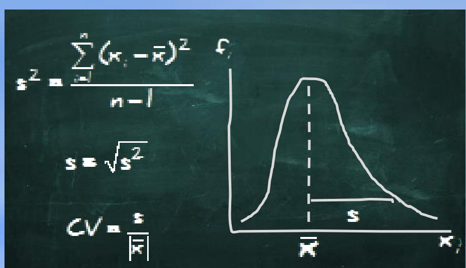
METODOLOGÍA DE LA INVESTIGACIÓN SOCIAL CUANTITATIVA

Pedro López-Roldán
Sandra Fachelli



UAB
Universitat Autònoma de Barcelona

METODOLOGÍA DE LA INVESTIGACIÓN SOCIAL CUANTITATIVA

Pedro López-Roldán
Sandra Fachelli

Bellaterra (Cerdanyola del Vallès) | Barcelona
Dipòsit Digital de Documents
Universitat Autònoma de Barcelona

Pedro López-Roldán
Centre d'Estudis Sociològics sobre la Vida Quotidiana i el Treball (http://quit.uab.cat)
Institut d'Estudis del Treball (http://iet.uab.cat/)
Departament de Sociologia. Universitat Autònoma de Barcelona
pedro.lopez.roldan@uab.cat

Sandra Fachelli
Departament de Sociologia i Anàlisi de les Organitzacions
Universitat de Barcelona
Grup de Recerca en Eduació i Treball (http://grupsderecerca.uab.cat/gret)
Departament de Sociologia. Universitat Autònoma de Barcelona
sandra.fachelli@ub.edu

Edición digital: http://ddd.uab.cat/record/129382

1ª edición, febrero de 2015

Edifici B · Campus de la UAB · 08193 Bellaterra
(Cerdanyola del Vallés) · Barcelona · España
Tel. +34 93 581 1676

Índice general

Metodología de la Investigación Social Cuantitativa

Pedro López-Roldán
Sandra Fachelli

PARTE III. ANÁLISIS

Capítulo III.11
Análisis factorial

Bellaterra (Cerdanyola del Vallès) | Barcelona
Dipòsit Digital de Documents
Universitat Autònoma de Barcelona

Cómo citar este capítulo:

López-Roldán, P.; Fachelli, S. (2016). Análisis factorial. En P. López-Roldán y S. Fachelli, *Metodología de la Investigación Social Cuantitativa*. Bellaterra (Cerdanyola del Vallès): Dipòsit Digital de Documents, Universitat Autònoma de Barcelona. 1ª edición, versión 3. Edición digital: http://ddd.uab.cat/record/142928

Capítulo acabado de redactar en octubre de 2016

Índice de contenidos

PARTE III
Capítulo 11

Análisis factorial

A lo largo de los capítulos III.11 y III.12 daremos cuenta de dos técnicas de análisis de datos multivariables: el análisis factorial (en dos versiones: componentes principales y correspondencias) y el análisis de clasificación que, siendo autónomas y siendo presentadas por separado, se concebirán como complementarias en el contexto de una metodología más general que denominamos de construcción tipológica y, en particular, de construcción de una tipología estructural y articulada (López-Roldán, 1994, 1996; López-Roldán y Fachelli, 2015), de especial interés y utilidad en la práctica de la investigación en ciencias sociales y específicamente en el ámbito de la sociología.

Son dos técnicas complementarias que ilustran bien la naturaleza y los principales conceptos de las técnicas de análisis multivariable en el tratamiento de modelo de análisis de interdependencia. Una analiza el espacio de las variables y la otra el espacio de los individuos, ambas con una finalidad general de reducción de la complejidad de los datos. En el caso del análisis factorial el objetivo es establecer las interrelaciones que se dan entre un conjunto de variables observadas que identifican una realidad social con la intención de reducir la complejidad que reflejan a lo que de forma latente es esencial o más significativo y así identificar unos pocos factores sintéticos que más diferencian a los individuos. En el caso del análisis de clasificación, el objetivo es unirlos para la formación de grupos homogéneos de individuos de acuerdo con sus similitudes desde un punto de vista social. En esta tarea, los factores del análisis factorial previo son variables que actuarán de criterios clasificatorios para obtener estas clasificaciones que identificaremos como tipologías.

En este capítulo dedicado al análisis factorial (AFA) presentaremos en un primer apartado el conjunto de procedimientos específicos que se identifican de forma genérica dentro del análisis factorial para así tener una primera visión general de la naturaleza y especificidad de las dos técnicas concretas que centrarán nuestro interés: el análisis factorial de componentes principales (ACP) y el análisis factorial de correspondencias (ACO). A continuación desarrollaremos los rasgos básicos de cada uno de los procedimientos, su formulación matemática, las representaciones gráficas y todos aquellos elementos necesarios para la lectura o interpretación de los resultados de este tipo de análisis.

A lo largo del capítulo reproduciremos ejemplos concretos de investigaciones donde hemos participado o ejercicios ilustrativos que contribuyan a la comprensión de los diversos conceptos implicados. Los resultados de estas técnicas de análisis se presentarán con las tablas y gráficos que se pueden obtener con los softwares SPSS, SPAD y R; al final del capítulo se detalla cómo obtener esos resultados con la aplicación informática.

1. Análisis factorial: características y clasificación

El análisis factorial tiene sus orígenes en el s. XIX, pero es sobre todo a partir de los trabajos de Spearman (1904) que adopta su primera formulación. El origen y desarrollo de análisis factorial se vincula principalmente con los trabajos realizados en el campo de la psicología ante el problema planteado de medir los factores de la inteligencia humana. Los investigadores en este campo constataron que algunos de los diferentes test psicológicos y de inteligencia estaban muy relacionados entre sí y que, por tanto, no medían diferentes aspectos, factores o dimensiones de la inteligencia. Spearman determinó que como resultado de varios test había un factor general (factor G) de la inteligencia y un cierto número de factores específicos. Posteriormente con los trabajos de Burt (1940) y Thurstone (1947) y otros se perfeccionó el llamado método del Análisis Factorial de Factores Principales donde a través de un modelo lineal y a partir de un conjunto extenso de variables iniciales permitía obtención de un conjunto de factores comunes (hipotéticos) y otros factores específicos que incluían las características propias de cada variable y un error aleatorio.

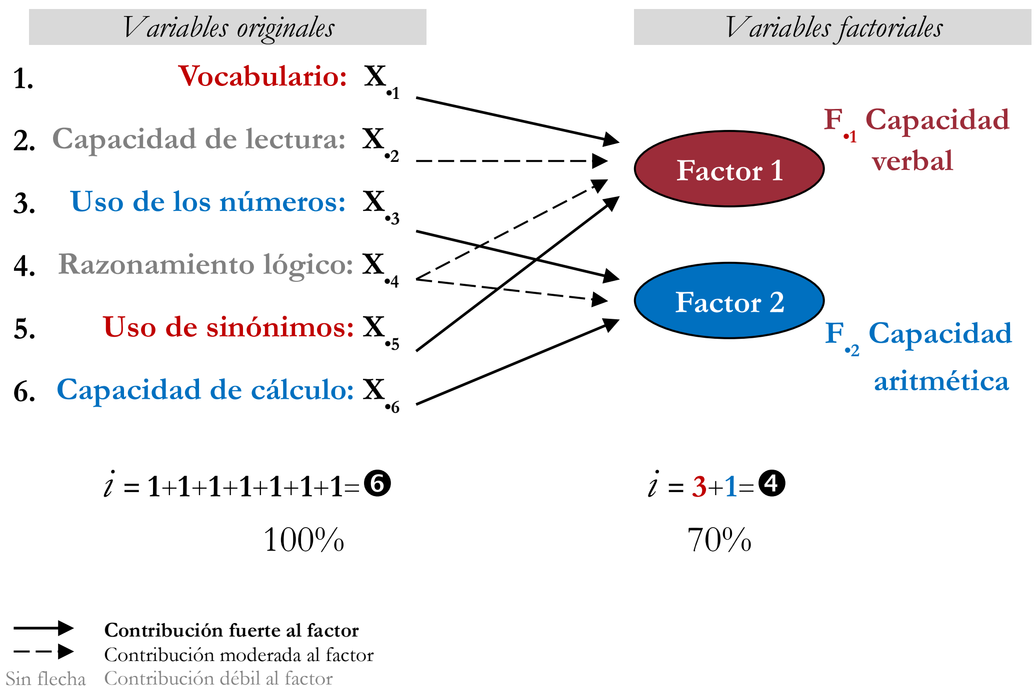
Así, por ejemplo, el análisis factorial permitía explicar las correlaciones encontradas entre las distintas pruebas aplicadas a un grupo de sujetos: pruebas relativas al vocabulario, a la capacidad de lectura, al uso de sinónimos, al uso de los números y a la capacidad de cálculo aritmético a través de dos tipos diferentes de test. Se podía poner de manifiesto la hipótesis según la cual estas seis variables-test no miden seis sino un número inferior de factores que explican las diferencias o la variabilidad de las respuestas. Dado que hay correlaciones entre ellas se pueden formar dos grupos de variables o factores principales que miden dos aspectos de la inteligencia: la capacidad verbal y la capacidad aritmética.

Principios metodológicos y matemáticos diferentes fueron planteados en las aportaciones de Pearson (1901) y de Hotelling (1933) y otros, los cuales desarrollaron el llamado Análisis Factorial de Componentes Principales donde no se realiza la distinción entre factores comunes y específicos, cada factor finalmente explicará una parte del total de la varianza inicial aportada por las variables introducidas. Kaiser (1958) introdujo el procedimiento de la rotación varimax que implicaba la rotación ortogonal con el objetivo de ganar en interpretabilidad de los factores obtenidos. Otra extensión de estos procedimientos dio lugar al Análisis Factorial de Correspondencias, desarrollado en la tradición francesa durante los años 60 por Benzécri (1973).

Vemos pues que bajo la denominación de análisis factorial se incluye una variedad de procedimientos de estadísticos. Todos ellos podemos decir que están destinados al análisis de dimensionalización, es decir, la determinación de las dimensiones (nuevas variables o factores) subyacentes a un conjunto de información (campo de variables o espacio de atributos) que definen una problemática científica específica, múltiple y compleja. En este sentido el análisis factorial está destinado a la búsqueda de un nuevo conocimiento, es decir, está destinado al descubrimiento o la constatación de las dimensiones latentes a los fenómenos (sociales).

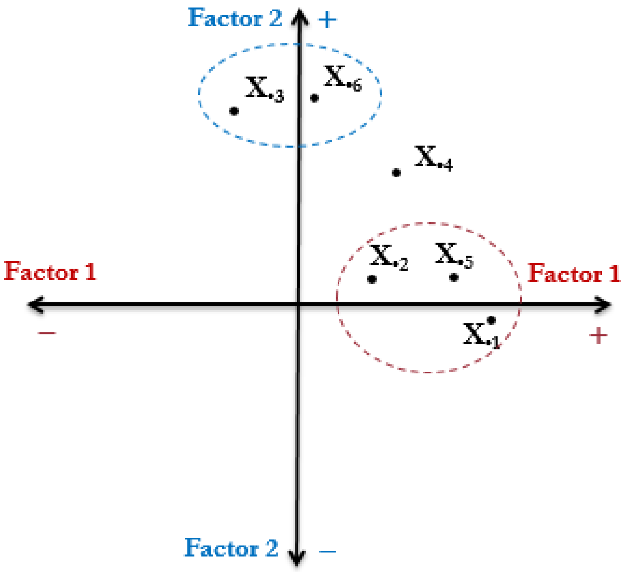
Estos procedimientos consideran como principio que hay una serie de relaciones latentes entre un conjunto de variables manifiestas que identifican un fenómeno social. Para encontrar estas relaciones latentes, o causas no manifiestas de su relación, o, si se quiere expresar así, la explicación de su estructuración, el análisis factorial proporciona un número reducido de factores o de dimensiones subyacentes al conjunto de variables originales. Estos factores no son más que nuevas variables que se obtienen a partir de la combinación, con una contribución diferente, de las variables originales. La relación entre variables originales y factores puede esquematizarse según se muestra en el Gráfico III.11.1 y se expresa geométricamente en una distribución espacial según muestra el Gráfico III.11.2.

Gráfico III.11.1. Esquema de la lógica del análisis factorial



Las técnicas de análisis factorial muestran la existencia de variables que conjuntamente miden el mismo fenómeno, o varias de ellas distintos fenómenos subyacentes a la realidad estudiada. Es decir, los factores o dimensiones que se encuentran son combinaciones (lineales) de las variables originales, cada una de estas variables contribuirá en mayor o menor medida a configurar las nuevas variables factoriales. Por tanto, el modelo matemático sobre el que se edifica el análisis factorial es el modelo lineal y se trata de explicar las correlaciones, la variabilidad, de un conjunto de información con un número menor de variables nuevas subyacentes que salen de la combinación de las variables originales. En términos geométricos los factores se corresponden con unos ejes, perpendiculares entre sí, que mejor se ajustan a la forma de la nube de puntos que configuran un conjunto de individuos en el espacio de las variables. Este ajuste consiste en elegir aquellos ejes en el espacio que de forma progresiva hacen máxima la proyección de los puntos sobre estos ejes[1].

Gráfico III.11.2. Gráfico factorial

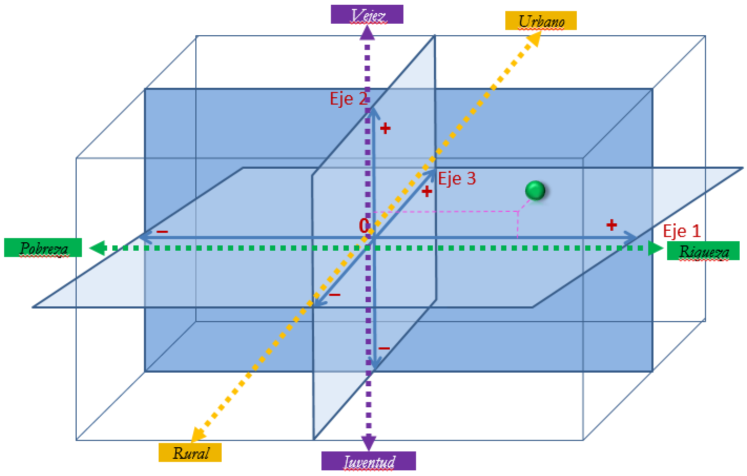


Con el análisis factorial además se logra el objetivo de este tipo de técnicas como es la reducción de la información original tratada. Es la aplicación de la estrategia o el principio de parsimonia científica o de economía de la información: todo modelo debe procurar ser lo más simple posible en la interpretación o explicación de los datos. La máxima de este tipo de técnicas se expresa en la afirmación "*pérdida de información y ganancia en significación*". Esta reducción de información es fundamental ya que sintetiza o realza lo significativo subyacente en el conjunto de información. Cuando se hace estas técnicas asumen una pérdida parcial del total de la variabilidad de la información inicial contenida en las variables, pero lo hacen como estrategia destinada a ganar en significación, es decir, en inteligibilidad, densidad y estructura explicativa.

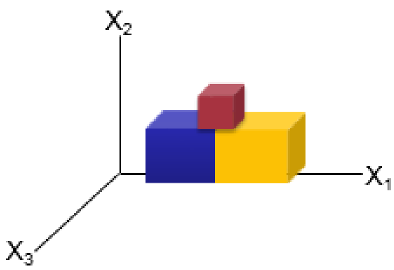
Uno de los aspectos más interesantes de estas técnicas es la expresión gráfica de los resultados del análisis. De hecho se trata de técnicas que traducen geométricamente el procedimiento de análisis implicado. La información de individuos y variables se representa gráficamente a través de diagramas de dispersión en espacios vectoriales que podemos visualizar en gráficos de dos y tres dimensiones. La técnica lo que busca es reducir el espacio multidimensional inicial de representación de la información a unas pocas dimensiones que den cuenta de la mayor parte de ella, gráficamente, que conserve la mayor parte de la información sobre la forma de la nube de puntos (Gráfico III.11.3)[2].

[1] A lo largo del texto emplearemos habitualmente de forma equivalente las expresiones; ejes, dimensiones componentes, factores o nuevas variables.
[2]Este video ilustra nuestros comentarios: https://www.youtube.com/watch?v=BfTMmoDFXyE.

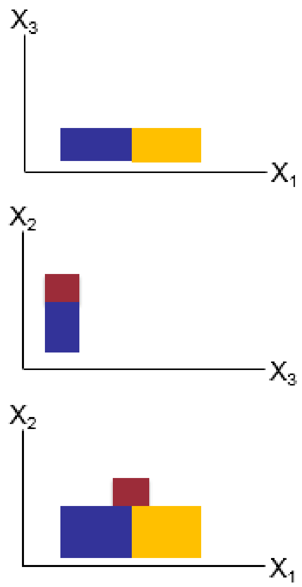
Gráfico III.11.3. Representaciones gráficas en el espacio vectorial



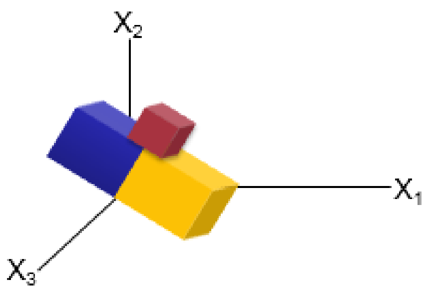
Representación de una nube de puntos en el espacio original de 3D



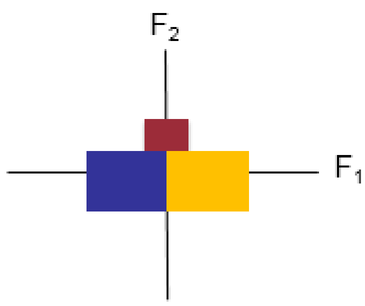
Reducción de dimensiones Proyección en el plano de 2D



Representación de una nube de puntos en el espacio original de 3D



Reducción de dimensiones Factores principales en 2D



La interpretación que se hace de los nuevos factores tan sólo es posible en el contexto de una problemática previamente establecida que justifique la pertinencia de las variables que son analizadas y de las posibles relaciones entre ellas en el seno de una teoría o de un modelo más o menos elaborado, y, por tanto, en un contexto teórico que permite la conceptualización de los resultados que se obtienen.

Desde el punto de vista de la investigación social la aplicación de este tipo de técnicas proporciona valiosos resultados:

1. Ante todo nos permite el tratamiento de numerosa información, de forma multidimensional, con la que el investigador/a puede identificar y explicar los fenómenos sociales con toda su complejidad y sus múltiples manifestaciones. La descripción conjunta de un grupo numeroso de variables da lugar a una síntesis e interrelación de la información. Esta síntesis e interrelación reduce y estructura las variables consideradas.
2. El análisis factorial es especialmente adecuado cuando se utilizan informaciones distintas y relevantes para el estudio y que se relacionan entre sí sin un modelo preciso que establece su asociación y el grado validez.
3. La búsqueda de factores o fenómenos latentes a partir de otros manifiestos es una característica común de los diversos procedimientos de análisis factorial. Según los objetivos de investigación podemos distinguir entre:
   - El Análisis Factorial Exploratorio que pretende buscar o descubrir los factores a partir de la tarea interpretativa y atribuyendo un significado a posteriori a los factores.
   - El Análisis Factorial Confirmatorio que implica la explicitación previa de un modelo determinado sobre los factores subyacentes y que se somete a confirmación con los datos observados.
   - También permite la construcción de medidas complejas o índices a partir de un conglomerado de variables o ítems que constituyen un factor determinado.
4. Sola o complementaria con otras técnicas, el análisis factorial permite la consecución de otros fines que van más allá de la mera descripción o de un ejercicio puramente exploratorio:
   - porque permite conceptualizar nuevas realidades en el contexto de una teoría,
   - porque cubre fines en la construcción de tipologías al complementarse con las técnicas de clasificación automática,
   - porque algunas de estas técnicas permiten el establecimiento de relaciones de causalidad,
   - porque permiten o incorporan procedimientos de validación,
   - porque permiten resolver los problemas de (multi)colinealidad en los análisis de regresión.

El modelo matemático de análisis factorial se basa en tres elementos clave:

1) El primero consiste en el tratamiento de la matriz de datos X como un espacio vectorial euclídeo de p variables y n individuos que son posiciones puntuales que forman la nube de puntos. Este conjunto de variables engendra un espacio vectorial del cual se saca su base[3] que son los factores.

[3] La base de un espacio vectorial es un concepto técnico matemático y algebraico que proporciona vectores o variables linealmente independientes como veremos más adelante.

2) El segundo, en encontrar los factores (o nuevas variables) de este espacio vectorial, lo que equivale a descubrir unos nuevos ejes en este espacio donde la proyección de los puntos sobre ellos sea máxima. Esto equivale a afirmar que la identificación de los ejes se realiza haciendo que cada uno, de forma gradual y progresiva, proyecte sobre ellos el máximo de la inercia (varianza) de todos los puntos, en otras palabras, se consiga acumular la mayor cantidad posible de información.

3) Finalmente, otro elemento importante es la naturaleza lineal de la dependencia de los ejes y las variables observadas. Por un lado, los ejes se encuentran mediante la base del sistema vectorial, como dimensiones linealmente independientes y, por otro lado, la expresión de las variables originales en función de los factores o ejes se realiza mediante una expresión que especifica una relación lineal:

$$X = F \cdot A' \qquad \textit{Ecuación 1}$$

donde X es la matriz de datos originales, o de variables observadas, de orden ($n \times p$), de n individuos y p variables. F es la matriz obtenida de factores, de orden ($n \times p$) ó ($n \times m$) si hay reducción, con $m<p$, de n individuos y m factores. A' es la matriz traspuesta[4] de orden ($p \times p$) ó ($m \times p$) si hay reducción, de p variables y m factores.

$$\underset{\textit{Variables originales}}{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}} = \underbrace{\underset{\textit{Factores de carga}}{\begin{pmatrix} a_{11} \\ a_{22} \\ \vdots \\ a_{mp} \end{pmatrix}} \times \underset{\textit{Variables factoriales}}{\begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{pmatrix}}}_{\textit{Comunalidad}} + \underbrace{\underset{\textit{Errores Factores únicos}}{\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{pmatrix}}}_{\textit{Especificidad}}$$

Varianza total = Varianza común + Varianza específica

Para cada variable original se obtiene una ecuación lineal:

$$x_1 = a_{11}F_1 + a_{12}F_2 + \cdots + a_{1m}F_m + e_1$$
$$x_2 = a_{21}F_1 + a_{22}F_2 + \cdots + a_{2m}F_m + e_2$$

$$x_p = a_{p1}F_1 + a_{p2}F_2 + \cdots + a_{pm}F_m + e_p$$

La expresión de la *Ecuación 1* es clave a la hora de diferenciar los métodos de análisis factorial:

a) Análisis Factorial Exploratorio que pretende buscar o descubrir los factores pero no a partir de un modelo prefijado o cerrado que se impone en los datos sino a partir de la tarea interpretativa y atribuyendo un significado a posteriori los

[4] Una matriz traspuesta es una matriz transformada que simplemente dispone la información de las filas en columna y viceversa. Más adelante veremos que a esta matriz la llamaremos factorial o de saturaciones y será el material interpretativo principal de las relaciones entre las variables originales X y los factores F.

factores, si bien el modelo sustantivo debe estar igualmente presente en mayor o menor medida. En este caso los modelos son de interdependencia donde se puede distinguir dos tipos particulares a partir de la siguiente relación en el tratamiento de la varianza:

$$\textit{Varianza total} = \textit{Varianza común} + \underbrace{\textit{Varianza específica} + \textit{Varianza de error de medición}}_{\textit{Varianza no común}}$$

a.1) Análisis Factorial de Varianza Total. Se extraen los factores comunes explicando el total de la varianza. En la definición de los factores o componentes Y interviene el conjunto de variables X contribuyendo cada una en su totalidad, aunque con pesos diferentes, sin diferenciar una parte común y una parte específica. En este grupo se encuentran tres tipos particulares de técnicas[5]:

- El Análisis de Componentes Principales donde se relacionan variables cuantitativas.
- El Análisis de Correspondencias Múltiples donde se relacionan variables cualitativas.
- El Análisis de Componentes Categórico donde se pueden combinar variables cuantitativas y cualitativas.

a.2) Análisis Factorial de Factor Común. En este caso se extraen factores en relación a la varianza común (covarianza) de las variables y supone que existe un factor común subyacente a todas las variables. Así distingue la comunalidad (varianza compartida con otras variables) de la varianza única o específica (no compartida y propia de la variable original). En este grupo se encuentran las siguientes técnicas particulares:

- Método de mínimos cuadrados no ponderados. Minimiza la suma de los cuadrados de las diferencias entre las matrices de correlación observada y reproducida, ignorando las diagonales.
- Método de mínimos cuadrados generalizados. Minimiza la suma de los cuadrados de las diferencias entre las matrices de correlación observada y reproducida. Las correlaciones se ponderan por el inverso de su unicidad, de manera que las variables que tengan un valor alto de unicidad reciban una ponderación menor que aquéllas que tengan un valor bajo de unicidad.
- Método de máxima verosimilitud. Proporciona las estimaciones de los parámetros que con mayor probabilidad ha producido la matriz de correlaciones observada, si la muestra procede de una distribución normal multivariada. Las correlaciones se ponderan por el inverso de la unicidad de las variables, y se emplea un algoritmo iterativo.
- Factorización de ejes principales. Parte de la matriz de correlaciones original con los cuadrados de los coeficientes de correlación múltiple insertados en la diagonal principal como estimaciones iniciales de las comunalidades. Las saturaciones factoriales resultantes se utilizan para estimar de nuevo las comunalidades que reemplazan a las estimaciones

[5] Existe también el llamado Análisis Factorial Q donde se interrelacionan casos y no variables, con variables se denominaría como Análisis Factorial R.

previas de comunalidad en la diagonal. Las iteraciones continúan hasta que el cambio en las comunalidades, de una iteración a la siguiente, satisfaga el criterio de convergencia para la extracción.

- Alfa. Considera a las variables incluidas en el análisis como una muestra del universo de las variables posibles. Este método maximiza el Alfa de Cronbach para los factores.
- Factorización imagen. Desarrollado por Guttman y basado en la teoría de las imágenes. La parte común de una variable, llamada la imagen parcial, se define como su regresión lineal sobre las restantes variables, en lugar de ser una función de los factores hipotéticos.

b) Análisis Factorial Confirmatorio. En este caso se exige una hipótesis y una distinción entre la parte con que cada variable contribuye a la creación de la explicación común que genera los ejes y la parte específica no explicada por la totalidad según un modelo previo, tratándose así de un análisis de dependencia.

En este capítulo daremos cuenta de dos técnicas de análisis factorial de varianza total: el análisis de componentes principales y el análisis de correspondencias.

2. Análisis de componentes principales

2.1. Definición, objetivos y modelo de análisis

El análisis de componentes principales es un método algebraico/estadístico que trata de sintetizar y dar una estructura a la información contenida en una matriz de datos. El procedimiento consiste en homologar esta matriz en un espacio vectorial tratando de encontrar en él unos ejes o dimensiones (llamados también componentes o factores) que, siendo combinación lineal de las variables introducidas:

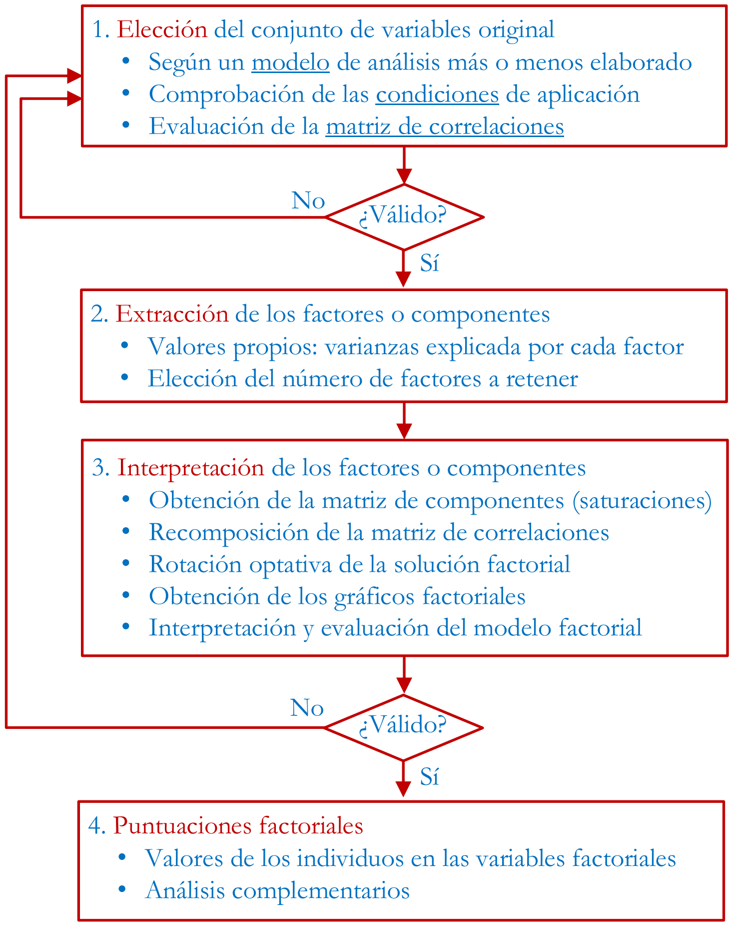
- No pierdan la información inicial al conservar la varianza total.
- No tengan correlación entre ellos, es decir, sean linealmente independientes, lo que asegura la estructuración de las variables iniciales.
- Tengan una importancia diferencial y conocida en la explicación de la varianza total.

El objetivo básico consiste en reducir el número de variables introducidas. Para ello se toman como nuevas variables los ejes o componentes encontrados, con la elección de un número y peso de los mismos suficiente para que la pérdida de varianza total sea aceptable, cubriendo así las finalidades del método, es decir, simplificar, reducir y estructurar la información inicial.

Para el investigador/a el análisis ni empieza el tratamiento descrito ni termina en él. Previamente se deberán definir las variables que entran en juego en el análisis, validarlas y teniendo en consideración un modelo, aunque sea elemental, de hipótesis, descriptivas o explicativas, que dé cuenta de la problemática considerada. De forma simultánea y posteriormente al proceso deberá tener criterios estadísticos pero sobre todo substantivos, para elegir el número de ejes y fundamentalmente para dar una identidad a estos ejes, así como para interpretar y proponer una estructura del conjunto de las variables y de las relaciones o agrupaciones entre ellas.

A este método se le suele clasificar entre los descriptivos pero, como destacaremos, permite ir más allá y llenar otros fines que los que se derivan de la naturaleza empirista y de rango menor y subsidiario que, como a nivel de análisis, es habitualmente atribuido a la descripción. Precisamente no situamos los procedimientos que nos sirven para estructurar y/o tipificar la realidad social como simplemente descriptivos.

Gráfico III.11.4. Etapas de un análisis de componente principales



El proceso de análisis de un ACP se puede dividir en 4 etapas básicas que se esquematizan en el Gráfico III.11.4. Los apartados siguientes se presentan de acuerdo con este orden del proceso:

1) En primer lugar es fundamental la elección del espacio de atributos o conjunto de variables original según un modelo de análisis más o menos elaborado. Esta elección deber ser coherente con las condiciones formales de aplicación de la técnica y deben cumplirse determinadas condiciones de intercorrelación entre las variables que se establecen mediante el examen de la matriz de correlaciones. Se

trata pues de una etapa previa que determina la pertinencia del ACP con las variables elegidas o la revisión del modelo para garantizarlas.

2) A continuación se extraen los ejes factoriales o componentes con el cálculo de los valores propios o varianzas incorporados a cada uno de los ejes que determinan la varianza explicada por cada uno de ellos y se decide el número de factores a retener en el análisis.
3) Se procede a la interpretación de los factores o componentes a partir del cálculo de la correlación entre éstos y las variables originales, de la comunalidad y recomposición de la matriz de correlaciones, con la posibilidad de realizar una rotación de la solución factorial para mejorar la interpretabilidad de los ejes factoriales.
4) Si se concluye la validez de los resultados podemos calcular de las puntuaciones factoriales o valores para los individuos en las variables factoriales para obtener análisis adicionales con otros procedimientos.

Para mostrar el proceso del ACP seguiremos el ejemplo del análisis factorial realizado para la construcción de la muestra estratificada de una encuesta[6]. A partir de datos censales del año 1986 se busca construir los estratos de la muestra para obtener una representación de la población mayor de 18 años de la Región Metropolitana de Barcelona. Se dispone de la información de la población clasificada en 3509 secciones censales y caracterizadas según 17 variables. Se recurrió al ACP con el objetivo de reducir la información contenida en esta matriz original buscando unas nuevas componentes que más explicaran la variabilidad del conjunto y, por tanto, mejor nos definieran las diferencias entre las distintas secciones a efectos de su clasificación posterior en estratos. Presentaremos los resultados del tratamiento de estos datos, su lectura e interpretación al hilo del proceso de análisis que se explica seguidamente.

2.2. Etapa 1: Elección de las variables originales

La elección del espacio atributos (Barton, 1985) o conjunto de variables original según un modelo de análisis más o menos elaborado es una etapa de crucial importancia en cualquier análisis de datos y en un ACP en particular. Cabe tener presente que las conclusiones de un análisis factorial no se podrán llevar más allá de la pertinencia del campo acotado por las variables definidas.

Aparte de esta primera exigencia general y fundamental sobre las variables previamente consideradas se efectuará una segunda pesquisa para decidir la conveniencia de realizar un ACP. El punto de partida de una análisis de componentes principales es la matriz de correlaciones, de hecho el análisis busca mostrar las correlaciones que se dan entre grupos de variables, por tanto, se debe constatar de antemano que las variables están manifiestamente relacionadas. Además se deben dar una serie de condiciones generales previas para la aplicación de la técnica.

[6] La muestra estratificada en la que nos referimos ha servido de base para la recogida (producción) de información de la *Encuesta Metropolitana de Barcelona* que hoy en día se denomina *Encuesta de Condiciones de Vida y Hábitos de la Población de Cataluña* (ECVHPC), elaborada en el seno del *Institut d'Estudis Regionals i Metropolitans de Barcelona* en colaboración con el *Institut d'Estadística de Catalunya* (www.enquestadecondicionsdevida.cat). Para una exposición más detallada del diseño y del proceso de construcción de las muestras construidas con esta metodología en los años 1990, 1995, 2000 y 2005 se pueden consultar los artículos Lozares y López-Roldán (1991a, 1991b), López-Roldán, Lozares y Domínguez (2000), López-Roldán y Lozares (2007).

2.2.1. *Las variables seleccionadas*

A continuación se presenta la relación de indicadores o variables de la Padrón de Habitantes de 1986 que se consideraron para realizar el ACP del ejemplo de análisis que utilizamos. Cabe destacar la particularidad de los datos de la matriz[7]. En la Tabla III.11.1 se presentan los valores de medias y desviaciones de las variables consideradas.

Tabla III.11.1. Variables utilizadas en el ACP

	Media	Desviación estándar
P1 % Menores de 15 años	20,2	6,4
P2 % Mayores de 65 años	13,2	6,8
P4 % de Inmigración fuera Cataluña	37,4	11,5
P6 % de Nuevos residentes municipio 81-86	5,8	3,2
P7 % de Analfabetos >10 años	5,2	4,4
P8 % de Titulados medios-superiores >20 años	9,7	10,2
P9 % de Escolarización 14-24 años	47,6	14,1
P11 % de Parados antes ocupados	14,9	5,0
P12 % de Paro busca primer empleo	9,2	4,8
P14 % de Activas >15 años	31,8	5,6
P15 % de Profesiones altas	18,1	13,3
P16 % de Profesiones bajas	4,2	3,7
P17 % de Terciario medio / comercio / hostelería	12,4	4,4
P18 % de Terciario alto / finanzas	4,3	3,4
P19 % de Agropecuario	1,0	3,0
P20 % de Vehículo privado trabajo	37,7	12,3
P23 Población Sección/Municipio	3,7	13,3

Fuente: Padrón Municipal de Habitantes de la Región Metropolitana de Barcelona 1986

Excepto en el caso de la última, la P23, las variables se obtienen calculando el porcentaje de personas de la sección censal que tienen una determinada característica sobre el total de personas de la sección censal.. Por tanto, la unidad elemental no son las personas (3.177.765 según el Padrón de 1986) sino la agregación de éstas en secciones censales, un total de 3509, de cada una de las cuales se dispone del porcentaje (o índice) correspondiente a cada variable. Por tanto, se trata en todos los casos de variables medidas con una escala cuantitativa.

En relación a los criterios de elección de las variables para ser incluidas en la ACP, en este caso, no hubo mucho margen de maniobra en la elección pues se trata de la información limitada que se recoge en el Padrón, pero relevantes para los objetivos de construcción de la muestra de la encuesta y, por tanto, los objetivos de la investigación. Entre esta información se seleccionaron finalmente 17 indicadores (de los 23 disponibles) relativos a características poblacionales, cultural-educativas, ocupacionales, profesionales y de movilidad como se pueden ver a continuación[8].

[7] La matriz se puede encontrar en la página web con el nombre RMB1986.sav.

[8] Inicialmente se consideraron 23 indicadores de la sociedad metropolitana. Las variables que empleamos aquí conservan su nombre original con la numeración de un conjunto inicial de 23. Las variables P3, P5, P10, P13, P21

2.2.2. *Condiciones de aplicación*

Para poder aplicar un análisis de componentes principales se deben cumplir las siguientes cuatro condiciones generales:

1. Tamaño muestral suficiente:
   - Se recomienda considerar desde un mínimo de 50 casos (tamaño muy pobre) a un nivel mínimo deseable de entre 200 y 300 casos, siendo 1000 un tamaño de muestra estadística excelente (Comrey y Lee, 1992).
   - También se recomienda un mínimo es de 5 casos por variable, si bien es preferible 10 o más (Hair et al., 2011: 101).
   - Como en todo ejercicio previo de análisis multidimensional es necesario analizar e inspeccionar las variables previamente para detectar, en otros aspectos, la presencia y la importancia de los valores perdidos que pueden reducir de forma significativa el tamaño de la muestra.
2. Normalidad. Si el objetivo es exploratorio no se trata de una condición restrictiva. Si se realiza inferencia estadística se asume normalidad multivariable de las variables y de sus combinaciones lineales. Se pueden transformar las variables con exceso de asimetría o curtosis (Tabachnick y Fidell, 2007: 613).
3. Linealidad. Las relaciones entre pares de variables han de ser lineales pues el ACP se basa en el análisis de las correlaciones lineales entre el conjunto de variables elegido. Con gráficos de dispersión se puede observar. Del mismo modo los factores y las variables originales mantienen este tipo de relación. Si no se satisface se puede también operar una transformación de las variables.
4. Ausencia de casos extremos que puedan influir en la solución factorial.

2.2.3. *Análisis de la matriz de correlaciones*

Para la realización de un ACP se debe constatar que las variables consideradas muestren de forma manifiesta un cierto grado de asociación entre ellas, es decir, estén correlacionadas. De hecho el punto de partida o la información con la que se realizan los cálculos de la ACP es la contenida en la matriz de correlaciones. Por tanto, un segundo paso dentro del proceso de análisis consiste en el cálculo de la matriz de correlaciones y en otras pruebas basadas en esta matriz para evaluar la pertinencia del modelo concebido. El ACP lo que hace es tratar de encontrar los factores que ayudan a explicar las correlaciones entre las variables, si estas correlaciones son bajas es poco probable que se formen factores comunes. A continuación reproducimos la matriz de correlaciones (*R*) entre las variables consideradas (Tabla III.11.2).

En la matriz de correlaciones se han destacado en color las correlaciones más importantes. En un primer proceso de análisis, descriptivo e ilustrativo de los resultados que se verán posteriormente, se trata de ver la existencia de correlaciones suficientemente importantes y realizar una lectura de éstas a partir de la distinción entre las correlaciones positivas y negativas.

y P22 fueron descartadas por varios criterios, principalmente por la redundancia de los indicadores o por baja correlación. Los indicadores provienen de variables cualitativas donde, por ejemplo, se distingue entre categorías profesionales altas, medias y bajas. Cada categoría se convierte en una variable con el % de personas de la sección que tienen esa característica, y se suprime una de ellas (en este caso las categorías medias) para evitar que sumen 100% y formen una combinación lineal perfecta que dificultaría los cálculos del ACP.

Tabla III.11.2. Matriz de correlaciones (R)

Correlaciones	P1	P2	P4	P6	P7	P8	P9	P11	P12	P14	P15	P16	P17	P18	P19	P20	P23
P1 Menores de 15 años	**1**	-0,83	0,39	0,33	0,20	-0,31	-0,26	0,05	0,20	0,17	-0,36	0,29	-0,29	-0,35	0,08	0,44	0,12
P2 Mayores de 65 años	**-0,83**	**1**	-0,46	-0,24	-0,15	0,28	0,22	-0,03	-0,27	-0,20	0,35	-0,28	0,34	0,31	-0,03	-0,41	-0,08
P4 Inmigración fuera Cataluña	0,39	-0,46	**1**	-0,15	0,52	-0,63	-0,56	0,52	0,53	-0,25	-0,67	0,46	-0,06	-0,58	-0,19	-0,16	-0,19
P6 Nuevos residentes municipio	0,33	-0,24	-0,15	**1**	-0,08	0,15	0,03	-0,24	-0,23	0,36	0,15	-0,12	-0,01	0,11	0,25	0,38	0,30
P7 Analfabetos >10 años	0,20	-0,15	0,52	-0,08	**1**	-0,55	-0,65	0,59	0,64	-0,31	-0,56	0,40	0,10	-0,54	0,05	-0,12	-0,03
P8 Titulados medios-superiores	-0,31	0,28	**-0,63**	0,15	-0,55	**1**	0,81	-0,60	-0,50	0,39	0,95	-0,46	-0,21	0,63	-0,11	0,20	-0,08
P9 Escolarización 14-24 años	-0,26	0,22	-0,56	0,03	**-0,65**	**0,81**	**1**	-0,65	-0,62	0,22	0,81	-0,48	-0,16	0,62	-0,16	0,13	-0,13
P11 Parados antes ocupados	0,05	-0,03	0,52	-0,24	0,59	**-0,60**	**-0,65**	**1**	0,52	-0,28	-0,61	0,35	0,24	-0,46	-0,17	-0,35	-0,20
P12 Paro busca primer empleo	0,20	-0,27	0,53	-0,23	**0,64**	-0,50	**-0,62**	0,52	**1**	-0,30	-0,53	0,45	-0,04	-0,51	-0,03	-0,09	-0,08
P14 Activas >15 años	0,17	-0,20	-0,25	0,36	-0,31	0,39	0,22	-0,28	-0,30	**1**	0,36	-0,22	-0,14	0,35	-0,13	0,21	-0,07
P15 Profesiones altas	-0,36	0,35	**-0,67**	0,15	-0,56	**0,95**	**0,81**	**-0,61**	-0,53	0,36	**1**	-0,50	-0,18	0,65	-0,12	0,18	-0,10
P16 Profesiones bajas	0,29	-0,28	0,46	-0,12	0,40	-0,46	-0,48	0,35	0,45	-0,22	-0,50	**1**	-0,03	-0,44	-0,01	-0,01	0,03
P17 Terciario, comercio, hostel.	-0,29	0,34	-0,06	-0,01	0,10	-0,21	-0,16	0,24	-0,04	-0,14	-0,18	-0,03	**1**	-0,05	0,02	-0,34	-0,05
P18 Terciario alto/finanzas	-0,35	0,31	-0,58	0,11	-0,54	**0,63**	**0,62**	-0,46	-0,51	0,35	**0,65**	-0,44	-0,05	**1**	-0,16	-0,04	-0,14
P19 Agropecuario	0,08	-0,03	-0,19	0,25	0,05	-0,11	-0,16	-0,17	-0,03	-0,13	-0,12	-0,01	0,02	-0,16	**1**	0,25	0,64
P20 Vehículo privado trabajo	0,44	-0,41	-0,16	0,38	-0,12	0,20	0,13	-0,35	-0,09	0,21	0,18	-0,01	-0,34	-0,04	0,25	**1**	0,36
P23 Población Sección/Municipio	0,12	-0,08	-0,19	0,30	-0,03	-0,08	-0,13	-0,20	-0,08	-0,07	-0,10	0,03	-0,05	-0,14	**0,64**	0,36	**1**
Significación unilateral	**P1**	**P2**	**P4**	**P6**	**P7**	**P8**	**P9**	**P11**	**P12**	**P14**	**P15**	**P16**	**P17**	**P18**	**P19**	**P20**	**P23**
P1 Menores de 15 años		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
P2 Mayores de 65 años	0,00		0,00	0,00	0,00	0,00	0,00	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,06	0,00	0,00
P4 Inmigración fuera Cataluña	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
P6 Nuevos residentes municipio	0,00	0,00	0,00		0,00	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,32	0,00	0,00	0,00	0,00
P7 Analfabetos >10 años	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,03
P8 Titulados medios-superiores	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
P9 Escolarización 14-24 años	0,00	0,00	0,00	0,02	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
P11 Parados antes ocupados	0,00	0,03	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
P12 Paro busca primer empleo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,01	0,00	0,02	0,00	0,00
P14 Activas >15 años	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,00
P15 Profesiones altas	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00
P16 Profesiones bajas	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00		0,05	0,00	0,37	0,37	0,02
P17 Terciario, comercio, hostel.	0,00	0,00	0,00	0,32	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,05		0,00	0,17	0,00	0,00
P18 Terciario alto/finanzas	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00		0,00	0,01	0,00
P19 Agropecuario	0,00	0,06	0,00	0,00	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,37	0,17	0,00		0,00	0,00
P20 Vehículo privado trabajo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,37	0,00	0,01	0,00		0,00
P23 Población Sección/Municipio	0,00	0,00	0,00	0,00	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,00	

Determinante de la matriz: 0,0000066

Las correlaciones importantes negativas que muestran la oposición, dos a dos en cada fila, entre estos grupos de variables son:

P1: Menores de 15 años	P2: Mayores de 65 años
P8: Titulados superiores	P4: Inmigración fuera Cataluña
P9: Escolarización 14-24	P7: Analfabetos
P8: Titulados superiores	P11: Parados antes ocupados
P9: Escolarización 14-24	P11: Parados antes ocupados
P15: Profesiones altas	P4: Inmigración fuera Cataluña

De forma complementaria se puede seguir el análisis mediante la descripción de las correlaciones importantes positivas que mostrarían la relación dos a dos entre estos grupos de variables:

P8: Titulados superiores	P9: Escolarización 14-24
P8: Titulados superiores	P15: Profesiones altas
P9: Escolarización 14-24	P15: Profesiones altas
P12: Parados primer empleo	P7: Analfabetos

De hecho la lectura de las correlaciones positivas lo que hace es reiterar y destacar las interrelaciones de las variables que se dan en el interior de cada grupo de las anteriores.

Estas correlaciones nos muestran la importancia de la relación positiva entre altos niveles educativos de la población con elevados porcentajes de categorías profesionales altas, y una relación inversa, negativa, con las variables relacionadas con el desempleo y la presencia de población de origen inmigrante pues los territorios que tienen las características anteriores de altos niveles educativos y profesionales es donde es menor la incidencia del desempleo y donde el porcentaje de inmigrantes es menor. También se pone de manifiesto la oposición entre secciones censales con alta proporción de población mayor que se contraponen evidentemente con las secciones que tienen más población joven.

2.2.4. *Evaluación del ACP*

Además de la matriz de correlaciones disponemos de otra información descriptiva inicial de la relación entre las variables que nos ayudan a determinar la pertenencia de del análisis de componentes. Se trata de evaluar a través de varios cálculos basados en la matriz de correlaciones: el determinante de la matriz, la significación de los coeficientes, las medidas de adecuación y la matriz anti-imagen, la idoneidad de un ACP. Pasamos a comentarlos.

En la Tabla III.11.2 de la matriz de correlaciones aparece como nota a pie el determinante de la matriz de correlaciones, en este caso el valor es de 0,0000066. El determinante de R es un indicador del grado de intercorrelación existente entre las variables. Un determinante bajo, como es el caso en el que nos encontramos, significa que hay variables con intercorrelaciones altas y que los datos son adecuados para realizar un ACP. Si el determinante fuese cero, una fila (o columna) de la misma sería idéntica a otra, lo que equivaldría a decir que la correlación que guardan las dos variables con todas las demás es idéntica, es decir, que ambas variables estarían tan asociadas que serían iguales. En la medida en que el determinante se acerque más a cero, sin que llegue a serlo, se puede afirmar que las variables están más asociadas y se favorecen las condiciones para la realización del ACP[9].

Una información adicional a considerar es la significación de los coeficientes de correlación. Como podemos ver la inmensa mayor parte de ellos lo son y nos reiteran la conclusión anterior de la pertenencia del conjunto de las variables introducidas. Si encontráramos numerosos coeficientes, presentes en alguna o algunas variables, que manifiestan una baja o no significativa correlación con las otras variables se trataría de reconsiderar estas variables suprimiéndolas del análisis[10].

Disponemos adicionalmente una serie de índices y de pruebas que nos permiten evaluar la fuerza de la relación entre las distintas variables. Así, el test de esfericidad de Bartlett se utiliza para probar la hipótesis nula según la cual la matriz de correlaciones de las variables observadas es la matriz identidad I:

$$H_0\text{: } R = I$$
$$H_A\text{: } R \neq I$$

[9] Si fuese 0, con señalamos en la nota anterior, se generaría un problema de cálculo en la resolución de las ecuaciones que se utilizan con la técnica. El SPSS mostraría una nota a pie de página en la matriz de correlaciones con el mensaje la matriz no es semidefinida positiva, un concepto técnico algebraico que revelaría esta situación.

[10] No obstante en este caso la significación estadística de los coeficientes se plantea con datos poblacionales, agrupados en secciones censales, donde el razonamiento inferencial de hecho no es pertinente.

Es decir, cuando los coeficientes de correlación entre las diferentes variables serían todos nulos (o no diferirían significativamente ser nulos), excepción hecha de la diagonal que expresa la correlación de las variables con sí mismas que, por lo tanto, sería una diagonal de unos. Esto significaría que las variables estarían incorrelacionadas y la nube de puntos tendería a la esfericidad, es decir, tendería a ser una esfera. La expresión geométrica de esta independencia consiste en que la nube de puntos adopta una forma esférica, por tanto, donde no se pueden encontrar direcciones privilegiadas de máxima variabilidad. En consecuencia, no se podrían encontrar los ejes factoriales ya que las variables originales son independientes y en este caso no vale la pena hacer el ACP. En el caso que estamos estudiando se debe rechazar la hipótesis nula de independencia o de esfericidad.

La expresión que estudiamos del estadístico de la prueba de Barlett está basada en una transformación del determinante de la matriz de correlaciones R y sigue una distribución de chi-cuadrado, con la siguiente fórmula, a partir de n casos p variables:

$$\chi^2 = -\left(n-1-\frac{2p+5}{6}\right)\cdot ln|R| \qquad \textit{Ecuación 2}$$

con $\upsilon = \dfrac{p^2-p}{2}$ grados de libertad.

En el ejemplo que estamos considerando se puede apreciar en la Tabla III.11.3 que aparece a continuación que el valor del estadístico es elevado (41.770) y el nivel de significación asociado, mínimo, 0,000, con 136 grados de libertad. En consecuencia, rechazamos la hipótesis nula y concluimos la existencia de interdependencia o intercorrelación significativa entre las variables. En general, un test no significativo de Bartlett evidenciaría la no relación entre las variables, R aproximaría a la matriz identidad y, por tanto, sería aconsejable reconsiderar el modelo factorial.

Tabla III.11.3. Medida de adecuación muestral de Kaiser-Meyer-Olkin y prueba de Barlett

Medida Kaiser-Meyer-Olkin de adecuación muestral		0,847
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	41770,155
	Grados de libertad	136
	Significación	0,000

Otro indicador de la relación entre las variables la expresa el coeficiente de correlación parcial ($rp_{jj'}$). La correlación parcial mide la correlación entre dos variables cuando permanecen constantes el resto de las variables del análisis, o haciendo que sus variaciones sean nulas; es decir, cuando se elimina la influencia de las otras variables. Si existen factores comunes a un conjunto de variables con una fuerte correlación, al eliminar el efecto de asociación que tienen las otras variables sobre dos de ellas, el coeficiente de correlación parcial entre dos variables debería ser pequeño (próximo a cero). Por tanto, 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. Se pueden asociar pequeños coeficientes de correlación parcial a fuertes correlaciones comunes en el conjunto de variables originales u observadas, o sea a la existencia de factores conjuntos comunes. Este hecho hace recomendable y factible la conveniencia de aplicar el ACP. Por el contrario, cuando dos variables comparten gran cantidad de información entre ellas, pero no la comparten con las restantes variables, no contribuyen a generar factores comunes y la correlación parcial entre ellas será elevada, siendo un indicador de la falta de idoneidad de las variables implicadas en el análisis.

El valor negativo del coeficiente de correlación parcial se denomina correlación anti-imagen. En la Tabla 3.4 aparecen reflejados estos valores en la llamada matriz de correlaciones anti-imagen. Si la proporción de coeficientes de correlación parcial elevados es considerable se desaconseja considerar el modelo factorial, es necesario por lo tanto que la mayor parte de los valores de esta matriz sean valores bajos, como el caso que aquí se ilustra (la práctica totalidad de los mismos se aproximan al valor 0), para aconsejar la realización de la ACP.

