



# TRABAJO DE FIN DE GRADO

Análisis Estadístico sobre la venta de vehículos de motor en  
el Sector Comercio

*(Statistical analysis about the sale of motor vehicles in the commercial sector)*

**Autor:** D. Mario Campillo Reyes

**Tutor/es:** D. Sergio Martínez Puertas

## GRADO EN ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS

Facultad de Ciencias Económicas y Empresariales

UNIVERSIDAD DE ALMERÍA

Curso académico: 2015/2016

Almería, Junio de 2016

# ÍNDICE

1. INTRODUCCIÓN.....	3
<b><u>CAPÍTULO 1: MARCO TEÓRICO</u></b>	
CRISIS ECONÓMICA ESPAÑOLA DEL PERIODO 2008-2016.....	5
2. SECTOR TERCIARIO A NIVEL NACIONAL.....	5
2.1 SECTOR COMERCIO.....	6
2.1.1 SECTOR DE LA AUTOMOCIÓN.....	7
<b><u>CAPÍTULO 2: METODOLOGÍA, ANÁLISIS Y RESULTADOS</u></b>	
3. PRESENTACIÓN.....	8
3.1 PROGRAMA SPSS: DEFINICIÓN Y UTILIDAD.....	9
3.2 VARIABLES CLAVE PARA EL ANÁLISIS.....	9
4. ANÁLISIS DE COMPONENTES PRINCIPALES.....	13
4.1 DETERMINACIÓN NÚMERO DE COMPONENTES, INTERPRETACIÓN Y ROTACIÓN.....	15
4.2 ANÁLISIS DE COMPONENTES PRINCIPALES POR COMUNIDADES 2014.....	23
5. ANÁLISIS CLÚSTER O DE CONGLOMERADO .....	25
5.1 ANÁLISIS JERÁRQUICO.....	27
5.1.1 DETERMINACIÓN DEL NÚMERO DE GRUPOS.....	28
5.1.2 INTERPRETACIÓN DE LOS RESULTADOS OBTENIDOS.....	33
5.2 ANÁLISIS NO JERÁRQUICO.....	35
5.2.1 DESCRIPCIÓN DETALLADA DEL PROCESO.....	38
6. CONCLUSIÓN.....	41
7. BIBLIOGRAFÍA.....	42

## **RESUMEN**

El objetivo principal que se persigue con la elaboración de dicho trabajo, es conocer más a fondo el sector automovilístico, a través de la aplicación de dos técnicas estadísticas descriptivas a una serie de variables clave, sacadas del Instituto Nacional de Estadística (INE).

De este modo, primero se hace un pequeño recorrido a través del sector terciario o de servicios, pasando por el sector comercio y terminando por el sector que nos concierne, el sector del automóvil, con información básicamente teórica.

En segundo lugar, llevaremos a cabo el estudio completo de dichas variables, gracias al programa utilizado, el SPSS. A través de dicho instrumento, aplicaremos dos métodos para llevar a cabo el estudio, análisis de componentes principales y análisis clúster o de conglomerados, y dentro de este, análisis jerárquico y no jerárquico. Analizando todas las gráficas y tablas pertinentes.

Y por último, realizaremos una breve conclusión a modo de resumen de los resultados obtenidos en dichas técnicas. Así como la bibliografía del proyecto.

## **ANÁLISIS ESTADÍSTICO SOBRE LA VENTA DE VEHÍCULOS DE MOTOS EN EL SECTOR COMERCIO**

### **1. INTRODUCCIÓN**

En el presente trabajo de fin de grado se realiza un análisis del sector de la automoción, a través de ciertas variables. Este trabajo está realizado por D. Mario Campillo Reyes del grado Administración y Dirección de Empresas de la Universidad de Almería, con el apoyo, como tutor de dicho trabajo, del docente D. Sergio Martínez Puertas, del Departamento de Estadística, del curso académico 2015 – 2016.

Entre los motivos que nos ha llevado a la elaboración de este trabajo se encuentra, entre otros: el interés personal acerca de la evolución y la afección de la crisis a esta parte de la

economía española, considerada de las más importantes, a través del análisis de dos métodos estadísticos descriptivos..

Debido a esto, el objetivo principal que se persigue con este proyecto es analizar el sector automovilístico a nivel nacional y de forma general (incluye tanto turismos, camiones, remolques y semirremolques, todoterrenos, ambulancias, microbuses...) tanto nuevos como de segunda mano, a través de varias variables clave para conocer cómo ha afectado la crisis a este sector, analizando los años anteriores, el inicio y los posteriores a esta.

Todos los datos necesarios para llevar a cabo dicho análisis lo hemos sacado del Instituto Nacional de Estadística (INE) a través de su página web, más concretamente en el INEbase, en el apartado comercio, dónde nos ofrecen todos los datos numéricos que necesitamos para realizar el estudio con la ayuda de dos métodos estadísticos descriptivos, análisis de Componentes Principales y análisis Cluster.

El siguiente trabajo presenta dos capítulos bien diferenciados, dónde en el primero, marco teórico, con una primera parte dónde se detalla brevemente el porqué de la crisis económica.

A continuación, se explica de forma resumida toda la información acerca de la situación del sector terciario o de servicios, el subsector comercio y dentro de este, el sector de la automoción. Toda la información necesaria para realizar este informe del sector ha sido obtenida de noticias y páginas de internet de diferentes periódicos económicos así como de la Asociación Española de Fabricantes de Vehículos y Camiones (ANFAC) y de la Asociación Nacional de Vendedores de Vehículos de motor, Reparación y Recambios (GANBAM).

En el capítulo dos se detalla la forma de analizar los datos que tenemos, explicando el programa utilizado así como las variables y las diferentes técnicas estadísticas utilizadas, como son componentes principales y cluster o de conglomerados.

En ambas se detalla la definición propia de cada técnica y el proceso que se lleva a cabo para obtener los resultados finales, incluidas todas las tablas y gráficas, los cuales son analizados y explicados para dar respuesta al objetivo principal de este proyecto de fin de carrera, ver cómo afecta la crisis a la venta de automóviles.

Por último, tenemos la conclusión final que hemos obtenido de todo el proceso llevado a cabo en este proyecto y el apartado de bibliografía, dónde se puede encontrar todas los link de interés, y de dónde se ha sacado todo lo necesario para realizar este proyecto de fin de grado.

## **CAPÍTULO 1: MARCO TEÓRICO**

En el presente trabajo vamos a llevar a cabo un estudio detallado del sector de ventas automovilísticas españolas, es decir, a través de una serie de variables clave, veremos la evolución que ha llevado a lo largo de los último años, más concretamente en el periodo 2005-2016 y por tanto, veremos cómo le afecta la crisis económica.

Para una mejor comprensión a la hora del análisis de los resultados obtenidos por los métodos estadísticos, vamos a detallar brevemente la explicación del porqué de la crisis española y además una explicación del sector servicio o terciario

### **Crisis económica española del periodo 2007-2016**

Dicha crisis, también conocida como recesión económica española, comenzó a finales del año 2007 y según varios estudios, concluyó en el año 2014. Ésta fue debida a dos sucesos: una crisis financiera mundial y la explosión de la burbuja inmobiliaria, que llevaba muchos años siendo el motor de la economía española. Esto provocó un nivel elevado de desempleo, así como la necesidad de una reestructuración, no sólo económica, sino también política y social. Como veremos más adelante, la crisis española afecto a miles de empresarios, así como a particulares y sus negocios, haciendo que la competencia se elevase, se redujera la productividad drásticamente, etc. Además, el tener el euro como moneda ha hecho que sea más complicado el salir de esta situación.

#### **1. Sector Terciario o de Servicios en España**

Dentro de la economía de España, la cual es considerada como una de las más grandes a nivel de la Unión Europea, el sector terciario o de servicio, que se refiere a aquellas actividades que se encargan de la distribución y el consumo de bienes y servicios, tiene una gran importancia, ya que tiene un peso en el Producto Interior Bruto del 51,2%, y el cual

da empleo a más de la mitad de la población española, según estudios llevados a cabo por el Instituto Nacional de Estadística (INE).

Este sector no ha dejado de crecer durante los últimos 50 años, aun pasando por una recesión, respecto de los demás sectores, ya sea en el número de empleados o el peso que tiene en el PIB, por lo que el sector servicio es considerado como el pilar que mantiene a la economía española. Este crecimiento es debido sobre todo a los avances tecnológicos, que han reducido la necesidad de mano de obra en los otros dos sectores, aumento del sector turismo, aumento de los servicios en las empresas, aumento de la población, entre muchos otros factores. Un dato importante es que, en los años de crisis en los que vivimos, el sector terciario es el que más puestos de trabajo ha otorgado a la población española, de ahí que sea considerado con tal importancia.

Por otro lado, queda aclarar que este aumento del sector servicio no es igual en todo el territorio nacional, es decir, el nivel no es el mismo en unas regiones que otras, por ejemplo, en Madrid el nivel será mucho más elevado que en cualquier otra comunidad autónoma por el gran número de personas y de empresas situadas en este punto ( 70% de los ciudadanos de esta ciudad está empleada en este sector), de ahí que otras comunidades sean más destacables en los sectores secundario y primario, por ejemplo, Castilla La Mancha y Galicia.

### **1.1 Sector Comercio**

Desglosando el sector terciario, podemos observar que está compuesto por subsectores o actividades como son el turismo, transporte, sector financiero, Tecnologías y la Comunicación (TIC) y el comercio. Tan solo el turismo y el comercio ocupan más de la mitad del PIB, en nuestro caso, vamos a desarrollar un poco el tema del comercio en España.

En primer lugar, comentar que su importancia se debe tanto al número de empresas y de empleados como de la cifra de negocios, así como su aporte al valor añadido. Este se compone a su vez de minoristas, que son los que venden el producto a los clientes finales y quien tiene mayor peso, y los mayoristas, que son quienes obtienen los productos necesarios a los fabricantes y quienes los distribuyen a los minoristas. También hay que distinguir entre comercio interior, nuestro caso, y comercio exterior.

En cuanto al PIB, el comercio en España tiene un peso del 14%, que se ha mantenido más o menos constante a lo largo de los últimos 15 años, por lo tanto podemos decir que el comercio interior español es elevado si lo comparamos con otros países como Alemania (9%) o Francia (12%). En lo que concierne al empleo, el comercio interior es uno de los más importantes en cuanto a generador de empleo (15% de la población activa y mismo porcentaje en población ocupada).

En cuanto a la actividad que más importancia tiene dentro del comercio español, es el sector de la automoción, objeto de análisis de este proyecto y el cual se desarrolla en el siguiente punto.

### *1.1.1 Sector de la Automoción*

Este sector ha ido evolucionando a lo largo de los años y constituye el tercer brazo motor de la economía española dentro del sector comercio, por detrás siempre del turismo y la industria agroalimentaria, generando el 7% del Producto Interior Bruto, que es un poco más bajo si lo comparamos con los años de antes de la crisis (10,4%). Todo esto es gracias a que España permite una fabricación muy competitiva, una concentración en nuestro territorio de todo tipo de plantas de producción del sector y cuenta con empresas de todas partes del mundo, que realizan enormes inversiones, lo que ha propiciado un ambiente inmejorable para el desarrollo de esta actividad, incluso para periodos futuros (2015-2019), además del impulso exportador, y el esfuerzo llevado a cabo por el mercado interior creando en España treinta y cuatro centros tecnológicos dedicados exclusivamente a las actividades de automoción y a programas específicos de formación y desarrollo. De ahí que España ocupe el segundo puesto de Europa y el decimosexto del mundo en cuanto a fabricación de coches y sea considerada como un referente en el sector automovilístico.

En la actualidad, este sector comienza a recuperarse después de los años de recesión vividos a partir del año 2007. En parte, esta recuperación viene dada por el aumento de la rentas de los particulares y como se ha mencionado en el párrafo anterior, gracias a la evolución de la tecnología y del resto de factores nombrados.

Los agentes que componen este sector son: constructores de vehículos, que centran su actividad en el desarrollo de motores, ensamblaje, diseño...; fabricantes de los diferentes componentes y equipos; distribuidores; desguaces y empresas recicladoras; talleres y los

concesionarios de vehículos, que es la parte que nos interesa, dónde se produce la comercialización y la venta de los productos, en nuestro caso vehículos, tanto de primera como de segunda mano.

