

# Kent Academic Repository

## Full text document (pdf)

### Citation for published version

Mar Molinero, Cecilio and Segovia, Maria-Manuela and Contreras, Ignacio (2009) Efficiency Assessment in the Automobile Insurance Sector. *Revista de Metodos Cuantitativos Para la Economía y la Empresa*, 7 . pp. 57-76. ISSN 1886-516X.

### DOI

### Link to record in KAR

<http://kar.kent.ac.uk/25434/>

### Document Version

Publisher pdf

#### Copyright & reuse

Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

#### Versions of research

The version in the Kent Academic Repository may differ from the final published version.

Users are advised to check <http://kar.kent.ac.uk> for the status of the paper. **Users should always cite the published version of record.**

#### Enquiries

For any further enquiries regarding the licence status of this document, please contact:

[researchsupport@kent.ac.uk](mailto:researchsupport@kent.ac.uk)

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at <http://kar.kent.ac.uk/contact.html>

## Evaluación de la eficiencia de una cartera de asegurados en el sector del automóvil

SEGOVIA GONZÁLEZ, M. MANUELA

Departamento de Economía, Métodos Cuantitativos e Historia Económica

Universidad Pablo de Olavide

Correo electrónico: [mmseggon@upo.es](mailto:mmseggon@upo.es)

CONTRERAS RUBIO, IGNACIO

Departamento de Economía, Métodos Cuantitativos e Historia Económica

Universidad Pablo de Olavide

Correo electrónico: [iconrub@upo.es](mailto:iconrub@upo.es)

MAR MOLINERO, CECILIO

Kent Business School

University of Kent (United Kingdom)

Correo electrónico: [C.Mar-Molinero@kent.ac.uk](mailto:C.Mar-Molinero@kent.ac.uk)

### RESUMEN

La situación actual del sector del seguro del automóvil exige a las compañías aseguradoras diseñar productos cada vez más ajustados al tipo de cliente al que se dirige. En el presente trabajo analizamos los 80.000 registros ofrecidos por una entidad aseguradora con una importante cuota en el mercado español. Tras identificar una serie de perfiles en los clientes asegurados, se estudia la eficiencia de cada uno de ellos mediante la aplicación de Análisis Envoltante de Datos (DEA) y estadística multivariante. Los resultados obtenidos permiten identificar aquellos clientes que mejor rendimiento ofrecen a la compañía. Mediante la aplicación de test estadísticos paramétricos y no paramétricos, se determinan, además, las similitudes y diferencias existentes entre grupos de clientes que comparten una determinada característica.

**Palabras clave:** eficiencia; seguro de automóvil; análisis envoltante de datos; estadística multivariante.

**Clasificación JEL:** C44; D81; G22.

**2000MSC:** 62H25; 62J05.

# Efficiency Assessment in the Automobile Insurance Sector

## ABSTRACT

Insurance companies have to take risk and cost into account when pricing car insurance policies in order to design new products to cover the risk of private use of cars. In this paper we use data from 80,000 car insurance policies in order to assess, once risk and cost have been taken into account, the policies that generate the highest returns for the company. After different profiles of clients are identified, including among the characteristics the regional location of the clients, we use Data Envelopment Analysis and multivariate statistics. The results obtained not only permit us to identify the best groups of clients for the company but also to determine similarities and differences among profiles of clients.

**Keywords:** efficiency; car insurance; data envelopment analysis; multivariate statistics.

**JEL classification:** C44; D81; G22.

**2000MSC:** 62H25; 62J05.



## 1. INTRODUCCIÓN

En los últimos años se han ido produciendo importantes cambios en el mercado asegurador del automóvil. En nueve o diez años, se ha pasado de una situación en la que este sector operaba con márgenes técnicos negativos, a una situación en la que se ha desatado una lucha intensa, por parte de las compañías, para mejorar el posicionamiento en el mercado e incrementar su cuota de negocio en este ramo. Este cambio tan brusco se puede atribuir principalmente a dos causas. En primer lugar, la publicación de los baremos de lesionados ha permitido aumentar la estabilidad y control técnico del ramo. En segundo lugar, la aparición de un nuevo canal de distribución, la venta por teléfono e internet. No obstante, el comportamiento del sector del seguro del automóvil es el de un sector totalmente maduro. Sus características principales son un bajo crecimiento, una excesiva oferta, poca innovación de productos, productos estandarizados con muy poca diferenciación (salvo en el precio) y una fuerte competencia.

Las razones expuestas anteriormente, unidas a esta fase de madurez que atraviesa el sector, ha originado que se estén practicando, por parte de algunas compañías, políticas muy agresivas. Dichas prácticas no están justificadas técnicamente y su único fin es adquirir cuota en el mercado, aun a costa de ofrecer prestaciones que suponen una nula rentabilidad para la compañía que efectúa dicha oferta. Todo esto no ha favorecido el control y buen desarrollo del negocio asegurador del automóvil. Esto ha hecho que la mayoría de las compañías hayan optado por mantener y elaborar productos novedosos, con amplias coberturas para el cliente y precios competitivos. Con esta estrategia pretenden lograr en todo momento una política comercial, que mantenga cotas altas de competitividad en sus productos y huir de las variaciones y grandes oscilaciones que se observan en el mercado en relación con las políticas de precio de algunas compañías de seguros.

En el ramo del seguro de automóvil operan un gran número de aseguradoras que ofrecen productos muy similares. Por ello, las normas de contratación de cada compañía es una de las actividades de mayor trascendencia en cuanto a los resultados de su negocio. Dichas normas deben buscar el equilibrio técnico-comercial, de tal forma que, por un lado, sirvan de filtro para los riesgos que mayor siniestralidad presentan a priori y, por otro lado, permitan mantener lo más abierto posible el mercado potencial asegurador. Existen estudios en los que se analiza el comportamiento del asegurado ante la incertidumbre que puede provocar la existencia de riesgos adicionales. Éstos pueden venir dados por la posibilidad de que el asegurador no cumpla con el pago de la indemnización acordada en caso de siniestro, o bien por otro tipo de riesgos secundarios que afectan a la riqueza final del individuo. Esta incertidumbre puede provocar en el cliente la elección de una mayor cobertura para protegerse así indirectamente. Los primeros que abordaron este tipo de cuestiones fueron Mayers y Smith (1983) y Doherty y Schlesinger (1983). Asimismo, otras cuestiones similares fueron tratadas por Doherty y Schlesinger (1990) y Mahul (2000), entre otros. Por otra parte, hay trabajos que muestran de forma empírica cómo

el asegurado cuando está satisfecho con el trato recibido y con el servicio que se le ha dado, posee una mayor predisposición a renovar su póliza con la actual aseguradora (Pujol y Bolancé, 2004). No obstante, cuando el asegurado decide contratar o renovar su póliza, el precio o prima del seguro es un factor determinante, aunque no es el único. Las tarifas del ramo de automóviles están estructuradas en función de una serie de variables y los riesgos de cada producto tienen sus particularidades en cuanto a la tarificación. En este sentido destacamos los trabajos de Boj *et al.* (2004) y Guillén *et al.* (2005), en los que podemos encontrar análisis estadístico multivariante aplicado a la selección de factores de riesgo en la tarificación y el cálculo actuarial. Las compañías de seguros, utilizando su propia experiencia y las estadísticas suministradas por la patronal de entidades aseguradoras y reaseguradoras española (UNESPA), son capaces de identificar, *a priori*, la posible siniestralidad de grupos homogéneos de asegurados. Para ello, tienen en cuenta las características que son más relevantes a la hora de tener una mayor o menor siniestralidad. Dichas categorías son: el grupo de tarificación del vehículo, la zona de circulación, el uso del vehículo, las circunstancias personales del conductor y las bonificaciones y penalizaciones. Siendo dichas características las que tendremos en cuenta para definir los diferentes tipos de clientes con los que realizaremos nuestro trabajo empírico. La información de que disponemos nos ha sido facilitada por una compañía del sector del automóvil que alcanza importantes cuotas en el mercado asegurador español.

