

**UNIVERSIDAD DE JAÉN**  
*Facultad de Ciencias Sociales y Jurídicas*

Trabajo Fin de Grado

# **ESTUDIO SOCIO- ECONÓMICO DE LOS MUNICIPIOS DE JAÉN**

**Alumno: Sonia Uceda Rodríguez**

**Julio, 2014**

“Estudio Socio-Económico de los  
Municipios de Jaén”

*A mi familia, con cariño.*

# Summary of socio-economic survey of municipalities in Jaén

The objective of this study is to see how it affects municipalities in Jaén unemployed people according to their educational attainment and sex, as well as students enrolled in public schools at different educational levels and the level of activity by cluster analysis and dimension reduction.

From this analysis it has been inferred that unemployment based on studies completed and sex affects mainly into two separate groups for men and women with secondary education, and other men and women with other types of studies.

As for the students enrolled in public schools by level of education affects different three groups: the first group would consist of students in public schools in the levels of infant, primary, professional qualification programs, high school, middle grade and secondary education for adults, the second group would consist of students in public schools in higher standards and specialized education and the third and last group would be composed of students enrolled in public schools for special education levels and education compulsory secondary.

Finally, in terms of registered unemployment in the towns of Jaén activity level we get a grouping into two: the first group would consist of household activities and fishing, the second group would consist of extraterritorial organizations, extractive industries and production and distribution of electricity, gas and water. As well as health and veterinary activities, social services and transport, storage and communications. The latter with a lower weight.

We have obtained that unemployment does not affect equally to all but there is discrimination on the basis of completed studies separating people with secondary studies other with any other educational level, students enrolled in public schools are separated into students made compulsory studies of other students to do any other type of study, and level of activity, unemployment does not affect equally to persons engaged in housework and other fishing activities.

## INDICE

<b>1.</b>	<b>Introducción</b>	<b>5</b>
1.1	Paro en los Municipios de Jaén	5
1.2	Nivel Educativo	5
1.3	Contratos registrados en la provincia de Jaén por nivel de actividad	6
<b>2.</b>	<b>Metodología</b>	<b>10</b>
2.1	Análisis de Componentes Principales	10
2.1.1	<i>Concepto ACP</i>	10
2.1.2	<i>Metodología ACP</i>	11
2.2	Análisis Factorial	12
2.2.1	<i>Concepto AF</i>	12
2.2.2	<i>Metodología AF</i>	13
2.3	Análisis de Correspondencias	14
2.3.1	<i>Concepto ACO</i>	14
2.3.2	<i>Metodología ACO</i>	15
2.4	Análisis Cluster	16
2.4.1	<i>Concepto AC</i>	16
2.4.2	<i>Metodología AC</i>	16
<b>3.</b>	<b>Análisis</b>	<b>18</b>
3.1	Paro registrado por nivel de estudios terminados y sexo	18
3.2	Alumnos en centros públicos por nivel educativo	26
3.3	Paro registrado por nivel de actividad	44
<b>4.</b>	<b>Resultados</b>	<b>51</b>
<b>5.</b>	<b>Conclusiones y recomendaciones</b>	<b>52</b>
<b>6.</b>	<b>Bibliografía</b>	<b>53</b>

# 1. INTRODUCCIÓN

## 1.1 PARO EN LOS MUNICIPIOS DE JAÉN

Los datos del desempleo en Jaén dan escalofríos. **La tasa de paro es la más alta de las provincias españolas y ha ascendido a casi el 41% según los datos del Instituto Nacional de Empleo.** Unas cifras y unas estadísticas que ponen de manifiesto la extremada debilidad del tejido productivo de la provincia que ha perdido, con la virulencia de la crisis económica, los pujantes focos de actividad fabril e industrial de Bailén, Linares, Mancha Real o La Carolina. Localidades que han visto como el paro se ha desbocado en los últimos meses con el cierre de las fábricas de cerámica industrial para la construcción, la desaparición de la industria automovilística linarense con Santana o el desplome del mercado de los muebles de cocina. **A lo que hay que unir la desastrosa cosecha de aceituna que tuvo una merma en producción; y en generación de empleo, de casi el 80%**

Así las cosas los últimos datos del paro en Jaén han dado un ligero respiro ya que **la lista descendió en 334 personas. Aun así siguen siendo 68.039 personas las que buscan un trabajo en una provincia de poco más de 650.000 habitantes.** Una bajada mínima, pero una bajada al fin y al cabo en un mercado laboral tan precario y deteriorado como el jienense que ha visto como en el último año, de septiembre de 2012 a septiembre de 2013, los parados han aumentado en más de 2.700 personas.

## 1.2 NIVEL EDUCATIVO

Como se puede observar (Figura 1.1), la población se gasta mucho menos cuando la enseñanza es pública. Independientemente de que sea pública, privada concertada o privada sin concierto, se puede observar claramente que en función del aumento de los estudios, la población tiende a tener un menor gasto dedicado a formación. Excepto en el nivel universitario.

En este estudio vamos a tratar de determinar si todo esto tiene una causa justificada, el paro, y si realmente afecta de igual modo a los municipios de Jaén.

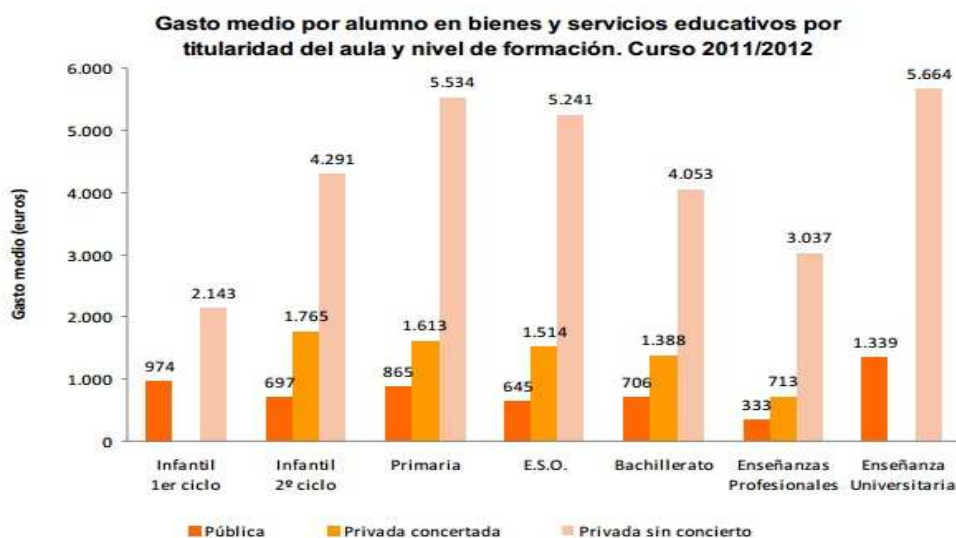


Figura 1.1: Extraído del Observatorio Económico de la Provincia de Jaén.

### 1.3 PARO Y CONTRATOS REGISTRADOS EN LA PROVINCIA DE JAÉN

La escasa campaña de aceituna y de producción de aceite que ha registrado la provincia jienense este año está empezando ya a reflejarse en las principales magnitudes económicas de Jaén. Así se refleja no sólo en la caída del empleo y el aumento del paro, sobre todo en agricultura (recordar que se estimaron en más de 5 millones los jornales perdidos en esta campaña agrícola), sino en aspectos del consumo y la demanda de crédito (caen las hipotecas y venta de viviendas, las de automóviles, etc.).

Según el informe de Analistas, la provincia de Jaén registra un notable aumento del nivel de desempleo en el primer trimestre del año, con una tasa interanual del 31,3%, lo que supone 29.600 parados más en términos absolutos. Este aumento en la cifra de parados, que también se debió a un incremento de los activos (2,8%), propicia un incremento de la tasa de paro de 8,9 puntos respecto al primer trimestre de 2012, frente a los 3,7 puntos en el conjunto de Andalucía.

Por su parte, el empleo ha registrado una caída del 10,6% en términos interanuales, frente al -5,8% del conjunto regional. Este comportamiento se ha debido en gran medida al fuerte descenso del empleo agrario, que se ha reducido casi un 46%, dada la negativa campaña de la recogida de aceituna. Así, la tasa de paro en el sector agrario se ha incrementado notablemente en el primer trimestre (19,5% en términos interanuales), hasta situarse en el 65,9% del total de activos en el sector. Del mismo modo, ha destacado la caída en la afiliación de extranjeros en

la provincia hasta abril (-35,3% en términos interanuales), siendo menos intensa la caída en el conjunto regional (-3,5%).

En lo que se refiere al sector construcción, cabría señalar la caída en las operaciones de compraventas entre enero y abril en la provincia, con un descenso del 10,4% en términos interanuales, debido tanto al descenso en las operaciones de vivienda nueva (-18%) como usada (-6%).

Por otro lado, la provincia de Jaén registra el mayor aumento relativo de los depósitos privados que, a tenor de la información del Banco de España, han crecido un 2,6% en el primer trimestre en términos interanuales (0,7% en Andalucía).

Por su parte, las exportaciones de la provincia en términos reales han caído un 27,8% en el periodo comprendido entre enero y abril, debido, sobre todo, a la caída en la venta de alimentos al exterior (-68,2%). En términos de valor, por el contrario, se ha observado un incremento de las ventas totales del 10,6%, con una caída muy significativa en alimentos, debido al descenso de la partida grasas y aceites (-64,4%), que tiene un peso sobre el total de alimentos del 70,5%.

En función de las contrataciones, el volumen de contratación se ha incrementado un 14,90% respecto de 2009, alcanzándose la cota más alta del último quinquenio.

La mayor parte de la contratación se formaliza con personas sin estudios o en niveles formativos más bajos.

La contratación a extranjeros se ha incrementado un 38,78%.

En el Sector Agricultura se formaliza el 64,46 % de la contratación provincial, seguido por Servicios con un 23,26%.

La temporalidad en la contratación tiene una representatividad del 98,15%

La movilidad interprovincial en la contratación tiene su base en la actividad agrícola.

Las ocupaciones más contratadas se encuadran en el sector agrícola.

De los contratos formalizados en las veinte ocupaciones más relevantes, el 72,01% se registran en ocupaciones del sector agrícola.

La contratación de ocupaciones del Sector Servicios presenta variaciones positivas, en especial las de trabajadores de cuidados personales y las relacionadas con la hostelería.

La mayor parte de las ocupaciones más demandadas por los desempleados se enmarcan en el Sector Servicios.

En la provincia, el sector económico que presenta un crecimiento más relevante en el número de activos es el agrícola (Figura 1.2) con un aumento del 11,20%, y en menor medida el de servicios con el 1,94%. El resto de sectores experimentan disminuciones de distinta significatividad, destacando el de Construcción con una variación interanual negativa del 9,67% y del 15,71% desde el año 2006.

Tabla 3: EVOLUCIÓN DE LA POBLACIÓN ACTIVA POR SECTOR ECONÓMICO					
	2006	2007	2008	2009	2010
<b>Activos (en miles)</b>					
Agricultura	34,9	38,6	39,9	45,5	50,6
Industria	38,4	37,2	35,3	31,5	28,1
Construcción	29,9	37,6	34,4	27,9	25,2
Servicios	143,8	158,9	169,1	159,4	162,5
Parados(*)	13,9	12,0	16,7	23,1	31,5

Figura 1.2: Extraído del SEPE: Servicio Público de Empleo Estatal

Indicadores de confianza empresarial. (Figura 1.3):

#### Porcentaje de respuestas en relación a su situación y expectativas

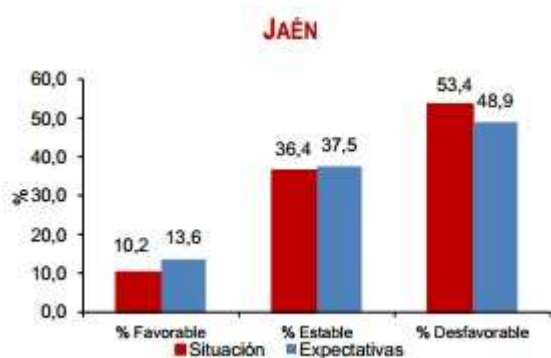


Figura 1.3: Extraído del Observatorio económico de la Provincia de Jaén

Lo que se quiere conseguir con este estudio es analizar dentro de los municipios de Jaén, cual es el perfil o perfiles de personas que más se adaptan en la actualidad a los parados por nivel de estudios terminados y sexo, al perfil o perfiles de alumnos en centros públicos por nivel

educativo, a los emigrantes más destacados en función de la edad y el sexo, y a qué personas corresponden los contratos registrados por nivel de estudios y sexo.

Para ver, en conclusión, cual es el perfil más afectado por los tiempos de crisis actuales, analizamos los datos citados anteriormente mediante técnicas inferenciales y de agrupación y reducción de la dimensión, para poder llegar a encontrar una aproximación de las mayores víctimas de nuestra sociedad actual dentro de los municipios de Jaén.

## 2. METODOLOGÍA

En este proyecto vamos a utilizar los cuatro tipos de análisis más relevantes en las técnicas de agrupación y reducción de la dimensión que son el análisis de componentes principales, análisis factorial, análisis de correspondencias y análisis cluster.

### 2.1 ANÁLISIS DE COMPONENTES PRINCIPALES

#### 2.1.1 CONCEPTO ACP

En estadística, el **análisis de componentes principales** (en español **ACP**, en inglés, **PCA**) es una técnica utilizada para reducir la dimensionalidad de un conjunto de datos. Intuitivamente la técnica sirve para hallar las causas de la variabilidad de un conjunto de datos y ordenarlas por importancia.

Técnicamente, el ACP busca la proyección según la cual los datos queden mejor representados en términos de mínimos cuadrados. El ACP se emplea sobre todo en análisis exploratorio de datos y para construir modelos predictivos. El ACP comporta el cálculo de la descomposición en autovalores de la matriz de covarianza, normalmente tras centrar los datos en la media de cada atributo.

El ACP construye una transformación lineal que escoge un nuevo sistema de coordenadas para el conjunto original de datos en el cual la varianza de mayor tamaño del conjunto de datos es capturada en el primer eje (llamado el Primer Componente Principal), la segunda varianza más grande es el segundo eje, y así sucesivamente. Para construir esta transformación lineal debe construirse primero la matriz de covarianza o matriz de coeficientes de correlación. Debido a la simetría de esta matriz existe una base completa de vectores propios de la misma. La transformación que lleva de las antiguas coordenadas a las coordenadas de la nueva base es precisamente la transformación lineal necesaria para reducir la dimensionalidad de datos. Además las coordenadas en la nueva base dan la composición en factores subyacentes de los datos iniciales.

