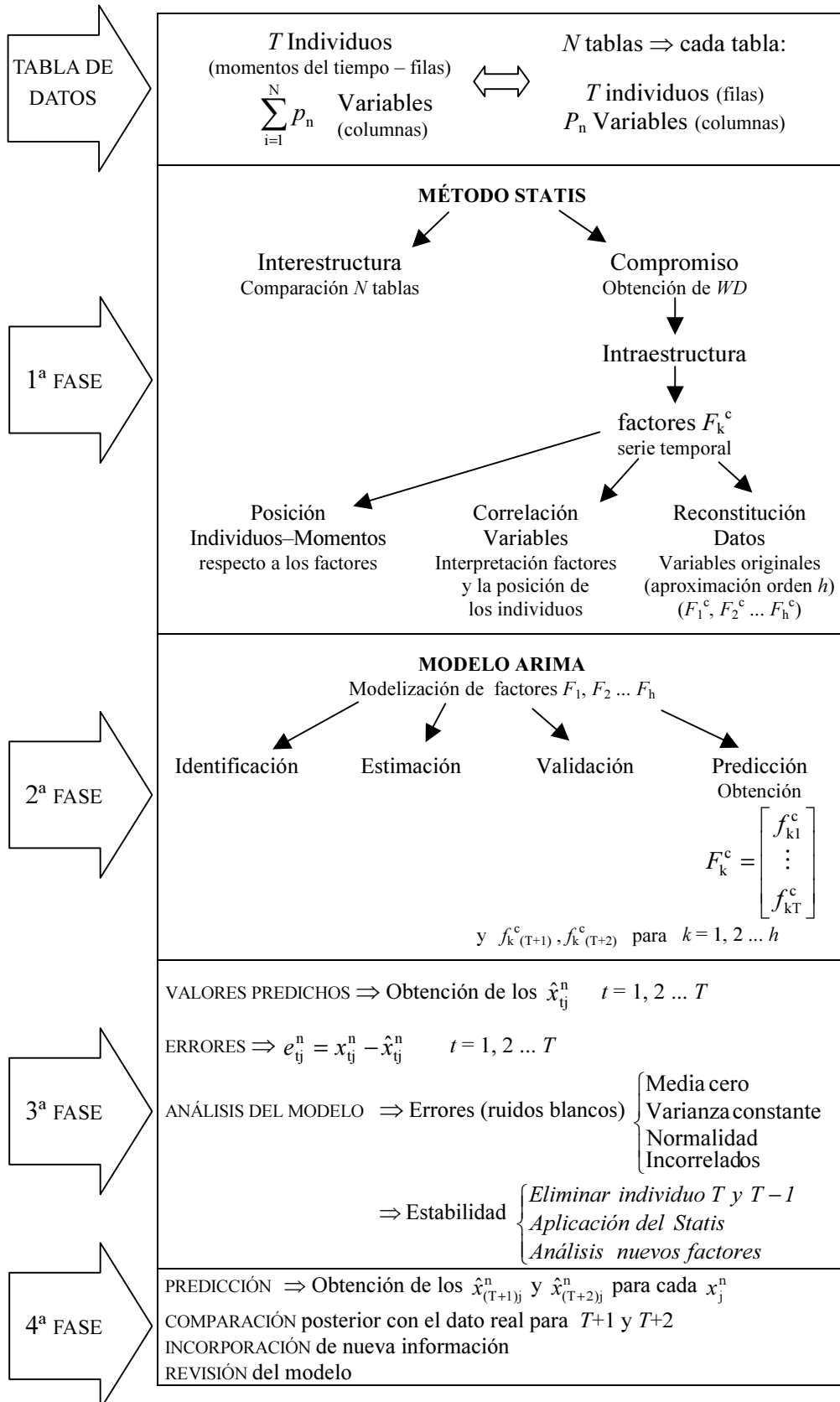
- Es posible predecir las: $\sum_{n=1}^N p_n$

variables estudiadas en las tablas sin necesidad de analizar más que h modelos, los correspondientes a los h factores considerados, con las consiguientes ventajas derivadas de la simplificación del proceso.

- Dado que las h series modelizadas son los factores resultado de la intraestructura del *Statis*, es posible caracterizar, a través de las correlaciones con las variables originales, y dotar de significado dichos factores, que reflejan la estructura o comportamiento común del fenómeno estudiado, durante el periodo considerado, a través de los años-individuos.
- El posicionamiento de los años-individuos sobre dichos factores permite describir los mismos en función de las variables: el cambio de posición, en la dirección de un factor, de dos años-individuos consecutivos refleja un cambio significativo en las variables que definen ese factor, detectando así las fluctuaciones o comportamientos más importantes sin necesidad de observar de manera individual los datos de partida, que podría ser un trabajo arduo dado el elevado número de variables consideradas.

A modo de resumen se expresa en el siguiente cuadro el proceso que relaciona la metodología STATIS con las técnicas de predicción.

Figura II.28
Proceso predictivo del STATIS



Capítulo III
Estudio del sector asegurador español

Nuestro trabajo se completa con este capítulo en el que se analiza un sector, a nuestro juicio, fundamental para cualquier economía, por las funciones que desempeña en el desenvolvimiento de la actividad económica. El objetivo que pretendemos con él tiene una doble vertiente: en primer lugar, comprobar la importancia práctica de la metodología descrita, ya que proporciona resultados más globales y a la vez más precisos que otros métodos estadísticos, y en segundo lugar, analizar un sector de especial relevancia para indicar el grado de desarrollo de cualquier economía.

En cuanto a la aplicación de la metodología STATIS para analizar información de este sector, a nuestro modo de ver, constituye una novedad importante desde el punto de vista empírico. Tradicionalmente, su empleo ha quedado relegado a otros ámbitos del conocimiento distintos de la Economía y las Finanzas, como son la Ecología y la Biología. Hasta ahora, las aplicaciones en el ámbito económico, además de escasas, se han limitado al campo de la Economía Agraria, y sólo en fechas recientes, se ha utilizado para otros estudios a los que ya hemos hecho referencia.

Por lo que se refiere al sector del seguro, éste siempre ha suscitado mucho interés, sobre todo por parte de los poderes públicos y de la legislación, con el fin de establecer medidas suficientes para la protección de los asegurados. Actualmente, la crisis de los sistemas tradicionales de Seguridad Social y la creciente cultura financiera de los usuarios han hecho del seguro un producto de consumo frecuente, lo que ha traído consigo un creciente interés por lo

que a él concierne. Como consecuencia, el estudio de este sector resulta interesante, no sólo para las entidades que desarrollan su actividad, sino también para los ciudadanos y las instituciones públicas.

El sector asegurador puede considerarse como uno de los más claros ejemplos de internacionalización de una actividad. Los procesos de fusión y absorción que se han producido entre las grandes aseguradoras mundiales en los últimos años, han modificado sustancialmente el panorama del sector en todos los mercados locales. Sin embargo, en el momento actual, dado que los grandes grupos europeos y mundiales ya están prácticamente conformados, no parece que se vayan a producir grandes modificaciones en este sentido (González de Frutos, 1999). Por este motivo, cuando se pretende estudiar dicho sector localmente, como en el caso español, conviene tener en cuenta que se trata de entidades sometidas al control de la legislación española, como luego se especificará, pero que pueden operar en mercados españoles o extranjeros.

Ambos objetivos nos inducen a estructurar el capítulo en dos partes, comenzando, obviamente, por la descripción del sector asegurador español, y analizando su atractivo a través del modelo desarrollado por Porter para el análisis de sectores industriales. Con ello se pretende conocer el sector e identificar las características y variables de las entidades aseguradoras que pueden analizarse empíricamente con el fin de obtener conclusiones relevantes.

En la segunda parte se comprueban, de forma empírica, las posibilidades descriptivas y predictivas de la metodología STATIS, descritas en el capítulo anterior, realizando el análisis de las variables relevantes del sector que nos ocupa. Con ello se pretende materializar la validez de los desarrollos teóricos, contribuyendo, de esta forma, al avance en el estudio de la Estadística, que es lo que, en definitiva, se pretende con un trabajo de investigación, el cual complementaremos con una profundización en el conocimiento del sector.

Esta forma de proceder, a nuestro juicio, permite poner de manifiesto las ventajas de emplear la metodología estadística para el tratamiento de datos en el ámbito de la Economía de la Empresa y en el ámbito financiero, así como la necesidad de complementarse que tienen las distintas disciplinas con el fin de crear conocimiento. Las conclusiones obtenidas pueden resultar interesantes tanto desde el punto de vista metodológico y estadístico, como desde el punto de vista económico para cualquier entidad del sector estudiado.

1. Análisis general del sector asegurador

El proceso que conlleva un estudio del sector del seguro incluye, en primer lugar, la descripción de las características generales de la actividad aseguradora, explicando la función económica y financiera que cumple dentro de cualquier economía, y los rasgos característicos de la entidad que tiene por objeto dicha actividad: la empresa aseguradora. Como consecuencia de este procedimiento, seguidamente se limita el sector asegurador español, analizando sus rasgos más característicos, y se completa este análisis general con la exposición teórica del modelo de más amplia difusión para el Análisis Estructural de un sector industrial, el modelo de Porter, que se aplicará al sector asegurador español, como paso previo para la formulación de una estrategia en toda empresa aseguradora.

1.1 Características de la actividad aseguradora

Comenzaremos por considerar y definir una serie de conceptos básicos relativos a la actividad aseguradora. El primero de ellos es el *riesgo*, que en su concepción más general y según UNESPA (1998b), significa la contingencia o proximidad de un daño, con connotaciones de inseguridad y peligro, pudiendo definirse como la posibilidad de que ocurra un daño económicamente desfavorable.

Una de las posibles estrategias frente al riesgo es el *seguro*, que puede definirse como un acuerdo o contrato por el que se establece el traspaso o transferencia, a cambio de un precio (la *prima*), de las consecuencias económicas desfavorables producidas por la realización del riesgo asegurado (el *siniestro*) a otra persona llamada *asegurador* (la entidad de seguros) distinta de la que puede sufrirlas (*asegurado*). El conjunto de documentos que recogen las condiciones y acuerdos que componen el contrato de seguro se denomina *póliza*, que es el documento probatorio de la existencia del contrato de seguro entre las partes, y de las condiciones y acuerdos en que se basa la relación contractual. La persona física o jurídica que firma la póliza, es decir, contacta con la entidad aseguradora, es el *tomador* del seguro.

Las obligaciones de la entidad aseguradora generalmente consisten en la *indemnización* en metálico de los daños causados en el siniestro, previa valoración de los mismos; la reposición de los bienes perdidos o dañados; o la prestación de un servicio (por ejemplo, el servicio de enterramiento, en los seguros de decesos). El coste de dichas obligaciones es la *siniestralidad*. Además, dicha entidad debe compensar una gran masa de riesgos para que el número de siniestros que se produzcan no traiga consigo desviaciones importantes o imprevistas de la siniestralidad, y sea lo más aproximado posible a la probabilidad teórica deducida de las estadísticas conforme a la ley de los grandes números. Así, el asegurado elimina la aleatoriedad por medio del seguro y el asegurador, lo hace por medio del agrupamiento de una gran masa de asegurados.

Los requisitos que tienen que cumplirse para que un riesgo sea objeto de seguro son los siguientes:

- El suceso debe ser incierto o aleatorio, es decir, no puede existir certidumbre de que el acontecimiento se va a producir, o como en el caso del seguro de vida, se conoce con certeza que ocurrirá pero se ignora cuando.

- Debe existir la posibilidad de riesgo, es decir, el suceso debe “poder suceder”.
- El riesgo debe ser concreto, de forma que no puede haber ambigüedad en sus características, que deben ser cuantificables.
- Los riesgos deben ser lícitos, es decir, no pueden ir contra las reglas morales o de orden público.
- Deben ser riesgos fortuitos o accidentales, es decir, el riesgo debe derivar de un acto independiente de la voluntad humana de producirlo.
- Debe ser susceptible de tratamiento estadístico con el fin de que el asegurador pueda establecer las *bases técnicas* adecuadas para la cobertura del seguro¹.

Los diferentes tipos de riesgos asegurables dan lugar a los distintos ramos del sector², y la clasificación de los mismos puede realizarse en virtud de diferentes criterios: según las consecuencias económicas del siniestro, según sus características y según el sistema utilizado para su valoración, entre otras. Sin embargo, con objeto de analizar el sector, creemos conveniente la distinción básica entre seguros de vida y seguros de no vida, atendiendo a la

¹ Las *bases técnicas*, recogidas en UNESPA (1998d), comprenderán, según la estructura y organización comercial de la entidad aseguradora, los siguientes apartados:

- *Información genérica del riesgo asegurable* conforme a la póliza respectiva, los factores de riesgo considerados en la tarifa, y los sistemas de tarificación utilizados.
- *Información estadística sobre el riesgo*, indicando si es homologada por la Administración, o indicando el tamaño de las muestras, las fuentes y el método de obtención de la misma.
- *Recargo de seguridad* sobre la prima de riesgo, en función de los datos estadísticos que se posean, destinado a cubrir las desviaciones desfavorables de la siniestralidad esperada.
- *Recargos de gestión*, necesarios para cubrir los gastos de administración, de adquisición y mantenimiento del negocio.
- *Recargo para beneficio o excedente*, destinado a remunerar los recursos financieros e incrementar la solvencia de la empresa.
- *El sistema de cálculo de las primas*, en función de las bases estadísticas y financieras, si proceden, estableciendo su equivalencia actuarial para fijar la *prima pura* que corresponde al riesgo cubierto.

² La clasificación completa de los ramos del seguro está en la Ley 30/95 de Ordenación y Supervisión de los Seguros Privados, y es la que recoge UNESPA (1998g).

legislación vigente. En el primero prevalece la idea de previsión, ya que trata de subsanar determinadas situaciones que pueden producirse (muerte, invalidez...), o que con seguridad se presentarán en un momento (jubilación), y por tanto, los recursos captados son auténtico ahorro. El asegurador debe colocar estos recursos para conseguir la máxima rentabilidad, y otorgar al asegurado prestaciones equivalentes a las que ofrece el mercado.

En el seguro no vida prevalece también la idea de ahorro–previsión, ya que va destinado a compensar la pérdida eventual que puede producirse en el patrimonio del asegurado. Sin embargo, en este caso, se trata de un ahorro finalista cuya misión es ir constituyendo la cantidad necesaria para reponer la pérdida sufrida, y no hay formación neta de capital, ya que todo el ahorro se invierte en la reposición de dicho capital siniestrado. Dentro de los seguros no vida se encuentran los seguros de automóviles, los seguros de salud (enfermedad y asistencia sanitaria), y el resto de los seguros.

Otra diferenciación a tener en cuenta es la que distingue entre seguros personales y seguros industriales. Así, se denominan seguros personales aquellos relativos a las pólizas vendidas a personas y, por tanto, se encuentran entre ellos los seguros de vida, de accidentes, de automóviles, de asistencia sanitaria y enfermedad, de decesos, de defensa jurídica, de multirriesgo del hogar y de las comunidades. El resto de los ramos se consideran seguros industriales y, por tanto, se refieren a las pólizas vendidas a empresas.

1.1.1 La función económica del seguro

Comenzamos analizando la función básica y principal del seguro en una economía, considerando luego el papel que desempeña dentro del sistema financiero. Según Martínez Rego (1991: 29-47) la función económica del

seguro es la reducción del riesgo, que se puede analizar empleando dos enfoques: el enfoque microeconómico y el enfoque macroeconómico.

ENFOQUE MICROECONÓMICO

Este enfoque hace referencia a las decisiones individuales de los sujetos económicos (economías domésticas y empresas, sobre todo). Es sabido que una de sus características esenciales es la aversión al riesgo, de tal manera que, conforme mejoran las condiciones económicas, estos sujetos son más reacios a cambiar de situación en su patrimonio. Por tanto, sería necesario constituir reservas de precaución para mantener ese nivel económico en caso de que un suceso afectase dicho patrimonio. Dado que el importe óptimo de dichas reservas sería difícil de estimar, y sus rendimientos financieros estarían limitados, los agentes están dispuestos a transferir dicho riesgo a otros agentes especializados, las entidades aseguradoras, a cambio de la prima. Estas entidades son neutras ante el riesgo, ya que pueden combinar los distintos riesgos individuales.

De esta manera, gracias al seguro, el riesgo financiero se estabiliza y se convierte en un componente más del coste total de producción o del presupuesto familiar, y facilita la obtención de fondos a las familias y a las empresas debido a la garantía de mantenimiento de potencial económico que el seguro supone para los prestamistas. Además, ciertos autores consideran al seguro como un factor de incremento de la productividad en la medida en que el pago de indemnizaciones aceleraría la reactivación de la producción después de haberse producido un siniestro.

ENFOQUE MACROECONÓMICO

Desde un punto de vista global, la función principal del seguro es la distribución equitativa, entre la mayoría de los miembros de una sociedad, de las pérdidas económicas sufridas por una minoría. Las consecuencias para el conjunto de la economía son las siguientes:

-
- Se reducen las reservas de precaución acumuladas que habría que destinar a reponer las pérdidas de toda la sociedad, lo que permite que estas reservas de tesorería se liberen quedando disponibles para fines de inversión. De esta manera, se consigue una mejor asignación de los recursos económicos.
 - Se promueve la creación de ahorro en la medida que las provisiones específicas que constituyen las empresas de seguros (las *provisiones técnicas*) van a dar lugar a importantes acumulaciones de capital que se colocan en los mercados financieros. Por ello, el seguro es un poderoso medio de captación y distribución del ahorro.
 - Mejora el riesgo de crédito de las empresas al eliminar, a través de una cobertura adecuada de seguro, una gran variedad de riesgos asociados a los negocios.
 - Se reduce el nivel de riesgo global de la sociedad, como resultado de las medidas llevadas a cabo por los aseguradores mediante la *prevención* y la *selección* de riesgos. Por *prevención* se entiende la adopción de medidas dirigidas a evitar la ocurrencia de un riesgo, y en caso de que ocurra, conseguir que sus consecuencias sean lo menos dañosas posibles. Por *selección* se entiende un conjunto de criterios y actuaciones aplicados por las entidades aseguradoras al objeto de determinar las condiciones de aceptación de riesgos, y orientar con ellas la composición de sus carteras en un sentido determinado. Con esta selección se persigue que las condiciones de los riesgos a asegurar permitan que la siniestralidad real se acerque a la siniestralidad esperada. Las entidades de seguros realizan la selección de riesgos a través de lo que denominan *normas de aceptación y suscripción*, siendo ésta una de las funciones más importantes en la actividad diaria de estas entidades.
 - Facilita la estabilidad social y contribuye al desarrollo del bienestar colectivo, como consecuencia del carácter liberador que la seguridad económica confiere respecto de las tensiones creadas por el riesgo.

- El seguro contribuye a la formación de la Renta Nacional a través del valor añadido que genera. El servicio del asegurador es el de un intermediario y, por lo tanto, su valor añadido puede medirse por la diferencia entre su cifra de negocio, representada por el total de primas recaudadas o el total de siniestros pagados, y el importe de los pagos efectuados a terceros.

1.1.2 El seguro dentro del sistema financiero

Además de la función económica, el seguro cumple una función financiera, teniendo en cuenta que el sector asegurador de cualquier economía está integrado dentro del denominado sistema financiero. Todo sistema financiero está formado por tres elementos fundamentales: los activos o instrumentos, las instituciones o intermediarios, y los mercados financieros. Su función fundamental es canalizar los recursos excedentes de los ahorradores o prestamistas hacia los demandantes o prestatarios de la manera más eficiente posible, es decir "canalizar el ahorro hacia la inversión" (García Delgado, 1993: 277). Si todos los agentes tuviesen información completa y perfecta, la existencia de intermediarios y de mercados organizados no sería necesaria: cada agente podría decidir donde colocar sus fondos, ya que conocería el mejor precio al que comprar y vender un activo financiero. Como los agentes no disponen de toda la información y además la información es costosa, es necesario contar con mercados en los que se intercambien activos financieros y en los que se determine su precio.

Además, no siempre es posible encontrar ahorradores dispuestos a adquirir activos emitidos por los demandantes de financiación, puesto que los deseos de los ahorradores e inversores pueden no ser coincidentes. En este caso, es necesaria la existencia de intermediarios financieros para adquirir los títulos emitidos por los demandantes de fondos, que los transforman en títulos

emitidos por ellos mismos, asumiendo cierto riesgo que intentarían cubrir aplicando un diferencial adecuado de precios.

No todos los intermediarios financieros desarrollan de la misma manera esas funciones, por lo que se distingue entre *intermediarios bancarios y no bancarios*. Los primeros, son los que pueden emitir pasivos que tienen consideración de dinero (los bancos, las cajas y las cooperativas de crédito). Los segundos, incluyen instituciones muy diversas y especializadas, tales como las sociedades de crédito hipotecario, las entidades de financiación (factoring y leasing), las sociedades mediadoras en el mercado de dinero, las instituciones de inversión colectiva y las compañías de seguros.

En la mayoría de los países los intermediarios financieros bancarios cuentan con una posición de fuerza dentro del mercado de crédito, si bien, a medida que se alcanzan mayores cotas de desarrollo económico, dicha posición va perdiendo importancia en términos relativos, como consecuencia de la introducción de formas de financiación directa, que permiten la realización de préstamos fuera de los canales del sistema crediticio, y que, por lo general, vienen unidas al propio proceso de desarrollo económico. En cambio, entre seguro y desarrollo económico existe una influencia mutua, ya que el desarrollo favorece la expansión del seguro y éste, a su vez, juega un papel muy importante en la consecución del desarrollo económico: la cobertura del riesgo y la financiación de actividades económicas, a través de los flujos, captados de las unidades económicas ahorradoras, que se colocan en el mercado financiero (es decir, la inversión de las provisiones técnicas).

La entidad aseguradora es, además, uno de los intermediarios financieros más eficientes del mercado debido a varios motivos (Martínez Rego, 1991: 59-61):

- Toda institución financiera tiene que mantener un pasivo o entrada de inputs continuada. Las empresas de seguros que cuenten con una cartera equilibrada pueden hacer provisiones bastante estables sobre sus entradas de capital, con la característica añadida de que esas

previsiones se van a poder desarrollar con suficiente independencia de las fluctuaciones del mercado.

- Las entidades financieras tienen que enfrentarse, asimismo, a la dificultad que supone conseguir un equilibrio entre sus depósitos, en gran parte a corto plazo, y su estrategia de inversión que tiene un horizonte temporal mucho más amplio. Por su parte, las aseguradoras, y especialmente las que se dedican al ramo de vida, cuentan con unos recursos que preceden de los depósitos de sus clientes a largo plazo y sus rescates están evaluados técnicamente, lo que les permite obtener con una mayor facilidad ese estado de equilibrio.
- Mientras que para las entidades financieras de depósito tradicionales la posibilidad de que se pueda producir una retirada masiva de fondos, como consecuencia de la aparición de determinadas circunstancias tanto endógenas como exógenas, es un problema al que se ven continuamente expuestas y que les afecta muy directamente, en el caso de las aseguradoras esta posibilidad es menor, y por tanto, sus inquietudes por este motivo se ven reducidas.
- Para acometer su actividad (Calvo *et al.*, 1999: 392), el sector asegurador necesita un capital fijo relativamente pequeño, y su capital circulante se lo prestan los propios clientes a cuenta del servicio que ha de empezar a prestar en ese momento (la seguridad), que cobra por adelantado. Por eso, desde el punto de vista financiero, se puede decir que el tomador de la póliza de seguros es el prestamista que proporciona el crédito al asegurador para que preste el servicio, siendo el asegurador un mero recolocador de fondos.
- En cuanto a las inversiones y desde una perspectiva amplia, las empresas de seguros pueden actuar con una mayor libertad que las grandes instituciones financieras ya que éstas, por lo general, están más influenciadas en sus decisiones de inversión, bien por factores de índole política o social, bien por intereses del grupo al que pertenecen.

Una vez especificada la función económica y financiera del seguro se van a considerar de manera detallada las características diferenciadoras, desde el punto de vista empresarial, de las entidades cuya actividad habitual es la asunción de riesgos: las empresas aseguradoras.

1.1.3 La empresa aseguradora

La entidad aseguradora, el asegurador, podría definirse como la persona jurídica que, constituida con arreglo a lo dispuesto por la legislación correspondiente, se dedica a asumir riesgos ajenos, es decir, a proporcionar protección o seguridad frente al riesgo mediante la percepción del precio llamado prima. Las características básicas de las empresas que se dedican al seguro son las siguientes:

— *Su finalidad.* El objetivo de la empresa es realizar contratos de seguros, de forma que agrupe un número suficiente de sujetos u objetos expuestos al riesgo, y poder obtener así una regularidad estadística que le permita evitar desviaciones importantes o imprevistas de la siniestralidad. La entidad aseguradora, con su carácter de permanencia, infunde confianza a los asegurados sobre las garantías de cobertura de los riesgos que aseguran.

— *El producto ofrecido, la seguridad, tiene un carácter intangible y su proceso de producción implica un alto grado de contacto personal.* Por este motivo, tradicionalmente las entidades aseguradoras han sido consideradas dentro del sector de los servicios dirigidos a personas (Hope y Mühlemann, 1997: 7). Los seguros tienen otras características de los servicios, como la heterogeneidad del resultado final obtenido, la simultaneidad entre producción y consumo, y la imposibilidad de almacenamiento.

Por tanto, como en todos los servicios, el diseño y control de su producción son complicados, ya que la presencia intensiva del cliente en la prestación

del servicio dificulta la estandarización del resultado obtenido que, en cada caso, será distinto según las circunstancias. En la mayoría de los ramos del seguro, esta heterogeneidad dificulta al cliente la comparación de los precios entre las diferentes entidades. De ahí que, debido a la presencia del cliente en el proceso, la producción de un seguro requiere un esfuerzo por parte de la entidad para conocer y analizar sus características específicas.

Como consecuencia de la simultaneidad, las entidades de seguros deben prestar mucha atención al medio en el que se desarrolla la contratación del seguro: aspectos decorativos de las oficinas y de tratamiento de los clientes anteriores pueden influir de manera decisiva en la contratación. El personal, que en el caso de los bienes, es el vendedor, en el caso de los servicios es el que produce o presta dicho servicio, por lo que la formación del mismo, para poder atender satisfactoriamente a todo tipo de clientes, es una cuestión fundamental. La simultaneidad de producción y consumo impide que los clientes puedan probar antes de adquirir el servicio. Este hecho tiene una importancia fundamental en el caso de los seguros, en cuanto que el cliente puede examinar las cláusulas y coberturas de su contrato de seguro, pero no tiene certeza de que producido el siniestro, la atención y trámites se lleven a cabo de manera eficaz.

Este aspecto adquiere connotaciones diferentes según los distintos ramos existentes. Por ejemplo, para los seguros de vida, a pesar de su complejidad, a medida que aumenta la cultura del cliente y el nivel de renta, los usuarios demandan más asesoramiento y más atención. Para el caso del automóvil, la elección se basa más en el precio que otros casos, ya que las pólizas de seguros del automóvil no son percibidas como un producto que reclama información detallada para su comprensión y la capacidad de diferenciarlos es, por tanto, pequeña. En el caso de los seguros de salud, la atención al paciente es fundamental para la diferenciación, y la competencia en precio es menor, al ser un servicio delicado relacionado con la salud.

Debido a la intangibilidad, el elemento subjetivo en la apreciación de lo que se compra es muy importante. Por ello, la imagen y prestigio de la marca juega un papel decisivo. Además, dicha intangibilidad impide proteger legalmente el producto ofrecido, y en el caso de los seguros, facilita la imitación entre entidades en cuanto a las coberturas y condiciones de las pólizas ofertadas al cliente. Por otra parte, dado que el prestigio se va a transmitir a través del personal que presta el servicio, la motivación de la fuerza humana es una cuestión de suma importancia en este sector (Ballarín, 1995: 253-5).

La imposibilidad de almacenamiento dificulta la adecuación de la oferta y la demanda, con lo que se plantean problemas a la hora de gestionar los recursos. En el caso de los seguros, la distribución tradicional a través de una red de agentes permite resolver estos problemas y adaptar los recursos humanos a las necesidades del momento.

Dado que la producción del seguro requiere el contacto personal con el cliente, el seguro debe estar próximo al mismo, lo que ha provocado el desarrollo de canales de distribución específicos. Tradicionalmente, han sido los mediadores, los agentes de seguros y los corredores de seguros, los mayoritariamente empleados. Los primeros son aquellas personas físicas o jurídicas que, mediante la celebración de un contrato de agencia con una entidad aseguradora, se comprometen frente a ésta a realizar las actividades de mediación (es decir, promoción, oferta y suscripción de las pólizas de seguros). Los corredores de seguros son las personas físicas o jurídicas que realizan la actividad mercantil de mediación en seguros privados, sin mantener vínculos que supongan afección con entidades aseguradoras o pérdida de independencia respecto a éstas, ofreciendo asesoramiento profesional imparcial a quienes demandan la cobertura de los riesgos.

Las oficinas de la entidad aseguradora, o la venta en oficinas propias, ha sido el canal de distribución que le ha seguido en importancia. También los bancos y cajas distribuyen a través de sus oficinas pólizas de seguros de

los ramos más afines a lo que es su actividad tradicional, fenómeno que se denomina bancaseguros, y que tiene una especial importancia en el sector. En los últimos años, para algunos ramos del seguro orientados a clientes-tipo y que necesitan poco asesoramiento, se han desarrollado nuevos canales de distribución que acercan el seguro al cliente a menor coste, reduciendo la presencia personal y contribuyendo a la estandarización del resultado obtenido. Éstos son los derivados del Marketing directo, denominación que recibe cualquier sistema directo de distribución que no implique contacto personal: mailing, teléfono, correo, Internet, etc.

— *La alteración o inversión del proceso productivo.* Una característica peculiar de la empresa de seguros en relación con cualquier otra empresa, es el hecho de que el ciclo de producción del seguro es inverso: el pago (la prima) se produce antes de la prestación del servicio (la indemnización por la ocurrencia del siniestro), es decir, en este caso los ingresos preceden a los gastos, determinándose el precio del servicio ofrecido sobre la base de métodos estadísticos antes de conocerse su verdadero coste (la siniestralidad). Por tanto, (Calvo *et al.*, 1999: 392) el tiempo juega a favor del asegurador, ya que ese coste se pospone y da lugar, entretanto, a un cúmulo de ahorro que forman las *provisiones técnicas*.

— *La administración de un patrimonio condicional,* que se refleja en esas *provisiones o reservas técnicas*, que obligatoriamente han de constituir las empresas para atender al pago de indemnizaciones cuando se produzca el siniestro, y garantizar así los compromisos adquiridos con los asegurados. Por tanto, las entidades de seguros además de las reservas propiamente dichas, que son comunes a cualquier tipo de entidades (las que forman parte de sus recursos propios y las constituidas para hacer frente a deudas y compromisos con terceros), tienen la obligación de calcular y contabilizar las siguientes *provisiones técnicas*³:

³ La denominación de estas provisiones es la que propone el nuevo Reglamento de Ordenación del Seguro Privado (ROSP), recogido en el Real Decreto 2486/1998 de 20 de noviembre.

-
- *La provisión para seguros de vida, antigua reserva matemática*, exclusiva del ramo de vida, que está destinada a conseguir un equilibrio futuro entre primas y riesgos. En la medida en que en los seguros de vida se cobran primas y cubren ciertos riesgos al asegurado durante muchos años, es necesario que la entidad aseguradora retenga las primas cobradas y las invierta hasta que llegue el momento de realizar ese pago futuro.
 - *La provisión para primas no consumidas* se constituye por las primas que el asegurador debe conservar al fin del ejercicio. Las primas cobradas por el asegurador suelen abarcar el período correspondiente a una anualidad. Si esta anualidad coincidiese exactamente con la fecha de inicio y fin del ejercicio económico de la entidad (1 de enero a 31 de diciembre), esta provisión no sería necesaria. Pero al existir pólizas de muy diversos vencimientos y haberse cobrado anticipadamente la prima de toda una anualidad, la entidad aseguradora tiene en su poder estas porciones de prima que, a 31 de diciembre debe trasladar a la cuenta de resultados del siguiente año.
 - *La provisión para primas pendientes de cobro* tiene por objeto conseguir que el asegurador pueda hacer frente al riesgo de impago de las primas emitidas que al cierre del ejercicio están pendientes de cobro. Conviene precisar que la nueva regulación elimina el carácter técnico de esta provisión.
 - *La provisión para prestaciones*, que sustituye a dos de las anteriores:
 - *La provisión para siniestros pendientes de declaración* surge de la obligación contable de las entidades aseguradoras de asignar éstos al año de su ocurrencia. Normalmente, la mayoría de los siniestros que gestionan los aseguradores han ocurrido unos días antes, por lo que el siniestro es abierto en el mismo año de ocurrencia. En este caso no se plantearía ningún problema, puesto que el siniestro es asignado al año en curso; el problema surge cuando la entidad aseguradora cierra los libros contables a final de año y éstos

deben contener los siniestros del año, siendo probable que exista un siniestro cuya comunicación está en camino y que no se podrá conocer en el momento de realizar el cierre contable. Por esta razón, aunque no se conozcan la cuantía de los siniestros ocurridos y no comunicados, es necesario crear una provisión que reserve una determinada cantidad para todos estos siniestros.

- *La provisión para siniestros pendientes de liquidación y pago* son las cantidades que han de conservarse para atender al pago de los siniestros pendientes de liquidación o de pago en el momento de finalizar el ejercicio. Por tanto, la administración de estas provisiones exige que las aseguradoras las inviertan con el fin de garantizar el cumplimiento de los principios fundamentales de congruencia, seguridad, liquidez y rentabilidad, y así, puedan atender adecuadamente la finalidad para la que fueron creadas.

La decisión de inversión constituye una fuente adicional de ingresos y determina una parte de la estrategia de las empresas aseguradoras, dada la importancia que tiene la gestión del patrimonio constituido por las provisiones.

- *La estabilidad.* Este concepto está íntimamente relacionado con la permanencia y es un requisito imprescindible para que pueda desarrollar su actividad y hacer frente a sus obligaciones futuras. Para lograr esta estabilidad las aseguradoras, además de la constitución de las *provisiones técnicas* que le permiten periodificar las principales fuentes de ingresos y gastos, cuentan con dos importantes mecanismos:
 - La constitución de la *provisión de estabilización*, integrada por los importes del recargo de seguridad incluido en las primas. Estas reservas permitirán a la empresa poder compensar el exceso de siniestralidad que se pueda producir.
 - El *reaseguro*, mecanismo que permite transferir o ceder (reaseguro cedido) parte de los riesgos asumidos directamente por la empresa a

otras compañías mediante unos contratos (contratos de reaseguro). A través de este instrumento técnico, la aseguradora puede conseguir la compensación estadística que necesita, igualando u homogeneizando los riesgos que componen su cartera de riesgos asegurados. Este instrumento permite, asimismo, que la empresa pueda obtener participaciones en el conjunto de riesgos homogéneos de otra entidad, y de esta manera, conseguir ampliar el número de riesgos iguales con los que trabaja (reaseguro aceptado).

- Dos conceptos relacionados con la permanencia y estabilidad de las empresas aseguradoras son *el margen de solvencia* y *el fondo de garantía*. En cada ejercicio económico, las entidades aseguradoras deberán disponer, como *margen de solvencia*, de un *patrimonio propio no comprometido* (deducidos los elementos inmateriales), de cuantía no inferior a unas determinadas cantidades que se fijan en función del tipo de actividad que desarrollan (seguros de vida y seguros no vida) y del volumen de su negocio. En cuanto al *fondo de garantía*, está constituido por la tercera parte de la cuantía mínima del margen de solvencia y no podrá ser inferior a unos determinados importes, fijados en cifras absolutas, también en función de los ramos del seguro en los que opere la entidad aseguradora.

— *La forma jurídica* que, en España como en la mayoría de los países, fija la legislación. Concretamente en la ley 30/95 de 8 de noviembre se establece que la actividad aseguradora únicamente podrá ser realizada por entidades que adopten la forma de Sociedad Anónima, Mutua, Delegación de una entidad aseguradora extranjera con establecimiento permanente en España, Sociedad Cooperativa de seguros y Mutualidad de Previsión Social.

En esta ley, adaptándose a la normativa comunitaria, se considera una nueva forma de ejercer la actividad aseguradora para las entidades aseguradoras no españolas que es la libre prestación de servicios, a través de la autorización administrativa única en el ámbito de los seguros de vida,

mientras que, para los seguros distintos del de vida, ya había sido reconocida por la ley 21/1990. La autorización administrativa única supone que las entidades aseguradoras españolas podrán operar en todo el ámbito del Espacio Económico Europeo, en régimen de derecho de establecimiento o en régimen de libre prestación de servicios, sometidas exclusivamente al control financiero de las autoridades españolas. En sentido contrario, se aplica a las entidades aseguradoras domiciliadas en cualquier estado miembro del Espacio Económico Europeo que quieran ejercer su actividad en el mercado español.

En el caso de las Sociedades Anónimas y las delegaciones extranjeras, el asegurado, a cambio de la prima, recibe únicamente los beneficios de la cobertura de los riesgos. Si se trata de una mutua, los asegurados son también propietarios de la entidad, por lo que obtienen una participación en los beneficios de la actividad aseguradora, generalmente vía reducción de primas en los ejercicios siguientes. Las entidades que operan en España en régimen de libre prestación de servicios ejercen su actividad de manera similar a las delegaciones extranjeras, pero no están sometidas al control de la Dirección General de Seguros. En cuanto a las mutualidades de previsión social, tienen objeto social exclusivamente asegurador, si bien, atendiendo a su especial naturaleza, pueden otorgar prestaciones sociales con otra autorización administrativa específica concedida a tal efecto.

Otra entidad que opera en el sector asegurador es el Consorcio de Compensación de Seguros, entidad de Derecho Público con personalidad jurídica propia adscrita a la Dirección General de Seguros, que tiene como función la cobertura, en régimen de compensación, de los riesgos extraordinarios sobre las personas, los bienes y otros riesgos especiales, que no están bajo garantía por ser motivados por causas anormales o de naturaleza extraordinaria excluidas por los aseguradores ordinarios.

Una vez considerados los rasgos peculiares de la actividad aseguradora, a través de la empresa aseguradora, el paso siguiente consiste en caracterizar el sector asegurador, concretando para el caso español.

1.2 El sector asegurador español

Por sector asegurador español (González de Frutos, 1999) se entiende el conjunto de entidades aseguradoras que, con independencia de donde ejerzan su actividad habitual, tienen domicilio social en territorio español y están sometidas al control de la Dirección General de Seguros, órgano dependiente del Ministerio de Economía y Hacienda, encargado de la supervisión de la actividad aseguradora.

La información relativa al sector es recogida en publicaciones, de organismos o entidades como la Fundación Mapfre Estudios, el Centro de Estudios del Seguro (CES), la Unión Española de Entidades Aseguradoras y Reaseguradoras (UNESPA), la Investigación Corporativa entre Entidades Aseguradoras (ICEA), el Instituto de Estudios Superiores, Financieros y de Seguros (INESE), el Instituto de Actuarios de España (IAE), las Tecnologías de la Información y Redes para las Entidades Aseguradoras (TIREA) o la Dirección General de Seguros (DGS). Además de las revistas especializadas en el sector, como *Actualidad Aseguradora*, *Actualidad Aseguradora Internacional*, o *Aseguradores*, otras revistas de carácter financiero pueden ser consultadas para obtener información sobre seguros, como la revista *Ranking*, *Mercado de Riesgos* o *Ausbanc*, y también otras publicaciones económicas más generales, como la *Revista de Información Comercial Española (ICE)* o los diarios económicos *Expansión* y *Cinco Días*.

En el caso que nos ocupa, es decir, el español, este sector ha experimentado en las últimas dos décadas una transformación importante, tanto cuantitativa como cualitativa, adaptándose a un mercado cada vez

más abierto e internacionalizado, manteniendo el adecuado equilibrio entre la competitividad y la solvencia. Comenzamos por la legislación reguladora del sector, que es la responsable en gran medida de dicha transformación, y que ha motivado cambios en el número de compañías, en el ritmo de crecimiento del sector, y en los resultados obtenidos en los distintos ramos. A continuación, nos referiremos a los indicadores más relevantes que lo caracterizan, y basándonos en ellos, se procede a especificar las perspectivas que tiene, el sector desde un punto de vista global.

1.2.1 Legislación

Al revisar la evolución legislativa del sector se observa como éste ha permanecido durante muchos años con una estructura cerrada en el mercado interior español, regulado bajo principios intervencionistas y sin un desarrollo de libre mercado. Esta regulación legislativa tiene sus antecedentes en las ordenanzas de Bilbao y en la regulación del primer Código de Comercio, sin que exista una reglamentación específica hasta la Ley de 1908 y su Reglamento. En 1953 se publica una nueva ley que estará vigente hasta la ley 33/84, norma que contempla los principios que regían la legislación de seguros en la Comunidad Europea. Por ello, al producirse la adhesión de España, sólo fueron necesarias algunas modificaciones legales y otras reglamentarias para conseguir la adaptación requerida, sin necesidad de promulgar ninguna otra ley. Con dicha ley se inicia el desarrollo de libre mercado en el sector.

Para la adaptación de las llamadas segundas directivas comunitarias, fue necesaria la aprobación de la ley 20/91. Esta adaptación no fue completa y es necesario esperar a la ley 30/95 de Ordenación y Supervisión de Seguros Privados (LOSSP), que adopta los principios del Derecho comunitario aunque no los traspone en su totalidad. Esta ley 30/95, que es la vigente

actualmente, prevé los requisitos que deben cumplir las entidades que quieran ejercer la actividad aseguradora y las condiciones necesarias para el ejercicio de su actividad. Su desarrollo reglamentario no se ha producido hasta el Reglamento de 20 de noviembre de 1998, que afecta sobre todo a las provisiones técnicas, en un intento de construir sistemas de gestión muy profesionalizados, intensificando la preocupación por la solvencia, y forzando el desarrollo de sistemas de autocontrol e información muy extensos, con el objeto de asegurar la continuidad de la entidad aseguradora y proteger al asegurado.

Aunque en condiciones de libre mercado, la actividad aseguradora está tutelada a través de la administración, tutela que ejerce a través de una serie de organismos, siendo el principal la ya mencionada Dirección General de Seguros (DGS), que tiene como principales funciones la autorización administrativa para el ejercicio de la actividad, la aprobación de productos aseguradores, la autorización de las entidades para actuar en los distintos ramos y el control económico-financiero de todas las entidades aseguradoras. Dependiendo de ella, la Comisión Liquidadora de Entidades Aseguradoras (CLEA), tiene como fin administrar y proceder a la liquidación de entidades aseguradoras en proceso de quiebra o suspensión de pagos en todos aquellos casos en que el activo de la empresa sea inferior al pasivo, no se lleve la contabilidad obligatoria, o se lleve de tal manera que sea imposible conocer cuál es la situación real de la entidad. Su papel se puede extender a liquidar también parcelas concretas de la actividad de las entidades que tengan problemas de rentabilidad, pero en ningún caso se procede a la intervención, como en el caso de las entidades bancarias.

También están reguladas legalmente las condiciones de acceso y las garantías de las entidades aseguradoras, en cuanto a la exclusividad del objeto social (operaciones de seguros y otras directamente relacionadas con ellas) y a los capitales mínimos necesarios para su ejercicio. Los requisitos de capital son diferentes para los distintos ramos: 1500 millones para seguros de vida, y seguros de crédito y caución; 350 millones para los

seguros de asistencia sanitaria, accidentes, enfermedad y decesos; y 500 millones para el resto de los ramos. Las garantías financieras mínimas también son diferentes para cada uno de los ramos. Además, la ley 30/95 establece el principio de especialización para las entidades de nueva creación, por el cual las entidades que quieran trabajar el ramo de vida sólo podrán trabajar en ese ramo, o en su caso, podrán también ser autorizadas para ejercer su actividad conjuntamente en los ramos de accidentes y enfermedad. Las entidades autorizadas antes de 1984 pueden seguir ejerciendo en todos los ramos para los que fueron autorizadas, pudiendo trabajar tanto para los seguros de vida como para los seguros distintos del de vida. De todos modos, el reglamento de 25 de noviembre de 1998 establece que se podrán suscribir riesgos accesorios de otro principal, bajo un mismo contrato, siempre que no se requieran mayores garantías financieras y con una serie de condiciones.

También es necesario considerar como regulación específica del sector el Plan Contable de Entidades Aseguradoras (PCEA) y las normas para la formulación de las cuentas de los grupos de entidades aseguradoras, recogidas en el Real Decreto 2014/1997, de 26 de diciembre.

Por tanto, la legislación y los organismos públicos condicionan, el ejercicio de la actividad y el abandono de la misma. Toda la legislación debe considerar que un excesivo control. Sobre las entidades de Seguros, en comparación con el resto de países europeos, puede beneficiar a las entidades europeas que ejercen su actividad en España, en régimen de libre prestación de servicios, y que no están sometidas a tan férreo seguimiento.

1.2.2 Indicadores económicos relevantes del sector

La racionalización y liberalización del sector, sobre todo con la ley 33/84, ha traído consigo una drástica reducción del número de competidores: el

número de compañías registradas en la Dirección General de Seguros descendió de las 700, en 1975; a las 475, en 1985; y a las 378, en 1998; lo que supone una reducción de aproximadamente el 46% en 20 años. La ley 30/95 de Ordenación y Supervisión del Seguro Privado (LOSSP) ha impulsado este proceso como consecuencia de las dificultades de algunas entidades para alcanzar los capitales mínimos que en ella se establecen y que dio lugar a muchos procesos de fusiones y adquisiciones.

Aún así, el número de entidades es elevado y el proceso debe continuar. Tal y como indica la Dirección General de Seguros (González de Frutos, 1999), el sector español del seguro tiene un problema estructural dado que hay muchos más operadores de los necesarios y por ello, es un mercado vulnerable. Para adaptar el sector a la realidad europea sería recomendable una drástica reducción del número de compañías.

Una vez analizado el número de entidades, es necesario considerar la progresión de la tasa de crecimiento, las cifras relativas a los costes y a los resultados, así como los canales de distribución más comunes, con el fin de caracterizar adecuadamente el sector.

a. Ritmo de crecimiento del sector

Para analizar la evolución general del sector asegurador y su ritmo de crecimiento es importante fijarse en los siguientes indicadores:

- El porcentaje de crecimiento del volumen de primas de unos años a otros.
- El porcentaje que las primas representan respecto al PIB, es decir, la penetración del seguro en la economía, que indica el peso económico del sector.
- El gasto per cápita en seguros, expresado como volumen de primas por habitante, denominado densidad de aseguramiento. Éste es un

buen indicador del nivel de desarrollo de una economía ya que, a mayor desarrollo, mayor densidad de aseguramiento.

- El porcentaje que el capital extranjero representa sobre el total del sector, es decir, la penetración extranjera en el sector.

La liberalización del sector asegurador español, junto con su reestructuración y la reducción del número de compañías, trajo consigo el crecimiento muy rápido del sector: en el período 78-90 se consiguieron tasas de crecimiento de un 12,4 por ciento anual por término medio. En la siguiente tabla se recogen los valores de los indicadores del sector, antes mencionados, para el periodo 1991-1998:

Tabla III.1
Indicadores del crecimiento y evolución del sector asegurador⁴

Años	% crecimiento	% primas/PIB	% gasto per capita	% capital extranjero
1991	23,7	3,83	53.859	—
1992	16	4,14	62.469	—
1993	9,1	4,38	68.085	—
1994	26,4	5,2	85.590	—
1995	1	4,87	86418	37,4
1996	12,7	5,1	97.383	37
1997	9,3	5,38	106.341	32,82
1998	7	5,41	—	29

Hasta 1995 el ritmo de crecimiento del sector había sido muy elevado (18,6% por término medio anualmente), lo que había provocado un aumento importante de la penetración del seguro en la economía y en la densidad de aseguramiento. En 1995, se observa un retroceso en el ritmo de crecimiento del sector, pero se puede decir que el año fue favorable en

⁴ Las fuentes de datos empleadas son las publicaciones Fundación Mapfre Estudios (1995), *El mercado español de seguros en 1995*; Fundación Mapfre Estudios (1996), *El mercado español de seguros en 1996*; Mapfre Estudios (1997), *El mercado español de seguros en 1997*. Para el año 1998 se han empleado los datos recogidos en UNESPA (1999b).

términos generales, ya que la caída del volumen de contratación es atribuible básicamente a que en, el año 1994, se realizaron importantes operaciones de vida a prima única, derivadas de la externalización de los fondos de las empresas, que produjeron un aumento coyuntural de las primas. Para los años 96, 97 y 98, se observan tasas de crecimiento menores (7,4% por término medio), aunque aún superiores a la del PIB, lo que provoca el incremento de la tasa de penetración y pone de manifiesto un progresivo incremento de la importancia del sector en la economía. Además, como se puede observar en la tabla, a falta del dato relativo a 1998, el crecimiento de la densidad de aseguramiento, aunque a ritmo menor que hasta el año 95, es indicativo de la evolución favorable del sector y del desarrollo de la economía, si bien, los déficits estructurales de aseguramiento en el caso español son evidentes, y los niveles alcanzados son bajos si se comparan con los de países del entorno, sobre todo en el caso de países europeos como Alemania, Dinamarca y el Reino Unido (UNESPA, 1999b). Por tanto, el mercado español de seguros aún está lejos de la saturación y es considerado como un mercado ideal para las compañías extranjeras (Barquero y Huertas, 1997: 423).

El capital extranjero representa la tercera parte del capital invertido en el sector, cifra que puede considerarse elevada pero que no resulta sorprendente, si tiene en cuenta el hecho ya mencionado de que la actividad aseguradora está muy internacionalizada, aunque se observa una tendencia decreciente en dicha tasa. A tenor de lo expuesto, dicho descenso no parece indicar la pérdida de interés de las entidades extranjeras por el mercado español, sino un cambio en la forma de acceso al mismo, fundamentalmente, las alianzas estratégicas o la libre prestación de servicios.

b. Ramos principales de seguros

Aunque la información de las diversas publicaciones aparece detallada para todos los ramos del seguro a los que hace referencia la ley, la caracterización

del sector puede llevarse a cabo satisfactoriamente sin necesidad de tanto desglose, por lo que sólo van a considerarse los más importantes por volumen de primas, que serán los tratados en la segunda parte del capítulo.

Por ello, la tabla III.2 sólo recoge la importancia relativa de esos ramos en el sector (% que representan sobre el total), y su comportamiento (% de crecimiento) desde el año 1991:

Tabla III.2
Importancia relativa y evolución de los grandes ramos⁵

Años	Vida (%)		No Vida				
			Autos (%)		Salud (%)		Total No Vida (%)
	%	Cto	%	Cto	%	Cto	Cto
1991	34,89	40,67	33,19	4,5	8,5	14	17,3
1992	32,13	23,6	34,95	5,3	8,7	19,2	12,5
1993	25,71	7,8	39,09	16,4	9,1	12,8	9,7
1994	34,75	64,7	33,32	4,7	7,8	9	7
1995	40,89	-4,6	29,27	1,7	8,8	11,2	5,2
1996	38,71	21,1	30,67	3,1	8,4	10,2	7
1997	43,55	17,8	27,68	-1,8	8,7	10,9	2,7

Vamos a considerar de manera particular cada uno de esos ramos agregados: seguros de vida y seguros no vida. Respecto a los seguros de vida, entre 1974 y 1990, el mercado se cuadruplicó y en la década de los 90, según se observa en la tabla, aún sigue ofreciendo un gran potencial de crecimiento (tasa media anual del 22,7%, entre 1990 y 1997). La importancia relativa de los seguros de vida dentro del sector, en general, es creciente puesto que su tasa de crecimiento es bastante superior al del resto de los ramos. A pesar de ello, el comportamiento es irregular de unos años a otros. La razón no hay que buscarla sólo en la evolución de la economía española,

⁵ Las fuentes de datos empleadas son las publicaciones Fundación Mapfre Estudios (1995), *El mercado español de seguros en 1995*; Fundación Mapfre Estudios (1996), *El mercado español de seguros en 1996*; Mapfre Estudios (1997). El porcentaje (%) representa la importancia de las primas en el ramos respecto al total del sector. La tasa de crecimiento (Cto), también en porcentaje, está calculada sobre el año anterior. Sólo hay datos disponibles hasta el año 1997.

que ha sido favorable, sino que hay que tener en cuenta la influencia que en este tipo de seguros tiene la fiscalidad. El entorno impositivo español ha sido muy cambiante en los últimos años y esto se refleja en ese crecimiento irregular. Más en concreto, a pesar de la falta de datos relativos al año 1998, las estimaciones parecen indicar que las cifras reflejan la inestabilidad fiscal que es, quizás, el efecto más temido por los aseguradores de vida.

Dentro de los seguros no vida destacan los seguros de salud (asistencia sanitaria y enfermedad) y los seguros de automóviles, por ser los de mayor importancia relativa. Estos ramos no vida están, en general, muy fuertemente vinculados a la actividad económica: a los procesos de consumo, algunos, y a los de formación de capital, otros. Al observar la tabla, se puede comprobar que el crecimiento ha sido inferior que en seguros de vida, ya que la tasa media anual ha sido del 8,67%, entre 1990 y 1997. El mayor componente de éstos sigue siendo el seguro de automóviles, a pesar de que su importancia ha ido descendiendo. La competencia en el ramo es muy fuerte debido, entre otras razones, al desarrollo de los canales de distribución directos, que ha traído consigo una reducción del precio de las primas, hecho que se manifiesta en la reducción del crecimiento, siendo negativo para el último año (tasa media anual del 4,71%, entre 1990 y 1997). Las estimaciones para el año 1998 parecen mostrar un suave crecimiento en la facturación, y ese dato puede ser equívoco ya que puede deberse al crecimiento del parque de vehículos o a la renovación del mismo. La caída de los precios y la competencia en este ramo sigue siendo fuerte y los márgenes muy estrechos.

Los seguros de salud, formados en buena parte por las pólizas de Asistencia Sanitaria⁶, presentan un fuerte crecimiento de las primas (12,43% de tasa media anual para el mismo periodo), lo cual es un claro reflejo del aumento de la renta de la población y de la insatisfacción que existe con el sistema público de atención sanitaria. Por esta razón, este tipo de seguro

⁶ En otros países europeos, las pólizas de enfermedad son más numerosas que las de asistencia sanitaria y así se considera en A.M. Cabezas Ares (1996: 3).

está sometido a una fuerte competencia desde la vertiente pública: la reducción de las listas de espera y las mejoras introducidas han impulsado a los aseguradores a implantar políticas constantes de ampliación y mejora de servicios. Para interpretar correctamente ese crecimiento de las primas hay que tener en cuenta también el aumento del precio, debido al incremento de las coberturas, que en cierta medida transforman el producto.

Del resto de ramos no vida, es el de los Multirriesgos del Hogar, el que sigue en importancia al de automóviles y al de salud, presentando gran potencial de crecimiento según todas las fuentes. Merece destacar también la evolución favorable del ramo de responsabilidad civil, debida al crecimiento de la prima más que al crecimiento de las pólizas, como consecuencia del reclamo por parte de los asegurados de coberturas nuevas, no obstante, sólo representa aproximadamente un 3% de los seguros no vida.

c. Costes y resultados

En el caso de las empresas aseguradoras los gastos del ejercicio de la actividad, o *gastos de gestión*, se han clasificado tradicionalmente en *gastos internos* y *gastos externos*. Los primeros son los gastos del ejercicio destinados a la administración general de la entidad, e incluyen el importe bruto de las remuneraciones anuales, la seguridad social, y las dotaciones a la amortización e impuestos no repercutibles, entre otros. Los externos son los gastos del ejercicio destinados a la producción o mejora de la imagen de la entidad, incluyendo comisiones, gastos de agencia, de publicidad, de propaganda y de relaciones públicas. No hay costes de mantenimiento, ni de productos terminados, dadas las características del sector.

La evolución, desde 1991, de estos gastos (en porcentaje sobre el volumen de primas) para los ramos más importantes del sector se recoge en la siguiente tabla:

Tabla III.3
Porcentaje de gastos internos y externos sobre primas⁷

Años	Vida		No Vida					
			Autos		Salud		Total No Vida	
	Gastinter	Gastexter	Gastinter	Gastexter	Gastinter	Gastexter	Gastinter	Gastexter
1991	8,24	5,36	16,36	12,96	13,11	3,67	17,12	14,85
1992	7,84	4,58	16,51	12,95	12,41	3,45	16,93	15,09
1993	7,78	5,62	16,57	13,08	13,67	3,33	17,35	15,43
1994	5,09	3,89	16,96	12,51	11,63	3,6	17,09	14,56
1995	6,06	4,26	16,79	11,61	12,01	3,54	17,10	14,69
1996	5,34	4,20	16,77	12,5	11,61	3,41	16,37	14,62
1997	5,31	3,85	17,12	12,62	12,53	3,38	16,61	14,88

Como puede observarse en el ramo de seguros de vida, los gastos externos e internos representan una cantidad inferior, sobre las primas, que para los ramos de no vida. Este hecho permite concluir que la productividad es mayor en las entidades dedicadas a los seguros de vida que para las que se dedican a los ramos no vida y considerando todo el periodo, se puede deducir que dicho gastos han ido disminuyendo siendo el descenso más significativo para las entidades de vida.

Dentro de los seguros no vida, es el ramo del automóvil en el que ambos grupos de gastos son importantes: el esfuerzo publicitario de las entidades dedicado a este ramo del seguro es grande y por ello el porcentaje que los gastos externos representan sobre las primas es bastante elevado. Además, los gastos generales de funcionamiento de la entidad también son importantes, y como resultado, los márgenes en este ramo son muy estrechos y la competencia en precios afecta a las entidades especialmente.

En el caso de los seguros de salud son los gastos de gestión interna los que deben someterse a un mayor control dado que representan un porcentaje bastante más elevado sobre las primas que los externos.

⁷ La fuente de datos empleada es la publicación *Información Estadística del Seguro Privado* (UNESPA, 1999a).

Como consecuencia, parece que el sector asegurador español no encuentra en estos últimos ejercicios una vía clara para conseguir una mayor eficiencia. Por tanto, sería necesario para acercarse a los competidores europeos, la implantación de estructuras mucho más eficientes que las existentes hoy por hoy en la mayoría de las entidades (Dirección General de Seguros, 1998: 107).

Además, es necesario considerar, como coste de los seguros lo que se denomina *siniestralidad*, es decir, en el caso de seguros no vida, el porcentaje que la *siniestralidad contable* (suma de los siniestros pagados durante el ejercicio más el incremento de las provisiones técnicas para prestaciones o siniestros pendientes) representa sobre las primas. En el caso de seguros de vida, la siniestralidad viene dada como la suma del importe de los siniestros, más el incremento de las reservas matemáticas. El *equilibrio técnico* es la suma de los porcentajes de gastos de gestión y siniestralidad sobre primas del ejercicio, y muestra la rentabilidad de las operaciones de seguro en cada ramo, sin tener en cuenta los posibles ingresos ajenos a la actividad, como podrían ser los financieros. El *resultado técnico* vendrá dado por diferencia entre el total de las primas y el equilibrio técnico.

Los datos relativos a la *siniestralidad*, al *equilibrio técnico* y a los *resultados* para los principales ramos se recogen a continuación:

Tabla III.4
Siniestralidad, equilibrio y resultado técnico del sector⁸

Años	Vida			No Vida								
	Sinies	Equi	R° tec	Autos			Salud			Total No Vida		
				Sinies	Equi	R° tec	Sinies	Equi	R° tec	Sinies	Equi	R° tec
1991	102,95	116,56	-16,56	78,56	107,88	-7,88	81,49	98,27	1,73	73,18	105,16	-5,16
1992	101,82	114,24	-14,24	73,36	102,82	-2,82	80,69	96,55	3,45	71,14	103,16	-3,16
1993	104,83	118,23	-18,23	74,35	104	-4	82,33	99,33	0,67	71,23	104,01	-4,01
1994	104,11	113,09	-13,09	74,45	103,92	-3,42	82,8	98,02	1,98	70,82	102,77	-2,77
1995	103,21	113,53	-13,53	74,48	102,88	-2,88	83,83	99,38	0,62	72,19	103,99	-3,99
1996	112,14	121,71	-21,71	74,35	103,62	-3,62	85,2	100,22	-0,22	70,77	101,76	-1,76
1997	119,3	128,46	-28,46	77,77	107,51	-7,51	82,29	98,2	1,8	72,51	103,99	-3,99

⁸ Véase nota anterior.

De esta tabla se deduce que, excepto el ramo de seguros de salud, y no en todos los años, todos tienen un resultado técnico negativo, es decir, los gastos propios de la actividad no pueden ser cubiertos con los ingresos, teniéndose que cubrir esas diferencias con otro tipo de ingreso que, en la mayoría de los casos, es el financiero. Con carácter general, se puede decir que el comportamiento ha sido distinto en los ramos vida y no vida, y en el caso de los seguros de vida, la siniestralidad ha ido creciendo, si bien, para estos seguros, ese crecimiento ha sido consecuencia de un aumento de las provisiones matemáticas en los últimos años. Como consecuencia, el equilibrio técnico ha aumentado y el resultado técnico se ha reducido aún más.

En el caso de los seguros no vida, la siniestralidad ha tenido un comportamiento heterogéneo, aunque si se comparan las cifras iniciales y finales del periodo, se observa que la siniestralidad ha descendido y con ella el equilibrio técnico, lo que ha traído consigo una mejora del resultado técnico.

Para el ramo del automóvil, la siniestralidad y el resultado técnico negativo no han mejorado, a pesar de que el coste medio de los siniestros se ha reducido como consecuencia, fundamentalmente, de la aplicación del nuevo Sistema para la Valoración de las Indemnizaciones para daños a las personas. En este sentido, hay que tener en cuenta la política de descuentos sobre las primas seguida por muchas aseguradoras, y la disminución de precios, lo que ha traído consigo incluso un resultado técnico aún menor en el año 97.

En salud, la siniestralidad es bastante elevada, pero el resultado técnico es positivo, debido a que los gastos externos e internos representan un porcentaje inferior sobre las primas que en el resto de los ramos. En el resto de los ramos no vida, aunque no se dan resultados de manera específica para los ramos que lo componen, sólo presenta resultado técnico positivo para la mayoría de los años el ramo de accidentes personales⁹.

⁹ Véase nota 4.

Se deduce pues que, en un contexto de tipos de interés decrecientes, los resultados del sector están bastante comprometidos y se refuerza la necesidad de que las compañías se apoyen en sus resultados técnicos para alcanzar un equilibrio sostenido a largo plazo.

d. Canales de distribución

Remitiéndonos a las características apuntadas de la actividad aseguradora, tradicionalmente, el canal de distribución mayoritario ha sido la mediación (agente o corredor), seguido de las oficinas de la propia entidad, aunque, esto ha cambiado en los últimos años, y varía según el ramo considerado. Según ICEA, en su informe para 1998 sobre canales de distribución (Aseguradores, 1999b: 21), actualmente los bancos y cajas son los principales canalizadores del negocio de vida (el fenómeno de la bancaseguros distribuye el 94% y los agentes el 2,64%), mientras que los mediadores son los mayores distribuidores del negocio asegurador en el caso de los ramos no vida. La importancia de este fenómeno se manifiesta también en el hecho de que, entre las quince aseguradoras más rentables, seis son aseguradoras bancarias; y entre las tres primeras por volumen de primas, la segunda y la tercera también son entidades de bancaseguros.

La distribución a través de mediadores emplea el 61% de agentes, el 19% de corredores y el 12% de oficinas de la entidad. Las empresas que trabajan sin mediadores, es decir, las de marketing directo, comercializan el 70% del negocio directamente, y el 17% a través de agentes. Se trata, sobre todo, de productos que se orientan a clientes-tipo y que necesitan poco asesoramiento, como por ejemplo en el ramo del automóvil y en los que es fácil la comparación de los precios de los distintos seguros. Como consecuencia de los menores gastos de esta forma de distribución, especialmente en este ramo, se ha producido un abaratamiento del precio de la prima y un incremento de la competencia.

1.2.3 Perspectivas

Una vez analizados los principales indicadores del sector, lo que ha permitido la caracterización del mismo, es necesario tener en cuenta una serie de acontecimientos que van a ser fundamentales para el desenvolvimiento de la actividad aseguradora en el futuro. Según la Dirección General de Seguros (González de Frutos, 1999) el momento actual se caracteriza por su transitoriedad debido a los cambios que se avecinan: moneda única, cambios legislativos, cambios en el comportamiento del consumidor y en los canales de distribución, entre otras circunstancias.

La implantación del euro plantea a la empresa aseguradora española ciertos retos, oportunidades y amenazas que vendrán motivados, entre otros factores, por la reducción de los costes de transacción y la eliminación de los costes por diferencias de cambio, así como por el aumento de la competencia en el sector. En cualquier caso se prevé que la introducción de la moneda única origine un entorno de estabilidad, con un recorte en los ingresos financieros originado por los bajos tipos de interés. Esto conducirá a tomar medidas que tiendan a incrementar las primas para lograr mantener los resultados, es decir, a un mayor control de la siniestralidad y a una reducción de los gastos generales y de distribución. Será, por tanto, necesario (Pérez Torres, 1999) la aplicación del concepto de eficacia en las entidades.

Con la introducción del euro, se espera que el aumento de la competencia entre las aseguradoras que se dirigen al segmento integrado por particulares, y pequeñas y medianas empresas, sea mayor que entre aquellas aseguradoras que cubren los riesgos del segmento constituido por grandes empresas, dado que la mayoría de estos seguros están contratados con aseguradoras extranjeras. Como consecuencia del impacto de la competencia en los mercados locales y de la necesidad de ser más competitivo, será necesario tener en cuenta los mercados exteriores incipientes para aumentar el negocio. Hasta ahora pocas son las empresas españolas que se han lanzado al mercado

exterior: el Grupo Mapfre, el Grupo Allianz-AGF y Catalana Occidente, pero su volumen de negocio fuera de nuestro país es muy pequeño. Además de por el aumento de la competencia, la estrategia de internacionalización de las instituciones aseguradoras se hará necesaria como respuesta a las exigencias de sus clientes corporativos, que tienen negocios en otros países.

Los cambios legislativos que más van a afectar al sector son los que traen consigo el Reglamento que desarrolla la Ley de Ordenación y Supervisión de Seguros Privados y el Reglamento de Exteriorización de los Compromisos por Pensiones. El primero traerá consecuencias en determinados ramos en cuanto establece métodos más prudenciales para el cálculo de las provisiones. El segundo impulsará sobre todo el ramo vida, como consecuencia de la obligación de exteriorizar los compromisos por pensiones de las empresas antes del 1 de enero del 2001.

Todo conduce a que las compañías financieras y de seguros deben acelerar sus programas de fidelización de la clientela. Según algunos estudios (Aseguradores, 1999a), el 45% de las compañías utiliza bases de datos por segmentos para identificar los mejores clientes y oportunidades de venta, pero sólo el 25% segmenta sus bases de datos para identificar a los clientes que pueden irse para contactar con ellos antes de que lo hagan. Por tanto, es necesario que las entidades aseguradoras tengan en cuenta dicho aspecto, lo que posiblemente afectará al sector. Se debe destacar también el nivel de información creciente de los asegurados en todos los ramos, debido a la prensa económica y a las organizaciones de consumidores (Fundación Mapfre Estudios, 1999b), lo que conduce a que los clientes muestren, cada vez más, una actitud más independiente. Éstos saben que existen distintas ofertas y buscan las mejores en calidad-precio lo que traerá consigo la reducción de los márgenes y el incremento de la competencia.

En cuanto a los canales de distribución, los más novedosos, los canales de distribución directos, permiten ofrecer más barato el producto, pero eliminan las relaciones personales, y éstas son las que construyen una fidelidad a

largo plazo. Por ese motivo, el empleo de este tipo de canales puede afectar al comportamiento de los asegurados y a ciertos ramos del sector, como es el caso del automóvil. En cuanto a los sistemas tradicionales de distribución, es necesario intensificar la comunicación con los mediadores para fidelizarlos con la empresa, dado que de ellos depende la imagen de la misma (Villanueva, 1999: 24). La confianza, la experiencia y el conocimiento profundo del sector por parte del corredor de seguros, y su actitud ante la resolución de siniestros, son las ventajas añadidas que ofrece este tipo de canal. Además, las aseguradoras ya tienen infraestructura informática suficiente para vender a través de Internet, pero para distribuir "on line" se tienen que diseñar productos específicos para este tipo de canal; se trata de una actividad de futuro y es necesario prepararse para estar en buena posición en los próximos años.

La preocupación de las empresas del sector respecto a las consecuencias del efecto 2000, ha llevado durante varios años a adoptar medidas para la adecuación de sus sistemas informáticos, al establecimiento de mecanismos para el tratamiento adecuado de posibles reclamaciones relacionadas con riesgos previsibles no cubiertos por las pólizas (fundamentalmente de responsabilidad civil), y para la asegurabilidad de los riesgos del milenio cuando hayan sido adoptadas por el asegurado las correspondientes medidas de prevención. La reducción de la incertidumbre respecto a dicho efecto creemos que influirá favorablemente en la evolución del sector.

Para resumir, las perspectivas son diferentes según los distintos ramos. En el caso de los seguros de vida, es previsible que su tasa de crecimiento siga aumentando debido al comportamiento demográfico de la población, lo que aumentará también la necesidad de los sistemas de previsión social alternativos. El entorno favorable de crecimiento económico también favorecerá el ahorro y la contratación de productos de previsión, y por tanto el seguro de vida. En lo que a las innovaciones de producto se refiere, merece destacar la aparición de un nuevo producto, los unit linked, que está revolucionando y reactivando este mercado. Este producto reúne las ventajas

del seguro de vida y de los fondos de inversión, ya que permite al inversor cambiar sus ahorros entre los fondos tantas veces como quiera sin tener que tributar por ello, eliminando por tanto los costes de cambio. Gracias a este producto, según los datos relativos al mes de septiembre de 1999, el crecimiento de las primas del seguro de vida hasta ese mes, respecto al mismo periodo del año anterior, ha sido del 31%. El ramo de vida, también se verá favorecido por el cambio de tratamiento fiscal de la ley del IRPF. Por ese motivo, el diseño de sistemas tributarios que impulsen el ahorro de particulares y empresas mediante este tipo de instrumento de ahorro provocará la evolución favorable de este ramo. Es obvio que con el envejecimiento de la población y el incremento de jubilados este ramo tendrá una gran oportunidad en la próxima década, lo cual también puede ser una amenaza para compañías incapaces de evaluar bien el riesgo de las rentas de supervivencia.

En cuanto al ramo del automóvil, el crecimiento del parque automovilístico puede impulsar este sector y superar la crisis de los últimos años debida a la bajada del precio de las pólizas y a la competencia. Este hecho también beneficiará a otros ramos relacionados con el automóvil, como el de responsabilidad civil, daños propios e incendios, robo, defensa jurídica y asistencia en viaje. Los ramos del automóvil y de responsabilidad civil dependen en gran medida de los incentivos y de las medidas gubernamentales (Meilán, 1999: 106) por lo que pueden verse afectados por los cambios legislativos.

En el caso de los seguros de salud, la competencia desde la vertiente pública impulsará a los aseguradores a implantar políticas constantes de ampliación y mejora de servicios, lo que puede traer aparejado un aumento del precio de la prima y, por tanto, un crecimiento del volumen de éstas. La nueva legislación hace más atractiva la contratación de este producto, a pesar de la eliminación en la ley de IRPF de la desgravación por este concepto. El deterioro de los sistemas tradicionales de seguridad social también impulsará este tipo de seguro. Respecto al resto de ramos, parece

que son el de responsabilidad civil y el de multirriesgos los que más posibilidades de crecimiento presentan.

Con el fin de analizar más en detalle el sector sería interesante conocer otras magnitudes, como el porcentaje que los gastos externos e internos representan sobre las primas en el caso de los seguros industriales y personales, o el equilibrio técnico y la siniestralidad. La agregación de este tipo de información sería costosa en cuanto que la información sobre seguros solo está disponible para los ramos desagregados. La aplicación de la metodología STATIS permitirá obtener conclusiones, sobre este tipo seguros, sin necesidad de disponer de los datos concretos. Por otra parte, la aplicación conjunta de los modelos *ARIMA* y de dicha metodología permitirá confirmar las tendencias señaladas al observar los datos presentados anteriormente, analizar un mayor número de años, poner de manifiesto relaciones entre las variables que no se observan directamente sobre los datos y, en definitiva, completar el estudio de la evolución del sector.

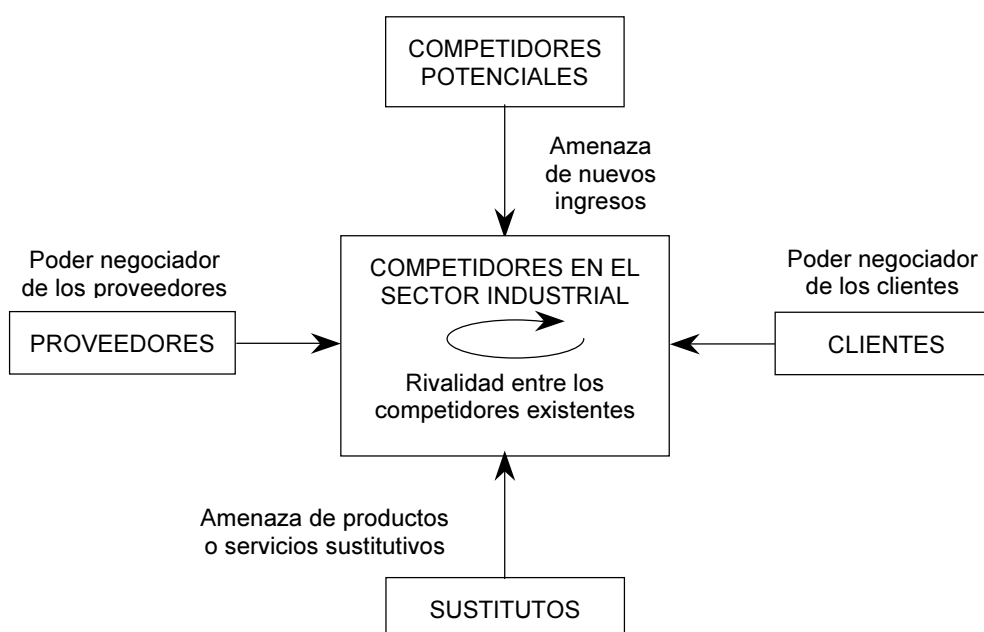
1.3 Análisis estructural del sector asegurador español

Con el fin de profundizar en el conocimiento del sector asegurador español, se va a analizar la capacidad del mismo para obtener beneficios mediante el modelo de Porter de las cinco fuerzas competitivas. El objetivo de este modelo es identificar las amenazas y oportunidades que el entorno presenta para cualquier organización del sector, determinando su capacidad para obtener beneficios. Si se añade a este Análisis Externo, el Análisis Interno de las fortalezas y debilidades de la propia organización, se estará en condiciones de formular la estrategia competitiva más conveniente.

1.3.1 El modelo de Porter para el análisis de sectores industriales

Este modelo parte de la idea, expresada por Porter (1982: 24), de que la capacidad de un sector para generar rentas superiores depende de la acción de cinco fuerzas competitivas básicas, que son las recogidas en la figura III.1:

Figura III.1
Determinantes del nivel de competencia de un sector industrial (Porter, 1982: 24)



La rivalidad entre competidores aparece en el centro puesto que se ve afectada por las otras fuerzas, que deben ser analizadas teniendo en cuenta si son lo suficientemente fuertes como para reducir o eliminar las ganancias de la industria. El análisis de un sector basado en este modelo es una poderosa herramienta para diagnosticar la situación actual y la posible evolución del sector. No obstante, hay que tener en cuenta sus limitaciones en cuanto ignora los cambios en la renta del consumidor, las estrategias de las entidades para modificar la demanda, y el papel del gobierno, excepto si este es proveedor o cliente. Por tanto, el modelo es básicamente cualitativo

y resulta adecuado si el objetivo principal es detectar la tendencia de la industria a aumentar o disminuir sus ganancias.

a. Las cinco fuerzas competitivas de un sector industrial

A continuación se van considerar cada una de estas fuerzas y los factores económicos que son relevantes en cada una de ellas.

COMPETIDORES POTENCIALES

Los competidores potenciales hacen referencia a las nuevas empresas que quieren entrar a competir en el sector o industria. La existencia de nuevos competidores es una amenaza para el sector pues reduce su atractivo, ya que el mercado debe repartirse entre más, disminuyendo la concentración e intensificando la competencia actual.

La decisión de entrar en un nuevo negocio depende de los niveles de rentabilidad que presente la industria, y del riesgo que ello suponga. Ahora bien, el acceso a la industria para estos nuevos competidores está condicionado por la existencia de dos tipos de factores: las barreras de entrada y la reacción de los competidores establecidos ante un nuevo ingreso.

— *Barreras de entrada*. Se definen (Ventura Victoria, 1994: 55) como “las características del sector que reducen la tasa de entrada por debajo de lo que sería necesario, para eliminar los beneficios extraordinarios de las empresas ya presentes en él”. Es decir, serían aquellos mecanismos que dificultan el ingreso de nuevas empresas a la industria, normalmente mediante la disminución de las expectativas de rentabilidad de los posibles nuevos competidores, lo que implica que la existencia de barreras de entrada frena la aparición de nuevos competidores, protegiendo a los ya instalados y preservando, por tanto, sus expectativas de rentabilidad.

Dejando a un lado la clasificación de las barreras de entrada que distingue entre barreras de entrada absolutas y relativas (las primeras son aquéllas que son imposibles de superar, mientras que las relativas son aquéllas que pueden superarse aunque ofrezcan niveles de dificultad diferentes), las principales barreras de entrada tal y como recoge Ventura Victoria (1994, 55-74) son:

- Las economías de escala y de alcance, que frenan el ingreso obligando al que pretende entrar a producir a gran escala, con los riesgos que ello conlleva, o a entrar a pequeña escala, soportando desventajas en costes.

Las economías de escala se ponen de manifiesto en la reducción de los costes unitarios a medida que aumenta el volumen de producción. Son varias las causas que provocan esta reducción de costes, una de las cuales es la división del trabajo en partes más elementales ganando en especialización. La adaptación de los equipos productivos, diseñados para conseguir dicha especialización, también se traduce en costes fijos elevados de tal manera que si la empresa no produce y vende un gran número de unidades no es rentable. Por ello, esas nuevas tecnologías presentan "indivisibilidades" que hacen que las empresas del mercado luchan por mantener su nivel de rentabilidad. La existencia de economías de escala juega un papel esencial para entender el número de empresas que a largo plazo forman cierto sector, pues la presencia de dichas economías determina que una empresa que quiera entrar a competir deberá alcanzar el tamaño mínimo para ser eficiente, o de lo contrario tendrá costes superiores a las empresas ya instaladas. Si este tamaño mínimo eficiente es relativamente grande, teniendo en cuenta el tamaño del mercado, el nuevo ingreso debe alcanzar una cuota de mercado sustancial o si no hay una desventaja de coste considerable, con lo que las economías de escala se convierten en una barrera de entrada.

Las economías de alcance surgen de la posibilidad de compartir recursos entre diferentes productos o actividades, de tal forma que se consigue un coste menor que la suma de los costes de producir o realizar las actividades por separado. Esta es una ventaja de las empresas existentes respecto a las nuevas entradas y en la medida que éstas existan, una empresa deberá diversificarse o de lo contrario tendrá una desventaja sustancial.

- Diferenciación de las empresas establecidas en función de la marca o prestigio logrado. La diferenciación crea una barrera para el ingreso obligando al que se quiere incorporar a realizar grandes inversiones para superar la fidelidad existente. La importancia del prestigio de una marca será mayor cuando el comprador no conoce a ciencia cierta la calidad del producto o servicio que adquiere hasta que lo usa, asumiendo un riesgo que aumenta con la importancia del producto, como sucede con los productos farmacéuticos, servicios médicos o bienes que exigen importantes desembolsos.
- Costes de cambio elevados a los que debe hacer frente el cliente al cambiar de proveedor, tales como reentrenamiento del personal, coste del nuevo equipo auxiliar, necesidad de asistencia técnica, rediseño del producto o costes. Al ser elevados, los proveedores de nuevo ingreso tendrán que ofrecer una gran reducción de precios o una mejora en el rendimiento del producto para que el comprador cambie de proveedor. Ya sea como consecuencia de este caso o del anterior, la fidelidad de los clientes hacia las empresas establecidas, supone una importante barrera de entrada para las empresas que pretenden irrumpir en el sector.
- La necesidad de invertir grandes recursos financieros para competir. Estas necesidades de capital no tienen por qué ser sólo para instalaciones productivas sino que pueden ser para financiar a los

clientes, mantener existencias o cubrir pérdidas iniciales, para publicidad o investigación y desarrollo.

- Dificultad para acceder a los canales de distribución si ya están cubiertos por las empresas existentes. Las nuevas empresas deben convencer a los canales de que acepten sus productos mediante promociones, publicidad compartida, etc., lo que implica un alto coste y una barrera de entrada para las empresas que pretenden entrar.
- Desventajas en costes diferentes de las economías de escala, como por ejemplo, tecnología de producto patentada, acceso favorable a las materias primas por haber contratado las fuentes más favorables y/o a precios favorables; ventajas de localización; subvenciones del gobierno, concedidas normalmente a empresas ya establecidas y el efecto experiencia.
- Política gubernamental: el gobierno puede limitar o impedir la entrada a determinadas industrias mediante licencias, requerimientos mínimos de capital, normas anticontaminantes, de seguridad y eficacia, etc. Estas limitaciones pueden disuadir a los nuevos competidores y convertirse en una barrera de entrada al sector.

— *Reacción de los competidores establecidos.* En la medida que los competidores actuales sean capaces de desarrollar una reacción fuerte ante los nuevos, éstos tienden a ser disuadidos. Las condiciones que señalan una alta probabilidad de rechazo por parte de los ya existentes son:

- Tradición de represalias en el sector, como pueden ser guerras de precios, campañas de publicidad masivas u ofertas especiales para disuadir a los nuevos entrantes. La reputación lograda por una empresa, basada en sus actuaciones pasadas para defender el mercado, puede afectar claramente a las expectativas de las empresas potencialmente entrantes. La creación de esa reputación agresiva requiere un planteamiento a largo plazo, en virtud de lo cual decide comportarse

de forma agresiva con los primeros entrantes a fin de mostrar una reputación que, a pesar de su coste, influya negativamente en los intentos de entrada de las nuevas empresas.

- Empresas establecidas con fuertes recursos para defenderse, como exceso de liquidez, capacidad de endeudamiento no utilizada o ventajas en canales de distribución o clientes. Tal y como manifiesta Ventura Victoria (1994, 58) cuando en un sector haya economías de escala y el tamaño mínimo eficiente sea grande con relación al mercado, la cuota de mercado que necesitan las empresas será grande y por tanto mayor la reacción esperada de las empresas instaladas ante las nuevas entradas.
- La presencia de activos específicos a la industria es otro factor que refuerza la posibilidad de luchar por el mercado y no acomodar la entrada de nuevos competidores, pues se trata de activos que no pueden ser fácilmente convertidos para otras actividades sin que se devalúen o carezcan de valor en absoluto. Esto fuerza a las empresas instaladas a defender su posición en el mercado, dadas las pérdidas que supone la salida de la industria.

RIVALIDAD DE LOS COMPETIDORES EXISTENTES

En este caso se hace referencia a la actuación de los competidores existentes en la industria en un determinado momento. Aquí se estudian las características básicas de la industria, que definen el marco genérico de la competencia, así como posibles acciones y reacciones de las empresas ya establecidas que puedan alterar la intensidad de la competencia. Tal intensidad es el resultado de la siguiente serie de factores estructurales:

— *Número de competidores y equilibrio entre ellos.* Este factor tiene que ver con la distribución de la cuota de mercado entre los distintos competidores. En este sentido, las industrias pueden ser concentradas o fragmentadas. Una empresa es concentrada cuando pocas empresas tienen

una cuota alta del mercado, quedando el resto para un número amplio de empresas. Por el contrario, se dice que una industria es fragmentada cuando existe un gran número de empresas medianas y pequeñas y ninguna tiene una participación importante en el mercado, de modo que es necesario acumular los volúmenes de ventas de un gran número de empresas para obtener una cuota de mercado significativa. A medida que sea mayor el número de competidores establecidos y el equilibrio entre los mismos, la intensidad de la competencia será mayor y por consiguiente en las industrias concentradas, la intensidad de la competencia se reduce.

— *Ritmo de crecimiento del sector.* Una primera aproximación (STRATEGOR, 1995: 24) al ritmo de crecimiento de un sector la ofrece la comparación de su tasa de crecimiento con la del crecimiento del PIB: un sector en crecimiento será aquél cuyo índice de expansión supera el ritmo de evolución del PIB. También puede compararse el ritmo de crecimiento en el sector con el ritmo de crecimiento del mismo sector en países vecinos.

Cuando las ventas se expanden globalmente, cada empresa puede crecer manteniendo, incluso ampliando, su participación inicial en la cuota de mercado. En cambio, cuando el ritmo de crecimiento se reduce, las empresas que quieren crecer o mantener su cifra de ventas deberán arrebatarla a las empresas competidoras lo que incrementará la intensidad de la competencia. En mercados con oscilaciones cíclicas la competencia también se verá incrementada, debido a la dificultad de adaptar la demanda a la oferta en las diferentes fases por las que atraviesa la economía.

— *Barreras a la movilidad.* Son aquellos obstáculos o dificultades que impiden a las empresas moverse de un segmento a otro del sector. La existencia de barreras a la movilidad protege la competencia dentro del segmento, por lo que su intensidad para el conjunto de la industria decrece. Son típicas de industrias oligopolistas en las que mediante acuerdos tácitos se reparte el mercado, o de industrias con alto grado de diferenciación de producto, como en el sector automovilístico.

— *Barreras de salida*. Son aquellos factores que impiden o dificultan el abandono del sector por parte de una empresa, incluso en el caso de resultados pobres o negativos. La existencia de barreras de salida fuerza a las empresas a luchar por sobrevivir y, por tanto, seguir compitiendo en la industria, por lo que la intensidad de la competencia aumenta. Las principales barreras de salida dependen de que se den algunos factores del tipo siguiente:

- Activos especializados, que son activos cuya utilidad fuera del sector es muy escasa por lo que tienen poco valor de liquidación o altos costes de reconversión.
- Costes fijos de salida elevados, relativos a aquellos que deben asumirse necesariamente para abandonar la actividad como, por ejemplo, las indemnizaciones laborales, los costes de reinstalación o liquidación de fondos a bajo precio, etc.
- Interrelaciones estratégicas que se dan entre los distintos negocios de una empresa y que la obligan a continuar en todos ellos sin posibilidad de abandonar ninguno. Sucede, por ejemplo, cuando hay que estar presente en los mismos negocios que está la empresa líder del sector.
- Barreras emocionales, que con frecuencia tienen extraordinaria fuerza, como son la identificación personal con el negocio, la lealtad hacia los empleados y otras razones. Este tipo de barreras hacen que se siga compitiendo a veces de forma irracional, frente a lo que sería un planteamiento estrictamente económico.
- Restricciones sociales y gubernamentales, que hacen referencia a las repercusiones de tipo social o político que puede tener la decisión de abandonar una actividad, los movimientos sociales y manifestaciones públicas que genera el cierre de una empresa que implican una considerable pérdida de imagen, pudiendo afectar a otros negocios de la misma. En otras ocasiones son las administraciones las que impiden o presionan ante el cierre de instalaciones bien por la pérdida de empleos o por otros efectos económicos.

— *La estructura de costes de las empresas.* Un mayor peso de los costes fijos sobre los variables dentro de la estructura de costes de las empresas hace tender a las empresas a operar a plena capacidad, para intentar reducir sus costes medios. Ello tiende a incrementar los volúmenes de producción con lo que la intensidad de la competencia aumenta. Lo mismo se podría decir ante la presencia de otros costes, como los de publicidad, o los de almacenamiento de materias primas y/o productos terminados.

— *Diferenciación de productos.* A medida que se dé un mayor nivel de diferenciación de productos, la intensidad de la competencia se reduce, ya que los clientes se fidelizan con los distintos productos. Si no hay diferenciación de producto, la elección de los compradores se basa en el precio, lo que da como resultado una intensa competencia en precio y servicio.

— *Costes de cambio.* De igual forma que el anterior, la existencia de coste de cambio de proveedores o clientes reduce la intensidad de la competencia, al vincularse más estrechamente unos y otros, mientras que la ausencia de costes de cambio intensifica la competencia.

— *Capacidad productiva instalada.* En un sector la competencia se recrudece cuando la capacidad productiva mínima recomendable para desempeñar la actividad es mayor, lo que fuerza a las empresas del sector a realizar movimientos competitivos para dar salida a grandes volúmenes de producción.

— *Diversidad de competidores.* Si los competidores difieren en estrategias, orígenes, personalidad, relaciones con sus compañías matrices, objetivos y formas de competir, pueden interferir continuamente unos sobre otros, provocando efectos intensificadores de la competencia. Los competidores extranjeros y las empresas de diferentes tamaños pueden acentuar esta situación.

— *Intereses estratégicos.* A medida que un mayor número de empresas estén interesadas simultáneamente en lograr el éxito en un sector, la

competencia se intensifica, ya que estarán dispuestas a desarrollar todo tipo de acciones que les conduzcan a ese fin, aunque tengan que sacrificar temporalmente sus resultados.

PRODUCTOS SUSTITUTIVOS Y COMPLEMENTARIOS

Los sustitutivos son aquellos productos que satisfacen las mismas necesidades de los consumidores que las que satisface el producto ofrecido por la industria. Una aproximación correcta a éstos debe contemplar todos los productos que cumplan dicha condición independientemente de la industria de la que provengan y en la medida en que en una industria aparezcan productos sustitutivos, el atractivo de la misma tenderá a decrecer y por tanto también sus expectativas de rentas superiores. El efecto de los productos sustitutivos sobre el atractivo de una industria depende de los siguientes factores:

— Los precios de los productos sustitutivos con relación a los de la industria. Si el precio de estos sustitutivos es alto difícilmente afectará a la demanda del producto del sector.