En cuanto a ventas, tanto el mercado de coches nuevos como de coches usados ha crecido notablemente durante los últimos años. España se puede considerar como la líder de Europa en cuanto a ventas de coches nuevos, con un crecimiento en 2014 respecto de años anteriores del 19%, por delante de Reino Unido (9,4%), Italia (4,3%) y Alemania (2,6%). En el mercado de segunda mano también se ha producido un crecimiento notable del 3% según la Asociación Nacional de Vendedores de Vehículos de motor, Reparación y Recambios (GANBAM).

En lo que se refiere al empleo, la industria de la automoción está volviendo a los niveles que presentaban a principios del año 2008, con solo 5 puntos de diferencia y 17% de diferencia con el año 2007. Aun así todavía queda mucho camino en cuanto a la generación de empleo. Como aportación importante, en 2014 se generó 20.000 puestos de trabajo en este sector y otros 80.000 en la industria auxiliar, según datos revelados por la Asociación Española de Fabricantes de Automóviles y Camiones (ANFAC).

Todo lo mencionado anteriormente es un pequeño resumen acerca del sector terciario, el subsector comercio y dentro de este, el sector de la automoción, con datos más bien generales para conocer un poco más acerca de la situación en la que se encuentran tales sectores y con esto, poder llevar a cabo nuestro análisis que será descrito con más detalle en los puntos siguientes.

## **CAPÍTULO 2: METODOLOGÍA, ANÁLISIS Y RESULTADOS**

### **3. Presentación**

Una vez informados acerca de la situación general del sector de la automoción en el capítulo anterior, procedemos a la realización del análisis de este sector a través de dos métodos estadísticos descriptivos con la ayuda del programa SPSS, el cual es descrito de forma más detallada en el siguiente apartado, y de las variables clave, que serán analizadas individualmente antes de llevar a cabo el estudio general.



Como información necesaria para entender el proyecto, cuando nos referimos a estadística descriptiva, hablamos de la parte de la estadística que se dedica a la recolección, ordenación, análisis y representación de un conjunto de datos con el objetivo de desarrollar de manera correcta las características del conjunto, todo a través de tablas y gráficos.

### **3.1 Programa SPSS: definición y utilidad**

El SPSS (en inglés: '*Statistical Package for the Social Sciences*') es un programa estadístico muy popular entre las organizaciones e instituciones, sobre todo de investigación de mercados.

Entre las ventajas, destacamos la facilidad de navegación, ya que ordena todos los cuadros de forma secuencial, de manera que el investigador puede seleccionar los análisis que mejor le convengan para su estudio. Pero como todo programa, tiene también sus inconvenientes, a continuación nombramos algunas de ellas: si se tiene pocos conocimientos de estadística, el usuario puede perderse debido a la complejidad de algunas tablas; la abundancia de información puede suponer también un serio problema y confundir más que aclarar, etc.

Su función principal es la de organizar datos y crear archivos electrónicos. Este programa estadístico nos permite el análisis de extensas y complejas bases de datos con el fin de llevar a cabo aplicaciones prácticas o investigaciones, además, se pueden realizar así métodos estadísticos complejos como básicos, todo con el fin de observar la tendencia que seguirán dichos datos, que permitirá a los investigadores planificar actividades a largo plazo.

En nuestro caso, las técnicas utilizadas con el SPSS son el análisis de Componentes Principales y el análisis Clúster, cuyas tablas y gráficos nos darán una serie de información que nos permitirá sacar unos resultados, cuyo estudio nos dará respuesta al objetivo buscado, la afección de la crisis a este sector.

### **3.2 Variables clave para el análisis**

Para la realización del presente estudio, se han utilizado en total siete variables clave. Estas son: el número de empresas, número de locales, volumen de negocio (cifra de negocio),

valor de la producción, productividad (en euros, €) y total, compuesto por empresas y profesionales, como una variable, y particulares, como otra.

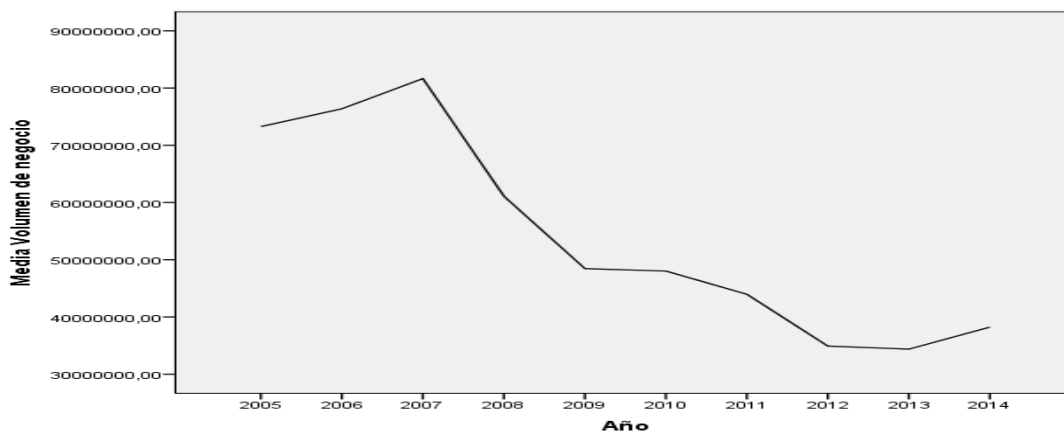
En nuestro caso, en vez de coger el total, hemos preferido diferenciar entre empresas y particulares como dos variables separadas.

A continuación, hemos realizado un pequeño análisis de cada una de las variables iniciales a través de un gráfico lineal simple que mide la evolución de éstas a lo largo de los últimos nueve años. De este modo, las variables objeto de análisis se han observado para el periodo considerado (2005-2014), teniendo un total de diez observaciones para cada variable.

Como dato importante, todas las variables objeto de estudio en este proyecto se miden en miles de euros según el Instituto Nacional de Estadística, con excepción de las variables número de empresas y número de locales

Un rango inherente a todas las variables que se puede ver claramente es el movimiento de descenso que se aprecia a partir del año 2007, año en el que la explota la burbuja inmobiliaria, es decir, cuando comienza la crisis económica.

**Gráfico 3.2.1: Volumen de negocio**

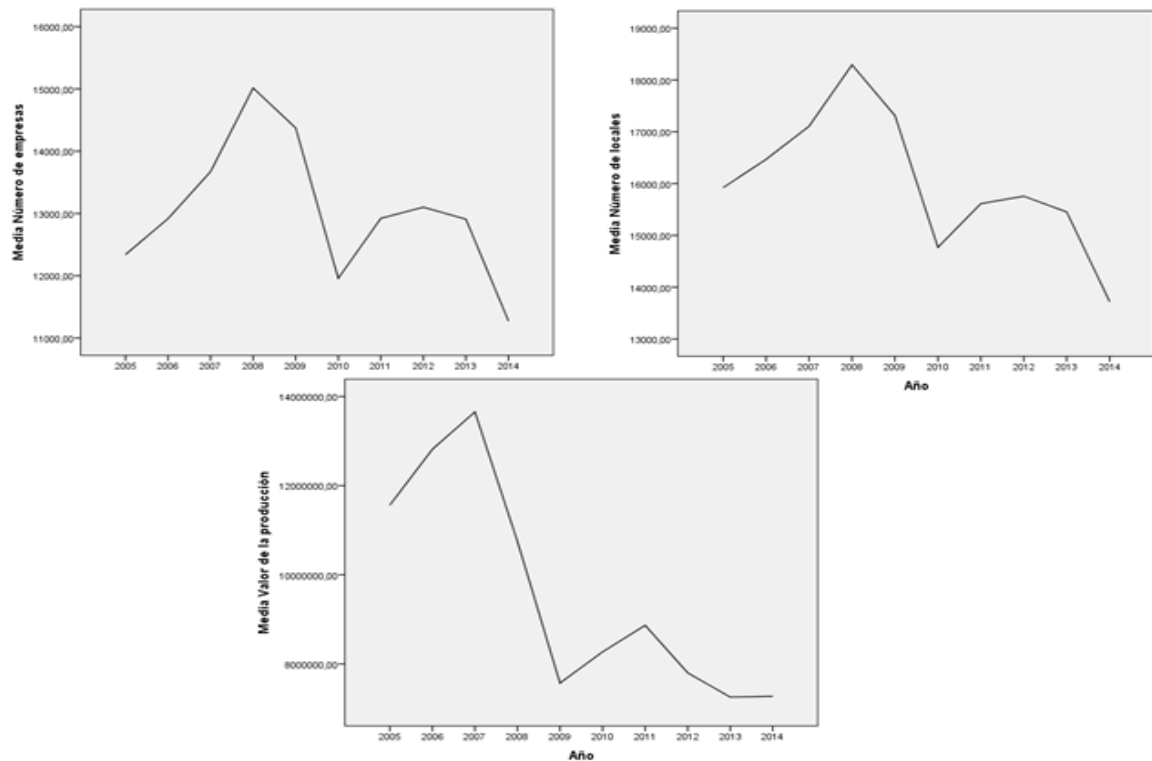


Fuente: Elaboración propia

En el gráfico 3.2.1 aparece lo referente a la variable volumen de negocio, es decir, a los ingresos que obtienen las organizaciones por la realización de sus actividades a lo largo del tiempo, en este caso en el periodo estudiado.

Si lo analizamos más detenidamente, podemos observar como la crisis ha afectado a los ingresos de las empresas automovilísticas hasta el 2013, año en el que comienza a recuperarse, aunque muy lentamente.

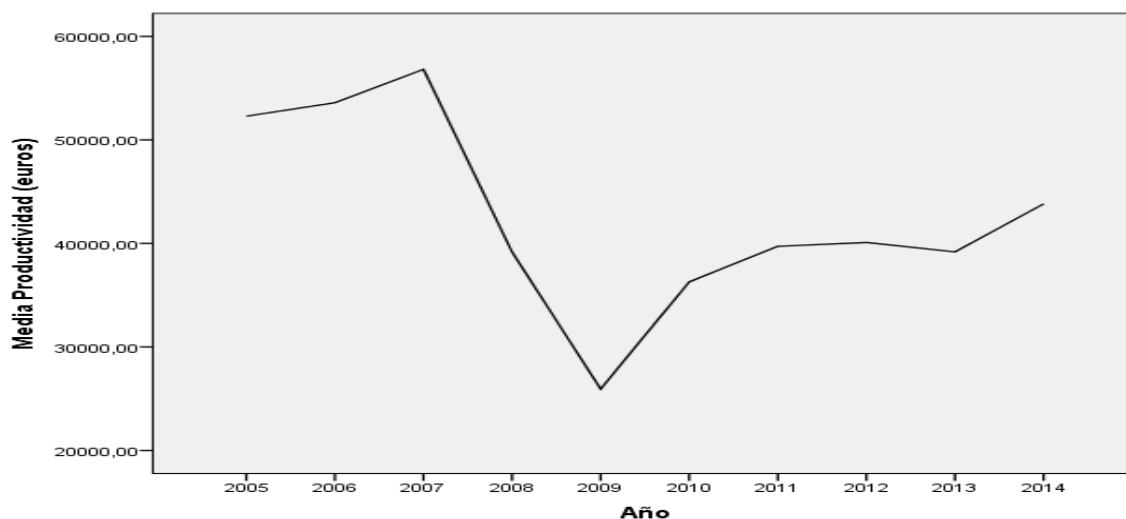
**Gráfico 3.2.2: número de empresas, de locales y valor de la producción**



Fuente: Elaboración Propia

El gráfico 3.2.2 presenta las variables número de locales, número de empresas y valor de la producción. Hemos decidido analizarlas en conjunto ya que siguen una tendencia temporal igual. Si nos fijamos bien, a partir del 2009-2010 sufren una subida repentina durante un año o dos y luego vuelven a descender.