Una de las ventajas del ACP para reducir la dimensionalidad de un grupo de datos, es que retiene aquellas características del conjunto de datos que contribuyen más a su varianza, manteniendo un orden de bajo nivel de los componentes principales e ignorando los de alto nivel. El objetivo es que esos componentes de bajo orden a veces contienen el aspecto "más importante" de esa información.

Existen dos formas básicas de aplicar el ACP:

1. Método basado en la matriz de correlación, cuando los datos no son dimensionalmente homogéneos o el orden de magnitud de las variables aleatorias medidas no es el mismo.
2. Método basado en la matriz de covarianzas, que se usa cuando los datos son dimensionalmente homogéneos y presentan valores medios similares.

### 2.1.2 METODOLOGÍA ACP

En este estudio vamos a aplicar el análisis de componentes principales a los datos recogidos del Instituto de Estadística y Cartografía del paro registrado por nivel de estudios y sexo en el último año (datos incluidos en el cd anexo), mediante el programa estadístico R Commander. Aplicamos este análisis debido a que esta técnica nos va a ayudar a reducir la dimensión de las variables de este conjunto de datos, de modo que las ordene por importancia y los datos queden mejor representados sin contar con todas las variables, sino con las más representativas.

Para ello comenzaremos observando el gráfico de sedimentación para ver qué variables llegan al autovalor 1 y en una primera observación saber de forma general las variables que podrían ser extraídas, es decir, qué variables explican mejor los datos.

Analizaremos las saturaciones o las cargas de cada uno de los ítems de estas variables o componentes. En este tipo de análisis lo principal es averiguar con cuántas variables podemos llegar a explicar el 80% del total de los datos.

Observaremos la importancia de los componentes, la proporción de la varianza y la proporción acumulada, incluyendo un biplot finalmente para ver de forma más clara los componentes principales dentro de este conjunto de datos y tomar la decisión correcta para reducir la dimensión de variables.

A continuación haremos un análisis cluster, detallado más abajo, para verificar si los datos realmente pueden quedar agrupados.

## 2.2 ANÁLISIS FACTORIAL

### 2.2.1 CONCEPTO AF

**Análisis factorial** es una técnica estadística de reducción de datos usada para explicar las correlaciones entre las variables observadas en términos de un número menor de variables no observadas llamadas factores. Las variables observadas se modelan como combinaciones lineales de factores más expresiones de error. El análisis factorial se originó en psicometría, y se usa en las ciencias del comportamiento tales como ciencias sociales, marketing, gestión de productos, investigación de operaciones y otras ciencias aplicadas que tratan con grandes cantidades de datos.

Existen dos tipos de análisis factorial:

El *análisis factorial exploratorio*, AFE, se usa para tratar de descubrir la estructura interna de un número relativamente grande de variables. La hipótesis *a priori* del investigador es que pueden existir una serie de factores asociados a grupos de variables. Las *cargas* de los distintos factores se utilizan para intuir la relación de éstos con las distintas variables. Es el tipo de análisis factorial más común.

El *análisis factorial confirmatorio*, AFC, trata de determinar si el número de factores obtenidos y sus *cargas* se corresponden con los que cabría esperar a la luz de una teoría previa acerca de los datos. La hipótesis *a priori* es que existen unos determinados factores preestablecidos y que cada uno de ellos está asociado con un determinado subconjunto de las variables. El análisis factorial confirmatorio entonces arroja un nivel de confianza para poder aceptar o rechazar dicha hipótesis.

El análisis factorial se utiliza para identificar factores que expliquen una variedad de resultados en diferentes pruebas. Por ejemplo, investigación en inteligencia halla que la gente que obtiene una nota alta en una prueba de habilidad verbal también se desempeña bien en pruebas que requieren habilidades verbales. Los investigadores explican esto mediante el uso de análisis factorial para aislar un factor a menudo llamado inteligencia cristalizada o inteligencia verbal, que representa el grado en el cual alguien es capaz de resolver problemas usando habilidades verbales.

Análisis factorial en psicología se asocia frecuentemente con la investigación sobre la inteligencia. Sin embargo, también se ha utilizado en un amplio rango de dominios, tales como personalidad, actitudes, creencias, etc. Está asociado a la psicometría, debido a que puede

evaluar la validez de un instrumento estableciendo si el instrumento de verdad mide los factores postulados.

### 2.2.2 METODOLOGÍA AF

En este estudio vamos a aplicar el análisis factorial a los datos recogidos del Instituto de Estadística y Cartografía de los alumnos en centros públicos por nivel educativo (datos incluidos en el cd anexo), mediante el programa estadístico SPSS. Aplicamos este análisis debido a que esta técnica nos va a ayudar a tratar de descubrir si existen factores asociados a las variables, y, en función de las cargas, podremos intuir la relación de estos factores con las distintas variables. Su propósito principal es sacar a la luz la estructura subyacente en una matriz de datos, para ello estudia las interrelaciones entre variables, no exigiendo una distinción entre variables dependientes o independientes y calcula un conjunto de dimensiones latentes (no observables) conocidas como factores. Por lo que vamos a tratar de expresar la información contenida en la matriz de datos en un número menor de dimensiones representadas por dichos factores, que estarán formados por las variables más correladas entre sí.

Para la estimación de las cargas y las comunalidades vamos a usar el método de componentes principales que es un método que proporciona normalmente buenos resultados siempre que sea cierta alguna condición como:

Que las correlaciones sean bastante grandes dando lugar a un valor pequeño de  $m$ , o que el número de variables sea grande.

También en ciertos pasos compararemos los resultados de este análisis con el análisis de factorización de ejes principales.

La comunalidad de una variable es la proporción de su varianza que puede ser explicada por el modelo factorial obtenido. Estudiando las comunalidades de la extracción podemos valorar cuáles de las variables son peor explicadas por el modelo.

Para la determinación del número de factores, es decir, para decidir el número de  $m$  factores que se deben extraer, tendremos en cuenta:

- Varianza total: Seleccionaremos  $m$  suficientemente grande para que la suma de las comunalidades o la suma de los autovalores constituya una proporción relativamente grande de la  $tr(S)$  o de  $p$ , deberá representar un 80% de la variabilidad total.
- Regla de Kaiser: Retendremos los factores cuyos autovalores, utilizando en nuestro caso la matriz de correlaciones  $R$ , sea 1.

- Gráfico de sedimentación: Elegiremos  $m$  como el número de autovalores en la zona de fuerte pendiente.

- Test de hipótesis: Asociado al método que vayamos a utilizar para la extracción de factores, en el caso en que rechacemos la hipótesis nula es un mal ajuste para  $S$  y se debe probar un valor superior de  $m$ .

En nuestro estudio nos hemos propuesto utilizar el método de rotación Varimax, cuyo objetivo es calcular las cargas rotadas que maximicen las varianzas de las cargas cuadráticas en cada columna de la matriz de cargas estimadas.

A continuación haremos un análisis cluster, detallado más abajo, para verificar si los datos realmente pueden quedar agrupados.

## 2.3 ANÁLISIS DE CORRESPONDENCIAS

### 2.3.1 CONCEPTO ACO

Es una técnica descriptiva o exploratoria cuyo objetivo es resumir una gran cantidad de datos en un número reducido de dimensiones, con la menor pérdida de información posible.

En esta línea, su objetivo es similar al de los métodos factoriales, salvo que en el caso del análisis de correspondencias el método se aplica sobre variables categóricas u ordinales.

El análisis de correspondencias simples se utiliza a menudo en la representación de datos que se pueden presentar en forma de tablas de contingencia de dos variables nominales u ordinales. Otras utilizaciones implican el tratamiento de tablas de proximidad o distancia entre elementos, y tablas de preferencias.

Si se trata de una tabla de contingencia de dos variables cualitativas, una variable cuyas categorías aparecen en filas y la otra variable cuyas categorías son representadas en columnas, el análisis de correspondencias consiste en resumir la información presente en las filas y columnas de manera que pueda proyectarse sobre un subespacio reducido, y representarse simultáneamente los puntos fila y los puntos columna, pudiéndose obtener conclusiones sobre relaciones entre las dos variables nominales u ordinales de origen

La extensión del análisis de correspondencias simples al caso de varias variables nominales (tablas de contingencia multidimensionales) se denomina Análisis de Correspondencias Múltiples, y utiliza los mismos principios generales que la técnica anterior. En general se orienta a casos en los cuales una variable representa ítems o individuos y el resto son variables cualitativas u ordinales que representan cualidades.

Entre la utilización del Análisis de Correspondencias Simple y Múltiple, estudios:

- Preferencias de consumo en Investigación de Mercados.
- Posicionamiento de empresas a partir de las preferencias de consumidores.
- Búsqueda de tipologías de individuos respecto a variables cualitativas (patrones de enfermedades en medicina, perfiles psicológicos, comportamiento de especies en biología, etc.).

El Análisis de Correspondencias tiene dos objetivos básicos:

Asociación entre categorías de columnas o filas: Medir la asociación de solo una fila o columna, para ver, por ejemplo, si las modalidades de una variable pueden ser combinadas.

Asociación entre categorías de filas y columnas: Estudiar si existe relación entre categorías de las filas y columnas.

El análisis de correspondencias solo requiere que los datos representen las respuestas a una serie de preguntas y que estén organizadas en categorías. Dependiendo si existen dos o más variables el análisis será simple o múltiple.

### 2.3.2 METODOLOGÍA ACO

En este estudio vamos a aplicar el análisis de correspondencias a los datos recogidos del Instituto de Estadística y Cartografía del paro registrado por nivel de actividad (datos incluidos en el cd anexo), mediante el programa estadístico R Commander. Aplicamos este análisis debido a que esta técnica nos va a ayudar a resumir una gran cantidad de datos en un número reducido de dimensiones, con la menor pérdida de información posible.

En la primera parte del análisis se mostrarán las dimensiones que explican todos los datos o toda la varianza. Observaremos con cuántas dimensiones se explican al menos un 80% de los datos.

Haremos un gráfico para observar mejor las puntuaciones de los perfiles por filas y columnas así como las contribuciones totales de cada perfil fila a la inercia de cada dimensión y las contribuciones relativas de cada dimensión la inercia del punto.

## 2.4 ANÁLISIS CLUSTER

### 2.4.1 CONCEPTO AC

El Análisis Cluster, conocido como Análisis de Conglomerados, es una técnica estadística multivariante que busca agrupar elementos (o variables) tratando de lograr la máxima homogeneidad en cada grupo y la mayor diferencia entre los grupos.

El Análisis Cluster tiene una importante tradición de aplicación en muchas áreas de investigación. Sin embargo, junto con los beneficios del Análisis Cluster existen algunos inconvenientes. El Análisis Cluster es una técnica descriptiva, atórica y no inferencial.

El Análisis Cluster no tiene bases estadísticas sobre las que deducir inferencias estadísticas para una población a partir de una muestra, es un método basado en criterios geométricos y se utiliza fundamentalmente como una técnica exploratoria, descriptiva pero no explicativa.

Las soluciones no son únicas, en la medida en que la pertenencia al conglomerado para cualquier número de soluciones depende de muchos elementos del procedimiento elegido. Por otra parte, la solución cluster depende totalmente de las variables utilizadas, la adición o destrucción de variables relevantes puede tener un impacto substancial sobre la solución resultante

Los algoritmos de formación de conglomerados se agrupan en dos categorías:

- Algoritmos de partición: Método de dividir el conjunto de observaciones en  $k$  conglomerados (clusters), en donde  $k$  lo define inicialmente el usuario.
- Algoritmos jerárquicos: Método que entrega una jerarquía de divisiones del conjunto de elementos en conglomerados.

Un método jerárquico aglomerativo parte con una situación en que cada observación forma un conglomerado y en sucesivos pasos se van uniendo, hasta que finalmente todas las situaciones están en un único conglomerado.

Un método jerárquico disociativo sigue el sentido inverso, parte de un gran conglomerado y en pasos sucesivos se va dividiendo hasta que cada observación queda en un conglomerado distinto.

### 2.4.2 METODOLOGÍA AC

En este estudio vamos a aplicar el análisis cluster a los datos anteriormente mencionados cuyo análisis previo será el de análisis de componentes principales para los datos de parados en los municipios de Jaén por nivel educativo y sexo y el análisis factorial para los datos recogidos de

los alumnos en centros públicos por nivel educativo en los municipios de Jaén, mediante el programa estadístico R Commander. Aplicamos este análisis debido a que esta técnica nos va a ayudar a encontrar un agrupamiento de las variables que refleje posibles similitudes entre subconjuntos de ellas y que sean diferentes entre los distintos subgrupos.

En primer lugar vamos a hacer un análisis de distancias y para ello haremos la matriz de distancias.

La matriz de distancias expresa el distanciamiento entre pares de casos. La medida de distancia seleccionada es la distancia euclídea al cuadrado.

En el siguiente paso realizaremos una tabla de agrupamiento, utilizando el método “Ward” para observar cómo se unen los cluster. Mostraremos los resultados utilizando también el método “Average”.

A continuación aplicaremos el análisis cluster y obtendremos un dendrograma que nos ayudará a decidir el número de grupos que podrían representar mejor la estructura de los datos teniendo en cuenta la forma en que se van anidando los cluster y la medida de similitud a la cual lo hacen.

## 3. ANÁLISIS

Comenzamos el análisis aplicando a cada grupo de datos los cuatro tipos de métodos mencionados anteriormente.

### 3.1 PARO POR NIVEL DE ESTUDIOS TERMINADOS Y SEXO

Aplicamos el método de análisis de componentes principales a los datos recogidos de paro por nivel de estudios terminados y sexo en los municipios de Jaén.

Aplicamos este análisis debido a que esta técnica nos va a ayudar a reducir la dimensión de las variables de este conjunto de datos, de modo que las ordene por importancia y los datos queden mejor representados sin contar con todas las variables, sino con las más representativas.

En este primer paso contamos con ocho variables que son:

HA: Hombres analfabetos

HP: Hombres con estudios primarios

HS: Hombres con estudios secundarios

HPS: Hombres con estudios postsecundarios

MA: Mujeres analfabetas

MP: Mujeres con estudios primarios

MS: Mujeres con estudios secundarios

MPS: Mujeres con estudios postsecundarios.