En el presente trabajo se busca estudiar la eficiencia de las pólizas vendidas en el sector del automóvil para uso privado teniendo en cuenta las características que determinan el nivel de riesgo, y los ingresos que de ellas se obtienen, a través del Análisis Envoltante de Datos (DEA).

De manera sintética, el DEA es una técnica no paramétrica de análisis de eficiencia de unidades de producción que a partir de múltiples *inputs* es capaz de producir múltiples *outputs* utilizando una estructura productiva homogénea. Denotamos cada unidad como UOA (*unit of assessment*) en vez del término tradicional DMU (*decision making unit*), siguiendo la opinión de Sarrico *et al.* (1997), ya que un perfil de asegurado no toma ninguna decisión. En este trabajo no se usa el concepto de DMU en el sentido tradicional de una unidad de gestión que toma sus propias decisiones. Por ello, utilizamos las siglas UOA, que extienden el concepto a unidad en evaluación. Para una introducción al análisis DEA véase, por ejemplo, Thanassoulis (2001) o Cooper *et al.* (2000).

Un elemento determinante en los resultados obtenidos a partir del DEA es la selección de los *inputs* y *outputs* que serán incluidos en el modelo. Cada combinación de *inputs* y *outputs* que se considere dará lugar a unos coeficientes de eficiencia distintos y, consecuentemente, a una ordenación diferente de las unidades de producción. Es más, una UOA puede ser eficiente o no dependiendo de cuales sean las variables que se incluyan en el modelo. Para salvar este inconveniente, proponemos aplicar la metodología propuesta en Serrano-Cinca *et al.* (2005) y

Serrano-Cinca y Mar-Molinero (2004). En ambos trabajos, se propone la utilización conjunta del análisis DEA y del Análisis de Componentes Principales (PCA) con el objetivo de guiar al decisor en la selección del modelo DEA. Asimismo, la metodología aplicada permitirá identificar las fortalezas y debilidades de una UOA en función de la forma en la que dicha unidad logra su nivel de eficiencia. Cada tipo de cliente es tratado como una UOA y lo que se pretende analizar es la eficiencia de cada unidad desde el punto de vista del rendimiento obtenido por la empresa aseguradora.

Para tal fin, propondremos dos *inputs* y dos *outputs* con los que evaluar nuestras unidades de producción. Como *inputs* incluimos en el análisis aquellas variables que suponen un coste para la compañía y que sintetizamos a través del número de siniestros total y el coste medio derivado de ellos. Como *outputs* consideramos los ingresos que generan dichos clientes a la compañía y que medimos a través de los valores de prima bruta y prima neta. En todos los casos, consideramos los valores por cada 100 clientes, de manera que eliminamos el efecto que pueda tener el que en cada grupo de clientes el número de observaciones sea distinto. La metodología que aplicamos consistirá en estimar todos los posibles modelos, obtenidos como combinación de las variables propuestas, y visualizar los resultados por medio del PCA. Los resultados obtenidos del PCA nos permitirán relacionar los modelos y las UOA, en tanto identificará similitudes y discordancias entre los diferentes casos.

El resto del trabajo se estructura de la siguiente manera. En la Sección 2 se presentan los datos con los que se analizan los diferentes tipos de clientes considerados mediante la aplicación de DEA y estadística multivariante. En la Sección 3 discutimos los resultados obtenidos y se obtienen algunos resultados adicionales. La Sección 4 está dedicada a conclusiones. Por último, al final del trabajo se recogen las referencias bibliográficas utilizadas en este trabajo.

## **2. ANÁLISIS DE EFICIENCIA DE LA CARTERA DE CLIENTES**

En la presente sección analizamos el comportamiento de diferentes tipos de clientes para la entidad aseguradora mediante la combinación de análisis DEA y técnicas de análisis multivariante.

### **2.1 Datos**

En primer lugar merece destacarse que resulta complejo, dadas las restricciones a que obliga la legislación, obtener información acerca de la cartera de clientes de una compañía aseguradora. No obstante, una compañía española con una amplia implantación a nivel nacional nos ha cedido los datos correspondientes al ejercicio del año 2002, sin incluir aquellos valores que eliminarían su carácter anónimo.

Para realizar nuestro estudio nos restringiremos a los individuos asegurados en dicho período que conduzcan un vehículo de tipo turismo y cuyo uso sea particular. El motivo por el que realizamos estas restricciones es que pretendemos realizar un estudio lo mas homogéneo posible. Si considerásemos otro tipo de vehículos, como es el caso de camiones, motos o taxis, se podrían distorsionar los resultados. Piénsese, por ejemplo, como los profesionales que trabajan con su vehículo están más expuestos al riesgo.

Después de llevar a cabo las restricciones anteriores y realizar una exhaustiva depuración de los datos nos quedamos con más de 80.000 asegurados. Sobre estos valores, realizamos una estratificación de la muestra según el siguiente detalle: sexo, edad, zona geográfica y tipo de turismo.

- La variable sexo se justifica por los previsibles diferentes comportamientos entre los hombres y las mujeres a la hora de conducir.
- La edad del conductor actúa como un elemento importante a considerar en los resultados del estudio. Distinguimos entre los individuos menores o iguales a 30 años y los individuos mayores de 30 años.
- La zona geográfica, ya que la climatología, la densidad de población y las infraestructuras actúan como componentes importantes en los resultados obtenidos. Se considerarán cuatro zonas distintas: la zona Norte, Centro, Mediterránea y Sur.
- El tipo de turismo, distinguiendo entre gama baja, media y alta. Un coche de gran tamaño y potencia puede provocar, a iguales circunstancias de accidente, mayores daños que uno pequeño. De la misma manera, también serán mayores los gastos de reparación de los daños propios del vehículo.

Los técnicos de la propia compañía aseguradora nos indicaron que esta estratificación es la que habitualmente realizan la mayoría de las compañías. En particular, cada una de ellas, al lanzar al mercado un producto, consideran además otras características. Esto es debido a que cuanto más se ajuste a la realidad de los clientes mayor éxito tendrá el nuevo producto, en el sentido de que aportará mayores beneficios económicos para la compañía y de satisfacción para el cliente.

La distribución de los asegurados en cuanto al tipo de turismo se presentan con unos porcentajes del 12,48%, 40,38% y 47,14%, para los vehículos de gama alta, media y baja, respectivamente. En cuanto al sexo tenemos que la proporción de hombres es muy superior a la de mujeres encontrando un 82,73% de hombres frente a un 17,27% de mujeres. La zona Sur es la que presenta un mayor porcentaje de asegurados con un 52,22%, seguida de la zona Mediterránea con un 18,31%, la zona Norte con un 17,67% y por último, la zona Centro con un 11,80%. Si consideramos la edad del conductor, nos encontramos con un menor porcentaje de

conductores jóvenes un 8,33% de asegurados menores o iguales a 30 años frente a un 91,67% de asegurados mayores de 30 años. Esta última característica se presenta de manera frecuente en la mayoría de las compañías del sector debido a las numerosas dificultades y a las grandes tarifas que aplican a la hora de asegurar a los individuos con poca edad y experiencia.