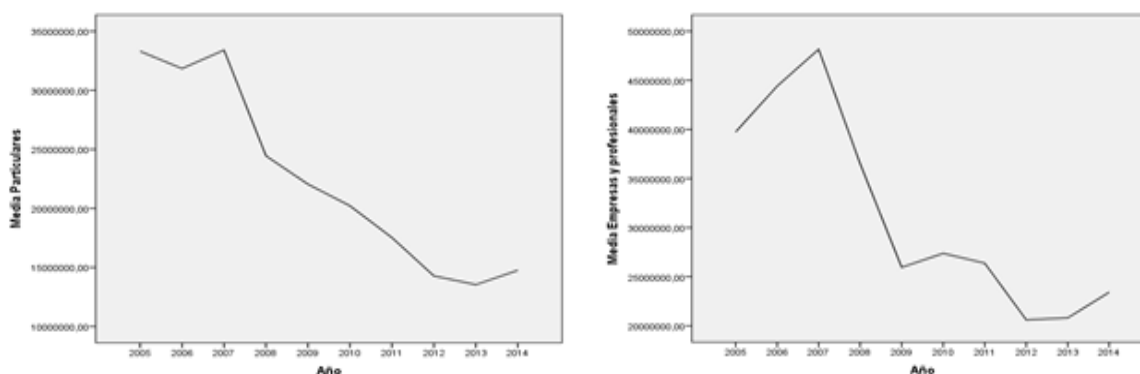
**Gráfico 3.2.3: Gráfico Productividad**



Fuente: Elaboración Propia

La productividad (gráfico 3.2.3) sigue una tendencia descendente desde el 2007 hasta el 2009, pero a diferencia del resto, a partir de este año comienza a verse una recuperación paulatina hasta la actualidad.

**Gráfico 3.2.4: Gráfica Particulares y empresas y profesionales**



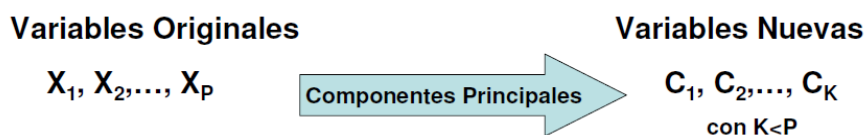
Fuente: Elaboración Propia

La gráfica de particulares y empresas y profesionales (gráfica 3.2.4) son las únicas que presentan un descenso prolongado desde el comienzo de la crisis en el año 2007 hasta el 2013, aunque la variable empresas y profesionales tienen un pequeño ascenso en el año 2009 pero este dura muy poco y vuelve a descender. En el 2013 ambas parecen recuperarse, pero muy lentamente

#### **4. Análisis de Componentes Principales**

El análisis de componentes principales es una técnica multivariante de reducción o simplificación de la dimensión y se aplica cuando las variables a estudiar son numéricas, es decir, cuantitativas. Se utiliza para pasar de un conjunto de variables originales a un nuevo conjunto menor de variables que llamaremos componentes principales, las cuales están incorrelacionadas entre sí.

**Figura 4.1: Esquema Componentes Principales**



Fuente: Elaboración propia

El objetivo principal que se persigue con esta técnica es, teniendo una elevada cantidad de variables iniciales, obtener un menor número de componentes, con las cuales poder realizar un mejor estudio posterior, ya que al tener muchas variables, en nuestro caso siete, y que están correlacionadas entre sí, no nos permite saber el papel que juegan, es decir, la importancia de cada una de ellas. El análisis de componentes principales nos permite obtener un nuevo conjunto de variables, las componentes principales, que están incorrelacionadas entre sí, lo que permite simplificar la interpretación de las variables originales.

En general, la extracción de componentes principales se efectúa sobre variables tipificadas para evitar problemas de escala.

**Tabla 4.1: Estadísticos descriptivos**

	Media	Desviación típica	N del análisis
Número de empresas	13048,1000	1100,62593	10
Número de locales	16040,7000	1318,48432	10
Volumen de negocio	54039709,0000	17804341,17917	10
Valor de la producción	9582392,3000	2415202,68837	10
Productividad (euros)	42691,7400	9271,51269	10
Empresas y profesionales	31336236,1000	10051077,78086	10
Particulares	22535668,2000	7921084,49236	10

Fuente: Elaboración Propia

Para explicar lo anteriormente dicho, podemos observar la tabla 4.1, estadísticos descriptivos, que está formada por tres columnas. La primera donde aparece la media de cada variable, la segunda recoge las desviaciones típicas y la última el número de observaciones que se ha realizado de cada una.

Se puede ver como hay tres variables que destacan mucho más que las otras, volumen de negocio, empresas y profesionales y particulares. De esta manera, si aplicásemos la técnica de componentes principales sin tipificar las variables, no estaríamos dando la misma importancia en un principio a cada una de las variables, sino que tendrían más importancia aquellas variables con mayor varianza y de este modo generamos problemas de escala. Por ello, para evitar que una variable tenga más peso en el análisis únicamente por su escala, necesitamos que todas ellas tengan la misma importancia, es decir, que todas las variables estén tipificadas, y por lo tanto tengan una varianza igual a uno.

El análisis de componentes principales permite obtener tanta componentes como variables, esto es, si disponemos de siete variables, en principio podemos extraer hasta siete componentes, pero quedarse con todas ellas no simplificaría el problema. Por ello, es necesario ordenar las componentes según la información que llevan incorporada. Para saber la cantidad de información que contiene cada componente utilizaremos su varianza, es decir, su valor propio. Por lo tanto cuando mayor sea ésta, mayor será la información que incorpora cada componente. Es importante tener en cuenta que las variables se encuentren tipificadas para su correcto estudio, como hemos comentado en el párrafo anterior, con lo que evitaremos errores derivados de escala. Otro dato importante es que las nuevas variables se expresan como una combinación lineal de las originales.

Algunos de los problemas con los que nos encontramos son, de forma resumida, los siguientes: saber el número de componentes principales a retener, por tanto hay que saber la importancia de cada componente; interpretar de forma correcta las nuevas variables, es decir, estudiar la relación entre las variables originales ( $X_p$ ) y las nuevas ( $C_k$ ) y si esta es difícil de llevar a cabo, habrá que realizar una rotación de los componentes principales.

#### **4.1 Determinación del número de componentes, la interpretación y rotación**

En primer lugar, calcularemos el valor propio, que es, como hemos comentado anteriormente, la varianza de cada una de nuestras variables. Con esto obtenemos el peso o importancia que tienen.

Existen varios criterios, con los cuales podemos determinar el número de componentes principales, pero en el presente trabajo consideraremos tres: criterio de la variabilidad total explicada, criterio de la media aritmética y gráfico de sedimentación, que son los más empleados.

A modo de pequeño resumen de los criterios mencionados, para una mejor comprensión de los datos analizados más adelante: el criterio de la media aritmética consiste en que el valor propio total del componente sea mayor que la media de los valores propios, que al estar las variables originales tipificadas, este toma valor igual a uno, de este modo, el criterio de la media aritmética consiste en retener aquellas componentes cuyo valor propio sea superior a la unidad. Con el criterio de la variabilidad total explicada se escogerán las componentes que supongan un determinado porcentaje elevado, en nuestro caso, hemos considerado a partir de un 95% en la mayoría de los casos y por último, el criterio del gráfico de sedimentación, donde las componentes a retener son las que se encuentran en la pendiente del gráfico.

En nuestro caso en particular, la aplicación de los tres criterios coincide y nos sugieren retener dos componentes principales ( $C_1$  Y  $C_2$ ), de las siete componentes posibles.

**Tabla 4.1.1: Varianza Total Explicada**

Componente	Autovalores iniciales			Sumas de extracción de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	4,860	69,434	69,434	4,860	69,434	69,434
2	1,866	26,662	96,096	1,866	26,662	96,096
3	,234	3,350	99,445			
4	,029	,420	99,865			
5	,008	,113	99,978			
6	,002	,022	100,000			
7	2,873E-6	4,105E-5	100,000			

Fuente: Elaboración propia

En la Tabla 4.1.1 aparece dos cuadros o secciones con tres columnas cada una. En el primer cuadro, aparecen recogidos, en la primera columna, total, los valores propios iniciales de los siete componentes principales que se pueden obtener. En la siguiente columna aparece el porcentaje de la varianza total y en la última el porcentaje acumulado total. En el segundo cuadro, denominado sumas de las saturaciones al cuadrado de la extracción, dónde se pueden ver las mismas columnas, pero en esta ocasión solamente aparecen recogidos, en el caso de realizar un análisis de componentes principales, como es nuestro caso, las componentes retenidas.

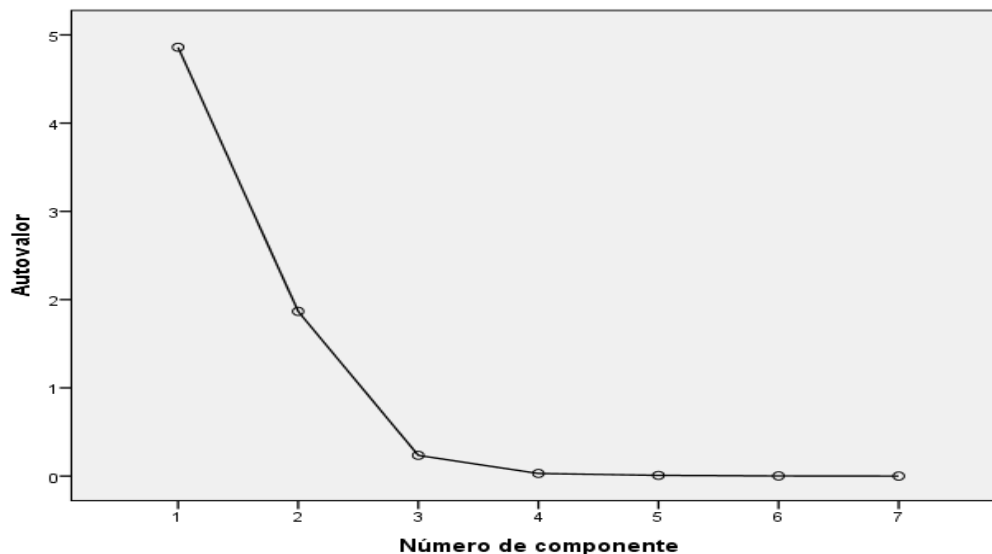
También aparece otra sección, llamada suma de las saturaciones al cuadrado de la rotación, la cual no aparece, por motivos de mejor visualización, pero que explicaremos detalladamente más adelante, en la parte dedicada a la rotación.

Si observamos la tabla que hemos obtenido, podemos ver que se puede justificar la elección de dos componentes principales gracias a la aplicación de dos de los criterios mencionados anteriormente, criterio de varianza total explicada y criterio de la media aritmética.

En el primer caso, se puede ver que el porcentaje acumulado de las dos primeras componentes es ya del 96,096%, suficiente para nuestro estudio, por lo tanto se desecha el resto. Según el criterio de la media, está aún más claro que el primero, ya que este criterio dice que sólo ha de cogerse los componentes cuyo valor propio sea superior a 1. En el cuadro se observa que solamente las dos primeras componentes (C1=4,860 y C2=1,866) tienen un valor propio superior a la media, por lo tanto el resto se desechan.



**Gráfico4.1.2: Gráfico de Sedimentación**



Fuente: Elaboración Propia

Y por último, el criterio del gráfico de sedimentación (Gráfico 4.1.2) corrobora la elección de dos componentes principales a retener, ya que como se puede observar en el gráfico, tanto la C1 como la C2 se encuentran en la pendiente mientras que el resto se mantienen horizontales, es decir, en la zona de sedimentación, y por lo tanto se desechan.

En definitiva y como hemos comentado al principio, los tres criterios utilizados llegan a la misma conclusión, son dos las componentes a retener.

Una vez que hemos seleccionado el número de componentes principales y desechado el resto, pasamos al siguiente punto, que consiste en saber interpretar de manera correcta dichas componentes, esto es, hemos obtenido dos nuevas variables que representan a nuestras siete variables originales y debemos interpretar correctamente el valor de estas nuevas variables. Aquí es dónde se determinará que variables son las que mejor interpretadas están por las componentes retenidas.

Para llevar a cabo dicha interpretación será necesario analizar la comunalidad de cada una de las variables.