Los resultados obtenidos con el programa estadístico R Commander son:

## Gráfico de Sedimentación

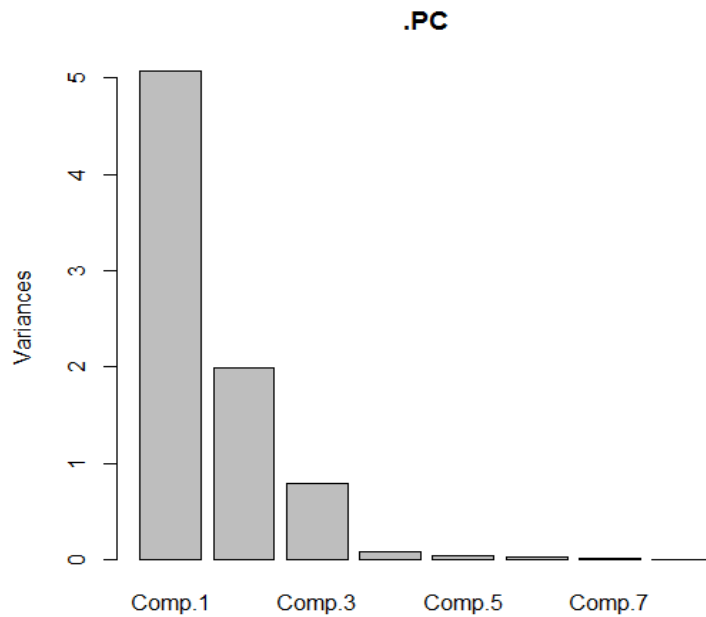


Figura 3.1.1: Extraída del programa estadístico R Commander

En primer lugar obtenemos el gráfico de sedimentación (Figura 3.1.1). En el eje vertical se encuentran las varianzas o también denominados autovalores, donde se puede observar que no llegan al autovalor 1 nada más que los dos primeros componentes, aunque se observa que el tercer componente también está cerca del autovalor 1. Por la regla de Kaiser extraeríamos los dos primeros componentes y en el gráfico de sedimentación observamos que es el primer componente el que claramente podría ser extraído.

A continuación vemos las saturaciones o las cargas de cada uno de los ítems de estos componentes (Figura 3.1.2). Podemos ver que ítems saturan o están más relacionados con cada uno de los componentes.

### Saturaciones de los ítems

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
HA	-0.43287223	-0.01437809	0.1351830	0.565724681	0.4236739	-0.2405349		
	0.41373632	-0.25605021						
HP	-0.38873693	0.10389601	-0.5068121	-0.125652903	0.2703531	0.3996475	0.21979349	
	0.53322038							
HPS	-0.42240654	0.13614188	-0.1456341	-0.687176341	0.0491364	-0.4235761		
	0.03313680	-0.35593303						
HS	-0.08346495	-0.68754053	-0.1403960	0.008362301	0.4115094	-0.1676007	-	
	0.54426800	0.08296392						
MA	-0.43244613	0.04473743	0.1995172	0.173451526	-0.4740345	-0.4316279	-	
	0.18139404	0.54514331						
MP	-0.42425893	0.11074276	-0.2419399	0.273274353	-0.3236848	0.4062982	-	
	0.44705681	-0.45256681						
MPS	-0.31710324	-0.11743298	0.7552602	-0.286434202	0.1351991	0.4602400	-	
	0.03518814	0.04306601						
MS	-0.09247587	-0.68534399	-0.1232345	-0.071251849	-0.4773844	0.1082425		
	0.49918492	-0.11265148						

Figura 3.1.2: Extraída del programa estadístico R Commander

Observamos que la primera componente tiene más carga en las variables HA, HP, HPS, MA, MP y MPS, y la segunda componente en las variables en HS y MS. Para ver si efectivamente estos ítems son los que más saturan, tendremos que compararlo con un biplot, un poco más abajo.

La ventana de resultados del R Commander nos devuelve los autovalores que vienen a continuación (Figura 3.1.3) donde efectivamente vemos que el primer componente tiene un autovalor de 5.0679, el segundo 1.9854 y el tercero tiene un autovalor de la unidad, se queda por debajo de 1.

### Autovalores

Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
5.067936328	1.985395570	0.790252130	0.078631618	0.035924676	0.024123281		
0.012818872	0.004917525						

Figura 3.1.3: Extraída del programa estadístico R Commander

Observamos la importancia de los componentes, la proporción de la varianza y la proporción acumulada.

### Importancia de los componentes

Importance of components:							
	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
Standard deviation	2.251208	1.4090407	0.88896126	0.280413299	0.189538061		
Proportion of Variance	0.633492	0.2481744	0.09878152	0.009828952	0.004490585		
Cumulative Proportion	0.633492	0.8816665	0.98044800	0.990276956	0.994767540		

Figura 3.1.4: Extraída del programa estadístico R Commander

De acuerdo al criterio de seleccionar autovalores hasta lograr explicar al menos un 80% de la variabilidad total el número de componentes a retener serían 2. La regla de Kaiser no crea ninguna duda debido a que  $\lambda_2=1.409$ , es más grande que la unidad.

Vemos que la proporción de varianza (Figura 3.1.4) que explica el primero es un 63.35% de la varianza y el segundo de 24.82%, el tercero ya explica solamente un 9.88%, el cuarto un 0.98%...

De igual modo se observa que la varianza acumulada explicada llega a pasar del 88% solamente con los dos primeros componentes.

El biplot muestra gráficamente los resultados (Figura 3.1.5).

## Biplot

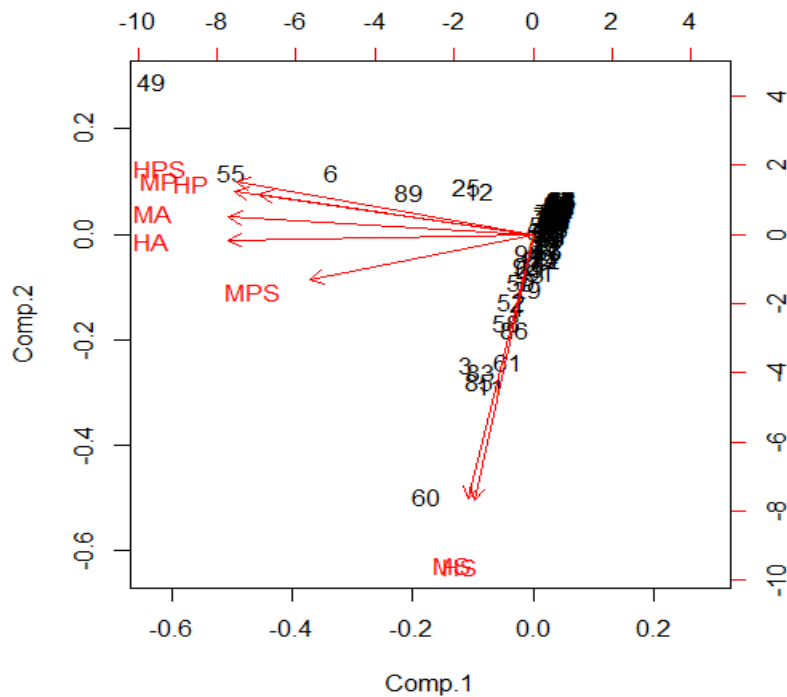


Figura 3.1.5: Extraída del programa estadístico R Commander

Las variables que tienden hacia la primera componente son HA, HP, HPS, MA, MP y MPS, y las variables que tienden hacia la segunda componente son HS y MS.

Para la interpretación de las componentes hemos utilizado tanto las cargas como el biplot que aparece en la última figura.

Como anteriormente hemos mencionado, con las cargas de cada una de las variables en los distintos componentes, vemos que también se cumple en el gráfico.

Podemos decir que la primera componente representa una característica que se puede definir como el paro registrado por nivel de estudios terminados y sexo en los municipios de Jaén caracterizado por mujeres y hombres analfabetos, mujeres y hombres con estudios primarios, mujeres y hombres con estudios postsecundarios.

La segunda componente se puede definir como el paro registrado por nivel de estudios terminados y sexo en los municipios de Jaén caracterizado por los hombres y mujeres con estudios secundarios.

De este análisis podemos deducir que en los municipios de Jaén, la mayoría de parados se podrían clasificar en dos grupos, los hombres y mujeres con estudios secundarios, y los hombres y mujeres sin estudios o con estudios primarios y postsecundarios.

Para verificar que efectivamente las variables pueden ser agrupadas vamos a realizar un análisis cluster.

Aplicamos este análisis debido a que esta técnica nos va a ayudar a encontrar un agrupamiento de las variables que refleje posibles similitudes entre subconjuntos de ellas y que sean diferentes entre los distintos subgrupos.

Se desea encontrar un agrupamiento de las variables que refleje posibles similitudes entre subconjuntos de ellas.

En primer lugar vamos a hacer un análisis de distancias y para ello hemos obtenido la matriz de distancias (Figura 3.1.6).

**Matriz de distancias**

HA	HP	HS	HPS	MA	MP	MS	
HP	152952.13						
HS	2862740.41	3263208.78					
HPS	409819.08	458960.04	3959757.23				
MA	46923.05	235346.51	3217579.72	280764.26			
MP	88276.17	25008.00	3277482.88	468064.08	146674.54		
MS	2560276.14	2918917.65	113009.13	3615693.62	2856483.07	2914146.61	
MPS	682149.64	1361500.06	3141212.19	1120448.77	584069.81	1130765.44	2839556.93

*Figura 3.1.6: Extraída del programa estadístico R Commander*

La matriz de distancias expresa el distanciamiento entre pares de casos. La medida de distancia seleccionada es la distancia euclídea al cuadrado.

Observamos en la matriz cada una de las distancias de cada variable, por ejemplo entre las variables HP y HA, hay una distancia de 152952.13, por lo que hay una distancia bastante grande; entre HS y HA hay también mucha distancia, 2862740.41, y así con todas las variables se observan grandes distancias. Las distancias más pequeñas entre las distintas variables son entre MA y HA; MP y HP.

En el siguiente paso realizamos una tabla de agrupamiento, utilizando el método “Ward”:

### Tabla de agrupamiento “Ward”

[,1] [,2]
[1,] -2 -6
[2,] -1 -5
[3,] -3 -7
[4,] 1 2
[5,] -4 4
[6,] -8 5
[7,] 3 6

Figura 3.1.7: Extraída del programa estadístico R Commander

En el paso 1 (Figura 3.1.7), se unen los cluster formados por una sola observación 2 y 6. Los signos negativos indican que cuando se unen está formados por una sola observación cada uno. En paso 2, se unen los cluster 1 y 5. En el paso 3 se unen los cluster 3 y 7, hasta el momento todos formados por una sola observación. En el paso 4, se unen los cluster 1 y 2 y así sucesivamente. Se unirán unos a otros cuando encajen en la numeración, por ejemplo, un cluster 1 y 2 se puede unir a un cluster formado por las observaciones 3 y 4.

A continuación se muestran los resultados utilizando el método “Average”:

### Tabla de agrupamiento “Average”

[,1] [,2]
[1,] -2 -6
[2,] -1 -5
[3,] -3 -7
[4,] 1 2
[5,] -4 4
[6,] -8 5
[7,] 3 6

Figura 3.1.8: Extraída del programa estadístico R Commander

Vemos que se obtienen los mismos resultados con el método Ward que con el método Average (Figura 3.1.8).

A continuación se aplica el análisis cluster y se obtiene un dendrograma (Figura 3.1.9).

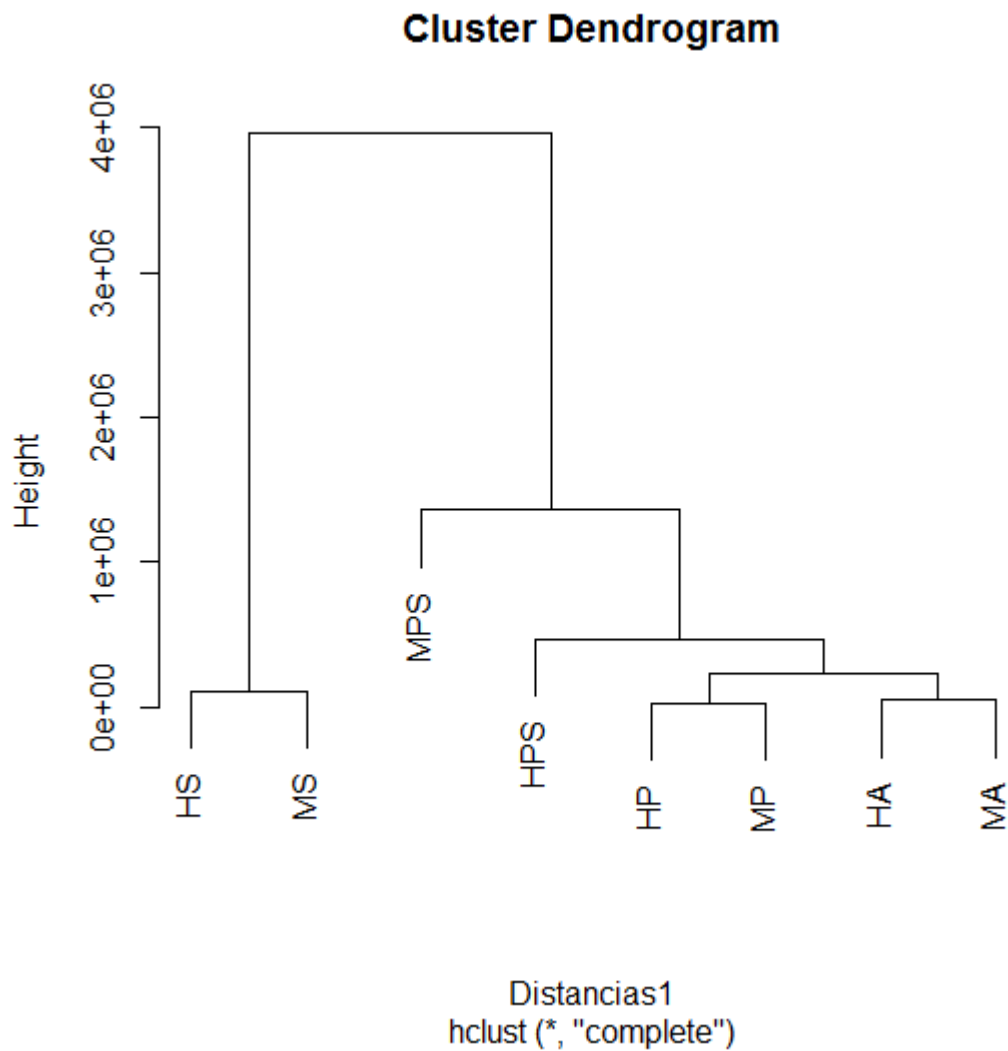


Figura 3.1.9: Extraída del programa estadístico R Commander

El dendrograma es una valiosa herramienta visual que puede ayudar a decidir el número de grupos que podrían representar mejor la estructura de los datos teniendo en cuenta la forma en que se van anidando los cluster y la medida de similitud a la cual lo hacen.