— El grado de obsolescencia que los productos sustitutivos incorporan en los productos de la industria, y por tanto afecta negativamente al sector.

— Si los costes de cambio para consumir productos alternativos son elevados, la existencia del producto sustitutivo no afectará gravemente al sector.

En determinadas ocasiones el nivel de precios propio de la industria marca el umbral a partir del cuál se pueden hacer económicamente rentables productos alternativos como ocurre, por ejemplo, con el petróleo como fuente de energía frente a otras energías alternativas.

En el caso de los productos complementarios, es decir, aquéllos que se emplean conjuntamente con el producto del sector, la situación es la inversa y pueden ofrecer nuevas oportunidades de beneficios.

PODER DE NEGOCIACIÓN DE PROVEEDORES Y CLIENTES

Este análisis es importante ya que va a determinar la capacidad que tienen las empresas de un sector para influir, de manera decisiva, en las industrias que le preceden o le siguen en el proceso de producción. En general puede afirmarse que, a medida que el poder de negociación de éstos es mayor, el atractivo de la industria disminuye, ya que serán ellos quienes impongan sus condiciones en las transacciones realizadas con las empresas de la industria analizada, captando así una parte del valor añadido generado por dicho sector. Por el contrario, una posición de debilidad de los proveedores y clientes facilita a las empresas establecidas en la industria la negociación de las mejores condiciones para sus transacciones.

Aunque el poder negociador de proveedores y clientes no es siempre homogéneo para todos ellos, existen algunos factores que inciden, de manera general, en dicho poder, favoreciendo a unos agentes u otros. Los factores más importantes (Navas López y Guerras Martín, 1998: 147) que deben considerarse para analizar el poder relativo de un cliente o proveedor son los siguientes:

- Grado de concentración de éstos en relación con el sector estudiado.
- Volumen de transacciones en relación con dicho sector.
- Grado de diferenciación de los productos o servicios objeto de transacción y su carácter almacenable o no.
- Costes de cambio de proveedor.
- Nivel de beneficios del cliente o proveedor en relación con la industria.
- Amenaza real de integración vertical hacia atrás o hacia delante.

-
- Importancia del producto o servicio vendido para la calidad de los productos del cliente.
 - Nivel de información del cliente o del proveedor respecto a la industria.

Así, un grupo de clientes tiene mucho poder si una gran proporción de las compras del sector es adquirida por ellos, y además éste presenta costes fijos altos que obligan a mantener la capacidad ocupada al máximo. Cuando los productos están poco diferenciados, los clientes, seguros de que siempre pueden encontrar proveedores alternativos, tendrán gran capacidad de negociación, circunstancia que también se produce si los gastos de cambiar de proveedor son muy bajos. Si los clientes están parcialmente integrados o plantean una amenaza creíble de integración hacia atrás, están en posición de exigir concesiones en la negociación, no sólo por dicha amenaza de integración sino por el conocimiento del negocio que poseen. En el caso de que la información de la que dispone el comprador sobre la oferta sea total, éste tendrá mayor ventaja negociadora que cuando la información es escasa, pues con información completa el comprador estará en posición de comparar mejor los diferentes productos ofrecidos por los vendedores. De manera análoga deben considerarse dichos factores para los proveedores.

b. Las estrategias genéricas

El análisis de las cinco fuerzas identifican las amenazas sobre las ganancias de la industria a las que todas las entidades del sector deben enfrentarse. En primer lugar las compañías pueden posicionarse frente a sus rivales, desarrollando una ventaja en costes (estrategia de liderazgo en costes) o diferenciándose (estrategia de diferenciación). En segundo lugar, las compañías pueden identificar un segmento de la industria en el que las cinco fuerzas son menos severas (estrategia de segmentación). En tercer lugar, las entidades pueden intentar cambiar alguna de las cinco fuerzas, aunque esta alternativa es difícil de llevar a cabo.

Con el fin de abordar posteriormente el concepto de grupo estratégico, es conveniente definir someramente estas estrategias. La estrategia de liderazgo en costes tiene como objetivo lograr un coste final mínimo respecto a la competencia, junto con una calidad aceptable y una política de precios que permita alcanzar un volumen de ventas y un crecimiento de la cuota de mercado rentables. El líder en costes no puede ignorar la posibilidad de diferenciar el producto ya que si éste no es aceptado por los compradores, puede verse obligado a reducir los precios en exceso para ganar cuota de mercado incurriendo en pérdidas. En esta estrategia el papel de la función de producción es fundamental, y persigue lograr la máxima eficiencia posible en la asignación de recursos y en la explotación de los mismos. La empresa debe llevar a cabo inversiones que tengan una mayor incidencia en la reducción del coste final, respecto a la competencia, tales como el diseño de productos estandarizados que facilite la fabricación a gran escala, o las inversiones continuadas en equipos productivos que generen una mayor eficiencia relativa.

La estrategia de diferenciación persigue lograr poder de mercado, esto es, conseguir que el mercado acepte un producto a un precio superior a la competencia o tenga más demanda a igualdad de precio. Una empresa tiene básicamente dos formas para diferenciarse: basada en las características intrínsecas del producto (calidad, diseño innovador, tecnología incorporada, grado de exclusividad) o basada en la imagen de marca. La empresa debe optar por la diferenciación que le permita fijar un precio superior al coste de diferenciar, teniendo en cuenta que si quiere obtener resultados superiores la empresa diferenciada debe mantener la proximidad en el coste respecto a sus competidores.

La última estrategia consiste en dirigirse a un grupo de compradores en particular, en un segmento de la línea del producto, o en un mercado geográfico. Esta estrategia se basa en la premisa de que la empresa puede así servir a su estrecho objetivo estratégico con más efectividad o eficacia que los competidores que consideran todo el mercado. Como consecuencia,

la empresa logra la diferenciación para satisfacer mejor las necesidades de un objetivo en particular, o costes inferiores al servir a éste, o ambos respecto al mercado limitado e incluso desde la perspectiva del mercado en su totalidad.

Estas estrategias pueden entenderse como las grandes vías diseñadas para obtener una ventaja competitiva sostenible frente a la competencia, y por tanto, mayores beneficios a largo plazo. La opción estratégica elegida determina la posición de la empresa dentro del sector, próxima a otras empresas del sector que siguen estrategias similares.

Para poder avanzar en el conocimiento del entorno competitivo resulta útil clasificar a las empresas de un sector en grupos homogéneos respecto a las principales dimensiones que configuran la estrategia, lo que da lugar al análisis de los **grupos estratégicos** (Ventura Victoria, 1994: 163). Este análisis resultará útil para valorar el atractivo de las distintas áreas, no sólo para las empresas que quieran entrar en un determinado sector, sino también para las empresas ya establecidas, en cuanto a sus decisiones sobre qué segmentos de mercado abordar, qué áreas abandonar y cómo repartir los recursos (STRATEGOR, 1995: 18).

Según considera Porter (1982: 148) un grupo estratégico es el conjunto de empresas de un sector que siguen una misma o similar estrategia. Este mismo autor señala que los grupos estratégicos no son equivalentes a los segmentos del mercado, sino que están definidos sobre la base de una concepción más amplia. Por tanto, un grupo estratégico es un dispositivo analítico diseñado como ayuda en el análisis estructural, y es un marco de referencia intermedio entre considerar un sector industrial en su conjunto y considerar cada empresa por separado.

Estos grupos existen por diversas razones: fuerzas y debilidades diferentes al inicio de las operaciones, diferentes épocas de ingreso en el negocio, etc. Sin embargo, una vez que se han formado los grupos, las empresas que se encuentran dentro del mismo grupo estratégico, por lo general, se parecen mucho unas a las otras y tienden a ser afectadas y a responder de igual forma a las circunstancias o movimientos en el sector industrial.

La delimitación de los grupos estratégicos pasa por la identificación previa de las variables o dimensiones estratégicas que permitan definir las características específicas de las empresas de cada grupo. Respecto a los resultados, tal y como manifiesta Grant (1996: 135) la hipótesis de que las diferencias de rentabilidad dentro de los grupos estratégicos son menores que las diferencias entre grupos estratégicos, no ha recibido soporte empírico consistente. La inconsistencia de los hallazgos empíricos puede reflejar el hecho de que los miembros de un grupo, aunque persiguen estrategias similares, no están necesariamente en competencia entre sí. Este hecho no significa que el concepto de grupo estratégico no sea útil, sino que el análisis de grupos estratégicos es un buen instrumento descriptivo para obtener una representación de la tipología de empresas en un sector, y los tipos de estrategias que han resultado viables. En definitiva, dicho análisis puede contribuir al entendimiento de la estructura, dinámica competitiva y evolución de un sector, así como a la comprensión de los aspectos de dirección estratégica propios del mismo. Aunque si a corto plazo, a largo plazo ninguna de las tipologías que puedan establecerse es fija, sino que evolucionan con mayor o menor velocidad en función de la madurez del sector y del comportamiento de las empresas que lo forman (STRATEGOR, 1995: 18).

Una vez definidos los grupos estratégicos hay que plantearse si es posible para una empresa trasladarse de un grupo a otro y a qué precio. Esto depende de las barreras a la movilidad, de forma que si tales barreras no existen, las empresas pueden trasladarse de los grupos con menor rendimiento a los grupos con mayor rendimiento, igualándose las

posibilidades de todas las empresas; mientras que si las barreras a la movilidad existen, actúan como barreras de entrada para el grupo correspondiente.

Las últimas tendencias en el Análisis Estratégico ponen mayor énfasis en las características cualitativas de las empresas para la formulación de una estrategia. Tal y como manifiesta Ventura Victoria (1994: 166) los grupos estratégicos y las barreras a la movilidad se explican por las diferencias de recursos y capacidades entre las empresas, que son las que motivan las diferentes formas de enfrentarse a las cinco fuerzas competitivas. Según este planteamiento, la empresa puede considerarse como un conjunto de recursos heterogéneos, que difieren de unas empresas a otras, y de movilidad imperfecta, que es el punto de partida de la **Teoría de los recursos y capacidades**.

Como resultado, las organizaciones deben dedicar atención preferente a identificar, desarrollar, proteger y desplegar aquellos recursos y capacidades que aseguren el logro de una ventaja competitiva sostenible y por tanto rentas superiores a largo plazo. Los recursos y las capacidades son los activos de la empresa considerados desde una perspectiva amplia. Los recursos pueden ser tangibles, como fábricas, instalaciones, o activos financieros; o intangibles, como reputación, prestigio o habilidades de los trabajadores. Pero los recursos poseídos por la empresa no explican por sí solos la ventaja competitiva de la misma y es necesario conocer la forma en que la empresa es capaz de explotar conjuntamente dichos recursos, lo que determinará las capacidades de la misma.

Estas capacidades se convierten en competencias distintivas cuando no son generalizables al resto de empresas. Según este planteamiento, la esencia de la estrategia no es sólo la estructura de los productos y mercados de

una empresa, sino también la dinámica de su comportamiento, y el objetivo es determinar y crear habilidades organizativas difíciles de imitar, que distingan a una empresa de sus competidores ante los clientes (Ventura Victoria, 1996: 83-118).

1.3.2 Aplicación al sector asegurador español

Tomando como referencia el modelo de Porter de las cinco fuerzas competitivas, se van a tratar de identificar las características estructurales que determinan la naturaleza de la competencia del sector del seguro español.

a. Las cinco fuerzas competitivas en el sector asegurador

Los factores que permiten analizar dichas fuerzas van a ser considerados de manera general para todo el sector, especificando, cuando sea conveniente, las características diferenciales para los seguros de vida, los seguros de automóviles y los seguros de salud, que, como ya se ha señalado, son los ramos con una mayor importancia relativa en el sector.

COMPETIDORES POTENCIALES

La existencia de competidores potenciales es una amenaza para el sector, dado que intensificará las condiciones de competencia actual, reduciendo su grado de atractivo. Las barreras de entrada y las posibles reacciones de los competidores ya establecidos, condicionan la entrada al sector de otros nuevos.

— *Barreras de entrada.* Consideraremos a continuación los factores mencionados, que pueden suponer una barrera de entrada en el sector en cualquiera de los ramos.

-
- *Las economías de escala y de alcance.* La existencia de economías de escala en el sector del seguro ha sido siempre objeto de controversias (Lecea Flores de Lemus, 1991: 44). Los estudios empíricos no son unánimes al respecto, pero la mayoría parece confirmar su existencia; aunque sean de menor cuantía que las que pueden existir en otros sectores industriales, porque el coste de lanzamiento de un producto de seguro es relativamente pequeño. En el sector del seguro tienen su origen en las funciones de distribución, automatización, inversión, y sobre todo en la imagen de marca y prestigio. Dada la importancia en el sector de esta última, una nueva empresa que entrase a competir necesitaría realizar un gran esfuerzo inversor para que su marca fuese reconocida. Sin embargo, esta barrera puede no ser importante para entidades de sectores afines a la actividad aseguradora, como por ejemplo las entidades bancarias, que ya tienen un prestigio y una imagen de marca en actividades similares.

En cuanto a las economías de alcance, no parecen existir ventajas en la diversificación del negocio y en la producción de seguros de diferentes ramos, posibilidad, que por otra parte, en algunos casos esta limitada por la legislación (como ocurre con los seguros de vida). Conviene señalar que las ventajas de diversificar si pueden ser importantes para entidades procedentes de otros sectores financieros que, aprovechando sus recursos ociosos y la gran cantidad de información de los clientes, se dediquen a la producción de seguros, sobre todo para los ramos personales. En este caso, hay una importante barrera de entrada legislativa en cuanto que las entidades del sector sólo pueden realizar operaciones de seguros y las directamente relacionadas con ellas. Por tanto, las entidades de otros sectores deben crear entidades dedicadas exclusivamente al negocio de seguros, o bien distribuir seguros de otras entidades; sólo en el primer caso la competencia aumenta.

- *La diferenciación de los productos.* Puesto que el seguro es un producto en el que el comprador no conoce a ciencia cierta su calidad

hasta que se "usa" (se produce el siniestro), el prestigio y la imagen de marca, conseguidos a través de la atención al cliente y la gestión eficaz de siniestros, son fundamentales para la adquisición del mismo. Por este motivo, la buena imagen necesaria para operar en el sector puede considerarse como una barrera de entrada importante. No obstante, es necesario en este caso separar los diferentes ramos.

Para los seguros de vida, dada su complejidad, los usuarios demandan más asesoramiento y más atención a la hora de contratar el producto, sobre todo a medida que aumenta el nivel de renta y la cultura de los mismos. La inversión en la formación del personal que presta el servicio es fundamental y el coste que ello supone es una barrera de entrada importante para competir en este ramo.

Por su parte, en el caso del automóvil, la elección se basa más en el precio que en otros casos, ya que las pólizas de seguros del automóvil no son percibidas como un producto que reclama información detallada para su comprensión, y la capacidad de diferenciarlos es por tanto, pequeña. Las barreras de entrada en este ramo son menores.

En el caso de los seguros de salud, la atención al paciente es fundamental para la diferenciación y la competencia en precio es menor al ser un servicio delicado relacionado con la salud. Por este motivo, la imagen y el prestigio vuelven a disuadir de entrar a los potenciales competidores de este ramo.

- *Necesidades financieras.* Los requisitos de capital y de solvencia ya mencionados, establecidos por ley, son una barrera para el ejercicio de la actividad. Esa es la razón por la que las nuevas empresas que pretendan acceder al sector deben contar con suficientes recursos financieros para cumplir con dichos requisitos.
- *Costes de cambio.* Los costes de cambio son importantes sobre todo en el caso de los seguros de vida, en los que los costes fiscales surgidos al cambiar de seguro suelen ser elevados desde el punto de

vista económico. En otros casos, como para el seguro del automóvil, el cambio no supone coste (más aún cuando se tiene en cuenta la baja siniestralidad que el cliente haya tenido incluso en otra compañía) y la fidelización a la marca es difícil de conseguir.

También hay costes de cambio para el asegurado en el ramo de los seguros de salud como consecuencia del vínculo afectivo que se puede crear con el personal sanitario. Sin embargo, estos costes de cambio son menores de lo que pudiera parecer, pues la mayoría de las compañías aseguradoras subcontratan los servicios sanitarios con los mismos centros médicos, y por tanto, el asegurado puede cambiar de entidad aseguradora sin dejar de ser atendido por el mismo personal.

- *Acceso a los canales de distribución.* La distribución del seguro por los canales tradicionales, mediadores u oficinas propias, requiere un esfuerzo importante por parte del nuevo competidor que pretende entrar en el sector: la instalación de oficinas propias es costosa y en general los competidores instalados tienen ubicaciones mejores; la captación de agentes o corredores debidamente preparados supone hacer un esfuerzo de motivación en cuanto que ya están trabajando con otras entidades. Los nuevos canales de distribución directa sí facilitan el acceso al sector en determinados ramos, como el del automóvil, pero no en otros ramos en los que se requiere un seguimiento del seguro más exhaustivo.
- *Desventajas en otros costes.* Las ventajas de localización de las empresas ya instaladas pueden ser una barrera importante para las que pretenden acceder al sector, en cuanto que la contratación de un seguro requiere la presencia del consumidor para su prestación. Por tanto, el desarrollo de los nuevos canales de distribución elimina, en parte, esta limitación, reduciendo dichas ventajas de las empresas instaladas.

En el caso del sector asegurador, la Curva de Experiencia tiene gran importancia, pues las entidades que adquieren contratos en primer lugar tienen, al aprender en la práctica, ciertas ventajas en el establecimiento de primas y precios (Williamson, 1991: 52), ya que conocen mejor el comportamiento de los asegurados (cuales son responsables y cuáles no, causando éstos más siniestros).

La imagen ya consolidada por las empresas existentes y los elevados gastos de publicidad de algunos ramos son una barrera de entrada a tener en cuenta (Balladín, 1985: 256), sobre todo en los ramos personales del seguro.

- *Política gubernamental.* Algunas de las anteriores barreras de entrada son precisamente consecuencia de la política gubernamental y de los requisitos legales de ejercicio de la actividad, por lo que no es necesario que se vuelvan a referir.

— *Reacción de los competidores instalados.* En el caso del seguro de vida, la entrada de nuevas compañías en el ramo, sobre todo como consecuencia de la bancaseguros y de los competidores extranjeros, ha traído como consecuencia un gran esfuerzo publicitario por parte de las ya existentes, con el fin de reforzar su imagen y fidelizar a los clientes. Este comportamiento de las entidades del sector puede disuadir las tentativas de entrada al sector de nuevas entidades.

Sin embargo, el ejemplo más claro de reacción de las empresas existentes ante los nuevos competidores es el ramo del automóvil: la entrada de nuevas empresas, que emplean los canales directos de distribución, ha traído consigo una gran competencia en precios. Mientras las empresas instaladas, debido a su capacidad, han reducido sus márgenes sin soportar pérdidas, las nuevas siguen teniendo resultados negativos (Dot Pallerés, 1999), de forma que, si estas persisten, la supervivencia de estas nuevos competidores puede verse amenazada seriamente.

Merece especial atención la amenaza que supone para el sector los nuevos competidores procedentes del sector financiero, bancos y cajas sobre todo, debido a su importancia en los últimos años. Este fenómeno, conocido como bancaseguros, está cambiando la posición de las compañías de seguros e intermediarios tradicionales, presentando gran relevancia en el caso de los seguros de vida, pues comercializa el 72% del volumen de primas del ramo. Sin embargo, dado que los seguros no vida presentan problemas específicos que los bancos en muchos casos no están preparados para resolver (regularización de siniestros, indemnizaciones, etc.), este tipo de seguros no ha sido comercializado masivamente por estas entidades y se ha reducido a los seguros de salud, automóviles y multirriesgos del hogar.

La forma en que las entidades bancarias y cajas irrumpen en el sector son las siguientes: creación de una entidad propia aseguradora, aprovechando la amplia cobertura geográfica de sus sucursales y oficinas; comprando o fusionándose con una aseguradora y reorientando la distribución a través de la red de sucursales bancarias; o creando una joint-venture con una aseguradora ya existente, con la ventaja de no tener que realizar grandes inversiones en términos de tiempo, capital y conocimientos, aprovechando la experiencia que aporta la entidad aseguradora. Como consecuencia de las dos últimas formas de acceso, la competencia en el sector no se ve incrementada, sino que se puede considerar como un nuevo canal de distribución, por lo que al considerar los competidores potenciales sólo se hace referencia al primer caso.

El fenómeno de la bancaseguros supone el ejemplo más claro de entrada en el sector asegurador de otro tipo de entidades, dada la facilidad de eludir algunas de las barreras ya comentadas:

- Al comercializar productos de seguros, los bancos y cajas aprovechan la capacidad instalada, en lo referente a redes informáticas y oficinas, obteniendo ventajas en cuanto a economías de escala y de alcance. El coste de distribución del seguro disminuye al aprovechar la red ya instalada.

- Las entidades financieras, bancos y cajas, aprovechan su imagen corporativa y con ella respaldan los productos de seguros que comercializan.
- Las necesidades financieras requeridas para operar en el ámbito asegurador también son un requisito que bancos y cajas cumplen, por lo que no constituyen una barrera de entrada.
- Por otra parte, la fidelidad de la clientela es mayor en el ámbito bancario que en el asegurador, con la ventaja añadida, para este tipo de entidades, de disponer de sus propias bases de datos sobre sus clientes, con amplia información relativa a éstos, pudiendo ofertar así productos de seguros de manera más fácil y acertada, según las características del cliente.
- Mientras que el mercado asegurador se encuentra aún en expansión, sobre todo en el ramo vida, el mercado bancario se considera un mercado bastante saturado, lo que ha incrementado el interés de este tipo de entidades por el negocio asegurador. Además, la venta de seguros de vida no precisa cualificación técnica diferente de la que generalmente poseen los empleados bancarios, por lo que frente a los bancos y cajas, las entidades aseguradoras tienen la ventaja de una mayor especialización, pudiendo ofrecer un mejor servicio postventa.

RIVALIDAD DE LOS COMPETIDORES EXISTENTES

La intensidad de la competencia en un sector viene determinada por el grado de rivalidad existente entre las empresas actuantes en un momento determinado. Los factores determinantes de dicha competitividad en caso del sector asegurador son los siguientes:

— *Número de competidores y equilibrio entre ellos.* Aunque como consecuencia de los procesos de racionalización y liberalización del sector, el número de compañías registradas en la Dirección General de Seguros

descendió casi un 46% en los últimos 20 años, el número de operadores en el mercado aún es muy numeroso (González de Frutos, 1999). Para adaptar el sector a la realidad europea, se recomienda una drástica reducción del número de compañías, lo que favorece la entrada de nuevos competidores, aumentando la competencia existente y poniendo en peligro la supervivencia de los actuales.

Además de estudiar el número de competidores, es necesario analizar cómo éstos se reparten el mercado, es decir, el grado de concentración del sector (Martínez Rego, 1991: 232-6). Para ello, se va a emplear el índice de concentración de Gini¹⁰. Este coeficiente expresa concentración máxima si toma el valor uno (habría una empresa que acumularía todas las primas y con una cuota del mercado del 100%), y un reparto equitativo de la cuota de mercado si toma el valor cero (todas las entidades aseguradoras tendrían la misma cuota de mercado).

Se han considerado, en las siguientes tablas los datos correspondientes a la cuota de mercado de la primera, de las cinco, diez, veinte y cincuenta empresas de mayor volumen de primas¹¹, desde el año 1995 al año 1998:

Tabla III.5
Cuota de mercado de las empresas de seguros en 1995

Primeras compañías	Cuota de mercado (%)
La primera	4,82
Las 5 primeras	18,95
Las 10 primeras	30,58
Las 20 primeras	48,43
Las 30 primeras	60,99
Las 50 primeras	77,62
Las 419 en total	100

¹⁰ Las empresas se ordenan de menor a mayor cuota de mercado. Los p_i expresan el porcentaje de empresas que presentan cierta cuota de mercado. Los q_i expresan el porcentaje de cuota de mercado que corresponde al p_i . El índice de Gini toma la siguiente expresión:

$$I_G = \frac{\sum_{i=1}^{n-1} (p_i - q_i)}{\sum_{i=1}^{n-1} p_i}$$

¹¹ El número de empresas que las publicaciones del sector consideran para estudiar la concentración son la primera, las cinco, diez, veinte, treinta y cincuenta primeras empresas por volumen de primas.

Tabla III.6
Cuota de mercado de las empresas de seguros en 1996

Primeras compañías	Cuota de mercado (%)
La primera	4,48
Las 5 primeras	19,25
Las 10 primeras	31,22
Las 20 primeras	50,32
Las 30 primeras	62,82
Las 50 primeras	79,65
Las 398 en total	100

Tabla III.7
Cuota de mercado de las empresas de seguros en 1997

Primeras compañías	Cuota de mercado (%)
La primera	4,12
Las 5 primeras	17,85
Las 10 primeras	30,75
Las 20 primeras	50,66
Las 30 primeras	64,3
Las 50 primeras	81,12
Las 392 en total	100

Tabla III.8
Cuota de mercado de las empresas de seguros en 1998

Primeras compañías	Cuota de mercado (%)
La primera	3,76
Las 5 primeras	15,57
Las 10 primeras	28,2
Las 20 primeras	48,62
Las 30 primeras	64,67
Las 50 primeras	82,64
Las 378 en total	100

Con estos datos, se ha calculado dicho índice para cada año y los resultados obtenidos son los de la tabla que se presenta a continuación:

Tabla III.9
Valores del índice de concentración de Gini

Años	Valor del índice
1995	0,373
1996	0,383
1997	0,256
1998	0,373

Este índice nos da una idea general del sector, considerando el conjunto de empresas, y de la tabla se deduce que la concentración del sector es baja, es decir, se trata de un sector bastante fragmentado, y por ello, muy vulnerable y competitivo. Además, en ninguno de los ramos las cuotas máximas superan los diez puntos porcentuales (Mercado de riesgos, 1999: 17). El índice va incrementándose como consecuencia de los procesos de fusión y adquisición antes mencionados, excepto para 1997, año en el que se rompe dicha tendencia coyunturalmente.

Considerando el último año para el que se dispone de información (tabla III.8), la primera empresa representa el 3,76% del mercado, mientras que las cinco y diez primeras empresas representan el 15,57% y el 28,2% del negocio total del sector, respectivamente. Las cincuenta primeras empresas (que suponen aproximadamente el 15% de las empresas del sector) agrupan en torno al 80% de la cuota de mercado para los años 1995, 1996, 1997 y 1998. Por tanto, el sector puede considerarse concentrado y algunos ramos, especialmente¹². Es más, en algunos casos, los diez primeros grupos aseguradores acaparan el 50% de la cuota de mercado y el resto se lo reparte un gran número de pequeñas entidades. Aunque la opinión generalizada es que la reestructuración del sector pasa por conseguir un mayor grado de concentración, las fusiones sólo son aconsejables si con ellas se obtienen beneficios adicionales ya que ser más grande no es la única manera de tener éxito (Actualidad Aseguradora, 1999b). Por otra parte, la dimensión de la compañía no sería signo de debilidad si todas las pequeñas compañías fueran solventes, rentables y estuvieran situadas en adecuados nichos de mercado (Barquero Cabrero y Huertas Colomina, 1997: 419).

¹² Tal y como indica M. Oliva Fures, M. Carles Lavilla (1999: 18), el mercado de seguros de salud sí está concentrado.

— Ritmo de crecimiento del sector. La liberalización del sector asegurador español, además de su reestructuración y de la reducción del número de compañías, trajo también consigo un crecimiento muy rápido del sector: en el período 78-90 se consiguieron tasas de un 12,4% anual por término medio. Para el periodo 91-98 se obtiene una tasa de crecimiento medio anual del 12,8%, superior al crecimiento del PIB, si bien, ha sido menor en los últimos años del período considerado (tabla III.1: 333). Esto trae como consecuencia un progresivo incremento de la importancia del sector en la economía, aunque los déficit estructurales de aseguramiento en el caso español aún son importantes, tal y como queda de manifiesto al comparar con los de países del entorno. Por tanto, el mercado español de seguros aún está lejos de la saturación.

Particularizando para los ramos señalados anteriormente, se puede decir que el crecimiento del ramo de vida, excepto para el año 95, es bastante superior al del resto de los ramos, lo que trae como consecuencia un incremento de su importancia en el sector. Por otra parte, con el envejecimiento de la población y el incremento de jubilados este ramo tendrá grandes oportunidades de crecimiento en la próxima década, aunque esto también puede constituir una amenaza para aquellas compañías incapaces de evaluar adecuadamente el riesgo de las rentas de supervivencia.

El mayor componente de los seguros no vida, los seguros de automóviles, ha ido perdiendo importancia debido al abaratamiento de los precios de las primas. La caída de los precios y la competencia sigue siendo fuerte, y los márgenes en este ramo son muy estrechos, lo que sería indicativo de la saturación del mismo.

Los seguros de salud o enfermedad, formados en buena parte por las pólizas de Asistencia Sanitaria, presentan un fuerte crecimiento, aunque este tipo de seguro está sometido a una fuerte competencia desde la vertiente pública que es necesario tener en cuenta.

Del resto de ramos no vida, el ramo de multirriesgos del hogar, tiene gran potencial de crecimiento, según todas las fuentes, destacando la evolución favorable del ramo de responsabilidad civil.

— *Barreras de salida.* Como ya indicamos, la existencia de barreras importantes que impidan o dificulten la salida de las empresas del sector, provoca un incremento de la competencia. En el caso que nos ocupa, vamos a analizar diferentes aspectos que pueden reflejar la existencia de estas barreras:

- En cuanto a los activos específicos, en este sector podría considerarse la imagen de marca lograda que, en el caso de las entidades aseguradoras, tiene especial importancia, si bien, su valor al salir de la industria podría ser recuperable en actividades afines.
- Respecto a los costes fijos de salida, podemos considerar aquellos derivados de los compromisos adquiridos a largo plazo como consecuencia de los distintos contratos firmados, sobre todo en los casos de entidades dedicadas a los seguros de vida, cuyos contratos crean una vinculación a largo plazo con los asegurados. Asimismo, la retribución del personal supone un coste a tener en cuenta a la hora de salir de determinado sector, más en el caso de aquellas entidades aseguradoras cuyo principal canal de distribución sean las oficinas y el personal propios.
- Las interrelaciones estratégicas entre las diferentes áreas de negocio o ramos de actividad, obligan a las entidades a seguir en todos ellos para no perjudicar la imagen global de la empresa que, en el caso de las entidades aseguradoras, tiene gran transcendencia.
- Las barreras emocionales, en lo relativo a la vinculación del personal con la entidad, que serán especialmente relevantes para entidades con sistemas tradicionales de distribución (mediadores y oficinas propias), en los que el esfuerzo realizado para fidelizar los recursos humanos, es muy elevado, dado que de ellos depende la imagen de la empresa.

- Las restricciones gubernamentales para abandonar la actividad deben tenerse en cuenta, pues se prevé la intervención del correspondiente órgano de control cuando se estime conveniente para salvaguardar los intereses de los asegurados o de otras entidades aseguradoras.

— *Estructura de costes de la empresa.* Para las empresas aseguradoras los costes del ejercicio de la actividad o gastos de *gestión* se clasifican en *gastos internos* y *gastos externos*. Como ya se ha indicado, los primeros son los gastos del ejercicio destinados a la administración general de la entidad, y en ellos se incluye el importe bruto de las remuneraciones anuales, la seguridad social y las dotaciones a la amortización e impuestos no repercutibles, entre otros. Dentro de los externos se incluyen los gastos del ejercicio destinados a la producción o mejora de la imagen de la entidad, así como las comisiones, los gastos de agencia, de publicidad, de propaganda y de relaciones públicas. Estos costes pueden considerarse fijos en cierto sentido, y son elevados en el caso de los ramos no vida, especialmente, en el ramo del automóvil, representando un alto porcentaje sobre las primas en el caso de los seguros de salud.

Por otra parte, excepto para el ramo de la salud, el resultado técnico era negativo, es decir, los gastos propios de la actividad no pueden ser cubiertos con los ingresos, teniéndose que cubrir esas diferencias con otro tipo de ingresos que, en la mayoría de los casos, es el financiero. Como consecuencia, en un contexto de tipos de interés decrecientes, los resultados del sector están bastante comprometidos y se refuerza la necesidad de que las compañías se apoyen en sus resultados técnicos para alcanzar un equilibrio sostenido a largo plazo, lo que podría aumentar la competencia en todo el sector.

— *Diferenciación de productos.* Teniendo en cuenta que el seguro es un producto en el que el comprador no conoce a ciencia cierta su calidad, el prestigio y la imagen de marca, conseguidos a través de la atención al cliente

y la gestión eficaz de siniestros, son fundamentales para la adquisición del mismo. La imagen conseguida permite diferenciar los productos y reducir la competencia en el sector.

Es necesario matizar que para algunos ramos, como es el del automóvil, la elección se basa más en el precio que en otros casos, ya que las pólizas de seguros del automóvil no son percibidas como un producto que reclama información detallada para su comprensión, y por tanto, la competencia es mayor.

— *Costes de cambio.* Los costes de cambio se hacen patentes, sobre todo, en el seguro de vida, en el que los costes fiscales, al cambiar de un seguro a otro, son bastante significativos, por lo que los clientes son reticentes a cambiar de asegurador. En otros casos, como el seguro del automóvil, el cambio no supone coste (de hecho las compañías tienen en cuenta la baja siniestralidad que el cliente haya tenido en otras compañías) y la fidelización a la marca es difícil de conseguir, lo que intensifica la competencia. En general, para el resto de los seguros no vida, los costes de cambio son bajos y como consecuencia la competencia se ve intensificada.

— *Diversidad de competidores.* El mercado español siempre ha sido muy atractivo para el capital extranjero. La existencia de competidores con formas de hacer diferentes incrementa considerablemente la competencia, así como la fragmentación del mercado, debido a la dificultad de llegar a acuerdos.

— *Intereses estratégicos.* La necesidad de abrirse hueco en el mercado lleva a ciertas compañías a políticas de precios bajos a pesar de las pérdidas. Este hecho es general en el ramo del automóvil, sobre todo, en compañías que acaban de instalarse y que comercializan seguros telefónicamente. Los bajos precios han incrementado la competencia de manera notable y han provocado una crisis del ramo en los últimos años.

PRODUCTOS SUSTITUTIVOS Y COMPLEMENTARIOS

Tal y como considera Barquero Cabrero y Huertas Colomina (1997: 424), en seguros los productos sustitutivos son el autoseguro (asumir los riesgos propios) y las compañías cautivas (formadas para asumir los riesgos del grupo al que pertenecen). La cobertura de riesgos mediante esta forma representa una amenaza potencial ya que los grandes grupos pueden establecer o comprar compañías de seguros para cubrir sus propios riesgos. La libre prestación de servicios puede animar a los grupos multinacionales a establecer compañías cautivas, obteniendo así unas primas más bajas, coberturas más amplias y los correspondientes ingresos financieros.

En cuanto al autoseguro, el problema que se plantea es invertir correctamente las provisiones realizadas para que la cobertura del seguro sea efectiva. Debido en parte a dicha dificultad, la forma de autoseguro empleada por las empresas para atender los compromisos de pensiones de sus empleados, se ha limitado desde la Administración en el sentido de que las últimas regulaciones legislativas del sector exigen a las empresas (excepto a las financieras) la exteriorización de sus compromisos por pensiones. En concreto, el reglamento de 27 de octubre de 1999 obliga a las empresas a que esa exteriorización se realice antes del 1 de enero del 2001, lo que impulsará el ramo de vida.

En el ramo de vida, el seguro y los planes de pensiones están vinculados por ser instrumentos de previsión social (Pérez Torres, 1999). Por tanto, dichos productos pueden considerarse como sustitutivos y la regulación de uno de esos productos influye en el otro (García de Quevedo Ruíz, 1999). La nueva ley del Impuesto sobre la Renta de las Personas Físicas mejora el tratamiento fiscal del seguro de vida, que antes estaba en clara desventaja respecto a los planes de pensiones, lo que puede afectar de manera favorable al mismo.

Para el caso de los seguros de automóvil, en el mercado europeo los nuevos métodos de distribución (venta por correo, telefónica, etc.)

combinados con una mayor desregulación, erosionarán las diferencias de precios entre países e incrementarán la competencia, y para este tipo de seguro, la posibilidad de adquirirlo en diferentes países no debe ser ignorada. Estos seguros adquiridos en el exterior podrían considerarse como productos sustitutivos.

En cuanto a los seguros de salud, el producto sustitutivo más claro es la seguridad social pública: las actuaciones de la Administración para mejorar ésta pueden afectar a la contratación de este tipo de seguros y, por tanto, impulsar a los aseguradores a implantar políticas constantes de ampliación y mejora de servicios.

PODER DE LOS CLIENTES Y PROVEEDORES

En general, los clientes particulares no tienen posibilidad de integración vertical dado que son muchos y muy dispersos. La integración, en este caso, es el autoaseguramiento de los individuos, que no pueden a bajo coste revelar la condición de aseguradores (Williamson, 1991: 110).

En cuanto a los clientes corporativos, sí existe esa posibilidad, pero no es sencillo dados los requisitos de acceso a la actividad. Pero, tal y como se ha indicado, las empresas cautivas pueden considerarse una forma de integración vertical y éstas si pueden plantear cierta amenaza para el sector, aunque no fuerte.

Por otra parte, los clientes cada vez tienen más información, y por tanto, su poder de negociación aumenta, si bien, este poder se ve limitado debido a los contratos de adhesión entre asegurador y asegurado que se establecen para la mayoría de los ramos del seguro.

Los proveedores en este sector no son lo que tradicionalmente se entiende como tal, aunque pueden tener dicha condición los suministradores de tecnología y de recursos humanos. Dada la importancia de las inversiones en tecnología, éstos sí que pueden tener cierto poder sobre el sector, aunque

no demasiado, dada la competencia existente en el sector tecnológico. En cuanto a los segundos, el poder negociador será mayor para aquellas empresas aseguradoras que tengan un sistema tradicional de distribución basado en una red de agentes, pues la imagen de la empresa depende de ellos y, por tanto, su poder negociador también será importante.

Teniendo en cuenta los factores analizados para cada una de las cinco fuerzas competitivas, a modo de resumen, las siguientes tablas tratan de sintetizar la influencia que dichas fuerzas pueden tener sobre la capacidad de generar beneficios del sector asegurador, considerando en particular los ramos de seguros de vida, automóviles y salud:

Tabla III.10
Ramo de seguros de vida

SEGUROS DE VIDA	
LAS 5 FUERZAS	AMENAZA SOBRE LOS BENEFICIOS DEL SECTOR
COMPETIDORES POTENCIALES	Alta
RIVALIDAD EXISTENTE	Media
PRODUCTOS SUSTITUTIVOS	Baja
PODER DE CLIENTES Y PROVEEDORES	Baja

Tabla III.11
Ramo de seguros de automóviles

SEGUROS DE AUTOMÓVILES	
LAS 5 FUERZAS	AMENAZA SOBRE LOS BENEFICIOS DEL SECTOR
COMPETIDORES POTENCIALES	Baja
RIVALIDAD EXISTENTE	Alta
PRODUCTOS SUSTITUTIVOS	Baja/Media
PODER DE CLIENTES Y PROVEEDORES	Baja

Tabla III.12
Ramo de seguros de salud

SEGUROS DE SALUD	
LAS 5 FUERZAS	AMENAZA SOBRE LOS BENEFICIOS DEL SECTOR
COMPETIDORES POTENCIALES	Baja
RIVALIDAD EXISTENTE	Media
PRODUCTOS SUSTITUTIVOS	Alta
PODER DE CLIENTES Y PROVEEDORES	Baja

Tabla III.13
El sector asegurador

SECTOR ASEGURADOR	
LAS 5 FUERZAS	AMENAZA SOBRE LOS BENEFICIOS DEL SECTOR
COMPETIDORES POTENCIALES	Baja/Media
RIVALIDAD EXISTENTE	Media/Alta
PRODUCTOS SUSTITUTIVOS	Baja
PODER DE CLIENTES Y PROVEEDORES	Baja

Con el fin de hacer frente a las fuerzas del mercado, cada entidad aseguradora adoptará una estrategia diferente, de forma que las distintas estrategias dan lugar a diferentes **grupos estratégicos**. Para la identificación de los mismos, se han desarrollado dos metodologías: la primera consiste en reconocer los grupos a priori, basándose en el conocimiento del sector que tenga el investigador o directivo. Se consideran ciertas variables de cada grupo y se confirman o rechazan, con posterioridad, las hipótesis planteadas. La segunda consiste en determinar los grupos desde un conjunto de variables, más o menos numerosas, que definen las dimensiones estratégicas. Después, mediante técnicas estadísticas, se reconocen las empresas con estrategias de actuación similares, describiendo el comportamiento de los grupos una vez éstos identificados. Este planteamiento es el que se va a emplear para la identificación de grupos estratégicos en la parte empírica del presente estudio.

Hacemos referencia a continuación a algunos de los trabajos que analizan el sector asegurador y los grupos estratégicos, dado que las conclusiones obtenidas pueden servirnos de referencia en la segunda parte de este capítulo. Los trabajos de Fiegenbaum y Thomas (1990; 1995) analizan el comportamiento de las entidades aseguradoras estadounidenses, identificando un grupo de empresas cuya estrategia comercial se basa en la diversificación (empresas diversificadas), otro grupo que concentra su actividad en el

seguro de vida (estrategia vida), y otro que se orienta a la venta de pólizas de personas (estrategia personal).

El estudio de Martínez Martínez (1993) relativo al sector asegurador español parte de la descripción de posibles estrategias a priori. Después, apoyándose en las opiniones de altos directivos del sector, identifica las empresas que más se ajustan a cada una de esas posibles estrategias y que, mediante alguna técnica de clasificación, pueden identificarse como grupo estratégico. Se distingue básicamente entre empresas que persiguen una estrategia de diversificación (que concentra un mix de ramos) y empresas que persiguen una estrategia de especialización (que se concentra en el ramo de automóviles).

El trabajo de Pinillos Acosta y Martín Peña (1998) trata de identificar los grupos estratégicos del sector asegurador español para el año 1991 y para el año 1995. Para cada año, estos autores consideran las variables que Fiegenbaum y Thomas consideran relevantes, desde el punto de vista estratégico, en el sector del seguro, analizándolas mediante un Método Factorial que permite obtener las dimensiones relevantes (los factores) para clasificar las empresas en grupos estratégicos. La aplicación del *Análisis Cluster* a los factores obtenidos permite obtener una tipología de empresas para cada año. Una vez comparadas y uniformadas las dos clasificaciones, se obtienen cuatro grupos estratégicos que pueden considerarse estables en el período: empresas dedicadas a los seguros de vida; empresas especializadas en seguros de vida que distribuyen sus productos a través de oficinas bancarias y de cajas de ahorro; empresas orientadas a los seguros no vida; y otras empresas de seguros.

Según la teoría que considera a la empresa como un conjunto de recursos heterogéneos, los grupos estratégicos identificados se explican por las diferencias en los recursos y capacidades de cada empresa aseguradora. El

análisis de los mismos en cada entidad, junto con el Análisis Estructural del sector, permitiría formular la estrategia adecuada (Ventura Victoria, 1994: 275). Parece que, en el sector asegurador, el dominio de las técnicas de transmisión de datos a distancia (de pagos automáticos y de transacciones electrónicas de fondos) y la tecnología de los procesos de producción del seguro (contratación, gestión de siniestros, distribución, contabilidad e inversiones) pueden convertirse en capacidades distintivas que conduzcan a una ventaja competitiva sostenible a largo plazo (Pérez Barbero, 1999: 121).

2. Análisis empírico

2.1 Introducción

En esta segunda parte del capítulo y con el fin profundizar en el conocimiento del sector asegurador, se analiza cierta información relativa al mismo mediante la metodología *STATIS*, la cual se aplicará en función de los fines, descriptivos o predictivos, que se persiguen conseguir. Las conclusiones obtenidas permiten además demostrar, de manera empírica, las posibilidades de la metodología empleada que ya se han expuesto teóricamente.

Para completar la descripción del sector, la información a considerar debe hacer referencia a una serie de variables relativas a las empresas que lo integran, que representen la mayor cuota de mercado posible, y que configuran la estructura básica del mismo. Las variables seleccionadas deben permitir agrupar las diferentes empresas, de tal manera que sea posible caracterizar fácilmente cada uno de los grupos obtenidos. Por otra parte, parece conveniente considerar diferentes años con el fin de poner en evidencia si los cambios en la estructura del sector, y por tanto de algunas entidades, son importantes o no a corto plazo.

Sin embargo, si lo que se desea es conocer la evolución del sector deben tomarse variables, de forma agregada para todas las empresas que lo conforman, que expliquen el comportamiento del mismo a más largo plazo y permitan predecir sus características en el futuro. Además, y teniendo en

cuenta los rasgos específicos de los principales ramos que ya se han señalado, parece razonable considerar dichas variables de manera particular para los más representativos.

Estas razones inducen a considerar información diferente según los objetivos que se desean alcanzar.

Antes de comenzar con el análisis empírico propiamente dicho, haremos una breve referencia a los estudios que se han realizado sobre el sector financiero, considerando especialmente el ámbito asegurador, a través de métodos estadísticos. Con ello queremos dejar patente la utilidad e importancia de la metodología propuesta para el análisis de uno de los sectores más relevantes de la economía, además de contribuir al avance de la Estadística como ciencia imprescindible para el análisis económico-financiero.

— Hasta los años cincuenta, los estudios de tipo financiero eran básicamente descriptivos y el objeto de los mismos era informar de una manera general de los instrumentos e instituciones financieras. El empleo de métodos estadísticos sólo viene siendo una práctica habitual en los últimos cuarenta años y diversos son los enfoques que han sido adoptados:

Los *Modelos de Regresión* y los *Modelos Econométricos* se encuentran prácticamente en todos los sectores del análisis financiero. Se pueden citar como ejemplo los trabajos basados en las publicaciones de Markowitz, de 1959, para la selección de carteras y sobre los que se ha constituido un campo de reflexión teórica interesante para la investigación estadística.

El *Análisis Multivariante* se ha convertido en una herramienta cada vez más utilizada en numerosos estudios exploratorios o descriptivos. Pueden citarse los trabajos de Herbst, de 1974, que emplea a la vez un *Análisis Factorial* y un modelo de *Regresión Múltiple*. El *Análisis Discriminante* y los *Modelos Logit* también son empleados de forma habitual.

Los *Modelos de Series Cronológicas* y sobre todo los *modelos ARIMA*, tienen un campo importante de aplicación de la Estadística en finanzas. Desde mediados de los setenta, muchos autores han recurrido a la metodología Box-Jenkins para estimar o prever ganancias, pérdidas y tasas, destacando ahora las tentativas de aplicación del *Análisis Espectral* a las tasas de interés o para chequear la eficacia de un mercado.

La *Teoría de la Decisión*, siguiendo una aproximación bayesiana en su desarrollo, es empleada sobre todo a la hora de seleccionar las distintas carteras.

Como trabajos más recientes, pueden citarse los de Cabedo Semper y Moya Clemente (1999), que combinan las técnicas de *Análisis Multivariante* con modelos autorregresivos, en concreto ARCH, para reducir las dimensiones del riesgo a una serie de factores con una mínima pérdida de información, permitiendo la formulación de un modelo predictivo sin considerar un número excesivo de parámetros; y el de Lambert y Lindsey (1999), que emplea modelos de regresión para el análisis de rendimientos financieros.

— En el ámbito del sector asegurador, dejando a un lado los métodos estadísticos para la organización de la actividad aseguradora (como la *Estadística Actuarial* y *Biometría*), técnicas descriptivas y de predicción se aplican de manera habitual para el análisis del sector. A modo de ejemplo pueden citarse los artículos Lecea Flores de Lemus (1991) en el que se estudia la organización del sector empleando índices estadísticos de concentración; los artículos de Lado Couste (1996) y Fiegenbaum y Thomas (1990) en los que se emplea el *Análisis Cluster* y el *Análisis de la Varianza* para la determinación de los grupos estratégicos que conforman el sector. Con el mismo objetivo, el artículo de Pinillos Costa y Martín Peña (1998) emplea el ACP y el *Análisis Discriminante*; la tesis de Martínez Martínez (1993) se apoya en esa última técnica y en el *Análisis Cluster* con el mismo fin. El trabajo de Nora Lado y Rivera Camino (1996) constituye un buen ejemplo de aplicación de técnicas estadísticas, tales como *Análisis de*

Componentes Principales y Análisis Discriminante, para analizar este sector, según la orientación al mercado de las empresas aseguradoras.

Según lo expuesto, la metodología STATIS aplicada al sector asegurador proporciona, a nuestro juicio, un avance en el estudio del mismo, permitiendo, de una forma simultánea, su descripción y su evolución. Asimismo, queda patente la dualidad existente entre el desarrollo estadístico y el económico, a los que con este trabajo hemos querido contribuir.

2.2 Estudio descriptivo del sector asegurador español

La descripción de la información relativa al sector del seguro en España se plantea en términos de obtener una estructura que caracterice el sector a través de los datos analizados, lo que se realiza a través de los métodos *Statis* y *Statis dual*. Posteriormente se determina la tipología de grupos de empresas del sector, es decir, los grupos estratégicos, mediante el *Statis* y el *Análisis Cluster*. Previamente se realiza la descripción de los datos que van a ser objeto de estudio.

2.2.1 Descripción de los datos

Para una adecuada descripción del sector, las fuentes de información que ya han sido citadas en la primera parte de este capítulo, relativas al sector asegurador, fueron revisadas con el fin de seleccionar las series de variables más homogéneas, y para el mayor número posible de entidades aseguradoras. Como consecuencia de este proceso, los datos se seleccionaron de la Revista Ranking; a nuestro juicio, dada la información del sector

disponible, no parece aconsejable el empleo de fuentes de información directas, pues el coste que ello supondría no se vería compensado por la información obtenida, sin olvidar que las entidades aseguradoras son reacias a proporcionar otra información distinta a la tienen la obligación de suministrar.

A continuación se procede a describir de forma detallada los individuos, las variables medidas en ellos y el periodo de tiempo considerado.

Los individuos: las entidades aseguradoras

La población objeto de estudio son las entidades aseguradoras, en cuanto son éstas las que llevan a cabo la actividad aseguradora y conforman el sector. En el caso español, se trata de aquellas entidades que están sometidas al control de la Dirección General de Seguros e inscritas en el Registro de Entidades Aseguradoras. De todos modos, el análisis de todas las que conforman el sector, dado que es un sector fragmentado, no parece aconsejable, y la selección de la muestra se ha realizado teniendo en cuenta las cuotas de mercado de las primeras empresas.

Así, en un primer momento se consideraron las 50 primeras empresas, por volumen de primas en el año 97, último año del que se disponía información en el momento de plantearnos el estudio. La elección de ese número de entidades se justifica por el hecho de que dichas empresas representan una cuota de mercado importante sobre el total del sector, en torno al 80%. Además, las 50 primeras empresas son un punto de referencia empleado frecuentemente en la mayoría de estudios sobre este sector.

Posteriormente, el número de entidades tuvo que quedar limitado a 48, pues fueron eliminadas las entidades que ocupan las posiciones 29 y 34 respecto al volumen de ventas: Royal&Sun Alliance Seguros y National-Nederlanden, Cía.Seg.Vida, N.V.; la razón de esta eliminación es que la primera es una entidad resultado de la fusión de otras dos compañías, y

por ello, los datos relativos a la misma para el año 1997 no son comparables con los de años anteriores; respecto a la segunda, no existe suficiente información disponible como para ser incluida en el estudio.

Una vez depurada la información, los datos finales corresponden a 48 entidades durante tres años 1995, 1996 y 1997. Los procesos de fusión y absorción han modificado la naturaleza de algunas compañías y por ello pocas entidades presentan datos homogéneos y comparables para más años, reduciéndose la cuota de mercado estudiada de manera considerable si nos limitásemos sólo a éstas. Por tanto, en 1997 la cuota de mercado estudiada fue el 78,86%, en el año 1996 el 75,21% y en 1995 el 71,81%.

La tabla siguiente recoge las distintas entidades consideradas en el estudio, ordenadas por orden alfabético, y que constituyen los individuos de nuestro estudio descriptivo.

Tabla III.14
Individuos (Entidades Aseguradoras)

Compañías	Etiquetas
Adeslas Cia. de Seguros	Adeslas
Aegon-Union Aseguradora, S.A. Seguros y Reaseguros	AegonUnionAsegurad
AGF Unión Fenix	AGFUnionFenix
Allianz-Ras Seguros y Reaseguros, S.A.	AllianzRas
Antares, Seguros de Vida y Pensiones	AntaresVida
Argentaria Vida, S.A. de Seguros y Reaseguros	ArgentariaVida
Ascat Vida, S. A. de Seguros y Reaseguros	AscatVida
Aseguradora Valenciana, S.A. Seguros y Reaseguros	AseguradValenciana
ASISA (Asistencia Sanitaria Interprovincial)	Asisa
Assicurazioni Generali, S.P.A. Del. España	AssicurazioniGenerali
Athena, Cia. Ibérica de Seguros y Reaseguros	Athena
Banesto Seguros, Cia. aseguradora	Banesto
Bankinter Seguros de Vida, S.A.	BankinterVida
Bansabadell Vida, S.A.	BansabadellVida
Barclays Vida y Pensiones	BarclaysVida
Biharko Vida y Pensiones	BiharkoVida
Bilbao Cia. Anma. de Seguros y Reaseguros	BilbaoCiaAnma
Caja Madrid Vida, S.A. de Seguros y Reaseguros	CajaMadridVida
Caser Grupo Asegurador	Casergrupo
Catalana Occidente	CatalanaOccidente
Crédito y caución, Cia. española Seguros y Reaseguros	CreditoyCaucion
Db Vida Cia. de Seguros y Reaseguros	DbVida
EuroSeguros, Cia de Seguros y Reaseguros	Euroseguros

Fiatc Mutua de Seguros y Reaseguros prima fija	FiactMutua
Finisterre, S.A. Cia. Seguros y Reaseguros	Finisterre
Intercaser, S.A. de Seguros y Reaseguros	Intercase
La Estrella, S.A. de Seguros y Reaseguros	LaEstrella
Lagun Aro Vida, S.A.	LagunAroVida
Mapfre Mutualidad de Seguros y Reaseguros a prima fija	MapfreMutualidad
Mapfre Seguros Generales, S.A. (Grupo)	MapfreSegGrales
Mapfre Vida, S.A.	MapfreVida
Multinacional Aseguradora, s.a	MultinacionalAsegurad
Musini, S.A. de Seguros y Reaseguros	Musini
Mutua General de Seguros	MutuaGeneralSegur
Mutua Madrileña Automovilística, Sdad Seg.	MutuaMadri
Ocaso, S.A. Cia. de Seguros y Reaseguros	Ocaso
Pelayo Mutua de Seguros y Reaseguros a prima fija	PelayoMutua
Plus Ultra, Cia. Anma. de Seguros y Reaseguros	PlusUltra
Sanitas, S.A.	Sanitas
Santa Lucia, S.A.	StaLucia
Santander Seguros y Reaseguros	Santander
Seguros Génesis, S.A. de Seguros y Reaseguros	SegurosGenesis
Vidacaixa, S.A. Seguros y Reaseguros	Vidacaixa
Vitalicio Seguros	Vitalicio
Winterthur Seguros generales, S.A.	WinterthurSegGrales
Winterthur Vida, S.A., Seguros sobre la Vida	WinterthurVida
Zurich Internacional (España)	ZurichInternacional
Zurich Seguros (sucursal España)	ZurichSegEspana

Las variables

Las variables que han sido consideradas para la caracterización del sector, se han seleccionado teniendo en cuenta las investigaciones y discusiones de ejecutivos y académicos del mundo de los seguros; algunas de ellas han sido ya empleadas en algunos estudios anteriores sobre el tema (Fiegenbaum y Thomas, 1990; Pinillos Costa y Martín Peña, 1998), pero la razón fundamental para su elección ha sido la opinión de expertos.

Con el fin de relacionar los resultados que la empresa aseguradora obtiene, con su estructura organizativa y con su orientación al mercado y al producto, se consideraron ciertas variables indicativas de los resultados obtenidos por las distintas entidades, tales como la cuota de mercado, el equilibrio técnico o la rentabilidad económica. Su eliminación posterior fue motivada por su

baja correlación con los factores obtenidos, no aportando ninguna información al estudio.

Finalmente las variables seleccionadas y medidas sobre las 48 entidades aseguradoras fueron las siguientes, clasificándose en dos grupos:

— **Variables de alcance:** hacen referencia al ámbito del producto, diversificación del mismo y tamaño del mercado. La que hace referencia a éste último son:

- *primas seguros industriales / primas totales* = INDUS

Esta variable indica, para cada empresa, la orientación al mercado, es decir, el porcentaje que de las primas totales corresponde a las pólizas vendidas a las empresas. Los ramos del transporte, incendios, otros daños, responsabilidad civil, asistencia de viajes, otros ramos, crédito y caución, pérdidas pecuniarias diversas, multirriesgo comercio, multirriesgo pymes, otros multirriesgos, se incluirían en este caso. El resto de los ramos se consideran dentro de los seguros personales.

Dado que el cien por cien de las primas sería la suma de las primas personales y las primas industriales, no es necesario incluir en el análisis otra variable (PERSO), que recoge las ramas personales, debido a que su correlación con la variable INDUS sería perfecta y negativa, y por ello su posición sería simétrica respecto a esta variable.

La orientación al producto, es decir, la elección de las líneas de productos, y la diversificación se manifiesta a través de cuatro variables:

- *primas seguros vida / primas totales* = VIDA

Para cada empresa representa el porcentaje que las primas de vida suponen sobre el total de las de la empresa.

- *primas seguros autos / primas totales* = AUTOS

Porcentaje que las primas de seguros de automóviles suponen respecto al total.

- $\text{primas seguros salud} / \text{primas totales} = \text{SALUD}$

Representa el porcentaje que las primas de enfermedad y asistencia sanitaria suponen sobre el total de las de la empresa.

- $\text{primas seguros resto} / \text{primas totales} = \text{RESTO}$

Es el porcentaje que los ramos no incluidos en las anteriores suponen sobre el total de las primas.

También se incluye como variable de alcance, aunque con un carácter distinto, la variable $\text{primas} / \text{empleados} = \text{PRIMEM}$.

Esta ratio resulta de dividir las primas totales entre el número de empleados, y se ha considerado para reflejar el impacto de la bancaseguros. Como es lógico, el número de empleados y el número de oficinas a través de los cuales se distribuye el producto de seguro, será mayor que en el caso de las compañías tradicionales de seguros, al aprovechar la red de sucursales. Según lo expuesto, las ratios primas/número de oficinas, y número de empleados/número de oficinas, estarían también indicadas para medir el impacto de la bancaseguros. Si bien, el dato del número de oficinas no estaba disponible para muchas de las compañías consideradas por lo que las mencionadas ratios no han podido analizarse.

Conviene precisar que en este caso, el número de empleados registrado en la fuente de información empleada es más reducido para las entidades de bancaseguros (debido a la forma de recopilar los datos), y por ello este tipo de entidades presentan un valor más alto en esta variable que el resto, en contra de lo que sería lógico. De cualquier modo, la variable cumple su función a la hora de detectar las entidades de bancaseguros.

— **Variables de desarrollo:** Estas variables informan acerca del uso de recursos y se han considerado las siguientes:

-
- *gastos internos / primas* = GINTPR

Para cada una de las compañías estudiadas, este cociente representa el porcentaje que los gastos internos representan sobre el total de primas. Como ya hemos indicado, los gastos internos son los que tienen lugar en el ámbito de la función administrativa de la entidad aseguradora.

- *gastos externos / primas* = GEXTPR

Es la proporción, en porcentaje, que los gastos externos suponen respecto a las primas. Los gastos externos son los generados como consecuencia de la comercialización de los productos. Con ambos ratios se pretende analizar el peso de los gastos derivados de la actividad propia de las empresas de seguros, y por tanto, la productividad en el uso de los recursos.

- *Margen de solvencia / margen mínimo exigido* = MMMMI

Esta ratio indica el número de veces que la empresa mantiene el margen mínimo de solvencia exigido. La solvencia es una cuestión fundamental en la empresa de seguros al resultar indispensable para que pueda cumplirse su objeto social. De ahí la vigilancia que sobre dicho nivel exigido mantienen las autoridades económicas y la importancia en el estudio del sector.

- *Inversiones financieras / Inversiones totales* = INFINIT

La decisión de inversión en las compañías de seguros constituye una fuente adicional de ingresos y determina una parte de la estrategia de las empresas aseguradoras. Por ello, resulta sumamente interesante estudiar la proporción que las inversiones financieras representan sobre el total.

- *Negocio neto / capitales propios* = NNC

Esta ratio representa la proporción que el negocio neto (volumen de primas más los niveles de reaseguro aceptado y menos el seguro

cedido) supone sobre los capitales propios. El negocio neto manifiesta la adecuación de las decisiones en el desarrollo del negocio asegurador, y los capitales propios, como parte importante de la estructura financiera de la empresa, también son decisión de la empresa.

- $\text{Negocio neto} / (\text{capitales propios} + \text{provisiones técnicas}) = \text{NNCPRO}$

Esta variable complementa a la anterior al recoger las provisiones técnicas, una parte de los cuales son de obligado cumplimiento, y otra, consecuencia de decisión de la empresa.

- $\text{Negocio neto} / \text{primas} = \text{NNPR}$

Como variable de desarrollo, debe estudiarse también la importancia del reaseguro en la actividad de las entidades aseguradoras puesto que un adecuado uso del mismo puede mejorar los resultados de la entidad.

Para analizar éste puede considerarse directamente el porcentaje que representa sobre las primas. El inconveniente de dicha variable es que toma valores negativos en aquellos casos en los que el reaseguro aceptado es menor que el reaseguro cedido, lo que dificulta la aplicación de técnicas factoriales; por ello, se considera en su lugar el porcentaje que el negocio neto representa respecto a las primas. Un porcentaje inferior al 100 por cien indica una mayor importancia del reaseguro cedido sobre el reaseguro aceptado, creciendo dicha importancia a medida que el valor de la ratio disminuye, y por tanto aumentando la necesidad del reaseguro en la entidad como sistema de cobertura de los riesgos asumidos.

La prima media (porcentaje de primas/número de pólizas) también debería considerarse como variable de desarrollo. Como no existen datos disponibles para las compañías consideradas, dicha variable no se ha incluido en el estudio. Una vez más, la homogeneidad de la información disponible condiciona el estudio.

A modo de resumen, la tabla siguiente recoge las variables que intervienen en el análisis y su etiqueta identificativa:

Tabla III.15
VARIABLES ANALIZADAS PARA CADA ENTIDAD ASEGURADORA

VARIABLES	ETIQUETAS
VARIABLES DE ALCANCE	
- primas seguros industriales / primas totales	INDUS
- primas seguros vida / primas totales	VIDA
- primas seguros autos / primas totales	AUTOS
- primas seguros salud / primas totales	SALUD
- primas seguros resto / primas totales	RESTO
- primas / empleados	PRIMEM
VARIABLES DE DESARROLLO	
- gastos internos / primas	GINTPR
- gastos externos / primas	GEXTPR
- Margen de solvencia / margen mínimo exigido	MMMMMI
- Inversiones financieras / Inversiones totales	INFINIT
- Negocio neto / capitales propios	NNC
- Negocio neto / capitales propios + provisiones técnicas	NNCPRO
- Negocio neto / primas	NNPR

Periodo de tiempo

Como se pondrá de manifiesto, la estructura del sector no se modifica de manera significativa a corto plazo, y es posible definir una estructura común para varios años sucesivos, obteniendo conclusiones más relevantes que si solamente se considerasen los datos relativos a un año. Los seleccionados, debido al motivo ya explicado de la homogeneidad de la información disponible para los individuos analizados, se recogen en la tabla siguiente:

Tabla III.16
Años estudiados

Años	Etiquetas	Operadores Asociados
Año 1995	95	W_1D / V_1M
Año 1996	96	W_2D / V_2M
Año 1997	97	W_3D / V_3M

Cuando la variable o el individuo se refiera a un año concreto, se adjuntará a la etiqueta original la correspondiente a dicho año.

Los datos sobre los individuos y las variables para estos años se recogen en las tablas a, b y c del anexo A.2 (pp. 568-70). Cada una de ellas hace referencia a un año y está formada por las 48 entidades aseguradoras, sobre las que se han observado las 13 variables indicadas. Como consecuencia, las 3 tablas están formadas por los mismos individuos y las mismas variables.

Si estas tablas se yuxtaponen, el resultado es una tabla total de datos de 48 filas (entidades-individuos) y 13×3 columnas (variables) que puede tratarse mediante el *Statis* (se analizan las tres nubes de individuos-entidades). Si las tablas se apilan, unas debajo de otras, el resultado es una tabla total, formada por 48×3 filas y 13 columnas, que es la considerada por el *Statis dual* (en este caso se analizan las tres nubes de variables).

2.2.2 Estructura del sector asegurador español: *Statis* y *Statis dual*

Una vez especificada la información, se procede a su análisis desde el punto de vista descriptivo, para conocer la estructura del sector. Atendiendo a los datos y a los objetivos, la metodología correcta será la correspondiente a los dos métodos que constituyen la metodología STATIS, ya que cuando los individuos y las variables son los mismos en las distintas tablas, la aplicación de ambos permite obtener resultados complementarios y conclusiones más completas.

La aplicación de dicha metodología se ha llevado a cabo con el paquete estadístico SPAD (módulo TM versión 3.5) y los resultados obtenidos se complementan con los obtenidos del módulo de Análisis Factorial y del módulo de Análisis Descriptivo de la misma versión del programa.

Para evitar la influencia de las unidades de medida, expresadas unas en porcentajes y otras en tantos unitarios, y de la varianza, que varía considerablemente de unas variables a otras (Anexo 2.1.1: p. 571), se van a centrar y tipificar las variables. Para cada nube de individuos, el operador considerado y analizado por el *Statis* será el $W_t D$ correspondiente. Cada nube de variables estará representada por el operador $V_t M$ (matriz de correlaciones) que será analizado por el *Statis dual*. Tanto los $W_t D$ como los $V_t M$ se van a considerar normados con el fin de equilibrar la influencia de todas las tablas en el análisis.

El desarrollo de este apartado se realiza describiendo, en primer lugar, los principales resultados obtenidos en cada etapa que constituye el *Statis* y el *Statis dual*, y a continuación se especifica la interpretación de los mismos y las conclusiones que se derivan de cada una de dichas etapas, respecto al sector y respecto al método empleado.

a. Resultados obtenidos

Para poder aplicar cualquiera de las versiones de la metodología STATIS es necesario que las nubes de individuos, para el *Statis*, y las nubes de variables, para el *Statis dual*, sean comparables; por ello, el primer paso es el análisis de dichas nubes por separado y la determinación de sus valores propios con el fin de estudiar su dimensión y su forma.

Los valores propios obtenidos se recogen en el anexo A.2.1.2 (pp. 572-4) y de ellos se deduce que las tres nubes de individuos consideradas (matrices W_1, W_2, W_3) tienen la misma dimensión y forma similar, puesto que el número de valores propios no nulos es el mismo y el reparto de dichos valores, dentro de cada nube, es similar. Por tanto, las diferencias entre nubes se deben al cambio de posición de los individuos de unas tablas a otras y tales diferencias se pondrán de manifiesto al aplicar el método *Statis*.

Al observar el epígrafe A.2.2.1 (pp. 586-8), donde se recogen los valores propios de las nubes de variables (matrices de correlación en este caso V_1M , V_2M , V_3M), se pueden hacer las mismas consideraciones. El tratamiento mediante el *Statis dual* permitirá obtener las variables que son responsables de dicha variación.

Conviene señalar que, dado que ambas versiones son un caso especial del ACP, podrían tenerse en cuenta ciertas hipótesis de partida sobre los datos, relativas a la normalidad de las variables. Si bien, el planteamiento reciente es considerar el ACP como una técnica de representación de datos óptima debido a las propiedades geométricas y algebraicas que cumple, que se utiliza sin necesidad de hipótesis generales de naturaleza estadística y sin necesidad de un modelo particular. Así es considerado por Lebart *et al.* (1995: 32) y es el que subyace en muchas de las aplicaciones del método en todos los ámbitos del conocimiento, incluido el económico-financiero.

1. LA INTERESTRUCTURA

Ambas versiones consideran una matriz S , de coeficientes RV entre los operadores representativos de las tablas de datos, analizan su estructura con la obtención de valores y vectores propios, y tratan de obtener una imagen euclídea de dichos operadores con el fin de comparar globalmente las nubes analizadas. Así, en el caso del *Statis* esa imagen representa la norma y los ángulos entre operadores $W_t D$ que se deducen de las nubes de individuos, y en el caso del *Statis dual* la imagen obtenida representa la norma y los ángulos entre operadores $V_t M$, que se deducen de las nubes de variables.

— La matriz S de coeficientes RV del *Statis* y del *Statis dual* se recoge en las tablas III.17a y III.17b:

Tabla III.17a
Matriz S de productos escalares entre $W_t D$ normado

Matr	95	96	97
95	1	0,832	0,834
96	0,832	1	0,927
97	0,834	0,927	1

Tabla III.17b
Matriz S de productos escalares entre $V_t M$ normado

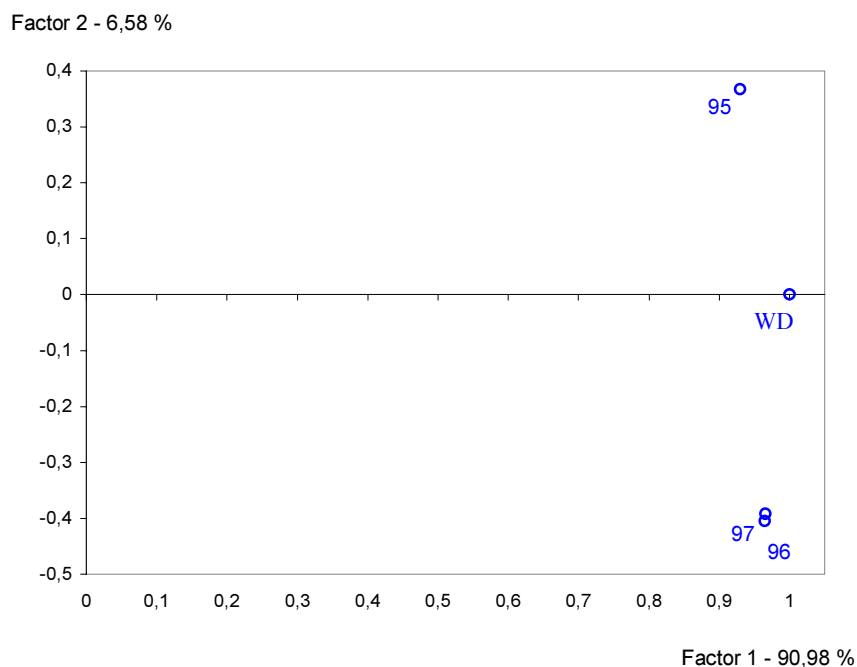
Matr	95	96	97
95	1	0,935	0,936
96	0,935	1	0,993
97	0,936	0,993	1

Respecto a estas matrices, los epígrafes A.2.1.3 (p. 575) y A.2.2.2 (p. 589) recogen la significatividad o no de los productos escalares que estas presentan, valorada a través del test de permutaciones. Un producto escalar suficientemente elevado, que no es debido solo al azar, se señala mediante un asterisco e indica que las tablas están correlacionadas de forma significativa (CISIA, 1997: 27). En este caso, ésta es la situación que se da tanto para las nubes de individuos como para las de variables, y por ello la aplicación del STATIS (la versión correspondiente) es correcta.

De forma breve se pueden comentar las matrices obtenidas. Así, la matriz de la tabla III.17a pone de manifiesto coeficientes RV más altos entre las nubes de los años 96 y 97 (igual a 0,927), que entre éstas y las del 95 (0,832 y 0,834 respectivamente), aunque también se pueden considerar elevados. Lo mismo ocurre con la matriz de la tabla III.17b, a pesar de que los valores son diferentes.

— La imagen euclídea en dos dimensiones de los operadores correspondientes a las diferentes tablas de datos, se obtiene al descomponer la matriz de la tabla III.9a en sus vectores y valores propios, cuya representación gráfica corresponde a la figura III.2:

Figura III.2
Primer plano de la interestructura del *Statis*



Las coordenadas, contribuciones y cosenos cuadrados de los operadores, a partir de las que se ha obtenido dicha representación, se recogen en el epígrafe A.2.1.4 (p. 580).

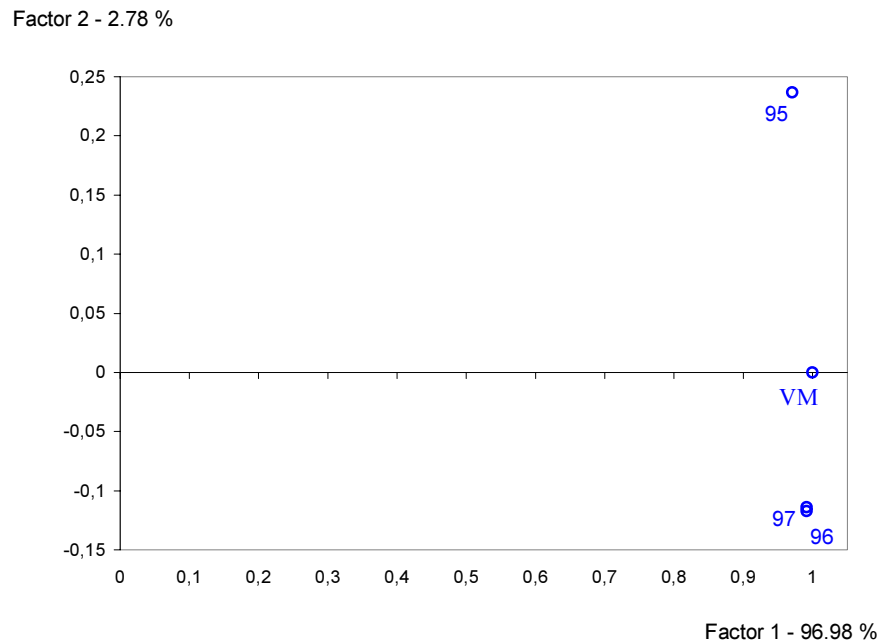
Como puede observarse este plano explica el 97,56% de la inercia total (casi toda acumulada por el primer factor lo que es debido a la semejanza de las tablas), y para su correcta interpretación es necesario tener en cuenta la norma de los operadores y el ángulo entre ellos. Dado que se trata de objetos normados, el operador compromiso *WD* tiene norma unitaria y se coloca a una distancia uno del origen sobre el primer eje. Los tres puntos correspondientes a los años estudiados están bien representados en cuanto que su norma está próxima a uno y su posición en el plano está lejos del origen.

Como era previsible según lo expuesto al analizar la matriz de la tabla III.17a, el vector correspondiente al año 1995 forma un ángulo elevado con el compromiso y con los vectores representativos de los otros dos

años, 96 y 97. Sin embargo, el ángulo entre el vector representativo del año 1997 y el del año 1996 es pequeño, es más, los dos vectores son prácticamente colineales.

La descomposición de la matriz de la tabla III.17b en sus vectores y valores propios permite obtener la imagen en dos dimensiones de los operadores asociados a las tablas de datos, que es la que aparece en la figura III.3:

Figura III.3
Primer plano de la interestructura del *Statis dual*



Las coordenadas, contribuciones y cosenos cuadrados de dichos operadores se recogen en el epígrafe A.2.2.3 (p. 592). El plano formado por los primeros ejes explica el 99,76% de la inercia total y al tratarse de objetos normados, el operador compromiso VM tiene norma unitaria y se coloca a una distancia uno del origen sobre el primer eje. Los tres puntos correspondientes a los años estudiados están bien representados en cuanto que su norma está próxima a uno y su posición en el plano está lejos del origen (la norma indica la correlación entre las variables de cada tabla y

esta es similar en todos). Como en el caso de las nubes de individuos, el vector correspondiente al año 1995 forma un ángulo grande con el compromiso, y también con los otros dos vectores representativos de los años 96 y 97. El ángulo entre el vector representativo del año 1997 y el del año 1996 es muy pequeño (los vectores son prácticamente colineales); y por tanto, el comportamiento de las variables en esos dos años es similar.

2. EL COMPROMISO

La matriz compromiso WD , en el caso del *Statis*, y VM , en el caso del *Statis dual*, se obtiene como media ponderada de las matrices $W_t D$ y $V_t M$, respectivamente, analizadas en la etapa de la interestructura.

Las ponderaciones empleadas para la obtención de uno y otro compromiso (WD y VM) son diferentes puesto que las matrices son de diferente naturaleza: VM es una matriz de correlación media de las matrices de correlación obtenidas, y WD es una matriz de productos escalares entre individuos, que representa la nube media de individuos; los resultados se expresan a continuación, según los epígrafes del anexo A.2.1.5 (p. 581) y A.2.2.4 (pp. 593-4):

Tabla III.18a
Ponderaciones *Statis*

	PONDE
NUBE 95	0,341
NUBE 96	0,354
NUBE 97	0,354

Tabla III.18b
Ponderaciones *Statis dual*

	PONDE
NUBE 95	0,298
NUBE 96	0,349
NUBE 97	0,353

Por ello tenemos que:

$$WD = 0,341 \frac{W_1 D}{\|W_1 D\|_{HS}} + 0,354 \frac{W_2 D}{\|W_2 D\|_{HS}} + 0,354 \frac{W_3 D}{\|W_3 D\|_{HS}}$$

$$VM = 0,298 \frac{V_1 M}{\|V_1 M\|_{HS}} + 0,349 \frac{V_2 M}{\|V_2 M\|_{HS}} + 0,353 \frac{V_3 M}{\|V_3 M\|_{HS}}$$

Como puede observarse, tanto para los operadores representativos de las nubes de individuos como de variables, las ponderaciones son diferentes pero hay una gran similitud entre ellas, lo que permite afirmar que el compromiso es un buen resumen de las tres tablas. Es necesario destacar que en ambos casos, debido a su estructura diferente, la nube correspondiente al año 95 es la que tiene una ponderación inferior y por tanto influirá en menor medida en la construcción de la estructura común. Como era previsible la ponderación de las nubes correspondientes a los años 96 y 97 es muy similar (igual en el caso de las nubes de individuos), dada la semejanza entre ellas manifestada al analizar la interestructura.

La imagen euclídea, de los individuos y de las variables, obtenida a partir de estas matrices es analizada en la siguiente etapa: la intraestructura.

3. LA INTRAESTRUCTURA

En primer lugar vamos a considerar los resultados obtenidos con el *Statis* y posteriormente los obtenidos con el *Statis dual* con el fin de establecer comparaciones. En cuanto al primero, la descomposición del operador WD en sus valores y vectores propios es similar al análisis de la nube de individuos en un ACP *centrado*. A partir de aquí se extrae la estructura común de los 48 individuos-compromiso (entidades aseguradoras), es decir, las coordenadas de dichos individuos sobre los ejes del compromiso (que dan lugar a los factores). Las posiciones de dichos individuos se interpretan, como en un ACP, según el significado asignado a los factores en función de sus correlaciones con las 13×3 variables (cuya representación es el círculo de correlaciones). Dado que todas las tablas tienen las mismas variables, la interpretación de los 3 puntos representativos de una de ellas (de diferente color según la tabla, como se observará en las representaciones obtenidas) permiten obtener conclusiones sobre la evolución de las 13 variables en el período considerado, en función de la correlación con los factores, así como la interpretación de éstos.

Los principales aspectos de las compañías y la relación entre variables se ponen de manifiesto reteniendo solamente los tres primeros factores obtenidos de la descomposición en vectores y valores propios del operador WD , que explican el 66,93% de la inercia total de la matriz: el primero el 35,65%; el segundo el 19,98% y el tercero el 11,29%, según se recoge en la tabla siguiente:

Tabla III.19
Primeros valores propios y porcentaje de inercia explicado por cada factor

Histograma de los 13 primeros valores propios				
NUMERO	VALOR PROPIO	PORCENTAJE	PORCENTAJE ACUMUL.	
1	0.8105	35.65	35.65	*****
2	0.4542	19.98	55.64	*****
3	0.2566	11.29	66.93	*****
4	0.1885	8.29	75.22	*****
5	0.1145	5.04	80.26	*****
6	0.0823	3.62	83.87	*****
7	0.0677	2.98	86.85	*****
8	0.0590	2.60	89.45	*****
9	0.0506	2.22	91.67	*****
10	0.0447	1.97	93.64	*****
11	0.0366	1.61	95.25	****
12	0.0288	1.27	96.52	***
13	0.0162	0.71	97.23	**

Como se puede observar, a partir de este factor los valores propios correspondientes descienden más rápidamente por lo que, según se explicó en el ACP, es aconsejable tomar los factores correspondientes a los valores anteriores a este decrecimiento. Además, como el total de valores propios distintos de cero es elevado (concretamente 32 según se observa en el epígrafe A.2.1.6: p. 583) el porcentaje de inercia explicado por los tres primeros es suficientemente elevado. Por otra parte, también es necesario considerar la interpretabilidad de los factores, en función de las variables con las que se relacionan, y por ello en este caso es aconsejable tomar el cuarto y el quinto factor, según se pondrá de manifiesto más tarde. Éstos reflejan tendencias no observables fácilmente en los datos originales, aunque expliquen poco porcentaje de inercia. Considerando los cinco se explica el 80,26% de la inercia total.

Estudio por variables

A partir de la siguiente tabla, en la que se recogen las correlaciones de las variables con los cinco primeros factores, realizaremos el estudio de las mismas según los planos factoriales considerados:

Tabla III.20
Correlaciones de las variables con los cinco primeros factores

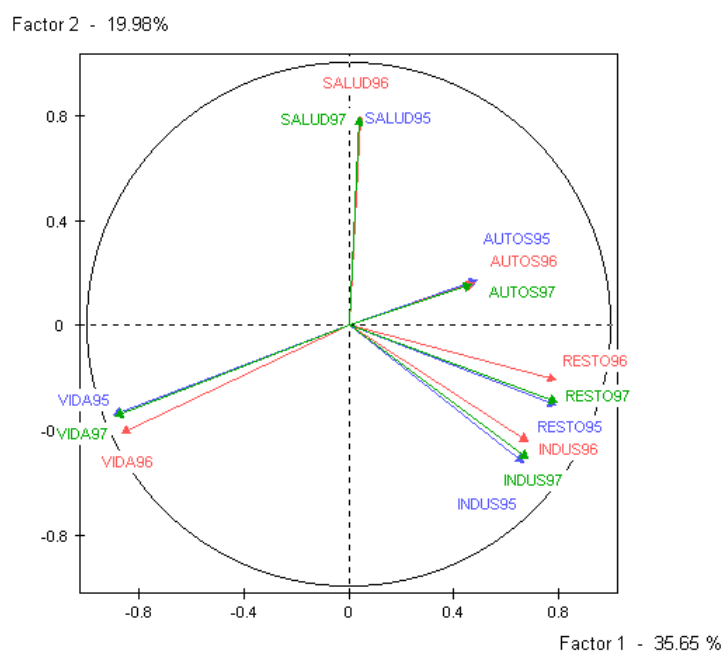
VARIABLES	CORRELACIONES				
	1	2	3	4	5
VIDA95	-0.90	-0.34	0.00	-0.07	0.10
AUTOS95	0.49	0.17	-0.61	0.51	-0.17
SALUD95	0.04	0.79	0.56	0.09	0.13
RESTO95	0.79	-0.31	0.10	-0.43	-0.09
INDUS95	0.67	-0.52	0.39	0.15	-0.10
PRIMEM95	-0.53	-0.11	0.05	-0.10	-0.53
GEXTPR95	0.54	0.05	-0.32	-0.70	0.05
GINTPR95	0.68	0.09	-0.08	0.00	0.31
MMMMI95	0.36	-0.28	0.27	0.14	0.12
INFINIT95	-0.56	-0.49	0.01	0.02	0.47
NNC95	-0.13	-0.13	0.28	-0.16	0.10
NNCPR95	0.08	0.84	0.34	-0.09	-0.02
NNPR95	-0.46	0.57	-0.52	-0.16	0.14
VIDA96	-0.86	-0.41	0.07	-0.04	0.10
AUTOS96	0.48	0.16	-0.61	0.53	-0.14
SALUD96	0.04	0.79	0.56	0.09	0.14
RESTO96	0.79	-0.21	0.01	-0.51	-0.12
INDUS96	0.68	-0.44	0.34	0.05	-0.15
PRIMEM96	-0.67	-0.18	0.12	-0.09	-0.51
GEXTPR96	0.63	0.07	-0.38	-0.63	0.02
GINTPR96	0.84	0.18	-0.21	0.08	0.21
MMMMI96	0.59	-0.44	0.20	0.27	0.18
INFINIT96	-0.56	-0.48	0.04	-0.11	0.47
NNC96	-0.66	0.10	0.04	-0.09	-0.30
NNCPR96	0.15	0.86	0.39	-0.11	-0.04
NNPR96	-0.55	0.55	-0.50	-0.15	0.11
VIDA97	-0.89	-0.34	0.00	-0.08	0.10
AUTOS97	0.47	0.16	-0.61	0.54	-0.13
SALUD97	0.04	0.79	0.56	0.09	0.14
RESTO97	0.79	-0.29	0.12	-0.47	-0.13
INDUS97	0.68	-0.50	0.41	0.09	-0.13
PRIMEM97	-0.68	-0.16	0.10	-0.10	-0.38
GEXTPR97	0.64	0.10	-0.34	-0.63	0.03
GINTPR97	0.85	0.14	-0.22	0.12	0.16
MMMMI97	0.66	-0.42	0.11	0.31	0.10
INFINIT97	-0.48	-0.54	0.00	-0.09	0.45
NNC97	-0.65	-0.01	0.02	-0.09	-0.10
NNCPR97	0.13	0.86	0.39	-0.10	0.03
NNPR97	-0.55	0.54	-0.50	-0.15	0.11

- Primer plano factorial (factores 1 y 2)

Las figuras III.4a y III.4b representan las correlaciones de las variables con el primer plano factorial, que es el engendrado por la primera y segunda componente principal de la matriz compromiso. Para una mejor caracterización de los ejes según las variables, se han representado las correlaciones de las variables y no las puntuaciones obtenidas del ACP *centrado*. Las puntuaciones, las correlaciones de todas las

variables y los ejes unitarios se recogen en el apartado A.2.1.7 de anexo (p. 585).

Figura III.4a
Círculo de correlaciones de las variables de alcance (excepto PRIMEM)
Factores 1 y 2

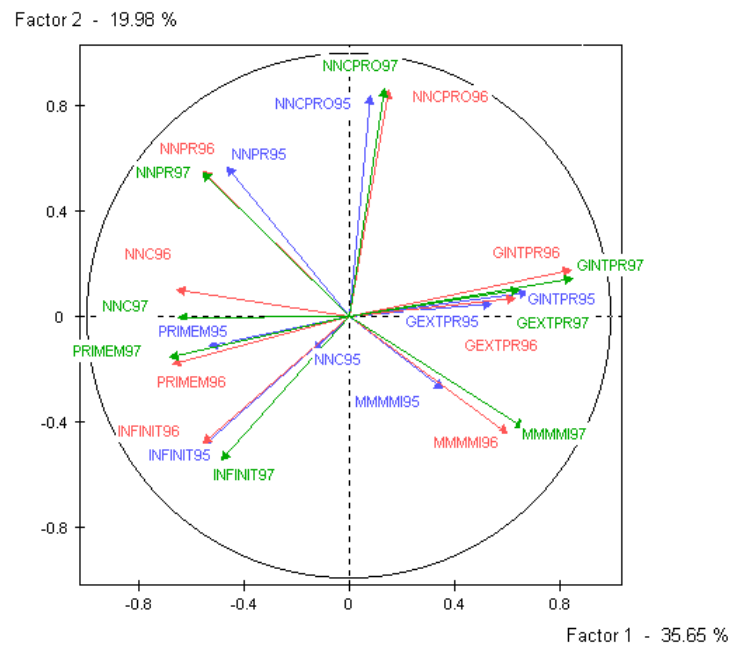


Esta figura representa las variables de alcance, excepto la variable PRIMEM. Como indicamos, estas variables permiten caracterizar los factores según la orientación al producto (variables VIDA, AUTOS, SALUD y RESTO), y al mercado (variable INDUS versus PERSO, tal y como se indicó). Cada variable está representada por tres vectores, cada uno de un color, con objeto de distinguir las variables según la tabla (año) a la que pertenecen. Todas las representaciones de variables resultado del *Statis* están elaboradas conforme a este criterio.

La figura III.4b, por su parte, recoge las variables indicativas del aprovechamiento de los recursos en las entidades estudiadas (variables de desarrollo) y del impacto de la bancaseguros (PRIMEM). Como se desarrolló (p. 149), en un ACP *centrado* las contribuciones relativas, que dan información de la calidad de representación de las variables, son proporcionales a los coeficientes de correlación. Por ello, las variables

con vectores próximos al círculo de correlaciones estarán bien representadas y su posición en el plano puede ser interpretada adecuadamente.

Figura III.4b
Círculo de correlaciones de las variables de desarrollo y PRIMEM
Factores 1 y 2



De la observación de la figura se deduce que, los vectores representativos de una misma variable pero en distintos años se agrupan según los ejes, lo que facilita la interpretación de los mismos (p. 220). Por otra parte, excepto las variables NNC95 y MMMMI95 de la figura III.4b, todas las variables están próximas al círculo de correlaciones y por tanto bien representadas.

El primer factor, explica el 35,65% de la inercia total, presentando una correlación bastante alta (superior a 0,5 e inferior a 0,8) y positiva con las variables de alcance que informan sobre la orientación al producto RESTO95, RESTO96 y RESTO97 (0,79 para todas ellas); y con las variables que dan información sobre la orientación al mercado INDUS95 (0,67), INDUS96 (0,68), INDUS97 (0,68). Las variables de desarrollo que presentan una correlación bastante alta y positiva con dicho factor

son GEXTPR95 (0,54), GEXTPR96 (0,63) y GEXTPR97 (0,64), así como con GINTPR95 (0,68), GINTPR96 (0,84), GINTPR97 (0,85). Las variables MMMMI96 y MMMMI97 también presentan una correlación importante: 0,59 y 0,66 respectivamente. Las de AUTOS, para todos los años, presentan una correlación en torno a 0,5 que aunque no es muy elevada si merece atención.