**Tabla 4.1.3: Comunalidad de una variable**

	Inicial	Extracción
Número de empresas	1,000	,978
Número de locales	1,000	,993
Volumen de negocio	1,000	,979
Valor de la producción	1,000	,982
Productividad (euros)	1,000	,880
Empresas y profesionales	1,000	,987
Particulares	1,000	,927

Fuente: Elaboración propia

Antes de analizar la tabla y para una mejor comprensión de dicho análisis, vamos a explicar en qué consiste las comunalidades, comenzando con su definición: es la *‘parte de la varianza de la variable explicada por las componentes retenidas, siendo esta como un criterio de representación de cada variable mediante las componentes principales’* (Análisis de datos multivariantes – Daniel Peña).

Al estar las variables tipificadas, el valor de su varianza es igual a uno, por tanto el valor de las comunalidades variará entre cero y uno, es decir, cuanto más cerca esté del uno mejor representada estará por las componentes retenidas y al contrario, cuanto más cerca de cero peor será dicha representación.

En dicha tabla 4.1.3, nos encontramos con dos columnas. La primera, denominada Inicial, se corresponde con las varianzas originales, que al estar tipificadas, todas valen uno. La segunda columna, extracción, se puede ver la representación, que en nuestro caso, es una muy buena, siendo la variable ‘número de locales’ la mejor representada, con un 0,993 y la peor ‘productividad’ con un 0,880. Así, las dos componentes principales retenidas representan de manera bastante satisfactoria al conjunto de variables estudiadas.

Una vez establecido dicho criterio, pasamos a la parte más interesante, saber cuáles de las variables se relacionan con cada componente retenida, es decir, al tener siete variables y dos componentes principales, habrá que saber cómo se relacionan las variables originales con las componentes, más concretamente asociaremos cada variable con una de las componentes retenidas y para ello, tendremos en cuenta las correlaciones entre variables y componentes. De este modo, asociando variables a cada una de las componentes, podremos interpretar de manera adecuada los valores de las componentes.

Para ello se analiza otra tabla, denominada la Matriz de Componentes, que muestra los coeficientes de correlación entre variables y componentes y asociaremos cada variable a la componente con la que tenga mayor correlación.

**Tabla 4.1.4: Matriz de Componentes**

	Componente	
	1	2
Número de empresas	,363	,920
Número de locales	,616	,784
Volumen de negocio	,985	-,090
Valor de la producción	,984	-,120
Productividad (euros)	,717	-,605
Empresas y profesionales	,988	-,103
Particulares	,960	-,078

Fuente: Elaboración propia

Como podemos ver en dicha tabla 4.1.4, la variable número de empresas presenta una correlación de 0,920 con la segunda componente (C2); el volumen de negocio, la productividad (euros), empresas y particulares se relacionan con la componente C1 con correlaciones de 0,985, 0,984, 0,988 y 0,960 respectivamente y por último, hay dos variables las cuales no está muy claro con que componente se relacionan, el número de locales y la productividad (euros). Es aquí cuando aparece el último apartado de este punto, la rotación.

Con la rotación de los componentes principales obtenidos, lo que se pretende es mejorar los resultados y así poder relacionar las variables con las componentes de forma más clara. Al realizar el análisis de rotación, obtenemos los siguientes resultados.

**Tabla 4.1.5: Variación Total Explicada Rotación**

Sumas de rotación de cargas al cuadrado		
Total	% de varianza	% acumulado
4,505	64,356	64,356
2,222	31,739	96,096

Fuente: Elaboración propia

En la tabla de la varianza total explicada (tabla 4.1.1), al rotar, aparece una nueva sección, como se puede ver en la tabla 4.1.5. En esta parte podemos observar como el porcentaje de la varianza de la primera componente ha empeorado sustancialmente (64,356%) pero la segunda componente ha mejorado (31,739%), el resultado final sigue siendo el mismo (96,096%).

**Tabla 4.1.6: Matriz de Componentes Rotados**

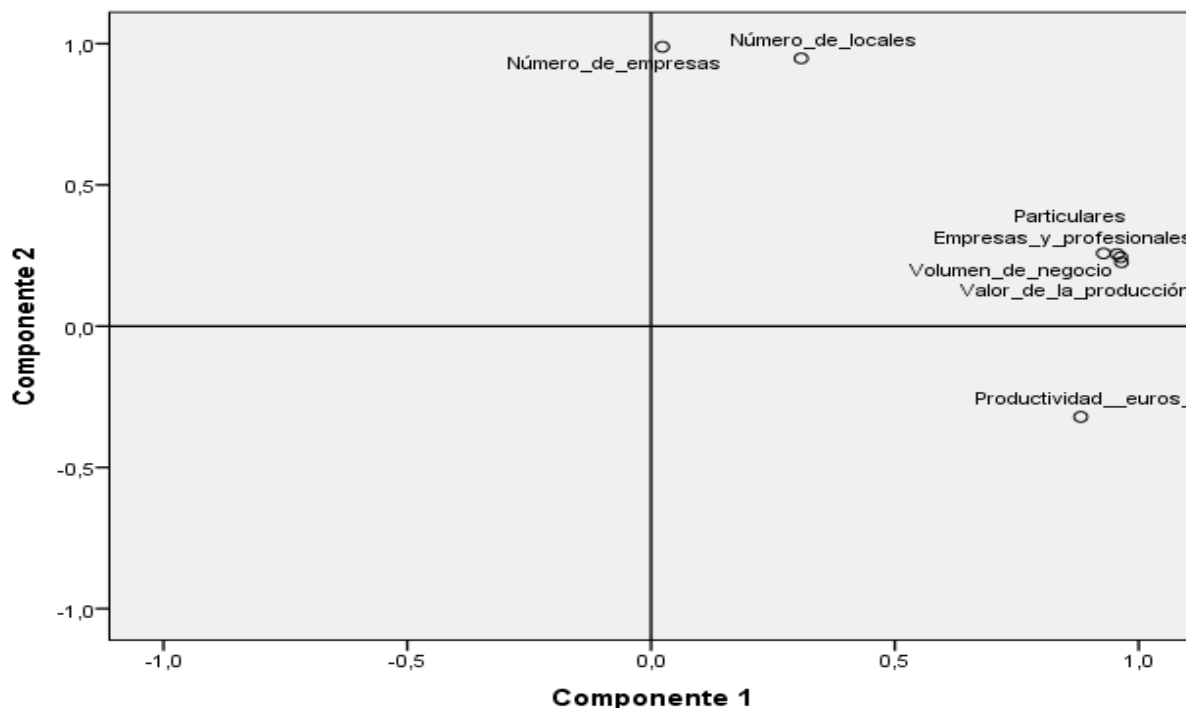
	Componente	
	1	2
Número de empresas	,023	,989
Número de locales	,308	,948
Volumen de negocio	,956	,255
Valor de la producción	,965	,226
Productividad (euros)	,881	-,321
Empresas y profesionales	,963	,244
Particulares	,928	,258

Fuente: Elaboración propia

La nueva tabla 4.1.6 muestra la matriz de componentes pero en esta ocasión, con los datos ya rotados, es decir, en la anterior se podía ver como la relación entre las variables número de locales y productividad y las componentes no estaba clara, al ser rotado, se ve claramente como la variable número de locales, con una correlación de 0,948, se relaciona

con la componente dos y la productividad, con una correlación de 0,881, aunque no es mucho pero si más claro que antes, con la primera componente.

**Gráfica 4.1.7: Gráfica de Componentes en espacio rotado**

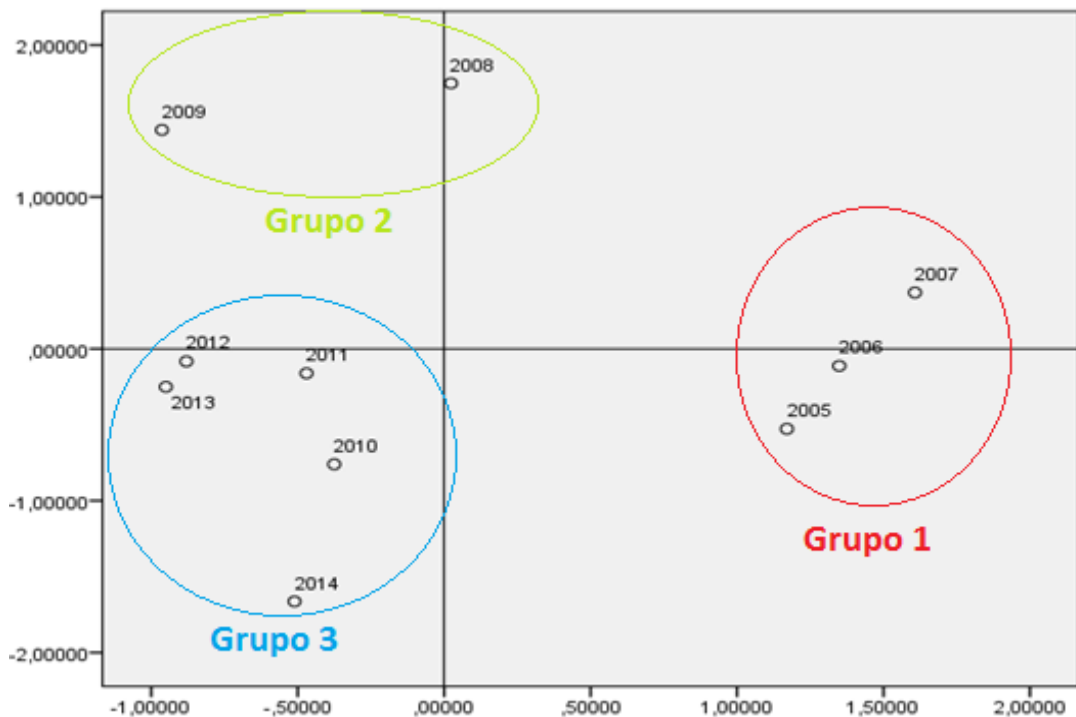


Fuente: Elaboración propia

Como se puede observar en la gráfica anterior, que representa las correlaciones entre componentes y variables, la componente dos claramente está asociada con correlaciones positivas con las variables número de empresas y número de locales, por eso, esta componente hace referencia a la expansión del negocio, esto es, valores altos de esta componente implicaría que ese año el sector destaca en número de locales y empresas. Por otra parte, la componente uno estaría asociada de manera directa (correlaciones positivas) con el resto de las variables, por lo tanto, esta componente se puede asociar a la rentabilidad de dichas empresas y en consecuencia valores altos de esta componente indican años de rentabilidad del sector.

Para entender un poco mejor el gráfico, la interpretación es la siguiente: digamos que la componente dos se relaciona de manera directa con el número de empresas y locales y la componente uno se relaciona de manera directa con en el resto de variables.

Gráfica 4.1.8: Gráfica de años en relación componentes



Fuente: Elaboración propia

La gráfica 4.1.8, representa el valor de la primera y segunda componente para el conjunto de años estudiados y debe interpretarse junto a la gráfica de los componentes rotados, explicada en el párrafo anterior. Los ejes siguen siendo los componentes (eje vertical: componente dos; eje horizontal: componente uno).

Si observamos bien, tenemos tres grupos, que relacionamos con los periodos de la crisis, siendo el primer grupo, de color rojo (2007, 2006, 2005), el perteneciente al ciclo anterior a la crisis; el grupo dos, de color verde (2008 y 2009), como el periodo de inicio y el tercer grupo, de color azul, que pertenece al ciclo posterior y comprende el resto de años, hasta el 2014 en este caso.

Interpretando el gráfico, podemos ver cómo los años del primer grupo destacan por tener unos valores muy altos en la componente C1 lo que equivale a años que destacan en las variables de volumen de negocio, productividad, valor de la producción, pero unos valores pequeños en relación al número de empresas y locales, por tanto podemos deducir que este grupo correspondiente a los años anteriores de la crisis económica, destacan en rentabilidad pero no en número de locales y empresas

El segundo grupo presenta unos valores altos en relación al anterior grupo en cuanto a número de locales y empresas, pero por lo que más destaca es por sus valores negativos en las variables que se relacionan con la rentabilidad, por tanto se ve claramente como fue en el año 2008 cuando la crisis estalló, de ahí que en el 2009, la rentabilidad de las empresas automovilísticas tuvieran sus peores momentos en muchos años, como demuestra el gráfico.