Cortando el gráfico con un segmento perpendicular a las ramas se obtiene una partición con un número de grupos igual a las ramas “cortadas”

El número de cluster o clases en que deseamos agrupar los datos nos llevará “cortar” el dendrograma del ejemplo verticalmente y ver a qué nivel de similitud se da dicho agrupamiento. En el ejemplo puede apreciarse un salto importante en la longitud de las líneas verticales que definen los grupos en 2. Cuando se observa en la escala vertical un amplio rango sin existencia de agrupamiento puede ser un indicio de que los cluster se encuentran separados a esos niveles de similaridad.

## Bannerplot

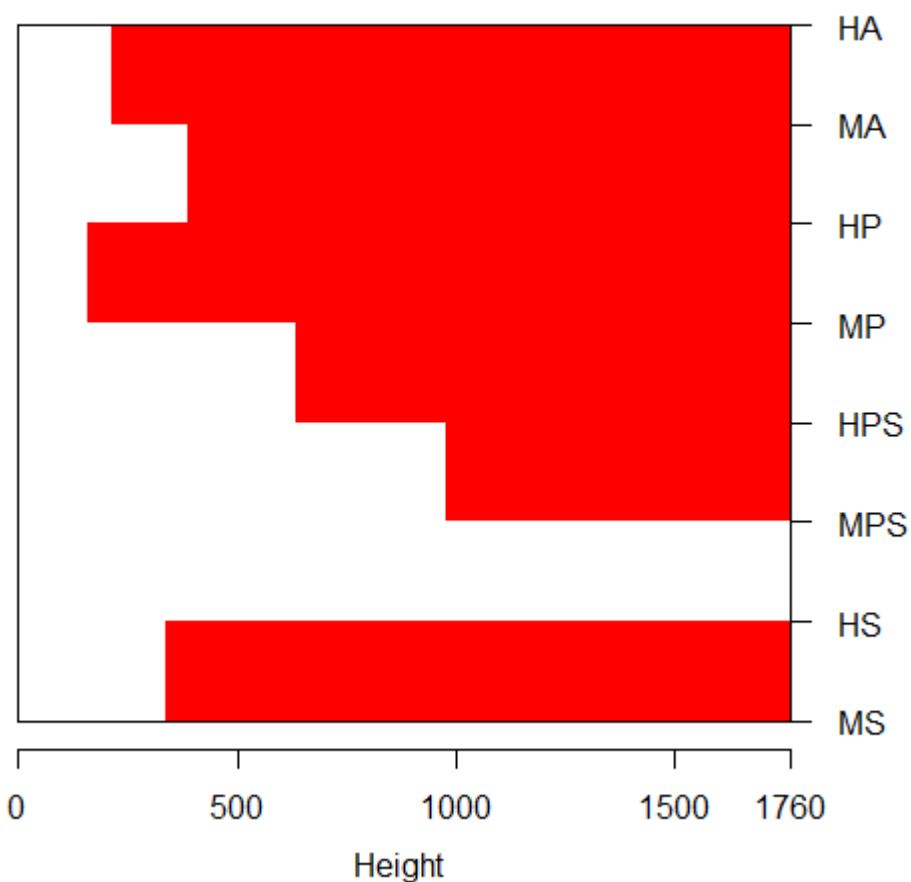


Figura 3.1.10: Extraída del programa estadístico R Commander

Según el dendrograma y el bannerplot (Figura 3.1.10), si decidimos establecer 2 cluster, estarán formados por los siguientes casos:

Cluster 1: Hombres y mujeres con estudios secundarios.

Cluster 2: Hombres y mujeres analfabetos, con estudios primarios y estudios postsecundarios.

Con los dos análisis verificamos el posible agrupamiento de las variables.

### 3.2 ALUMNOS EN CENTROS PÚBLICOS POR NIVEL EDUCATIVO

Aplicamos el método de análisis factorial a los datos recogidos de los alumnos en centros públicos por nivel educativo en los municipios de Jaén.

Aplicamos este análisis debido a que esta técnica nos va a ayudar a tratar de descubrir si existen factores asociados a las variables, y, en función de las cargas, podremos intuir la relación de estos factores con las distintas variables.

En este análisis contamos con diez variables que son:

INF: Infantil

PRI: Primaria

EDE: Educación Especial

ESO: Educación Secundaria Obligatoria

PCP: Programas de Cualificación Profesional Inicial

BCH: Bachillerato

GM: C.F Grado Medio

GS: C.F Grado Superior

ESA: Enseñanza Secundaria de Adultos

ERE: Enseñanzas de Régimen Especial

Para este análisis hemos relativizado los datos con el total de la población de cada municipio.

Veamos en primer lugar las hipótesis que asume el análisis factorial:

Sobre la *linealidad* se pueden realizar diagramas de dispersión por pares de variables para comprobarla. En nuestro estudio al contar con un número grande de variables se convierte en un procedimiento poco práctico por lo que no lo realizaremos.

*Matriz de correlaciones:* Fuera de la diagonal se puede observar si existen correlaciones entre pares de variables.

**Matriz de correlaciones<sup>a</sup>**

	INF	PRI	EDE	ESO	PCP	BCH	GM	GS	ESA	ERE
Correlación INF	1,000	,286	,037	-,104	-,290	-,338	-,287	-,205	-,207	-,207
PRI	,286	1,000	,045	,012	-,444	-,555	-,463	-,372	-,311	-,352
EDE	,037	,045	1,000	,034	,031	,053	-,015	,044	-,016	,008
ESO	-,104	,012	,034	1,000	,093	,144	-,008	-,077	-,129	-,144
PCP	-,290	-,444	,031	,093	1,000	,761	,645	,339	,389	,259
BCH	-,338	-,555	,053	,144	,761	1,000	,645	,421	,441	,383
GM	-,287	-,463	-,015	-,008	,645	,645	1,000	,507	,612	,351
GS	-,205	-,372	,044	-,077	,339	,421	,507	1,000	,245	,670
ESA	-,207	-,311	-,016	-,129	,389	,441	,612	,245	1,000	,267
ERE	-,207	-,352	,008	-,144	,259	,383	,351	,670	,267	1,000

a. Determinante = ,025

Figura 3.2.1: Extraída del programa estadístico SPSS.

Esta matriz nos muestra las correlaciones de las variables (Figura 3.2.1). Por ejemplo la variable PCP con la variable BCH tiene una correlación de 0.761, la variable PCP con la variable GM tiene una correlación de 0.645, y si observamos todas las correlaciones con cada una de las variables, vemos que algunas son altas, por lo que sí podemos realizar un análisis factorial.

Para que el análisis sea fructífero es conveniente que la matriz contenga grupos de variables que correlacionen fuertemente entre sí. Una matriz de correlaciones próxima a una matriz identidad indica que el análisis factorial conducirá a una solución deficiente.

*Determinante* = 0,25. Como se aproxima a 0 pero no llega a serlo, esto significa que las variables están linealmente asociadas y que es un buen síntoma de cara a la idoneidad al análisis.

La siguiente tabla (Figura 3.2.2) muestra, para cada una de las variables incluidas en el análisis, algunos estadísticos descriptivos univariados: la media, la desviación típica y el número de casos válidos para el análisis.

Como el análisis se basa en la matriz de correlaciones, las diferencias de escala y de variabilidad entre las variables carecen de relevancia.

**Estadísticos descriptivos**

	Media	Desviación típica	N del análisis
INF	,22357	,091553	98
PRI	,41301	,140521	98
EDE	,00232	,003543	98
ESO	,24539	,090919	98
PCP	,00462	,008572	98
BCH	,04421	,062920	98
GM	,01339	,027488	98
GS	,00555	,014649	98
ESA	,00459	,019552	98
ERE	,01275	,040174	98

Figura 3.2.2: Extraída del programa estadístico SPSS.

En cuanto a la *dependencia* utilizamos el test de Bartlett para realizar un test preliminar de completa independencia o esfericidad de las variables (Figura 3.2.3).

### KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.	,781
Prueba de esfericidad de Chi-cuadrado aproximado	342,796
Bartlett gl	45
Sig.	,000

Figura 3.2.3: Extraída del programa estadístico SPSS.

Muestra un p-valor significativo, por lo que rechaza la independencia de las variables.

Este test indica la presencia de correlaciones no nulas.

Cuanto más se aproxime a 1 el KMO supone que los datos se aproximan a una matriz diagonal, en nuestro caso tenemos un KMO por muy cercano a 0.8 por lo que se considera satisfactorio ya que está cercano a 1. Ahora vamos a calcular, basándonos también en el KMO, las medidas de adecuación muestral individuales para cada variable. Para ello, en la matriz de correlaciones anti-imagen, en cuya diagonal aparecen los MSA, un valor de MSA próximo a 1 indicara que la variable es adecuada para su tratamiento en el análisis factorial con el resto de variables.

La siguiente tabla ofrece la matriz de varianzas-covarianzas anti-imagen y la matriz de correlaciones anti-imagen (Figura 3.2.4). La matriz de correlaciones anti-imagen se utiliza como diagnóstico de la adecuación de los datos a un modelo factorial.

Matrices anti-imagen											
		INF	PRI	EDE	ESO	PCP	BCH	GM	GS	ESA	ERE
Covarianza anti-imagen	INF	,851	-,077	-,042	,090	,021	,039	,018	,002	,028	,041
	PRI	-,077	,640	-,062	-,052	,002	,124	,038	,031	-,001	,043
	EDE	-,042	-,062	,980	-,016	-,006	-,044	,033	-,045	-,001	,019
	ESO	,090	-,052	-,016	,882	,012	-,109	-,002	,014	,124	,095
	PCP	,021	,002	-,006	,012	,373	-,188	-,109	,015	,033	,029
	BCH	,039	,124	-,044	-,109	-,188	,308	-,041	-,008	-,049	-,059
	GM	,018	,038	,033	-,002	-,109	-,041	,347	-,137	-,214	,052
	GS	,002	,031	-,045	,014	,015	-,008	-,137	,448	,091	-,279
	ESA	,028	-,001	-,001	,124	,033	-,049	-,214	,091	,579	-,062
	ERE	,041	,043	-,019	,095	,029	-,059	,052	-,279	-,062	,504
Correlación anti-imagen	INF	,930 <sup>a</sup>	-,104	-,046	,104	,038	,076	,033	,003	,040	,063
	PRI	-,104	,910 <sup>a</sup>	-,078	-,070	,003	,279	,081	,058	-,001	,076
	EDE	-,046	-,078	,314 <sup>a</sup>	-,017	-,009	-,080	,057	-,067	-,002	,026
	ESO	,104	-,070	-,017	,433 <sup>a</sup>	,021	-,209	-,004	,022	,174	,142
	PCP	,038	,003	-,009	,021	,797 <sup>a</sup>	-,555	-,303	,036	,071	,066
	BCH	,076	,279	-,080	-,209	-,555	,799 <sup>a</sup>	-,124	-,021	-,116	-,151
	GM	,033	,081	,057	-,004	-,303	-,124	,796 <sup>a</sup>	-,349	-,478	,125
	GS	,003	,058	-,067	,022	,036	-,021	-,349	,710 <sup>a</sup>	,179	-,588
	ESA	,040	-,001	-,002	,174	,071	-,116	-,478	,179	,757 <sup>a</sup>	-,114
	ERE	,063	,076	,026	,142	,066	-,151	,125	-,588	-,114	,708 <sup>a</sup>

a. Medida de adecuación muestral

Figura 3.2.4: Extraída del programa estadístico SPSS.

En este contexto, un coeficiente de correlación parcial expresa el grado de relación existente entre dos variables tras eliminar el efecto de las restantes variables incluidas en el análisis.

Cuando las variables incluidas en el análisis comparten gran cantidad de información debido a la presencia de factores comunes, la correlación parcial entre cualquier par de variables debe ser reducida. Por el contrario, cuando dos variables comparten gran cantidad de información entre ellas, pero no la comparten con las restantes variables (ni consecuentemente con los factores comunes), la correlación parcial entre ellas será elevada, siendo esto un mal síntoma de cara a la idoneidad del análisis.

Como los elementos fuera de la diagonal de la matriz de correlaciones anti-imagen que presentan nuestros datos son pequeños, el modelo factorial es adecuado, debido a que un coeficiente de correlación parcial expresa el grado de relación lineal existente entre dos variables tras eliminar el efecto que sobre esta relación tiene el resto de variables. Respecto a la diagonal donde aparecen las medidas de adecuación muestral individuales encontramos que ninguna variable se encuentra con un MSA menor que 0.5, excepto dos que tienen un MSA de 0.314 y otra con un MSA de 0.433. Habrá que tener especial cuidado con esas dos.

La siguiente tabla (Figura 3.2.5) contiene las comunalidades asignadas inicialmente a las variables (inicial) y las comunalidades reproducidas por la solución factorial (extracción). La comunalidad de una variable es la proporción de su varianza que puede ser explicada por el modelo factorial obtenido. Estudiando las comunalidades de la extracción podemos valorar cuáles de las variables son peor explicadas por el modelo. En nuestro estudio, hemos utilizado el método de extracción de componentes principales, y obtenemos que la variable peor explicada es INF. En una nota a pie de tabla se indica que, para llegar a esta solución factorial, se ha utilizado un método de extracción denominado componentes principales. Dicho método de extracción, que es el que actúa por defecto, asume que es posible explicar el 100% de la varianza observada y, por ello, todas las comunalidades iniciales son iguales a la unidad (que es justamente la varianza de una variable en puntuaciones típicas).

Un valor moderado de las comunalidades está entre 0.4 y 0.7. Si una variable tiene una comunalidad inferior a 0.4 puede ser por dos causas:

- a) No está relacionada con otras variables
- b) Sugiere que un factor adicional debiera ser explorado

Nuestras comunalidades muestran un valor inferior a 0.4 que tras la rotación serán estudiados para eliminarlos o para incluirlos en un nuevo factor.

Respecto a las variables que habían presentado un valor límite en el MSA, muestra el valor más pequeño de las comunalidades, además tras la extracción su valor tampoco mejora.

### Comunalidades

	Inicial	Extracción
INF	1,000	,280
PRI	1,000	,475
EDE	1,000	,726
ESO	1,000	,642
PCP	1,000	,697
BCH	1,000	,787
GM	1,000	,716
GS	1,000	,716
ESA	1,000	,483
ERE	1,000	,708

Método de extracción: Análisis de Componentes principales.

