

Trabajo Fin de Grado

Análisis y clasificación de los ciudadanos de Zaragoza respecto a su comportamiento medioambiental

Autora

Sheila Continente Cáncer

Directores

Pilar Gargallo Valero

Jesús Ángel Miguel Álvarez

Facultad de Economía y Empresa

Universidad de Zaragoza

Junio 2016

Resumen

Este trabajo pretende analizar el comportamiento medioambiental de los ciudadanos de Zaragoza a partir de la información recogida a través de una encuesta realizada a una muestra de 200 individuos, estratificada por género y junta municipal. Con el fin de resumir la información contenida en la encuesta y poder reducir la dimensión de las 33 variables en 5 factores, se aplica la técnica multivariante del Análisis Factorial. Los 5 indicadores encontrados revelan aspectos importantes sobre la concienciación, la actitud y la responsabilidad de los individuos en el cuidado del medioambiente. A partir de ellos, se utiliza el Análisis Clúster para realizar una clasificación de los individuos en tres grupos diferentes: el primero está formado por aquellos ciudadanos con pocos recursos cuya mayor preocupación son sus necesidades básicas; el segundo grupo de individuos que muestra interés por el medioambiente pero no asumen su propia responsabilidad y, el tercer grupo constituido por personas que tienen una actitud activa y responsable con el medioambiente. Finalmente, se determina el perfil de cada grupo en función de algunas de sus características socio-económicas.

Palabras clave: Medioambiente, Diseño muestral y Tratamiento de encuestas, Análisis Factorial, Análisis Clúster

Abstract

This study analyzes the environmental behavior of citizens of Zaragoza from the information obtained by a survey of a sample of 200 individuals, stratified by gender and municipal board. The multivariate technique of Factor Analysis is applied for summarizing survey data and for reducing the dimension of 33 variables into 5 factors. The 5 indicators found reveal the most important aspects of awareness, attitude and responsibility of individuals in taking care for the environment. From these indicators, the Cluster Analysis is used to make a classification of individuals into three different groups: the first consists of those citizens with few resources whose main concern is their basic needs; a second group of individuals, which show interest on the environment but that they do not assume their own responsibility, and the third group, who have an active and responsible attitude to the environment. Finally, the profile of each group based on some of their socio-economic characteristics is determined.

Key words: Environment, Design and Surveys Sampling, Factor Analysis, Cluster Analysis

CONTENIDO

1. INTRODUCCIÓN.....	9
2. TÉCNICAS UTILIZADAS	11
2.1. ANÁLISIS FACTORIAL	11
2.1.1. <i>Formulación del problema.....</i>	<i>11</i>
2.1.2. <i>Análisis de la matriz correlación</i>	<i>12</i>
2.1.3. <i>Extracción de factores.....</i>	<i>13</i>
2.1.4. <i>Determinación del número de factores</i>	<i>14</i>
2.1.5. <i>Interpretación de factores</i>	<i>15</i>
2.1.6. <i>Rotación de factores.....</i>	<i>15</i>
2.1.7. <i>Cálculo puntuaciones factoriales.....</i>	<i>16</i>
2.2. ANÁLISIS CLÚSTER.....	16
2.2.1. <i>Formulación del problema.....</i>	<i>17</i>
2.2.2. <i>Métodos de clasificación.....</i>	<i>17</i>
2.2.2.1. Clúster jerárquico.....	18
2.2.2.1.1. <i>Medida de asociación</i>	<i>18</i>
2.2.2.1.2. <i>Métodos de conglomeración</i>	<i>18</i>
2.2.2.1.3. <i>Número de grupos</i>	<i>19</i>
2.2.2.2. Clúster K-medias	19
2.2.2.3. Clúster bietápico	20
3. ELECCIÓN DEL CONJUNTO DE DATOS	21
3.1. DETERMINACIÓN DE LA MUESTRA.....	21
3.2. DISEÑO DE LA ENCUESTA.....	22
3.3. ANÁLISIS DESCRIPTIVO.....	23
4. RESULTADOS DEL ANÁLISIS FACTORIAL	24
5. RESULTADOS ANÁLISIS CLÚSTER.....	30
6. CONCLUSIONES	42
7. BLIBLIOGRAFÍA	46

ÍNDICE DE FIGURAS

Figura 1. Gráfico de sedimentación.....	28
Figura 2. Resumen del modelo bietápico	31
Figura 3. Diagrama de sectores en función del tamaño de los clústeres	32
Figura 4. Importancia del predictor	32
Figura 5. Comparación de clústeres respecto de cada factor.....	33
Figura 6. Barras de error del factor “comportamiento saludable” por clúster.....	34
Figura 7. Barras de error del factor “compromiso ecológico” por clúster	34
Figura 8. Barras de error del factor “actitud respetuosa” por clúster	34
Figura 9. Barras de error del factor “ideología ecológica” por clúster.....	34
Figura 10. Barras de error del factor “posición ahorradora” por clúster	34
Figura 11. Diagrama de barras de la variable género por clúster	35
Figura 12. Barras de error de la variable edad por clúster.....	36
Figura 13. Diagrama de barras de la variable tipo de hogar por clúster.....	36
Figura 14. Diagrama de barras de la variable n° de miembros del hogar por clúster....	37
Figura 15. Diagrama de barras de la variable nivel de estudios por clúster	37
Figura 16. Diagrama de barras de la variable ingresos mensuales por clúster.....	38
Figura 17. Barras de error del problema del desgaste del medio ambiente por clúster.	39
Figura 18. Barras de error del problema de la productividad por clúster.....	39
Figura 19. Barras de error del problema población desempleada por clúster	39
Figura 20. Barras de error del problema del fracaso escolar por clúster	39
Figura 21. Barras de error del problema del sistema de pensiones por clúster	39
Figura 22. Barras de error de la política “prohibición de las bolsas de plástico en los comercios” por clúster	40

Figura 23. Barras de error de la política “regulación de la emisión de gases de los vehículos” por clúster	40
Figura 24. Barras de error de la política “prohibición a la empresas de emitir humos y gases” por clúster.....	40
Figura 25. Barras de error de la política “prohibición de la tala de árboles en bosques y ciudades” por clúster	40
Figura 26. Barras de error de la política “obligación a las empresas a tener plantas de tratamientos de aguas” por clúster.....	40

ÍNDICE DE TABLAS

Tabla 1. Población original estratificada por sexo y junta municipal	21
Tabla 2. Muestreo estratificado por junta municipal y sexo.....	22
Tabla 3. Matriz de correlaciones.....	26
Tabla 4. Prueba de KMO y Bartlett.....	26
Tabla 5. Matriz anti-imagen	26
Tabla 6. Comunalidades	27
Tabla 7. Varianza total explicada	28
Tabla 8. Matriz de componente	29
Tabla 9. Matriz de componente rotado	29
Tabla 10. Porcentaje de ciudadanos en cada intervalo de ingresos mensuales por clúster	44
Tabla 11. Porcentaje de ciudadanos en cada intervalo de estudios por clúster	44

1. INTRODUCCIÓN

Hoy en día uno de los temas que cobra mayor interés es la preocupación por el medioambiente. Diversos estudios avalan la crisis que vive nuestro planeta, y por ello, es una cuestión que nos incumbe a todos.

Cada vez más países se unen a la lucha contra el cambio climático, un ejemplo de ello, es la Cumbre del Clima celebrada en noviembre de 2015, donde representantes de cerca de 200 países adoptaron el primer acuerdo global para atajar el calentamiento desencadenado por el hombre, con sus emisiones de gases de efecto invernadero.

Debido a este problema tan acuciante, la sociedad requiere que todos sus individuos estén concienciados, desde las grandes empresas, gobiernos, hasta el ciudadano de a pie. Por esta razón, hay una necesidad de comprender mejor las creencias y actitudes de los miembros de la población en general.

Diversos trabajos establecen que cuanto más desarrollado es un país, mayor es su preocupación por el entorno y el medioambiente. Esta teoría fue enunciada por *Inglehart* en los años 70 y desarrollada en los 90. Si lo pensamos detenidamente, este planteamiento tiene mucha lógica ya que un país pobre tiene otras muchas prioridades, empezando por el bienestar de sus ciudadanos. En cambio, un país desarrollado en el cual sus ciudadanos tienen aseguradas las necesidades básicas, su prioridad debería ser el cuidado y protección del medioambiente. Hay un gran debate en cuanto a esta teoría, uno de los trabajos que hace referencia a él es *García, 2004*.

Además, podemos pensar que el nivel de educación de una sociedad también será indicativo de su compromiso por el medio ambiente. Ya que una sociedad más culta y más formada estará más informada sobre los problemas medioambientales y los riesgos que conlleva.

Si llevamos esta teoría a una ciudad de un país desarrollado como es Zaragoza, podría plantearse la pregunta de si los ciudadanos con más renta y mayor nivel de estudios están más concienciados con la protección y cuidado del medioambiente.

Por todo ello, el objeto de este trabajo es crear una serie de indicadores que describan las tipologías de los ciudadanos de Zaragoza respecto de la concienciación y

comportamiento medioambiental. Además, se quiere clasificar a los encuestados en diferentes grupos homogéneos que sigan parecidas pautas en relación al comportamiento medioambiental. Para su realización, se han utilizado técnicas estadísticas multivariantes, como son el Análisis Factorial y el Análisis Clúster.

Para ello, se realizó una encuesta con 33 preguntas acerca de concienciación y comportamiento medioambiental (véase anexo I), a ciudadanos de Zaragoza. Dicha encuesta estaba dividida en cuatro apartados: datos generales, principios morales, concienciación medioambiental y comportamiento de la unidad familiar.

Dada la imposibilidad de realizar la encuesta al total de la población, se tomó una muestra representativa de ciudadanos de Zaragoza estratificada por sexo y por la junta municipal a la que pertenecían.

A continuación se realiza una explicación de las técnicas estadísticas utilizadas, para después seguir con los resultados obtenidos y las conclusiones.

2. TÉCNICAS UTILIZADAS

En este epígrafe se explican las técnicas estadísticas en las que se ha basado este estudio, análisis factorial y clúster.

2.1. ANÁLISIS FACTORIAL

El Análisis Factorial es una técnica estadística multivariante de reducción de la dimensión de los datos. En el análisis factorial todas las variables son independientes, no existe a priori una dependencia conceptual de unas variables sobre otras. Su propósito último consiste en buscar el número mínimo de dimensiones capaces de explicar el máximo de información contenida en los datos, con la menor distorsión posible.

Dado un conjunto de variables, mediante un análisis factorial, podemos reducir las variables a analizar, en un conjunto menor de indicadores inobservables, denominados factores comunes, que explican todas las covarianzas o correlaciones entre las variables. Cualquier porción de la varianza inexplicada por los factores comunes se asigna a términos de error residuales que llamaremos factores únicos o específicos.

Este análisis tiene sentido si se cumplen dos condiciones: parsimonia e interpretabilidad. Según el principio de parsimonia el número de factores debe ser lo más reducido posible y éstos deben ser susceptibles de interpretación sustantiva.

2.1.1. Formulación del problema

En la etapa de formulación del problema se seleccionan las variables a analizar y los elementos de la población en que dichas variables van a ser observadas. Es muy importante que las variables elegidas contengan los aspectos más relevantes de la temática que se quiere estudiar. Por ello, es recomendable que tengamos una idea bastante clara de cuáles son los factores comunes que queremos medir, para de acuerdo con ellos, elegir las variables. En caso contrario, corremos el riesgo de encontrar factores espurios o que los factores queden mal estimados por una mala selección de las variables.

2.1.2. Análisis de la matriz correlación

Una vez obtenidos los datos, el siguiente paso a realizar es hallar y examinar la matriz de correlaciones muestrales. Para que el análisis factorial sea adecuado, las variables tienen que estar altamente intercorrelacionadas, ya que de lo contrario el análisis no sería apropiado. Además, también se espera que las variables que tienen correlación muy alta entre sí la tengan con el mismo factor o factores.

Esta relación entre las variables se puede medir con distintos indicadores, a continuación citamos algunos.

1. Test de esfericidad de Bartlett

Contrasta, bajo la hipótesis de normalidad multivariante, si la matriz de correlación de las variables observadas es la identidad. Si una matriz de correlación es la identidad significa que las intercorrelaciones entre las variables son cero. Si se confirma la hipótesis nula significa que las variables no están intercorrelacionadas y, por tanto, el análisis factorial carecerá de sentido.

2. Matriz de correlación anti-imagen

Esta matriz de correlación proporciona fuera de la diagonal los coeficientes de correlación parcial. Si los coeficientes son bajos indicará que las variables comparten factores comunes ya que se eliminan los efectos lineales de las otras variables. Las correlaciones parciales son estimaciones de las correlaciones entre los factores únicos y deberían ser próximos a cero cuando el Análisis Factorial es adecuado, ya que, estos factores se supone que están incorrelados entre sí.