Y por último, el tercer grupo, que presenta valores negativos tanto en la expansión de las empresas, es decir, se cierran muchos negocios y locales y por tanto, una de las consecuencias es la subida en el nivel de desempleo. Pero no se queda ahí, también presenta valores nefastos en relación a la rentabilidad, que aun siendo mejor como se ve en 2010, debido sobre todo a la medidas económicas tomadas por el gobierno, llevando a cabo el PLAN E, el cual afectaba a actividades de I+D, medioambiente, y al sector de la automoción. Pero luego vuelve a descender a niveles muy bajos, como vemos en el gráfico, sobre todo en 2012.

Según estudios, y viendo el gráfico, en 2014 ya se ven signos de mejora y se prevé una subida significativa de estos niveles en 2015 y 2016.

#### **4.2 Análisis de componentes principales por comunidades 2014**

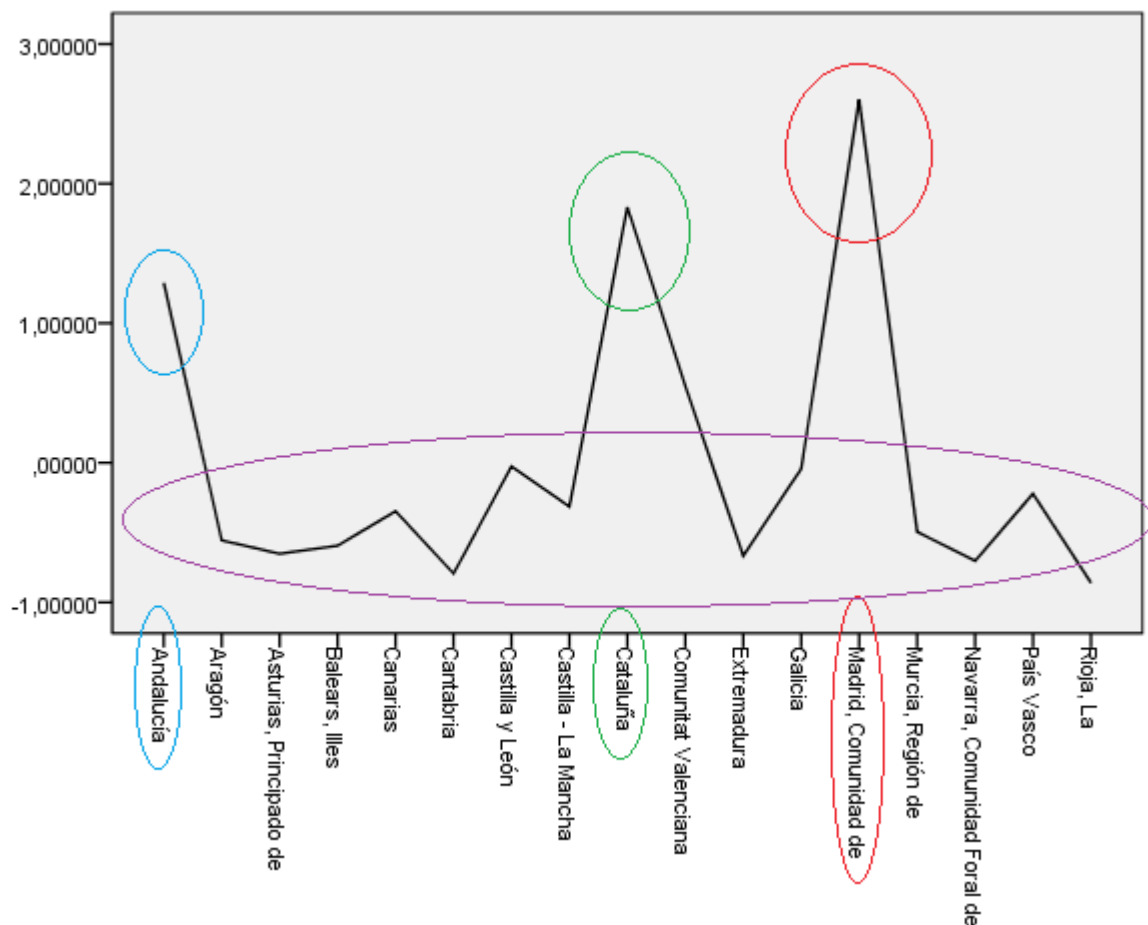
En este apartado, hemos decidido llevar a cabo el mismo procedimiento que en el apartado anterior, pero en este caso, por comunidades y el periodo de estudio es sólo el año 2014., con lo que las observaciones objeto de estudio son las comunidades y además el número de variables ya no es siete, si no cinco: sueldos y salarios, volumen de negocio, número de locales, inversión bruta en bienes materiales y personal ocupado.

Como comentábamos en el punto 4.1, primero debemos calcular el número de componentes principales a retener, a través del SPSS y la interpretación de ciertos criterios como son la variabilidad total explicada, la media aritmética o el gráfico de sedimentación, los cuales están explicados en el punto anterior.

Para no volver a copiar todas las tablas resultantes del proceso y expandir demasiado el trabajo, vamos a comentar de forma resumida los resultados obtenidos.

Al llevar a cabo los diferentes criterios, hemos obtenido las tablas de varianza total explicada, así como los valores propios y el gráfico de sedimentación. Dichos resultados nos sugieren, excepto el gráfico de sedimentación que nos sugiere dos, que sólo se retenga una componente principal. Con la primera componente, explicamos un 88,837% de la varianza total y al llevar a cabo el proceso tanto con una como con dos componentes nos dimos cuenta que el retener dos no tenía sentido pues todas las variables analizadas se encuentran asociadas muy claramente de forma directa con la primera componente y por tanto, al no asociar ninguna variable a la segundo componente, podemos desechar dicha componente.

**Gráfica 4.2.1: Comunidades según las variables**



Fuente: Elaboración propia



Al tener sólo una componente asociada de forma directa con todas las variables, podemos interpretar dicha componente como un indicador de la importancia del sector para cada comunidad autónoma y de este modo si representamos el valor de dicho indicador para todas las comunidades obtenemos el gráfico 4.2.1, en el cual aparecen las diferentes comunidades autónomas de España y donde podemos ver que tenemos cuatro grupos de comunidades bien diferenciados.

Por un lado, tenemos la comunidad de Madrid, la cual es la que más destaca en cuanto a inversiones, salarios, etc. Esto es normal, debido a que es la capital de la península, y es dónde se encuentran la mayoría de las sedes importantes de todas las empresas automovilísticas y también de otros sectores. Luego tenemos Cataluña, la cual también destaca por ser otra de las comunidades más importantes de España, dónde su economía se basa prácticamente en el sector industrial y en el sector servicio. Y la tercera comunidad, Andalucía, la cual comprende una gran parte del territorio español y es muy importante en cuanto al sector servicios, sobre todo por el turismo y el comercio de frutas y verduras.

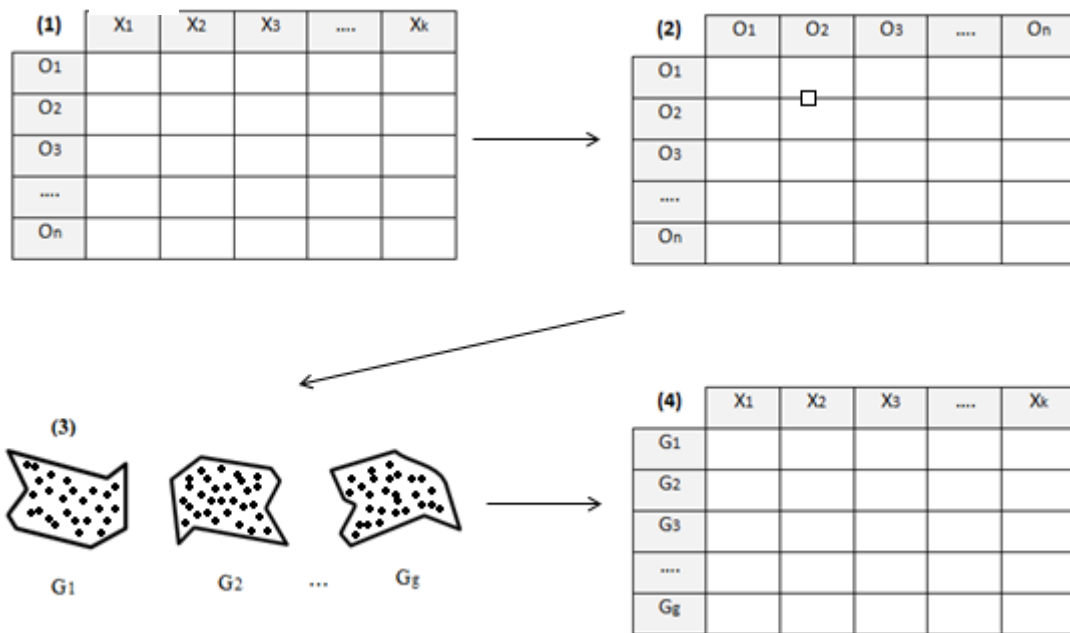
Por último, tenemos el resto de comunidades autónomas, las cuales no destacan y por tanto están todas agrupadas en un mismo grupo.

## **5. Análisis Clúster o de Conglomerados**

Este análisis, también conocido como análisis de conglomerados, es otra técnica multivariante, pero esta vez, de clasificación. Esta herramienta nos permite clasificar las diferentes observaciones en distintos grupos con lo que cada grupo o clúster esté compuesto por observaciones homogéneas, es decir, que sean parecidas entre sí y que los grupos sean diferentes unos de otros respecto a las variables consideradas.

Para comprender un poco mejor el proceso que se lleva a cabo al utilizar este tipo de técnica hemos incluido un pequeño esquema (figura 5.1) en el que podemos ver la secuencia que se sigue, compuesta por cuatro pasos.

**Figura 5.1: Esquema análisis clúster o de conglomerados**



Fuente: Elaboración propia

Primeramente, tenemos un número determinado de observaciones ( $O_n$ ) de los que tenemos información sobre unas variables determinadas ( $X_k$ ).

El siguiente paso consiste en establecer un indicador que nos explique en qué medida cada par de observaciones se parece entre sí. Esta medida recibe el nombre de distancia o similaridad.

En el paso tres, lo que se pretende es crear grupos y que cada uno de ellos contenga aquellas observaciones que más se parezcan entre sí, siempre teniendo en cuenta la medida de la distancia que hemos calculado en el paso anterior.

Para la creación de estos grupos tenemos dos técnicas de conglomerados, el jerárquico y el no jerárquico, que a su vez dentro de estas se pueden utilizar distintos métodos, como por el ejemplo, método del centroide, de Ward, de vecino más cercano, etc.

En nuestro caso, hemos considerado para el proyecto utilizar ambas técnicas, de forma que si con dichas técnicas obtenemos resultados muy similares, las conclusiones derivadas de los mismos quedan corroboradas por dos técnicas distintas. Estas técnicas serán detalladas en los apartados posteriores, incluyendo así tanto definición como la explicación del resultado y proceso llevado a cabo en cada caso.

Y para finalizar el proceso, en el último paso es dónde debemos describir los grupos que hemos obtenido y compararlos unos con los otros, por lo que sólo hace falta ver qué valores toman las variables utilizadas en dicho análisis en cada uno de los grupos obtenidos.