Con el lado negativo del factor, están altamente correlacionadas (superior a 0,8) las variables de alcance VIDA95 (-0,9), VIDA96 (-0,86) y VIDA97 (-0,89) y las variables indicativas de la bancaseguros (PRIMEMP95 con una correlación de -0,53; PRIMEMP96 con un valor de -0,67; y PRIMEMP97 con una correlación de -0,68). Teniendo en cuenta la relación INDUS versus PERSO, es necesario mencionar que esta última variable tendría una correlación alta con esta parte del eje. En cuanto a variables de desarrollo, presentan una correlación bastante alta las variables INFINIT95 (-0,56), INFINIT96 (-0,56) y INFINIT97 (-0,48); NNC96 (-0,66) y NNC97 (-0,65), NNPR96 y NNPR97 (-0,55 para ambas).

Merece la pena destacar que las variables de alcance relativas a las primas de seguros relacionados con la salud presentan una correlación prácticamente nula con este factor (0,04 para las tres: SALUD95, SALUD96, y SALUD97). También es necesario tener en cuenta el hecho de que las variables de AUTOS presenten una correlación positiva con este factor, aunque no sea importante.

El segundo factor, que explica el 19,98% de la inercia, presenta una correlación bastante alta y positiva con las variables de alcance SALUD (0,79 en todos los años), lo que indica la permanencia en el tiempo del comportamiento de esas variables. Este hecho también se manifiesta al observar las correlaciones de estas variables con los cinco factores retenidos: respecto al factor correspondiente, la correlación es la misma para todos los años. En cuanto a las variables de desarrollo, las variables NNCPRO95, NNCPRO96 y NNCPRO97 presentan una correlación positiva alta con este segundo factor (respectivamente, 0,84, 0,86 y

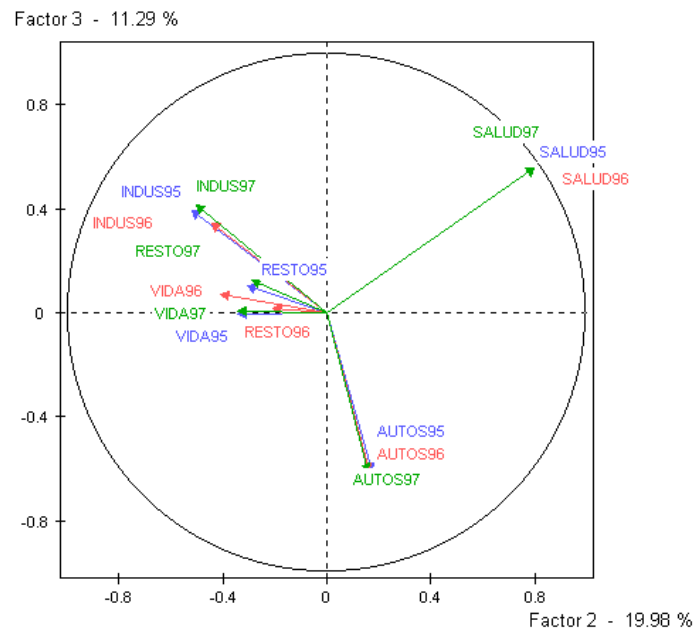
0,86). También superan el valor 0,5 las correlaciones de las variables NNPR95 (0,57), NNPR96 (0,55) y NNPR97 (0,54).

El lado negativo del segundo factor no presenta correlaciones especialmente elevadas con ninguna variable. En torno a $-0,5$ presentan correlaciones las variables INDUS95 ($-0,52$), INDUS96 ($-0,44$) e INDUS97 ($-0,5$), así como INFINIT95 ($-0,49$), INFINIT96 ($-0,48$) e INFINIT97 ($-0,54$).

Se van a analizar a continuación otros planos. Con el fin de que los mismos sean estables se ha representado en la figura siguiente las variables de alcance, excepto PRIMEM, en el plano formado por los factores 2 y 3, de inercia similar y que conforman el que denominaremos segundo plano factorial.

- Segundo plano factorial (factores 2 y 3)

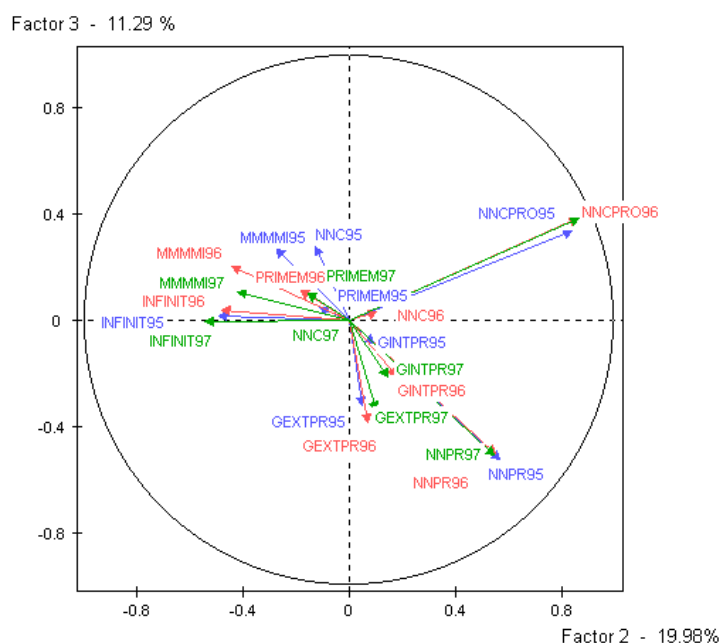
Figura III.5a
Círculo de correlaciones de las variables de alcance (excepto PRIMEM)
Factores 2 y 3



Este plano representa el 31,27% de la inercia total. Aunque ese valor no es elevado, el plano se considera en cuanto su interpretación va a permitir obtener conclusiones importantes acerca del sector asegurador.

Respecto a la figura III.5a podemos decir que solamente las variables de SALUD y AUTOS están correctamente representadas. La figura III.5b representa el resto de variables en los mismos factores.

Figura III.5b
Círculo de correlaciones de las variables de desarrollo y PRIMEM
Factores 2 y 3



En el caso de la figura III.5b, solo están adecuadamente representadas las variables relativas al reaseguro (NNPR95, NNPR96 y NNPR97), así como las que consideran la productividad de los capitales propios y las provisiones.

Puesto que el segundo factor ya ha sido analizado, sólo se van a considerar ahora las correlaciones con el tercer factor, que explica el 11,29% de la inercia total. Al analizar dichas correlaciones, recogidas en la tabla III.20, puede observarse una correlación bastante alta y positiva de las variables SALUD para los tres años (0,56 en todos los casos). Con la parte negativa de dicho factor se correlacionan las variables AUTOS95, AUTOS96 y AUTOS97, que presentan las tres una correlación de $-0,61$. Merece especial atención el hecho de que las variables de VIDA presenten una correlación prácticamente nula con

este factor (0 para las variables VIDA95 y VIDA97, y 0,07 para la variable VIDA96). En cuanto a las de desarrollo, las variables NNPR95, NNPR96 y NNPR97 presentan una correlación medianamente importante (en torno a 0,5) con el lado negativo de dicho factor.

- Tercer plano factorial (factores 4 y 5)

Con el mismo fin de estabilidad, se han representado las variables en los factores 4 y 5 aunque el plano resultante tenga poca inercia. Este tercer plano factorial explica el 13,33% de la inercia total. A pesar de ese bajo porcentaje se pueden extraer conclusiones interesantes. Las figuras III.6a y III.6b recogen las correlaciones de todas las variables con dichos factores.

Figura III.6a
Círculo de correlaciones de las variables de alcance, excepto PRIMEM
Factores 4 y 5

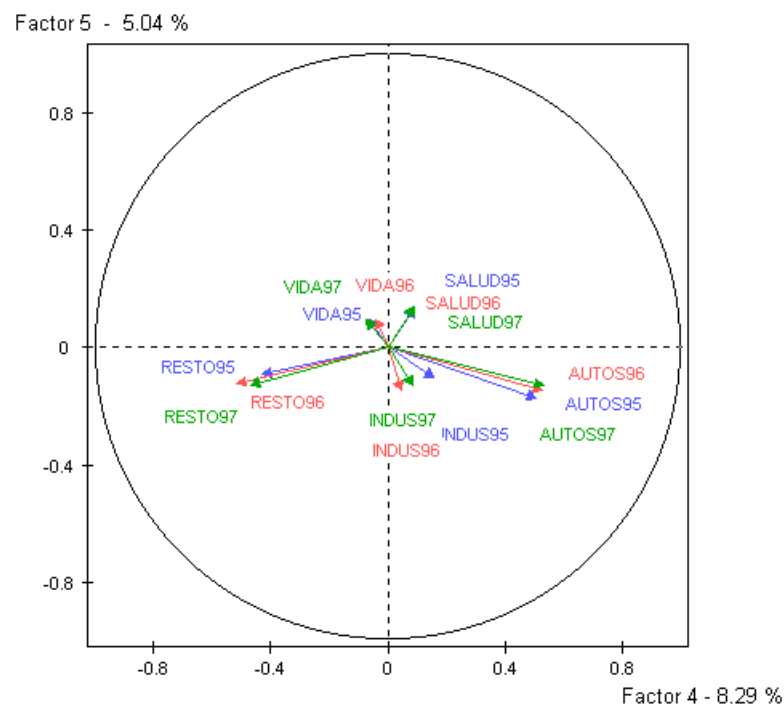
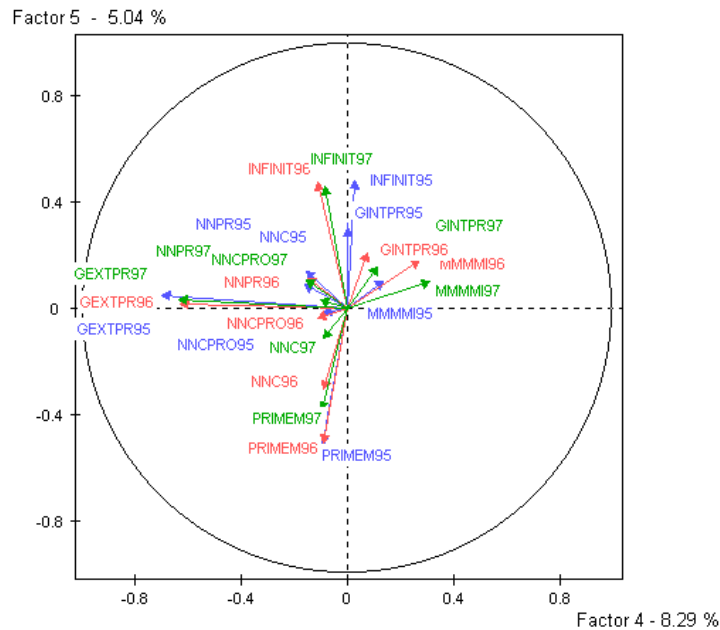


Figura III.6b
Círculo de correlaciones de las variables de desarrollo y PRIMEM
Factores 4 y 5



Debido a la baja inercia del plano pocas variables se representan correctamente: como variables de alcance podemos considerar las de AUTOS95, AUTOS96 y AUTOS97, PRIMEM95, PRIMEM96 y PRIMEM97, así como RESTO95, RESTO96 y RESTO97. En cuanto a las variables de desarrollo, solo aparecen bastante bien representadas las relativas a los gastos externos y a las inversiones financieras.

El cuarto factor solo presenta una correlación bastante alta, en el lado positivo, con las variables AUTOS95 (0,51), AUTOS96 (0,53) y AUTOS97 (0,54). En la parte negativa de dicho factor destacan las correlaciones de las variables GEXTPR95 (-0,7), GEXTPR96 (-0,63) y GEXTPR97 (-0,63). Las variables RESTO95, RESTO96 y RESTO97 presentan una correlación en torno al 0,5 con la parte negativa de este factor.

El quinto factor, con el 5,04% de la inercia, presenta las mayores correlaciones positivas con las variables relativas a las inversiones financieras. Así tenemos que las variables INFINIT95, INFINIT96 e INFINIT97 presentan una correlación, respectivamente, de 0,47, 0,47 y 0,45,

valores que no son muy elevados. Con el lado negativo de dicho factor presentan correlaciones importantes las variables que reflejan el fenómeno de bancaseguros: PRIMEM95 (-0,53), PRIMEM96 (-0,51) y PRIMEM97 (-0,38), correlación que va disminuyendo con el tiempo.

Los rasgos del sector, puestos de manifiesto a través de las correlaciones de las variables con los distintos factores, serán interpretados según los resultados descritos y van a permitir caracterizar la estructura común de dicho sector.

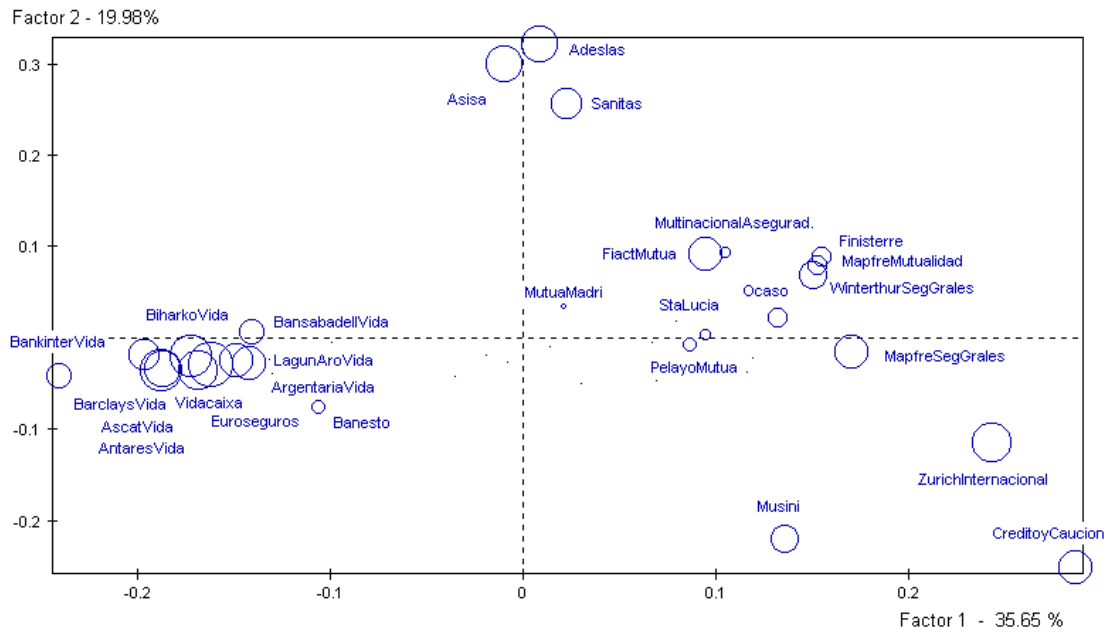
Estudio por individuos

La representación de los individuos en los planos permitirá definir las características de dichos individuos. Según se expuso al considerar la interpretación de los factores de un ACP la representación de los individuos (entidades) en este plano permitirá la caracterización de cada uno de ellos: el alejamiento de un individuo en la dirección marcada por el lado positivo o negativo de un factor indica porcentajes superiores a la media en las variables que caracterizan dicho lado, y valores inferiores a la media para las variables que caracterizan el lado opuesto del mismo factor, respectivamente. Además, un coseno cuadrado elevado para un individuo, con independencia del plano en el que esto se produzca, permite señalar que dicho individuo está bien representado por la estructura común del sector.

- Primer plano factorial (factores 1 y 2)

Las figuras III.7a y III.7b muestran los 48 individuos-compromiso (entidades aseguradoras) en el primer plano factorial. Las coordenadas, contribuciones absolutas y cosenos cuadrados de dichos individuos para los cinco primeros ejes se recogen en el epígrafe A.2.1.6 (p. 584). El icono representativo de cada individuo es proporcional a su coseno cuadrado, es decir, a la calidad de representación. Por tanto los círculos más grandes indican que la calidad de representación en el plano es mejor y más fidedigna será la interpretación de su posición.

Figura III.7a
Individuos-compromiso de mayor inercia
Factores 1 y 2



Dado que una representación de todos los individuos en el mismo gráfico no proporcionaría una visión adecuada de todos ellos, se tomó la decisión de hacer dos representaciones gráficas para cada plano. El criterio empleado para seleccionar los individuos que debía destacarse en una y en otra representación debía cumplir un único requisito: facilitar la interpretación de las representaciones obtenidas y la comparación con los resultados de los análisis posteriores; no obstante, todas las representaciones tienen en cuenta a todos los individuos y los no destacados están representados por puntos muy pequeños con el fin de ofrecer una imagen global.

Así el gráfico de la figura III.7a destaca la posición de los individuos de mayor inercia en el primer plano (más separados del origen) y por tanto, los que tienen una interpretación más clara según el significado otorgado a los factores.

Esta representación resalta casi todos los individuos que en la descomposición de la inercia de la matriz compromiso (anexo A.2.1.6:

p. 582) presentan un mayor porcentaje. Los que no han sido destacados en este primer plano posiblemente sean los de más inercia en el segundo o en el tercero.

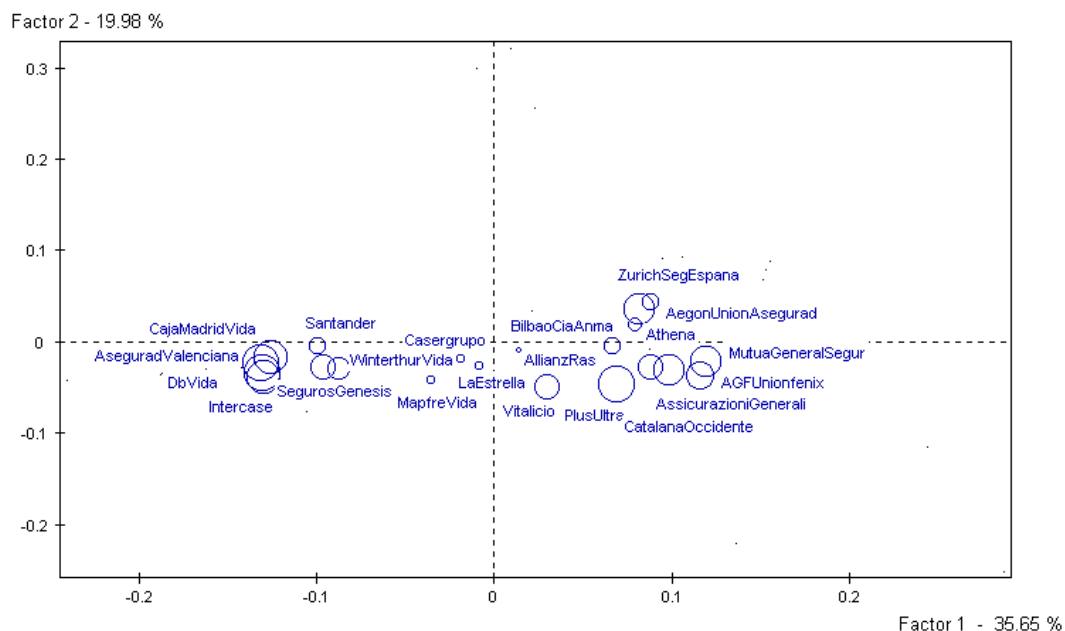
El plano, tal y como se deduce de la observación de la figura III.7a y de la interpretación de los cosenos cuadrados del epígrafe A.2.1.6 (p. 584) no representa adecuadamente las entidades Santa Lucía, Pelayo Mutua, Mapfre Mutualidad, Multinacional Aseguradora, Mutua Madrileña y Ocaso, siendo necesario analizar su posición en otros planos para poder caracterizarlos.

Por otra parte, solo tres de los individuos de dicha representación presentan una posición destacada respecto al segundo factor: Asisa, Adeslas y Sanitas. El resto de individuos se diferencian respecto al primer factor, encontrando entidades como Bankinter Vida, Bansabadell Vida, Lagun Aro Vida, Ascat Vida, Barclays Vida, Biharko Vida, Vidacaixa, Antares Vida, Argentaria Vida y Euroseguros, que se encuentran situadas en el extremo negativo del factor, y que pertenecen todas a grupos bancarios o de Cajas de Ahorros. Dada la posición más extrema de Bankinter Vida esta sería la entidad mejor definida por las características señaladas.

En el lado opuesto del factor destacan, bastante bien representadas, entidades como Musini, Zurich Internacional y Crédito y Caución, en el lado positivo del mismo, y con una coordenada prácticamente nula con el segundo factor. Entidades como Fiact Mutua, Winterthur Seguros Generales, Mapfre Seguros Generales y Finisterre también están posicionadas en el lado positivo de este primer factor, pero su posición es menos extrema.

La agrupación de individuos resultado del Análisis Cluster permitirá caracterizar de manera más precisa los individuos, sobre todo para analizar los de la figura III.7b que son los que tienen un comportamiento menos definido respecto a los factores caracterizados, pues ocupan las posiciones centrales del plano.

Figura III.7b
Otros individuos-compromiso
Factores 1 y 2



Casi todos los individuos están bien representados, excepto Bilbao Cia. Anma., Athena, Caser Grupo, La Estrella, Mapfre Vida, Allianz Ras. Conviene señalar que los cuatro últimos individuos no van a ser representados tampoco de forma adecuada en ninguno de los otros planos, como se deduce al analizar sus cosenos cuadrados (anexo A.2.1.6: p. 584).

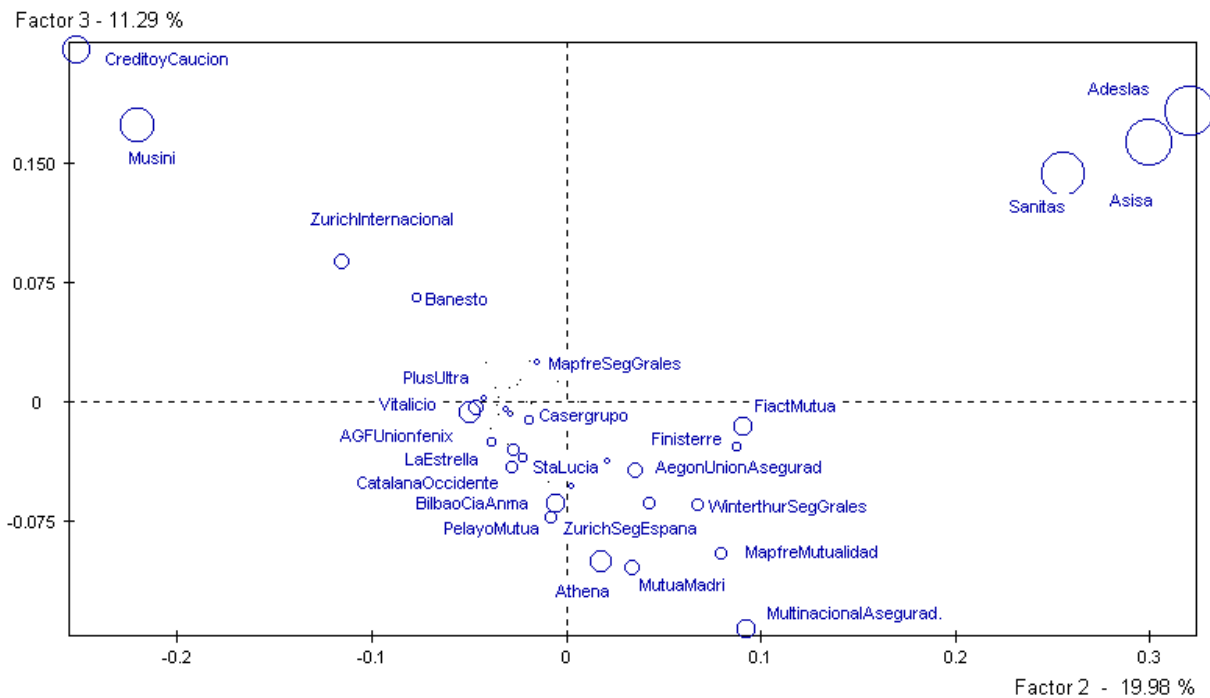
Entre los individuos bien representados, en la parte negativa del primer factor encontramos a Santander, Caja Madrid Vida, Aseguradora Valenciana, Db Vida, Intercaser, Seguros Génesis y Winterthur Vida.

Las características que definen Winterthur Seguros Generales o Mapfre Seguros Generales (de la figura III.7a) serán también las que definen en este caso entidades como Plus Ultra, Zurich Seguros España, Mutua General de Seguros, AGF Unión y Fenix, Assicurazioni Generali, Catalana Occidente o Vitalicio. Si bien, puesto que estos últimos individuos ocupan posiciones más centrales respecto al primer factor, dichas características están menos acentuadas.

- Segundo plano factorial (factores 2 y 3)

Vamos a revisar el segundo y tercer planos factoriales para caracterizar los individuos que no estaban bien representados en el anterior, así como para establecer diferencias entre los individuos que parecen tener los mismos rasgos debido a que ocupan una posición similar en ese primer plano. La figura III.8 solo destaca aquellos individuos de mayor inercia y que están bien representados en el segundo plano (factores 2 y 3). Como ya indicamos, este plano no representa a las entidades con porcentajes importantes de seguros de vida. Por ello, en general, van a ser estas entidades las que están mal representadas.

Figura III.8
Individuos-compromiso de mayor inercia
Factores 2 y 3



Con una posición claramente definida aparecen, en el lado positivo del segundo factor, entidades como Asisa, Adeslas y Sanitas. En situación opuesta respecto a dicho segundo factor aparecen entidades como Crédito y Caución, y Musini.

El tercer factor permite diferenciar a Zurich Internacional de las otras dos entidades que ocupan una posición similar en el primer plano: Crédito y Caución, y Musini. Asimismo este factor permite representar mejor algunas entidades que aparecían mal representadas o con una posición central que dificultaba su caracterización, como Mutua Madrileña, Multinacional Aseguradora, Mapfre Mutualidad, Pelayo Mutua, Zurich Seguros España, Winterthur Seguros Generales, Athena y Bilbao Cia. Anma.

- Tercer plano factorial (factores 4 y 5)

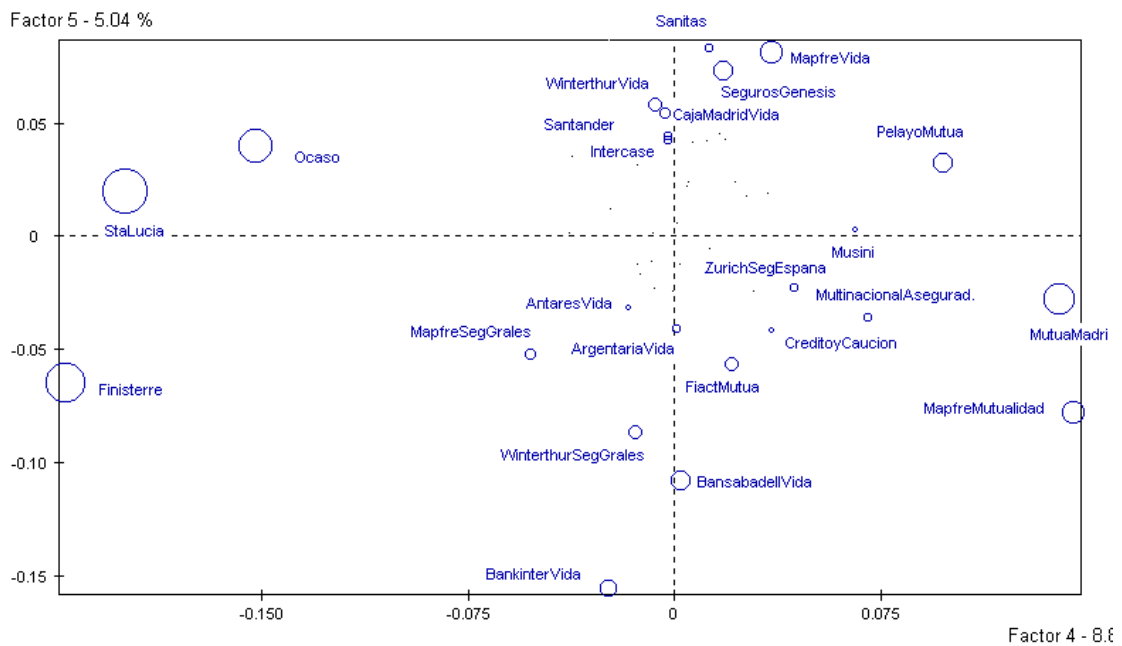
Aunque de baja inercia, este plano permite representar adecuadamente ciertas entidades, además de poner de manifiesto algunas tendencias difícilmente observables en los otros casos. Así, el cuarto factor establece diferencias entre entidades que parecían similares al analizar el primer plano factorial: mientras Mutua Madrileña, Mapfre Mutualidad y Pelayo Mutua se sitúan en la parte positiva de dicho factor, en el lado negativo ocupan una posición destacada entidades como Ocaso, Santa Lucia y Finisterre.

El quinto factor establece una clara diferencia entre las entidades como Mapfre Vida, Seguros Génesis y Winterthur Vida, situadas en el extremo positivo de dicho factor, frente a otras como BankinterVida, BansabadellVida, que se sitúan en el extremo opuesto.

A continuación se detallan los resultados obtenidos con la aplicación del *Statis dual*, que van a permitir matizar la interpretación otorgada a los factores. Este método analiza el operador VM (matriz de correlación en este caso), cuya descomposición en valores y vectores propios es similar al ACP *normado* de la nube de variables (pues éstas están centradas y reducidas). Así, se extrae la estructura de variables común a todas las tablas, y las correlaciones de las 13 variables—compromiso con los factores que permiten su caracterización. La representación de los (48×3) individuos (las 48 entidades en los 3 años)

es interpretada según el significado otorgado a los factores, y los 3 puntos representativos de cada entidad, (uno por cada tabla anual) permiten definir sus trayectorias en un espacio óptimo e interpretar éstas de acuerdo con el significado de cada factor y de cada plano.

Figura III.9
Individuos-compromiso de mayor inercia
Factores 4 y 5



De la misma forma que se ha realizado para el *Statis*, en el *Statis dual* comenzamos describiendo los resultados obtenidos para las variables, a través de sus correlaciones con los factores obtenidos, con el fin de caracterizarlos. Posteriormente, la representación de los individuos sobre dichos ejes permitirá interpretar la posición de cada uno de ellos, así como su cambio de posición de unos años a otros.

Las principales relaciones entre las variables se ponen de manifiesto reteniendo los tres primeros factores obtenidos en la descomposición de la matriz VM , que explican el 69,02% de la inercia. A partir de ese valor los valores descienden rápidamente. Si bien, en el caso de considerar cinco factores, el porcentaje explicado aumenta hasta el

84,23%, lo que enriquece las conclusiones que se pueden obtener ya que en este caso los dos factores son interpretables en función de las variables con las que se relacionan. Los valores propios obtenidos se recogen en la siguiente tabla:

Tabla III.21
Valores propios y porcentajes de inercia explicado por cada factor

Histograma de los 13 primeros valores propios				
NUMERO	VALOR PROPIO	PORCENTAJE	PORCENTAJE ACUMULADO	
1	4.8399	37.23	37.23	*****
2	2.6433	20.33	57.56	*****
3	1.4892	11.46	69.02	*****
4	1.1461	8.82	77.83	*****
5	0.8315	6.40	84.23	*****
6	0.6197	4.77	89.00	*****
7	0.4643	3.57	92.57	*****
8	0.4155	3.20	95.76	*****
9	0.2710	2.08	97.85	*****
10	0.1325	1.02	98.87	***
11	0.0896	0.69	99.56	**
12	0.0574	0.44	100.00	*
13	0.0002	0.00	100.00	*

Representación de las variables

La siguiente tabla recoge las correlaciones de las variables con los cinco primeros factores: las coordenadas, contribuciones y cosenos cuadrados de todas las variables—compromiso se recogen en el epígrafe A.2.2.5 (p. 595):

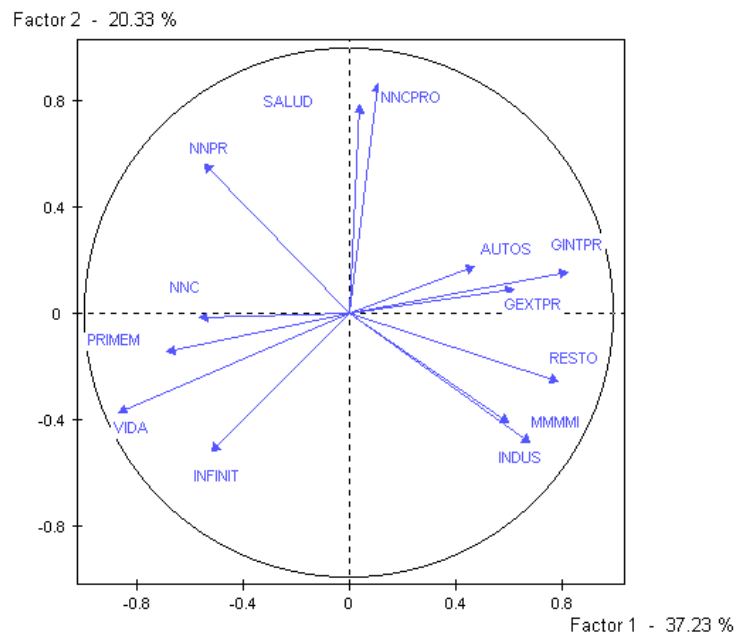
Tabla III.22
Correlaciones de las variables con los cinco primeros factores

VARIABLE	CORRELACIONES				
	1	2	3	4	5
VIDA	-0.87	-0.37	0.00	0.01	0.18
AUTOS	0.47	0.18	-0.61	0.40	-0.38
SALUD	0.04	0.78	0.54	0.16	0.13
RESTO	0.78	-0.26	0.13	-0.51	-0.02
INDUS	0.68	-0.49	0.41	0.02	-0.15
PRIMEM	-0.69	-0.15	0.14	-0.20	-0.43
GEXTPR	0.62	0.09	-0.31	-0.66	0.13
GINTPR	0.82	0.15	-0.18	0.04	0.03
MMMMI	0.60	-0.41	0.18	0.31	-0.03
INFINIT	-0.52	-0.52	-0.01	-0.01	0.38
NNC	-0.57	-0.02	0.19	-0.30	-0.51
NNCPRO	0.10	0.87	0.39	-0.09	0.02
NNPR	-0.54	0.56	-0.51	-0.16	0.06

- Primer plano factorial (factores 1 y 2)

La figura III.10 representa el círculo de correlaciones formado por las 13 variables–compromiso en el primer plano obtenido, el de los factores 1 y 2, que representa el 57,56% de la inercia total.

Figura III.10
Círculo de correlaciones de las variables–compromiso
Factores 1 y 2



El primer factor presenta una correlación positiva alta con las variables–compromiso de alcance RESTO (0,78), INDUS (0,68); y con las variables de desarrollo GEXTPR (0,62), MMMMI (0,6), y más alta con la variable GINTPR (0,82). Por otra parte, la variable VIDA presenta una correlación alta y negativa con el mismo factor (-0,87). La parte negativa de este factor también está correlacionada, aunque en menor medida, con las variables PRIMEM (-0,69), y con las variables de desarrollo NNC (-0,57), NNPR (-0,54) e INFINIT (-0,52).

El factor 2 está correlacionado en su parte positiva con las variables NNCPRO (0,87) y SALUD (0,78). Aunque en menor media, la variable NNC también está correlacionada con la parte positiva de este factor (0,56).

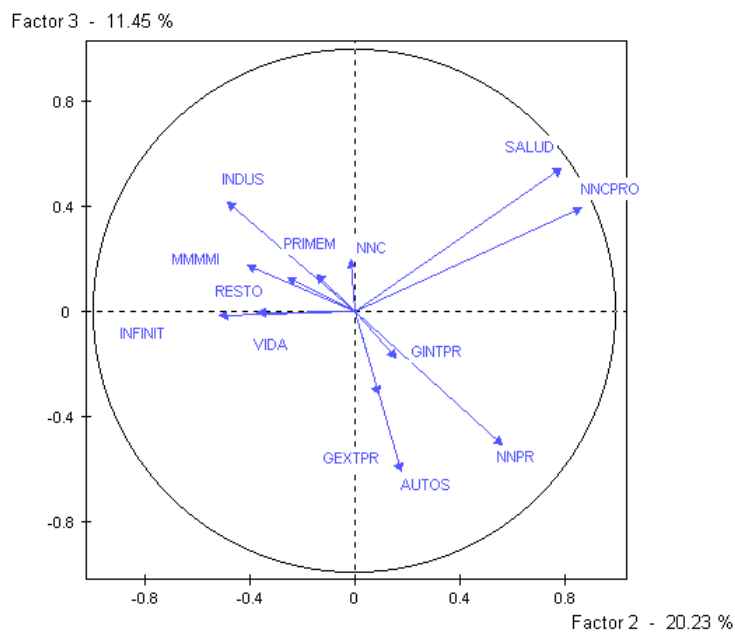
En cuanto al extremo opuesto del factor, el variable INFINIT es la única que presenta una correlación negativa de cierta importancia ($-0,52$).

Por tanto, este primer plano factorial del *Statis dual* tiene una interpretación similar al obtenido anteriormente. Además su inercia también parecida: un 55,64% en el primer caso frente al 57,64 % en el caso del *Statis dual*.

- Segundo plano factorial (factores 2 y 3)

La figura III.11 recoge el segundo plano factorial, formado por los factores 2 y 3 que explica el 31,78% de la inercia.

Figura III.11
Círculo de correlaciones de las variables-compromiso
Factores 2 y 3



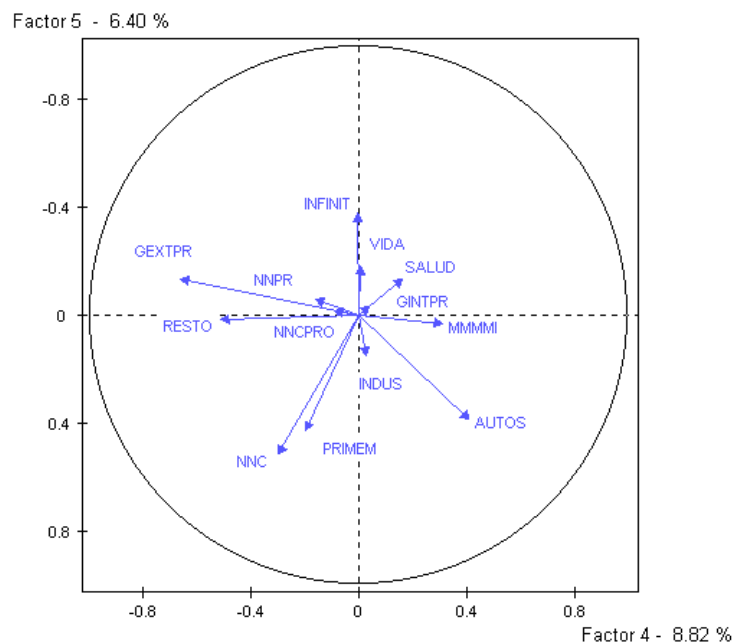
Como el segundo factor ya ha sido comentado, únicamente se va a prestar atención al tercer factor. Éste presenta una correlación bastante alta y negativa con las variables AUTOS ($-0,61$) y NNPR ($-0,51$). El extremo positivo se correlaciona con la variable SALUD, aunque esta correlación no sea demasiado importante ($0,54$). Es significativo el

hecho de que la variable VIDA tenga una correlación prácticamente nula con dicho factor. Como consecuencia, el segundo plano tiene un significado similar al segundo plano obtenido con el método *Statis*.

- Tercer plano factorial (factores 4 y 5)

La figura III.12 presenta el círculo de correlación del tercer plano factorial, formado por los factores 4 y 5. Estos dos factores, aunque de inercia baja (en total 15,22%), son considerados puesto que reflejan ciertas tendencias que no se manifiestan con los otros factores.

Figura III.12
Círculo de correlaciones de las variables–compromiso
Factores 4 y 5



El factor 4 presenta correlaciones importantes con la variable de desarrollo GEXTPR (-0,66), y con la variable de alcance RESTO (-0,51). El factor 5 presenta una correlación positiva con la variable NNC que, aunque no es muy elevada, sí puede considerarse importante (0,51). La variable PRIMEM (-0,43) no presenta una correlación muy importante

con este factor, pero dado que la correlación de las otras variables es mucho menor, ese valor puede ser interesante.

Según lo expuesto, los factores obtenidos con el *Statis dual* tienen un significado similar a los del *Statis*.

En cuanto a los individuos, las coordenadas de los mismos, sus contribuciones y los cosenos cuadrados se recogen en el apartado A.2.2.5 (pp. 596-8). Su representación da lugar a las mismas figuras que las que representan las trayectorias de los individuos (figuras III.13a hasta III.14d, y las recogidas en el epígrafe A.2.2.6 del anexo, pp. 602-5), por lo que las principales conclusiones coinciden con las expuestas al analizar dichas trayectorias.

Conviene precisar que a partir de dichas coordenadas es posible representar los individuos por años, y analizar su situación para el año correspondiente. El inconveniente de este tipo de gráficas es el gran número de individuos, aunque siempre es posible hacer representaciones parciales. Puesto que el objetivo de la aplicación no es éste, en el apartado A.2.2.7 (pp. 599-601) del anexo sólo se recogen, a modo de ejemplo, las gráficas correspondientes al primer plano factorial, para los tres años.

4. TRAYECTORIAS

Dados los inconvenientes que presentan las trayectorias de los individuos y las variables obtenidas, respectivamente, por el *Statis* y por el *Statis dual* (p. 220) se propone como solución alternativa el análisis de la descomposición de las distancias, por filas o por columnas. En nuestro caso, el estudio de las trayectorias a través de esa descomposición puede completarse combinando adecuadamente los resultados obtenidos al aplicar simultáneamente las dos versiones de la metodología STATIS, ya que los factores tienen significados similares.

- Trayectorias de las variables (Resultado del *Statis*)

La descomposición de las distancias, por variables, entre las matrices de correlación permiten poner de manifiesto las variables responsables de las diferencias entre dichas tablas. La interpretación se matiza con las trayectorias de las figuras III.4a hasta III.6b, obtenidas con el *Statis*.

Para una correcta interpretación de estas descomposiciones hay que tener en cuenta que el cambio, importante o no, es en términos relativos. Un variable presenta un cambio significativo cuando acumula un porcentaje muy superior al resto de variables. El epígrafe A.2.2.2 (p. 591) recoge dicha descomposición de distancias. Así, las diferencias entre el año 95 y 97 se deben principalmente a las variables PRIMEM (10,6%); GINTPR (8,9%); MMMMI (12,1%); y NNC (35,8%). Las diferencias entre el año 95 y el 96 se deben a las variables PRIMEM (10,5%); GINTPR (9%); MMMMI (11%) y NNC (38,3%). Sin embargo, las variables VIDA (11,3%), RESTO (10,1%), MMMMI (20%), INFINIT (9,6%), NNC (35,8%) y NNPR (9,8%), son las responsables de las diferencias entre las matrices de correlación del año 96 y 97.

La trayectoria de cada variable vendría dada por los tres vectores que representan cada variable (en cada año) en dichas figuras, y cuya descripción ya se ha realizado. Conviene volver a señalar que la característica general en todos los casos es la agrupación de las variables por factores, por lo que a pesar de las variaciones de unos años a otros en algunas variables, éstas no influyen en la interpretación otorgada a los factores.

- Trayectorias de los individuos (Resultado del *Statis dual*)

Las trayectorias de los individuos que pueden considerarse más significativas se manifiestan al observar el análisis de los porcentajes obtenidos en la descomposición de la distancia, según líneas, del epígrafe A.2.1.3 (p. 579). Igual que para las variables, conviene señalar que para una correcta interpretación de estas descomposiciones hay

que tener en cuenta que el cambio, significativo o no, se interpreta en términos relativos: un individuo presenta un cambio importante cuando acumula un porcentaje muy superior al resto de individuos.

El cambio de posición en el periodo considerado, es decir, del año 95 al 97 es especialmente importante para entidades como Barclays Vida (5,6%), Antares Vida (5,57%), Santander (4,6%), Banesto (15,9%), Multinacional Aseguradora (11,6%), Crédito y Caución (5,4%), Bankinter Vida (5,9%) y Zurich Internacional (3,2%). Del 95 al 96 experimentan cambios importantes entidades como Vidacaixa (2%), Banesto (23,2%), Multinacional Aseguradora (12%), Crédito y Caución (6,8%), Bankinter Vida (5,5%), Musini (4%), Zurich Internacional (4,2%). Entre el año 96 y el año 97 son las compañías Vidacaixa (4,6%), Antares Vida (4,4%), Santander (5,5%), Argentaria Vida (4,2%), Banesto (6,1%), Crédito y Caución (4,1%), Bankinter Vida (5,3%), Musini (7,5%) y Bansabadell Vida (3,2%) las que presentan un mayor porcentaje y por tanto las responsables, en mayor medida, de las diferencias entre las tablas relativas a esos dos años.

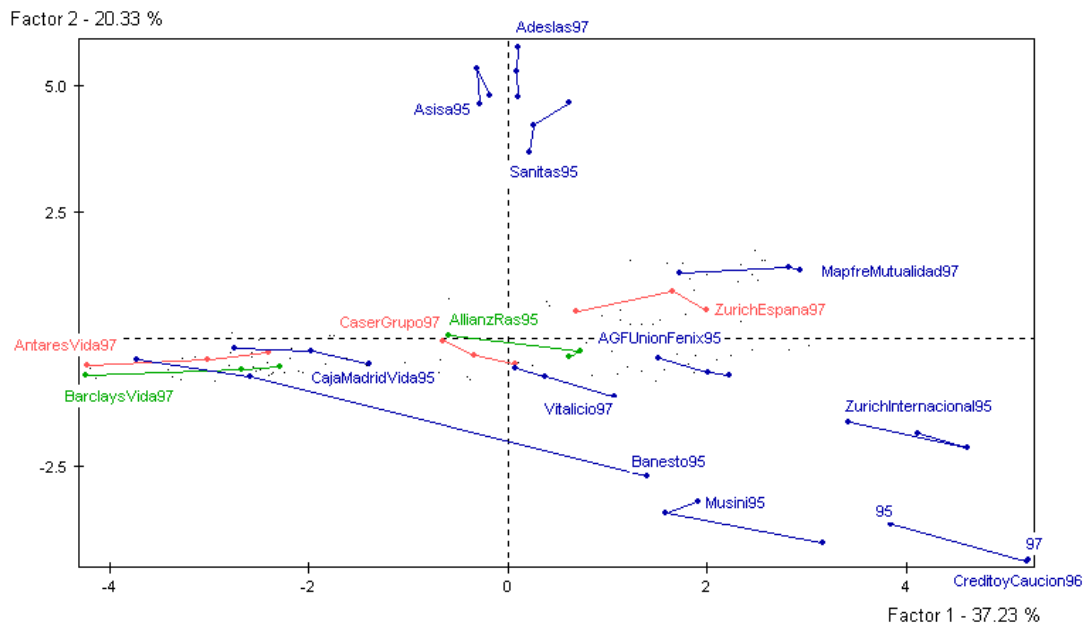
Además de dichas descomposiciones, el análisis de las trayectorias, en este caso, se realiza teniendo en cuenta las coordenadas, en los ejes factoriales, obtenidas con el *Statis dual* para todas las entidades y para los tres años considerados. Esto es posible debido a la relación, ya puesta de manifiesto, entre los significados de los factores en ambos análisis. La trayectoria de cada entidad estará formada por los tres puntos que la representan, es decir, los correspondientes a los años 1995, 1996 y 1997.

A continuación se muestran las representaciones gráficas que presentan las trayectorias de mayor longitud, y el resto se recogen en epígrafe A.2.2.8 del anexo (pp. 602-5). Cada figura destaca una serie de individuos, de tal forma que son necesarias cuatro (identificadas con las letras a, b, c y d detrás del número correspondiente) por cada plano para observar las trayectorias

de las 48 entidades. El único objetivo de esta forma de representar las trayectorias es facilitar su interpretación; con el mismo fin, para cada uno de los individuos, sólo se ha etiquetado el punto inicial (año 95) ó final (año 97) de la trayectoria, y se han empleado colores diferentes para diferenciar unas de otras.

Así, la figura III.13a destaca las trayectorias del primer grupo de individuos en el primer plano factorial.

Figura III.13a
Trayectorias de individuos
Factores 1 y 2



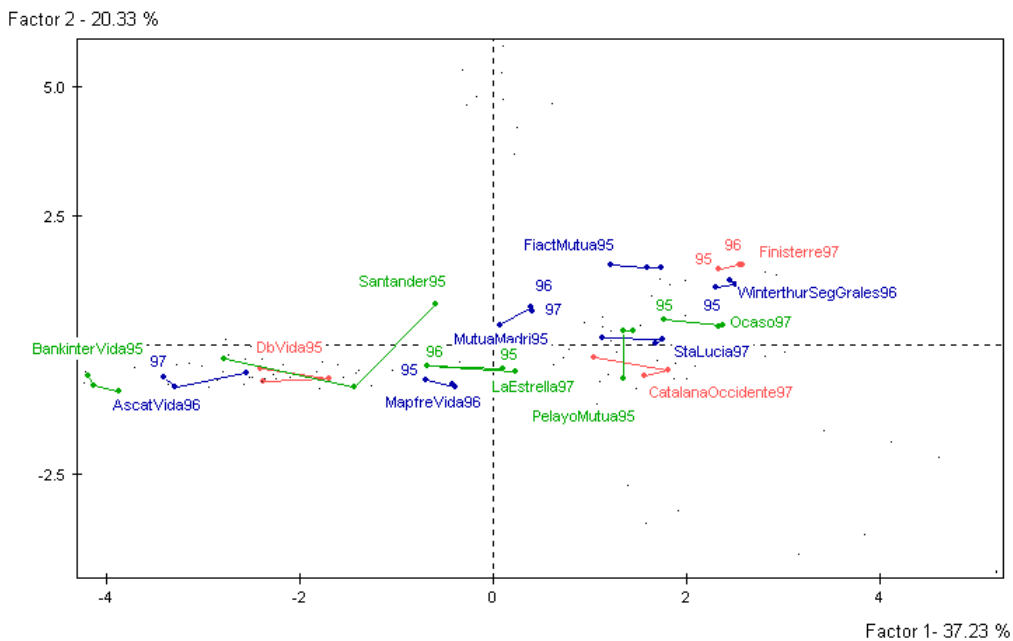
En esta figura se observa que los cambios de posición más importantes son las que presentan las entidades Banesto, Antares Vida y Barclays Vida: Banesto cambia del lado positivo al lado negativo del primer factor; Antares Vida y Barclays Vida se mueven desde la parte central hasta el extremo negativo del mismo factor.

Otras tres entidades que presentan trayectorias significativas, y en el mismo sentido son Crédito y Caución, Zurich España y Mapfre Mutualidad, las cuales extreman sus posiciones respecto a la parte positiva del primer

factor. Musini y Zurich Internacional presentan trayectorias cambiantes en el periodo; si bien, la posición final es más extrema en el lado positivo del factor primero. Las trayectorias de Allianz Ras, Vitalicio, Caser Grupo, y AGF Unión y Fénix son cortas y ocupan la posición central. Adeslas y Sanitas extreman un poco más sus posiciones hacia el lado positivo del segundo factor.

La figura III.13b presenta las trayectorias sobre el mismo plano de otro conjunto de individuos.

Figura III.13b
Trayectorias de individuos
Factores 1 y 2



Todos presentan trayectorias cortas y cambiantes en el periodo, excepto Santander, que extrema su posición en la parte negativa del primer factor; Ascat Vida extrema su posición hacia la parte negativa del primer factor, lo contrario que en el caso de Bankinter Vida.

Las figuras III.13c y III.13d presentan las trayectorias del resto de individuos en el primer plano factorial.

Figura III.13c
Trayectorias de individuos
Factores 1 y 2

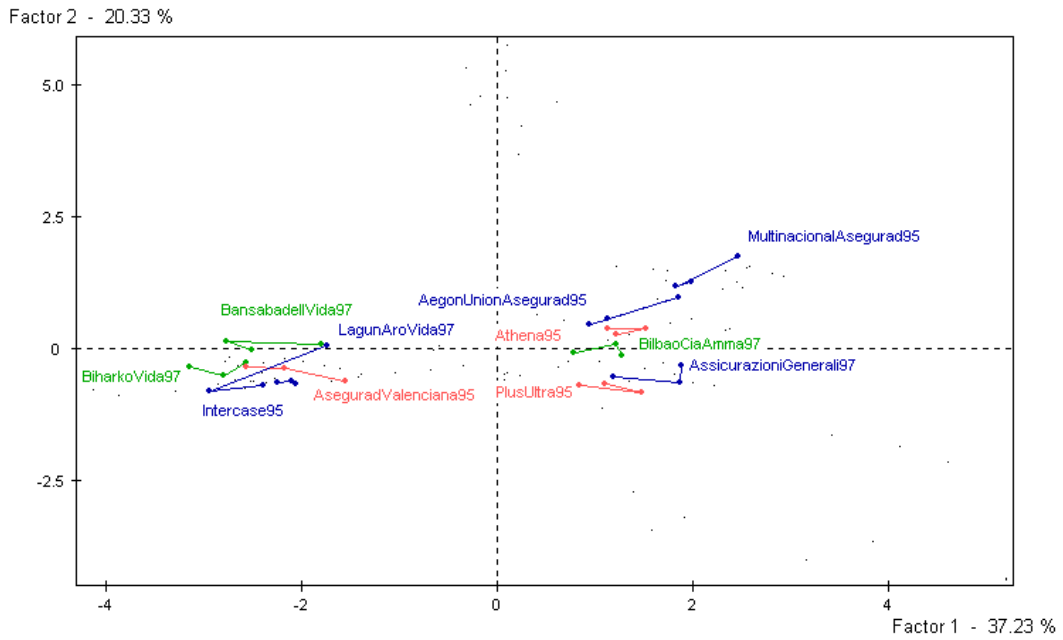
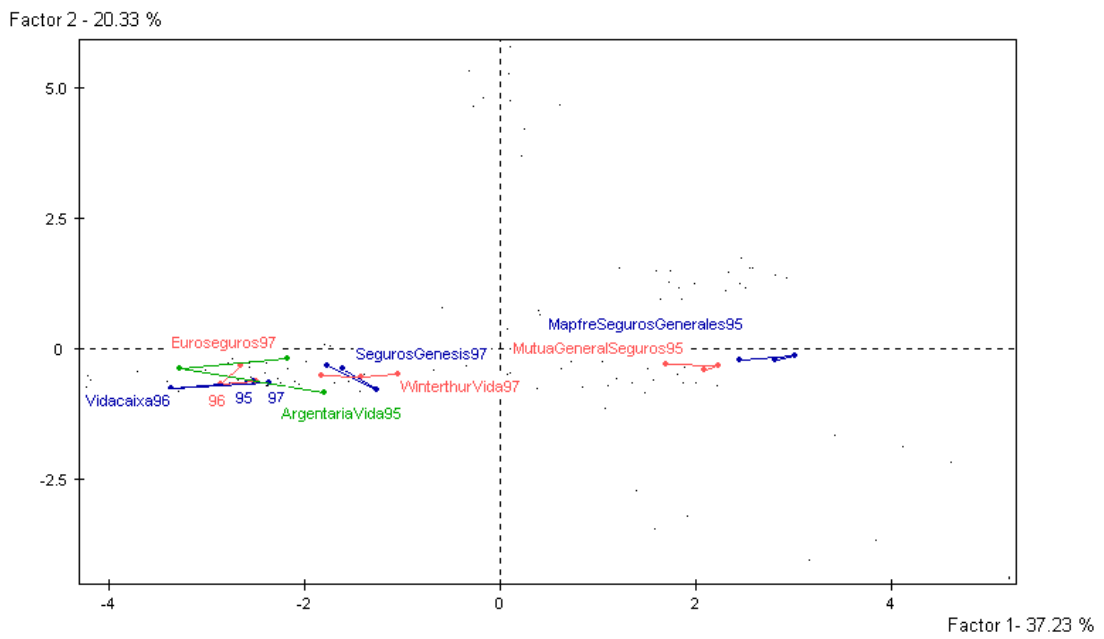


Figura III.13d
Trayectorias de individuos
Factores 1 y 2



En estas figuras ninguna trayectoria es especialmente importante. La mayoría son cortas y presentan movimientos heterogéneos. Si bien, es posible destacar un movimiento hacia el extremo negativo del primer factor de Bansabadell Vida, y la evolución de la trayectoria de Lagun Aro Vida en sentido contrario.

En el anexo A.2.2.8 (pp. 602-5) se recogen las trayectorias de los 48 individuos sobre el segundo y tercer factor, que conforman el segundo plano. Pocas trayectorias se interpretan adecuadamente pues son pocos los individuos con una calidad de representación adecuada, siendo muchas trayectorias pequeñas y centrales. En concreto, en dicho plano, los individuos con altos porcentajes de seguros de vida no tienen representaciones adecuadas (cosenos cuadrados recogidos en el epígrafe A.2.2.6: pp. 596-8). Cabe destacar la trayectoria de Pelayo Mutua (figura e), que cambia su posición de la parte central a la parte negativa del tercer factor, y la trayectoria de Multinacional Aseguradora (figura f), que se mueve en sentido contrario.

Aunque la inercia del tercer plano factorial es pequeña, las trayectorias de ciertos individuos bien representados, como Ocaso, Mapfre Mutualidad, Pelayo, Finisterre, Santa Lucia, Multinacional Aseguradora y Musini, si son importantes. Las figuras III.14a hasta III.14d recogen las mencionadas trayectorias.

Las trayectorias que ocupan las posiciones centrales no tienen una interpretación clara, y suelen corresponder a individuos cuya calidad de representación no es buena en el plano considerado (anexo A.2.2.6: pp. 596-8). Destacan las trayectorias de Barclays Vida y Antares Vida, que se mueven a lo largo del factor cinco, extremando su posición negativa. La aseguradora Musini presenta cambios respecto al mismo factor: pasa de una posición central a una posición positiva del año 95 al 96, para volver a una posición central en el 97.

Figura III.14a
Trayectorias de individuos
Factores 4 y 5

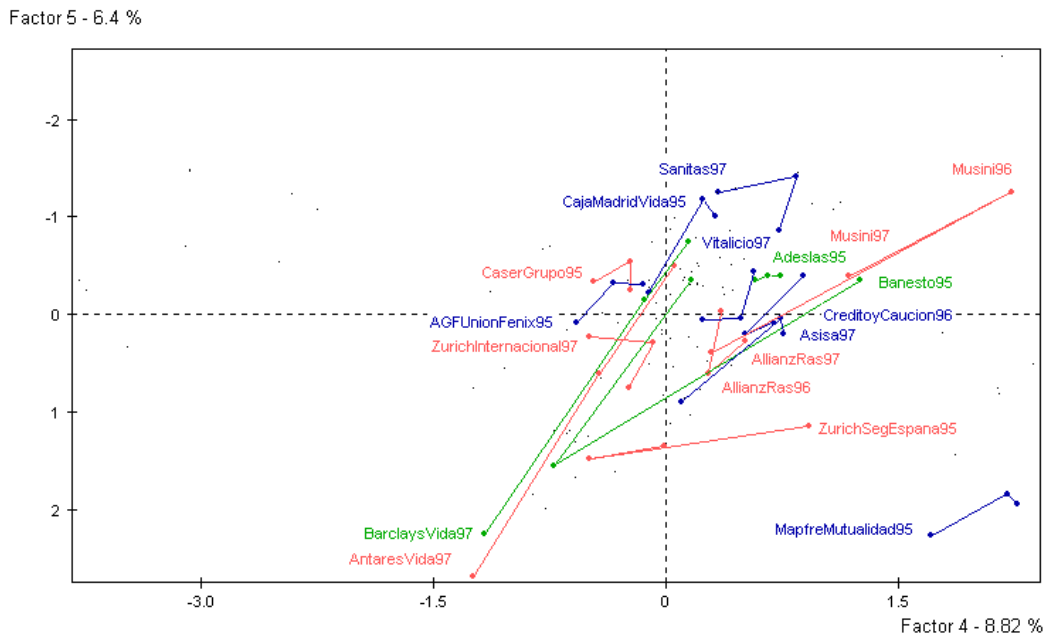


Figura III.14b
Trayectorias de individuos
Factores 4 y 5

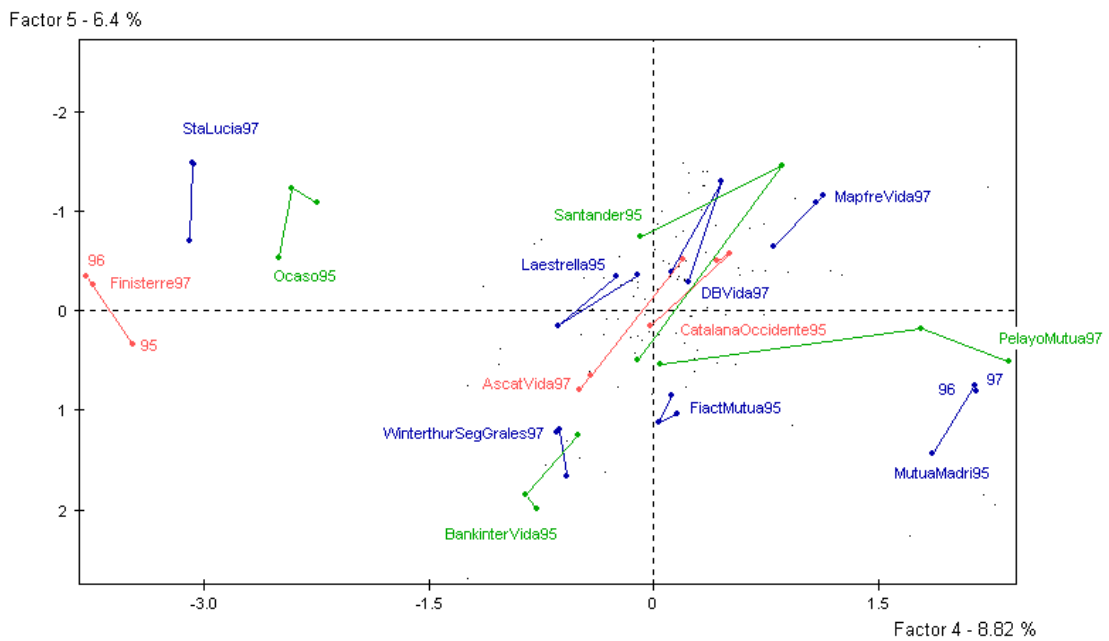


Figura III.14c
Trayectorias de individuos
Factores 4 y 5

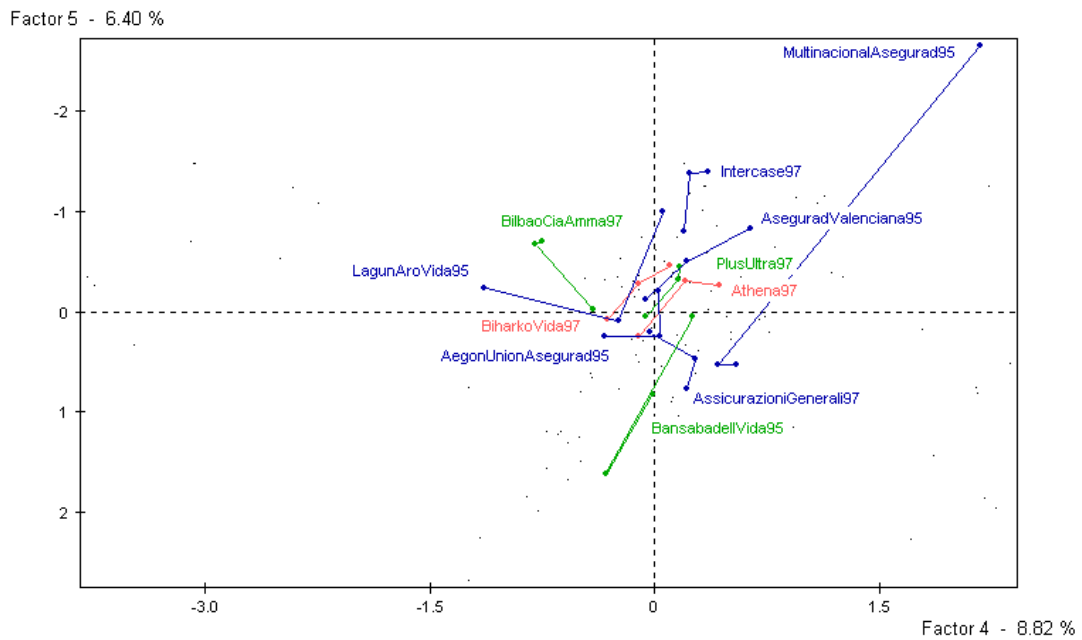
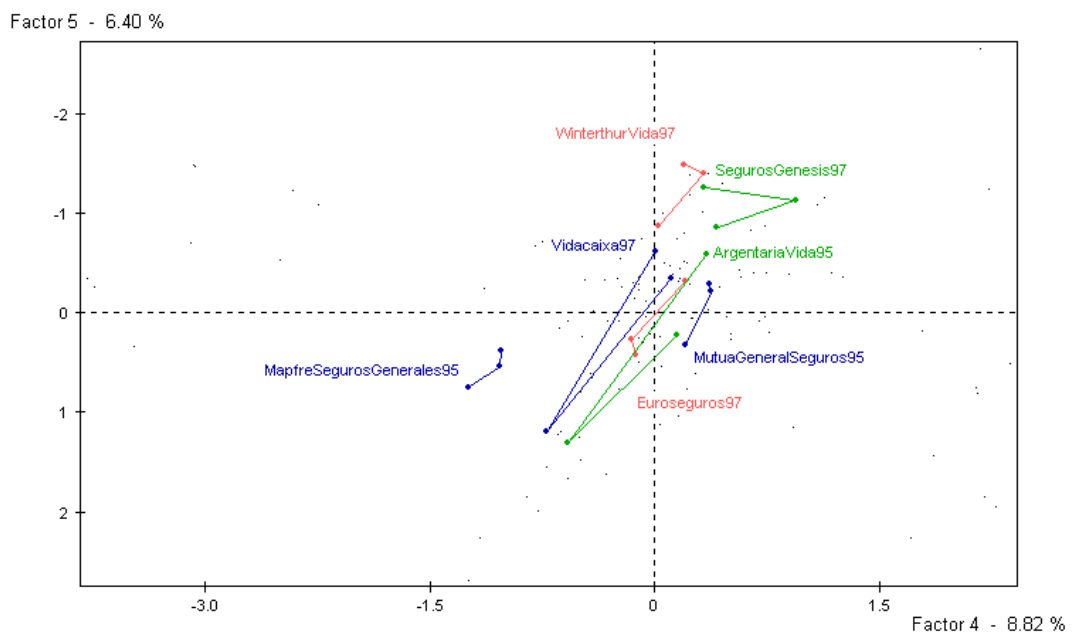


Figura III.14d
Trayectorias de individuos
Factores 4 y 5



Respecto al cuarto factor, es Zurich Seguros España el individuo que presenta un mayor cambio: primero pasa del lado positivo al negativo (entre el 95 y el 96), para ocupar una posición central respecto a ese factor en el último año considerado. También se observa un cambio de posición de la entidad Pelayo Mutua respecto al factor cuarto.

Asimismo destaca el cambio de posición de Multinacional Aseguradora, hacia la parte central y respecto al factor quinto.

b. Interpretación de resultados

Una vez descritos los resultados de la aplicación de ambas versiones vamos a proceder a su interpretación; se expresan respecto al sector asegurador en primer lugar, y respecto al método en segundo lugar, para cada una de las etapas.

1. INTERESTRUCTURA

Con relación al sector se pueden extraer las siguientes conclusiones:

— Como consecuencia de los análisis previos de las nubes de individuos y variables, dada su semejanza en dimensión y forma, se deduce que la estructura del sector representada por dichas matrices no varía de manera destacada de unos años a otros, y por tanto tiene sentido buscar una estructura común, a pesar de que haya individuos o variables que lógicamente modifiquen su comportamiento. Es decir, la estructura del sector del seguro no se modifica a corto plazo de manera significativa.

— Al analizar las matrices S obtenidas, se deduce que las compañías se comportan de manera más parecida, respecto a las variables analizadas, entre los años 1996 y 1997, que con respecto al año 1995. Esto indica que

el comportamiento del sector es más estable durante estos dos años, en los que apenas se modifica su estructura.

— La misma conclusión se obtiene al analizar la imagen euclídea del *Statis*: puesto que los objetos están bien representados, se detecta un cambio del comportamiento de las compañías aseguradoras, respecto a las variables observadas, del año 1995 al año 1996, pero parece que dicho comportamiento se mantiene respecto al año 1997. El estudio de la intraestructura permite poner de manifiesto las entidades responsables de las diferencias entre el año 95 y los otros dos, dado que esas diferencias no se deben ni a la dimensión ni a la forma de la nube.

— La imagen euclídea del *Statis dual* también pone de manifiesto un cambio o evolución en el comportamiento de las variables relativas a las empresas del sector entre el año 1995 y los otros dos. El estudio de la intraestructura permitirá explicar qué variables son las que originan tales diferencias.

Respecto al método se pueden hacer las siguientes precisiones:

— Es necesario señalar que la nube de individuos y la nube de variables obtenidas a partir de una misma tabla tienen los mismos porcentajes de inercia para todos los valores propios. El valor propio correspondiente también es el mismo pues, aunque en un caso se analiza la matriz $W_t = Z_t' M Z_t$ y en el otro la matriz $V_t = Z_t D Z_t'$, en ambos se han normado los operadores.

— El análisis de las matrices RV da una idea aproximada de las diferencias entre nubes, de individuos o de variables, según corresponda.

— Las matrices S obtenidas por ambas versiones son diferentes, pero permiten obtener conclusiones similares. Lo mismo ocurre con las dos imágenes euclídeas obtenidas, lo que corrobora la complementariedad de las dos versiones del *STATIS* en esta etapa. Esto se debe a que si el

comportamiento de los individuos varía de una tabla a otra será debido a que presentan valores diferentes para el conjunto de variables, y ese hecho se acaba de demostrar al comparar las nubes de variables que se deducen de las respectivas tablas.

2. COMPROMISO

Respecto al sector, en esta etapa no se obtienen conclusiones diferentes a las ya mencionadas respecto a la interestructura. Respecto al método, se puede señalar que el análisis de la interestructura ya permite intuir cuáles serán las nubes más ponderadas, y las que menos, en la construcción del compromiso. Aunque las ponderaciones de las nubes de individuos y de variables obtenidas por el *Statis* y por el *Statis dual* son diferentes, guardan cierta semejanza: las nubes menos ponderadas son las correspondientes al mismo año en ambos casos, al ser las que más se diferencian respecto a la estructura común.

3. INTRAESTRUCTURA

La interpretación de los resultados obtenidos en esta etapa permite caracterizar los factores, según las variables con las que están correlacionados, y la posición de los individuos sobre los mismos. En primer lugar se van a considerar los resultados del *Statis*:

— El lado positivo de este primer factor parece establecer una relación entre un alto porcentaje de negocio dedicado a seguros no vida, en especial los industriales, y altos porcentajes de gastos sobre primas, tanto externos como internos. Además, dado que las correlaciones de las variables representativas de estos gastos van creciendo con el tiempo, esta relación parece acentuarse. Teniendo en cuenta las variables que representan las veces que el margen mínimo de solvencia es mantenido, parece que son también las entidades que dedican su negocio a los seguros no vida, las que

presentan un mayor valor en esta variable. Por tanto, un alto porcentaje del negocio procedente de seguros no vida, principalmente de carácter industrial, parece ir acompañado de un mejor nivel de solvencia, aunque también una menor productividad en el uso de recursos pues sus gastos externos e internos son, en términos relativos, mayores.

— El lado negativo del ese factor, permite relacionar los altos porcentajes de seguros vida con el mantenimiento de altos porcentajes de inversiones financieras, y con una alta productividad de los capitales propios. Como las variables que representan al reaseguro también están correlacionadas de manera importante con este lado del factor, parece lógico deducir que altos porcentajes de seguros de vida, se relacionan con altos porcentajes de negocio neto sobre primas, y por tanto poca importancia del reaseguro cedido y de este sistema de cobertura de riesgos. La correlación negativa de las variables representativas de la bancaseguros con este factor indica que las entidades que emplean este canal de distribución se dedican sobre todo al negocio de seguros de vida.

— Todo lo anterior permite resaltar la eficacia del primer factor para caracterizar las entidades según su estrategia de producto vida/no vida. Así, un alto porcentaje de primas procede del seguro de vida, implica: altos porcentajes de inversiones financieras, debido a la necesidad de invertir convenientemente las provisiones matemáticas; alta productividad de los capitales propios, que no va acompañada de alta productividad de los capitales propios más las provisiones pues éstas son muy elevadas; alta productividad de los recursos que se traduce en gastos externos e internos no muy altos con relación a las primas y bajos respecto al sector, pero con un margen de solvencia mantenido menor que otro tipo de entidades. En este tipo de entidades el reaseguro cedido no tiene mucha importancia, aunque si la tiene otro fenómeno propio del sector: la bancaseguros, por lo que se deduce que las entidades bancarias que distribuyen productos de seguros se dedican sobre todo al ramo de vida. La razón hay que buscarla en el hecho de que los conocimientos necesarios para la distribución de

este tipo de seguros son similares a los que poseen los empleados de las entidades bancarias, circunstancia que no se da cuando se distribuyen otro tipo de seguros que requieren un servicio y unos conocimientos más especializados.

Estas características se oponen a las de las entidades en las que un alto porcentaje de su negocio proviene de seguros no vida, destinados a las empresas sobre todo: menor productividad de los capitales propios y de los recursos, aunque mayor margen de solvencia mantenido; baja incidencia de la bancaseguros como canal de distribución, y menor importancia de las inversiones financieras. Parece ser que es en este tipo de entidades en las que el reaseguro cedido si es importante como sistema de cobertura de los riesgos.

— Las variables de alcance relativas a los seguros de salud presentan una correlación prácticamente nula con el primer factor del *Statis*, lo que parece indicar que su estructura es diferente del resto de entidades. Por otra parte, la correlación no muy elevada de las variables AUTOS con la parte positiva de este factor sería indicio de que las entidades con porcentajes importantes obtenidos como resultado de seguros de automóviles tienen una estructura parecida al resto de entidades de no vida (y no salud), pero con ciertas diferencias.

Por ese motivo podríamos denominar al primer factor con el nombre: ORIENTACIÓN AL PRODUCTO: VIDA/NO VIDA.

— El segundo factor del *Statis* permite caracterizar sobre todo a las entidades con porcentajes importantes de seguros de salud, que presentan porcentajes de inversiones financieras respecto a las inversiones totales y más bien bajos, y altas productividades de los capitales propios más provisiones. Además, la importancia del reaseguro cedido para entidades de estas características es pequeña puesto que los valores de negocio neto sobre primas son bajos. Dado que las variables que indican el porcentaje

de los seguros de salud en las diversas entidades presentan una posición central respecto al primer factor, se puede decir que los niveles de gastos de estas entidades están en una posición intermedia entre los niveles alcanzados en las entidades dedicadas mayoritariamente a otro tipo de seguros no vida, y las que se dedican a seguros de vida.

El segundo factor puede recibir el nombre de ORIENTACIÓN AL RAMO DE SALUD.

— Aunque el segundo plano (factores 2 y 3) no tiene mucha inercia, es considerado, ya que permite obtener conclusiones importantes acerca del sector asegurador. En concreto, el tercer factor permite caracterizar de una manera más precisa las entidades con altos porcentajes de seguros de automóviles, que no habían sido bien caracterizadas anteriormente. Estas entidades presentan porcentajes de gastos externos e internos relativamente altos, pero no tan elevados como si se consideran otras entidades de seguros no vida, ni niveles tan altos de los márgenes de solvencia mantenidos sobre los mínimos exigidos. También parece claro que son entidades en las que el reaseguro cedido no es importante, y que no presentan una gran productividad de los capitales propios más provisiones. Por tanto, el tercer factor se puede designar como ORIENTACIÓN AL RAMO DEL AUTOMOVIL.

— Teniendo en cuenta la interpretación hecha para el segundo factor, que también forma parte del segundo plano, éste permite diferenciar perfectamente las entidades dedicadas a los seguros del automóvil, y las que se dedican mayoritariamente al seguro de salud tal y como hemos señalado. Por otra parte, la correlación nula del plano con las variables de vida permite concluir que este plano no aporta nada al conocimiento de las entidades que tienen altos porcentajes de primas procedentes de los seguros de ese ramo.

— El tercer plano se interpreta teniendo en cuenta que está formado por los factores 4 y 5. El primero de ellos parece establecer una clara diferencia,

en cuanto a la variable gastos externos, entre las entidades que se dedican mayoritariamente a seguros no vida (distintos de los de salud y de los automóviles), y las entidades que se dedican a los de automóviles. Así los gastos externos son más elevados para las primeras que para estas últimas entidades. Si bien, tal y como se manifestó al interpretar el primer factor en ambos, los gastos internos son elevados, en comparación con las medias del sector.

El cuarto factor se puede denominar RESTO DE SEGUROS NO VIDA.

La interpretación del factor 5 no es demasiado clara, si bien es el que permite diferenciar las entidades que, dentro del seguro de vida, distribuyen sus productos a través de oficinas bancarias, y que presentan un porcentaje inferior de inversiones financieras. Por ese motivo podríamos decir que es el factor BANCASEGUROS.

— Estas interpretaciones se matizan con los resultados obtenidos, respecto a las variables, con el *Statis dual*, y con los obtenidos al aplicar el *Análisis Cluster*, destacando la importancia del factor y poniendo de manifiesto, de forma empírica, la complementariedad de los métodos que ya se ha expresado en la parte teórica de este estudio.

— Los rasgos generales del sector son los que se han puesto de manifiesto a través de las interpretaciones dadas, que permiten caracterizar la estructura común del mismo, algunas de las cuales ya se habían señalado cuando se observaban directamente los datos relativos al sector en la primera parte de este capítulo.

Teniendo en cuenta la denominación y el significado de los factores, se caracterizan a continuación los individuos que ocupan posiciones destacadas sobre los mismos:

— Por su posición respecto al primer factor, entidades como Bankinter Vida, Bansabadell Vida, Lagun Aro Vida, Ascat Vida, Barclays Vida, Biharko

Vida, Vidacaixa, Antares Vida, Argentaria Vida y Euroseguros se caracterizan por una productividad mayor de los recursos que la media del sector (bajos porcentajes de gastos externos e internos), pero con unos márgenes de solvencia mínima mantenidos poco elevados. El porcentaje que las inversiones financieras representan sobre el total de inversiones es más elevado que para el resto de entidades del sector, lo mismo que ocurre con la prima por empleado, por lo que posiblemente serán entidades en las que la bancaseguros juega un papel importante como canal de distribución. Tampoco son entidades en las que el reaseguro sea un sistema importante para la cobertura de riesgos, pero sí presentan una productividad por encima de la media de los capitales propios. Dada la posición extrema de Bankinter Vida esta sería la entidad mejor definida por las características señaladas. Todas estas entidades pertenecen a grupos bancarios o de Cajas de Ahorros.

— Los rasgos que definen a entidades como el Santander, Caja Madrid Vida, Aseguradora Valenciana, Db Vida, Intercaser, Seguros Génesis y Winterthur Vida son básicamente los mismos que los ya señalados para las entidades anteriores, pero menos acentuados. Las diferencias entre unos y otros se manifestarán en el tercer plano.

— En el lado opuesto del factor destacan, bastante bien representadas, entidades como Musini, Zurich Internacional y Crédito y Caución. Estas entidades también presentan una coordenada, respecto al lado negativo del segundo factor, de cierta importancia. Por todo ello serán entidades en las que el porcentaje de primas procedentes de seguros no vida, especialmente el procedente de seguros industriales, es muy importante. Además serán entidades en las que el porcentaje de negocio procedente de seguros de salud es prácticamente nulo. Los gastos externos e internos, sobre todo éstos, serán elevados respecto al resto de entidades del sector asegurador, pero los márgenes de solvencia mantenidos serán, sobre el mínimo, bastante más elevados. Sin embargo, dada su posición en el plano, la productividad de los capitales propios y de los capitales propios más las provisiones es inferior a la media.

— El resto de entidades que están bien representadas en el plano ocupan una posición destacada sobre el primer factor, en el lado positivo del mismo, pero presentan una coordenada prácticamente nula con el segundo factor. Por este motivo entidades como Fiact Mutua, Winterthur Seguros Generales, Mapfre Seguros Generales y Finisterre, pueden ser caracterizadas por mantener bajos (o nulos) porcentajes de negocio procedentes de seguros de vida y de seguros de salud. La cifra de gastos externos e internos, sobre las primas, sería bastante alta, pero los márgenes de solvencia mantenidos, sobre el mínimo, no serían muy elevados. La productividad de los capitales propios sería inferior a la media, pero la de los capitales propios más las provisiones sería un poco más elevada. El porcentaje que las inversiones financieras representan sobre las totales tomaría un valor bajo, lo mismo que la prima por empleado.

— Por otra parte, solo tres de los individuos de dicha representación presentan una posición destacada respecto al segundo factor: Asisa, Adeslas y Sanitas, que son entidades que dedican casi el 100% de su negocio a seguros de salud. Como consecuencia de la interpretación de este factor, es posible caracterizar dichas aseguradoras como entidades con alta productividad de sus capitales propios más provisiones, porcentajes de gastos externos e internos intermedios respecto a las medias del sector, y márgenes de solvencia mantenidos no muy elevados.

— Casi todos los individuos están bien representados, excepto Bilbao Cia. Anma., Athena, Caser Grupo, La Estrella, Mapfre Vida, Allianz Ras. Conviene señalar que los cuatro últimos individuos tampoco van a ser representados de forma adecuada en los otros planos, hecho que se deduce al analizar sus cosenos cuadrados. Por este motivo, es posible señalar que sus características difieren de la estructura común al resto de entidades, al menos en los años considerados.

— Las características apuntadas para Winterthur Seguros Generales o Mapfre Seguros Generales son las que definen en este caso a entidades

como Plus Ultra, Zurich Seguros España, Mutua General de Seguros, AGF Unión y Fenix, Assicurazioni Generali, Catalana Occidente o Vitalicio, pero con las características menos acentuadas ya que ocupan posiciones más centrales respecto al primer factor.

— El tercer factor permite diferenciar a Zurich Internacional de las otras dos entidades que ocupan una posición similar en el primer plano: Crédito y Caución, y Musini. Teniendo en cuenta la interpretación del tercer factor, parece que la diferencia entre dichas entidades se debe a la existencia en Zurich Internacional de una parte de negocio procedente de seguros de automóviles, circunstancia que no se produce en las otras dos entidades.

— Asimismo, el mismo factor permite establecer diferencias entre entidades que en el primer plano se situaban en el lado positivo del primer factor, algunas mal representadas. Las diferencias se establecen básicamente respecto a la importancia, en estas entidades, del negocio procedente de las primas de seguros de automóviles, en entidades como Mutua Madrileña, Multinacional Aseguradora, Mapfre Mutualidad, Pelayo Mutua, Zurich Seguros España, Winterthur Seguros Generales, Athena y Bilbao Cia. Anma. Dada la correlación de este factor con las variables que hacían referencia al reaseguro, es posible añadir que, en este tipo de entidades, este sistema de cobertura de riesgos no juega un papel muy importante.

— El cuarto factor establece diferencias entre entidades que parecían similares al analizar el primer plano factorial: mientras Mutua Madrileña, Mapfre Mutualidad y Pelayo Mutua se sitúan en la parte positiva de dicho factor, en el lado negativo ocupan una posición destacada entidades como Ocaso, Santa Lucia y Finisterre. Según la interpretación del factor, hay una clara diferencia, en cuanto a la variable gastos externos, entre las entidades que se dedican mayoritariamente a seguros no vida (distintos de los de salud y de los automóviles), como Santa Lucia, Finisterre y Ocaso, y las entidades que se dedican a los seguros de automóviles. Así, los gastos externos son más elevados para las primeras que para entidades como

Mutua Madrileña, Mapfre Mutualidad y Pelayo Mutua. Si bien, tal y como se manifestó al interpretar el primer factor, en ambos casos los gastos son elevados en comparación con las medias del sector.

— El quinto factor establece una clara diferencia entre las entidades que se dedicaban mayoritariamente a los seguros de vida como Mapfre Vida, Seguros Génesis y Winterthur Vida, y otras como Bankinter Vida, Bansabadell Vida: las primeras presentan porcentajes superiores en la variable inversiones financieras sobre inversiones totales, pero una prima por empleado menor. Las características opuestas parecen definir a las otras entidades.

La interpretación del dendograma obtenido al aplicar el *Análisis Cluster* permitirá matizar las conclusiones obtenidas sobre las características de los individuos, en cuanto que el mismo permite clasificarlos teniendo en cuenta, de una manera global, todos los factores comentados, aún así, con lo expuesto, la capacidad explicativa de los factores obtenidos como resultado del método *Statis* queda ya demostrada.

Con relación a los resultados obtenidos al aplicar el *Statis dual* se pueden hacer las siguientes consideraciones:

— Teniendo en cuenta la correlación con las variables, el primer plano factorial obtenido con el *Statis dual* tiene una interpretación similar al del *Statis*, y también una inercia similar: un 57,64% en el primer caso frente al 55,64% en el caso del *Statis*. Como consecuencia, su interpretación permite confirmar el significado otorgado a los factores, y por tanto las características básicas de la estructura del sector.

— En los planos segundo y tercero la situación es semejante: el segundo explica el 31,78% de la inercia siendo del 31,27% en el caso del *Statis*; para el tercero la diferencia es un poco más grande (13,33% para el *Statis* y para el *Statis dual* 15,22%) pero tampoco es importante. En ambos casos la interpretación es similar a la obtenida para los factores del *Statis*.

— Debido a que los factores obtenidos por ambas versiones están dotados del mismo significado, la representación de los individuos del *Statis dual* permite analizar las trayectorias de éstos más representativas en un espacio óptimo, sin necesidad de emplear individuos suplementarios.

— El posicionamiento de los individuos en los planos del *Statis dual* permite representar los individuos por años (cada tabla) y analizar su situación. El inconveniente de este tipo de gráficas es el gran número de individuos, aunque siempre es posible hacer representaciones parciales. Este aspecto concreto del método puede ser analizado en estudios posteriores.

Como consecuencia de esta etapa, se pueden hacer algunas apreciaciones más sobre la metodología STATIS:

— Como en un ACP, la selección de factores depende de su interpretación y de las conclusiones que permiten obtener. Así, en nuestro caso, se han tomado en consideración planos que, a pesar de su poca inercia, aportan información relevante para el sector.

— La aplicación del método permite caracterizar mejor cierto tipo de seguros de los cuales conseguir información sería una tarea complicada. Tal es el caso de los seguros industriales, para los cuales las publicaciones no aportan información agregada como tal, sino de manera particular para cada uno de los ramos que están integrados dentro de esa categoría. Además, permite confirmar las características ya expuestas para otros tipos de seguros.

— La estructura común del sector se pone de manifiesto de manera sencilla a través de las correlaciones entre los factores y las variables.

— Como consecuencia, la aplicación de la metodología STATIS caracteriza el sector de manera más precisa que con otras técnicas factoriales, reafirmando, por tanto, su validez con fines descriptivos.

4. TRAYECTORIAS

Las trayectorias de las variables, obtenidas como resultado del *Statis* y de la descomposición de distancias, permiten afirmar que, aunque hay ciertas variables que modifican su comportamiento de unos años a otros, en general las correlaciones de las variables respecto a los factores permanecen estables, comportamiento que facilita la interpretación de los factores.

Las trayectorias de los individuos, analizadas a través de la descomposición de las distancias y del posicionamiento de los mismos en los factores del *Statis dual*, permiten detectar las entidades aseguradoras que modifican su comportamiento en el periodo considerado:

— En el primer plano factorial los cambios de posición más importantes son las que presentan las entidades Banesto, Antares Vida y Barclays Vida. Para Banesto, que cambia del lado positivo al lado negativo del primer factor, ese cambio implica un aumento del porcentaje de su negocio en seguros de vida, así como una disminución considerable de los gastos externos e internos (respecto a las primas), y del margen de solvencia mantenido. El cambio de posición de las otras dos entidades debe ser interpretado como un reforzamiento de su actividad en el ramo de seguros de vida, y un aumento de la productividad de sus capitales propios.

— Otras tres entidades que presentan trayectorias significativas son Crédito y Caución, Zurich España y Mapfre Mutuality, extremando sus posiciones respecto a la parte positiva del primer factor, lo que puede interpretarse como un aumento del porcentaje de gastos externos e internos (sobre primas), y un aumento de los márgenes de solvencia mantenidos. En concreto, Musini y Zurich Internacional presentan trayectorias cambiantes, pero su posición al final del periodo es más extrema respecto al primer factor, por lo que su trayectoria se puede interpretar como en el caso anterior. Las trayectorias de Allianz Ras, Vitalicio, Caser Grupo, y AGF Unión y Fénix no tienen una clara interpretación debido a su posición central.

— Adeslas y Sanitas extreman sus posiciones hacia el lado positivo del segundo factor, lo que permite destacar la consolidación de su actividad en el ramo de los seguros de salud.

— La aseguradora Santander extrema su posición en la parte negativa del primer factor, como consecuencia seguramente de un aumento de las primas procedentes del seguro de vida, y de una disminución de los gastos externos e internos. La trayectoria de Ascat Vida se explica teniendo en cuenta el aumento de la productividad de los capitales propios y de la prima por empleado, siendo la situación contraria para Bankinter Vida.

— Bansabadell Vida extrema su posición hacia el lado negativo del primer factor, lo que es un indicio claro del aumento de la prima por empleado y de la productividad de sus capitales propios. La trayectoria de Lagun Aro Vida, en sentido contrario, es el resultado de un descenso de las mismas variables.