En la Tabla 1 se muestran los distintos perfiles o tipos de clientes que se obtienen al considerar la estratificación anterior.

		<i>Alta (A)</i>		<i>Intermedia (I)</i>		<i>Baja (B)</i>	
		<i>Mujer (F)</i>	<i>Hombre (M)</i>	<i>Mujer (F)</i>	<i>Hombre (M)</i>	<i>Mujer (F)</i>	<i>Hombre (M)</i>
<i>Norte (N)</i>	>30 (+)	AFN+	AMN+	IFN+	IMN+	BFC+	BMN+
	≤30 (-)	-	AMN-	IFN-	IMN-	BFC-	BMN-
<i>Centro (C)</i>	>30 (+)	AFC+	AMC+	IFC+	IMC+	BFC+	BMC+
	≤30 (-)	-	AMC-	IFC-	IMC-	BFC-	BMC-
<i>Mediterránea (M)</i>	>30 (+)	AFM+	AMM+	IFM+	IMM+	BFM+	BMM+
	≤30 (-)	-	AMM-	IFM-	IMM-	BFM-	BMM-
<i>Sur (S)</i>	>30 (+)	AFS+	AMS+	IFS+	IMS+	BFS+	BMS+
	≤30 (-)	AFS-	AMS-	IFS-	IMS-	BFS-	BMS-

**Tabla 1.** Perfiles de clientes.

Es importante destacar que en lugar de trabajar con 48 perfiles, que sería el resultado de aplicar la estratificación anterior, lo haremos con 45. Esto es debido a que existen tres perfiles en los que no disponemos de suficiente tamaño muestral para que la representación de ellos sea significativa. Los perfiles no considerados en nuestro estudio son los correspondientes a las mujeres con coche de gama alta, menores de 30 años de las zonas Norte, Centro y Mediterránea y que en la tabla se denotan con “-“.

## 2.2 Análisis DEA

Nuestro objetivo principal será proporcionar a la compañía el comportamiento de cada uno de los perfiles seleccionados. Todo el estudio se desarrollará desde el punto de vista de la entidad aseguradora, pretendiendo comparar la eficiencia de cada uno de los tipos de cliente recogidos en la Tabla 1.

Para tal fin aplicamos DEA, ya que permite estudiar la eficiencia relativa de cada una de las unidades, esto es, una unidad será o no eficiente en función de los resultados obtenidos por el resto, realizando comparaciones múltiples a través de modelos basados en la programación lineal. Un elemento determinante del resultado que se obtendrá con la aplicación del DEA es la combinación de *inputs* y *outputs* que se considere finalmente en el modelo.

En todos los trabajos existentes en la literatura en los que se aplica el DEA al sector asegurador, el objetivo que se persigue es analizar el comportamiento de las diferentes



compañías aseguradoras dentro de la industria del seguro. Algunos trabajos realizados en este campo son, entre otros, Cummins y Rubio-Misas (1998), Fecher *et al.* (1993), Fukuyama (1997), Tone y Sahoo (2005) o Kao y Hwang (2008). Ninguno de los trabajos citados analiza el problema desde el punto de vista del cliente, en todos ellos se estudia la eficiencia de las diferentes aseguradoras. Aun así, en ellos se considera como *outputs* de los modelos el número de primas o los ingresos derivados de las mismas, que pueden considerarse, sin duda, un elemento determinante en este tipo de análisis.

En el presente trabajo definimos a la unidad evaluada como un tipo de riesgo para la compañía. La información relevante para la aseguradora sobre el mismo puede sintetizarse en dos grupos de variables: aquellas que suponen un coste, de manera directa o indirecta y que, en cualquier caso, valores bajos de estas variables identifican a los buenos riesgos. Por otro lado, han de tenerse en cuenta aquellos conceptos que suponen un ingreso para la compañía. En este segundo caso valores altos identifican a los mejores riesgos. Los primeros conceptos se asocian con los *inputs* del modelo mientras que los segundos se identifican con los *outputs*.

Para los primeros hemos considerado dos variables, una estimación de la probabilidad del siniestro y el coste medio de los mismos. Para los segundos consideramos las primas ingresadas a partir de cada tipo de cliente. De manera sintética, las variables consideradas son las siguientes:

*Input A*: número de siniestros por cada 100 asegurados.

*Input B*: coste medio de los siniestros por cada 100 asegurados (euros).

*Output 1*: ingresos por prima bruta por cada 100 asegurados (euros).

*Output 2*: ingresos por prima neta por cada 100 asegurados (euros).

Los ingresos de prima neta, *output 2*, se obtienen como ingresos de prima bruta menos las bonificaciones asignadas por la compañía. De esta forma tendremos en cuenta en nuestro modelo el comportamiento de los asegurados en años anteriores a través de las bonificaciones asignadas a cada cliente. Esta variable captura de alguna forma la calidad y fidelidad de los clientes en tanto las compañías ofrecen este tipo de bonificaciones con el objetivo de retener a sus mejores asegurados. Algunos trabajos relacionando las primas con la experiencia pasada de los individuos en cuanto a la siniestralidad son, entre otros, Lemaire (1985), Henriet y Rochet (1986) y Dionne y Vanasse (1992).

De manera resumida, en la Tabla 2 se muestra la información de *inputs* y *outputs* recogidos en los modelos. En todos los casos consideramos los valores por cada 100 asegurados para evitar el efecto que pueda tener el diferente número de asegurados en cada categoría.

	<i>Inputs</i>		<i>Outputs</i>	
	A	B	1	2
<i>Máx</i>	76,92	51.009,10	142.989,31	99.631,13
<i>Mín</i>	21,28	1.125,38	48.812,08	30.016,87
<i>Media</i>	37,49	17.774,16	88.362,13	58.729,66
<i>Desv. Std.</i>	9,70	8.743,68	26.381,49	18.732,99

**Tabla 2.** Valores de *inputs* y *outputs*.

Para la evaluación de las unidades, hemos aplicado el modelo con rendimientos a escala constante y orientación al output, modelo CCR-O. La orientación al output se justifica porque la empresa aseguradora no puede controlar ni el riesgo del siniestro ni su costo medio, aunque sí que puede controlar los ingresos que los clientes generan a través de la política de tarifas.

La manera tradicional de evaluar las unidades consiste en resolver el modelo que incluye simultáneamente las cuatro variables, *inputs* A y B y *outputs* 1 y 2. La metodología que aplicaremos propone resolver todos los modelos que surgen de combinar las cuatro variables descritas. Con dos *inputs* y dos *outputs*, pueden obtenerse 9 modelos diferentes. La notación utilizada para los modelos es la que aparece en la Tabla 3.

<i>Modelo</i>	<i>Inputs</i>	<i>Outputs</i>
A1	A	1
AB1	A,B	1
B1	B	1
A2	A	2
AB2	A,B	2
B2	B	2
A12	A	1,2
AB12	A,B	1,2
B12	B	1,2

**Tabla 3.** Modelos DEA.

La razón principal para resolver todos los posibles modelos es doble. Por un lado, como se dijo, la eficiencia de una unidad dependerá de la combinación de *inputs* y *outputs* que se considere, por lo que tiene sentido evaluar a cada UOA en función de todos los posibles modelos que puedan construirse. Por otro, estamos interesados en estudiar las fortalezas o debilidades de cada unidad con respecto a cada variable y esta información puede obtenerse estudiando los coeficientes correspondientes a los distintos modelos. En la Tabla 4 se representan los valores de eficiencia de las 45 unidades de producción en los 9 modelos considerados. Este valor aparece expresado en porcentaje.