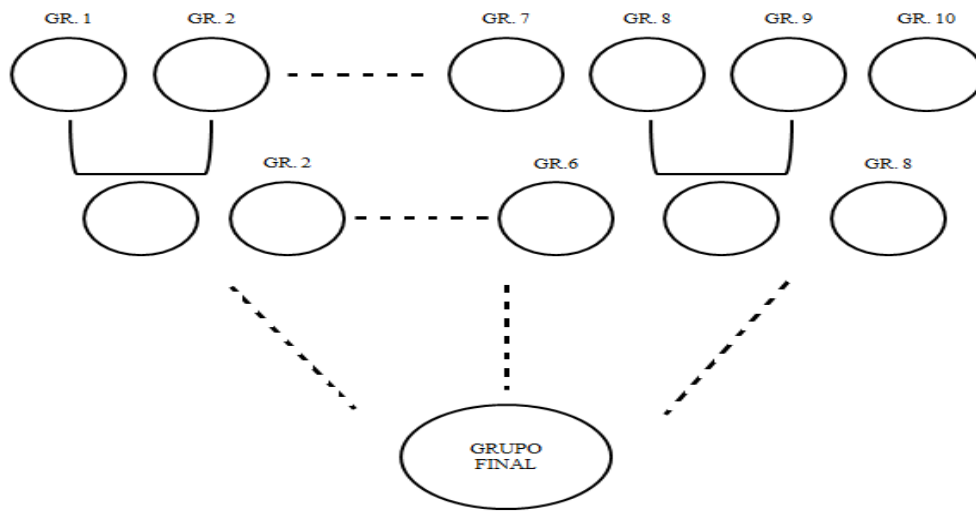
### **5.1 Análisis Clúster Jerárquico**

En esta técnica se consideran, inicialmente, a todas las observaciones como un grupo, es decir, en nuestro caso, al principio tenemos 10 grupos, ya que son 10 años los que hemos observado para cada una de las siete variables incluidas en el análisis.

A continuación, este método lo que hace es ir secuencialmente agrupando los distintos grupo en otros mayores, fusionándolos según la distancia a la que se encuentren, que será definida de antemano mediante otro tipo de técnica específica, con el fin de que todos los individuos iniciales confluyan en un único grupo final.

En este proyecto, se ha utilizado este método tanto para averiguar el número de grupos o conglomerados más adecuado como para un estudio del comportamiento de nuestras variables, a diferencia del análisis no jerárquico, en el cual, ya sabemos el número de grupos de antemano por tanto sólo realizamos el análisis de los resultados obtenidos. Esto se debe a que si no se tiene a priori una idea del número de grupos o conglomerados, utilizar la técnica de análisis jerárquicos es la mejor opción para determinar el número de grupos más adecuado, y una vez tengamos el resultado, emplear el número de grupos obtenido con la técnica no jerárquica en un análisis no jerárquico, ya que este debe aplicarse sabiendo de antemano el número de conglomerados.

**Figura 5.1.1: Esquema general análisis jerárquico**



Fuente: Elaboración propia

Para una mejor comprensión del funcionamiento de este tipo de métodos, hemos incluido un esquema muy general, (figura 5.1.1) en el cual aparece recogido el proceso completo de esta técnica.

Como se puede ver, y como se explicaba en el apartado anterior, tenemos una serie de grupos iniciales (los diez años considerados) y según las distancias calculadas estos individuos se unen para formar otros grupos mayores y así sucesivamente hasta agruparlos a todos en un grupo final.

A continuación, veremos el proceso que hay que llevar a cabo para determinar el número de grupos, ya que se requiere una explicación previa de cada paso a seguir, así como el cálculo de la distancia de proximidad o similaridad, los criterios necesarios para establecer las fusiones entre los individuos y el criterio utilizado para determinar el número de grupos más razonable, que utilizaremos para dar una explicación detallada de cómo ha afectado la crisis a este sector junto con las variables de estudio.

### *5.1.1 Determinación del número de grupos*

Cómo en el caso de la técnica de análisis de componentes principales, para obtener unos buenos resultados, lo primero es tipificar todas las variables para evitar problemas derivados de la escala.

Una vez hecho esto, habrá que calcular las medidas de proximidad o similitud, las cuales serán definidas como la distancia existente entre dos elementos. Estas medidas son necesarias para saber si dichos grupos están cercanos entre sí.

En principio existen muchas alternativas para elegir el cálculo de dicha distancia, como por ejemplo, coseno, distancia de Chebychev, correlación de Pearson, bloque... pero en nuestro caso en particular, vamos a utilizar la distancia euclídea al cuadrado, por ser la más usada y clara.

**Tabla 5.1.1.1: Tabla Matriz de Proximidades**

Caso	Distancia euclídea al cuadrado									
	1:2005	2:2006	3:2007	4:2008	5:2009	6:2010	7:2011	8:2012	9:2013	10:2014
1:2005	,000	1,016	4,187	13,077	21,180	11,975	11,863	18,685	20,122	19,707
2:2006	1,016	,000	1,214	10,925	23,133	17,006	15,132	22,699	24,732	26,554
3:2007	4,187	1,214	,000	11,316	28,278	26,056	22,273	30,682	33,444	37,843
4:2008	13,077	10,925	11,316	,000	6,363	17,654	11,065	14,537	16,991	30,746
5:2009	21,180	23,133	28,278	6,363	,000	9,935	6,289	6,890	7,863	20,327
6:2010	11,975	17,006	26,056	17,654	9,935	,000	1,551	3,402	3,012	2,783
7:2011	11,863	15,132	22,273	11,065	6,289	1,551	,000	,991	1,315	5,246
8:2012	18,685	22,699	30,682	14,537	6,890	3,402	,991	,000	,155	5,469
9:2013	20,122	24,732	33,444	16,991	7,863	3,012	1,315	,155	,000	4,318
10:2014	19,707	26,554	37,843	30,746	20,327	2,783	5,246	5,469	4,318	,000

Fuente: Elaboración propia

La distancia euclídea al cuadrado se obtiene a través de vectores, realizando una serie de operaciones teniendo como resultado la tabla 5.1.1.1, matriz de proximidades, que muestra la distancia de cada pareja.

Esta tabla no es necesaria para llevar a cabo el análisis clúster, pero de esta forma el estudio es más completo. Como se puede ver, la distancia entre el 2012 y 2013 es muy pequeña (0,155) por tanto la técnica jerárquica comenzará agrupando ambas observaciones en un mismo grupo, y así sucesivamente, iremos fusionando los grupos más cercanos entre sí hasta que todas las observaciones queden agrupadas en un único grupo. Como resultado de esta fusión total, la técnica jerárquica nos proporcionará las herramientas necesarias para determinar el número de grupos más adecuado, es decir nos permitirá analizar cuando las fusiones de los grupos dejan de tener sentido pues se encuentran a una distancia muy considerable.

Otro elemento importante a tener en cuenta en un análisis jerárquico, además de la medida de distancia seleccionada, es la forma de fusionar los grupos. Existen varios procedimientos para llevar a cabo dicha actuación, como pueden ser, vecino más próximo, vecino más lejano, método de Ward, etc.

En el presente trabajo vamos a utilizar el método de agrupación de centroides por ser uno de los más utilizados, aun así cualquier método de fusión de grupos nos serviría, ya que no se ha llegado a establecer preferencias por ningún método.

Para la explicación de esta técnica nos sirve también la figura 5.1.1. Este método comienza por agrupar aquellos grupos más próximos entre sí. A continuación, el grupo que se forma toma el valor que se obtiene de la media de dichas variables, es decir, el grupo queda representado por el centroide que lo representa. Una vez hecho esto, se vuelve a calcular la matriz de distancias, ya que se ha reducido el número de grupos y se han formado otros nuevos. Este proceso se repetirá una y otra vez hasta que todos los individuos están en un solo grupo.

Una vez llegados hasta aquí, para terminar, hay que determinar el número de grupos o conglomerados, ya que existen numerosas posibilidades. Habrá que ver que solución es el más adecuado y razonable para nuestro análisis.

**Tabla 5.1.1.2: Historial de conglomeración**

Etapa	Clúster combinado		Coeficientes	Primera aparición del clúster de etapa		Etapa siguiente
	Clúster 1	Clúster 2		Clúster 1	Clúster 2	
1	8	9	,155	0	0	3
2	1	2	1,016	0	0	5
3	7	8	1,114	0	1	4
4	6	7	2,382	0	3	6
5	1	3	2,447	2	0	9
6	6	10	3,802	4	0	8
7	4	5	6,363	0	0	8
8	4	6	11,509	7	6	9
9	1	4	16,947	5	8	0

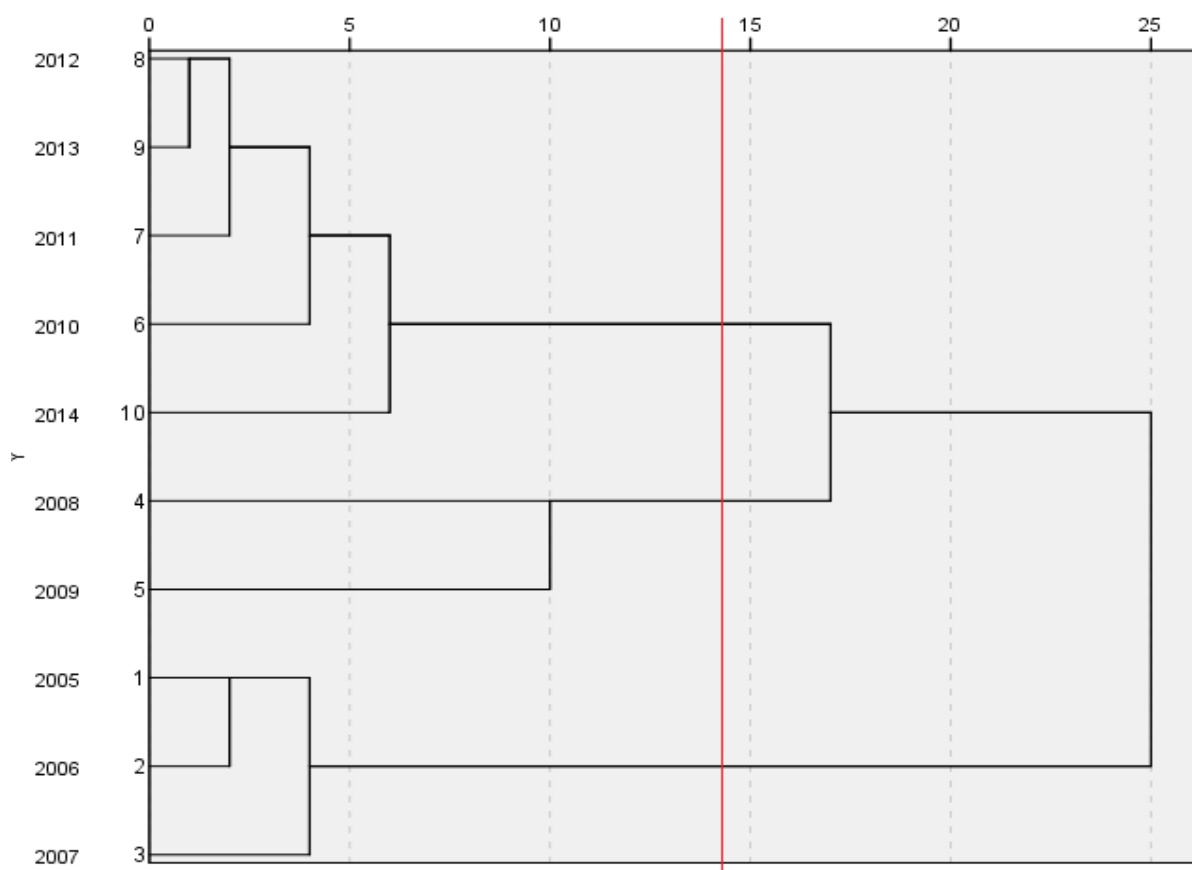
Fuente: Elaboración propia

La siguiente tabla 5.1.1.2 se le denomina historial de conglomeración. La interpretación de dicha tabla es muy importante y de gran utilidad, junto con otra gráfica, que denominamos, dendrograma, ya que el estudio en conjunto de estas dos tablas es lo que nos permite saber el número de grupos que se deben retener.

Analizando la tabla superior por partes, podemos ver que está formada por varias secciones. La primera columna, llamada etapa, nos muestra el número de etapas que se han llevado a cabo hasta agrupar a todos los grupos en uno sólo. La sección, clúster combinado, nos indica que grupos se han fusionado, en el primera caso, por ejemplo, tenemos el ocho y el nueve, por tanto se forma un nuevo grupo que se denotará con el menor número de los dos y será utilizado en otras etapas. La columna de coeficientes nos indica la distancia que hay entre las parejas, y como se puede observar, la primera agrupación es la que hemos comentado en el apartado de la tabla de las distancias.