— En el caso del segundo plano, cabe destacar la trayectoria de Pelayo Mutua cuyo movimiento puede ser debido al incremento de la importancia de los seguros de automóviles en su actividad. La trayectoria de Multinacional Aseguradora, en sentido contrario, responde a una reducción de ese ramo de seguros en su negocio.

— Aunque la inercia del tercer plano factorial es pequeña, este permite representar adecuadamente las trayectorias de ciertos individuos. Así, respecto al cuarto factor, es Zurich Seguros España el individuo que presenta un mayor cambio probablemente debido a las modificaciones en sus gastos externos, que han ido aumentando en el período. La evolución Pelayo Mutua, respecto al mismo factor, se debe a su cambio de estructura, en cuanto refuerza su atención hacia los seguros del automóvil, como ya se había manifestado.

— Destaca el cambio de posición, respecto al factor quinto, de Multinacional Aseguradora, como consecuencia de un aumento de la

productividad de los capitales propios, que en el año 95 era negativa. También destacan las trayectorias de Barclays Vida y Antares Vida, cuyo cambio de posición respecto al mismo factor puede venir motivado por una disminución del porcentaje de inversiones financieras sobre las totales, y por un aumento de la prima por empleado y de la productividad de los capitales propios. La aseguradora Musini también presenta cambios debidos, posiblemente, a las modificaciones en la parte de su negocio destinada a seguros diferentes de vida, salud y automóviles.

Con el fin de completar el estudio del sector, en el siguiente apartado se procede al establecimiento de grupos de entidades aseguradoras homogéneas en función de las variables consideradas, y tomando como base las coordenadas factoriales obtenidas anteriormente con el *Statis*.

2.2.3 Determinación de grupos estratégicos en el sector asegurador español: *Statis* y *Cluster*

El objetivo de este epígrafe es poner de manifiesto de manera empírica la complementariedad entre el método *Statis* y el *Análisis Cluster* en la descripción y clasificación de una serie de individuos, que son en este caso las entidades aseguradoras. La aplicación del *Análisis Cluster* sobre los factores obtenidos permite agrupar dichas entidades, y el posicionamiento posterior de los grupos sobre los factores ayuda a su caracterización, profundizando en el conocimiento de la estructura del sector.

Cuando se intenta establecer grupos estratégicos en una industria o sector, el primer problema a resolver es la selección de las dimensiones relevantes para definir tales grupos, ya que éstas varían según el tipo de industria de que se trate. Entre las más utilizadas en los estudios empíricos están el grado de especialización en productos y mercados, el grado de integración

vertical, la calidad, la imagen de marca, el liderazgo tecnológico, etc. (Ventura Victoria, 1994: 163). Si bien, dada la variedad de opciones al alcance de las empresas, se plantean problemas metodológicos a la hora de detectar las variables independientes con mayor capacidad de particionar el conjunto de las empresas.

Por este motivo, en algunos casos, en lugar de considerar directamente las variables estratégicas de un sector o industria, para la clasificación de los grupos se consideran los factores obtenidos a partir del tratamiento previo de dichas variables por un Método Factorial (Pinillos Costa y Martín Peña, 1998: 43), aplicando posteriormente el *Análisis Cluster* para establecer los grupos. Esta forma de proceder enlaza con la opinión de Lebart *et al.* (1995: 147), según la cual los métodos de clasificación no pueden mostrar siempre la importancia de ciertas tendencias existentes, y el empleo previo de Métodos Factoriales resuelve este problema; el posicionamiento de las clases o grupos sobre los ejes factoriales permite la identificación de ciertas tendencias o fenómenos latentes. Además, cuando la clasificación se hace sobre el subespacio definido por los factores más significativos (que recogen la mayor parte de la información), los grupos obtenidos consideran la dimensión real de la nube de puntos, contribuyendo a su homogeneidad.

Como sucede en este caso, y según los mismos autores Lebart *et al.* (1995: 185-6), cuando el número de individuos es muy grande las gráficas resultantes de un Método Factorial resultan difíciles de interpretar si se representan todos a la vez. La organización y clasificación de los individuos en grupos homogéneos, mediante la aplicación posterior de *Técnicas Cluster*, permite matizar la interpretación de las posiciones de los individuos y corrigen ciertas deformaciones debidas a la proyección en planos de, como mucho, tres dimensiones. De esa manera ambos tipos de métodos se validan mutuamente.

En este caso, el Método Factorial empleado es el *Statis*, que con el *Análisis Cluster*, permitirá determinar la tipología de grupos correspondiente.

A diferencia de otros métodos empleados en algunos de los estudios sobre grupos, como el *Análisis de Componentes Principales* (ACP), cuando se consideran variables para más de un año el *Statis* permite obtener una única clasificación de individuos, y no una para cada año considerado, aprovechando así mejor la información disponible, y eliminando el problema de la comparación posterior de los distintos dendogramas obtenidos (Pinillos Acosta, 1998). Por otra parte, entre los métodos apropiados para tablas de tres entradas, el *Statis* es el adecuado cuando el objeto de estudio es el establecimiento de una tipología de individuos (p. 19). Desde el punto de vista metodológico tampoco sería correcto, con el fin de obtener un único dendograma, aplicar un ACP a todas las variables pues no se tendría en cuenta que pertenecen a años diferentes.

Una vez hechas estas consideraciones, se procede a la aplicación conjunta de *Statis* y *Cluster*.

La información de partida es la descrita anteriormente, es decir, las 48 entidades aseguradoras de la tabla III.14 (pp. 388-9), sobre las que se han observado las variables especificadas en la tabla III.15 (p. 394), para los años 1995, 1996 y 1997. Estas variables son las que, desde el punto de vista estratégico, mejor pueden definir la estructura y los grupos de entidades del sector. Dado que, en general, los trabajos realizados no permiten establecer diferencias entre las variables de resultados de unos grupos a otros, (STRATEGOR, 1995: 20), en este caso tampoco se ha considerado ninguna variable de este tipo.

Después de analizar dicha información mediante el método *Statis* (apartado anterior), se aplica el *Análisis Cluster* sobre los factores obtenidos en la etapa de la intraestructura. Se obtiene así una clasificación de entidades que se interpretará convenientemente. Su descripción detallada y su posicionamiento en los planos factoriales pondrá fin al estudio descriptivo.

1. APLICACIÓN DEL METODO FACTORIAL: el método *Statis*

Para la posterior aplicación del *Análisis Cluster* es necesario hacer una serie de precisiones con relación a los resultados obtenidos:

- *La interestructura*

De esta primera etapa, y dado que la conclusión obtenida fue la semejanza de nubes, se deduce que de una manera global, las entidades analizadas no han cambiado su posición estratégica en los años comentados. Por ese motivo tiene sentido un intento de clasificación de las entidades que considere, de manera conjunta, la información de los tres años.

- *El compromiso*

En esta etapa, cada nube de individuos es ponderada adecuadamente con el fin de obtener los valores medios del período para los individuos analizados, respecto a las variables observadas, y que conforman la estructura del sector en dicho espacio de tiempo.

- *La intraestructura*

La estructura común para todos los años, obtenida en la etapa anterior, es descrita a través de los factores, interpretados según las correlaciones con las distintas variables. Esos factores son, por tanto, las dimensiones relevantes del sector y los que permiten clasificar adecuadamente los individuos analizados (entidades aseguradoras, en este caso). Así, esas dimensiones, en cuanto que son el resultado de una Método Factorial, son independientes y resumen el máximo de información posible.

El número de factores a considerar como relevante depende de su inercia, como ya hemos indicado, y también de la información que proporcionan. En este caso, ya se ha justificado la elección de cinco para describir el sector, siendo éstos, precisamente, las dimensiones relevantes para la realización de grupos estratégicos; la comparación

de dendogramas, cuya diferencia es el número de factores de partida considerados, permite matizar la interpretación ya otorgada a dichos factores y justificar la necesidad de incluir esos cinco. Los valores de los mismos se recogen en el epígrafe A.2.1.6 del anexo (p. 584).

2. ANÁLISIS CLUSTER

La aplicación al caso del sector asegurador se basa en las etapas descritas en el capítulo II (p. 247), que configuran el desarrollo de esta técnica:

- *Selección de las variables*

La primera etapa consiste en elegir las características o variables relevantes para identificar los grupos o *clusters*. En este caso, ya se ha indicado que se consideran como tales los cinco primeros factores obtenidos a partir de la infraestructura del *Statis*, y que teniendo en cuenta su significado y su importancia en términos de inercia explicada, constituyen las dimensiones básicas para caracterizar el sector asegurador. También se van a analizar los dendogramas obtenidos al emplear dos, tres y cuatro factores, ya que las diferencias entre ellos permitirán añadir información sobre sus características.

- *Elección de la medida de asociación*

La determinación de los grupos (tanto de individuos como de variables) lleva implícita una medida que permite analizar la proximidad óptima de los objetos en estudio. Cuando se agrupan individuos, la proximidad se suele expresar en términos de distancia, mientras que en el caso de las variables se trata de medidas del tipo de coeficientes de correlación. Como partimos de coordenadas factoriales y el objeto es el agrupamiento de individuos, la medida de asociación empleada ha sido la distancia euclídea.

- *Elección de la técnica*

Una vez que se ha especificado la medida de agrupamiento, el siguiente paso corresponde a la selección de la *técnica cluster* más adecuada para el estudio. En este caso vamos a emplear un *Método jerárquico*, en concreto el Método de Ward. La selección de este método se debe a que es el que proporciona mejores resultados para clasificaciones a partir de factores y es, por ello, el único contemplado en el paquete estadístico SPAD (versión 3.5).

- *Validación e interpretación de resultados*

En primer lugar se pretende validar el número de factores empleados en la clasificación, comparando los diferentes dendogramas obtenidos. Posteriormente se intentará justificar el número de grupos seleccionados y finalmente se comentará más detenidamente el dendograma global (figura III.18) con el fin de obtener conclusiones sobre el sector y los grupos estratégicos. Para la validación de resultados, pueden emplearse también otras técnicas de general aceptación para dicho propósito, como el *Análisis de la Varianza* o el *Análisis Discriminante*; si bien en este caso, la aplicación de este tipo de técnicas excede nuestro objetivo.

— La comparación de dendogramas, obtenidos con distinto número de factores de partida, permite justificar que los cinco elegidos son los adecuados, y que son los que deben tenerse en cuenta para el establecimiento de los grupos, matizando además el significado que se otorgó a los mismos al analizar sus correlaciones con las variables en la etapa de la intraestructura del método *Statis*. La comparación entre el dendograma a partir de dos factores (figura h, epígrafe A.2.3: p. 606), y el obtenido a partir de tres factores (figura i, del mismo epígrafe, p. 607), permite matizar el significado del tercer factor. Los tres grandes grupos establecidos por ambos dendogramas son los mismos: entidades dedicadas

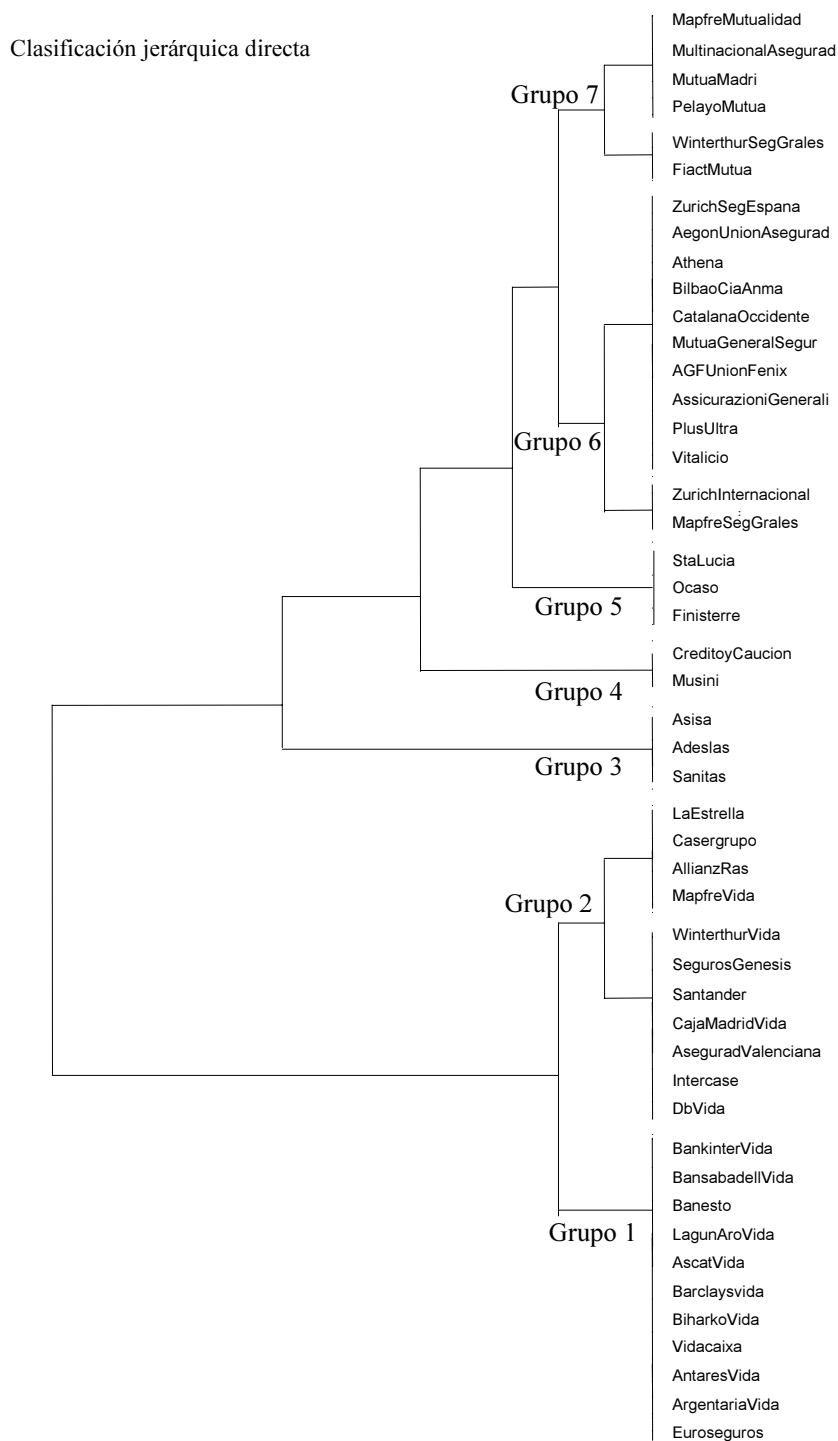
a los seguros de vida, entidades dedicadas a seguros de salud, y las que podríamos designar como entidades dedicadas a otros seguros no vida. Si bien, en el caso de emplear tres factores, dentro del último grupo mencionado, se hace una distinción entre entidades según la importancia que tienen en su negocio las primas procedentes de los seguros de automóviles. Este hecho permite reafirmar el significado que se había otorgado al tercer factor, relacionado con los seguros de automóviles. Como consecuencia, el dendograma con tres factores establece grupos estratégicos de una manera más completa, de acuerdo con la estructura del sector.

Entre el dendograma con tres factores comentado y el obtenido con cuatro factores (figura j, epígrafe A.2.3: p. 608) también existen diferencias que hacen posible subrayar la importancia del nuevo factor incluido, el cuarto. Así, los tres grandes grupos son los mismos, con las mismas entidades, pero el nuevo conjunto permite matizar los formados con las entidades de otros seguros no vida. De esta manera se identifica un nuevo grupo formado por entidades que no dedican su negocio a los seguros de automóviles, como son Santa Lucía, Ocaso y Finisterre. Por otra parte, parece que la mayoría de las entidades con una estructura mutualista, se clasifican aparte, y son entidades en las que gran parte de su negocio procede de los seguros de automóviles.

Al comparar el dendograma que se acaba de comentar y el obtenido con los cinco primeros factores (figura III.15), que se recoge a continuación, se completa la interpretación del nuevo factor incluido. Mientras los tres grupos generales son los mismos, las diferencias se establecen respecto a las entidades incluidas en esos grupos. Dada su peculiar estructura, aquellos individuos mal representados por la estructura común del sector (La Estrella, Caser Grupo, Allianz Ras y Mapfre Vida, como ya se comentó previamente) se consideran dentro del grupo de seguros de vida, mientras que en los anteriores dendogramas estaban clasificados en el grupo de otros seguros no vida. Por otra parte, este último dendograma permite establecer diferencias dentro de las entidades de vida que no se habían establecido con los anteriores. Parece ser que ciertas entidades, debido a

su consideración como entidades de bancaseguros, tienen una estructura similar entre ellas. Como consecuencia de lo anterior, creemos que es el último de los dendogramas el mejor para el establecimiento de los grupos.

Figura III.15
Clasificación de las entidades aseguradoras con cinco factores



— En cuanto al número de grupos a considerar es necesario hacer una precisión. Como señalan cierto autores (Fiegenbaum y Thomas, 1995: 463), “las compañías analizan el comportamiento de organizaciones de referencia similares (grupos estratégicos) en el mismo ambiente competitivo, en búsqueda de nuevas opciones estratégicas”. Dado que los grandes grupos parecen claros, el nivel de desagregación considerado va a depender, por tanto, de cada entidad. Según la afirmación comentada parece que lo más adecuado sería fijarse en el grupo más pequeño que pueda considerarse, para tomarlo como referencia, pero sin olvidar las características del grupo más general al que pertenece.

Como en nuestro caso no se trata de establecer el grupo de referencia similar a una entidad concreta, sino de establecer los grandes grupos estratégicos del sector, nos ocupamos de analizar el dendograma de la figura III.15, considerando los tres grandes grupos que se han obtenido con todos los dendogramas analizados y haciendo referencia a los subgrupos establecidos por el último de ellos, que es el que proporciona una clasificación más estricta.

La interpretación de este dendograma da lugar a los siguientes grupos:

- *Entidades de seguros de vida*

Está formado por dos grupos: el 1 que incluye las entidades de vida y con una importancia especial del canal de distribución bancario (bancaseguros); el grupo 2, que lo forman otras entidades de vida.

Así tenemos:

Grupo 1:

- Bankinter Vida
- Bansabadell Vida
- Banesto
- Lagun Aro Vida
- Ascat Vida
- Barclays Vida

- Biharko Vida
- Vidacaixa
- Antares Vida
- Argentaria Vida
- Euroseguros

Según las interpretaciones hechas al analizar los resultados del *Statis*, se puede afirmar que la entidad Banesto se incluye en este grupo debido al cambio de organización que experimentó entre el año 95 y 97, y que se puso de manifiesto en el estudio de las trayectorias. Este hecho constituye un buen ejemplo de complementariedad entre los resultados de ambos métodos.

Grupo 2:

- La Estrella
- Caser Grupo
- Allianz Ras
- Mapfre Vida

Estas entidades, sobre las que ya se ha hecho alguna consideración, tienen unas características especiales y aparecen en este grupo porque parte de su negocio proviene de seguros de vida, aunque sus características difieren de las generales del sector. También se incluyen en el grupo 2 otras aseguradoras:

- Winterthur Vida
- Seguros Génesis
- Santander
- Caja Madrid Vida
- Aseguradora Valenciana
- Intercaser
- Db Vida

Algunas de estas entidades también distribuyen sus productos a través de la bancaseguros, pero su estructura es diferente a las del

grupo 1. El posicionamiento del grupo sobre los factores permitirá establecer los rasgos diferenciadores de las mismas.

- *Entidades de seguros de salud*

Originan el grupo 3 y las aseguradoras incluidas en él son:

- Asisa
- Adeslas
- Sanitas

La claridad con que se manifiesta este grupo en todos los dendogramas permite resaltar las características diferenciadoras de las entidades que se dedican a este tipo de seguros

- *Entidades de otros seguros no vida*

Se pueden distinguir dos grupos: el primero es el dedicado exclusivamente a seguros que no son de salud y autos, que va a denominarse grupo 4. Lo forman dos entidades con características especiales debido a los ramos a los que dedican prácticamente su negocio: los seguros de tipo industrial y muy especializados. Dichas entidades son:

- Crédito y Caución
- Musini

El segundo grupo mencionado es el dedicado a seguros generales (que pueden estar autorizados en toda clase de ramos, teniendo en cuenta las limitaciones que se comentaron para los seguros de vida).

Dentro de éste se distinguen los grupos 5, 6 y 7:

Grupo 5:

- Santa Lucia
- Ocaso
- Finisterre

Estas entidades no cuentan con primas procedentes de seguros de automóviles.

Grupo 6:

- Zurich Seguros España
- Aegón Unión Aseguradora
- Athena
- Bilbao Cia. Anma.
- Catalana Occidente
- Mutua General de Seguros
- AGF Unión y Fénix
- Assicurazioni Generali
- Plus Ultra
- Vitalicio
- Zurich Internacional
- Mapfre Seguros Generales

En estas aseguradoras las primas de seguros de automóviles no tiene especial relevancia; no obstante, las dos últimas entidades parecen tener unas características un tanto diferentes.

Grupo 7:

- Mapfre Mutualidad
- Multinacional Aseguradora
- Mutua Madrileña
- Pelayo Mutua
- Winterthur Seguros Generales
- Fiact Mutua

Aquí se incluyen entidades en las que los seguros de automóviles tienen una importancia especial, algunas con una estructura mutualista. La descripción más detallada de estos grupos se realiza a continuación.

3. DESCRIPCIÓN Y POSICIONAMIENTO DE LOS GRUPOS EN LOS PLANOS FACTORIALES

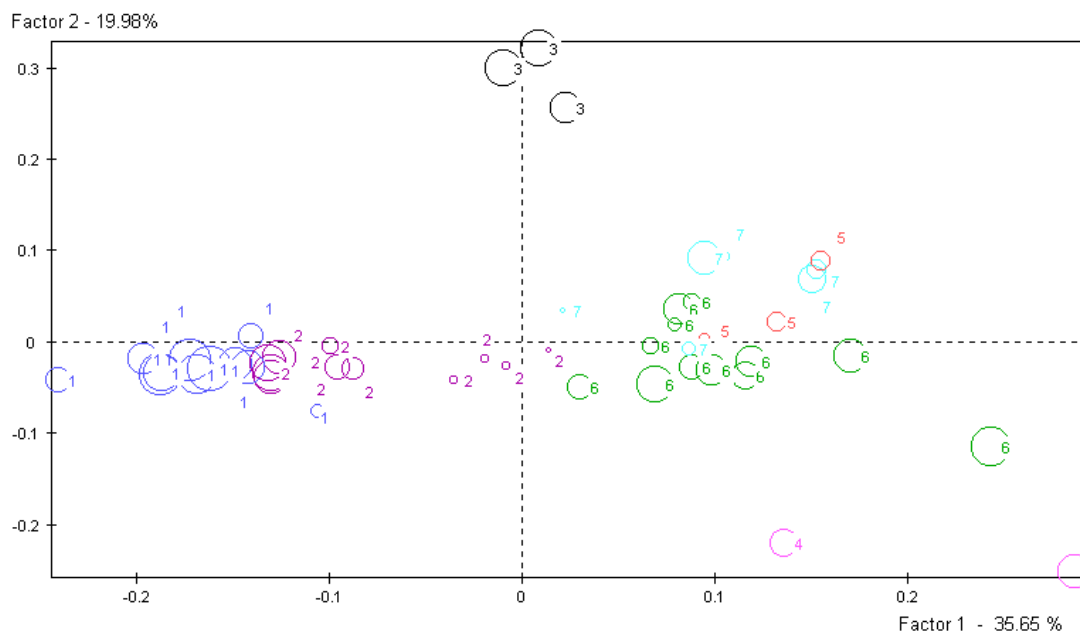
El posicionamiento de los grupos obtenidos en los planos factoriales que se han determinado en la infraestructura permite su análisis de acuerdo con la estructura común del sector, teniendo en cuenta la interpretación otorgada a los factores que los forman, y con la posibilidad de examinar la homogeneidad de dichos grupos.

Con esta fase se ponen de manifiesto, de manera más explícita, las ventajas del empleo del método *Statis* y del *Análisis Cluster* en el análisis de los grupos estratégicos. A nuestro juicio, la descripción de los grupos por medio de los factores permite definir las características en términos de variables concretas de cada grupo, facilitando a otras entidades del sector, no incluidas en el estudio, la selección de su grupo de referencia.

Por otra parte, el empleo de un Método Factorial previo elimina la necesidad de comparar las medias de los grupos considerados con las obtenidas para todo el sector, tal y como se hace en algunos estudios de grupos estratégicos (Pinillos Costa y Martín Peña, 1998: 62-70). Esto se debe a las características propias del análisis: el alejamiento de un individuo del origen y en la dirección de un factor significa que presenta valores superiores a la media para las variables relacionadas con ese extremo del factor, y valores por debajo de la media para las variables que se relacionan con el extremo opuesto.

Las figuras III.19, III.20 y III.21 representan los grupos en el primer, segundo y tercer planos factoriales, respectivamente.

Figura III.16
Posicionamiento de los grupos obtenidos
Factores 1 y 2



La interpretación de esta representación gráfica permite extraer conclusiones muy interesantes. Como era de suponer, los grupos 1 y 2, como grupos de vida que son, se colocan en el extremo negativo del primer factor. Las diferencias entre ellos vendrán marcadas por otros factores y otros planos, como se verá más adelante. Dada su posición en este caso, serán entidades caracterizadas por altos porcentajes de inversiones financieras, debido a la necesidad de invertir convenientemente las provisiones matemáticas; alta productividad de los capitales propios (que no va acompañada de alta productividad de los capitales propios más las provisiones pues éstas son muy elevadas); alta productividad de los recursos que se traduce en gastos externos e internos no muy altos con relación a las primas, pero con un margen de solvencia mantenido menor que otro tipo de entidades. Para estas entidades el reaseguro cedido no tiene mucha importancia, aunque si la tiene otro fenómeno propio del sector: la bancaseguros (cuya importancia depende del grupo considerado).

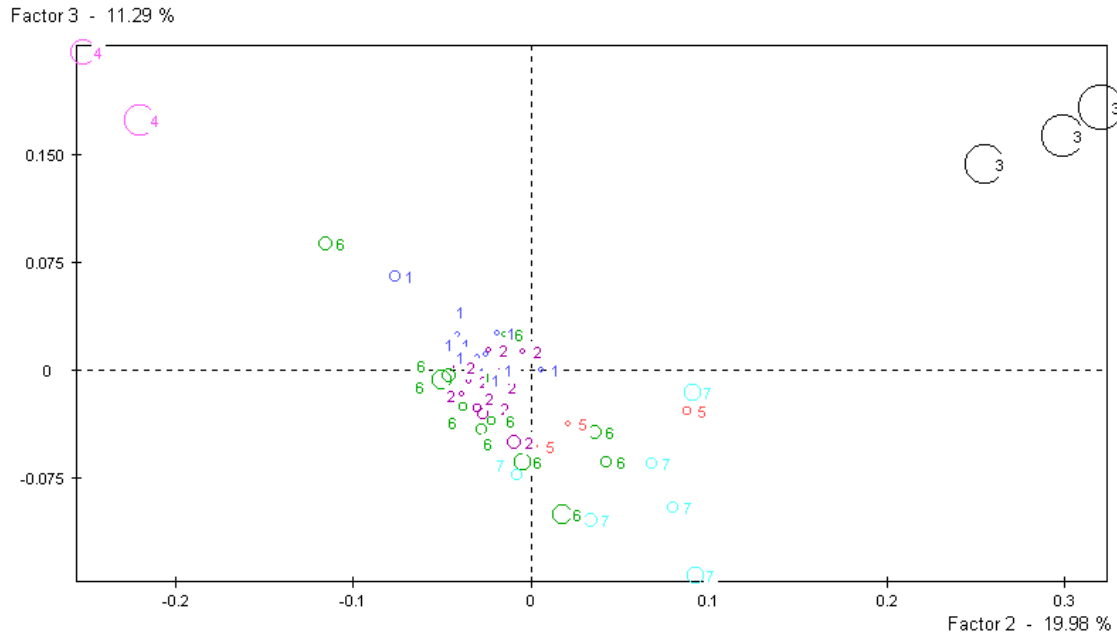
Conviene señalar que dentro del grupo 2 hay comportamientos heterogéneos que se ponían ya de relieve con el dendograma, respecto a las entidades La Estrella, Caser Grupo, Allianz Ras y Mapfre Vida. Como se deduce de la representación obtenida, dada su coordenada negativa con este factor, su estructura parece estar más próxima a las entidades de vida que a otro tipo de entidades.

El grupo de entidades de la salud, grupo 3, aparece claramente posicionado sobre el segundo factor. Por ello, teniendo en cuenta las interpretaciones del mismo, serán entidades con altas productividades de los capitales propios y de las provisiones y en las que la importancia del reaseguro cedido no es significativa. Por otra parte, dada su posición respecto al primer factor, los porcentajes de gastos externos e internos sobre las primas están en torno a la media del conjunto del sector.

El plano permite definir adecuadamente el grupo 4, formado por Musini, y Crédito y Caución. Según su posición en el plano, son entidades en las que el porcentaje de primas procedentes de seguros no vida, en concreto de los seguros industriales, es muy importante. Los gastos externos e internos, sobre todo estos últimos, serán elevados respecto al resto de entidades del sector asegurador, y también los márgenes de solvencia mantenidos sobre el mínimo. Sin embargo, la productividad de los capitales propios y de los capitales propios más las provisiones es inferior a la media.

Los grupos 5, 6 y 7 presentan una posición similar en el plano, pues están situados todos ellos en el extremo positivo del primer factor. Por ello, pueden caracterizarse por ser entidades con un alto porcentaje de negocio dedicado a seguros no vida, en especial los industriales, y altos porcentajes de gastos sobre primas, tanto externos como internos. Según la interpretación ya realizada de este lado del factor, parece que son entidades con márgenes de solvencia mantenidos elevados sobre el mínimo exigido.

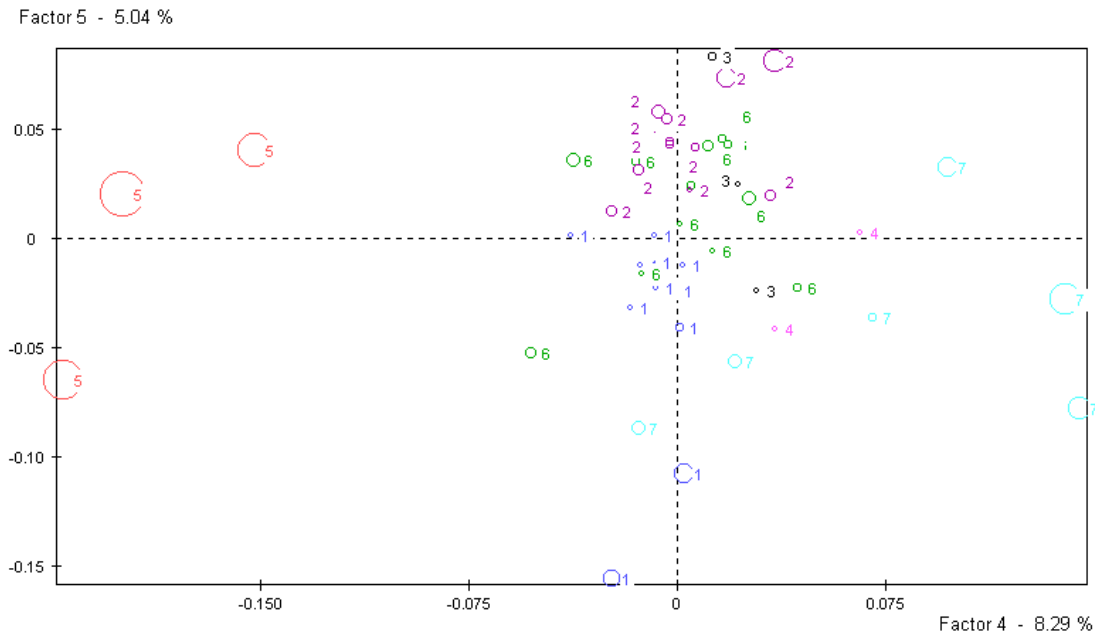
Figura III.17
Posicionamiento de los grupos obtenidos
Factores 2 y 3



Además del grupo 3 y 4, cuyas características diferenciadoras ya habían sido puestas de manifiesto al analizar el plano anterior, esta representación gráfica permite establecer las diferencias entre el grupo 7, y los otros dos grupos de características similares, los grupos 5 y 6. Teniendo en cuenta la interpretación del factor, serán entidades que no destacan por altos porcentajes de gastos externos e internos, ni de elevados márgenes de solvencia mantenidos sobre los mínimos exigidos, y en las que las primas procedentes de seguros de automóviles son importantes. Si bien, parece claro que son entidades en las que el reaseguro cedido tampoco es importante, y que no presentan una gran productividad de los capitales propios más las provisiones.

El grupo 5 aparece en este caso en situación opuesta al grupo 7, respecto al cuarto factor. Es este factor el que permite caracterizar al grupo 5 como un conjunto de entidades que no presentan primas procedentes del negocio de automóviles y además, a diferencia del grupo 7, presentan unos gastos externos más elevados.

Figura III.18
Posicionamiento de los grupos obtenidos
Factores 4 y 5



Los grupos 1 y 2 se diferencian respecto al quinto factor. Por ello el grupo 1 tendrá unos porcentajes más altos de primas por empleado y menores porcentajes de inversiones financieras sobre las totales. Parece que son las entidades de vida con menores inversiones financieras y que emplean la bancaseguros como canal de distribución. Como ya indicamos, las entidades bancarias que distribuyen productos de seguros se dedican sobre todo al seguro de vida, ya que los conocimientos necesarios para la distribución de este tipo de seguros pueden considerarse similares a los que poseen los empleados de las entidades bancarias, circunstancia que no se da cuando se distribuyen otro tipo de seguros.

Con todo lo expuesto, a nuestro juicio, el procedimiento describe el sector y permite la identificación de los grupos estratégicos de manera satisfactoria. Es necesario mencionar la coincidencia entre los grupos identificados y los grandes ramos del sector, hecho que permite conjugar las conclusiones obtenidas con los resultados del Análisis Estructural (tablas III.10, III.11, III.12 y III.13 de este capítulo). Por tanto, las amenazas que

allí se especificaron sobre las ganancias de los grandes ramos, son las que afectan a estos grupos, y los hechos que fueron analizados en aquel caso como barreras de entrada pueden considerarse como barreras a la movilidad de los distintos grupos.

2.3 Evolución del sector asegurador español

Para completar el estudio del sector asegurador español analizamos su evolución con el fin de determinar las perspectivas del mismo. Comenzamos con una descripción de los datos a analizar y los criterios empleados para la selección de los mismos. En la primera fase del análisis se aplica la metodología *STATIS* (versión *Statis*) a la tabla de datos descrita, con el fin de explicar de manera general, en función de las variables seleccionadas, el comportamiento del sector en los diferentes años, y predecir cada una de esas variables sin necesidad de modelizarlas de manera particular. Posteriormente, en una segunda fase, los factores obtenidos del *Statis*, considerados como valores de carácter temporal, son modelizados teniendo en cuenta su naturaleza estocástica, lo cual se realiza mediante modelos *ARIMA*.

En la tercera fase, se van a considerar los valores predichos para las variables originales, obtenidos a través de las fórmulas de reconstitución de los datos a partir de los valores de los factores modelizados, valorando posteriormente la fiabilidad de los resultados obtenidos. La última fase pretende obtener los valores de las variables originales para los dos años siguientes a los considerados en la tabla de datos.

2.3.1 Descripción de los datos

Comenzamos describiendo los individuos (años), las variables analizadas correspondientes a los grandes ramos del sector y las diferentes tablas de doble entrada que componen la tabla total de tres entradas. De las diferentes fuentes de datos consultadas, y dada la necesidad de considerar un periodo elevado de años, sólo las publicaciones de UNESPA (1998a; 1999a) disponen de información suficiente para analizar la evolución del sector.

Los individuos: los años

La primera cuestión a tener en cuenta con relación al estudio evolutivo es que los individuos se corresponden con los años en que éste se realiza. Hemos considerado un período de once años, desde 1987 hasta 1997, ambos inclusive, debiéndose dicha selección a la información más actual disponible en las publicaciones citadas. Para profundizar más en el conocimiento del sector, hubiese sido conveniente disponer de información para un mayor número de años, lo que proporcionaría series más largas y, como consecuencia, mejores predicciones; sin embargo, de nuevo, la información disponible condiciona el estudio. En definitiva, las filas de las tablas vienen expresadas por los años, cuya identificación se realiza mediante etiquetas con las dos últimas cifras del año.

Las variables

Las variables del estudio han sido seleccionadas después de observar las publicaciones más importantes sobre el sector, tales como ICEA, INESE, Dirección General de Seguros, o UNESPA. Conviene aclarar que, con un mismo fin predictivo, hubiera sido posible considerar la información empleada anteriormente para analizar la estructura del sector, pero la misma presentaba dos restricciones importantes: el número de años era muy

reducido, y no era posible disponer de datos para todas las empresas del mismo.

Por tanto, se han tenido en cuenta los diferentes estudios y fundamentalmente opiniones de expertos en el tema. Todo ello ha inducido a pensar que, con carácter general, estas variables se analizan cuando se trata de estudiar las perspectivas y la evolución del sector asegurador.

Según este criterio, además de las que consideramos en este trabajo, hubiera sido necesario incluir otras variables, como por ejemplo la importancia del capital extranjero (es decir, la penetración extranjera en el sector), la evolución del número de compañías, el coste medio de los siniestros (siniestralidad/número de siniestros), etc. Sin embargo, la información relativa a estas variables no estaba disponible para un número de años suficiente, tomando finalmente las siguientes:

- *Gastos internos/primas adquiridas* = *gastinter*

Esta variable expresa, en porcentaje, el cociente entre los gastos de gestión interna y las primas adquiridas para cada uno de los años considerados en cada uno de los ramos agregados. Como ya hemos indicado, como gastos internos se consideran todos los de administración general de la entidad.

- *Gastos externos/primas adquiridas* = *gastexter*

Este cociente, en porcentaje, expresa la importancia de los gastos externos sobre las primas adquiridas, considerando aquí todos aquellos gastos de gestión propios de la entidad aseguradora para la producción o mejora de la imagen de la entidad.

- *Siniestralidad/primas adquiridas* = *sinies*

Esta variable se calcula de manera diferente en los ramos vida y no vida. En el caso de los seguros no vida, es el cociente (en porcentaje) entre la siniestralidad contable (siniestros y gastos pagados durante el

ejercicio más el incremento de las provisiones técnicas para prestaciones o siniestros pendientes y otras provisiones técnicas) y las primas adquiridas. En el caso de los ramos de vida es el cociente, en porcentaje, entre la suma del importe de los siniestros más el incremento de las reservas matemáticas, por un lado, y las primas adquiridas, por el otro. La siniestralidad es una variable de las más importantes para cualquier estudio del sector.

- $(\text{sinies} + \text{gastinter} + \text{gastexter}) / \text{prim. adquirid} = \text{equi}$

Además del análisis individual de las variables que lo componen, la mayoría de los estudios sobre el sector analizan esta variable. Expresada en porcentaje, muestra la rentabilidad de las operaciones de seguro en cada ramo sin tener en cuenta otros ingresos, ajenos a la actividad, como podrían ser los financieros.

- $\text{Primas adquiridas} / \text{Producto Interior Bruto} = \text{pib}$

Recoge el porcentaje que las primas del ramo agregado suponen sobre el Producto Interior Bruto. Este cociente expresa la penetración del seguro en la economía e indica el peso económico del ramo considerado.

- $\text{Primas adquiridas ramo} / \text{Primas totales} = \text{prim}$

Esta variable expresa el porcentaje de las primas del ramo respecto de las primas totales. Mide la evolución de las primas de cada ramo en función de los otros, y así la importancia dentro del sector.

La siguiente tabla resume las variables comentadas, en orden alfabético, y las etiquetas correspondientes:

Tabla III.23
Variables

Variables	Etiquetas
$(\text{sinies} + \text{gastinter} + \text{gastexter}) / \text{prim. adquirid}$	equi
<i>Gastos externos/primas adquiridas</i>	gastexter
<i>Gastos internos/primas adquiridas</i>	gastinter
<i>Primas adquiridas/Producto Interior Bruto</i>	pib
<i>Primas adquiridas ramo/Primas totales</i>	prim
<i>Siniestralidad/primas adquiridas</i>	sinies

Estas variables constituyen las columnas de las tablas analizadas, medidas cada una de ellas para cada año considerado.

Las tablas: los ramos agregados

Tomando como referencia los resultados obtenidos al analizar la estructura del sector en los epígrafes anteriores, y dado que sus características varían según el tipo de seguro considerado, parece conveniente analizar la evolución del mismo teniendo en cuenta tales diferencias, especificando las variables analizadas para los siguientes ramos agregados:

- VIDA, que se refiere única y exclusivamente a los seguros de vida.
- AUTOS, que recoge las primas del ramo de seguros de automóviles.
- SALUD, que hace referencia a las primas de enfermedad y asistencia sanitaria, y
- RESTO, que agrupa a los ramos no incluidos en las anteriores.

Cada uno de estos ramos da lugar a una tabla de doble entrada (por tanto, 4 en total) formada por 11 filas (los individuos-años, del 87 al 97), y 6 columnas (las variables mencionadas), aplicándose el *Statis* sobre la tabla resultante (de 11 filas \times 24 columnas). Las etiquetas empleadas para la identificación de las diferentes tablas se recogen a continuación:

Tabla III.24
Ramos de seguros

Tablas	Etiquetas	Operadores Asociados
VIDA	VI	$W_1 D$
AUTOS	AU	$W_2 D$
SALUD	SA	$W_3 D$
RESTO	RE	$W_4 D$

Cuando las variables hagan referencia a una tabla concreta su etiqueta irá precedida de las siglas VI, AU, SA o RE según corresponda. Los datos a tratar están tabulados en apartado A.3 del anexo (tablas a, b, c y d: pp. 609-10).

2.3.2 Comportamiento del sector

La primera etapa del análisis evolutivo pretende comparar de manera global las diferentes tablas, es decir, los distintos ramos agregados del sector (etapa de interestructura), para obtener unas dimensiones o factores que expliquen el comportamiento del sector en el período considerado (etapa de intraestructura). La modelización de esos factores y la reconstitución, a partir de los mismos, de las variables originales de las tablas permitirá hacer previsiones.

Para evitar la influencia de las unidades de medida y de las varianzas de las variables, cuyos valores varían considerablemente de unas a otras según se deduce de los resultados expresados en el Anexo A.3.1.1 (p. 611), éstas se van a centrar y a tipificar. Para cada nube de individuos, el operador utilizado y analizado será el $W_t D / \|W_t D\|$ correspondiente, que se va a considerar normado con el fin de equilibrar la influencia de todas las tablas en el análisis.

Comenzamos describiendo las distintas etapas del método y sus resultados, interpretando posteriormente los mismos.

a. Resultados obtenidos

La aplicación del *Statis* es posible dadas las características de las nubes de individuos obtenidas a partir de las tablas de datos. Tal y como se deduce al observar los datos (apartado A.3.1.2: pp. 612-3), las cuatro nubes de individuos consideradas (matrices W_1 , W_2 , W_3 , W_4) tienen la misma dimensión (es decir, el número de valores propios no nulos es el mismo), y similar forma (el reparto, dentro de cada una de ellas, de dichos valores es parecido). Si bien, ya se pone en evidencia que la tabla más diferente es la relativa a los seguros de vida, circunstancia que se pondrá de relieve a lo largo de todo el estudio.

Las diferencias entre nubes se deben al cambio de posición de los individuos de unas tablas a otras, que en nuestro caso son cambios en el comportamiento de los años respecto a las variables analizadas, y su análisis va a permitir caracterizar el sector asegurador en el periodo considerado.

1. INTERESTRUCTURA

Para comparar globalmente las nubes, y establecer semejanzas y diferencias entre las mismas, se construye la matriz S de coeficientes RV entre los operadores representativos de las tablas de datos. La obtención de sus valores y vectores propios permite obtener una imagen euclídea de dichos operadores, que refleja la norma y el ángulo entre ellos, en un espacio de dos dimensiones.

En este caso, dicha matriz de coeficientes se recoge en la siguiente tabla:

Tabla III.25
Matriz de coeficientes RV

Matr	VI	AU	SA	RE
VI	1			
AU	0,554	1		
SA	0,434	0,529	1	
RE	0,489	0,675	0,677	1

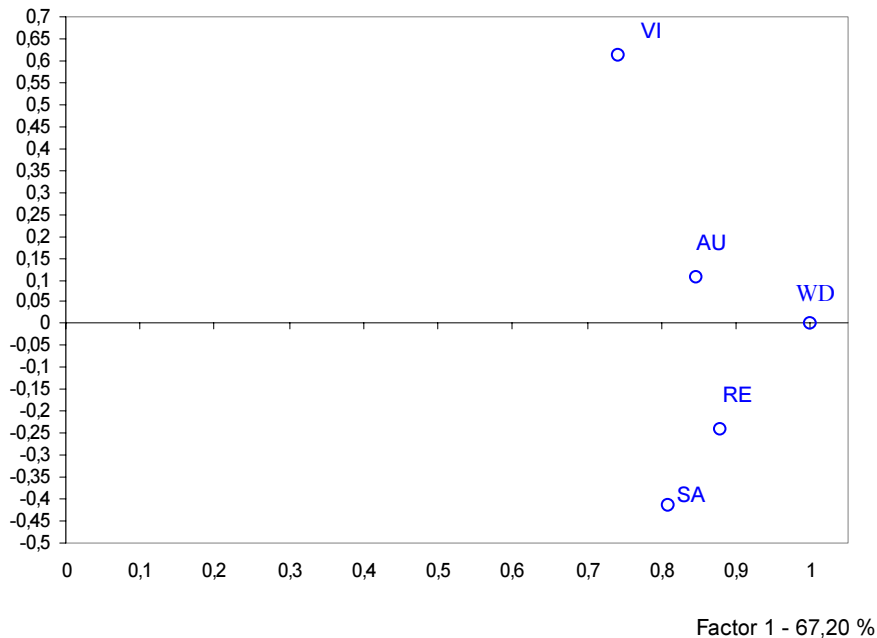
En el epígrafe A.3.1.2 (p. 614) se recogen los tests de permutaciones para dicha matriz, marcando con un asterisco los productos escalares suficientemente elevados, que no son debido solo al azar, e indican que las tablas están correlacionadas de manera importante. Para este caso todos son significativos, excepto el correspondiente al coeficiente RV entre la nube de vida y de salud ($RV = 0,434$), por lo que las características de esas nubes son especialmente diferentes. A nuestro juicio, ese valor no es lo suficientemente pequeño como para excluir del estudio una de las dos tablas dado que, como se demostrará, el modelo final que se va obtener es válido para cumplir los objetivos planteados, y la eliminación de una de ellas supondría un sesgo importante en el análisis del sector.

De dicha matriz se destacan los coeficientes RV entre las nubes relativas a los seguros de automóviles y el resto de seguros ($RV = 0,675$), y entre esta última y los seguros de salud ($RV = 0,677$), que son más elevados que los otros. Sin embargo, el coeficiente más bajo ($RV = 0,434$) es el que presentan las nubes de seguros de vida y de salud. La interpretación de estos resultados permite extraer conclusiones interesantes respecto al sector, como se pondrá de manifiesto.

La imagen euclídea en dos dimensiones, obtenida al descomponer dicha matriz en vectores y valores propios, permite obtener la representación gráfica siguiente (figura III.19). Las coordenadas, contribuciones y los cosenos cuadrados se recogen en el epígrafe A.3.1.4 (p. 616).

Figura III.19
Primer plano de la interestructura

Factor 2 - 15,32 %



Este plano explica el 82,52% de la inercia total (casi toda, el 67,2% acumulada por el primer factor, dada la semejanza de las tablas). Conviene precisar que si las tablas fuesen aún más parecidas, como en las interestructuras de las figuras III.2 (p. 399) y III.3 (p. 400), estos porcentajes de inercia superarían el 95% de la inercia total.

Como se trata de objetos normados, el operador compromiso *WD* tiene norma unitaria y se coloca a una distancia uno del origen sobre el primer eje. Los puntos correspondientes a los operadores estudiados están bien representados en cuanto que su norma está próxima a uno y su posición en el plano está lejos del origen. El vector correspondiente a la tabla *VIDA* forma un ángulo bastante elevado con el compromiso y con los otros vectores, *AUTOS*, *RESTO* y *SALUD*. Por otra parte, el ángulo formado entre estos tres es más pequeño, especialmente entre los relativos a *RESTO* y *SALUD*, lo que indica una estructura más parecida entre ellos, respecto a las variables analizadas, en los años estudiados.

2. COMPROMISO

En esta etapa se obtienen las ponderaciones para las distintas tablas en la construcción del compromiso, cuyos resultados se recogen a continuación:

Tabla III.26
Ponderaciones del *Statist*

	PONDE
NUBE VI	0,275
NUBE AU	0,315
NUBE SA	0,301
NUBE RE	0,326

Por ello, tenemos que el compromiso esta representado por la matriz WD , de tal manera que:

$$WD = 0,275 \frac{W_1 D}{\|W_1 D\|_{HS}} + 0,315 \frac{W_2 D}{\|W_2 D\|_{HS}} + 0,301 \frac{W_3 D}{\|W_3 D\|_{HS}} + 0,326 \frac{W_4 D}{\|W_4 D\|_{HS}}$$

Como era previsible las ponderaciones de las nubes AUTOS, SALUD y RESTO son bastante similares, dada la semejanza entre ellas manifestada al analizar la interestructura. Debido a su estructura diferente, la nube correspondiente a los seguros de VIDA es la que tiene una ponderación inferior y por tanto influirá en menor medida en la construcción del compromiso.

La imagen euclídea de los individuos-compromiso (años) y de las variables es analizada en la siguiente etapa, la intraestructura.

3. INTRAESTRUCTURA

La descomposición del operador WD en sus valores y vectores propios es similar al análisis de la nube de individuos en un ACP *centrado*. A partir de aquí se extrae la estructura común de los 11 individuos-compromiso (años), es decir, las coordenadas de dichos individuos en los ejes del compromiso (que dan lugar a los factores). Las posiciones de dichos individuos

se interpretan, como en un ACP, según el significado asignado a los factores, en función de sus correlaciones con las 24 variables (6 variables en cada tabla \times 4 tablas), que se representan en los círculos de correlaciones.

Los valores propios correspondientes se recogen en la siguiente tabla:

Tabla III.27
Valores propios y porcentajes de inercia

Histograma de los 10 primeros valores propios				
NUMERO	VALOR PROPIO	PORCENT	PORCENT ACUMUL	
1	0.7331	36.93	36.93	*****
2	0.5815	29.29	66.22	*****
3	0.2638	13.29	79.51	*****
4	0.1996	10.05	89.57	*****
5	0.1055	5.31	94.88	*****
6	0.0539	2.72	97.60	*****
7	0.0260	1.31	98.91	***
8	0.0126	0.64	99.55	**
9	0.0058	0.29	99.84	*
10	0.0032	0.16	100.00	*

Las principales características de los individuos-compromiso (años) y su relación con las variables, se ponen de manifiesto reteniendo los tres primeros factores, correspondientes a los tres primeros valores propios del operador *WD*, que explican el 79,51% de la inercia total de la matriz. A partir de este factor los valores propios correspondientes descienden más rápidamente, por lo que no se consideran significativos, además de no tener una clara interpretación.

Estudio por variables

La tabla III.28 recoge las correlaciones de las variables con esos tres primeros factores, mientras que las puntuaciones y los ejes unitarios se expresan en el apartado A.3.1.7 (p. 619 del anexo).

Tabla III.28
Correlaciones de las variables con los primeros factores

VARIABLES	CORRELACIONES		
	1	2	3
Viequi	-0.35	0.33	0.79
Vigastexter	-0.70	-0.62	-0.05
Vigastinter	-0.43	-0.87	0.20
Vipib	-0.23	0.96	-0.03
Viprim	0.14	0.98	-0.08
Visinies	-0.15	0.59	0.71
AUequi	-0.11	-0.06	0.19
AUGastexter	0.87	-0.19	-0.06
AUGastinter	0.69	0.66	-0.02
AUpib	-0.94	-0.03	-0.09
AUprim	-0.09	-0.98	0.02
AUsinies	-0.41	-0.08	0.21
SAequi	-0.18	0.11	-0.86
SAGastexter	0.84	-0.25	-0.26
SAGastinter	0.11	-0.69	-0.38
SAPib	-0.86	0.50	-0.01
SAprim	-0.93	-0.27	0.03
SASinies	-0.37	0.39	-0.74
REequi	-0.91	-0.30	-0.04
REGastexter	-0.81	0.36	-0.32
REGastinter	-0.46	-0.08	-0.53
REpib	-0.65	0.46	0.31
REprim	0.45	-0.84	0.20
REsinies	-0.73	-0.56	0.22

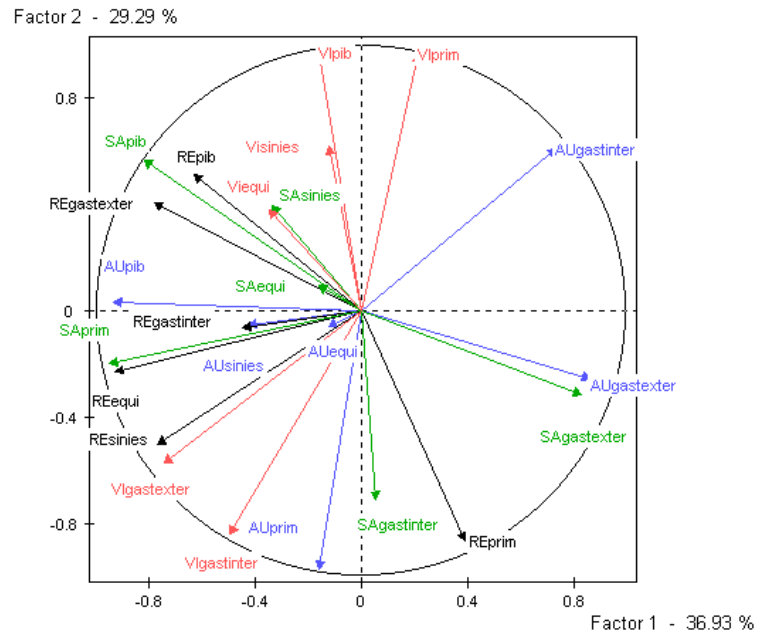
En las siguientes figuras se representan las variables en los diferentes planos factoriales.

- Primer plano factorial (factores 1 y 2)

El primer plano, figura III.20, explica el 66,22% de la inercia total y está formado por los factores primero y segundo, de forma que explica el 36,93% de la inercia de la matriz compromiso. El extremo positivo de dicho factor está correlacionado con las variables AUGastexter (0,87), AUGastinter (0,69) y SAGastexter (0,84); la parte negativa presenta una correlación muy elevada con AUpib (-0,94), SAprim (-0,93), REequi (-0,91), y alta con las variables SAPib (-0,86), REGastexter (-0,81), RESinies (-0,73). Asimismo, este extremo presenta una correlación importante con VIGastexter (-0,7) y REpib (-0,65).

El segundo factor, que explica el 29,29% de la inercia, presenta una correlación positiva muy alta con VIPrim (0,98) y VIPib (0,96) y, en menor medida, con AUGastinter (0,66). La parte negativa del eje está correlacionada con las variables AUprim (-0,98), VIGastinter (-0,87), VIGastexter (-0,62), SAGastinter (-0,69) y RESinies (-0,56).

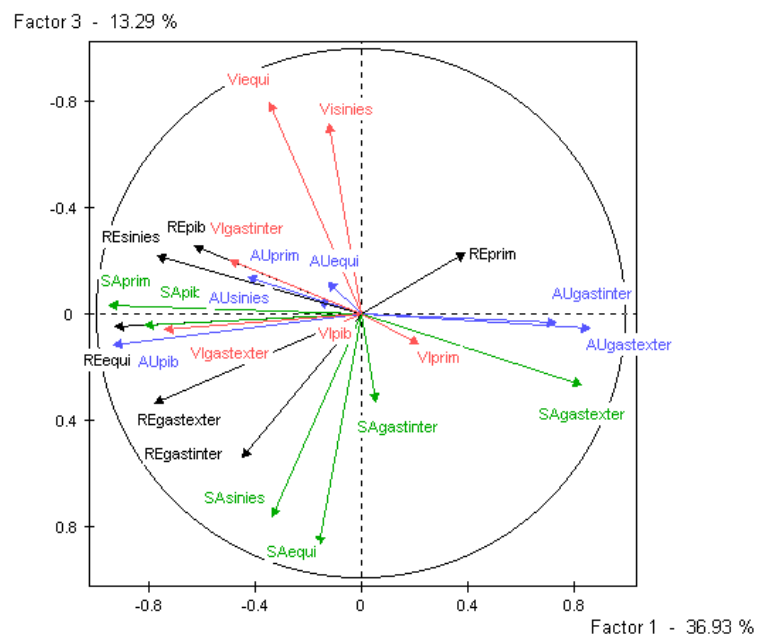
Figura III.20
Círculo de correlaciones. Factores 1 y 2



- Segundo plano factorial (factores 1 y 3)

A continuación se representa el segundo plano, formado en este caso por los factores 1 y 3, que explica el 50,22% de la varianza total.

Figura III.21
Círculo de correlaciones. Factores 1 y 3

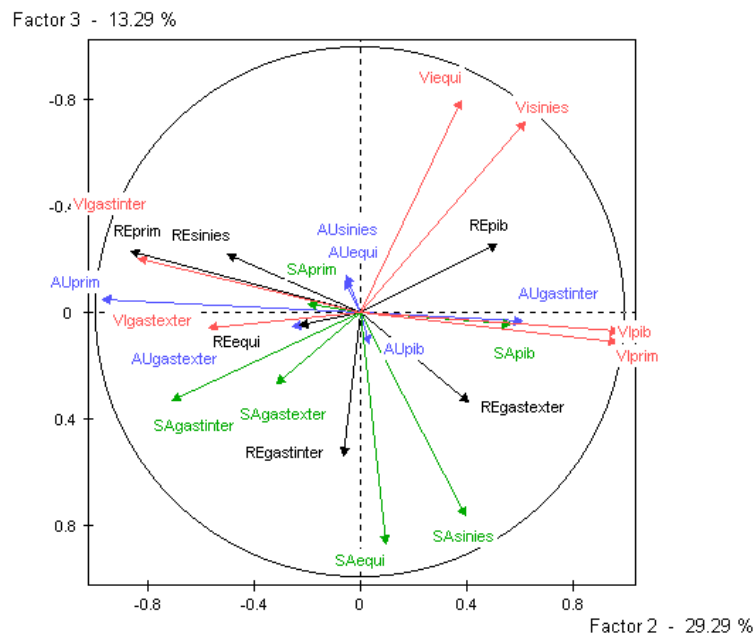


Dado que el primer factor ya ha sido tenido en cuenta, ahora solo se va a considerar las correlaciones del tercer factor. Este representa el 13,29% de la inercia total y presenta una correlación positiva elevada con las variables VI_{equi} (0,79), VI_{sinies} (0,71). El extremo negativo de este factor está correlacionado con las variables SA_{equi} (-0,86), SA_{sinies} (-0,74) y $RE_{gastinter}$ (-0,53).

- Tercer plano factorial (factores 2 y 3)

A continuación se representan los factores 2 y 3, con el objetivo de facilitar su interpretación, aunque sus correlaciones ya han sido descritas anteriormente.

Figura III.22
Círculo de correlaciones
Factores 2 y 3



Este plano explica el 42,58% de la inercia total. El segundo factor muestra de manera clara la oposición entre la importancia de los seguros de vida en el sector (VI_{prim}), vinculada además a su importancia en la economía (VI_{pib}) y correlacionada en sentido positivo, frente a la importancia de los seguros de automóviles (AU_{prim}), y el

resto de los seguros (REprim), que se correlacionan en sentido negativo. El mismo factor permite destacar de nuevo la correlación negativa de las variables relativas a los gastos en los seguros de vida (VIgastinter y VIgastexter).

También se pone de manifiesto de manera clara la fuerte correlación, en sentido negativo, del tercer factor con casi todas las variables relativas a los seguros de salud, excepto las que explican su importancia en el sector y en la economía, y en oposición a las variables relativas a los seguros de vida que recogen la siniestralidad y el equilibrio técnico.

Por otra parte, las variables relativas a los seguros de automóviles, excepto la anteriormente mencionada, apenas están correlacionadas con el plano.

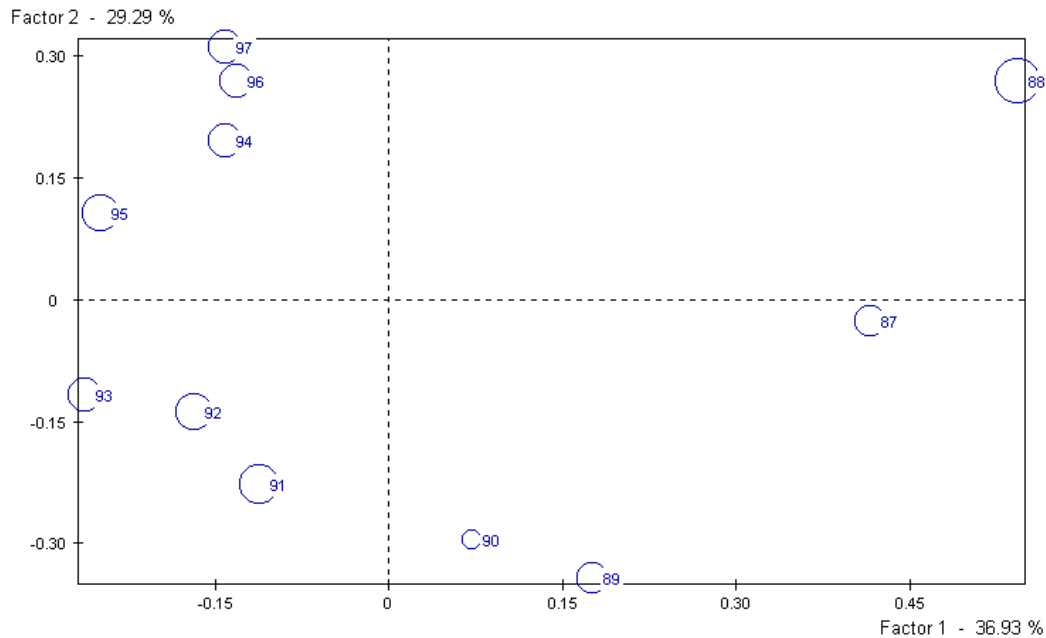
Estudio por individuos

La proyección de los individuos en los planos anteriores permite analizar su posición y luego su interpretación según el significado de los factores. Las representaciones obtenidas consideran los individuos de manera proporcional a su contribución relativa (coseno cuadrado) y se muestran en las figuras siguientes. Las puntuaciones, contribuciones absolutas y cosenos cuadrados se expresan en el epígrafe A.3.1.6 (p. 618).

- Primer plano factorial (factores 1 y 2)

Del análisis de la figura anterior se puede deducir que los cambios más significativos se producen entre los años 1987 y 1988 (respecto al segundo factor), y entre ese año y 1989 (respecto al mismo factor en sentido contrario, y respecto al primer factor hacia una posición más central). Por ello, el individuo-año 1988 presenta valores más altos que el año 1989 en las variables correlacionadas con el primer cuadrante del gráfico (extremo positivo del primer y del segundo factor); y valores más bajos para las variables correlacionadas con el cuarto cuadrante (extremo positivo del primer factor y negativo del segundo factor).

Figura III.23
Los individuos-compromiso (años). Factores 1 y 2



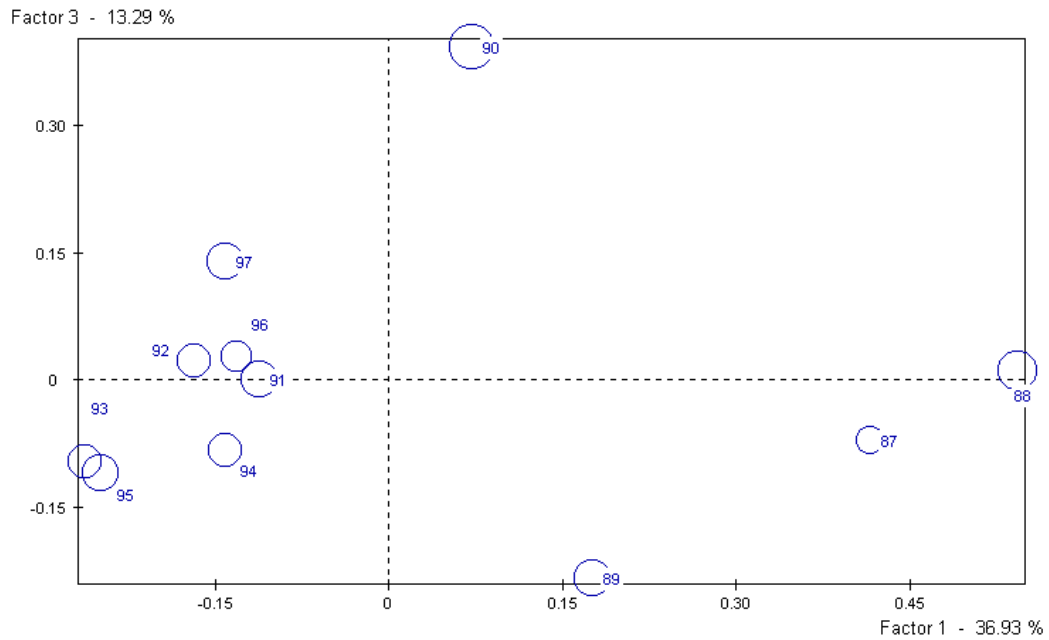
El gráfico permite detectar cambios significativos entre el año 1989 y el año 1997, dado que éste se sitúa en la parte positiva del segundo factor, mientras que el primero está en la parte negativa. Ese cambio de posición se ha ido produciendo gradualmente, respecto al segundo factor, y de forma paulatina entre años consecutivos. Esas fluctuaciones han ido acompañadas de otras respecto al primer factor, de tal manera que los años correspondientes al principio y fin del periodo estudiado, presentan posiciones opuestas respecto a dicho factor.

- Segundo plano factorial (factores 2 y 3)

La figura III.24 representa en el segundo plano factorial los individuos-compromiso (años).

El cambio de posición más importante se manifiesta entre los años 1988 y 1989, respecto al tercer factor, y entre ese año y 1990 respecto al mismo factor, pero en sentido contrario. El resto de los años experimentan cambios, siempre respecto a dicho factor, sin una tendencia clara respecto a un extremo u otro del mismo.

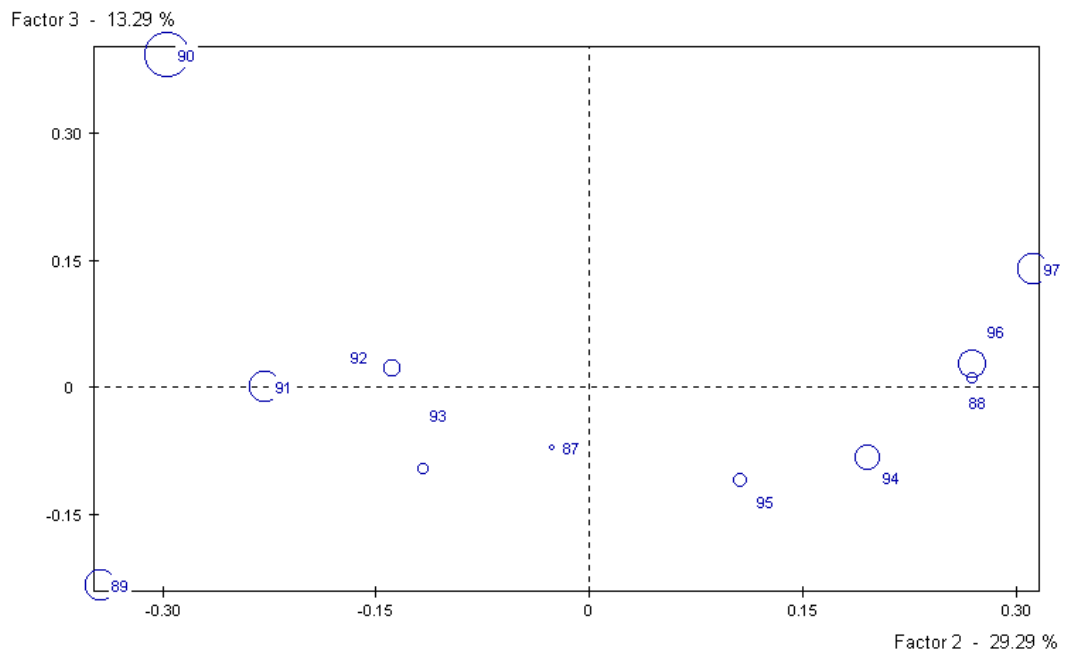
Figura III.24
Los individuos-compromiso (años). Factores 1 y 3



- Tercer plano factorial (factores 2 y 3)

La figura III.25 representa los individuos en el tercer plano factorial.

Figura III.25
Los individuos-compromiso (años). Factores 2 y 3



La calidad de representación de los individuos en este plano es baja en general, y sólo los individuos correspondientes a los años 89, 90 y 97 están adecuadamente representados. Además, la oposición entre el año 89 y 90 respecto al tercer factor, y la oposición entre éstos y el año 97, respecto al segundo factor, ya se han puesto de manifiesto en los otros planos.

4. TRAYECTORIAS

Lo que en otros casos equivale a las trayectorias de los individuos, en este caso sería la representación del valor de cada año respecto a cada ramo. Esos valores no pueden considerarse como trayectorias puesto que se trata del valor (para un mismo año) de los distintos ramos agregados. Como en este caso el objetivo es la predicción, no se va a profundizar en este tipo de cuestiones, que pueden ser consideradas en trabajos posteriores.

b. Interpretación de resultados

Los resultados anteriormente descritos en cada etapa permiten hacer una serie de consideraciones, que son las siguientes:

1. INTERESTRUCTURA

— El análisis de esta matriz permite comparar las distintas nubes de individuos (años) y obtener ya ciertas conclusiones interesantes respecto al sector: la semejanza entre la nube de individuos correspondiente a la tabla del RESTO de seguros, y las de AUTOS (coeficiente $RV = 0,675$) y SALUD ($RV = 0,677$) pone de manifiesto un comportamiento similar de esos ramos en el periodo considerado, respecto a las variables observadas.

La nube relativa a los seguros de vida es la que presenta coeficientes RV más pequeños con todas las nubes, lo que indicaría que el comportamiento

de este ramo de los seguros difiere de los otros, al menos entre 1987 y 1997, lo que no resulta sorprendente teniendo en cuenta las diferencias entre los dos grandes ramos agregados del sector: seguros de vida y seguros de no vida.

La interpretación de los ángulos formados por los vectores correspondientes a cada tabla permiten confirmar las conclusiones obtenidas al analizar la tabla de coeficientes RV .

— Del análisis de la matriz S se pueden extraer conclusiones importantes: aunque lo suficientemente semejantes para la aplicación del método, el comportamiento de los años estudiados, respecto a las variables analizadas, en el caso de los seguros de vida difiere del comportamiento de los seguros no vida, especialmente si se comparan con los de salud. Entre éstos, también existen diferencias aunque su comportamiento se asemeje más en conjunto. La imagen euclídea obtenida permite confirmar dichas conclusiones.

— Con relación al método, se puede decir que la interestructura permite poner de manifiesto su capacidad explicativa cuando se consideran tablas relativas a las diferentes partes de un todo, que es el sector asegurador en este caso, con el fin de obtener conclusiones globales acerca de la estructura del mismo a través de una representación sencilla y fácil de interpretar.

2. COMPROMISO E INTRAESTRUCTURA

La interpretación de los resultados obtenidos al analizar los factores y planos de la intraestructura, permite hacer ciertas consideraciones sobre el sector. Es necesario precisar que mientras las correlaciones entre las variables, y éstas con los factores, permiten establecer relaciones entre ellas en el periodo estudiado, las posiciones de los individuos permiten caracterizar los distintos años según los valores de las variables.

— En cuanto a las primeras, comenzando por los gastos externos, el primer factor permite decir que, en el caso del automóvil ($AUGastexter$), éstos presentan un comportamiento similar a los gastos internos ($AUGastinter$),

pero que difiere en ciertos aspectos dado que su correlación con el segundo factor es de signo contrario. Una situación parecida se produce en el caso de los seguros de salud, ya que dichas variables (SAgastinter y SAgastexter) están correlacionadas positivamente con el primer factor, pero con diferente intensidad. Sin embargo, los gastos externos e internos en el resto de seguros no vida (REgastexter y REgastinter) si parecen tener un comportamiento parejo. En el caso de los seguros de vida, los gastos externos e internos (VIgastexter y VIgastinter) también presentan comportamientos similares en el periodo considerado, teniendo en cuenta la correlación negativa y bastante alta de ambos tipos de gastos con el segundo factor.

— Las correlaciones negativas del primer factor con las variables de gastos externos, siniestralidad y equilibrio técnico del resto de seguros no vida (REgastexter, REsinies y REequi) indica que para este tipo de seguros, y con el fin de mantener el equilibrio resultado de la actividad aseguradora, se debe incidir especialmente en esas dos variables, gastos externos y siniestralidad, más que en los gastos internos.

— Las variables indicativas del peso de todos los seguros no vida en la economía (AUpib, SApib y REpib) presentan un comportamiento similar en el periodo considerado, teniendo en cuenta su correlación fuerte y en sentido negativo con el primer factor. Por otra parte, teniendo en cuenta su correlación negativa y alta con el mismo factor, la importancia de los seguros de salud dentro del sector (SAprim) se ve favorecida por el aumento de su importancia en el ámbito económico general (SApib). No ocurre lo mismo con los automóviles ni con el resto de seguros no vida.

— En el caso de los seguros de salud y de los automóviles, teniendo en cuenta las correlaciones con el primer factor, en el primer caso, y con el segundo factor, en el segundo caso, los gastos externos (SAgastexter y AUGastexter) disminuyen con el aumento de la importancia de éstos en la economía (SApib y AUpib). Sin embargo, para el resto de seguros no vida,

considerando el segundo factor, dichos gastos (REgastexter) aumentan con la importancia del ramo en la economía (REpib).

— Las correlaciones de las variables relativas a los seguros de vida con el segundo factor ponen de manifiesto que, lógicamente, el incremento de la importancia en el sector del ramo vida va acompañado de un descenso de la importancia de los ramos no vida, y esto se pone de manifiesto en cuanto que las variables indicativas de esa importancia (VIprim por un lado, y AUprim, SAprim y REprim, por el otro) presentan correlaciones opuestas con los dos factores considerados en el plano, y por tanto ocupan posiciones casi simétricas respecto al origen.

— Dado que las variables VIpib y VIprim presentan una correlación positiva y fuerte con el segundo factor y casi nula con el primero, un crecimiento o decrecimiento de la importancia de los seguros de vida en la economía, señalado por la primera variable, trae consigo un incremento o decrecimiento, respectivamente, de la importancia de ramo dentro del sector (medido por la segunda variable). Las variables VIsinies y VIEqui están correlacionadas prácticamente de la misma manera que las anteriores con esos dos factores, lo que indica que los cambios en la variable relativa a la importancia de los seguros de vida dentro de la economía y dentro del sector, van acompañados con movimientos en el mismo sentido en la siniestralidad y en el equilibrio técnico de ese ramo, aunque no siempre, dada su correlación diferente con el tercer factor. Teniendo en cuenta la correlación con este factor, se puede añadir que para este tipo de seguros los gastos externos e internos (VIgastexter y VIgastinter), presentan un comportamiento diferente respecto al equilibrio, y éste se ve influenciado sobre todo por la siniestralidad del ramo. Dada su oposición en el primer plano factorial, es posible afirmar que un aumento de la importancia de los seguros de vida en la economía y en el sector trae consigo una reducción de los gastos, tanto externos como internos.

— Teniendo en cuenta la correlación negativa y alta con el tercer factor, podemos decir que el comportamiento de la siniestralidad y el equilibrio técnico en el caso de los seguros de salud (SAsinies y SAequi) es opuesto al caso de los seguros de vida. Es decir, aunque también en los seguros de salud la siniestralidad es la principal responsable del equilibrio del ramo, un aumento de estas variables en el periodo permite deducir un descenso en las mismas variables en el caso de los seguros de vida (VIsinies y VIEqui), y viceversa.

— La correlación negativa, con el segundo factor, de la variable representativa de la importancia de los seguros de automóviles (AUprim), va acompañado de una correlación positiva de la variable de gastos internos (AUGastinter), lo que parece indicar que estos se reducen a medida que crece la importancia del ramo relativa del ramo o, al contrario, aumentan cuando disminuye dicha importancia.

— Las variables relativas al equilibrio técnico y la siniestralidad no están bien representadas (AUequi y AUsinies) en ninguno de los dos planos, por lo que no es posible hacer ninguna afirmación respecto a las mismas.

— Dado que las variables no se correlacionan exactamente con los factores ni por tipo de variable (como por ejemplo, variables de gastos o de siniestralidad), ni por tabla a la que pertenecen (variables de vida, autos, salud o resto), es difícil asignar a los factores una denominación que resuma el significado de cada uno de ellos.

No obstante, con el fin de asignarles una denominación, se puede tener en cuenta que el primer factor se correlaciona, positiva o negativamente, sobre todo, con las variables relativas a los seguros de automóviles y del resto de seguros, que hacen referencia a los gastos, y que en el caso de éstos últimos evolucionan en el mismo sentido que la siniestralidad y el equilibrio técnico; el segundo factor con las variables relativas a los seguros de vida, cuya importancia en el sector y en la economía ha evolucionado en

sentido contrario a los gastos, y a la importancia de dos de los otros ramos: automóviles y resto de seguros no vida; y el tercero con los seguros de vida y de salud, con relación al equilibrio técnico y a la siniestralidad, y que han evolucionado en sentido contrario en un tipo y otro de seguros, en el periodo estudiado.

De todos modos, éstos cumplen su función explicativa pues permiten detectar las variables relaciones entre las variables en los años observados, y los cambios más significativos en los valores de las variables de unos años a otros, como se va a poner de manifiesto al interpretar las posiciones de los individuos-compromiso.

Respecto a estos individuos-compromiso (que son los años), la representación en los planos permite detectar los cambios más significativos entre ellos sin necesidad de analizar una por una todas las variables, teniendo en cuenta que el alejamiento de un individuo en la dirección de un factor implica que éste presenta valores por encima de la media para las variables correlacionadas con el extremo en el que se aleja el individuo, y valores por debajo de la media para las variables correlacionadas con el otro extremo.

— Como el individuo-año 1989 presenta valores más altos que el año 1988 en las variables correlacionadas con el cuarto cuadrante del gráfico (extremo negativo del primer factor y positivo del segundo); y valores más bajos para las variables correlacionadas con el primer cuadrante (extremo positivo del primer factor y positivo del segundo factor), se puede decir que las variaciones más significativas en 1989 son el aumento de la importancia del resto de los seguros en el sector (REprim) y de los seguros de salud (SAprim), así como de los gastos internos para este tipo de seguros (SAgastinter). Por el mismo motivo, se habrá producido una disminución de los gastos internos en el ramo del automóvil (AUGastinter) y un aumento de la importancia de los seguros de vida en el sector (AUprim). La reducción del equilibrio técnico y de la siniestralidad (VIEqui y VISinies) en los seguros de vida también se produce, aunque en menor medida.

— El cambio de posición, respecto al segundo factor, entre el año 1987 y 1991 revela el aumento de los gastos internos y externos en el ramo vida (VIgastinter y VIgastexter), un aumento de la importancia del ramo de salud en el sector (SAprim), acompañado de un descenso en sus gastos externos (SAgastexter). También es posible destacar el descenso de los gastos externos en el ramo del automóvil (AUgastexter), así como el aumento de la siniestralidad y el equilibrio técnico en el resto de los seguros no vida (RESinies y REequi).

— El cambio de posición de 1989 a 1997, se debe a un aumento de la importancia de los seguros de vida, salud y del resto de seguros en la economía (VIpib, SApib y REpib). En el caso de los seguros de vida y salud, este cambio ha venido acompañado de un aumento de la importancia de los mismos en el sector (VIprim y SAprim) en contraste con lo que sucede con el resto de seguros (REprim), que disminuyen su importancia relativa dentro del mismo. En el caso de los seguros de salud también se ha producido una disminución de los gastos internos (SAgastinter), igual que en el caso de los seguros de vida (VIgastinter). En este último caso también han disminuido los gastos externos (VIgastexter), al contrario de lo que ha sucedido en el resto de seguros no vida (REgastexter).

— Los cambios respecto al tercer factor entre el año 1989 y 1990 se deben a un aumento bastante pronunciado de la siniestralidad y del equilibrio técnico para los seguros de vida (VISinies y VIEqui), y un descenso notable en las mismas variables en el caso de los seguros de salud (SAsinies y SAequi). Teniendo en cuenta, el tercer plano mostrado, también es posible poner de manifiesto un descenso en las variables de los gastos en los seguros de salud (SAgastexter y SAgastinter), y en los gastos internos del resto de seguros no vida (REgastinter).

— Con relación al tercer plano, la oposición entre el año 1990 y 1997 vuelve a poner de manifiesto una mayor importancia de los seguros de vida en el sector y en la economía en el último año mencionado, acompañado

de un descenso de los gastos internos y externos en este tipo de seguros. Asimismo, se ha producido un descenso de la importancia de los seguros de automóviles (AUprim).

Es interesante hacer especial referencia al periodo 1991-1997, puesto que fue el considerado en el estudio general del sector de la primera parte de este capítulo. Así, para el último año mencionado los valores de las variables relativas a importancia de los seguros de vida en la economía y en el sector son más elevados, al igual que la siniestralidad y el equilibrio en ese caso, y los gastos internos en el ramo del automóvil. Lo mismo sucede con la importancia en la economía del ramo de salud y con los gastos externos del resto de seguros no vida. Sin embargo, ha descendido la importancia de los automóviles y del resto de seguros no vida en el sector, el equilibrio y la siniestralidad en éstos últimos, los gastos internos y externos en el ramo de vida y los internos en el ramo de salud.

— Con el fin de comparar los valores iniciales y finales de las variables en el periodo considerado, es necesario comparar las posiciones sobre el plano entre el año 1987 y 1997. Así, en el caso de los seguros de vida, los hechos más destacados son el incremento de la importancia de este tipo de seguros en el sector y en la economía (VIprim y VIpib), el aumento de la siniestralidad y el equilibrio técnico (VISinies y VIEqui), aunque en menor medida, teniendo en cuenta el cambio de posición respecto al tercer factor; y el comportamiento cambiante de los gastos externos e internos (VIGastexter y VIGastinter), que finalmente presentan valores más altos que al inicio del periodo.

Respecto a los seguros de automóviles, a pesar del aumento de su importancia en la economía (AUpib), en el sector su peso ha disminuido (AUprim). Los gastos externos también han disminuido (AUGastexter), pero los gastos internos (AUGastinter) no han sufrido importantes modificaciones, respecto al inicio del periodo.

En el caso de los seguros de salud, entre los años indicados, se ha producido un aumento de su importancia en la economía y en el sector (SApib y SAprim), el cual ha ido acompañado de un descenso de los gastos internos (SAGastinter) y externos (SAGastexter).

Los gastos externos han aumentado, entre el año 1987 y 1997, para el resto de seguros de no vida (REgastexter), así como su importancia en la economía (REpib). Si bien, su peso en el sector ha disminuido (REprim), al contrario que su siniestralidad y equilibrio técnico (RESinies y REequi).

El hecho de que sobre algunas variables no se hayan hecho apreciaciones se debe a que las posiciones en el plano de los años 87 y 97 no lo permiten, debido posiblemente a que presentan valores similares para esos dos años.

— Conviene resaltar el hecho de que algunas de estas tendencias, respecto a los años analizados, ya habían sido puestas de manifiesto al observar la información global expresada en el apartado anterior de este capítulo, lo que demuestra la coherencia de los resultados obtenidos con la aplicación del método *Statis* a la tabla de individuos–compromiso (años). El método, por tanto, detecta los cambios más importantes, por lo que a nuestro juicio, los resultados obtenidos pueden emplearse como un excelente punto de partida para analizar un periodo de tiempo largo cuando el número de variables a estudiar es elevado.

2.3.3 Análisis temporal

En este apartado se realiza el análisis temporal de los factores obtenidos anteriormente y que son representativos del sector según las consideraciones expuestas. Los resultados de los mismos se expresan en la tabla III.21 y su

estudio se realiza a través de la metodología Box-Jenkins para procesos *ARIMA*. Conviene volver a señalar, que la misma ofrece algunas ventajas con respecto a otros métodos de series temporales, como la suavización exponencial, la descomposición de series y el análisis de la tendencia simple: extrae mucha información de la serie y la consigue utilizando para ello un número reducido de parámetros. Permite, asimismo, una mayor flexibilidad en la elección del modelo correcto, incluyendo un proceso hace posible examinar una gran variedad de modelos en la búsqueda del más adecuado.

Tabla III.29
Factores temporales

Año	Factor1 Gastos Autos y Resto	Factor2 Importancia Vida	Factor3 Siniestralidad y equilibrio Vida y Salud
87	0,42	-0,03	-0,07
88	0,54	0,27	0,01
89	0,18	-0,34	-0,23
90	0,07	-0,30	0,39
91	-0,11	-0,23	0,00
92	-0,17	-0,14	0,02
93	-0,26	-0,12	-0,10
94	-0,14	0,20	-0,08
95	-0,25	0,11	-0,11
96	-0,13	0,27	0,03
97	-0,14	0,31	0,14

IDENTIFICACIÓN

El primer requisito que debe tenerse en cuenta para la identificación del modelo es que la serie objeto de estudio sea estacionaria, tanto en media como en varianza. En el caso de que no sea estacionaria en media es necesario diferenciar la serie y determinar el número de diferencias (d). Si la serie no es estacionaria en varianza, es necesario tomar logaritmos, es decir, tomar $\lambda = 0$. Ninguna de las series consideradas presenta componente estacional dado su carácter anual.

Factor 1 (Gastos- Automóviles y Resto)

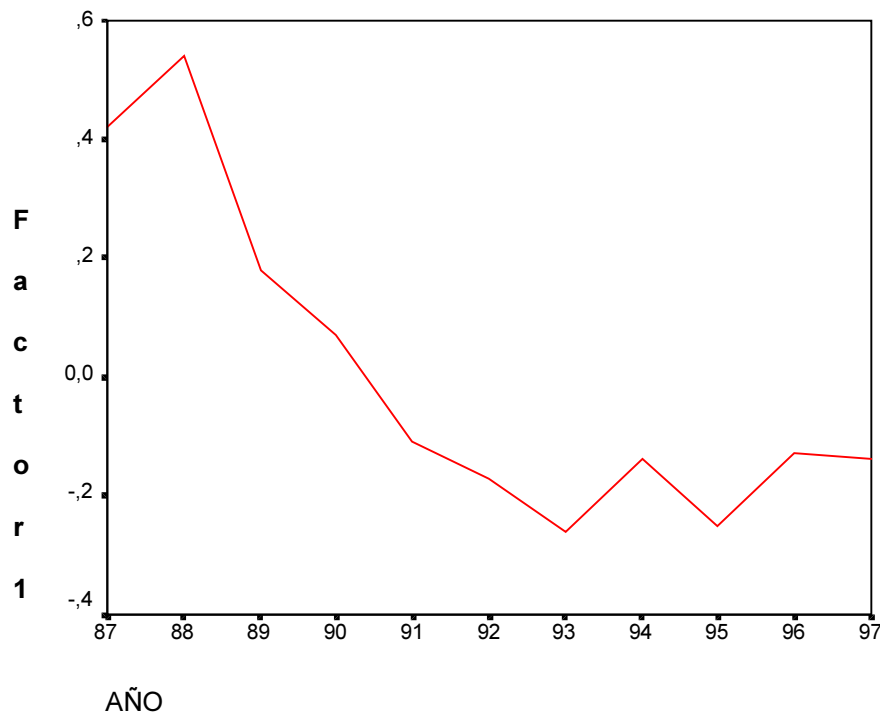
- Identificación de d

Los instrumentos que permiten analizar la estacionariedad en media son el gráfico de la serie y su correlograma muestral.

Gráfico de la serie temporal

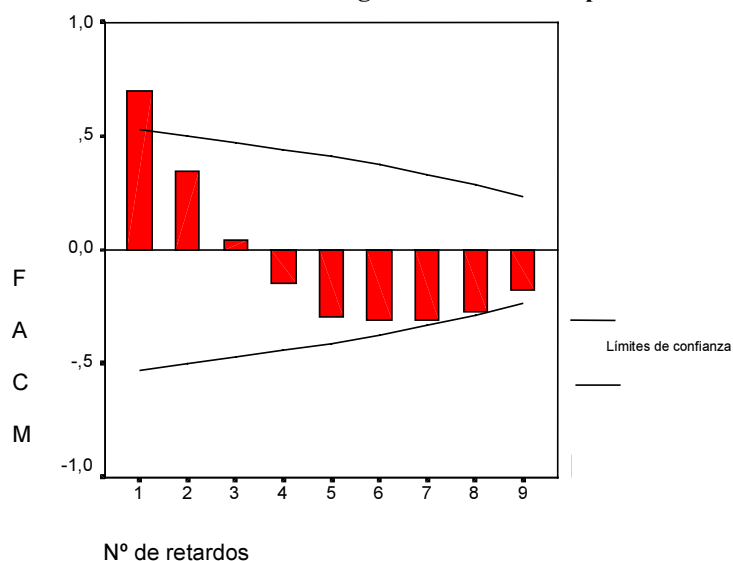
La figura III.26 recoge dicho gráfico para el primer factor.