3. Medida de adecuación de la muestra KMO

KMO es un índice que toma valores entre 0 y 1 y que se utiliza para comparar las magnitudes de los coeficientes de correlación observados con las magnitudes de los coeficientes de correlación parcial de forma que, cuanto más pequeño sea su valor, mayor es el valor de los coeficientes de correlación parciales y, por lo tanto, menos deseable es realizar un Análisis Factorial. Kaiser, Meyer y Olkin aconsejan que si $KMO \geq 0,75$ la idea de realizar un análisis factorial es buena, si $0,75 > KMO \geq 0,5$ la idea es aceptable y si $KMO < 0,5$ es inaceptable.

4. Medida de adecuación de la muestra MSA

Es una medida de adecuación muestral para cada variable, tiene una forma similar al índice KMO. Sólo se incluyen los coeficientes de la variable que se desea comprobar. Un valor bajo de MSA indica que las hipótesis hechas por el modelo del Análisis Factorial son poco compatibles para el caso de esa variable.

La experiencia práctica nos dice que no es seguro emplear sólo estos índices si no que deberían complementarse con otras fuentes como pueden ser las comunalidades de cada variable, los residuos del modelo y la interpretabilidad de los factores obtenidos a la hora de tomar la decisión de eliminar una variable del estudio.

2.1.3. Extracción de factores

Si hemos concluido la fase anterior exitosamente, es decir, que el análisis factorial es una técnica adecuada para nuestros datos, podremos pasar a determinar el método más favorable para la extracción de los factores. Existen varios métodos cada uno con sus ventajas e inconvenientes

1. Método de las Componentes Principales

El método consiste en estimar las puntuaciones factoriales mediante las puntuaciones tipificadas de las k primeras componentes principales y la matriz de cargas factoriales mediante las correlaciones de las variables originales con dichas componentes. Este método tiene la ventaja de proporcionar siempre una solución. Tiene el inconveniente, sin embargo, de que al no estar basado en el modelo de Análisis Factorial puede llevar a estimadores muy sesgados de la matriz de cargas factoriales, particularmente, si existen variables con comunalidades bajas.

2. Método de los Ejes Principales

Método de estimación iterativo en el que, como estimación inicial de la comunalidad, la matriz de correlaciones original se reduce sustituyendo los unos de su diagonal por las estimaciones de la correlación múltiple al cuadrado entre cada variable y todas las demás. La matriz reducida se auto-descompone y se corrigen las estimaciones iniciales de la comunalidad por las nuevas estimaciones resultantes. El proceso continúa hasta que no existe diferencia entre las estimaciones de las comunalidades entre dos pasos sucesivos o se alcanza alguno de los criterios de parada.

3. Método de la Máxima Verosimilitud

Método basado en la hipótesis de normalidad multivariante. Consiste en aplicar el método de la máxima verosimilitud. Tiene la ventaja, sobre los dos anteriores, de que las estimaciones obtenidas no dependen de la escala de medida de las variables. Además tiene todas las propiedades estadísticas de este método y, en particular, es asintóticamente insesgada, eficiente y normal si las hipótesis del modelo factorial son ciertas. Su principal inconveniente radica en que, al realizarse la optimización de la función de verosimilitud por métodos iterativos, si las variables originales no son normales, puede haber problemas de convergencia sobre todo en muestras finitas.

4. Método de mínimos cuadrados no ponderados

Método que produce, para un número fijo de factores, una matriz de coeficientes que minimiza la suma de las diferencias al cuadrado entre las matrices de correlación observada, y la reproducida, eliminando, en dichas diferencias, los elementos de la diagonal.

5. Método de mínimos cuadrados generalizados

Método que minimiza el mismo criterio que el anterior pero ponderando las correlaciones inversamente por la varianza del factor específico. Este método permite, además, aplicar contrastes de hipótesis para determinar el número de factores.

6. Método de factorización por imágenes

Método consistente en aplicar el método del eje principal a la matriz de correlaciones, obtenida a partir de las partes predichas de las diversas regresiones lineales de cada una de las variables sobre las demás (dicha parte recibe el nombre de imagen de la variable).

2.1.4. Determinación del número de factores

La siguiente etapa es determinar el número de factores necesarios para explicar la estructura de los datos originales. La matriz factorial puede presentar un número de factores superior al que necesitamos. Por tanto, uno de los problemas que se plantean consiste en ver qué factores conviene conservar para cumplir con el principio de parsimonia. Para ello, existen varias reglas y criterios, uno de los más utilizados es la regla de Kaiser: *"conservar aquellos factores cuyos valores propios son mayores a la*

unidad”. La determinación “a priori”, el criterio del porcentaje de la varianza, el gráfico de Sedimentación y las pruebas de significación son otros de los criterios que se utilizan.

2.1.5. Interpretación de factores

En la fase de interpretación juega un papel preponderante la teoría existente sobre el tema. Para interpretar los factores, primero identificamos las variables cuyas correlaciones con el factor son las más elevadas en valor absoluto, a continuación intentamos dar un nombre a cada factor, el nombre debe asignarse en función de la estructura de sus correlaciones con las variables. Si dicha correlación es positiva (resp. negativa) la relación entre el factor y dicha variable es directa (resp. inversa).

Dos estrategias más pueden ayudar a interpretar los factores: a) ordenarlos y b) eliminar las cargas bajas. Se puede ordenar la matriz factorial de tal forma que las variables con cargas altas para el mismo factor aparezcan juntas. La eliminación de las cargas factoriales bajas también facilita la interpretación de los resultados al suprimir información redundante.

2.1.6. Rotación de factores

El objetivo de la rotación es que los factores obtenidos sean más fáciles de interpretar. Hay varios procedimientos de Rotación de Factores, rotación ortogonal (Varimax, Quartimax, Equamax,..) y rotación oblicua (Oblimin, Promax,..). Estos métodos intentan aproximar la solución obtenida al Principio de Estructura Simple (Thurstone, 1935) según el cual la matriz de cargas factoriales debe reunir las siguientes características:

- 1) Cada factor debe tener unos pocos pesos altos y los otros próximos a cero.
- 2) Cada variable no debe estar saturada más que en un factor.
- 3) No deben existir factores con la misma distribución, es decir, dos factores distintos deben presentar distribuciones diferentes de cargas altas y bajas.

Conviene señalar que tanto en la rotación ortogonal, como en la rotación oblicua la comunalidad de cada variable no se modifica, es decir, la rotación no afecta a la bondad de ajuste de la solución factorial: aunque cambie la matriz factorial, las especificidades

no cambian y, por tanto, las comunalidades permanecen inalteradas. Sin embargo, cambia la varianza explicada por cada factor cuantificada a través de su varianza.

2.1.7. Cálculo puntuaciones factoriales

Una vez determinados nuestros factores finales el siguiente paso es calcular la matriz de puntuaciones factoriales. Las puntuaciones factoriales nos permiten conocer, entre otras cuestiones, qué sujetos son los más raros o extremos, dónde se ubican ciertos grupos de la muestra, en qué factor sobresalen unos sujetos y en qué factor no. El Análisis Factorial es en otras ocasiones un paso previo a otros análisis, como por ejemplo, Regresión Múltiple o Análisis Clúster, en los que se sustituye el conjunto de variables originales por los factores obtenidos. Por ello, es necesario conocer los valores que toman los factores en cada observación.

Existen diversos métodos de estimación de la matriz. Sería deseable que los factores estimados cumplieran determinadas propiedades: cada factor estimado tenga correlación alta con el verdadero factor, cada factor estimado tenga correlación nula con los demás factores verdaderos, los factores estimados sean incorrelacionados dos a dos, los factores estimados sean estimadores insesgados de los verdaderos factores. Sin embargo, por la propia naturaleza de los factores comunes, el problema de su estimación es complejo. Se puede demostrar que los factores no son, en general, combinación lineal de las variables originales. Además, en la mayoría de las situaciones, no existirá una solución exacta ni siquiera será única.

Para el cálculo de la matriz existen varios métodos, cada uno con ventajas e inconvenientes. El programa SPSS nos da a elegir entre tres métodos: regresión, Bartlett y Anderson-Rubin.

2.2. ANÁLISIS CLÚSTER

El análisis clúster es una técnica de clasificación que distribuye los diferentes objetos de un conjunto de datos en grupos o conglomerados. Estos grupos deben tener dos características: los elementos de un mismo grupo tienen que ser similares unos de otros y, por el contrario, los elementos pertenecientes a distintos grupos deben tener un comportamiento diferente. Por tanto, podemos decir que los grupos deben estar cohesionados internamente y cada grupo estar aislado externamente del resto de grupos.

El análisis clúster no es un método inferencial como el análisis factorial, es una técnica eminentemente exploratoria. Normalmente, no se tienen que realizar supuestos acerca de las propiedades de la distribución de los datos. Por ello, este análisis es recomendable para conjuntos de datos de los que queremos extraer información sin imponer restricciones. Puede resultar muy útil para elaborar hipótesis sobre lo que se pretende estudiar sin tener teorías preconcebidas.

2.2.1. Formulación del problema

El primer paso para realizar el análisis clúster es la selección de las variables que queremos analizar y los individuos de nuestro conjunto de datos. Hay que tener especial cuidado con las variables que vamos a introducir ya que todas deben ser relevantes respecto del tema de estudio que queremos realizar. De lo contrario, la conclusión que obtendremos podría estar condicionada por dichas variables y no resultar útil. Como ya se ha comentado anteriormente (2.1.7.) puede resultar interesante factorizar previamente las variables y realizar el análisis clúster con factores en lugar de con variables.

Las variables pueden ser cuantitativas y cualitativas pero a la hora de realizar el análisis clúster las variables elegidas tienen que estar medidas en las mismas unidades, para ello podemos estandarizar los datos, variable por variable y así conseguir que todas las variables tengan la misma importancia en el análisis. Además, hay que dedicar especial atención a la multicolinealidad, ya que las variables que son multicolineales están implícitamente ponderadas con más fuerza.

Es preciso señalar que la bondad de ajuste de los resultados obtenidos está relacionada directamente con la calidad de los datos considerados. Además, el análisis clúster es muy sensible a la presencia de individuos muy diferentes del resto (atípicos).

2.2.2. Métodos de clasificación

Para distribuir los individuos de nuestra base de datos en diferentes grupos, SPSS nos ofrece tres procedimientos: clúster jerárquico, clúster de K-medias y clúster bietápico. Dependiendo de nuestra base de datos será más recomendable usar un procedimiento u otro. Por ejemplo, si sabemos cuántos clústeres queremos obtener, podríamos usar el clúster de K-medias ya que se debe decidir el número de clústeres a priori. Si por el contrario, no lo sabemos y nuestro tamaño muestral es elevado usaríamos el clúster bietápico y si el tamaño muestral consta de pocos individuos el clúster jerárquico.

2.2.2.1. Clúster jerárquico

Este método se caracteriza porque su proceso para la determinación de los conglomerados es de forma jerárquica. Puede ser aglomerativo o divisivo. En el clúster aglomerativo cada individuo comienza siendo un grupo, sucesivamente se van uniendo los grupos más similares hasta que todos los individuos están en un mismo grupo (o clúster). Por tanto, el analista debe decidir en qué etapa le interesa más detener el proceso. El clúster divisivo funciona al contrario, comienza con un solo grupo formado por todos los individuos y termina siendo cada individuo un grupo. En este tipo de métodos una vez que un individuo ha sido asignado a un determinado clúster permanece inamovible.

Para realizar el análisis debemos tener en cuenta tres cuestiones:

1. Seleccionar el criterio para determinar la similaridad o distancia entre individuos.
2. Seleccionar el criterio para determinar qué clústeres se unirán en las sucesivas etapas.
3. El número de clústeres que necesitamos para representar nuestros datos.

2.2.2.1.1. Medida de asociación

Puede ser en función de la similaridad o de la distancia entre casos. La diferencia entre una y otra radica en que la similaridad mide cómo de semejantes son dos individuos y la distancia el recorrido que hay entre ellos. Si elegimos una distancia como medida de asociación (por ejemplo, la distancia euclídea) los grupos formados contendrán individuos semejantes de forma que la distancia entre ellos ha de ser pequeña. Si elegimos una medida de similaridad (por ejemplo, el coeficiente de correlación) los grupos formados contendrán individuos con una similaridad alta entre sí.

2.2.2.1.2. Métodos de conglomeración

Hay diversos métodos de conglomeración para determinar cómo uniremos los grupos, a continuación mencionamos algunos:

Método del vecino más próximo. Mide la proximidad entre dos grupos calculando la distancia entre sus individuos más próximos o la similitud entre sus individuos más similares.

Método del vecino más alejado. Mide la proximidad entre dos grupos calculando la distancia entre sus individuos más lejanos o la similitud entre sus individuos menos similares.

Método medio entre grupos. Mide la proximidad entre dos grupos como la media de las distancias o similitudes entre individuos de ambos grupos.

Método medio dentro de los grupos. Mide la proximidad entre dos grupos con la distancia media existente entre los miembros del grupo unión de los dos grupos.

Método de la agrupación de centroides. Mide la proximidad entre dos grupos calculando la distancia entre sus medias para todas las variables.

Método de la agrupación de medianas. Mide la proximidad entre dos grupos calculando la distancia entre sus medianas.