Un primer análisis de los resultados obtenidos permite comprobar cómo la eficiencia obtenida por una UOA particular depende de cómo se especifique el modelo, tal y como puede verse en la Tabla 4.

Considerando el modelo general, modelo AB12, solo dos unidades son eficientes: IFN- y AFS-. Todas coinciden con ser mujeres, con coches de gama media o alta y jóvenes (menores de 30). Estos dos grupos de clientes son los que mejor comportamiento tendrían para la compañía si consideramos todas las variables explicativas. Podemos ver cómo este buen comportamiento se debe, no obstante, a distintos factores en cada caso.

Para AFS-, el coeficiente de eficiencia de 100 en el modelo general contrasta con menores puntuaciones obtenidas en los modelos en los que el único *input* considerado es el A, número de siniestros por cada 100 asegurados. Esto indica que el buen resultado obtenido se debe al comportamiento de este tipo de clientes en las restantes variables o, lo que es lo mismo, a pesar de mostrar unos malos registros en cuanto al número de accidentes, el buen comportamiento en el *input* restante, así como en la cuantía de las primas, son suficientes para compensarlo y hacer que la unidad sea considerada eficiente.

En cuanto a la unidad IFN-, se puede ver como todos sus coeficientes son 100 excepto en aquellos modelos en los que el único *input* es el coste de los siniestros (B1, B2 y B12). En este caso, la eficiencia de la unidad se explica por un buen comportamiento en el *input* A, así como por una buena combinación en las variables que determinan la eficiencia global.

En la siguiente sección se realiza un análisis multivariante sobre los coeficientes DEA en el que se identifican las principales características de los valores de la Tabla 4. En dicha tabla nos encontramos con un gran volumen de información acerca de las unidades de producción eficientes, aunque también nos podemos encontrar con información redundante ya que algunos modelos DEA pueden ser equivalentes (Serrano-Cinca y Mar-Molinero, 2004). Para lograr todo lo anterior, se aplicarán técnicas de análisis multivariante, tales como, el análisis factorial y análisis de regresión múltiple.

### **2.3 Análisis factorial**

El análisis factorial es una técnica de reducción de datos estándar que elimina la información redundante, resalta la información más significativa del conjunto de datos y ayuda a visualizar las principales relaciones existentes entre cada una de las observaciones (Dunteman, 1989).

En los valores recogidos en la Tabla 4, los diferentes modelos DEA pueden considerarse como variables y los perfiles definidos anteriormente como observaciones de nuestro estudio, de forma que disponemos de 45 observaciones y 9 variables. Puede ocurrir que con un conjunto más pequeño de datos podamos explicar las relaciones existentes entre cada uno de los perfiles considerados. Para tal fin, aplicamos la técnica del análisis factorial.

<i>UOA</i>	<i>A1</i>	<i>AB1</i>	<i>B1</i>	<i>A2</i>	<i>AB2</i>	<i>B2</i>	<i>A12</i>	<i>AB12</i>	<i>B12</i>
AFN+	60	60	31	54	54	25	60	60	31
AFC+	34	37	24	33	35	21	34	37	24
AFM+	43	43	15	42	42	13	43	43	15
AFS+	34	34	17	34	34	15	34	34	17
AFS-	67	100	100	75	100	100	75	100	100
AMN+	58	58	24	51	51	19	58	58	24
AMN-	86	86	39	85	85	35	86	86	39
AMC+	44	44	25	41	41	21	44	44	25
AMC-	34	34	10	36	36	9	36	36	10
AMM+	47	47	20	44	44	17	47	47	20
AMM-	47	60	45	51	61	44	51	61	45
AMS+	40	41	25	39	40	22	40	41	25
AMS-	53	53	25	57	57	24	57	57	25
IFN+	48	48	24	42	42	19	48	48	24
IFN-	100	100	59	100	100	53	100	100	59
IFC+	36	45	34	32	38	28	36	45	34
IFC-	61	63	39	61	64	35	61	64	39
IFM+	38	38	16	35	35	13	38	38	16
IFM-	51	51	21	52	52	19	52	52	21
IFS+	27	31	21	26	28	18	27	31	21
IFS-	39	48	35	42	49	34	42	49	35
IMN+	51	51	25	42	42	18	51	51	25
IMN-	55	55	25	55	55	22	55	55	25
IMC+	40	46	31	36	40	24	40	46	31
IMC-	35	35	15	37	37	14	37	37	15
IMM+	44	44	24	38	38	19	44	44	24
IMM-	34	34	16	36	36	15	36	36	16
IMS+	27	31	21	25	27	17	27	31	21
IMS-	31	31	17	34	34	17	34	34	17
BFC+	39	43	28	34	36	21	39	43	28
BFC-	51	63	46	49	58	40	51	63	46
BFC+	29	38	30	26	32	24	29	38	30
BFC-	47	59	44	51	60	43	51	60	44
BFM+	37	41	26	33	35	21	37	41	26
BFM-	35	38	24	37	39	23	37	39	24
BFS+	24	29	20	22	25	17	24	29	20
BFS-	25	33	26	28	33	25	28	33	26
BMN+	48	48	26	40	40	19	48	48	26
BMN-	40	40	23	40	40	20	40	40	23
BMC+	38	44	29	33	37	23	38	44	29
BMC-	28	37	30	28	35	27	28	37	30
BMM+	40	42	26	35	36	20	40	42	26
BMM-	25	26	16	28	29	16	28	29	16
BMS+	27	31	21	24	27	17	27	31	21
BMS-	26	29	19	28	31	18	28	31	19

**Tabla 4.** Coeficientes de eficiencia DEA (%).

Para estudiar el número de factores que se deben incluir en el modelo, se empezó por realizar un análisis de componentes principales. Los valores propios obtenidos aparecen en la Tabla 5. Es importante apuntar que el mínimo valor para la extracción de autovalores ha sido tomado como 0,8, siguiendo la recomendación de Joliffe (1972). Se obtienen dos autovalores que son mayores al límite expuesto. La primera componente aporta, con gran diferencia, la mayor parte de información, explicando un 86,03% de la variabilidad total. Considerando las

dos primeras componentes se explica un 99,17%, por lo que muestran una adecuada representación de los datos. Por esta razón, el modelo factorial contiene tan solo dos factores.

<i>Componente</i>	<i>Autovalor</i>	<i>% de varianza</i>	<i>Acumulado %</i>
PC1	7,74	86,03	86,03
PC2	1,18	13,14	99,17

**Tabla 5.** Resultados del análisis de componentes principales.

En la Tabla 6 exponemos las correlaciones existentes entre cada uno de los modelos y los dos factores. Los resultados se muestran ordenados respecto a las correlaciones con el primer factor. Podemos observar que el primer factor está fuertemente correlacionado y de forma positiva con todos los modelos. Dicho factor representa una “medida global de eficiencia”, tomando el máximo valor cuando consideramos el modelo completo AB12, seguido de los modelos con dos *inputs* (AB1, AB2). Para interpretar el segundo factor tenemos que tener en cuenta tanto el valor de las correlaciones como el signo de las mismas. Encontramos que dicho factor está correlacionado de forma positiva con los modelos B1, B2 y B12 y de forma negativa con los modelos A1, A2 y A12. La correlación positiva ocurre con aquellos modelos cuyo único *input* es B y la correlación negativa con aquellos modelos cuyo único *input* es A. Puede concluirse que la segunda componente discrimina entre modelos cuya eficiencia se concentra en el coste del siniestro y aquellos que la derivan del número de siniestros. Dicho factor puede denotarse como “coste del siniestro *versus* número de siniestros”.