Figura 3.2.5: Extraída del programa estadístico SPSS.

A continuación observamos la varianza total explicada.

Componente	Varianza total explicada								
	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	3,907	39,067	39,067	3,907	39,067	39,067	3,386	33,859	33,859
2	1,279	12,791	51,858	1,279	12,791	51,858	1,784	17,838	51,697
3	1,045	10,448	62,306	1,045	10,448	62,306	1,061	10,609	62,306
4	,976	9,764	72,070						
5	,781	7,806	79,876						
6	,664	6,640	86,515						
7	,533	5,329	91,845						
8	,388	3,879	95,724						
9	,224	2,244	97,968						
10	,203	2,032	100,000						

Método de extracción: Análisis de Componentes principales.

Figura 3.2.6: Extraída del programa estadístico SPSS.

En la tabla de porcentajes de varianza explicada (Figura 3.2.6) se ofrece un listado de los autovalores de la matriz de varianzas-covarianzas y del porcentaje de varianza que representa cada uno de ellos. Los autovalores expresan la cantidad de la varianza total que está explicada por cada factor, y los porcentajes de varianza explicada asociados a cada factor se obtienen dividiendo su correspondiente autovalor por la suma de los autovalores (la cual coincide con el número de variables). Por defecto, se extraen tantos factores como autovalores mayores que 1 tiene la matriz analizada.

De acuerdo al primer criterio (varianza total), seleccionamos los autovalores hasta lograr explicar al menos un 80% de la variabilidad total. El número de componentes a retener serían 3, llegando a explicar un 63.306%. La primera componente logra explicar un 33.859%, la segunda un 17.838% y la tercera un 10.609%, entre las tres explican un 62.306% de la variabilidad total.

Según el criterio de Kaiser, retendremos aquellos autovalores que presenten un valor por encima de 1, coincidiendo con el primer criterio retendríamos 3 factores.

La tabla muestra también, para cada factor con autovalor mayor que 1, la suma de las saturaciones al cuadrado. Las sumas de cuadrados de la columna Total (que coinciden con los autovalores cuando se utilizan otros métodos de extracción), pueden ayudarnos, a determinar el número idóneo de factores.

La información de esta tabla puede utilizarse para tomar una decisión sobre el número idóneo de factores que deben extraerse.

Tal y como muestra la columna de porcentajes acumulados (% acumulado), con los 10 factores que es posible extraer se consigue explicar el 100% de la varianza total, pero con ello no se consigue el objetivo de reducir el número de dimensiones necesarias para explicar los datos.

Según el criterio de Kaiser, retendremos aquellos autovalores que presenten un valor por encima de 1, coincidiendo con el primer criterio retendríamos 3 factores.

Según el tercer criterio, observamos el gráfico de sedimentación (Figura 3.2.7).

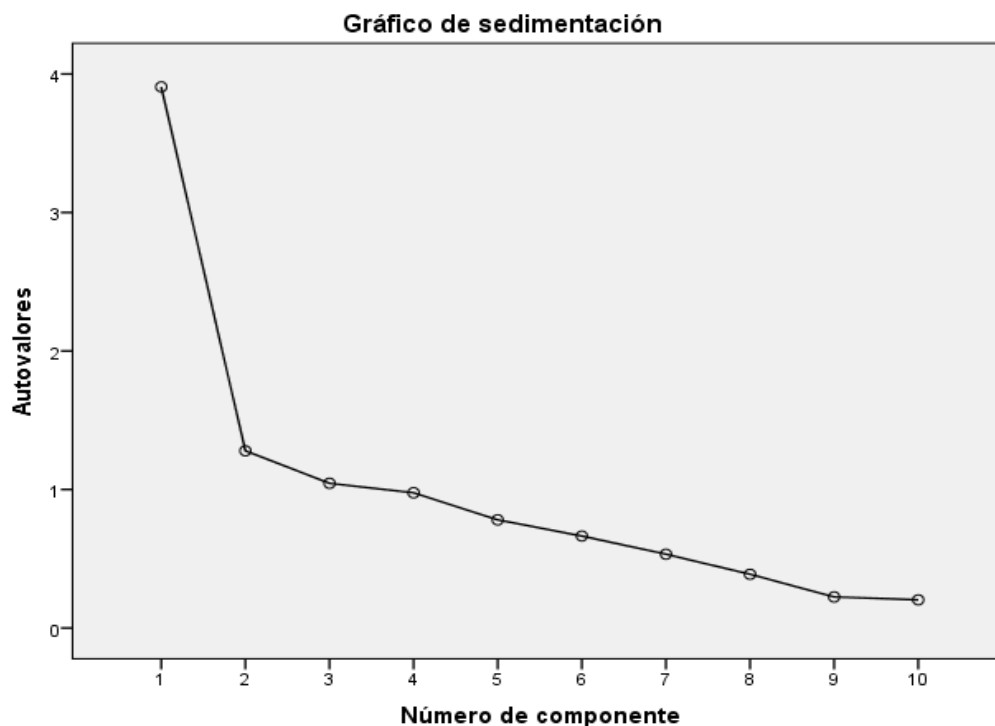


Figura 3.2.7: Extraída del programa estadístico SPSS.

El gráfico anterior es el gráfico de sedimentación, que sirve para determinar el número óptimo de factores. Originalmente propuesto por Cattell (1966), consiste simplemente en una representación gráfica del tamaño de los autovalores.

Según hemos señalado ya, un autovalor indica la cantidad de varianza explicada por una componente principal. Tanto la tabla de porcentajes de varianza explicada como el gráfico de sedimentación ofrecen los autovalores ordenados de mayor a menor: el primer autovalor es el mayor de los posibles, el segundo autovalor es el segundo mayor, y así sucesivamente. Si un autovalor se aproxima a cero, esto significa que el factor correspondiente a ese autovalor es incapaz de explicar una cantidad relevante de la varianza total. Por tanto, un factor al que corresponde un autovalor próximo a cero se considera un factor residual y carente de sentido en el análisis.

Al representar todos los autovalores según tamaño, es posible formarse muy rápidamente una idea sobre si la cantidad de varianza asociada a cada uno de ellos es relevante para el análisis o si por el contrario se trata solo de varianza residual. Los autovalores residuales se encuentran en la parte derecha del gráfico, formando una planicie de apenas inclinación, frente a la fuerte pendiente formada por el autovalor que explica la mayor parte de la varianza disponible. Por ello, es conveniente inspeccionar el gráfico de sedimentación de izquierda a derecha, buscando el punto de inflexión en el que los autovalores dejan de formar una pendiente significativa y comienzan a describir una caída de muy poca inclinación.

Aplicando el análisis de factorización de ejes principales obtenemos los siguientes resultados:

**Comunalidades**

	Inicial	Extracción
INF	,149	,156
PRI	,360	,366
EDE	,020	,006
ESO	,118	,157
PCP	,627	,663
BCH	,692	,854
GM	,653	,723
GS	,552	,653
ESA	,421	,625
ERE	,496	,704

Método de extracción:

Factorización de Ejes principales.

*Figura 3.2.8: Extraída del programa estadístico SPSS.*

**Varianza total explicada**

Factor	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	3,907	39,067	39,067	3,560	35,601	35,601	2,958	29,584	29,584
2	1,279	12,791	51,858	,848	8,485	44,085	1,434	14,339	43,923
3	1,045	10,448	62,306	,498	4,982	49,068	,514	5,145	49,068
4	,976	9,764	72,070						
5	,781	7,806	79,876						
6	,664	6,640	86,515						
7	,533	5,329	91,845						
8	,388	3,879	95,724						
9	,224	2,244	97,968						
10	,203	2,032	100,000						

Método de extracción: Factorización de Ejes principales.

Figura 3.2.9: Extraída del programa estadístico SPSS.

Se puede observar (Figura 3.2.8) que en este caso las variables peor explicadas serían INF, PRI, EDE y ESO, y la tabla de la varianza total explicada (Figura 3.2.9) nos muestra que si extraemos los tres primeros factores llegaríamos a explicar el 49.068% de la varianza.

En la próxima tabla (Figura 3.2.10) se encuentra la solución factorial propiamente dicha. Contiene las correlaciones entre las variables originales (o saturaciones) y cada uno de los factores. Conviene señalar que esta matriz cambia de denominación dependiendo del método de extracción elegido. En este caso se denomina matriz de componentes porque en nuestro estudio hemos utilizado el método de componentes principales como método de extracción (es el método que actúa por defecto).

**Matriz de componentes<sup>a</sup>**

	Componente		
	1	2	3
INF	-,461	-,251	,071
PRI	-,685	-,026	,074
EDE	,011	,057	,850
ESO	,002	,756	,265
PCP	,774	,314	-,021
BCH	,845	,266	,051
GM	,836	,050	-,121
GS	,672	-,429	,283
ESA	,626	-,055	-,297
ERE	,609	-,531	,235

Método de extracción: Análisis de componentes principales.

a. 3 componentes extraídos

Figura 3.2.10: Extraída del programa estadístico SPSS.

Comparando las saturaciones relativas de cada variable en cada factor, podemos apreciar que extrae 3 componentes. Las variables INF, PRI, PCP, BCH, GM, GS, ESA y ERE saturan en un único factor porque constituyen un mismo grupo de variables dentro de la matriz de correlaciones. La variable ESO satura en el segundo componente y la variable EDE satura en el tercer y último componente.

Vamos a estudiar ahora la *matriz de correlaciones reproducidas*:

Se obtiene a partir de la solución factorial calculada.

La matriz de correlaciones reproducidas contiene las correlaciones que es posible reproducir utilizando tan solo la información contenida en la solución factorial. En concreto, la matriz reproducida se obtiene post-multiplicando la matriz factorial por su traspuesta.

En la diagonal (Figura 3.2.11) se encuentran las comunalidades finales, la cuales deben estar por encima de 0.5 aunque en nuestro caso tenemos algunas variables que presentan un valor por debajo de este límite, aunque no muy alejadas, por lo que deben de ser estudiadas.

Correlaciones reproducidas											
	INF	PRI	EDE	ESO	PCP	BCH	GM	GS	ESA	ERE	
Correlación reproducida	INF	,280 <sup>a</sup>	,327	,041	-,172	-,437	-,453	-,406	-,182	-,296	-,131
	PRI	,327	,475 <sup>a</sup>	,054	-,002	-,539	-,582	-,583	-,428	-,449	-,386
	EDE	,041	,054	,726 <sup>a</sup>	,268	,009	,068	-,091	,224	-,249	,176
	ESO	-,172	-,002	,268	,642 <sup>a</sup>	,233	,217	,008	-,248	-,119	-,338
	PCP	-,437	-,539	,009	,233	,697 <sup>a</sup>	,736	,665	,379	,473	,300
	BCH	-,453	-,582	,068	,217	,736	,787 <sup>a</sup>	,714	,468	,499	,385
	GM	-,406	-,583	-,091	,008	,665	,714	,716 <sup>a</sup>	,506	,557	,454
	GS	-,182	-,428	,224	-,248	,379	,468	,506	,716 <sup>a</sup>	,360	,704
	ESA	-,296	-,449	-,249	-,119	,473	,499	,557	,360	,483 <sup>a</sup>	,341
	ERE	-,131	-,386	,176	-,338	,300	,385	,454	,704	,341	,708 <sup>a</sup>
Residual <sup>b</sup>	INF		-,041	-,004	,068	,146	,115	,119	-,024	,089	-,076
	PRI	-,041		-,009	,014	,095	,026	,120	,056	,138	,034
	EDE	-,004	-,009		-,234	,022	-,015	,076	-,180	,233	-,168
	ESO	,068	,014	-,234		-,141	-,073	-,016	,171	-,011	,194
	PCP	,146	,095	,022	-,141		,025	-,020	-,041	-,085	-,041
	BCH	,115	,026	-,015	-,073	,025		-,069	-,047	-,059	-,002
	GM	,119	,120	,076	-,016	-,020	-,069		,001	,055	-,103
	GS	-,024	,056	-,180	,171	-,041	-,047	,001		-,115	-,033
	ESA	,089	,138	,233	-,011	-,085	-,059	,055	-,115		-,074
	ERE	-,076	,034	-,168	,194	-,041	-,002	-,103	-,033	-,074	

Método de extracción: Análisis de Componentes principales.

a. Comunalidades reproducidas

b. Los residuos se calculan entre las correlaciones observadas y reproducidas. Hay 26 (57,0%) residuales no redundantes con valores absolutos mayores que 0,05.

Figura 3.2.11: Extraída del programa estadístico SPSS.

Además la tabla anterior también contiene la matriz residual, es decir, los residuos del análisis factorial. Cada residuo expresa la diferencia existente entre la correlación observada entre dos variables y la correlación reproducida por la estructura factorial para esas dos variables. Si el

análisis ha sido fructífero, la mayoría de las correlaciones reproducidas se parecerán a las correlaciones observadas y los residuos serán muy pequeños, como es nuestro caso. La tabla incluye una nota a pie de tabla que contabiliza el número de residuos mayores que 0.05 y el porcentaje que ese número representa sobre el total de correlaciones no redundantes de la matriz, en nuestro caso abarca un 57%.

A continuación podemos proceder a la rotación de los factores para hacer una mejor interpretación:

Vamos a aplicar el método de rotación Varimax utilizando el análisis de componentes principales (Figura 3.2.12).

**Matriz de componentes rotados<sup>a</sup>**

	Componente		
	1	2	3
INF	-,529	,030	-,005
PRI	-,634	-,265	,046
EDE	-,078	,130	,838
ESO	,287	-,604	,440
PCP	,829	,064	,078
BCH	,863	,152	,138
GM	,787	,301	-,081
GS	,383	,730	,190
ESA	,577	,264	-,284
ERE	,290	,781	,117

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Varimax con Kaiser.

a. La rotación ha convergido en 6 iteraciones.

*Figura 3.2.12: Extraída del programa estadístico SPSS.*

Como todas nuestras variables muestran una complejidad 1, esto es síntoma de una buena elección ya que hay diversas cargas grandes en cada factor, además, es un motivo para no dudar de la existencia de una estructura factorial en nuestros datos.

Nuestros factores son fuertes y estables ya que son deseables 5 o más variables con cargas de 0.5 para ser esto un indicativo característico de un factor sólido.

Se puede ver de una forma más clara con el gráfico de componentes de factores rotados (Figura 3.2.13) y en la matriz de transformación de las componentes (Figura 3.2.14).