Figura III.26
Representación del primer factor



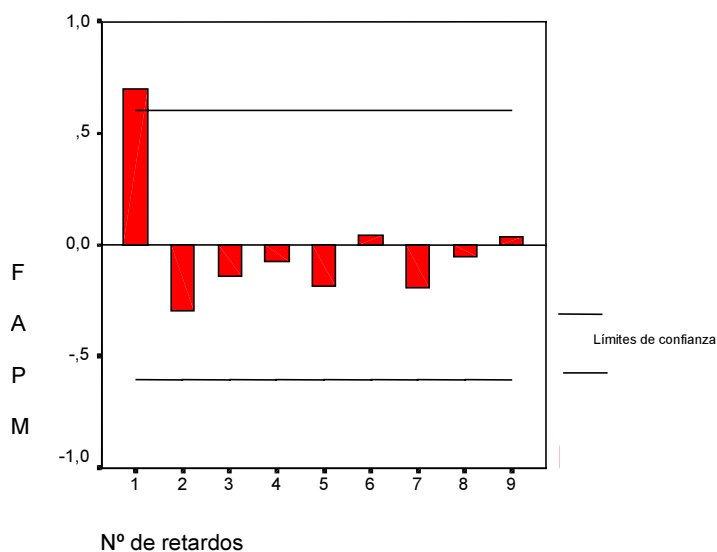
Como los tres primeros valores presentan un valor muy alto, la serie parece presentar una tendencia definida decreciente, aunque no para los últimos años analizados. Por ello, para una correcta identificación es necesario analizar detenidamente el otro instrumento mencionado, el correlograma muestral (FACM), que aparece en la figura III.27. El apartado A.3.2.1 (p. 620) recoge los valores de las autocorrelaciones para los diferentes retardos, así como los valores de las autocorrelaciones parciales.

Figura III.27
Correlograma muestral del primer factor



A la vista de esta representación, se puede decir que la serie analizada es estacionaria en media en cuanto que dicha función tiende rápidamente a cero a medida que aumenta el desfase temporal considerado, es decir, el orden del retardo. Además, la representación en la figura siguiente del correlograma parcial muestral (FAPM), con un primer retardo lejos de la unidad, permite extraer la misma conclusión.

Figura III.28
Correlograma parcial muestral del primer factor



Por este motivo, y siguiendo el desarrollo teórico expuesto, cuando la serie temporal es estacionaria en media no se debe diferenciar y por tanto el valor de d es cero. Por otra parte, puesto que los instrumentos considerados ofrecen una clara evidencia del cuál es el grado apropiado de diferenciación, no parece necesario, a nuestro juicio, recurrir a la *sobrediferenciación*.

- Identificación de λ

Como ya hemos indicado, cuando la dispersión de la serie permanece invariable a lo largo del tiempo se dice que la serie es estacionaria en varianza y no es necesario realizar ninguna transformación en la serie. El valor del parámetro λ es 1. En este caso, dado que la serie presenta valores negativos, la estacionariedad en varianza no puede ser eliminada tomando logaritmos (mediante la transformación de Box-Cox) y, por ello, el único valor posible para λ es precisamente 1.

- Transformación de la serie

Teniendo en cuenta que los valores de los parámetros d y λ deben ser cero y uno, respectivamente, la serie analizada es la original, es decir, $W_t = X_t$

- Identificación del proceso generador de la serie

Se trata de averiguar los procesos *ARMA* susceptibles de generar la serie temporal mediante la comparación de las Funciones de Autocorrelación Muestral (FACM) y Autocorrelación Parcial Muestral (FAPM) de la serie X_t (figuras III.27 y III.28) con las teóricas recogidas en el anexo (A.1.1: p. 567). Dichas funciones se obtienen a partir de la serie original y no del proceso estocástico que las genera, por lo que es necesario tener en cuenta, al nivel de significación $\alpha = 0,05$, las correspondientes bandas de significatividad de las autocorrelaciones.

En muchos casos el resultado de esa comparación puede dar lugar a la existencia de diversos modelos, por lo que es necesario considerar todos, discriminando aquellos que incumplan alguna de las etapas posteriores de la metodología. En este caso, como el primer retardo de la FACM es el más elevado, superando las bandas de significación, y la FAPM presenta los valores más elevados también en el primer retardo, podemos proponer un modelo $AR(1)$. Además, dado que los otros retardos no son significativos en ninguna de esas funciones, no parece necesario considerar otros modelos para ser evaluados.

- Identificación del término independiente

Además de determinar los posibles procesos que pueden generar la serie también debe determinarse si el modelo debe incluir constante o no. La constante en un modelo AR coincide con la esperanza del proceso. En este caso, dado que la serie considerada, es un factor resultado del *Statis* tiene media cero. Por ese motivo, si el proceso identificado es un AR no será necesario incluir constante.

Por otra parte, considerando el estadístico t se tiene:

$$t = \frac{\overline{W}}{\hat{\sigma}_{\overline{W}}} = 0,0033$$

lo que permite asegurar que no se puede rechazar la hipótesis nula de que el parámetro sea cero. También podemos considerar el parámetro en el modelo y eliminar dicho parámetro si como resultado de la estimación éste no resulta significativo.

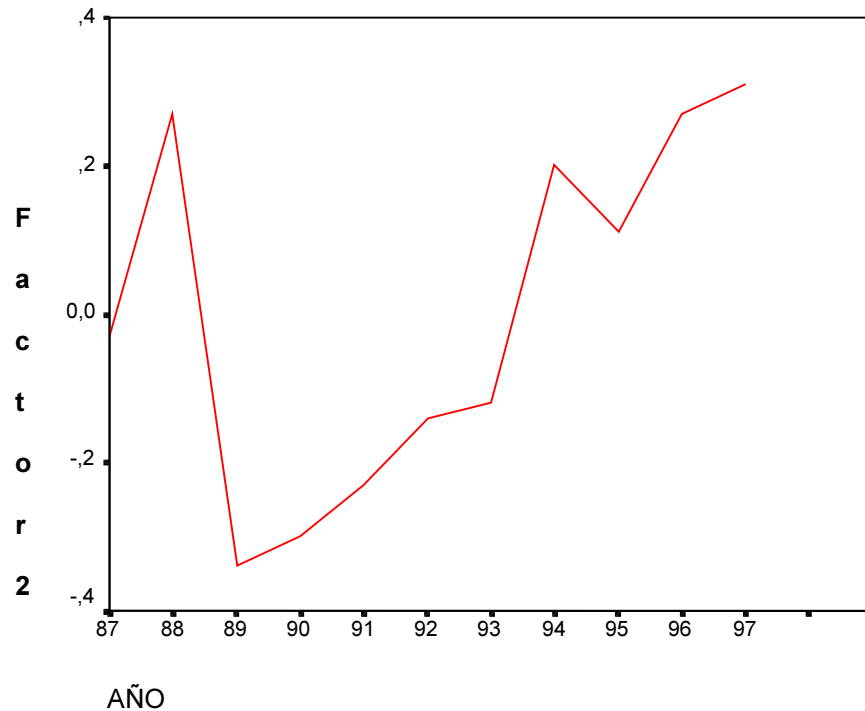
Factor 2 (Importancia –Vida)

Gráfico de la serie temporal

La figura III.29 recoge dicho gráfico para el segundo factor. De su observación se deduce que el factor considerado presenta cierta tendencia, razón por la cual la serie no parece ser estacionaria en

media. En cuanto a la estacionariedad en varianza, si se cumple puesto que la dispersión (excepto para el valor correspondiente al año 89), no varía de manera significativa a lo largo del tiempo.

Figura III.29
Representación del segundo factor



Para un correcto análisis es necesario considerar detenidamente, el correlograma muestral (FACM), que se representa en la figura III.30. En el apartado A.3.2.2 (p. 624) se recogen los valores de las autocorrelaciones para los diferentes retardos, así como los valores de las autocorrelaciones parciales, correspondientes a dicha representación.

De esta representación se deduce que la serie analizada es estacionaria en media en cuanto que dicha función tiende rápidamente a cero y sus retardos no superan los límites de confianza. Por otra parte, la representación de la figura siguiente, correspondiente al correlograma parcial muestral (FAPM), tampoco presenta ningún retardo significativo.

Figura III.30
Correlograma muestral del segundo factor

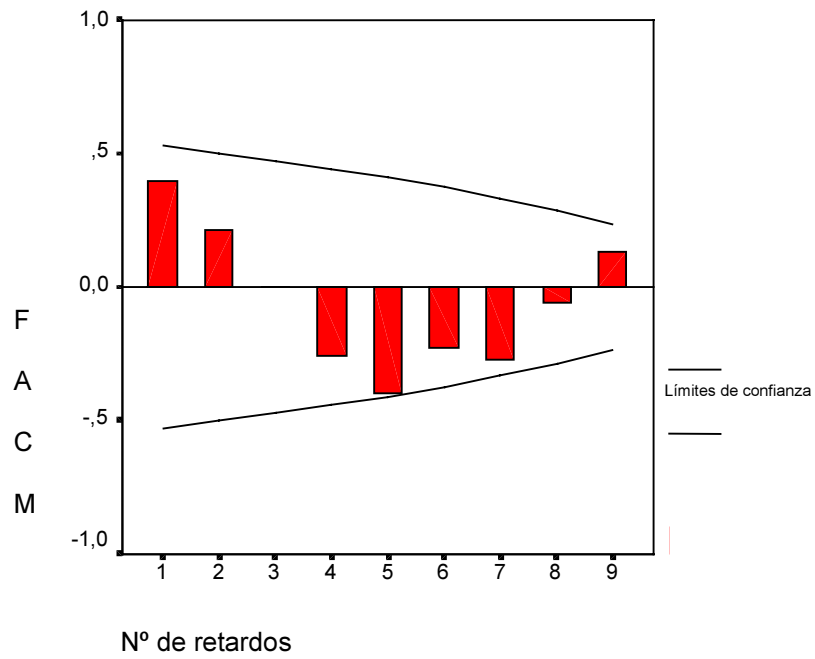
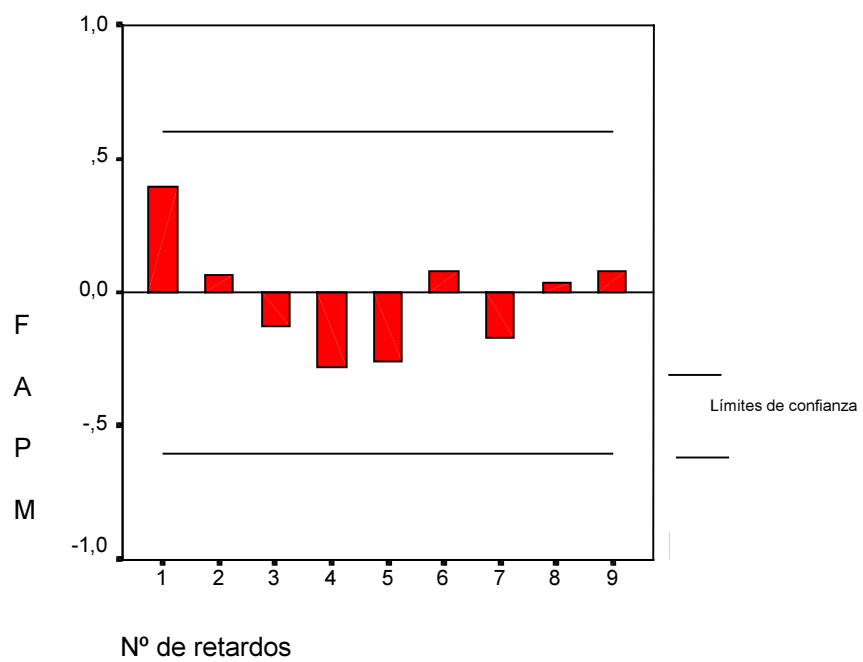


Figura III.31
Correlograma muestral parcial del segundo factor



En definitiva, como el primer retardo está lejos de uno, la serie temporal puede considerarse estacionaria en media y, por tanto, no se va a diferenciar; en este caso el valor de d es cero.

- Identificación de λ

Como la dispersión de la serie permanece invariable a lo largo del tiempo, se puede decir que la serie es estacionaria en varianza, y no es necesario realizar ninguna transformación en la serie ($\lambda = 1$).

- Transformación de la serie

Dado que, según hemos concluido, los valores de los parámetros d y λ deben ser cero y uno respectivamente, la serie analizada es la original, por tanto, $W_t = X_t$.

- Identificación del proceso generador de la serie

Los procesos *ARMA* susceptibles de generar la serie temporal se plantean como resultado de comparar las funciones de Autocorrelación y Autocorrelación Parcial Muestrales de la serie X_t (figuras III.30 y III.31) con las teóricas recogidas en el anexo (epígrafe A.1.1: p. 567). Para la correcta interpretación de dichas gráficas es necesario tener en cuenta, al nivel de significación 0,05 las correspondientes bandas de significatividad de las autocorrelaciones. Aunque el primer retardo de la FAPM es el más elevado, ni en esta ni en la FACM se alcanzan las bandas de significación, de lo que se deduce que la serie analizada es un ruido blanco.

- Identificación del término independiente

Como en el caso anterior, considerando el estadístico t se tiene:

$$t = \frac{\bar{W}}{\hat{\sigma}_{\bar{w}}} = 3,49E - 9$$

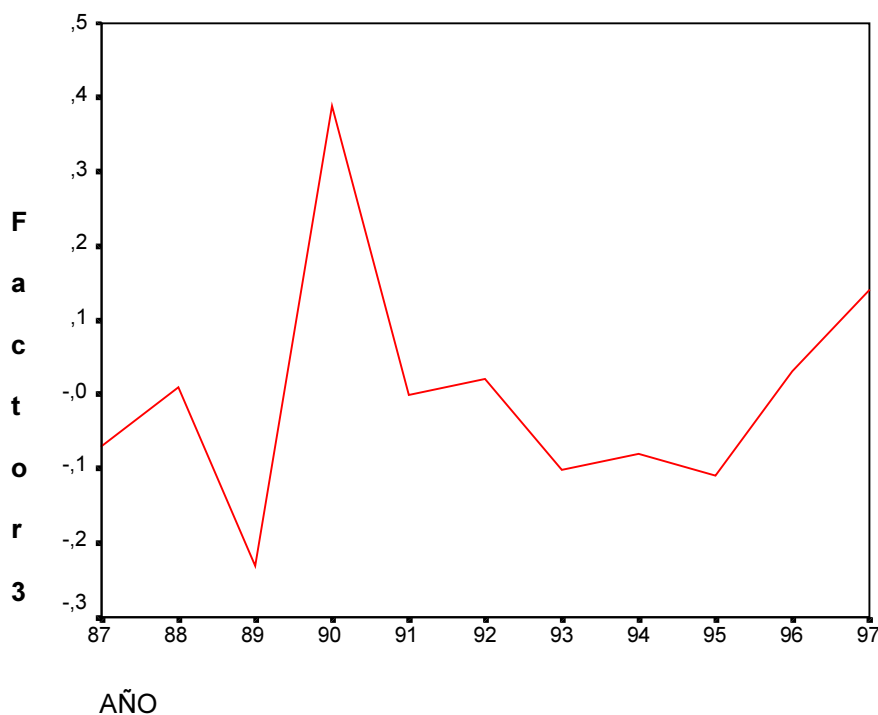
luego no se puede rechazar la hipótesis nula de que el parámetro sea cero, y el proceso identificado es un ruido blanco sin deriva (media). Además, teniendo en cuenta que no puede rechazarse la hipótesis nula de normalidad (anexo A.3.2.2 Identificación, p. 624) es un ruido blanco gaussiano.

Factor 3 (Siniestralidad y Equilibrio – Vida y Salud)

Gráfico de la serie temporal

La figura III.32 recoge la evolución temporal de dicho factor.

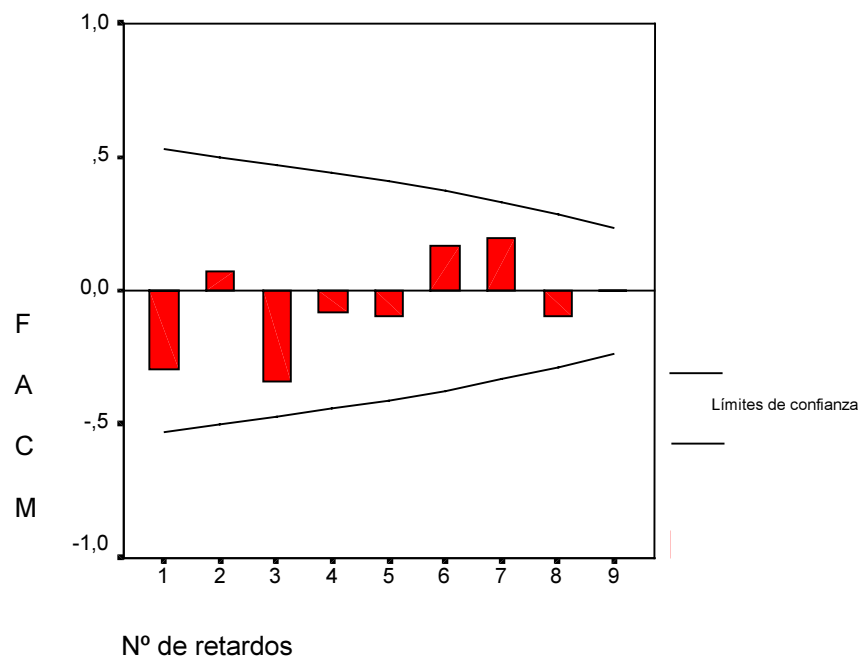
Figura III.32
Representación del tercer factor



Excepto para los años 89 y 90, que presentan un valores que oscilan bastante, la serie no parece presentar una tendencia definida, creciente o decreciente, por lo que la serie podría considerarse estacionaria en media. Dada la forma especial que presenta, para una correcta

identificación, se utiliza el correlograma muestral (FACM), que aparece en la figura III.33, y que se ha obtenido a partir de los valores de las autocorrelaciones para los diferentes retardos (anexo A.3.2.3 Identificación, p. 626).

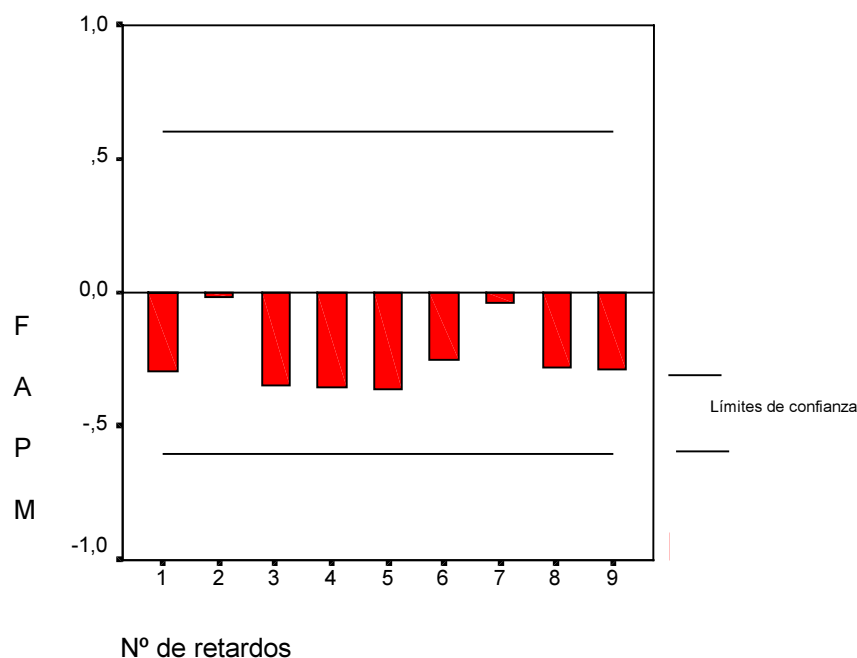
Figura III.33
Correlograma muestral del tercer factor



Según esta representación, se puede decir que la serie analizada es estacionaria en media en cuanto que los retardos no son significativos y además presentan valores bajos, que tienden a cero; la representación del correlograma parcial muestral (FAPM), recogida en la siguiente figura, permite obtener la misma conclusión en cuanto que su primer retardo está lejos de uno.

Por este motivo, podemos decir que la serie temporal es estacionaria en media y, por tanto, no es necesario su diferenciación; en definitiva el valor de d es cero.

Figura III.34
Correlograma muestral parcial del tercer factor



- Identificación de λ

Al observar la figura III.32, se deduce que la dispersión de la serie permanece invariable a lo largo del tiempo, por lo que se dice que la serie es estacionaria en varianza y no es necesario realizar ninguna transformación en la serie ($\lambda = 1$).

- Transformación de la serie

Por las razones expuestas, $d = 0$ y $\lambda = 1$, es decir, $W_t = X_t$.

- Identificación del proceso generador de la serie

La comparación de las Funciones de Autocorrelación y Autocorrelación Parcial muestrales de la serie X_t (figuras III.33 y III.34), con las teóricas, recogidas en el epígrafe A.1.1 (p. 567) del anexo, permite afirmar qué procesos *ARMA* son susceptibles de generar la serie. Su correcta interpretación implica tener en cuenta, al nivel de significación

0,05, las correspondientes bandas de significatividad de ambas funciones.

Su observación, puesto que ninguno de los retardos es significativo y sus valores se alejan bastante de las bandas de significatividad establecidas, permite concluir que el proceso es un ruido blanco.

- Identificación del término independiente

Considerando el estadístico t se tiene: $t = \frac{\bar{W}}{\hat{\sigma}_{\bar{w}}} = -1,49E - 9$

y se deduce que no se puede rechazar la hipótesis nula de que el parámetro sea cero, y el proceso identificado es un ruido blanco sin deriva, y gaussiano (Anexo A.3.2.3 Identificación, p. 626).

ESTIMACIÓN

Una vez identificados los posibles procesos generadores de la serie temporal objeto de estudio, se trata de determinar los parámetros de dichos modelos. El proceso de estimación se realiza mediante el programa estadístico SPSS (versión 7.5.2S), y el algoritmo utilizado es el de Marquardt. Dado que los factores dos y tres son ruidos blanco, solo es necesario la estimación de los parámetros del modelo propuesto para el factor 1:

- Modelo $AR(1)$:

VARIABLES en el modelo:

	B	S (B)	VALOR T	APPROX. PROB.
AR1	,83776762	,14528043	5,7665552	,00027060
Constante	,06826330	,22038315	,3097483	,76380423

- Modelo $AR(1)$, sin término independiente:

VARIABLES EN EL MODELO:

	B	S(B)	VALOR T	PROBAB. APROX.
AR1	,83885839	,14106355	5,9466699	,00014187

VALIDACIÓN

En esta fase se trata de establecer si existe o no adecuación entre los datos y los modelos estimados, analizándolos según las directrices de la metodología *ARIMA*. De la misma manera que en la etapa anterior, se realiza para el factor 1.

- Análisis de los coeficientes estimados

La significatividad de los parámetros se estudia teniendo en cuenta la distribución muestral de los estadísticos que se emplean en su estimación y, por tanto, analizando los valores de la distribución t de Student.

En los modelos considerados observamos como, excepto para los modelos $AR(1)$, los valores de la t y la probabilidad aproximada, al nivel 0,05, permiten rechazar la hipótesis nula de que los parámetros son cero. Los dos modelos $AR(1)$ considerados tienen el parámetro del modelo significativo a ese nivel, pero no ocurre lo mismo con el término independiente; entonces el único modelo que se analiza es el $AR(1)$ sin dicho término. Los resultados obtenidos en la estimación se detallan, para este modelo, en el epígrafe A.3.2.1 (Estimación, p. 621).

- Análisis de los residuos del modelo

Si los residuos constituyen ruidos blancos gaussianos se distribuyen normalmente y son independientes entre sí. Vamos a analizar dichos residuos para el modelo propuesto.

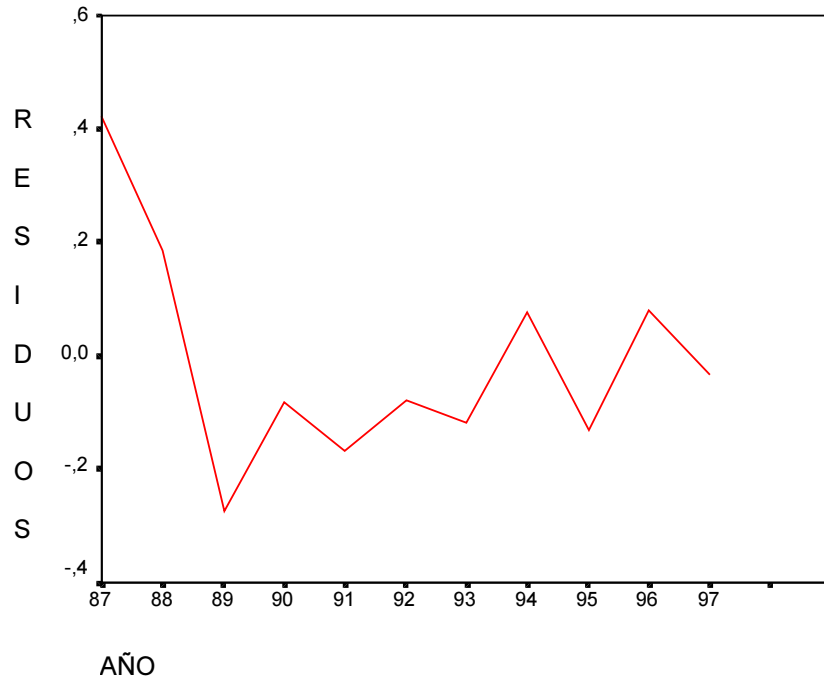
Media cero

En principio, dado que el procedimiento de estimación que se utiliza es no lineal, los residuos no tienen por qué tener media cero, aunque se suele analizar esta característica. Considerando los resultados incluidos en el epígrafe A.3.2.1 (Validación, p. 622), el cociente entre la media de los residuos y su desviación típica es $-0,54$. Por ello, podemos concluir que, para el nivel de significación habitual, la media de los residuos no difiere significativamente de cero.

Varianza Constante

El análisis del gráfico de los residuos permite llegar a la conclusión de si éstos tienen o no varianza constante (homoscedasticidad). La figura III.35 recoge dicha evolución temporal:

Figura III.35
Residuos del modelo $AR(1)$ sin término independiente



Excepto para los primeros valores, podemos decir que la dispersión de los residuos a lo largo del tiempo permanece constante por lo que no parece necesario realizar ninguna transformación en la serie original.

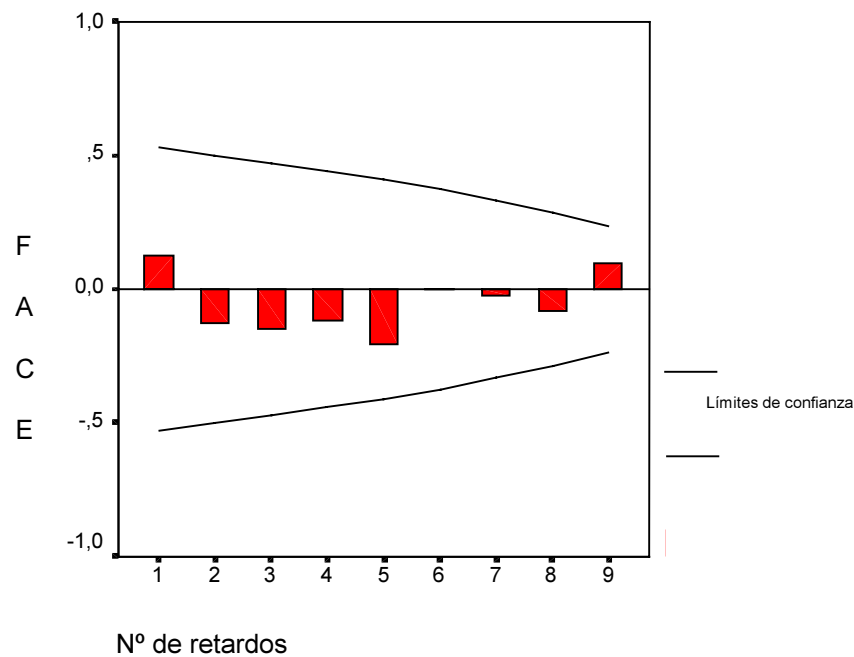
Distribución Normal

La normalidad de los residuos va a ser analizada a través del contraste de Kolgomorov-Smirnov. Como se recoge en el apartado A.3.2.1 del anexo (Validación, p. 622), al nivel de significación habitual, no es posible rechazar la hipótesis nula de normalidad, por lo que debe concluirse que estos tienen una distribución normal.

Residuos Incorrelados

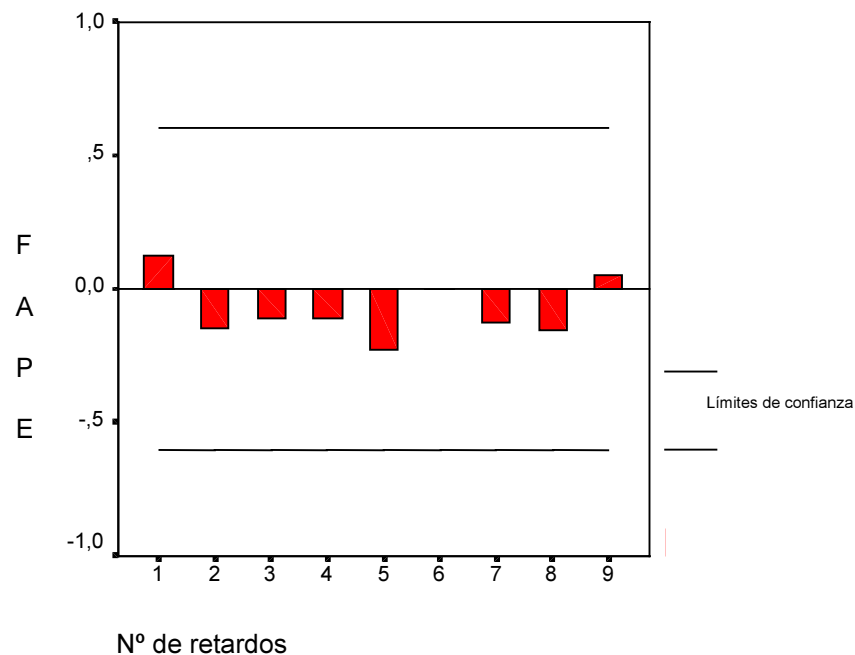
Es necesario considerar las funciones de autocorrelación (FACE) y de autocorrelación parcial (FAPE) de los residuos. Así, los residuos serán independientes, y por tanto un ruido blanco, cuando los coeficientes de ambas sean significativamente distintos de cero. Las figuras siguientes presentan, respectivamente, ambas funciones.

Figura III.36
Función de Autocorrelación de los residuos del modelo $AR(1)$
sin término independiente



De manera individual, puede observarse que ninguno de los retardos es significativo al nivel indicado. Por otra parte, se debe analizar el estadístico de Box y Ljung, que permite considerar de manera global los residuos. Este contraste toma como hipótesis nula la de incorrelación, y los valores de este estadístico se recogen en el mencionado apartado A.3.2.1 (Validación, p. 622) del anexo. Así, al nivel de significación 0,05, en ningún caso es posible rechazar la hipótesis nula, y por ello debe concluirse que, globalmente, los residuos están incorrelados.

Figura III.37
Función de Autocorrelación parcial de los residuos del modelo $AR(1)$
sin término independiente



Puesto que el análisis de los residuos es el principal criterio a tener en cuenta de cara a evaluar la adecuación de un determinado modelo, y en este caso, este análisis ha sido satisfactorio, podemos decir que el modelo $AR(1)$ es adecuado, pasando a la fase de predicción. Si bien, con el fin de completar el análisis es conveniente finalizar la fase de validación considerando las dos cuestiones siguientes.

- Introducción de parámetros adicionales en el modelo y bondad del ajuste

La introducción de parámetros adicionales ya se ha realizado con la consideración de los otros modelos, observando que esos nuevos parámetros no eran significativos. La bondad del ajuste, dado que no tenemos modelos alternativos, no va a ser considerada.

- Análisis de la estabilidad

Como el número de observaciones no es muy elevado, en lugar de aplicar el test de Chow, para analizar la estabilidad de este modelo vamos a eliminar las tres últimas observaciones (los años 95, 96 y 97, lo que representa el 27,3% del total de las consideradas) y volver a estimar el modelo. Los resultados obtenidos son los siguientes:

Variables en el modelo:				
	B	S(B)	VALOR T	PROBAB. APROX.
AR1	,81865739	,17181445	4,7647761	,00204808

El parámetro obtenido varía en menos de 0,1, por tanto, se puede concluir que el modelo es estable.

PREDICCIÓN

Factor 1

Como existe adecuación entre los datos y el modelo, podemos obtener los valores del factor para los individuos del 87 al 97, es decir los valores predichos, que se recogen en la siguiente tabla:

Tabla III.30
Valores predichos para el primer factor

AÑO	87	88	89	90	91	92	93	94	95	96	97
Valor predicho	,00	,356	,453	,151	,059	-,092	-,143	-,218	-,117	-,210	-,109

Esos valores, junto con los obtenidos para los otros dos factores permiten reconstituir la tabla de datos original. Si bien, los correspondientes a los años 98 y 99 servirán para predecir las variables en esos años, es decir, uno y dos periodos hacia adelante.

Tabla III.31
Valores del primer factor para el 98 y el 99

AÑO	98	99
Valor predicho	-,11744	-,09852

Otros valores que aparecen como resultado del proceso de estimación son los indicadores de validez de las predicciones. Para el modelo *AR*, toman el valor de $-7,287$ para AIC y el valor $-6,889$ para SBC (epígrafe A.3.2.1, Estimación, p. 621). Estos indicadores permitirían resolver el problema de elección del modelo apropiado para predecir en el supuesto de que varios de los obtenidos fuesen válidos. Por tanto, en este caso, dichos valores no nos proporcionan gran información.

Factores 2 y 3

Dado que el segundo factor y el tercer factor son ruidos blancos, los valores predichos para este coinciden con los del factor original. Para los individuos-años 98 y 99, la predicción puntual para el factor fuera del horizonte temporal considerado permanece invariable (e igual a la media del proceso).

2.3.4 Validación del modelo

A partir de la fórmula de reconstitución de los datos y con los valores predichos obtenidos para los tres factores considerados, se obtienen los correspondientes para las variables originales, en el horizonte temporal

considerado, de acuerdo con la fórmula de reconstitución (capítulo II, apartado 3.2.5 del presente estudio) que se expresa de nuevo a continuación para la variable X_j de la tabla n :

$$\hat{X}_j^n = \frac{\sum_{k=1}^h \hat{F}_k^c \frac{g_{kj}^n}{\sqrt{\lambda_k^c}}}{\sqrt{\alpha_n \beta_n}} S_{x_j^n} + \bar{X}_j^n \Rightarrow \hat{x}_{tj}^n = \frac{\sum_{k=1}^h \hat{f}_{kt}^c \frac{g_{kj}^n}{\sqrt{\lambda_k^c}}}{\sqrt{\alpha_n \beta_n}} S_{x_j^n} + \bar{X}_j^n$$

Dichos valores se presentan en el epígrafe A.3.3.1 (p. 628) y su comparación con los realmente observados permite obtener los errores del modelo a través de la expresión:

$$e_{tj}^n = x_{tj}^n - \hat{x}_{tj}^n \quad t = 1, 2 \dots T$$

que son, por tanto, los errores cometidos al obtener la variable X_j de la tabla n a partir de los valores predichos de los factores y no a partir de sus valores originales. Para el caso que nos ocupa dichos errores se recogen en el mencionado apartado A.3.3.2 del anexo (p. 629). Para identificar el error correspondiente a cada variable se ha considerado la etiqueta original de dicha variable precedida de la letra E .

ANÁLISIS DEL MODELO

El análisis de dichos errores permitirá probar la validez del modelo propuesto. Si éste es válido, se obtendrán las predicciones para los años 98 y 99 de las variables originales, a partir de los valores de predicción de los tres factores y de la fórmula de reconstitución de los datos.

Para analizar las características de los errores obtenidos vamos a comenzar por el análisis de sus representaciones gráficas. Este análisis permite determinar, en todos los casos, la ausencia de una tendencia o comportamiento definido. Las figuras a, b, c y d del apartado A.3.3.3 del anexo (pp. 630-631), que recogen los errores para todas las variables analizadas, permiten llegar a dicha conclusión y, en todos los casos la

varianza parece mantenerse constante a lo largo del tiempo. Por ese motivo, este primer análisis ya permite llegar a la conclusión de que los errores pueden tener un comportamiento de ruido blanco.

Media cero

En el apartado A.3.3.3 del anexo (p. 632) se recogen los cocientes entre la media y la desviación típica de los residuos. Al observar dichos resultados se puede deducir que, dado que ninguno de esos valores superan el valor 2 (en términos absolutos), la media de las variables estudiadas, en este caso los errores, no difieren significativamente de cero.

Normalidad

Para el nivel de significación habitual ninguno de los valores p , permite rechazar al nivel de significación 0,05, la hipótesis nula de normalidad.

Errores Incorrelados

Como puede deducirse al observar el epígrafe A.3.3.3 del anexo (pp. 633-56), las funciones de autocorrelación y autocorrelación parcial de los errores de todas las variables, tienen el comportamiento de ruido blanco puesto que ningún retardo, al nivel habitual, es significativo. Por otra parte, el contraste global de Box-Ljung no permite rechazar la hipótesis nula de ausencia de correlación en ningún caso.

ESTABILIDAD DEL MODELO

La eliminación, en primer lugar, del último individuo (97) y, posteriormente, de los dos últimos 97 y 96, y la aplicación de nuevo, en ambos casos, del método *Statis*, permite obtener unos nuevos factores que no difieren de los

anteriores de manera significativa tal y como lo muestra las gráficas del anexo A.3.3.3 (figuras e-j: pp. 658-9, 661-2). Por tanto, los procesos *ARIMA* para dichos factores serían similares a los estimados para los factores originales, y también las predicciones realizadas. Es decir, el modelo de predicción propuesto se considera estable.

Como consecuencia, los errores tienen un comportamiento de ruido blanco, y por tanto el modelo puede emplearse para la predicción.

2.3.5 Predicción e interpretación de resultados

Los valores predichos de las variables originales, para el año 98 y 99, se recogen en las tablas III.32 a III.35. Como hemos indicado estos valores se han obtenido a partir de los factores estimados y de la fórmula de reconstitución. Así, tenemos que la predicción para esos años viene dada por las siguientes expresiones:

- para un periodo hacia delante (año 98):

$$\hat{X}_{(T+1)j}^n = \frac{\sum_{k=1}^h \hat{f}_{k(T+1)}^c \frac{g_{kj}^n}{\sqrt{\lambda_k^c}}}{\sqrt{\alpha_n \beta_n}} S_{x_j^n} + \bar{X}_j^n \quad ,, \quad j = 1, 2 \dots p_n; n = 1, 2 \dots N$$

- para dos periodos hacia delante (año 99):

$$\hat{X}_{(T+2)j}^n = \frac{\sum_{k=1}^h \hat{f}_{k(T+2)}^c \frac{g_{kj}^n}{\sqrt{\lambda_k^c}}}{\sqrt{\alpha_n \beta_n}} S_{x_j^n} + \bar{X}_j^n \quad ,, \quad j = 1, 2 \dots p_n; n = 1, 2 \dots N$$

Tabla III.32
Variables relativas a los seguros de vida

AÑO	VIEqui	VIgastexter	VIgastinter	VIpib	VIprim	Visinies
98	116,49	4,33	6,53	1,67	37,54	105,65
99	116,44	4,31	6,51	1,67	37,56	105,62

Tabla III.33
Variables relativas a los seguros de automóviles

AÑO	AUequi	AUgastexter	AUgastinter	AUpib	AUprim	AUsinies
98	104,58	13,20	16,82	1,34	31,42	74,56
99	104,57	13,22	16,83	1,34	31,41	74,52

Tabla III.34
Variables relativas a los seguros de salud

AÑO	SAequi	SAGastexter	SAGastinter	SAPib	SAPrim	SASinies
98	98,11	3,72	12,59	0,69	15,73	81,81
99	98,10	3,73	12,59	0,68	15,69	81,79

Tabla III.35
Variables relativas al resto de seguros

AÑO	REequi	REgastexter	REgastinter	REpib	REprim	RESinies
98	104,22	31,76	22,38	0,65	15,31	50,10
99	104,08	31,71	22,36	0,65	15,33	50,01

Más que los valores concretos, lo que resulta especialmente interesante es analizar la tendencia, a la baja o al alza, de las distintas variables. Según dichas tablas, respecto a los seguros de vida, se observa cierta tendencia a la reducción de su peso en la economía y en el sector, acompañado de un descenso del equilibrio y la siniestralidad del ramo. Mientras tanto, los gastos externos e internos muestran una tendencia ascendente.

La importancia, en el sector y en la economía, de los seguros del automóvil parece disminuir, lo mismo que la siniestralidad y el equilibrio técnico en el ramo. Además, mientras los gastos externos tienen a aumentar, los internos tienden a disminuir.

La tendencia de los seguros de salud es a la reducción de su importancia en el sector y en la economía, lo mismo que su siniestralidad, y al contrario que los gastos internos y externos del ramo. Respecto al equilibrio del ramo, solo es posible afirmar que presenta cierta tendencia al alza.

En el resto de los ramos no vida, parece observarse una tendencia al alza de su importancia en el sector, pero no en la economía. Los gastos

internos tienden a aumentar, al contrario que los externos; mientras tanto, el equilibrio y la siniestralidad presentan cierta tendencia, aunque no muy clara, a disminuir.

La comparación de los datos reales de los años 1998 y 1999 con los de las tablas anteriores, tomados de la misma fuente de información para que haya homogeneidad, permitirá revisar el modelo, añadiendo información adicional, con el fin de predecir a más largo plazo. A falta de dichos datos, una manera de aproximarse a la validez del modelo es considerar las predicciones obtenidas para las variables relativas al equilibrio: teniendo en cuenta que dichas variables se obtienen como resultado de sumar las relativas a los gastos internos, gastos externos y siniestralidad, este hecho también se cumple, excepto por razones de redondeo, con los valores predichos obtenidos. Por otra parte, para ambos años, la importancia de los diferentes ramos agregados en el sector (variables prim) suma el cien por cien del sector.

Una vez descrito el sector, la predicción mediante el procedimiento desarrollado permite hacer una serie de consideraciones:

- Los factores obtenidos, además de resumir la estructura general del sector en el periodo considerado, constituyen las dimensiones relevantes para predecir, recogiendo la mayor parte de la información de la tabla de datos original, en concreto el 79,51% del total.
- Al obtenerse de un método factorial, las dimensiones restantes de poca inercia recogen las posibles fluctuaciones, irregularidades, errores o tendencias puntuales de los datos, que no deben tenerse en cuenta para predecir ya que solo distorsionarían los resultados obtenidos.
- Aunque la utilidad descriptiva del procedimiento, respecto a los individuos (años), es especialmente interesante para detectar los

cambios más importantes en los periodos intermedios, la descripción de las diferencias entre el principio y el final del periodo, a través del método, permite demostrar la validez del mismo.

- La estabilidad de los factores obtenidos indica el mantenimiento, por lo menos a corto plazo, de la estructura descrita y la relación explicada entre las variables, lo que favorece el proceso de toma de decisiones.
- El primero de ellos, que hace referencia sobre todo a los gastos en el caso del automóvil y del resto de seguros, sigue un proceso $AR(1)$, es decir, los valores obtenidos están correlacionados con cualquier valor pasado o futuro de la serie, aunque esta correlación sea tanto más pequeña cuanto más distantes en el tiempo sean las observaciones.
- Dado que los demás factores son ruidos blancos gaussianos, de media cero, no tienen influencia en la elaboración de las predicciones, reforzando aún más la importancia del primer factor obtenido.
- Otra ventaja del procedimiento propuesto es la reducción del número de series a modelizar, pudiendo predecir veinticuatro variables tratando tan solo tres series, es decir, sin necesidad de analizar de manera individual todas las variables. La otra posibilidad, la predicción mediante Modelos Multivariantes de Series Temporales, complicaría en exceso el proceso, dado el número de parámetros a estimar.
- A nuestro juicio, y teniendo en cuenta además el poder explicativo de estos factores ya puesto de manifiesto respecto a las variables y a los individuos, el procedimiento propuesto es altamente satisfactorio, ya que permite la descripción del fenómeno estudiado además, y como paso previo, a la predicción.

CONCLUSIONES

De acuerdo con lo que hemos visto, consideramos que este trabajo ha logrado alcanzar el objetivo inicialmente planteado: la comprobación de que la metodología STATIS, bien de manera individual o bien en combinación con otras técnicas, permite un satisfactorio tratamiento de la información disponible, cuando se trata de describir, clasificar y predecir un fenómeno de naturaleza económica. A nuestro juicio, el cumplimiento de este objetivo y el procedimiento seguido para su logro, permiten demostrar la versatilidad de la metodología expuesta para analizar fenómenos económicos, sirviendo el caso analizado como ejemplo de su aplicación en un campo en el que se ha empleado escasamente hasta el momento, así como confirmar, de manera empírica, la importancia del sector asegurador español.

Y lo anterior puede afirmarse, a la luz de la metodología empleada y del sector concreto al que se aplica, según se ha constatado en los tres capítulos del presente estudio y que, sintéticamente, nos permiten indicar lo que sigue:

— Con relación al Capítulo I, y refiriéndonos a las Técnicas Descriptivas Multidimensionales, nos parece importante expresar que la distinción entre técnicas para tablas de doble entrada y para tablas de tres entradas pone de manifiesto una cuestión fundamental: en ningún caso estas últimas pueden analizarse mediante técnicas para tablas de doble entrada, aunque en realidad sean una sucesión de las mismas; el tratamiento de manera

individual de cada una de éstas por las técnicas que le son propias, plantearía el problema de comparación de los resultados individuales obtenidos, dejando de lado las relaciones que pueden existir entre las de doble entrada.

La distinción de técnicas, basada en el tipo de tabla de tres entradas analizada (según los tres modos, objetos, variables u ocasiones, estén o no totalmente cruzados), en el tipo de datos (cuantitativos, cualitativos o mixtos), y en los objetivos perseguidos, permite observar las características distintivas de las mismas, facilitando la tarea de seleccionar la que es conveniente aplicar en cada caso, e identificando diferencias y semejanzas entre técnicas próximas cuando se trata de caracterizar una en concreto, como ocurre en este estudio con el STATIS.

Una de las primeras conclusiones se concreta en que la metodología STATIS es la adecuada cuando las variables objeto de estudio son de tipo cuantitativo, y la estrategia global requiere análisis separados y conjuntos de los tres modos, o bien, cuando se trata de establecer una tipología de objetos o variables, equilibrando la influencia de cada uno de los grupos (de variables o de individuos).

Otra cuestión importante deducida de la revisión realizada es la comparación con otros métodos. En este sentido el STATIS es comparable con el Análisis Factorial Múltiple (AFM), teniendo en cuenta que ambos son métodos asimétricos y debido a las etapas de las que constan. Mientras el STATIS compara los datos entre los distintos momentos del tiempo (interestructura), y realiza un análisis detallado describiendo las relaciones entre variables e individuos (intraestructura), el AFM hace especial énfasis en las relaciones de la intraestructura, condicionando a ésta el análisis de la interestructura. Como consecuencia de esas diferencias en el procedimiento, aunque las principales dimensiones comunes son puestas de relieve por ambos métodos, el STATIS minimiza la importancia de las estructuras marginales que no participan en la estructura común, debiéndose emplear si

el objeto del estudio es obtener una visión global de las semejanzas entre los grupos. Conviene añadir que, si el modo común en las tablas son las variables, la aplicación del AFM únicamente es posible invirtiendo éstas, mientras que el STATIS podría ser aplicado sin modificaciones (*Statis dual*).

También es importante destacar que, al tener los métodos asimétricos una especial relevancia cuando las ocasiones se refieren a momentos del tiempo diferentes, es conveniente la comparación del STATIS con métodos que consideran el aspecto temporal de la información. Éste presenta ciertas ventajas respecto al *Doble Análisis en Componentes Principales*, ya que en algunas etapas de éste existen dificultades para encontrar un óptimo global, y para relacionar los análisis correspondientes al espacio de los individuos y al de las variables. Respecto al *Análisis de Series Cronológicas Multidimensionales*, ambos métodos proporcionan resultados similares (si se compara con el *Statis dual*), con la ventaja de que en el STATIS las tablas de datos no tienen porque tener el mismo número de individuos.

En cuanto a las Técnicas de Predicción, sabemos que cumplen un papel fundamental en la toma de decisiones en casi todos los ámbitos económicos (gestión de empresas, administraciones públicas, etc.) y, por tanto, su conocimiento resulta imprescindible para resolver los problemas que se plantean en el desenvolvimiento de dicha actividad. Todo ello permite afirmar que su clasificación ayuda a la selección de la que es apropiada en cada caso y permite poner de manifiesto bastante consenso entre las diferentes tipologías planteadas por los distintos autores.

De la revisión realizada y en la que se han destacado las basadas en información subjetiva, en el análisis aislado de series, o en las relaciones entre las variables, son precisamente en estas últimas técnicas en las que la aplicación del STATIS puede resolver ciertos problemas, ayudando al cumplimiento de su función predictiva. En concreto, el excesivo número de parámetros a estimar que surge en algunos *Modelos de Series Temporales*

Multivariantes puede ser reducido mediante el empleo de un Método Factorial, y es en este contexto en el que la metodología STATIS tiene una función predictiva que cumplir.

Estas cuestiones, desarrolladas ampliamente en el Capítulo I, han permitido definir el marco general que ayuda a caracterizar y organizar las diversas Técnicas Descriptivas (Multidimensionales) y Predictivas existentes. De esta forma hemos contextualizado la metodología sometida a examen en el presente estudio, sentando las bases para su desarrollo.

— Aunque se trata de un método ampliamente conocido, del desarrollo realizado sobre el *Análisis de Componentes Principales* (ACP), en la primera parte del Capítulo II, se pueden extraer algunas consideraciones generales, destacando de forma especial las diferencias entre la versión *normada* y la *no-normada*, que son, a nuestro entender, sumamente interesantes dado su empleo generalizado.

A pesar de otras opiniones al respecto, los planteamientos recientes sobre el ACP, en la línea del trabajo de Pearson de 1901, lo consideran como una técnica de representación de datos óptima, que puede emplearse de manera independiente, sin hipótesis de partida, debido a las propiedades geométricas y algebraicas que cumple. Sin embargo, el ACP no es siempre la herramienta tan eficaz que se pretende que sea, y en algunas situaciones, la identificación de componentes y la atribución de un significado concreto a las mismas puede llevar a conclusiones erróneas. Incluso en estos casos, es un método adecuado para reducir la dimensionalidad de los datos, aunque sólo es posible cuando las variables iniciales presentan coeficientes de correlación no nulos. Esta cuestión debe considerarse a través del Test de esfericidad de Barlett, o, en su defecto, analizando los valores propios de la matriz de correlaciones; en este último caso si éstos son todos similares, la nube es esférica y el ACP no permite obtener conclusiones relevantes.

Otra cuestión destacable con relación a esta técnica es que la imagen obtenida de los individuos con el ACP, de menores dimensiones que la original, refleja de forma adecuada las distancias originales entre los mismos, y mantiene una cantidad importante de información de la tabla de datos de partida. Las nuevas dimensiones constituyen la mejor aproximación, en el sentido de los mínimos cuadrados, de la matriz de datos original por otra de menor rango, pudiendo emplearse la última con fines predictivos, para reconstruir la primera.

La posición de cada individuo se caracteriza teniendo en cuenta la interpretación que de las nuevas dimensiones puede hacerse, en función de las variables con las que se correlacionan. Si todas las variables están correlacionadas positivamente entre ellas, dos a dos, formando ángulos agudos, existe una estructura común al conjunto de individuos, que queda reflejada por la primera componente, y que es la dirección que une el origen con el centro de gravedad, circunstancia en la que se basa la interestructura del STATIS.

Generalmente cuando las unidades de las variables que intervienen en el estudio son razonablemente comparables, y también sus varianzas, es preferible un ACP *no-normado* ya que cada componente extrae la k -ésima parte más grande del total de la varianza. En otro caso, las componentes extraídas reflejan las variables originales, en orden descendente de varianza, siendo entonces conveniente aplicar el ACP *normado*.

Así, en el ACP *normado*, todas las variables tienen la misma influencia en la determinación de las distancias entre individuos, y el producto escalar entre variables coincide con el coseno del ángulo que forman y con su coeficiente de correlación. Para el ACP *no-normado* las distancias están influidas por las unidades de medida y por las varianzas de las variables, siendo su producto escalar la covarianza. Las coordenadas de las variables con los ejes, en el ACP *normado*, coinciden con el coeficiente de correlación

de éstas con los factores, circunstancia que no se produce en el ACP *no-normado*.

Con respecto a la metodología STATIS, es necesario comenzar poniendo de relieve la escasa atención que dicha técnica ha suscitado por parte de la literatura española de Análisis de Datos. Por ese motivo, su desarrollo amplio y detallado, en la segunda parte del Capítulo II, pretende dar una visión completa de la misma, y contribuir, en la medida de lo posible, a su comprensión y difusión.

La primera cuestión a destacar respecto a la metodología analizada, es que existen dos versiones diferentes, aunque similares, para el estudio de individuos y de variables. De las diferentes fases de que consta se pueden extraer las siguientes conclusiones:

- En relación con la interestructura, debemos destacar que parte de la matriz de productos escalares entre operadores, y su descomposición en valores y vectores propios permite obtener una imagen euclídea óptima; este proceso permite su comparación de una manera sencilla y recoge gran parte de la información de la tabla de datos original. Esta característica de la interestructura viene motivada por la semejanza en dimensión y forma de las nubes analizadas, la cual debe comprobarse de forma previa y también se pone de manifiesto mediante coeficientes RV elevados.

- El compromiso, por su parte, tiene como objetivo la construcción de la nube (individuos o variables) que represente de la mejor manera posible las nubes analizadas, y que es obtenida como media ponderada de éstas. Esta nube constituye la estructura común de los individuos, o de las variables, siempre y cuando las tablas de partida sean semejantes. Es preciso señalar que, el sistema de ponderaciones de esta metodología tiende a reducir la influencia, en la construcción del compromiso, de aquellas nubes, de individuos o de variables, que más se alejan de la estructura común.

- La etapa de la intraestructura tiene como fin describir la estructura común obtenida en la fase anterior, a través de la imagen euclídea de los individuos–compromiso o de las variables–compromiso. Dicha imagen es la que se obtendría al realizar un ACP *centrado* de las nubes originales con la ponderación obtenida, de individuos o de variables, por lo que sus resultados e interpretación siguen las pautas de dicho análisis: los factores obtenidos se caracterizan a través de las correlaciones con las variables, y la posición de los individuos en los planos se interpreta teniendo en cuenta ese significado.

- Cuando las tablas analizadas corresponden a distintos momentos del tiempo, es interesante estudiar la evolución del comportamiento de los individuos, o de las variables, en el periodo considerado. Por ello, la cuarta fase del método se encarga de la construcción y análisis de las trayectorias, es decir, de los puntos representativos de un mismo individuo, o de la misma variable, cada uno referido a un momento del tiempo diferente, el correspondiente a cada tabla. En el *Statis*, las trayectorias de los individuos se obtienen considerando el operador de cada tabla como elemento suplementario del compromiso WD y en el *Statis dual*, son las trayectorias de las variables las que se obtienen como elementos suplementarios de VM.

Aunque la etapa de la intraestructura, tanto en una versión como en la otra, es un ACP de las tablas analizadas, la ponderación asignada a cada tabla para su análisis permite considerar sus características distintivas, analizando de forma más precisa las relaciones entre individuos o variables.

El segundo capítulo se completa con un examen exhaustivo de las posibilidades que la metodología STATIS ofrece para describir y predecir, algunas de las cuales no habían sido contempladas hasta el momento, o al menos, desde el punto de vista aquí considerado. Puesto que descripción y

predicción son fundamentales para el conocimiento de un fenómeno, el desarrollo de procedimientos que contribuyan a llevarlas a cabo de manera satisfactoria resultan de sumo interés; en este sentido, esta parte del presente estudio parece especialmente destacable.

- Cuando los individuos y las variables que forman las tablas de datos son los mismos, la aplicación simultánea de las dos versiones de la metodología STATIS, permite deducir las siguientes conclusiones:

Teniendo en cuenta la diferente naturaleza de las matrices analizadas en la etapa de la interestructura, la representación obtenida no es la misma en uno y en otro caso, pero las conclusiones son similares. Como es lógico pensar, si el comportamiento de los individuos varía de una tabla a otra, y por tanto sus nubes son diferentes (*Statis*), se debe a que éstos no presentan los mismos valores respecto a las variables analizadas, hecho que se refleja al comparar las nubes de éstas (*Statis dual*).

Por el mismo motivo, la ponderación de una nube de individuos para la construcción del compromiso del *Statis*, no es la misma que la ponderación que la nube de variables respectiva recibe en la construcción del compromiso del *Statis dual*, pero generalmente las nubes de menor ponderación, o las de mayor, son las mismas en ambos casos.

En la etapa de la intraestructura del *Statis*, si las T correlaciones de cada una de las variables se agrupan en torno al mismo factor, los factores respectivos del *Statis dual* tendrán un significado similar. Como consecuencia, es posible sustituir el cálculo de trayectorias de las variables, como individuos suplementarios, que haría el *Statis dual*, por las coordenadas de las variables, obtenidas con el *Statis*, en la etapa de la intraestructura; asimismo, las trayectorias de los individuos pueden analizarse a partir de los resultados obtenidos en la intraestructura del *Statis dual*. Las trayectorias así obtenidas,

constituyen una nueva posibilidad para el cálculo de las mismas, aunque esta solución sólo sea factible cuando pueden aplicarse las dos versiones.

- Cuando el objetivo de un estudio es la clasificación, además de la descripción, de una serie de individuos, y las variables observadas son cuantitativas y referidas a momentos del tiempo diferentes, la aplicación de la metodología STATIS, como paso previo para la clasificación mediante el *Análisis Cluster*, permite mejorar los resultados obtenidos. Al tomar como punto de partida el subespacio definido por los factores que recogen la mayor parte de la información, los grupos obtenidos consideran la dimensión real de la nube de puntos, y corrigen ciertas deformaciones debidas a la proyección en planos de pocas dimensiones. Es obvio que, en algunas ocasiones, los métodos de clasificación no pueden mostrar siempre la importancia de ciertas tendencias existentes, y el empleo previo de métodos factoriales permite el posterior posicionamiento de las clases o grupos sobre los ejes factoriales, así como la caracterización de esos grupos, identificando las tendencias o fenómenos latentes.

La clasificación a partir de factores facilita la elección del número de variables a considerar para la clasificación, y la distancia para medir la semejanza entre individuos: los primeros factores obtenidos, que no están correlacionados entre sí, son las dimensiones a considerar para el establecimiento de los grupos, siempre y cuando expliquen gran parte de la información de la tabla de datos original; por ese motivo, la medida de asociación entre individuos más apropiada es la distancia euclídea usual.

La elección de la Técnica de Clasificación apropiada depende de la información disponible, siendo recomendable la aplicación de una *Técnica Cluster* cuando hay poca o ninguna información sobre los posibles grupos. Dentro de dichas técnicas, teniendo en cuenta la

cantidad de métodos existentes que pueden aplicarse, la elección de uno u otro depende en muchos casos del criterio del investigador, destacando la utilidad del Método de Ward cuando se toman como punto de partida las puntuaciones factoriales.

- Si el objetivo final de un estudio es la predicción, uno de los problemas que más frecuentemente se plantean es la necesidad de reducir el número de parámetros, que crece exponencialmente conforme aumenta al número de variables relativas al fenómeno considerado. Cuando la naturaleza de la información disponible lo permite, la aplicación previa de la metodología STATIS consigue obtener nuevas variables, construidas a partir de las originales, que recogen las características principales de éstas, así como una representación más sencilla del fenómeno estudiado. Esta solución es mucho más interesante que la modelización univariante de cada serie, en cuanto que permite establecer relaciones entre variables, y contribuye al conocimiento y descripción del fenómeno analizado; también es una alternativa a los modelos *VARMA*, en la que la estimación de un gran número de parámetros altamente correlacionados entre sí complica en exceso el modelo.

Las relaciones existentes entre el *Análisis Factorial de un proceso*, basado en la descomposición de Karhunen-Loeve, y la fórmula de reconstitución de los datos, en un *Análisis de Componentes Principales*, son la base de este planteamiento. De esta manera, los factores obtenidos en un ACP pueden interpretarse como las variables aleatorias, no correlacionadas entre sí, que permiten reconstituir las variables originales y que constituyen la mejor aproximación de orden h de éstas.

Cuando la información de partida respecto al fenómeno estudiado viene dada por un conjunto de individuos (momentos del tiempo sucesivos), en los que se han observado una serie de variables, agrupadas en

diferentes tablas según su naturaleza, la reducción de la dimensionalidad no puede hacerse a través de un ACP, sino que es necesaria la aplicación de un método propio para tablas de tres entradas. Más concretamente, si el objetivo es estudiar la estructura común de las diferentes partes que configuran un fenómeno, expresadas sus peculiaridades por las diferentes tablas, la metodología STATIS proporciona la solución adecuada, aplicando la versión *Statist* si todas ellas están formadas por los mismos individuos.

El posicionamiento de los individuos (que en este caso son los distintos momentos del tiempo) sobre los ejes del compromiso da lugar a los factores, y permite describir la evolución del fenómeno estudiado, detectando los cambios más relevantes en las variables consideradas. El carácter estocástico de los factores obtenidos hace posible su tratamiento mediante la metodología *ARIMA*, de Box-Jenkins, aprovechando además sus ventajas, de todos conocidas, como técnica predictiva.

Una vez modelizados los h factores, los valores obtenidos para ellos permiten obtener los valores predichos para las variables de la tabla de datos original, en el periodo estudiado, empleando las fórmulas de reconstitución de los datos. Las diferencias entre las variables realmente observadas y estos valores predichos permiten definir los errores cometidos. Muchas de las pautas para el tratamiento de los residuos, que plantea la metodología Box-Jenkins en su fase de validación son, a nuestro juicio, válidas para enjuiciar dichos errores. Así, son consideradas cuestiones como su media nula, normalidad, incorrelación y estabilidad, y cuya evaluación satisfactoria permite concluir que el modelo propuesto es válido, y también las predicciones obtenidas de las variables originales fuera del horizonte temporal. En concreto, la estabilidad de los factores permite obtener conclusiones relativas a la permanencia de la estructura descrita en el periodo considerado.

Según nuestro criterio, y teniendo en cuenta el poder explicativo de estos factores respecto a las variables y a los individuos, el procedimiento propuesto es altamente satisfactorio, ya que permite una descripción y predicción válida para el análisis de información.

Con este Capítulo II creemos que, se alcanzan varios objetivos parciales que contribuyen a conseguir el objetivo principal señalado. El desarrollo del ACP, además de reparar en cuestiones que habitualmente se dan por conocidas, sirve de complemento imprescindible para comprender el desarrollo teórico de la metodología STATIS. Dicho desarrollo, y las nuevas posibilidades descriptivas, de clasificación y de predicción abordadas, contribuyen a difundir la misma, ya que estos nuevos planteamientos permiten aprovechar al máximo sus propiedades teóricas.

— Del desarrollo del Capítulo III se extraen conclusiones interesantes respecto al fenómeno estudiado, es decir, el sector asegurador español, cuyo tratamiento, a través de la metodología mencionada, sirve para demostrar empíricamente su utilidad en el análisis de datos económicos. Estas conclusiones se esquematizan con relación a las características generales de la actividad y del sector, a la estructura y grupos de entidades que lo configuran, y a la evolución del mismo.

- El sector asegurador, que tradicionalmente ha suscitado el interés de los poderes públicos con el fin de establecer medidas de protección del asegurado, ha despertado en épocas recientes un especial interés por parte de los ciudadanos al convertirse el seguro en un producto de consumo habitual, debido a la crisis de los sistemas tradicionales de Seguridad Social y a su creciente cultura financiera.

La importante función que el seguro desempeña en la actividad económica general, y en el ámbito financiero en particular, también justifica el interés por su estudio. La primera, que es la reducción del

riesgo, puede ser entendida desde el punto de vista microeconómico, como medio de estabilización del riesgo de los agentes económicos, facilitando la obtención de fondos por parte los prestamistas; en cuanto al enfoque macroeconómico, se hace referencia a la distribución equitativa de las pérdidas económicas de una minoría entre la mayoría, contribuyendo así a la creación de ahorro, a la estabilidad social y a la formación de la Renta Nacional, entre otras consecuencias.

Respecto al aspecto financiero y teniendo en cuenta que se trata de una actividad de intermediación financiera no bancaria, su importancia aumenta con el desarrollo económico debido a sus funciones, de cobertura del riesgo y de financiación de actividades económicas a través de la inversión de las provisiones técnicas. Además, se trata de uno de los intermediarios más eficientes del mercado, ya que el equilibrio que debe mantener entre depósitos e inversión es más fácil de conseguir que para otros, dada su vinculación a largo plazo con el cliente, y en los que la retirada masiva de fondos es menos probable.

Atendiendo a los rasgos que caracterizan a las entidades cuya actividad habitual es la realización de contratos de seguros, es decir, las aseguradoras, podemos afirmar que se trata de empresas de servicios en las que los elementos subjetivos, como la imagen y el prestigio de marca, juegan un papel decisivo. Dado que su proceso productivo está invertido, puesto que los ingresos preceden a los gastos, la gestión financiera de los fondos es fundamental para atender a la prestación del servicio cuando sea necesario, convirtiéndose los ingresos financieros en una parte importante de su cuenta de resultados. La permanencia en el tiempo y la estabilidad de las aseguradoras son dos conceptos estrechamente relacionados con esa actividad habitual.

- El sector ha experimentado en las últimas dos décadas una transformación importante, tanto cuantitativa como cualitativa,

impulsada sobre todo por los cambios legislativos. Así, ha pasado de ser un sector basado en principios intervencionistas, con una estructura cerrada en el mercado interior, a ser un sector dirigido a un mercado cada vez más abierto e internacionalizado, en el que los poderes públicos tutelan el acceso, ejercicio y cese de la actividad aseguradora, con el fin de mantener el adecuado equilibrio entre competitividad y solvencia.

La racionalización y liberalización del sector ha traído consigo, además de una drástica reducción del número de compañías (aproximadamente el 46% en 20 años), un rápido crecimiento del mismo, alcanzándose tasas de crecimiento de un 12,4% por término medio anual entre 1978 y 1990, y de un 18,6% entre 1991 y 1995, lo que ha provocado un aumento importante de la densidad de aseguramiento y de la penetración del seguro en la economía. Aún así, los niveles de aseguramiento en el caso español son bajos si se comparan con los de países de su entorno como Alemania, Dinamarca o el Reino Unido, por lo que el mercado español de seguros aún está lejos de la saturación y es considerado como un mercado ideal para las compañías extranjeras. La presencia de éstas siempre ha sido importante, alcanzando hasta la tercera parte del capital invertido en el sector, aunque se observa una tendencia decreciente que puede reflejar un cambio en la forma de acceso al mercado español en beneficio de otras, como las alianzas estratégicas o la libre prestación de servicios.

Por ramos, y en términos globales, el crecimiento más importante lo han experimentado los seguros de vida, que han tenido una tasa de crecimiento medio anual del 22,7% entre 1990 y 1997, frente al crecimiento medio anual del 8,67% en los seguros no vida. Entre éstos, y refiriéndonos a los grandes ramos, el crecimiento más importante se ha producido en los seguros de salud, con un 12,43% de crecimiento medio anual entre esos años; el otro de los grandes ramos, el del automóvil, tan solo presenta un crecimiento medio del

4,71%, experimentando tasas de crecimiento negativas en el último año. Estas disparidades se deben a las características diferentes de dichos mercados: mientras el mercado de vida y de salud es un mercado en expansión debido a la estructura demográfica de la población y la crisis de los sistemas de Seguridad Social tradicionales, el mercado del automóvil está saturado y en él la competencia es especialmente fuerte, hecho basado sobre todo en el desarrollo de nuevos canales de distribución, favoreciendo las reducciones del precio de las primas.

- Con relación al Análisis Estructural del sector y revisando las cinco fuerzas competitivas a las que hace referencia el modelo de Porter, se ponen de manifiesto ciertas cuestiones relativas a las amenazas y oportunidades que el entorno presenta para cualquier entidad que opere en el mismo, determinando su capacidad para obtener beneficios.

Respecto a los competidores potenciales, la diferenciación del producto necesaria para competir en el sector, sobre todo en términos de imagen de marca y prestigio, los costes que surgen cuando el cliente cambia de aseguradora en algunos ramos, y los requisitos para acceder a la actividad, en términos de capitales y exclusividad, constituyen barreras de entrada importantes, por lo que no son una fuerte amenaza para los beneficios del sector.

En cualquier caso, la situación es diferente según el ramo considerado. Así, en el caso de los seguros de vida, la amenaza de nuevos competidores, sobre todo procedentes del sector bancario (fenómeno denominado bancaseguros), es especialmente alta, puesto que las entidades financieras pueden sortear con éxito las barreras mencionadas, consiguiendo además una mayor eficiencia en el empleo de sus recursos. En el caso de los seguros del automóvil, aunque los bajos costes de cambio harían pensar en una alta amenaza

por parte de los competidores potenciales, la saturación del mercado, el comportamiento agresivo que tradicionalmente han adoptado los ya instalados ante las nuevas entradas es, a nuestro juicio, suficiente para disuadir a los competidores potenciales, que tendrían que ofrecer precios bajos y soportar pérdidas iniciales. Para el caso de los seguros de salud, teniendo en cuenta los costes de cambio y los requisitos para acceder a la actividad, no parece que las amenazas de nuevos competidores sean importantes.

La rivalidad existente en el sector puede considerarse bastante importante: teniendo en cuenta que ya que en casi todos los grandes ramos, excepto para los seguros de salud, han soportado resultados técnicos negativos en los últimos años, es decir, los gastos propios de la actividad no pueden ser cubiertos con los ingresos, teniéndose que cubrir esas diferencias con otro tipo de ingreso que, en la mayoría de los casos, es el financiero. En un contexto de tipos de interés decrecientes, como el actual, los resultados del sector están bastante comprometidos, reforzándose la necesidad de que las compañías se apoyen en sus resultados técnicos para alcanzar el equilibrio sostenido a largo plazo, lo que podría aumentar la competencia en todo el sector.

Considerando la situación para los diferentes ramos, en el caso de los seguros de vida, a pesar de que los costes de cambio y la diferenciación del producto reducirían el nivel de competencia, la diversidad de competidores, procedentes muchos de ellos de la bancaseguros. Su pequeña dimensión incrementa la competencia, aún cuando el ritmo de crecimiento ha sido elevado y el mercado está lejos de la saturación. En el caso de los seguros de automóviles la rivalidad es muy alta, puesto que los gastos externos e internos son bastante elevados, la competencia se basa en el precio, los costes de cambio de seguro del automóvil son bajos, y el mercado está saturado. La competencia es menor para los seguros de salud, ya que se trata de un ramo bastante concentrado, en el que el ritmo de

crecimiento ha sido elevado y presenta bastantes perspectivas de crecimiento.

La amenaza de los productos sustitutivos se puede considerar baja ya que el autoseguro y las compañías cautivas son situaciones poco frecuentes. Si bien, conviene especificar que para el seguro de vida, dado que constituye un instrumento de ahorro, su comportamiento se ve afectado por la existencia de otros productos de ahorro similares, como los planes de pensiones o los fondos de inversión; de todos modos, considerando el nuevo tratamiento fiscal establecido por la ley del IRPF para los seguros de vida, la amenaza de los sustitutivos mencionados es baja. En el caso del automóvil, la posibilidad de adquirir este tipo de seguros en el exterior si puede afectar de manera importante al ramo. Teniendo en cuenta que el sustitutivo más claro de los seguros de salud es la Seguridad Social pública, las actuaciones de la administración para mejorar ésta pueden afectar a la contratación de este tipo de seguros.

El poder de negociación de los proveedores, que en este caso son básicamente los suministradores de tecnología, es muy bajo dada la fragmentación del sector informático sobre todo. También pueden considerarse como proveedores el personal que la entidad emplea para la prestación del servicio, y su poder si puede ser importante, dado que son ellos los encargados de la transmisión de la imagen y prestigio de la entidad.

El poder de negociación de los clientes está muy limitado en el sector: los individuales, no pueden a bajo costo adoptar la condición de aseguradores; en cuanto a los corporativos, sí existe esa posibilidad pero no es sencillo dado los requisitos de acceso a la actividad, y las citadas empresas cautivas, que pueden considerarse una forma de integración vertical, solo representan cierta amenaza para el sector. Además, aunque los asegurados cada vez tienen más información, y por tanto su poder de negociación aumenta, los contratos de

adhesión entre asegurador y asegurado que se establecen para la mayoría de los ramos del seguro limitan ese poder.

- Atendiendo al estudio descriptivo realizado, a través de las dos versiones de la metodología STATIS, se puede deducir que las entidades más importantes del sector por volumen de primas, que representan aproximadamente el 80% de la cuota de mercado, permiten identificar los rasgos generales y las dimensiones estratégicas relevantes del sector y de las empresas en el periodo 1995-1997. Las conclusiones obtenidas son fundamentales para conocer mejor el sector y su estructura básica que, según se ha demostrado, no varía en exceso de unos años a otros; es decir, las aseguradoras no cambian de manera trascendental su comportamiento a corto plazo respecto a las variables estratégicas consideradas.

Como consecuencia de este estudio se deduce que, las entidades que dedican su negocio de forma prioritaria al seguro de vida presentan altos porcentajes de inversiones financieras y gastos externos e internos no muy altos con relación a las primas y bajos respecto al sector. Sin embargo, su margen de solvencia, sobre el mínimo, es más bien bajo, y el reaseguro cedido no tiene mucha importancia como mecanismo de cobertura de riesgos. Muchas de estas entidades emplean el canal de distribución bancario (bancaseguros), ya que los conocimientos necesarios para la distribución de este tipo de seguros son similares a los que poseen los empleados bancarios. Las que no emplean este canal de distribución presentan, en general, porcentajes superiores de inversiones financieras que el resto de aseguradoras de vida.

Estas características se oponen a las de las entidades que dedican su negocio de manera prioritaria a los seguros no vida, dedicados en su mayor parte a empresas, ya que en este caso la productividad de los capitales propios y de los recursos es menor, aunque su margen de

solvencia mantenido, sobre el mínimo, es mayor. La baja incidencia de la bancaseguros, la menor importancia de las inversiones financieras y el empleo frecuente del reaseguro cedido son otras de sus características.

Mientras que las entidades que se dedican mayoritariamente a los seguros de automóviles, dado su carácter de seguros no vida, tienen características similares a las que se acaban de exponer, los rasgos de las entidades cuyo objeto principal son los seguros de salud difieren bastante de las expuestas. Así, éstas últimas presentan unas inversiones financieras poco importantes respecto a las inversiones totales, y altas productividades de los capitales propios más las provisiones. El reaseguro cedido no es importante y los gastos, tanto internos como externos, alcanzan valores intermedios entre el nivel alcanzado en las entidades dedicadas a seguros de vida y el de las dedicadas a otros seguros no vida.

Las entidades con altos porcentajes de seguros de automóviles presentan gastos externos e internos relativamente altos, pero no tan elevados como los de entidades que se dedican a otros seguros no vida de carácter industrial, y tampoco niveles tan altos del margen de solvencia mantenido. También parece claro que son entidades en las que el reaseguro cedido no es importante y en las que la productividad de los capitales propios más las provisiones no es muy elevada.

Teniendo en cuenta lo anterior, se puede afirmar que la decisión básica que caracteriza la estructura de cualquier entidad es el tipo de producto que va a comercializar: seguro de vida o no vida. Las diferencias expuestas justifican las distinciones que los poderes públicos y las legislaciones hacen entre esos dos tipos de seguros. Además, dentro de los seguros no vida, hay claras diferencias entre aquellas entidades que se dedican de forma prioritaria a los seguros de salud, o de automóviles y el resto de entidades que comercializan seguros no vida.

-
- Dado que las características de las entidades dependen de la naturaleza de su negocio, el establecimiento de una tipología de las mismas en función de la dicha naturaleza, permite hacer grupos de entidades de características similares, lo que permite analizar cada entidad, tomando como referencia las entidades con las que se agrupa, siendo posible establecer semejanzas y diferencias entre las mismas. El procedimiento seguido, aplicación de la metodología *Statis* para determinar las dimensiones relevantes y posteriormente el *Análisis Cluster*, permite la obtención de grupos homogéneos, cuyo análisis permite validar dicho procedimiento para el establecimiento de grupos estratégicos en otros sectores; los principales obtenidos se exponen a continuación.

El grupo formado por Bankinter Vida, Bansabadell Vida, Lagun Aro Vida, Ascat Vida, Barclays Vida, Biharko Vida, Vidacaixa, Antares Vida, Argentaria Vida y Euroseguros, se caracteriza por una productividad mayor de los recursos que la media del sector (bajos porcentajes de gastos externos e internos) y por unos márgenes de solvencia mínima mantenidos más bien bajos. El porcentaje que las inversiones financieras representan sobre el total de inversiones es más elevado que para el resto de entidades del sector. Lo mismo que ocurre con la prima por empleado, por lo que posiblemente serán entidades en las que la bancaseguros juega un papel importante como canal de distribución. Tampoco son entidades en las que el reaseguro sea un sistema importante para la cobertura de riesgos, pero si presentan una productividad por encima de la media de los capitales propios.

EL segundo grupo obtenido y formado por Santander, Caja Madrid Vida, Aseguradora Valenciana, Db Vida, Intercaser, Seguros Génesis y Winterthur Vida, se caracteriza básicamente por los mismos rasgos que los ya señalados para las entidades anteriores, pero menos acentuados y en las que el fenómeno de la bancaseguros es menos importante. Asisa, Adeslas y Sanitas, que son entidades que dedican casi el 100% de su negocio a seguros de salud forman otro grupo, en

el que existe una alta productividad de sus capitales propios y provisiones, porcentajes de gastos externos e internos intermedios respecto a las medias del sector y márgenes de solvencia mantenidos no muy elevados.

Musini y Crédito y Caución son entidades en las que el porcentaje de primas procedentes de seguros no vida, especialmente el procedente de seguros industriales, es muy importante, presentando gastos externos e internos y márgenes de solvencia elevados respecto al resto de entidades del sector asegurador, pero valores inferiores a la media de productividad de los capitales propios, y de los capitales propios y provisiones.

El resto de entidades no vida, pueden caracterizarse por presentar altos porcentaje de negocio dedicados a los seguros industriales y altos porcentajes de gastos sobre primas, tanto externos como internos. Si bien hay diferencias: mientras Santa Lucia, Ocaso y Finisterre son entidades que no tienen primas procedentes de seguros de automóviles, y que por tanto, presentan unos gastos externos más elevados; otras como Mapfre Mutualidad, Fiact Mutua, Mutua Madrileña, Multinacional Aseguradora, Pelayo Mutua ó Winterthur Seguros Generales tienen porcentajes importantes de este tipo de seguros, y gastos más bajos. Plus Ultra, Zurich Seguros España, Mutua General de Seguros, AGF Unión y Fenix, Assicurazioni Generali, Catalana Occidente ó Vitalicio, son compañías en las que el seguro de automóviles no tiene especial relevancia y pueden catalogarse como entidades de seguros generales ya que dedican su negocio a una amplia variedad de ramos.

Parece que Caser Grupo, La Estrella, Mapfre Vida y Allianz Ras presentan unas características que difieren de la estructura común de las entidades consideradas, al menos en los años estudiados, aunque con cierta similitud a las entidades de seguros de vida.

Una vez que se han formado los grupos, las empresas que se encuentran dentro del mismo tienden a ser afectadas y a responder de igual forma a las circunstancias o movimientos en el sector industrial, confirmando la utilidad del concepto de grupo estratégico como instrumento descriptivo que contribuye al entendimiento de su estructura y dinámica competitiva.

- A pesar de que la estructura descrita y, por tanto, los grupos obtenidos, parece mantenerse entre 1995 y 1997, del análisis de las trayectorias se deduce que hay entidades que han modificado su comportamiento en el periodo considerado. Este es el caso de Banesto, que ha reorientado su negocio hacia el ramo de vida, reduciendo de forma considerable sus gastos externos e internos (respecto a las primas) y su margen de solvencia mantenido, acercando su posición a las entidades de bancaseguros que distribuyen productos de vida; Antares Vida y Barclays Vida también han reforzado su actividad en el mismo ramo, aumentando la productividad de sus capitales propios y de la prima por empleado. En el caso de la aseguradora Santander, el mismo hecho ha dado lugar también a una reducción de los gastos.

Entidades como Pelayo Mutua o Multinacional Aseguradora han cambiado su estrategia hacia el ramo del automóvil: la primera reforzando su actividad en dicho ramo, y la segunda en sentido contrario, acompañada, en este caso, de un importante aumento de la productividad de los capitales propios, que en el año 95 era negativa. Otros cambios a destacar son los que han experimentado Crédito y Caución, Zurich España y Mapfre Mutualidad, en las que se ha producido un aumento del porcentaje de gastos externos e internos (sobre primas), y un aumento de los márgenes de solvencia mantenidos.

-
- Una vez descrita la estructura del sector a través de las principales entidades que lo configuran, con el fin de completar el estudio se analizaron las perspectivas del mismo, prediciendo en la medida de lo posible el comportamiento de los principales indicadores que dan cuenta de su evolución. Dadas las diferencias entre los grandes ramos del sector, éstas se han tenido en cuenta para llevar a cabo dicha tarea, considerando dichos indicadores de manera particular para cada uno de ellos. Las conclusiones obtenidas del análisis empírico, realizado a través de *Statis* y la metodología Box-Jenkins, son las que se exponen a continuación.

Del análisis de las correlaciones entre variables y factores obtenidos en la etapa del compromiso se obtienen conclusiones relativas a la estructura y relaciones entre las variables estudiadas en el periodo 1987-1997. En el ramo de seguros de vida, parece que el incremento de su importancia en la economía, ha ido acompañado de un aumento del peso de dicho ramo en el sector, y viceversa, produciéndose un movimiento en sentido contrario de los gastos externos e internos. Por otra parte, en este caso dichos gastos no son los que determinan el equilibrio técnico, sino que es la siniestralidad, lo que parece lógico teniendo en cuenta que este concepto incluye el incremento de las provisiones matemáticas, exclusivas de este ramo del seguro. Como éstas suelen incrementarse cuando aumenta el número de primas, la mayor importancia del ramo influye en cierta medida en ese incremento.

Lógicamente, el incremento de la importancia de los seguros de vida en el sector trae consigo un descenso de la importancia de los no vida. Al considerar los diferentes ramos por separado, en los seguros de automóviles, los gastos externos e internos no presentan un comportamiento parejo; tampoco el incremento de su importancia en la economía parece influir en el comportamiento del ramo en el sector. Además, mientras los gastos internos parece que disminuyen

cuando aumenta la importancia del sector, los gastos externos lo hacen cuando aumenta ésta en la economía. En este caso, el equilibrio y la siniestralidad del ramo tampoco parecen estar influidos de manera especial por los gastos, ni tampoco por el comportamiento del ramo en el sector y en la economía.

Respecto a los gastos externos e internos en los seguros de salud, tampoco su comportamiento parece estar ligado. Sin embargo, para este tipo de seguros no vida si se observa que el incremento de su importancia en la economía influye en el mismo sentido que su comportamiento respecto al sector, afectando ambos hechos favorablemente a los gastos externos, ya que éstos tienden a disminuir. Al igual que los casos anteriores, el equilibrio está influenciado sobre todo por la siniestralidad y, en menor medida, por los gastos externos e internos. Conviene añadir que, en el periodo considerado, la siniestralidad y el equilibrio para este ramo ha tenido una evolución un tanto diferente a la siniestralidad y el equilibrio en el caso de los seguros de vida.

Respecto al resto de seguros no vida, los gastos internos y externos presentan una tendencia similar, siendo precisamente éstos últimos, junto con la siniestralidad, los que influyen en el nivel de equilibrio técnico del ramo. El aumento de la importancia de estos ramos en la economía no trae consigo un aumento de su peso en el sector, aunque si parece influir en el aumento de los gastos externos y, en menor medida, en los internos.

El estudio de las posiciones de los individuos (años) respecto a los factores del compromiso y la comparación de dichas posiciones, facilitan la identificación de las fluctuaciones más importantes del periodo. Así, el cambio de posición entre 1988 y 1989 permite deducir un incremento de la importancia en el sector de los seguros de vida, salud y resto de seguros no vida; el aumento de los gastos

internos, en el caso del automóvil y de los de salud; y el descenso de la siniestralidad y el equilibrio técnico, en los seguros de vida.

Entre 1989 y 1990 los principales cambios fueron experimentados por las variables de siniestralidad y equilibrio técnico en los seguros de vida, que aumentaron, y por las mismas variables en los seguros de salud, que disminuyeron. Los gastos, externos e internos, en este último ramo también disminuyeron, al igual que los internos en el resto de seguros no vida.

Entre 1991 y 1997 se produjo un aumento de la importancia de los seguros de vida, en el sector y en la economía, y de la siniestralidad y del equilibrio de ese ramo. Asimismo, también se ha detectado un aumento de la importancia de los seguros de salud en la economía, de los gastos internos en el ramo del automóvil, y de los gastos externos en el caso del resto de seguros no vida. Sin embargo, se ha producido un descenso del peso de los seguros del automóvil y del resto de seguros no vida en el sector, así como la reducción de los gastos internos y externos en el ramo de vida, de los internos en el caso de los de salud, y de la siniestralidad y del equilibrio técnico en el resto de seguros no vida.

A modo de resumen, se puede señalar que el sector ha sufrido transformaciones internas importantes en el periodo considerado, produciéndose un incremento notable de la importancia de los seguros de vida y salud y un descenso en el caso de los del automóvil. Estos cambios en el ambiente competitivo han influido en el comportamiento de las aseguradoras, algunas de las cuales han reorientado sus negocios hacia los ramos más rentables, lo que repercutirá en su organización interna. Como consecuencia, la estructura de las mismas se irá transformado, induciendo cambios en el sector, que influirán en las decisiones de las empresas.

-
- Una vez completado el proceso descriptivo, la predicción de valores concretos para los indicadores analizados en los años 98 y 99 constituyen el final del proceso de estudio del sector. Los factores empleados para tal fin, puesto que son el resultado de un método factorial, recogen la información relevante del fenómeno estudiado, dejando de lado las posibles fluctuaciones, irregularidades, errores o tendencias puntuales de los datos, que solo distorsionarían los resultados. Puesto que los otros son ruidos blancos, merece destacarse la importancia del primer factor, que es un proceso autorregresivo de orden uno. Por tanto, sus valores están correlacionados con cualquier valor pasado o futuro de la serie, aunque esta correlación sea tanto más pequeña cuanto más distantes en el tiempo sean las observaciones. Por esta razón, la predicción se ha limitado a un horizonte temporal de dos años.

Como en todo proceso de predicción, más que los valores concretos predichos, es necesario hacer referencia a las tendencias, al alza o la baja, que se deducen de dichos valores. Parece que los seguros de vida y de salud tienden a reducir su peso en la economía y en el sector, al contrario que los seguros de automóviles. El resto de seguros no vida parece que van a aumentar su importancia en el sector, a pesar de que su peso en la economía tiende a disminuir.

Respecto a los gastos, los internos tienden a aumentar en todos los grandes ramos considerados, excepto en el de los automóviles; los externos también presentan una tendencia alcista en los seguros de vida, automóviles y salud, al contrario que para el resto de seguros no vida, donde tienden a disminuir. En cuanto a la siniestralidad y el equilibrio técnico, parece que se va a comportar de manera positiva en todos los ramos considerados, teniendo en cuenta la tendencia a la baja que presenta en todos ellos.

A tenor de los comportamientos previstos, nos parece conveniente realizar las siguientes matizaciones:

- Parece que algunas de las tendencias señaladas no están de acuerdo con las perspectivas indicadas para algunos ramos del sector. Así, mientras los ramos de vida y salud tienen unas perspectivas muy favorables de crecimiento, considerando la estructura demográfica de la población y la crisis de los sistemas tradicionales de Seguridad Social, la tendencia señalada a través de los valores pronosticados es a la baja. Sin embargo, las tendencias pronosticadas sobre la reducción de la siniestralidad y del equilibrio técnico si son acordes con las creencias imperantes en el sector, teniendo en cuenta que una de las recomendaciones en todos los foros sobre el mismo es la atención preferente a la siniestralidad. El incremento de la importancia del ramo del automóvil, debido al aumento de la facturación y del volumen de primas, si parece estar también de acuerdo con las perspectivas para ese ramo, evolución favorable que es debida, probablemente, al crecimiento del parque de vehículos y no a la recuperación de los precios de la prima.
- La comparación de los datos reales de los años 1998 y 1999 con los estimados, siempre y cuando la fuente de información sea la misma, permitirá revisar el modelo, y añadir información adicional, con el fin de predecir a más largo plazo. Hasta el momento, dado que esa información ha dejado de ser de acceso público, no se puede disponer de ella. Esta limitación de información relativa al sector, además de su retraso, ha sido la tónica general que ha condicionado todo el proceso de estudio. Aparte de la información de obligada aportación, las entidades aseguradoras son bastante reticentes a proporcionar datos de su actividad, aunque se les solicite información no confidencial.

Teniendo en cuenta las conclusiones de este Capítulo III, consideramos oportuno afirmar la validez del proceso seguido para analizar un sector, ya que su Análisis General y Empírico permite profundizar en el conocimiento del mismo, además de proporcionar un ejemplo de aplicación de la metodología expuesta.

Como consecuencia de lo aquí se recoge, a nuestro modo de ver, este trabajo puede constituir un marco de referencia a tener en cuenta, pudiendo ser empleadas sus conclusiones como punto de partida tanto desde el punto de vista estadístico, como desde el punto de vista empresarial y empírico para las empresas e interesados en el sector asegurador:

— Respecto a la metodología STATIS, otros trabajos pueden analizar cuestiones puntuales que este trabajo ha dejado planteadas, como por ejemplo las correlaciones entre los factores obtenidos con el *Statis* y con el *Statis dual*, cuando se aplican simultáneamente, cuya relación se ha puesto de manifiesto, pero no ha sido matemáticamente demostrada. También es posible plantearse el empleo de diferentes ponderaciones de las tablas analizadas en la etapa de la interestructura, dando mayor importancia a las más próximas en el tiempo.