Método medio de Ward. Trata de ir agrupando de forma jerárquica elementos de modo que se minimice la distancia euclídea al cuadrado.

2.2.2.1.3. Número de grupos

Para decidir el número de grupos que necesitamos hay varias opciones. Dado que SPSS proporciona las distancias a las que se forma cada clúster, un modo de determinar el número de grupos consiste en localizar en qué iteraciones dichas distancias pegan grandes saltos. Otra opción es calcular la tasa de variación entre los coeficientes de conglomeración obtenidos en las sucesivas etapas, nos detendremos cuando el incremento en la tasa de variación sea claramente elevado a la anterior. En definitiva hay que conseguir un número de clústeres que el analista pueda interpretar con facilidad.

2.2.2.2. Clúster *K*-medias

El análisis de conglomerados de *K* medias se basa en las distancias entre individuos de un conjunto de variables. El analista debe decidir a priori el número de grupos que desea obtener (*K*). Una vez decidido el número de grupos, *K*, se seleccionan los *K* individuos más distantes entre sí. En la siguiente etapa se van asignando cada caso al centro más próximo y actualizando el valor de los centros conforme se incorporan nuevos casos.

Una vez distribuidos todos los individuos en K grupos, se inicia un proceso iterativo para calcular los centroides finales de los K conglomerados.

Una desventaja de este tipo de clúster es la sensibilidad a los atípicos. Posiblemente, si en nuestros datos hay algún atípico será seleccionado como uno de los centros iniciales. En consecuencia ese clúster estará formado por un número reducido de individuos nada representativos. Para evitar esto, antes de realizar el análisis clúster se debe analizar la base de datos en busca de atípicos y eliminarlos del análisis inicial.

2.2.2.3. Clúster bietápico

Cuando tenemos conjunto de datos bastante amplio o necesitamos un procedimiento que pueda formar rápidamente grupos sobre la base de cualquier dato (continuo, categórico), ninguno de los dos procedimientos anteriores es válido. El clúster jerárquico requiere una matriz de distancias entre todos los pares de individuos y el clúster de K medias arrastrar los individuos dentro y fuera de los grupos, además de saber a priori el número de grupos. Para dichas exigencias se diseñó el clúster bietápico. Este clúster no requiere un análisis profundo para su realización (lo cual es muy útil en conjunto de datos muy grandes).

El algoritmo se basa en una medida de distancia que ofrece los mejores resultados, suponiendo todas las variables independientes. Esto raramente ocurre en la práctica, las variables tienden a tener relaciones más o menos fuertes con el resto. Sin embargo, este algoritmo se comporta razonablemente bien aunque tal supuesto no se cumpla. Dado que el análisis no implica la comprobación de hipótesis ni cálculo de los niveles de significación es perfectamente aceptable para agrupar datos cuyo rendimiento no es el esperado. Consta de dos etapas:

1) Preclasificación en grupos pequeños. La primera etapa de este procedimiento es formar conglomerados pequeños. El objetivo es reducir el tamaño de la matriz de distancias entre todos los pares de individuos posibles. Estos “preclústeres” son grupos formados por los individuos de nuestra población que se utilizan en lugar de los datos originales. El algoritmo decide si un individuo debe unirse a un “preclúster” o empezar uno nuevo. Una vez completada esta preclasificación, todos los individuos que componen un “preclúster” son tratados como una entidad. En consecuencia, el tamaño de

la matriz de distancias ya no depende del número de individuos si no del número de “preclústeres”

2) Agrupación jerárquica de “preclústeres”. En esta etapa, SPSS utiliza el algoritmo de agrupamiento jerárquico en los “preclústeres” obtenidos en la etapa anterior. Este algoritmo selecciona el número óptimo de clústeres basándose en el criterio de Schwarz (BIC) o en el criterio de información de Akaike (AIC). El número de grupos óptimo será el que minimice el valor del BIC de ese número de grupos y el de sus adyacentes.

3. ELECCIÓN DEL CONJUNTO DE DATOS

3.1. DETERMINACIÓN DE LA MUESTRA

La población de Zaragoza cuenta con 670.571 ciudadanos de los cuales 346.660 son mujeres y 323.911 hombres (Véase Tabla 1). Dada la imposibilidad de encuestar al total de la población se realizó un muestreo aleatorio. Para que la realización de las encuestas no fuera un trabajo tedioso se fijó un error del 7% y un nivel de confianza del 95% que dio como resultado un tamaño muestral de 196 ciudadanos (Véase Tabla 2). Con el objetivo de realizar una muestra lo más representativa posible de la población el muestreo se realizó en proporción al número de ciudadanos que residían en cada junta municipal de Zaragoza y al sexo.

Tabla 1. Población original estratificada por sexo y junta municipal

Junta municipal	Hombres	Mujeres	Total
Casco Histórico	22.837	24.042	46.879
Centro	23.954	29.457	53.411
Delicias	53.090	57.655	110.745
Universidad	23.272	27.049	50.321
Casablanca	20.808	20.632	41.440
San José	31.844	35.520	67.364
Las Fuentes	21.050	21.794	42.844
La Almozara	11.982	12.746	24.728
Miralbueno	5.879	6.008	11.887
Oliver-Valdefierro	15.762	15.803	31.565
Torrero-La Paz	18.822	19.517	38.339
ACTUR-Rey Fernando	29.216	29.923	59.139
El Rabal	38.772	39.764	78.536
Santa Isabel	6.623	6.750	13.373
TOTAL	323.911	346.660	670.571

Tabla 2. Muestreo estratificado por junta municipal y sexo

Junta municipal	Hombres	Mujeres	Total
Casco Histórico	7	7	14
Centro	7	8	15
Delicias	15	17	32
Universidad	7	8	15
Casablanca	6	6	12
San José	9	10	19
Las Fuentes	6	6	12
La Almozara	4	4	8
Miralbueno	2	2	4
Oliver-Valdefierro	5	5	10
Torrero-La Paz	6	6	12
ACTUR-Rey Fernando	8	9	17
El Rabal	11	11	22
Santa Isabel	2	2	4
TOTAL	95	101	196

3.2. DISEÑO DE LA ENCUESTA

Decidida ya nuestra base de datos, se realizó la encuesta en diferentes lugares de Zaragoza hasta conseguir la muestra deseada (finalmente se obtuvieron 200 encuestas). La encuesta estaba dividida en cuatro apartados, véase anexo I:

Apartado 1. Datos generales. Con el propósito de poder interpretar los resultados obtenidos en función de variables como los ingresos o el nivel estudios se incluyó preguntas acerca de aspectos personales.

Apartado 2. Valores. En él se incluyen preguntas acerca de los principios morales del encuestado.

Apartado 3. Concienciación. En él se incluyen cuestiones sobre el grado de implicación del encuestado en temas medioambientales y además los problemas de la economía española y políticas medioambientales que consideran más importantes.

Apartado 4. Comportamiento de la unidad familiar. Este apartado está a su vez dividido en 5 bloques:

- Reciclaje. Cuestiones sobre la actitud hacia el reciclaje del encuestado.

- Ahorro de energía y agua. Cuestiones acerca del grado de eficiencia del hogar del encuestado.
- Consumo. Cuestiones sobre consumo ecológico.
- Transporte. Cuestiones acerca de los medios de transporte que utiliza el encuestado.

3.3. ANÁLISIS DESCRIPTIVO

Con el objetivo de conocer mejor nuestras variables realizamos un análisis descriptivo univariante de cada una de las variables. Dicho análisis se desarrolla en el anexo II. **Como visión general del análisis unidimensional**, podemos concluir que respecto a los datos generales, los encuestados que predominan son de clase media, de edad entre 40 y 50 años, con estudios universitarios, de grado medio o formación profesional y sus hogares son biparentales o en pareja.

Podemos decir que, en relación a los principios morales, hay un consenso general acerca de que todos debemos cuidar y respetar el medioambiente. Sin embargo, respecto al trato igualitario hacia las personas hay disparidad de opiniones. El liderazgo y el dinero son considerados importantes si bien hay otras cuestiones a las que les otorgan mayor relevancia.

En cuando a la concienciación, la mayoría de los encuestados no colabora con organizaciones o proyectos para el sostenimiento y cuidado del medioambiente. Sin embargo, sí se preocupa de las noticias e informaciones acerca del medioambiente. Además consideran que son el gobierno y las empresas los que más deben contribuir al cuidado y protección del medio ambiente.

Atendiendo a los problemas de la economía española el problema más importante para los encuestados, por rotunda mayoría, es la población desempleada y el menor el desgaste del medio ambiente. En cuanto a la productividad hay divergencia de opiniones y respecto a los problemas del fracaso escolar y sistema de pensiones hay dos tipos de encuestados los que otorgan una importancia elevada y los que les otorgan una importancia residual.

Respecto a las políticas medioambientales las más importantes son que las empresas tengan plantas de tratamiento de aguas, que prohíban la emisión de humos y gases en las empresas y su regulación para los vehículos.

Por último, en relación al comportamiento de la unidad familiar podemos deducir que los encuestados reciclan los residuos y medicamentos, aunque piensan que el espacio y la distancia a los contenedores dificultan el reciclaje, no piensan que los hogares desconozcan como se debe reciclar, aprovechan al máximo la luz solar y son eficientes energéticamente. Por otro lado, respecto al transporte, los encuestados no utilizan la bicicleta, conducen eficientemente y acerca del transporte público hay disparidad de opiniones.

Por otro lado, **se realizó un análisis bidimensional** de algunas variables respecto de la variable sexo, para poder observar si existen diferencias en el comportamiento entre hombres y mujeres. En consecuencia, podemos concluir que las mujeres piensan que el trato debe ser igualitario en mayor medida que los hombres, además colaboran más con las organizaciones medioambientales, reciclan más y usan con más frecuencia el transporte urbano.

4. RESULTADOS DEL ANÁLISIS FACTORIAL

Con ayuda del análisis factorial, determinaremos las bases que nos permitirán describir e interpretar la actitud y compromiso hacia el medio ambiente de los encuestados.

Dado nuestro número elevado de variables iniciales (33), realizamos varios análisis factoriales intermedios en función de los distintos apartados en los que se divide la encuesta. Finalmente obtenemos, **en esta primera fase**, 11 factores. Decidimos no incorporar en dicho análisis algunas variables que o bien no discriminan o bien las utilizaremos en una etapa posterior para correlacionarlas con cada factor. A continuación se detallan los 11 factores obtenidos:

1. Factor igualdad. Este factor está relacionado con las actitudes del encuestado hacia el trato igualitario de personas, la importancia del liderazgo y del dinero.
2. Factor concienciación ecológica. Este factor está relacionado con la preocupación por el medioambiente y el grado de colaboración del encuestado.

3. Factor contribución público vs privado. Este factor está relacionado con la opinión del encuestado hacia qué sector debería contribuir más al medioambiente.
4. Factor reciclaje. Este factor está relacionado con la acción de reciclar del encuestado.
5. Factor barrera al reciclaje. Este factor está relacionado con la actitud del encuestado a los posibles obstáculos que surgen del reciclaje.
6. Factor eficiencia energética. Este factor está relacionado con el comportamiento del encuestado respecto del ahorro energético.
7. Factor actitud saludable. Este factor está relacionado con el comportamiento del encuestado respecto de hábitos considerados saludables.
8. Factor consumo ecológico. Este factor está relacionado con el comportamiento del encuestado respecto del consumo de productos con trascendencia ecológica.
9. Factor transporte público vs privado. Este factor está relacionado con el comportamiento del encuestado respecto del transporte urbano frente al propio.
10. Factor bicicleta. Este factor está relacionado con el comportamiento del encuestado respecto de la bicicleta.
11. Factor conducción eficiente. Este factor está relacionado con el comportamiento del encuestado respecto de la conducción del vehículo propio.

En el anexo III se describen los diferentes pasos realizados para la obtención de nuestros 11 factores.

En la segunda fase realizamos un análisis factorial con los 11 factores obtenidos en el análisis previo. Este análisis tiene el objetivo de reducir la dimensión en mayor medida y así conseguir los indicadores últimos que nos permitan interpretar y caracterizar el comportamiento de los distintos encuestados.

A continuación se muestran los resultados que ofrece SPSS al analizar la matriz de correlaciones, es decir, los coeficientes de correlación de Pearson entre cada par de variables.