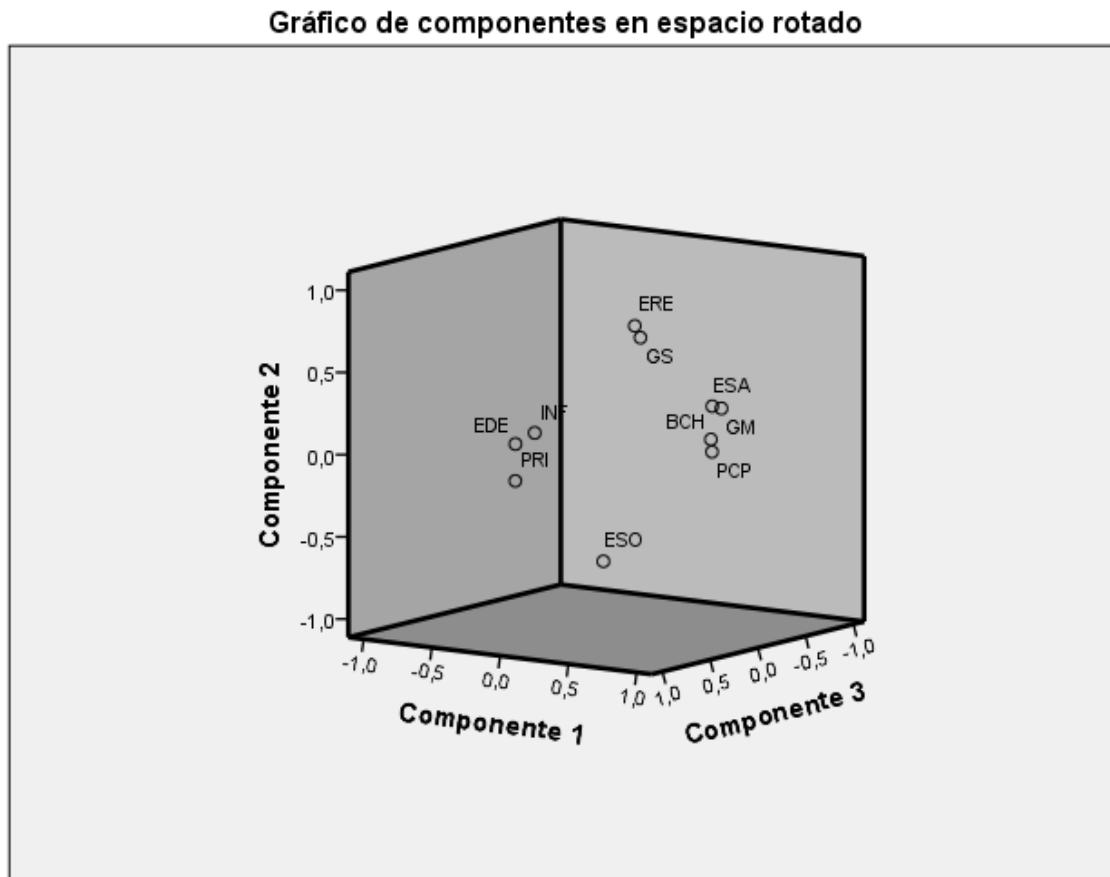


Figura 3.2.13: Extraída del programa estadístico SPSS.

**Matriz de transformación de las componentes**

Componente	1	2	3
1	,896	,443	,029
2	,423	-,873	,243
3	-,132	,206	,970

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Varimax con Kaiser.

Figura 3.2.14: Extraída del programa estadístico SPSS.

Ahora vamos a proceder con el método de rotación Equamax (Figura 3.2.15) que trata de maximizar una media ponderada de los criterios Varimax y Quartimax.

**Matriz de factores rotados<sup>a</sup>**

	Factor		
	1	2	3
INF	-,346	-,165	,096
PRI	-,505	-,327	,059
EDE	,002	,019	-,072
ESO	,113	-,160	-,345
PCP	,779	,168	-,165
BCH	,841	,281	-,261
GM	,768	,306	,200
GS	,274	,760	,002
ESA	,604	,142	,490
ERE	,172	,820	,053

Método de extracción: Factorización del eje principal.

Método de rotación: Normalización Equamax con Kaiser.

a. La rotación ha convergido en 5 iteraciones.

Figura 3.2.15: Extraída del programa estadístico SPSS.

Ambos métodos de rotación ortogonal coinciden en la composición de las cargas de los factores por lo que es un buen síntoma de que hemos realizado un fructífero análisis factorial.

De igual modo se puede observar más claramente en la matriz de los factores ya rotados (Figura 3.2.16) y el gráfico de factor en espacio total rotado (Figura 3.2.17).

**Matriz de transformación de los factores**

Factor	1	2	3
1	,840	,541	,019
2	-,532	,819	,214
3	-,101	,190	-,977

Método de extracción: Factorización del eje principal.

Método de rotación: Normalización Equamax con Kaiser.

Figura 3.2.16: Extraída del programa estadístico SPSS.

Gráfico de factor en espacio factorial rotado

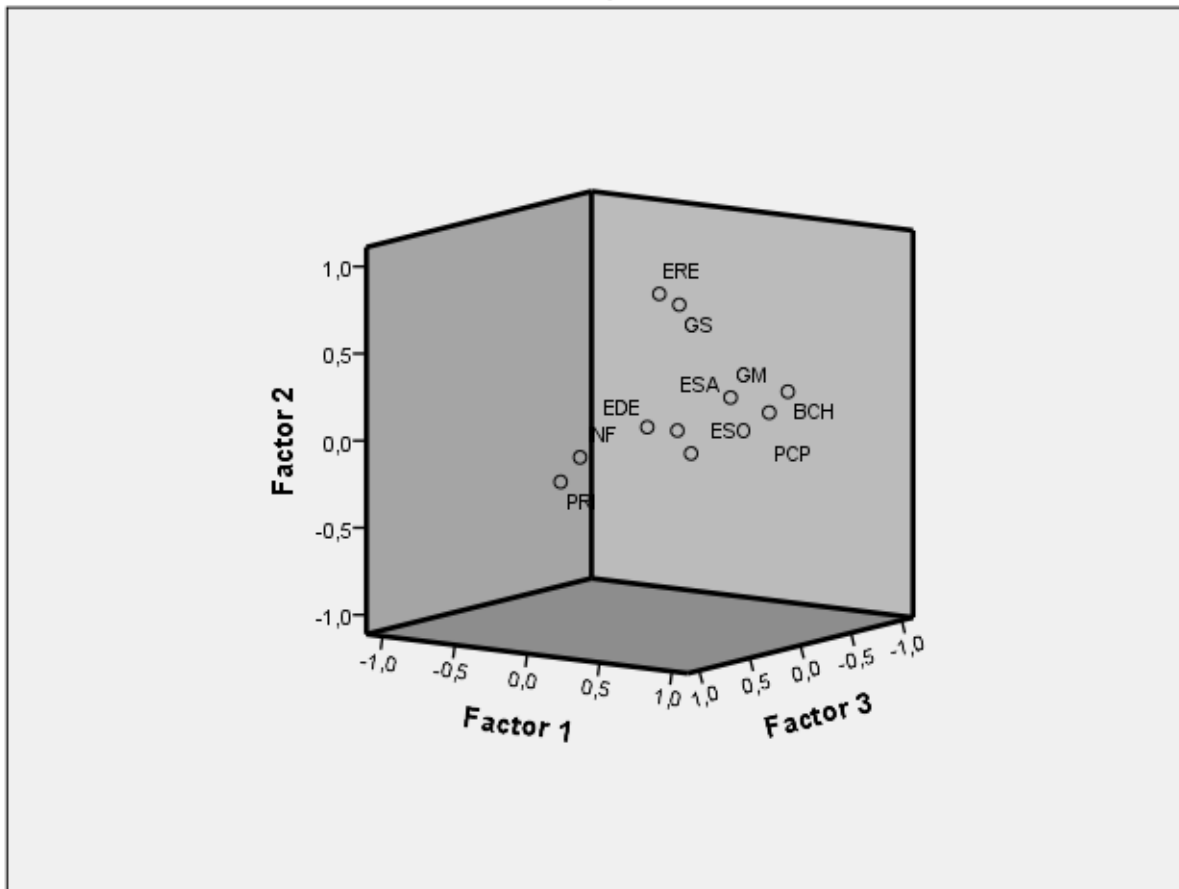


Figura 3.2.17: Extraída del programa estadístico SPSS.

En conclusión del análisis factorial obtenemos, después de rotar los factores, podemos apreciar tres factores. Las variables INF, PRI, PCP, BCH, GM y ESA cargan en el primer factor. Las variables GS y ERE saturan en el segundo factor y las variables EDE y ESO constituyen el tercer y último factor.

Para verificar que efectivamente las variables pueden ser agrupadas vamos a realizar un análisis cluster.

Aplicamos este análisis debido a que esta técnica nos va a ayudar a encontrar un agrupamiento de las variables que refleje posibles similitudes entre subconjuntos de ellas y que sean diferentes entre los distintos subgrupos.

Se desea encontrar un agrupamiento de las variables que refleje posibles similitudes entre subconjuntos de ellas.

En primer lugar vamos a hacer un análisis de distancias y para ello hemos obtenido la matriz de distancias (Figura 3.2.18).

**Matriz de Distancias**

INF	PRI	EDE	ESO	PCP	BCH	GM	GS	ESA	
PRI	5.527684								
EDE	5.611616	18.440564							
ESO	1.830842	5.440414	6.593206						
PCP	5.562277	18.365701	0.008819	6.474297					
BCH	4.727699	16.576585	0.556181	4.992129	0.464758				
GM	5.356267	17.981339	0.086851	6.152835	0.058048	0.334152			
GS	5.545587	18.349793	0.022725	6.478635	0.019700	0.476140	0.060590		
ESA	5.622009	18.461321	0.039101	6.565715	0.031664	0.470024	0.053976	0.044296	
ERE	5.473934	18.153854	0.168516	6.364272	0.152769	0.450043	0.155011	0.105625	0.159619

Figura 3.2.18: Extraída del programa estadístico R Commander.

La matriz de distancias expresa el distanciamiento entre pares de casos. La medida de distancia seleccionada es la distancia euclídea al cuadrado.

Observamos en la matriz cada una de las distancias de cada variable, por ejemplo entre las distancias más altas tenemos a las variables EDE y PRI con una distancia de 18.440, entre ESA y PRI existe también una de las distancias más altas que es 18.461, así como con todos los pares de variables que contengan la variable PRI, excepto el caso de la distancia entre ESO y PRI, que la distancia es menor. Las distancias más pequeñas entre las distintas variables son entre GS y BCH, ESA y BCH, GS y EDE... Aunque la distancia más pequeña es entre PCP y EDE con una distancia de 0.008, muy cercana a 0.

En el siguiente paso realizamos una tabla de agrupamiento, utilizando el método “Ward” (Figura 3.2.19).

### Tabla de agrupamiento “Ward”

[,1] [,2]
[1,] -3 -5
[2,] -8 1
[3,] -9 2
[4,] -7 3
[5,] -10 4
[6,] -6 5
[7,] -1 -4
[8,] -2 7
[9,] 6 8

Figura 3.2.19: Extraída del programa estadístico R Commander.

Según la tabla anterior, en el paso 1, se unen los cluster formados por una sola observación 3 y 5. Los signos negativos indican que cuando se unen está formados por una sola observación cada uno. En paso 2, se unen los cluster 8 y 1. En el paso 3 se unen los cluster 9 y 2. En el paso 4, se unen los cluster 7 y 3, así sucesivamente. Se unirán unos a otros cuando encajen en la numeración, por ejemplo, un cluster 1 y 2 se puede unir a un cluster formado por las observaciones 3 y 4.

A continuación se muestran los resultados utilizando el método “Average”:

### Tabla de agrupamiento “Average”

[,1] [,2]
[1,] -3 -5
[2,] -8 1
[3,] -9 2
[4,] -7 3
[5,] -10 4
[6,] -6 5
[7,] -1 -4
[8,] -2 7
[9,] 6 8

Figura 3.2.20: Extraída del programa estadístico R Commander.

Vemos que se obtienen los mismos resultados con el método Ward que con el método Average (Figura 3.2.20).

A continuación se aplica el análisis cluster y se obtiene el dendrograma (Figura 3.2.21).

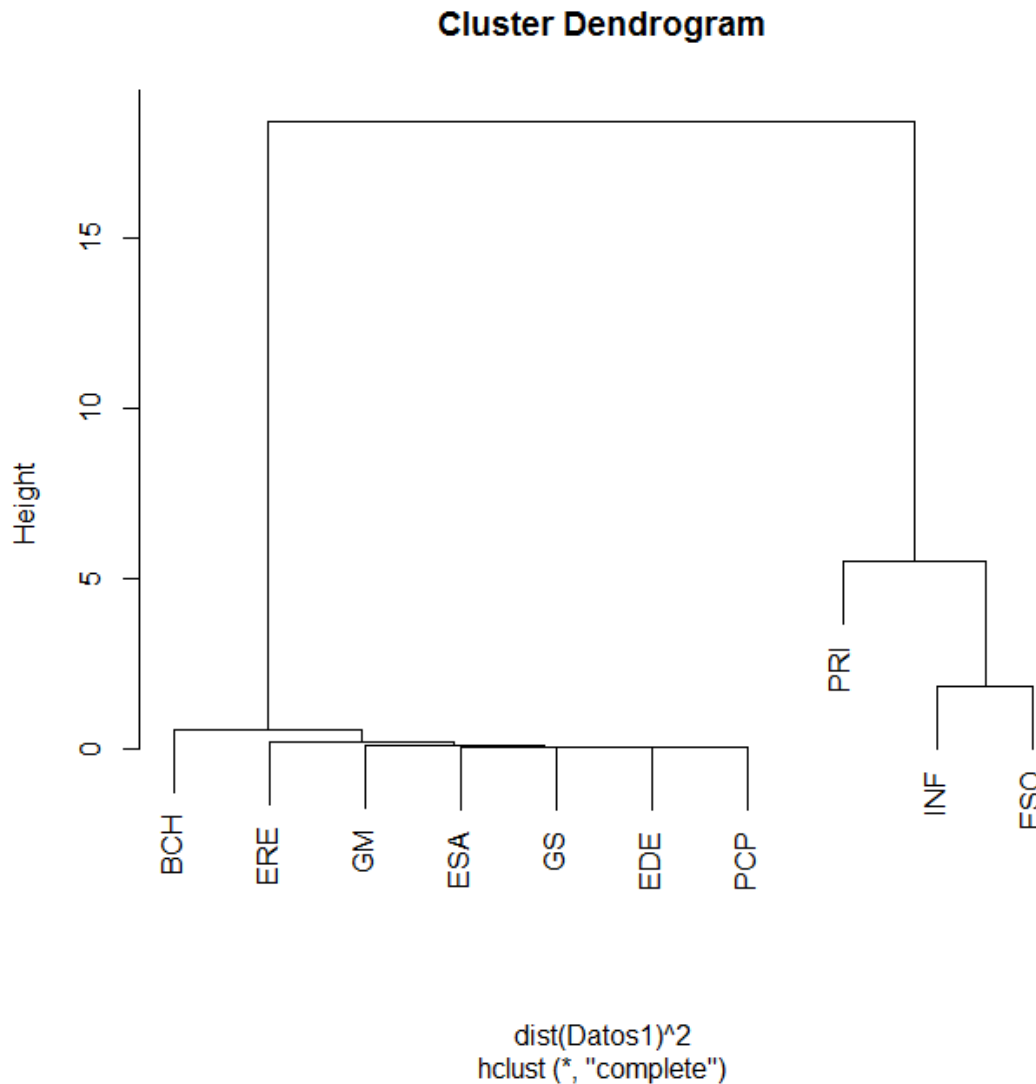


Figura 3.2.21: Extraída del programa estadístico R Commander.