— El proceso señalado para la obtención de grupos estratégicos, cuando la información de la que se dispone se refiere a periodos diferentes, puede ser completado con otros métodos estadísticos que permitan validar las clasificaciones obtenidas.

— Ciertos aspectos relativos al *Análisis Factorial de un proceso* pueden ampliarse teniendo en cuenta algunas de las cuestiones planteadas en artículos como el de Aguilera *et al.* (1997), que plantea un modelo de *Regresión Lineal* a partir de las componentes principales asociadas con ciertas variables predictoras.

— En lo tocante al procedimiento predictivo propuesto, será necesario su revisión más profunda, poniendo especial énfasis en algunos aspectos relativos a su validez, elaborando indicadores que permitan enjuiciar de una manera más objetiva los resultados obtenidos.

— En cuanto al sector, se puede analizar el mismo considerando una mayor desagregación de algunos ramos, ya que aumentar el número de años o el número de entidades estudiadas choca con la limitación de la información disponible pero puede contribuir de manera significativa a su comprensión. Este proceso puede ser interesante, sobre todo para los seguros clasificados como resto de seguros no vida, en los que se ha hecho menos hincapié, habida cuenta de los ramos tan dispares que bajo ese concepto se han considerado.

Para finalizar este estudio sólo nos resta añadir, teniendo en cuenta las críticas y opiniones que cualquier procedimiento predictivo suscita siempre, las palabras de uno de los más prestigiosos estudiosos de la previsión en economía. Así, según Granger (1989: 163) "El buen predictor no es el más inteligente, sino el que, únicamente, tiene su ignorancia mejor organizada". Por tanto, cualquier procedimiento que permita ordenar la información y emplearla de manera útil, constituye un paso más en el avance de la Estadística y del conocimiento, y en ese sentido es como deben ser interpretados los resultados obtenidos en el presente trabajo. Interpretación que nos proponemos modificar teniendo en cuenta la aparición de futuros trabajos. Lo recogido en estas páginas es el intento de solidificar una buena base sobre la que anclar, edificar, agrandar, corregir el conocimiento estadístico y su aplicación a los más diversos campos.

BIBLIOGRAFÍA

-
- Abascal, E.; Grande, I. (1989), *Métodos Multivariantes para la investigación comercial*. Barcelona: Ariel Economía.
- Actualidad Internacional Aseguradora, (1999a), “Balance del seguro en 1998”, *Actualidad Aseguradora Internacional* (en línea): INESE (ref. de 20 de enero de 1999). Accesible a través de World Wide Web: <<http://www.inese.es/editoria/aa/portada.htm>.
- Actualidad Internacional Aseguradora, (1999b), “Un sector en transición”, *Actualidad Aseguradora Internacional* (en línea): INESE (ref. de 14 de octubre de 1999). Accesible a través de World Wide Web: <<http://www.inese.es/editoria/aa/inter1.htm>.
- Actualidad Internacional Aseguradora, (1999c), “Los unit linked tiran del seguro de vida, que creció un 31% hasta septiembre”, *Actualidad Aseguradora Internacional* (en línea): INESE (ref. de 16 de noviembre de 1999). Accesible a través de World Wide Web: <<http://www.inese.es/editoria/aa/adirecto.htm>.
- Aguilera, A. M.; Ocaña, F. A.; Valderrama, M. J. (1997), “An approximated principal component prediction model for continuous-time stochastic processes”, *Applied Stochastic Models and Data Analysis*, 13, pp. 61-12.
- Alcaide, A.; Álvarez, N. (1990), *Econometría. Modelos deterministas y estocásticos*. Madrid: Ed. Centro de Estudios Ramón Areces.
- Álvarez Esteban, R.; Gómez Riocerzo, M.; Huerga Castro, C.; Vallejo Pascual, E. (1996) “ Rasgos básicos del sector agrario castellano-leonés: un nuevo planteamiento espacial y temporal: Metodología STATIS”, *Actas del 5º Congreso de Economía Regional de Castilla y León*. Salamanca: Ed. Junta de Castilla y León, pp. 23-37.
- Amanieu, M.; Guelorget, O.; Nougier-Soule, J. (1981), “Analyse de la macrofaune beuthique d’une lagune littorale méditerranéenne”, *Vie Milieu*, 31, pp. 303-312.

-
- Anderson, T. W. (1958), *An introduction to Multivariate Statistical Analysis*. New York: Ed. John Wiley & Sons.
- Anderson, T. W. (1963), "The use of factor analysis in statistical analysis of multiple time series", *Psychometrika*, 28, pp. 1-25.
- Aseguradores (1999a), "La fidelidad de los clientes en cuestión", *Aseguradores*, 308, p. 17.
- Aseguradores (1999b), "Estadística sobre los canales de distribución", *Aseguradores*, 308, p. 21.
- Aseguradores (2000), "Perspectivas del Seguro y de la Economía para el año 2000", *Aseguradores*, 311, pp. 6-9.
- Aznar, A.; Trivez, F. J. (1993a), *Métodos de predicción en economía I. Fundamentos, Input-Output, Modelos econométricos y métodos no paramétricos de series temporales*. Barcelona: Ed. Ariel Economía.
- Aznar, A.; Trivez, F. J. (1993b), *Métodos de predicción en economía II. Análisis de series temporales*. Barcelona: Ed. Ariel Economía.
- Baena, V.; Juez, J. M. R. (1999), "Ranking", *Revista Ranking*, 132, pp. 87-112.
- Baena, V.; Lozano, L. (1997), "Ranking", *Revista Ranking*, 108, pp. 116-139.
- Baena, V.; Lozano, L. (1998), "Ranking", *Revista Ranking*, 120, pp. 75-98.
- Ballarín, E. (1985), *Estrategias competitivas para la banca*. Barcelona: Ed. Ariel, S.A.
- Baran, E. (1995), "Dynamique spatio-temporelle des peuplements de poissons estuariens en Guinée", *Thèse de troisième cycle*. Université de Bretagne Occidentale.
- Barquero Cabrero, J. D.; Huertas Colomina, F. J. (1997), *Manual de banca, finanzas y seguros*. Barcelona: Ediciones Gestión 2000.
- Bernard, M. C.; Díaz Llanos, F. J.; Escoufier, Y. (1978), "La méthode STATIS: une application à l'évolution des Campagnes Languedociennes". Centre de Recherches en Informatique et Gestion.
- Bernard, M. C.; Lavit, Ch. (1985), "L'évolution des sociétés rurales du Languedoc-Roussillon (1954-1975): application à la méthode STATIS", *110ème Congrès national des Sociétés savantes (Géographie)*, Montpellier. pp. 183- 196.

-
- Bisquerra Alzina, R. (1989), *Introducción conceptual al análisis multivariable*, I. Barcelona: PPU.
- Boucher de la Rupelle, J. M. (1980), “Techniques Graphiques en Analyse de données Mutidimensionnelles”, *Thèse de troisième cycle*. Montpellier: Université des Sciences et Techniques du Languedoc.
- Boumaza, R. (1998), “Analyse en composantes principales de distributions gaussiennes multidimensionnelles”, *Revue de Statistique Appliquée*, XLVI(2), pp. 5-20.
- Bouroche, J. M.; Saporta, G. (1987), *L'analyse des données. Collection Que sais-je?*. Paris: Ed. P.U.F.
- Bove, G.; Di Ciaccio, A. (1994), “ A user-oriented overview of multiway methods and software”, *Computational Statistics and Data Analysis*, 18, pp. 15-37.
- Box, G. E. P.; Tiao, G. C. (1977), “ A canonical analysis of multiple time series”, *Biometrika*, 64, 2, pp. 355-365.
- Box, G. P.; Jenkins, G. M. (1976) *Times series analysis. Forecasting and control*. San Francisco: Ed. Holden-Day.
- Braun, J. M. (1973), “Séries Chronologiques multiples, recherche d'indicateurs”, *Revue de statistique appliquée*, XXI(1), pp. 81-106.
- Bueno Campos, E. (1991), *Dirección Estratégica de la Empresa: metodología, técnicas y casos*. Ed.: Pirámide.
- Cabedo Semper, J.; Moya Clemente, I. (1999), “Cálculo del valor en riesgo mediante la metodología ARCH-Factorial”, *Revista Europea de Dirección y Economía de la Empresa*, vol. 8, 4, pp. 109-204.
- Cabezas Ares, A. M. (1996), “El sector de los seguros de salud: una visión estratégica”. *Trabajo de investigación dirigido por Juan Ventura Victoria. Departamento de Dirección y Economía de la Empresa*. Universidad de León.
- Cailliez, F.; Pages, J. P. (1976), *Introduction à l'analyse des données*, Paris: Ed. SMASH.
- Calvo, A.; Rodríguez Saiz, L.; Parejo, J.A.; Cuervo, A. (1999), *Manual del sistema financiero español*. Barcelona: Ed. Ariel Economía.
- Campos Guitart, A. (1999), “Los riesgos de un proyecto de implantación del Euro”, *Actualidad Aseguradora Internacional* (en línea): INESE (ref. de

-
- 20 de enero de 1999). Accesible a través de World Wide Web: <<http://www.inese.es/editoria/aa/opinion.htm>.
- Canals, J. (1990), *Estrategias del sector bancario. El reto de 1993*. Barcelona: Ed. Ariel Economía.
- Caridad, J. M^a.; (1998), *Econometría: Modelos Econométricos y Series Temporales*, I y II. Barcelona: Ed. Reverté.
- Carrier, A. (1985), “Application de l’analyse factorielle des évolutions et de l’analyse intra-période”, *Statistique et Analyse des Données*, 10, 1, pp. 27-53.
- Carroll, J. D. (1968), “Generalization of canonical correlation analysis to three or more sets of variables”, *Proc. of American Psychological Association*, pp.227-228.
- Casin, Ph. (1996), “L’analyse en composantes principales généralisée”, *Revue de Statistique Appliquée*, XLIV (3), PP. 63-81.
- Cegarra, J. G. (2000), “Hacia la nueva organización”, *Gestión, Revista de economía*, 12, pp. 12-20.
- Chatfield, C.; Collins, A. J. (1980), *Introduction to Multivariate Analysis*. New York: Ed. Chapman and Hall.
- Chessel, D.; Doledec, S. (1995), *ADE Versión 4*. Université Lyon 1, Villeurbanne cedex: Ed. HyperCard © Stacks and Programme library for the Analysis of Environmental Data. URA CNRS 1451.
- Chessel, D.; Hanafi, M. (1996), “Analyses de la co-inertie de K nuages de points”, *Revue de Statistique Appliquée*, XLIV(2), pp. 35-60.
- Círculo de empresarios (1999), “La gestión del conocimiento como ventaja competitiva”, *Documentos Círculo*, 99/05.
- CISIA (1997), *SPAD-TM Versión 3*. Saint-Mandé: Ed. CISIA
- Coppi, R. (1994), “An introduction to multiway data and their analysis”, *Computational Statistics and Data Analysis*, 18, pp. 3-13.
- Cuadras, C. M. (1980), *Métodos de Análisis Multivariante*. Barcelona: Ed. Editorial Universitaria de Barcelona.
- De la Fuente, J. M.; Azofra, V.; Hernangómez, J.; Gutiérrez, J.; de Andrés, P. (1995), “Análisis del comportamiento estratégico de las instituciones financieras españolas, 1987-1991”, *Documento de trabajo del FIES*, 117.

-
- Deville, J. C. (1974), “Méthodes statistiques et numériques de l’analyse harmonique”, *Annales de l’INSEE*, 15, pp. 3-97.
- Deville, J. C. (1978), “Analyse et prévision de séries chronologiques multiples non stationnaires”, *Statistique et Analyse des données*, 3, pp. 19-29.
- Dirección General de Seguros (1998a), *Memoria Estadística 1995*. Madrid: Ed. Secretaría General Técnica. Ministerio de Economía y Hacienda.
- Dirección General de Seguros (1998b), *Seguros y Fondos de Pensiones. Informe 1997*. Madrid: Ed. Secretaría General Técnica. Ministerio de Economía y Hacienda.
- Dot Pallerés, J. M. (1999), “La bancaseguros y la Venta Telefónica como canales alternativos”. *Encuentro sobre el sector asegurador español ante el cambio de siglo*. Santander: Universidad Internacional Menéndez Pelayo.
- Droesbeke, J.-J.; Fichet, B.; Tassi, P. (1994), *Modélisation ARCH. Théorie statistique et applications dans le domaine de la finance*. A, S, U., Bruselas: Ed. Ellipses.
- Durán Herrera, J. J. (1996), *Multinacionales Españolas I: Algunos casos relevantes*. Madrid: Ed. Pirámide.
- Durán Herrera, J. J. (1999), *Multinacionales Españolas en Iberoamérica. Valor Estratégico*. Madrid: Ed. Pirámide.
- Ecoutin, J. M. (1989), “Analyse des variations spatio-temporelles des peuplements de poissons de la baie Cocody (Côte d’Ivoire)”; *Mémoire de diplôme universitaire de Méthodes statistiques des sciences de la nature et de la vie*. Montpellier: Ed. USTL.
- Escofier, B.; Pagès, J. (1992), *Análisis factoriales simples y múltiples. Objetivos, métodos e interpretación*. Bilbao: Ed. Servicio editorial de la Universidad del País Vasco.
- Escoufier, Y. (1970), “Echantillonnage dans une population de variables aléatoires réelles”, *Publications de L’ISUP*, XIX (4), pp. 1-47.
- Escoufier, Y. (1973), “Le traitement des variables vectorielles”, *Biometrics*, 29, pp. 751-760.
- Escoufier, Y. (1976), “Opérateur associé à un tableau de données”, *Annales de l’INSEE*, 22-23, pp. 165 - 178.

-
- Escoufier, Y. (1977), "Operators related to a data matrix". Barra, J. R. (ed), *Recent developments in statistics*, North-Holland Publishing Company, pp. 125-131.
- Escoufier, Y. (1980), "L'analyse conjointe de plusieurs matrices", *Biométrie et Temps*, 59, pp. 59-76.
- Escoufier, Y. (1985), "Analyse conjointe de plusieurs tableaux de données", *Statistique et Analyse de Données*, 10, 1, pp. 1-10.
- Escoufier, Y.; L'Hermier des Plantes, H. (1978), "A propos de la comparaison graphique des matrices de variance", *Biometrical J.*, 20, 5, pp. 477-483.
- Escoufier, Y.; Pages, J. P.; Cazes, P. (1976), "Operateurs et analyse des tableaux à plus de deux dimensions", *Cahiers du BURQ*, 25, pp. 61-89.
- Escuder Vallés, R. (1987), *Métodos Estadísticos aplicados a la economía*. Barcelona: Ariel Economía.
- Estrategia, Estructura, Decisión, Identidad (STRATEGOR) (1995), *Política general de Empresa*. Barcelona: Ed. Masson.
- Fernández, R. (1999), "La industria aseguradora frente al reto europeo", *Actualidad Aseguradora Internacional* (en línea): INESE (ref. de 15 de marzo de 1999). Accesible a través de World Wide Web: <<http://www.inese.es/editoria/aa/document.htm>>.
- Fiegenbaum, A.; Thomas, H. (1990), "Strategic Groups and performance: the U. S. insurance industry, 1970-84", *Strategic Management Journal*, 11, pp. 197-215.
- Fiegenbaum, A.; Thomas, H. (1995), "Strategic Groups as reference groups: theory, modeling and empirical examination of industry and competitive strategy", *Strategic Management Journal*, 16, pp. 461-476.
- Florentino, J. (2000), "Seguros on line", *Ausbanc*, 116, pp. 83-86.
- Florido Castillo, M. (1999), "Una visión desde arriba", *Actualidad Aseguradora Internacional* (en línea): INESE (ref. de 7 de mayo de 1999). Accesible a través de World Wide Web: <<http://www.inese.es/editoria/aa/opinion.htm>>.
- Flury, B. N. (1984), "Common Principal Components in K Groups", *Journal of the American Statistical Association*, 79, 388, pp. 892-898.
- Foucart, T. (1978), "Sur les suites de tableaux de contingence Indexes par le temps", *Statistique et Analyse de Données*, 2, pp. 67-84.

-
- Foucart, T. (1983) “Une nouvelle approche de la Méthode STATIS”, *Revue de Statistique Appliquée*, XXXI (2), pp. 61-75.
- Fundación Mapfre Estudios, (1991a), “La distribución comercial del seguro: sus estrategias y riesgos”, *Cuadernos de la Fundación Mapfre*, 5. Madrid: Fundación Mapfre Estudios.
- Fundación Mapfre Estudios, (1991b), “Los seguros en una Europa cambiante: 1990-1995”, *Cuadernos de la Fundación Mapfre*, 4. Madrid: Fundación Mapfre Estudios.
- Fundación Mapfre Estudios, (1994), *El mercado español de seguros en 1994*. Madrid: Fundación Mapfre Estudios.
- Fundación Mapfre Estudios, (1995), *El mercado español de seguros en 1995*. Madrid: Fundación Mapfre Estudios.
- Fundación Mapfre Estudios, (1996), *El mercado español de seguros en 1996*. Madrid: Fundación Mapfre Estudios.
- Fundación Mapfre Estudios, (1997), *El mercado español de seguros en 1997*. Madrid: Fundación Mapfre Estudios.
- García de Quevedo Ruiz, J. C. (1999), “El ahorro privado en previsión social. Situación actual y perspectivas”. *Encuentro sobre el sector asegurador español ante el cambio de siglo*. Santander: Universidad Internacional Menéndez Pelayo.
- García Delgado, J. L. (1993), *Lecciones de Economía Española*. Madrid: Ed.: Civitas.
- Gay, B. (1998), “La rentabilidad del sector cae a los niveles de 1995”, *Revista Ranking*, 120, pp. 24-30.
- Gay, B. (1999), “Seguros: comprar, vender, crecer”, *Revista Ranking*, 132, pp. 28-36.
- Génard, M.; Souty, M.; Holmes, S. (1994), “Correlations among quality parameters of peach fruit”, *Journal of the science of food and agriculture*, 66, pp. 241-245.
- Girard, S.; Chalmond, B.; Dinten, J.-M. (1998), “Une ACP non linéaire basée sur l’approximation par variété”, *Revue de Statistique Appliquée*, XLVI(3), pp. 5-19.
- Glaçon, F. (1981), “Analyse conjointe de plusieurs matrices de données. Comparasion de différentes méthodes”, *Thèse de troisième cycle*, Grenoble.

-
- González de Frutos, P. (1999), “El sector asegurador español ante el cambio de siglo”, *Conferencia de Inaguración del Encuentro sobre el sector asegurador español ante el cambio de siglo*. Santander: Universidad Internacional Menéndez Pelayo.
- Gower, J. C. (1975), “Generalized Procrustes Analysis”, *Psychometrika*, 40, pp. 33-51.
- Gower, J. C. (1989), “Generalized Canonical Analysis”. Coppi, R.; Bolasco, S. (ed), *Multway Data Analysis*. Amsterdam: Ed. North-Holland, pp. 221-232.
- Granger, C. W. J. (1989), *Forecasting in business and economics*, 2ª Edición. San Diego: Ed. Academic Press.
- Granger, C. W. J.; Newbold, P. (1986), *Forecasting Economic Time Series*. 2ª Edición. San Diego: Ed. Academic Press.
- Granola, J. (1985), “Corporate modeling at allstate insurance company”. Cumming, J. D. (ed), *Strategic planning and modeling in property-liability insurance*. Boston: Ed.: Kluwer-Nijhoff Publishing, pp. 139-155.
- Grant, R. M. (1995), *Dirección Estratégica: conceptos, técnicas y aplicaciones*. Ed.: Civitas.
- Greenacre, M. J. (1999), “Correspondence Analysis in Marketing Research”, *Documentación del curso de la 5ª Applied Statistic Week*. Barcelona: Universidad Pompeu Fabra, pp. 1-9.
- Groupe d'étude et de réflexion interrégional (GERI) (1996), *L'analyse des données évolutives. Méthodes et applications*. Paris: Ed. Technip.
- Gutiérrez, R.; González, A. (1991), *Estadística Multivariante*, I y II. Granada: Ed. Facultad de Ciencias de Granada.
- Harman, H. H. (1980), *Análisis Factorial Moderno*. Madrid: Ed. Saltés.
- Harshman, R. A.; Lundy, M. E. (1984) “ The PARAFAC Model for three-way for three-way factor Anaysis and Multidimensional Scaling”. Coppi, R.; Bolasco, S. (ed), *Multway Data Analysis*. Amsterdam: Ed. North-Holland, pp. 122-215.
- Harshman, R. A.; Lundy, M. E. (1994) “ PARAFAC: Parallel Factor Analysis”, *Computational Statistics and Data Analysis*, 18, pp. 39-72.
- Heskett, J. L. (1988), *La gestión en las empresas de servicios*. Esplugues de Llobregat: Ed. Plaza @ Janés.

-
- Hope, Ch.; Mühleman, A. (1997), *Service Operations Management. Strategy, design and delivery*. London: Ed. Prentice Hall.
- Hotelling, H. (1936) "Relation between two sets of variables", *Biometrika*, 28, pp. 129-149.
- Huerga Castro, M. C.; Vallejo Pascual, E.; Abad González, J. (1998), "Análisis comparativo de las provincias de Castilla y León según sus exportaciones e importaciones en el periodo 1990-96", *Actas del 6º Congreso de Economía Regional de Castilla y León*. Salamanca: Ed. Junta de Castilla y León, pp. 410-428.
- ICEA (1999), "El seguro del automóvil: últimos datos. Situación de las bonificaciones" ICEA. Accesible a través de la World Wide Web: <<http://www.icea.es/ultima%20hora/12-11-99.htm>>.
- Judez Asensio, L. (1989), *Técnicas de análisis de datos multidimensionales*. Madrid: Ed. Ministerio de Agricultura, Pesca y Alimentación.
- Khodja, H.; Lubès-Niel, H.; Sabatier, R.; Masson, J.-M.; Servat, E.; Paturel, J.-E. (1998), "Analyse spatio-temporelle de données pluviométriques en Affique de l'ouest. Recherche d'une rupture en moyenne. Une alternative intéressante: Les Tests de permutations", *Revue de Statistique Appliquée*, XLVI(1), pp. 95-110.
- Kiers, H.A.L. (1988), "Comparison of anglo-saxon and french tree-mode methods", *Statistique et Analyse des Données*, 13, 3, pp. 14-32.
- Kiers, H.A.L. (1989), *Three-way methods for the analysis of qualitative and quantitative two-way data*, University of Leiden: Ed. DSWO Press.
- Kiers, H.A.L. (1991), "Hierarchical relations among three-way methods", *Psychometrika*, 56, pp. 449-470.
- Kroonenber, P. M. (1994), "The TUCKALS line. A suite of programs for three-way data Analysis", *Computational Statistics and Data Analysis*, 18, pp. 73-96.
- L'Hermier des Plantes, H. (1976), "Structuration des Tableaux à Trois Indices de la Statistique", *Thèse de 3ème cycle*, Université des Sciences et Techniques du Languedoc.
- L'Hermier des Plantes, H.; Thiebaut, B. (1977), "Étude de la pluviosité au moyen de la Méthode STATIS", *Revue de Statistique Appliquée* XXV (2), pp. 57-81.
- Lado Cousté, N. (1996), "Tipos estratégicos, orientación al mercado y resultados económicos: Análisis empírico del sector asegurador español",

-
- Cuaderno de la fundación Mapfre*, 31. Madrid: Fundación Mapfre Estudios.
- Lado Cousté, N.; Rivera Camino, J. (1996), "Forms of market strategies in the European insurance sector", *Working Paper* 96-33. Universidad Carlos III de Madrid.
- Lafosse, R. (1997), "Analyse de concordance de deux tableaux: monogamies, simultanées et découpages", *Revue de Statistique Appliquée*, XLV(3), pp. 45-72.
- Lafosse, R.; Hanafi, M. (1997), "Concordance d'un tableau avec K tableaux: définition de k+1 uples synthétiques", *Revue de Statistique Appliquée*, XLV(4), pp. 111-126.
- Lambert, P.; Lindsey, J. K. (1999), "Analysing financial returns by using regression models based on non-symmetric stable distributions", *Journal of the Royal Statistical Society. Series C: Applied Statistics*, vol. 48, 3.
- Landaluce Calvo, M. J.; Valencia Garcia, O. (1998) "Estudio comparativo de la metodología STATIS y AFM: Una aplicación", *Actas del 6º Congreso de Economía Regional de Castilla y León*. Salamanca: Ed. Junta de Castilla y León.
- Landaluce Calvo, M. I. (1995) "Estudio de la estructura de gasto medio de las Comunidades Autónomas españolas. Una aplicación del análisis factorial múltiple". *Tesis de tercer ciclo*. Universidad del País Vasco, pp. 442-455.
- Lavit, Ch. (1985), "Application de la methode STATIS", *Statistique et Analyse de Données*, 10, 1, pp. 103-116.
- Lavit, Ch. (1988), *Analyse conjointe de tableaux quantitatifs*, Paris: Ed. Masson.
- Lavit, Ch.; Escoufier, Y.; Sabatier, R.; Traissac, P. (1994), "The ACT (STATIS method)", *Computational Statistics and Data Analysis*, 18, pp. 97-119.
- Lavit, Ch.; Pernin, M-O. (1987), "Solution using STATIS". Jansen, J.; Marcotorchino, F.; Proth, J. M. (ed), *Data Analysis: the ins and outs of solving real problems*. New York: Ed. Plenum Press.
- Lebart, L.; Morineau, A.; Fénelon, J. P. (1981), *Traitement des Données Statistique*. Paris: Ed. Dunod.
- Lebart, L.; Morineau, A.; Piron, M. (1995), *Statistique Exploratoire Multidimensionnelle*. Paris: Ed. Dunod.

-
- Lebart, L.; Morineau, A.; Tabard, N. (1977), *Techniques de la description Statistique*. Paris: Ed. Dunod.
- Lebart, L.; Morineau, A.; Warwick, K. (1984), *Multivariate Descriptive Statistical Analysis*. New York: Ed. John Wiley & Sons.
- Lecea Flores de Lemus, A. de (1991), “Los sectores de la banca y del seguro españoles ante el mercado único”, *Revista de Economía ICE*, 690, pp. 33-50.
- Manzano Martos, A. (1993), “La legislación española de seguros y su adaptación a la normativa comunitaria”, *Cuadernos de la Fundación Mapfre*, 16. Madrid: Fundación Mapfre Estudios.
- Martínez Martínez, M. A. (1993), “Organización y estrategia de la Empresa. El ejemplo del sector asegurador español”, *Tesis doctoral*. Universidad de Alcalá de Henares.
- Martínez Rego, A. (1991), “Dimensión y eficiencia de la empresa aseguradora en España”, *Tesis doctoral*. Universidad Autónoma de Madrid.
- Martín-Guzmán, M. P.; Martín Pliego, F. J. (1989), *Curso básico de Estadística Económica*. Madrid: Ed. AC.
- Mc Cleary, R.; Hay, R. A. (1982), *Applied time series analysis for the social sciences*. Beverly Hills: Ed. Sage Publications.
- Mc Donald, R. P. (1984) “The Invariant Factor Model for multimode data”. Coppi, R.; Bolasco, S. (ed), *Multway Data Analysis*. Amsterdam: Ed. North-Holland, pp. 285-307.
- Meilán, J. (1999), “Los equipos sanos son los más rentables”, *Ausbanc*, 111, p. 106.
- Mercado de riesgos (1999), “Mapfre lidera la mayor parte de los ramos del sector seguros”, *Mercado de riesgos*, 45, pp. 17-25.
- Millán Aguilar, A. (1999), “Aportaciones del nuevo Reglamento de Ordenación del Seguro Privado”, 104, pp. 48-60.
- Mills, T. (1993), *The econometric modelling of financial time series*. Cambridge: Ed. Cambridge University Press.
- Morineau, A.; Sammartino, A-E.; Gettler-Summa, M.; Pardoux, C. (1994), “Analyses des données et modélisation des séries temporelles. Application à la prévision des ventes de périodiques”. *Revue de Statistique Appliquée*, XLII (4), pp. 61-81.

-
- Morrison, D. F., (1967) *Multivariate Statistical Methods. Second Edition*. Tokyo: Ed. Mc Graw Hill.
- Muñoz, A. (1999), “Un ejercicio irrepetible”, *Memoria Anual de UNESPA 1998* (en línea): UNESPA, 1999 (ref. de 14 de octubre de 1999). Accesible a través de World Wide Web: <<http://www.unespa.es/memoria/home.htm>.
- Mures Quintana, M. J. (1991), “Procesos Estocásticos de difusión lognormal multivariante. Aplicación a la modelización del gasto público en España”. *Tesis doctoral*. Universidad de Granada.
- Navas López, L.; Guerras Martín, L. A. (1998), *La dirección estratégica de la empresa: teoría y aplicaciones*. Ed.: Civitas.
- Oliva Fures, M.; Carles Lavilla, M. (1999), “Concentración y beneficios en la industria de seguros sanitarios privados”, *Documento de trabajo*, 99-04. Universitat Rovira i Virgili, pp. 1-23.
- Otter, P. W.; Schuur, J. F. (1982), “Principal Component Analysis in multivariate forecasting of economic time-series”. Anderson, O. D. (ed), *Time series Analysis, Theory and Practice*, 1. New York: Ed. North-Holland Publishing Company, pp. 323-347.
- Pagès, J. (1996) “Eléments de comparaison entre l’analyse factorielle multiple et la méthode STATIS”, *Revue de statistique Appliquée*, XLIV (4), pp. 81-95.
- Pardoux, C. (1989), “Apport de l’analyse factorielle à l’étude d’un processus”, *Revue de Statistique Appliquée*, XXXVII (4) , pp. 41-60.
- Pardo-Vivero, A. (1999), “¿Qué hubo, como le fue -al mundo- en 1998?”, *Actualidad Aseguradora Internacional* (en línea): INESE (ref. de 20 de enero de 1999). Accesible a través de World Wide Web: <<http://www.inese.es/editoria/aa/opinion.htm>.
- Paul, M. (1997), “El seguro en clave fiscal”, *Revista Ranking*, 108, pp. 24-33.
- Peña, D. (1989), *Estadística: Modelos y Métodos*, II. Madrid: Ed. Alianza Universidad Textos.
- Peña, D. (1996), “El futuro de los métodos estadísticos”, *Documento de trabajo*, 96-04, pp.1-11.
- Peña, D.; Box, G. E. P. (1984), “Hidden relationships in multivariate time series”, *Proc. of Busi. Econ. St. Am. Statistic Association*, pp. 494-499.

-
- Peña, D.; Box, G. E. P. (1987), "Identifying a simplifying structure in time series", *Journal of the American Statistical Association*, 82, 399, pp. 836-843.
- Peña, D.; Poncela, P. (1996), "Pooling Information and forecasting with dynamic factor analysis", *Working Paper*, 96/63. Madrid: Ed. Dpto. de Estadística y Econometría. Universidad Carlos III de Madrid.
- Pérez Barbero, P. A. (1999), "Competitividad, tecnologías de vanguardia y calidad, son las claves para liderar el mercado", *Revista Ranking*, 132, pp. 121-123.
- Pérez Hugalde, C. (1988), *Evolución de las relaciones entre la productividad del trabajo en la agricultura, las estructuras agrarias y el desarrollo económico en España 1960-1979. Un análisis mediante métodos estadísticos multivariantes*. Madrid: Ed. Ministerio de Agricultura, Pesca y Alimentación. Colección Tesis Doctorales.
- Pérez Hugalde, C.; García Velazquez, A.; Júdez, L. (1986), "Análisis de la evolución de los resultados económicos de los sectores agrario y no agrario en el proceso de desarrollo (1960-1980)", *Revista de Estudios Agro-sociales*, 137 extra, pp. 317-351.
- Pérez Pérez, L.; Pérez Hugalde, C. (1990), "Nota sobre la dinámica comarcal de la dimensión de las explotaciones agrarias en Aragón", *Investigación Agraria (Economía)*, 5(1), pp. 95-108.
- Pérez Torres, J. L. (1999), "Aseguradoras del siglo XXI: Hacia un sistema de organización", *Actualidad Aseguradora Internacional* (en línea): INESE (ref. de 20 de enero de 1999). Accesible a través de World Wide Web: <<http://www.inese.es/editoria/aa/opinion.htm>>.
- Pernin, M-O. (1986), "Contribution à la méthodologie d'analyse de données longitudinales. Exemple de la croissance chez l'êtré humain". *Thèse de troisième cycle*. Université Claude Bernard Lyon I.
- Pindyck, R. S.; Rubinfeld, D. L. (1976), *Modelos Económicos*. Barcelona: Ed. Labor Universitaria.
- Pinillos Costa, M. J.; Martín Peña, M. L. (1998), "Grupos estratégicos en el sector asegurador, 1991-1995: El impacto de la bancaseguros", *Documentos de trabajo del FIES*, 140.
- Place, M-C. (1980), "Contribution algorithmique à la mise en oeuvre de la méthode STATIS", *Thèse de 3ème cycle*, Université des Sciences et Techniques du Languedoc.

-
- Pontier, J.; Pernin, M. O. (1989), "Relations entre analyse canonique complète et méthode LONGI", *Revue de Statistique Appliquée*, XXXVII (4), pp. 67-82.
- Porter, M. E. (1982), *Estrategia competitiva. Técnicas para el análisis de los sectores industriales y de la competencia*. México: Ed. Cecsá.
- Prieto Pérez, E. (1976), "Las entidades de seguros como intermediarios financieros", *Anales del Instituto de Actuarios Españoles*, pp. 81-92.
- Pulido, A. (1989a), *Modelos Económicos*. Madrid: Ed. Pirámide.
- Pulido, A. (1989b), *Predicción económica y empresarial*. Madrid: Ed. Pirámide.
- Queysanne, M. (1975), *Algebra básica*, Barcelona: Ed. Vicens-Vivens.
- Rala Esparteiro, M. de N. (1994), *Las relaciones banca-seguros*. Madrid: Centro de Estudios del Seguro.
- Revista Ranking (1996), "Ranking", *Revista Ranking*, 97, pp. 85-113.
- Robert, P.; Escoufier, Y. (1976), "A unifying tool for linear multivariate statistical methods: the RV-coefficient", *Applied Statistics*, 25, 3, pp. 257-265.
- Saporta, G. (1981), "Méthodes exploratoires d'analyse de données temporelles", *Cahiers du BURO*, 37-38. Paris: Ed. Université Pierre et Marie Curie.
- Saporta, G. (1985), "Data analysis for numerical and categorical individual time series", *Applied Stochastic Models and Data Analysis*, 1, pp. 109-119.
- Saporta, G. (1990), *Probabilités, Analyse des Données et Statistique*, Paris: Ed. Technip.
- Sastre, J. A. (1999), "Preocupación por los cambios estratégicos", *Actualidad Aseguradora Internacional* (en línea): INESE (ref. de 7 de mayo de 1999). Accesible a través de World Wide Web: <<http://www.inese.es/editoria/aa/opinion.htm>>.
- Simier, M.; Blanc, L.; Pellegrin, F.; Nandris, D. (1999), "Approche simultanée de K couples de tableaux: Application à l'étude des relations pathologie végétale-environnement", *Revue de Statistique Appliquée*, XLVII(1), pp. 31-46.

-
- Simon, C. (1993), *Présentation de méthodes d'analyse de tableaux à trois dimensions et applications à l'étude de la satisfaction de la clientèle de France Télécom*. Université Pierre et Marie Curie: Mémoire ISUP.
- Snyder, C. W.; Law, H. G. (1984) "Overview of multimode Analytic Methods", *Research methods for multimode data analysis*. New York: Ed. Praeger. pp. 3-25.
- Stevens, J. (1986), *Applied Multivariate Statistics for the social sciences*. New Jersey: Ed. LEA.
- Suárez Llanos, C. (1999), "Las provisiones técnicas en el nuevo Reglamento". *Encuentro sobre el sector asegurador español ante el cambio de siglo*. Santander: Universidad Internacional Menéndez Pelayo.
- Tenenhaus, H.; Prieuret, B., (1971) *Analyse des séries chronologiques multidimensionnelles*. Ottawa: Ed. Universidad de Ottawa.
- Tennenhaus, M. (1999), "L'Approche PLS", *Revue de Statistique Appliquée*, XLVII(2), pp. 5-40.
- Thioulouse, J.; Chessel, D. (1987), "Les analyses multitableaux en écologie factorielle". *I- De la typologie d'état à la typologie de fonctionnement par l'analyse triadique*, Acta Oecologica. Oecolo. Gener., 8, 4, pp. 463-480.
- Thioulouse, J.; Chessel, D.; Dolédec, S.; Olivier, J-M. (1997), "ADE-4: a multivariate analysis and graphical display software", *Statistics and Computing*, 7, pp. 75-83.
- Tipping, M. E.; Bishop, Ch. M. (1999), "Probabilistic principal component analysis", *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, vol. 61, 3, pp. 611-622.
- Tucker, L. R. (1966), "Some mathematical notes on three-mode factor analysis", *Psychometrika*, 31, 3, pp. 279-311.
- Tucker, L. R. (1972), "Relations between multidimensional scaling and three-mode factor analysis", *Psychometrika*, 37, 1, pp. 3-27.
- UNESPA (1997), *Información Estadística del seguro privado*. Datos 1986-95. Madrid: Ed. Editorial Aseguradora.
- UNESPA (1998a), *Información Estadística del seguro privado*. Datos 1996. Madrid: Ed. Editorial Aseguradora.

-
- UNESPA (1998b), “Qué es el riesgo y qué es el seguro”, *Nociones de seguros: UNESPA*, 1998 (ref. de 29 de diciembre de 1998). Accesible a través de World Wide Web: <http://www.unespa.es/segesp/pagina_1.htm.
- UNESPA (1998c), “La póliza”, *Nociones de seguros: UNESPA*, 1998 (ref. de 29 de diciembre de 1998). Accesible a través de World Wide Web: <http://www.unespa.es/segesp/pagina_5.htm.
- UNESPA (1998d), “Las bases técnicas del seguro”, *Nociones de seguros: UNESPA*, 1998 (ref. de 29 de diciembre de 1998). Accesible a través de World Wide Web: <http://www.unespa.es/segesp/pagina_6.htm.
- UNESPA (1998e), “El siniestro”, *Nociones de seguros: UNESPA*, 1998 (ref. de 29 de diciembre de 1998). Accesible a través de World Wide Web: <http://www.unespa.es/segesp/pagina_9.htm.
- UNESPA (1998f), “Las garantías financieras del seguro”, *Nociones de seguros: UNESPA*, 1998 (ref. de 29 de diciembre de 1998). Accesible a través de World Wide Web: <http://www.unespa.es/segesp/pagina_11.htm.
- UNESPA (1998g), “Clasificación de los ramos del seguro”, *Nociones de seguros: UNESPA*, 1998 (ref. de 29 de diciembre de 1998). Accesible a través de World Wide Web: <http://www.unespa.es/segesp/pagina_12.htm.
- UNESPA (1998h), “Los seguros de vida”, *Nociones de seguros: UNESPA*, 1998 (ref. de 29 de diciembre de 1998). Accesible a través de World Wide Web: <http://www.unespa.es/segesp/pagina_13.htm.
- UNESPA (1998i), “Los seguros de accidentes”, *Nociones de seguros: UNESPA*, 1998 (ref. de 29 de diciembre de 1998). Accesible a través de World Wide Web: <http://www.unespa.es/segesp/pagina_15.htm.
- UNESPA (1998j), “Los seguros de enfermedad”, *Nociones de seguros: UNESPA*, 1998 (ref. de 29 de diciembre de 1998). Accesible a través de World Wide Web: <http://www.unespa.es/segesp/pagina_17.htm.
- UNESPA (1999a), *Información Estadística del seguro privado*. Cifras provisionales 1997. Madrid: Ed. Editorial Aseguradora.
- UNESPA, (1999b) “El seguro en 1998”, *Memoria Anual de UNESPA 1998* (en línea): UNESPA, 1999 (ref. de 14 de octubre de 1999. Accesible a través de World Wide Web: <<http://www.unespa.es/memoria/home.htm>.
- UNESPA, (1999c) “El seguro en Europa”, *Memoria Anual de UNESPA 1998* (en línea): UNESPA, 1999 (ref. de 14 de octubre de 1999.

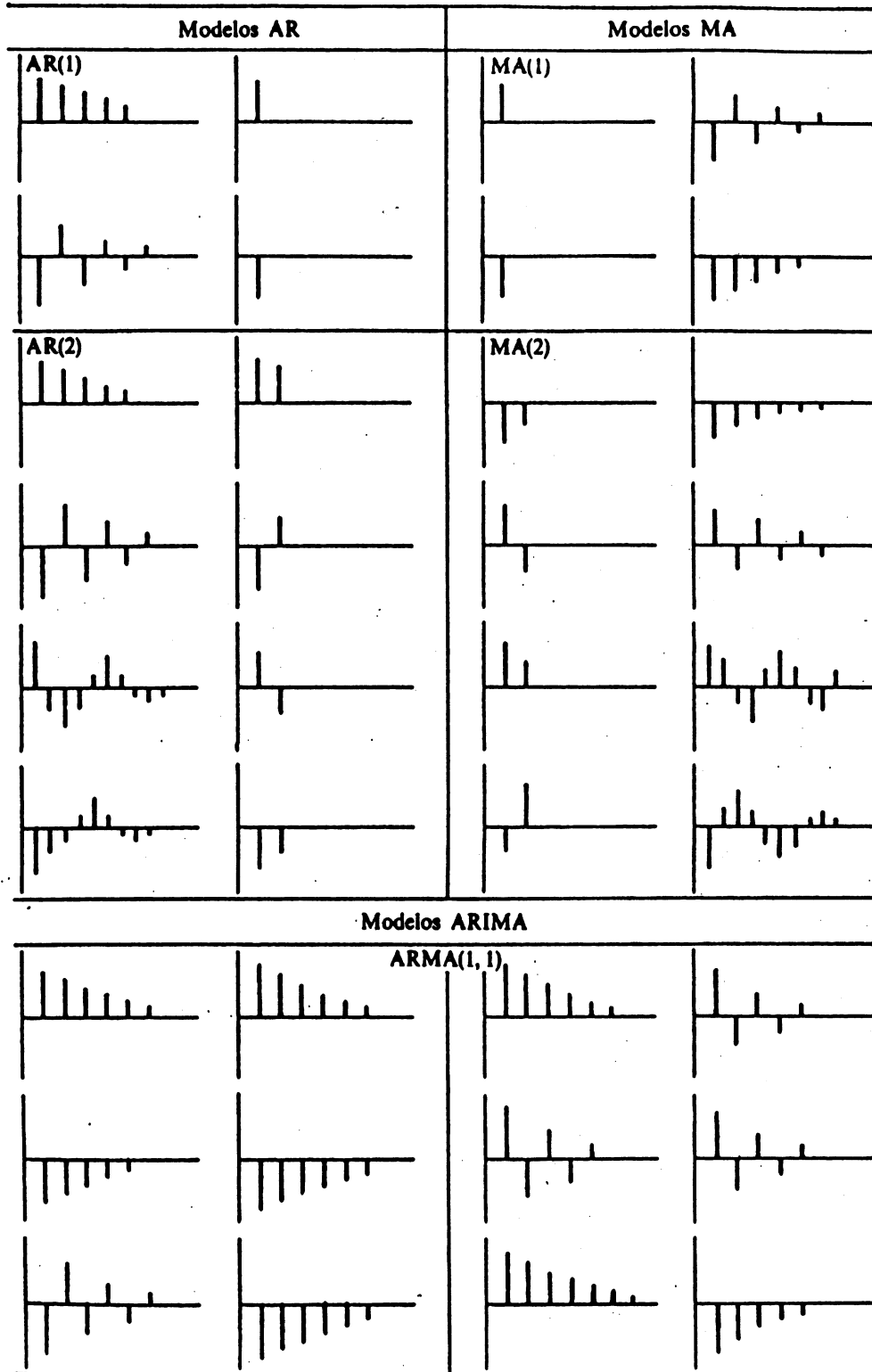
-
- Accesible a través de World Wide Web:
<<http://www.unespa.es/memoria/home.htm>.
- UNESPA, (1999d) “Legislación Comunitaria”, *Memoria Anual de UNESPA 1998* (en línea): UNESPA, 1999 (ref. de 14 de octubre de 1999). Accesible a través de World Wide Web:
<<http://www.unespa.es/memoria/home.htm>.
- UNESPA, (1999e) “La normativa del Euro”, *Memoria Anual de UNESPA 1998* (en línea): UNESPA, 1999 (ref. de 14 de octubre de 1999). Accesible a través de World Wide Web:
<<http://www.unespa.es/memoria/home.htm>.
- UNESPA, (1999f) “Las nuevas reglas del juego”, *Memoria Anual de UNESPA 1998* (en línea): UNESPA, 1999 (ref. de 14 de octubre de 1999). Accesible a través de World Wide Web:
<<http://www.unespa.es/memoria/home.htm>.
- UNESPA, (1999g) “Seguro y previsión colectiva”, *Memoria Anual de UNESPA 1998* (en línea): UNESPA, 1999 (ref. de 14 de octubre de 1999). Accesible a través de World Wide Web:
<<http://www.unespa.es/memoria/home.htm>.
- UNESPA, (1999h) “El ramo del automovil”, *Memoria Anual de UNESPA 1998* (en línea): UNESPA, 1999 (ref. de 14 de octubre de 1999). Accesible a través de World Wide Web:
<<http://www.unespa.es/memoria/home.htm>.
- UNESPA, (1999i) “Asegurar el efecto 2000”, *Memoria Anual de UNESPA 1998* (en línea): UNESPA, 1999 (ref. de 14 de octubre de 1999). Accesible a través de World Wide Web:
<<http://www.unespa.es/memoria/home.htm>.
- UNESPA, (1999j) “Nuevos retos”, *Memoria Anual de UNESPA 1998* (en línea): UNESPA, 1999 (ref. de 14 de octubre de 1999). Accesible a través de World Wide Web: <<http://www.unespa.es/memoria/home.htm>.
- Van den Wollenberg, A. L. (1977), “Redundancy analysis. An alternative for canonical correlation analysis”, *Psychometrika*, 42, pp. 207-219.
- Ventura Victoria, J. (1994), *Análisis competitivo de la empresa: un enfoque estratégico*. Madrid: Ed. Civitas.
- Ventura Victoria, J. (1996), *Análisis dinámico de la estrategia empresarial: un ensayo interdisciplinar*. Oviedo: Servicio de Publicaciones de la Universidad de Oviedo.

-
- Verneaux, J.; Chessel, D. (1994) “Approche de la stabilité des structures mésologiques et biologiques des cours d’eau du bassin du Doubs para la méthode ACT-STATIS”, *Symposium Relations entre le Poisson et son Habitat: théories, méthodes et applications*. Lyon.
- Villanueva, E. (1999), “Estamos haciendo un gran esfuerzo para devolver la calidad de servicio a los mediadores”, *Aseguradores*, 308, pp. 24-25.
- Wei, W.W.S. (1990), *Time series analysis. Univariate and Multivariate Methods*. Redwood City, California: Ed. Addison-Wesley Publishing Company.
- Williamson, O. E. (1991), *Mercados y jerarquías: su análisis y sus implicaciones antitrust*. México: Ed. Fondo de Cultura Económica Contemporánea.
- Zurita, A. (1999), “Hacia una nueva confrontación del sector asegurador”. *Ausbanc*, 111, pp. 102-110.

ANEXO

A.1 CUADROS GENERALES

A.1.1 Funciones de autocorrelación y autocorrelación parcial teóricas



A.2 DATOS DEL SECTOR ASEGURADOR ANALIZADOS DESDE EL PUNTO DE VISTA DESCRIPTIVO¹

Tabla a
Individuos y variables para el año 1995 (Ranking, 1996: 85-113)

ETIQUETAS	VIDA95	AUTOS95	SALUD95	RESTO95	INDUS95	PRIMEM95	GINTPR95	GEXTPR95	MMMM95	INFINIT95	NNC95	NNCPR95	NNPR95
Adeslas	0	0	99,8	0,2	0,2	79,91	14,28	1,58	2,61	0,59	6,93	3,17	100
AegonUnionAsegurad	13,36	45,78	9,18	31,68	15,25	66,77	14,11	16,86	2,05	0,83	4,99	0,69	91,26
AGFUnionFenix	31,2	21,7	1,7	45,4	33,2	55,20	25,04	14,01	1,97	0,80	9,28	0,28	88,94
AllianzRas	53,4	36,3	0	10,3	6,5	139,20	7,79	7,87	2,50	0,92	15,28	0,75	94,83
AntaresVida	100	0	0	0	0	1390,38	0,78	3,37	2,12	0,98	8,73	0,82	99,93
ArgentariaVida	100	0	0	0	0	431,52	3,24	3,84	1,78	0,88	2,08	0,14	85,65
AscatVida	100	0	0	0	0	1243,28	1,4	0,77	1,08	0,99	9,76	0,37	99
AseguradValenciana	100	0	0	0	0	181,71	7,11	2,27	4,35	0,97	3,65	0,54	96,25
Asisa	0	0	99,79	0,21	0	118,24	7,9	0,11	0,89	0,37	6,89	2,51	100
AssicurazioniGenerali	21,21	35,6	2,47	40,72	27,5	58,50	16,84	13,54	6,86	0,59	2,62	0,39	90,03
Athena	9,12	68,02	1,73	21,13	6,24	46,63	18,25	20,74	2,66	0,94	3,05	0,54	94,03
Banesto	74	0	0	26	20	145,87	12,96	9,87	13,52	0,92	0,28	0,18	56,88
BankinterVida	100	0	0	0	0	9175,33	0,68	0	1,12	0,96	11,06	0,87	99,84
BansabadellVida	100	0	0	0	0	3051,21	0,51	0,99	1,33	0,58	9,69	0,44	99,82
BarclaysVida	100	0	0	0	0	715,66	3,13	3,03	1,23	0,99	6,68	0,30	99,19
BiharkoVida	100	0	0	0	0	1407,41	1,18	1,2	1,04	0,98	10,23	0,70	99,57
BilbaoCiaAnma	29,9	39,5	0	30,6	14	56,40	16,98	17,21	2,32	0,89	3,49	0,41	93,62
CajaMadridVida	100	0	0	0	0	129,15	14,54	4,08	1,87	1,00	1,98	0,35	94,89
Casergupo	57,57	9,15	0,11	33,17	16,2	73,98	13,74	11,41	2,08	0,83	2,87	0,24	92,59
CatalanaOccidente	32,4	32,4	0	35,2	12,9	64,85	17,11	13,9	5,41	0,78	2,33	0,33	97,65
CreditoyCaucion	0	0	0	100	100	66,00	14,08	5,65	7,02	0,81	0,52	0,19	22,89
DbVida	100	0	0	0	0	1215,34	1,33	2,29	1,24	0,90	9,53	0,33	99,46
Euroseguros	97,26	0	0	2,74	0,13	1676,51	1,29	0,63	1,79	0,96	7,31	0,37	98,68
FiactMutua	3,41	51,59	25,08	19,92	12,94	51,73	12,56	14,53	2,15	0,46	8,07	0,93	89,46
Finisterre	3,69	0	1,97	94,34	5,09	202,70	11	38,35	2,19	0,29	4,10	1,42	92,24
Intercase	99,64	0	0	0,36	0	566,73	2,93	2,98	1,30	0,99	3,24	0,17	99,36
LaEstrella	46,81	21,33	3	28,86	14,87	69,33	14,18	11,56	2,03	0,94	6,83	0,25	91,57
LagunAroVida	100	0	0	0	0	3050,85	14,11	16,86	1,07	0,98	19,70	0,92	99,08
MapfreMutualidad	0	92,06	0	7,94	4,05	59,90	16	5,99	5,60	0,24	2,86	0,67	94,14
MapfreSegGrales	0	14,53	0	85,47	43,3	45,04	17,88	14,29	2,06	0,54	3,59	0,74	77,38
MapfreVida	94,3	0	5,2	0,5	0	68,74	12,3	5,9	7,15	0,82	4,15	0,23	98,65
MultinacionalAsegurad	0	86,01	0	13,99	5,09	54,86	14,38	12,97	0,33	0,72	-805,55	1,33	98,79
Musini	24,23	0	0	75,77	73,35	157,56	8,21	1,06	4,00	0,93	0,27	0,06	32,65
MutuaGeneralSegur	27,35	35,66	0,83	36,16	17,16	32,55	22,79	13,45	6,12	0,74	4,84	0,43	90,32
MutuaMadri	2,28	94,82	0	2,9	0	239,95	7,19	0	2,37	0,76	1,92	0,35	88,94
Ocaso	7,55	0	0,4	92,05	3,69	42,16	19,53	26,27	3,07	0,78	3,35	0,96	100,24
PelayoMutua	1,52	41,23	0	57,25	49,68	55,61	14,5	5,71	2,90	0,92	3,59	0,47	95,97
PlusUltra	34,3	28,4	0,4	36,9	25,2	79,28	13,49	11,84	3,48	0,90	4,20	0,48	81,09
Sanitas	0	0	99,67	0,33	0	81,52	15,49	6,29	4,45	0,81	4,43	2,39	100
StaLucia	14,92	0	0	85,08	1,26	258,62	7,44	35,51	2,86	0,83	2,62	0,75	99,48
Santander	96	0	0	4	1	1046,64	25,89	6,61	2,27	0,95	0,17	1,81	94,26
SegurosGenesis	100	0	0	0	0	100,78	9,95	1,11	1,55	0,95	5,04	0,40	99,47
Vidacaixa	100	0	0	0	0	1708,99	1,58	2,31	1,69	0,97	10,12	0,31	100
Vitalicio	46,92	25,3	1,42	26,36	19,07	106,46	9,83	7,59	3,26	0,84	4,53	0,28	89,82
WinterthurSegGrales	0	50,3	0,1	49,6	27,6	49,28	17,73	17,09	1,99	0,21	3,69	0,73	92,55
WinterthurVida	100	0	0	0	0	241,03	5,81	6,81	1,56	0,95	4,50	0,24	98,33
ZurichInternacional	0	21	0	79	65,4	54,38	19,37	14,88	8,60	0,72	2,42	0,55	70,25
ZurichSegEspana	0	65,6	0,1	34,3	15,3	115,71	9,95	1,11	3,29	0,74	16,65	0,97	94,77

¹ En primer lugar se recogen las medias y varianzas de las variables analizadas para los tres años considerados. Los epígrafes siguientes hasta A.2.1.6, excepto el A.2.1.4, recogen los resultados obtenidos al aplicar el método *Statis* mediante el módulo SPAD-TM (versión 3.5). Dicho epígrafe A.2.1.4 ha sido calculado mediante la hoja de cálculo EXCEL. El epígrafe A.2.1.7 contiene las coordenadas de las variables obtenidas al aplicar un ACP *centrado* a las nubes de individuos, ponderadas por los coeficientes correspondientes, mediante el módulo de Análisis Factorial del SPAD.

Tabla b
Individuos y variables para el año 1996 (Baena y Lozano, 1997: 116-139)

ETIQUETAS	VIDA96	AUTOS96	SALUD96	RESTO96	INDUS96	PRIMEM96	GINTPR96	GEXTPR96	MMMMI96	INFINIT96	NNC96	NNCPR96	NNPR96
Adeslas	0	0	100	0	0	86,33	11,93	1,71	2,89	0,62	6,27	2,95	100
AegonUnionAsegurad	15,04	40,35	14,71	29,9	14,38	77,02	15,25	15,76	3,55	0,55	3,16	0,64	92,33
AGFUnionFenix	32,33	22,75	1,13	43,79	32,41	64,72	23,6	13,63	3,23	0,84	4,01	0,28	87,4
AllianzRas	25,7	55,48	0	18,82	11,52	91,89	12,36	12,97	4,42	0,91	7,35	0,44	90,18
AntaresVida	100	0	0	0	0	1633,86	0,56	2,62	3,08	0,99	10,63	0,76	100,04
ArgentariaVida	100	0	0	0	0	2268,44	1,06	0,82	1,52	0,84	10,97	0,48	97,31
AscatVida	100	0	0	0	0	2231,45	1,19	1,06	1,27	1,00	9,39	0,32	94,15
AseguradValenciana	100	0	0	0	0	402,13	3,76	1,59	2,97	0,97	7,47	0,62	97,86
Asisa	0	0	99,81	0,19	0	122,28	6,85	0,11	0,94	0,40	6,59	2,48	100
AssicurazioniGenerali	34,5	28,14	2,09	35,27	23,24	83,36	13,01	12,99	6,29	0,62	4,90	0,40	79,71
Athena	9,76	66,03	1,61	22,6	7,06	54,18	16,86	18,4	3,26	0,94	2,97	0,57	94,62
Banesto	98,9	0	0	1,1	0,2	3825,67	0,47	0,53	2,59	0,98	10,14	0,88	99,86
BankinterVida	100	0	0	0	0	4888,25	1,08	0	1,06	0,95	8,87	0,36	99,63
BansabadellVida	100	0	0	0	0	2763,07	0,86	0,42	1,25	0,49	7,42	0,32	99,72
BarclaysVida	100	0	0	0	0	924,41	3,22	2,07	1,84	0,99	8,47	0,31	99,67
BiharkoVida	100	0	0	0	0	1272,00	1,46	1,68	1,31	0,98	6,85	0,38	99,37
BilbaoCiaAnma	30,7	31,6	1,7	36	11,9	63,65	15,69	20,45	2,31	0,87	3,35	0,42	94,54
CajaMadridVida	100	0	0	0	0	286,57	7,12	2,07	1,11	0,99	4,80	0,51	97,86
Casergupo	70,1	5,81	0,47	23,62	11,84	119,50	8,54	8,24	2,22	0,77	4,59	0,32	93,97
CatalanaOccidente	35	31,1	1,3	32,6	17,3	70,61	17,35	13,08	6,21	0,78	1,85	0,30	94,86
CreditoyCaucion	0	0	0	100	100	70,27	13,43	5,69	8,46	0,82	0,44	0,15	23,01
DbVida	100	0	0	0	0	487,97	4,83	4,39	2,47	0,91	2,78	0,12	99,36
Euroseguros	96,71	0	0,02	3,27	0,12	1820,68	1,29	0,77	2,21	0,96	7,81	0,36	98,66
FiactMutua	3,73	48,32	23,54	24,41	13,71	56,09	11,96	14,35	2,81	0,47	6,05	0,87	85,53
Finisterre	3,66	0	2,08	94,26	4,66	194,23	12,13	37,53	2,09	0,53	4,53	1,46	93,87
Intercase	99,99	0	0	0,01	0	533,33	2,98	4,25	1,25	0,99	3,14	0,15	99,44
LaEstrella	57,84	17,07	2,45	22,64	11,42	96,63	11,08	9,44	1,84	0,96	9,36	0,29	94,6
LagunAroVida	100	0	0	0	0	1966,00	0,8	3,12	2,13	1,00	7,30	0,29	98,28
MapfreMutualidad	0	87,59	0	12,41	4,27	54,86	18,36	5,76	7,03	0,17	2,06	0,61	94,64
MapfreSegGrales	0	19,47	0	80,53	40,19	48,23	17,89	13,23	2,77	0,54	3,88	0,71	80,31
MapfreVida	94,2	0	5,2	0,6	0	73,49	11,88	5,79	7,66	0,84	3,98	0,22	98,78
MultinacionalAsegurad	0	75,15	0	24,85	9,19	53,13	16,66	13,94	1,93	0,67	2,79	0,82	92,69
Musini	25,31	0	0	73,69	73,69	145,79	8,36	1,74	8,75	0,93	0,19	0,05	28,1
MutuaGeneralSegur	30,32	32,93	0,93	35,82	17,02	34,97	22,69	13,35	6,17	0,74	3,20	0,40	90,72
MutuaMadri	2,28	93,82	0	3,9	0	254,76	6,81	0	2,67	0,74	1,78	0,33	99,83
Ocaso	8,06	0	0,42	91,52	3,61	44,31	18,87	25,87	3,87	0,82	2,82	0,87	99,2
PelayoMutua	1,4	84,26	0	14,34	3,34	50,46	16,01	5,47	4,91	0,92	2,38	0,45	97,58
PlusUltra	31,4	31,6	0,4	36,6	25,1	88,58	12,98	12,44	4,21	0,88	3,43	0,44	80,37
Sanitas	0	0	99,7	0,3	0	86,68	17,11	3,36	1,88	0,81	3,39	1,98	100
StaLucia	13,85	0	0	86,15	1,27	267,46	7,41	35,34	3,38	0,85	2,06	0,67	99,47
Santander	96,33	0	0	4,67	0,81	590,50	7,62	0,1	2,52	0,94	1,52	0,15	94,27
SegurosGenesis	100	0	0	0	0	111,93	9,58	1,29	5,50	0,97	4,86	0,39	99,41
Vidacaixa	100	0	0	0	0	2230,63	1,57	2,86	2,21	0,97	11,97	0,29	100,06
Vitalicio	53,36	21,6	1,26	23,78	17,26	118,99	8,45	6,96	5,17	0,77	5,26	0,29	89,96
WinterthurSegGrales	0	55,01	0,05	44,94	25,25	96,75	18,64	17,6	1,73	0,42	4,51	0,74	93,73
WinterthurVida	100	0	0	0	0	276,00	7,74	6,39	2,54	0,92	3,21	0,22	98,47
ZurichInternacional	0	14,3	0	85,7	74,9	64,69	16,29	15,05	9,12	0,65	1,83	0,48	69,86
ZurichSegEspana	0	66	0,1	33,9	14,3	57,97	19,31	16,21	3,26	0,74	9,21	0,93	96,85

Tabla c
Individuos y variables para el año 1997 (Baena y Lozano, 1998: 75-98)

ETIQUETAS	VIDA97	AUTOS97	SALUD97	RESTO97	INDUS97	PRIMEM97	GINTPR97	GEXTPR97	MMMMI97	INFINIT97	NNC97	NNCPR97	NNPR97
Adeslas	0	0,00	100,00	0,00	0	99,15	10,57	1,84	1,16	0,49	5,68	2,98	100,00
AegonUnionAsegurad	19,31	42,33	12,78	25,58	14,16	81,00	13,85	15,06	3,29	0,82	3,98	0,64	95,19
AGFUnionFenix	29	26,37	0,00	44,63	31	62,81	24,58	14,32	4,26	0,81	3,41	0,25	86,23
AllianzRas	30,31	52,60	0,00	17,09	10,73	102,47	10,78	12,05	4,95	0,92	7,35	0,42	90,55
AntaresVida	100	0,00	0,00	0,00	0	4805,29	0,23	1,06	3,01	1,00	24,54	0,86	99,93
ArgentariaVida	100	0,00	0,00	0,00	0	1578,92	1,24	1,71	1,56	0,66	7,07	0,28	96,23
AscatVida	100	0,00	0,00	0,00	0	2740,38	0,78	0,84	1,14	0,98	14,46	0,39	99,25
AseguradValenciana	100	0,00	0,00	0,00	0	880,14	3,17	1,47	2,13	0,98	12,24	0,62	98,72
Asisa	0	0,00	99,75	0,25	0	137,60	6,66	0,1	1,16	0,48	4,46	2,07	100,00
AssicurazioniGenerali	33,16	28,77	2,08	35,99	24,17	76,47	13,48	12,04	4,84	0,79	6,81	0,37	88,92
Athena	11,61	65,07	1,44	21,88	7,41	57,16	15,77	14,85	3,26	0,95	2,82	0,54	94,57
Banesto	96,78	0,00	0,00	3,22	0,7	2272,67	0,83	0,99	2,20	0,98	6,09	0,39	97,18
BankinterVida	100	0,00	0,00	0,00	0	6593,25	1,01	0	1,04	0,97	7,36	0,30	99,27
BansabadellVida	100	0,00	0,00	0,00	0	1397,13	2,62	2,63	1,35	0,54	4,10	0,19	99,46
BarclaysVida	100	0,00	0,00	0,00	0	5331,74	1,37	0,91	1,12	0,99	19,57	0,45	99,83
BiharkoVida	100	0,00	0,00	0,00	0	1754,93	1,28	1,14	1,12	0,98	11,66	0,37	108,43
BilbaoCiaAnma	33,6	28,04	1,70	36,66	19,3	60,88	16,64	19,73	2,42	0,87	3,09	0,38	94,78
CajaMadridVida	100	0,00	0,00	0,00	0	1374,07	1,74	1,16	1,18	0,97	9,87	0,70	99,45
Casagrupo	74,48	4,35	0,20	20,97	7,26	155,28	7,37	6,95	1,81	0,68	6,24	0,35	96,06
CatalanaOccidente	38,1	29,00	0,00	32,90	17,6	79,11	16,39	12,94	6,50	0,80	1,87	0,30	94,57
CreditoYCaucion	0	0,00	0,00	100,00	100	72,68	16,17	6,93	10,13	0,82	0,39	0,14	22,95
DbVida	100	0,00	0,00	0,00	0	1198,63	1,38	1,96	3,17	0,93	8,77	0,31	99,73
Euroseguros	96,77	0,00	0,01	3,22	0,09	2172,40	1,1	0,8	1,53	0,81	9,70	0,38	99,07
FiactMutua	4,69	44,66	23,98	26,67	16,09	54,59	12,99	13,98	2,93	0,47	5,37	0,83	87,16
Finisterre	3,44	0,00	1,86	94,70	4,31	199,36	11,7	37,38	1,94	0,48	4,75	1,35	93,36
Intercase	100	0,00	0,00	0,00	0	552,78	3,11	3,52	1,31	0,98	2,97	0,15	99,26
LaEstrella	46	25,64	3,40	24,96	14,3	83,28	13,29	11,46	3,12	0,94	6,89	0,22	93,67
LagunAroVida	100	0,00	0,00	0,00	0	1519,10	1,38	4,67	1,10	1,00	3,96	0,22	97,73
MapfreMutualidad	0	87,23	0,00	12,77	4,51	51,67	20,88	5,53	8,03	0,17	1,69	0,57	95,22
MapfreSegGrales	0	18,59	0,00	81,41	40,22	49,12	18,22	13,3	2,80	0,61	4,03	0,65	81,46
MapfreVida	92,5	0,00	6,90	0,60	0	76,31	12,28	5,36	7,72	0,85	3,69	0,21	98,92
MultinacionalAsegurad	0	80,92	0,00	19,08	10,99	45,71	21,89	14,42	1,51	0,69	2,18	0,76	93,75
Musini	23	0,00	0,00	77,00	63,9	148,03	8,57	1,94	8,20	0,94	0,24	0,04	24,12
MutuaGeneralSegur	30,85	30,97	1,11	37,07	17,14	35,99	23,16	12,95	6,38	0,77	3,07	0,37	90,42
MutuaMadri	3,08	92,33	0,00	4,59	0	257,87	7,25	0	3,14	0,74	1,54	0,31	99,84
Ocaso	9,93	0,00	0,41	89,66	3,47	44,46	19,46	25,08	4,34	0,79	2,70	0,80	99,31
PelayoMutua	1,35	97,47	0,00	1,18	1	47,18	16,61	5,3	6,42	0,87	1,85	0,41	95,99
PlusUltra	38,9	28,36	0,50	32,24	21,5	94,87	12,86	12,05	4,18	0,88	3,45	0,45	83,53
Sanitas	0	0,00	99,80	0,20	0	97,81	17,53	10,12	2,15	0,72	3,83	2,25	100,00
StaLucia	14,17	0,00	0,00	85,83	1,2	276,07	7,35	35,35	3,52	0,87	1,84	0,62	99,40
Santander	99,01	0,00	0,00	0,99	0,19	2860,33	2,72	0,07	1,45	0,84	9,00	0,52	98,95
SegurosGenesis	100	0,00	0,00	0,00	0	124,51	10,7	3,55	1,72	0,97	5,37	0,33	99,31
Vidacaixa	100	0,00	0,00	0,00	0	2,08	1,93	3,75	2,09	0,98	10,84	0,22	100,12
Vitalicio	37,44	28,46	1,82	32,28	23,2	94,12	12,26	9,38	5,71	0,92	3,61	0,20	86,31
WinterthurSegGrales	0	53,18	0,00	46,82	24,65	89,59	17	17,6	1,60	0,37	4,41	0,73	93,62
WinterthurVida	100	0,00	0,00	0,00	0	186,98	12,27	9,07	2,36	0,90	2,75	0,16	98,09
ZurichInternacional	0	11,75	0,00	88,25	77,4	66,77	16,85	15,27	7,72	0,70	1,97	0,51	73,45
ZurichSegEspana	0	66,12	0,10	33,78	13,7	63,02	18,45	16,63	6,95	0,76	11,44	0,97	96,28

A.2.1 Resultados obtenidos de la aplicación del método *Statis*

A.2.1.1 Medias y varianzas de las variables analizadas en los tres años

VARIABLES	MEDIA	VARIANZA
VIDA95	48.382	1825.51
AUTOS95	19.089	726.36
SALUD95	7.353	583.85
RESTO95	25.176	927.14
INDUS95	13.254	448.51
PRIMEM95	627.696	2065515.10
GEXTPR95	9.089	74.46
GINTPR95	11.216	45.36
MMMMI95	3.046	5.75
INFINIT95	0.801	0.04
NNC95	-11.280	13439.76
NNPCRO95	0.683	0.40
NNPR95	90.913	241.24
VIDA96	50.455	1867.10
AUTOS96	19.341	766.29
SALUD96	7.479	584.19
RESTO96	22.725	844.83
INDUS96	10.790	358.23
PRIMEM96	652.162	1140196.84
GEXTPR96	8.594	76.28
GINTPR96	10.103	44.18
MMMMI96	3.414	4.55
INFINIT96	0.800	0.04
NNC96	5.121	8.72
NNPCRO96	0.593	0.31
NNPR96	91.857	232.20
VIDA97	49.322	1847.28
AUTOS97	19.629	803.33
SALUD97	7.455	584.53
RESTO97	23.593	922.34
INDUS97	11.879	422.84
PRIMEM97	837.620	2115631.34
GEXTPR97	8.734	73.02
GINTPR97	10.203	51.44
MMMMI97	3.390	5.49
INFINIT97	0.801	0.04
NNC97	5.936	22.30
NNPCRO97	0.574	0.30
NNPR97	92.492	241.83

A.2.1.2 Comparación de las nubes de individuos

NUBE 1- AÑO 1995
 REPARTO DE LA INERCIA SEGUN LOS EJES PRINCIPALES

Norma de $W_i = 257.50$

Traza de la matriz = 2.3728

Histograma de los 47 primeros valores propios

NUMERO	VALOR PROPIO	PERCENTAJE	PERCENTAJE ACUMUL.
1	0.7496	31.59	31.59
2	0.4805	20.25	51.84
3	0.2910	12.26	64.11
4	0.2139	9.02	73.12
5	0.1709	7.20	80.32
6	0.1442	6.08	86.40
7	0.1161	4.89	91.29
8	0.0932	3.93	95.22
9	0.0582	2.45	97.67
10	0.0270	1.14	98.81
11	0.0198	0.83	99.64
12	0.0085	0.36	100.00
13	0.0000	0.00	100.00
14	0.0000	0.00	100.00
15	0.0000	0.00	100.00
16	0.0000	0.00	100.00
17	0.0000	0.00	100.00
18	0.0000	0.00	100.00
19	0.0000	0.00	100.00
20	0.0000	0.00	100.00
21	0.0000	0.00	100.00
22	0.0000	0.00	100.00
23	0.0000	0.00	100.00
24	0.0000	0.00	100.00
25	0.0000	0.00	100.00
26	0.0000	0.00	100.00
27	0.0000	0.00	100.00
28	0.0000	0.00	100.00
29	0.0000	0.00	100.00
30	0.0000	0.00	100.00
31	0.0000	0.00	100.00
32	0.0000	0.00	100.00
33	0.0000	0.00	100.00
34	0.0000	0.00	100.00
35	0.0000	0.00	100.00
36	0.0000	0.00	100.00
37	0.0000	0.00	100.00
38	0.0000	0.00	100.00
39	0.0000	0.00	100.00
40	0.0000	0.00	100.00
41	0.0000	0.00	100.00
42	0.0000	0.00	100.00
43	0.0000	0.00	100.00
44	0.0000	0.00	100.00
45	0.0000	0.00	100.00
46	0.0000	0.00	100.00
47	0.0000	0.00	100.00

NUBE 2- AÑO 1996
 REPARTO DE LA INERCIA SEGUN LOS EJES PRINCIPALES

Norma de $W_2 = 293.42$

Traza de la matriz = 2.0823

Histograma de los 47 primeros valores propios

NUMERO	VALOR PROPIO	PERCENTAJE PROPIO	PERCENTAJE ACUMUL.
1	0.8334	40.02	40.02
2	0.4263	20.47	60.50
3	0.2368	11.37	71.87
4	0.1868	8.97	80.84
5	0.1375	6.60	87.44
6	0.0754	3.62	91.07
7	0.0621	2.98	94.05
8	0.0469	2.25	96.30
9	0.0377	1.81	98.11
10	0.0223	1.07	99.18
11	0.0106	0.51	99.69
12	0.0065	0.31	100.00
13	0.0000	0.00	100.00
14	0.0000	0.00	100.00
15	0.0000	0.00	100.00
16	0.0000	0.00	100.00
17	0.0000	0.00	100.00
18	0.0000	0.00	100.00
19	0.0000	0.00	100.00
20	0.0000	0.00	100.00
21	0.0000	0.00	100.00
22	0.0000	0.00	100.00
23	0.0000	0.00	100.00
24	0.0000	0.00	100.00
25	0.0000	0.00	100.00
26	0.0000	0.00	100.00
27	0.0000	0.00	100.00
28	0.0000	0.00	100.00
29	0.0000	0.00	100.00
30	0.0000	0.00	100.00
31	0.0000	0.00	100.00
32	0.0000	0.00	100.00
33	0.0000	0.00	100.00
34	0.0000	0.00	100.00
35	0.0000	0.00	100.00
36	0.0000	0.00	100.00
37	0.0000	0.00	100.00
38	0.0000	0.00	100.00
39	0.0000	0.00	100.00
40	0.0000	0.00	100.00
41	0.0000	0.00	100.00
42	0.0000	0.00	100.00
43	0.0000	0.00	100.00
44	0.0000	0.00	100.00
45	0.0000	0.00	100.00
46	0.0000	0.00	100.00
47	0.0000	0.00	100.00

NUBE 3- AÑO 1997
 REPARTO DE LA INERCIA SEGUN LOS EJES PRINCIPALES

Norma de $W_3 = 296.90$

Traza de la matriz = 2.0579

Histograma de los 47 primeros valores propios

NUMERO	VALOR PROPIO	PORCENTAJE	PORCENTAJE ACUMUL.
1	0.8351	40.58	40.58
2	0.4289	20.84	61.42
3	0.2396	11.64	73.06
4	0.1756	8.53	81.60
5	0.1253	6.09	87.68
6	0.0908	4.41	92.10
7	0.0559	2.72	94.81
8	0.0434	2.11	96.92
9	0.0348	1.69	98.61
10	0.0128	0.62	99.24
11	0.0092	0.45	99.68
12	0.0065	0.32	100.00
13	0.0000	0.00	100.00
14	0.0000	0.00	100.00
15	0.0000	0.00	100.00
16	0.0000	0.00	100.00
17	0.0000	0.00	100.00
18	0.0000	0.00	100.00
19	0.0000	0.00	100.00
20	0.0000	0.00	100.00
21	0.0000	0.00	100.00
22	0.0000	0.00	100.00
23	0.0000	0.00	100.00
24	0.0000	0.00	100.00
25	0.0000	0.00	100.00
26	0.0000	0.00	100.00
27	0.0000	0.00	100.00
28	0.0000	0.00	100.00
29	0.0000	0.00	100.00
30	0.0000	0.00	100.00
31	0.0000	0.00	100.00
32	0.0000	0.00	100.00
33	0.0000	0.00	100.00
34	0.0000	0.00	100.00
35	0.0000	0.00	100.00
36	0.0000	0.00	100.00
37	0.0000	0.00	100.00
38	0.0000	0.00	100.00
39	0.0000	0.00	100.00
40	0.0000	0.00	100.00
41	0.0000	0.00	100.00
42	0.0000	0.00	100.00
43	0.0000	0.00	100.00
44	0.0000	0.00	100.00
45	0.0000	0.00	100.00
46	0.0000	0.00	100.00
47	0.0000	0.00	100.00

A.2.1.3 Situación de las nubes de individuos unas respecto de otras

PRODUCTOS ESCALARES ENTRE NUBES NORMADAS Coeficiente RV entre matrices

Matr	95	96	97
95	1.000		
96	0.832	1.000	
97	0.834	0.927	1.000

TEST DE PERMUTACIONES

* indica que las nubes son similares

Matr	95	96	97
95	*		
96	*	*	
97	*	*	*

INTERPRETACION DE LAS DISTANCIAS ENTRE NUBES

DISTANCIA EUCLIDEA ENTRE NUBES

Matr	95	96	97
95	0.000		
96	0.580	0.000	
97	0.576	0.383	0.000
	95	96	97

DESCOMPOSICION DE LAS DISTANCIAS AL CUADRADO SEGUN LOS INDIVIDUOS

- ENTRE LA NUBE 95 Y LAS OTRAS

LINEA	MEDIA	NUBE	
		96	97
Euroseguros	1.1	1.1	1.2
MapfreMutualidad	1.7	1.6	1.8
Vidacaixa	1.5	2.0	1.1
Casergupo	0.5	0.3	0.8
Vitalicio	0.6	0.2	1.0
BarclaysVida	3.2	0.9	5.6
WinterthurSegGrales	0.7	0.7	0.6
AllianzRas	1.4	1.4	1.4
MapfreVida	0.9	1.0	0.8
AntaresVida	3.5	1.4	5.7
Stalucia	0.5	0.5	0.5
MutuaMadri	0.7	0.6	0.7
AGFUnionFenix	0.8	0.6	1.0
CatalaOccident	1.0	1.1	0.9
Asisa	1.0	1.0	1.0
AscatVida	1.9	1.8	2.0
MapfreSegGrales	0.9	1.0	0.8
LaEstrella	0.5	0.9	0.1
Intercase	0.9	0.8	1.0
AegonUnionAsegurad	0.5	0.8	0.2
CajaMadridVida	1.0	0.4	1.6
Ocaso	0.9	0.8	1.0
PlusUltra	0.4	0.5	0.3
Adeslas	0.9	1.1	0.7
Sanitas	0.7	0.6	0.7
DbVida	1.2	1.5	1.0
Santander	3.6	2.6	4.6
AssicurazioniGenerali	1.3	1.0	1.5
ArgentariaVida	1.9	2.8	1.0
Banesto	19.5	23.1	15.9
ZurichSegEspana	1.3	1.2	1.4
SegurosGenesis	0.7	0.9	0.5
BiharkoVida	1.2	1.1	1.4
FiactMutua	0.2	0.2	0.3
BilbaoCiaAnma	0.4	0.3	0.4
AseguradValenciana	0.6	0.4	0.9
MultinacionalAsegura	11.8	12.0	11.6
WinterthurVida	1.0	0.7	1.3
CreditoyCaucion	6.1	6.8	5.4
Athena	0.5	0.4	0.5
BankinterVida	5.7	5.5	5.9
MutuaGeneralSegur	1.2	1.2	1.3
PelayoMutua	2.2	1.9	2.5
Musini	3.4	4.0	2.8
LagunAroVida	2.0	2.6	1.4
BansabadellVida	1.7	1.6	1.9
Finisterre	0.9	0.9	0.9
ZurichInternacional	3.7	4.2	3.2

- ENTRE LA NUBE 96 Y LAS OTRAS

LINEA	MEDIA	NUBE	
		95	97
Euroseguros	1.2	1.1	1.3
MapfreMutualidad	1.5	1.6	1.5
Vidacaixa	3.3	2.0	4.6
Casergrupo	0.4	0.3	0.5
Vitalicio	0.8	0.2	1.5
BarclaysVida	5.1	0.9	9.4
WinterthurSegGrales	0.8	0.7	0.8
AllianzRas	0.8	1.4	0.2
MapfreVida	0.9	1.0	0.7
AntaresVida	5.4	1.4	9.4
StaLucia	0.6	0.5	0.7
MutuaMadri	0.5	0.6	0.4
AGFUnionFenix	0.8	0.6	0.9
CatalaOccident	1.1	1.1	1.0
Asisa	1.0	1.0	1.1
AscatVida	2.2	1.8	2.6
MapfreSegGrales	1.1	1.0	1.1
LaEstrella	1.5	0.9	2.2
Intercase	0.6	0.8	0.4
AegonUnionAsegurad	1.2	0.8	1.7
CajaMadridVida	1.1	0.4	1.7
Ocaso	0.8	0.8	0.8
PlusUltra	0.6	0.5	0.7
Adeslas	0.9	1.1	0.8
Sanitas	0.6	0.6	0.6
DbVida	1.5	1.5	1.5
Santander	4.0	2.6	5.5
AssicurazioniGenerali	1.1	1.0	1.2
ArgentariaVida	3.5	2.8	4.2
Banesto	14.6	23.1	6.1
ZurichSegEspana	1.2	1.2	1.2
SegurosGenesis	1.1	0.9	1.2
BiharkoVida	1.2	1.1	1.4
FiactMutua	0.3	0.2	0.4
BilbaoCiaAnma	0.5	0.3	0.6
AseguradValenciana	0.7	0.4	0.9
MultinacionalAsegura	6.3	12.0	0.7
WinterthurVida	0.8	0.7	0.8
CreditoCaucion	5.4	6.8	4.1
Athena	0.6	0.4	0.8
BankinterVida	5.4	5.5	5.3
MutuaGeneralSegur	1.0	1.2	0.8
PelayoMutua	1.4	1.9	1.0
Musini	5.8	4.0	7.5
LagunAroVida	2.1	2.6	1.5
BansabadellVida	2.4	1.6	3.2
Finisterre	0.8	0.9	0.7
ZurichInternacional	3.5	4.2	2.7

- ENTRE LA NUBE 97 Y LAS OTRAS

LINEA	MEDIA	NUBE	
		95	96
Euroseguros	1.2	1.2	1.3
MapfreMutualidad	1.7	1.8	1.5
Vidacaixa	2.8	1.1	4.6
Casergrupo	0.7	0.8	0.5
Vitalicio	1.3	1.0	1.5
BarclaysVida	7.5	5.6	9.4
WinterthurSegGrales	0.7	0.6	0.8
AllianzRas	0.8	1.4	0.2
MapfreVida	0.8	0.8	0.7
AntaresVida	7.5	5.7	9.4
StaLucia	0.6	0.5	0.7
MutuaMadri	0.6	0.7	0.4
AGFUnionFenix	1.0	1.0	0.9
CatalaOccident	0.9	0.9	1.0
Asisa	1.1	1.0	1.1
AscatVida	2.3	2.0	2.6
MapfreSegGrales	0.9	0.8	1.1
LaEstrella	1.1	0.1	2.2
Intercase	0.7	1.0	0.4
AegonUnionAsegurad	0.9	0.2	1.7
CajaMadridVida	1.7	1.6	1.7
Ocaso	0.9	1.0	0.8
PlusUltra	0.5	0.3	0.7
Adeslas	0.8	0.7	0.8
Sanitas	0.7	0.7	0.6
DbVida	1.2	1.0	1.5
Santander	5.1	4.6	5.5
AssicurazioniGenerali	1.4	1.5	1.2
ArgentariaVida	2.6	1.0	4.2
Banesto	11.0	15.9	6.1
ZurichSegEspana	1.3	1.4	1.2
SegurosGenesis	0.9	0.5	1.2
BiharkoVida	1.4	1.4	1.4
FiactMutua	0.3	0.3	0.4
BilbaoCiaAnma	0.5	0.4	0.6
AseguradValenciana	0.9	0.9	0.9
MultinacionalAsegura	6.1	11.6	0.7
WinterthurVida	1.1	1.3	0.8
CreditoCaucion	4.8	5.4	4.1
Athena	0.6	0.5	0.8
BankinterVida	5.6	5.9	5.3
MutuaGeneralSegur	1.0	1.3	0.8
PelayoMutua	1.7	2.5	1.0
Musini	5.2	2.8	7.5
LagunAroVida	1.4	1.4	1.5
BansabadellVida	2.5	1.9	3.2
Finisterre	0.8	0.9	0.7
ZurichInternacional	2.9	3.2	2.7

LA SUMA DE LAS DISTANCIAS AL CUADRADO SE DESCOMPONE SEGUN LAS LINEAS

Euroseguros	=	1.16	%
MapfreMutualidad	=	1.68	%
Vidacaixa	=	2.09	%
Casergrupo	=	0.54	%
Vitalicio	=	0.75	%
BarclaysVida	=	4.31	%
WinterthurSegGrales	=	0.69	%
AllianzRas	=	1.18	%
MapfreVida	=	0.89	%
AntaresVida	=	4.58	%
Stalucia	=	0.56	%
MutuaMadri	=	0.62	%
AGFUnionFenix	=	0.82	%
CatalaOccident	=	1.00	%
Asisa	=	1.02	%
AscatVida	=	2.02	%
MapfreSegGrales	=	0.94	%
LaEstrella	=	0.78	%
Intercase	=	0.82	%
AegonUnionAsegurad	=	0.70	%
CajaMadridVida	=	1.12	%
Ocaso	=	0.87	%
PlusUltra	=	0.45	%
Adeslas	=	0.88	%
Sanitas	=	0.65	%
DbVida	=	1.29	%
Santander	=	3.93	%
AssicurazioniGenerali	=	1.26	%
ArgentariaVida	=	2.32	%
Banesto	=	17.13	%
ZurichSegEspana	=	1.26	%
SegurosGenesis	=	0.82	%
BiharkoVida	=	1.26	%
FiactMutua	=	0.24	%
BilbaoCiaAnma	=	0.42	%
AseguradValenciana	=	0.69	%
MultinacionalAsegura	=	9.79	%
WinterthurVida	=	0.98	%
CreditoyCaucion	=	5.75	%
Athena	=	0.53	%
BankinterVida	=	5.64	%
MutuaGeneralSegur	=	1.17	%
PelayoMutua	=	1.95	%
Musini	=	4.17	%
LagunAroVida	=	1.90	%
BansabadellVida	=	1.99	%
Finisterre	=	0.90	%
ZurichInternacional	=	3.51	%

A.2.1.4 Estudio de la nube de nubes: interestructura no-centrada

DESCOMPOSICION DE LA INERCIA SEGUN LOS EJES PRINCIPALES

Traza de la matriz = 3
 Histograma de los 3 primeros valores propios

NUMERO	VALOR PROPIO	PORCENTAJE	ACUMUL.
95	2.7294	90.98	90.98
96	0.1975	6.58	97.56
97	0.0729	2.44	100.00

COORDENADAS, CONTRIBUCIONES Y COSENO CUADRADOS DE LAS NUBES SOBRE LOS EJES PRINCIPALES

NUBE	DISTO	COORDENADAS			CONTRIBUCIONES			COSENO CUADRADOS					
		1	2	3	1	2	3	1	2	3			
95	1	0.93	0.36	-0.00	0.00	0.32	0.68	0.00	0.00	0.86	0.13	0.00	0.00
96	1	0.96	-0.17	-0.19	0.00	0.34	0.16	0.49	0.00	0.93	0.03	0.04	0.00
97	1	0.97	-0.17	0.19	0.00	0.34	0.15	0.50	0.00	0.94	0.03	0.04	0.00

A.2.1.5 Estudio de la nube de nubes tomando como origen la nube comprometido

CONSTRUCCION DE LA NUBE COMPROMISO = MEDIA PONDERADA DE LAS NUBES CON LOS COEFICIENTES

NUBE 95 = 0.341
 NUBE 96 = 0.354
 NUBE 97 = 0.354

PRODUCTOS ESCALARES Y DISTANCIAS ENTRE LA NUBE COMPROMISO Y LAS DIFERENTES NUBES

NUBE 95 = PRODUCTO ESCALAR = 0.930; DISTANCIA = 0.374
 NUBE 96 = PRODUCTO ESCALAR = 0.965; DISTANCIA = 0.265
 NUBE 97 = PRODUCTO ESCALAR = 0.966; DISTANCIA = 0.261

DESCOMPOSICION DE LA INERCIA SEGUN LAS NUBES

NUBE 95 = 50.25 %
 NUBE 96 = 25.19 %
 NUBE 97 = 24.56 %

REPRESENTACION GRAFICA DE LA NUBE DE NUBES (INTERESTRUCTURA CENTRADA)

DESCOMPOSICION DE LA INERCIA SEGUN LOS EJES PRINCIPALES

Traza de la matriz = 0.2779

Histograma de los 3 primeros valores propios

NUMERO	VALOR PROPIO	PORCENTAJE	PORCENTAJE ACUMUL.
95	0.1983	71.35	71.35
96	0.0733	26.37	97.73
97	0.0063	2.27	100.00

COORDENADAS, CONTRIBUCIONES Y COSENO CUADRADOS DE LAS NUBES SOBRE LOS EJES PRINCIPALES (INTERESTRUCTURA CENTRADA)

NUBE	COORDENADAS			CONTRIBUCIONES			COSENO CUADRADOS		
	1	2	3	1	2	3	1	2	3
95	0.14	-0.37	0.00	0.04	0.00	0.00	0.69	0.00	0.31
96	0.07	0.18	-0.19	0.05	0.00	0.00	0.16	0.49	0.35

A.2.1.6 Estudio de la nube comprometido**DESCOMPOSICION DE LA INERCIA SEGUN LAS LINEAS**