	<i>Factor</i>	
	<i>F1</i>	<i>F2</i>
AB12	0,9938	-0,0900
AB1	0,9923	-0,0825
AB2	0,9872	-0,0872
A2	0,9343	-0,3335
A12	0,9293	-0,3651
A1	0,9016	-0,4205
B1	0,8712	0,4875
B12	0,8712	0,4875
B2	0,8536	0,5131

**Tabla 6.** Análisis factorial, matriz de correlaciones.

Las puntuaciones de cada perfil respecto a cada factor ofrecen una medida de eficiencia de cada uno de los conceptos anteriores, eficiencia global y eficiencia coste frente al número de siniestros. La Figura 1 muestra la representación de las puntuaciones en los dos factores (F1 y F2). Los perfiles con mayores puntuaciones en el primer factor son aquellos con mayor eficiencia global, destacando los perfiles AFS- y IFN-. Por otro lado en el extremo izquierdo tenemos a IFS+ y AFC+, que es un tipo de cliente que muestra una baja eficiencia en todos los modelos considerados. Es importante destacar que la eficiencia se incrementa según nos

desplazamos de izquierda a derecha en la Figura 1, lo que confirma la interpretación del primer factor como una medida de eficiencia global.

El segundo factor se identifica como el valor que distingue entre la eficiencia debida al coste y al número de los siniestros. En la parte superior derecha de la Figura 1 destaca AFS- con un coeficiente de eficiencia igual a 100 salvo en los modelos que contienen únicamente el *input* A. Por el contrario, en la parte inferior derecha podemos ver los perfiles AMN- y IFN- que ven reducida su eficiencia cuando el único *input* es B.

Pueden estudiarse las relaciones entre cada modelo y el valor obtenido en las puntuaciones factoriales. Para ello, se ha realizado un análisis de regresión en el que como variable dependiente hemos tomado la eficiencia de cada uno de los modelos y como variables independientes las puntuaciones de los dos primeros factores (Schiffman *et al.*, 1981). De esta forma, podremos constatar que representar los distintos modelos de eficiencia en términos de sus puntuaciones factoriales resulta adecuado y estaremos reduciendo la dimensionalidad de nuestro problema. Asimismo, podremos obtener modelos similares de eficiencia, puesto que el ángulo entre cualesquiera dos vectores está relacionado con la correlación existente entre la eficiencia generada por los dos modelos en cuestión. En la Tabla 7 mostramos el coeficiente de determinación ajustado y los coeficientes de regresión junto con los estadísticos de contraste para ver si son significativos o no. En nuestro caso todos los valores de  $R^2$  son altos, todos son superiores a 0,98.

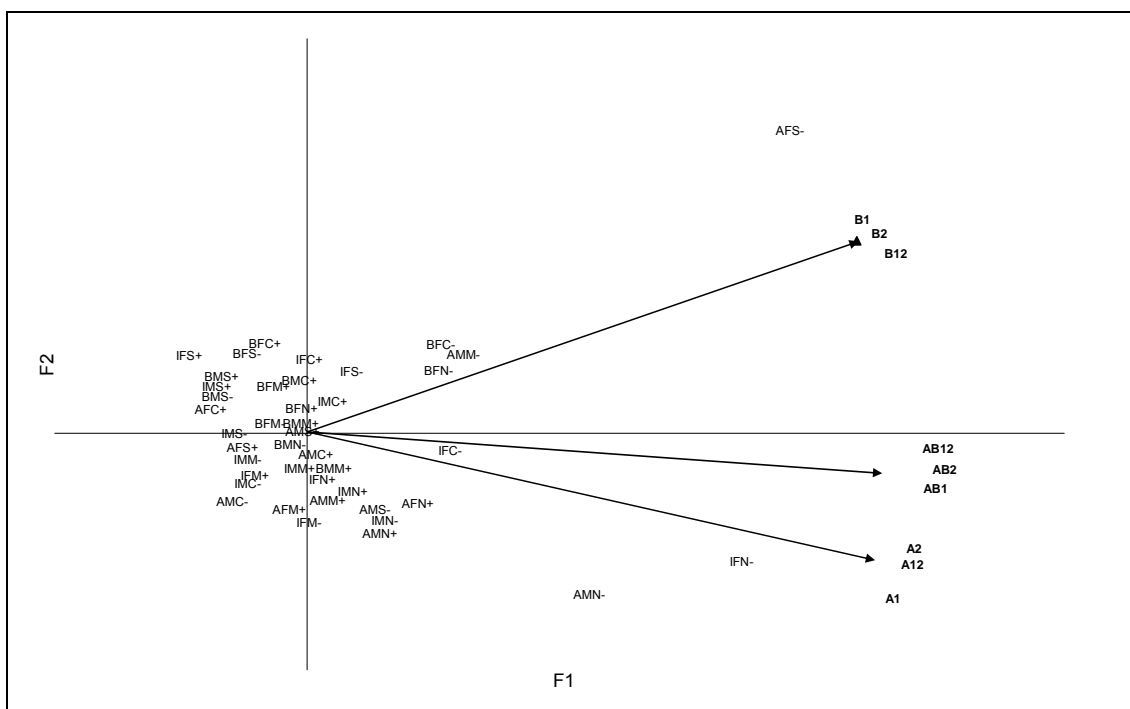
	<i>Directional cosines</i>		<i>F</i>	<i>Aj R<sup>2</sup></i>
	<i>φ1</i>	<i>φ2</i>		
A1	0,902	-0,420	1996,172	0,989
	57,263**	-26,707**		
AB1	0,992	-0,083	2454,588	0,991
	69,825**	-5,806**		
B1	0,871	0,487	6338,496	0,997
	98,257**	54,978**		
A2	0,934	-0,334	1304,407	0,983
	48,104**	-17,172**		
AB2	0,987	-0,087	1162,574	0,981
	48,033**	-4,243**		
B2	0,854	0,513	2536,086	0,991
	61,041**	36,691**		
A12	0,929	-0,365	6854,379	0,997
	108,978**	-42,808**		
AB12	0,871	0,487	6338,496	0,997
	98,257**	54,978**		
B12	0,994	-0,090	4915,243	0,996
	98,745**	-8,940**		

\*\*Significativo al nivel 0,01.

**Tabla 7.** Análisis ProFit. Resultados de la regresión lineal.

En la Figura 1 representamos igualmente los vectores normalizados obtenidos al realizar la regresión. Pueden diferenciarse tres grupos de modelos. En la gráfica, los modelos se representan con vectores normalizados y los perfiles de clientes mediante las etiquetas correspondientes.