Tabla III.11.4. Matriz de correlaciones anti-imagen

Correlaciones	P1	P2	P4	P6	P7	P8	P9	P11	P12	P14	P15	P16	P17	P18	P19	P20	P23
P1 Menores de 15 años	**0,77**[a]	0,61	0,06	-0,24	-0,09	0,10	-0,06	-0,01	0,11	-0,07	0,03	-0,08	0,13	0,14	0,02	-0,18	0,08
P2 Mayores de 65 años	0,61	**0,71**[a]	0,30	0,03	-0,11	0,08	0,18	-0,07	0,25	0,25	-0,21	0,00	-0,12	0,07	0,02	0,19	0,03
P4 Inmigración fuera Cataluña	0,06	0,30	**0,90**[a]	-0,06	-0,13	0,00	0,04	-0,02	0,02	0,13	0,14	-0,08	0,16	0,24	0,21	0,19	0,17
P6 Nuevos residentes municipio	-0,24	0,03	-0,06	**0,73**[a]	-0,10	-0,03	0,19	0,05	0,18	-0,15	-0,12	0,08	-0,22	-0,12	-0,10	-0,11	-0,15
P7 Analfabetos >10 años	-0,09	-0,11	-0,13	-0,10	***0,92***[a]	0,00	0,15	-0,21	-0,31	0,09	-0,05	-0,04	-0,03	0,10	-0,09	-0,02	0,06
P8 Titulados medios-superiores	0,10	0,08	0,00	-0,03	0,00	***0,85***[a]	-0,23	-0,03	-0,10	-0,16	-0,76	-0,06	0,08	0,00	-0,05	0,00	-0,05
P9 Escolarización 14-24 años	-0,06	0,18	0,04	0,19	0,15	-0,23	**0,87**[a]	0,25	0,35	0,36	-0,19	0,11	-0,06	-0,10	0,11	0,06	0,13
P11 Parados antes ocupados	-0,01	-0,07	-0,02	0,05	-0,21	-0,03	0,25	**0,93**[a]	-0,04	0,02	0,10	0,03	-0,10	0,05	0,20	0,10	0,12
P12 Paro busca primer empleo	0,11	0,25	0,02	0,18	-0,31	-0,10	0,35	-0,04	**0,87**[a]	0,20	-0,03	-0,09	0,08	0,05	0,04	0,00	0,09
P14 Activas >15 años	-0,07	0,25	0,13	-0,15	0,09	-0,16	0,36	0,02	0,20	**0,75**[a]	-0,07	0,06	-0,03	-0,14	0,13	0,06	0,11
P15 Profesiones altas	0,03	-0,21	0,14	-0,12	-0,05	-0,76	-0,19	0,10	-0,03	-0,07	**0,84**[a]	0,09	0,10	-0,05	0,09	-0,14	0,10
P16 Profesiones bajas	-0,08	0,00	-0,08	0,08	-0,04	-0,06	0,11	0,03	-0,09	0,06	0,09	**0,96**[a]	0,01	0,04	0,08	-0,04	-0,05
P17 Terciario, comercio, hostel.	0,13	-0,12	0,16	-0,22	-0,03	0,08	-0,06	-0,10	0,08	-0,03	0,10	0,01	**0,76**[a]	0,07	0,00	0,13	0,07
P18 Terciario alto/finanzas	0,14	0,07	0,24	-0,12	0,10	0,00	-0,10	0,05	0,05	-0,14	-0,05	0,04	0,07	**0,94**[a]	0,11	0,17	0,08
P19 Agropecuario	0,02	0,02	0,21	-0,10	-0,09	-0,05	0,11	0,20	0,04	0,13	0,09	0,08	0,00	0,11	**0,67**[a]	-0,01	-0,43
P20 Vehículo privado trabajo	-0,18	0,19	0,19	-0,11	-0,02	0,00	0,06	0,10	0,00	0,06	-0,14	-0,04	0,13	0,17	-0,01	**0,83**[a]	-0,18
P23 Población Sección/Municipio	0,08	0,03	0,17	-0,15	0,06	-0,05	0,13	0,12	0,09	0,11	0,10	-0,05	0,07	0,08	-0,43	-0,18	**0,68**[a]

a. Medidas de adecuación de muestral (MSA)

La diagonal de la matriz de correlaciones anti-imagen contiene la medida de adecuación muestral de Kaiser-Meier-Olkin (KMO) para cada variable, un índice que se utiliza para comparar, para dos variables (j,j'), los coeficientes de correlación simples ($r_{jj'}$), los que hemos utilizado hasta ahora, con los de correlación parcial ($rp_{jj'}$), que tienen en cuenta las reflexiones precedentes:

$$KMO = \frac{\sum_{j=1}^{p}\sum_{j'=1}^{p} r_{jj'}^2}{\sum_{j=1}^{p}\sum_{j'=1}^{p} r_{jj'}^2 + \sum_{j=1}^{p}\sum_{j'=1}^{p} rp_{jj'}^2} \qquad \text{Ecuación 3}$$

siempre que $j \neq j'$.

La fórmula de hecho es relativamente simple: divide las correlaciones simples entre la suma de éstas y las correlaciones parciales[11], y se podría expresar como $r/_{r+rp}$. En la medida que la expresión $\sum_{j=1}^{p}\sum_{j'=1}^{p} r_{jj'}^2$ (suma de los coeficientes de correlación parcial al cuadrado entre todos los pares de variables) sea pequeña, tienda a cero, será señal de que el conjunto de los coeficientes de correlación parcial tienden a cero, se da una fuerte relación asociativa en el conjunto de variables y aconseja más el ACP. En el límite, cuando esta expresión de las correlaciones parciales es cero, el índice KMO toma el valor 1. En el caso contrario, cuando las correlaciones entre las variables son nulas, el índice vale cero; y, en general, cuando se dan valores pequeños de KMO indican que las correlaciones entre pares de variables no son explicadas por otras variables y es necesario reconsiderar el modelo factorial.

Por tanto, el índice varía entre 0 y 1, y según los sugerentes calificativos de Kaiser (1974) podemos establecer una escala con la siguiente valoración:

Valor de KMO	Valoración
0,9	Maravilloso
0,8	Meritorio
0,7	Intermedio
0,6	Mediocre
0,5	Miserable
Menos de 0,5	Inaceptable

El valor en este caso (ver la Tabla III.11.3) es bastante maravilloso (0,847), y nos permite considerar como válida la aplicación de una ACP. Por otra parte, la evaluación del índice KMO se puede ver a partir de cada una de las variables individualmente. La diagonal de la matriz de correlaciones anti-imagen contiene la medida de adecuación muestral (*Measures of Sampling Adequacy,* MSA) En este caso se incluye un único sumatorio de los pares de coeficientes donde aparece la variable en cuestión según la fórmula siguiente:

$$MSA_{x_{+j}} = \frac{\sum_{j\neq j'} r_{jj'}^2}{\sum_{j\neq j'} r_{jj'}^2 + \sum_{j\neq j'} rp_{jj'}^2} \qquad \textit{Ecuación 4}$$

que son los valores que aparecen en la diagonal de la matriz anti-imagen de correlaciones (Tabla III.11.4)[12]. En este caso los valores deben ser elevados para un buen ACP, y es de utilidad para considerar la posibilidad de suprimir alguna variable. En el ejemplo de los datos del Padrón de 1986 para la Región Metropolitana de

[11] La correlación parcial se define como $rp_{yx\cdot z} = \dfrac{r_{yx} - r_{yz}\cdot r_{xz}}{\sqrt{1-r_{yz}^2}\cdot\sqrt{1-r_{xz}^2}}$.

[12] Los valores de la diagonal de la matriz de covarianza anti-imagen se obtienen restando a 1 la correlación múltiple al cuadrado entre cada variable y las restantes variables del análisis. Representan, por tanto, una estimación de la unicidad de cada variable, es decir, una estimación de lo que cada variable tiene como propio y no comparte con las demás.

Barcelona ya se ha efectuado la selección y se puede observar que todas las variables tienen valores altos que refuerzan la idea de la pertinencia de la realización de una ACP.

En la práctica del análisis se trata de probar la inclusión o no de las variables candidatas a formar parte del ACP según el valor resultante de las medidas de adecuación muestral. Como criterio general se puede exigir un mínimo de 0,5 si al mismo tiempo el KMO alcanza un valor mínimo de 0,7 y, como comentaremos más tarde, el porcentaje de varianza explicada por los factores, alcanza el 70%.

2.3. Etapa 2: Extracción de los factores o componentes

Hasta ahora sólo hemos constatado la existencia de intercorrelación entre las variables, condición necesaria para la aplicación del procedimiento de la ACP. La realización de un ACP implica encontrar unos ejes factoriales o componentes con el cálculo de los valores propios o varianzas incorporados a cada uno de los ejes y decidir el número de ejes o factores a retener en el análisis. Se trata de encontrar las dimensiones latentes en el campo de variables considerado en relación a los individuos correspondientes. En nuestro ejemplo considerando 17 variables referidas a 3509 secciones censales de la Región Metropolitana de Barcelona.

Veamos cómo se calculan los ejes factoriales. Estos ejes son vectores en el espacio vectorial formado por las variables introducidas, es decir, partimos del espacio vectorial euclídeo engendrado por las p variables con las que se caracterizan un total de n unidades, configurando así una nube de puntos N_n^p. Para encontrar estos ejes o vectores se imponen varias condiciones:

a) Que acumulen o expliquen la máxima inercia o varianza total del sistema, por tanto, que la proyección de los puntos sobre estos nuevos ejes sea máxima.
b) Que esta varianza extraída para cada uno de los ejes se haga de manera jerárquica, es decir, sean ejes que de forma sucesiva y jerárquica acumulen la mayor cantidad de inercia; el análisis nos llevará además a retener unos pocos ejes, a costa de una pérdida de inercia o varianza, pero logrando el doble objetivo de "reducir la información y ganar en significación", en sencillez y relevancia explicativa.
c) Que sean base del sistema vectorial, es decir, que sean linealmente independientes (la correlación entre los factores es cero).
d) Y como condición técnica suplementaria que estos vectores sean unitarios.

A continuación mostraremos en el proceso de extracción de los factores estas cuatro condiciones. Antes definiremos el concepto de inercia, similar al de varianza, y presentaron la idea de cambio de ejes que conlleva la ACP.

2.3.1. *Extracción de los factores*

El punto de partida en un ACP es la matriz original X con el conjunto de variables originales p que son elegidas para el análisis. Este conjunto de variables configura una variabilidad total que no es más una medida de la información total contenida en ellas.

Esta variabilidad total se expresa a través de la matriz de varianzas y covarianzas *V* o bien a través de la matriz de correlaciones R. Consideraremos en general esta última en los análisis con el software pues facilita la interpretación de los resultados.

El objetivo es constatar en ese conjunto de variabilidad inicial, de diferencias entre los individuos, que se dan redundancias informativas, es decir, existen correlaciones entre grupos de variables que nos permite expresar el 100% de la información total inicial de las *p* variables en solo un número reducido de *m* factores o componentes que expresan la mayor parte del 100% de la información.

La redundancia informativa o las intercorrelaciones se concentran en los factores. Esta concentración no implica exactamente una agrupación de variables en factores, sino una combinación (lineal) de la mayor parte de la variabilidad de cada una de ellas en unos pocos factores que permite substituir la complejidad de *p* dimensiones por un número menor de *m* dimensiones. Así, por ejemplo, podríamos considerar *p=20* variables iniciales cuya variabilidad total inicial (20), una unidad por variable configurando 20 dimensiones iniciales, se convierte y reduce a 15 (lo que implica conservar el 75%) pero que se expresaría solamente en términos de, por ejemplo, 3 nuevas variables que son los factores. Pasamos así de un espacio vectorial de 20 dimensiones a otro de 3, con un coste del 25% de pérdida de información, pero que, como destacaremos, es el 25% menos importante pues son factores secundarios de diferenciación.

La información total inicial, el total de variabilidad, se expresa en el concepto de inercia. En este sentido podemos equiparar la idea de variabilidad con la de inercia. Desde un punto de vista geométrico la variabilidad o la inercia son distancias, dispersiones en el espacio. Así se define la inercia I_C en un espacio vectorial de *p* dimensiones que configura una nube de puntos con *n* individuos N_n^p , respecto de un centro *C*, como[13]:

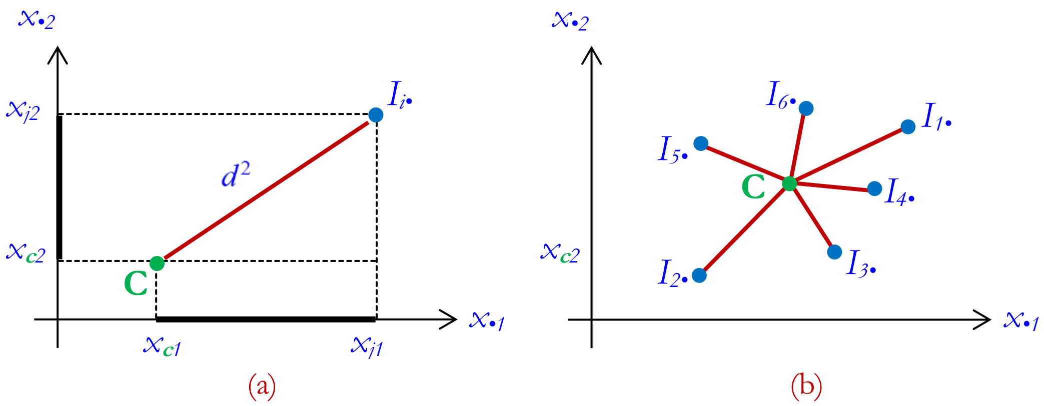
$$I_C = \sum_{i=1}^{n} d^2(I_{i\bullet}, C) = \sum_{i=1}^{n}\sum_{j=1}^{p}(x_{ij} - \overline{x}_{cj})^2 \qquad \textit{Ecuación 5}$$

es decir, es la suma de todas las distancias de cada individuo $I_{i\bullet}$ al centro *C*, definido éste por el punto que se obtiene de la media de todas las variables. En el Gráfico III.11.5 se representa en un espacio de dos dimensiones y para un individuo (gráfico a) y para un conjunto de individuos (gráfico b). La inercia es la distancia (cuadrática) d^2.

La inercia es pues toda la variabilidad de nuestros datos y es la suma de todas las varianzas multiplicada por el número de individuos: $I_C = n \cdot Tr(V)$, es decir, el número de casos por la traza de la matriz de varianzas y covarianzas (la suma de las varianzas de las *p* variables).

[13] La inercia es un concepto de la física y nos referimos de hecho a la "inercia con respecto al centro de masas" que no es más que la dispersión respecto de la media.

Gráfico III.11.5. Representación geométrica de la inercia

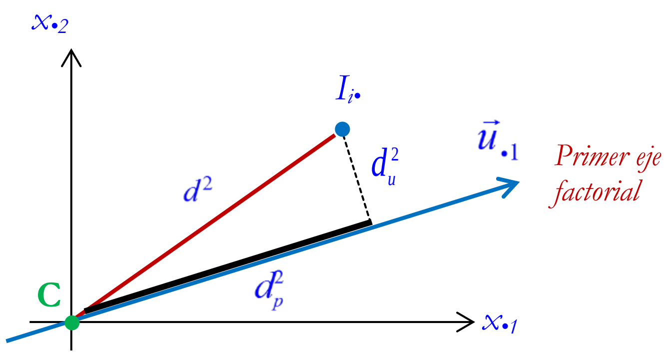


Cuando se colocan todos los individuos en el espacio la existencia de redundancias informativas y de correlaciones entre las variables se expresa en disposiciones alineadas de los individuos en ese espacio marcando direcciones (deformaciones) en el mismo, es decir, dispersiones en una dirección determinada. Encontrar los factores significa dar cuenta de estas direcciones en el espacio por donde hacer pasar el eje factorial o nueva variable factorial $F_{\bullet k}$.

Para encontrar esos ejes se utilizan los vectores directores unitarios que nos indican por donde hacer pasar el eje $\vec{u}_{\bullet j}$ de tal manera que se cumplan las condiciones anunciadas anteriormente: que acumulen la mayor parte de la inercia, de forma jerárquica y sean linealmente independientes.

Acumular la mayor parte de la inercia significa geométricamente que la proyección de los puntos-individuos sobre los nuevos ejes factoriales buscados sea la máxima: es la distancia d_p^2 del Gráfico III.11.6, por tanto, que la distancia del individuo al eje d_h^2 sea la más pequeña posible.

Gráfico III.11.6. Descomposición de la inercia



Así se procede a la descomposición de la inercia total, como suma de la distancia al eje más la proyección: el triángulo de la figura muestra de hecho que es la aplicación del Teorema de Pitágoras: $d^2 = d_p^2 + d_u^2$, extendido a todos los puntos de la nube.

En la figura se ha representado la dirección por donde pasa el primer eje factorial $\vec{u}_{\bullet 1}$. Éste es el primero que consigue acumular la mayor inercia posible, el que consigue mayor proyección o cercanía de los puntos. El segundo eje es el que consigue proyectar sobre él la mayor parte de la inercia restante que no acumula el primer eje, y así sucesivamente con el tercero, cuatro, etc. Los ejes o vectores que cumplen con estas características se denominan vectores propios ($\vec{u}_{\bullet j}$), y la inercia que acumulan, valores propios (λ_k), que como veremos inmediatamente expresan la varianza explicada por el factor, su importancia[14].

Como resultado por tanto de este proceso, y considerando la matriz de correlaciones R, se concluye que la inercia total se reparte de manera jerárquica en p valores propios λ_k con $k=1\ldots p$, asociados a cada eje o vector propio, siendo la suma de todos ellos igual la inercia total e igual al valor p. Esto nos permitirá interpretar cada valor propio como la proporción varianza explicada (VE):

$$VE(\lambda_k) = \frac{\lambda_k}{\sum_{k=1}^{p} \lambda_k} = \frac{\lambda_k}{p} \qquad \textit{Ecuación 6}$$

---

14 En términos algebraicos los vectores y los valores propios de obtienen diagonalizando la matriz de varianzas y covarianzas (Lozares y López-Roldán, 2000: 82-90). Encontrar los vectores propios implica hacer máximas las proyecciones lo que se expresa maximizando la suma de los productos escalares al cuadrado de todos los vectores-individuos y el vector propio. El cálculo en términos matriciales implica que el primer valor propio es el que maximiza la expresión $n \cdot U_{\bullet 1}^{'} \cdot V \cdot U_{\bullet 1}$. Esta expresión con los datos estandarizados implica considerar la matriz de correlaciones R en vez de V: $n \cdot U_{\bullet 1}^{'} \cdot R \cdot U_{\bullet 1}$. La resolución de la maximización implica obtener el primer valor propio λ_1 que maximiza la expresión $V \cdot U_{\bullet 1} = \lambda_1 \cdot U_{\bullet 1}$ o bien $(V - \lambda_1 I) \cdot U_{\bullet 1} = 0$ que es el mayor valor propio de V y así se continúa con el resto de vectores y valores propios maximizando la varianza residual y considerando vectores perpendiculares entre sí (linealmente independientes). Obtenidos los vectores propios se verifica que la matriz de varianzas y covarianzas V se puede expresar como: $V = U \cdot D_\lambda \cdot U^{'}$, donde U es la matriz de vectores propios y D_λ es la matriz diagonal de valores propios:

$$V = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1p} \\ u_{21} & u_{22} & \cdots & u_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ u_{p1} & u_{p2} & \cdots & u_{pp} \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix} \cdot \begin{pmatrix} u_{11} & u_{21} & \cdots & u_{p1} \\ u_{12} & u_{22} & \cdots & u_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ u_{1p} & u_{2p} & \cdots & u_{pp} \end{pmatrix}$$

De donde se deriva que la traza de V es la suma de los valores propios: $\mathrm{Tr}(V) = \lambda_1 + \lambda_2 + \ldots + \lambda_p = \sum_{k=1}^{p} \lambda_k$. Con este resultado y sabiendo como vimos que la inercia es $I_C = n \cdot Tr(V)$ se concluye que la inercia total es $I_C = n \cdot Tr(V) = n \cdot \sum_{k=1}^{p} \lambda_k$. Si consideramos las variables estandarizadas entonces $I_C = Tr(R) = \sum_{k=1}^{p} \lambda_k = p$.

2.3.2. *Reducción del número factores o componentes*

Hemos visto por tanto que con la extracción de los factores mediante el ACP, a partir de la matriz de correlaciones en particular, se obtiene que el primer factor o la primera componente es la combinación lineal de las variables originales que tiene en cuenta la mayor cantidad de varianza del conjunto, la segunda será la siguiente con más varianza pero no correlacionada con el anterior, y así sucesivamente. La varianza total en cada ACP será igual, si consideramos las variables estandarizadas (con media 0 y varianza 1), al número de variables p incluidas en el análisis.

Este resultado nos muestra que el número de vectores y valores propios coincide con el número de variables introducidas en el análisis, p. Considerar todos los valores propios o toda la varianza incorporada a los ejes, el 100% inicial, no nos reporta ninguna ganancia. Uno de los objetivos de la ACP es precisamente reducir el espacio de atributos (de variables) inicial a un número m, con $m<p$. Esta reducción implicará una pérdida de la inercia o de la varianza total de sistema de variables iniciales. Pero esta pérdida de varianza, o pérdida de información, implica una ganancia en sencillez, una reducción y una síntesis de la estructura significativa que se deriva de las variables originales consideradas y que se expresa en términos de unas nuevas variables que son los principales factores o dimensiones de diferenciación de los individuos.

En el ejemplo que estamos considerando, con 17 variables, si éstas están estandarizadas, la varianza total es de 17. Todas las variables en un inicio determinan un espacio de 17 dimensiones donde cada variable contribuye individualmente con un valor de 1, se dice que la comunalidad inicial de cada variable es 1, es "lo que aportan al común". Así, en un inicio, la proporción que representa cada variable sobre la variabilidad total es de 1 dividido por el número de variables (en este caso 1/17 es aproximadamente el 6%). Esta idea se refleja en la columna inicial de la Tabla III.11.5 de comunalidades que se adjunta, obtenida con el software SPSS.

Tabla III.11.5. Tabla de comunalidades inicial y de la extracción del ACP

	Inicial	Extracción
P1 Menores de 15 años	1,000	0,823
P2 Mayores de 65 años	1,000	0,829
P4 Inmigración fuera Cataluña	1,000	0,728
P6 Nuevos residentes municipio 81-86	1,000	0,711
P7 Analfabetos >10 años	1,000	0,571
P8 Titulados medios-superiores >20 años	1,000	0,820
P9 Escolarización 14-24 años	1,000	0,815
P11 Parados antes ocupados	1,000	0,700
P12 Paro busca primer empleo	1,000	0,598
P14 Activas >15 años	1,000	0,669
P15 Profesiones altas	1,000	0,850
P16 Profesiones bajas	1,000	0,435
P17 Terciario medio / comercio / hostelería	1,000	0,711
P18 Terciario alto / finanzas	1,000	0,650
P19 Agropecuario	1,000	0,764
P20 Vehículo privado trabajo	1,000	0,652
P23 Población Sección/Municipio	1,000	0,776

Cuando el ACP extrae los ejes factoriales, obtiene tantos ejes (componentes o dimensiones) como variables hay. Los valores propios, es decir, la varianza asociada a cada eje, también llamados autovalores, se reproducen en la Tabla III.11.6 de varianza total explicada. Se observa cómo estos valores que corresponden a cada factor o componente están ordenados de forma decreciente. La suma de todos los valores propios es igual a 17, el número de variables, pero ahora este total no es resultado de acumular o sumar la información o la variabilidad 1 de cada variable, sino que es el resultado de sumar la variabilidad acumulada por la importancia decreciente de cada factor. Ahora, la proporción que representa cada factor sobre la variabilidad total no es 1 dividido sobre el número de variables, los valores que se obtienen dan la importancia relativa y decreciente de los factores o componentes. Así se presenta en la columna % de la varianza y % acumulado que se presentan en la la Tabla III.11.6.

Como se puede observar el primer factor representa el 36% de toda la varianza, el segundo el 17,3%, el tercero del 11,1% y así de forma sucesiva y decreciente. Ahora se trata de determinar cuántos ejes nos quedamos, cuantos factores o componentes retendremos el análisis para explicar la estructura subyacente contenida al conjunto de las 17 variables iniciales.

Tabla III.11.6. Tabla de valores propios (autovalores) y de varianza explicada de la extracción de las componentes del ACP

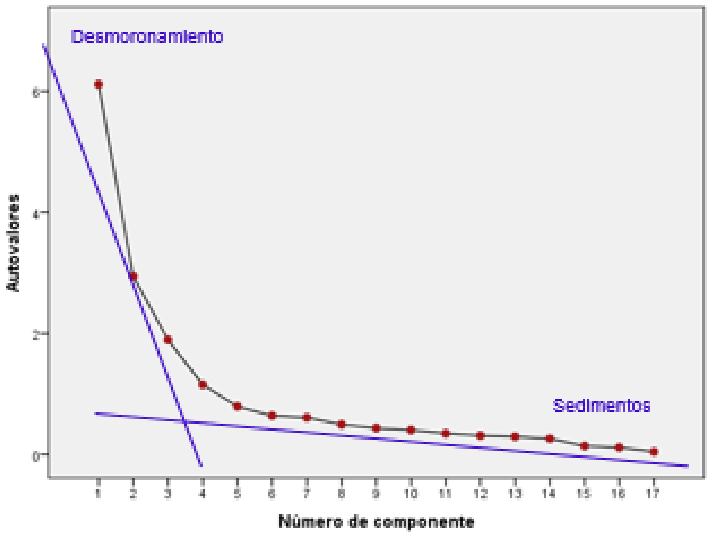
Componente o factor	Valores propios (autovalores) iniciales			Varianza explicada (sumas de las saturaciones al cuadrado)		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	6,112	35,952	35,952	6,112	35,952	35,952
2	2,943	17,310	53,261	2,943	17,310	53,261
3	1,894	11,139	64,400	1,894	11,139	64,400
4	1,153	6,782	71,182	1,153	6,782	71,182
5	0,793	4,666	75,848			
6	0,640	3,763	79,611			
7	0,609	3,580	83,191			
8	0,495	2,911	86,102			
9	0,437	2,571	88,674			
10	0,404	2,376	91,050			
11	0,349	2,051	93,101			
12	0,313	1,840	94,941			
13	0,296	1,741	96,681			
14	0,262	1,544	98,225			
15	0,141	0,828	99,053			
16	0,116	0,682	99,735			
17	0,045	0,265	100,000			

Existen diversos criterios alternativos para determinar el número de componentes a retener. Señalaremos cuatro criterios en nuestro análisis:

1. Considerar todos aquellos factores que tienen un valor propio superior a 1, pues supone considerar un factor que mejora la varianza proporcionada en un inicio para cada variable sola (Kaiser, 1960).
2. Considerar el número de ejes que acumulan en torno al 70% de la varianza total, cantidad que se considera equilibrada entre la pérdida de información (del 30%) y la ganancia en significación (el 70% retiene los principales factores de variabilidad).
3. Representar gráficamente los distintos factores y los valores propios asociados y observar el comportamiento de la curva resultante (gráfico de sedimentación). El número de ejes a retener viene determinado por el cambio de pendiente de la curva, donde está presente el cambio de continuidad de la curva. Es el llamado *scree test* (Cattell, 1966), en términos más coloquiales "test del codo", donde se sitúa el codo del brazo imaginario que dibuja la forma de la curva es el punto que determina el número de componentes.
4. Un cuarto criterio adicional y principal, según nuestro criterio, que debe presidir siempre nuestro razonamiento tiene que ver con la interpretabilidad y la pertinencia substantiva de los ejes obtenidos que en un momento pueden llevarnos a considerar más o menos en función del propio contenido de los factores y sus implicaciones en el análisis del fenómeno.

En el caso que nos ocupa nos encontramos con una coincidencia de los tres primeros criterios mencionados. Si consideramos los cuatro primeros ejes, en total estos acumulan el 71,2% de la varianza total; además los cuatro son los que tienen un valor propio por encima de la unidad; y tal como se observa en el gráfico de sedimentación (Gráfico III.11.7), el cambio de pendiente de la curva se produce a partir de la cuarta componente o factor.

Gráfico III.11.7. Gráfico de sedimentación del ACP



Por tanto, las 17 variables con las que hemos caracterizado la realidad social de la Región Metropolitana de Barcelona se pueden expresar en función de cuatro factores o componentes, cuatro dimensiones o cuatro nuevas variables, cada una de las cuales posee una importancia diferente, y en conjunto representan 71% de la variabilidad total. Por lo tanto, el 29% restante es información que es descartada, la menos relevante pues son factores secundarios de diferenciación, y no es contemplado en el análisis.

El efecto de pérdida de información se produce en conjunto y también se refleja en cada una de las variables. En la tabla de comunalidad anterior (Tabla III.11.5) vemos en la columna Extracción como cada una de las variables pasa de una comunalidad inicial de 1 a una comunalidad tras la extracción inferior a 1, y diferente en cada caso. Como destacaremos más tarde, los valores altos de comunalidad de las variables nos indican la mayor presencia de estas variables en los factores y en los resultados del análisis.

El paso siguiente será dar sentido interpretativo o substantivo a estos cuatro factores, se trata de identificarlos e interpretarlos según unos conceptos o criterios previos derivados de nuestro modelo de análisis, del conocimiento del fenómeno o de la investigación en torno al mismo. Pasamos a comentarlo dentro de la tercera etapa del proceso del análisis de componentes principales.

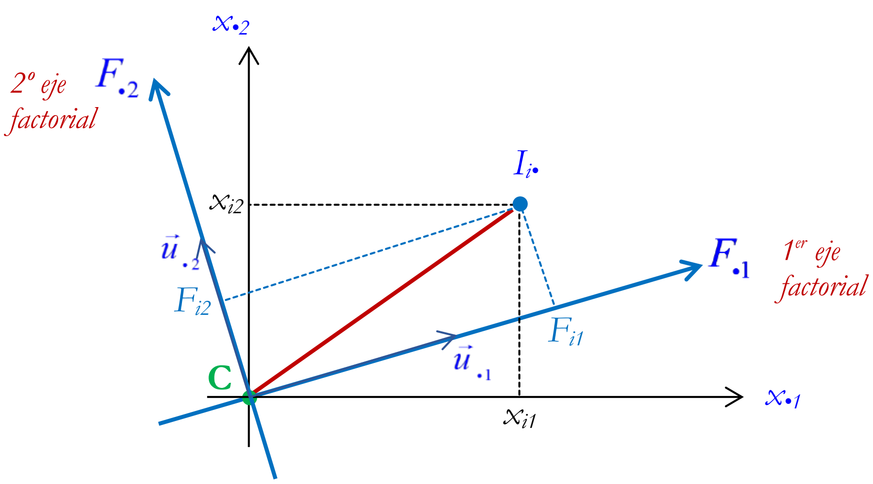
2.4. Etapa 3: Interpretación de los factores

2.4.1. *Relación entre los factores y las variables originales*

La interpretación de los factores o componentes que se obtienen a la ACP se hace mediante el análisis de la relación entre las variables originales y las componentes. De esta forma logramos, por un lado, recomponer las variables originales en los nuevos ejes para mostrar la estructura del espacio de atributos original; del otro, nos permite dar identidad a los factores de diferenciación del análisis. Para llegar a este resultado final es necesario precisar los conceptos de matriz de saturaciones y de comunalidad.

Con el análisis factorial obtenemos nuevas coordenadas para los individuos en los nuevos ejes o factores $F_{\bullet 1}, F_{\bullet 2}, \ldots, F_{\bullet p}$, lo cuales se obtienen como hemos visto a partir de las direcciones en el espacio que marcan los vectores propios $\vec{u}_{\bullet 1}, \vec{u}_{\bullet 2}, \ldots, \vec{u}_{\bullet p}$.

Gráfico III.11.8. Representación del cambio de ejes



Es decir, supone un cambio de sistema de referencia: pasamos de las coordenadas $x_{\bullet 1}, x_{\bullet 2}, \ldots, x_{\bullet p}$ de los individuos en las variables originales observadas a un nuevo sistema que es el espacio de los factores. En el caso de dos variables este cambio de ejes se puede representar como en el Gráfico III.11.8. Matemáticamente supone una transformación que se expresa como:

$$F = X \cdot U \qquad \textit{Ecuación 7}$$

Si estandarizamos las variables factoriales en la Ecuación 7 ésta se puede transformar en $F = X \cdot U \cdot D_{\lambda}^{1/2}$ donde $U \cdot D_{\lambda}^{1/2}$ es la denominada matriz de saturaciones o componentes, también llamada matriz factorial, que identificamos con A:

$$F = X \cdot U \cdot D_{\lambda}^{1/2} = X \cdot A \qquad \textit{Ecuación 8}$$

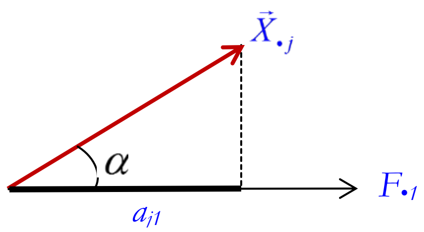
De la Ecuación 8 se deriva:

$$X = F^{S} \cdot A' \qquad \textit{Ecuación 9}$$

que es la expresión que nos permite interpretar los factores en función de las variables originales. Cada elemento de la matriz de saturaciones se denomina factor de carga (*factor loading*) y tiene la expresión: $a_{jk} = u_{jk}\sqrt{\lambda_k}$ y se interpreta como la proyección de la variable original $x_{\bullet j}$ en el nuevo factor $F_{\bullet k}$ (ver Gráfico III.11.9), es decir, con las variables originales estandarizadas, miden la correlación entre las variables originales y los factores:

$$a_{jk} = \cos \alpha_{x_{\bullet j}, F_{\bullet k}} = r_{x_{\bullet j}, F_{\bullet k}} \qquad \textit{Ecuación 10}$$

Gráfico III.11.9. Representación geométrica de la relación entre variables originales y los nuevos ejes



La Tabla III.11.7 presenta la matriz de saturaciones que se obtienen del ejemplo de análisis de la Región Metropolitana de Barcelona. Este el principal material informativo para la interpretación e identificación de los factores que se obtienen.

Tabla III.11.7. Matriz de saturaciones o de componentes (A)

	Componente			
	1	2	3	4
P15 Profesiones altas	0,906	-0,016	-0,124	-0,115
P8 Titulados medios-superiores >20 años	0,882	0,033	-0,153	-0,134
P9 Escolarización 14-24 años	0,863	-0,017	-0,182	-0,190
P4 Inmigración fuera Cataluña	-0,782	0,030	-0,340	-0,019
P18 Terciario alto / finanzas	0,773	-0,138	-0,132	0,127
P7 Analfabetos >10 años	-0,751	-0,076	0,041	0,016
P12 Paro busca primer empleo	-0,733	-0,054	-0,119	-0,207
P11 Parados antes ocupados	-0,728	-0,352	-0,076	0,202
P16 Profesiones bajas	-0,623	0,100	-0,080	-0,176
P20 Vehículo privado trabajo	0,159	0,772	0,040	-0,171
P1 Jóvenes <15 años	-0,389	0,756	-0,299	0,102
P2 Viejos mayores de 65 años	0,382	-0,727	0,391	-0,025
P6 Nuevos residentes municipio 81-86	0,186	0,599	0,162	0,540
P19 Agropecuario	-0,043	0,384	0,777	-0,107
P23 Población Sección/Municipio	-0,007	0,485	0,726	-0,116
P17 Terciario medio / comercio / hostelería	-0,110	-0,463	0,373	0,587
P14 Activas >15 años	0,410	0,357	-0,347	0,503

En color se han destacado los valores más altos asociados a cada uno de los factores o componentes. Cada fila de esta matriz contiene los coeficientes o factores de carga que expresan la variable original en función de los factores estandarizados. Así, por ejemplo, la variable P15 se puede expresar como:

$$\vec{P}_{15} = 0{,}906 \cdot \vec{F}_1 - 0{,}016 \cdot \vec{F}_2 - 0{,}124 \cdot \vec{F}_3 - 0{,}115 \cdot \vec{F}_4$$

que proviene de la *Ecuación 9*, y donde $\vec{F}_1, \vec{F}_2, \vec{F}_3, \vec{F}_4$ son los vectores o nuevos ejes, y los coeficientes son los a_{jk} de la matriz de saturaciones nos indican el peso que tiene cada factor en cada una de las variables, como en una ecuación de regresión.

Así pues, la matriz de saturaciones o de componentes se puede interpretar en cuatro sentidos:

a) En primer lugar, como hemos destacado, los factores de carga o valores de la matriz de saturaciones son los coeficientes de correlación entre las variables originales y los factores. Cuanto más elevado es el coeficiente más relevante es la variable para configurar el factor y también mayor será su proyección sobre el eje.

b) Los coeficientes son también las coordenadas de las variables originales en el espacio de las variables factoriales. Obtendremos así los gráficos factoriales que son una ayuda muy útil y visual para la interpretación del contenido de los factores obtenidos en el análisis.

c) Por otro lado, si consideramos el producto $A'A$, es decir, si elevamos al cuadrado y sumamos los coeficientes de cada columna de la matriz de saturaciones obtenemos los valores propios, es decir, la varianza incorporada en cada eje o varianza explicada por el factor o componente:

$$\lambda_k = \sum_{j=1}^{p} a_{jk}^2 \qquad \text{Ecuación 11}$$

En el ejemplo que tratamos, si consideramos la columna del primer factor rotado F_1, obtenemos el primer valor propio λ_1 a partir de las 17 variables originales:

$$\lambda_1 = \sum_{j=1}^{17} a_{j1}^2 = 0{,}906^2 + 0{,}882^2 + 0{,}863^2 - 0{,}782^2 + \cdots - 0{,}110^2 - 0{,}410^2 = 6{,}072$$

d) Por su parte, si consideramos el producto AA', es decir, si elevamos al cuadrado y sumamos los coeficientes de cada fila de la matriz de saturaciones obtenemos los valores de comunalidad de cada variable, es decir, la parte de cada variable explicada por el conjunto de los factores. Si retenemos m factores, con $m<p$, entonces es la comunalidad referida a los m factores es:

$$h_j^2 = \sum_{k=1}^{m} a_{jk}^2 \qquad \text{Ecuación 12}$$

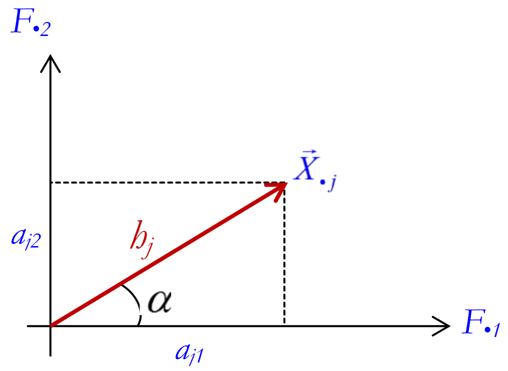
En el ejemplo que tratamos, si consideramos la comunalidad de la variable P15:

$$h_{P15}^2 = \sum_{k=1}^{4} a_{P15k}^2 = 0{,}906^2 - 0{,}016^2 - 0{,}124^2 - 0{,}115^2 = 0{,}850$$

En este sentido se interpreta que los factores de carga son los coeficientes resultantes de una ecuación de regresión múltiple, donde la variable original sería la variable dependiente y los factores las variable independientes. Como estos factores están incorrelacionados, los coeficientes no dependerán el uno del otro y representarán las contribuciones únicas de cada factor o la correlación entre ambos, factor y variable. De esta forma podemos calcular la proporción de varianza de cada variable explicada por el modelo de m factores.

Las comunalidades y su distribución entre los factores son interesantes desde el punto de vista de la interpretación, pues es la parte de la contribución de cada variable a estructurar el sistema de 4 factores. En consecuencia, comunalidades altas tienen importancia en la "creación" de los cuatro factores. Variables con comunalidad baja contribuyen poco a forma o estructurar el espacio, contribuyen poco a dispersar la nube de puntos. De hecho geométricamente (véase el Gráfico III.11.10 para el caso de dos ejes factoriales) la comunalidad se interpreta como la longitud del vector-variable en el espacio de las componentes (la distancia h_j en el gráfico). Una comunalidad alta implica un alejamiento del centro de coordenadas, una comunalidad baja significa una ubicación cercana al centro.

Gráfico III.11.10. Representación geométrica de la comunalidad



La mayor parte de las variables del ejemplo contribuyen fuertemente a la creación de los factores elegidos, según se puede ver en la Tabla III.11.5. Tan sólo la variable P16 tiene una comunalidad inferior a 0,5. Once variables tienen valores superiores a 0,7 y las variables con mayor comunalidad, superior a 0,8, son: P15 (profesiones altas), P1 y P2 (la edad) y P8, P9 (titulación superior y escolarización).

Procederemos a continuación a interpretar los factores que se obtienen. Este ejercicio analítico-interpretativo lo podríamos realizar con la Tabla III.11.7 anterior pero disponemos de una opción adicional destinada a la interpretación: la rotación de los factores, que presentaremos y aplicaremos previamente en el apartado siguiente.

2.4.2. *Interpretación de los factores*

La interpretación de los factores o componentes nos permitirá dar identidad a los ejes factoriales en función de las variables originales, descubriendo así la estructura de relaciones del espacio de atributos original e identificando la estructura emergida. La interpretación puede hacerse a partir de la matriz de saturaciones o de componentes original que hemos visto o bien puede hacerse a partir de la matriz de saturaciones que resulta de la rotación de los ejes.

La rotación de los ejes factoriales obtenidos en la extracción inicial se realiza con el objetivo de identificar con más claridad la relación o el modelo subyacente que se establece entre los factores y las variables. La rotación de los factores consiste en una transformación de la matriz factorial original en otra más simple que adecua mejor los ejes al aproximarse a las variables correlacionadas, se facilita así la interpretación de la estructura de los datos, acentuando el carácter de los factores, sin alterar la bondad de ajuste de la solución factorial. Las comunalidades y los porcentajes de varianza explicada se mantienen inalterados, simplemente se redistribuye la varianza explicada entre los factores.

Las rotaciones pueden ser ortogonales (rectangulares) u oblicuas. La rotación ortogonal se hace mediante un giro de los ejes manteniendo ángulos de 90º, con ejes perpendiculares. Esta transformación mantiene las distancias y deja inalterable la comunalidad pero no los valores propios. Dentro de las rotaciones ortogonales hay además diferentes métodos:

- Varimax, que minimiza el número de variables que tienen un factor o componente de saturación sobre una variable, acentuando así las que lo tienen más elevado. Las componentes quedan más limpias pues tienen sobre ellas las variables que más peso tienen eliminando sobre dichas componentes las variables intermedias. Es el procedimiento más utilizado.
- Quartimax, que minimiza el número de factores que corresponden a una variable. Se trata de que cada variable proyecte factores o componentes diferentes, dentro de los límites del método. La lógica de la reducción de un análisis exploratorio no va en esta dirección, se utiliza cuando las variables introducidas poseen determinados grados de independencia supuesta.
- Equimax, se trata de una combinación de las dos técnicas precedentes.

La rotación oblicua no conserva la ortogonalidad y, en consecuencia, tampoco se mantiene la comunalidad de cada variable, rompiendo así la incorrelación de los factores. Este procedimiento de rotación y sus procedimientos particulares (Oblimin, Oblimax, Promax) es más útil con modelos previos, en una lógica de análisis factorial confirmatorio y no se emplean en el ACP.

En el ejemplo de los datos censales de la Región Metropolitana de Barcelona se ha procedido a una rotación ortogonal por el procedimiento varimax. Como resultado de la transformación la cantidad global de varianza explicada por los cuatro esos se mantiene invariable e igual al 71,2%. Lo que sí cambia es la distribución de la varianza entre los factores o componentes. Este efecto de redistribución se implica calcular de nuevo los valores propios y ver la redistribución porcentual, esta es la información que se ha destacado en color en la Tabla III.11.8, donde aparecen los valores iniciales sin rotar y los nuevos como resultado de la rotación.

Tabla III.11.8. Tabla de valores propios (autovalores) y de varianza explicada tras la rotación de las componentes del ACP

Componente o factor	Valores propios (autovalores) iniciales			Varianza explicada					
				Sin rotación			Con rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	6,112	35,952	35,952	6,112	35,952	35,952	6,072	35,719	35,719
2	2,943	17,310	53,261	2,943	17,310	53,261	2,199	12,935	48,654
3	1,894	11,139	64,400	1,894	11,139	64,400	2,082	12,247	60,901
4	1,153	6,782	71,182	1,153	6,782	71,182	1,748	10,281	71,182
5	0,793	4,666	75,848						
6	0,640	3,763	79,611						
7	0,609	3,580	83,191						
8	0,495	2,911	86,102						
9	0,437	2,571	88,674						
10	0,404	2,376	91,050						
11	0,349	2,051	93,101						
12	0,313	1,840	94,941						
13	0,296	1,741	96,681						
14	0,262	1,544	98,225						
15	0,141	0,828	99,053						
16	0,116	0,682	99,735						
17	0,045	0,265	100,000						

Como consecuencia de la rotación disminuye mínimamente el peso del primer eje, y un poco más el segundo y aumentan los dos últimos. Para una lectura del conjunto de los cuatro componentes y de su distribución interna podemos calcular el tanto por ciento que representa cada eje sobre el total del cuatro retenidos:

	Varianza explicada recalculada		
Factor	Total	% de la varianza	% acumulado
1	6,072	50,2	50,2
2	2,199	18,2	68,4
3	2,082	17,2	75,6
4	1,747	14,4	100,0

Los factores rotados dan lugar a nuevos valores de carga o proyecciones sobre los ejes que se recogen en la matriz factorial rotada. La matriz de saturaciones o de componentes se modifica y da lugar a la nueva que se presenta en la Tabla III.11.9.

Finalmente procederemos a dar identidad a los ejes factoriales obtenidos ya la estructura subyacente al espacio de atributos a partir de la lectura de la tabla de componentes o saturada tras la rotación varimax. Recordemos que esta matriz nos proporciona las correlaciones entre las variables originales y las componentes; son por tanto las proyecciones de las variables sobre las componentes.

Tabla III.11.9. Matriz de saturaciones o de componentes rotados del ACP

	Componente			
	1	2	3	4
P15 Profesiones altas	0,915	0,064	-0,092	-0,016
P8 Titulados medios-superiores >20 años	0,892	0,123	-0,094	-0,002
P9 Escolarización 14-24 años	0,880	0,133	-0,134	-0,072
P4 Inmigración fuera Cataluña	-0,771	0,218	-0,292	-0,009
P18 Terciario alto / finanzas	0,763	-0,146	-0,190	0,105
P7 Analfabetos >10 años	-0,749	-0,045	-0,006	-0,092
P12 Paro busca primer empleo	-0,737	-0,293	-0,258	-0,075
P11 Parados antes ocupados	-0,710	0,163	-0,106	-0,237
P16 Profesiones bajas	-0,605	0,232	-0,007	-0,124
P20 Vehículo privado trabajo	-0,161	-0,811	0,041	0,161
P1 Jóvenes <15 años	0,385	-0,690	0,033	-0,451
P2 Viejos mayores de 65 años	-0,400	0,628	0,050	0,516
P6 Nuevos residentes municipio 81-86	0,163	0,614	0,401	0,295
P19 Agropecuario	-0,016	0,087	0,874	0,069
P23 Población Sección/Municipio	-0,052	-0,010	0,872	0,011
P17 Terciario medio / comercio / hostelería	0,128	0,044	0,330	0,764
P14 Activas >15 años	0,366	0,108	-0,219	0,689

Rotación varimax con normalización de Kaiser

Una ayuda inestimable para la interpretación de los factores y de la información de la matriz de saturaciones es la representación gráfica que relaciona los diferentes factores en representaciones bidimensionales o tridimensionales. En estos gráficos, llamados gráficos factoriales, se consideran las nuevas variables factoriales o componentes como los ejes sobre el plano del cual se proyectan las respectivas coordenadas o valores de carga para cada variable en relación a cada una de las componentes. En ellos se trata de identificar los grupos de variables que se forman teniendo en cuenta que aquellas que tienen mayor carga o proyección aparecerán al final de un eje. Se adjuntan dos representaciones gráficas de las componentes rotadas, de los factores 1 y 2 (Gráfico III.11.11) y de los factores 3 y 4 (Gráfico III.11.12).

Junto con las representaciones gráficas se pueden sugerir varios criterios para encontrar la identidad de los ejes factoriales:

a) Empezar por dar identidad al eje de mayor peso teniendo este dato presente para darle la importancia que le corresponde.
b) Dentro de las variables con máxima comunalidad elegir las de mayor valor sobre el eje analizado y mínima sobre los demás, tanto valores positivos como negativos.
c) Establecer una escala de variables que recorran todo el eje, independientemente de su comunalidad. Primero para las variables más cercanas a él, y después para otros siempre que se mantengan constantes en relación a los otros ejes. Este criterio es primordial pues marca la identidad latente de la dimensión que se busca.
d) Agrupar las variables por proximidad, sea en todos los ejes, sea en pares de ellos. Así obtendremos una agrupación de las mismas. Si además ubicamos estos grupos sobre la representación de los ejes y contribuirá a comprender mejor la estructura de la realidad analizada.
e) Establecer agrupaciones de variables según comunalidades.

A continuación apuntaremos las principales características identificativas de los ejes factoriales o de las dimensiones básicas que dan cuenta de la diversidad social de la Región Metropolitana de Barcelona a partir de los datos del Padrón de Habitantes de 1986.

El factor 1 representa el 36% de la varianza, al 50% si tenemos en cuenta sólo los cuatro ejes factoriales. Tiene en consecuencia un peso considerable y será decisivo a la hora de configurar la estructura de las variables elegidas. En uno de los extremos de esta dimensión aparecen categorías profesionales y niveles educativos altos: *profesiones altas*, *títulos medios y superiores*, *escolarización entre 14 y 24 años*, *terciario-alto y finanzas*. En el otro extremo tenemos categorías profesionales y niveles educativos bajos junto a elevados índices migratorios y de paro: *analfabetos con más de 10 años*, *presencia migratoria*, *desempleados antes ocupados*, *profesiones bajas*.

Ambos extremos dan contenido a una dimensión cargada por variables que van en la dirección de la categoría profesional y el nivel educativo que nos permite definir el primer eje como categoría socioprofesional y cultural. El hecho de que aparezcan variables referidas a la inmigración y al empleo hace perfilar mejor la naturaleza del eje, pues estas variables se arropan con las categorías bajas. Se trata de un eje que marca también la integración en el mundo laboral y el origen, fenómeno muy vinculado a la categoría sociocultural. Podríamos por tanto hablar de una dimensión de categoría social entendida como un compuesto de integración laboral ocupacional, categoría socioprofesional y cultural y origen inmigrante.

Gráfico III.11.11. Gráfico de componentes en el espacio rotado de los factores 1 y 2

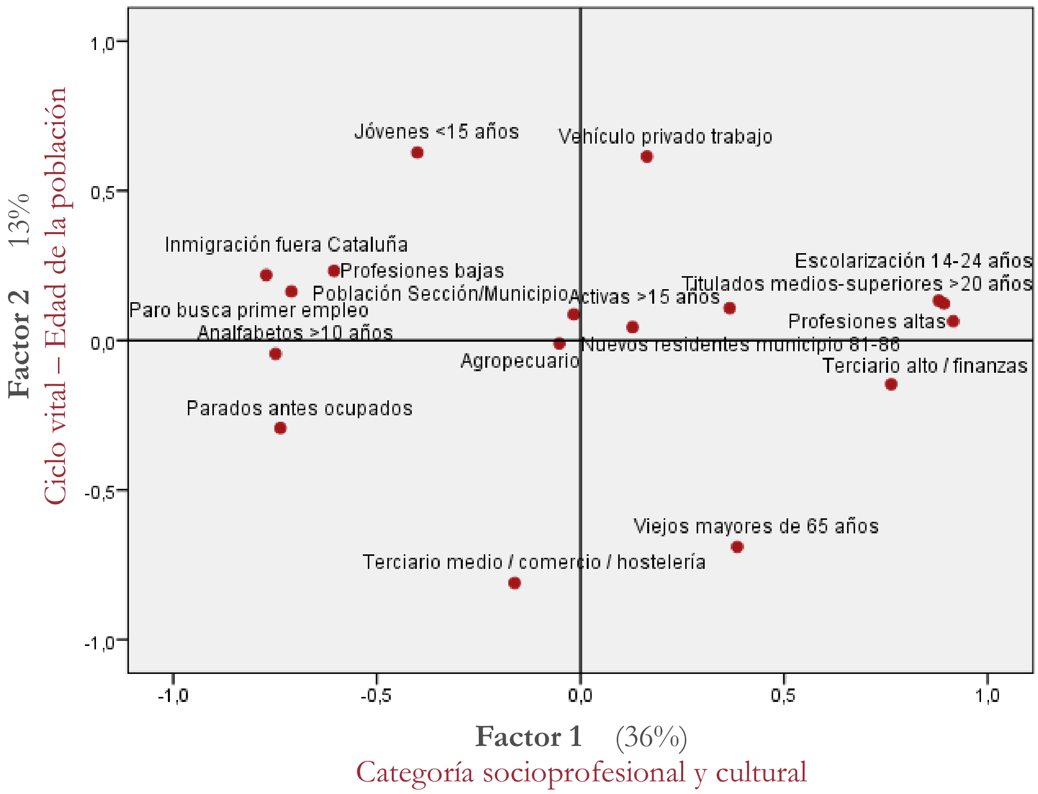
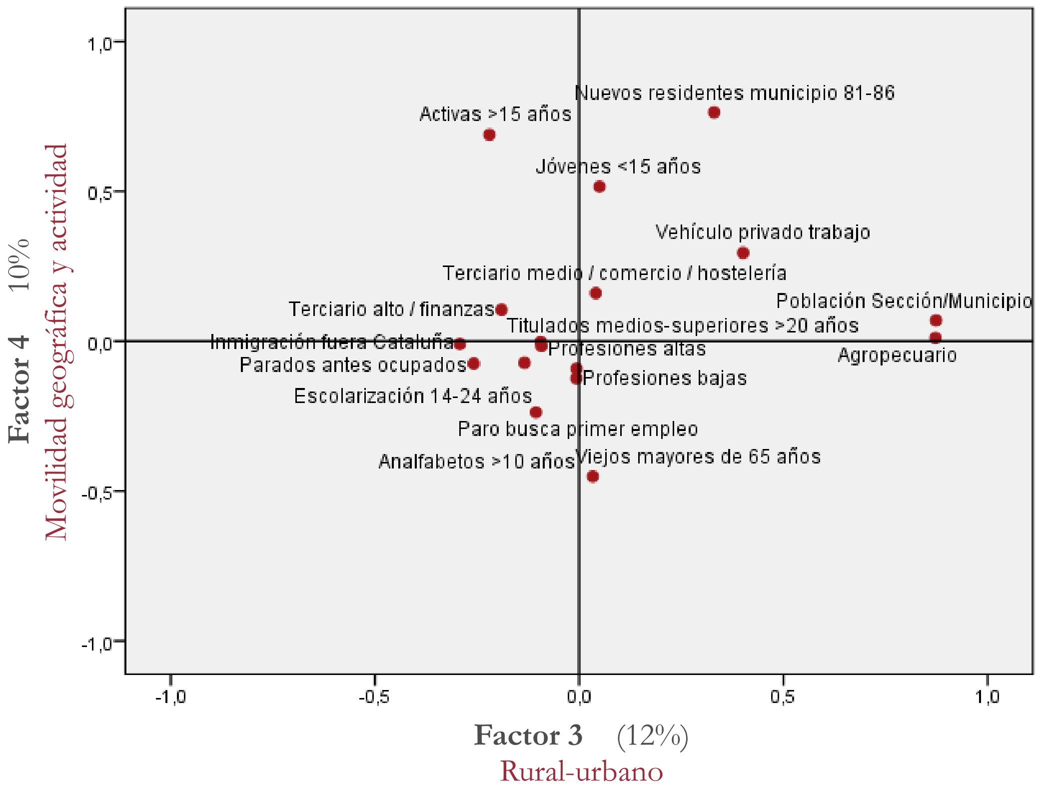


Gráfico III.11.12. Gráfico de componentes en el espacio rotado de los factores 3 y 4



El factor 2 no llega a representar el 20% de la varianza explicada por el conjunto de los cuatro factores. Básicamente se trata de un eje que va de un polo en el que se sitúan variables como *jóvenes menores de 15 años* y uti*lizó del vehículo privado para ir al trabajo*, y en el otro *mayores de 65 años* y *terciario medio, comercio y hostelería*. La identidad del eje está marcada básicamente por la edad, por lo tanto diferencia las secciones censales con una mayor proporción de población joven o infantil, más periféricas o residenciales que precisan del vehículo privado para ir al trabajo, de las secciones con una mayor presencia de personas mayores, propia de zonas urbanas o más antiguas donde predomina el comercio. Se trata pues de un eje marcado por el ciclo vital y/o de la edad de la población.