Euroseguros	=	1.27 %
MapfreMutualidad	=	3.66 %
Vidacaixa	=	1.59 %
Casergrupo	=	0.27 %
Vitalicio	=	0.30 %
BarclaysVida	=	2.12 %
WinterthurSegGrales	=	2.04 %
AllianzRas	=	0.47 %
MapfreVida	=	1.16 %
AntaresVida	=	2.56 %
Stalucia	=	2.65 %
MutuaMadri	=	1.93 %
AGFUnionFenix	=	1.16 %
CatalaOccident	=	0.71 %
Asisa	=	5.33 %
AscatVida	=	1.75 %
MapfreSegGrales	=	1.84 %
LaEstrella	=	0.38 %
Intercase	=	1.19 %
AegonUnionAsegurad	=	0.57 %
CajaMadridVida	=	1.02 %
Ocaso	=	2.14 %
PlusUltra	=	0.38 %
Adeslas	=	6.11 %
Sanitas	=	4.34 %
DbVida	=	1.06 %
Santander	=	1.40 %
AssicurazioniGenerali	=	0.74 %
ArgentariaVida	=	1.37 %
Banesto	=	2.85 %
ZurichSegEspana	=	1.41 %
SegurosGenesis	=	0.89 %
BiharkoVida	=	1.41 %
FiactMutua	=	1.10 %
BilbaoCiaAnma	=	0.62 %
AseguradValenciana	=	1.02 %
MultinacionalAsegura	=	4.35 %
WinterthurVida	=	0.81 %
CreditoyCaucion	=	8.86 %
Athena	=	1.15 %
BankinterVida	=	5.02 %
MutuaGeneralSegur	=	1.07 %
PelayoMutua	=	1.64 %
Musini	=	5.14 %
LagunAroVida	=	1.37 %
BansabadellVida	=	1.83 %
Finisterre	=	3.98 %
ZurichInternacional	=	3.94 %

REPRESENTACION GRAFICA DE LA NUBE COMPROMISO
DESCOMPOSICION DE LA INERCIA SEGUN LOS EJES PRINCIPALES

Traza de la matriz: 2.2732
Histograma de los 47 primeros valores propios

NUMERO	VALOR PROPIO	PORCENTAJE	PORCENTAJE ACUMUL.
1	0.8105	35.65	35.65
2	0.4542	19.98	55.64
3	0.2566	11.29	66.93
4	0.1885	8.29	75.22
5	0.1145	5.04	80.26
6	0.0823	3.62	83.87
7	0.0677	2.98	86.85
8	0.0590	2.60	89.45
9	0.0506	2.22	91.67
10	0.0447	1.97	93.64
11	0.0366	1.61	95.25
12	0.0288	1.27	96.52
13	0.0162	0.71	97.23
14	0.0146	0.64	97.87
15	0.0114	0.50	98.37
16	0.0077	0.34	98.71
17	0.0056	0.25	98.96
18	0.0050	0.22	99.18
19	0.0043	0.19	99.36
20	0.0028	0.12	99.48
21	0.0024	0.11	99.59
22	0.0023	0.10	99.69
23	0.0019	0.09	99.78
24	0.0014	0.06	99.84
25	0.0012	0.05	99.89
26	0.0007	0.03	99.92
27	0.0006	0.03	99.95
28	0.0005	0.02	99.97
29	0.0003	0.01	99.98
30	0.0002	0.01	99.99
31	0.0001	0.00	99.99
32	0.0001	0.00	100.00
33	0.0000	0.00	100.00
34	0.0000	0.00	100.00
35	0.0000	0.00	100.00
36	0.0000	0.00	100.00
37	0.0000	0.00	100.00
38	0.0000	0.00	100.00
39	0.0000	0.00	100.00
40	0.0000	0.00	100.00
41	0.0000	0.00	100.00
42	0.0000	0.00	100.00
43	0.0000	0.00	100.00
44	0.0000	0.00	100.00
45	0.0000	0.00	100.00
46	0.0000	0.00	100.00
47	0.0000	0.00	100.00

COORDENADAS, CONTRIBUCIONES Y COSEÑOS CUADRADOS DE LOS INDIVIDUOS-COMPROMISO SOBRE LOS EJES PRINCIPALES

LINEA	DISTO	COORDENADAS					CONTRIBUCIONES					COSEÑOS CUADRADOS				
		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Euroseguros	0.03	0.16	-0.03	0.01	0.00	-0.02	0.03	0.00	0.00	0.00	0.01	0.91	0.03	0.00	0.00	0.02
MapfreMutualidad	0.08	-0.15	0.08	-0.10	-0.14	-0.08	0.03	0.01	0.04	0.11	0.05	0.28	0.08	0.11	0.25	0.07
Vidacaixa	0.04	-0.17	-0.04	0.00	0.01	-0.01	0.03	0.00	0.00	0.00	0.00	0.78	0.04	0.00	0.00	0.00
Casergupo	0.01	0.02	-0.02	-0.01	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.06	0.06	0.02	0.09	0.02
Vitalicio	0.01	-0.03	-0.05	-0.01	-0.03	0.02	0.00	0.01	0.00	0.00	0.00	0.14	0.37	0.01	0.10	0.05
ParclaysVida	0.05	0.19	-0.03	0.00	0.01	-0.01	0.04	0.00	0.00	0.00	0.00	0.73	0.03	0.00	0.00	0.00
WinterthurSegGrales	0.05	-0.15	0.07	-0.06	0.01	-0.09	0.03	0.01	0.02	0.00	0.07	0.49	0.10	0.09	0.00	0.16
AllianzRas	0.01	-0.01	-0.01	-0.05	-0.03	0.02	0.00	0.00	0.01	0.01	0.00	0.02	0.01	0.24	0.11	0.03
MapfreVida	0.03	0.03	-0.04	0.00	-0.04	0.08	0.00	0.00	0.00	0.01	0.06	0.05	0.07	0.00	0.05	0.25
AntaresVida	0.06	0.20	-0.02	0.03	0.02	-0.03	0.05	0.00	0.00	0.00	0.01	0.66	0.01	0.01	0.00	0.02
Stalucia	0.06	-0.09	0.00	-0.05	0.20	0.02	0.01	0.00	0.01	0.21	0.00	0.15	0.00	0.05	0.66	0.01
MutuaMadri	0.04	-0.02	0.03	-0.10	-0.14	-0.03	0.00	0.00	0.04	0.10	0.01	0.01	0.03	0.25	0.44	0.02
AGFUnionFenix	0.03	-0.12	-0.04	-0.03	0.01	0.04	0.02	0.00	0.00	0.00	0.01	0.51	0.05	0.02	0.01	0.05
CatalaOccident	0.02	-0.09	-0.03	-0.04	-0.01	-0.04	0.01	0.00	0.01	0.00	0.02	0.48	0.05	0.10	0.01	0.11
Asisa	0.12	0.01	0.30	0.16	-0.03	-0.02	0.00	0.20	0.10	0.00	0.01	0.00	0.74	0.22	0.01	0.00
AscatVida	0.04	0.19	-0.04	0.01	0.01	-0.02	0.04	0.00	0.00	0.00	0.00	0.88	0.03	0.00	0.00	0.01
MapfreSegGrales	0.04	-0.17	-0.02	0.02	0.05	-0.05	0.04	0.00	0.00	0.01	0.02	0.70	0.01	0.01	0.07	0.07
LaEstrella	0.01	0.01	-0.03	-0.03	0.01	0.03	0.00	0.00	0.00	0.00	0.01	0.01	0.08	0.10	0.02	0.11
Intercase	0.03	0.13	-0.04	-0.02	0.00	0.04	0.02	0.00	0.00	0.00	0.02	0.62	0.06	0.01	0.00	0.07
AegonUnionAsegurad	0.01	-0.08	0.04	-0.04	0.00	0.01	0.01	0.00	0.01	0.00	0.00	0.52	0.10	0.14	0.00	0.00
CajaMadridVida	0.02	0.13	-0.02	0.00	0.00	0.05	0.02	0.00	0.00	0.00	0.03	0.68	0.01	0.00	0.00	0.13
Ocaso	0.05	-0.13	0.02	-0.04	0.15	0.04	0.02	0.00	0.01	0.12	0.01	0.36	0.01	0.03	0.48	0.03
PlusUltra	0.01	-0.07	-0.05	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.00	0.55	0.25	0.00	0.00	0.07
Adeslas	0.14	-0.01	0.32	0.18	-0.02	0.02	0.00	0.13	0.08	0.00	0.01	0.00	0.74	0.24	0.00	0.00
Sanitas	0.10	-0.02	0.26	0.14	-0.01	0.08	0.00	0.14	0.03	0.00	0.06	0.01	0.66	0.21	0.00	0.07
DbVida	0.02	0.13	-0.03	-0.01	0.00	0.02	0.02	0.00	0.00	0.00	0.00	0.71	0.05	0.00	0.00	0.02
Santander	0.03	0.10	0.00	0.01	0.00	0.04	0.01	0.00	0.00	0.00	0.02	0.31	0.00	0.00	0.00	0.06
AssicurazioniGenerali	0.02	-0.10	-0.03	-0.01	-0.01	-0.01	0.01	0.00	0.00	0.00	0.00	0.58	0.06	0.00	0.01	0.00
ArgentariaVida	0.03	0.15	-0.03	0.01	0.00	-0.04	0.03	0.00	0.00	0.00	0.01	0.70	0.02	0.00	0.00	0.05
Banesto	0.06	0.11	-0.08	0.07	0.00	-0.01	0.01	0.01	0.02	0.00	0.00	0.17	0.09	0.07	0.00	0.00
ZurichSegEspaña	0.03	-0.09	0.04	-0.06	-0.04	-0.02	0.01	0.00	0.02	0.01	0.00	0.24	0.06	0.13	0.06	0.02
SegurosGenesis	0.02	0.10	-0.03	-0.01	-0.02	0.07	0.01	0.00	0.00	0.00	0.05	0.45	0.04	0.00	0.02	0.26
BiharkoVida	0.03	0.17	-0.02	0.00	0.01	0.00	0.04	0.00	0.00	0.00	0.00	0.92	0.01	0.00	0.00	0.00
FiactMutua	0.03	-0.10	0.09	-0.02	-0.02	-0.06	0.01	0.02	0.00	0.00	0.03	0.36	0.33	0.01	0.02	0.13
BilbaoCiaAnma	0.01	-0.07	0.00	-0.06	0.04	0.04	0.01	0.00	0.02	0.01	0.01	0.31	0.00	0.29	0.10	0.09
AseguradValenciana	0.02	0.13	-0.02	0.01	-0.01	0.04	0.02	0.00	0.00	0.00	0.02	0.75	0.02	0.01	0.00	0.07
MultinacionalAsegura	0.10	-0.11	0.09	-0.14	-0.07	-0.04	0.01	0.02	0.08	0.03	0.01	0.11	0.09	0.21	0.05	0.01
WinterthurVida	0.02	0.09	-0.03	-0.03	0.01	0.06	0.01	0.00	0.00	0.00	0.03	0.42	0.05	0.04	0.00	0.18
CreditoCaucion	0.20	-0.29	-0.25	0.22	-0.04	-0.04	0.10	0.14	0.19	0.01	0.02	0.41	0.32	0.24	0.01	0.01
Athens	0.03	-0.08	-0.02	-0.10	-0.02	0.05	0.01	0.00	0.04	0.00	0.02	0.24	0.01	0.38	0.01	0.08
BankinterVida	0.11	0.24	-0.04	0.02	0.02	-0.16	0.07	0.00	0.00	0.00	0.21	0.51	0.01	0.01	0.00	0.21
MutuaGeneralSegur	0.02	-0.12	-0.02	-0.04	-0.02	0.04	0.02	0.00	0.01	0.00	0.02	0.59	0.02	0.05	0.01	0.07
PelayoMutua	0.04	-0.09	-0.01	-0.07	-0.10	0.03	0.01	0.00	0.02	0.05	0.01	0.20	0.00	0.14	0.25	0.03
Musini	0.12	-0.14	-0.22	0.17	-0.07	0.00	0.02	0.11	0.12	0.02	0.00	0.16	0.41	0.26	0.04	0.00
LagunAroVida	0.03	0.14	-0.03	-0.01	0.04	0.00	0.02	0.00	0.00	0.01	0.00	0.65	0.02	0.00	0.05	0.00
BansabadelVida	0.04	0.14	0.01	0.00	0.00	-0.11	0.02	0.00	0.00	0.00	0.10	0.47	0.00	0.00	0.00	0.28
Finisterre	0.09	-0.16	0.09	-0.03	0.22	-0.06	0.03	0.02	0.00	0.26	0.04	0.27	0.08	0.01	0.54	0.05
ZurichInternacional	0.09	-0.24	-0.12	0.09	0.01	-0.02	0.07	0.03	0.03	0.00	0.00	0.66	0.15	0.09	0.00	0.00

A.2.1.7 Las variables (ACP centrado de las nubes ponderadas)

COORDENADAS, CORRELACIONES Y EJES UNITARIOS DE LAS VARIABLES SOBRE LOS CINCO PRIMEROS EJES

VARIABLES	COORDENADAS					CORRELACIONES VARIABLE-FACTOR					EJES UNITARIOS INICIALES				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
EJES															
VIDA95	0.03	-0.01	0.00	0.00	0.00	-0.90	-0.34	0.00	-0.07	0.10	0.25	-0.13	0.00	-0.04	0.07
AUTOS95	-0.02	0.01	-0.02	0.02	-0.01	0.49	0.17	-0.61	0.51	-0.17	-0.14	0.06	-0.30	0.29	-0.12
SALUD95	0.00	0.03	0.02	0.00	0.00	0.04	0.79	0.56	0.09	0.13	-0.01	0.29	0.28	0.05	0.10
RESTO95	-0.03	-0.01	0.00	-0.02	0.00	0.79	-0.31	0.10	-0.43	-0.09	-0.22	-0.11	0.05	-0.25	-0.07
INDUS95	-0.02	-0.02	0.01	0.01	0.00	0.67	-0.52	0.39	0.15	-0.10	-0.18	-0.19	0.19	0.09	-0.08
PRIMEM95	0.02	0.00	0.00	0.00	-0.02	-0.53	-0.11	0.05	-0.10	-0.53	0.15	-0.04	0.02	-0.06	-0.39
GEXTPR95	-0.02	0.00	-0.01	-0.03	0.00	0.54	0.05	-0.32	-0.70	0.05	-0.15	0.02	-0.16	-0.40	0.04
GINTPR95	-0.02	0.00	0.00	0.00	0.01	0.68	0.09	-0.08	0.00	0.31	-0.19	0.03	-0.04	0.00	0.23
MMMI95	-0.01	-0.01	0.01	0.00	0.00	0.36	-0.28	0.27	0.14	0.12	-0.10	-0.10	0.13	0.08	0.09
INFINIT95	0.02	-0.02	0.00	0.00	0.02	-0.56	-0.49	0.01	0.02	0.47	0.15	-0.18	0.01	0.01	0.35
NNCPR95	0.00	0.00	0.01	-0.01	0.00	-0.13	-0.13	0.34	-0.16	0.10	0.04	0.05	0.14	-0.09	0.07
NNCPR95	0.00	0.03	0.01	0.00	0.00	0.08	0.84	0.38	-0.09	-0.02	-0.02	0.31	0.17	-0.05	-0.02
NNPR95	0.02	0.02	-0.02	-0.01	0.00	-0.46	0.57	-0.52	-0.16	0.14	0.13	0.21	-0.26	-0.09	0.10
VIDA96	0.03	-0.01	0.00	0.00	0.00	-0.86	-0.41	0.07	-0.04	0.10	0.23	-0.14	0.03	-0.02	0.07
AUTOS96	-0.02	0.01	-0.02	0.02	0.00	0.48	0.16	-0.61	0.53	0.14	-0.13	0.06	-0.29	-0.10	0.10
SALUD96	0.00	0.03	0.02	0.00	0.00	0.04	0.79	0.56	0.09	0.14	-0.01	0.28	0.26	0.05	0.10
RESTO96	-0.03	-0.01	0.00	-0.02	0.00	0.79	-0.21	0.01	-0.51	-0.12	-0.21	-0.07	0.01	-0.28	-0.09
GEXTPR96	-0.02	-0.02	0.01	0.00	0.00	0.68	-0.44	0.34	0.05	-0.15	-0.18	-0.16	0.16	0.03	-0.10
INDUS96	0.02	-0.01	0.00	0.00	-0.02	-0.67	-0.18	0.12	-0.09	-0.51	0.18	-0.06	0.05	-0.05	-0.36
PRIMEM96	-0.02	0.00	0.00	0.00	0.00	0.63	0.07	-0.38	-0.63	0.02	-0.17	0.03	-0.18	-0.34	0.01
GINTPR96	-0.02	0.00	-0.01	-0.02	0.00	0.84	0.18	-0.21	0.08	0.21	-0.22	0.06	-0.10	0.04	0.15
MMMI96	-0.03	0.01	-0.01	0.00	0.01	0.59	-0.44	0.20	0.27	0.18	-0.16	-0.16	0.10	0.15	0.13
INFINIT96	0.02	-0.02	0.00	0.00	0.02	-0.56	-0.48	0.04	-0.11	0.47	0.15	-0.17	0.02	-0.06	0.33
NNCPR96	0.02	0.00	0.00	0.00	-0.01	-0.66	0.10	0.04	-0.09	-0.30	0.17	0.04	0.02	-0.05	-0.21
NNCPR96	-0.01	0.03	0.01	0.00	0.00	0.15	0.86	0.39	-0.11	-0.04	-0.04	0.30	0.18	-0.06	-0.03
NNPR96	0.02	0.02	-0.02	-0.01	0.00	-0.55	0.55	-0.50	-0.15	0.11	0.15	0.19	-0.23	-0.08	0.08
VIDA97	0.03	-0.01	0.00	0.00	0.00	-0.89	-0.34	0.00	-0.08	0.10	0.23	-0.12	0.00	-0.04	0.07
AUTOS97	-0.02	0.01	-0.02	0.02	0.00	0.47	0.16	-0.61	0.54	-0.13	-0.12	0.06	-0.28	0.29	-0.09
SALUD97	0.00	0.03	0.02	0.00	0.00	0.04	0.79	0.56	0.09	0.14	-0.01	0.28	0.26	0.05	0.10
RESTO97	-0.03	-0.01	0.00	-0.02	0.00	0.79	-0.29	0.12	-0.47	-0.13	-0.21	-0.10	0.06	-0.25	-0.09
INDUS97	-0.02	-0.02	0.01	0.00	0.00	0.68	-0.50	0.41	0.09	-0.13	-0.18	-0.18	0.19	0.05	-0.09
PRIMEM97	0.02	-0.01	0.00	0.00	-0.01	-0.68	-0.16	0.10	-0.10	-0.38	0.18	-0.05	0.05	-0.05	-0.27
GEXTPR97	-0.02	0.00	-0.01	-0.02	0.00	0.64	0.10	-0.34	-0.63	0.03	-0.17	0.04	-0.16	-0.34	0.02
GINTPR97	-0.03	0.00	-0.01	0.00	0.01	0.85	0.14	-0.22	0.12	0.16	-0.22	0.05	-0.10	0.06	0.11
MMMI97	-0.02	-0.01	0.00	0.01	0.00	0.66	-0.42	0.11	0.31	0.10	-0.17	-0.15	0.05	0.17	0.07
INFINIT97	0.02	-0.02	0.00	0.00	0.02	-0.48	-0.54	0.00	-0.09	0.45	0.13	-0.19	0.00	-0.05	0.32
NNC97	0.02	0.00	0.00	0.00	0.00	-0.65	-0.01	0.02	-0.09	-0.10	0.17	0.00	0.01	-0.05	-0.07
NNCPR097	0.00	0.03	0.01	0.00	0.00	0.13	0.86	0.39	-0.10	0.03	-0.04	0.30	0.18	-0.05	0.02
NNPR97	0.02	0.02	-0.02	-0.01	0.00	-0.55	0.54	-0.50	-0.15	0.11	0.15	0.19	-0.23	-0.08	0.08

A.2.2 Resultados obtenidos de la aplicación del método *Statis dual*²

A.2.2.1 Comparación de las matrices de correlación (nubes de columnas)

NUBE 1- AÑO 1995													
MATRIZ DE CORRELACION													
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1.000												
2	-0.586	1.000											
3	-0.323	-0.148	1.000										
4	-0.628	0.054	-0.210	1.000									
5	-0.468	-0.076	-0.161	0.716	1.000								
6	0.450	-0.268	-0.114	-0.303	-0.239	1.000							
7	-0.459	0.136	-0.162	0.653	0.088	-0.257	1.000						
8	-0.575	0.342	0.069	0.449	0.363	-0.439	0.483	1.000					
9	-0.223	0.038	-0.045	0.315	0.431	-0.267	0.107	0.358	1.000				
10	0.626	-0.371	-0.303	-0.310	-0.110	0.220	-0.320	-0.289	-0.086	1.000			
11	0.174	-0.363	0.046	0.041	0.044	0.074	-0.072	-0.081	0.150	0.059	1.000		
12	-0.395	-0.048	0.825	-0.113	-0.257	0.010	0.046	0.188	-0.175	-0.382	-0.145	1.000	
13	0.279	0.047	0.152	-0.553	-0.849	0.197	0.011	-0.189	-0.533	0.009	-0.059	0.282	1.000

REPARTO DE LA INERCIA SEGUN LOS EJES PRINCIPALES			
Norma V _{1M} = 5.478			
Traza de la matriz = 2.3731			
Histograma de los 13 primeros valores propios			
NUMERO	VALOR PROPIO	PORCENTAJE	ACUMULADO
1	0.7497	31.59	31.59
2	0.4806	20.25	51.84
3	0.2911	12.26	64.11
4	0.2139	9.02	73.12
5	0.1709	7.20	80.32
6	0.1442	6.08	86.40
7	0.1161	4.89	91.29
8	0.0932	3.93	95.22
9	0.0582	2.45	97.67
10	0.0270	1.14	98.81
11	0.0198	0.83	99.64
12	0.0085	0.36	100.00
13	0.0000	0.00	100.00

² Dado que las variables y los individuos son los mismos que los analizados mediante el *Statis*, en este caso no es necesario recoger de nuevo sus medias y sus desviaciones típicas. Los epígrafes siguientes, hasta el A.2.2.5, excepto el A.2.2.3, recogen los resultados obtenidos al aplicar el método *Statis dual* mediante el programa SPAD-TM. Dicho epígrafe A.2.2.3 ha sido calculado mediante la hoja de cálculo EXCEL. El epígrafe A.2.2.6 contiene las coordenadas de los individuos obtenidas al aplicar un ACP centrado a las nubes de variables debidamente ponderadas, mediante el módulo de Análisis Factorial del SPAD. Las representaciones gráficas recogidas en el A.2.2.7, así como las que aparecen en el texto, son el resultado de esta última técnica.

NUBE 2 - AÑO 1996

MATRIZ DE CORRELACION

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1.000												
2	-0.616	1.000											
3	-0.335	-0.150	1.000										
4	-0.622	0.087	-0.191	1.000									
5	-0.436	0.086	-0.141	0.684	1.000								
6	0.596	-0.371	-0.162	-0.399	-0.306	1.000							
7	-0.586	0.235	-0.168	0.787	0.221	-0.434	1.000						
8	-0.786	0.550	0.098	0.563	0.445	-0.701	0.568	1.000					
9	-0.278	0.176	-0.177	0.393	0.591	-0.397	0.129	0.438	1.000				
10	0.642	-0.401	-0.309	-0.315	-0.214	0.307	-0.298	-0.452	-0.143	1.000			
11	0.471	-0.283	0.020	-0.448	-0.389	0.646	-0.331	-0.566	-0.511	0.243	1.000		
12	-0.470	-0.046	0.874	0.015	-0.163	-0.142	0.098	0.167	-0.235	-0.430	0.113	1.000	
13	0.230	0.026	0.131	-0.475	-0.787	0.240	-0.042	-0.226	-0.642	0.065	0.421	0.223	1.000

REPARTO DE LA INERCIA SEGUN LOS EJES PRINCIPALES

Norma $V_2M = 6.243$

Traza de la matriz = 2.0823

Histograma de los 13 primeros valores propios

NUMERO	VALOR PROPIO	PORCENTAJE	ACUMULADO
1	0,8334	40.02	40.02
2	0,4263	20.47	60.50
3	0,2368	11.37	71.87
4	0,1868	8.97	80.84
5	0,1375	6.60	87.44
6	0,0754	3.62	91.07
7	0,0621	2.98	94.05
8	0,0469	2.25	96.30
9	0,0377	1.81	98.11
10	0,0223	1.07	99.18
11	0,0106	0.51	99.69
12	0,0065	0.31	100.00
13	0,0000	0.00	100.00

A.2.2.2 *Situación de las nubes de variables (matrices de correlación) unas respecto las otras*

PRODUCTOS ESCALARES ENTRE MATRICES DE CORRELACION (SIN NORMAR)

Matr	95	96	97
95	30.017		
96	31.995	38.975	
97	32.400	39.163	39.905

PRODUCTOS ESCALARES ENTRE MATRICES DE CORRELACION NORMADAS
Coeficientes RV entre matrices

Matr	95	96	97
95	1.000		
96	0.935	1.000	
97	0.936	0.993	1.000

TEST DE PERMUTACIONES

* indica que las nubes son similares

Matr	95	96	97
95	*		
96	*	*	
97	*	*	*

INTERPRETACION DE LAS DISTANCIAS ENTRE MATRICES DE CORRELACION

DISTANCIA EUCLIDEA ENTRE MATRICES DE CORRELACION

Matr	95	96	97
95	0.000		
96	2.237	0.000	
97	2.263	0.744	0.000

DESCOMPOSICION DE LA DISTANCIA AL CUADRADO SEGÚN LAS VARIABLES

- ENTRE LA MATRIZ DE CORRELACION 95 Y LAS OTRAS

COLUMNA	MEDIA	MATRIZ	
		96	97
VIDA	4.7	3.8	5.6
AUTOS	2.4	1.9	3.0
SALUD	0.8	0.5	1.0
RESTO	5.9	6.2	5.7
INDUS	5.1	5.4	4.8
PRIMEM	10.5	10.5	10.6
GEXTPR	3.6	3.5	3.8
GINTPR	9.0	9.0	8.9
MMMMI	11.6	11.0	12.1
INFINIT	2.2	1.7	2.7
NNC	37.1	38.3	35.8
NNCPRO	2.3	2.9	1.8
NNPR	4.8	5.4	4.2

- ENTRE LA MATRIZ DE CORRELACION 96 Y LAS OTRAS

COLUMNA	MEDIA	MATRIZ	
		95	97
VIDA	7.6	3.8	11.3
AUTOS	4.6	1.9	7.4
SALUD	1.8	0.5	3.2
RESTO	8.1	6.2	10.1
INDUS	5.9	5.4	6.4
PRIMEM	5.9	10.5	1.4
GEXTPR	3.4	3.5	3.3
GINTPR	6.6	9.0	4.2
MMMMI	15.5	11.0	20.0
INFINIT	6.3	1.7	10.8
NNC	24.0	38.3	9.6
NNCPRO	2.6	2.9	2.4
NNPR	7.6	5.4	9.8

- ENTRE LA MATRIZ DE CORRELACION 97 Y LAS OTRAS

+-----+-----+-----+-----+				
MATRIZ				
+-----+-----+-----+-----+				
COLUMNA	MEDIA	95	96	
+-----+-----+-----+-----+				
VIDA	8.5	5.6	11.3	
AUTOS	5.2	3.0	7.4	
SALUD	2.1	1.0	3.2	
RESTO	7.9	5.7	10.1	
INDUS	5.6	4.8	6.4	
PRIMEM	6.0	10.6	1.4	
GEXTPR	3.6	3.8	3.3	
GINTPR	6.6	8.9	4.2	
MMMMI	16.1	12.1	20.0	
INFINIT	6.8	2.7	10.8	
NNC	22.7	35.8	9.6	
NNCPRO	2.1	1.8	2.4	
NNPR	7.0	4.2	9.8	
+-----+-----+-----+-----+				

LA SUMA DE LAS DISTANCIAS AL CUADRADO SE DECOMPONE SEGUN LAS VARIABLES:

VIDA	=	5.08	%
AUTOS	=	2.69	%
SALUD	=	0.89	%
RESTO	=	6.15	%
INDUS	=	5.15	%
PRIMEM	=	10.04	%
GEXTPR	=	3.62	%
GINTPR	=	8.71	%
MMMMI	=	12.00	%
INFINIT	=	2.66	%
NNC	=	35.62	%
NNCPRO	=	2.35	%
NNPR	=	5.04	%

A.2.2.3 Estudio de la nube de nubes: interestructura no-centrada

DESCOMPOSICION DE LA INERCIA SEGUN LOS EJES PRINCIPALES

Traza de la matriz = 3

Histograma de los 3 primeros valores propios

NUMERO	VALOR PROPIO	PORCENTAJE	PORCENTAJE ACUMULADO
95	2.9096	96.98	96.98
96	0.0834	2.78	99.76
97	0.0070	0.24	100.00

COORDENADAS, CONTRIBUCIONES Y COSENO CUADRADOS DE LAS NUBES SOBRE LOS EJES PRINCIPALES

NUBE	DISTO	COORDENADAS			CONTRIBUCIONES			COSENO CUADRADOS					
		1	2	3	1	2	3	1	2	3			
95	1	0.97	0.24	0.00	0.00	0.32	0.67	0.00	0.00	0.64	0.05	0.00	0.00
96	1	0.99	-0.12	-0.06	0.00	0.34	0.16	0.49	0.00	0.98	0.01	0.00	0.00
97	1	0.99	-0.11	0.06	0.00	0.34	0.16	0.50	0.00	0.98	0.01	0.00	0.00

A.2.2.4 Estudio de la nube de las matrices de correlación tomando como origen la nube compromiso

CONSTRUCCION DE LA MATRIZ DE CORRELACION COMPROMISO = MEDIA PONDERADA DE LAS MATRICES DE CORRELACION PONDERADA CON LOS COEFICIENTES

NUBE 95 = 0.298
 NUBE 96 = 0.349
 NUBE 97 = 0.353

MATRIZ DE CORRELACION COMPROMISO

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1.000												
2	-0.596	1.000											
3	-0.328	-0.150	1.000										
4	-0.619	0.053	-0.198	1.000									
5	-0.452	0.068	-0.148	0.706	1.000								
6	0.553	-0.336	-0.145	-0.366	-0.284	1.000							
7	-0.546	0.191	-0.147	0.726	0.172	-0.388	1.000						
8	-0.721	0.492	0.080	0.515	0.417	-0.603	0.547	1.000					
9	-0.326	0.189	-0.146	0.410	0.565	-0.359	0.147	0.455	1.000				
10	0.611	-0.362	-0.324	-0.282	-0.140	0.296	-0.290	-0.394	-0.106	1.000			
11	0.410	-0.314	-0.008	-0.295	-0.253	0.487	-0.268	-0.424	-0.270	0.224	1.000		
12	-0.420	-0.051	0.862	-0.044	-0.202	-0.088	0.089	0.163	-0.221	-0.427	0.014	1.000	
13	0.287	0.032	0.134	-0.548	-0.842	0.224	-0.029	-0.230	-0.607	0.029	0.257	0.238	1.000

PRODUCTOS ESCALARES Y DISTANCIAS ENTRE LA MATRIZ COMPROMISO Y LAS DIFERENTES MATRICES DE CORRELACION

NUBE 95 : PRODUCTO ESCALAR = 31.549 , DISTANCIA = 1.558
 NUBE 96 : PRODUCTO ESCALAR = 36.962 , DISTANCIA = 0.747
 NUBE 97 : PRODUCTO ESCALAR = 37.411 , DISTANCIA = 0.769

PRODUCTOS ESCALARES Y DISTANCIAS ENTRE LA MATRIZ COMPROMISO Y LAS DIFERENTES MATRICES DE CORRELACION NORMALIZADAS

NUBE 95 : PRODUCTO ESCALAR = 0.971
 NUBE 96 : PRODUCTO ESCALAR = 0.991
 NUBE 97 : PRODUCTO ESCALAR = 0.991

DESCOMPOSICION DE LA INERCIA SEGUN LAS MATRICES DE CORRELACION

MATRIZ DE CORRELACION 95 = 67.85 %
 MATRIZ DE CORRELACION 96 = 15.61 %
 MATRIZ DE CORRELACION 97 = 16.54 %

REPRESENTACION GRAFICA DE LA NUBE DE MATRICES DE CORRELACION (INTERESTRUCTURA CENTRADA)

DESCOMPOSICION DE LA INERCIA SEGUN LOS EJES PRINCIPALES

Traza de la matriz = 3.5781

Histograma de los 3 primeros valores propios

NUMERO	VALOR PROPIO	PORCENTAJE	PORCENTAJE ACUMUL.
1	3.3017	92.28	92.28
2	0.2764	7.72	100.00
3	0.0000	0.00	100.00

COORDENADAS DE LAS MATRICES DE CORRELACION SOBRE LOS EJES PRINCIPALES (INTERESTRUCTURA CENTRADA)

NUBE	COORDENADAS			CONTRIBUCION			COSENO CUADRADOS					
	1	2	3	0	1	2	3	0	1	2	3	0
95	2.43	1.56	0.00	0.00	0.74	0.00	0.26	0.00	1.00	0.00	0.00	0.00
96	0.56	-0.65	0.38	0.00	0.13	0.51	0.36	0.00	0.75	0.25	0.00	0.00
97	0.59	-0.68	-0.37	0.00	0.14	0.49	0.37	0.00	0.77	0.23	0.00	0.00

A.2.2.5 Estudio de la matriz compromiso

REPRESENTACION GRAFICA DE LA NUBE COMPROMISO
DESCOMPOSICION DE LA INERCIA SEGUN LOS EJES PRINCIPALES

Traza de la matriz : 2.1621
Histograma de los 13 primeros valores propios

NUMERO	VALOR PROPIO	PORCENTAJE	PORCENTAJE ACUMULADO
1	0,8050	37.23	37.23
2	0,4396	20.33	57.56
3	0,2477	11.46	69.02
4	0,1906	8.82	77.83
5	0,1383	6.40	84.23
6	0,1031	4.77	89.00
7	0,0772	3.57	92.57
8	0,0691	3.20	95.76
9	0,0451	2.08	97.85
10	0,0220	1.02	98.87
11	0,0149	0.69	99.56
12	0,0095	0.44	100.00
13	0,0000	0.00	100.00

COORDENADAS, CONTRIBUCIONES Y COSENO CUADRADOS DE LAS VARIABLES-COMPROMISO SOBRE LOS EJES PRINCIPALES

VARIABLE	DISTO					COORDENADAS					CONTRIBUCION					COSENO CUADRADOS				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
VIDA	1.00	-0.87	-0.37	0.00	0.01	0.18	0.16	0.05	0.00	0.00	0.04	0.76	0.14	0.00	0.00	0.03				
AUTOS	1.00	0.47	0.18	-0.61	0.40	-0.38	0.05	0.01	0.25	0.14	0.17	0.22	0.03	0.37	0.16	0.15				
SALUD	1.00	0.04	0.78	0.54	0.16	0.13	0.00	0.23	0.20	0.02	0.02	0.00	0.62	0.30	0.03	0.02				
RESTO	1.00	0.78	-0.26	0.13	-0.51	-0.02	0.13	0.03	0.01	0.23	0.00	0.62	0.07	0.02	0.26	0.00				
INDUS	1.00	0.68	-0.49	0.41	0.02	-0.15	0.10	0.09	0.12	0.00	0.03	0.46	0.24	0.17	0.00	0.02				
PRIMEM	1.00	-0.69	-0.15	0.14	-0.20	-0.43	0.10	0.01	0.01	0.03	0.22	0.47	0.02	0.02	0.04	0.18				
GEXTPR	1.00	0.62	0.09	-0.31	-0.66	0.13	0.08	0.00	0.07	0.38	0.02	0.38	0.01	0.10	0.43	0.02				
GINTPR	1.00	0.82	0.15	-0.18	0.04	0.03	0.14	0.01	0.02	0.00	0.00	0.68	0.02	0.03	0.00	0.00				
MMMMI	1.00	0.60	-0.41	0.18	0.31	-0.03	0.07	0.06	0.02	0.08	0.00	0.36	0.17	0.03	0.10	0.00				
INFINIT	1.00	-0.52	-0.52	-0.01	-0.01	0.38	0.06	0.10	0.00	0.00	0.17	0.27	0.27	0.00	0.00	0.14				
NNC	1.00	-0.57	-0.02	0.19	-0.30	-0.51	0.07	0.00	0.03	0.08	0.31	0.32	0.00	0.04	0.09	0.26				
NNCPRO	1.00	0.10	0.87	0.39	-0.09	0.02	0.00	0.28	0.10	0.01	0.00	0.01	0.75	0.16	0.01	0.00				
NNPR	1.00	-0.54	0.56	-0.51	-0.16	0.06	0.06	0.12	0.17	0.02	0.00	0.30	0.31	0.26	0.03	0.00				

A.2.2.6 Los individuos (ACP centrado de las matrices ponderadas)

COORDENADAS, CONTRIBUCIONES Y CONSENSOS CUADRADOS DE LOS INDIVIDUOS EN LOS PRIMEROS CINCO EJES

INDIVIDUOS	COORDENADAS					CONTRIBUCIONES					COSENSOS CUADRADOS						
	P.REL	DISTO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Euroseguros95	0.69	7.06	-2.49	-0.62	0.05	0.20	-0.32	0.9	0.1	0.0	0.0	0.1	0.88	0.05	0.00	0.01	0.01
MapfireMutualidad95	0.69	16.71	1.72	1.29	-1.55	1.71	2.26	0.4	0.4	1.1	1.8	4.3	0.18	0.10	0.14	0.17	0.31
VidaCaixa95	0.69	7.20	-2.53	-0.62	-0.08	0.11	-0.35	0.9	0.1	0.0	0.0	0.1	0.89	0.05	0.00	0.00	0.02
Casagrupo95	0.69	1.27	0.07	-0.48	-0.34	-0.47	-0.34	0.0	0.1	0.1	0.1	0.1	0.00	0.18	0.09	0.17	0.09
Vitalicio95	0.69	0.79	0.07	-0.59	-0.23	0.24	0.06	0.0	0.1	0.0	0.0	0.0	0.01	0.44	0.07	0.07	0.00
BarclaysVida95	0.69	6.42	-2.28	-0.55	-0.22	0.15	-0.74	0.7	0.1	0.0	0.0	0.5	0.81	0.05	0.01	0.00	0.09
WinterthurSegrales95	0.69	12.74	2.30	1.13	-0.80	-0.58	1.66	0.8	0.3	0.3	0.2	2.3	0.42	0.10	0.05	0.03	0.22
AllianzRas95	0.69	1.57	-0.59	0.06	-0.58	0.35	-0.03	0.0	0.0	0.2	0.1	0.0	0.22	0.00	0.22	0.08	0.00
MapfreVida95	0.69	6.04	-0.69	-0.67	-0.13	0.80	-0.64	0.1	0.1	0.0	0.4	0.3	0.08	0.07	0.00	0.11	0.07
AntaresVida95	0.69	6.76	-2.40	-0.26	0.16	0.06	-0.50	0.8	0.0	0.0	0.0	0.2	0.85	0.01	0.00	0.00	0.04
Stalucia95	0.69	13.85	1.13	0.16	-0.65	-3.10	-0.70	0.2	0.0	0.2	5.8	0.4	0.09	0.00	0.03	0.69	0.04
MutuaMadril95	0.69	10.75	0.07	0.40	-1.55	1.86	1.43	0.0	0.0	1.1	2.1	1.7	0.00	0.01	0.22	0.32	0.19
AGFUnionFenix95	0.69	6.17	1.51	-0.38	-0.42	-0.58	0.09	0.3	0.0	0.1	0.2	0.0	0.37	0.02	0.03	0.05	0.00
CatalaOccident95	0.69	2.96	1.04	-0.25	-0.82	-0.02	0.16	0.2	0.0	0.3	0.0	0.0	0.37	0.02	0.23	0.00	0.01
Asisa95	0.69	29.36	-0.28	4.64	2.53	0.70	0.08	0.0	5.7	3.0	0.3	0.0	0.00	0.73	0.22	0.02	0.00
AscatVida95	0.69	7.52	-2.56	-0.55	-0.05	0.20	-0.52	0.9	0.1	0.0	0.0	0.2	0.87	0.04	0.00	0.01	0.04
MapfreSegrales95	0.69	10.25	2.45	-0.20	0.63	-1.24	0.76	0.9	0.0	0.2	0.9	0.5	0.59	0.00	0.04	0.15	0.06
LaEstrella95	0.69	1.48	0.09	-0.46	-0.52	-0.25	-0.34	0.0	0.1	0.1	0.0	0.1	0.01	0.14	0.18	0.04	0.08
Intercases95	0.69	6.67	-2.25	-0.65	-0.29	0.20	-0.79	0.7	0.1	0.0	0.0	0.5	0.76	0.06	0.01	0.01	0.09
AegonUnionAsegurad95	0.69	2.74	0.94	0.45	-0.74	-0.33	0.25	0.1	0.1	0.3	0.1	0.1	0.32	0.07	0.20	0.04	0.02
CajaMadridVida95	0.69	4.77	-1.39	-0.48	-0.36	0.32	-1.01	0.3	0.1	0.1	0.1	0.8	0.40	0.05	0.03	0.02	0.21
Ocaso95	0.69	11.40	1.78	0.50	-0.49	-2.50	-0.53	0.5	0.1	0.1	3.8	0.2	0.28	0.02	0.02	0.55	0.02
PlusUltra95	0.69	1.70	0.84	-0.71	-0.04	-0.06	0.05	0.1	0.1	0.0	0.0	0.0	0.42	0.30	0.00	0.00	0.00
Adelas95	0.69	31.81	0.11	4.77	2.76	0.74	-0.40	0.0	6.0	3.6	0.3	0.1	0.00	0.72	0.24	0.02	0.00
Sanitas95	0.69	23.37	0.22	3.69	2.31	0.73	-0.86	0.0	3.6	2.5	0.3	0.6	0.00	0.58	0.23	0.02	0.03
DbVida95	0.69	6.73	-2.41	-0.45	-0.11	0.12	-0.39	0.8	0.1	0.0	0.0	0.1	0.86	0.03	0.00	0.00	0.02
Santander95	0.69	10.24	-0.60	0.80	0.17	-0.09	-0.74	0.1	0.2	0.0	0.0	0.5	0.03	0.06	0.00	0.00	0.05
AssicurazioniGenerali95	0.69	5.83	1.88	-0.34	-0.26	0.22	0.77	0.5	0.0	0.0	0.0	0.5	0.61	0.02	0.01	0.01	0.10
ArgentariaVida95	0.69	5.55	-1.81	-0.83	0.03	0.35	-0.58	0.5	0.2	0.0	0.1	0.3	0.59	0.12	0.00	0.02	0.06
Banesto95	0.69	23.32	1.39	-2.72	1.43	1.26	-0.35	0.3	1.9	1.0	1.0	0.1	0.08	0.32	0.09	0.07	0.01
ZurichSegEspaña95	0.69	5.25	0.68	0.54	-0.57	0.92	1.15	0.1	0.1	0.2	0.5	1.1	0.09	0.06	0.06	0.16	0.25
SegurosGenesis95	0.69	4.99	-1.77	-0.32	-0.29	0.41	-0.86	0.4	0.0	0.0	0.1	0.6	0.63	0.02	0.02	0.03	0.15
BiharkoVida95	0.69	7.41	-2.57	-0.27	0.10	0.10	-0.47	1.0	0.0	0.0	0.0	0.2	0.89	0.01	0.00	0.00	0.03
FiactMutua95	0.69	6.16	1.22	1.56	-0.35	0.16	1.03	0.2	0.6	0.1	0.0	0.9	0.24	0.40	0.02	0.00	0.17
BilbaoCiaAnma95	0.69	2.84	0.78	-0.07	-1.07	-0.41	-0.02	0.1	0.0	0.5	0.1	0.0	0.22	0.00	0.41	0.06	0.00
AseguradValenciana95	0.69	4.78	-1.56	-0.62	0.04	0.64	-0.83	0.3	0.1	0.0	0.3	0.6	0.51	0.08	0.00	0.09	0.14
MultinacionalAsegurad95	0.69	51.87	2.47	1.75	-2.74	2.17	-2.65	0.9	0.8	3.5	2.8	5.9	0.12	0.06	0.15	0.09	0.14
WinterthurVida95	0.69	4.95	-1.83	-0.50	-0.41	0.02	-0.88	0.5	0.1	0.1	0.0	0.6	0.67	0.05	0.03	0.00	0.15
CreditoCaucion95	0.69	42.63	3.84	-3.67	3.49	0.11	0.90	2.1	3.5	5.7	0.0	0.7	0.35	0.32	0.29	0.00	0.02
Athena95	0.69	7.15	1.13	0.37	-1.75	-0.11	0.25	0.2	0.0	1.4	0.0	0.1	0.18	0.02	0.43	0.00	0.01
BankinterVida95	0.69	39.10	-4.20	-0.58	0.81	-0.78	1.99	2.5	0.1	0.3	0.4	3.3	0.45	0.01	0.02	0.02	0.10
MutuaGeneralSegur95	0.69	5.48	1.69	-0.28	-0.62	0.20	0.33	0.4	0.0	0.2	0.0	0.1	0.52	0.01	0.07	0.01	0.02
PelayoMutua95	0.69	6.41	1.34	-0.63	-0.10	0.05	0.54	0.3	0.1	0.0	0.0	0.2	0.28	0.06	0.00	0.00	0.05
Musini95	0.69	25.48	1.91	-3.21	2.77	0.29	0.39	0.5	2.7	3.6	0.1	0.1	0.14	0.40	0.30	0.00	0.01
LagunAroVida95	0.69	7.96	-1.74	0.06	-0.35	-1.14	-0.24	0.4	0.0	0.1	0.8	0.0	0.38	0.00	0.02	0.16	0.01
BansabadellVida95	0.69	10.37	-2.50	-0.01	0.15	-0.02	0.83	0.9	0.0	0.0	0.0	0.6	0.60	0.00	0.00	0.00	0.07
Finisterre95	0.69	23.46	2.34	1.47	-0.18	-3.47	0.33	0.8	0.6	0.0	7.3	0.1	0.23	0.09	0.00	0.51	0.00

COORDENADAS, CONTRIBUCIONES Y CONSENSOS CUADRADOS DE LOS INDIVIDUOS EN LOS PRIMEROS CINCO EJES (Continuación)

IDENTIFICADOR	INDIVIDUOS					COORDENADAS					CONTRIBUCIONES					COSENSOS CUADRADOS				
	P.REL	DISTO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5			
ZurichInternacional95	0.69	17.84	3.42	-1.65	1.23	-0.24	0.76	1.7	0.7	0.7	0.0	0.5	0.66	0.15	0.08	0.00	0.03			
Euroseguros96	0.69	8.82	-2.85	-0.67	0.24	-0.15	0.27	1.2	0.1	0.0	0.0	0.1	0.92	0.05	0.01	0.00	0.01			
MapfireMutualidad96	0.69	25.51	2.81	1.42	-1.67	2.20	1.85	1.1	0.5	1.3	2.9	2.9	0.31	0.08	0.11	0.19	0.13			
VidaCaixa96	0.69	14.72	-3.37	-0.74	0.35	-0.72	1.20	1.6	0.1	0.1	0.3	1.2	0.77	0.04	0.01	0.04	0.10			
Casergupo96	0.69	1.53	-0.34	-0.33	-0.20	-0.23	-0.54	0.0	0.0	0.0	0.0	0.2	0.08	0.07	0.03	0.03	0.19			
Vitalicio96	0.69	1.62	0.37	-0.74	0.00	0.48	0.05	0.0	0.1	0.0	0.0	0.0	0.09	0.34	0.00	0.14	0.00			
BarclaysVida96	0.69	8.24	-2.68	-0.60	-0.01	-0.14	-0.14	1.0	0.1	0.0	0.0	0.0	0.87	0.04	0.00	0.00	0.00			
WinterthurSegGrales96	0.69	12.37	2.51	1.17	-1.06	-0.63	1.19	0.9	0.4	0.5	0.2	1.2	0.51	0.11	0.09	0.03	0.11			
AllianzRas96	0.69	4.17	0.73	-0.25	-0.91	0.28	0.61	0.1	0.0	0.4	0.0	0.3	0.13	0.02	0.20	0.02	0.09			
MapfreVida96	0.69	8.08	-0.39	-0.80	-0.17	1.08	-1.08	0.0	0.2	0.0	0.7	1.0	0.02	0.08	0.00	0.14	0.15			
AntaresVida96	0.69	11.56	-3.02	-0.40	0.57	-0.43	0.61	1.3	0.0	0.2	0.1	0.3	0.79	0.01	0.03	0.02	0.03			
Stalucia96	0.69	18.22	1.75	0.12	-0.85	-3.07	-1.48	0.4	0.0	0.3	5.7	1.8	0.17	0.00	0.04	0.52	0.12			
MutuaMadrri96	0.69	13.27	0.39	0.75	-2.10	2.14	0.76	0.0	0.1	2.1	2.8	0.5	0.01	0.04	0.33	0.35	0.04			
AGFUnionFenix96	0.69	7.79	2.01	-0.65	-0.36	-0.34	-0.32	0.6	0.1	0.1	0.1	0.1	0.52	0.06	0.02	0.01	0.01			
Asisa96	0.69	5.90	1.81	-0.48	-0.78	0.50	-0.57	0.5	0.1	0.3	0.2	0.3	0.56	0.04	0.10	0.04	0.05			
AscatVida96	0.69	38.24	-0.31	5.34	2.92	0.51	0.21	0.0	7.5	4.0	0.2	0.0	0.00	0.75	0.22	0.01	0.00			
MapfreSegrales96	0.69	13.04	3.02	-0.12	0.53	-1.04	0.55	1.3	0.0	0.1	0.7	0.2	0.70	0.00	0.02	0.08	0.02			
LaEstrella96	0.69	4.12	-0.68	-0.40	-0.29	-0.63	0.16	0.1	0.0	0.0	0.2	0.0	0.11	0.04	0.02	0.10	0.01			
Intercase96	0.69	7.89	-2.11	-0.63	-0.54	0.23	-1.38	0.6	0.1	0.1	0.0	1.6	0.56	0.05	0.04	0.01	0.24			
AegonUnionAsegurad96	0.69	5.38	1.86	0.97	-0.62	0.03	0.24	0.5	0.2	0.0	0.0	0.0	0.64	0.17	0.07	0.00	0.01			
CajaMadridVida96	0.69	6.33	-1.98	-0.23	-0.25	0.24	-1.18	0.6	0.0	0.0	0.0	1.2	0.62	0.01	0.01	0.01	0.22			
Ocaso96	0.69	15.10	2.38	0.39	-0.57	-2.41	-1.23	0.8	0.0	0.2	3.5	1.3	0.37	0.01	0.02	0.39	0.10			
PlusUltra96	0.69	3.37	1.48	-0.83	-0.10	0.16	-0.131	0.3	0.2	0.0	0.0	0.1	0.65	0.20	0.00	0.01	0.03			
Adelas96	0.69	39.39	0.09	5.28	3.14	0.66	-0.40	0.0	7.3	4.6	0.3	0.1	0.00	0.71	0.25	0.01	0.00			
Sanitas96	0.69	27.61	0.25	4.22	2.15	0.84	-1.41	0.0	4.7	2.2	0.4	1.7	0.00	0.64	0.17	0.03	0.07			
DbVida96	0.69	6.08	-1.69	-0.63	-0.53	0.45	-1.29	0.4	0.1	0.1	0.1	1.4	0.47	0.07	0.05	0.03	0.27			
Santander96	0.69	6.59	-1.43	-0.80	-0.33	0.86	-1.46	0.3	0.2	0.1	0.4	1.8	0.31	0.10	0.02	0.11	0.32			
AssicurazioniGenerali96	0.69	5.31	1.87	-0.65	0.17	0.27	0.48	0.5	0.1	0.0	0.0	0.2	0.66	0.08	0.01	0.01	0.04			
ArgentinaVida96	0.69	13.31	-3.27	-0.37	0.51	-0.58	1.30	1.5	0.0	0.1	0.2	1.4	0.80	0.01	0.02	0.03	0.13			
Banesto96	0.69	19.85	-3.72	-0.41	0.90	-0.72	1.56	2.0	0.0	0.4	0.3	2.0	0.70	0.01	0.04	0.03	0.12			
ZurichSegEspana96	0.69	10.39	1.65	0.95	-1.09	-0.49	1.49	0.4	0.2	0.6	0.1	1.8	0.26	0.09	0.11	0.02	0.21			
SegurosGenesis96	0.69	6.23	-1.26	-0.78	-0.11	0.94	-1.13	0.2	0.2	0.0	0.5	1.1	0.26	0.10	0.00	0.14	0.21			
BiharkoVida96	0.69	8.49	-2.80	-0.52	0.00	-0.11	-0.28	1.1	0.1	0.0	0.0	0.1	0.92	0.03	0.00	0.00	0.01			
FiaactMutua96	0.69	7.32	1.60	1.50	-0.08	0.03	1.12	0.4	0.6	0.0	0.0	1.0	0.35	0.31	0.00	0.00	0.17			
BilbaoCiaAnma96	0.69	4.61	1.22	0.08	-0.17	-0.80	-0.68	0.2	0.0	0.6	0.4	0.4	0.33	0.00	0.30	0.14	0.10			
AseguradValenciana96	0.69	6.30	-2.18	-0.39	0.19	0.22	-0.50	0.7	0.0	0.0	0.0	0.2	0.75	0.02	0.01	0.01	0.04			
MultinacionalAsegurad96	0.69	9.32	1.82	1.18	-1.68	0.42	0.53	0.5	0.4	1.3	0.1	0.2	0.36	0.15	0.30	0.02	0.03			
WinterthurVida96	0.69	5.02	-1.42	-0.53	-0.57	0.32	-1.39	0.3	0.1	0.2	0.1	1.6	0.40	0.06	0.07	0.02	0.39			
CreditoCaucion96	0.69	63.96	5.20	-4.41	3.85	0.74	0.04	3.9	5.1	6.9	0.3	0.0	0.42	0.30	0.23	0.01	0.00			
Athena96	0.69	7.91	1.52	0.39	-1.79	0.20	-0.30	0.3	0.0	1.5	0.0	0.1	0.29	0.02	0.40	0.01	0.01			
BankinterVida96	0.69	26.45	-4.13	-0.77	0.54	-0.85	1.84	2.5	0.2	0.1	0.4	2.8	0.65	0.02	0.01	0.03	0.13			
MutuaGeneralSegur96	0.69	7.73	2.23	-0.33	-0.70	0.38	-0.21	0.7	0.0	0.2	1.0	0.0	0.64	0.01	0.06	0.02	0.01			
PelayoMutua96	0.69	10.77	1.34	0.28	-1.91	1.78	0.18	0.3	0.0	1.7	1.9	0.0	0.17	0.01	0.34	0.29	0.00			
Musini96	0.69	31.62	1.58	-3.44	2.15	2.23	-1.25	0.4	3.1	2.2	3.0	1.3	0.08	0.37	0.15	0.16	0.05			
LagunAroVida96	0.69	9.56	-2.94	-0.82	0.11	-0.24	0.09	1.2	0.2	0.0	0.0	0.0	0.90	0.07	0.00	0.01	0.00			
BansabadellVida96	0.69	14.98	-2.77	0.13	0.23	-0.33	1.62	1.1	0.0	0.0	0.1	2.2	0.51	0.00	0.00	0.01	0.17			
Finisterre96	0.69	25.15	2.55	1.54	-0.20	-3.79	-0.35	0.9	0.6	0.0	8.7	0.1	0.26	0.09	0.00	0.57	0.00			
ZurichInternacional96	0.69	31.93	4.61	-2.15	1.74	-0.08	0.29	3.1	1.2	1.4	0.0	0.1	0.67	0.15	0.09	0.00	0.00			

COORDENADAS, CONTRIBUCIONES Y CONSENSOS CUADRADOS DE LOS INDIVIDUOS EN LOS PRIMEROS CINCO EJES (Continuación)

INDIVIDUOS	COORDENADAS					CONTRIBUCIONES					COSENSOS CUADRADOS						
	P. REL	DISTO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Euroseguros97	0.69	7.90	-2.66	-0.33	0.20	-0.12	0.43	1.0	0.0	0.0	0.0	0.2	0.89	0.01	0.00	0.00	0.02
MapfireMutualidad97	0.69	27.41	2.94	1.37	-1.61	2.27	1.95	1.2	0.5	1.2	3.1	3.2	0.31	0.07	0.09	0.19	0.14
VidaCaixa97	0.69	8.36	-2.36	-0.66	-0.16	0.01	-0.61	0.8	0.1	0.0	0.0	0.3	0.66	0.05	0.00	0.00	0.04
Casergupo97	0.69	2.41	-0.66	-0.03	-0.20	-0.22	-0.25	0.1	0.0	0.0	0.0	0.1	0.18	0.00	0.02	-0.02	0.03
Vitalicio97	0.69	3.43	1.07	-1.13	-0.16	0.57	-0.43	0.2	0.3	0.0	0.2	0.2	0.33	0.38	0.01	0.09	0.05
BarclaysVida97	0.69	26.83	-4.24	-0.72	0.76	-1.17	2.26	2.6	0.1	0.3	0.8	4.2	0.67	0.02	0.02	0.05	0.19
WinterthurSegGrales97	0.69	12.68	2.45	1.27	-1.00	-0.64	1.22	0.9	0.4	0.5	0.2	1.2	0.47	0.13	0.08	0.03	0.12
AllianzRas97	0.69	3.29	0.62	-0.36	-0.85	0.51	0.27	0.1	0.0	0.3	0.2	0.1	0.12	0.04	0.22	0.08	0.02
MapfreVida97	0.69	7.67	-0.42	-0.76	-0.15	1.13	-1.15	0.0	0.1	0.0	0.8	1.1	0.02	0.07	0.00	0.17	0.17
AntaresVida97	0.69	32.03	-4.21	-0.52	1.27	-1.25	2.68	2.5	0.1	0.8	0.9	6.0	0.55	0.01	0.05	0.05	0.22
Stalucia97	0.69	17.83	1.68	0.03	-0.85	-3.07	-1.47	0.4	0.0	0.3	5.7	1.8	0.16	0.00	0.04	0.53	0.12
MutuaMadi97	0.69	12.03	0.41	0.65	-1.97	2.15	0.81	0.0	0.1	1.8	2.8	0.6	0.01	0.04	0.32	0.39	0.05
AGUnionFenix97	0.69	7.82	2.22	-0.71	-0.45	-0.14	-0.30	0.7	0.1	0.1	0.0	0.1	0.63	0.06	0.03	0.00	0.01
CatalaOccident97	0.69	4.74	1.56	-0.59	-0.67	0.42	-0.50	0.3	0.1	0.2	0.1	0.2	0.51	0.07	0.10	0.04	0.05
Asisa97	0.69	32.12	-0.18	4.80	2.64	0.89	-0.39	0.0	6.1	3.3	0.5	0.1	0.00	0.72	0.22	0.02	0.00
AscatVida97	0.69	13.34	-3.40	-0.62	0.37	-0.50	0.79	1.7	0.1	0.1	0.2	0.5	0.87	0.03	0.01	0.02	0.05
MapfreSegrales97	0.69	11.14	2.82	-0.21	0.45	-1.02	0.38	1.1	0.0	0.1	0.6	0.1	0.71	0.00	0.02	0.09	0.01
LaEstrella97	0.69	1.76	0.24	-0.51	-0.59	-0.11	-0.36	0.0	0.1	0.2	0.0	0.1	0.03	0.15	0.20	0.01	0.07
Intercases97	0.69	7.63	-2.06	-0.67	-0.48	0.36	-1.39	0.6	0.1	0.1	0.1	1.6	0.56	0.06	0.03	0.02	0.25
AegonUnionAsegurad97	0.69	2.65	1.13	0.56	-0.71	0.03	-0.21	0.2	0.1	0.2	0.0	0.0	0.48	0.12	0.19	0.00	0.02
CajaMadridVida97	0.69	8.35	-2.74	-0.17	0.25	-0.11	-0.23	1.1	0.0	0.0	0.0	0.0	0.90	0.00	0.01	0.00	0.01
Ocaso97	0.69	13.69	2.33	0.36	-0.55	-2.25	-1.08	0.8	0.0	0.1	3.1	1.0	0.40	0.01	0.02	0.37	0.09
PlusUltra97	0.69	2.17	1.10	-0.69	-0.18	0.17	-0.44	0.2	0.1	0.0	0.0	0.2	0.56	0.22	0.02	0.01	0.09
Adelas97	0.69	43.93	0.10	5.78	3.10	0.57	-0.35	0.0	8.8	4.5	0.2	0.1	0.00	0.76	0.22	0.01	0.00
Sanitas97	0.69	30.60	0.61	4.68	2.23	0.34	-1.25	0.1	5.8	2.3	0.1	1.3	0.01	0.72	0.16	0.00	0.05
DbVida97	0.69	6.75	-2.38	-0.69	0.07	0.24	-0.29	0.8	0.1	0.0	0.0	0.1	0.84	0.07	0.00	0.01	0.01
Santander97	0.69	8.65	-2.78	-0.26	0.29	-0.10	0.50	1.1	0.0	0.0	0.0	0.2	0.90	0.01	0.01	0.00	0.03
AssicurazioniGenerali97	0.69	2.18	1.19	-0.55	-0.15	-0.03	0.21	0.2	0.1	0.0	0.0	0.0	0.65	0.14	0.01	0.00	0.02
ArgentariaVida97	0.69	7.39	-2.18	-0.18	0.04	0.15	0.23	0.7	0.0	0.0	0.0	0.0	0.64	0.00	0.00	0.00	0.01
Baesto97	0.69	7.88	-2.59	-0.75	0.18	0.17	-0.35	1.0	0.1	0.0	0.0	0.1	0.85	0.07	0.00	0.00	0.02
ZurichSegEspaña97	0.69	11.62	1.99	0.57	-0.84	-0.01	1.35	0.6	0.1	0.3	0.0	1.5	0.34	0.03	0.06	0.00	0.16
SegurosGenesis97	0.69	5.52	-1.61	-0.38	-0.45	0.33	-1.25	0.4	0.0	0.1	0.1	1.3	0.47	0.03	0.04	0.02	0.28
BiharkoVida97	0.69	10.73	-3.14	-0.35	-0.10	-0.31	0.08	1.4	0.0	0.0	0.1	0.0	0.92	0.01	0.00	0.01	0.00
FiactoMutua97	0.69	7.11	1.75	1.49	-0.06	0.12	0.85	0.4	0.6	0.0	0.0	0.6	0.43	0.31	0.00	0.00	0.10
BilbaoCiaAnma97	0.69	4.42	1.27	-0.13	-0.97	-0.75	-0.71	0.2	0.0	0.4	0.3	0.4	0.36	0.00	0.21	0.13	0.11
AseguradValenciana97	0.69	8.22	-2.57	-0.35	0.28	-0.06	-0.12	1.0	0.0	0.0	0.0	0.0	0.81	0.02	0.01	0.00	0.00
MultinacionalAsegurad97	0.69	11.93	1.99	1.27	-1.93	0.55	0.53	0.6	0.4	1.7	0.2	0.2	0.33	0.13	0.31	0.03	0.02
WinterthurVida97	0.69	5.09	-1.04	-0.47	-0.77	0.20	-1.47	0.2	0.1	0.3	0.0	1.8	0.21	0.04	0.12	0.01	0.43
CreditoCaucion97	0.69	61.34	5.21	-4.36	3.74	0.76	0.20	3.9	5.0	6.5	0.4	0.0	0.44	0.31	0.23	0.01	0.00
Athena97	0.69	6.30	1.22	0.28	-1.59	0.43	-0.26	0.2	0.0	1.2	0.1	0.1	0.24	0.01	0.40	0.03	0.01
BankinterVida97	0.69	24.98	-3.88	-0.90	0.39	-0.50	1.25	2.2	0.2	0.1	0.2	1.3	0.60	0.03	0.01	0.01	0.06
MutuaGeneralSegur97	0.69	7.06	2.09	-0.41	-0.63	0.36	-0.28	0.6	0.0	0.2	0.1	0.1	0.62	0.02	0.06	0.02	0.01
PelayoMutua97	0.69	14.46	1.45	0.29	-2.06	2.38	0.51	0.3	0.0	2.0	3.4	0.2	0.15	0.01	0.29	0.39	0.02
Musini97	0.69	40.07	3.17	-4.03	3.14	1.18	-0.39	1.4	4.3	4.6	0.8	0.1	0.25	0.40	0.25	0.03	0.00
LagunAroVida97	0.69	8.04	-2.40	-0.71	-0.29	0.06	-0.99	0.8	0.1	0.0	0.0	0.8	0.72	0.06	0.01	0.00	0.12
BansabadellVida97	0.69	8.74	-1.79	0.08	-0.28	0.25	0.05	0.5	0.0	0.0	0.0	0.0	0.37	0.00	0.01	0.01	0.00
Finsterre97	0.69	25.60	2.58	1.56	-0.26	-3.74	-0.26	1.0	0.6	0.0	8.5	0.1	0.26	0.10	0.00	0.55	0.00
ZurichInternacional97	0.69	25.24	4.12	-1.86	1.57	-0.49	0.23	2.4	0.9	1.2	0.1	0.0	0.67	0.14	0.10	0.01	0.00

Figura a
Los individuos en el año 1995
Factores 1 y 2

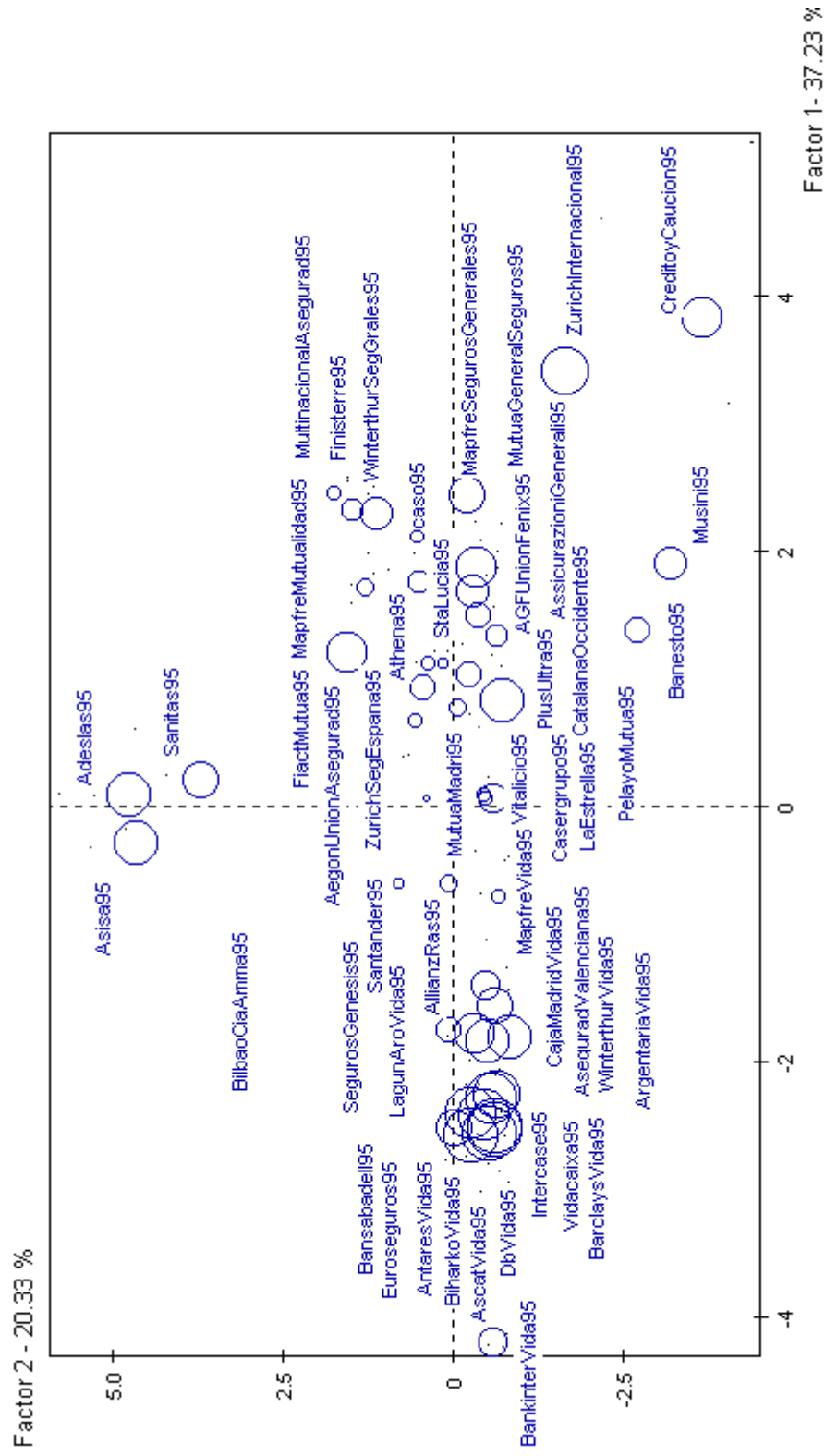


Figura b
Los individuos en el año 1996
Factores 1 y 2

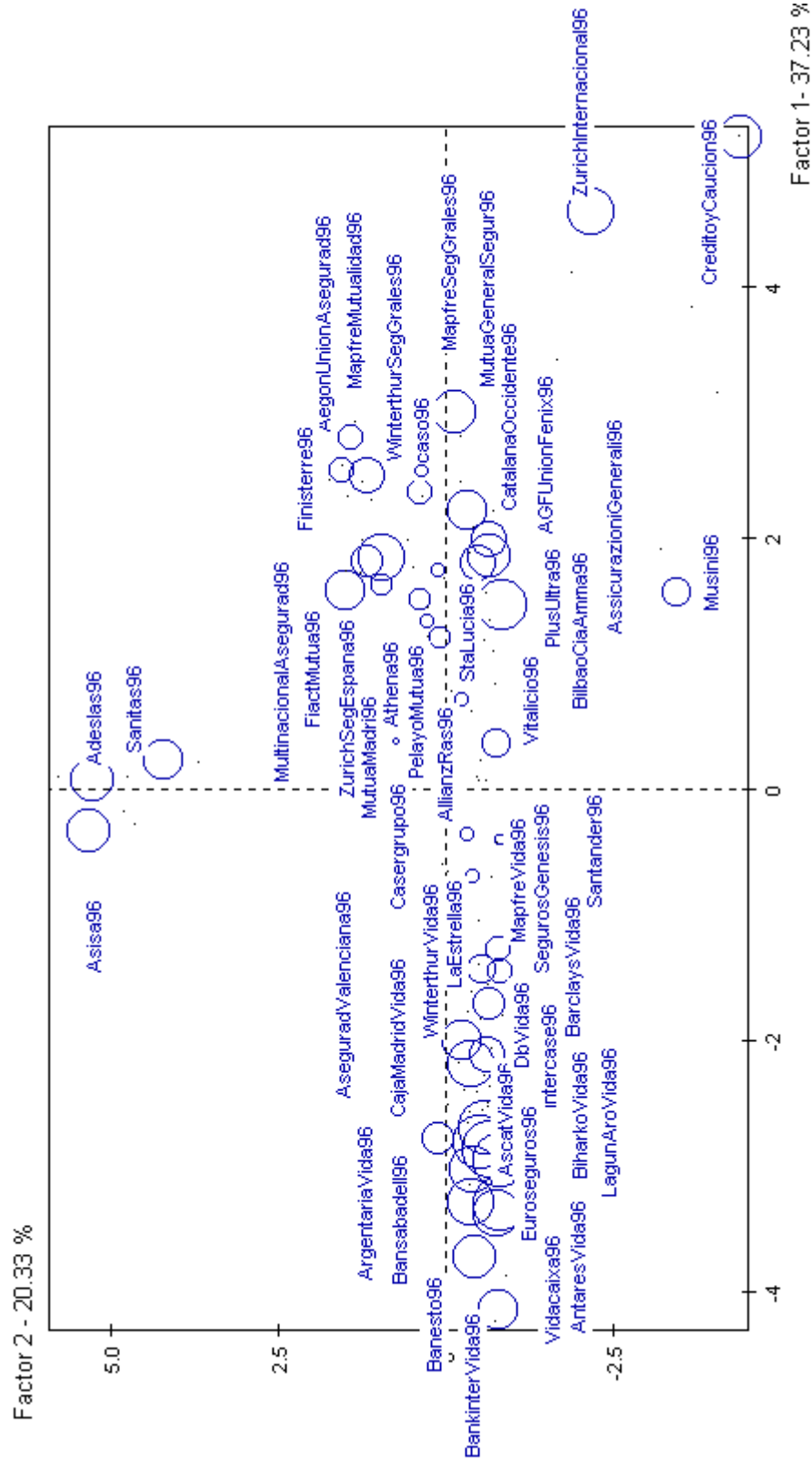


Figura c
Los individuos en el año 1997
Factores 1 y 2

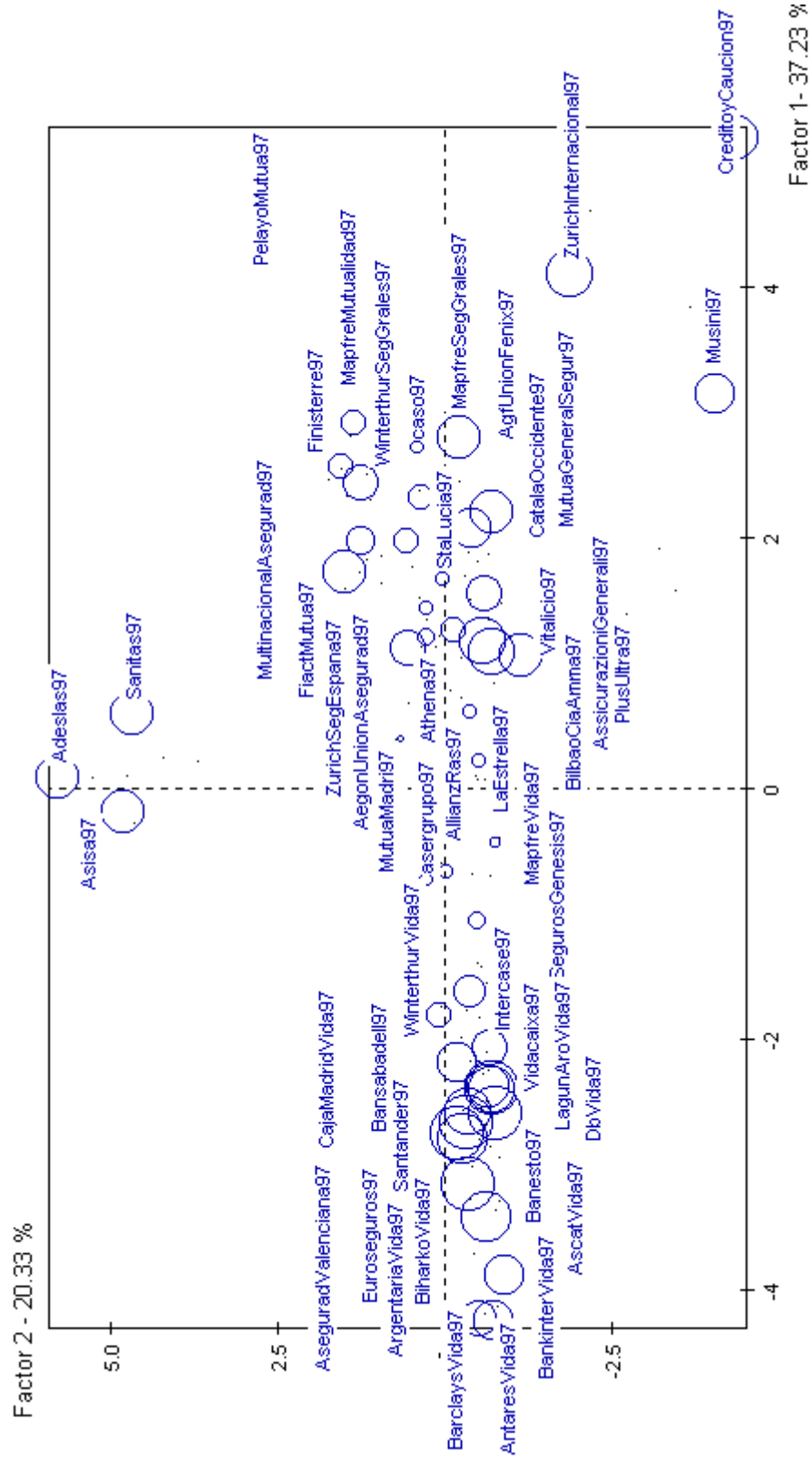


Figura d
Trayectorias de los individuos
Factores 2 y 3

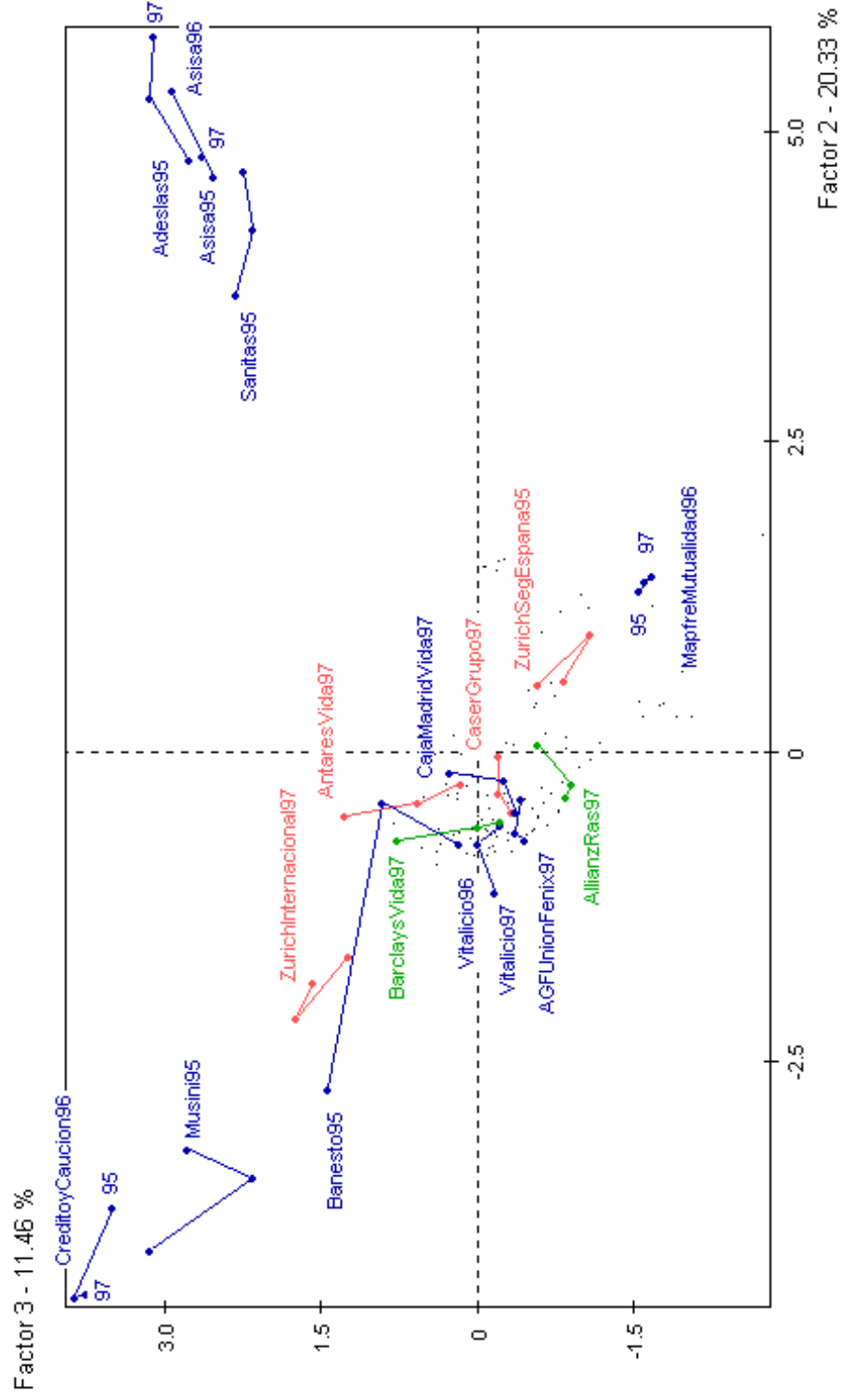


Figura e
Trayectorias de los individuos
Factores 2 y 3

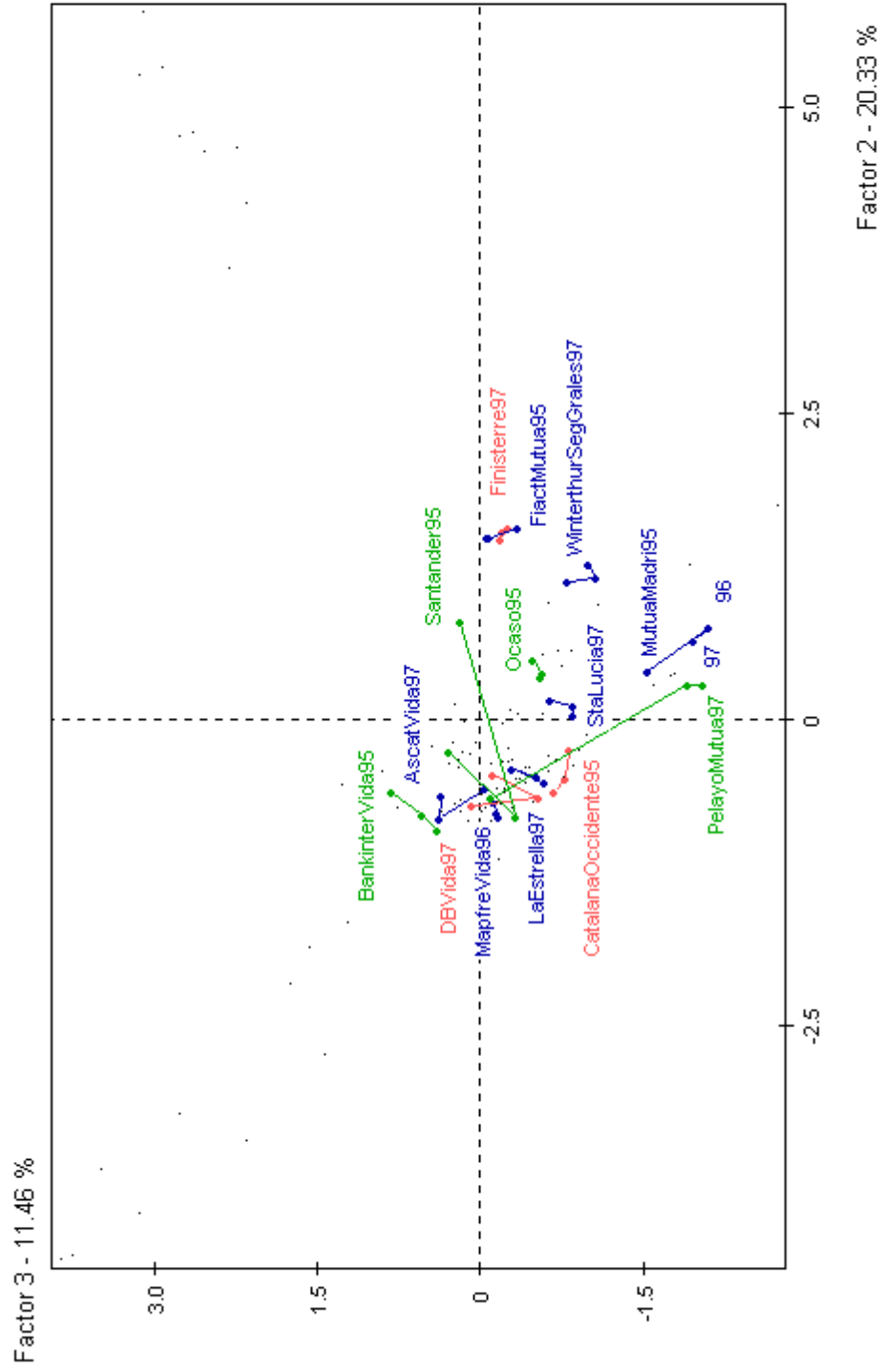


Figura f
Trayectorias de los individuos
Factores 2 y 3

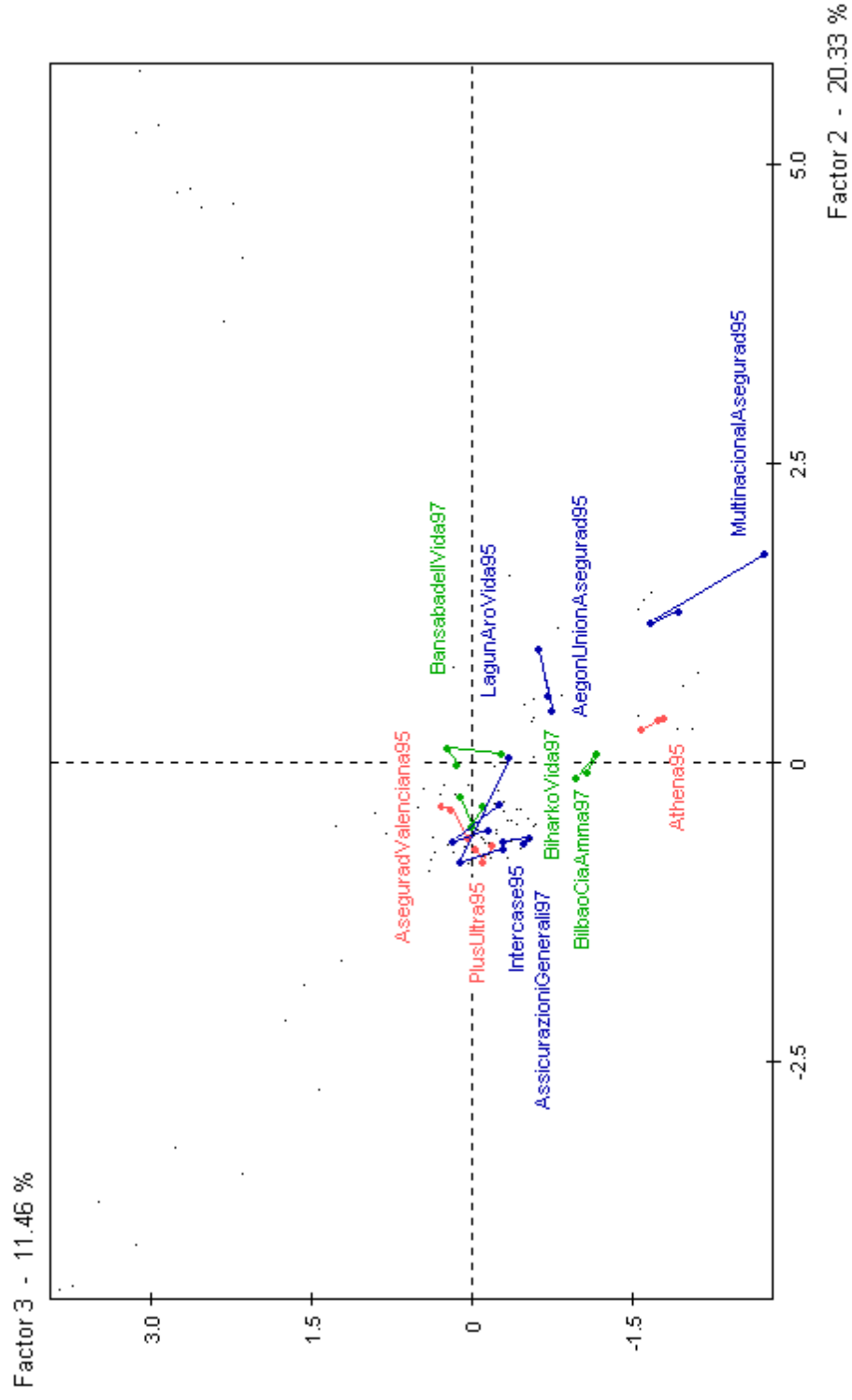
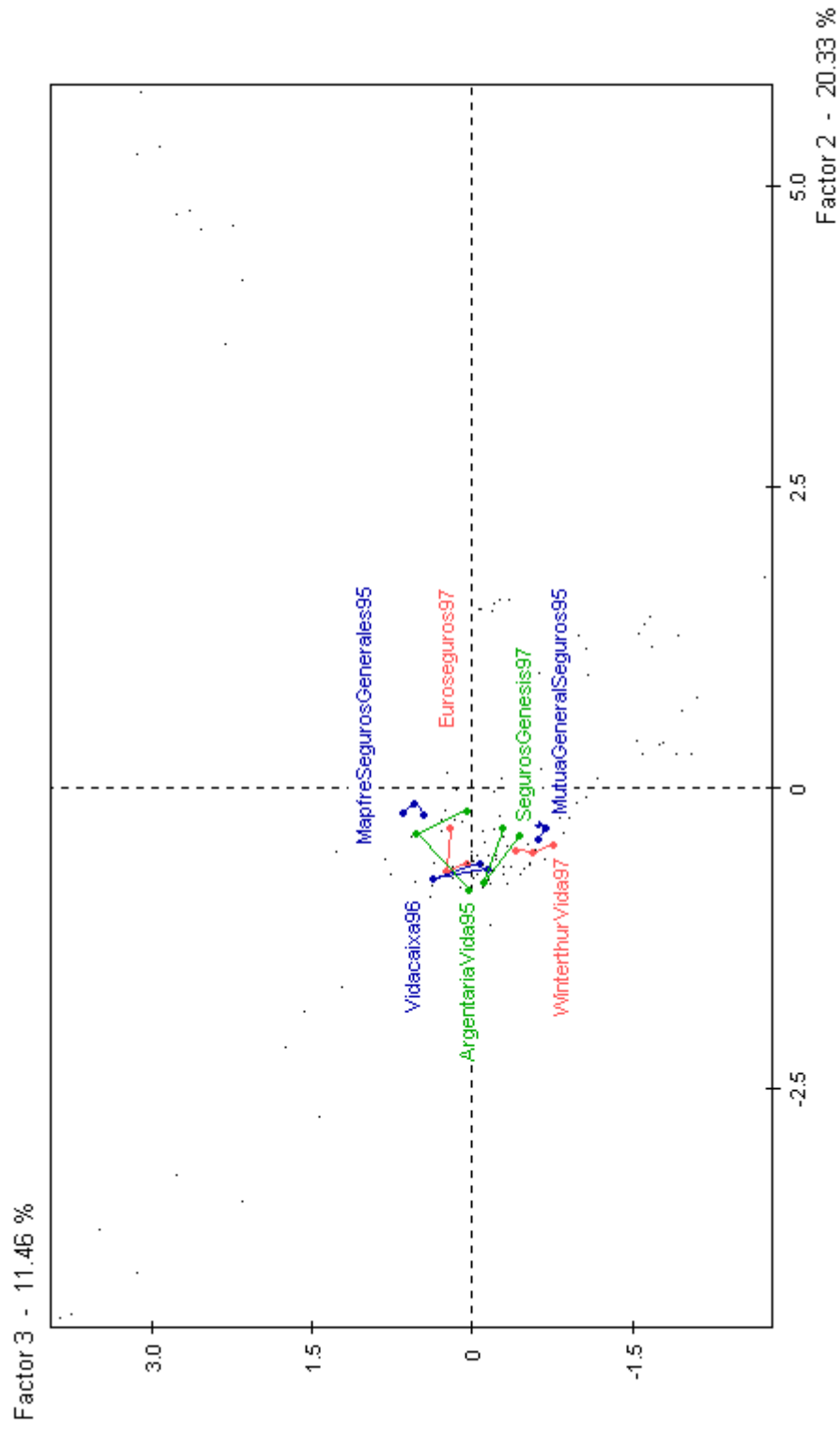


Figura g
Trayectorias de los individuos
Factores 2 y 3



A.2.3 Análisis Cluster

Figura h
Dendograma con dos factores

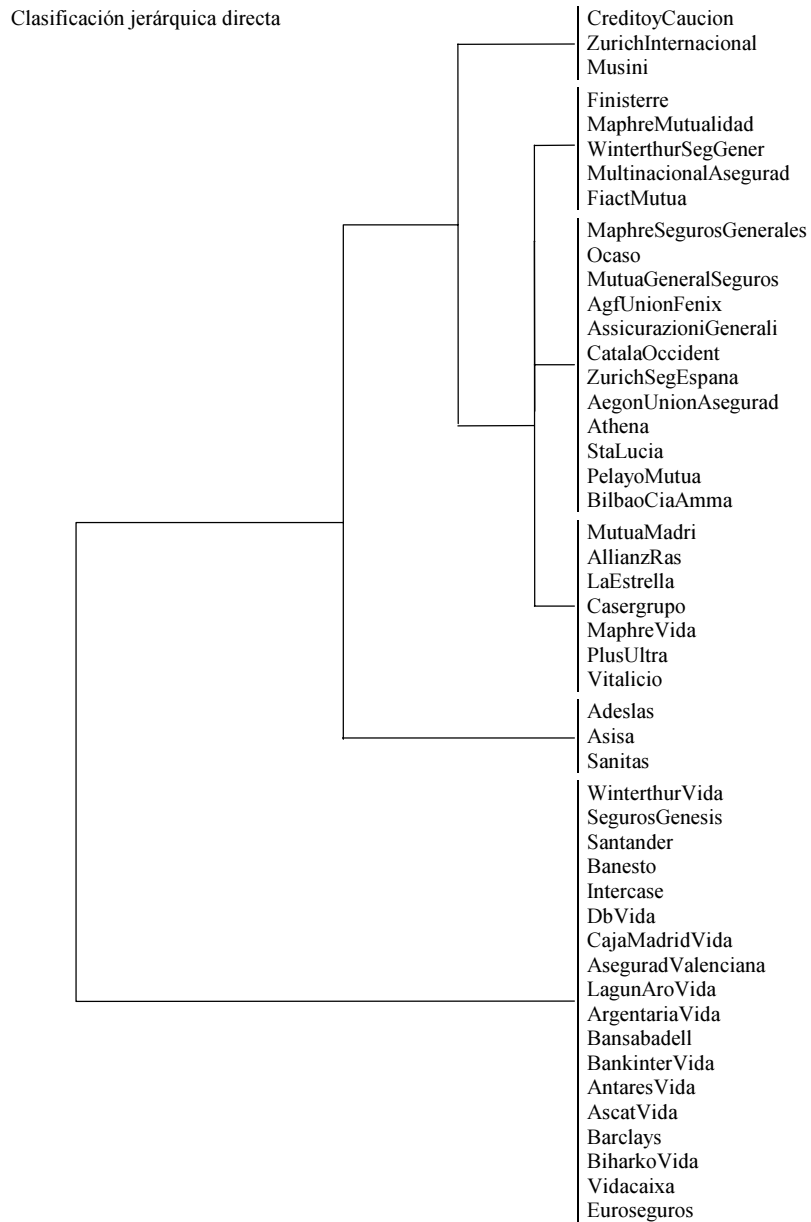


Figura i
Dendrograma con tres factores

Clasificación jerárquica directa

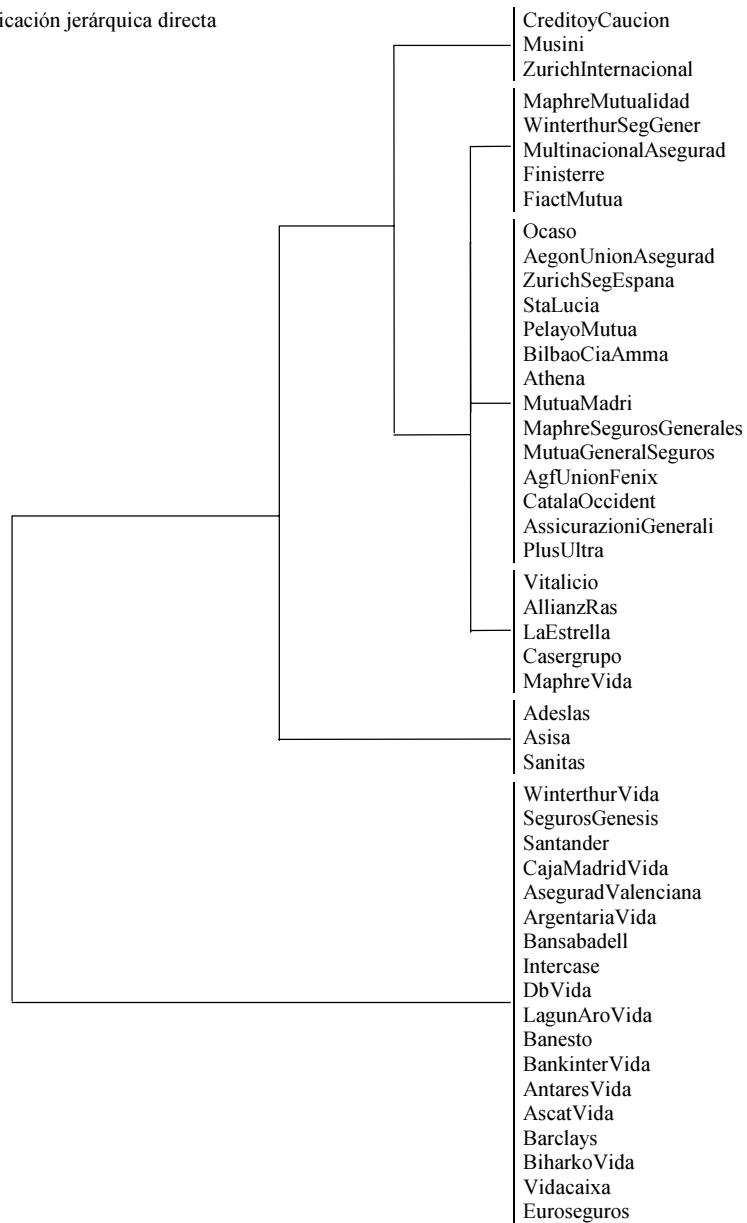
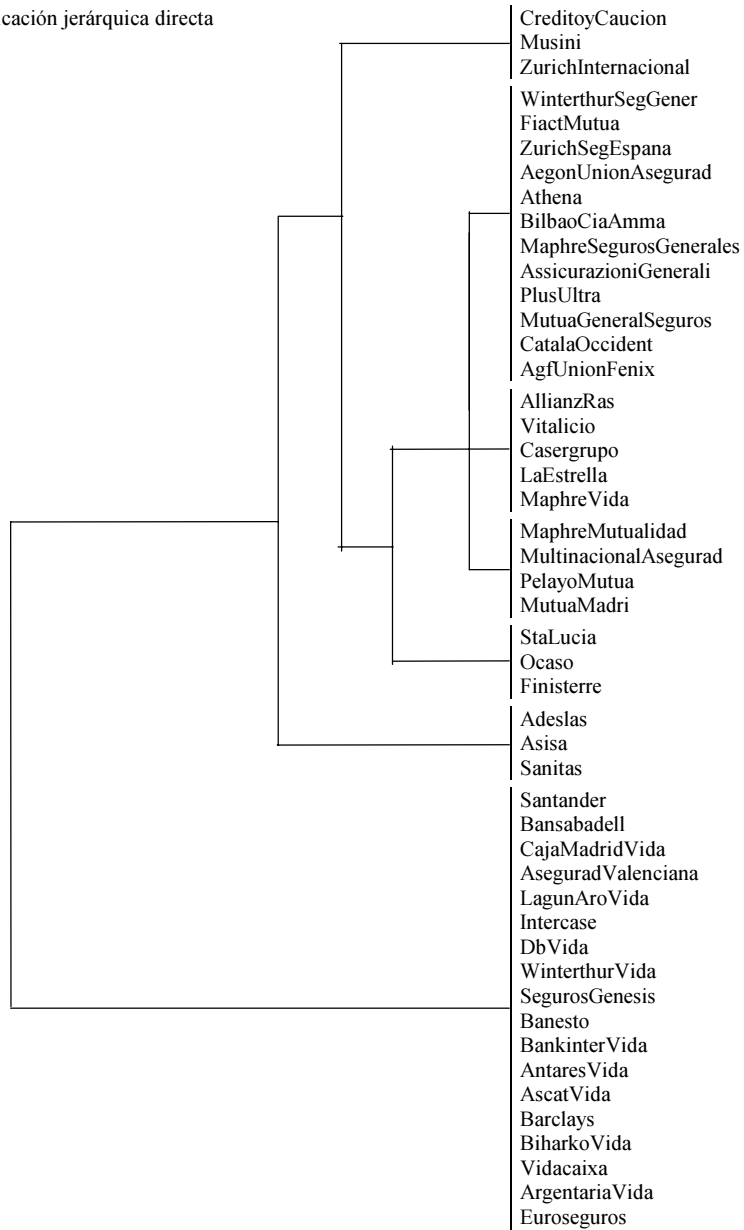


Figura j
Dendograma con cuatro factores

Clasificación jerárquica directa



A.3 DATOS DEL SECTOR ASEGURADOR ANALIZADOS DESDE EL PUNTO DE VISTA PREDICTIVO³

Tabla a
Variables relativas a los seguros de vida (UNESPA, 1998a; UNESPA, 1999a)

AÑOS	Vlequi	Vgastexter	Vgastinter	Vipib	Viprim	Vsinies
87	113,03	2,66	4,87	1,26	37,47	105,50
88	112,41	2,48	3,54	2,02	48,45	106,38
89	101,21	5,03	7,85	0,99	29,77	88,32
90	125,29	4,66	8,81	0,93	27,44	111,82
91	116,56	5,36	8,24	1,20	31,21	102,95
92	114,24	4,58	7,84	1,38	33,22	101,82
93	118,23	5,62	7,78	1,44	32,85	104,83
94	113,09	3,89	5,09	2,23	42,96	104,11
95	113,53	4,26	6,06	1,97	40,58	103,21
96	121,71	4,20	5,34	2,26	43,57	112,14
97	128,46	3,85	5,31	2,53	46,99	119,30

Tabla b
Variables relativas a los seguros de automóviles (UNESPA, 1998a; UNESPA, 1999a)

AÑOS	AUequi	AUgastexter	AUgastinter	AUpib	AUprim	AUsinies
87	92,22	13,93	17,14	1,06	31,57	61,15
88	109,13	16,00	18,08	1,05	25,35	75,05
89	108,89	14,94	16,62	1,22	36,63	77,33
90	106,80	13,66	16,56	1,27	37,60	76,58
91	107,88	12,96	16,36	1,38	35,95	78,56
92	102,82	12,95	16,51	1,42	34,21	73,36
93	104,00	13,08	16,57	1,53	34,84	74,35
94	103,92	12,51	16,96	1,51	28,98	74,45
95	102,88	11,61	16,79	1,42	29,18	74,48
96	103,62	12,50	16,77	1,39	26,66	74,35
97	107,51	12,62	17,12	1,29	23,97	77,77

³ En primer lugar se recogen los datos analizados para los ramos agregados considerados. Los siguientes epígrafes están ordenados según las cuatro fases descritas en la parte teórica. La primera fase está calculada con el módulo TM del SPAD (versión 3.5) y con el Módulo Factorial del mismo programa (excepto el apartado A.3.1.4 que se ha llevado a cabo con la hoja de cálculo EXCEL). La segunda fase, es decir, la modelización *ARIMA* de los factores se lleva a cabo con el SPSS (versión 7.5. 2S). En la tercera fase se ha empleado de nuevo la hoja de cálculo EXCEL para la obtención de los valores predichos a partir de la fórmula de reconstitución de los datos. El análisis de los errores, dado que tienen naturaleza estocástica, se ha llevado a cabo también con el SPSS. Para el cálculo, en la cuarta fase, de las predicciones se ha vuelto a emplear la hoja EXCEL.