Matriz de correlaciones

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10	Factor 11
Factor 1	1,000	,000	,044	-,247	-,073	-,117	,079	-,052	-,355	,031	,169
Factor 2	,000	1,000	,318	,416	-,239	,063	,074	,402	-,049	,363	,091
Factor 3	,044	,318	1,000	,300	-,227	-,015	,112	,099	-,055	,141	,058
Factor 4	-,247	,416	,300	1,000	,000	,092	,116	,313	,111	,101	,185
Factor 5	-,073	-,239	-,227	,000	1,000	,082	-,205	-,064	,102	-,095	-,100
Factor 6	-,117	,063	-,015	,092	,082	1,000	,000	,000	,260	,126	,016
Factor 7	,079	,074	,112	,116	-,205	,000	1,000	,000	,123	-,081	,219
Factor 8	-,052	,402	,099	,313	-,064	,000	,000	1,000	-,109	,243	,118
Factor 9	-,355	-,049	-,055	,111	,102	,260	,123	-,109	1,000	,000	,000
Factor 10	,031	,363	,141	,101	-,095	,126	-,081	,243	,000	1,000	,000
Factor 11	,169	,091	,058	,185	-,100	,016	,219	,118	,000	,000	1,000

Tabla 3. Matriz de correlaciones**Prueba de KMO y Bartlett**

Medida Kaiser-Meyer-Olkin de adecuación de muestreo		,608
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	277,767
	gl	55
	Sig.	,000

Tabla 4. Prueba de KMO y Bartlett**Matrices anti-imagen**

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10	Factor 11
Factor 1	,477 ^a	-,067	-,078	,274	-,044	,022	-,117	,068	,338	-,058	-,215
Factor 2	-,067	,678 ^a	-,152	-,306	,197	-,042	-,017	-,260	,024	-,274	,035
Factor 3	-,078	-,152	,688 ^a	-,245	,171	,026	-,046	,075	,037	-,057	,041
Factor 4	,274	-,306	-,245	,600 ^a	-,155	-,036	-,081	-,173	-,038	,067	-,191
Factor 5	-,044	,197	,171	-,155	,568 ^a	-,068	,193	-,018	-,082	,034	,072
Factor 6	,022	-,042	,026	-,036	-,068	,616 ^a	,013	,031	-,221	-,116	-,023
Factor 7	-,117	-,017	-,046	-,081	,193	,013	,536 ^a	,017	-,172	,117	-,159
Factor 8	,068	-,260	,075	-,173	-,018	,031	,017	,701 ^a	,137	-,139	-,083
Factor 9	,338	,024	,037	-,038	-,082	-,221	-,172	,137	,524 ^a	-,051	-,042
Factor 10	-,058	-,274	-,057	,067	,034	-,116	,117	-,139	-,051	,649 ^a	,026
Factor 11	-,215	,035	,041	-,191	,072	-,023	-,159	-,083	-,042	,026	,537 ^a

a. Medidas de adecuación de muestreo (MSA)

Tabla 5. Matriz anti-imagen

Atendiendo a la medida de adecuación muestral de Kaiser-Meter-Olkin (KMO) (Tabla 4) se observa que el valor del KMO = 0,608 por lo que al encontrarse entre 0,5 y 0,75 podemos concluir que la idea de realizar un análisis factorial es buena. Además, el test de esfericidad de Bartlett contrasta la hipótesis nula de que la matriz de correlaciones

observada es en realidad una matriz identidad, y como observamos en nuestra tabla se rechaza la hipótesis nula, lo que nos indica que existen relaciones significativas entre las variables.

Observando la matriz de correlación anti-imagen (Tabla 5) vemos que los coeficientes de correlación parcial son valores muy próximos a cero. Esto significa que las variables incluidas en el análisis comparten gran cantidad de información debido a la presencia de factores comunes, por tanto, es un buen síntoma de cara a la idoneidad del análisis. Además si nos fijamos en la diagonal de la matriz vemos que los valores son altos por lo que es otro indicativo de la correcta adecuación muestral.

Como consecuencia de lo anterior podemos concluir que el Análisis factorial es una técnica adecuada para analizar nuestros datos. De entre los diversos métodos de extracción que ofrece SPSS se optó por componentes principales.

Comunalidades

	Inicial	Extracción
Factor 1	1,000	,730
Factor 2	1,000	,670
Factor 3	1,000	,520
Factor 4	1,000	,724
Factor 5	1,000	,634
Factor 6	1,000	,689
Factor 7	1,000	,633
Factor 8	1,000	,614
Factor 9	1,000	,685
Factor 10	1,000	,684
Factor 11	1,000	,734

Método de extracción: análisis de componentes principales.

Tabla 6. Comunalidades

Observando las comunalidades (Tabla 6) vemos que son altas. Esto quiere decir que la variable que está muy relacionada con las restantes variables del análisis, tenderá a compartir su información con un factor común.

A continuación vemos como los tres criterios utilizados de determinación de factores corroboran la misma solución: cinco factores.

Varianza total explicada

Componente	Autovalores iniciales			Sumas de extracción de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	2,266	20,605	20,605	2,266	20,605	20,605
2	1,646	14,961	35,565	1,646	14,961	35,565
3	1,334	12,125	47,690	1,334	12,125	47,690
4	1,040	9,451	57,141	1,040	9,451	57,141
5	1,032	9,379	66,520	1,032	9,379	66,520
6	,839	7,631	74,152			
7	,700	6,363	80,515			
8	,693	6,299	86,813			
9	,535	4,863	91,676			
10	,515	4,679	96,355			
11	,401	3,645	100,000			

Método de extracción: análisis de componentes principales.

Tabla 7. Varianza total explicada

Gráfico de sedimentación

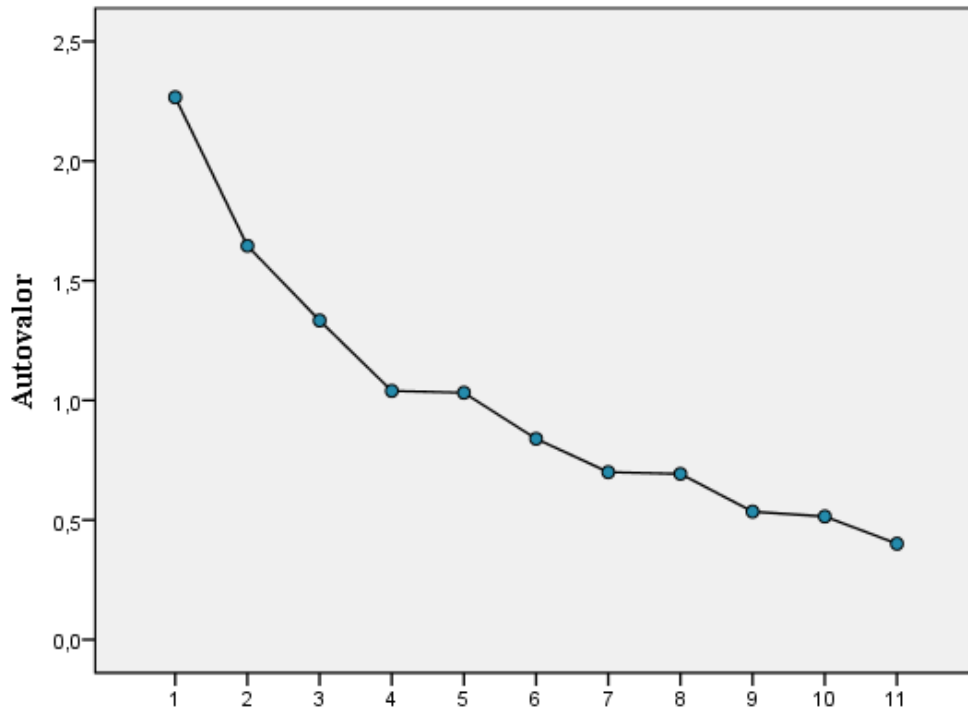


Figura 1. Gráfico de sedimentación

Como podemos observar, la tabla de la varianza total explicada (Tabla 7) nos ofrece cinco factores. El primer factor explica un 21%, el segundo un 15%, el tercero un 12%, el cuarto un 10% y el quinto factor un 9% por lo que explicaríamos en total un 67%.

El gráfico de sedimentación (Figura 1) nos muestra los valores propios asociados a cada factor, permitiéndonos prescindir de aquellos que no son capaces de explicar una cantidad relevante de la varianza total. Si utilizamos la regla de Kaiser, vemos que los factores con un valor propio mayor que la unidad son los primeros cinco factores. En consecuencia podemos concluir que el resto de factores tienen la consideración de residuales, y por tanto, los factores que tendremos en cuenta serán los cinco primeros.

Matriz de componente

	Componente				
	1	2	3	4	5
Factor 1		-,731		,414	
Factor 2	,796				
Factor 3	,555				
Factor 4	,661				
Factor 5					-,434
Factor 6		,520		,475	,421
Factor 7			,750		
Factor 8	,602				
Factor 9		,719			
Factor 10	,481		-,429		,432
Factor 11			,498	,516	

Método de extracción: análisis de componentes principales.

Tabla 8. Matriz de componente

Matriz de componente rotado

	Componente				
	1	2	3	4	5
Factor 1		-,797			
Factor 2	,664				
Factor 3			,660		
Factor 4	,703				
Factor 5			-,774		
Factor 6					,763
Factor 7				,680	
Factor 8	,774				
Factor 9		,719			
Factor 10					,643
Factor 11				,791	

Método de rotación: Varimax con normalización Kaiser.

Tabla 9. Matriz de componente rotado

Como vemos en la matriz de componentes no rotados (Tabla 8) los factores no cumplen el principio de estructura simple, es decir, hay variables que están saturadas de más de un factor. Para intentar solucionarlo, aplicamos la técnica de rotación factorial Varimax. En consecuencia, vemos que en la matriz ya rotada (Tabla 9) todas las variables están saturadas de un solo factor facilitando su interpretación.

- El factor 1 podemos denominarlo como “*Compromiso ecológico*” ya que se relaciona fuertemente con los factores 2, 4 y 8 que recogen aspectos vinculados con una actitud y comportamientos beneficiosos para el medio ambiente.
- El factor 2 puede denominarse como “*Actitud respetuosa*” ya que se relaciona fuertemente con los factores 1 y 9 que recogen aspectos vinculados con el correcto comportamiento de un ciudadano.
- El factor 3 podemos denominarlo como “*Ideología ecológica*” ya que se relaciona fuertemente con los factores 3 y 5 que recogen aspectos vinculados con la opinión de que el sector privado es quién debe contribuir más al cuidado del medio ambiente y por la convicción de que no hay impedimentos para reciclar.
- El factor 4 puede denominarse como “*Comportamiento saludable*” ya que se relaciona fuertemente con los factores 7 y 11 que recogen aspectos vinculados con la actitud saludable y la conducción eficiente.
- El factor 5 puede denominarse como “*Posición ahorradora*” ya que se relaciona fuertemente con los factores 6 y 10 que recogen aspectos vinculados con el ahorro energético y el ahorro de combustible al usarse la bicicleta.

Una vez reducida nuestra dimensión a cinco factores, en el apartado siguiente, los utilizamos para construir grupos de encuestados lo más homogéneos posibles, en función de dichos factores, mediante la técnica del análisis clúster.

5. RESULTADOS ANÁLISIS CLÚSTER

Como se ha indicado anteriormente en el apartado 2.2.2., el procedimiento para clasificar los individuos de la población depende de nuestro conjunto de datos. Puesto que no sabíamos a priori el número de grupos que deseábamos obtener, y nuestra base de datos era razonablemente amplia, optamos por el clúster bietápico. Para aplicar este

método es recomendable que las variables sean independientes, por lo tanto, al emplear los factores de la sección anterior nos aseguramos que son incorreladas.

Para hacer la clasificación de nuestros individuos introducimos los cinco factores obtenidos del análisis factorial (apartado 2.1.7.): “*Compromiso ecológico*”, “*actitud respetuosa*”, “*ideología ecológica*”, “*comportamiento saludable*” y “*posición ahorradora*”.

A continuación se muestra en la Figura 2 el resumen del modelo y el gráfico de calidad de los clústeres. Como se puede observar en la tabla del resumen del modelo se han obtenido 3 clústeres basados en nuestros 5 factores. Además, el gráfico que le sigue, establece si la calidad de los clústeres obtenidos es insuficiente, en cuyo caso habría que plantearse si es adecuado seguir con el análisis. Puesto que en este caso la medida de silueta de la cohesión y separación es suficiente proseguimos con el análisis.

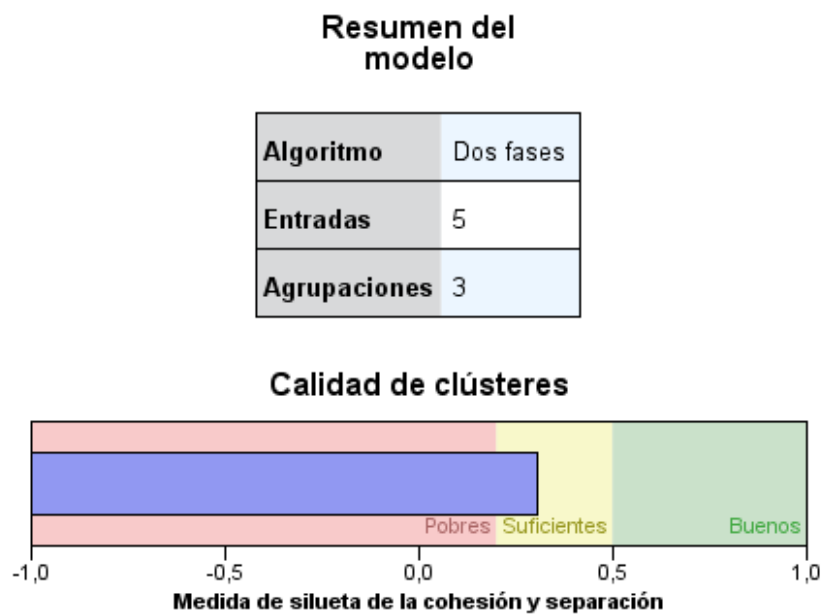


Figura 2. Resumen del modelo bietápico

En la Figura siguiente (3) se muestra un diagrama de sectores con el tamaño de cada clúster. El primer clúster contiene el 19% (38), el segundo contiene el 37% (74) y el tercero el 44% de los encuestados (88).