Un primer grupo de modelos es el formado por A1, A2 y A12, en los que el único *input* es A y aproxima la eficiencia con respecto al número de siniestros. Un segundo grupo formado por B1, B2 y B12, en el que el único *input* es B, permite ver la eficiencia con respecto al coste medio del siniestro. Por último, un tercer grupo formado por los restantes modelos mide la eficiencia global de la unidad. De cada uno de estos grupos se ha seleccionado un modelo para indicar la dirección de mejora de cada una de las eficiencias mencionadas anteriormente, respectivamente, A12, B12 y AB12.

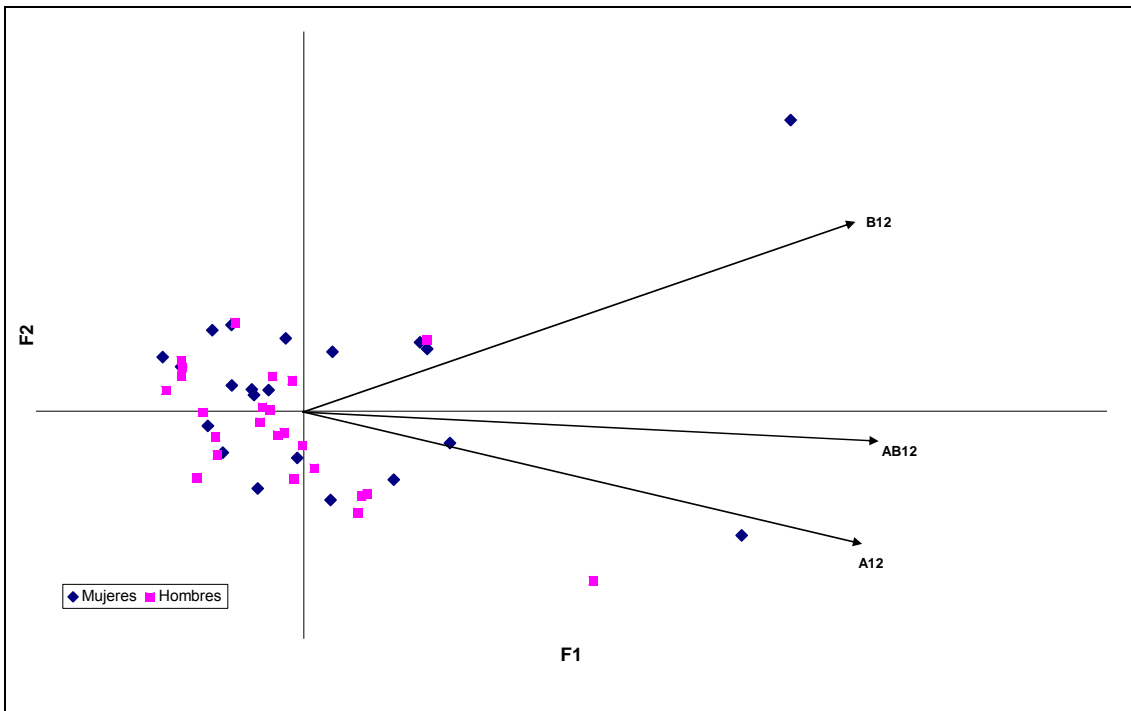


**Figura 1.** Gráfico de cargas factoriales, F1 vs. F2.

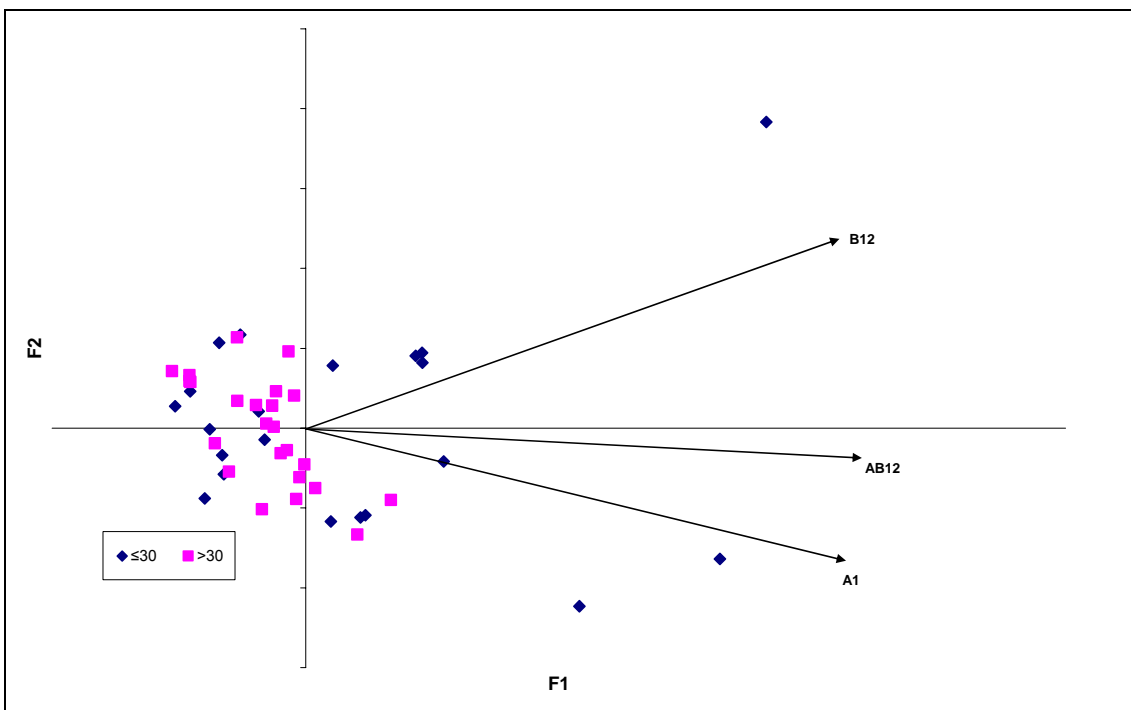
### 3. DISCUSIÓN DE LOS RESULTADOS

A partir de los valores presentados en la Sección 2 pueden obtenerse algunas conclusiones sobre el comportamiento de los perfiles de clientes considerados.

En las Figuras 2 y 3 se representan, respectivamente, los valores de las 45 unidades con respecto a los factores F1 y F2, diferenciando los asegurados por sexo y edad. Puede apreciarse como los clientes femeninos (Figura 2) y menores de 30 (Figura 3) tienen un mejor comportamiento para la compañía, esto es, obtienen mayores valores de eficiencia.