Las diferencias en la notación decimal (punto o coma, según el caso) se deben a las diferencias en los diferentes programas empleados: mientras el SPAD emplea el punto el SPSS y la hoja de cálculo EXCEL emplea la coma.

Tabla c
Variables relativas a los seguros de salud (UNESPA, 1998a; UNESPA, 1999a)

AÑOS	SAequi	SAgastexter	SAgastinter	SApib	SAprim	SAsinies
87	98,28	4,03	13,38	0,46	13,69	80,87
88	96,28	4,55	11,78	0,47	11,35	79,95
89	102,42	4,67	14,12	0,49	14,88	83,63
90	91,60	3,81	12,34	0,52	15,45	75,44
91	98,27	3,67	13,11	0,64	16,65	81,49
92	96,55	3,45	12,41	0,71	17,04	80,69
93	99,33	3,33	13,67	0,77	17,58	82,33
94	98,02	3,60	11,63	0,79	15,21	82,80
95	99,38	3,54	12,01	0,80	16,46	83,83
96	100,22	3,41	11,61	0,83	16,03	85,20
97	98,20	3,38	12,53	0,88	16,30	82,29

Tabla d
Variables relativas al resto de seguros (UNESPA, 1998a; UNESPA, 1999a)

AÑOS	REequi	REgastexter	REgastinter	REpib	REprim	RESinies
87	94,02	28,64	21,15	0,58	17,27	44,23
88	87,80	28,59	22,13	0,62	14,85	37,08
89	102,08	29,36	22,02	0,62	18,72	50,69
90	103,76	27,46	21,02	0,66	19,51	55,28
91	106,20	30,56	22,94	0,62	16,19	52,70
92	111,15	32,57	22,83	0,64	15,52	55,75
93	109,60	35,40	23,61	0,65	14,73	50,59
94	105,77	33,48	23,83	0,67	12,85	48,46
95	111,84	34,53	23,83	0,67	13,79	53,48
96	99,96	31,81	21,14	0,71	13,74	47,02
97	104,78	33,84	20,85	0,69	12,74	50,10

A.3.1 Primera fase

A.3.1.1 Medias y varianzas de las variables analizadas

VARIABLES	MEDIA	VARIANZA
VIequi	116,16	53,7289
VIgastexter	4,24	0,9946
VIgastinter	6,43	2,9929
VIpib	1,65	0,3158
VIprim	37,68	52,6046
VIsinies	105,49	59,8070
AUequi	104,52	22,4089
AUgastexter	13,34	1,5356
AUgastinter	16,86	0,2262
AUpib	1,32	0,0255
AUprim	31,36	23,1323
AUsinies	74,31	21,8996
SAequi	98,05	7,4387
SAgastexter	3,77	0,2164
SAgastinter	12,60	0,7403
SApib	0,67	0,0248
SAprim	15,51	3,0800
SAsinies	81,68	6,6373
REequi	103,36	53,0086
REgastexter	31,48	7,3425
REgastinter	22,30	1,3610
REpib	0,65	0,0014
REprim	15,45	5,1647
REsinies	49,58	29,0640

A.3.1.2 Comparación de las nubes de individuos (Método Statist)

NUBE 1 - VI
 REPARTO DE LA INERCIA SEGUN LOS EJES PRINCIPALES

Traza de la matriz = 1.4617
 Histograma de los 10 primeros valores propios

NUMERO	VALOR PROPIO	PORCENT.	PORCENT. ACUMUL.
1	0.8968	61.35	61.35
2	0.4214	28.83	90.19
3	0.1344	9.20	99.38
4	0.0063	0.43	99.82
5	0.0027	0.18	100.00
6	0.0000	0.00	100.00
7	0.0000	0.00	100.00
8	0.0000	0.00	100.00
9	0.0000	0.00	100.00
10	0.0000	0.00	100.00

NUBE 2 - AU
 REPARTO DE LA INERCIA SEGUN LOS EJES PRINCIPALES

Traza de la matriz = 1.7296
 Histograma de los 10 primeros valores propios

NUMERO	VALOR PROPIO	PORCENT.	PORCENT. ACUMUL.
1	0.7473	43.21	43.21
2	0.5506	31.84	75.04
3	0.3680	21.28	96.32
4	0.0529	3.06	99.38
5	0.0108	0.62	100.00
6	0.0000	0.00	100.00
7	0.0000	0.00	100.00
8	0.0000	0.00	100.00
9	0.0000	0.00	100.00
10	0.0000	0.00	100.00

NUBE 3 - SA
 REPARTO DE LA INERCIA SEGUN LOS EJES PRINCIPALES

Traza de la matriz = 1.6595
 Histograma de los 10 primeros valores propios

NUMERO	VALOR PROPIO	PORCENT.	PORCENT. ACUMUL.
1	0.7995	48.18	48.18
2	0.5201	31.34	79.52
3	0.2990	18.02	97.54
4	0.0245	1.47	99.01
5	0.0164	0.99	100.00
6	0.0000	0.00	100.00
7	0.0000	0.00	100.00
8	0.0000	0.00	100.00
9	0.0000	0.00	100.00
10	0.0000	0.00	100.00

NUBE 4 - RE
 REPARTO DE LA INERCIA SEGUN LOS EJES PRINCIPALES

Traza de la matriz = 1.6501
 Histograma de los 10 primeros valores propios

NUMERO	VALOR PROPIO	PORCENT.	PORCENT. ACUMUL.
1	0.8543	51.77	51.77
2	0.4269	25.87	77.64
3	0.2895	17.55	95.18
4	0.0625	3.79	98.97
5	0.0170	1.03	100.00
6	0.0000	0.00	100.00
7	0.0000	0.00	100.00
8	0.0000	0.00	100.00
9	0.0000	0.00	100.00
10	0.0000	0.00	100.00

A.3.1.3 Situación de las nubes de individuos unas respecto de otras

PRODUCTOS ESCALARES ENTRE NUBES NORMADAS
Coeficientes RV entre matrices

Matr	VI	AU	SA	RE
VI	1.000			
AU	0.554	1.000		
SA	0.434	0.529	1.000	
RE	0.489	0.675	0.677	1.000

TEST DE PERMUTACIONES

* indica que las nubes son similares

Matr	VI	AU	SA	RE
VI	*			
AU	*	*		
SA		*	*	
RE	*	*	*	*

INTERPRETACION DE LAS DISTANCIAS ENTRE NUBES

DISTANCIA EUCLIDEA ENTRE NUBES

Matr	VI	AU	SA	RE
1	0.000			
2	0.945	0.000		
3	1.064	0.971	0.000	
4	1.011	0.806	0.804	0.000

DESCOMPOSICION DE LAS DISTANCIAS AL CUADRADO SEGUN LOS INDIVIDUOS

- ENTRE LA NUBE VI Y LAS OTRAS

LINEA	MEDIA	NUBE VI		
		AU	SA	RE
87	13.1	23.4	4.5	11.3
88	18.5	19.8	21.1	14.7
89	16.4	17.8	14.2	17.2
90	9.9	4.9	16.9	7.8
91	3.8	3.2	4.3	3.8
92	4.0	2.6	5.2	4.2
93	5.6	2.1	5.0	9.5
94	5.1	5.7	2.7	7.0
95	7.2	4.8	3.9	12.9
96	4.6	4.4	7.5	1.9
97	11.9	11.4	14.7	9.7

- ENTRE LA NUBE AU Y LAS OTRAS

LINEA	MEDIA	NUBE AU		
		VI	SA	RE
87	27.8	23.4	27.4	32.6
88	15.0	19.8	12.5	12.6
89	12.8	17.8	14.4	6.2
90	12.3	4.9	21.5	10.4
91	5.3	3.2	5.0	7.8
92	2.5	2.6	2.3	2.6
93	4.0	2.1	3.2	6.5
94	3.3	5.7	0.9	3.3
95	4.8	4.8	1.4	8.2
96	4.5	4.4	4.3	4.9
97	7.8	11.4	7.0	4.9

- ENTRE LA NUBE SA Y LAS OTRAS

LINEA	MEDIA	NUBE SA		
		VI	AU	RE
87	13.0	4.5	27.4	7.1
88	15.3	21.1	12.5	12.2
89	16.0	14.2	14.4	19.4
90	20.2	16.9	21.5	22.2
91	3.6	4.3	5.0	1.6
92	4.0	5.2	2.3	4.4
93	5.0	5.0	3.2	6.6
94	2.9	2.7	0.9	5.0
95	4.4	3.9	1.4	8.1
96	7.6	7.5	4.3	10.9
97	8.1	14.7	7.0	2.6

- ENTRE LA NUBE RE Y LAS OTRAS

LINEA	MEDIA	NUBE RE		
		VI	AU	SA
87	17.0	11.3	32.6	7.1
88	13.2	14.7	12.6	12.2
89	14.3	17.2	6.2	19.4
90	13.5	7.8	10.4	22.2
91	4.4	3.8	7.8	1.6
92	3.7	4.2	2.6	4.4
93	7.6	9.5	6.5	6.6
94	5.1	7.0	3.3	5.0
95	9.7	12.9	8.2	8.1
96	5.9	1.9	4.9	10.9
97	5.7	9.7	4.9	2.6

LA SUMA DE LAS DISTANCIAS AL CUADRADO SE DESCOMPONE SEGUN LAS LINEAS

87	= 16.87 %
88	= 15.97 %
89	= 15.08 %
90	= 13.78 %
91	= 4.24 %
92	= 3.63 %
93	= 5.47 %
94	= 4.07 %
95	= 6.38 %
96	= 5.42 %
97	= 9.10 %

A.3.1.4 Estudio de la nube de nubes: interestructura no-centrada

DESCOMPOSICION DE LA INERCIA SEGUN LOS EJES PRINCIPALES

Traza de la matriz = 4
 Histograma de los 4 primeros valores propios

NUMERO	VALOR PROPIO	PORCENTAJE	ACUMULADO
VI	2.6880	67.20	67.20
AU	0.6130	15.32	82.52
SA	0.4303	10.76	93.28
RE	0.2692	6.72	100.00

COORDENADAS, CONTRIBUCIONES Y COSENO CUADRADOS DE LAS NUBES SOBRE LOS EJES PRINCIPALES

NUBE	DISTO	COORDENADAS				CONTRIBUCIONES				COSENO CUADRADOS						
		1	2	3	4	1	2	3	4	1	2	3	4			
VI	1	0.74	0.61	-0.27	0.05	0.00	0.20	0.61	0.17	0.08	0.00	0.55	0.37	0.08	0.04	0.00
AU	1	0.85	0.11	0.46	-0.25	0.00	0.27	0.02	0.45	0.14	0.00	0.72	0.01	0.30	0.07	0.00
SA	1	0.81	-0.41	-0.36	-0.22	0.00	0.24	0.27	0.30	0.12	0.00	0.65	0.17	0.13	0.06	0.00
RE	1	0.88	-0.24	0.12	0.39	0.00	0.28	0.09	0.03	0.16	0.00	0.77	0.06	0.01	0.08	0.00

A.3.1.15 Estudio de la nube de nubes tomando como origen la nube compromiso

CONSTRUCCION DE LA NUBE COMPROMISO = MEDIA PONDERADA DE LAS NUBES CON LOS COEFICIENTES

NUBE VI = 0.275
 NUBE AU = 0.315
 NUBE SA = 0.301
 NUBE RE = 0.326

PRODUCTOS ESCALARES Y DISTANCIAS ENTRE LA NUBE COMPROMISO Y LAS DIFERENTES NUBES

NUBE VI = PRODUCTO ESCALAR = 0.740; DISTANCIA = 0.721
 NUBE AU = PRODUCTO ESCALAR = 0.847; DISTANCIA = 0.553
 NUBE SA = PRODUCTO ESCALAR = 0.808; DISTANCIA = 0.620
 NUBE RE = PRODUCTO ESCALAR = 0.877; DISTANCIA = 0.495

DESCOMPOSICION DE LA INERCIA SEGUN LAS NUBES

NUBE VI = 35.72 %
 NUBE AU = 21.00 %
 NUBE SA = 26.43 %
 NUBE RE = 16.85 %

REPRESENTACION GRAFICA DE LA NUBE DE NUBES (INTERESTRUCTURA CENTRADA)

DESCOMPOSICION DE LA INERCIA SEGUN LOS EJES PRINCIPALES

Traza de la matriz = 1.4554
 Histograma de los 4 primeros valores propios

NUMERO	VALOR PROPIO	PORCENT.	POURCENT. ACUMUL.
1	0.6223	42.76	42.76
2	0.4404	30.26	73.02
3	0.2711	18.63	91.64
4	0.1216	8.36	100.00

COORDENADAS, CONTRIBUCIONES Y COSENOS CUADRADOS DE LAS NUBES SOBRE LOS EJES PRINCIPALES (INTERESTRUCTURA CENTRADA)

NUBE	COORDENADAS				CONTRIBUCIONES				COSENOS CUADRADOS					
	DISTO	1	2	3	0	1	2	3	0	1	2	3	0	
VI	0.52	-0.65	-0.27	0.05	0.13	0.00	0.68	0.17	0.01	0.14	0.00	0.82	0.14	0.00
AU	0.31	-0.11	0.43	-0.28	0.19	0.00	0.02	0.41	0.28	0.28	0.00	0.04	0.59	0.25
SA	0.38	0.36	-0.42	-0.23	0.15	0.00	0.21	0.40	0.20	0.19	0.00	0.34	0.46	0.14
RE	0.25	0.23	0.09	0.37	0.22	0.00	0.08	0.02	0.51	0.39	0.00	0.21	0.03	0.56

A.3.1.6 Estudio de la nube comprometido

DESCOMPOSICION DE LA INERCIA SEGUN LAS LINEAS

87	=	15.21 %
88	=	20.80 %
89	=	12.85 %
90	=	12.67 %
91	=	4.03 %
92	=	3.30 %
93	=	6.08 %
94	=	4.55 %
95	=	5.19 %
96	=	6.64 %
97	=	8.69 %

REPRESENTACION GRAFICA DE LA NUBE COMPROMISO
 DESCOMPOSICION DE LA INERCIA SEGUN LOS EJES PRINCIPALES

Traza de la matriz: 1.9852
 Histograma de los 10 primeros valores propios

NUMERO	VALOR PROPIO	PORCENT.	PORCENT. ACUMUL.
1	0.7331	36.93	36.93
2	0.5815	29.29	66.22
3	0.2638	13.29	79.51
4	0.1996	10.05	89.57
5	0.1055	5.31	94.88
6	0.0539	2.72	97.60
7	0.0260	1.31	98.91
8	0.0126	0.64	99.55
9	0.0058	0.29	99.84
10	0.0032	0.16	100.00

COORDENADAS, CONTRIBUCIONES Y COSENO CUADRADOS DE LOS INDIVIDUOS-COMPROMISO SOBRE LOS EJES PRINCIPALES (FACTORES A MODELIZAR)

LINEA	DISTO	COORDENADAS			CONTRIBUCIONES			COSENO CUADRADOS		
		1	2	3	1	2	3	1	2	3
87	0.30	0.42	-0.03	-0.07	0.24	0.00	0.02	0.57	0.00	0.02
88	0.41	0.54	0.27	0.01	0.40	0.12	0.00	0.72	0.17	0.00
89	0.26	0.18	-0.34	-0.23	0.04	0.20	0.21	0.12	0.46	0.22
90	0.25	0.07	-0.30	0.39	0.01	0.15	0.59	0.02	0.35	0.62
91	0.08	-0.11	-0.23	0.00	0.02	0.09	0.00	0.16	0.65	0.00
92	0.07	-0.17	-0.14	0.02	0.04	0.03	0.00	0.43	0.29	0.01
93	0.12	-0.26	-0.12	-0.10	0.09	0.02	0.03	0.57	0.11	0.08
94	0.09	-0.14	0.20	-0.08	0.03	0.07	0.03	0.22	0.42	0.08
95	0.10	0.05	0.11	0.11	0.08	0.03	0.05	0.50	0.11	0.10

A.3.1.7 *Coordenadas de las variables (ACP centrado de las nubes ponderadas)*

COORDENADAS DE LAS VARIABLES ACTIVAS SOBRE LOS TRES PRIMEROS EJES

VARIABLES	COORDENADAS			CORRELACIONES			EJES UNITARIOS		
	1	2	3	1	2	3	1	2	3
VIequi	-0.09	0.09	0.21	-0.35	0.33	0.79	-0.11	0.11	0.40
VIgastexter	-0.18	-0.16	-0.01	-0.70	-0.62	-0.05	-0.21	-0.21	-0.02
VIgastinter	-0.11	-0.23	0.05	-0.43	-0.87	0.20	-0.13	-0.30	0.10
VIpib	-0.06	0.25	-0.01	-0.23	0.96	-0.03	-0.07	0.33	-0.01
VIprim	0.04	0.25	-0.02	0.14	0.98	-0.08	0.04	0.33	-0.04
Visinies	-0.04	0.15	0.19	-0.15	0.59	0.71	-0.04	0.20	0.36
AUequi	-0.03	-0.02	0.06	-0.11	-0.06	0.19	-0.04	-0.02	0.11
AUgastexter	0.26	-0.06	-0.02	0.87	-0.19	-0.06	0.31	-0.08	-0.03
AUgastinter	0.21	0.20	-0.01	0.69	0.66	-0.02	0.24	0.26	-0.01
AUpib	-0.28	-0.01	-0.03	-0.94	-0.03	-0.09	-0.33	-0.01	-0.06
AUprim	-0.03	-0.30	0.01	-0.09	-0.98	0.02	-0.03	-0.39	0.01
AUsinies	-0.12	-0.02	0.06	-0.41	-0.08	0.21	-0.15	-0.03	0.12
SAequi	-0.05	0.03	-0.25	-0.18	0.11	-0.86	-0.06	0.04	-0.48
SAgastexter	0.24	-0.07	-0.07	0.84	-0.25	-0.26	0.28	-0.09	-0.14
SAgastinter	0.03	-0.20	-0.11	0.11	-0.69	-0.38	0.04	-0.26	-0.21
SApib	-0.25	0.15	-0.00	-0.86	0.50	-0.01	-0.29	0.19	-0.01
SAprim	-0.27	-0.08	0.01	-0.93	-0.27	0.03	-0.31	-0.10	0.02
SAsinies	-0.11	0.11	-0.21	-0.37	0.39	-0.74	-0.13	0.15	-0.42
REequi	-0.27	-0.09	-0.01	-0.91	-0.30	-0.04	-0.32	-0.12	-0.02
REgastexter	-0.24	0.11	-0.10	-0.81	0.36	-0.32	-0.28	0.14	-0.19
REgastinter	-0.14	-0.02	-0.16	-0.46	-0.08	-0.53	-0.16	-0.03	-0.31
REpib	-0.20	0.14	0.09	-0.65	0.46	0.31	-0.23	0.18	0.18
REprim	0.14	-0.25	0.06	0.45	-0.84	0.20	0.16	-0.33	0.12
REsinies	-0.22	-0.17	0.07	-0.73	-0.56	0.22	-0.25	-0.22	0.13

A.3.2 Segunda fase: Estudio temporal

A.3.2.1 Modelo ARIMA para el factor 1 (Factor1)

IDENTIFICACIÓN

FUNCIONES DE AUTOCORRELACIÓN Y AUTOCORRELACIÓN PARCIAL DEL PRIMER FACTOR (Factor1)

- AUTOCORRELACIONES:

Re.	Auto-corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	,702	,264										7,045	,008
2	,345	,251						I*****				8,940	,011
3	,045	,237						I*				8,976	,030
4	-,148	,221						***I				9,421	,051
5	-,295	,205						*****I				11,498	,042
6	-,305	,187						*****I				14,165	,028
7	-,311	,167						*****I				17,613	,014
8	-,274	,145						*****I				21,201	,007
9	-,179	,118						*****I				23,482	,005

Símbolos: - AUTOCORRELACIONES: * Límites del error estándar: .
Total Casos: 11

- AUTOCORRELACIONES PARCIALES:

Auto-Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	,702	,302								
2	-,291	,302								
3	-,137	,302								
4	-,072	,302								
5	-,186	,302								
6	,047	,302								
7	-,190	,302								
8	-,049	,302								
9	,035	,302								

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total Casos: 11

TERMINO INDEPENDIENTE

- ESTADÍSTICOS DESCRIPTIVOS

	N	Mínimo	Máximo	Media (w)	Desviación (σ_w)	w/ σ_w
Factor1	11	-,26	,54	9,091E-04	,2706	,0033

ESTIMACIÓNVARIABLE: **Factor1**

Número de diferencias no-estacionales: 0
 No hay componente estacional en el modelo.

PARÁMETROS:

AR1 _____ < valor obtenido a partir de la estimación>
 95,00 porcentaje de confianza para la obtención de intervalos.
 Algoritmo empleado en la estimación: Algoritmo de Melard

CRITERIOS DE FINALIZACIÓN DEL PROCESO:

Parámetro Epsilon: ,001
 Máximo de la Constante de Marquardt: 1,00E+09
 Porcentaje de SSQ: ,001
 Número máximo de iteraciones: 10

VALORES INICIALES:

AR1 ,70157
 Constante de Marquardt = ,001
 Suma ajustada de cuadrados = ,29299762

DESCRIPCIÓN DE ITERACIONES:

Iteración	Sum. Ajust. de Cuadrados	Constante de Marquardt
1	,27711724	,00100000
2	,27683267	,00010000
3	,27682606	,00001000

Fin del proceso de estimación en la iteración 4 pues la suma de los cuadrados decrece menos del ,001 por ciento.

PARÁMETROS FINALES:

Número de residuos 11
 Error Estándar ,1574317
 Log de verosimilitud 4,6434723
 AIC -7,2869445
 SBC -6,8890493

ANÁLISIS DE LA VARIANZA:

Grados de libertad	Suma ajustada de Cuadrados	Varianza Residual
Residuos 10	,27682589	,02478475

VARIABLES EN EL MODELO:

B	S(B)	T-RATIO	APPROX. PROB.
AR1 ,83885839	,14106355	5,9466699	,00014187

MATRIZ DE COVARIANZAS:

AR1
 AR1 ,01989893

MATRIZ DE CORRELACIONES:

AR1
 AR1 1,0000000

VALIDACIÓN

ANÁLISIS DE LOS RESIDUOS

- MEDIA CERO

Descriptivos:	Media	-1,05E-02
	Desviación típica	,1925507
	Media/Desviación	-,054

- NORMALIDAD

Prueba de Kolmogorov-Smirnov para una muestra (La distribución de contraste es la Normal)

N :	11
Parámetros normales:	Media -1,05E-02
	Desviación típica ,1925507
Diferencias más extremas:	Absoluta ,182
	Positiva ,182
	Negativa -,115
Z de Kolmogorov-Smirnov:	,603
Sig.asintót. (bilateral):	,860

- FUNCIONES DE AUTOCORRELACIÓN Y DE AUTOCORRELACION PARCIAL DE LOS RESIDUOS

- Autocorrelaciones:

Auto-	Estandar	Re. corre	Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	,128	,264	I***	,236	,627
2	-,125	,251	**I	,483	,785
3	-,145	,237	***I	,859	,835
4	-,119	,221	**I	1,150	,886
5	-,210	,205	****I	2,196	,821
6	,001	,187	*	2,196	,901
7	-,022	,167	*	2,214	,947
8	-,079	,145	**I	2,511	,961
9	,094	,118	I**	3,141	,958

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

- Autocorrelaciones parciales:

Auto-	Estandar	Parcia	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	,128	,302	I***
2	-,144	,302	***I
3	-,112	,302	**I
4	-,107	,302	**I
5	-,228	,302	****I
6	,002	,302	*
7	-,126	,302	***I
8	-,155	,302	***I
9	,052	,302	I*

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

PREDICCIÓN

Tabla e
Valores predichos

AÑO	87	88	89	90	91	92	93	94	95	96	97
Valor predicho	,00	,356	,453	,151	,059	-,092	-,143	-,218	-,117	-,210	-,109

-NORMALIDAD

Prueba de Kolmogorov-Smirnov para una muestra (La distribución de contraste es la Normal)

N : 11
 Parámetros normales: Media 8,467E-10
 Desviación típica ,2424
 Diferencias más extremas: Absoluta ,159
 Positiva ,144
 Negativa -,159
 Z de Kolmogorov-Smirnov: ,527
 Sig.asintót. (bilateral): ,944

PREDICCIÓN

Tabla f
Valores predichos

AÑO	87	88	89	90	91	92	93	94	95	96	97
Valor predicho	-,03	,27	-,34	-,3	-,23	-,14	-,12	,2	,11	,27	,31

A.3.2.3 Modelo ARIMA para el factor 3 (Factor3)

IDENTIFICACIÓN

FUNCIONES DE AUTOCORRELACIÓN Y AUTOCORRELACIÓN PARCIAL DEL TERCER FACTOR (Factor3)

- AUTOCORRELACIONES:

Re.	Auto- Estandar										Box-Ljung	Probabilidad	
	corre	Error	-1	-.75	-.5	-.25	0	.25	.5	.75			1
1	-,297	,264					*****I					1,260	,262
2	,075	,251					I*					1,349	,509
3	-,337	,237					*****I					3,373	,338
4	-,082	,221					**I					3,513	,476
5	-,093	,205					**I					3,718	,591
6	,167	,187					I***					4,511	,608
7	,201	,167					I****					5,959	,545
8	-,093	,145					**I					6,374	,605
9	-,003	,118					*					6,375	,702

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

- AUTOCORRELACIONES PARCIALES:

Parcia	Auto- Estand.										
			-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	-,297	,302					*****I				
2	-,015	,302					*				
3	-,349	,302					*****I				
4	-,350	,302					*****I				
5	-,359	,302					*****I				
6	-,252	,302					*****I				
7	-,037	,302					*I				
8	-,279	,302					*****I				
9	-,290	,302					*****I				

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

TERMINO INDEPENDIENTE

- ESTADÍSTICOS DESCRIPTIVOS

	N	Mínimo	Máximo	Media (w)	Desviación (σ_w)	w/ σ_w
Factor3	11	-,23	,39	-1,69E-10	,1611	-1,049E-9

-NORMALIDAD

Prueba de Kolmogorov-Smirnov para una muestra (La distribución de contraste es la Normal)

N : 11
 Parámetros normales: Media 1,69E-10
 Desviación típica ,1611
 Diferencias más extremas: Absoluta ,244
 Positiva ,244
 Negativa -,156
 Z de Kolmogorov-Smirnov: ,810
 Sig.asintót. (bilateral): ,528

PREDICCIÓN

Tabla g
Valores predichos

AÑO	87	88	89	90	91	92	93	94	95	96	97
Valor predicho	,07	,01	,23	-,39	,00	-,02	,1	,08	,11	,03	,14

A.3.3 Tercera fase: Validación del modelo**A.3.3.1 Valores predichos a través de la formula de reconstitución****Tabla h**
Variables relativas a los seguros de Vida

AÑO	VIequi	VIGastexter	VIGastinter	VIpib	VIprim	Visinies
87	115,29	4,26	6,44	1,64	37,49	104,58
88	116,13	3,75	5,63	1,79	40,48	106,64
89	111,25	4,16	6,57	1,37	35,51	100,56
90	119,08	4,32	7,13	1,41	34,84	107,72
91	115,26	4,37	6,82	1,49	35,74	104,12
92	116,20	4,41	6,79	1,57	36,32	105,04
93	115,08	4,45	6,72	1,60	36,56	103,95
94	116,53	4,26	6,17	1,83	39,25	106,06
95	115,63	4,25	6,25	1,75	38,61	105,10
96	117,95	4,19	6,10	1,87	39,76	107,60
97	119,00	4,07	6,01	1,88	40,12	108,85

Tabla i
Variables relativas a los seguros de Automóviles

AÑO	AUequi	AUGastexter	AUGastinter	AUpib	AUprim	AUsinies
87	104,40	13,36	16,85	1,32	31,52	74,20
88	104,24	13,68	17,10	1,26	29,56	73,49
89	104,01	14,02	16,90	1,25	33,09	73,11
90	105,24	13,55	16,79	1,29	33,19	74,79
91	104,58	13,48	16,79	1,31	32,71	74,28
92	104,65	13,27	16,77	1,34	32,25	74,59
93	104,46	13,22	16,76	1,35	32,12	74,48
94	104,41	13,03	16,86	1,36	30,25	74,55
95	104,34	13,18	16,87	1,34	30,73	74,32
96	104,57	13,00	16,89	1,35	29,86	74,69
97	104,70	13,09	16,94	1,33	29,60	74,66

Tabla j
Variables relativas a los seguros de salud

AÑO	SAequi	SAGastexter	SAGastinter	SAPib	SAPrim	SAsinies
87	98,35	3,79	12,66	0,67	15,52	81,89
88	97,92	3,88	12,43	0,64	14,70	81,59
89	98,70	4,06	13,04	0,57	14,86	81,57
90	96,15	3,79	12,60	0,62	15,46	79,79
91	97,94	3,83	12,78	0,64	15,54	81,33
92	97,96	3,74	12,68	0,67	15,77	81,54
93	98,52	3,74	12,74	0,68	15,83	82,04
94	98,59	3,66	12,48	0,72	15,78	82,44
95	98,63	3,72	12,57	0,70	15,65	82,33
96	98,12	3,63	12,36	0,73	15,73	82,14
97	97,60	3,65	12,27	0,72	15,54	81,69

Tabla k
Variables relativas al resto de seguros

AÑO	REequi	REGastexter	REGastinter	REpib	REprim	REsinies
87	103,47	31,56	22,39	0,65	15,32	49,53
88	100,04	30,94	22,05	0,64	16,51	47,01
89	101,08	30,34	22,33	0,62	14,98	48,34
90	102,90	30,08	21,79	0,65	15,24	50,97
91	103,56	31,05	22,29	0,64	14,97	50,20
92	104,41	31,49	22,35	0,65	15,02	50,57
93	104,78	31,84	22,52	0,65	14,91	50,43
94	104,44	32,39	22,51	0,66	15,60	49,59
95	103,97	32,08	22,49	0,65	15,48	49,42
96	104,14	32,27	22,37	0,66	15,86	49,54
97	103,25	31,89	22,18	0,66	16,17	49,20

A.3.3.2 Errores

Tabla l
Errores de las variables relativas a los seguros de vida

AÑO	EVIequi	EVIgastexter	EVIgastinter	EVIpib	EVIprim	EVISinies
87	-2,26	-1,6	-1,57	-0,38	-0,02	0,92
88	-3,72	-1,27	-2,09	0,23	7,97	-0,26
89	-10,04	0,87	1,28	-0,38	-5,74	-12,24
90	6,21	0,34	1,68	-0,48	-7,39	4,1
91	1,3	0,99	1,42	-0,29	-4,53	-1,17
92	-1,96	0,17	1,05	-0,2	-3,1	-3,22
93	3,15	1,17	1,06	-0,16	-3,71	0,88
94	-3,44	-0,37	-1,08	0,41	3,72	-1,95
95	-2,1	0,01	-0,19	0,22	1,96	-1,89
96	3,76	0,01	-0,76	0,4	3,81	4,54
97	9,46	-0,22	-0,7	0,65	6,87	10,45

Tabla m
Errores de las variables relativas a los seguros de automóviles

AÑO	EAUequi	EAUgastexter	EAUgastinter	EAUpib	EAUprim	EAUsinies
87	-12,18	0,57	0,29	-0,26	0,05	-13,05
88	4,89	2,32	0,98	-0,21	-4,21	1,56
89	4,88	0,92	-0,28	-0,04	3,54	4,22
90	1,56	0,11	-0,23	-0,01	4,41	1,79
91	3,3	-0,52	-0,43	0,07	3,25	4,28
92	-1,83	-0,32	-0,26	0,08	1,96	-1,23
93	-0,46	-0,14	-0,19	0,18	2,72	-0,13
94	-0,49	-0,52	0,1	0,15	-1,27	-0,1
95	-1,46	-1,57	-0,08	0,08	-1,56	0,16
96	-0,95	-0,5	-0,12	0,03	-3,2	-0,34
97	2,81	-0,47	0,18	-0,04	-5,63	3,11

Tabla n
Errores de las variables relativas a los seguros de salud

AÑO	ESAequi	ESAgastexter	ESAgastinter	ESApib	ESAprim	ESAsinies
87	-0,07	0,24	0,72	-0,21	-1,83	-1,02
88	-1,64	0,67	-0,65	-0,17	-3,35	-1,64
89	3,72	0,61	1,08	-0,07	0,02	2,06
90	-4,56	0,02	-0,26	-0,09	-0,01	-4,35
91	0,33	-0,16	0,34	0	1,1	0,15
92	-1,41	-0,29	-0,28	0,04	1,28	-0,85
93	0,81	-0,42	0,94	0,09	1,74	0,29
94	-0,56	-0,07	-0,85	0,07	-0,57	0,36
95	0,75	-0,18	-0,57	0,1	0,81	1,5
96	2,09	-0,23	-0,75	0,1	0,3	3,06
97	0,6	-0,26	0,26	0,16	0,76	0,59

Tabla o
Errores de las variables relativas al resto de seguros

AÑO	EREequi	EREGastexter	EREGastinter	EREpib	EREprim	EREsinies
87	-9,45	-2,92	-1,24	-0,07	1,95	-5,3
88	-12,24	-2,36	0,08	-0,03	-1,67	-9,93
89	0,99	-0,98	-0,3	0	3,74	2,35
90	0,86	-2,62	-0,77	0,02	4,26	4,3
91	2,64	-0,49	0,65	-0,02	1,22	2,5
92	6,74	1,08	0,47	-0,01	0,5	5,18
93	4,83	3,56	1,09	0	-0,18	0,16
94	1,32	1,1	1,32	0,01	-2,74	-1,13
95	7,87	2,45	1,34	0,02	-1,69	4,06
96	-4,18	-0,46	-1,23	0,05	-2,12	-2,53
97	1,53	1,95	-1,33	0,02	-3,43	0,9

A.3.3.3 *Análisis del modelo*

ANÁLISIS DE LOS ERRORES

- REPRESENTACIONES GRÁFICAS

Figura a
Representaciones de los errores de las variables relativas a los seguros de vida

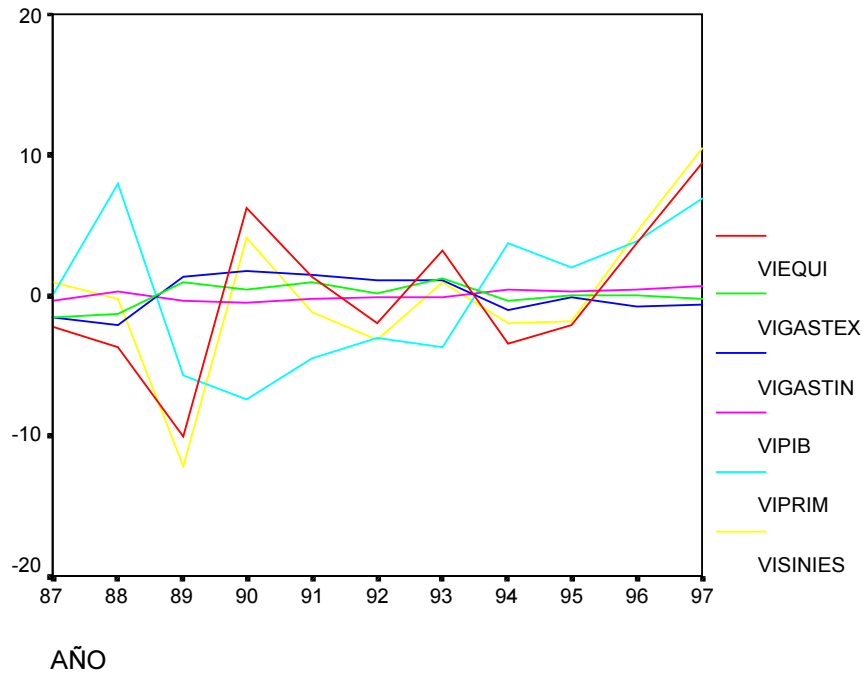


Figura b
Representaciones de los errores de las variables relativas a los seguros de automóviles

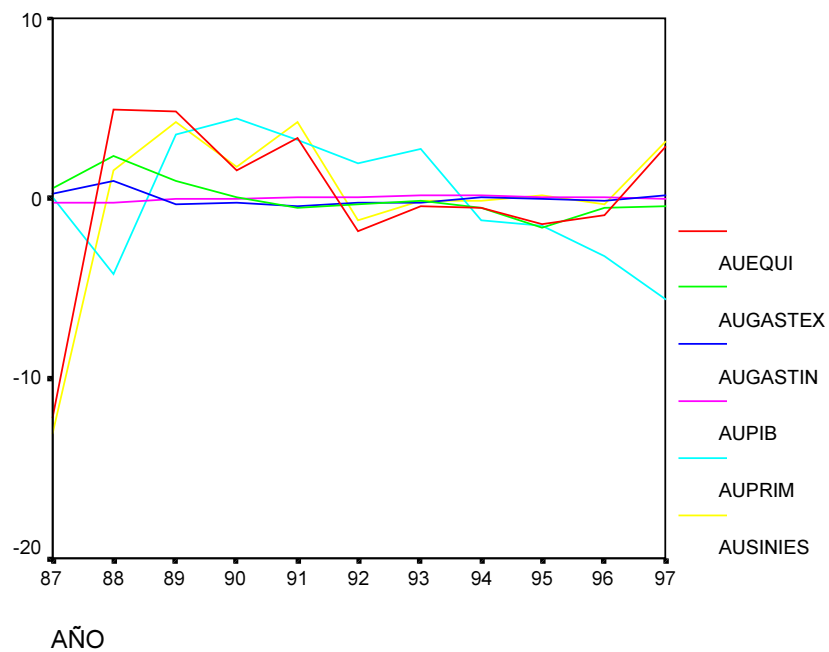


Figura c
Representaciones de los errores de las variables relativas a los seguros de salud

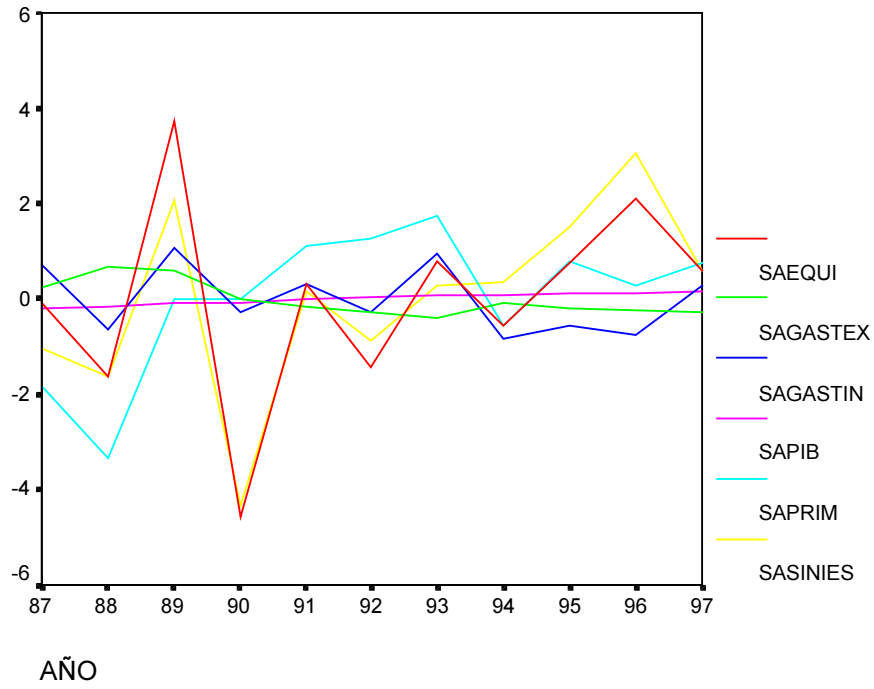
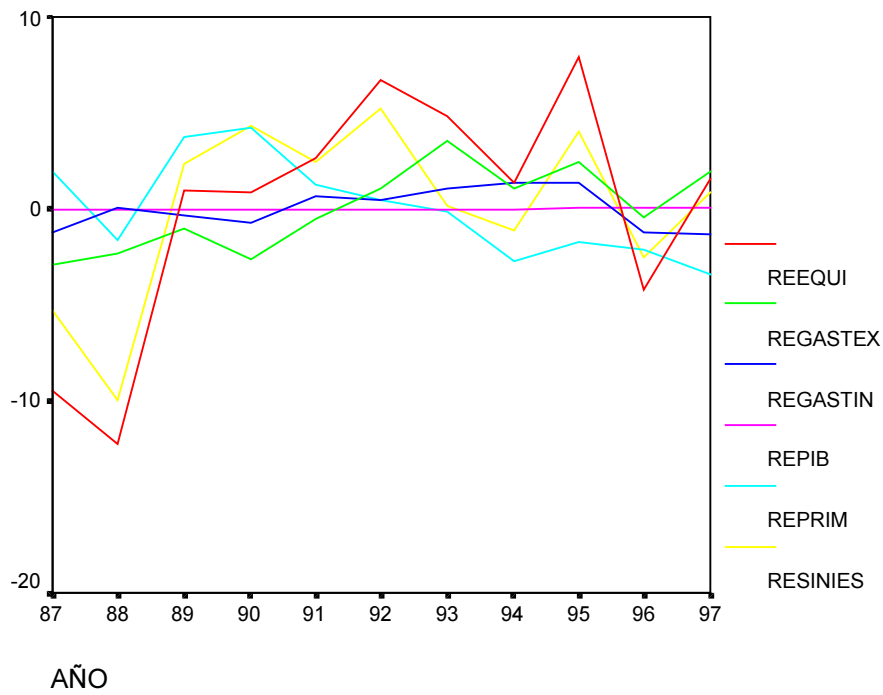


Figura d
Representaciones de los errores de las variables relativas al resto de seguros



- MEDIA CERO

Descriptivos:

	Media	Desv.típ.	Media/Desv.típ.
EVIequi	0,03	5,42	0,01
EVIgastex	0,01	0,87	0,01
EVIgastin	0,01	1,34	0,01
EVIpib	0,00	0,39	0,00
EVIprim	-0,01	5,24	0,00
EVIsinies	0,02	5,63	0,00
EAUequi	0,01	4,74	0,00
EAUgastex	-0,01	1,01	-0,01
EAUgastin	0,00	0,39	-0,01
EAUpib	0,00	0,14	0,01
EAUprim	0,01	3,43	0,00
EAUsinies	0,02	4,73	0,01
ESAequi	0,01	2,14	0,00
ESAgastex	0,00	0,36	-0,01
ESAgastin	0,00	0,70	0,00
ESApib	0,00	0,12	0,01
ESAprim	0,02	1,49	0,01
ESAsinies	0,01	1,99	0,01
EREequi	0,08	6,32	0,01
EREGastex	0,03	2,17	0,01
EREGastin	0,01	1,04	0,01
EREpib	0,00	0,03	0,01
EREprim	-0,01	2,59	-0,01
EREsinies	0,05	4,57	0,01

- NORMALIDAD

Prueba de Kolmogorov-Smirnov para una muestra (La distribución de contraste es la Normal)

Tabla p
Errores de las variables relativas a los seguros de vida

	EVIequi	EVIgastex	EVIgastin	EVIpib	EVIprim	EVIsinies
Z de Kolmogorov	,626	,505	,787	,683	,586	,637
Sig. asintót. bilateral	,828	,960	,566	,740	,883	,812

Tabla q
Errores de las variables relativas a los seguros de automóviles

	EAUequi	EAUgastex	EAUgastin	EAUpib	EAUprim	EAUsinies
Z de Kolmogorov	,855	,717	,690	,623	,565	1,009
Sig. asintót. bilateral	,457	,683	,727	,833	,907	,260

Tabla r
Errores de las variables relativas a los seguros de salud

	ESAequi	ESAgastex	ESAgastin	ESApib	ESAprim	ESAsinies
Z de Kolmogorov	,571	,684	,632	,532	,724	,544
Sig. asintót. bilateral	,9	,738	,82	,94	,671	,928

Tabla s
Errores de las variables relativas al resto de seguros

	EREequi	EREGastex	EREGastin	EREpib	EREprim	EREsinies
Z de Kolmogorov	,917	,466	,514	,514	,639	,487
Sig. asintót. bilateral	,37	,982	,954	,955	,808	,972

- FUNCIONES DE AUTOCORRELACIÓN Y DE AUTOCORRELACION PARCIAL DE LOS ERRORES

EVIequi

- Autocorrelaciones:

Re.	Auto- Estandar										Box-Ljung	Probabilidad	
	corre	Error	-1	-.75	-.5	-.25	0	.25	.5	.75			1
1	-,291	,264					*****I					1,211	,271
2	-,185	,251					****I					1,752	,416
3	,140	,237					I***					2,105	,551
4	-,209	,221					****I					2,996	,559
5	-,087	,205					**I					3,177	,673
6	,161	,187					I***					3,919	,688
7	,183	,167					I****					5,115	,646
8	-,116	,145					**I					5,758	,674
9	-,034	,118					*I					5,840	,756

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

- Autocorrelaciones Parciales:

Re.	Auto- Estand.										
	Parcia		-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	-,291	,302					*****I				
2	-,294	,302					*****I				
3	-,018	,302					*				
4	-,258	,302					*****I				
5	-,271	,302					*****I				
6	-,110	,302					**I				
7	,193	,302					I****				
8	,040	,302					I*				
9	-,029	,302					*I				

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

EVIGastexter

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	-,193	,264					****I					,532	,466
2	-,066	,251			.		*I					,600	,741
3	-,177	,237			.		****I					1,161	,762
4	,045	,221			.		I*					1,203	,878
5	-,176	,205			.		****I					1,940	,857
6	-,007	,187			.		*					1,941	,925
7	,058	,167			.		I*					2,061	,956
8	-,068	,145			.		*I					2,282	,971
9	,044	,118			.		I*					2,421	,983

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

- Autocorrelaciones parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	-,193	,302				****I				
2	-,107	,302			.	**I				
3	-,222	,302			.	****I				
4	-,055	,302			.	*I				
5	-,243	,302			.	*****I				
6	-,172	,302			.	***I				
7	-,062	,302			.	*I				
8	-,219	,302			.	****I				
9	-,100	,302			.	**I				

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

EVIGastinter

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	,079	,264					I**					,089	,766
2	-,049	,251					*I					,127	,939
3	-,131	,237					***I					,433	,933
4	-,198	,221					****I					1,234	,872
5	-,295	,205					*****I					3,305	,653
6	,089	,187					I**					3,534	,739
7	-,063	,167					*I					3,674	,816
8	-,024	,145					*					3,702	,883
9	,069	,118					I*					4,038	,909

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

- Autocorrelaciones parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	,079	,302				I**				
2	-,055	,302				*I				
3	-,124	,302				**I				
4	-,185	,302				****I				
5	-,299	,302				*****I				
6	,083	,302				I**				
7	-,177	,302				****I				
8	-,135	,302				***I				
9	-,045	,302				*I				

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

EVIpib

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	,016	,264					*					,004	,951
2	,244	,251		.			I*****					,953	,621
3	,077	,237		.			I**					1,060	,787
4	-,178	,221		.			****I					1,707	,789
5	-,185	,205		.			****I					2,525	,773
6	-,054	,187		.			*I					2,609	,856
7	-,281	,167		.			*****I					5,427	,608
8	-,063	,145		.			*I					5,615	,690
9	-,016	,118		.			*					5,633	,776

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

- Autocorrelaciones Parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	,016	,302		.		*				
2	,244	,302		.		I*****				
3	,075	,302		.		I*				
4	-,255	,302		.		*****I				
5	-,252	,302		.		*****I				
6	,059	,302		.		I*				
7	-,140	,302		.		***I				
8	-,093	,302		.		**I				
9	,022	,302		.		*				

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

EVIprim

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	-,085	,264					**I					,103	,748
2	,100	,251					I**					,262	,877
3	-,044	,237					*I					,296	,961
4	-,197	,221					****I					1,088	,896
5	-,182	,205					****I					1,875	,866
6	,102	,187					I**					2,173	,903
7	-,169	,167					***I					3,195	,866
8	-,066	,145					*I					3,404	,907
9	,047	,118					I*					3,563	,938

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

- Autocorrelaciones parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	-,085	,302				**I				
2	,094	,302				I**				
3	-,028	,302				*I				
4	-,215	,302				****I				
5	-,220	,302				****I				
6	,115	,302				I**				
7	-,135	,302				***I				
8	-,207	,302				****I				
9	-,025	,302				*I				

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

EVIsinies

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	-,399	,264					*****I					2,282	,131
2	-,112	,251					**I					2,480	,289
3	,220	,237					I****					3,345	,341
4	-,150	,221					***I					3,806	,433
5	-,091	,205					**I					4,004	,549
6	,209	,187						I****				5,253	,512
7	,027	,167						I*				5,280	,626
8	-,200	,145					****I					7,196	,516
9	,009	,118					*					7,202	,616

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

- Autocorrelaciones parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	-,399	,302				*****I				
2	-,323	,302				*****I				
3	,043	,302					I*			
4	-,071	,302					*I			
5	-,167	,302					***I			
6	,054	,302					I*			
7	,188	,302					I****			
8	-,054	,302					*I			
9	-,201	,302					****I			

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

EAUequi

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	-,022	,264					*					,007	,932
2	-,240	,251			.		*****I					,921	,631
3	-,103	,237			.		**I					1,111	,774
4	-,250	,221			.		*****I					2,393	,664
5	-,001	,205			.		*					2,393	,793
6	-,010	,187			.		*					2,396	,880
7	,002	,167			.		*					2,396	,935
8	,114	,145			.		I**					3,011	,934
9	,115	,118			.		I**					3,955	,914

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

- Autocorrelaciones parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	-,022	,302				*				
2	-,240	,302			.	*****I				
3	-,122	,302			.	**I				
4	-,342	,302			.	*****I				
5	-,123	,302			.	**I				
6	-,252	,302			.	*****I				
7	-,194	,302			.	****I				
8	-,152	,302			.	***I				
9	-,038	,302			.	*I				

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

EAUgastexter

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	,368	,264						I*****				1,937	,164
2	,145	,251						I***				2,270	,321
3	-,109	,237						**I				2,480	,479
4	-,003	,221						*				2,481	,648
5	-,227	,205						*****I				3,709	,592
6	-,259	,187						*****I				5,627	,466
7	-,348	,167						*****I				9,955	,191
8	-,167	,145						***I				11,292	,186
9	,063	,118						I*				11,574	,238

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

- Autocorrelaciones parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	,368	,302						I*****		
2	,011	,302						*		
3	-,191	,302						****I		
4	,115	,302						I**		
5	-,276	,302						*****I		
6	-,154	,302						***I		
7	-,175	,302						****I		
8	-,050	,302						*I		
9	,192	,302						I****		

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

EAUgastinter

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	-,248	,264					*****I					,879	,348
2	-,004	,251			.		*					,880	,644
3	-,153	,237			.		***I					1,296	,730
4	-,068	,221			.		*I					1,390	,846
5	-,139	,205			.		***I					1,853	,869
6	,127	,187			.		I***					2,311	,889
7	,008	,167			.		*					2,313	,940
8	-,130	,145			.		***I					3,122	,926
9	,078	,118			.		I**					3,561	,938

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

- Autocorrelaciones parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	-,248	,302				*****I				
2	-,070	,302			.	*I				
3	-,183	,302			.	***I				
4	-,173	,302			.	***I				
5	-,258	,302			.	*****I				
6	-,051	,302			.	*I				
7	-,059	,302			.	*I				
8	-,268	,302			.	*****I				
9	-,113	,302			.	**I				

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

EAUpib

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	,188	,264					I****					,505	,477
2	-,111	,251				**I						,702	,704
3	-,080	,237				**I						,816	,846
4	-,033	,221				*I						,838	,933
5	-,069	,205				*I						,953	,966
6	-,376	,187			*	*****I						4,996	,544
7	-,162	,167				***I						5,935	,547
8	-,134	,145				***I						6,795	,559
9	,104	,118				I**						7,565	,578

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

- Autocorrelaciones parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	,188	,302				I****				
2	-,152	,302				***I				
3	-,028	,302				*I				
4	-,029	,302				*I				
5	-,076	,302				**I				
6	-,381	,302			*	*****I				
7	-,050	,302				*I				
8	-,257	,302				*****I				
9	,096	,302				I**				

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

EAUprim

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	-,035	,264					*I					,018	,894
2	,159	,251					I***					,419	,811
3	-,005	,237					*					,420	,936
4	-,147	,221					***I					,863	,930
5	-,183	,205					****I					1,662	,894
6	,018	,187					*					1,671	,947
7	-,214	,167					****I					3,315	,854
8	-,138	,145					***I					4,225	,836
9	,058	,118					I*					4,465	,878

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

- Autocorrelaciones parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	-,035	,302				*I				
2	,158	,302				I***				
3	,006	,302				*				
4	-,177	,302				****I				
5	-,204	,302				****I				
6	,061	,302				I*				
7	-,154	,302				***I				
8	-,218	,302				****I				
9	,045	,302				I*				

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

EAUsinies

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	-,009	,264					*					,001	,973
2	-,268	,251			.	*****I						1,144	,565
3	-,146	,237			.	***I						1,525	,677
4	-,225	,221			.	*****I						2,562	,633
5	,040	,205			.		I*					2,600	,761
6	,018	,187			.		*					2,609	,856
7	,037	,167			.		I*					2,659	,915
8	,075	,145			.		I*					2,926	,939
9	,077	,118			.		I**					3,354	,949

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

- Autocorrelaciones parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	-,009	,302				*				
2	-,268	,302			.	*****I				
3	-,163	,302			.	***I				
4	-,340	,302			.	*****I				
5	-,113	,302			.	**I				
6	-,242	,302			.	*****I				
7	-,149	,302			.	***I				
8	-,151	,302			.	***I				
9	-,024	,302			.	*				

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

ESAEqui

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	-,528	,264					*****I					3,988	,046
2	,346	,251						I*****				5,893	,053
3	-,198	,237					****I					6,591	,086
4	,123	,221						I**				6,901	,141
5	-,150	,205					***I					7,438	,190
6	-,125	,187					**I					7,885	,247
7	,085	,167						I**				8,143	,320
8	-,028	,145					*I					8,181	,416
9	-,025	,118					*					8,224	,512

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

- Autocorrelaciones parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	-,528	,302				*****I				
2	,093	,302					I**			
3	,024	,302					*			
4	,006	,302					*			
5	-,107	,302					**I			
6	-,345	,302					*****I			
7	-,099	,302					**I			
8	,110	,302						I**		
9	-,024	,302						*		

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

ESAgastexter

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	,670	,264										6,415	,011
2	,211	,251						I*****				7,121	,028
3	-,087	,237					**I					7,256	,064
4	-,218	,221					****I					8,224	,084
5	-,213	,205					****I					9,309	,097
6	-,164	,187					***I					10,080	,121
7	-,209	,167					****I					11,648	,113
8	-,267	,145					*****I					15,050	,058
9	-,174	,118					. ***I					17,206	,046

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

- Autocorrelaciones parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	,670	,302								I*****
2	-,431	,302								*****I
3	,000	,302								*
4	-,097	,302								**I
5	-,001	,302								*
6	-,076	,302								**I
7	-,246	,302								*****I
8	-,068	,302								*I
9	,122	,302								I**

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

ESAgastinter

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	-,404	,264					*****I					2,330	,127
2	,385	,251						I*****				4,682	,096
3	-,407	,237					*****I					7,639	,054
4	,391	,221						I*****.				10,758	,029
5	-,387	,205					*****I					14,331	,014
6	,182	,187						I****				15,281	,018
7	-,228	,167					*****I					17,142	,017
8	,073	,145						I*				17,397	,026
9	-,143	,118					***I					18,868	,026

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

- Autocorrelaciones parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	-,404	,302				*****I				
2	,265	,302					I*****			
3	-,238	,302				*****I				
4	,170	,302					I***			
5	-,142	,302				***I				
6	-,170	,302				***I				
7	,018	,302				*				
8	-,214	,302				****I				
9	-,036	,302				*I				

ESApib

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	,666	,264										6,345	,012
2	,423	,251										9,186	,010
3	,217	,237										10,025	,018
4	-,007	,221										10,026	,040
5	-,215	,205										11,127	,049
6	-,321	,187										14,072	,029
7	-,364	,167										18,809	,009
8	-,343	,145										24,409	,002
9	-,332	,118										32,285	,000

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

- Autocorrelaciones parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	,666	,302								
2	-,037	,302								
3	-,091	,302								
4	-,192	,302								
5	-,200	,302								
6	-,073	,302								
7	-,064	,302								
8	-,025	,302								
9	-,135	,302								

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

ESAprim

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	,392	,264						I*****				2,202	,138
2	,134	,251						I***				2,488	,288
3	-,148	,237					***I					2,877	,411
4	-,170	,221					***I					3,466	,483
5	-,313	,205					*****I					5,804	,326
6	-,018	,187					*					5,814	,444
7	-,072	,167					*I					5,998	,540
8	-,109	,145					**I					6,560	,585
9	-,136	,118					***I					7,876	,547

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

- Autocorrelaciones parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	,392	,302						I*****		
2	-,024	,302						*		
3	-,227	,302						*****I		
4	-,039	,302						*I		
5	-,241	,302						*****I		
6	,215	,302						I****		
7	-,169	,302						***I		
8	-,196	,302						****I		
9	-,020	,302						*		

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

ESAsinies

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	-,117	,264					**I					,196	,658
2	,282	,251					I*****					1,461	,482
3	,028	,237					I*					1,475	,688
4	-,048	,221					*I					1,522	,823
5	-,136	,205					***I					1,964	,854
6	-,277	,187					*****I					4,154	,656
7	,022	,167					*					4,172	,760
8	-,136	,145					***I					5,051	,752
9	-,103	,118					**I					5,813	,758

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

- Autocorrelaciones parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	-,117	,302				**I				
2	,272	,302				I*****				
3	,091	,302				I**				
4	-,125	,302				***I				
5	-,209	,302				****I				
6	-,311	,302				*****I				
7	,068	,302				I*				
8	,097	,302				I**				
9	-,130	,302				***I				

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

EREequi

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	-,095	,264					**I					,129	,719
2	-,326	,251			.	*****I						1,814	,404
3	-,046	,237			.		*I					1,851	,604
4	-,062	,221			.		*I					1,930	,749
5	-,047	,205			.		*I					1,982	,852
6	,059	,187			.		I*					2,082	,912
7	-,034	,167			.		*I					2,122	,953
8	-,008	,145			.		*					2,126	,977
9	,103	,118			.		I**					2,877	,969

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

- Autocorrelaciones Parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	-,095	,302				**I				
2	-,338	,302			.	*****I				
3	-,136	,302			.		***I			
4	-,230	,302			.		*****I			
5	-,194	,302			.		****I			
6	-,127	,302			.		***I			
7	-,207	,302			.		****I			
8	-,164	,302			.		***I			
9	-,073	,302			.		*I			

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

EREgastexter

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	,457	,264						I*****				2,990	,084
2	-,048	,251					*I					3,027	,220
3	-,202	,237					****I					3,756	,289
4	-,099	,221					**I					3,956	,412
5	-,098	,205					**I					4,186	,523
6	-,210	,187					****I					5,449	,488
7	-,198	,167					****I					6,849	,445
8	-,157	,145					***I					8,028	,431
9	,013	,118					*					8,040	,530

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

- Autocorrelaciones parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	,457	,302						I*****		
2	-,326	,302					*****I			
3	-,034	,302					*I			
4	,031	,302					I*			
5	-,164	,302					***I			
6	-,175	,302					****I			
7	-,037	,302					*I			
8	-,171	,302					***I			
9	,081	,302					I**			

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

EREgastinter

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	,208	,264					I****					,617	,432
2	-,335	,251			. *****I							2,405	,300
3	-,190	,237			. ****I							3,054	,383
4	-,058	,221			. *I							3,122	,538
5	-,131	,205			. ***I							3,531	,619
6	,273	,187			. I*****							5,657	,463
7	,262	,167			. I*****							8,106	,323
8	-,155	,145			. ***I							9,256	,321
9	-,302	,118			. *.****I							15,764	,072

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

- Autocorrelaciones Parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	,208	,302				I****				
2	-,396	,302			. *****I					
3	-,014	,302			. *					
4	-,164	,302			. ***I					
5	-,199	,302			. ****I					
6	,363	,302			. I*****					
7	-,067	,302			. *I					
8	-,073	,302			. *I					
9	-,109	,302			. **I					

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

ERepib

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	,255	,264					I*****					,933	,334
2	,016	,251		.			*					,937	,626
3	,143	,237			.		I***					1,305	,728
4	,106	,221			.		I**					1,535	,820
5	-,106	,205			.		**I					1,805	,875
6	-,477	,187			***	*****I						8,302	,217
7	-,132	,167			.		***I					8,922	,258
8	-,070	,145			.		*I					9,152	,330
9	-,165	,118			.		***I					11,092	,269

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

- Autocorrelaciones Parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	,255	,302				I*****				
2	-,052	,302		.		*I				
3	,164	,302		.		I***				
4	,029	,302		.		I*				
5	-,145	,302		.		***I				
6	-,478	,302		.	*****I					
7	,085	,302		.		I**				
8	-,068	,302		.		*I				
9	,045	,302		.		I*				

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

EREprim

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	-,033	,264					*I					,016	,901
2	,031	,251		.			I*					,030	,985
3	-,095	,237		.			**I					,193	,979
4	-,225	,221		.			****I					1,226	,874
5	-,249	,205		.			****I					2,701	,746
6	,160	,187		.			I***					3,433	,753
7	-,094	,167		.			**I					3,746	,808
8	-,014	,145		.			*					3,756	,878
9	,020	,118		.			*					3,784	,925

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

- Autocorrelaciones parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	-,033	,302		.		*I				
2	,030	,302		.		I*				
3	-,094	,302		.		**I				
4	-,234	,302		.		****I				
5	-,280	,302		.		*****I				
6	,141	,302		.		I***				
7	-,118	,302		.		**I				
8	-,167	,302		.		***I				
9	-,099	,302		.		**I				

Símbolos: Autocorrelaciones: * Límites del error estándar: .
 Total casos: 11

EREsinies

- Autocorrelaciones:

Re.	Auto- corre	Estandar Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Probabilidad
1	-,029	,264					*I					,012	,912
2	-,298	,251			.	*****I						1,428	,490
3	-,127	,237			.	***I						1,718	,633
4	-,074	,221			.	*I						1,829	,767
5	-,008	,205			.	*						1,831	,872
6	,043	,187			.	I*						1,884	,930
7	,013	,167			.	*						1,889	,966
8	-,013	,145			.	*						1,897	,984
9	,039	,118			.	I*						2,005	,991

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

- Autocorrelaciones Parciales:

Auto- Parcia	Estand. Error	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	-,029	,302				*I				
2	-,300	,302			.	*****I				
3	-,162	,302			.	***I				
4	-,204	,302			.	****I				
5	-,145	,302			.	***I				
6	-,105	,302			.	**I				
7	-,106	,302			.	**I				
8	-,100	,302			.	**I				
9	-,032	,302			.	*I				

Símbolos: Autocorrelaciones: * Límites del error estándar: .
Total casos: 11

ESTABILIDAD
ELIMINACIÓN DEL AÑO-INDIVIDUO 97 (Resultado del método Statist)

- CONSTRUCCION DE LA NUBE COMPROMISO = MEDIA PONDERADA DE LAS NUBES POR LOS COEFICIENTES
 VI = 0.276
 AU = 0.315
 SA = 0.302
 RE = 0.318

- HISTOGRAMA DE LOS 9 VALORES PROPIOS

NUMERO	VALOR PROPIO	PERCENT.	PERCENT. ACUMUL.
1	0.7669	39.60	39.60
2	0.5362	27.68	67.28
3	0.2674	13.81	81.09
4	0.2068	10.68	91.76
5	0.0879	4.54	96.30
6	0.0455	2.35	98.65
7	0.0131	0.68	99.33
8	0.0069	0.36	99.69
9	0.0060	0.31	100.00

- COORDENADAS, CONTRIBUCIONES Y COSENO CUADRADOS DE LAS LINEAS SOBRE LOS EJES PRINCIPALES (FACTORES A MODELIZAR)

LINEA	DISTO	COORDENADAS			CONTRIBUCIONES			COSENO CUADRADOS		
		1	2	3	1	2	3	1	2	3
87	0.30	0.40	-0.07	-0.01	0.21	0.01	0.00	0.53	0.02	0.00
88	0.45	0.61	0.17	0.02	0.48	0.05	0.00	0.82	0.06	0.00
89	0.25	0.08	-0.34	-0.34	0.01	0.21	0.42	0.02	0.44	0.44
90	0.26	-0.01	-0.35	0.37	0.00	0.22	0.51	0.00	0.45	0.52
91	0.08	-0.18	-0.18	-0.02	0.04	0.06	0.00	0.41	0.39	0.01
92	0.07	-0.21	-0.06	0.05	0.06	0.01	0.01	0.67	0.06	0.03
93	0.13	-0.30	-0.01	-0.08	0.12	0.00	0.02	0.70	0.00	0.04
94	0.11	-0.09	0.29	-0.01	0.01	0.16	0.00	0.08	0.78	0.00
95	0.12	-0.22	0.22	-0.05	0.07	0.09	0.01	0.43	0.43	0.02
96	0.17	-0.07	0.32	0.08	0.01	0.19	0.02	0.03	0.60	0.03

- NUEVOS FACTORES

Figura e
Nuevo Factor 1

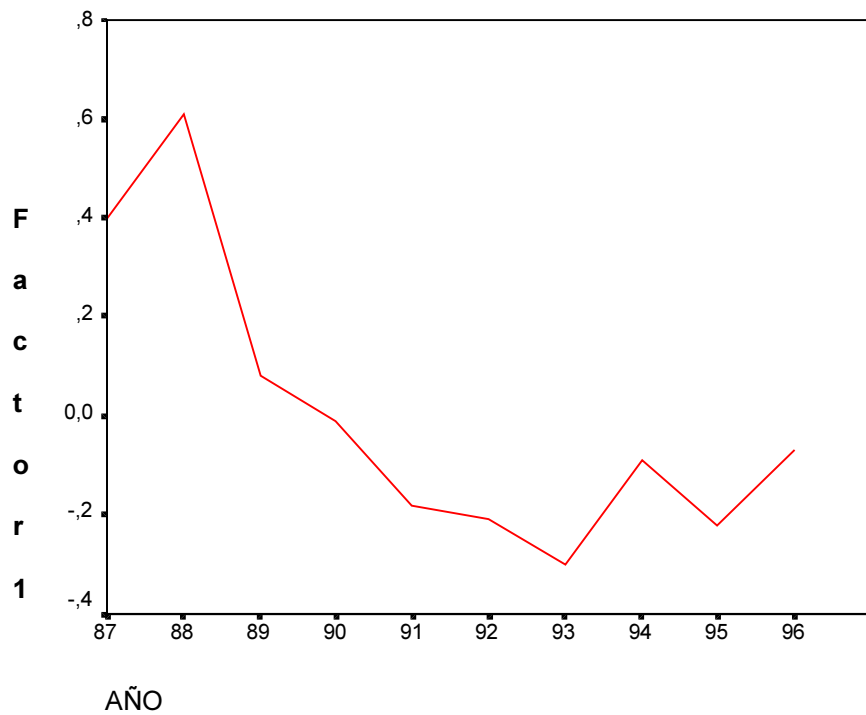


Figura f
Nuevo Factor 2

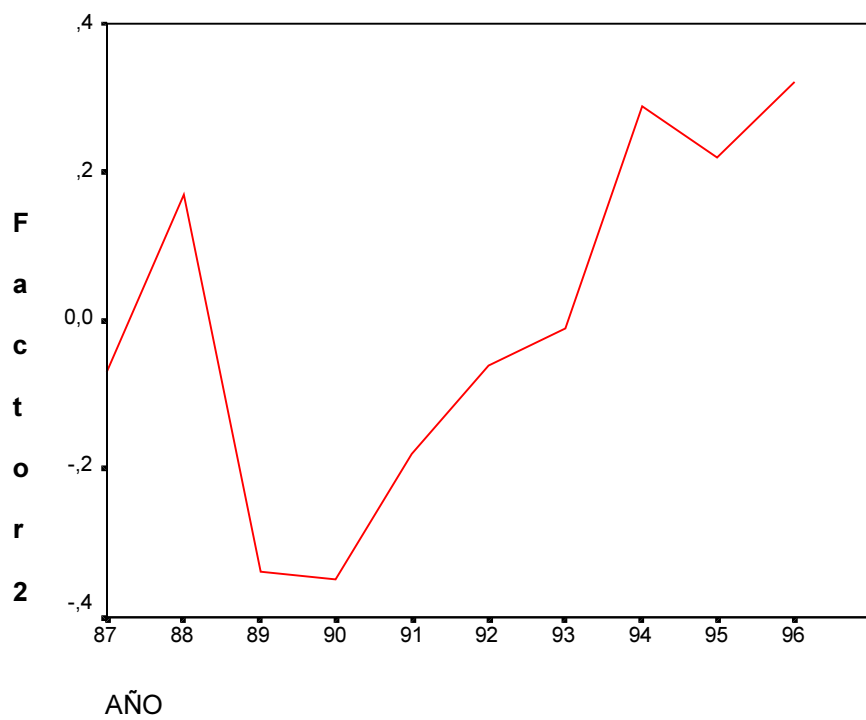
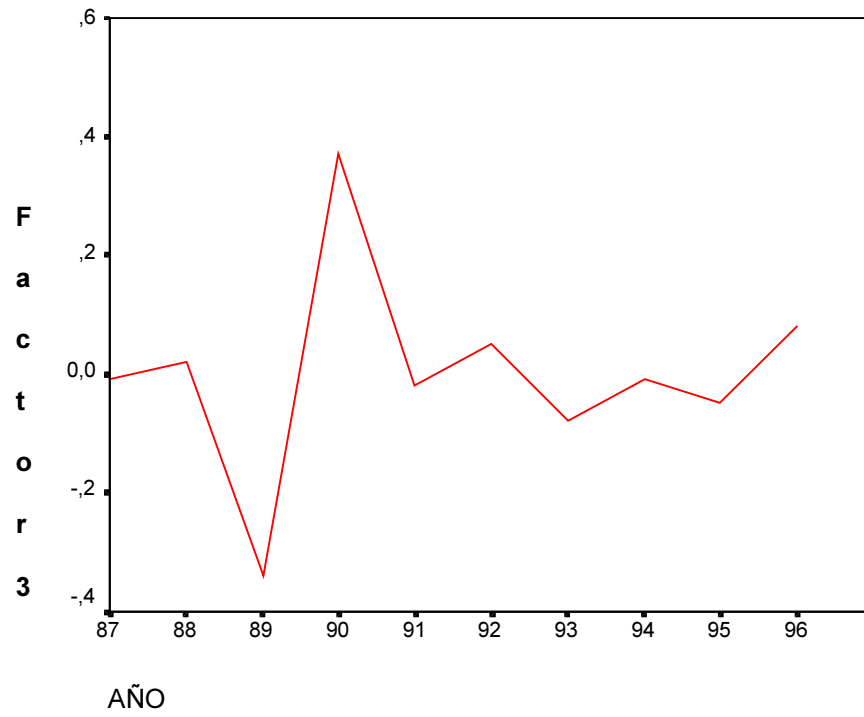


Figura g
Nuevo Factor 3



ELIMINACIÓN DEL AÑOS-INDIVIDUOS 96 Y 97 (Resultado del método Statist)

- CONSTRUCCION DE LA NUBE COMPROMISO = MEDIA PONDERADA DE LAS NUBES POR LOS COEFICIENTES

VI = 0.272
 AU = 0.309
 SA = 0.305
 RE = 0.315

- HISTOGRAMA DE LOS 9 VALORES PROPIOS

NUMERO	VALOR PROPIO	FORCENT.	FORCENT. ACUMUL.
1	0.8044	43.36	43.36
2	0.4702	25.34	68.70
3	0.2912	15.70	84.39
4	0.2102	11.33	95.72
5	0.0494	2.66	98.39
6	0.0135	0.73	99.12
7	0.0094	0.51	99.62
8	0.0070	0.38	100.00

- COORDENADAS, CONTRIBUCIONES Y COSENOS CUADRADOS DE LAS LINEAS SOBRE LOS EJES PRINCIPALES (FACTORES A MODELIZAR)

LINEA	COORDENADAS			CONTRIBUCIONES			COSENOS CUADRADOS			
	DISTO	1	2	3	1	2	3	1	2	3
87	0.30	0.40	-0.06	0.02	0.20	0.01	0.00	0.52	0.01	0.00
88	0.47	0.63	0.17	-0.05	0.49	0.06	0.01	0.84	0.06	0.01
89	0.25	0.05	-0.28	0.39	0.00	0.17	0.52	0.01	0.32	0.60
90	0.27	-0.03	-0.37	-0.36	0.00	0.29	0.44	0.00	0.50	0.47
91	0.07	-0.20	-0.13	0.03	0.05	0.04	0.00	0.55	0.23	0.01
92	0.07	-0.22	-0.01	-0.07	0.06	0.00	0.01	0.75	0.00	0.06
93	0.14	-0.32	0.05	0.06	0.13	0.01	0.01	0.73	0.02	0.03
94	0.14	-0.08	0.35	-0.04	0.01	0.26	0.01	0.05	0.89	0.01
95	0.14	-0.22	0.29	0.01	0.06	0.18	0.00	0.33	0.58	0.00

- NUEVOS FACTORES

Figura h
Nuevo Factor 1

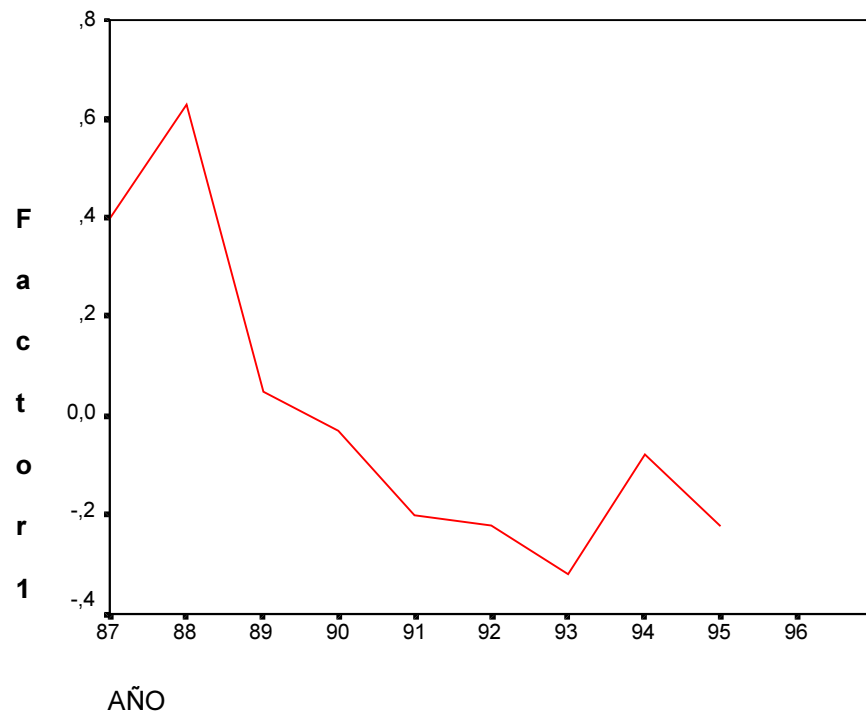


Figura i
Nuevo Factor 2

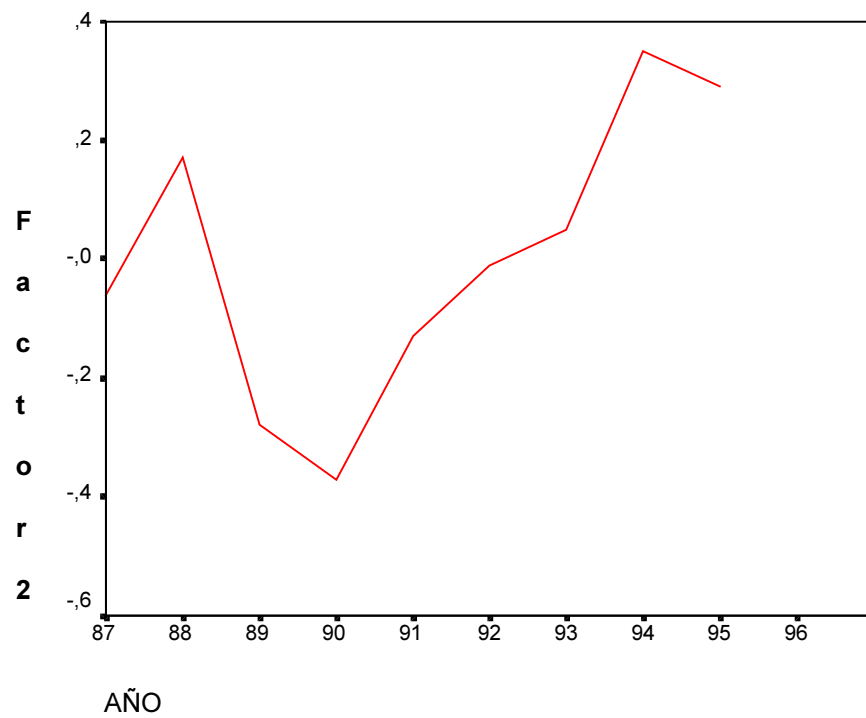


Figura j
Nuevo Factor 3