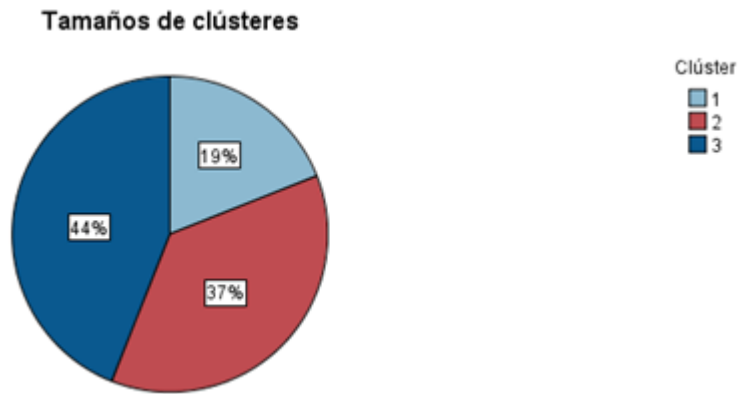


Figura 3. Diagrama de sectores en función del tamaño de los clústeres

En la Figura 4 se recoge el grado de importancia que ha tenido cada factor en la asignación de los individuos a los diferentes clústeres. Como podemos observar el factor que más influye en la distribución de los encuestados es “*Ideología ecológica*”, le siguen los factores “*Comportamiento saludable*” y “*Compromiso ecológico*”. Los factores con menor peso son “*Posición ahorradora*” y “*Actitud respetuosa*”.

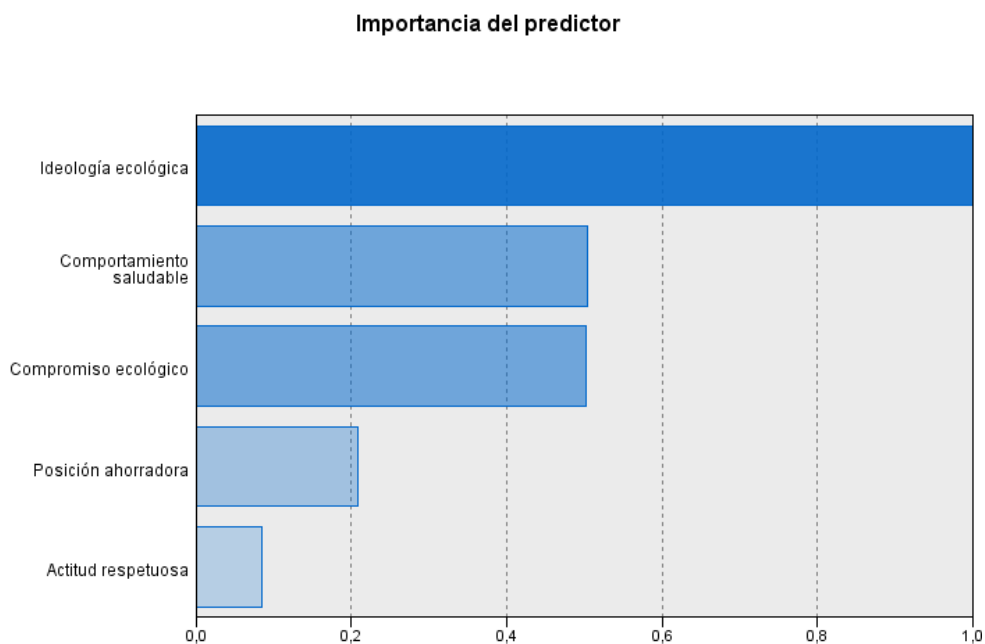


Figura 4. Importancia del predictor

La Figura 5 muestra la comparación entre cada clúster respecto a los diferentes factores. Además queda corroborado por las Figuras 6 a 11 que representan como se distribuye cada factor entre los clústeres. En consecuencia podemos advertir que en el **primer clúster** se hallan ciudadanos que valoran alto los factores “*posición ahorradora*” y “*actitud respetuosa*”, bajo los factores “*comportamiento saludable*” y “*compromiso ecológico*” e intermedio el factor “*ideología ecológica*”. Mientras que en el **segundo clúster** se encuentran los ciudadanos que valoran alto los factores “*ideología ecológica*” y “*comportamiento saludable*”, bajo los factores “*posición ahorradora*” y “*actitud respetuosa*” e intermedio el factor “*compromiso ecológico*”. Por último, el **tercer clúster** está formado por ciudadanos que valoran alto el factor “*compromiso ecológico*”, bajo el factor “*ideología ecológica*” e intermedio los factores “*comportamiento saludable*”, “*posición ahorradora*” y “*actitud respetuosa*”.

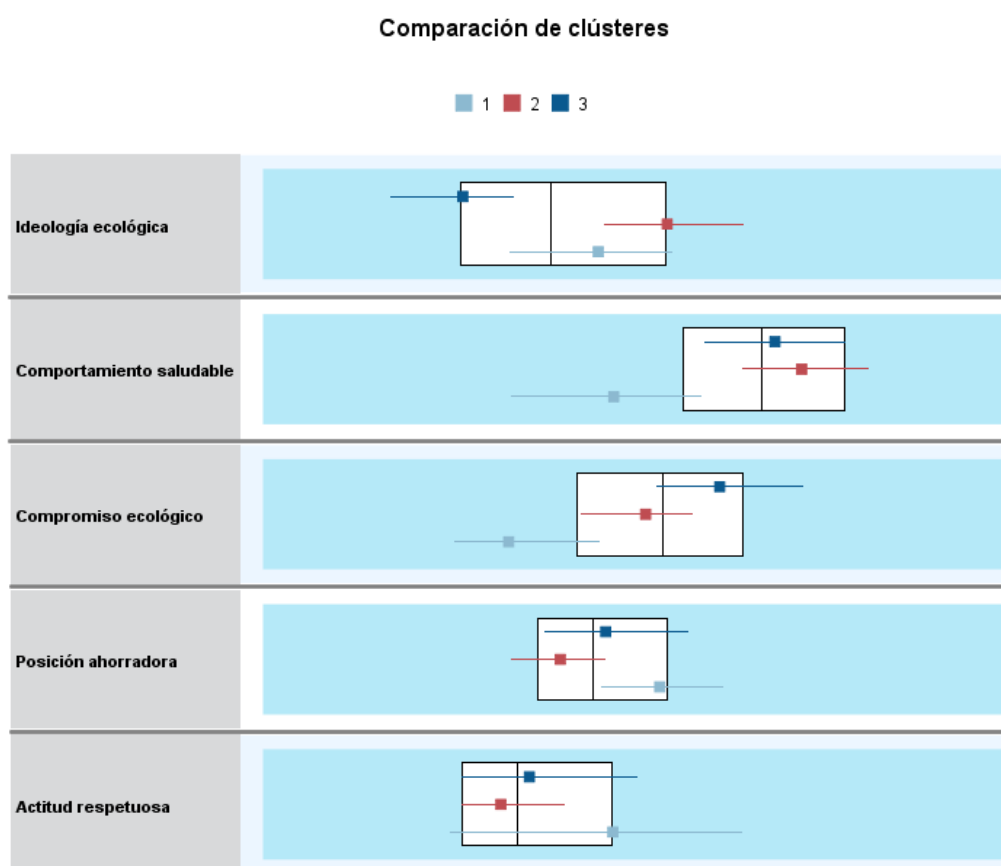


Figura 5. Comparación de clústeres respecto de cada factor

A continuación se muestran los gráficos de barras de error de cada factor. En ellos se puede apreciar cómo se distribuyen de forma diferente entre los tres clústeres.

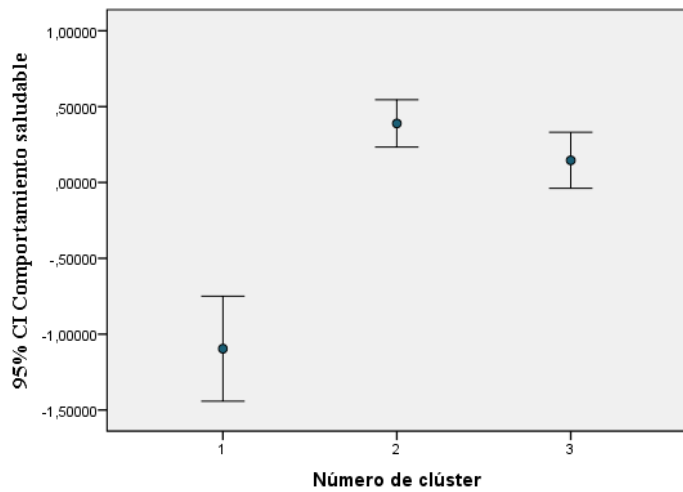


Figura 6. Barras de error del factor “comportamiento saludable” por clúster

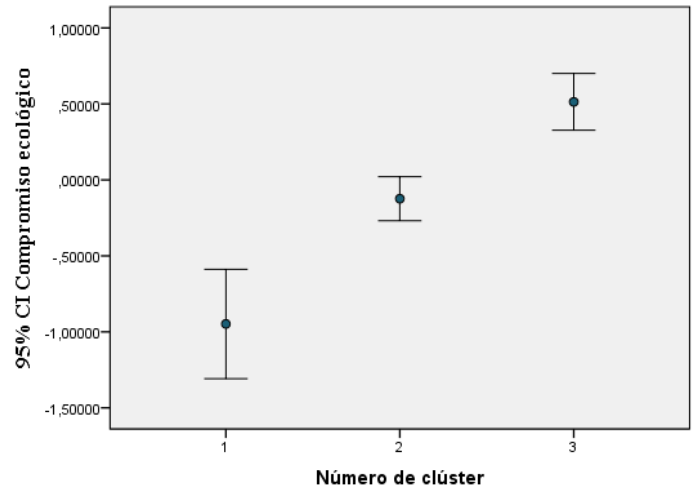


Figura 7. Barras de error del factor “compromiso ecológico” por clúster

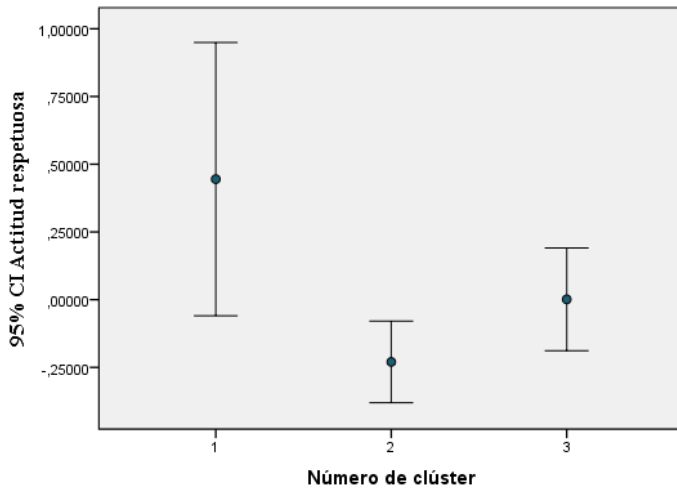


Figura 8. Barras de error del factor “actitud respetuosa” por clúster

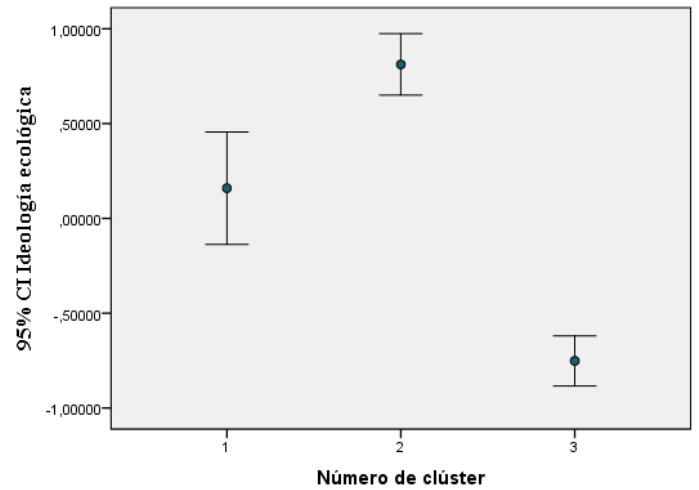


Figura 9. Barras de error del factor “ideología ecológica” por clúster

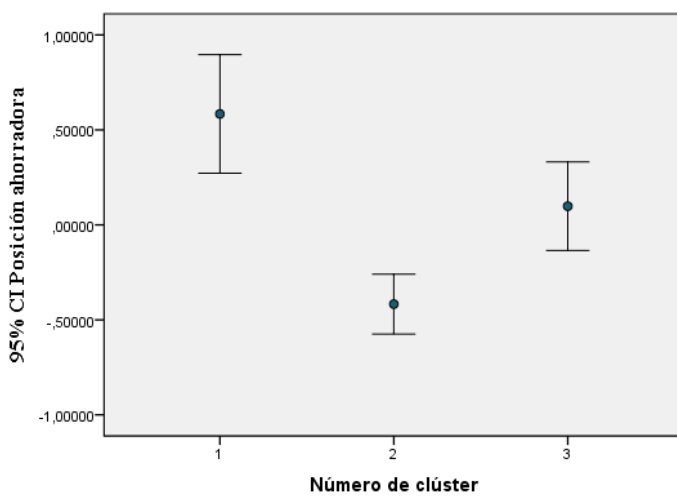


Figura 10. Barras de error del factor “posición ahorradora” por clúster

Con la finalidad de posteriormente sacar las conclusiones pertinentes, se cruzan las variables cualitativas que no incluimos en el análisis factorial y que en consecuencia no están incluidas en la clasificación de los clústeres. (Género del encuestado, edad, tipo y nº de miembros del hogar, nivel de estudios, barrio en el que se encuentra el hogar, ingresos mensuales).