El factor 3 tiene un peso del 17% en el conjunto de los cuatro ejes. La naturaleza del eje aparece bien definida pues las variables que se proyectan en uno de los extremos tienen poca incidencia sobre los demás. En uno de los lados aparecen nítidamente las variables que identifican a los *municipios pequeños* y el *sector agropecuario*, del otro, el resto poblacional. Se trata de una dimensión metropolitana o de “metropolización", de oposición rural-urbano, que diferencia a un grupo de secciones de zonas rurales y periféricas del ámbito urbano metropolitano mayoritario.

El factor 4 representa el 14% restante. Uno de los extremos aparece definido por: *nuevos residentes*, *mujeres activas* y *jóvenes*, en el otro podemos ver que se disponen las variables: *parados antes ocupados* y *mayores*. Son perfiles no tan nítidos pero se podría calificar la dimensión de dinamismo social que recoge elementos de movilidad geográfica y actividad.

2.5. Etapa 4: Puntuaciones factoriales

Finalmente resulta de interés calcular las puntuaciones factoriales de cada caso, individuo o unidad considerada, es decir, los valores de los individuos en los ejes o nuevas variables factoriales. Estas nuevas variables factoriales podrán ser reutilizadas posteriormente en otros tipos de análisis, en particular para proceder a hacer un análisis de clasificación o de regresión lineal.

Conocida la matriz inicial X (de casos y variables iniciales), las componentes de los nuevos ejes U es posible a través de la ecuación $Y=X\cdot U$ que nos permite calcular las puntuaciones factoriales o componentes de los individuos a los nuevos ejes.

Estas puntuaciones factoriales nos permiten igualmente visualizar los casos o individuos en el espacio factorial de dos o tres dimensiones. Cuando los individuos son pocos y tienen una clara identidad (son agrupaciones como territorios, asociaciones,…) se puede etiquetar y analizar su ubicación. En nuestro ejemplo son muchos casos y carecen en general de identidad al corresponderse con unidades muy pequeñas de territorio. En el 0 y el Gráfico III.11.14 se reproducen los gráficos de dispersión de las 3509 secciones censales en el espacio factorial que definen, primero bidimensional de los factores 1 y 2, y en segundo término la representación gráfica tridimensional de los factores 1, 2 y 3.

Gráfico III.11.13. Gráfico de los individuos (secciones censales) en el espacio factorial rotado de los factores 1 y 2

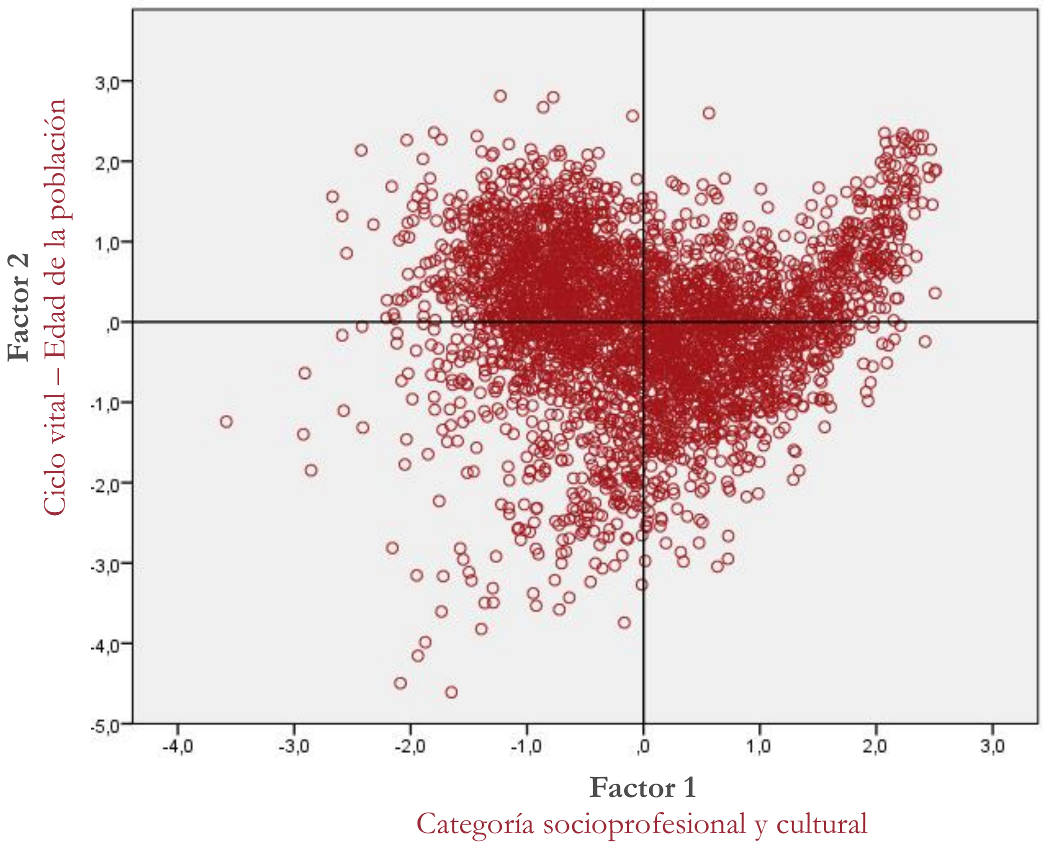
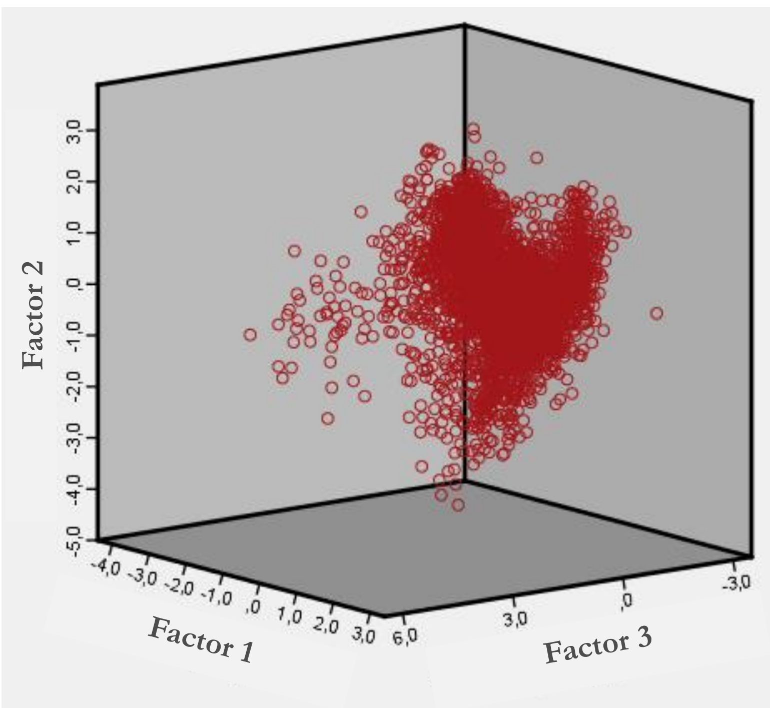


Gráfico III.11.14. Gráfico de los individuos (secciones censales) en el espacio factorial rotado de los factores 1, 2 y 3



En este caso un complemento gráfico de enorme interés son los mapas para dibujar zonas sociales o de estratificación social a partir de una unidad como la sección censal que se define precisamente como unidad territorial. En el próximo capítulo daremos cuenta de la estratificación y de los mapas sociales al hablar del análisis de clasificación. No obstante, los factores también pueden representarse en mapas sociales. Presentamos seguidamente un segundo ejemplo sobre la estratificación social de la ciudad de Buenos Aires que ilustrará este resultado.

2.6. Ejemplos de aplicación

2.6.1. *Los ejes de diferenciación de la ciudad de Buenos Aires*

Se presenta seguidamente la estructura socio-habitacional de la Ciudad de Buenos Aires inspirado en el modelo del Arquitecto Torres que ha sido reducido para adaptarlo a un esquema comparativo de tres censos15. El análisis se realiza a partir de información agregada en unidades geográficas denominados radios censales.

Con el análisis de componentes principales se pretende reducir la información original para obtener factores que sinteticen, por combinación de 16 variables censales iniciales seleccionados, la estructura socio-habitacional de la ciudad. Las variables utilizadas para el análisis se presentan a continuación.

Variables	**2010**	
	Media	Desviación
% hogares con hacinamiento <0,5	32,2	9,7
% hogares con hacinamiento 1,5-2	5,4	2,6
% hogares con hacinamiento 2-3	8,6	7,7
% hogares con hacinamiento +3	1,5	2,8
% población extranjera	12,5	9,8
% personas que nunca asistió a la escuela	1,0	1,0
% personas con estudios primarios	21,0	8,4
% personas con estudios superior	10,4	3,1
% personas con estudios universitarios	27,3	11,6
% de departamentos	68,9	27,4
% de inquilinatos	3,0	6,3
% de ranchos	0,3	1,4
% de inquilinos	29,9	11,1
% de propietarios de la vivienda y del terreno	56,4	13,0
Densidad del radio	27.490	19.275
Total (nº de radios censales)	**3552**	

Fuente: Censo de Población de la Ciudad de Buenos Aires 2010.

[15] Para conocer la evolución de los censos se puede consultar "Trazando el mapa social de Buenos Aires: dos décadas de cambios en la Ciudad", *Población de Buenos Aires. 2015*, vol. 12, nº 21, p. 7-39. http://ddd.uab.cat/pub/artpub/2015/132095/pobbueair_a2015n21p7iSPA_postprint.pdf. El Arquitecto Horacio A. Torres fue urbanista, prestigioso investigador del Consejo Nacional de Investigaciones Científicas y Técnicas de la Argentina (CONICET) y Profesor-Investigador de la Facultad de Arquitectura, Diseño y Urbanismo de la Universidad de Buenos Aires (FADU-UBA) que contribuyó al desarrollo del análisis del espacio urbano de la Ciudad así como del Gran Buenos Aires.

Aplicamos el Análisis de Componentes principales y retenemos tres dimensiones, linealmente independientes y ordenadas jerárquicamente, con las que identificamos la estructura de relaciones de las variables originales y que acumulan la mayor parte de la varianza explicada, expresando así los factores principales de diferenciación social.

En la tabla siguiente se recoge la información de la varianza explicada por los tres factores retenidos que alcanzan el 74,1% . El valor del KMO es de 0,871 y usando los calificativos sugeridos por Kaiser (1974) estaríamos en una situación de un ajuste "meritorio".

	Valores propios (autovalores) iniciales		
Componente o factor	Total	% de la varianza	% acumulado
1	4,211	28,1	28,1
2	3,948	26,3	54,4
3	2,952	19,7	74,1
Total		**74,1**	

A continuación se trata de interpretar los factores obtenidos. Los valores destacados en negrita en tabla siguiente son las variables que contribuyen en mayor medida a conformar cada factor.

Relación entre variables originales y factores del ACP. Componentes rotados

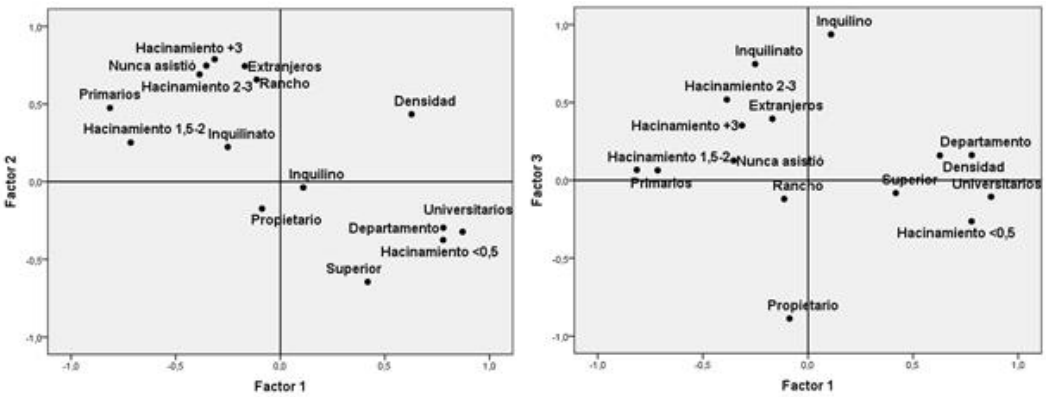
2010	Factor 1	Factor 2	Factor 3
Universitarios	**0,871**	-0,323	-0,106
Primarios	**-0,814**	0,475	0,067
Departamento	**0,779**	-0,296	0,162
Hacinamiento <0,5	**0,778**	-0,375	-0,263
Hacinamiento 1,5-2	**-0,715**	0,252	0,065
Densidad	**0,627**	0,434	0,160
Hacinamiento +3	-0,314	**0,788**	0,353
Nunca asistió	-0,354	**0,747**	0,127
Extranjeros	-0,169	**0,744**	0,395
Hacinamiento 2-3	-0,386	**0,691**	0,519
Rancho	-0,113	**0,657**	-0,119
Superior	0,418	**-0,645**	-0,082
Inquilino	0,110	-0,038	**0,937**
Propietario	-0,088	-0,173	**-0,888**
Inquilinato	-0,251	0,223	**0,747**

Fuente: Fachelli, Goicoechea y López-Roldán (2015)

El Factor 1 se interpreta como una dimensión de diferenciación socio-económica junto a densidad poblacional. A partir de los indicadores considerados vemos como se oponen en el eje dos perfiles claramente opuestos. El nivel más bajo de hacinamiento se superpone con el hecho de habitar en departamentos y en zonas altamente pobladas junto con la alta presencia de personas con estudios universitarios y superiores, atributos que expresan un perfil característico de los lugares de la ciudad donde habitan las personas de más alta posición social. En el polo opuesto, observamos el perfil contrapuesto, el de personas con una posición social caracterizada, en particular, por niveles medios y altos de hacinamiento y bajos niveles educativos, en zonas de la ciudad con una baja densidad de población. Resumimos así esta dimensión bajo la etiqueta de nivel socioeconómico y densidad poblacional.

Los gráficos factoriales siguientes presentan la posición en el espacio factorial, después de aplicar una rotación *varimax*, de las variables que intervienen en la conformación de los tres factores en cada uno de los años bajo estudio.

Gráficos factoriales en el espacio rotado. Censo 2010



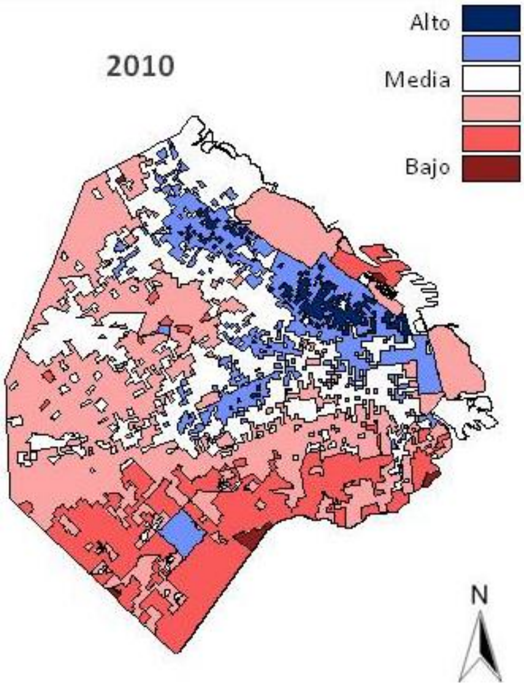
Fuente: Fachelli, Goicoechea y López-Roldán (2015)

Dicha identificación se justifica, principalmente, por la alta correlación entre la dimensión de nivel sociohabitacional y nivel socio-económico, corroborada a partir de varias investigaciones precedentes (Goldemberg, Fisherman y Torres, 1967; Torres, 1983, 1993 y 1999; Goicoechea, 2014).

La representación de dicho factor en el mapa de Buenos Aires da cuenta de la tradicional lógica de diferenciación de la ciudad a partir de la presencia de los sectores poblacionales de mayor nivel socio-económico localizados en la zona norte; mientras que los menores niveles socioeconómicos se localizan en la zona sur.

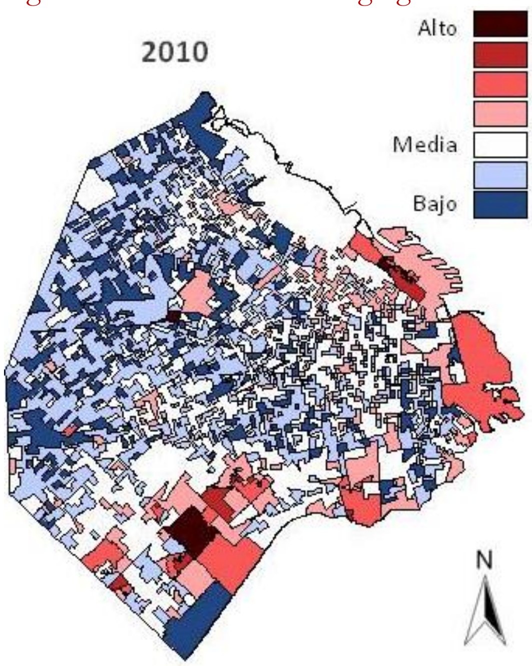
El Factor 2, independiente del primero, viene marcado sobre todo por la polarización entre zonas de la ciudad donde sobresale la presencia de viviendas de tipo rancho, donde también se concentran tres rasgos adicionales: la mayor proporción de población residente extranjera, los más altos niveles de hacinamiento y los niveles más bajos de estudios. El extremo opuesto, el de propietarios, perfil mayoritario en la ciudad, es el que también tiende a concentrar a la población con mayores niveles educativos y sin hacinamiento. Es decir, se trata de una diferenciación entre radios donde predominan los propietarios de la vivienda y el terreno, frente a los radios donde se destaca la alta proporción de viviendas particulares tipo rancho. Ambos perfiles implican principalmente rasgos de propiedad pero marcando diferencias sociales notables. Por tanto, se trata de una dimensión que identificamos como de segregación socio-residencial.

Mapa social según el Factor 1: nivel socioeconómico y densidad



Fuente: Fachelli, Goicoechea y López-Roldán (2015)

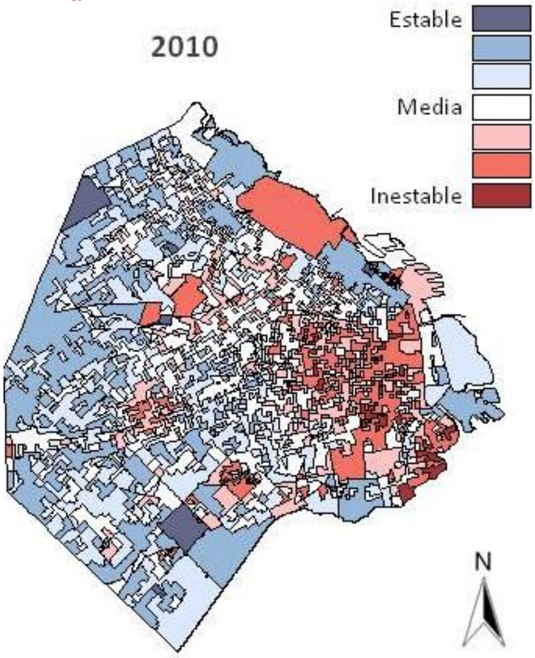
Mapa social según el Factor 2: nivel de segregación sociorresidencial



Fuente: Fachelli, Goicoechea y López-Roldán (2015)

Finalmente, el Factor 3 destaca la oposición entre inquilinos y propietarios, por tanto, de radios de la ciudad donde sobresale la presencia de situaciones estables de residencia (tenencia en propiedad) y la inestabilidad que implica habitar una vivienda que no es propia. Denominamos a esta tercera dimensión como estabilidad residencial. Se identifica una lógica de distribución de la estabilidad residencial en concordancia con la lógica centro – periferia, ya señalada por el Arquitecto Torres.

Mapa social según el Factor 3: nivel de estabilidad residencial



Fuente: Fachelli, Goicoechea y López-Roldán (2015)

En el capítulo siguiente completaremos el análisis de dimensionalización con el de tipificación de la ciudad en términos de tipos o estratos socioresidenciales que nos permitirá dibujar la conformación del perfil socio-espacial de la Ciudad de Buenos Aires.

3. Análisis de componentes principales con SPSS

3.1.1. *Estructura social del municipio de Alcobendas*

Para ilustrar el uso del software trataremos un primer ejemplo utilizando los datos relativos a un estudio realizado en el municipio de Alcobendas sobre su estructura social[16]. La matriz de datos Alcobendas.sav contiene los datos del estudio que fueron publicados en el anexo de la publicación referenciada. Se trata de un total de 36 secciones censales del ayuntamiento madrileño caracterizadas a partir de 15 variables, si bien la matriz original contenía un total de 27. Para hacer el análisis de componentes principales se utilizan, según el estudio, estas 15 variables de las que presentamos a

[16] Ayuntamiento de Alcobendas (1992). *Vivir en Alcobendas. Estructura social y conflicto.* Alcobendas: Ayuntamiento de Alcobendas.

continuación con el nombre y la identificación que le hemos asignado así como los valores de la media y de la desviación típica.

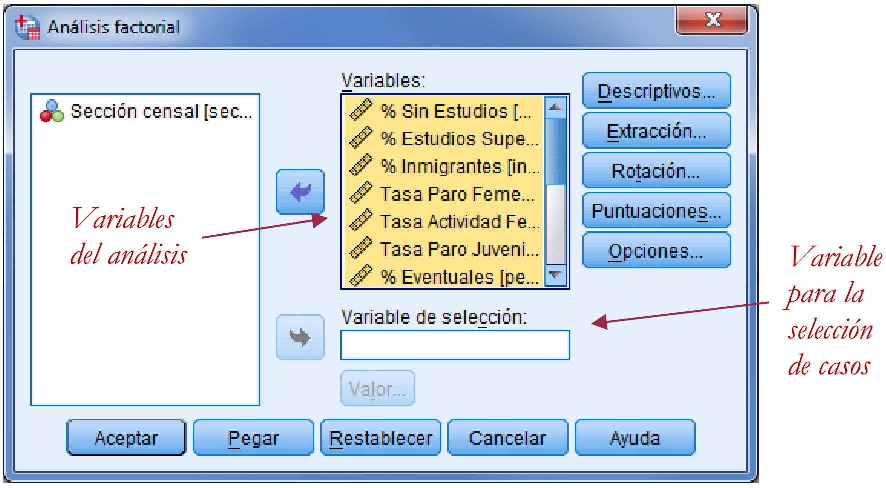
	seccion	sinest	super	inmigr	tparom	tactm	paroju	peven	pdtore	poper	ptecn	pevenm	ptecnm	padmim	pdomem	ppobjo
1	sc1	31	2	34	35	22	13	17	1	43	3	27	6	16	20	35
2	sc2	28	1	32	36	23	18	16	1	41	6	26	10	18	25	36
3	sc3	30	3	32	30	24	20	19	2	43	8	30	15	15	25	46
4	sc4	29	2	30	30	28	17	16	1	41	6	22	8	21	7	43
5	sc5	27	2	32	29	24	12	15	1	42	5	21	7	25	26	21
6	sc6	33	2	32	35	26	19	19	0	43	6	28	11	15	33	45
7	sc7	26	2	27	35	21	14	18	0	39	7	27	12	19	21	28
8	sc8	32	2	37	30	24	13	15	0	40	7	21	10	16	32	43
9	sc9	34	2	34	31	25	16	18	1	45	8	31	11	19	32	37
10	sc10	33	1	39	38	26	18	23	0	44	4	31	9	9	36	43
11	sc11	32	1	32	22	25	11	24	0	45	5	46	8	11	47	34
12	sc12	27	0	33	35	21	16	21	0	45	3	35	5	16	32	22
13	sc13	39	1	38	33	22	14	18	0	48	3	21	3	16	24	31
14	sc14	34	2	34	28	25	15	20	0	44	6	26	10	18	24	28
15	sc15	34	2	37	30	24	16	21	0	49	5	25	9	14	23	28
16	sc16	32	2	28	28	27	15	19	0	41	4	27	4	8	23	45
17	sc17	28	3	32	31	22	7	19	0	35	16	36	15	18	21	39
18	sc18	36	1	35	35	26	18	16	0	46	3	27	6	14	36	40
19	sc19	33	2	34	35	22	15	21	1	50	4	36	10	18	25	39
20	sc20	33	1	36	41	16	16	17	0	45	4	47	11	15	43	26
21	sc21	23	3	34	24	31	13	14	1	39	9	23	13	20	26	17
22	sc22	34	1	28	32	20	14	16	0	48	5	34	9	19	22	34
23	sc23	19	6	26	20	31	9	11	2	27	15	19	21	34	13	16
24	sc24	23	6	26	27	27	14	11	0	27	14	21	21	31	13	20
25	sc25	22	4	27	34	26	4	12	1	34	10	23	12	30	20	19
26	sc26	14	4	24	23	33	9	10	1	22	19	16	22	33	15	15
27	sc27	20	6	32	27	32	12	11	1	31	11	19	14	37	15	9
28	sc28	27	3	33	23	26	12	20	0	39	8	41	8	27	30	15
29	sc29	20	3	7	27	21	11	11	1	35	9	20	11	30	26	10
30	sc30	11	16	26	15	42	8	5	1	18	23	9	32	39	8	6
31	sc31	13	14	22	16	35	7	6	3	18	22	8	27	40	11	11
32	sc32	29	4	35	26	31	15	20	0	39	8	30	10	25	31	22
33	sc33	3	34	8	10	33	2	4	8	3	35	7	39	25	14	11
34	sc34	6	32	11	14	31	3	4	7	2	34	10	28	27	22	10
35	sc35	7	26	9	9	27	2	3	7	3	33	7	32	13	32	21
36	sc36	8	21	8	11	28	4	8	6	4	19	15	8	11	55	31

Estadísticos descriptivos

	Media	Desviación estándar	N de análisis
sinest % Sin Estudios	25,28	9,543	36
super % Estudios Superiores	6,03	8,772	36
inmigr % Inmigrantes	28,44	9,009	36
tparom Tasa Paro Femenino	27,36	8,247	36
tactm Tasa Actividad Femenina	26,31	5,019	36
paroju Tasa Paro Juvenil (15-21)	12,28	4,949	36
peven % Eventuales	14,94	5,722	36
pdtore % Directores	1,31	2,175	36
poper % Operarios	34,94	14,010	36
ptecn % Técnicos	10,75	8,949	36
pevenm % Mujeres Eventuales	24,78	10,040	36
ptecnm % Mujeres Técnicas	13,53	8,647	36
padmim % Mujeres Administrativas	21,17	8,564	36
pdomem % Mujeres en Servicio Doméstico	25,22	10,420	36
ppobjo Población -15 / +65	27,11	12,203	36

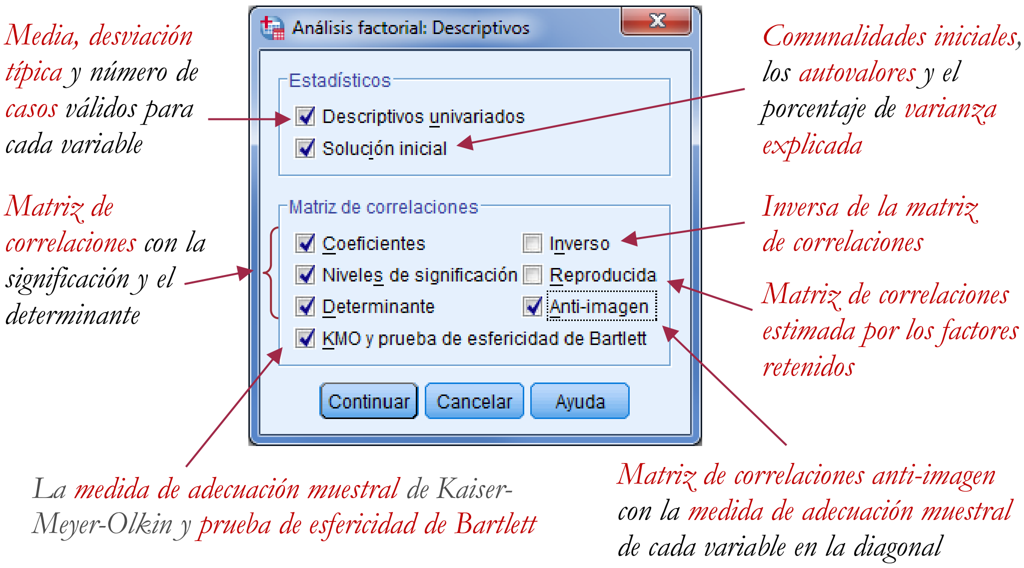
Los principales resultados de tablas y gráficos que se obtienen del ACP con el software SPSS se presentan a continuación y se exponen las características fundamentales del procedimiento para realizar el análisis a través de los cuadros de diálogo del menú. Estos mismos resultados se pueden obtener ejecutando el programa de sintaxis ACP-Albodendas.sps.

El procedimiento de análisis factorial corresponde al comando **FACTOR** y a través del menú se accede por: Analizar / Reducción de dimensiones / Factor.



En este caso ya se han pasado las variables que se utilizarán para hacer el análisis (todas menos la Sección Censal que identifica los casos) en el cuadro de Variables. Disponemos inicialmente también de la opción de realizar una selección de casos según el valor de un número entero. A continuación se trata de especificar las diversas opciones de instrucciones de este procedimiento a través de los 5 botones que se presentan: Descriptivos, Extracción, Rotación, Puntuaciones y Opciones.

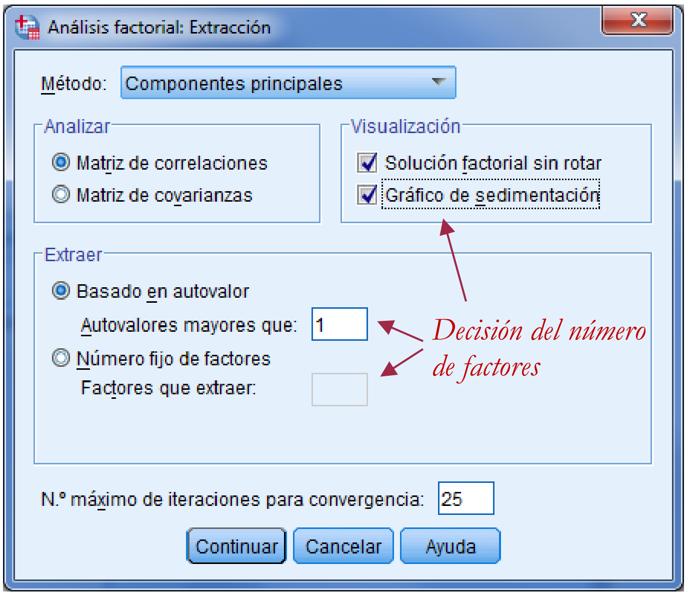
En primer lugar, pediremos las especificaciones de Descriptivos.



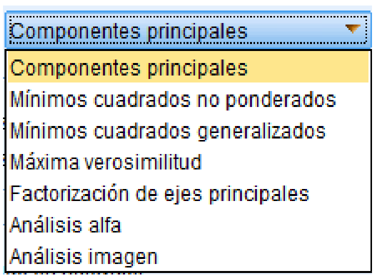
Con Descriptivos univariados se obtienen la media, la desviación típica y el número de casos válidos para cada variable. La solución inicial muestra las comunalidades iniciales, los autovalores y el porcentaje de varianza explicada.

Las opciones Coeficientes, Niveles de significación y Determinante corresponden a la matriz de correlaciones de las variables. Las opciones de la medida de adecuación muestral de Kaiser-Meyer-Olkin y la prueba de esfericidad de Bartlett nos permiten evaluar las variables del ACP. El KMO evalúa si las correlaciones parciales entre las variables son pequeñas, mientras que el test contrasta si la matriz de correlaciones es la matriz identidad (una diagonal de unos y el resto de ceros) que indicaría que el modelo factorial es inadecuado. Se puede pedir también la inversa de la matriz de correlaciones y la matriz de correlaciones reproducida a partir de las variables factoriales retenidas, donde se muestran igualmente las correlaciones de los residuos (la diferencia entre la correlación observada y la estimada). Finalmente se especifica la matriz de correlaciones (y covarianzas) anti-imagen que contiene los negativos de los coeficientes de correlación (covarianza) parcial. La mayoría de los elementos no diagonales deben ser pequeños, y la información más relevante está en la diagonal donde se muestra la medida de adecuación muestral de cada variable cuyos valores son la contribución individual de cada una al valor global del KMO, por tanto, de forma equivalente al KMO se deben considerar valores suficientemente altos y, en todo caso, superiores a 0,5. Si una variable no cumple con este criterio mínimo deberá reconsiderarse y eliminarse del análisis.

A continuación detallaremos el procedimiento de extracción de los factores.



En nuestro caso se trata de hacer un análisis factorial de componentes principales, que es el método por defecto ya seleccionado, si bien en el desplegable se puede optar por otros procedimientos de análisis factorial:



Junto al análisis de componentes principales se encuentran:
- Método de mínimos cuadrados no ponderados. Método de extracción de factores que minimiza la suma de los cuadrados de las diferencias entre las matrices de correlación observada y reproducida, ignorando las diagonales.
- Método de Mínimos cuadrados generalizados. Método de extracción de factores que minimiza la suma de los cuadrados de las diferencias entre las matrices de correlación observada y reproducida. Las correlaciones se ponderan por el inverso de su exclusividad, de manera que las variables que tengan un valor alto de exclusividad reciban una ponderación menor que aquéllas que tengan un valor bajo de exclusividad.
- Método de máxima verosimilitud. Método de extracción factorial que proporciona las estimaciones de los parámetros que con mayor probabilidad ha producido la matriz de correlaciones observada, si la muestra procede de una distribución normal multivariada. Las correlaciones se ponderan por el inverso de la exclusividad de las variables, y se emplea un algoritmo iterativo.
- Factorización de ejes principales. Método para la extracción de factores que parte de la matriz de correlaciones original con los cuadrados de los coeficientes de correlación múltiple insertados en la diagonal principal como estimaciones iniciales de las comunalidades. Las cargas factoriales resultantes se utilizan para estimar de nuevo las comunalidades que reemplazan a las estimaciones previas de comunalidad en la diagonal. Las iteraciones continúan hasta que el cambio en las comunalidades, de una iteración a la siguiente, satisfaga el criterio de convergencia para la extracción.
- Alfa. Método de extracción factorial que considera a las variables incluidas en el análisis como una muestra del universo de las variables posibles. Este método maximiza el Alfa de Cronbach para los factores.
- Factorización imagen. Método para la extracción de factores, desarrollado por Guttman y basado en la teoría de las imágenes. La parte común de una variable, llamada la imagen parcial, se define como su regresión lineal sobre las restantes variables, en lugar de ser una función de los factores hipotéticos.

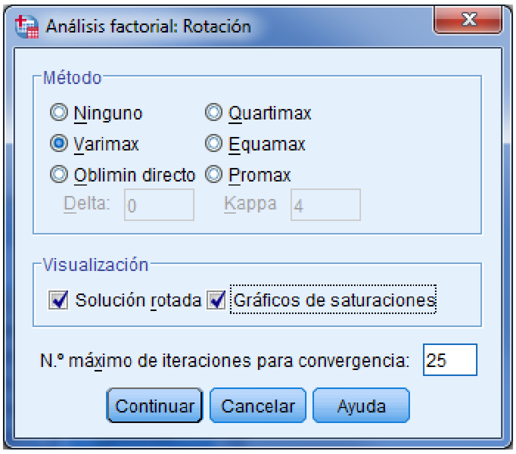
El análisis de componentes principales lo efectuaremos a partir de la matriz de correlaciones, más fácil de interpretar. No obstante esta opción se recomienda especialmente cuando las variables tengan escalas distintas, mientras que la matriz de varianzas y covarianzas se puede emplear para varios grupos con varianza distintas para cada variable.

Solicitaremos que genere la representación gráfica de los valores propios o autovalores, es decir, la varianza asociada a cada factor, el llamado Gráfico de Sedimentación, como ayuda visual en la decisión del número de factores a partir de localizar el cambio de pendiente de la curva decreciente que dibuja. Indicaremos también que nos presente los resultados de la solución factorial no rotada (la matriz de componentes o de saturaciones sin rotar, las comunalidades y los autovalores de la solución factorial).

El criterio por defecto es que se extraigan los factores o componentes con un valor propio o autovalor mayor de 1 (criterio de Kaiser) como se especifica en Basado en autovalor. Si posteriormente se observa en el análisis que este criterio no se corresponde con una decisión basada en la interpretación de los datos se puede volver a ejecutar precisando el número de factores decidido en la opción Número fijo de factores a extraer.

Finalmente se permite especificar el número máximo de pasos que el algoritmo puede seguir para estimar la solución, este valor por defecto no es necesario cambiarlo.

El siguiente paso será pedir una rotación de la solución factorial con el objetivo de mejorar la interpretación de los resultados.



No es un paso obligado, podemos comparar los resultados con y sin rotación y optar por la que consideremos más adecuada para la interpretación de los resultados.

Los métodos disponibles son los siguientes:

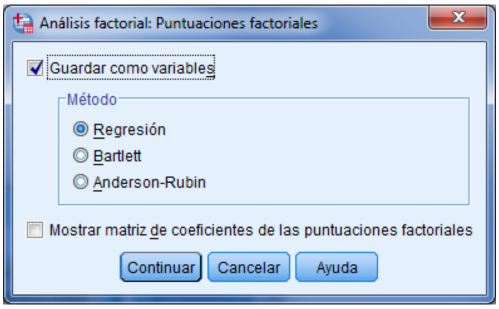
- Método Varimax. Rotación ortogonal que minimiza el número de variables que tienen saturaciones altas en cada factor. Simplifica la interpretación de los factores.
- Criterio Oblimin directo. Rotación oblicua (no ortogonal). Si delta es igual a cero, el valor prdetermnado, las soluciones son las más oblicuas. A medida que delta se va haciendo más negativo, los factores son menos oblicuos. Se puede introducir un número menor o igual que 0,8.
- Método Quartimax. Rotación que minimiza el número de factores necesarios para explicar cada variable. Simplifica la interpretación de las variables observadas.
- Método Equamax. Método de rotación que es combinación del método varimax, que simplifica los factores, y el método quartimax, que simplifica las variables. Se minimiza tanto el número de variables que saturan alto en un factor como el número de factores necesarios para explicar una variable.
- Rotación Promax. Rotación oblicua que permite que los factores estén correlacionados. Esta rotación se puede calcular más rápidamente que una rotación oblimin directa, por lo que es útil para conjuntos de datos grandes.

De los diferentes métodos de rotación escogeremos el procedimiento Varimax. Asimismo indicaremos que nos muestre los datos correspondientes a la solución rotada. Para rotaciones ortogonales esta opción extrae la matriz de configuración rotada y matriz de transformación de factor. Para rotaciones oblicuas, las matrices de correlaciones de factor, estructura y patrón.

Solicitaremos los Gráficos de saturaciones, la representación gráfica factorial, expresión gráfica de la matriz de componentes o de saturaciones, que se presenta en dos o tres dimensiones, dependiendo de si hay dos o más de dos componentes retenidos, de las variables originales en el espacio de los factores o componentes a partir de la solución rotada.

Como hemos pedido la solución rotada y los gráficos de saturaciones nos proporcionan los resultados rotados. En el caso de querer ver representada en primer término la solución no rotada, hay que tener marcada la opción del método Ninguno y pedir los Gráficos de saturaciones. Si optamos por comparar los resultados sin y con rotación veremos que buena parte de los resultados que proporciona el software serán comunes, como todos los relacionados con la matriz de correlaciones, pero otros cambiarán y habrá que analizar cómo lo hacen. En particular, habrá que observar la tabla de varianza explicada, la matriz de saturaciones o de componentes y las representaciones gráficas.

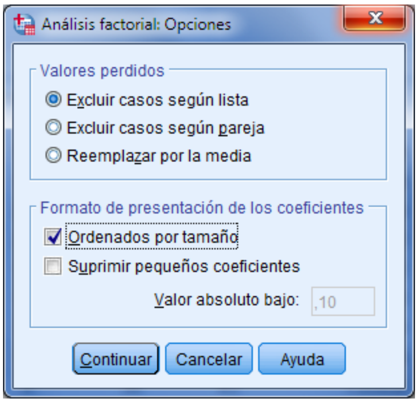
Como resultado de un análisis de componentes principales disponemos de tantas variables como factores o componentes hemos retenido el análisis. Si queremos utilizar posteriormente para realizar otros análisis es necesario guardarlas como nuevas variables en la matriz de datos. Esto es lo que hacemos cuando lo especificamos a través de Puntuaciones marcando la opción Guardar como variables.



Por defecto elegimos el método de la Regresión; en este caso las puntuaciones tienen media 0 y una varianza igual al cuadrado de la correlación múltiple entre las puntuaciones factoriales estimadas y los valores factoriales verdaderos. Las puntuaciones puede correlacionarse incluso si los factores son ortogonales. Con la opción de puntuaciones de Bartlett éstas tienen media 0 y se minimiza la suma de cuadrados de los factores únicos sobre el rango de las variables. La alternativa del método de Anderson-Rubin es una modificación del método de Bartlett, que asegura la ortogonalidad de los factores estimados, teniendo las puntuaciones resultantes una media 0, una desviación típica de 1 y no correlacionan entre sí.

Las variables se guardarán con el nombre que asigna por defecto el SPSS de FAC1_1, FAC2_1,… y así sucesivamente hasta el número total de componentes retenidos en relación al número de análisis realizado, al inicio el 1.

Finalmente dentro de Opciones consideraremos la exclusión de los valores perdidos (*missing values*) según lista, alternativamente si optamos por excluir según pareja excluye los perdidos de las parejas de variables que no puede calcular las correlaciones, y si optamos por reemplazar sustituye el valor perdido por la media.



Especificaremos que los coeficientes se presenten por orden de tamaño lo que nos facilitará la lectura de los factores de carga de la matriz de saturaciones o de componentes. Y no contemplaremos que se supriman los coeficientes cuyos valores absolutos sean menores que un valor especificado.

La mayor parte de las posibilidades que permite este procedimiento se pueden realizar a través del menú como hemos visto. Pero la utilización del lenguaje de comandos nos permitiría además otras posibilidades no proporcionadas en el cuadro de diálogo y solamente ejecutables en la ventana de sintaxis como: especificar gráficos factoriales individuales, especificar el nombre de las variables factoriales que se quieren guardar, especificar el número de puntuaciones factoriales que se van a guardar, especificar valores diagonales, especificar los criterios de convergencia para la iteración durante la extracción, guardar matrices de correlación o matrices de carga factorial para su análisis posterior y leer y analizar matrices de correlación o matrices de carga factorial..

En particular resulta una necesidad ejecutar la sintaxis para obtener gráficos bidimensionales si la solución factorial comporta 3 o más factores, pues los gráficos en tres dimensiones son más difíciles de leer. En este caso después de preparar la ejecución por el menú se puede optar por Pegar la sintaxis y añadir a la línea de la instrucción del comando **FACTOR** donde se especifica el gráfico:

/PLOT EIGEN ROTATION

las parejas de dimensiones. Por ejemplo, en el caso de 3 factores, las parejas (1,2) (1,3), de la siguiente forma:

/PLOT EIGEN ROTATION (1,2) (1,3)

Los resultados de las distintas especificaciones que hemos detallado se presentan seguidamente.

De la información derivada de la matriz de correlaciones se concluye que reunimos las condiciones para una adecuada aplicación de la técnica y para la interpretación de los resultados. El KMO es, en la terminología de Kaiser, meritorio, de 0,8 redondeando a un decimal. Y todas las medidas de adecuación muestral de las variables originales que se observan en la diagonal de la matriz de correlaciones anti-imagen son aceptables.

Matriz de correlaciones[a]

		sinest % Sin Estudios	super % Estudios Superiores	inmigr % Inmigrantes	tparom Tasa Paro Femenino	tactm Tasa Actividad Femenina	paroju Tasa Paro Juvenil (15-21)	peven % Eventuales	pdtore % Directores	poper % Operarios	ptecn % Técnicos	pevenm % Mujeres Eventuales	ptecnm % Mujeres Técnicas	padmim % Mujeres Administrativas	pdomem % Mujeres en Servicio Doméstico	ppobjo Población -15 / +65
Correlación	sinest % Sin Estudios	**1,000**	-0,868	0,847	0,844	-0,643	0,833	0,894	-0,820	0,960	-0,915	0,751	-0,816	-0,520	0,248	0,700
	super % Estudios Superiores	-0,868	**1,000**	-0,796	-0,834	0,535	-0,782	-0,812	0,940	-0,926	0,935	-0,705	0,801	0,246	-0,114	-0,491
	inmigr % Inmigrantes	0,847	-0,796	**1,000**	0,745	-0,316	0,747	0,786	-0,812	0,844	-0,783	0,640	-0,621	-0,275	0,077	0,489
	tparom Tasa Paro Femenino	0,844	-0,834	0,745	**1,000**	-0,696	0,776	0,746	-0,772	0,867	-0,850	0,678	-0,720	-0,368	0,128	0,571
	tactm Tasa Actividad Femenina	-0,643	0,535	-0,316	-0,696	**1,000**	-0,443	-0,603	0,381	-0,617	0,615	-0,670	0,649	0,613	-0,444	-0,528
	paroju Tasa Paro Juvenil (15-21)	0,833	-0,782	0,747	0,776	-0,443	**1,000**	0,763	-0,730	0,843	-0,830	0,591	-0,662	-0,398	0,142	0,613
	peven % Eventuales	0,894	-0,812	0,786	0,746	-0,603	0,763	**1,000**	-0,756	0,893	-0,860	0,858	-0,801	-0,575	0,359	0,654
	pdtore % Directores	-0,820	0,940	-0,812	-0,772	0,381	-0,730	-0,756	**1,000**	-0,867	0,855	-0,660	0,688	0,115	-0,014	-0,365
	poper % Operarios	0,960	-0,926	0,844	0,867	-0,617	0,843	0,893	-0,867	**1,000**	-0,952	0,764	-0,809	-0,400	0,164	0,576
	ptecn % Técnicos	-0,915	0,935	-0,783	-0,850	0,615	-0,830	-0,860	0,855	-0,952	**1,000**	-0,729	0,915	0,421	-0,270	-0,574
	pevenm % Mujeres Eventuales	0,751	-0,705	0,640	0,678	-0,670	0,591	0,858	-0,660	0,764	-0,729	**1,000**	-0,673	-0,478	0,471	0,496
	ptecnm % Mujeres Técnicas	-0,816	0,801	-0,621	-0,720	0,649	-0,662	-0,801	0,688	-0,809	0,915	-0,673	**1,000**	0,537	-0,447	-0,589
	padmim % Mujeres Administrativas	-0,520	0,246	-0,275	-0,368	0,613	-0,398	-0,575	0,115	-0,400	0,421	-0,478	0,537	**1,000**	-0,671	-0,787
	pdomem % Mujeres en Servicio Doméstico	0,248	-0,114	0,077	0,128	-0,444	0,142	0,359	-0,014	0,164	-0,270	0,471	-0,447	-0,671	**1,000**	0,364
	ppobjo Población -15 / +65	0,700	-0,491	0,489	0,571	-0,528	0,613	0,654	-0,365	0,576	-0,574	0,496	-0,589	-0,787	0,364	**1,000**
Sig. (unilateral)	sinest % Sin Estudios		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,072	0,000
	super % Estudios Superiores	0,000		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,074	0,254	0,001
	inmigr % Inmigrantes	0,000	0,000		0,000	0,030	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,052	0,329	0,001
	tparom Tasa Paro Femenino	0,000	0,000	0,000		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,014	0,228	0,000
	tactm Tasa Actividad Femenina	0,000	0,000	0,030	0,000		0,003	0,000	0,011	0,000	0,000	0,000	0,000	0,000	0,003	0,000
	paroju Tasa Paro Juvenil (15-21)	0,000	0,000	0,000	0,000	0,003		0,000	0,000	0,000	0,000	0,000	0,000	0,008	0,205	0,000
	peven % Eventuales	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,000	0,000	0,000	0,000	0,000	0,016	0,000
	pdtore % Directores	0,000	0,000	0,000	0,000	0,011	0,000	0,000		0,000	0,000	0,000	0,000	0,252	0,467	0,014
	poper % Operarios	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,000	0,000	0,008	0,169	0,000
	ptecn % Técnicos	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,000	0,005	0,055	0,000
	pevenm % Mujeres Eventuales	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,002	0,002	0,001
	ptecnm % Mujeres Técnicas	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,003	0,000
	padmim % Mujeres Administrativas	0,001	0,074	0,052	0,014	0,000	0,008	0,000	0,252	0,008	0,005	0,002	0,000		0,000	0,000
	pdomem % Mujeres en Servicio Doméstico	0,072	0,254	0,329	0,228	0,003	0,205	0,016	0,467	0,169	0,055	0,002	0,003	0,000		0,015
	ppobjo Población -15 / +65	0,000	0,001	0,001	0,000	0,000	0,000	0,000	0,014	0,000	0,000	0,001	0,000	0,000	0,015	

a. Determinante = 0,000

Prueba de KMO y Bartlett

Medida Kaiser-Meyer-Olkin de adecuación de muestreo		0,799
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	767,734
	gl	105
	Sig.	0,000

Matrices anti-imagen

		sinest % Sin Estudios	super % Estudios Superiores	inmigr % Inmigrantes	tparom Tasa Paro Femenino	tactm Tasa Actividad Femenina	paroju Tasa Paro Juvenil (15-21)	peven % Eventuales	pdtore % Directores	poper % Operarios	ptecn % Técnicos	pevenm % Mujeres Eventuales	ptecnm % Mujeres Técnicas	padmim % Mujeres Administrativas	pdomem % Mujeres en Servicio Doméstico	ppobjo Población -15 / +65
Correlación anti-imagen	sinest % Sin Estudios	**0,771**[a]	-0,516	-0,516	0,358	0,514	-0,199	0,097	0,594	-0,718	-0,211	0,236	0,302	-0,232	-0,213	-0,642
	super % Estudios Superiores	-0,516	**0,801**[a]	0,263	-0,274	-0,454	0,051	-0,003	-0,761	0,400	-0,150	-0,184	-0,111	0,398	0,157	0,555
	inmigr % Inmigrantes	-0,516	0,263	**0,827**[a]	-0,494	-0,636	0,166	-0,146	-0,235	0,151	0,005	-0,153	-0,151	0,195	0,029	0,337
	tparom Tasa Paro Femenino	0,358	-0,274	-0,494	**0,836**[a]	0,681	-0,259	0,142	0,351	-0,188	0,113	0,088	0,010	-0,241	0,047	-0,400
	tactm Tasa Actividad Femenina	0,514	-0,454	-0,636	0,681	**0,654**[a]	-0,256	-0,128	0,504	-0,160	0,075	0,422	-0,005	-0,469	-0,137	-0,494
	paroju Tasa Paro Juvenil (15-21)	-0,199	0,051	0,166	-0,259	-0,256	**0,912**[a]	-0,173	-0,155	0,101	0,405	0,083	-0,430	0,046	-0,048	-0,028
	peven % Eventuales	0,097	-0,003	-0,146	0,142	-0,128	-0,173	**0,907**[a]	0,099	-0,297	-0,212	-0,571	0,329	0,335	0,182	-0,026
	pdtore % Directores	0,594	-0,761	-0,235	0,351	0,504	-0,155	0,099	**0,764**[a]	-0,430	-0,126	0,258	0,213	-0,178	-0,190	-0,513
	poper % Operarios	-0,718	0,400	0,151	-0,188	-0,160	0,101	-0,297	-0,430	**0,811**[a]	0,552	-0,183	-0,462	0,184	0,363	0,476
	ptecn % Técnicos	-0,211	-0,150	0,005	0,113	0,075	0,405	-0,212	-0,126	0,552	**0,869**[a]	-0,023	-0,795	-0,015	0,166	0,010
	pevenm % Mujeres Eventuales	0,236	-0,184	-0,153	0,088	0,422	0,083	-0,571	0,258	-0,183	-0,023	**0,833**[a]	-0,144	-0,368	-0,541	-0,224
	ptecnm % Mujeres Técnicas	0,302	-0,111	-0,151	0,010	-0,005	-0,430	0,329	0,213	-0,462	-0,795	-0,144	**0,838**[a]	-0,025	0,095	-0,093
	padmim % Mujeres Administrativas	-0,232	0,398	0,195	-0,241	-0,469	0,046	0,335	-0,178	0,184	-0,015	-0,368	-0,025	**0,679**[a]	0,539	0,667
	pdomem % Mujeres en Servicio Doméstico	-0,213	0,157	0,029	0,047	-0,137	-0,048	0,182	-0,190	0,363	0,166	-0,541	0,095	0,539	**0,600**[a]	0,344
	ppobjo Población -15 / +65	-0,642	0,555	0,337	-0,400	-0,494	-0,028	-0,026	-0,513	0,476	0,010	-0,224	-0,093	0,667	0,344	**0,657**[a]

a. Medidas de adecuación de muestreo (MSA)

La tabla de comunalidades nos presenta la información de qué parte de cada variable se conserva después de la extracción de los factores y de retener, como veremos, 2 factores. Esto nos indica qué variables están mejor representadas en la información retenidas y el grado en que serán más o menos protagonistas en la configuración del contenido de los factores. Así, por ejemplo, la variable sinest (% sin estudios) con un valor de 0,938 marcará de forma notable el carácter de los factores, mientras que el papel de la variable tactm (la tasa de actividad femenina) con un 0,642 será menos importante.