En la otra parte de la tabla, tenemos la sección, primera aparición del clúster de etapa, en la cual se indica para cada grupo la etapa en la que salen por primera vez los elementos agrupados, de ahí que en el primer caso, se ponga cero, ya que son las observaciones originales. Y por último, la columna, próxima etapa, nos indica cuando un grupo volverá a utilizarse, tomando como ejemplo el primer caso, vemos como es en la tercera etapa cuando vuelve a aparecer.

**Tabla 5.1.1.3: Dendograma**



Fuente: Elaboración propia

Por último, tenemos el Dendograma (tabla 5.1.1.3), que junto con el historial de conglomerados se puede determinar el número de grupos más razonable para seguir adelante con el estudio.

En cuanto a su interpretación, tenemos que tener en cuenta que el análisis jerárquico lo que hace es ir fusionando los grupos más cercanos entre sí y así sucesivamente, pero llega un punto en el que los grupos fusionados están muy distantes, por ejemplo, mientras en la primera etapa la distancia era de 0'155, en la etapa 7 es de 11,509 unidades, por lo que hemos señalado con una línea roja el punto donde parece razonable dejar de fusionar.

Observando el historial de conglomerados y el dendograma, podemos concluir que el número de grupos razonable sean tres, los formados por los años 2005, 2006 y 2007; otro formado por los años 2008 y 2009 y por último el formado por los años restantes.



**Tabla 5.1.1.4: conglomerado de permanencia**

Caso	3 clústeres
1: 2005	1
2: 2006	1
3: 2007	1
4: 2008	2
5: 2009	2
6: 2010	3
7: 2011	3
8: 2012	3
9: 2013	3
10: 2014	3

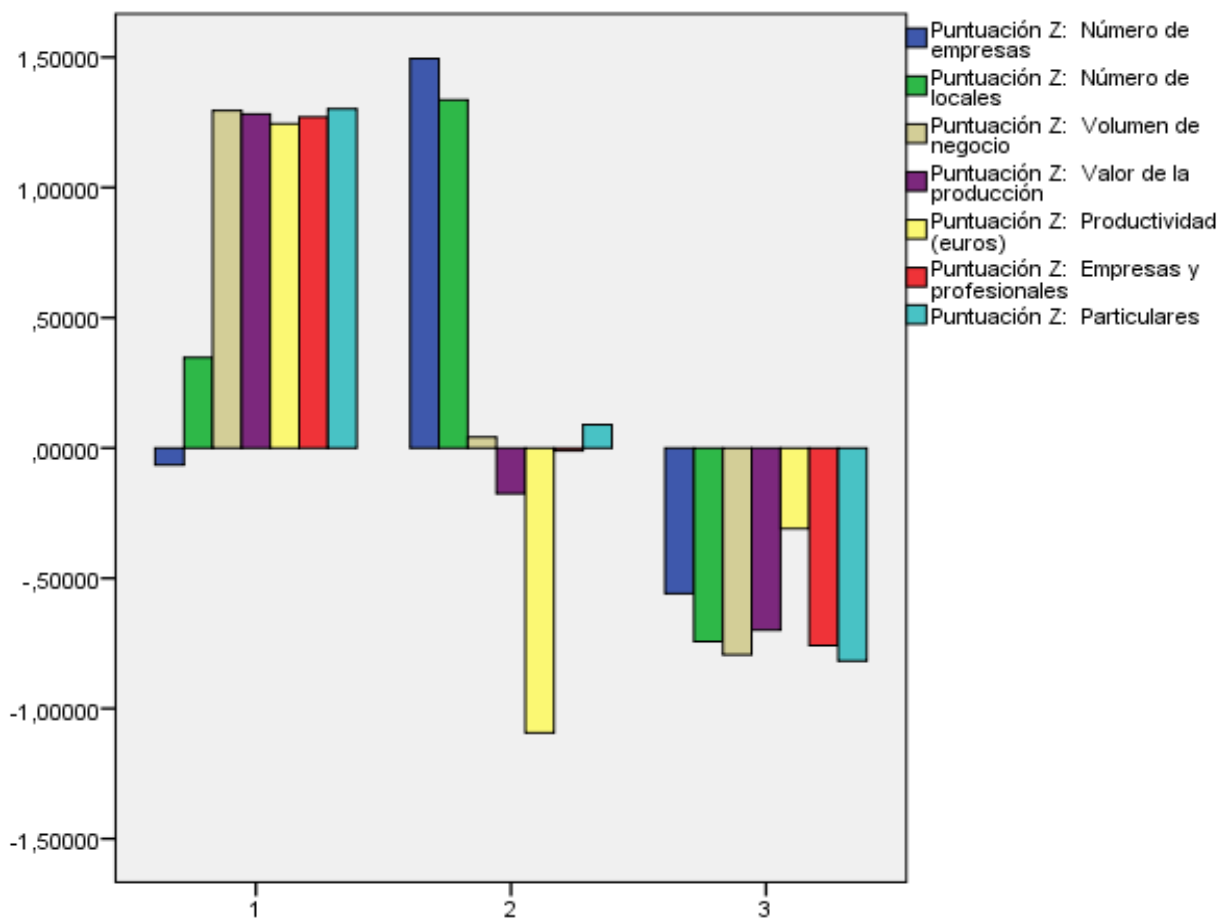
Fuente: Elaboración propia

Una vez determinado los grupos, vamos a analizar los grupos obtenidos y para ello analizaremos tabla 5.1.1.4. Dicha tabla nos aclara a que conglomerado pertenece cada año, y se denomina, conglomerado de pertenencia. En ella, podemos observar que el análisis clúster ha agrupado los años 2005, 2006 y 2007 en un mismo grupo (denominado aquí grupo 1). En el grupo 2, tendríamos los años 2008 y 2009 y finalmente en el grupo 3 tendríamos el resto de años.

### *5.1.2 Interpretación de los resultados obtenidos*

Una vez hemos explicado todo el proceso del análisis clúster y hemos determinado el número de grupos adecuado, así como hemos observados los grupos obtenidos procedemos al análisis de los mismos y para ello llevaremos a cabo la realización e interpretación de gráfico de barras para cada grupo con las variables de análisis tipificadas.

Gráfico 5.1.2.1: Diagrama de barras de conglomerados



Fuente: Elaboración propia

En dicho gráfico, las variables utilizadas no son las originales si no las variables tipificadas, de ahí que las variables puedan tomar valores negativos, es decir, por debajo de la media.

Como se puede observar en el diagrama, tenemos tres gráficos de barras, uno para cada caso, cada cual agrupa a los diferentes años y explica cómo se comportan las variables objeto de estudio.

Por tanto, si interpretamos dichos resultados, en el caso uno (años 2005, 2006 y 2007) que hace referencia justo al periodo anterior a la crisis, se ve como la variable número de locales y empresas se encuentran bastante por debajo del resto, llegando a ser negativo el número de empresas. Con esto podemos decir que durante esos años había poca

competencia y por tanto los niveles de producción, cifra de ventas... eran bastante elevados.

Luego pasamos al caso dos, que representa los años (2008 y 2009) más crudos del periodo de recesión que vivimos actualmente. Como se puede observar, y a diferencia del caso anterior, las variables número de locales y empresas se ha elevado considerablemente por encima de la media y del resto de variables, las cuales se han reducido hasta niveles incluso negativos como es el caso de valor de la producción y la productividad.

Y por último, el caso número tres, que agrupa desde el año 2010 hasta el 2014, dónde todas las variables se encuentran muy por debajo de la media.

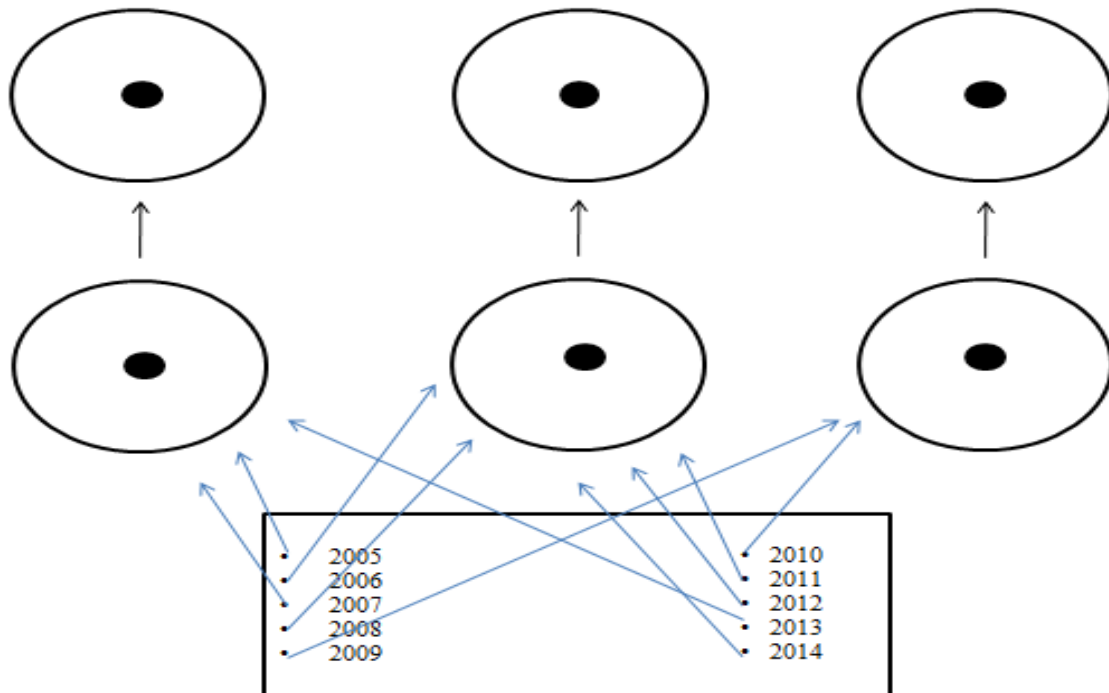
Como conclusión final, al ver el diagrama de barras, podemos interpretar y concluir que antes de la crisis, al haber pocas empresas y por tanto poca competencia, la rentabilidad de dichas empresas era bastante buena, teniendo niveles de volumen de negocio y productividad bastante altos. Al entrar la crisis, y por tanto bajar los precios, el número de empresas y locales aumenta, como se ve en el gráfico, haciendo que la competitividad sea mayor, y junto con el mal momento económico del país, hace que se reduzca considerablemente la productividad, así como la cifra de negocios y la venta a particulares y empresas, y por tanto, se mantenga esa situación a lo largo de los años siguientes, como se puede observar en el caso tres, aunque acabe afectando también a las empresas y locales que se acaban reduciendo de manera drástica.

## **5.2 Análisis Clúster no Jerárquico**

Otra de las técnicas que se pueden utilizar, como ya hemos explicado al principio del apartado cinco, es el método clúster no jerárquico. Con esta técnica se pretende conseguir el mismo resultado que el método jerárquico, pero con ciertas diferencias en el proceso.

En el no jerárquico, el número de grupos inicial se establece a priori, ya que como hemos comentado anteriormente, hemos utilizado la técnica de análisis jerárquica para el cálculo de los grupos, los cuales ahora se utilizan en este método. Por tanto, a cada grupo se les van asignando los diferente individuos, siendo estos los más homogéneos posibles dentro de cada grupo pero lo más diferentes posibles al resto de grupos iniciales.

Figura 5.2.1: Esquema general análisis clúster no jerárquico



Fuente: Elaboración propia

La figura 5.2.1 nos muestra un esquema básico y general de cómo funciona el análisis clúster no jerárquico. Simplemente es aplicar el párrafo anterior al dibujo.

El primer paso, como en el caso del jerárquico, hay que tipificar todas las variables para darles la misma importancia y evitar así la influencia de las unidades de medida.

Al principio tenemos los grupos, conocidos de antemano y en este caso tres. El programa calcula los centroides iniciales, que son los puntos negros que se pueden apreciar dentro de cada conglomerado, que se conocen como semillas. Estas pueden ser fijadas por la persona que lleva a cabo el estudio o investigación, o pueden ser elegidas aleatoriamente por medio del programa SPSS.

El siguiente paso consiste en colocar cada individuo u observación en el grupo cuyo centroide esté más cerca según la distancia euclídea. Este paso es representado por las flechas azules.