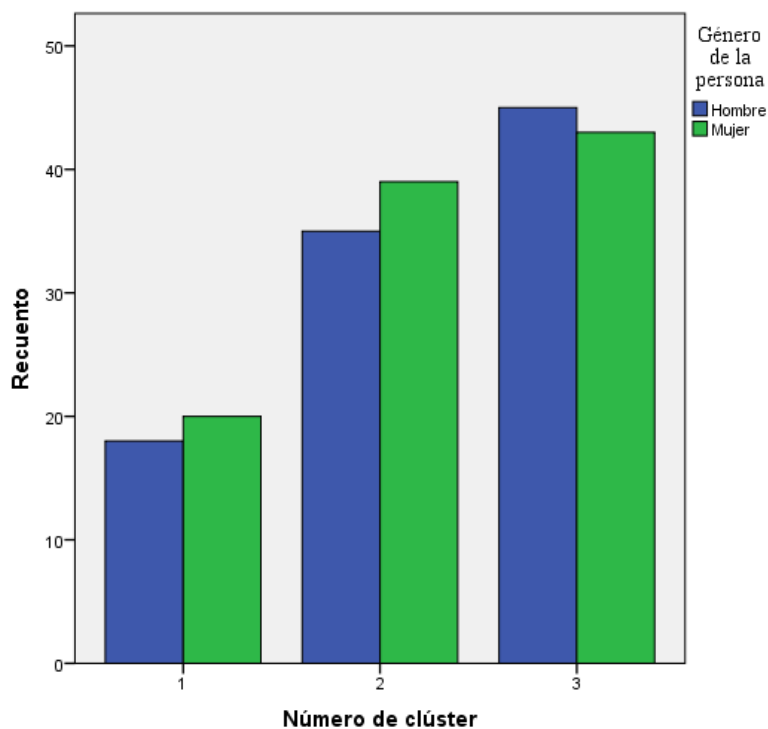


Figura 11. Diagrama de barras de la variable género por clúster

Observando la Figura 11 vemos que en los grupos no se encuentran diferencias por género, siendo prácticamente iguales el peso de mujeres y hombres en cada clúster. Sin embargo, si nos fijamos en la Figura 12 sí que apreciamos diferencias. Vemos que en el primer clúster se encuentran ciudadanos con un rango de edad amplio, no obstante, en el segundo clúster observamos como se compone de encuestados de menor edad y en el tercero de encuestados de edad más avanzada.

En la Figura 13 no se ven diferencias aparentes, los tres clústeres tienen semejante composición, en los tres predomina el hogar biparental y en pareja. En la Figura 14, acerca del número de miembros que compone el hogar tampoco se encuentran diferencias importantes.

Por el contrario, en la Figura 15 sobre el nivel de estudios sí se ven diferencias, si comparamos los 3 clústeres en proporción a su tamaño, en el primer clúster se encuentran más ciudadanos con un nivel de estudios bajo (primarios, ESO) que en los

otros dos clústeres, y menos ciudadanos con estudios universitarios que en el resto de clústeres. Comparando el segundo clúster con el tercero observamos que aún siendo de estructura similar en el segundo hay más universitarios y menos ciudadanos con estudios medios que en el tercero.

En la Figura 16 en relación a los ingresos mensuales también encontramos diferencias importantes comparando los clústeres en proporción a su tamaño. Observamos que en el primer clúster se hallan más encuestados con menor capacidad adquisitiva que en el resto de clústeres, y menos gente con una capacidad adquisitiva alta que en los otros dos clústeres. Si comparamos el segundo clúster con el tercero vemos que en el tercero se encuentran más ciudadanos con niveles de renta elevados que en el segundo.

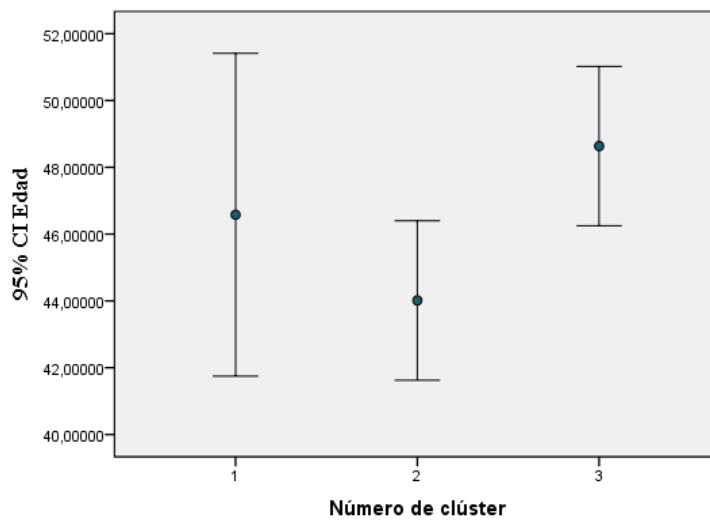


Figura 12. Barras de error de la variable edad por clúster

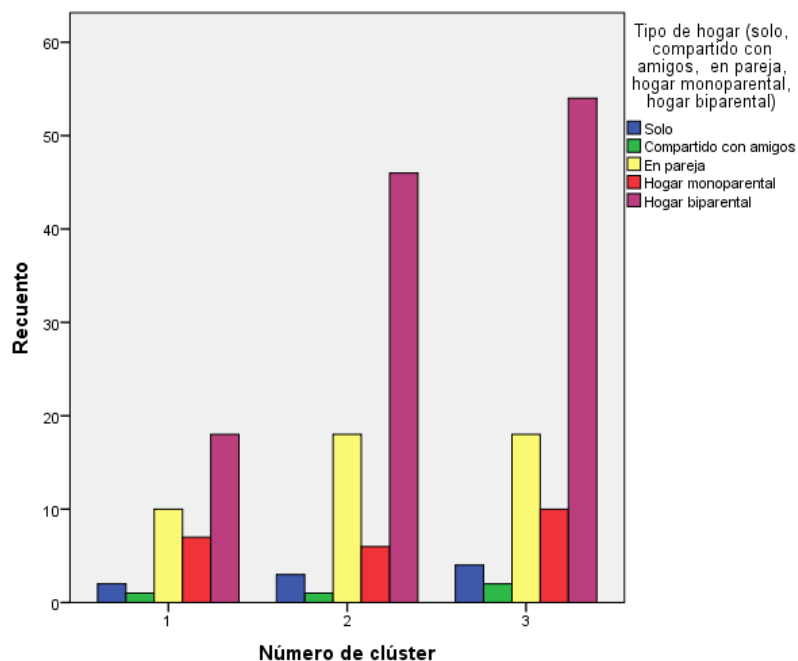


Figura 13. Diagrama de barras de la variable tipo de hogar por clúster

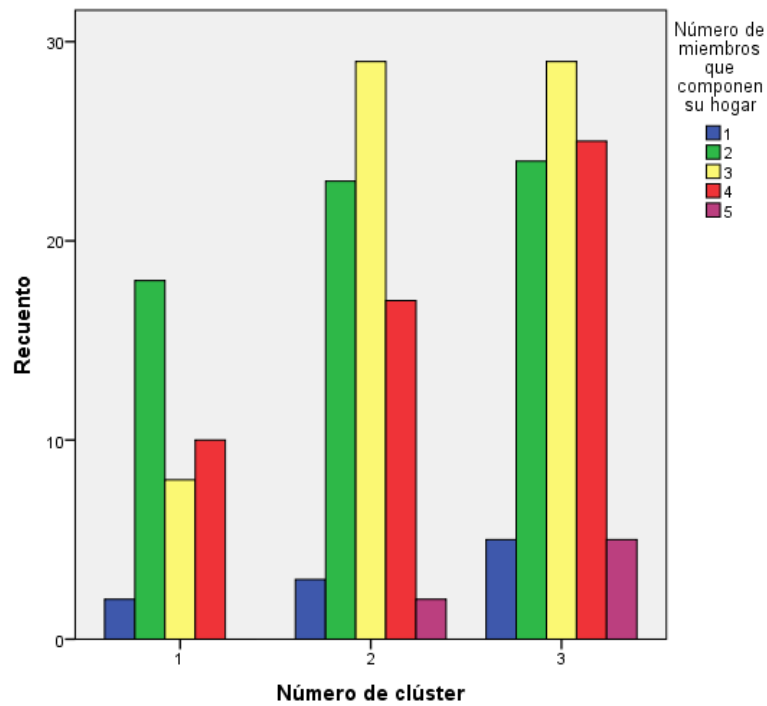


Figura 14. Diagrama de barras de la variable nº de miembros del hogar por clúster

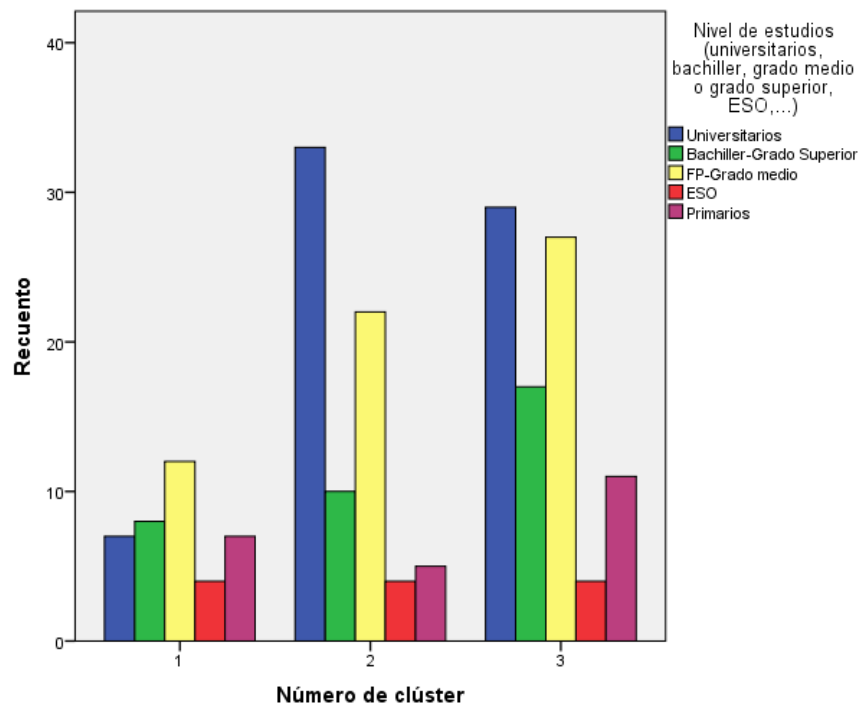


Figura 15. Diagrama de barras de la variable nivel de estudios por clúster

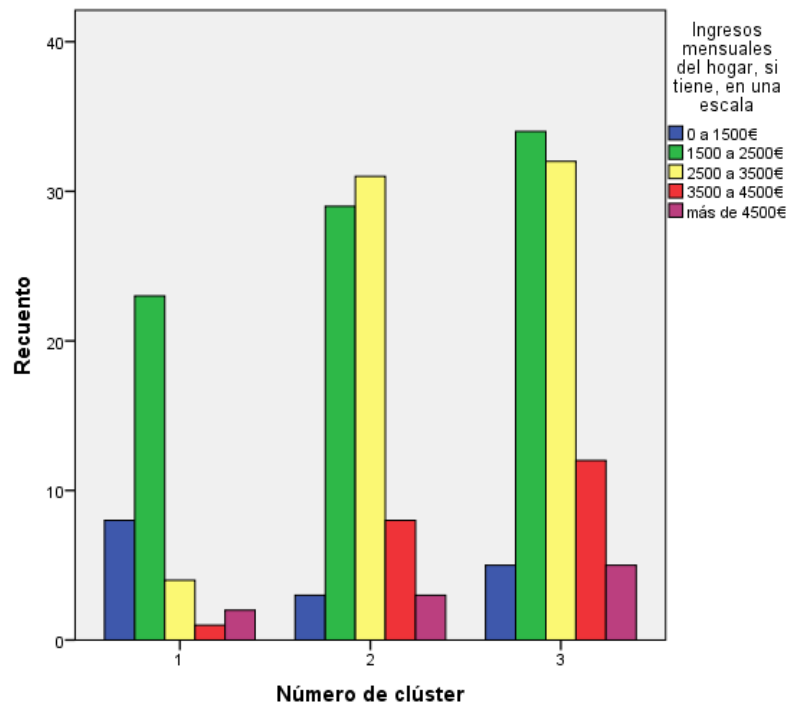


Figura 16. Diagrama de barras de la variable ingresos mensuales por clúster

Además de estas variables cualitativas, no incluimos en el análisis factorial las variables 15 y 16 (problemas de la economía española y políticas para el cuidado del medio ambiente) para poder determinar qué problemas y qué políticas se consideraban más importantes por los encuestados, según el clúster al que pertenecen.