Comunalidades

	Inicial	Extracción
sinest % Sin Estudios	1,000	0,938
super % Estudios Superiores	1,000	0,925
inmigr % Inmigrantes	1,000	0,784
tparom Tasa Paro Femenino	1,000	0,806
tactm Tasa Actividad Femenina	1,000	0,642
paroju Tasa Paro Juvenil (15-21)	1,000	0,747
peven % Eventuales	1,000	0,875
pdtore % Directores	1,000	0,901
poper % Operarios	1,000	0,961
ptecn % Técnicos	1,000	0,933
pevenm % Mujeres Eventuales	1,000	0,704
ptecnm % Mujeres Técnicas	1,000	0,784
padmim % Mujeres Administrativas	1,000	0,875
pdomem % Mujeres en Servicio Doméstico	1,000	0,717
ppobjo Población -15 / +65	1,000	0,647

Método de extracción: análisis de componentes principales.

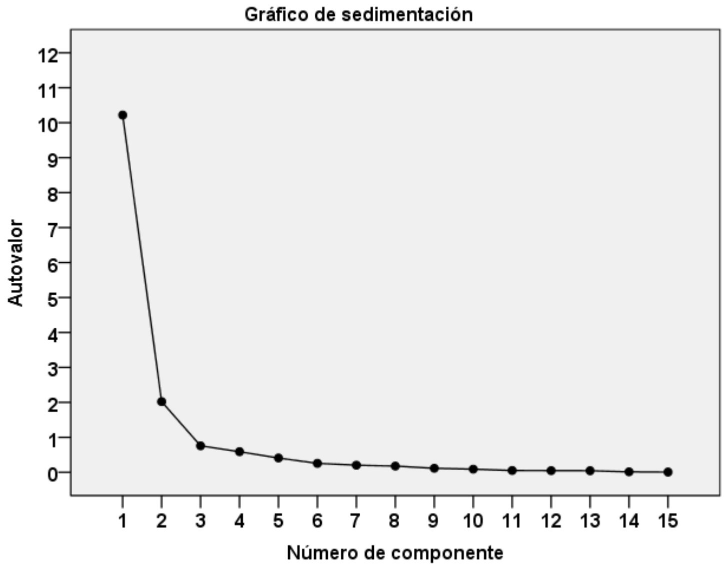
En el análisis se toma la decisión de retener dos factores pues se cumplen simultáneamente diversos criterios: con los dos factores se alcanza el 70% de la varianza explicada (llegando al 81,596%), corresponden a valores propios o autovalores por encima de 1 y, según el gráfico de sedimentación que se presenta seguidamente, se corresponde con el cambio de pendiente de la curva (donde está el codo de la curva).

Varianza total explicada

	Autovalores iniciales			Sumas de extracción de cargas al cuadrado			Sumas de rotación de cargas al cuadrado		
Componente	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	10,22	68,120	68,120	10,218	68,120	68,120	8,527	56,847	56,847
2	2,021	13,476	81,596	2,021	13,476	81,596	3,712	24,749	81,596
3	0,757	5,048	86,644						
4	0,589	3,924	90,568						
5	0,409	2,724	93,292						
6	0,256	1,708	94,999						
7	0,204	1,362	96,361						
8	0,177	1,177	97,538						
9	0,115	0,768	98,306						
10	0,090	0,601	98,907						
11	0,051	0,342	99,249						
12	0,048	0,318	99,566						
13	0,045	0,297	99,863						
14	0,014	0,091	99,954						
15	0,007	0,046	100,000						

Método de extracción: análisis de componentes principales.

No obstante, el primer factor es el principal y acumula la mayor parte de la varianza, casi llegando al 70% él solo, por ello se puede concluir que las diferencias entre las secciones censales del municipio madrileño se expresan fundamentalmente con la primera dimensión, introduciendo la segunda un matiz de menor relevancia.



Las tablas siguientes con la matriz de saturaciones o de componentes, antes y después de la rotación varimax, nos permite dar contenido e interpretar los factores de diferenciación social del municipio a partir de las variables originales consideradas.

Matriz de componente[a]

	Componente 1	Componente 2
sinest % Sin Estudios	**0,967**	-0,054
poper % Operarios	**0,963**	-0,186
ptecn % Técnicos	**-0,959**	0,119
peven % Eventuales	**0,934**	0,059
super % Estudios Superiores	**-0,912**	0,305
tparom Tasa Paro Femenino	**0,886**	-0,146
ptecnm % Mujeres Técnicas	**-0,876**	-0,127
paroju Tasa Paro Juvenil (15-21)	**0,849**	-0,162
pdtore % Directores	**-0,838**	0,446
inmigr % Inmigrantes	**0,828**	-0,315
pevenm % Mujeres Eventuales	**0,826**	0,149
ppobjo Población -15 / +65	**0,697**	0,401
tactm Tasa Actividad Femenina	**-0,695**	-0,398
pdomem % Mujeres en Servicio Doméstico	0,339	**0,776**
padmim % Mujeres Administrativas	-0,554	**-0,753**

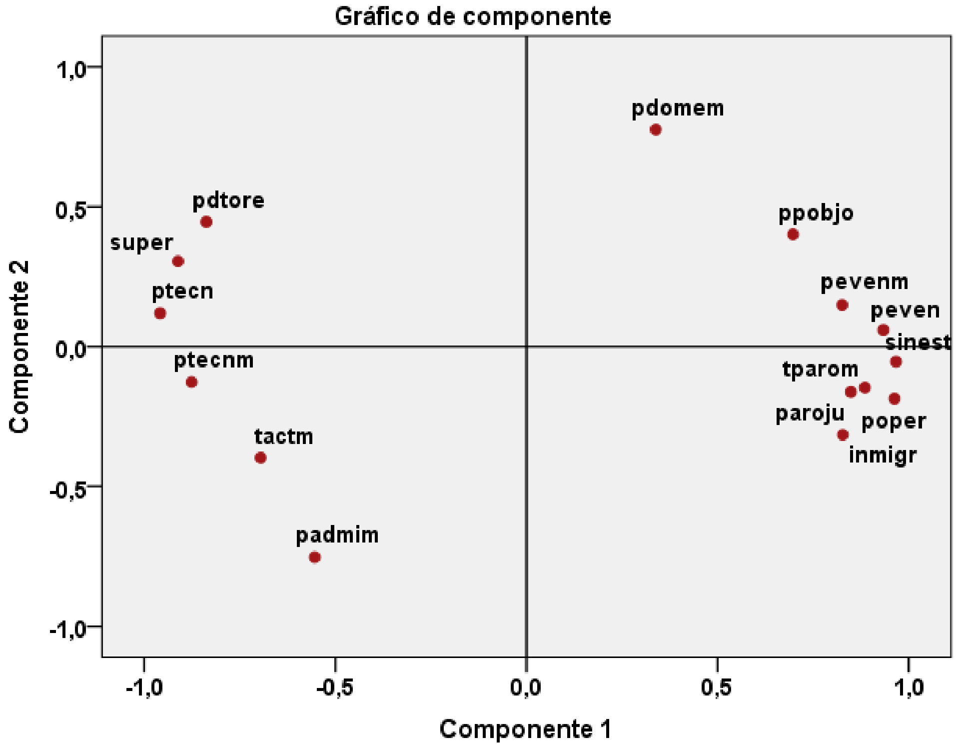
Método de extracción: análisis de componentes principales.
a. 2 componentes extraídos.

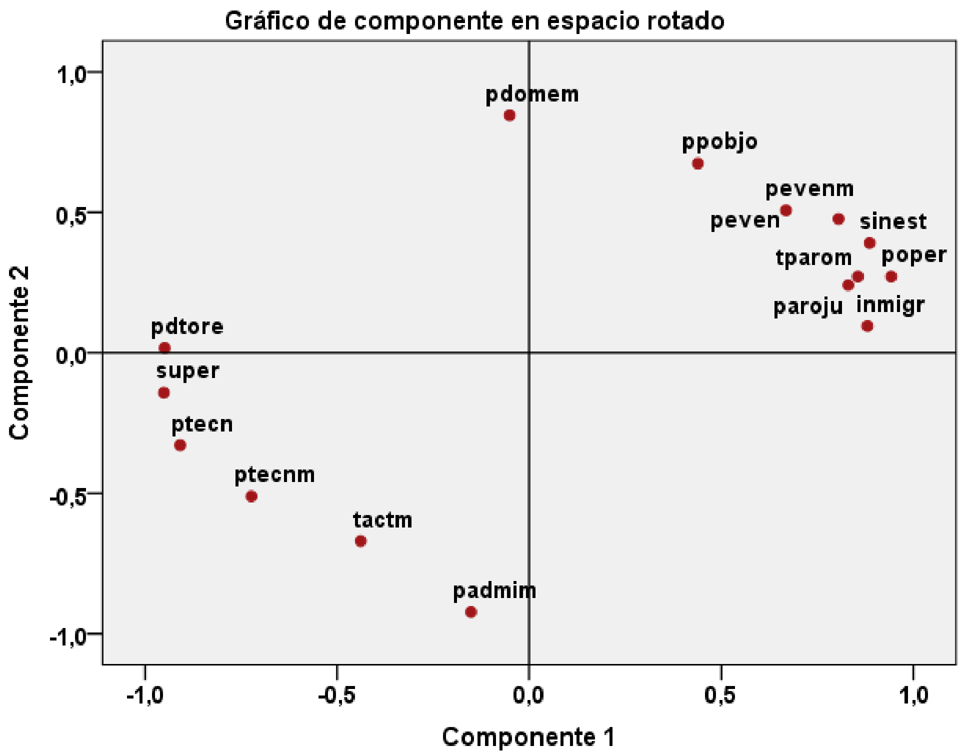
Matriz de componente rotado[a]

	Componente 1	Componente 2
super % Estudios Superiores	**-0,951**	-0,142
pdtore % Directores	**-0,949**	0,017
poper % Operarios	**0,942**	0,272
ptecn % Técnicos	**-0,908**	-0,329
sinest % Sin Estudios	**0,886**	0,391
inmigr % Inmigrantes	**0,881**	0,095
tparom Tasa Paro Femenino	**0,856**	0,272
paroju Tasa Paro Juvenil (15-21)	**0,830**	0,241
peven % Eventuales	**0,805**	0,477
ptecnm % Mujeres Técnicas	**-0,723**	-0,511
pevenm % Mujeres Eventuales	**0,668**	0,508
padmim % Mujeres Administrativas	-0,152	**-0,923**
pdomem % Mujeres en Servicio Doméstico	-0,051	**0,845**
ppobjo Población -15 / +65	0,439	**0,674**
tactm Tasa Actividad Femenina	-0,439	**-0,670**

Método de extracción: análisis de componentes principales.
Método de rotación: Varimax con normalización Kaiser.
a. La rotación ha convergido en 3 iteraciones.

Esta información se representa y se interpreta igualmente a través del gráfico de dispersión entre los factores y las variables originales (gráficos factoriales o gráficos de componentes). Se presenta un gráfico antes de la rotación y otro después de rotar.



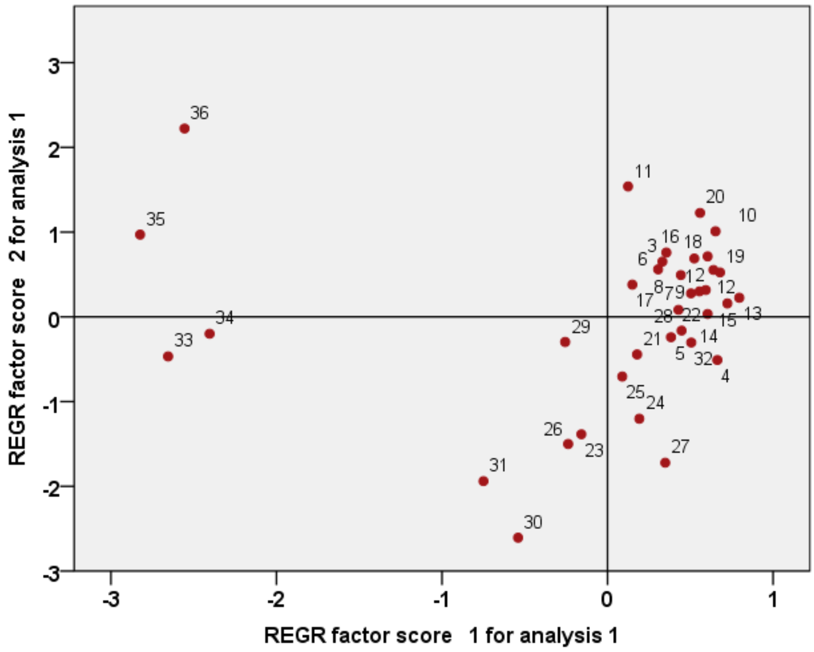


Como resultado de la rotación, la varianza total explicada, que permanece constante (81,596%), se redistribuye dando mayor peso al segundo factor o componente, pasando de un 13,5% a un 24,7%. En la matriz de componente y gráficamente se observa como las variables ppobjo y tactm, que al inicio están mayormente proyectadas sobre el primer eje, ahora, tras la rotación, se acercan el segundo eje y se convierten en características más marcadas para dar cuenta de la segunda dimensión. No obstante el contenido de ambos factores no se altera en lo fundamental. La primera componente o factor, el que corresponde al eje horizontal, con un 57% de varianza explicada, opone niveles ocupacionales, educativos y de actividad para diferenciar las secciones, a las derecha del gráfico (el polo positivo de la dimensión) que se caracterizan por un perfil de predominio de ocupaciones bajas, eventualidad contractual, desempleo, bajos niveles educativos y un alto porcentaje de población inmigrante. En el otro extremo de la dimensión. A la izquierda, se ubica el perfil contrario, la carencia de los rasgos que acabamos de citar, y donde aparecen como perfil de este polo el alto porcentaje de trabajadores que son directores o técnicos, con altos niveles educativos y donde es mayor la tasa de actividad femenina. Se trata, por tanto, de un factor de desigualdad socioprofesional y educativa que divide socialmente al territorio, un división que podemos etiquetar igualmente de estratificación social al diferenciar los perfiles de clase trabajadores frente a los de clase media.

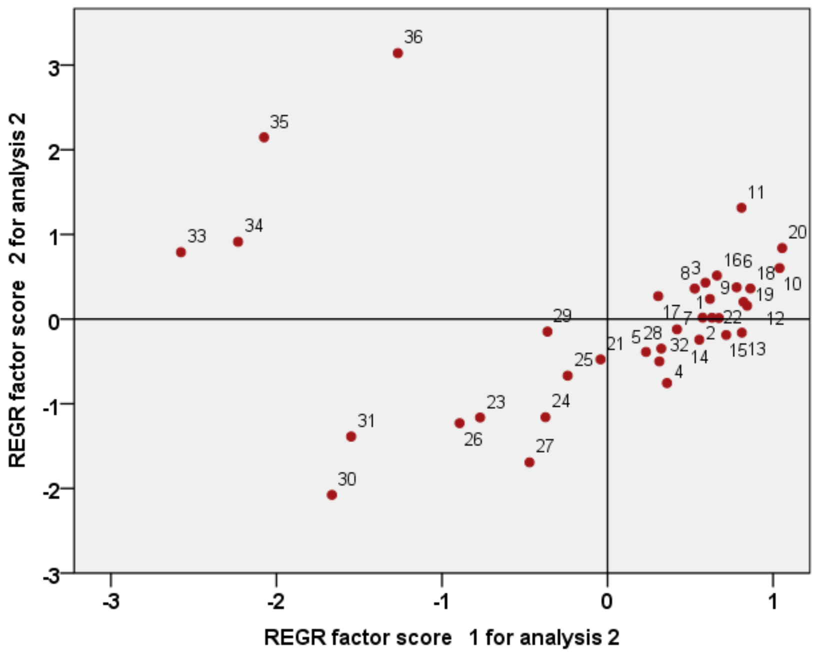
El segundo factor, el que corresponde al eje vertical, con un 25% de varianza explicada, introduce una segunda dimensión de diferenciación, independiente de la anterior, donde se contraponen dos rasgos de la actividad femenina: la proporción de mujeres ocupadas en el trabajo doméstico (zona superior) frente a la proporción de mujeres ocupadas como administrativas (zona inferior). Se trata de una diferenciación afecta tanto a la derecha como a la izquierda de la primera dimensión, por tanto, es una división presente tanto en las secciones censales de niveles socioprofesionales altos y bajos que los dividen internamente.

Para acabar de ver el sentido de esta división será importante analizar la distribución de las unidades en este espacio social construido. En el próximo capítulo de clasificación retomaremos esta cuestión. Pero una primera visualización de lo que encontraremos se puede presentar a partir del análisis factorial de componentes principales. Al ejecutar el ACP hemos guardado la variables factoriales. Si con las dos variables factoriales realizamos un gráfico de dispersión obtenemos, tanto en el caso sin rotar como rotado, la distribución de las 36 secciones censales de Alcobendas. Observamos primero como cuatro secciones de Alcobendas (de la 33 a la 36) se identifican con el perfil socioprofesional alto que hemos descrito frente al resto del municipio, especialmente en el primer gráfico. Como resultado de la rotación, las secciones censales del lado derecho más próximas a este grupo social más acomodado se ven proyectadas más claramente hacia la izquierda de la primera dimensión mientras las secciones 35 y 36 lo hacen sobre la zona superior de la segunda dimensión. Esto es, la primera dimensión deja de mostrar una realidad tan polarizada, y marca ahora un mayor continuum de estratificación social el perfil de clase trabajadora y de clase media definido por los extremos. Por su parte, la segunda dimensión diferencia más claramente las secciones entre las que se da una mayor proporción de mujeres ocupadas en servicio doméstico y las que lo tienen de ocupadas como administrativas, rasgo que afecta sobre todo a las secciones de perfil más acomodado, especialmente las secciones 35 y 36 por un lado, y las secciones 23, 24, 26, 27, 30 y 31 por otro.

Secciones censales en el espacio factorial sin rotar



Secciones censales en el espacio factorial rotado



3.1.2. *Los ejes de diferenciación de la ciudad de Buenos Aires*

Para reproducir los resultados presentados en el apartado 2.6.1 sobre la estructura socio-habitacional de la Ciudad de Buenos Aires se trabaja con la matriz de datos CABA-Censo2010.sav que contienen la información censal a partir de considerar 3552 radios censales en que se divide el territorio de la capital. Con el programa de instrucciones de SPSS ACP-CABA.sps se puede reproducir el análisis[17].

A partir de las 15 variables siguientes:

Estadísticos descriptivos

	Media	Desviación estándar	N de análisis
H12_Hasta0.50personas1 % Hacinamiento <0,5	32,1741	9,7	3.552
H12_1.501.99personas1 % Hacinamiento 1,5-2	5,4363	2,6	3.552
H12_2.003.00personas1 % Hacinamiento 2-3	8,6045	7,7	3.552
H12_Másde3.00personas1 % Hacinamiento +3	1,5100	2,8	3.552
P3_Otropaís1 % Extranjeros	12,5280	9,8	3.552
P5_Nuncaasistió1 Nunca asistió	0,9540	1,0	3.552
P6_Primario1 Primarios	20,9690	8,4	3.552
P6_Superiornouniversitario1 Superior	10,4311	3,1	3.552
P6_Universitario1 Univesitarios	27,3107	11,6	3.552
V2_Departamento1 Departamento	68,8852	27,4	3.552
V2_Piezaeninquilinato1 Inquinilato	2,9571	6,3	3.552
V2_Rancho1 Rancho	0,2561	1,4	3.552
V9_Inquilino1 Inquilino	29,8813	11,1	3.552
V9_Propietariodelaviviendaydelterreno1 Propietario	56,3918	13,0	3.552
Densidad Densidad	27.490,2036	19.274,9	3.552

Se realiza una análisis de componentes principales donde se obtiene un valor de KMO de 0,871.

Prueba de KMO y Bartlett

Medida Kaiser-Meyer-Olkin de adecuación de muestreo		0,871
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	48.842,210
	gl	105
	Sig.	0,000

Se toma la decisión de conservar tres ejes factoriales que son rotados mediante el procedimiento Varimax. La tabla de comunalidades es la siguiente.

[17] En la publicación siguiente se comentan los resultados: Fachelli, S.; Goicoechea, M. E.; López-Roldán, P. (2015). Trazando el mapa social de Buenos Aires: dos décadas de cambios en la Ciudad, *Población de Buenos Aires. 2015*, vol. 12, nº 21, p. 7-39. http://ddd.uab.cat/pub/artpub/2015/132095/pobbueair_a2015n21p7iSPA_postprint.pdf,

Comunalidades

	Inicial	Extracción
H12_Hasta0.50personas1 % Hacinamiento <0,5	1,000	0,816
H12_1.501.99personas1 % Hacinamiento 1,5-2	1,000	0,579
H12_2.003.00personas1 % Hacinamiento 2-3	1,000	0,895
H12_Másde3.00personas1 % Hacinamiento +3	1,000	0,844
P3_Otropaís1 % Extranjeros	1,000	0,738
P5_Nuncaasistió1 Nunca asistió	1,000	0,699
P6_Primario1 Primarios	1,000	0,893
P6_Superiornouniversitario1 Superior	1,000	0,597
P6_Universitario1 Univesitarios	1,000	0,874
V2_Departamento1 Departamento	1,000	0,721
V2_Piezaeninquilinato1 Inquinilato	1,000	0,671
V2_Rancho1 Rancho	1,000	0,459
V9_Inquilino1 Inquilino	1,000	0,892
V9_Propietariodelaviviendaydelterreno1 Propietario	1,000	0,826
Densidad Densidad	1,000	0,607

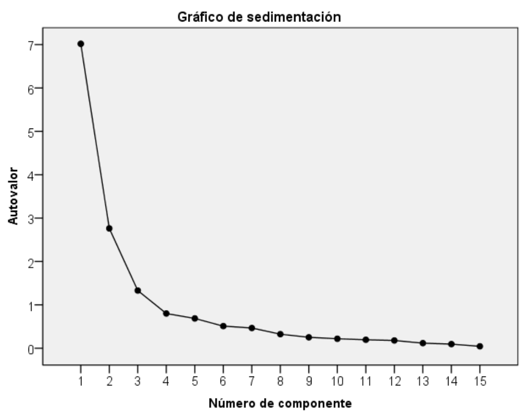
Método de extracción: análisis de componentes principales.

La varianza explicada por los tres factores es:

Varianza total explicada

Componente	Autovalores iniciales			Sumas de extracción de cargas al cuadrado			Sumas de rotación de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	7,019	46,790	46,790	7,019	46,790	46,790	4,211	28,071	28,071
2	2,763	18,419	65,209	2,763	18,419	65,209	3,948	26,320	54,391
3	1,330	8,865	74,074	1,330	8,865	74,074	2,952	19,683	74,074
4	0,800	5,336	79,410						
5	0,687	4,577	83,987						
6	0,510	3,398	87,385						
7	0,466	3,108	90,494						
8	0,324	2,161	92,654						
9	0,252	1,678	94,333						
10	0,219	1,460	95,793						
11	0,195	1,302	97,095						
12	0,179	1,195	98,290						
13	0,117	0,782	99,072						
14	0,095	0,632	99,705						
15	0,044	0,295	100,000						

Método de extracción: análisis de componentes principales.



Matriz de componente rotado[a]

	Componente 1	2	3
P6_Universitario1 Univesitarios	0,871	-0,323	-0,106
P6_Primario1 Primarios	-0,814	0,475	0,067
V2_Departamento1 Departamento	0,779	-0,296	0,162
H12_Hasta0.50personas1 % Hacinamiento <0,5	0,778	-0,375	-0,263
H12_1.501.99personas1 % Hacinamiento 1,5-2	-0,715	0,252	0,065
Densidad Densidad	0,627	0,434	0,160
H12_Másde3.00personas1 % Hacinamiento +3	-0,314	0,788	0,353
P5_Nuncaasistió1 Nunca asistió	-0,354	0,747	0,127
P3_Otropaís1 % Extranjeros	-0,169	0,744	0,395
H12_2.003.00personas1 % Hacinamiento 2-3	-0,386	0,691	0,519
V2_Rancho1 Rancho	-0,113	0,657	-0,119
P6_Superiornouniversitario1 Superior	0,418	-0,645	-0,082
V9_Inquilino1 Inquilino	0,110	-0,038	0,937
V9_Propietariodelaviviendaydelterreno1 Propietario	-0,088	-0,173	-0,888
V2_Piezaeninquilinato1 Inquinilato	-0,251	0,223	0,747

Método de extracción: análisis de componentes principales.
Método de rotación: Varimax con normalización Kaiser.

a. La rotación ha convergido en 5 iteraciones.

Con se explica en Fachelli, S.; Goicoechea, M. E.; López-Roldán, P. (2015) el Factor 1 se interpreta como una dimensión de diferenciación socio-económica junto a densidad poblacional, el Factor 2 como nivel de segregación sociorresidencial y Factor 3 como nivel de estabilidad residencial.

► Ejercicio 1. Propuesto

A partir de la matriz de datos CABA-Censo2010Torres.sav y siguiendo la forma de proceder de los ejercicios de análisis de componentes principales presentados anteriormente reproduce un análisis de ACP seleccionando las variables definidas por el modelo que se presenta a continuación:

Reproducción Modelo Torres	
A. Hacinamiento (más de dos personas por cuarto)	
H12_Hasta0.50personas1*	H12_1.501.99personas1
H12_0.510.99personas1	H12_2.003.00personas1
H12_1.001.49personas1	H12_Másde3.00personas1
B. Tipo de tenencia de vivienda	**D. Origen**
V9_Propietariodelaviviendaydelterreno1	P3_Argentina1*
V9_Propietariosólodelavivienda1	P3_Otropaís1
V9_Inquilino1	**E. Tipo de vivienda**
V9_Ocupanteporpréstamo1	V2_Casa1*
V9_Ocupanteportrabajo1	V2_Rancho1
V9_Otrasituación1*	V2_Casilla1
C. Educación	V2_Departamento1
P5_Nuncaasistió1	V2_Piezaeninquilinato1
P6_Primario1	V2_Piezaenhotelfamiliaropensión1
P6_Secundario1	V2_Localnoconstruidoparahabitación1
P6_Superiornouniversitario1	V2_Viviendamóvil1
P6_Universitario1	V2_Personasviviendoenlacalle1
* Variables eliminadas para evitar combinación lineal perfecta	

Las variables marcadas con un * no se deben considerar junto con todas las otra de cada bloque otras pues suman el 100% generando un problema de cálculo al formar una combinación lineal perfecta.

La variable H12 se basa en una escala con el nivel de hacinamiento del hogar (hacinamiento extremo: 3 personas por cuarto, sin contar baño y cocina). El tipo de vivienda de rancho o casilla corresponden a viviendas muy precarias (tipo barraca). Inquilinato significa una vivienda precaria donde viven varias familias alquilando habitaciones.

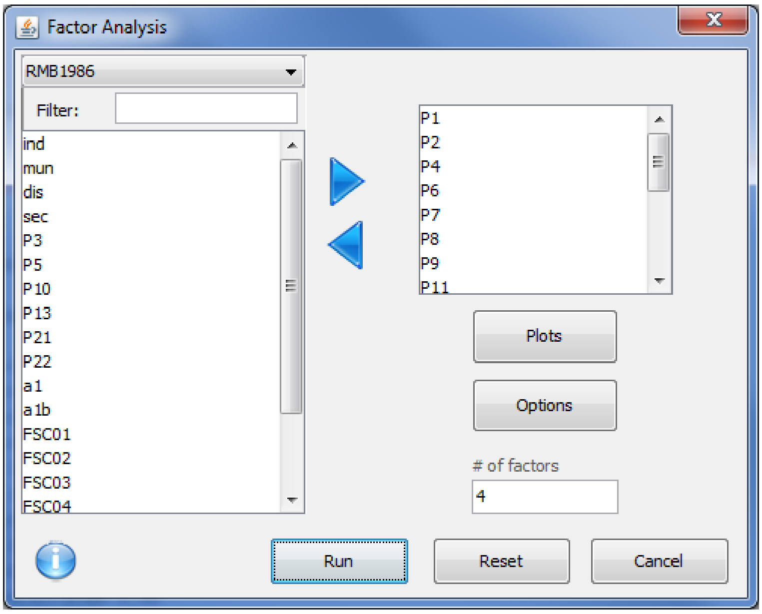
Interpreta los resultados obtenidos inicialmente en el ACP. Con estos resultados ¿se puede mejorar el modelo? Analiza la matriz anti-imagen para valorar la eliminación de variables con valores bajos y ten en cuenta al mismo tiempo que no baje la varianza explicada. Interpreta los nuevos resultados obtenidos.

4. Análisis de componentes principales con R

Existen diferentes alternativas para realizar un análisis factorial de componentes en R.

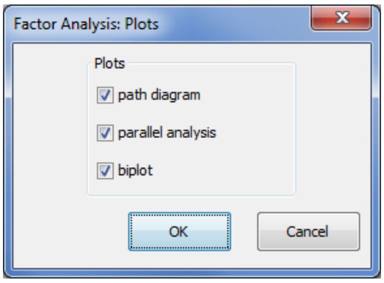
A través de Deducer podemos ejecutar con el menú el procedimiento Factor analysis del menú Psych. Para ello se requiere tener instalado y cargado el Reliability and factor analysis plugin. Se ejecuta la función `principal`.

Reproduciremos el ejemplo de análisis de los datos censales de la Región Metropolitana de Barcelona. Consideramos las variables P1, P2, P4, P6, P7, P8, P9, P11, P12, P14, P15, P16, P17, P18, P19, P20 y P23 que colocaremos en el recuadro de variables de la derecha:

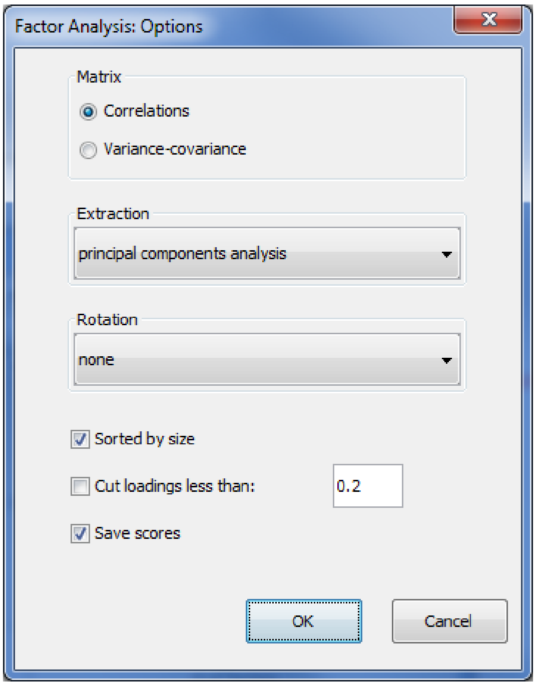


Hemos elegido que se extraigan 4 factores si bien esta es una decisión que se establece posteriormente cuando se ejecuta el procedimiento. Veremos que entre los resultados de la ejecución nos sugerirá el número de factores.

En Plots marcaremos las distintas opciones: path diagram que representa la relación entre variables originales y factores, parallel analysis para determinar gráficamente el número de factores (*scree plot*) y biplot para representar la nube de puntos de los individuos en el espacio cada pareja de factores junto con las variables originales.



En Options podemos especificar el método de extracción: elegiremos principal components analysis partiendo de la Correlation Matrix. En Rotation podemos elegir diversas alternativas de rotación de la solución factorial inicial. Es una alternativa no obligatoria del análisis destinada a facilitar la interpretación de los factores. En un primer momento realizaremos el análisis sin la rotación, opción none, y contrastaremos posteriormente los resultados con una rotación varimax que elegiremos del desplegable.



Para facilitar la lectura de la tabla de saturaciones o de factores de carga (*standardized loadings*) podemos elegir la opción Sorted by size. También podríamos filtrar los valores más bajos. Finalmente tenemos la opción de guardar la variables factoriales con Save scores: se creará una nueva matriz de datos de nombre RMB1986.1 con las variables factoriales del análisis y puntuaciones generadas con o sin rotación, según hayamos especificado una u otra opción.

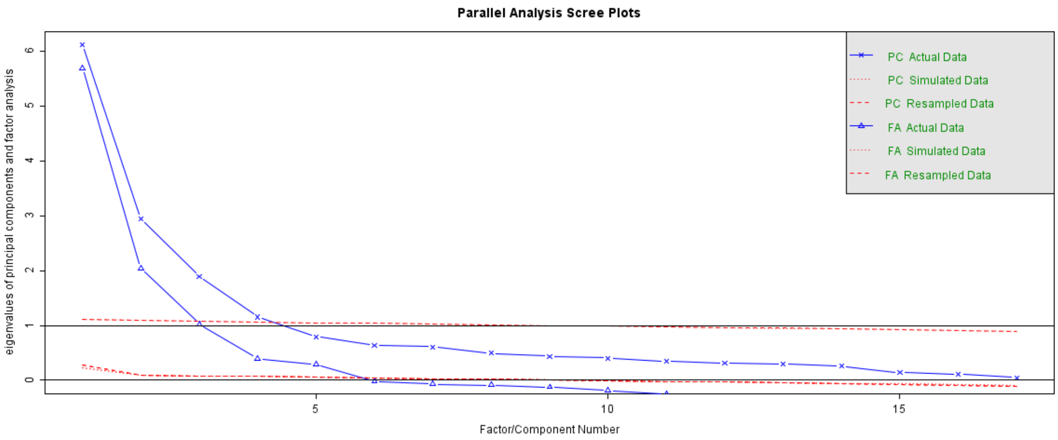
Los resultados que se obtienen se presentan a continuación. Nos aparece en primer lugar la tabla que relaciona las variables originales con los 4 factores (denominados PC1 a PC4), ordenadas según la mayor correlación o la contribución a cada factor. Las columnas h2 y u2 suman siempre 1 y se corresponden con la comunalidad (h2), lo que aporta cada variable al conjunto de los 4 factores y la parte excluida (u2) por el hecho de retener solamente cuatro factores.

Los cuatro factores acumulan el 71% de la varianza explicada como se puede ver después de la tabla de saturaciones o factores de carga. Se presentan primeramente los valores propios (denominados SS loadings), la proporción de varianza explicada retenida por los 4 primeros factores y el recálculo de esta proporción sobre el 100% que representan los cuatro. Así vemos como el primer factor representa algo más de la mitad del conjunto de los cuatro (51%).

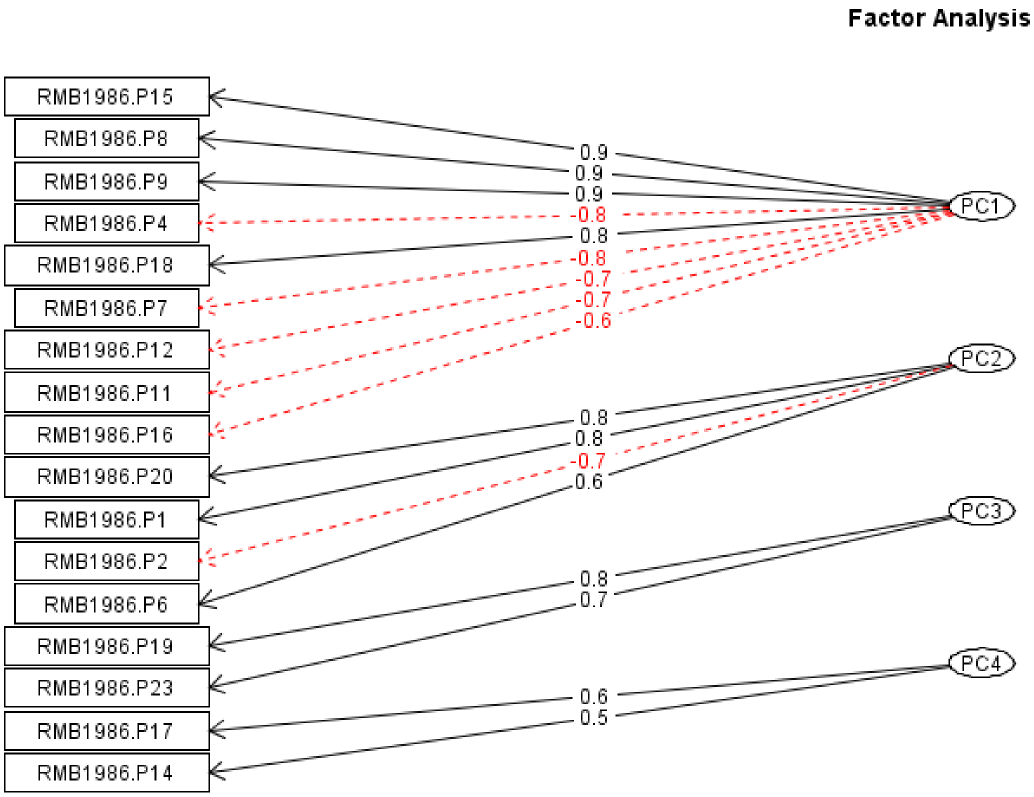
El número de 4 factores es el mensaje que aparece a continuación y que puede verse gráficamente en el Parallel analysis scree plot y cuya conclusión aparece más tarde en los resultados del análisis: *Parallel analysis suggests that the number of components = 4*.

Finalmente aparece la significación del modelo (prob<0) y la prueba de normalidad de Shapiro-Wilk que arroja un resultado significativo, de no normalidad (p-value<0,000).

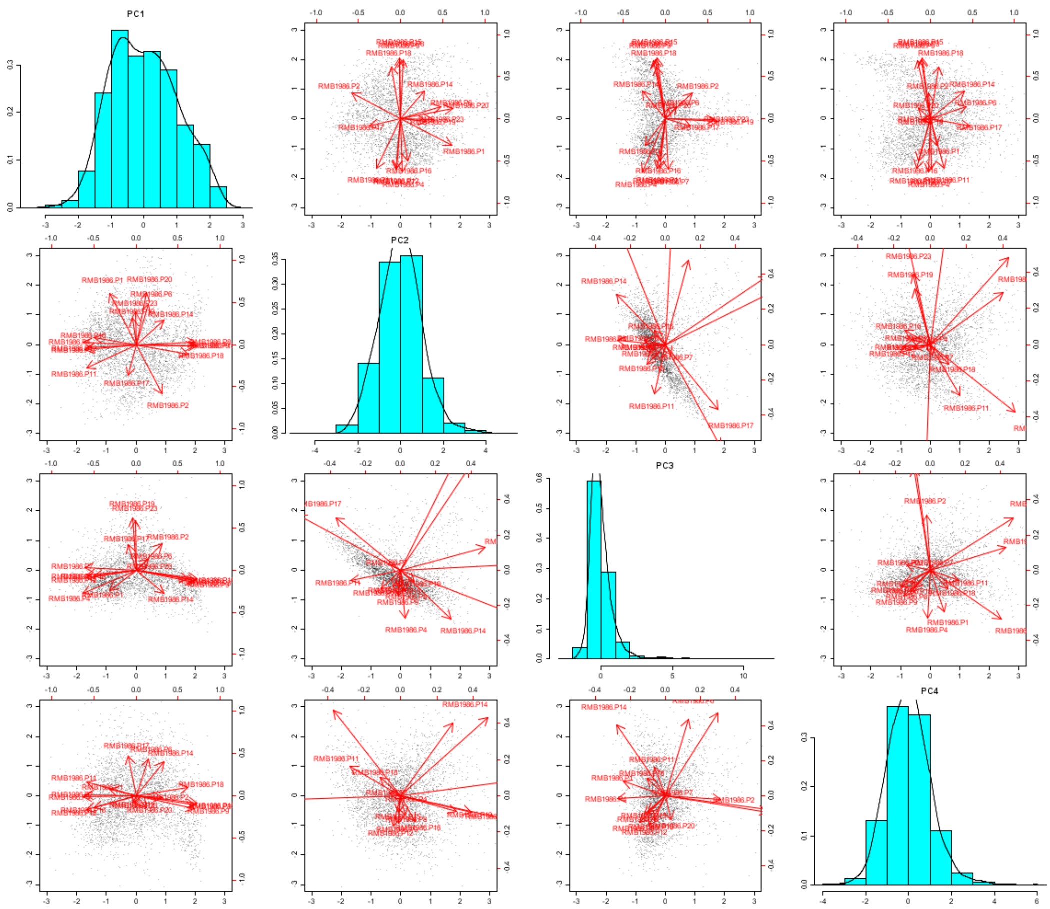
```
> pr.model<-data.frame(RMB1986$P1, RMB1986$P2, RMB1986$P4,
RMB1986$P6, RMB1986$P7, RMB1986$P8, RMB1986$P9, RMB1986$P11,
RMB1986$P12, RMB1986$P14, RMB1986$P15, RMB1986$P16, RMB1986$P17,
RMB1986$P18, RMB1986$P19, RMB1986$P20, RMB1986$P23)
> pr.model1<-principal(pr.model, nfactors= 4 , rotate= "none" ,
covar= FALSE , scores=TRUE )
> print.psych(pr.model1 , sort=T )
Principal Components Analysis
Call: principal(r = pr.model, nfactors = 4, rotate = "none", covar = FALSE,
    scores = TRUE)
Standardized loadings (pattern matrix) based upon correlation matrix
            item   PC1   PC2   PC3   PC4   h2   u2
RMB1986.P15   11  0.91 -0.02 -0.12 -0.11 0.85 0.15
RMB1986.P8     6  0.88  0.03 -0.15 -0.13 0.82 0.18
RMB1986.P9     7  0.86 -0.02 -0.18 -0.19 0.81 0.19
RMB1986.P4     3 -0.78  0.03 -0.34 -0.02 0.73 0.27
RMB1986.P18   14  0.77 -0.14 -0.13  0.13 0.65 0.35
RMB1986.P7     5 -0.75 -0.08  0.04  0.02 0.57 0.43
RMB1986.P12    9 -0.73 -0.05 -0.12 -0.21 0.60 0.40
RMB1986.P11    8 -0.73 -0.35 -0.08  0.20 0.70 0.30
RMB1986.P16   12 -0.62  0.10 -0.08 -0.18 0.44 0.56
RMB1986.P20   16  0.16  0.77  0.04 -0.17 0.65 0.35
RMB1986.P1     1 -0.39  0.76 -0.30  0.10 0.82 0.18
RMB1986.P2     2  0.38 -0.73  0.39 -0.03 0.83 0.17
RMB1986.P6     4  0.19  0.60  0.16  0.54 0.71 0.29
RMB1986.P19   15 -0.04  0.38  0.78 -0.11 0.76 0.24
RMB1986.P23   17 -0.01  0.48  0.73 -0.12 0.78 0.22
RMB1986.P17   13 -0.11 -0.46  0.37  0.59 0.71 0.29
RMB1986.P14   10  0.41  0.36 -0.35  0.50 0.67 0.33

                       PC1  PC2  PC3  PC4
SS loadings           6.11 2.94 1.89 1.15
Proportion Var        0.36 0.17 0.11 0.07
Cumulative Var        0.36 0.53 0.64 0.71
Proportion Explained  0.51 0.24 0.16 0.10
Cumulative Proportion 0.51 0.75 0.90 1.00

Test of the hypothesis that 4 components are sufficient.

The degrees of freedom for the null model are  136  and the objective function
was  11.93
The degrees of freedom for the model are 74  and the objective function was  2.45
The total number of observations was  3509  with MLE Chi Square =  8574.72  with
prob <  0

Fit based upon off diagonal values = 0.98
> RMB1986.1 <- cbind(pr.model, pr.model1$scores)
> JavaGD(width=800, height=600, ps=12)
> par(mfrow=c(2,1))
> pr.model2<- fa.diagram(pr.model1)
> pr.model3<- fa.parallel(pr.model)
Parallel analysis suggests that the number of factors =  5  and the number of
components =  4
> JavaGD(width=500, height=500, ps=12)
> biplot(pr.model1)
> print(mshapiro.test(t(na.omit(as.matrix(pr.model)))))
        Shapiro-Wilk normality test
data:  Z
W = 0.7916, p-value < 2.2e
```



El gráfico siguiente representa el vínculo entre las variables originales y los factores, y reproduce los datos de la tabla de factores de carga anterior.

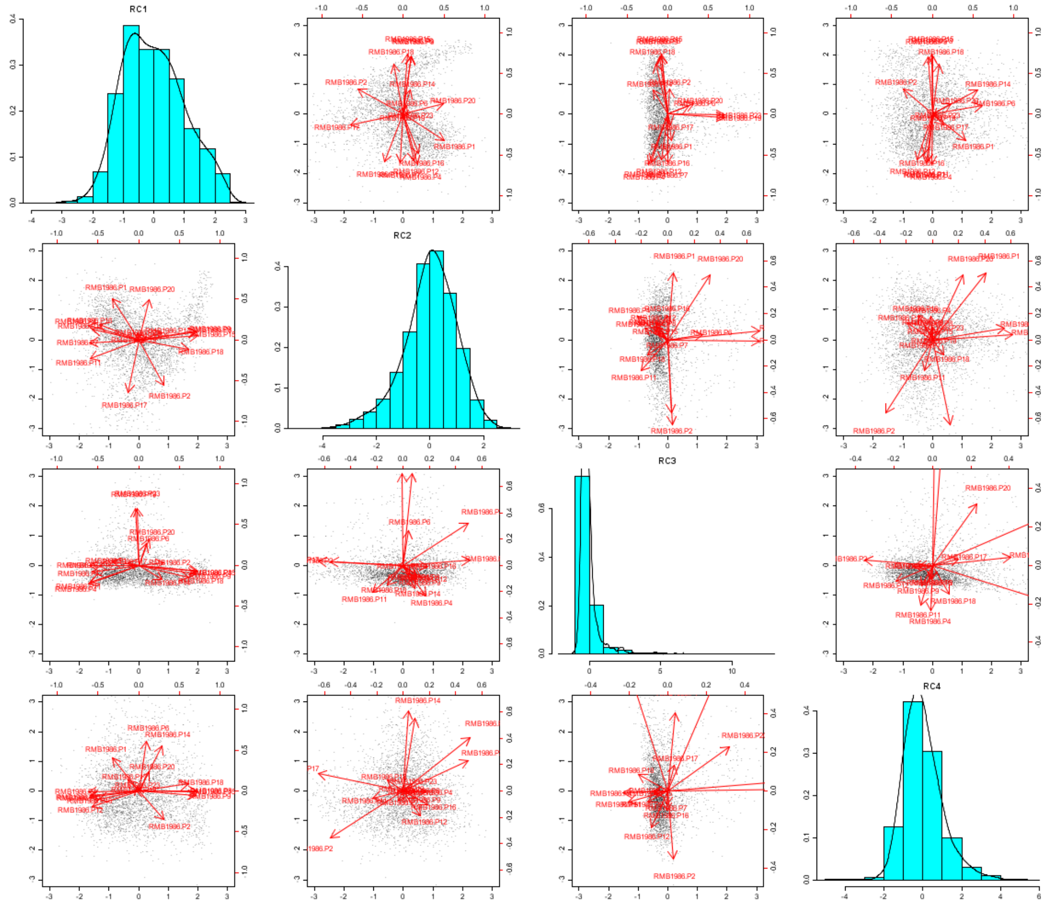


Los resultados del análisis de componentes se completan con la representación gráfica matricial entre pares de factores como aparece en el gráfico siguiente. Nos remitimos a la interpretación del contenido de los factores que se realizó en el texto al presentar el ACP con este ejemplo.



Si ejecutamos de nuevo el análisis con la rotación varimax de los factores los resultados son los siguientes. La única diferencia con los resultados anteriores es la nueva matriz de factores de carga y las representaciones gráficas de los biplot que muestran pequeñas variaciones para reforzar el carácter de los últimos factores.

```
            item   RC1   RC2   RC3   RC4   h2   u2
RMB1986.P15   11  0.92  0.06 -0.09 -0.02 0.85 0.15
RMB1986.P8     6  0.89  0.12 -0.09  0.00 0.82 0.18
RMB1986.P9     7  0.88  0.13 -0.13 -0.07 0.81 0.19
RMB1986.P4     3 -0.77  0.22 -0.29 -0.01 0.73 0.27
RMB1986.P18   14  0.76 -0.15 -0.19  0.11 0.65 0.35
RMB1986.P7     5 -0.75 -0.04 -0.01 -0.09 0.57 0.43
RMB1986.P11    8 -0.74 -0.29 -0.26 -0.07 0.70 0.30
RMB1986.P12    9 -0.71  0.16 -0.11 -0.24 0.60 0.40
RMB1986.P16   12 -0.60  0.23 -0.01 -0.13 0.44 0.56
RMB1986.P17   13 -0.16 -0.81  0.04  0.17 0.71 0.29
RMB1986.P2     2  0.38 -0.69  0.03 -0.44 0.83 0.17
RMB1986.P1     1 -0.40  0.63  0.05  0.51 0.82 0.18
RMB1986.P20   16  0.16  0.62  0.40  0.29 0.65 0.35
RMB1986.P23   17 -0.02  0.09  0.87  0.07 0.78 0.22
RMB1986.P19   15 -0.05 -0.01  0.87  0.01 0.76 0.24
RMB1986.P6     4  0.13  0.05  0.33  0.76 0.71 0.29
RMB1986.P14   10  0.37  0.11 -0.22  0.69 0.67 0.33
```



Algunos otros resultados posibles de un ACP, algunos necesarios para validar el análisis, no aparecen entre los que hemos visto. Los pediremos a continuación a través de instrucciones de R.

En primer lugar pediremos los estadísticos descriptivos de las variables del análisis (menú de de Deducer: Analysis/Descriptives):

	Mean	St. Deviation	Valid N
P1	20.178	6.36	3509
P2	13.172	6.83	3509
P3	86.104	114.70	3509
P4	37.393	11.47	3509
P6	5.842	3.21	3509
P7	5.181	4.40	3509
P8	9.674	10.22	3509
P9	47.650	14.09	3509
P11	14.890	5.05	3509
P12	9.166	4.76	3509
P14	31.775	5.61	3509
P15	18.116	13.26	3509
P16	4.208	3.71	3509
P17	12.398	4.40	3509
P18	4.289	3.35	3509
P19	0.986	2.97	3509
P20	37.658	12.32	3509
P23	3.676	13.25	3509

A continuación la matriz de correlaciones a través del menú de Deducer: Analysis/Correlation especificando que extraiga el coeficiente y la significación.