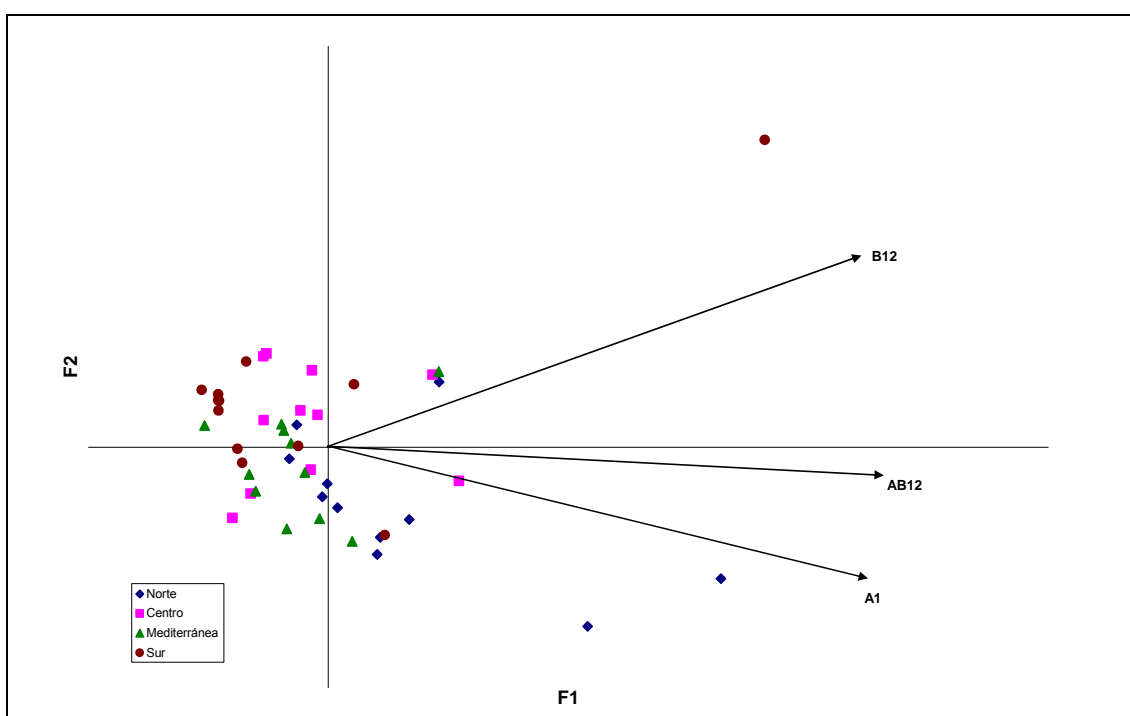
**Figura 2.** F1 vs. F2. Los perfiles se identifican por el sexo de los clientes.



**Figura 3.** F1 vs. F2. Los perfiles se diferencian por edad.



Con respecto a las zonas de circulación de los vehículos, se comprueba que existen diferencias significativas tanto si se considera una medida de eficiencia global como si se considera la eficiencia respecto al número de siniestros. Estas diferencias, que se aprecian en la Figura 4, se corroboran a partir de los resultados obtenidos con el test de Kruskal-Wallis (p-valor 0,006 y 0,003, respectivamente). Con el test de Dunn de comparaciones múltiples, comprobamos que existen diferencias significativas con ambas medidas de eficiencia entre la zona Norte y las zonas Sur, Mediterránea y Centro, siendo más eficientes los conductores de la zona Norte.



**Figura 4.** F1 vs. F2. Los perfiles se identifican por la zona geográfica.

En cuanto a la gama del vehículo, también se detectan diferencias en el modelo A12 a través del test de Kruskal-Wallis (p-valor 0,017). Los resultados del test de Dunn de comparaciones múltiples muestran que existen diferencias significativas entre los coches de gama alta y baja, siendo superior en los primeros. La eficiencia en cuanto al número de siniestros es mayor en los coches de gama alta.

Junto a los anteriores estudios globales para detectar diferencias entre los diferentes tipos de clientes, es posible realizar estudios parciales. Para ello, consideramos una variable y segmentamos la información con respecto a cualquier otra. Los resultados obtenidos, así como el test utilizado en cada caso, se recogen en la Tabla 8. En cada caso, pueden considerarse las tres medidas de eficiencia utilizadas hasta ahora, eficiencia global, eficiencia respecto al número de siniestros y eficiencia respecto al coste y que representamos, respectivamente, por los modelos AB12, A12 y B12.

<i>Variable</i>	<i>Segmentación</i>	<i>Modelo</i>	<i>Test (p-valor)</i>	<i>Más eficiente / Menos eficiente</i>
Sexo				
	Edad ≤30	B12	U (0.009)	Mann-Whitney Mujeres/Hombres
Edad				
	Mujeres	AB12	U (0.009)	Mann-Whitney ≤30 / >30
		A12	U (0.018)	Mann-Whitney ≤30 / >30
		B12	U (0.006)	Mann-Whitney ≤30 / >30
Área Geográfica				
	Hombres	AB12	Kruskal-Wallis (0,048)	Norte/Sur
		A12	Kruskal-Wallis (0,030)	Norte/Sur
	Gama intermedia	A12	Kruskal-Wallis (0,046)	Norte/Sur
	Gama baja	AB12	Kruskal-Wallis (0,034)	Centro/Medit., Norte/Sur
		A12	Kruskal-Wallis (0,019)	Norte/Sur, Centro/Sur
		B12	Kruskal-Wallis (0,029)	Centro/Sur, Centro/Medit.
	Edad >30	AB12	ANOVA (0,0005)	Norte/Sur, Norte/Centro, Norte/Medit., Centro/Sur, Medit./Sur
		A12	ANOVA (0,0005)	Norte/Centro, Norte/Sur, Medit./Sur
		B12	ANOVA (0,003)	Centro/Medit., Centro/Sur
Gama				
	Hombres	AB12	Kruskal-Wallis (0,049)	Alta/Baja
		A12	Kruskal-Wallis (0,028)	Alta/Baja
	Edad ≤30	A12	Kruskal-Wallis (0,032)	Alta/Baja
	Zona Sur	AB12	Kruskal-Wallis (0,030)	Alta/Baja
		A12	Kruskal-Wallis (0,037)	Alta/Baja

**Tabla 8.** Resultados de los análisis parciales.

En los siguientes análisis trabajaremos a un nivel de confianza del 95%. En primer lugar, comparamos la eficiencia entre hombres y mujeres. Para ello, segmentamos la información con respecto a las variables edad, zona geográfica y gama del coche. Si consideramos por separado los dos grupos de edad, obtenemos que en el grupo de menores de 30 años existen diferencias significativas en cuanto al comportamiento de hombres y mujeres si lo evaluamos por medio del modelo B12, siendo las mujeres más eficientes para la compañía. Este resultado se obtiene a partir del test no paramétrico de la U de Mann-Whitney, en tanto no se verifican condiciones de normalidad, obteniendo un p-valor igual a 0,009. Para los individuos mayores de 30 años no se han observado diferencias significativas con respecto al sexo, al igual que si tomamos las distintas zonas de circulación o las distintas gamas del vehículo.

En segundo lugar, comparamos la eficiencia con respecto a la edad. Al segmentar por la variable sexo, observamos que en cuanto a los hombres no existen diferencias significativas. Sí aparecen diferencias en el grupo de mujeres, siendo más eficientes las menores de 30 años.

En tercer lugar, comparamos la eficiencia con respecto a la zona geográfica en la que reside el conductor. Segmentando por la variable sexo, obtenemos que en los hombres existen diferencias significativas en cuanto a la zona de circulación, tanto en la eficiencia global como

en cuanto al número de siniestros (Kruskall-Wallis, 0,048 y 0,030, respectivamente), entre las zonas Norte y Sur. Aparecen diferencias de igual forma cuando segmentamos por la gama del vehículo. En los vehículos de gama baja aparecen diferencias significativas en los tres modelos de eficiencia. En cuanto a los coches de gama intermedia, las diferencias las encontramos entre las zonas Norte y Sur según el modelo A12. Segmentando por la variable edad, confirmamos que existen diferencias significativas para el grupo de conductores mayores de 30 años en los tres casos considerados.

En cuarto lugar compararemos la eficiencia con respecto a los distintos tipos de vehículos. Si segmentamos por la variable sexo o zona geográfica, encontramos que en los hombres y en la zona Sur existen diferencias significativas en cuanto a la eficiencia debida a los modelos AB12 y A12. En cuanto a la edad, las diferencias se encuentran respecto del modelo A12, para el grupo de conductores menores de 30 años. En todos los casos, la eficiencia de la gama alta es superior a la gama baja.