Para la variable referente a los problemas de la economía española, vemos que el problema de la **población desempleada** es el más importante para todos los encuestados independientemente del clúster al que pertenezcan (Figura 19). En la Figura 17 se observa que los encuestados de los clústeres 2 y 3 valoran más importante el problema del **desgaste del medioambiente** que los individuos del clúster 1. Fijándonos en la Figura 18 vemos que los individuos de los clústeres 2 y 3 otorgan una mayor importancia al problema de la **productividad**. Observamos que en la Figura 20 los individuos del clúster 1 valoran como más importante el problema del **fracaso escolar** que los otros dos clústeres. En la Figura 21 se ve como los individuos del clúster 1 otorgan mayor importancia al problema del **sistema de pensiones** que los otros dos clústeres.

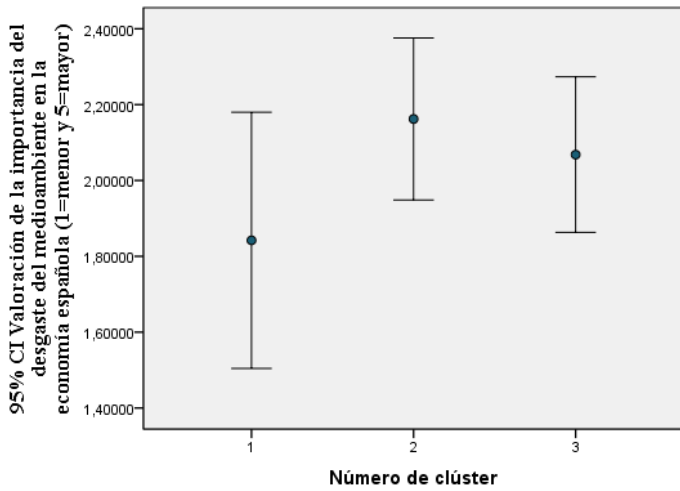


Figura 17. Barras de error del problema del desgaste del medio ambiente por clúster

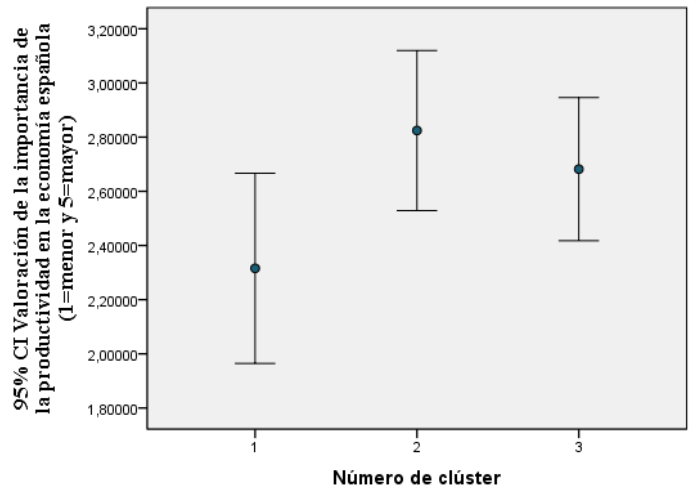


Figura 18. Barras de error del problema de la productividad por clúster

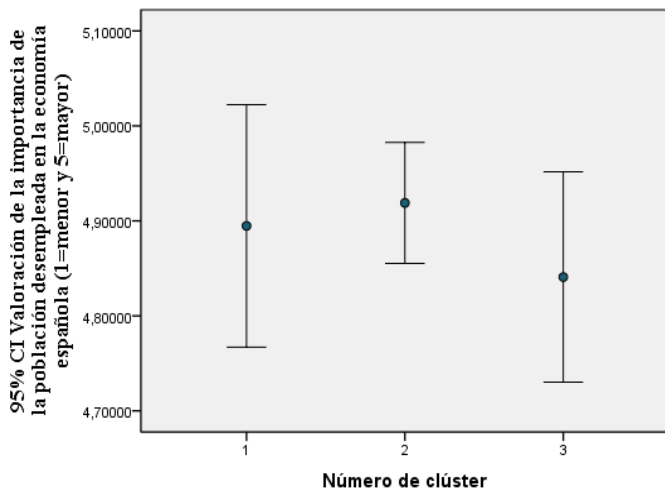


Figura 19. Barras de error del problema de la población desempleada por clúster

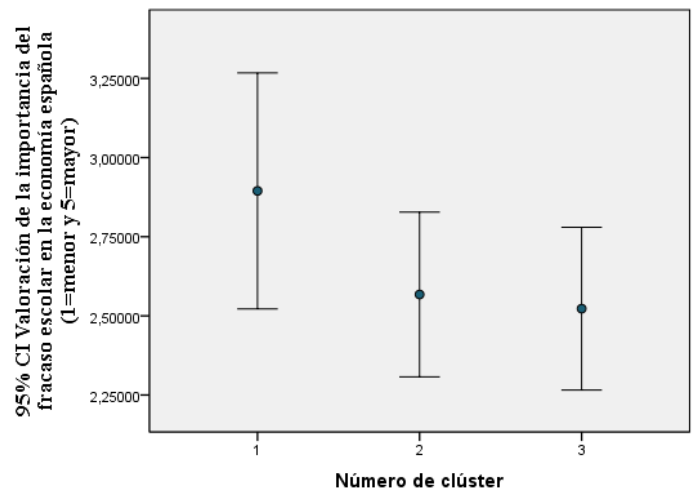


Figura 20. Barras de error del problema del fracaso escolar por clúster

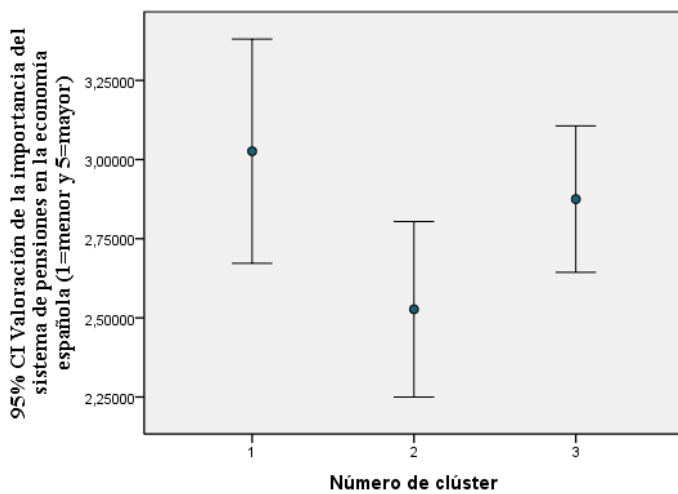


Figura 21. Barras de error del problema del sistema de pensiones por clúster

Para la variable relativa a las políticas medioambientales apenas observamos diferencias entre clústeres.

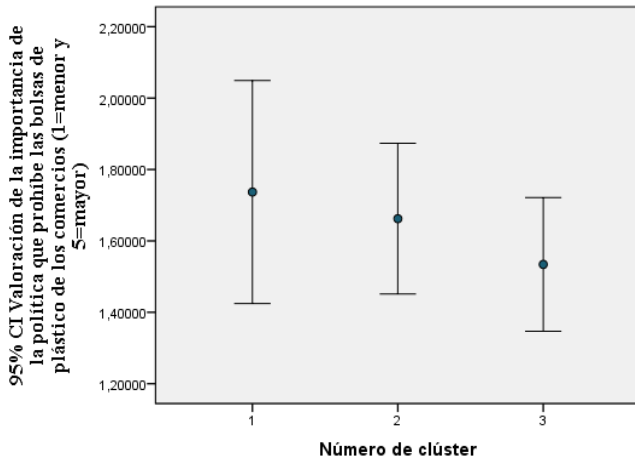


Figura 22. Barras de error de la política “prohibición de las bolsas de plástico en los comercios” por clúster

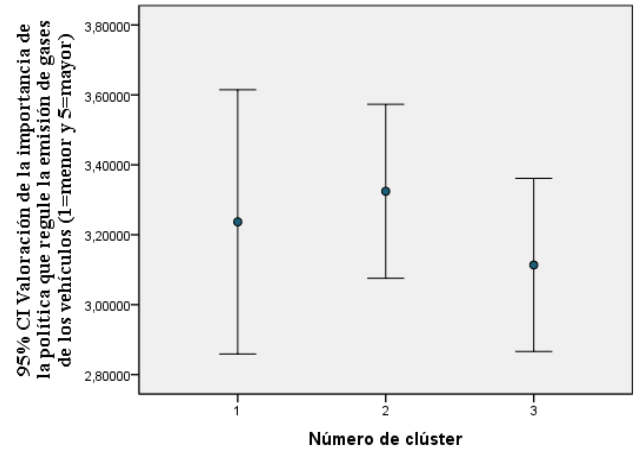


Figura 23. Barras de error de la política “regulación de la emisión de gases de los vehículos” por clúster

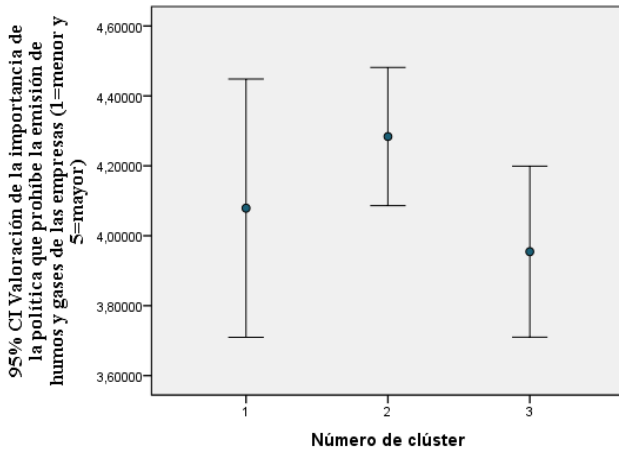


Figura 24. Barras de error de la política “prohibición a la empresas de emitir humos y gases” por clúster

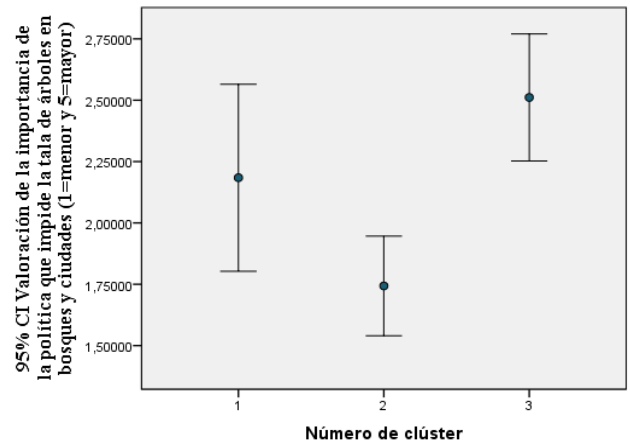


Figura 25. Barras de error de la política “prohibición de la tala de árboles en bosques y ciudades” por clúster

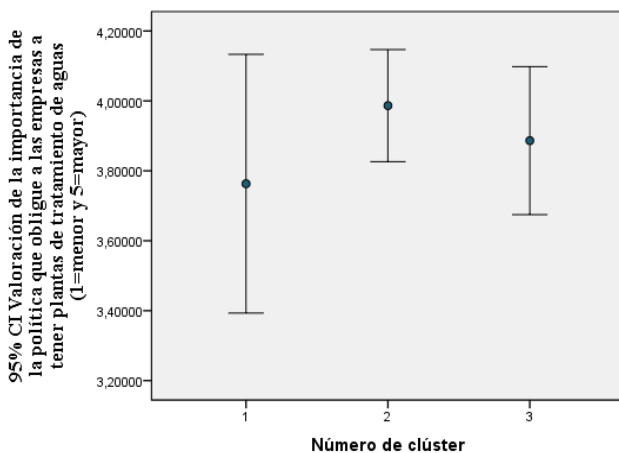


Figura 26. Barras de error de la política “obligación a las empresas a tener plantas de tratamientos de aguas” por clúster

En consecuencia, la visión general de cada clúster es la siguiente:

Primer clúster. Es el clúster de menor tamaño, en cuanto al género apenas vemos diferencias, 20 mujeres frente a 18 hombres. En él se encuentran ciudadanos con un rango de edad amplio, además, es el que posee mayor porcentaje de ciudadanos con estudios bajos y renta mensual baja. Los hogares predominantes son el biparental y en pareja compuestos por 2, 3 o 4 miembros. El problema al que mayor importancia otorgan es la población desempleada, le sigue el fracaso escolar y el sistema de pensiones. En relación a las políticas medioambientales la más valorada es la política de prohibir la emisión de humos y gases de las empresas, seguida de la de obligar a las empresas a tener plantas de tratamiento de aguas y regular la emisión de gases de los vehículos.