El dendrograma es una valiosa herramienta visual que puede ayudar a decidir el número de grupos que podrían representar mejor la estructura de los datos teniendo en cuenta la forma en que se van anidando los cluster y la medida de similitud a la cual lo hacen.

Cortando el gráfico con un segmento perpendicular a las ramas se obtiene una partición con un número de grupos igual a las ramas “cortadas”.

El número de cluster o clases en que deseamos agrupar los datos nos llevará “cortar” el dendrograma del ejemplo verticalmente y ver a qué nivel de similitud se da dicho agrupamiento. En el ejemplo puede apreciarse un salto importante en la longitud de las líneas verticales que definen los grupos en 2. Cuando se observa en la escala vertical un amplio rango sin existencia de agrupamiento puede ser un indicio de que los cluster se encuentran separados a esos niveles de similitud.

### Bannerplot

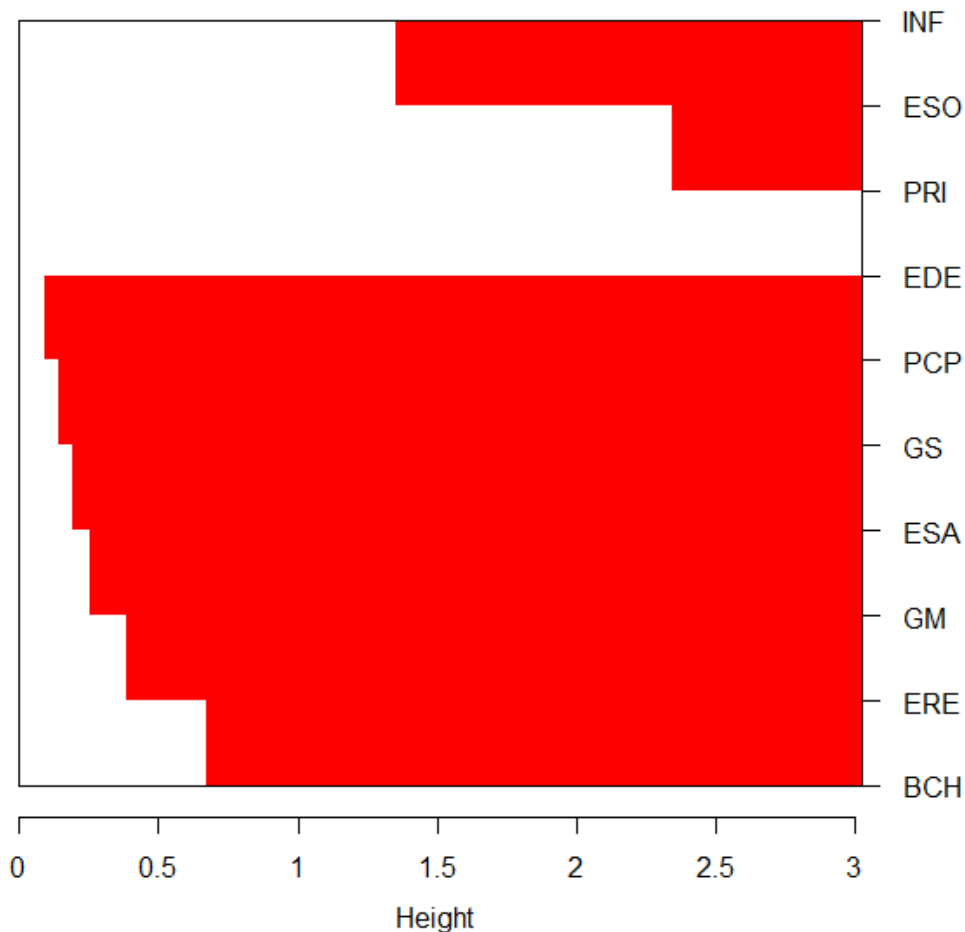


Figura 3.2.22: Extraída del programa estadístico R Commander.

Según el dendrograma y el bannerplot (Figura 3.2.22), si decidimos establecer 2 cluster, estarán formados por los siguientes casos:

Cluster 1: Alumnos en centros públicos por niveles educativos de infantil, primaria y educación secundaria obligatoria.

Cluster 2: Alumnos en centros públicos por niveles educativos de educación especial, programas de cualificación profesional inicial, bachillerato, grado medio, grado superior, enseñanza secundaria de adultos y enseñanzas de régimen especial.

Por lo que este análisis establece dos grupos de variables y el análisis factorial nos proporciona una división de tres componentes o factores. Lo que significa que tenemos dos opciones de agrupamiento.

### 3.3 PARO POR NIVEL DE ACTIVIDAD

Aplicamos el método de análisis de correspondencias a los datos recogidos del paro por nivel de actividad en los municipios de Jaén.

Aplicamos este análisis debido a que esta técnica nos va a ayudar a resumir una gran cantidad de datos en un número reducido de dimensiones, con la menor pérdida de información posible.

En este análisis contamos con diez variables que son:

AGCS: Agricultura, ganadería, caza y selvicultura.

P: Pesca

IE: Industrias extractivas

IM: Industria manufacturera

PYD: Producción y distribución de energía eléctrica, gas y agua

C: Construcción

COM: Comercio; reparación de vehículos de motor, motocicletas y ciclomotores y artículos personales y de uso doméstico

H: Hostelería

TAC: Transporte, almacenamiento y comunicaciones

IF: Intermediación financiera

AIA: Actividades inmobiliarias y de alquiler; servicios empresariales

ADM: Administración pública, defensa y seguridad social obligatoria

ED: Educación

AS: Actividades sanitarias y veterinarias, servicio social

OA: Otras actividades sociales y de servicios prestados a la comunidad; servicios personales

AH: Actividades de los hogares

OE: Organismos extraterritoriales

SE: Sin empleo anterior

Los resultados obtenidos al realizar el análisis de correspondencias mediante R Commander son los que se muestran en la siguiente figura: (Figura 3.3.1).

**Puntuaciones y contribuciones de los perfiles por filas y columnas**

Principal inertias (eigenvalues):																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Value	0.449291	0.088152	0.070055	0.060033	0.033086	0.024983	0.01833	0.00682	0.004375	0.003691	0.002577	0.001999	0.001434	0.00128	0.000427	0.000239 4e-06
Percentage	58.59%	11.5%	9.14%	7.83%	4.31%	3.26%	2.39%	0.89%	0.57%	0.48%	0.34%	0.26%	0.19%	0.17%	0.06%	0.03% 0%
Rows:																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Mass	0.113928	0.000880	0.024977	0.011399	0.000461	0.056199	0.003981	0.002179	0.001048	0.024684	0.029545	0.001844	0.009262	0.002179	0.003897	0.004400
ChiDist	1.834797	0.889763	0.354695	0.689685	1.387073	0.386495	0.554293	0.511026	1.210686	0.407138	0.549882	0.561147	0.921863	0.876626	0.627266	1.541444
Inertia	0.383537	0.000697	0.003142	0.005422	0.000887	0.008395	0.001223	0.000569	0.001536	0.004092	0.008934	0.000581	0.007871	0.001675	0.001534	0.010456
Dim. 1	-2.728442	0.479407	0.251534	0.594985	0.575976	0.500928	0.316906	0.501822	0.546599	0.525057	0.472698	0.575910	0.338783	0.532642	0.426581	0.792605
Dim. 2	0.422137	-0.328498	-0.050436	0.813431	1.492121	0.177093	1.007185	0.645000	0.078866	0.378100	0.965923	0.074305	0.113118	1.184390	-0.061174	1.488849
	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Mass	0.000587	0.003227	0.004023	0.001634	0.002347	0.001257	0.001467	0.049200	0.003772	0.003017	0.000671	0.010184	0.000461	0.000629	0.000754	0.000880

ChiDist 0.845671 0.933880 0.789839 0.818968 0.600999 0.832404 1.375831 0.769932 0.781836  
 0.558931 0.843151 0.681962 0.937718 0.899636 1.311811 0.754423  
 Inertia 0.000420 0.002814 0.002510 0.001096 0.000848 0.000871 0.002777 0.029166 0.002306  
 0.000943 0.000477 0.004736 0.000405 0.000509 0.001298 0.000501  
 Dim. 1 0.244562 0.680832 0.594691 0.385845 0.539829 0.707990 -0.513126 0.630533 0.656260  
 0.633194 0.627252 0.338462 0.521019 0.318283 0.616177 0.668795  
 Dim. 2 -0.957731 0.616293 -0.540867 0.241789 -0.749461 -0.127290 0.234174 0.907505 -0.395808  
 0.771450 0.573616 -0.065427 0.724840 -0.403456 -0.994963 1.163592  
           33   34   35   36   37   38   39   40   41   42   43   44   45   46  
 47   48  
 Mass 0.001341 0.002766 0.001341 0.000503 0.003395 0.002556 0.000293 0.000377 0.000377  
 0.010561 0.001215 0.004778 0.001760 0.000880 0.001676 0.166045  
 ChiDist 0.642004 0.742832 0.852818 1.050850 0.745860 0.657168 1.174905 1.118978 1.226535  
 0.981796 1.180546 0.674860 0.821639 1.257231 1.382474 0.624550  
 Inertia 0.000553 0.001526 0.000975 0.000555 0.001888 0.001104 0.000405 0.000472 0.000567  
 0.010180 0.001694 0.002176 0.001188 0.001391 0.003204 0.064768  
 Dim. 1 0.277479 0.517310 0.453711 -0.044806 0.348311 0.664013 0.352116 0.021024 -0.004384  
 0.563675 0.604240 0.541767 0.378652 0.672480 0.707707 -0.000524  
 Dim. 2 0.526626 0.863203 -1.052570 0.856205 -0.833122 0.862546 0.286508 0.483154 -1.413428  
 0.947811 -0.303330 -0.038308 -0.244352 0.657518 -1.085685 -2.044034  
           49   50   51   52   53   54   55   56   57   58   59   60   61   62  
 63   64  
 Mass 0.001886 0.000712 0.007837 0.000671 0.000084 0.135657 0.002012 0.001341 0.008759  
 0.006244 0.035915 0.010645 0.001006 0.004065 0.001048 0.002473  
 ChiDist 0.594021 0.889760 0.841149 0.692576 2.075157 0.584365 0.630611 0.893421 0.386246  
 0.553756 0.768621 0.457080 1.957016 1.211100 1.039401 0.814576  
 Inertia 0.000665 0.000564 0.005545 0.000322 0.000361 0.046325 0.000800 0.001070 0.001307  
 0.001915 0.021218 0.002224 0.003852 0.005963 0.001132 0.001641  
 Dim. 1 0.416172 0.234556 0.663649 0.572093 0.760348 0.259095 0.465340 0.379483 0.390146  
 0.572769 0.533095 0.530625 0.244457 0.585204 0.493465 0.643716  
 Dim. 2 -0.132357 0.650057 0.447124 -0.183257 -1.606747 0.754974 0.452597 -0.351535 0.319367  
 1.007711 0.895589 0.022282 0.560182 -0.209868 -1.473833 0.188816  
           65   66   67   68   69   70   71   72   73   74   75   76   77   78  
 79   80

Mass 0.003981 0.003185 0.004275 0.010812 0.001467 0.002347 0.005574 0.005113 0.004275  
 0.002137 0.000671 0.001634 0.004065 0.001425 0.001886 0.002431  
 ChiDist 0.482212 0.675174 0.786390 1.236747 0.759825 0.740336 0.855210 0.557022 0.822611  
 0.698711 0.957949 0.739127 0.873598 0.682797 1.416656 0.570140  
 Inertia 0.000926 0.001452 0.002643 0.016538 0.000847 0.001286 0.004077 0.001586 0.002893  
 0.001043 0.000615 0.000893 0.003102 0.000664 0.003785 0.000790  
 Dim. 1 0.451829 0.405595 0.575981 0.465469 0.379072 0.406933 0.473874 0.674072 0.371574  
 0.679691 0.358638 0.642041 0.605111 0.604317 -0.178918 0.226454  
 Dim. 2 0.161591 0.436683 0.691397 -0.477911 -0.612037 -0.834356 -0.419521 0.216272 0.277279  
 0.146392 -0.720992 0.587213 -0.187457 -0.124562 -1.283478 -0.927951

81 82 83 84 85 86 87 88 89 90 91 92 93 94  
 95 96 97

Mass 0.000838 0.016470 0.002137 0.018104 0.010142 0.001090 0.000545 0.065042 0.002221  
 0.006119 0.009765 0.002305 0.006328 0.000629 0.004987 0.000210 0.002598  
 ChiDist 1.192787 0.393404 0.818648 0.460931 0.949769 0.898862 0.846618 0.320805 0.918673  
 0.503225 0.867757 0.602304 0.545805 1.484064 0.483409 1.110361 0.739725  
 Inertia 0.001192 0.002549 0.001432 0.003846 0.009149 0.000880 0.000390 0.006694 0.001875  
 0.001549 0.007353 0.000836 0.001885 0.001385 0.001165 0.000258 0.001422  
 Dim. 1 0.645525 0.482045 0.420519 0.513104 0.541295 0.327264 0.246316 0.247022 0.566801  
 0.449698 0.555885 0.646060 0.500862 0.146795 0.383670 0.846049 0.487893  
 Dim. 2 -0.648397 0.026605 0.156601 0.305913 -0.569388 1.259053 -0.722200 0.188685 0.836360  
 0.573057 0.299767 0.622902 0.722449 1.779255 -0.221191 0.698984 0.238242

Columns:

AGCS P IE IM PYD C COM H TAC IF AIA ADM ED  
 AS OA AH

Mass 0.087173 0.000251 0.002095 0.145903 0.002766 0.096024 0.102526 0.053528 0.039687  
 0.010938 0.079873 0.084236 0.041280 0.079835 0.045429 0.003395  
 ChiDist 0.751239 1.513474 1.511423 0.715692 1.898277 0.644844 0.569278 0.833458 1.249356  
 1.300657 0.700448 0.714502 1.307421 1.254343 1.255386 1.389651  
 Inertia 0.049197 0.000576 0.004787 0.074734 0.009967 0.039929 0.033226 0.037184 0.061947  
 0.018504 0.039188 0.043003 0.070561 0.125611 0.071595 0.006555

```

Dim. 1  0.609798 -1.974103 -1.616983 0.628380 -1.936273  0.531071  0.409512  0.496245 -
1.815716 -1.854992 0.366493 0.598850 -1.875402 -1.841268 -1.848599 -1.934897
Dim. 2  0.369659 -1.478019  1.062911 0.806487  0.594609 -0.723489 -1.323756 -0.786311
0.158389 -0.174963 -1.818748 0.083956 -0.024152  0.405178 -0.175425 -0.837334

      OE   SE
Mass  0.000084 0.124977
ChiDist 2.244560 0.799020
Inertia 0.000422 0.079790
Dim. 1 -1.850999 0.580955
Dim. 2  1.028654 1.657011

```

Figura 3.3.1: Extraída del programa estadístico R Commander.