#### **4. CONCLUSIONES**

En el presente trabajo hemos estudiado el comportamiento de los diferentes tipos de asegurados que integran la cartera de clientes de una compañía aseguradora. A partir de los datos proporcionados por una aseguradora española con una elevada cuota de mercado, hemos analizado el comportamiento de cada tipo de cliente en términos de eficiencia. Se ha llevado a cabo aplicando conjuntamente el análisis envolvente de datos y técnicas de estadística multivariante.

Uno de los principales determinantes del análisis DEA es la elección de los *inputs* y *outputs* que forman el modelo. La elección de una u otra variable determina el resultado que se obtiene del análisis. Para salvar dicha dificultad se han aplicado todos los posibles modelos que se tienen con los *inputs* y *outputs* seleccionados. La gran cantidad de información obtenida se estudia y sintetiza con técnicas de estadísticas multivariante como el análisis factorial y análisis de regresión.

El primer resultado que se obtiene es la existencia de tres indicadores de eficiencia, “eficiencia global”, “eficiencia en el número” y “eficiencia en el coste de los siniestros”. En todos los casos, la medida de eficiencia hace referencia al resultado obtenido por la compañía para ese tipo de cliente.

En términos generales, se identifican los comportamientos más interesantes para la empresa en asegurados jóvenes, mujeres, ubicados en la zona Norte y que conducen vehículos de gama alta. Además de las comparaciones directas entre las variables, se han obtenido resultados parciales segmentando la información según las diferentes variables consideradas en

el estudio, lo que permite hacer comparaciones entre cada grupo de asegurados considerando por separado las tres medidas de eficiencia.

Los resultados obtenidos del análisis permitirían a la empresa aseguradora optimizar la composición de su cartera de clientes. Cada unidad puede considerarse como un tipo de riesgo para la compañía, el trabajo identifica aquellos más interesantes, clientes con altos valores de eficiencia, en los que los valores de las primas son suficientemente altos como para compensar los valores de siniestralidad y coste derivado de los mismos. De esta manera, la aseguradora puede optar por concentrar sus esfuerzos en atraer a este tipo de clientes incluso ajustando el valor de las primas.

De igual modo, se identifican a los clientes menos eficientes para la compañía. En este caso, el valor de las primas que se ingresa en comparación con la siniestralidad y coste derivado no resulta rentable para la compañía en términos relativos, esto es, la rentabilidad que obtiene la empresa aseguradora es menor que la que obtiene de clientes con mayores coeficientes de eficiencia. En este caso, la empresa puede reconsiderar el papel de este tipo de clientes o ajustar al alza sus primas para igualar su rendimiento al de clientes más eficientes.

## **BIBLIOGRAFÍA**

- BOJ, E., CLARAMUNT, M.M. Y FORTIANA, J. (2004): *Análisis multivariante aplicado a la selección de factores de riesgo en la tarificación*. Cuadernos de la Fundación MAPRE estudios. Instituto de Ciencias del Seguro. Madrid.
- COOPER, W.W., SEIFORD, L.M. Y TONE K. (2000): *Data Envelopment Analysis: a comprehensive text with models, applications, references*. Kluwer Academic Publisher, Boston.
- CUMMINS, J.D. Y RUBIO-MISAS, M. (1998): "M. Efficiency in the Spanish insurance industry". *Working paper, The Wharton School, University of Pennsylvania*.
- DIONNE, G. Y VANASSE, C. (1992): "Automobile insurance ratemaking in the presence of asymmetrical information". *Journal of Applied Econometrics*, 7 (2); pp.149–166.
- DOHERTY, N.A. Y SCHLESINGER, H. (1983): "Optimal insurance in incomplete markets". *Journal of Political Economy*, 91 (6); pp.1045–1054.
- DOHERTY, N.A. Y SCHLESINGER, H. (1990): "Rational insurance purchasing: consideration of contract non-performance". *Quarterly Journal of Economics*, 105; pp.143–153.
- DUNTEMAN, G.H. (1989): "Principal components analysis". *Series: Quantitative Applications in the Social Sciences*, nº 69. Series Editor: Michael S. Lewis-Beck, University of Towa.
- FECHER, F., KESSLER D., PERELMAN S. Y PESTIEAU P. (1993): "Productive performance of the French insurance industry". *Journal of Productivity Analysis*, 4; pp.77–93.

- FUKUYAMA, H. (1997): "Investigating productive efficiency and productivity changes of Japanese life insurance companies". *Pacific-Basin Finance Journal*, 5; pp.481–509.
- GUILLEN, M., AYUSO, M., BERMÚDEZ, L. Y MORILLO, I. (2005): *El Seguro de automóviles: estado actual y perspectiva de la técnica actuarial*. Fundación MAPFRE estudios. Instituto de Ciencias del Seguro. Madrid.
- HENRIET, D. Y ROCHET, J.C. (1986): "La logique des systèmes Bonus-Malus en assurance automobile: une approche théorique". *Annales d'Économie et de Statistique*, 1; pp.133–152.
- JOLIFFE, I.T. (1972): "Discarding variables in Principal Components Analysis". *Applied Statistics*, 21; pp.160–173.
- KAO, C. Y HWANG, S.N. (2008). "Efficiency decomposition in two-stage data envelopment analysis: an application to non-life insurance companies in Taiwan". *European Journal of Operational Research*, 185(1); pp.418–429.
- LEMAIRE, J. (1985): *Bonus-Malus systems in automobile insurance actuarial models*. Kluwer Academic Publishers, Boston.
- MAYERS, D. Y SMITH, C.W. (1983): "The interdependence of individual portfolio decisions and the demand for insurance". *Journal of Political Economy*, 91; pp.304–311.
- MAHUL, O. (2000): "Optimal insurance design with random initial wealth". *Economics Letters*, 69; pp.353–358.
- PUJOL, M. Y BOLANCÉ, C. (2004): *La matriz valor fidelidad en el análisis de los asegurados en el ramo del automóvil*. Fundación MAPFRE estudios. Instituto de Ciencias del Seguro. Madrid.
- SARRICO, C.S., HOGAN, S.M., DYSON, R.G. Y ATHANASSOPOULOS, A.D. (1997): "Data envelopment analysis and university selection". *Journal of the Operational Research Society*, 48; pp.1163–1177.
- SCHIFFMAN, J.F., REYNOLDS, M.L. Y YOUNG, F.W. (1981): *Introduction to Multidimensional Scaling: Theory, Methods and Applications*. Academic Press: London.
- SERRANO-CINCA, C., FUERTES CALLEN, Y. Y MAR-MOLINERO, C. (2005): "Measuring DEA efficiency in Internet companies". *Decision Support Systems*, 38; pp.557–573.
- SERRANO-CINCA, C. Y MAR-MOLINERO, C. (2004): "Selecting DEA specifications and ranking units via PCA". *Journal of the Operational Research Society*, 55; pp.521–528.
- THANASSOULIS, E. (2001): *Introduction to the theory and application of data envelopment analysis: A foundation text with integrated software*, Kluwer Academic Publishers, Boston.
- TONE, K. Y SAHOO, B.K. (2005): "Evaluating cost efficiency and returns to scale in the Life Insurance Corporation of India using data envelopment analysis". *Socio-Economic Planning Sciences*, 39; pp.261–285.