Pearson's product-moment correlation

		P1	P2	P4	P6	P7	P8	P9	P11	P12	P14	P15	P16	P17	P18	P19	P20	P23
P1	cor	1	-0.8274	0.387	0.3291	0.197	-0.3104	-0.2558	0.05181	0.1992	0.1723	-0.3565	0.2867	-0.2948	-0.3507	0.07509	0.4419	0.1192
	p-value		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0021	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P2	cor	-0.8274	1	-0.4586	-0.2403	-0.1539	0.2769	0.2209	-0.03082	-0.2707	-0.1978	0.3493	-0.282	0.3434	0.305	-0.02594	-0.4106	-0.08168
	p-value	0.0000		0.0000	0.0000	0.0000	0.0000	0.0000	0.0679	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1245	0.0000	0.0000
P4	cor	0.387	-0.4586	1	-0.1514	0.5226	-0.6283	-0.5614	0.5157	0.525	-0.2502	-0.6688	0.4595	-0.05721	-0.5779	-0.1854	-0.1625	-0.1933
	p-value	0.0000	0.0000		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	7e-04	0.0000	0.0000	0.0000	0.0000
P6	cor	0.3291	-0.2403	-0.1514	1	-0.0841	0.1482	0.03359	-0.2382	-0.2303	0.3603	0.1458	-0.1224	-0.008013	0.1069	0.2453	0.3822	0.299
	p-value	0.0000	0.0000	0.0000		0.0000	0.0000	0.0467	0.0000	0.0000	0.0000	0.0000	0.0000	0.6351	0.0000	0.0000	0.0000	0.0000
P7	cor	0.197	-0.1539	0.5226	-0.0841	1	-0.549	-0.6476	0.5917	0.6357	-0.308	-0.5585	0.4042	0.1038	-0.5433	0.04839	-0.1209	-0.03101
	p-value	0.0000	0.0000	0.0000	0.0000		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0041	0.0000	0.0663
P8	cor	-0.3104	0.2769	-0.6283	0.1482	-0.549	1	0.8053	-0.6046	-0.4967	0.3864	0.9489	-0.4601	-0.2137	0.6296	-0.1077	0.1976	-0.07832
	p-value	0.0000	0.0000	0.0000	0.0000	0.0000		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P9	cor	-0.2558	0.2209	-0.5614	0.03359	-0.6476	0.8053	1	-0.6545	-0.6238	0.218	0.8127	-0.4825	-0.1621	0.6224	-0.1578	0.1292	-0.1345
	p-value	0.0000	0.0000	0.0000	0.0467	0.0000	0.0000		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P11	cor	0.05181	-0.03082	0.5157	-0.2382	0.5917	-0.6046	-0.6545	1	0.5226	-0.278	-0.6126	0.347	0.2447	-0.4608	-0.1741	-0.3466	-0.2044
	p-value	0.0021	0.0679	0.0000	0.0000	0.0000	0.0000	0.0000		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P12	cor	0.1992	-0.2707	0.525	-0.2303	0.6357	-0.4967	-0.6238	0.5226	1	-0.3003	-0.5288	0.4456	-0.03664	-0.5066	-0.03427	-0.09495	-0.08045
	p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		0.0000	0.0000	0.0000	0.0300	0.0000	0.0424	0.0000	0.0000
P14	cor	0.1723	-0.1978	-0.2502	0.3603	-0.308	0.3864	0.218	-0.278	-0.3003	1	0.3649	-0.2159	-0.1365	0.3546	-0.1288	0.2098	-0.07177
	p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P15	cor	-0.3565	0.3493	-0.6688	0.1458	-0.5585	0.9489	0.8127	-0.6126	-0.5288	0.3649	1	-0.4978	-0.1806	0.654	-0.121	0.1816	-0.0961
	p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P16	cor	0.2867	-0.282	0.4595	-0.1224	0.4042	-0.4601	-0.4825	0.347	0.4456	-0.2159	-0.4978	1	-0.02772	-0.4405	-0.005731	-0.005649	0.03477
	p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		0.1006	0.0000	0.7343	0.7380	0.0394
P17	cor	-0.2948	0.3434	-0.05721	-0.008013	0.1038	-0.2137	-0.1621	0.2447	-0.03664	-0.1365	-0.1806	-0.02772	1	-0.04715	0.01598	-0.3409	-0.05042
	p-value	0.0000	0.0000	7e-04	0.6351	0.0000	0.0000	0.0000	0.0000	0.0300	0.0000	0.0000	0.1006		0.0052	0.3439	0.0000	0.0028
P18	cor	-0.3507	0.305	-0.5779	0.1069	-0.5433	0.6296	0.6224	-0.4608	-0.5066	0.3546	0.654	-0.4405	-0.04715	1	-0.1646	-0.0408	-0.1418
	p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0052		0.0000	0.0156	0.0000
P19	cor	0.07509	-0.02594	-0.1854	0.2453	0.04839	-0.1077	-0.1578	-0.1741	-0.03427	-0.1288	-0.121	-0.005731	0.01598	-0.1646	1	0.2466	0.6362
	p-value	0.0000	0.1245	0.0000	0.0000	0.0041	0.0000	0.0000	0.0000	0.0424	0.0000	0.0000	0.7343	0.3439	0.0000		0.0000	0.0000
P20	cor	0.4419	-0.4106	-0.1625	0.3822	-0.1209	0.1976	0.1292	-0.3466	-0.09495	0.2098	0.1816	-0.005649	-0.3409	-0.0408	0.2466	1	0.3552
	p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.7380	0.0000	0.0156	0.0000		0.0000
P23	cor	0.1192	-0.08168	-0.1933	0.299	-0.03101	-0.07832	-0.1345	-0.2044	-0.08045	-0.07177	-0.0961	0.03477	-0.05042	-0.1418	0.6362	0.3552	1
	p-value	0.0000	0.0000	0.0000	0.0000	0.0663	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0394	0.0028	0.0000	0.0000	0.0000	

HA: two.sided

Podemos calcular el determinante de la matriz de correlaciones:

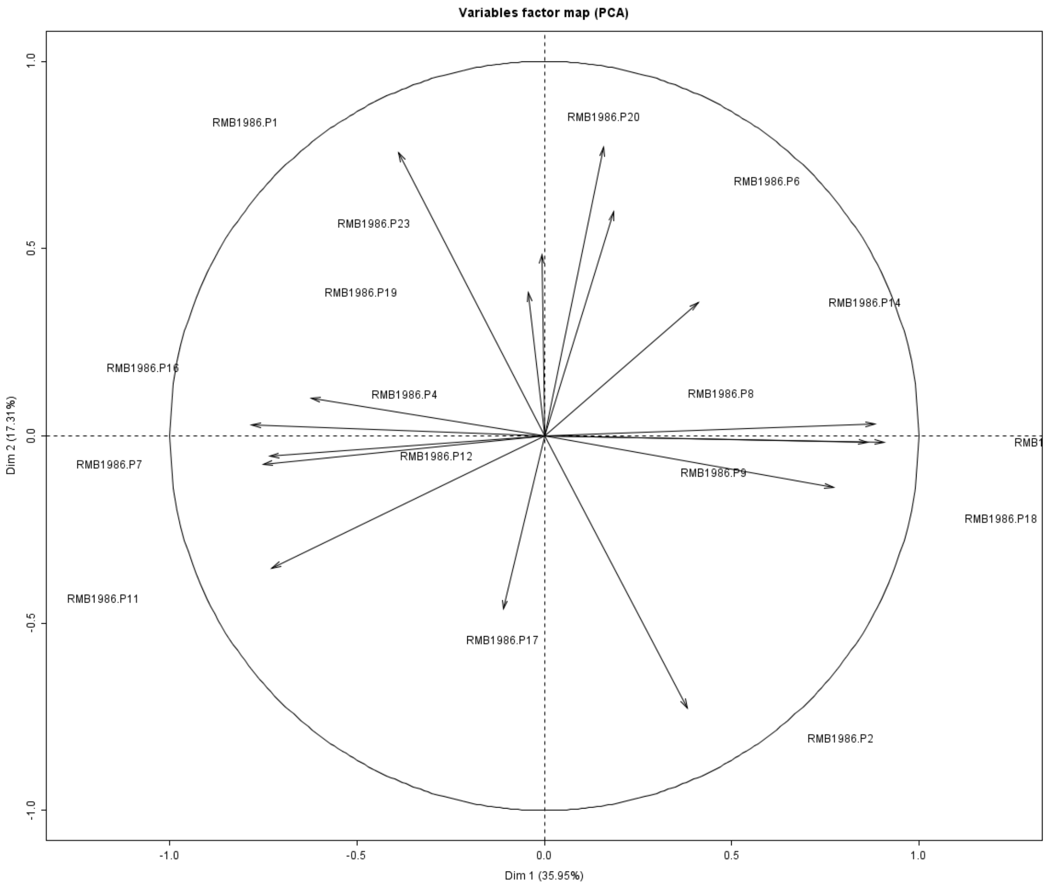
```
# Matriz de correlaciones
R=cor(DF)
det(R)
[1] 6.594877e-06
```

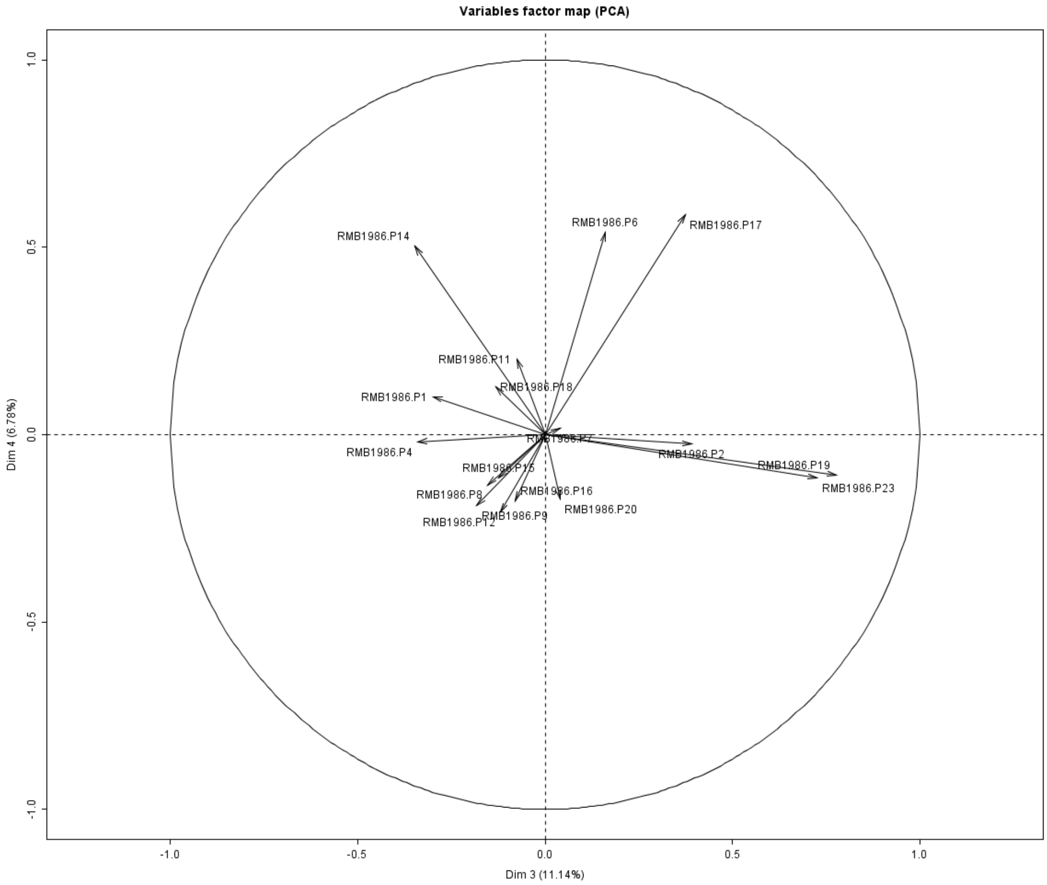
Así como realizar la prueba de Barlett y solicitar el valor del KMO y las medidas de adecuación muestral de cada variables (MSA):

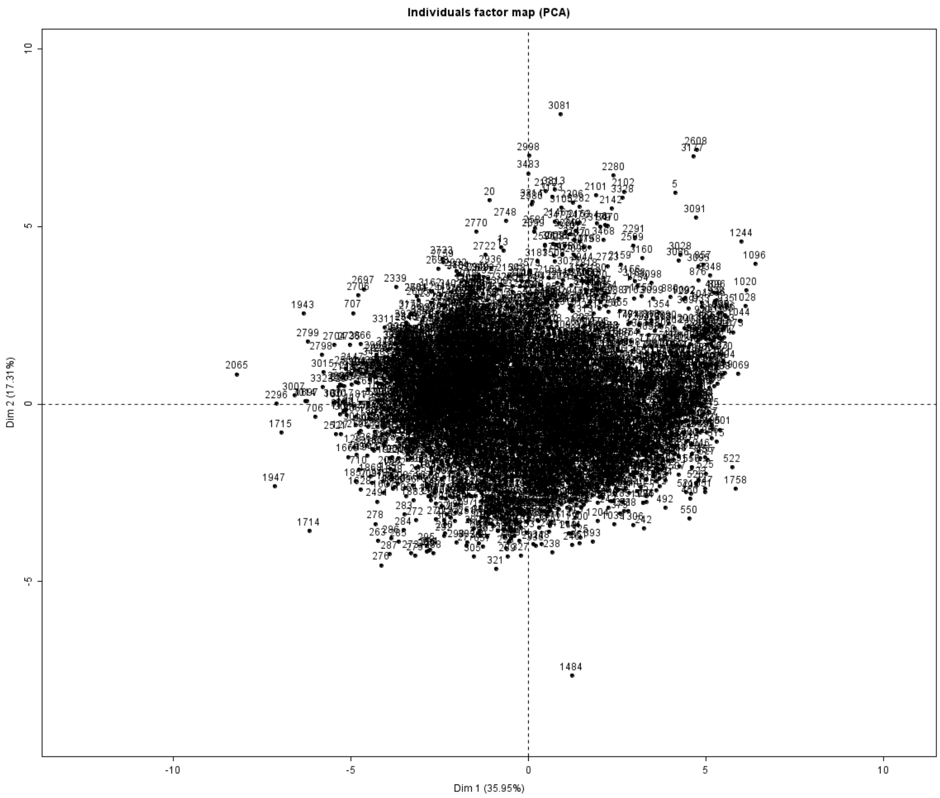
```
> # Test de Bartlett, KMO y Medida de adecuación muestral, psych
library
> KMO(R)
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = R)
Overall MSA =  0.85
MSA for each item =
RMB1986.P1  RMB1986.P2  RMB1986.P4  RMB1986.P6  RMB1986.P7  RMB1986.P8
0.77        0.71        0.90        0.73        0.92        0.85
RMB1986.P9 RMB1986.P11 RMB1986.P12 RMB1986.P14 RMB1986.P15 RMB1986.P16
0.87       0.93        0.87        0.75        0.84        0.96
RMB1986.P17 RMB1986.P18 RMB1986.P19 RMB1986.P20 RMB1986.P23
0.76        0.94        0.67        0.83        0.68
> cortest.bartlett(R,n=3509)
$chisq
[1] 41770.15

$p.value
[1] 0

$df
[1] 136
```

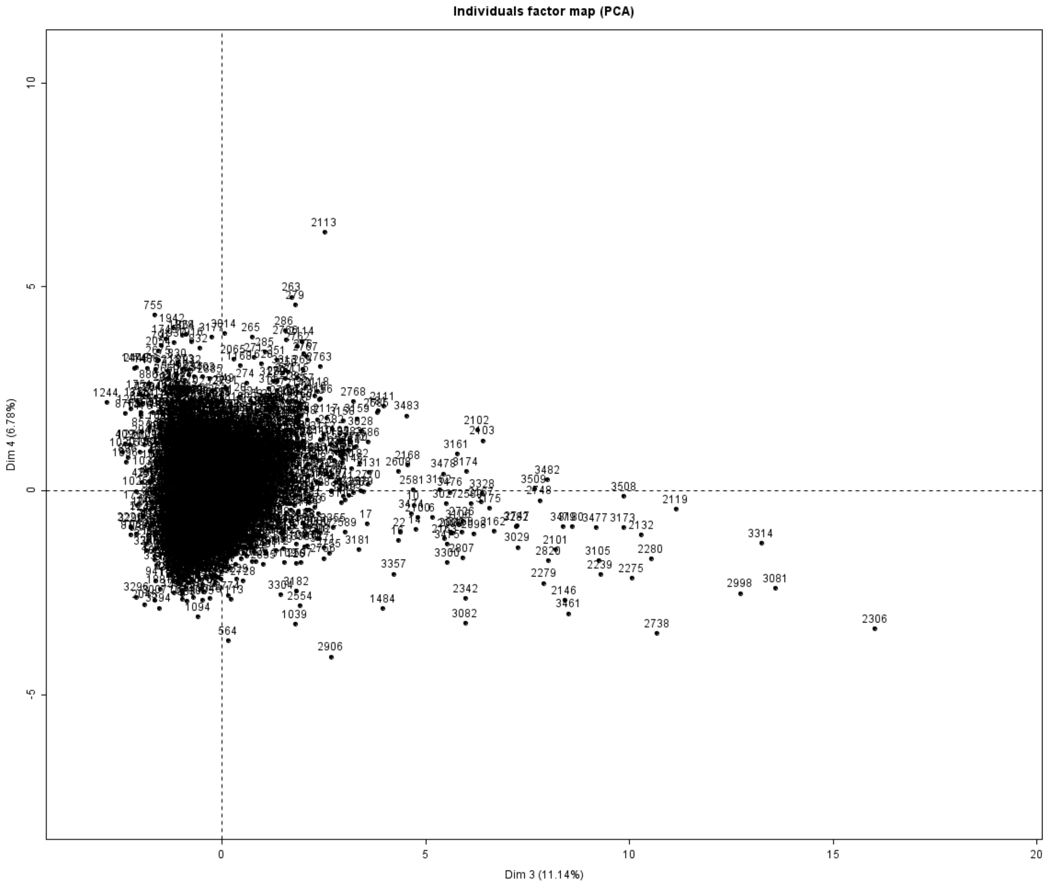
Para obtener los gráficos factoriales podemos instalar el paquete FactoMineR, que también realiza el análisis factorial de componentes principales (función `PCA`), y solicitar los gráficos con las varaibles originales y los individuos.

```
# FactoMineR package
install.packages("FactoMineR")
library(FactoMineR)
ACP=PCA(DF, ncp=4, graph=F)
# Gráficos factoriales
plot.PCA(ACP, axes=c(1,2), choix="var")
plot.PCA(ACP, axes=c(3,4), choix="var")
plot.PCA(ACP, axes=c(1,2), choix="ind")
plot.PCA(ACP, axes=c(3,4), choix="ind")
```




Variables factor map (PCA)
RMB1986.P14
RMB1986.P6
RMB1986.P17
RMB1986.P11
RMB1986.P18
RMB1986.P1
RMB1986.P7
RMB1986.P4
RMB1986.P2
RMB1986.P19
RMB1986.P15
RMB1986.P23
RMB1986.P8
RMB1986.P16
RMB1986.P9
RMB1986.P20
RMB1986.P12
Dim 4 (6.78%)
Dim 3 (11.14%)


Individuals factor map (PCA)
Dim 2 (17.31%)
Dim 1 (35.95%)


Individuals factor map (PCA)
Dim 3 (11.14%)
Dim 4 (6.78%)

5. Análisis de correspondencias

El análisis de correspondencia es similar al análisis de componentes principales con la particularidad de tratar variables cualitativas medidas a nivel nominal. Pertenece también al grupo de técnicas que en la presentación del capítulo denominamos como de Análisis Factorial de Varianza Total, donde se extraen los factores comunes explicando el total de la varianza.

El análisis multivariable de datos cualitativos o categóricos ha sido objeto de tratamiento más tardío y siguiendo los modelos o los métodos que se han formulado inicialmente con variables métricas. Hoy en día disponemos de una amplia gama de técnicas que posibilitan el tratamiento de variables cualitativas, y su introducción en los paquetes estadísticos ha facilitado la creciente generalización de su uso entre los científicos sociales. El análisis log-lineal, los modelos logit o la regresión logística son procedimientos de análisis habituales de la investigación social que pueden complementar la técnica que aquí presentamos.

La formulación original del análisis de correspondencias es obra de Jean-Paul Benzécri (1963), quien genera la tradición de la denominada Escuela Francesa de l'*Analyse des Données* (Benzécri, 1973; Lebart et al., 2004; Crivisqui, 1993; Greenacre, 2008) y ha sido implementado en varios softwares como el SPAD (*Système Portable pour l'Analyse des Données*). Esta perspectiva es la que presentaremos principalmente en estas páginas[18].

La tradición holandesa con el grupo de *Data Theory Scaling System Group* (DTSS) de la Universidad de Leiden (Gifi, 1981) es un enfoque diferente que se ha implementado en el programa SPSS donde se habla de técnicas de Escalamiento Óptimo (*Optimal Scaling*). Complementaremos el análisis obtenido en SPAD con la implementación del SPSS.

El análisis de correspondencias es una técnica estadística destinada al estudio de la asociación o interrelación entre variables cualitativas. Existen dos formulaciones principales: el análisis de correspondencias simples (ACS), que considera tan sólo dos variables cualitativas (o una tabla de dos dimensiones, que llamaremos tabla de correspondencias, donde las tablas de contingencia son un caso particular) y el análisis de correspondencias múltiples (ACM), que es una generalización al caso de más de dos variables.

Como técnica de análisis factorial trata la tabla de correspondencias de forma que reduce la cantidad de información original de los datos y expresa las fuentes principales de las asociaciones (las correspondencias) entre las filas y columnas como factores de variabilidad. Conceptualmente, el análisis de correspondencias es similar al análisis factorial de componentes principales, pero a partir de tablas de números positivos y tratando de forma equivalente filas y columnas a partir de la métrica del chi-cuadrado. Se presenta primero el análisis de correspondencias simples y a continuación la versión multivariable, con varios ejemplos que se tratan tanto con el SPAD como con el SPSS.

[18] Correa (2008) ha destacado diversos antecedentes del análisis de correspondencias a través de: Escalamiento óptimo de la Escuela Americana, Escalamiento dual canadiense, el Análisis de escalograma israelí, el Método de cuantificación japonés, los Promedios recíprocos, el Análisis Canónico Generalizado o el Análisis de varianza.

5.1. Análisis de correspondencias simples

El objetivo del ACS es analizar, describir y representar gráficamente las relaciones existentes entre dos variables donde se considera el nivel de medición nominal, por lo tanto, donde se pueden incluir todo tipo de variables que adopten un conjunto relativamente reducido de valores y con una mínima frecuencia. Se pueden considerar así tanto variables nominales y ordinales como variables cuantitativas; en este último caso directamente si la variable numérica es discreta con una unos pocos valores, o discretizando (agrupando) la variable continua. En todos los casos son tratadas con el mínimo nivel de medida nominal.

El análisis se realiza a partir de una tabla de correspondencias (conjunto de números positivos dispuestos en una tabla de doble entrada, en filas y columnas), donde las tablas de contingencia no son más que un caso particular. El ACS se entiende como una técnica destinada a obtener representaciones gráficas entre las proximidades geométricas entre puntos-línea y puntos-columna que traducen las asociaciones estadísticas entre las filas y las columnas de la tabla, teniendo en cuenta las diferencias de efectivos existentes entre estas filas y columnas.

El análisis de correspondencias se inscribe dentro de las técnicas de análisis factorial y por lo tanto el objetivo consiste en extraer unos factores (o dimensiones) que permitan reducir la dimensión de la tabla informativa inicial para obtener otra que, perdiendo una parte de la información, gane en significación, sea más fácil de interpretar y pueda ser representada gráficamente. Por tanto, el análisis y la descripción de las relaciones entre las dos variables implica encontrar un espacio de pocas dimensiones, generalmente dos en el ACS, donde se pueden representar y analizar las asociaciones entre las categorías (valores o modalidades) de cada variable. Esta representación nos da un gráfico donde cada categoría de las variables se representa por un punto; las distancias entre los puntos de categorías reflejan las relaciones o asociaciones entre ellas: las categorías que están asociadas, que son similares, aparecerán representadas en el espacio próximas las unas de las otras.

Para establecer la proximidad o asociación entre las categorías de una variable (que pueden estar en filas o en columnas) se considera su comparación a partir de sus perfiles, ya sean perfiles de fila −puntos fila− o perfiles de columna −puntos columna−. Cada perfil no es más que la distribución de frecuencias relativas o porcentajes, es decir, las distribuciones condicionales por por fila o por columna. La comparación entre dos perfiles se define a partir de una métrica, de una medida de distancia que evalúa sus diferencias. Esta medida en el ACS es la distancia de chi-cuadrado (χ^2). Si dos perfiles o distribuciones porcentuales son idénticos la distancia entre ellos será cero, a medida que las distribuciones difieran entre sí se pondrán de manifiesto variaciones entre ellas, y también con respecto a un comportamiento promedio del conjunto de los datos, así como los patrones de asociación.

A partir de aquí se procede a aplicar una serie de operaciones algebraicas propias de esta técnica de análisis factorial y se generan unos resultados (numéricos y gráficos) que nos permiten describir e interpretar las relaciones entre las variables.

Podemos utilizar el ACS para analizar tablas de contingencia y obtener una representación gráfica que ilustrará la relación entre las variables, en este sentido la podemos considerar como una técnica complementaria del análisis de la distribución de los datos de una la tabla de contingencia y de la interpretación de los estadísticos de asociación. Esta función de complemento es de cierto de interés cuando la tabla de contingencia contiene numerosas casillas, sino la ganancia que reporta la aplicación de la técnica no es muy relevante. Sin embargo, un ejemplo sencillo de una tabla de contingencia nos ayudará a entender el funcionamiento de la ACS.

Así, el uso de la ACS es especialmente interesante cuando el número de categorías de las variables es grande, y en este caso la representación será de gran ayuda para revelar la estructura de dependencia entre las variables. Posteriormente veremos toda la potencialidad de la técnica cuando tratamos el análisis multivariable de variables cualitativas, en la formulación del análisis de correspondencias múltiples, que de hecho implicará el tratamiento simultáneo de múltiples tablas de contingencia.

Pero, como hemos comentado, el ACS se puede utilizar para analizar cualquier tipo de tabla de doble entrada de números positivos, ampliando así sus posibilidades de utilización. Así podemos aplicar el análisis de correspondencias simples con los siguientes tipos de datos:

- Tablas de contingencia con el cruce de dos variables, por ejemplo, relacionando el conocimiento de la lengua y la edad agrupada en intervalos.
- Tablas que representan casos por variables. Pueden ser casos que representan agregados de unidades como los municipios, las comarcas, los países, etc. y relacionarlas con variables como el total o el porcentaje de ocupados por sectores de actividad, o el producto interior bruto de esos mismos sectores o la proporción de niveles educativos, por ejemplo. También es posible cruzar dos variables inicialmente para configurar las filas o las columnas, por ejemplo, podemos analizar el porcentaje de personas de una región, diferenciando el porcentaje de varones del de mujeres, según el nivel educativo alcanzado.
- Tablas que son matrices de distancias que expresan las proximidades en los objetos o categorías de ambas variables. En particular si procesamos una matriz de distancias por carretera entre los municipios de un territorio podemos obtener una representación gráfica que se aproximará en buena medida al mapa geográfico del territorio en cuestión.
- Tablas cuadradas que son la expresión de matrices de transición o tablas de movilidad entre origen y destino (la ocupación de padres e hijos/as), de la relación entre dos momentos en el tiempo de la misma variable (la ocupación en el primer trabajo y actualmente), o entre dos tipos de unidades con la misma información (la situación de actividad entre los miembros de una pareja).

Presentaremos algunos ejemplos de aplicación como los que hemos sugerido para ilustrar el uso de la técnica. Pero antes presentaremos los aspectos más técnicos de la formulación del ACS.

5.1.1. *Formulación del ACS*

A partir de una matriz o tabla de dole entrada de $I \times J$ casillas, el ACS busca representar las relaciones entre las distintas categorías a partir de un número reducido de dimensiones. Como resultado de la aplicación de esta técnica, el número máximo de dimensiones que se obtiene es igual al $mín(I,J)-1$, y se trata de poder sintetizar estas relaciones en un número menor (una, dos o tres dimensiones habitualmente) que expresan las fuentes de la asociación.

Los estadísticos que se derivan del análisis permiten determinar hasta qué punto las categorías de fila y de columna están representadas en un espacio de menor dimensión. Si esto es así, se trata de examinar qué categorías en fila son similares y cuáles no, lo mismo para las categorías en columna y, por último, qué categorías en fila están asociadas con qué categorías en columna. Es decir, se trata de ver las correspondencias entre filas y columnas, o correspondencias a partir de cualquier tabla de números positivos dispuestos en filas y columnas (tabla de correspondencias). Cuando existe asociación entre las variables, el ACS revela la naturaleza de la relación, especialmente cuando se dispone de un número elevado de categorías.

A partir de una tabla $N(I, J)$, constituida por I líneas, indexadas por i, con $i=1...I$, y por J columnas, indexadas por j, con $j=1...J$, que cruza dos variables cualitativas, Y y X, consideraremos la notación siguiente. El conjunto de datos de la tabla con las frecuencias absolutas mediante:

N(I,J): Tabla de frecuencias absolutas

Y \ X	1	2	…	j	…	J	Total
1	n_{11}	n_{12}	…	n_{1j}	…	n_{1J}	n_{1+}
2	n_{21}	n_{22}	…	n_{2j}	…	n_{2J}	n_{2+}
⋮	⋮	⋮		⋮		⋮	⋮
i	n_{i1}	n_{i2}	…	n_{ij}	…	n_{iJ}	n_{i+}
⋮	⋮	⋮		⋮		⋮	⋮
I	n_{I1}	n_{I2}	…	n_{Ij}	…	n_{IJ}	n_{I+}
Total	n_{+1}	n_{+2}	…	n_{+j}	…	n_{+J}	n_{++}

Donde n_{ij} es la frecuencia absoluta de cada casilla que surge de la distribución conjunta, siendo los totales marginales de fila y columna n_{i+} y n_{+j}, y n o n_{++} el total de casos.

Las frecuencias relativas f_{ij} expresadas como proporciones sobre el total de casos (o porcentajes si se multiplican por 100) mediante:

F(I,J): Tabla de frecuencias relativas

Y \ X	1	2	…	j	…	J	Total
1	f_{11}	f_{12}	…	f_{1j}	…	f_{1J}	f_{1+}
2	f_{21}	f_{22}	…	f_{2j}	…	f_{2J}	f_{2+}
⋮	⋮	⋮		⋮		⋮	⋮
i	f_{i1}	f_{i2}	…	$\boldsymbol{f_{ij}}$	…	f_{iJ}	$\boldsymbol{f_{i+}}$
⋮	⋮	⋮		⋮		⋮	⋮
I	f_{I1}	f_{I2}	…	f_{Ij}	…	f_{IJ}	f_{I+}
Total	f_{+1}	f_{+2}	…	$\boldsymbol{f_{+j}}$	…	f_{+J}	$\boldsymbol{f_{++}}$

Donde la frecuencia relativa genérica f_{ij} se determina por: $f_{ij} = \frac{n_{ij}}{n}$, y donde se cumple que: $\sum_{i=1}^{I}\sum_{j=1}^{J} f_{ij} = 1$.

La frecuencia relativa marginal de fila es $f_{i+} = \frac{n_{i+}}{n}$, siendo igualmente $f_{i+} = \sum_{j=1}^{J} f_{ij}$, valor que expresa el peso o la masa del punto-fila i, verificándose que $\sum_{i=1}^{I} f_{i+} = 1$.

La frecuencia relativa marginal de columna es $f_{+j} = \frac{n_{+j}}{n}$, siendo igualmente $f_{+j} = \sum_{i=1}^{I} f_{ij}$ $f_{i+} = \sum_{j=1}^{J} f_{ij}$, valor que expresa el peso o la masa del punto-columna j, verificándose que $\sum_{j=1}^{J} f_{+j} = 1$.

Estos totales de las tablas se corresponden con el perfil medio respecto del cual se posicionan las distintas categorías de las filas o columnas.

Asimismo podemos definir las frecuencias relativas que se obtienen de las distribuciones condicionales por fila o columna, que llamaremos perfiles, mediante: $f_{ij}^F = \frac{n_{ij}}{n_{i+}}$ es el perfil de fila, igual a $f_{ij}^F = \frac{f_{ij}}{f_{i+}}$ y $\sum_{j=1}^{J} f_{ij}^F = 1$ dando lugar a la tabla F^F.
$f_{ij}^C = \frac{n_{ij}}{n_{+j}}$ es el perfil de columna, igual a $f_{ij}^C = \frac{f_{ij}}{f_{+j}}$ y $\sum_{i=1}^{I} f_{ij}^C = 1$ dando lugar a la tabla F^C.

$$F^F = \begin{pmatrix} f_{11}/f_{1+} & f_{12}/f_{1+} & \cdots & f_{1J}/f_{1+} \\ f_{21}/f_{2+} & f_{22}/f_{2+} & \cdots & f_{2J}/f_{2+} \\ \vdots & \vdots & \ddots & \vdots \\ f_{I1}/f_{I+} & f_{I2}/f_{I+} & \cdots & f_{IJ}/f_{I+} \end{pmatrix} \qquad F^C = \begin{pmatrix} f_{11}/f_{+1} & f_{12}/f_{+2} & \cdots & f_{1J}/f_{+J} \\ f_{21}/f_{+1} & f_{22}/f_{+2} & \cdots & f_{2J}/f_{+J} \\ \vdots & \vdots & \ddots & \vdots \\ f_{I1}/f_{+1} & f_{I2}/f_{+2} & \cdots & f_{IJ}/f_{+J} \end{pmatrix}$$

Tabla de perfiles de fila *Tabla de perfiles de columna*

Si consideramos como ejemplo la tabla de contingencia que relaciona el nivel de ingresos (IngresosH)[19] con el nivel ocupacional dominante del hogar (OCUPAFAM)[20], según los datos del estudio 3041 del CIS, las diferentes distribuciones comentadas se recogen en la Tabla III.11.10.

Tabla III.11.10. Tabla de correspondencias entre la ocupación dominante del hogar y el nivel de ingresos: frecuencias absolutas y relativas, perfiles fila y columna

(a) Tabla de frecuencias absolutas

Frecuencias absolutas		OCUPAFAM Ocupación dominante de los padres y las madres									Total
		Directores y gerentes	Técnicos y profesionales	Profesionales de apoyo	Adminis-trativos	Trabajador de servicios	Trabajadores cualificados primario	Trabajadores cualificados industria	Operadores y montadores	Ocupaciones elementales	
IngresosH Ingresos del hogar	Bajos	5	11	25	5	71	127	113	50	67	474
	Medios	13	31	61	22	111	118	154	118	59	687
	Altos	25	61	66	9	74	48	71	57	18	429
Total		43	103	152	36	256	293	338	225	144	1590

(b) Tabla de frecuencias relativas

Frecuencias relativas		OCUPAFAM Ocupación dominante de los padres y las madres									Total
		Directores y gerentes	Técnicos y profesionales	Profesionales de apoyo	Adminis-trativos	Trabajador de servicios	Trabajadores cualificados primario	Trabajadores cualificados industria	Operadores y montadores	Ocupaciones elementales	
IngresosH Ingresos del hogar	Bajos	0,003	0,007	0,016	0,003	0,045	0,080	0,071	0,031	0,042	0,398
	Medios	0,008	0,019	0,038	0,014	0,070	0,074	0,097	0,074	0,037	0,294
	Altos	0,016	0,038	0,042	0,006	0,047	0,030	0,045	0,036	0,011	0,270
Total		0,027	0,065	0,096	0,023	0,161	0,184	0,213	0,142	0,091	1,000

(c) Tabla de perfiles de fila

Perfiles fila		OCUPAFAM Ocupación dominante de los padres y las madres									
		Directores y gerentes	Técnicos y profesionales	Profesionales de apoyo	Adminis-trativos	Trabajador de servicios	Trabajadores cualificados primario	Trabajadores cualificados industria	Operadores y montadores	Ocupaciones elementales	Total
IngresosH Ingresos del hogar	Bajos	0,011	0,023	0,053	0,011	0,150	0,268	0,238	0,105	0,141	1,000
	Medios	0,019	0,045	0,089	0,032	0,162	0,172	0,224	0,172	0,086	1,000
	Altos	0,058	0,142	0,154	0,021	0,172	0,112	0,166	0,133	0,042	1,000
Total		0,027	0,065	0,096	0,023	0,161	0,184	0,213	0,142	0,091	1,000

(d) Tabla de perfiles de columna

Perfiles columna		OCUPAFAM Ocupación dominante de los padres y las madres									Total
		Directores y gerentes	Técnicos y profesionales	Profesionales de apoyo	Adminis-trativos	Trabajador de servicios	Trabajadores cualificados primario	Trabajadores cualificados industria	Operadores y montadores	Ocupaciones elementales	
IngresosH Ingresos del hogar	Bajos	0,116	0,107	0,164	0,139	0,277	0,433	0,334	0,222	0,465	0,298
	Medios	0,302	0,301	0,401	0,611	0,434	0,403	0,456	0,524	0,410	0,432
	Altos	0,581	0,592	0,434	0,250	0,289	0,164	0,210	0,253	0,125	0,270
	Total	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Fuente: *Centro de Investigaciones Sociológicas, Estudio 3041 de 2014*.

El análisis de la relación entre las variables se realiza comparando los perfiles. Si realizamos la comparación por columna se trata de mirar en qué medida los perfiles de las diferentes columnas difieren entre sí y, por tanto, con respecto al perfil medio o

[19] Los valores de esta variable agrupan los de la variable original P45 con el criterios siguientes: bajos (hasta 900€), medios (entre 901€ y 1.800€) y altos (más de 1.800€). Hemos tomado tres valores con un objetivo didáctico y así poder representar los datos originales en un espacio tridimensional como veremos a continuación.

[20] Es la variable generada en el capítulo III.2 tomando el nivel ocupacional más alto entre el padre y la madre.

marginal de columna. Así, para cada columna j, su perfil f_{ij}/f_{+j} se relaciona con el del total f_{i+}.

Por ejemplo, entre los directores y gerentes el 58% tiene ingresos altos (superiores a 1.800€ mensuales):

$$f_{13}/f_{+3} = 0,016/0,270 = n_{13}/n_{+3} = 25/429 = 0,058$$

mientras que para el conjunto de la muestra este valor es del 27%, $f_{3+} = 0,270$. Son valores que difieren entre sí, el comportamiento de un perfil se aleja del comportamiento medio, del marginal, mostrando una correspondencia, una asociación, entre ser director o gerente y tener altos ingresos.
Si la condición de independencia perfecta se cumple esto implica que los dos valores, el de la columna y el marginal, serían los mismos. En general, para cada tipo de distribución, la condición de independencia se expresa así[21]: comparando perfiles fila: $\frac{f_{ij}}{f_{i+}} = f_{+j}$, comparando perfiles columna: $\frac{f_{ij}}{f_{+j}} = f_{i+}$, o comparando frecuencias relativas totales: $f_{ij} = f_{i+} \cdot f_{+j}$.

La información de los perfiles se puede representar gráficamente. De hecho, cada perfil se puede considerar como un punto y un vector en el espacio. En particular, los perfiles se pueden representar geométricamente ya que las coordenadas del perfil (ya sea de filas o de columnas) definen un punto en el hiperplano del espacio definido por las columnas (para los puntos-fila) o de filas (para los puntos-columna), de ecuación $\sum_{j=1}^{J} \frac{n_{ij}}{n_{i+}} = 1$ para las filas y $\sum_{i=1}^{I} \frac{n_{ij}}{n_{+j}} = 1$ para las columnas. Ello es así dado que la suma de los elementos de cada perfil es 1 (o 100 si el expresáramos en tanto por ciento) en el sistema de coordenadas definido respectivamente por las $j_1, j_2,...,j_J$ o las $i_1, i_2, ..., i_I$. Es decir, cualquier perfil se puede considerar como un punto (o vector) que se representa en un diagrama de dispersión configurando una nube de puntos. Lo veremos a continuación con un ejemplo. Planteamos primero de forma general las características de esta representación, dualmente, para las filas y para las columnas.

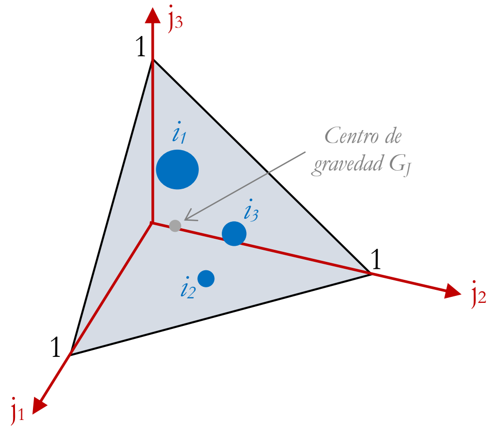
A partir de la tabla de frecuencias relativas $F(I, J)$ tenemos dos nubes de puntos:

- La nube de puntos $\mathbb{N}^I$ formada por los puntos-fila (o categorías de la variable dispuesta en filas), en el espacio vectorial $\mathbb{R}^J$ definido por las variables-columna (o categorías de la variable dispuesta en columna), con coordenadas $f_{ij}^F = \frac{f_{ij}}{f_{i+}}$, cada uno de los que está afectado de un peso o masa igual a f_{i+}. y que gráficamente se expresa con círculos de diferente tamaño (0). Estos I puntos se sitúan en un subespacio de J-1 dimensiones, lugar de puntos cuya suma de coordenadas es igual a 1, ya que sus J coordenadas verifican la relación: $\sum_{j=1}^{J} \frac{f_{ij}}{f_{i+}} = 1$. Las proximidades entre puntos se corresponden con las similitudes entre los perfiles. El centro de la nube de puntos se corresponde con el perfil medio, es decir, con el total o marginal

[21] El cociente de contingencia, la proporción condicional (perfil fila o perfil columna) dividida la proporción total, da en ambos casos los mismo valores, lo que significa equivalencia y simetría del análisis.

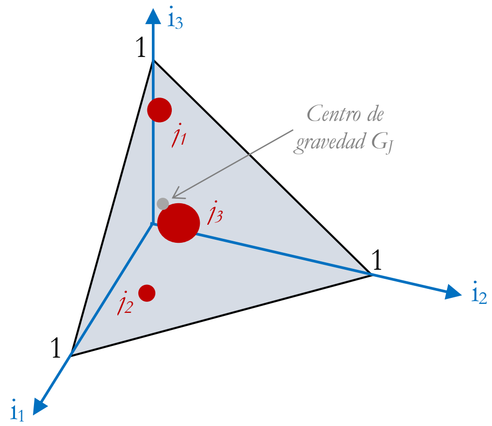
de las filas de la tabla, y se denomina centro de gravedad, G_J en este caso: $G_J = \sum_{i=1}^{I} f_{i+} \cdot f_{ij}^F = \sum_{i=1}^{I} f_{ij} = f_{+j}$.

Nube $\mathbb{N}^I$ de puntos-fila en el espacio vectorial $\mathbb{R}^J$ de las columnas



– Dualmente, en el espacio $\mathbb{R}^I$, tenemos la nube $\mathbb{N}^J$ de los *j* puntos-columna con coordenadas $f_{ij}^C = \dfrac{f_{ij}}{f_{+j}}$ y afectados cada uno por la masa o peso f_{+j}. Estos puntos se sitúan en un espacio de *I-1* dimensiones (Gráfico III.11.15). El centro de la nube de puntos es el perfil medio o centro de gravedad G_I : $G_I = \sum_{j=1}^{J} f_{+j} \cdot f_{ij}^C = \sum_{j=1}^{J} f_{ij} = f_{i+}$.

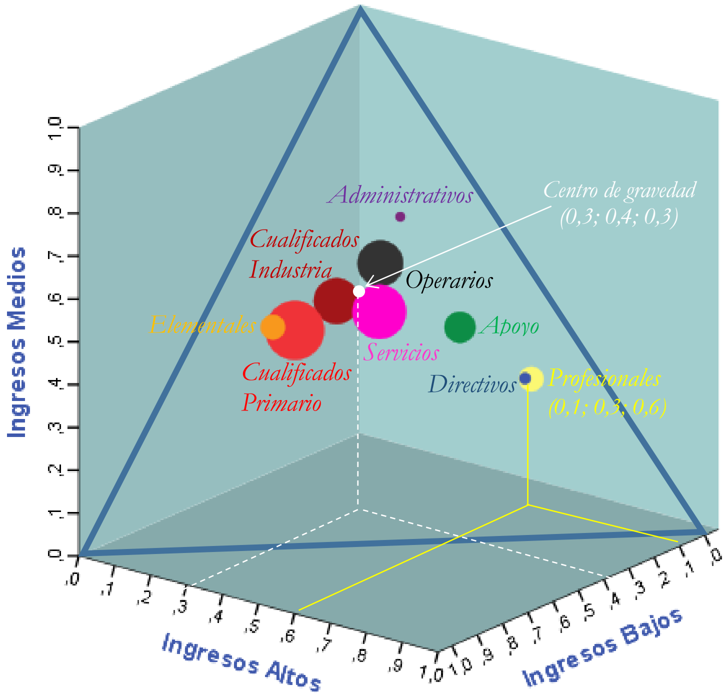
Gráfico III.11.15. Nube $\mathbb{N}^J$ de puntos-columna en el espacio vectorial $\mathbb{R}^I$ de las filas



En un espacio de tres dimensiones, todos los perfiles se sitúan en un triángulo equilátero formado por tres vértices a partir de tres puntos unidad: (1,0,0), (0,1,0) y (0,0,1). Por tanto, en el espacio tridimensional todos los puntos se encuentran en el interior de un triángulo que se representa en un plano bidimensional.

Veamos este resultado gráfico con el ejemplo anterior de la Tabla III.11.10. En particular, si representamos los puntos-columna de las 9 categorías de ocupación en el espacio de 3 dimensiones definido por los tres niveles de ingresos se obtiene el Gráfico III.11.16.

Gráfico III.11.16. Nube $\mathbb{N}^9$ de puntos-columna (ocupación dominante del hogar) en el espacio vectorial $\mathbb{R}^3$ de las filas (nivel de ingresos del hogar)



Las coordenadas de cada punto vienen dadas por los datos que se obtienen del perfil de columnas de la subtabla (d) de la Tabla III.11.10. En el caso por ejemplo de los profesionales éstos se distribuyen entre los que tienen bajos ingresos (0,1), medios (0,3) y altos (0,6). Otro punto de referencia del gráfico es centro de gravedad que corresponde a la media o marginal de los ingresos: bajos (0,3), medios (0,4) y altos (0,3).

La distribución de las distintas ocupaciones sigue una disposición donde podemos observar que los ingresos más altos se corresponden con las ocupaciones más altas de profesionales y directores, en la zona de la derecha, e inversamente, cómo las categorías ocupacionales inferiores, en contraposición, se sitúan en la zona de la izquierda. También se dibuja una zona intermedia, entre la derecha y la izquierda, de ocupaciones e ingresos medios. Se configuran así las correspondencias entres ocupaciones e ingresos a lo largo de su disposición gráfica de derecha-izquierda para expresar la existencia de una dimensión o factor de asociación principal que los vincula: a más nivel ocupacional más ingresos.

Además podemos observar cómo algunas categorías ocupacionales del centro del gráfico además tienden a desplazarse hacia el norte, circunstancia que apunta hacia la existencia de una segunda dimensión que relaciona los ingresos medios con algunas ocupaciones como destacaremos más adelante. Notemos por último que los puntos-columna representados adoptan tamaños distintos en función de su peso o masa, es decir, según sean más o menos frecuentes. Estos resultados los veremos expresados en la información que obtendremos de la aplicación del análisis de correspondencias a esta tabla. Veremos asimismo que es posible representar simultáneamente los puntos-columna y los puntos fila, lo que nos permitirá ver más claramente las correspondencias en las categorías de las filas y las columnas.

La proximidad (semejanza) o la desviación (diferencia) entre dos perfiles se define a partir de una métrica, de una medida de distancia. Mientras que el análisis factorial de componentes principales esta medida es la distancia euclídea, aquí consideramos la distancia de chi-cuadrado (χ^2). Así, la distancia entre dos puntos-fila i, i' en el espacio $\mathbb{R}^J$ será:

$$d^2(i,i') = \sum_{j=1}^{J} \frac{1}{f_{+j}} \cdot \left(\frac{f_{ij}}{f_{i+}} - \frac{f_{i'j}}{f_{i'+}} \right)^2 \qquad \textit{Ecuación 13}$$

De manera equivalente se define la distancia entre dos puntos-columna j, j' en el espacio $\mathbb{R}^I$:

$$d^2(j,j') = \sum_{i=1}^{I} \frac{1}{f_{i+}} \cdot \left(\frac{f_{ij}}{f_{+j}} - \frac{f_{ij'}}{f_{+j'}} \right)^2 \qquad \textit{Ecuación 14}$$

La distancia así definida depende de las diferencias de las frecuencias tomadas dos a dos, para cada i o j, elevadas al cuadrado y ponderadas por la frecuencia marginal de filas o columnas, equilibrando la importancia relativa de cada categoría-variable columna o fila.

La métrica χ^2 así definida verifica una importante propiedad matemática: la de la equivalencia distribucional (Teorema de Huygens). Es decir, si dos puntos-fila se confunden o superponen en $\mathbb{R}^J$, y si se les considera como un solo punto afectado por la suma de sus masas respectivas, entonces, las distancias entre parejas de puntos, no sólo en $\mathbb{R}^J$, sino también en $\mathbb{R}^I$, permanecen inalteradas. Que la distancia entre filas o columnas no se altere si se agrupan las columnas o las filas de perfil similar es una característica de interés cuando se consideran criterios de (re)codificación de los valores de las variables, favoreciendo así la robustez y la estabilidad de los resultados, pues no se pierde información uniendo categorías homogéneas ni se gana desacomponiéndolas en subgrupos. Propiedad que no posee la distancia euclídea.

Otra consecuencia de esta propiedad es que siempre que el peso de una categoría no sea excesivamente elevado, si se descompone una categoría en dos, la representación gráfica de las dos nuevas categorías tendrá como centro de gravedad aproximado el centro de gravedad de las categorías iniciales.

Esta distancia tiene un gran parecido con una suma de cuadrados de la distancia euclídea, pero no lo es estrictamente, y, por tanto, no permite aplicar los mismos cálculos que para un análisis de factorial de componentes[22]. No obstante, es posible aplicar una transformación que modifique la escala de los ejes teniendo en cuenta el peso de cada categoría.

[22] La distancia euclídea entre las filas i e i' obtenida sobre $N(I, J)$ sólo reflejaría la diferencia entre las frecuencias marginales n_{i+} y $n_{i'+}$, y se manifestaría en un efecto de tamaño. La distancia euclídea entre los perfiles F^F o F^C refleja su semejanza.

Mediante el ACS se trata de describir las proximidades entre perfiles de filas (o perfiles de columnas), teniendo en cuenta la diferencia de efectivos entre estas filas (o columnas). El ACS trata de representar cada nube de puntos en un subespacio de dimensión igual a $mín(I,J)-1$. Para cada nube de puntos pueden determinarse los ejes principales que indican las direcciones de dicha nube, donde se acumula mayor inercia o dispersión. El proceso de cálculo del ACS implica dos partes diferenciables: la obtención de la matriz de varianzas y covarianzas a partir de la definición de la distancia de Benzécri, y la extracción de los factores y proyección de cada categoría en los factores que genera las representaciones gráficas. La medida del valor de dispersión a lo largo de un eje es el valor propio, la suma de todos ellos es la traza, idéntico valor para ambos nubes, por lo que es posible representar en un mismo espacio.

El análisis factorial tiene por objetivo encontrar el subespacio vectorial de dimensión reducida que hace pasar, por transformación geométrica, del conjunto de puntos original a las proyecciones en el nuevo espacio de forma que se exprese el máximo de información (significación) con un número menor de dimensiones. El análisis consiste pues, como el análisis factorial de componentes principales, en encontrar los ejes principales de inercia a partir de la diagonalización de la matriz de varianzas y covarianzas, con la ventaja adicional de poder encontrar relaciones de tipo no- lineal.

El Gráfico III.11.17 esquematiza el proceso de un análisis de correspondencias. Partiendo de la tabla de correspondencias inicial se trata analizar de forma simétrica dos nubes de puntos en dos espacios.

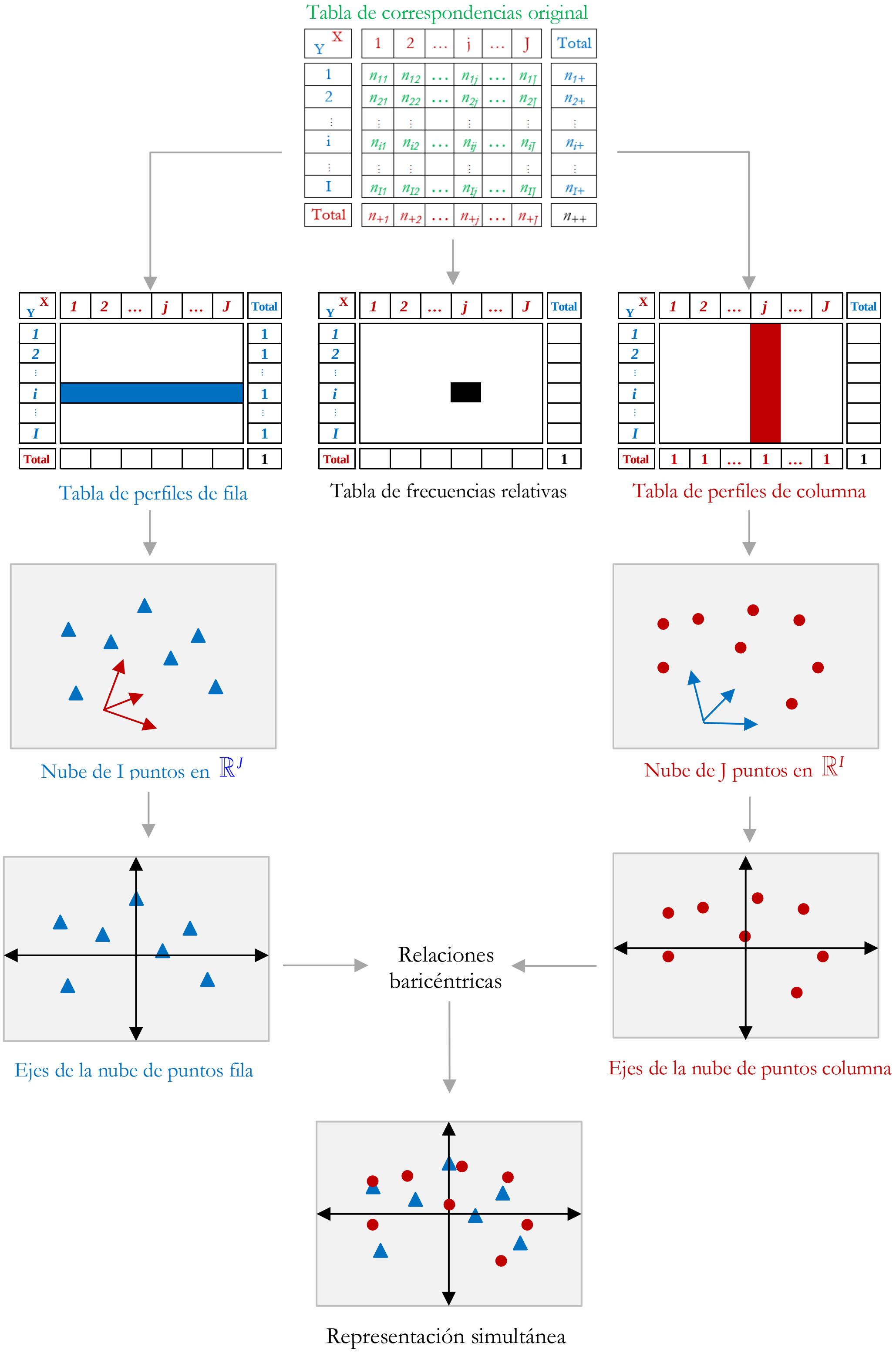
A partir del centro de gravedad de estos espacios podemos determinar la varianza o inercia total I_{G_J} en el espacio $\mathbb{R}^J$ a partir de la suma de las distancias de cada punto respecto al centro de masas o de gravedad (Gráfico III.11.18):

$$I_{G_J} = \sum_{i=1}^{I} f_{i+} \cdot d^2(i, G_J) = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{\left(f_{ij} - f_{i+} \cdot f_{+j}\right)^2}{f_{i+} \cdot f_{+j}} \qquad \textit{Ecuación 15}$$

Expresión de la inercia de la nube de puntos $\mathbb{N}^I$ respecto al centro G_j que coincide con la distancia de chi-cuadrado χ^2, salvo un coeficiente n ya que se consideran frecuencias relativas: $\chi^2 = n \cdot I_{G_J}$. Este valor es idéntico al que se obtendría de la inercia de la nube de puntos $\mathbb{N}^J$ respecto de G_i, por lo que:

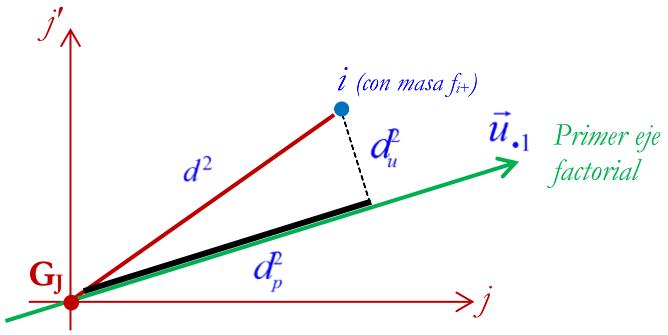
$$\chi^2 = n \cdot I_{G_J} = n \cdot I_{G_I} \qquad \textit{Ecuación 16}$$

Gráfico III.11.17. Esquema del ACS. Transformación de la tabla de correspondencias



A partir de la matriz de varianzas y covarianzas que se obtiene se trata de diagonalizarla generando los vectores propios y los valores propios. Esta operación algebraica implica que la inercia explicada por los ejes sea máxima, es decir, que la proyección de todos los puntos i sobre el eje, la distancia d_p^2, es máxima, y mínima la distancia d_u^2 de cada punto al vector o eje factorial que se obtiene (Gráfico III.11.18).