En la primera parte que nos da el análisis se muestra que con 16 dimensiones se explican todos los datos o toda la varianza.

En el primer eje, o la primera dimensión, la varianza explicada es del 58.59%, en el segundo eje es de 11.5%, en el tercero 9.14% y así hasta llegar al último eje, la dimensión 16, en el cual la varianza explicada es de 0.03%.

Como se puede observar, con las 4 primeras dimensiones se explica más del 80% de los datos por lo que con un mapa de 4 dimensiones se podría ver todo muy bien.

En los datos anteriores y en la siguiente figura (Figura 3.3.2) se muestran las puntuaciones de los perfiles por filas y columnas así como las contribuciones totales de cada perfil fila a la inercia de cada dimensión y las contribuciones relativas de cada dimensión la inercia del punto.

### Gráfico de puntuaciones y contribuciones de los perfiles por filas y columnas

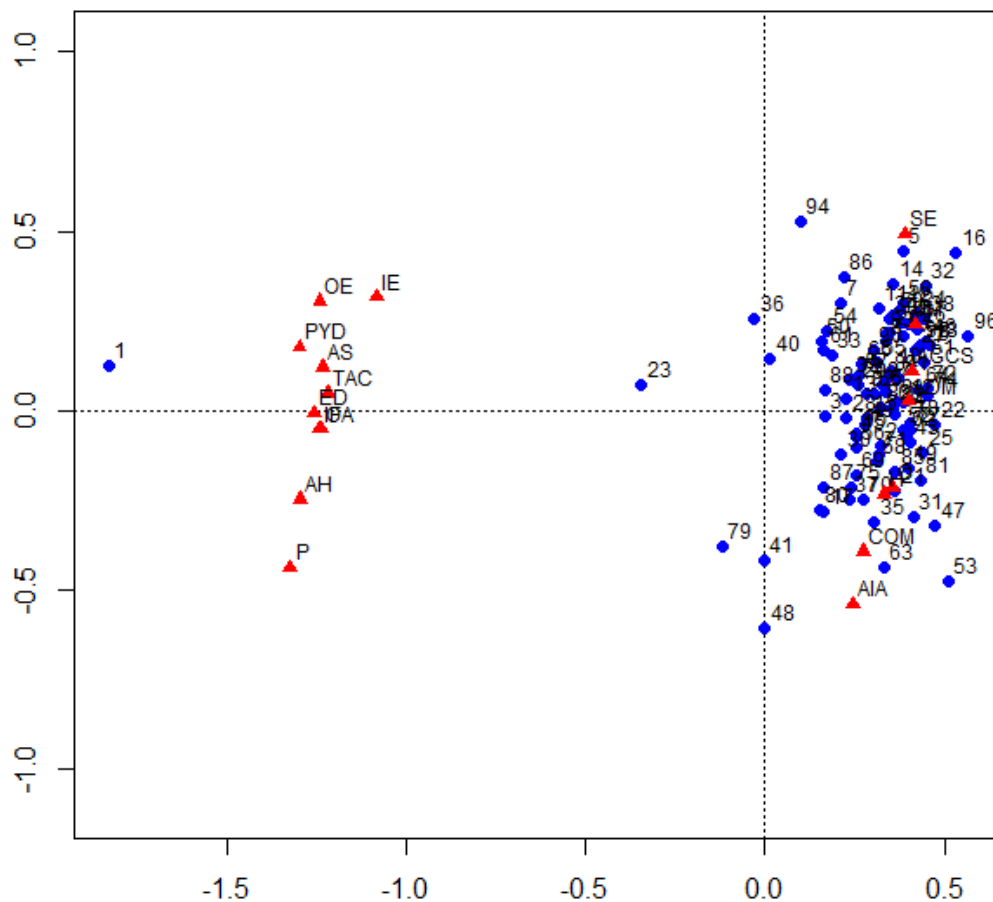


Figura 3.3.2: Extraída del programa estadístico R Commander.

Se observa que la primera dimensión discrimina en función de AH y P. Además, los puntos fila que más contribuyen la inercia de la primera dimensión es P. Dicha dimensión es, junto a la segunda, las que más contribuyen a explicar la inercia de cada uno de dichos puntos.

La segunda dimensión (cuyo poder discriminante es igual a la dimensión anterior, ver la figura 3.3.2) discrimina por OE, IE, PYD; y un poco menos pero también por AS y TAC, aunque ya se encuentran próximos a 0.

La primera dimensión estaría formada por actividades del hogar y pesca.

La segunda dimensión estaría formada por organismos extraterritoriales, industrias extractivas y producción y distribución de energía eléctrica, gas y agua. Así como por actividades sanitarias

y veterinarias, servicio social y transporte, almacenamiento y comunicaciones. Estas últimas con un peso menor.

## 4. RESULTADOS

Los resultados obtenidos en este análisis son:

De los datos que incluyen la tasa de paro por nivel de estudios terminados y sexo podemos decir que las variables se pueden dividir en dos componentes: la primera componente representa el paro registrado por nivel de estudios terminados y sexo en los municipios de Jaén caracterizado por mujeres y hombres analfabetos, mujeres y hombres con estudios primarios, mujeres y hombres con estudios postsecundarios.

La segunda componente se puede definir como el paro registrado por nivel de estudios terminados y sexo en los municipios de Jaén caracterizado por los hombres y mujeres con estudios secundarios.

Utilizando un segundo análisis, obtenemos el mismo agrupamiento.

De los datos analizados que incluyen a los alumnos en centros públicos por nivel educativo obtenemos una clasificación en tres factores:

El primer factor estaría formado por los alumnos en centros públicos en los niveles de infantil, primaria, programas de cualificación profesional, bachillerato, grado medio y educación secundaria para adultos.

El segundo factor estaría formado por los alumnos en centros públicos en los niveles de grado superior y enseñanzas de régimen especial.

El tercer y último factor estaría formado por los alumnos matriculados en centros públicos por los niveles educación especial y educación secundaria obligatoria.

Utilizando un segundo análisis obtenemos una agrupación de dos grupos formados por:

El primer grupo estaría constituido por los alumnos en centros públicos en los niveles de infantil, educación primaria y educación secundaria obligatoria, y el segundo grupo formado por programas de cualificación profesional, bachillerato, grado medio, grado superior, educación secundaria para adultos, enseñanzas de régimen especial y educación especial

Por último, de los datos que contienen al paro registrado en los municipios de Jaén por nivel de actividad obtenemos un agrupamiento en dos dimensiones.

La primera dimensión estaría formada por actividades del hogar y pesca.

La segunda dimensión estaría formada por organismos extraterritoriales, industrias extractivas y producción y distribución de energía eléctrica, gas y agua. Así como por actividades sanitarias y veterinarias, servicio social y transporte, almacenamiento y comunicaciones. Estas últimas con un peso menor.

## 5. CONCLUSIONES

Como conclusiones hemos obtenido que tanto las variables del paro por nivel de estudios terminados y sexo como las variables de los alumnos matriculados en centros públicos por nivel educativo y el paro en función del nivel de actividad en los municipios de Jaén, pueden ser agrupados de modo que te obtenga una forma más sencilla de recoger los datos para un estudio o análisis posterior.

Hemos conseguido llegar a agrupar un número considerable de variables en grupos de 2 en el caso del paro por estudios terminados y sexo y el paro en función del nivel de actividad, y en un grupo de 3, pudiendo reducirse a dos, a las variables de los alumnos matriculados en centros públicos por nivel educativo.

Con este agrupamiento concluimos que el paro en los municipios de Jaén no afecta de igual modo a todos sino que existe una discriminación en función de los estudios terminados separando a las personas con estudios secundarios de las demás con cualquier otro nivel educativo, los estudiantes matriculados en centros públicos se separan en alumnos que realizan estudios obligatorios del resto de alumnos que realicen cualquier otro tipo de estudio, y en el nivel de actividad , el paro no afecta del mismo modo a las personas dedicadas a las labores del hogar y la pesca que el resto de actividades.

Hoy en día con la crisis del país, los municipios de Jaén se ven gravemente influidos por la gran tasa de paro a nivel de estudios terminados, en los alumnos en centros públicos por nivel educativo y en el nivel de actividad.

## 6. BIBLIOGRAFÍA

- Agudo.A, “¿De quién son los parados de Jaén?”, Jaén, ABCandalucía.es, 3 de octubre de 2013. Dirección URL: <http://sevilla.abc.es/andalucia/jaen/20131003/sevi-quien-parados-jaen-201310022043.html>
- Análisis de Componentes Principales. (2014). En Wikipedia, la enciclopedia libre. Consultado el 15 de junio de 2014, a partir [http://es.wikipedia.org/wiki/An%C3%A1lisis\\_de\\_componentes\\_principales](http://es.wikipedia.org/wiki/An%C3%A1lisis_de_componentes_principales)
- Análisis de Correspondencias. (2013). En Wikipedia, la enciclopedia libre. Consultado el 15 de junio de 2014, a partir [http://es.wikipedia.org/wiki/An%C3%A1lisis\\_de\\_correspondencias](http://es.wikipedia.org/wiki/An%C3%A1lisis_de_correspondencias)
- Análisis Factorial. (2007). En Wikipedia, la enciclopedia libre. Consultado el 15 de junio de 2014, a partir [http://es.wikipedia.org/wiki/An%C3%A1lisis\\_factorial](http://es.wikipedia.org/wiki/An%C3%A1lisis_factorial)
- De la Fuente Fernández, S. “*Análisis Factorial*” (2011), Madrid, Universidad Autónoma de Madrid. Consultado el 23 de junio de 2014, a partir <http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/MULTIVARIANTE/FACTORIAL/analisis-factorial.pdf>
- De la Fuente Fernández, S. “*Análisis Correspondencias Simples y Múltiples*” (2011), Madrid, Universidad Autónoma de Madrid. Consultado el 23 de junio de 2014, a partir <http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/REDUCIR-DIMENSION/CORRESPONDENCIAS/correspondencias.pdf>

- Equipo de Trabajo de la Dirección Provincial de Jaén “*Informe del Mercado de Trabajo General y Colectivos*”, Jaén, Servicio Público de Empleo Estatal, abril de 2011. Dirección URL: [http://www.sepe.es/contenido/conocenos/publicaciones/pdf/IMT\\_2011\\_Datos2010\\_Jaen.pdf](http://www.sepe.es/contenido/conocenos/publicaciones/pdf/IMT_2011_Datos2010_Jaen.pdf)
  
- Estadística Multivariante. (2013). En Wikipedia, la enciclopedia libre. Consultado el 15 de junio de 2014, a partir [http://es.wikipedia.org/wiki/Estad%C3%ADstica\\_multivariante](http://es.wikipedia.org/wiki/Estad%C3%ADstica_multivariante)
  
- Instituto de Estadística y Cartografía de Andalucía, Datos recogidos del SIMA, Consultado el 26 de febrero de 2014 a partir <http://www.juntadeandalucia.es/institutodeestadisticaycartografia/sima/index2.htm>
  
- Kessler, M. “*Componentes Principales con R*”, Cartagena, Universidad Politécnica de Cartagena. Consultado el 12 de marzo de 2014, a partir <http://filemon.upct.es/~mathieu/organizacion/practicas/practicasconR/practacp.pdf>
  
- Martín Mesa, A. Duro Cobo, J.J. & Alcalá Olid, F. (2013), “*Observatorio Económico de la Provincia de Jaén*”, Jaén, Diputación Provincial de Jaén, Universidad de Jaén y Cátedra de Planificación Estratégica Territorial. Dirección URL: [http://www.dipujaen.es/export/observatorio\\_economico/Numero200.pdf](http://www.dipujaen.es/export/observatorio_economico/Numero200.pdf)
  
- Redacción, “*La Actividad Económica de Jaén se contraerá un -2,3% este año, según el Informe de Analistas Económicos*”, Jaén, Multipress, Noticiasdjaen.es, 22 de julio de 2013. Dirección URL: [http://www.multipressjaen.com/index.php?option=com\\_content&view=article&id=3133:la-actividad-economica-de-jaen-se-contraera-un-23-este-ano-segun-el-ultimo-informe-de-analistas-economicos&catid=16:economia&Itemid=23](http://www.multipressjaen.com/index.php?option=com_content&view=article&id=3133:la-actividad-economica-de-jaen-se-contraera-un-23-este-ano-segun-el-ultimo-informe-de-analistas-economicos&catid=16:economia&Itemid=23)

- Ruíz Molina, J.C. “*Apuntes del profesor*” (2013), asignatura: Técnicas de Agrupación y Reducción de la Dimensión. Universidad de Jaén.
  
- Terrádez Gurrea, M. “*Análisis de Componentes Principales*”, Catalunya, Universitat Oberta de Catalunya. Consultado el 20 de junio de 2014, a partir [http://www.uoc.edu/in3/emath/docs/Componentes\\_principales.pdf](http://www.uoc.edu/in3/emath/docs/Componentes_principales.pdf)
  
- Terrádez Gurrea, M. “*Análisis de Conclomerados*”, Catalunya, Universitat Oberta de Catalunya. Consultado el 21 de junio de 2014, a partir <http://www.uoc.edu/in3/emath/docs/Cluster.pdf>