Una vez hecho esto, y para terminar, se vuelve a calcular los centroides de cada grupo de acuerdo a los nuevos individuos que se han clasificado. Si el cambio en dichas semillas, es decir, la distancia entre los nuevos y los viejos centroides, es mayor que un valor de

convergencia establecido de antemano entonces se vuelve a realizar el paso de clasificación de las observaciones, así hasta que se cumpla el criterio de convergencia, es decir, que quede estable, o se supere un determinado número de iteraciones.

### 5.2.2 Descripción detallada del proceso

En este análisis vamos a utilizar el método de K-medias de conglomerados, el cual sigue los mismos pasos explicados en el apartado anterior.

**Tabla 5.2.2.1: Centroides iniciales**

	Clúster		
	1	2	3
Puntuación Z: Número de empresas	-1,61190	1,78707	,56777
Puntuación Z: Número de locales	-1,76089	1,70825	,80873
Puntuación Z: Volumen de negocio	-,88791	,39767	1,55106
Puntuación Z: Valor de la producción	-,95552	,48158	1,68639
Puntuación Z: Productividad (euros)	,12180	-,37841	1,52295
Puntuación Z: Empresas y profesionales	-,78718	,51391	1,67399
Puntuación Z: Particulares	-,98206	,24138	1,37177

Fuente: Elaboración propia

La tabla 5.2.2 muestra la primera del proceso, dónde aparecen las semillas iniciales de cada uno de los tres conglomerados. En este caso, éstas han sido obtenidas aleatoriamente por el programa.

**Tabla 5.2.2.2: Historial de iteraciones**

Iteración	Cambiar en centros de clústeres		
	1	2	3
1	1,560	1,261	1,043
2	,000	,000	,000

Fuente: Elaboración propia

A continuación, tenemos una de las tablas más importantes del método K-medias. Dicha tabla 5.2.2.2 nos muestra el número de repeticiones, ya que éste es un método iterativo, que se ha llevado a cabo hasta que se han estabilizado las distancias, que también aparecen en el cuadro. Como podemos ver, sólo se ha realizado una iteración, ya que a la segunda repetición las distancias ya son estables (0,000) y por lo tanto no es necesario seguir con el proceso, ya se ha alcanzado el criterio de convergencia.

**Tabla 5.2.2.3: Clúster de pertenencia**

Número del caso	años1	Clúster	Distancia
1	2005	3	1,011
2	2006	3	,174
3	2007	3	1,043
4	2008	2	1,261
5	2009	2	1,261
6	2010	1	1,010
7	2011	1	,831
8	2012	1	,935
9	2013	1	,794
10	2014	1	1,560

Fuente: Elaboración propia

Esta tabla 5.2.2.3 nos indica, para cada individuo u observación, el conglomerado al que pertenece. Por tanto, el grupo tres está formado por los años 2005, 2006 y 2007, el grupo dos por el 2008 y 2009 y el resto pertenecen al grupo uno. También aparece una columna, que nos indica la distancia, por ejemplo, la distancia del año 2005 a la semilla del grupo tres es de 1,011.

**Tabla 5.2.2.4: Centroides finales**

	Clúster		
	1	2	3
Puntuación Z: Número de empresas	-,55905	1,49497	-,06490
Puntuación Z: Número de locales	-,74290	1,33623	,34734
Puntuación Z: Volumen de negocio	-,79397	,04185	1,29539
Puntuación Z: Valor de la producción	-,69895	-,17529	1,28178
Puntuación Z: Productividad (euros)	-,30954	-1,09304	1,24460
Puntuación Z: Empresas y profesionales	-,75796	-,01069	1,27039
Puntuación Z: Particulares	-,81714	,08972	1,30209

Fuente: Elaboración propia

La tabla 5.2.2.4 nos proporciona información muy importante acerca de las características de los diferentes grupos, ya que con su interpretación podemos ver que variables tienen valores bajos y cuales altos, es decir, en que destaca cada uno, en nuestro caso en particular, nos permite diferenciar los periodos o ciclos de la crisis.

En primer lugar, la tabla nos muestra los centroides finales, los cuales pueden ser también negativos ya que dichas variables están tipificadas. Además, los datos numéricos que aparecen en dicha tabla nos mide la altura de las barras del gráfico 5.2.2.1 que veremos más adelante.

Dicho todo esto, vamos a analizar un poco más detalladamente los datos que aparecen. Si observamos al grupo uno, se puede ver que todas las variables están por debajo de la media, es decir, son negativas, por tanto se intuye que estos años son los pertenecientes al periodo posterior a la crisis económica, en el cual, las pocas empresas que hay tienen valores de volumen de negocios y productividad muy por debajo de los años anteriores a dicha recesión.

El grupo dos presenta datos un poco mejores pero aun así siguen siendo negativos, salvo el número de empresas y locales. En este grupo se clasifican los años de comienzo de la crisis, ya que es cuando explota la burbuja y los valores de las variables, sobre todo la productividad (-1,09304), comienzan a descender, aunque siga habiendo gran número de empresas.

Y por último, el grupo uno que incluye los años pre-crisis claramente, ya que si observamos los valores de todas las variables, excepto del número de empresas, se encuentran muy por encima de la media.

**Tabla 5.2.2.4: ANOVA**

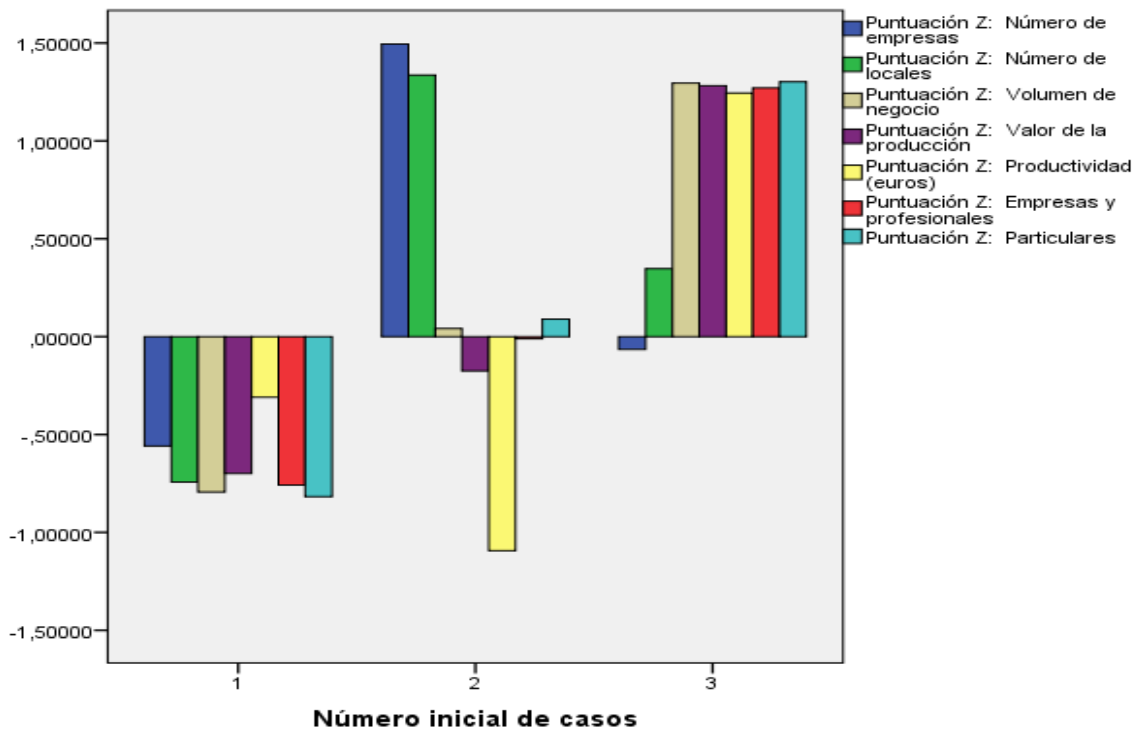
	Clúster		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
Puntuación Z: Número de empresas	3,023	2	,422	7	7,160	,020
Puntuación Z: Número de locales	3,346	2	,330	7	10,151	,009
Puntuación Z: Volumen de negocio	4,095	2	,116	7	35,368	,000
Puntuación Z: Valor de la producción	3,716	2	,224	7	16,602	,002
Puntuación Z: Productividad (euros)	3,758	2	,212	7	17,721	,002
Puntuación Z: Empresas y profesionales	3,857	2	,184	7	21,002	,001
Puntuación Z: Particulares	4,221	2	,080	7	52,858	,000

Fuente: Elaboración propia

Y ya terminando, tenemos la tabla ANOVA (tabla 5.2.2.4) que nos indica la importancia que ha tenido cada variables en la formación de los conglomerados. Dicha información se encuentra en la columna 'F'. Por tanto, vemos que las variables particulares, volumen de negocio y empresas son las que más han influido en la formación de los grupos.



**Gráfico 5.2.2.1: Gráfico de barras análisis no jerárquico**



Fuente: Elaboración propia

Y como en el método jerárquico, como resultado final obtenemos un gráfico de barras (figura 5.2.2.5) en el que se puede ver de forma más expresiva y clara los datos analizados en la figura 5.2.2.4, la tabla de centroides finales.

En conclusión, podemos ver como el gráfico de barras, que es el resultado final del análisis clúster no jerárquico, coincide totalmente con el gráfico del primer método, con la única diferencia de que se han invertido los grupos, es decir, ahora el grupo uno es el grupo tres, el grupo dos sigue siendo el mismo y el grupo tres es el grupo uno. Ambos análisis llegan a la misma conclusión por tanto se corroboran uno del otro.

## **6. Conclusión**

En dicho proyecto se ha llevado a cabo un estudio o análisis del sector servicio, más concretamente del sector automovilístico, a través de un programa estadístico, SPSS. Utilizando para ello dos técnicas descriptivas, como son componentes principales y clúster o de conglomerados.

Podemos concluir que el sector comercio, y dentro de este el sector automovilístico, es un pilar fundamental de la economía española, junto con el turismo y la industria agroalimentaria, y más en estos últimos años y los que vienen.

En cuanto a los métodos utilizados para nuestro análisis, podemos ver como todas estas técnicas nos dan unos resultados que tienen la misma interpretación. Es decir, todas las técnicas nos diferencian tres periodos clave, como son los años anteriores a la crisis, los años de inicio y los posteriores a esta. En los cuales, cada periodo presenta unas características, como es tener elevados niveles de productividad y/o cifra de negocios o reducidos niveles de empresas y locales que nos permiten saber a qué periodo pertenece.

También hemos llevado a cabo un estudio particular con variables parecidas sobre las diferentes comunidades autónomas españolas, pero esta vez sólo hemos realizado un análisis de componentes principales y sólo un año como periodo de estudio, el 2014. El cual nos permite saber que comunidades destacan más y en qué, en nuestro caso, Madrid, como la que más destaca en inversión, sueldos y salarios..., Cataluña y Andalucía respectivamente.

## **7. Bibliografía**

Instituto Nacional de Estadística (INE). Servicios. Sector Comercio

Ezequiel Uriel, Joaquín Aldás (2005). Análisis Multivariante Aplicado. Thomson.

Daniel Peña (2002). Análisis de Datos Multivariantes. S.A. Mcgraw-Hill/Interamericana de España.

Junta de Andalucía. El sector comercial en España y Andalucía. Monográfico 2013

El Mundo. Sección Macroeconomía. Noticia: Los servicios tiran del PIB (21/05/2014)

Asociación Española de Fabricantes de Equipos y Componentes para Automoción (SERNAUTO). Agentes del sector automoción

El Economista. Ecomotor. Noticia: El sector de la automoción, un motor de empleo cualificado en España (22/10/2014)

El Confidencial. Sección Motor. Noticia: El sector de la automoción, locomotora de la economía española (29/12/2014)

Marca España. Industrias avanzadas. Automóvil, motor económico.

M. Belén Castañeda, Alberto F.C, Yadira N., Wietse de Vries (2010). Procesamiento de datos y análisis estadísticos utilizando SPSS. EdiPUCRS.

Informe de situación sobre el sector de automoción en España. CCOO de Industria. (2015)