Gráfico III.11.18. Descomposición de la inercia



El proceso de obtención de los ejes factoriales es equivalente si consideramos los puntos-columna en el espacio de las filas. En ambos casos, como resultado, el primer eje factorial es el vector propio correspondiente al primer y mayor valor propio, λ_1, siendo por tanto el factor que más inercia explica. El segundo y sucesivos ejes son los que acumulan el resto de la inercia de forma ordenada y decreciente en importancia. Por tanto, la varianza explicada por un factor es λ_k / I_G, que mide la parte de la inercia total explicada por cada eje factorial, siendo la inercia total explicada la suma de todos los valores propios: $I_G = \sum_{k=1}^{K} \lambda_k$. Estos valores son todos positivos e inferiores o iguales a la unidad.

Como en ambos espacios las inercias asociadas a los ejes son las mismas se generan los mismos valores propios lo que permitirá la representación simultánea de las proyecciones de los dos nubes de puntos. Los factores que se generan para las filas y las columnas del mismo rango se relacionan entre sí a través de las fórmulas de transición o relaciones baricèntriques, como destacamos anteriormente. Así la proyección de un punto-fila en $\mathbb{R}^J$ o un punto-columna en $\mathbb{R}^I$, salvo un coeficiente $1/\sqrt{\lambda_k}$, está en el baricentro de los puntos-columna o puntos-fila sobre este eje, y cada punto afectado por los pesos que expresan su importancia relativa. Este hecho posibilita la representación simultánea y la interpretación de las relaciones entre filas y columnas.

5.1.2. *Interpretación del ACS*

Nos centraremos en este apartado en todos los elementos interpretativos de la aplicación de un análisis de correspondencias simples a través de diversos indicadores que aparecen en tablas y la representación de los gráficos factoriales. Tomaremos el ejemplo de la tabla de contingencia que relaciona la categoría ocupacional dominante del hogar y el nivel de ingresos y presentaremos los resultados que genera el software

SPAD. Junto a este ejemplo interpretaremos otros que nos ayudaran a interpretar este tipo de resultados.

El proceso de interpretación implica fundamentalmente dos aspectos: determinar el número de factores que sintetizan la mayor parte de la inercia explicada y dar identidad conceptual a los mismos como expresión de las fuentes de asociación entre las categoría de filas y columnas de la tabla de correspondencias.

Antes de proceder al análisis de los resultados de esta técnica con nuestro primer ejemplo conviene tener en cuenta que la tabla de contingencia debe cumplir con los requisitos o condiciones de aplicación que se exigen para la prueba estadística de independencia de chi-cuadrado que vimos en el capítulo III.6. Es decir, que la frecuencia mínima esperada en cada casilla sea 1 y 5 en el 80%. En la Tabla III.11.11 se presentan los datos de la tabla de contingencia donde se señala que la frecuencia mínima esperada es de 9,71. La relación entre ambas variables arroja un valor de chi-cuadrado de 180,568, significativo, con una intensidad de asociación de 0,238 según el estadístico de la V de Cramer.

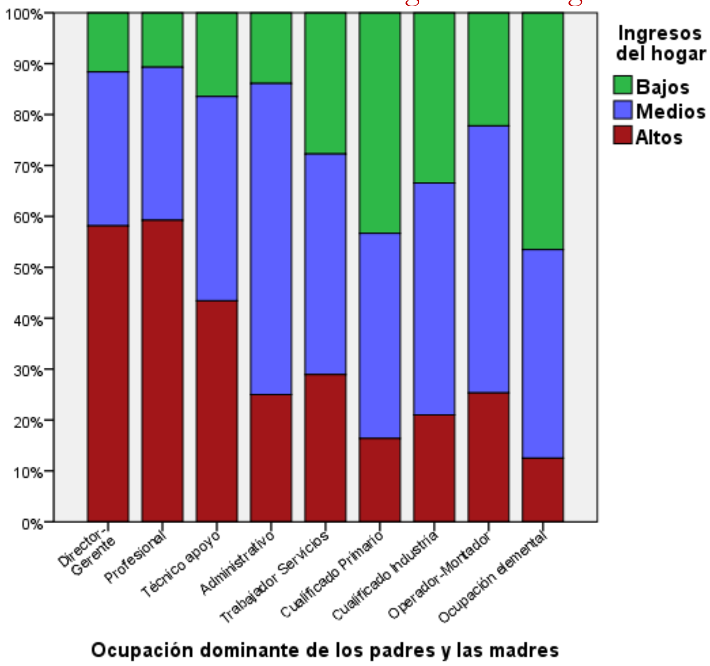
La relación de asociación que se observa muestra un patrón donde los mayores niveles ocupacionales (directores y gerentes, técnicos y profesionales) alcanzan mayores niveles de ingresos, mientras que los hogares con bajos niveles ocupacionales (trabajadores cualificados y no cualificados) destacan por la mayor proporción de ingresos bajos. Así se muestra en la Tabla III.11.11 y en el Gráfico III.11.19.

Tabla III.11.11. Tabla de contingencia entre la ocupación dominante del hogar y el nivel de ingresos

			IngresosH Ingresos del hogar			Total
			1 Bajos	2 Medios	3 Altos	
OCUPAFAM Ocupación dominante de los padres y las madres	1 Director-Gerente	Frecuencia	5	13	25	43
		% por fila	11,6%	30,2%	58,1%	100,0%
	2 Profesional	Frecuencia	11	31	61	103
		% por fila	10,7%	30,1%	59,2%	100,0%
	3 Técnico apoyo	Frecuencia	25	61	66	152
		% por fila	16,4%	40,1%	43,4%	100,0%
	4 Administrativo	Frecuencia	5	22	9	36
		% por fila	13,9%	61,1%	25,0%	100,0%
	5 Trabajador de Servicios	Recuento	71	111	74	256
		Frecuencia	27,7%	43,4%	28,9%	100,0%
	6 Cualificado Sector Primario	% por fila	127	118	48	293
		Frecuencia	43,3%	40,3%	16,4%	100,0%
	7 Cualificado Sector Industrial	% por fila	113	154	71	338
		Frecuencia	33,4%	45,6%	21,0%	100,0%
	8 Operador-Montador	% por fila	50	118	57	225
		Frecuencia	22,2%	52,4%	25,3%	100,0%
	9 Ocupación elemental	Frecuencia	67	59	18	144
		% por fila	46,5%	41,0%	12,5%	100,0%
	Total	Frecuencia	474	687	429	1.590
		% por fila	**29,8%**	**43,2%**	**27,0%**	**100,0%**

Fuente: *Centro de Investigaciones Sociológicas, Estudio 3041 de 2014*.
Los datos no incluyen la categoría “Sin datos”. χ^2 =180,568 (0,000), V de Cramer = 0,238. Frecuencia mínima esperada 9,71.

Gráfico III.11.19. Ocupación dominante del hogar según el nivel de ingresos del hogar



No obstante, esta tendencia mayoritaria se rompe puntualmente y de forma moderada en el orden de categorías ocupacionales original de la variable. Vemos en concreto como los trabajadores cualificados de la industria y sobre todo los operadores y montadores tienen menores niveles de ingresos bajos y, en consecuencia, mayores niveles medios e incluso altos. Por último las categorías intermedias de administrativos y trabajadores de servicios registran los mayores niveles de ingresos medios, sobre todo los primeros.

Veamos a continuación cómo se reflejan estas conclusiones de la lectura de la tabla de contingencia en términos de un análisis de correspondencias.

El primer resultado se refiere al número de dimensiones o factores en ACS. Como vimos el total de factores posibles es el mínimo de filas y columnas menos 1. En este caso tenemos que el valor mínimo entre 3 niveles de ingresos y 9 ocupaciones es 3, menos 1, da como resultado 2 factores. En la Tabla III.11.12 aparecen los dos valores propios que se obtienen del ACS de la tabla de contingencia que se interpretan como la inercia o varianza acumulada por cada eje factorial.

Tabla III.11.12. Tabla de valores propios y de inercia (varianza) explicada del ACS

	Valores propios		
Factor	Valor	% de la varianza	% acumulado
1	0,0993	87,40	87,40
2	0,0143	12,60	100,00
Total	0,1136	100,00	

Una primera conclusión es pues que el primer factor acumula la mayor parte de la inercia, el 87%, por lo que nos expresa la fuente o patrón principal de asociación, prácticamente único, entre las variables. Se suele considerar como unos de los criterios para decidir el número de factores a considerar que éstos alcancen un porcentaje suficiente del total, un valor de referencia que se suele contemplar es llegar por lo menos el 70% de la varianza explicada. Vemos como en nuestro caso excedemos sobradamente este valor con un solo factor. De forma muy secundaria, con el 13%, se sitúa el segundo factor que expresa una segunda fuente de variabilidad pero que hay que valorar como poco relevante. La suma de los valores propios representa la inercia total de la nube y es de 0,1136 en esta tabla. Surge de dividir el estadístico chi-cuadrado (en esta tabla un valor de 180,56), que mide la diferencia entre los efectivos observados y los efectivos teóricos o esperados que se obtendrían en promedio si las dos variables fueran independientes, entre el número de casos (1590).

Para dar contenido a estos factores necesitamos leer la información de las tablas que se obtienen (Tabla III.11.13)[23]. Esta información se divide para cada variable o conjunto de categorías dispuestas en fila y en columna. En cada una de ellas aparece en primer lugar el peso relativo (la frecuencia relativa de cada categoría sobre el total de casos válidos, en porcentaje) que nos valora la importancia de cada categoría y la distancia (la distancia cuadrática euclidiana de cada categoría al origen, al 0 o valor promedio) que nos indica qué tan dispersa o periférica es la categoría, cuanto mayor sea este valor más inercia acumulará y más determinante será en la configuración del carácter de los factores.

Las coordenadas nos permiten representar gráficamente las categorías, los puntos-fila y los puntos-columna, en el espacio factorial indicando la posición que ocupan para cada eje retenido.

Las contribuciones absolutas constituyen la información más relevante para dar cuenta del contenido de los factores obtenidos. Es la contribución en porcentaje (cada columna suma 100%, $\sum_{i=1}^{I} CTA_{ik} = 100$) de cada punto o categoría a la inercia total explicada por el eje factorial de rango k, y expresa la proporción de varianza explicada del factor k debida a una categoría. Esta cantidad es igual al cuadrado de la i-ésima coordenada del eje de rango k:

$$CTA_{ik} = \frac{f_{i+} \cdot y_{ik}^2}{\lambda_k} \qquad \textit{Ecuación 17}$$

Se trata de la contribución de la categoría i a la generación del eje k y se obtiene con el producto de la frecuencia relativa de la categoría por sus coordenadas sobre el eje k, dividido por el valor propio del eje k. La suma de las contribuciones de todas las modalidades activas sobre un eje es igual al 100% (suma en columna).

Si todas las contribuciones fueran iguales la contribución sería el 100% dividido por el número de categorías (en este caso 33% para los ingresos y 11% para las ocupaciones). Por lo tanto, las categorías o modalidades (de fila o columna) con contribuciones superiores a este valor son las que tienen una influencia superior a la media. En nuestro

[23] Son las tablas que genera el software SPAD. En él las contribuciones absolutas se etiquetan “contribuciones”, las contribuciones relativas se denominan “cosenos cuadrados”, la variable en fila se identifica como “casos” y la variable en columna “frecuencias activas”.

ejemplo observamos cómo el Eje 1 o primer factor viene marcado por los valores de ingresos altos (60,9) e ingresos bajos (38,0) de la variable en filas. Si nos fijamos en las coordenadas correspondientes a estos valores constatamos que se da una oposición entre ambas características pues los ingresos altos se sitúan en el lado positivo y los bajos en el negativo, pasando por un lugar intermedio próximo al cero de los niveles medios.

Tabla III.11.13. Coordenadas, contribuciones absolutas y relativas del ACS.
Variable en filas (IngresosH)

Categoría	Peso relativo	Distancia al origen	Coordenadas		Contribuciones absolutas		Contribuciones relativas	
			Eje 1	Eje 2	Eje 1	Eje 2	Eje 1	Eje 2
Ingresos bajos	29,81	0,14	-0,36	-0,12	38,0	32,2	0,89	0,11
Ingresos medios	43,21	0,02	-0,05	0,14	1,1	55,7	0,12	0,88
Ingresos altos	26,98	0,23	0,47	-0,08	60,9	12,1	0,97	0,03

Variable en columnas (OCUPAFAM)

Categoría	Peso relativo	Distancia al origen	Coordenadas		Contribuciones absolutas		Contribuciones relativas	
			Eje 1	Eje 2	Eje 1	Eje 2	Eje 1	Eje 2
Director-Gerente	2,70	0,510	0,69	-0,17	13,13	5,28	0,95	0,05
Profesional	6,48	0,548	0,72	-0,17	33,96	12,48	0,95	0,05
Técnico apoyo	9,56	0,162	0,40	-0,01	15,63	0,03	1,00	0,00
Administrativo	2,26	0,161	0,12	0,38	0,34	23,09	0,09	0,91
Trabajador Servicios	16,10	0,003	0,05	0,01	0,44	0,12	0,96	0,04
Cualificado Primario	18,43	0,105	-0,31	-0,10	17,54	13,65	0,90	0,10
Cualificado Industria	21,26	0,019	-0,13	0,03	3,87	1,26	0,96	0,04
Operador-Montador	14,15	0,040	0,05	0,19	0,30	37,51	0,05	0,95
Ocupación elemental	9,06	0,173	-0,40	-0,10	14,80	6,59	0,94	0,06

Si ahora miramos las categorías ocupacionales, de forma equivalente, podemos identificar las contribuciones absolutas más importantes y el lugar que ocupan en el primer eje. Constatamos cómo directores y gerentes, profesionales y técnicos de apoyo tienen las mayores contribuciones y se ubican en el lado positivo; los trabajadores cualificados del sector primario junto a las ocupaciones elementales son las categorías que sobresalen respecto del resto en el lado negativo. El resto de las categorías corresponden a contribuciones bajas con coordenadas próximas a cero en la primera dimensión. En conclusión, el primer factor, que representa el 87% de la varianza explicada, nos está expresando las correspondencias entre ingresos y ocupación, diferenciando las ocupaciones más altas a las que se asocian altos ingresos de las ocupaciones inferiores que tienen menores ingresos.

El segundo factor, con el 13% de la varianza explicada, introduce un matiz muy secundario en los patrones de asociación para diferenciar y destacar la correspondencia entre los ingresos medios y las categorías de administrativos y operadores-montadores (lado positivo) frente a las categorías extremas tanto bajas como altas de ocupación e ingresos (lado negativo). Esta contraposición, como destacaremos en la representación gráfica, se identifica con el llamado efecto Guttman.

Las contribuciones relativas son el cuadrado del coseno del ángulo que forman un punto y el eje factorial. Su interpretación es equivalente a la del coeficiente de correlación entre una categoría y cada factor, midiendo la contribución relativa del factor o eje k en la posición de una categoría, es decir, dando cuenta de la calidad de representación de las categorías activas sobre el eje. Cuando una categoría se proyecta mucho sobre uno de los ejes (su correlación es alta) y poco o nada sobre los otros se trata de un punto cuya inercia se acumula en un solo factor. En otros casos su inercia se repartirá entre diferentes factores. La lectura de la información de la tabla hay que hacerla teniendo en cuenta que la suma de los valores por fila suma en total 1 en cada fila ($\sum_{k=1}^{K} CTR_{ik} = 1$). Los valores más altos indican dónde están mejor representadas las categorías en cuestión. La contribución de la categoría i sobre el eje k se obtiene a partir de sus coordenadas divididas por el cuadrado de la distancia al origen:

$$CTR_{ik} = \frac{y_{ik}^2}{d^2(i, G_J)} = \cos^2(i,k) \qquad \textit{Ecuación 18}$$

En el ejemplo que nos ocupa podemos ver cómo, por ejemplo, la categoría de ingresos altos aparece reflejada casi por completo en el eje 1 y casi nada en el eje 2, comportamiento extremo que también se puede observar en el resto de puntos-fila y puntos-columna en el reparto entre ambos ejes.

Finalmente nos quedan las representaciones gráficas que nos permiten observar de forma visual los patrones de asociación que hemos destacado con las tablas. Estas representaciones se puede realizar en cada espacio vectorial por separado, puntos-fila en $\mathbb{R}^J$ o puntos-columna en $\mathbb{R}^I$, o bien simultáneamente que es la opción más interesante para observar las relaciones entre las categorías de filas y columnas.

Para interpretar los resultados de un ACS conviene leer los datos de las contribuciones con las consiguientes representaciones teniendo en cuenta los siguientes aspectos:

- Cada punto representa la distribución condicional de cada categoría f_{ij}^F o f_{ij}^C
- El origen de coordenadas de los ejes factoriales, el centro de gravedad, representa la distribución marginal f_{i+} o bien f_{+j}.
- La distancia entre dos perfiles o puntos-filas o puntos-columna es la imagen para esta proyección de la distancia de chi-cuadrado, esta imagen debe interpretarse teniendo en cuenta los efectos de perspectiva asociados a la proyección.
- En la medida en que la relevancia de un punto depende de su contribución absoluta y que a éstos se les asocia mayor importancia cuanto mayor es su masa o peso (número de casos de la categoría), hay que tener presente que un punto alejado del centro del gráfico puede no tener un peso importante si su masa es pequeña; y también un punto cercano al centro podría tener una masa importante y tener en consecuencia una contribución relevante que no se refleja en su distancia del centro.

Como procedimiento de lectura de la información gráfica se pueden seguir las pautas siguientes:

- Buscar en primer lugar los puntos o categorías con mayor contribución absoluta.
- De estos puntos se distinguen entre los positivos y los negativos que definirán cada una de las polaridades del eje.

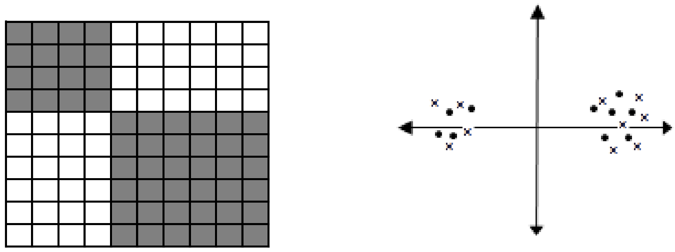
- Se estudia seguidamente la calidad de la representación de estos puntos, su contribución relativa, viendo en qué eje tiene mayor y menor presencia. Los valores altos son los que definirán el significado del eje. Como expresan la correlación con el eje, un valor alto significa que el punto está alineado o muy cercano al eje y caracteriza sobre todo esta dimensión.
- Otros puntos que tengan una contribución absoluta baja, es decir, cercanos al centro y con poco peso explicativo en la configuración del factor, pero con una contribución relativa importante sobre el eje, se situarán cerca de éste y ayudan a mostrar el recorrido y la identidad de la dimensión.
- Una vez analizado cada eje se interrelacionan para dar cuenta de la estructura de relaciones y el orden jerárquico que representa cada eje.
- Cuando una categoría tiene un perfil coincidente con el perfil medio ubicará en el centro del espacio próximo al origen. De la misma forma un alejamiento de este comportamiento medio implica un alejamiento del origen. Este punto central de referencia es de interés pues representa el perfil promedio de la información tratada, una especie de tipo ideal medio que define un modelo de comportamiento genérico.
- Si dos filas o columnas tienen una estructura o perfiles similares, estarán próximos en el espacio.
- Si un punto-fila y un punto-columna están próximos, su interpretación como asociación es válida cuando su disposición en el espacio está alejada del origen. De lo contrario la posible relación puede deberse más al hecho de que otros puntos fila, por ejemplo más alejados y situados en lugares o extremos opuestos, tienen una importancia sobre el punto-columna.

En el análisis de correspondencias existe la posibilidad, que veremos más adelante, de representar categorías suplementarias o ilustrativas, y es una ayuda muy interesante para la interpretación. Se pueden utilizar con diferentes objetivos:

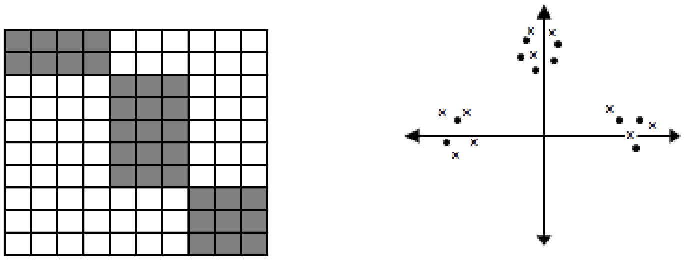
- Para representar centros de gravedad de algunas clases, sobre todo cuando el número de líneas es elevado.
- Para representar las características identificativas de grupos de individuos.
- En general para introducir información exógena que facilite la interpretación de los factores y de las categorías representadas.
- Para representar un elemento perturbador según un análisis previo que distorsionaba la estructura de relaciones obtenida.

Por último destacaremos desde el punto de vista gráfico algunas nubes de puntos características que obedecen a una disposición de los datos en la tabla de correspondencias:

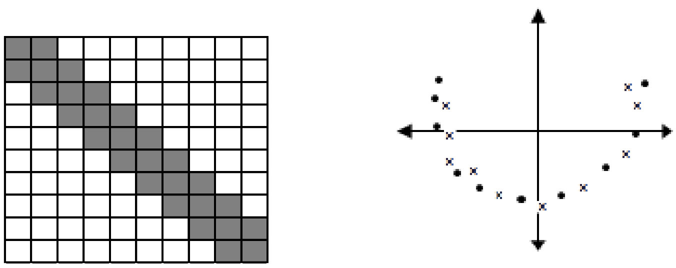
– Cuando se pueden ordenar las filas y columnas de forma que las casillas con más efectivos configuran dos bloques o subtablas, como las de la imagen adjunta, entonces el gráfico factorial se configura con dos nubes de puntos opuestos.



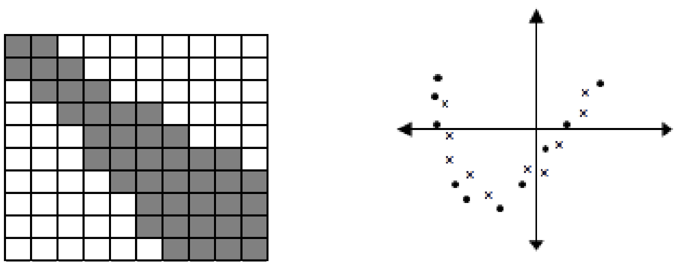
– Igualmente se pueden configuran tres bloques o subtablas:



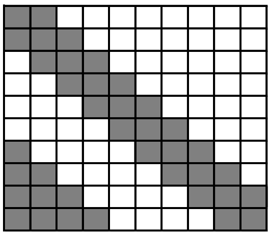
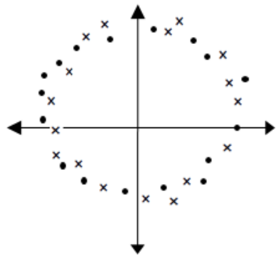
– El denominado efecto Guttman configura una nube de puntos en forma de parábola o arco. Esto sucede cuando existe una fuerte asociación entre las dos variables y los efectivos se disponen en la diagonal o bien estos efectivos de la tabla por filas y columnas se reordenan con las frecuencias más altas en la diagonal. Se configura un primer factor que opone los valores extremos, quedando en el centro los valores medios, mientras que el segundo factor opone los valores extremos ante los medios.



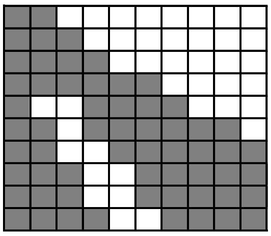
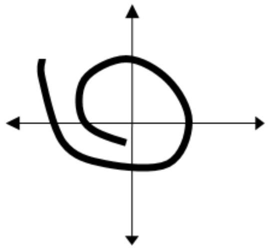
– También se puede configurar una parábola asimétrica con la concentración de efectivos en una diagonal con forma de trapecio. A medida que aumentan los valores de una variable aumenta la dispersión de los valores de la otra variable.



- Es posible una configuración circular cuando se relacionan en el mismo sentido los valores altos y bajos.

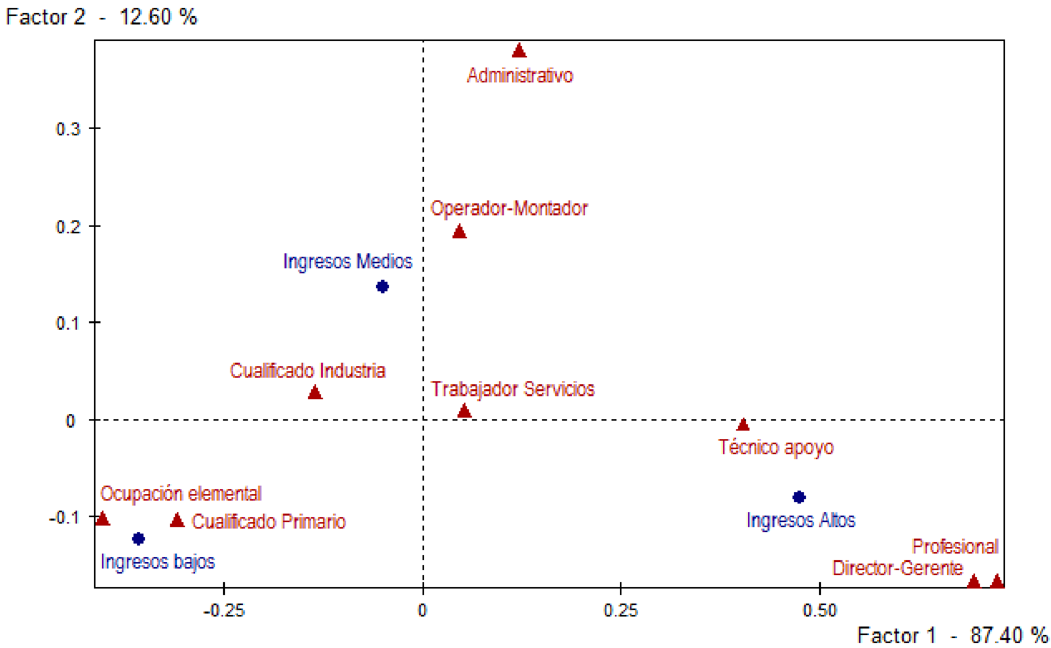
- Una combinación de los dos casos anteriores daría lugar a la configuración de una forma espiral.

En el ejemplo que analizamos obtenemos la representación factorial del Gráfico III.11.20. En él podemos apreciar la forma característica del efecto Guttman que opone mayores y menores niveles de ingresos y ocupación en la primera dimensión (eje horizontal), y la contraposición de los niveles medios de los extremos en la segunda (eje vertical). A pesar de la distribución espacial de los puntos con coordenadas que implican un alejamiento del centro tanto en la dirección derecha-izquierda como norte-sur, es importante no olvidar la importancia de cada factor. En este análisis obtenemos que la primera dimensión acumula con el 86% la mayor parte de la varianza, por tanto, que se configura una realidad prácticamente unidimensional expresada en el primer factor. La lectura derecha-izquierda es la primordial, por lo que los alejamientos de los puntos hacia el norte o hacia el sur son aspectos secundarios del patrón de asociación que muestra el primer factor[24].

[24] Las representaciones gráficas, aun siendo el resultado del ACS de carácter unidimensional, requieren un mínimo de dos dimensiones para por poder dibujarse.

Gráfico III.11.20. Gráfico factorial del ACS. Ingresos y ocupación



La posibilidad de representar simultáneamente los puntos-filas y los puntos-columnas descansa sobre la base de las fórmulas de transición entre los dos espacios factoriales de $\mathbb{R}^J$ y $\mathbb{R}^I$. Estas fórmulas de cambio de base se pueden emplear también para colocar en el espacio factorial categorías suplementarios o ilustrativas que no forman parte de la tabla de correspondencias original y que no juegan ningún papel en la determinación de los ejes, pero que se pueden proyectar sobre los ejes factoriales y ayudan de manera importante a interpretar los resultados de un ACS. Si consideramos las categorías de una variable como líneas adicionales de la tabla de contingencia, con coordenadas suplementarias f_{ij}^s / f_{i+}^s, la proyección de cada nuevo punto-fila sobre los ejes factoriales obtenidos a partir de la tabla de contingencia original se expresa como:

$$y_{ik}^S = \frac{1}{\sqrt{\lambda_k}} \cdot \sum_{j=1}^{J} \frac{f_{ij}^S}{f_{i+}^S} \cdot z_{jk} \qquad \textit{Ecuación 19}$$

5.1.3. *Ejemplos de aplicación del ACS*

Complementaremos el ejemplo visto en el apartado anterior con otros dos adicionales que analizarán tablas de correspondencias mayores o de formatos distintos.

En primer lugar consideraremos una tabla con un mayor número de categorías analizando la relación entre la distribución del nivel educativo de la población española y la provincia de residencia[25]. Se trata de la información de una tabla de correspondencias con 52 filas de provincias y 12 columnas de niveles de estudios (Tabla III.11.14) con la distribución porcentual por fila.

[25] Los datos se obtuvieron del Censo de Población del 2011 desde la web del Instituto Nacional de Estadística y se refieren a personas ocupadas de 16 y más. La matrices de datos España-Provincias x Educación.sba y España-Provincias x Educación.sav contienen esta información en forma de tabla de correspondencias.

Tabla III.11.14. Provincia de residencia y nivel educativo

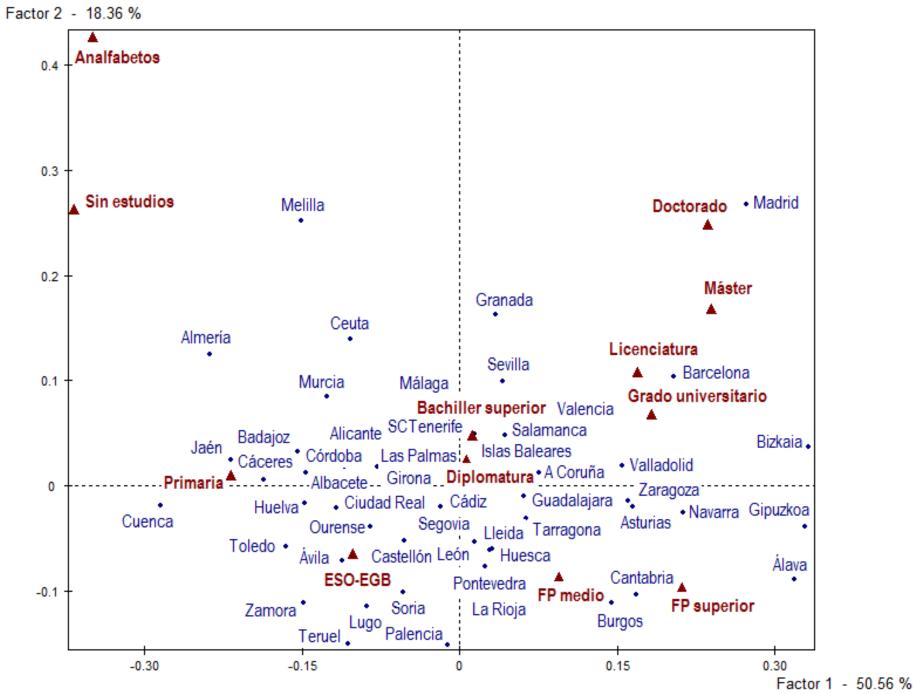
Nivel de estudios / Provincia	Analfabetos	Sin estudios	Primaria	ESO-EB	Bachiller Superior	FP medio	FP superior	Diplomatura	Grado Universitario	Licenciatura	Máster	Doctorado	Total	
Araba/Álava	0,2	1,0	5,6	22,8	10,6	10,7	18,5	11,2	2,1	15,0	1,5	0,9	**100,0**	**136.115**
Albacete	0,2	2,8	12,9	31,9	12,1	7,5	7,1	11,2	1,6	10,3	1,6	0,7	**100,0**	**138.810**
Alicante/Alacant	0,4	2,4	10,1	32,1	14,6	7,8	7,7	9,5	2,1	10,8	1,5	0,9	**100,0**	**610.375**
Almería	1,3	4,4	13,9	29,8	13,3	6,4	7,1	10,7	1,7	9,6	1,0	0,7	**100,0**	**242.575**
Ávila	0,2	1,7	10,3	34,9	13,8	8,3	7,3	10,4	1,5	10,3	0,9	0,5	**100,0**	**59.135**
Badajoz	0,4	3,4	11,8	31,7	12,0	7,3	7,3	12,3	1,1	10,8	1,1	0,8	**100,0**	**220.395**
Illes Balears	0,3	1,6	8,6	29,6	17,9	8,3	7,4	9,9	2,5	11,4	1,7	0,7	**100,0**	**419.935**
Barcelona	0,3	1,6	6,2	24,6	14,8	8,8	10,8	10,3	3,1	15,0	3,1	1,4	**100,0**	**2.247.220**
Burgos	0,1	1,2	9,2	24,9	12,9	11,1	15,1	10,5	1,7	11,6	1,1	0,6	**100,0**	**150.080**
Cáceres	0,6	3,2	11,2	35,1	10,3	7,0	6,7	12,3	1,4	10,5	1,0	0,6	**100,0**	**135.410**
Cádiz	0,4	2,8	10,6	26,4	11,8	9,8	11,2	12,5	1,4	10,8	1,4	0,8	**100,0**	**365.340**
Castellón/Castelló	0,3	1,8	9,6	34,0	13,0	8,1	8,9	9,7	1,3	11,3	1,4	0,8	**100,0**	**219.510**
Ciudad Real	0,4	3,1	11,5	32,6	10,7	8,0	9,4	10,5	1,4	10,2	1,4	0,7	**100,0**	**171.960**
Córdoba	0,3	3,7	11,3	31,2	9,7	8,6	8,8	11,6	1,3	11,2	1,1	1,1	**100,0**	**259.495**
A Coruña	0,2	2,1	8,8	28,2	12,9	8,4	11,5	9,9	1,5	13,8	1,7	1,2	**100,0**	**435.730**
Cuenca	0,4	3,3	15,8	34,9	11,1	7,0	6,0	10,0	1,3	9,1	0,9	0,3	**100,0**	**73.710**
Girona	0,6	2,3	9,1	31,9	15,1	8,8	8,1	8,7	2,5	10,6	1,4	0,9	**100,0**	**302.495**
Granada	0,4	2,7	9,4	27,7	11,9	7,1	7,7	11,5	1,8	15,0	2,1	2,6	**100,0**	**285.020**
Guadalajara	0,4	1,3	7,4	29,4	17,0	8,3	10,3	9,7	1,9	12,4	1,2	0,8	**100,0**	**104.650**
Gipuzkoa	0,2	1,0	5,6	21,9	10,8	11,1	17,1	10,9	2,5	16,2	1,7	1,0	**100,0**	**296.865**
Huelva	0,5	3,6	12,1	31,0	11,1	8,1	9,5	12,1	1,4	9,2	0,8	0,6	**100,0**	**163.560**
Huesca	0,2	1,5	8,5	28,9	15,0	9,9	10,4	12,1	1,8	9,4	1,3	0,9	**100,0**	**89.565**
Jaén	0,7	4,1	13,4	32,0	10,8	8,3	7,5	11,2	1,0	9,5	1,1	0,6	**100,0**	**203.115**
León	0,2	1,2	7,8	30,8	16,4	8,8	9,4	11,6	1,6	10,2	1,0	1,1	**100,0**	**176.125**
Lleida	0,4	1,5	7,7	30,8	13,8	10,6	9,7	10,5	2,0	10,7	1,4	0,8	**100,0**	**180.005**
La Rioja	0,3	1,2	8,9	30,2	12,4	10,2	11,4	11,1	1,6	11,1	0,9	0,8	**100,0**	**129.290**
Lugo	0,1	2,2	9,8	35,4	12,7	7,5	10,8	9,4	1,0	9,3	1,0	0,6	**100,0**	**132.975**
Madrid	0,2	1,3	5,8	20,6	17,7	6,6	8,4	10,5	2,9	21,2	3,0	1,8	**100,0**	**2.769.990**
Málaga	0,3	2,8	10,3	27,9	15,3	7,4	8,3	10,6	2,4	12,5	1,3	1,0	**100,0**	**523.870**
Murcia	1,1	3,6	11,5	29,6	13,2	8,1	8,1	10,2	1,7	10,5	1,3	1,2	**100,0**	**514.680**
Navarra	0,3	1,0	7,0	24,7	11,5	10,7	13,6	11,9	2,1	13,8	2,0	1,3	**100,0**	**266.990**
Ourense	0,1	2,1	10,2	33,0	15,5	7,8	8,3	9,7	1,2	10,3	1,1	0,8	**100,0**	**112.425**
Asturias	0,1	0,8	6,2	26,1	16,0	9,0	12,1	12,5	1,5	13,1	1,4	1,2	**100,0**	**390.910**
Palencia	0,2	1,0	10,3	31,4	11,5	10,9	10,7	11,9	1,3	9,5	0,8	0,6	**100,0**	**65.210**
Las Palmas	0,4	2,4	11,5	27,6	15,8	7,9	9,7	10,1	2,5	10,3	1,3	0,7	**100,0**	**398.585**
Pontevedra	0,2	1,6	9,6	31,2	12,2	8,2	12,0	9,7	1,4	11,7	1,4	0,8	**100,0**	**347.505**
Salamanca	0,1	1,2	8,9	30,5	12,3	8,2	7,5	11,6	1,6	13,9	1,7	2,3	**100,0**	**127.430**
S.C. de Tenerife	0,4	2,3	10,4	26,7	16,1	7,8	10,7	9,2	2,2	11,5	1,4	1,2	**100,0**	**356.120**
Cantabria	0,1	0,9	6,0	28,2	13,5	10,8	13,5	10,8	1,9	12,0	1,5	0,8	**100,0**	**227.730**
Segovia	0,3	1,5	9,7	30,9	14,8	8,1	8,8	10,2	1,5	12,8	0,9	0,6	**100,0**	**63.960**
Sevilla	0,4	2,7	9,8	26,2	11,3	8,7	9,3	11,8	1,5	15,2	1,6	1,5	**100,0**	**632.140**
Soria	0,1	1,2	10,0	32,4	14,4	8,6	8,9	12,3	1,6	9,3	0,9	0,4	**100,0**	**38.530**
Tarragona	0,4	2,3	8,4	28,2	12,6	10,5	11,5	10,1	2,4	11,1	1,9	0,8	**100,0**	**304.195**
Teruel	0,5	1,7	9,1	36,3	12,8	9,8	9,4	10,3	1,3	7,8	0,7	0,4	**100,0**	**53.445**
Toledo	0,3	2,8	12,0	34,3	13,5	8,1	8,1	9,0	1,3	9,2	1,0	0,4	**100,0**	**246.965**
Valencia/València	0,3	1,4	8,0	27,8	13,8	8,4	9,2	10,9	2,0	15,1	1,8	1,4	**100,0**	**941.145**
Valladolid	0,2	0,9	7,5	25,8	12,7	10,2	10,2	12,7	1,6	15,1	1,4	1,6	**100,0**	**211.660**
Bizkaia	0,2	0,9	5,7	21,3	11,8	10,0	15,0	10,9	2,3	18,6	2,1	1,1	**100,0**	**461.585**
Zamora	0,1	1,2	11,4	36,5	13,7	7,8	7,1	10,6	1,1	9,5	0,7	0,4	**100,0**	**63.955**
Zaragoza	0,1	0,9	6,6	26,3	14,4	10,6	10,8	12,3	2,0	12,8	1,8	1,4	**100,0**	**394.585**
Ceuta	0,9	3,4	10,7	26,7	16,8	7,0	7,2	12,5	1,4	11,4	1,2	0,8	**100,0**	**26.435**
Melilla	1,4	5,9	10,1	23,4	16,7	6,6	7,6	14,3	0,7	12,1	0,7	0,6	**100,0**	**25.035**
Total	**0,3**	**1,9**	**8,4**	**26,9**	**14,2**	**8,4**	**9,8**	**10,7**	**2,2**	**14,1**	**1,9**	**1,2**	**100,0**	**17.514.550**

Fuente: INE, Censo de Población 2011.

EL ACS de esta tabla supone considerar un número de dimensiones totales de $mín(I,J)\text{-}1 = mín(52,12)\text{-}1 = 11$ factores. En el Gráfico III.11.21 se representan los dos primeros factores que acumulan el 69% de la varianza explicada. De estos resultados concluimos la configuración de un primer factor principal que recorre el primer eje oponiendo los niveles educativos más altos en la derecha y los más bajos en la izquierda. Esta primera dimensión de menor a mayor nivel educativo se corresponde con una distribución de las distintas provincias según perfil característico en relación a las demás y al perfil medio que corresponde al centro del gráfico (este es el punto del perfil de España en conjunto). El lector podrá entretenerse en su mirada particular. Destacaremos como Madrid, Barcelona y País Vasco junto a otras provincias próximas destacan por situarse en la zona de la derecha mostrando que son les territorios con mayor proporción de estudios superiores: universitarios y de formación profesional de grado superior. Por el contrario provincias como Cuenca, Almería, Jaén o Cáceres son las que más se alejan del centro hacia la izquierda en este primer factor por ser las que tienen una mayor presencia de niveles educativos bajos.

El segundo factor polariza los niveles medios (ESO-EGB, Formación profesional) frente a los muy altos y los muy bajos. Con este segundo factor configura un espacio social bidimensional que posiciona a las distintas provincias en los lugares donde su perfil característico difiere del comportamiento general y se aproxima al de otras provincias homólogas. Estos diferentes espacios se pueden identificar en los distintos cuadrantes que configura el gráfico: mayor presencia de muy bajos niveles educativos (cuadrante izquierdo superior), de niveles bajos-medios (cuadrante izquierdo inferior), niveles medio-altos (cuadrante inferior derecho) y de niveles altos (cuadrante superior izquierdo). Una quinta zona se puede identificar si consideramos el espacio alrededor del centro del gráfico donde aparecen las provincias que tienen una distribución más parecida a la del conjunto del país.

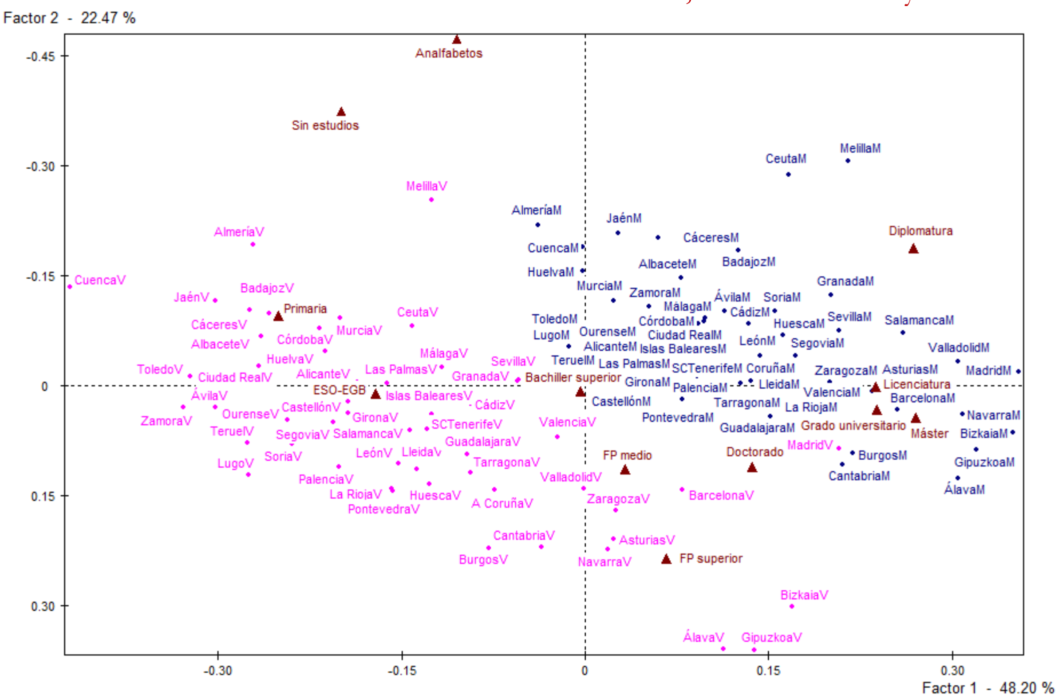
Gráfico III.11.21. Gráfico factorial del ACS. Provincias y Nivel educativo



Nos podemos preguntar a continuación en qué medida ese "mapa o espacio social" se modifica como resultado de considerar los niveles educativos alcanzados por varones y mujeres. Además de los comportamientos diferenciados por provincias que expresan distintos niveles de desarrollo del país, la expansión educativa de los últimos años se ha producido con una mayor incorporación de las mujeres a los estudios universitarios, en un contexto donde los varones tienden a abandonar más los estudios y cursan en mayor medida formación profesional. ¿Cómo se pueden reflejar estos aspectos mediante un análisis de correspondencias simples? Manejaremos tres variables, provincia, educación y sexo, si bien un ACS implica tratar tan solo dos. Ello será posible mediante la disposición de más filas en la tabla de correspondencias resultado de desdoblar cada provincia en dos: con el porcentaje de cada nivel de estudios sobre el total de varones y sobre el total de mujeres[26].

El Gráfico III.11.22 es elocuente y muestra una nueva imagen respecto del análisis anterior. Algunas tendencias se reiteran: el mayor y menor nivel educativo asociado a los distintos territorios y la configuración del carácter de los dos factores (en este caso con el 71% de la inercia), pero destaca ante todo la separación que se produce entre el perfil educativo de varones y mujeres en todas las provincias marcando dos espacios sociales claramente separados. Por un lado las mujeres alcanzan mayores niveles educativos, especialmente universitarios, mientras que los varones destacan por un perfil predominante de estudios medios y bajos, en relación al conjunto y a las mujeres, y también por la mayor proporción de formación profesional.

Gráfico III.11.22. Gráfico factorial del ACS. Provincias, Nivel educativo y Sexo



[26] La matrices de datos España-Provincias x Educación-Sexo.sba y España-Provincias x Educación-Sexo.sav contienen esta información en forma de tabla de correspondencias.

Un tratamiento similar al de los datos que acabamos de ver se puede obtener con otras tablas de correspondencias de números positivos. Por ejemplo podríamos disponer en filas países o cualquier otra división territorial y analizar el producto interior bruto distribuido por sectores de actividad. En este caso en vez del recuento de casos y los perfiles o porcentajes asociados disponemos del valor de los bienes y servicios producidos.

Para ilustrar el funcionamiento de la técnica presentaremos otro ejercicio tratando otro tipo de tabla de correspondencias, realizaremos un ACS destinado al análisis de una matriz de distancias, la que se obtiene de considerar los kilómetros por carretera entre las 47 capitales de provincia españolas peninsulares (Tabla III.11.15). El resultado de la aplicación de un ACS a este tipo de tabla de correspondencias, una matriz cuadrada y simétrica, con las distancias en el espacio físico entre ciudades, no es más "el mapa" de España (Gráfico III.11.23), con ciertas particularidades dadas la escasa información que se maneja (47 puntos del espacio). Lo primero que destaca es que las referencias cardinales habituales de la península, este-oeste y norte-sur, aparecen descolocadas.

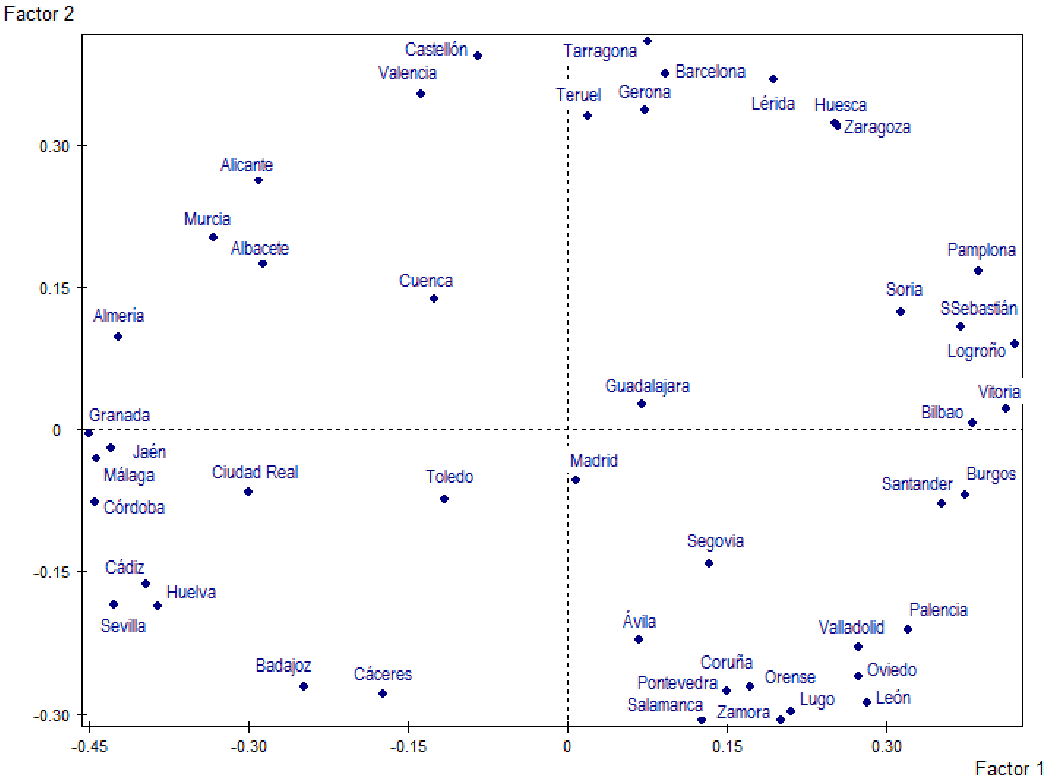
Ello es debido a que el principal factor que emerge para caracterizar las distancias entre las ciudades es la altura, España es algo más alta que ancha en distancias entre capitales de provincia[27], por lo que, en primer lugar, se disponen las ciudades en las posiciones relativas que marca el conocido factor norte-sur. En segundo lugar, el segundo factor muestra la orientación este-oeste de las ciudades peninsulares. Evidentemente los puntos cardinales son convenciones y podemos convenir mirar el gráfico cambiando las direcciones que marcan. Si intercambiamos en el gráfico los ejes factoriales observaremos como, con la salvedad de los desajustes propios de la falta de información, se reproduce la fisonomía habitual de representación de la España peninsular.