Segundo clúster. En este clúster tampoco se aprecian diferencias en cuanto al género, 35 hombres frente a 39 mujeres. En él predominan ciudadanos con un rango de edad que oscila entre 42 a 45 años, además, es el que posee mayor porcentaje de universitarios y la renta mensual predominantes es de 1500 a 3500€. Los hogares predominantes son el biparental y en pareja compuestos por 3, 2 o 4 miembros. El problema al que mayor importancia otorgan es la población desempleada, le sigue la productividad y el fracaso escolar. En cuanto a las políticas medioambientales la más valorada es la política de prohibir la emisión de humos y gases de las empresas, seguida de la de obligar a las empresas a tener plantas de tratamiento de aguas y regular la emisión de gases de los vehículos.

Tercer clúster. Es el clúster de mayor tamaño, en relación al género no apreciamos diferencias significativas, 45 hombres frente a 43 mujeres. En él predominan los ciudadanos con un rango de edad que oscila entre 47 a 52 años, además, la mayor parte de los ciudadanos tienen estudios medios y es el que posee mayor porcentaje de rentas elevadas (más de 3500€). Los hogares predominantes son el biparental y en pareja compuestos por 3, 4 o 2 miembros. El problema al que mayor importancia otorga es la población desempleada, le sigue el sistema de pensiones y la productividad. En cuanto a las políticas medioambientales la más valorada es la política de prohibir la emisión de humos y gases de las empresas, seguida de la de obligar a las empresas a tener plantas de tratamiento de aguas y regular la emisión de gases de los vehículos.

6. CONCLUSIONES

Mediante el análisis factorial se ha logrado crear 5 indicadores que describen las diferentes pautas y comportamientos de una muestra de 200 ciudadanos de Zaragoza estratificada por su sexo y por la junta municipal a la que pertenecen.

Los cinco indicadores creados se definen de la siguiente manera, un primer indicador denominado **“Compromiso ecológico”** que incluye cuestiones acerca de la concienciación ecológica, reciclaje y consumo sostenible. El segundo indicador llamado **“Actitud respetuosa”** recoge el comportamiento del ciudadano en relación a otros y a su entorno. El tercer indicador se denomina **“Ideología ecológica”** y está relacionado con cuestiones sobre la opinión del encuestado acerca de quién debería contribuir más al cuidado del medioambiente y la existencia o no de barreras al reciclaje. El cuarto indicador denominado **“Comportamiento saludable”** incluye hábitos sobre una vida sana como es la alimentación ecológica, conducción eficiente, etc. Por último, el quinto indicador llamado **“Posición ahorradora”** recoge cuestiones de ahorro en el hogar, como es el uso eficiente de la energía y el uso de la bicicleta.

A continuación con los indicadores mencionados se procedió a realizar una análisis clúster bietápico y se obtuvieron 3 clústeres:

- Primer clúster. Este grupo es el de menor tamaño (38) y está compuesto por ciudadanos que tienen una posición ahorradora y una actitud respetuosa. En cambio, su comportamiento saludable y compromiso ecológico es deficiente. Por último, respecto de la ideología ecológica no se posicionan acerca de quién debería contribuir más al medioambiente.
- Segundo clúster. Este grupo consta de 74 ciudadanos cuya ideología ecológica se basa en la opinión de que debe ser el sector privado quién contribuya más al medioambiente, además, tienen un comportamiento saludable. Sin embargo, no tienen una actitud respetuosa ni se preocupan del ahorro energético. Por último, no están muy comprometidos con el medioambiente si bien más que los del primer clúster.
- Tercer clúster. Este grupo es el de mayor tamaño (88) y está formado por ciudadanos comprometidos con el medioambiente. No obstante, respecto a su

ideología ecológica piensan que quién debe contribuir más al medio ambiente es el gobierno y consideran que existen barreras al reciclaje. Por último, respecto a su comportamiento saludable, posición ahorradora y actitud respetuosa son favorables aunque no destacan.

Además, **comparando con diferentes variables cualitativas** podemos sacar las siguientes conclusiones:

De acuerdo a los datos generales, en el primer clúster nos encontramos con ciudadanos de un rango amplio de edades, en cambio, en el segundo y tercer clúster se pueden apreciar algunas diferencias, si bien no muy significativas. En el segundo se encuentran ciudadanos con una media de edad menor que en el tercero.

Teniendo en cuenta que Zaragoza es una ciudad de un país desarrollado es lógico pensar que la mayoría de los encuestados tengan una clase social media y unos estudios mínimos. Siguiendo este razonamiento, podemos deducir algunas cuestiones. En cuanto al nivel de estudios en los tres grupos el más numeroso son los estudios medios (Bachiller-Grado Superior y Formación profesional-Grado medio), no obstante, podemos apreciar que en el primer grupo hay un porcentaje significativamente mayor de ciudadanos con estudios bajos (ESO y primarios) y menor de universitarios. En cambio, en el segundo y tercer grupo es al contrario, nos encontramos con un porcentaje de ciudadanos universitarios mucho mayor que los ciudadanos con estudios bajos.

Si observamos los ingresos mensuales del hogar, también podemos deducir algunas cuestiones. En los tres grupos el nivel adquisitivo predominante es el medio (1500 a 2500€ y 2500 a 3500€), sin embargo encontramos diferencias importantes en los demás niveles. En el primer grupo hay un porcentaje significativamente mayor de ciudadanos con rentas inferiores (0 a 1500€) y menor de rentas altas (más de 3500€). En cambio, en el segundo y tercer grupo es a la inversa, vemos que se encuentra un porcentaje mayor de rentas elevadas y menor de rentas inferiores. Hay que destacar que en el tercer grupo el porcentaje de rentas elevadas es un 20% frente al 15% del segundo y 8% del primero. Podemos pensar que este porcentaje más alto de renta se debe a que en él se encuentran ciudadanos de edad más avanzada que llevan más tiempo trabajando y en consecuencia su salario será más elevado por antigüedad (ver tablas 10 y 11).

		Ingresos mensuales		
		Intervalo	Frecuencia	Porcentaje
CLÚSTER 1		0 a 1500€	8	21%
		1500 a 3500€	27	71%
		más de 3500€	3	8%
CLÚSTER 2		0 a 1500€	3	4%
		1500 a 3500€	60	81%
		más de 3500€	11	15%
CLÚSTER 3		0 a 1500€	5	6%
		1500 a 3500€	66	75%
		más de 3500€	12	20%

Tabla 10. Porcentaje de ciudadanos en cada intervalo de ingresos mensuales por clúster

		Nivel de estudios		
		Intervalo	Frecuencia	Porcentaje
CLÚSTER 1		Superiores	7	18%
		Medios	20	53%
		Inferiores	11	29%
CLÚSTER 2		Superiores	33	45%
		Medios	32	43%
		Inferiores	9	12%
CLÚSTER 3		Superiores	29	33%
		Medios	44	50%
		Inferiores	15	17%

Tabla 11. Porcentaje de ciudadanos en cada intervalo de estudios por clúster

Antes de pasar a las cuestiones sobre los problemas de la economía española y las políticas medioambientales, **podemos intentar explicar la razón por la cual se distribuyen los indicadores**, de la forma descrita anteriormente.

Advertimos que el primer grupo se compone de ciudadanos que pese a ser respetuosos y ahorrar energéticamente no otorgan importancia al medioambiente. Una teoría sobre este comportamiento podría ser que al tener unos ingresos mensuales bajos, su ahorro energético y el uso de la bicicleta se debe a la necesidad del ahorro en el hogar y no a la preocupación por la eficiencia energética. Además, el consumo de productos ecológicos no es económico, por tanto, podemos pensar que en este grupo el nivel de ingresos condiciona su comportamiento.

Fijándonos en el segundo grupo vemos que pese a tener una ideología bien definida sobre que es el sector privado (hogares y empresas) quién debe de contribuir más al cuidado del medioambiente y que no existen barreras al reciclaje, su compromiso ecológico y posición ahorradora son deficientes. En consecuencia, podemos decir que este grupo aunque tiene los recursos y la información necesaria (nivel de estudios alto e ingresos medios) acerca de la importancia del cuidado del medioambiente no asume su responsabilidad. Además, podemos añadir que su comportamiento saludable puede deberse más por su propio beneficio que por la preocupación por el medioambiente. Esta actitud puede deberse a un problema en la educación que no ha sabido inculcar a sus ciudadanos los valores medioambientales necesarios, y por ello, debería estar más enfocada a la importancia del medioambiente y la problemática que lo rodea.

Observando al último grupo, vemos que se caracterizan por su alto compromiso ecológico, pese a que su ideología es que es el gobierno quién debe contribuir más al cuidado del medioambiente y que existen barreras al reciclaje. Por tanto, podemos deducir que estos ciudadanos piensan que aunque los hogares deben contribuir en todo lo posible, el gobierno debe estar comprometido y ayudar a eliminar las barreras que impiden el reciclaje. Es alentador ver como el grupo de mayor tamaño es el concienciado con el medioambiente, sin embargo, aun queda un largo camino por recorrer en la concienciación del medioambiente.

Una vez deducidas estas cuestiones, **intentamos explicar si hay relación entre lo comentado anteriormente y la valoración de los problemas de la economía española y las políticas medioambientales más importantes.**

En cuanto a los problemas de la economía española, una vez constatado que para el total de los ciudadanos el problema más importante es el desempleo, observamos el resto. Vemos que el problema del fracaso escolar es más importante para los ciudadanos del primer clúster que para los otros, esto puede ser consecuencia de que al ser el grupo con menos renta y estudios están más preocupados porque sus hijos consigan avanzar en la escala social. Respecto al problema de la productividad quién más alto lo valora es el grupo 2, seguido del 3 y por último el 1. Podemos pensar que los ciudadanos del grupo 2 los cuales la mayoría tienen estudios superiores estarán más concienciados de la importancia de la productividad, para un país como España. En relación al desgaste del medioambiente, los tres grupos lo valoran como poco importante, ya que consideran

más relevantes el resto de problemas. Sin embargo, el grupo 2 seguido del 3 le otorgan mayor importancia que el 1, esto es debido a que en estos clústeres se encuentran ciudadanos más concienciados con el medioambiente. En lo referente al sistema de pensiones, vemos que en este caso el grupo que mayor importancia le otorga es el 1 seguido del 3 y por último el 2. Podemos pensar que al haber en el grupo 1 y 3 ciudadanos de mayor edad que en el 2, éstos estarán más preocupados por su futuro cercano que los del segundo grupo que lo ven lejos.

En cuanto a las políticas medio ambientales apenas observamos diferencias entre clústeres, la mayoría de los ciudadanos les dan mayor importancia a las políticas de obligar a las empresas a tener plantas de tratamiento de aguas, prohibir la emisión de humos y gases de las empresas y regular la emisión de gases de los vehículos. Y menor importancia a prohibir las bolsas de plástico en los comercios e impedir la tala de árboles en los bosques y ciudades.

Para finalizar, concluimos que nuestros encuestados con mayor nivel de estudios y renta más elevada están más concienciados y comprometidos con el cuidado del medioambiente como comentábamos en la introducción en la teoría de Inglehart.

7. BLIBLIOGRAFÍA

Anja Kollmuss & Julian Agyman Tufts University, USA. (2002) Mind the Gap: why do people act environmentally and what are the barriers to pro-environmental behavior?

González López, A. (2002) La preocupación por la calidad del medio ambiente. Un modelo cognitivo sobre la conducta ecológica

Páramo, Pablo; Gómez, Francisco. Actitudes hacia el medio ambiente: su medición a partir de la teoría de facetas. Revista Latinoamericana de Psicología, vol. 29, núm. 2, 1997, pp. 243-266

Cifras de Zaragoza. Datos demográficos del padrón municipal (2015). Ayuntamiento de Zaragoza

Understanding society. Early findings from the first wave of the UK's household longitudinal study

Salvador Figueras, M y Gargallo Valero, P (2006): "Análisis Factorial", [en línea] 5campus.com, Estadística <http://www.5campus.com/leccion/factorial>

Oltra Altgado, C (2006): “Sociedad y medio ambiente. Ciudadanos y científicos ante el proceso de refoma medio ambiental de la sociedad”

M Norusis (2008). Cluster Analysis