Tabla III.11.15. Matriz de distancias por carretera en km. entre las capitales de provincia españolas peninsulares

	Albacete	Alicante	Almería	Ávila	Badajoz	Barcelona	Bilbao	Burgos	Cáceres	Cádiz	Castellón	Ciudad	Córdoba	Coruña	Cuenca	Gerona	Granada	Guadalajara	Huelva	Huesca	Jaén	León	Lérida	Logroño	Lugo	Madrid	Málaga	Murcia	Orense	Oviedo	Palencia	Pamplona	Pontevedra	Salamanca	SSebastián	Santander	Segovia	Sevilla	Soria	Tarragona	Teruel	Toledo	Valencia	Valladolid	Vitoria	Zamora	Zaragoza
Albacete	0	171	369	366	525	540	646	488	504	617	256	207	354	860	142	640	363	309	506	495	264	584	515	578	762	251	473	150	772	702	491	598	874	463	691	644	338	492	473	443	242	240	191	444	602	499	423
Alicante	171	0	294	537	696	515	817	659	675	688	231	378	525	1031	313	615	353	480	703	570	415	855	490	653	933	422	482	75	943	873	662	673	1045	634	766	815	509	609	548	417	317	411	166	615	739	670	498
Almería	369	294	0	663	604	809	958	800	651	484	525	407	332	1172	511	909	166	621	516	830	228	896	802	899	1074	563	219	219	1084	1014	803	970	1186	763	1032	956	650	422	794	711	577	526	460	756	914	811	758
Ávila	366	537	663	0	318	717	401	243	229	618	532	256	457	538	282	817	534	173	552	490	435	255	558	358	440	115	644	516	443	373	168	437	545	97	475	369	67	493	261	649	417	137	467	121	357	159	418
Badajoz	525	696	604	318	0	1022	694	536	89	342	805	318	272	772	555	1122	438	459	251	798	378	496	866	676	674	401	436	675	645	614	461	755	747	299	768	662	385	217	579	935	703	368	716	414	650	361	726
Barcelona	540	515	809	717	1022	0	620	583	918	1284	284	811	908	1118	562	100	868	563	1140	274	804	784	156	468	1020	621	997	590	1027	902	669	437	1129	778	529	693	650	1046	453	98	409	692	349	663	530	759	296
Bilbao	646	817	958	401	694	620	0	158	605	1058	607	585	795	644	562	720	829	396	939	322	730	359	464	152	546	395	939	796	605	304	244	159	707	395	119	108	355	993	257	555	488	466	633	280	66	376	324
Burgos	488	659	800	243	536	583	158	0	447	900	524	427	637	535	404	683	671	238	781	359	572	201	427	115	437	237	781	638	447	322	86	203	549	237	232	156	197	775	141	518	372	308	517	122	114	218	287
Cáceres	504	675	651	229	89	918	605	447	0	369	701	324	319	683	451	1018	485	355	323	694	423	407	762	595	585	297	506	654	556	525	372	650	658	210	679	573	296	264	490	831	599	264	636	325	561	272	622
Cádiz	617	688	484	618	342	1284	1058	900	369	0	873	464	263	1072	708	1384	335	721	219	1060	367	796	1128	999	974	663	265	613	945	914	761	1070	1047	599	1132	1056	750	125	894	1059	873	583	808	714	1014	661	988
Castellón	256	231	525	532	805	284	607	524	701	873	0	463	610	1026	305	384	584	396	856	355	520	725	259	455	928	417	713	306	938	868	596	458	1040	629	551	680	504	762	383	186	152	437	65	563	574	665	283
Ciudad	207	378	407	256	318	811	585	427	324	464	463	0	201	799	244	911	278	248	433	587	179	511	655	526	696	190	388	357	699	641	424	597	701	353	659	583	277	339	421	649	397	119	398	377	541	415	515
Córdoba	354	525	332	457	272	908	795	637	319	263	610	201	0	995	445	1008	166	458	232	797	104	733	865	736	897	400	187	444	921	851	640	807	977	529	869	793	497	138	631	796	596	320	545	578	751	616	725
Coruña	860	1031	1172	538	772	1118	644	535	683	1072	1026	799	995	0	776	1218	1043	667	1006	905	944	334	973	650	98	609	1153	1010	175	340	450	738	121	473	763	547	560	947	676	1064	907	675	961	455	549	411	833
Cuenca	142	313	511	282	555	562	562	404	451	708	305	244	445	776	0	662	479	486	677	406	380	500	472	464	678	167	615	292	689	618	407	535	791	379	636	560	254	583	359	464	153	187	220	360	518	415	334
Gerona	640	615	909	817	1122	100	720	683	1018	1384	384	911	1008	1218	662	0	968	663	1240	374	904	884	256	568	1120	721	1097	690	1127	1002	769	537	1229	878	629	793	750	1146	553	198	509	792	449	763	630	859	396
Granada	363	353	166	534	438	868	829	671	485	335	584	278	166	1043	479	968	0	492	350	831	99	761	861	770	945	434	129	278	955	885	674	841	1058	631	903	827	521	256	665	770	605	397	519	627	758	682	759
Guadalajara	309	480	621	173	459	563	396	238	355	721	396	248	458	667	486	663	492	0	690	339	393	391	407	278	569	58	602	459	579	509	298	349	681	270	441	394	145	596	173	476	244	129	410	251	352	306	267
Huelva	506	703	516	552	251	1140	939	781	323	219	856	433	232	1006	677	1240	350	690	0	1029	336	730	1097	968	908	632	313	628	879	821	695	1039	981	533	1101	962	619	94	863	1029	842	552	791	648	919	595	957
Huesca	495	570	830	490	798	274	322	359	694	1060	355	587	797	905	406	374	831	339	1029	0	732	560	118	244	807	397	941	611	803	626	442	163	905	554	255	430	426	935	229	209	253	468	398	439	256	535	72
Jaén	264	415	228	435	378	804	730	572	423	367	520	179	104	944	380	904	99	393	336	732	0	668	779	671	846	335	209	340	856	786	575	742	957	532	804	728	422	242	566	707	506	298	455	528	686	583	660
León	584	855	896	255	496	784	359	201	407	796	725	511	733	334	500	884	761	391	730	560	668	0	628	316	236	333	877	734	271	118	130	404	373	197	433	293	245	671	342	719	573	392	685	134	315	135	488
Lérida	515	490	802	558	866	156	464	427	762	1128	259	655	865	973	472	256	861	407	1097	118	779	628	0	312	875	465	1009	583	871	703	513	281	973	622	373	537	494	1003	297	91	319	536	324	507	374	603	140
Logroño	578	653	899	358	676	468	152	115	595	999	455	526	736	650	464	568	770	278	968	244	671	316	312	0	352	336	880	694	562	391	201	88	664	352	169	225	229	874	105	403	336	407	481	237	86	333	172
Lugo	762	933	1074	440	674	1020	546	437	585	974	928	696	897	98	678	1120	945	569	908	807	846	236	875	352	0	511	1055	912	95	242	352	640	148	375	665	449	462	849	578	966	809	577	863	357	451	313	735
Madrid	251	422	563	115	401	621	395	237	297	663	417	190	400	609	167	721	434	58	632	397	335	333	465	336	511	0	544	401	521	451	240	407	623	212	469	393	87	538	231	534	302	71	352	193	351	248	325
Málaga	473	482	219	644	436	997	939	781	506	265	713	388	187	1153	615	1097	129	602	313	941	209	877	1009	880	1055	544	0	407	1065	995	784	951	1153	756	1013	937	631	219	775	899	715	507	648	737	895	792	869
Murcia	150	75	219	516	675	590	796	638	654	613	306	357	444	1010	292	690	278	459	628	611	340	734	583	694	912	401	407	0	922	852	641	714	1024	613	807	794	488	534	598	492	385	390	241	594	752	649	539
Orense	772	943	1084	443	645	1027	605	447	556	945	938	699	921	175	689	1127	955	579	879	803	856	271	871	562	95	521	1065	922	0	337	361	650	102	346	679	544	467	820	574	962	805	580	873	356	561	284	731
Oviedo	702	873	1014	373	614	902	304	322	525	914	868	641	851	340	618	1002	885	509	821	626	786	118	703	391	242	451	995	852	337	0	248	463	390	315	423	207	363	789	463	835	694	510	803	252	340	253	604
Palencia	491	662	803	168	461	669	244	86	372	761	596	424	640	450	407	769	674	298	695	442	575	130	513	201	352	240	784	641	361	248	0	289	463	162	318	201	158	636	213	604	444	305	592	47	200	143	373
Pamplona	598	673	970	437	755	437	159	203	650	1070	458	597	807	738	535	537	841	349	1039	163	742	404	281	88	640	407	951	714	650	463	289	0	752	440	92	267	370	945	176	372	356	478	501	325	93	421	175
Pontevedra	874	1045	1186	545	747	1129	707	549	658	1047	1040	701	977	121	791	1229	1058	681	981	905	957	373	973	664	148	623	1153	1024	102	390	463	752	0	448	781	666	569	922	676	1064	907	682	975	458	663	386	833
Salamanca	463	634	763	97	299	778	395	237	210	599	629	353	529	473	379	878	631	270	533	554	532	197	622	352	375	212	756	613	346	315	162	440	448	0	469	363	164	474	325	713	514	234	564	115	351	62	482
SSebastián	691	766	1032	475	768	529	119	232	679	1132	551	659	869	763	636	629	903	441	1101	255	804	433	373	169	665	469	1013	807	679	423	318	92	781	469	0	227	429	1007	268	464	449	540	594	351	118	450	268
Santander	644	815	956	369	662	693	108	156	573	1056	680	583	793	547	560	793	827	394	962	430	728	293	537	225	449	393	937	794	544	207	201	267	666	363	227	0	359	837	297	628	528	464	673	248	174	344	397
Segovia	338	509	650	67	385	650	355	197	296	750	504	277	497	560	254	750	521	145	619	426	422	245	494	229	462	87	631	488	467	363	158	370	569	164	429	359	0	560	194	585	389	158	439	111	311	182	354
Sevilla	492	609	422	493	217	1046	993	775	264	125	762	339	138	947	583	1146	256	596	94	935	242	671	1003	874	849	538	219	534	820	789	636	945	922	474	1007	837	560	0	769	949	748	458	697	589	825	536	863
Soria	473	548	794	261	579	453	257	141	490	894	383	421	631	676	359	553	665	173	863	229	566	342	297	105	578	231	775	598	574	463	213	176	676	325	268	297	194	769	0	388	231	302	376	210	191	306	157
Tarragona	443	417	711	649	935	98	555	518	831	1059	186	649	796	1064	464	198	770	476	1029	209	707	719	91	403	966	534	899	492	962	835	604	372	1064	713	464	628	585	949	388	0	311	605	251	598	489	694	231
Teruel	242	317	577	417	703	409	488	372	599	873	152	397	596	907	153	509	605	244	842	253	506	573	319	336	809	302	715	385	805	694	444	356	907	514	449	528	389	748	231	311	0	340	145	441	422	537	181
Toledo	240	411	526	137	368	692	466	308	264	583	437	119	320	675	187	792	397	129	552	468	298	392	536	407	577	71	507	390	580	510	305	478	682	234	540	464	158	458	302	605	340	0	372	258	422	296	396
Valencia	191	166	460	467	716	349	633	517	636	808	65	398	545	961	220	449	519	410	791	398	455	685	324	481	863	352	648	241	873	803	592	501	975	564	594	673	439	697	376	251	145	372	0	545	576	600	326
Valladolid	444	615	756	121	414	663	280	122	325	714	563	377	578	455	360	763	627	251	648	439	528	134	507	237	357	193	737	594	356	252	47	325	458	115	351	248	111	589	210	598	441	258	545	0	236	96	367
Vitoria	602	739	914	357	650	530	66	114	561	1014	574	541	751	549	518	630	758	352	919	256	686	315	374	86	451	351	895	752	561	340	200	93	663	351	118	174	311	825	191	489	422	422	576	236	0	332	258
Zamora	499	670	811	159	361	759	376	218	272	661	665	415	616	411	415	859	682	306	595	535	583	135	603	333	313	248	792	649	284	253	143	421	386	62	450	344	182	536	306	694	537	296	600	96	332	0	463
Zaragoza	423	498	758	418	726	296	324	287	622	988	283	515	725	833	334	396	759	267	957	72	660	488	140	172	735	325	869	539	731	604	373	175	833	482	268	397	354	863	157	231	181	396	326	367	258	463	0

27 No obstante, la distancia entre los dos puntos extremos del este (Cabo de Creus) y el oeste (Cabo Touriñán) es mayor que entre los dos puntos extremos del norte (Estaca de Bares y Punta de Tarifa). Nuestros datos no reflejan esos extremos de la orografía.

Gráfico III.11.23. Gráfico factorial del ACS. Distancias entre ciudades de España



5.2. Análisis de correspondencias múltiples

El ACM es una generalización del análisis factorial de correspondencias simples que hace posible el estudio simultáneo de las relaciones de asociación entre múltiples variables cualitativas. Por ello es especialmente adecuado para el tratamiento de encuestas sociológicas y análisis multidimensional de la realidad social, y su aplicación se extiende a muchos otros diseños de investigación.

Como el ACS se trata de describir interrelaciones. El resultado de un ACM es también la obtención de unos factores como nuevas y reducidas dimensiones que se generan a partir de un modelo de análisis expresado en un conjunto más o menos numeroso de variables cualitativas. Estos factores o dimensiones expresan en gráficos de proyección factoriales la existencia de un continuo que opone categorías extremas pasando por un centro. Las asociaciones entre las variables se reflejan en la proximidad en el espacio. Como técnica de análisis factorial con variables categóricas permite la representación gráfica como un importante elemento de ayuda a la interpretación en destacar geométricamente las interacciones entre las variables y por tanto de estructurar la información. El objetivo de este tipo de técnica factorial es la ordenación escalar tanto de los individuos como de las categorías de las variables analizadas.

La generalización multivariable del análisis de correspondencias es posible a partir de considerar una tabla de correspondencias particular, resultado de construir una matriz de datos con la unión de todas las tablas de contingencia que relacionan cada pareja variables entre sí. Es una simple extensión del dominio de aplicación del ACS, pero con procedimientos de cálculo y reglas de interpretación específicas como veremos a continuación.

A partir de una matriz de datos original X, si consideramos un conjunto n individuos (indexados por $i=1...n$), sobre los que observamos p variables cualitativas x (indexadas por $j=1...p$). Cada variable cualitativa $x_{\cdot j}$ tiene un número de modalidades o categorías c_j posibles igual a J_j, y donde cada individuo posee una y sólo una categoría en cada variable. Así, cada variable $x_{\cdot j}$ puede descomponerse en J_j categorías. Si un individuo i ha respondido a la pregunta o variable j con la modalidad $c_j=c_{jo}$, entonces este individuo tendrá el valor 1 para esta categoría, $x_{ijc}=1$ si $c_j=c_{jo}$, y 0 para el resto de las categorías de la variable j considerada, $x_{ijc}=0$ si $c_j \neq c_{jo}$.

Por este procedimiento denominado de codificación disyuntiva completa se obtiene una matriz lógica, disyuntiva o binaria D con n filas, una por individuo, y c columnas ($c=\sum_{cj=1}^{J_j} c_j$). La Matriz o Tabla Disyuntiva D asociada a la matriz de datos original X es una transformación de ésta que expresa cada categoría de respuesta de cada una de las variables en términos de una nueva variable binaria o disyuntiva según la respuesta del individuo i: si para la variable j, el individuo i responde el valor de la categoría c, la nueva variable binaria tendrá para este individuo el valor 1, de forma abreviada $d_{ijc}=1$, de lo contrario tendrá el valor 0, $d_{ijc}=0$. Veamos un ejemplo.

En los datos que se adjuntan a continuación, en la matriz X:

$$X = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 2 & 1 \\ 1 & 3 & 2 \\ 2 & 1 & 2 \\ 1 & 2 & 1 \end{pmatrix}$$

x_{ij} es el valor que toma el individuo i ($i=1...n$, con $n=5$) en la variable j ($j=1...p$, con $p=3$). Imaginemos que la primera columna es la variables sexo (1 varón y 2 mujer), la segunda el nivel de estudios (1 bajo, 2 medio y 3 alto) y la tercera el voto (1 sí vota y 2 no vota). El conjunto de categorías que componen cada variable cualitativa j puede ser identificado con el índice c ($c_j=1...J_j$). Por ejemplo, para la segunda variable su valores variarán entre $c_2=1$ (primer valor correspondiente a estudios bajos) hasta $c_2=3$ (último valor correspondiente a estudios altos). La codificación disyuntiva completa implica considerar tres nuevas variables codificadas con 0 y 1: por ejemplo, un individuo con estudios medios se identifica con los códigos $0,1,0$, el primer 0 indica que no tiene estudios bajos, el 1 que tiene estudios medios y el último 0 que no tiene estudios altos. Operando esta codificación para todos los individuos y variables se obtiene la Matriz Disyuntiva D:

$$D = \begin{pmatrix} 10 & 100 & 01 \\ 01 & 010 & 10 \\ 10 & 001 & 01 \\ 01 & 100 & 01 \\ 10 & 010 & 10 \end{pmatrix}$$

La Matriz de Burt B o de correspondencias múltiples asociada a la matriz disyuntiva D se obtiene a partir de la contabilidad del número de veces que aparecen las combinaciones de respuesta entre cada par de variables, es decir, se trata de todas las posibles tablas de contingencia que se pueden formar con las p variables cualitativas consideradas, incluyendo la resultante de cruzar cada variable consigo misma. Como resultado se obtiene una matriz de $p \times p$ bloques (o tablas de contingencia), simétrica, donde la diagonal principal coincide con las frecuencias simples de cada una de las variables j. El término general $b_{jc,j'c'}$ de la tabla de Burt identifica la frecuencia con que aparece la categoría c de la variable j y la categoría c' de la variable j'. Esta matriz se obtiene multiplicando de la matriz disyuntiva por su traspuesta: $B=D'D$. Con los datos del ejemplo se obtiene esta matriz:

$$B = D'D = \begin{pmatrix} 3 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 2 & 1 & 1 & 0 & 2 & 1 \\ 1 & 1 & 2 & 0 & 0 & 0 & 2 \\ 1 & 1 & 0 & 2 & 0 & 2 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 2 & 0 & 2 & 0 & 2 & 0 \\ 1 & 1 & 2 & 0 & 1 & 0 & 3 \end{pmatrix}$$

con las tablas de contingencia de las tres variables cruzadas entre sí: en la diagonal por sí mismas, dando a las frecuencias de cada categoría de cada variable con ceros adyacentes, y las tres tablas de contingencia repetidas en el triángulo inferior y superior pero intercambiando filas por columnas.

La extensión del ACS al ACM se basa en la propiedad siguiente: si se consideran n individuos con dos variables de I y J categorías mutuamente excluyentes, entonces es equivalente someter al ACS la tabla de contingencia *N(I, J)* cruzando las dos variables que analizar la tabla disyuntiva *D* de *n* líneas e *I+J* columnas, o analizar la tabla de Burt de *I+J* líneas e *I+J* columnas. El ACM resulta pues de la aplicación a más de dos variables de un ACS para tablas del tipo *D* o *B*[28].

La aplicación de este procedimiento para obtener una ACM tiene como consecuencia que deban considerarse propiedades matemáticas particulares y algunas reglas de interpretación diferentes. Estos aspectos los presentamos a continuación:

– A partir de la Tabla de Burt se procede a la aplicación de esta técnica que busca la obtención de los valores propios que obtienen de diagonalizar la matriz de varianzas y covarianzas:

$$V = \frac{1}{p} D^{-1} B \qquad \textit{Ecuación 20}$$

– El número de factores en el ACM es es igual a la diferencia entre el número de categorías del conjunto de variables activas (*m*): $m - p$.

– La suma de los valores propios obtenidos es igual a la diferencia entre el número de categorías del conjunto de variables activas (*m*) y el número de estas variables (*p*), dividido por este número: $\sum_{j=1}^{p} \lambda_j = \frac{m-p}{p}$.

– Como en el análisis de correspondencias simples, los valores propios (λ_k) se pueden expresar como proporción de la varianza total explicada. Sin embargo, a diferencia del ACS y también del ACP, el valor total de esta varianza y los porcentajes que representan cada uno de los valores propios tienen un valor y una significación diferente, en general más pesimista. La proporción de inercia explicada por los ejes factoriales es débil, la codificación binaria introduce un ”ruido” que reduce la parte de explicación debida a cada valor propio. Dado el reducido valor de varianza explicada a que dan lugar los valores propios directamente obtenidos, Benzécri (1979: 377) ha propuesto una fórmula de cálculo de corrección de los valores propios mediante las siguientes operaciones:

1) Calcular la inversa del número de variables intervinientes, *1/p*.
2) Seleccionar todos aquellos valores propios que superen o igualen el valor *1/p*.
3) Calcular unos nuevos valores propios corregidos (λ_k^C) mediante una transformación que se expresa en la fórmula:

$$\lambda_k^C = \left(\frac{p}{p-1}\right)^2 \left(\lambda_k - \frac{1}{p}\right)^2 \qquad \textit{Ecuación 21}$$

4) A partir de estos nuevos valores propios calcular de nuevo el porcentaje de inercia o varianza total explicada.

Siguiendo a Greenacre (2008: 187-191, 98-201 y 274), podemos introducir una nueva modificación en el cálculo del porcentaje de inercia o varianza explicada que, siguiendo la corrección introducida Benzécri ajuste su cálculo eliminando la parte

[28] La legitimidad de esta presentación del ACM reside sobre la demostración de que el ACM es un caso particular del análisis canónico generalizado realizado en el caso en que las relaciones entre las variables se consideran por bloques.

de la inercia que corresponde a las tablas de la diagonal de la matriz de Burt. Cuando se procede de esta forma el autor denomina al ACM como análisis de correspondencias conjunto (ACCo). Los resultados son idénticos, tan sólo los datos están afectados por un factor de escala. En términos de inercia explicada se procede a transformar los valores propios de la misma forma que Benzécri (*Ecuación 21*). Pero se pone el nuevo valor corregido en relación con un nuevo total, no es la inercia total de la matriz de Burt, *I(B)*, sino la inercia que resulta de extraer la inercia de las tablas de la diagonal de esta matriz[29]. La inercia (total corregida) que se corresponde con las tablas situadas fuera de la diagonal se calcula con:

$$I_T^C = \frac{p}{p-1} \times \left(I(B) - \frac{m-p}{p^2} \right) \qquad \textit{Ecuación 22}$$

donde *m* es el número de categorías de las variables, *p* es el número de variables e *I(B)* la inercia total de la matriz de Burt. Con este nuevo total se calculan los porcentajes que corresponden a cada valor propio[30].

- Cada categoría representada en el espacio es el punto medio de todos los individuos que tienen esa característica, ponderado por el coeficiente $1/\sqrt{\lambda}$. El conjunto de categorías de cada variable forma un subespacio con un centro de gravedad *G*. Cada punto-individuo se sitúa en el baricentro de los puntos que representan las categorías que él posee, cada categoría con la misma ponderación. La subnube de puntos que representan las diversas categorías de una variable constituye un subespacio vectorial con el centro de gravedad G. De la misma manera, cada punto-categoría está situado, por afinidades cercanas, en el baricentro de los individuos que tienen esta modalidad.
- La inercia explicada por una categoría es mayor cuanto menor es su efectivo. De ello se deduce que es conveniente evitar que las categorías tengan una frecuencia muy baja. Se puede garantizar, por ejemplo, un mínimo de un 5%, si bien este aspecto debe estar sujeto a criterios de pertenencia conceptual y al tamaño de la muestra, y después de comprobar el comportamiento de la categoría en el conjunto de relaciones que mantiene con el resto de modalidades.
- Por otra parte, la inercia debida a una variable también es mayor cuantas más categorías compongan dicha variable. En este sentido es recomendable disponer de variables con un número equilibrado de categorías. Aquí de nuevo debemos considerar criterios de pertenencia conceptual y hay que observar los efectos (la estabilidad o no) que origina una (re)codificación en más o menos categorías.

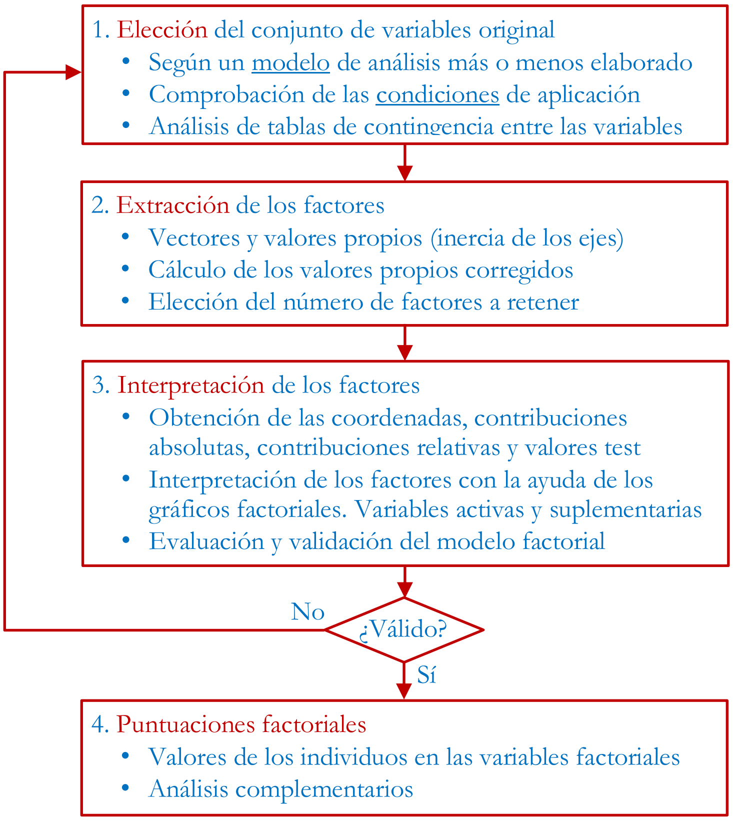
El proceso general de un análisis de correspondencias múltiples lo podemos esquematizar como se muestra en el Gráfico III.11.24. Daremos cuenta de cada etapa del proceso de análisis con un ejemplo de análisis destinado a estudiar la segmentación del mercado de trabajo español.

---

[29] En palabra del autor (Greenacre, 2008:196): *Queda claro que, en la matriz de Burt, la inclusión de las tablas de la diagonal degrada los resultados globales del ACM, ya que intentamos visualizar de forma innecesaria las tablas de la diagonal que, además, presentan inercias muy elevadas. De hecho, las tablas de la diagonal tienen los máximos valores de inercia que se pueden alcanzar. Ignorando las tablas de la diagonal podríamos mejorar el cálculo de la inercia explicada por el mapa.*

[30] En la página web se encuentra en archivo de Excel VPCorregidos.xlsx que mediante una plantilla permite realizar todos los cálculos que hemos comentado, solamente será necesario introducir tres parámetros: la relación de valores propios, el número de variables y el número de categorías. Lo comentaremos más adelante.

Gráfico III.11.24. Etapas de un análisis de correspondencias múltiples



5.2.1. *Elección del conjunto de variables original*

Como siempre es una etapa crucial del análisis definir el modelo de análisis que justifique la selección de las variables originales que definirán los factores. El ejemplo que proponemos busca estudiar el proceso de segmentación del mercado de trabajo de la población asalariada desde la perspectiva del empleo a partir de los datos de la Encuesta de Población Activa correspondientes al cuarto trimestre de 2014[31]. El modelo que aplicamos plantea la existencia de tres dimensiones iniciales que configuran distintas posiciones en el mercado laboral en función del grado de estabilidad de la ocupación, el nivel de cualificación y las características de las empresas[32]. Los indicadores o variables originales que se emplean en el análisis se presentan en la Tabla III.11.16. Se relacionan junto con sus categorías y la distribución de frecuencias.

---

[31] El análisis de segmentación lo concebimos en dos fases siguiendo la lógica metodológica de lo que hemos dado en llamar la construcción estructural y articulada de tipologías (López-Roldán, 1996a) en donde en una primera fase se analizan las variables originales para determinar los factores de diferenciación mediante un análisis factorial, en este caso de correspondencias, y en una segunda, utilizando estos factores o perfiles como criterios clasificatorios, se procede a clasificar los individuos en grupos homogéneos o tipos de la tipología, en este caso tipo de segmentos de una tipología de segmentación laboral. Aquí daremos cuenta de la primera parte factorial y en el capítulo siguiente completaremos el análisis con el proceso clasificatorio.

[32] Una cuarta dimensión es relevante en este análisis, el nivel salarial, pero los datos de la EPA publicados en la web del INE no lo proporciona. Tampoco disponemos de un indicador importante en este tipo de análisis sobre las características de las empresas como es el tamaño de las mismas. No obstante los resultados mostrarán las pautas fundamentales de este tipo de análisis. La matriz EPA4t2014.sav contiene la información original de la encuesta, EPA4t2014S.sba y EPA4t2014S.sav las variables empleadas en el análisis de segmentación.

Tabla III.11.16. Variables activas utilizadas en el ACM

Dimensión	Variables	Frecuencia	%
Estabilidad en el empleo	**Tipo de contrato** (CONTRATO)		
	Indefinido	36.228	13,2
	Temporal	11.608	37,4
	Duración del contrato (DURACION)		
	Indefinido	36.228	75,7
	Hasta 1 mes	5.622	11,8
	Hasta 6 meses	2.921	6,1
	Más de 6 meses	2.602	5,4
	Sin datos[1]	463	1,0
	Tiempo en la empresa (ANTIGUEDAD)		
	Hasta 1 año	8.083	16,9
	2-3 años	5.199	10,9
	4-10 años	13.326	27,9
	11-20 años	10.862	22,7
	Más de 20 años	10.366	21,7
	Tipo de jornada (JORNADA)		
	Completa	39.472	13,2
	Parcial	8.364	37,4
Nivel de cualificación	**Ocupación** (OCUPACION)		
	Directivos	1.188	2,5
	Profesionales	8.590	18,0
	Técnicos apoyo	5.081	10,6
	Administrativos	5.587	11,7
	Trabajador servicios	10.524	22,0
	Cualificado primario	525	1,1
	Cualificado industria	5.069	10,6
	Operadores	3.854	8,1
	Elementales	7.080	14,8
	Sin datos[1]	338	0,7
	Nivel de estudios (EDUCACION)		
	Primaria incompleta	557	1,2
	Primaria	2.596	5,4
	Secundaria 1ª etapa	13.279	27,8
	Secundaria 2ª General	6.176	12,9
	Secundaria 2ª Profesional	4.775	10,0
	Superior	20.453	42,8
Tipo de empresa	**Sector de actividad** (ACTIVIDAD)[2]		
	Primario	1.464	3,1
	Industria 1	2.518	5,3
	Industria 2	2.862	6,0
	Industria 3	2.089	4,4
	Construcción	2.291	4,8
	Comercio	10.034	21,0
	Transporte-Comunicaciones	3.283	6,9
	Financiero-Profesionales	5.588	11,7
	Administración	14.061	29,4
	Otros servicios	3.646	7,6
	Titularidad (EMPRESA)		
	Pública	39.472	13,2
	Privada	8.364	37,4
	Total	**47.836**	**100,0**

Fuente: INE. Encuesta de Población Activa, 4º trimestre de 2014.

[1] Categorías con asignación aleatoria por baja frecuencia (1% o menos).

[2] La agrupación de sectores se puede consultar en la información metodológica de la EPA. En particular, Industria 1 es: alimentación, textil, cuero, madera y papel; Industria 2: extractivas, refino de petróleo, industria química, farmacéutica, industria del caucho y materias plásticas, e Industria 3: construcción de maquinaria, equipo eléctrico y material de transporte. Instalación y reparación industrial.

Junto a las variables activas se han considerado tres variables suplementarias: edad, sexo y nacionalidad, que tienen un papel ilustrativo de ayuda a la interpretación de los factores, a modo de variables independientes que se relacionan con la estructura emergente de relaciones entre las variables originales del análisis factorial que actuaría como resultado o "variable dependiente" (Tabla III.11.17).

Tabla III.11.17. Variables suplementarias del ACM

Variables	Frecuencia	%
Edad (EDAD)		
16 a 19 años	229	0,5
20 a 24 años	2.045	4,3
25 a 29 años	4.232	8,8
30 a 34 años	5.250	11,0
35 a 39 años	7.224	15,1
40 a 44 años	7.274	15,2
45 a 49 años	7.314	15,3
50 a 54 años	6.766	14,1
55 a 59 años	5.001	10,5
60 a 64 años	2.501	5,2
Sexo (SEXO)		
Varón	24.356	50,9
Mujer	23.480	49,1
Nacionalidad (NACIONALIDAD)		
Española	39.472	92,3
Doble nacionalidad	5.081	2,1
Extranjera	5.587	5,6
Total	**47.836**	**100,0**

Fuente: INE. Encuesta de Población Activa, 4º trimestre de 2014.

En la selección de las variables y su categorización, además del criterio teórico-conceptual, es necesario tener en cuenta el porcentaje de casos de cada categoría. Como hemos señalado, una frecuencia muy baja puede tener como consecuencia la acumulación "en exceso" de una parte importante de la inercia total y tener un protagonismo que no merece, derivado no tanto de las características del fenómeno estudiado como por efecto de la aplicación de la técnica. Por ello, para garantizar la robustez y estabilidad del análisis, es conveniente analizar los efectos que una determinada codificación de nuestras categorías puede tener en la configuración de los factores del ACM. Como criterio general podemos contemplar categorías con una frecuencia relativa mínima no inferior al 5%, no obstante, si la muestra es importante, este valor puede ser inferior.

Adicionalmente existe la posibilidad de tratar las categorías con muy baja frecuencia, por debajo del 2%, ya provengan de los mismos valores de la variable como de la ausencia de información en categorías como "no sabe", "no contesta" o "no pertinente" que se suelen tratar como valores perdidos. El tratamiento consiste en la asignación de un valor a esas categorías que neutralice su posible efectivo disperso y particular o, si se trata de valores perdidos, nos obligue a eliminar del análisis los casos correspondientes. No obstante, siempre podemos excluir esos casos del análisis si no son muchos. Si por el contrario queremos mantenerlos podemos optar por la asignación de un valor determinado mediante, de hecho, un ejercicio de imputación de un valor que nos posibilita que la tabla de datos conserve su propiedad de codificación disyuntiva completa. Se puede optar por atribuir al azar a los casos que tienen una categoría muy poco frecuente a una de las categorías de la variable con frecuencia suficiente de tal manera que resulte un comportamiento de la categoría en cuestión como la media de todos los casos (como si fuera un punto en el centro del gráfico factorial que se obtiene)[33], o bien se puede optar por asignar un determinado valor como la moda o el valor de una categoría específica[34].

En el proceso de selección de las variables y su categorización es recomendable igualmente efectuar un análisis de las relaciones bivariables entre las diferentes variables mediante tablas de contingencia. Además de poder realizar un análisis descriptivo parcial que ayuda a ir construyendo una imagen progresiva de las interrelaciones entre las distintas variables y la construcción de una primera representación multidimensional, podemos utilizarlo para tener en cuenta un criterio de codificación basado en criterios estadísticos. Se trata de aplicar el criterio exigido en las tablas de contingencia cuando se realiza la prueba de independencia de chi-cuadrado, es decir, que la frecuencia mínima esperada por casilla sea 1 y 5 en el 80%. Con ello garantizamos una mínima y suficiente distribución de frecuencias entre las diversas casillas o, en caso contrario, nos sugiere la necesidad de agrupar las categorías con baja frecuencia que generan esta situación.

Finalmente destacaremos que en el ACM, al igual que en el ACS, existe la posibilidad de diferenciar las variables activas, las que generan los factores, de las suplementarias o ilustrativas, aquellas que pueden relacionarse con la solución factorial posicionándose en el espacio factorial para ayudar a interpretar los resultados pero sin intervenir en la configuración de los factores (Tabla III.11.17). Junto a las variables suplementarias se pueden tratar también como suplementarias una o más categorías de una variable o incluso a los propios individuos. Decidir qué variables y categorías cumplen el papel activo o suplementario es una decisión que corresponde a esta fase inicial de configuración de nuestro modelo de análisis concreto y operativo.

La selección de las variables no es una tarea simple y va a exigir la realización de diversos ejercicios de análisis de correspondencias para acabar de encontrar tanto las variables que mejor expresan el modelo de análisis como la categorización más adecuada. Se trata, como se destaca en el esquema del Gráfico III.11.24, de volver a esta etapa con nuevas propuestas de revisión de las variables consideradas inicialmente.

5.2.2. *Extracción de los factores*

Una vez decididas las variables iniciales se procede a construir la tabla de correspondencias, es decir, la matriz de Burt, para a continuación aplicar la técnica y obtener los vectores propios (factores) y valores propios (la inercia explicada por los factores). Con los factores generados se trata de decidir el número de factores que retendremos. Ya hemos señalado que en esa tarea es necesario efectuar un cálculo que transforme los valores propios iniciales en unos nuevos valores propios corregidos que expresen el peso real que tienen los factores. La Tabla III.11.18 recoge esta información.

En número total de factores que se pueden obtener en este análisis es el número de categorías menos el de variables. Inicialmente disponemos de 42 categorías (con

[33] Estrategia que sigue el software SPAD.
[34] Estrategia que sigue el software SPSS.

frecuencia mayor que 0) asociadas a las 8 variables elegidas. No obstante, dos de ellas, por su baja frecuencia, serán neutralizadas mediante una asignación aleatoria para garantizar la robustez del análisis[35]. Por tanto, en este caso, tras ventilar las dos categorías de baja frecuencia "sin datos" señaladas en la Tabla III.11.16, el número de dimensiones posibles son $m-p = 40-8 = 32$.

A partir de los 32 valores propios iniciales seleccionamos aquellos que son superiores a $1/p$, es decir, $1/8=0{,}125$, un total de 13, y aplicamos la transformación de Benzécri de la *Ecuación 22*. En la Tabla III.11.18 vemos la distribución del nuevo porcentaje de varianza explicada por los 13 primeros factores.

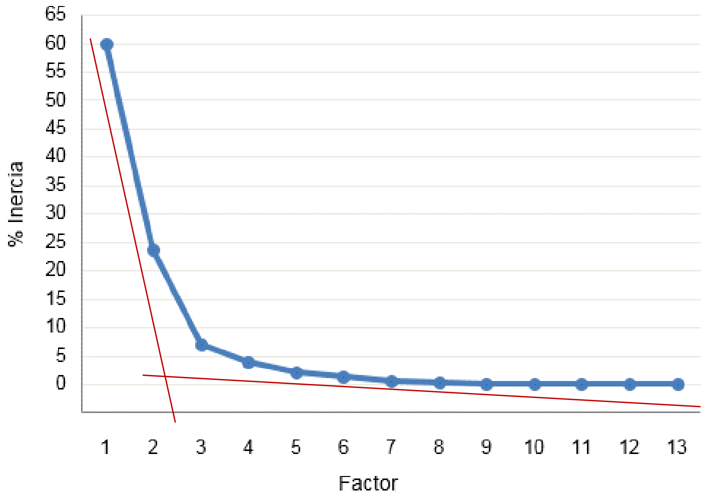
Tabla III.11.18. Tabla de valores propios y de inercia (varianza) explicada del ACM. Valores iniciales y corregidos

	Valores propios iniciales			Valores propios corregidos		
Factor	Valor	% de la varianza	% acumulado	Valor	% de la varianza	% acumulado
1	0,377	9,4%	9,4	**0,083**	**59,9%**	**59,9%**
2	0,284	7,1%	16,5	**0,033**	**23,8%**	**83,8%**
3	0,212	5,3%	21,8	0,010	7,2%	90,9%
4	0,190	4,8%	26,6	0,006	4,0%	95,0%
5	0,174	4,3%	30,9	0,003	2,3%	97,2%
6	0,164	4,1%	35,0	0,002	1,4%	98,6%
7	0,152	3,8%	38,8	0,001	0,7%	99,3%
8	0,147	3,7%	42,5	0,001	0,5%	99,8%
9	0,136	3,4%	45,9	0,000	0,1%	99,9%
10	0,133	3,3%	49,2	0,000	0,1%	100,0%
11	0,129	3,2%	52,5	0,000	0,0%	100,0%
12	0,128	3,2%	55,7	0,000	0,0%	100,0%
13	0,126	3,1%	58,8	0,000	0,0%	100,0%
14	0,124	3,1%	61,9			
15	0,124	3,1%	65,0			
16	0,123	3,1%	68,1			
17	0,121	3,0%	71,1			
18	0,118	3,0%	74,1			
19	0,117	2,9%	77,0			
20	0,115	2,9%	79,9			
21	0,111	2,8%	82,6			
22	0,100	2,5%	85,2			
23	0,096	2,4%	87,6			
24	0,088	2,2%	89,8			
25	0,085	2,1%	91,9			
26	0,083	2,1%	94,0			
27	0,069	1,7%	95,7			
28	0,055	1,4%	97,1			
29	0,049	1,2%	98,3			
30	0,041	1,0%	99,3			
31	0,027	0,7%	100,0			
32	0,001	0,0%	100,0			
Total	**4,000**	**100,0**		**0,139**	**100,0**	

[35] Los resultados del ACM obtenidos con el software SPAD fijan un criterio por defecto del 2% para tratar de esta forma las categorías con muy baja frecuencia. Aquí hemos fijado el umbral en el 1% por disponer de una muestra con un tamaño muestral importante.

Para decidir cuántos factores retener siempre consideraremos como prioritario el criterio sustantivo y de interpretabilidad de los factores. El contenido de los mismos lo veremos en seguida. Junto a éste consideraremos un criterio cuantitativo fijando como mínimo alcanzar el 70% de la varianza explicada. Adicionalmente estos criterios se pueden complementar con el comportamiento del gráfico de sedimentación (Gráfico III.11.25) donde el cruce de las dos tendencias de la curva marca el número de factores a retener. Por todos estos criterios el número de factores que se han considerado ha sido dos. Ambos acumulan el 84% de la varianza explicada.

Gráfico III.11.25. Gráfico de sedimentación del ACM



5.2.3. *Interpretación de los factores*

La información sobre los factores obtenidos en un ACM es similar a la que vimos con el ACS: coordenadas de las categorías, contribuciones absolutas, contribuciones relativas y valores test nos ayudan a dar identidad a los factores e interpretarlos con ayuda inestimable de los gráficos factoriales.

La Tabla III.11.19 recoge las coordenadas de las categorías, las contribuciones absolutas y las contribuciones relativas de las variables activas. La Tabla III.11.20 presenta los valores test de las variables activas e suplementarias.

Como en el ACS, en la Tabla III.11.19 aparece en primer lugar el peso relativo (la frecuencia relativa de cada categoría sobre el total de casos válidos, en porcentaje) que nos valora la importancia de cada categoría y la distancia de cada categoría al origen, al promedio, indicando qué tan dispersa o periférica es la categoría, sabiendo que cuanto mayor sea este valor más inercia acumulará y más determinante dar el carácter de los factores. A diferencia del ACS, la información de las categorías no se divide en filas y columnas, se considera el conjunto de todas ellas en una misma tabla. Las coordenadas o factores de carga nos permiten representar gráficamente las categorías de las variables en el espacio factorial indicando la posición que ocupan para cada eje retenido.

Tabla III.11.19. Coordenadas, contribuciones absolutas y relativas de las variables activas del ACM.

Categoría	Peso relativo	Distancia al origen	Coordenadas Eje 1	Coordenadas Eje 2	Contribuciones absolutas Eje 1	Contribuciones absolutas Eje 2	Contribuciones relativas Eje 1	Contribuciones relativas Eje 2
Tipo de contrato								
Indefinido	9,5	0,32	-0,41	0,33	4,3	3,6	0,54	0,34
Temporal	3,0	3,12	1,29	-1,03	13,5	11,3	0,54	0,34
					17,8	**14,9**		
Duración del contrato								
Indefinido	9,5	0,32	-0,41	0,33	4,2	3,6	0,53	0,34
Hasta 1 mes	1,5	7,34	1,39	-0,87	7,7	4,0	0,26	0,10
Hasta 6 meses	0,8	14,68	1,49	-0,97	4,7	2,6	0,15	0,06
Más de 6 meses	0,7	16,65	0,89	-1,47	1,5	5,4	0,05	0,13
					18,1	**15,5**		
Tiempo en la empresa								
Hasta 1 año	2,1	4,92	1,39	-0,81	10,8	4,8	0,39	0,13
2-3 años	1,4	8,20	0,60	0,04	1,3	0,0	0,04	0,00
4-10 años	3,5	2,59	-0,07	0,30	0,0	1,1	0,00	0,03
11-20 años	2,8	3,40	-0,41	0,34	1,3	1,1	0,05	0,03
Más de 20 años	2,7	3,61	-0,86	-0,12	5,3	0,1	0,20	0,00
					18,7	**7,2**		
Tipo de jornada								
Completa	10,3	0,21	-0,18	0,05	0,9	0,1	0,15	0,01
Parcial	2,2	4,72	0,84	-0,24	4,1	0,5	0,15	0,01
					4,9	**0,5**		
Ocupación								
Directivos	0,3	38,11	-0,68	0,35	0,4	0,1	0,01	0,00
Profesionales	2,3	4,55	-0,88	-1,18	4,7	11,1	0,17	0,31
Técnicos apoyo	1,3	8,34	-0,33	0,10	0,4	0,0	0,01	0,00
Administrativos	1,5	7,50	-0,37	0,11	0,5	0,1	0,02	0,00
Trabajador servicios	2,8	3,53	0,25	0,06	0,5	0,0	0,02	0,00
Cualificado primario	0,1	83,52	0,64	0,11	0,2	0,0	0,00	0,00
Cualificado industria	1,3	8,36	0,39	0,66	0,5	2,1	0,02	0,05
Operadores	1,0	11,29	0,19	0,90	0,1	2,9	0,00	0,07
Elementales	1,9	5,72	0,90	0,15	4,0	0,2	0,14	0,00
					11,3	**16,5**		
Nivel de estudios								
Primaria incompleta	0,1	84,88	0,89	0,32	0,3	0,1	0,01	0,00
Primaria	0,7	17,43	0,68	0,46	0,8	0,5	0,03	0,01
Secundaria 1ª etapa	3,5	2,60	0,53	0,49	2,6	3,0	0,11	0,09
Secundaria 2ª General	1,6	6,75	-0,03	0,22	0,0	0,3	0,00	0,01
Secundaria 2ª Profesional	1,2	9,02	0,27	0,25	0,2	0,3	0,01	0,01
Superior	5,3	1,34	-0,51	-0,51	3,7	4,9	0,19	0,20
					7,6	**9,0**		
Sector de actividad								
Primario	0,4	31,67	1,38	-0,14	1,9	0,0	0,06	0,00
Industria 1	0,7	18,00	0,25	0,85	0,1	1,7	0,00	0,04
Industria 2	0,7	15,71	-0,04	0,78	0,0	1,6	0,00	0,04
Industria 3	0,5	21,90	-0,04	0,66	0,0	0,8	0,00	0,02
Construcción	0,6	19,88	0,68	0,38	0,7	0,3	0,02	0,01
Comercio	2,6	3,77	0,48	0,42	1,6	1,6	0,06	0,05
Transporte-Comunicaciones	0,9	13,57	-0,11	0,37	0,0	0,4	0,00	0,01
Financiero-Profesionales	1,5	7,56	0,06	0,29	0,0	0,4	0,00	0,01
Administración	3,7	2,40	-0,80	-1,01	6,2	13,2	0,27	0,42
Otros servicios	1,0	12,12	0,69	0,22	1,2	0,2	0,04	0,00
					11,8	**20,2**		
Titularidad								
Pública	3,0	3,21	-0,97	-1,09	7,4	12,3	0,29	0,37
Privada	9,5	0,31	0,30	0,34	2,3	3,8	0,29	0,37
					9,8	**16,2**		
Total	100,0				**100,0**	**100,0**		

Las contribuciones absolutas son la principal información para identificar el contenido de los factores. Cada valor de la tabla es el porcentaje de cada categoría *i* con la que contribuye al total de la inercia explicada por el eje factorial *k*. Si todas las contribuciones fueran iguales cada contribución sería el 100% dividido por el número de categorías (100 entre 40 da 2,5%), por lo que las categorías con contribuciones superiores a este valor son las que tienen una influencia superior a la media. Por otra parte, la suma de las contribuciones de las categorías de una variable constituye la contribución de dicha variable.

En nuestro ejemplo observamos cómo la primera dimensión se configura por las contribuciones de la contratación temporal en el polo positivo frente a la contratación indefinida en el polo negativo. La eventualidad de la contratación vemos que se asocia con los contratos de duración determinada y con llevar poco tiempo en la empresa, así como con jornadas de trabajo parciales. Si nos fijamos en las variables de cualificación, la polaridad de contratación indefinida se corresponde en mayor medida con las ocupaciones de profesionales y niveles de estudios universitarios, mientras la contratación temporal se asocia principalmente con las categorías ocupacionales más elementales. El tipo de empresa aparece en este primer factor para destacar la contratación indefinida que en mayor medida se da en el sector público. Podemos hablar por tanto de una primera dimensión que expresa la segmentación laboral asociada doblemente a la estabilidad y a la cualificación.

La segunda dimensión está protagonizada específicamente por el perfil de trabajador/a de la administración pública con altos niveles de cualificación, son técnicos y profesionales con niveles de estudios superiores. Junto a este perfil que ocupa el polo negativo de la dimensión se ubican también las categorías de la ocupación temporal configurando así un doble perfil en este extremo. El resto de las categorías ocupa de forma generalizada valores moderados en la polaridad positiva, o próximos a cero, en general del sector privado, y donde podemos destacar el perfil de trabajador estable de la industria con cualificaciones medias-bajas. Se dibuja pues un segunda dimensión de sector público y privado.

Las contribuciones relativas informan sobre la participación cada categoría en los diferentes ejes, siendo la suma de las contribuciones relativas, para todos los factores, igual a la unidad (suma de todas las columnas para todos los factores posibles). En la Tabla III.11.19 podemos ver que las categorías de indefinido, temporal, público, privado, profesionales y administración son las mejor aparecen reflejadas en el conjunto de los dos primeros ejes factoriales que hemos conservado.

Por su parte la tabla de los valores-test equivalen a las coordenadas transformadas en puntuaciones típicas *z*. Por lo tanto, todo valor que no se encuentre dentro del intervalo $(-1{,}96\ ,\ 1{,}96)$ será significativo, es decir, la categoría en cuestión para el eje factorial considerado será importante en amyor o menor grado en la interpretación o caracterización del eje, siendo su contribución absoluta y relativa de mayor o menor intensidad según se aleje más o menos de este intervalo. Este valor-test es especialmente útil para valorar los individuos y las variables adicionales o suplementarios.

Tabla III.11.20. Coordenadas y valores test de las variables activas y suplementarias del ACM.

Variables activas

Categoría	Frecuencia	Distancia al origen	Coordenadas		Valores test	
			Eje 1	Eje 2	Eje 1	Eje 2
Tipo de contrato						
Indefinido	36.228	0,32	-0,41	0,33	-99,90	99,90
Temporal	11.608	3,12	1,29	-1,03	99,90	-99,90
Duración del contrato						
Indefinido	36.228	0,32	-0,41	0,33	-99,90	99,90
Hasta 1 mes	5.622	7,34	1,39	-0,87	99,90	-69,20
Hasta 6 meses	2.921	14,68	1,49	-0,97	83,29	-53,95
Más de 6 meses	2.602	16,65	0,89	-1,47	45,66	-77,80
Tiempo en la empresa	463					
Hasta 1 año		4,92	1,39	-0,81	99,90	-79,43
2-3 años	8.083	8,20	0,60	0,04	45,92	2,71
4-10 años	5.199	2,59	-0,07	0,30	-9,98	40,30
11-20 años	13.326	3,40	-0,41	0,34	-48,89	39,81
Más de 20 años	10.862	3,61	-0,86	-0,12	-98,62	-14,14
Tipo de jornada						
Completa	39.472	0,21	-0,18	0,05	-84,26	24,40
Parcial	8.364	4,72	0,84	-0,24	84,26	-24,40
Ocupación						
Directivos	1.188	38,11	-0,68	0,35	-23,15	13,10
Profesionales	8.590	4,55	-0,88	-1,18	-90,29	-99,90
Técnicos apoyo	5.081	8,34	-0,33	0,10	-24,14	7,91
Administrativos	5.587	7,50	-0,37	0,11	-29,38	9,16
Trabajador servicios	10.524	3,53	0,25	0,06	29,90	6,76
Cualificado primario	525	83,52	0,64	0,11	17,18	3,96
Cualificado industria	5.069	8,36	0,39	0,66	30,34	50,45
Operadores	3.854	11,29	0,19	0,90	13,12	59,00
Elementales	7.080	5,72	0,90	0,15	83,04	14,48
Nivel de estudios						
Primaria incompleta	557	84,88	0,89	0,32	21,22	7,68
Primaria	2.596	17,43	0,68	0,46	35,40	24,05
Secundaria 1ª etapa	13.279	2,60	0,53	0,49	71,87	66,94
Secundaria 2ª General	6.176	6,75	-0,03	0,22	-2,70	18,45
Secundaria 2ª Profesional	4.775	9,02	0,27	0,25	19,69	18,17
Superior	20.453	1,34	-0,51	-0,51	-95,97	-96,78
Sector de actividad						
Primario	1.464	31,67	1,38	-0,14	53,61	-5,49
Industria 1	2.518	18,00	0,25	0,85	12,88	43,65
Industria 2	2.862	15,71	-0,04	0,78	-2,40	42,89
Industria 3	2.089	21,90	-0,04	0,66	-1,70	31,07
Construcción	2.291	19,88	0,68	0,38	33,24	18,56
Comercio	10.034	3,77	0,48	0,42	53,55	46,96
Transporte-Comunicaciones	3.283	13,57	-0,11	0,37	-6,35	21,68
Financiero-Profesionales	5.588	7,56	0,06	0,29	4,87	23,03
Administración	14.061	2,40	-0,80	-1,01	-99,90	-99,90
Otros servicios	3.646	12,12	0,69	0,22	43,09	13,51
Titularidad						
Pública	11.368	3,21	-0,97	-1,09	-99,90	-99,90
Privada	36.468	0,31	0,30	0,34	99,90	99,90
Total	47.836					

Variables suplementarias

Categoría	Frecuencia	Distancia al origen	Coordenadas		Valores test	
			Eje 1	Eje 2	Eje 1	Eje 2
Edad						
16 a 19 años	229	207,89	1,64	-0,64	24,86	-9,67
20 a 24 años	2.045	22,39	1,13	-0,52	52,34	-24,18
25 a 29 años	4.232	10,30	0,59	-0,28	40,39	-19,21
30 a 34 años	5.250	8,11	0,21	0,02	16,44	1,56
35 a 39 años	7.224	5,62	0,04	0,05	3,48	5,02
40 a 44 años	7.274	5,58	-0,06	0,10	-6,01	9,63
45 a 49 años	7.314	5,54	-0,17	0,09	-15,64	8,40
50 a 54 años	6.766	6,07	-0,29	0,03	-25,72	2,47
55 a 59 años	5.001	8,57	-0,39	0,03	-28,97	1,91
60 a 64 años	2.501	18,13	-0,40	0,07	-20,42	3,60
Sexo						
Varón	24.356	0,96	0,02	0,18	3,83	39,88
Mujer	23.480	1,04	-0,02	-0,19	-3,83	-39,88
Nacionalidad						
Española	44.166	0,08	-0,06	-0,01	-42,47	-10,98
Doble nacionalidad	981	47,76	0,43	0,21	13,66	6,77
Extranjera	2.689	16,79	0,76	0,16	40,67	8,53
Total	47.836					

Si observamos las categorías de las variables suplementarias podemos apreciar cómo de forma sistemática las distintas edades se ubican a lo largo del primer eje desde las coordenadas positivas más alejadas del centro, las que corresponden a los más jóvenes, hasta los valores negativos más alejados de las edades más maduras. El origen inmigrante diferencia también a los autóctonos en posiciones centrales frente a las personas de nacionalidad extranjera o doble nacionalidad que se ubican primordialmente en el lado positivo de la precariedad del empleo. El sexo no marca diferencias en esta primera dimensión, por lo que cabe concluir que la estabilidad-inestabilidad del primer afecta por igual a ambos sexos. De forma modesta en la segunda dimensión las mujeres tienden a situarse en valores negativos por su mayor presencia en la administración pública mientras que los varones tienden a alejarse hacia valores positivos por su mayor presencia en los sectores industriales.

Finalmente podemos observar las representaciones de los gráficos factoriales en los dos primeros ejes donde se reflejan los distintos comentarios realizados en relación a la información de las tablas. En el Gráfico III.11.26 se representan las variables activas y en el Gráfico III.11.27 las suplementarias. Podrían haberse representado simultáneamente, pero en aras de una mejor visualización de las categorías en el espacio se ha preferido presentarlas por separado, pero nos podemos imaginar perfectamente las correspondencias entre uno y otro. El Gráfico III.11.28 posiciona los individuos en el espacio factorial lo que implica cuantificar o asignar un valor a cada individuo según cada una de las dos variables factoriales, caracterizadas por tener ser linealmente independientes entre sí, tener una métrica continua y estar tipificadas (tienen media es cero y la desviación típica uno). Estas variables factoriales se pueden utilizar para realizar análisis complementarios, gráficos con el caso anterior o utilizando estas variables con otras técnicas. En particular este será el caso de la metodología que seguiremos al combinar el análisis factorial con el análisis de clasificación con el objetivo de la construcción de tipologías: las variables factoriales se considerarán como las variables-criterio con las que clasificar las unidades en grupos homogéneos.

Gráfico III.11.26. Gráfico factorial del ACM. Variables activas

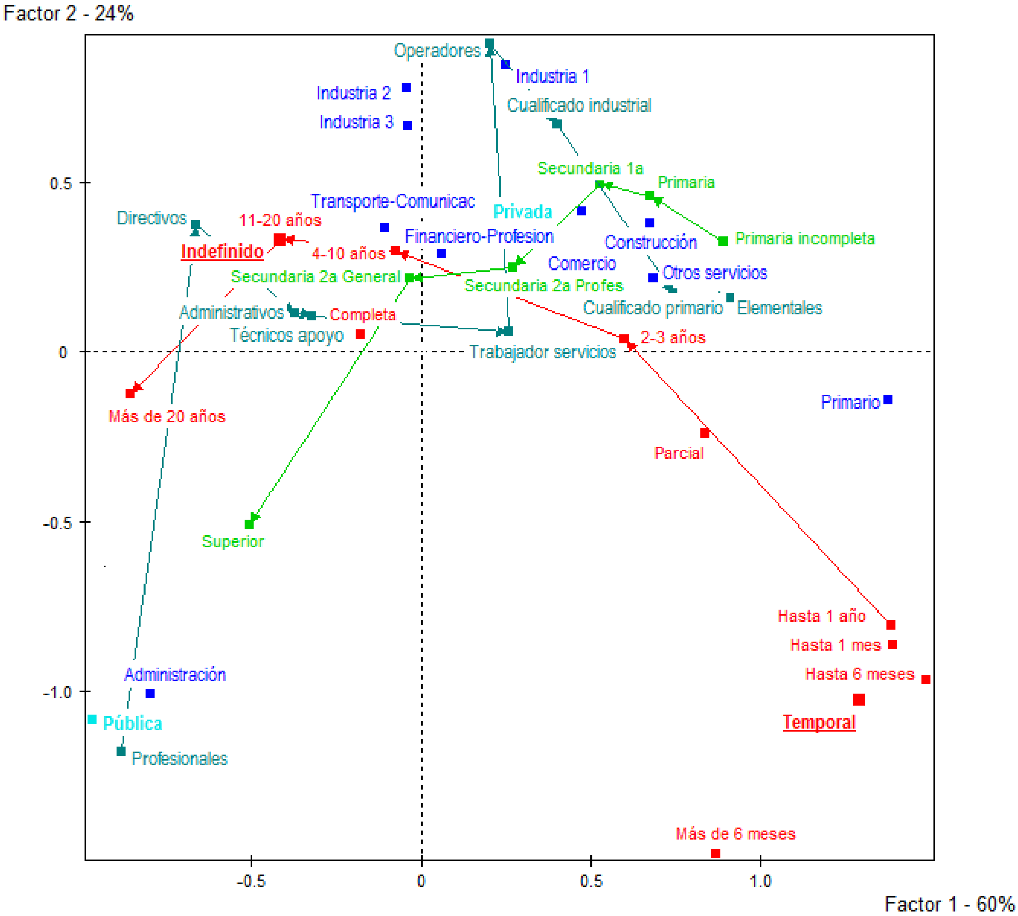


Gráfico III.11.27. Gráfico factorial del ACM. Variables suplementarias

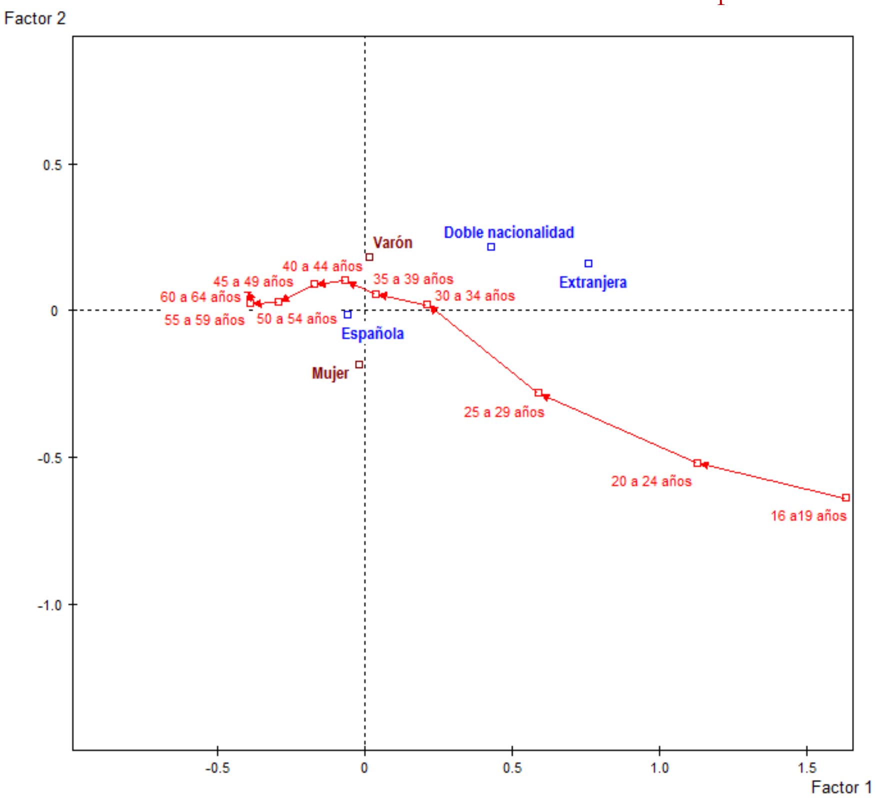
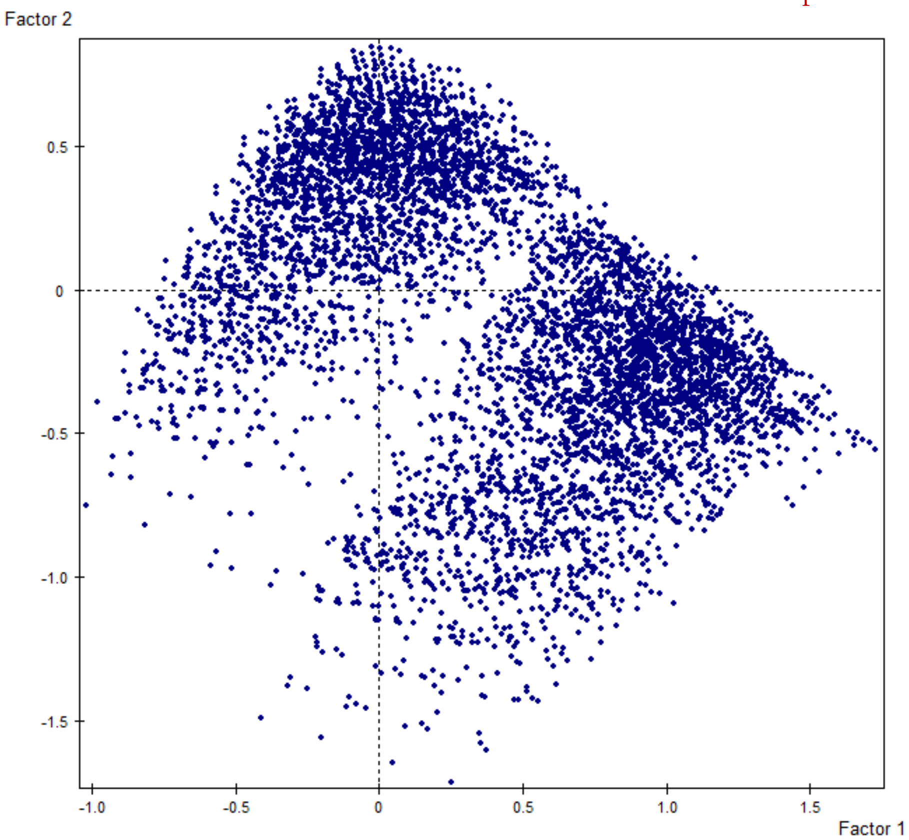


Gráfico III.11.28. Gráfico factorial del ACM. Individuos en el espacio de los factores



5.2.4. *Ejemplos de aplicación del ACM*

En este apartado presentaremos algunos ejemplos que complementarán el que acabamos de ver para así profundizar y mejorar el conocimiento y la capacidad interpretativa de los resultados de la técnica.

5.2.4.1. *Bourdieu, el espacio social y el gusto probable*

El ACM fue desarrollado por la Escuela Francesa de Análisis de Datos y son numerosos los estudios sociológicos que han utilizado esta metodología. Bourdieu cooperó con el Instituto Nacional de Estadísticas y de Estudios Económicos (INSEE) francés en las encuestas de fuerza laboral durante los años 60. Esa colaboración continuó en el Centro Europeo de Sociología dirigido por Raymond Aron. Hacia finales de los sesenta Bourdieu y Darbel encuentran la técnica de regresión limitada para sus análisis estructurales, pues consideran que la compleja estructura social de interrelaciones no puede ser reducida a una combinación de "efectos puros" de variables independientes (Lebaron, 1991: 6). Es así que se inclinan hacia el Análisis de Correspondencias.

Un trabajo clásico desarrollado por Pierre Bourdieu y sus colegas le dio mucha visibilidad a la técnica (Rouanet, Ackermann y Le Roux, 2001). Concretamente a partir de la obra La Distinción de 1979 donde se analizan los gustos y los estilos de vida de la burguesía francesa. No obstante fue la obra L'anatomie du goût (Bourdieu y Saning-Martin, 1976) la primera publicación donde Bourdieu aplica el Análisis Geométrico de Datos (Lebaron, 1991: 7).

Bourdieu en términos teóricos utiliza conceptos, ahora clásicos, que sin embargo tienen una vinculación muy estrecha con la metodología utilizada, como espacio social, habitus, estilo de vida y gusto.

El concepto de Espacio Social lo utiliza para referirse a las diferencias de clases (y de fracciones). Es el espacio que distribuye o diferencia a aquellos que están mejor provistos de capital económico y cultural de aquellos que están menos provistos. El espacio social es una representación abstracta, un mapa, para comprender la realidad social. El Habitus no sólo es una estructura estructurante, la cual organiza prácticas y la percepción sobre las prácticas, sino también una estructura estructurada: el principio de división entre la lógica de clases la cual organiza la percepción del mundo social. Es en sí mismo el producto de internalización de la división entre clases sociales.

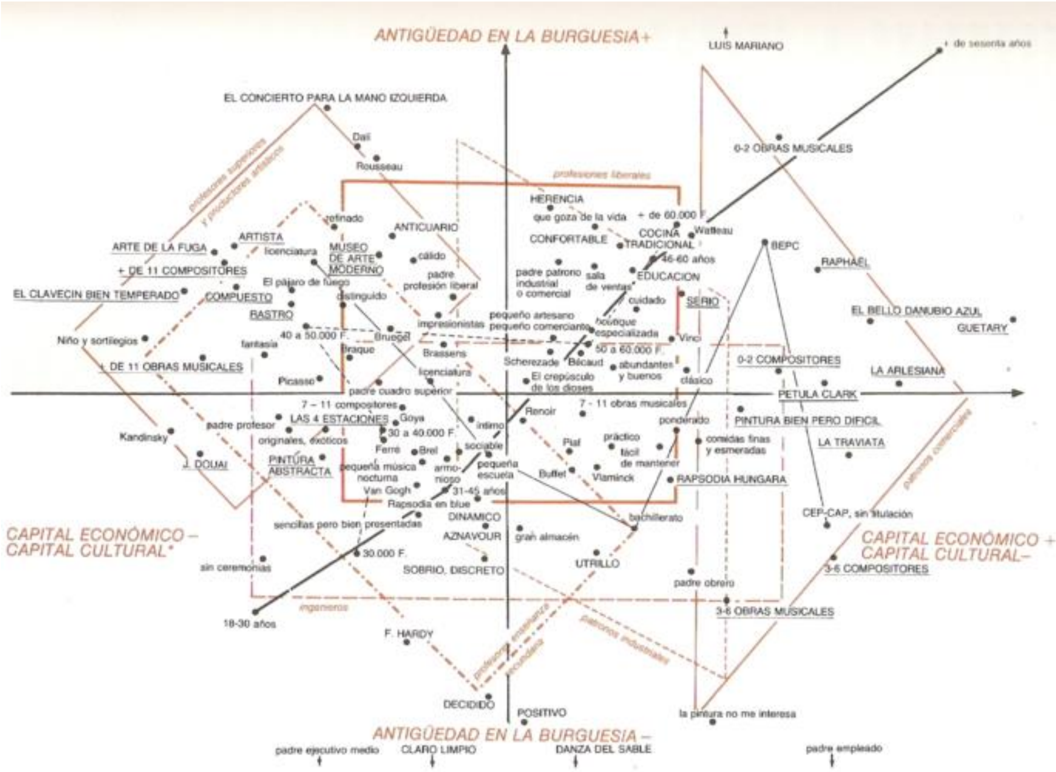
Los Estilos de vida son productos sistemáticos del habitus, que son percibidos en su relación mutua, a través de esquemas de hábitos. Se vuelven sistemas de signos que son socialmente calificados (distinguido, vulgar, etc.). Por su parte el Gusto es el operador práctico de la transmutación de cosas en signos distintos y distintivos, de una distribución continua en oposiciones discontinuas. Las diferencias inscritas a nivel físico se trasladan al orden simbólico, se trata de distinciones significantes.

Los datos que utilizó en el estudio de la distinción fueron recolectados a través de una encuesta realizada con dos muestras complementarias utilizando el mismo cuestionario con el objetivo de brindar una visión sintética del espacio social francés como una estructura global y estudiar en él dos subsectores en mayor profundidad: el espacio de la clase dominante y el espacio de la clase media (pequeña burguesía) (Lebaron, 1991: 7).

El tipo de preguntas que se realizan en la encuesta sobre gustos es muy amplio. Concretamente para el análisis de la clase dominante se realiza una selección de variables que serán utilizadas en el análisis como variables activas e ilustrativas. Entre las variables activas figuran las propiamente relacionadas con el gusto (entre paréntesis indicamos el número de categorías): Decoración del hogar (12), Amigos (12), Platos que se sirven a los amigos (6), Estilos de muebles (6), Cantantes preferidos (12), Obras de música clásica (15), Visita a museos (4) y Pintura (5). Entre las variables ilustrativas se encuentran las sociodemográficas y de origen social como la edad del encuestado, el nivel de instrucción, los ingresos y la profesión del padre.

La matriz de datos que se analiza de individuos y propiedades permite analizar a través del análisis de correspondencias dos nubes de puntos: la nube de los individuos y la nube de las propiedades. La interpretación se basa en un estudio conjunto de las dos nubes (Rouanet, Ackermann y Le Roux, 2001: 139). La primera impresión de los resultados obtenidos que brinda el gráfico presentado en la página 259 de la edición en castellano de La Distinción en 1988, puede resultar algo confuso. Se trata de la superposición de dos gráficos que representan las variantes del gusto dominante. Un gráfico representa el espacio de las propiedades (variables) plasmados en los ejes de inercia 1 y 2, y el otro gráfico (superpuesto) representa el espacio de los individuos de las diferentes fracciones que es delimitado por figuras triangulares o cuadradas.

Gráfico de las variantes del gusto dominante en el espacio social de los factores 1 y 2



Fuente: P. Bourdieu, La Distinción (1988: 259)

En el gráfico se diferencia la importancia de las contribuciones absolutas de las variables en los factores utilizando letras mayúsculas subrayadas para mostrar la mayor contribución al primer factor y las mayúsculas sin subrayado para mostrar las mayores contribuciones al segundo factor. A continuación describimos la interpretación general de los factores encontrados en el análisis seguido por la explicación que el propio Bourdieu realiza en La Distinción sobre la propia dimensión de análisis y aspectos del gusto asociados a ella.

El factor 1 presenta la distribución de capital, económico y cultural. Así el extremo derecho representa mayor capital económico y menor capital cultural, situación que expresa lo opuesto en el extremo izquierdo.

> Se trata de *"conjuntos coherentes de preferencias que encuentran su principio en unos sistemas de disposiciones distintos y distintivos, definidos tanto por la relación que mantienen entre ellos como por la relación que les une con sus condiciones sociales de producción. Los indicadores que miden el capital cultural (del que ya se sabe que varía más o menos en razón inversa de los indicadores del capital económico) aportan la contribución más alta a la constitución del primer factor (que resume el 5,8% de la inercia total contra el 3,5% y el 3,2% para el segundo y el tercero respectivamente)": se tiene así, de una parte, a aquellos que, teniendo los ingresos más débiles, poseen la competencia más alta; que conocen el mayor número de obras musicales (6%) y de compositores (7,7%), dicen preferir las obras que requieren las más "pura" disposición*

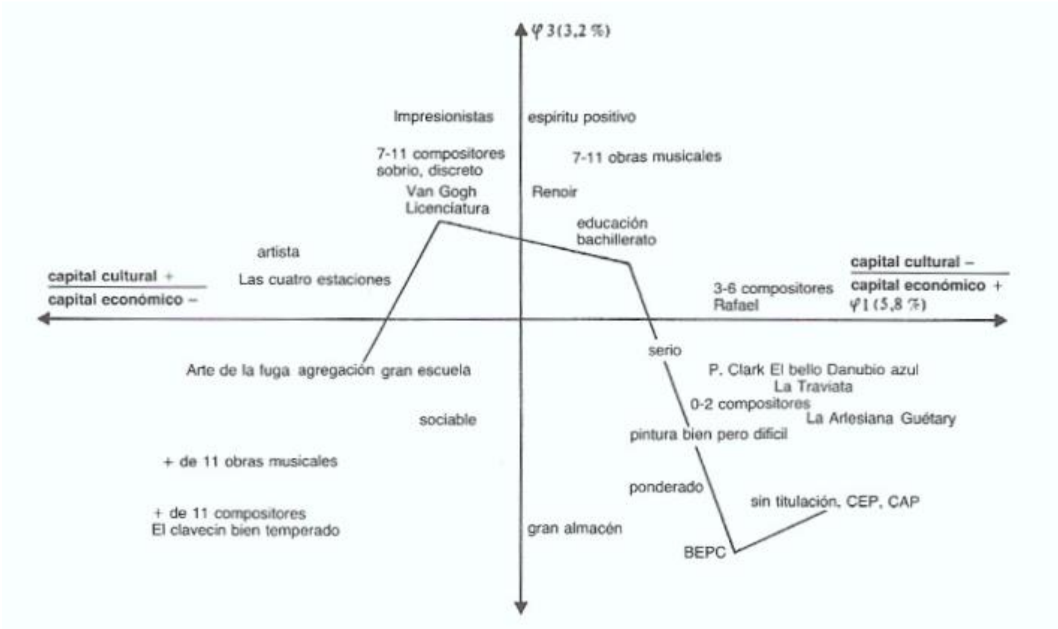
> *estética, como El clavecín bien temperado (1,8 %) o el Arte de la fuga (1,7 %), y la más general, la más capaz de aplicarse a unos campos menos consagrados, como la canción y el cine, o incluso a la cocina o a la decoración de la casa; que se interesan por la pintura abstracta, frecuentan el Museo de arte moderno y esperan de sus amigos que sean artistas (2,4%); por la parte contraria, los que disponen de los ingresos más altos y poseen la competencia más baja, conocen pocas obras musicales y pocos compositores, prefieren amigos serios (1,5%) y dedican sus preferencias a obras de cultura burguesa de segunda fila, depreciadas o clásicas -La Arlesiana (3%), El Bello Danubio azul (2,9%), La Traviata (2,1%), Rapsodia húngara, Buffet, Vlaminck, Utrillo, Raphaél (2,3%), Watteau, Vinci- y a la opereta -Guétary (1,8 %), Luis Mariano- o a la canción más divulgada -Petula Clark (2,2 %)."* (Bourdieu, 1979: 258-260)

El factor 2 expresa una dimensión vinculada a la antigüedad en la burguesía, oponiendo la fracción tradicional en el norte del cuadro con los advenedizos en el sur.

> *"Los indicadores de las disposiciones asociadas a la mayor o menor antigüedad en la burguesía son los que aportan las más altas contribuciones absolutas al segundo factor: huellas incorporadas de una trayectoria social y de un modo de adquisición del capital cultural, las disposiciones éticas y estéticas que se manifiestan principalmente en la relación con la cultura legítima y en los matices del arte de vivir cotidiano separan a unos individuos que tienen más o menos el mismo volumen de capital cultural. … Este segundo factor evidentemente distribuye las fracciones según la proporción de sus miembros que son originarios de la burguesía o de cualquier otra clase: de un lado, los miembros de profesiones liberales y los profesores de enseñanza superior (y, en menor grado, los cuadros del sector privado), y del otro los ingenieros, los cuadros del sector público y los profesores de la enseñanza secundaria, categorías que representan unas vías de acceso privilegiadas (con la mediación del éxito escolar) a la clase dominante, dividiéndose los patronos en partes casi iguales entre los dos polos. Los primeros, agrupados junto a los valores positivos del segundo factor, tienen en común el haber adquirido (inicialmente) su capital por familiarización en el seno de su propia familia y presentan signos de una antigua pertenencia a la burguesía tales como la posesión de muebles heredados (3,1%) y la frecuentación de anticuarios (2,4%), la predilección por un hogar confortable y por una cocina tradicional (1,5%), la frecuentación de los museos del Louvre y de arte moderno (1,8 %), el gusto por El concierto para la mano izquierda, del que se sabe que casi siempre se asocia con la práctica del piano. Los otros, que deben lo esencial de su capital a la escuela y a los aprendizajes tardíos que una alta cultura escolar favorece e implica, se-oponen a los precedentes por su inclinación hacia unos amigos decididos (2,6%) y con espíritu positivo (3,6%) y no, como en el otro polo, refinados o artistas; por su gusto hacia los hogares claros y limpios (3,2%), sobrios y discretos (1,6%) y hacia unas obras de cultura burguesa media, como La danza del sable (5,1 %), Utrillo y Van Gogh, o, en otro orden de cosas, Jacques Brel o Aznavour, Buffet y la Rapsodia en blue, otros tantos índices de una ascensión en curso. Se caracterizan por unas elecciones prudentes y en consecuencia relativamente homogéneas: nunca descienden hasta obras sospechosas de trivialidad o vulgaridad, como La Arlesiana o El bello Danubio azul, sólo raramente se aventuran hasta obras ya un poco menos "canónicas", como El niño y los sortilegios, elegidas con frecuencia por los intermediarios culturales y los productores artísticos."* Bourdieu (1979: 261-262)

El tercer factor presenta el análisis de correspondencias poniendo en relación el factor 1 con el 3 pero simplificando la carga de información con respecto a las categorías de las variables. El autor selecciona sólo las modalidades cuyas contribuciones absolutas eran iguales o superiores a 1,5 entre las variables activas y entre las ilustrativas elige representar sólo la variable educación.

Gráfico de las variantes del gusto dominante en el espacio social de los factores 1 y 3



Fuente: P. Bourdieu, La Distinción (1988: 262)

Así el tercer factor opone dos realidades, un grupo más cuestionador de los gustos burgueses y el resto de las fracciones, más afianzadas en la clase y menos cuestionadora de su propio gusto.

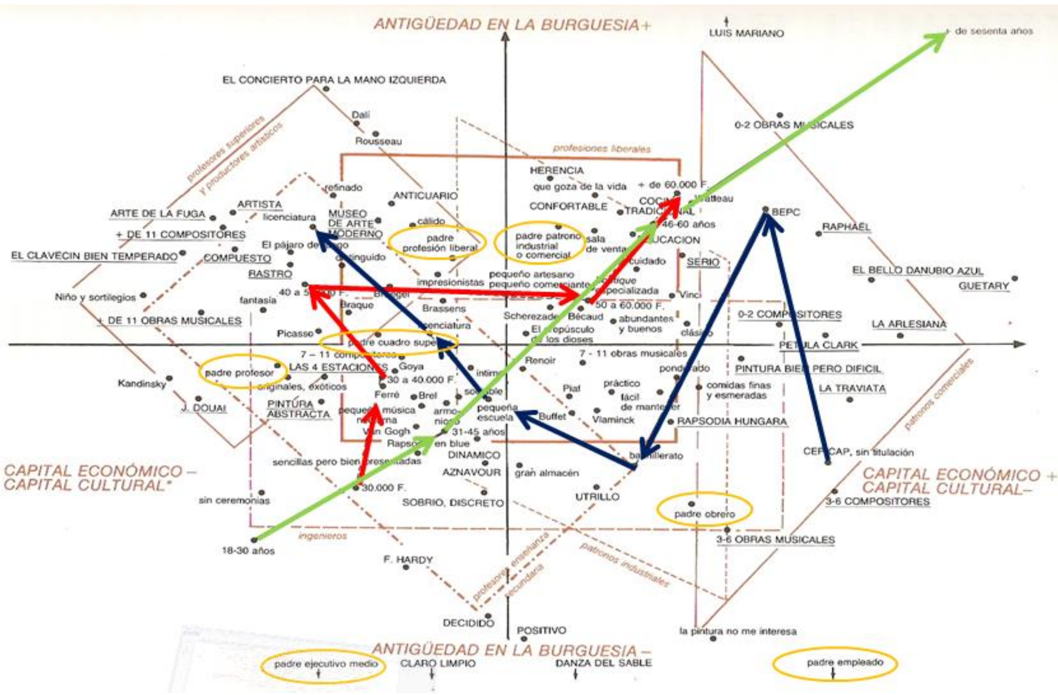
> *El tercer factor que, por parte de los individuos, opone a la mayoría de los profesores y sobre todo de los artistas -más inclinados aún que los profesores a marcar su rechazo del gusto burgués- y a los patronos comerciales, la parte más típicamente burguesa (por su origen, su residencia y su formación) de las profesiones liberales, de los industriales y de los cuadros, tiende sobre todo a caracterizar el "gusto burgués" de estas categorías oponiéndolo al gusto de todas las demás fracciones y principalmente al "gusto intelectual", a la vez más seguro y más audaz, pero también, secundariamente, a un gusto negativamente definido que acumula algunos rasgos del gusto medio y del gusto popular (el de los grandes comerciantes). Gusto modal o a la moda -como lo testimonia la adhesión al juicio en favor de los pintores impresionistas (4,2%), confirmado por la elección de Van Gogh (2,1%) o Renoir (2,1%)- y fundado en una competencia media (conocimiento de 7 a 11 obras, 3,3%, y de 7 a 11 compositores, 3,2%), el gusto burgués o mundano es fundamentalmente un gusto tradicional (con la preferencia por las comidas que caen dentro de la tradición francesa, 1,3%, o por las compras en los anticuarios, 1,0%, o por los amigos educados, 1,5%) y una especie de hedonismo temperado (con la elección, por ejemplo, de un hogar confortable pero también sobrio y discreto, 1,8%, e íntimo, 1,2%) y mesurado hasta en sus audacias (con la elección de El pájaro de fuego o de la Rapsodia en blue, 1,3 %, o la preferencia por unos amigos dotados de un espíritu positivo, 1,7 % -por oposición a artistas-). Este gusto se define sobre todo por oposición a un conjunto de indicadores que caracterizan una cultura a la vez más "académica" (conocimiento de 12 compositores y más, 3%, conocimiento de 12 obras o más, 1,9%, preferencia por Vinci, 1,6%, etc.) y más audaz*

> *-relativamente- (con la elección de Kandinsky, 1,4%, y Picasso, 1,3%) pero también más ascética (con la preferencia por Goya o por El clavecín bien temperado, compras en los rastros, etc.).* Bourdieu (1979, 263-266)

Así, expresa Lebaron (1991: 8), el análisis resulta de una importante fuerza sociológica que muestra la existencia de una homología estructural entre el "espacio de los estilos de vida" y el "espacio de las posiciones sociales".

El gráfico siguiente intenta echar luz al posicionamiento de algunas variables de interés como son las ilustrativas y que son utilizadas para arribar a la decisión de denominar a los factores.

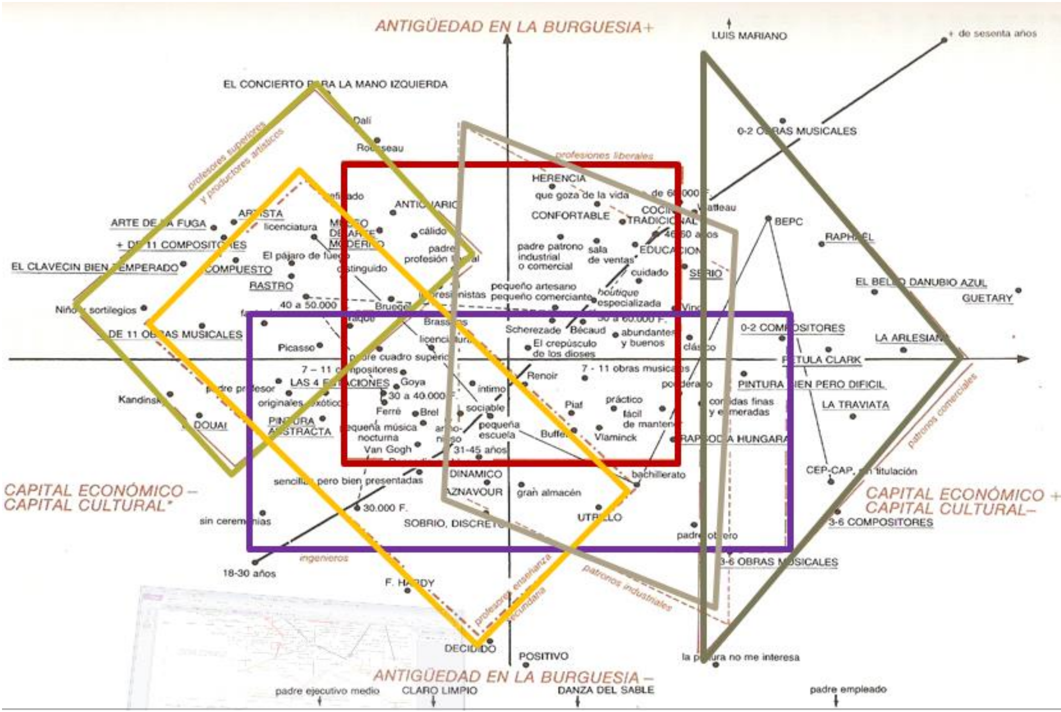
Gráfico de las variantes del gusto dominante en el espacio social de los factores 1 y 2 resaltando las variables ilustrativas



Así el nivel de instrucción, en azul, cruza desde el extremo derecho hacia el izquierdo, reforzando la interpretación sobre la distribución del capital cultural, los perceptores de bajos ingresos (flechas rojas) se encuentran en el cuadrante inferior izquierdo y los que tienen más altos ingresos se posicionan en el cuadrante inferior derecho. Ello contribuye a identificar los polos de capital económico. La antigüedad en la burguesía puede ser interpretada a partir de la ayuda inconmensurable que aporta el hecho de contar con la profesión del padre (círculos amarillos). Finalmente la edad del encuestado (flechas verdes) cruza el espacio social en diagonal desde el extremo inferior hacia el superior derecho.

Por su parte en el gráfico siguiente se representa el espacio de los individuos. En él se encuentran las nubes dibujadas a mano a partir de los contornos de varias subnubes de individuos, a modo de "factores estructurantes" como afirma Lebaron (1991:7) pues las nubes de individuos son estructuradas por factores externos en línea con la metodología utilizada (Le Roux y Rouanet, 2004: 33).

Gráfico de las variantes del gusto dominante en el espacio social de los factores 1 y 2 resaltando el espacio de los individuos



De esta manera los individuos que forman el corazón de la burguesía (profesiones liberales e ingenieros) quedan posicionados en el centro, los primeros apenas más vinculados a la antigua burguesía y los segundos a la nueva. Escorados a la derecha quedan los patronos comerciales y los industriales, los primeros más asociados a capital económico y menos capital cultural. Por su parte hacia la izquierda quedan los profesores superiores y productores artísticos y también aunque más asociados a la nueva burguesía y con menos capital económico que los anteriores encontramos a los profesores de enseñanza secundaria.

Consideramos este ejemplo una excelente demostración del oficio del sociólogo en cuanto a interpretar la realidad y los elementos que la estructuran partiendo de variables enterante cualitativas que conforman un espacio social donde se pueden analizar los aspectos cualitativos con un tratamiento métrico cuantitativo. Sobre este aspecto resulta interesante la reflexión que hace Bourdieu (2009) *"…constato correlaciones asombrosas, me sorprende ver hasta qué punto en el espacio de La Distinción, en un punto -y en el interior de ese punto minúsculo- encuentro todo un espacio de diferencias entre el alumno de la HEC, de la ENA, de la Escuela Normal, diferencias sistemáticas, no aleatorias… es evidente que son diferencias estadísticas, no significa que estemos condenados a tener un gusto determinado a partir de una postura determinada… pero tenemos gustos probables"*.

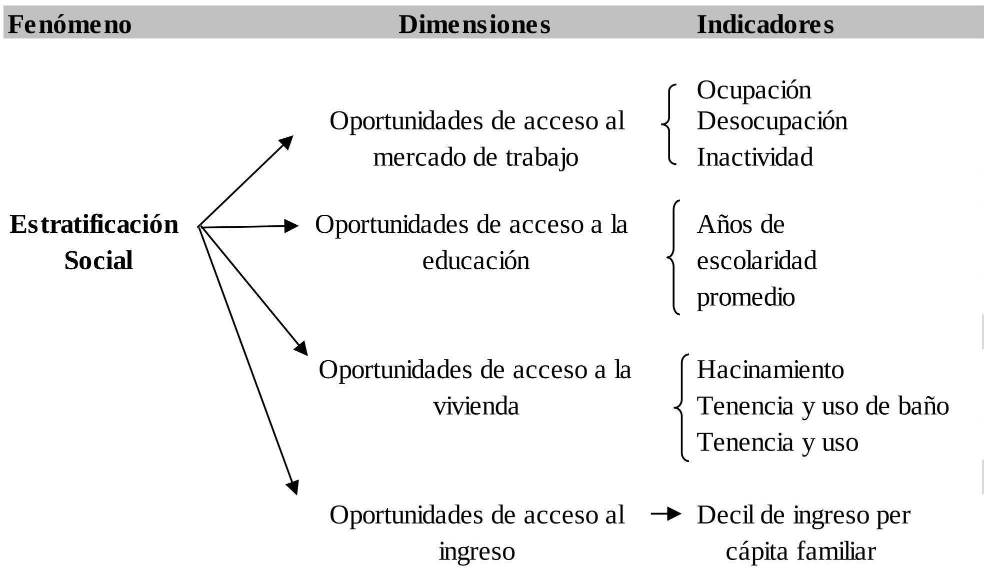
5.2.4.2. *Un modelo de estratificación para Argentina: el espacio de los factores*

Presentamos a continuación una síntesis del análisis desarrollado en Fachelli (2009) donde se conforma un modelo de estratificación social en Argentina desde una perspectiva multidimensional, incluyendo en forma simultánea indicadores de ocupación, ingreso, vivienda y educación, como elementos relevantes para estratificar la sociedad.

En términos operativos nos centramos en los bienes primarios sociales más básicos de la vida, como acceder a una ocupación estable, a una educación básica, a unos ingresos mínimos, a proteger la salud, a una vivienda, a vivir no hacinados y a tener seguridad social en la vejez. No obstante, en función de la información que nos brinda la base de datos a utilizar, seleccionamos aquellos bienes a los que podemos acceder empíricamente, y como no disponemos de los indicadores de salud y seguridad social para toda la población urbana, los bienes primarios que finalmente consideramos son los siguientes:

1. Acceso al mercado de trabajo.
2. Acceso a la educación.
3. Acceso a la vivienda
4. Acceso al ingreso.

La cantidad de bienes primarios que podrían tomarse en cuenta es muy amplia y en ese sentido el modelo puede ser diversificado y mejorado. El esquema del modelo de estratificación social que hemos definido se presenta a continuación y puede observarse en forma completa en Fachelli (2009).



El proceso descrito anteriormente nos conduce a analizar un cambio que hace referencia a la propia conformación de la estratificación como fenómeno social, en este caso se trata de una estructura bastante estable pues en los cuatro años analizados hemos encontrado los mismos patrones. El cuadro siguiente presenta los porcentajes de la varianza explicada para cada dimensión[36]:

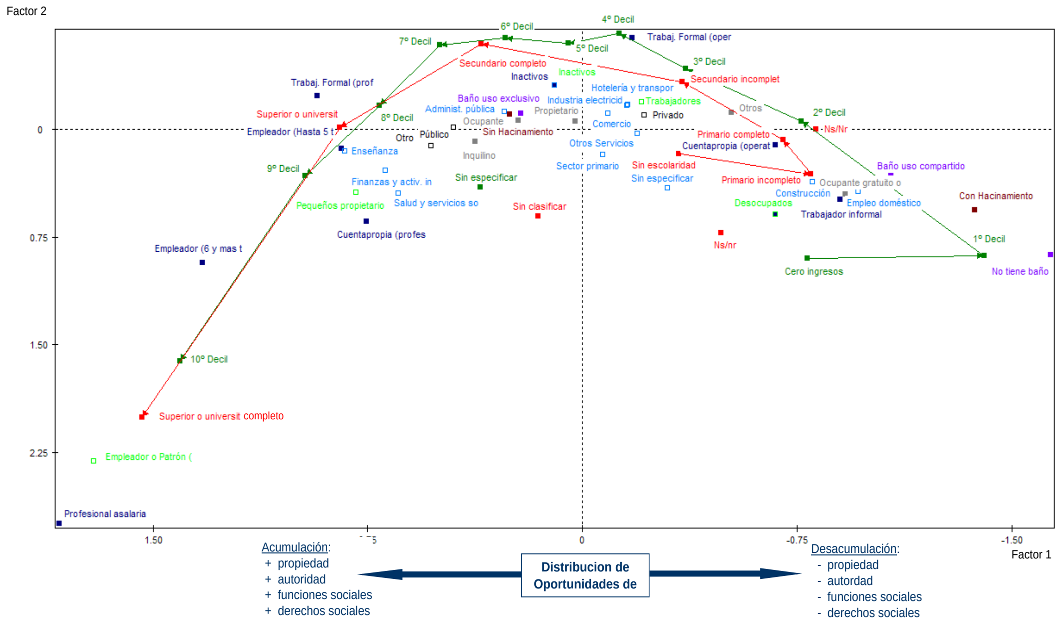
Varianza explicada ajustada de cada dimensión

Período	Estabilidad	Post Crisis	Recuperación	
			Incipiente	Consolidada
Eje factorial	1997	2002	2003	2006
1 = 1ra. Dimensión	64,7%	62,7%	60,3%	64,8%
2 = 2da. Dimensión	9,6%	10,3%	8,9%	9,6%
3 = 3ra. Dimensión	4,9%	5,3%	6,6%	5,0%
Varianza explicada	79,2%	78,2%	75,9%	79,4%

Fuente: Fachelli et al (2012)

Los tres gráficos siguientes presentan la distribución de variables en el espacio factorial.

1ra. Dimensión: Distribución de oportunidades de acumulación/desacumulación de bienes primarios. Año 1997



Los años 2002, 2003 y 2006 muestran la misma distribución global, aunque las coordenadas donde se posiciona cada variable presentan un leve movimiento.

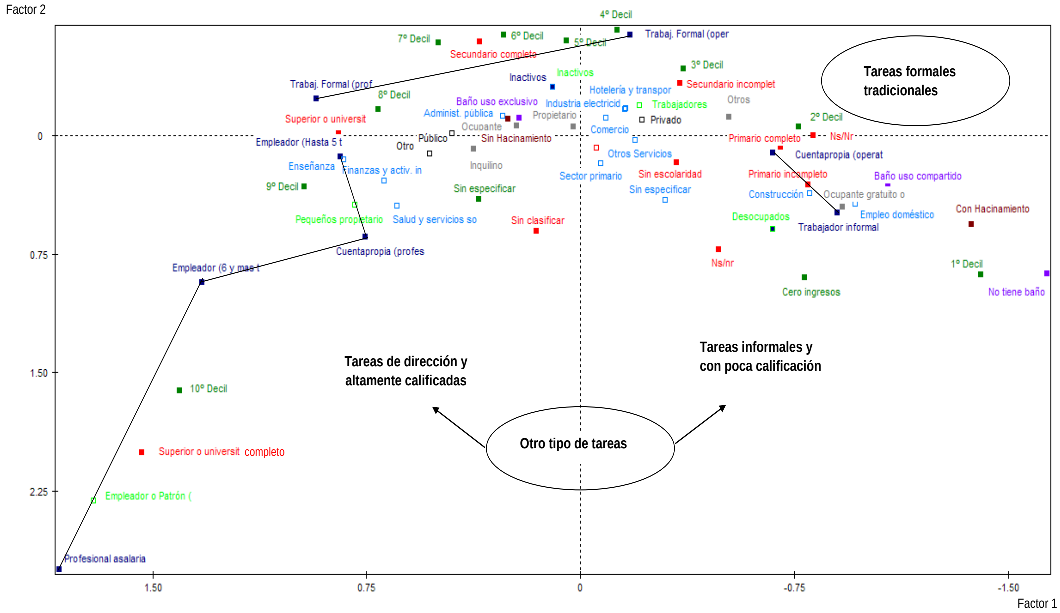
La primera dimensión expresa la posición de los hogares con respecto a la distribución de bienes primarios (poseedores o no poseedores de dichos bienes) con un porcentaje en torno al 63% de la varianza explicada, según el año analizado. La segunda dimensión está vinculada al mercado de trabajo y la transformación que se ha dado de la sociedad salarial (ligada a la industria y al trabajo formal) a la sociedad terciarizada vinculada en mayor medida al sector servicios, que a su vez expresa dos tipos de tareas opuestas -las precarizadas e informales por un lado y las muy calificadas por otro- con un porcentaje en torno al 9% de varianza explicada[37]. Finalmente, la tercera

[36] Se pueden hacer distintas correcciones de la varianza explicada. Aquí se utiliza a Greenacre (2008).

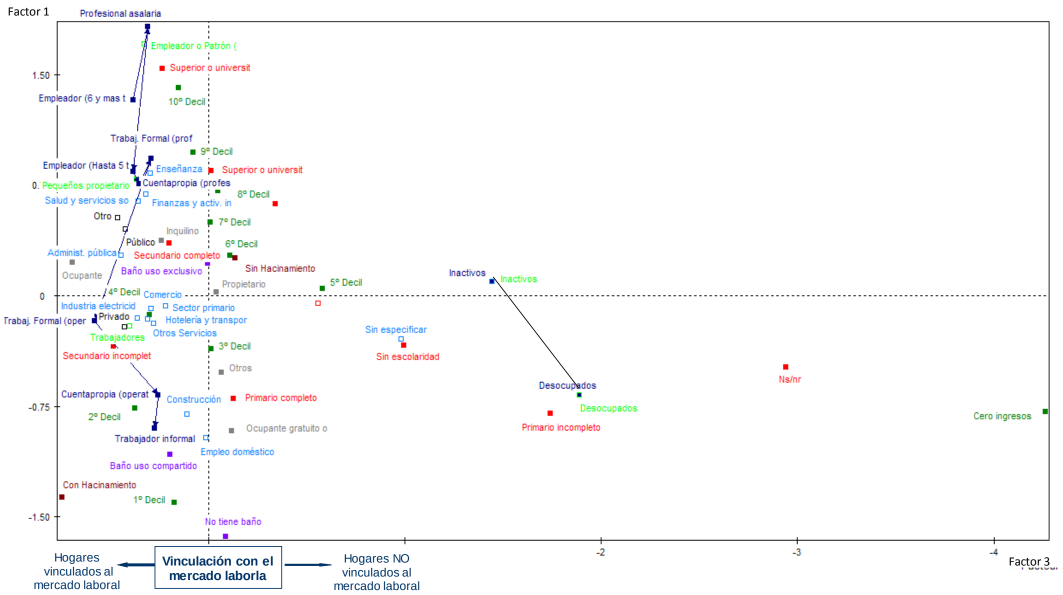
[37] Las características de esta dimensión va en la línea con los hallazgos de varios investigadores sobre el fenómeno que han denominado "heterogeneidad estructural" caracterizada principalmente por dos trayectorias opuestas, que

dimensión, con un 5% de varianza explicada, diferencia los hogares que tienen algún miembro ocupado laboralmente de aquellos hogares con personas inactivas o desocupadas.

2da. Dimensión: Inserción en el mercado laboral, tareas tradicionales vs. otro tipo de tareas. Año 1997



3ra. Dimensión: Mercantilización-Desmercantilización. Año 1997



al analizarse a nivel global e intergeneracional parece no haber movimiento pues éste es neutralizado por un efecto de composición entre aquellos grupos que ascienden y aquellos que descienden (Kessler y Espinosa, 2003; Filgueiras, 2007; Chávez Molina y Molina Derteano, 2009; Salvia, y Quartulli, 2010; Salvia y Vera, 2010; Salvia et al., 2008).

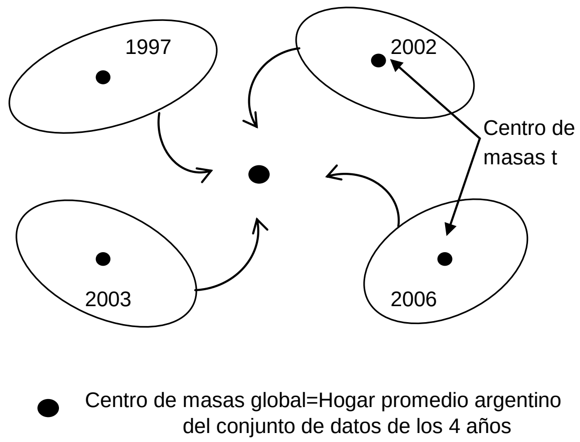
En el análisis de estratificación social se empleó igualmente la técnica del Análisis de Correspondencias Múltiples Condicional (ACMC). Este es un método que se deriva del ACM y estudia, por una parte, las relaciones entre las variables cualitativas definidas en una misma población y, por la otra, induce una estructura a partir de las relaciones del comportamiento observado entre esas variables (Escofier y Pagès, 1992). Este método permite introducir como condicionamiento una variable exterior y eliminar del análisis la parte ligada a esta variable (Escofier, 1990: 13).

De esta manera, en nuestro análisis la variable condicional "t" son los años 1997, 2002, 2003 y 2006. El ACM Condicional suprime la variable "t" y pone en relación todos los hogares para todos los períodos. Este análisis requirió la unión de las cuatro bases de datos (116.218 hogares sin expandir, que representan 27.631.632 hogares urbanos).

Se generan así nuevas nubes de puntos, nuevos centros de masa y un nuevo hogar promedio argentino conformado por la información de los cuatro años. También se recalculan las inercias y con ello se generan nuevos estratos y nuevas distancias inter e intra-estratos.

Para comprender el procedimiento en forma muy simplificada, presentamos el gráfico siguiente:

ACMC: esquema del procedimiento



Fuente: Fachelli (2009) sobre la base de Escofier, pág.16.

Este análisis se utilizó para comparación global con cada uno de los años analizados por separado y para validar el análisis realizado para cada año en particular. De esta manera se ratificaron las dimensiones obtenidas presentadas previamente. Para mayor detalle consultar Fachelli (2009; 2013).

6. Análisis de correspondencias con SPSS

Para ilustrar el uso del software SPSS con el objetivo de realizar un análisis de correspondencias presentaremos dos ejemplos. Uno de correspondencias simples donde analizaremos los datos de la Tabla III.11.10, la tabla de correspondencias entre la ocupación dominante del hogar y el nivel de ingresos, y otro de correspondencias múltiples con un ejemplo sencillo de tres variables que vimos en el capítulo III.5 y III.6

de tablas de contingencia y log-lineal donde se relacionan la actitud, los estudios y el sexo.

En SPSS, inicialmente el ACS se identificó como "análisis de correspondencias" y se realizaba con el procedimiento **ANACOR**. El análisis de correspondencias múltiples había sido llamado "análisis de homogeneidad" y se correspondía con el procedimiento **HOMALS** (*HOMogeneity analysis vía Alternating Least Squares*). Estos procedimientos se incluyen dentro de un módulo del SPSS llamado *Categories* y permiten la realización de un tipo de análisis llamado *optimal scaling*. Con la versión 13 de SPSS el análisis de correspondencias simples pasa a ser el comando **CORRESPONDENCE** y el análisis de correspondencias múltiples pasa a llamarse "Análisis de correspondencias múltiple" y se corresponde con el comando **MULTIPLE CORRESPONDENCE**.

Desde esta perspectiva, que difiere en algunos aspectos de la que presentaremos con SPAD o R, el análisis de correspondencias es una técnica destinada a cuantificar los datos nominales (categóricos) mediante la asignación de valores numéricos a los casos (objetos) y a las categorías de las variables, de manera que los objetos de la misma categoría estén cerca unos de otros y los objetos de categorías diferentes estén alejados unos de otros. Cada objeto se encuentra lo más cerca posible de los puntos de categoría para las categorías que se aplican a dicho objeto. De este modo, las categorías dividen los objetos en subgrupos homogéneos. Las variables se consideran homogéneas cuando clasifican objetos de las mismas categorías en los mismos subgrupos.

El *optimal scaling* o análisis de escalamiento óptimo incluye un conjunto de técnicas comunes desarrolladas para SPSS por el grupo *Data Theory Scaling System Group* (DTSS), formado por miembros de los departamentos de educación y de psicología de la Faculty of Social ANF Behavioral Sciences de la University of Leiden de Holanda (http://www.datatheory.nl/index.html).

El escalamiento óptimo asigna cuantificaciones numéricas a las categorías de cada variable mediante un criterio de optimización, lo que permite utilizar las variables así cuantificadas en procedimientos donde se requiere una métrica. La cuantificación óptima de cada variable escalada obtiene mediante un método iterativo denominado mínimos cuadrados alternantes donde se utilizan sucesivamente las cuantificaciones actuales para encontrar una solución final.

Las técnicas de escalamiento óptimo son:

- El análisis de correspondencias simples (**CORRESPONDENCE**, antiguo **ANACOR**) para el análisis de una tabla de contingencia de dos dimensiones o, más en general, una tabla de doble entrada de números positivos.
- El análisis de correspondencias múltiples (**MULTIPLE CORRESPONDENCE**, antiguo **HOMALS**) para el análisis de la relación de múltiples variables cualitativas.
- El análisis de componentes principales categórico o no lineal (**CATPCA**, *Categorical Principal Componentes Analysis*, antes llamado **PRINCALS**, *PRincipal Componentes analysis vía Alternating Least Squares*), analiza y reduce la dimensionalidad de un conjunto de variables en las que se combinan niveles de medida cualitativos y cuantitativos.
- El análisis de correlación canónica no lineal (**OVERALS**, Over Alternating Least Squares), también permite la utilización tanto de variables cualitativas como

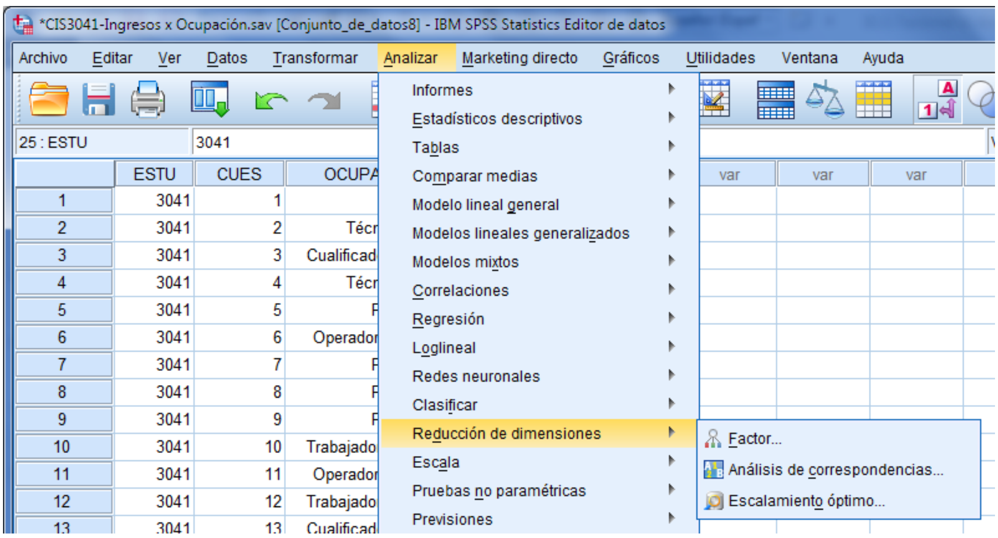
cuantitativas pero aquí el objetivo es analizar las relaciones existentes entre dos o más conjuntos de variables para establecer sus similitudes a través de variables canónicas de cada conjunto con las puntuaciones asignadas a los objetos.

- El análisis de regresión categórica (**CATREG**, *CATegorical regresión with optimal scaling using Alternating Least Squares*) destinado a predecir los valores de una variable dependiente cualitativa a partir de una combinación de variables independientes cualitativas.

Con excepción del análisis de regresión categórica, se caracterizan por ser técnicas de análisis factorial destinadas a la reducción de dimensiones a partir de un conjunto inicial de variables para describir estructuras y modelos de relación entre el conjunto de ellas. Todas ellas, como técnicas de escalamiento óptimo, permiten detectar relaciones no lineales y buscan obtener la máxima correlación entre las variables. Siguen los principios del análisis de componentes principales y del análisis de correlación canónica, adaptados a la utilización de variables categóricas o mixtas.

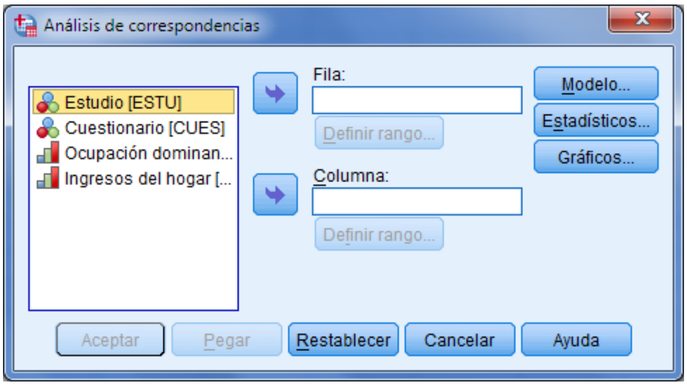
Como técnicas factoriales los resultados incluyen puntuaciones óptimas o cuantificaciones óptimas tanto para las categorías de cada variable (cuantificación de categorías) como de los individuos u objetos (cuantificación de objetos), por tanto, se derivan variables continuas; es por eso que se llaman técnicas de cuantificación de datos cualitativos (como también lo es la técnica del Escalamiento Multidimensional No Métrico). Una cuantificación es óptima en el sentido de que las categorías están separadas entre ellas a la dimensión o dimensiones consideradas tanto como sea posible y, a la vez, dentro de cada categoría los individuos están lo más próximos posible, es decir, con puntuaciones lo más homogéneas entre sí. Finalmente todas ellas permiten la representación gráfica para visualizar los resultados como ayuda a la interpretación de las estructuras subyacentes.

Seguidamente daremos cuenta de los dos procedimientos de análisis de correspondencias, simples y múltiples. Estos procedimientos se localizan en el menú de Analizar/Reducción de dimensiones:

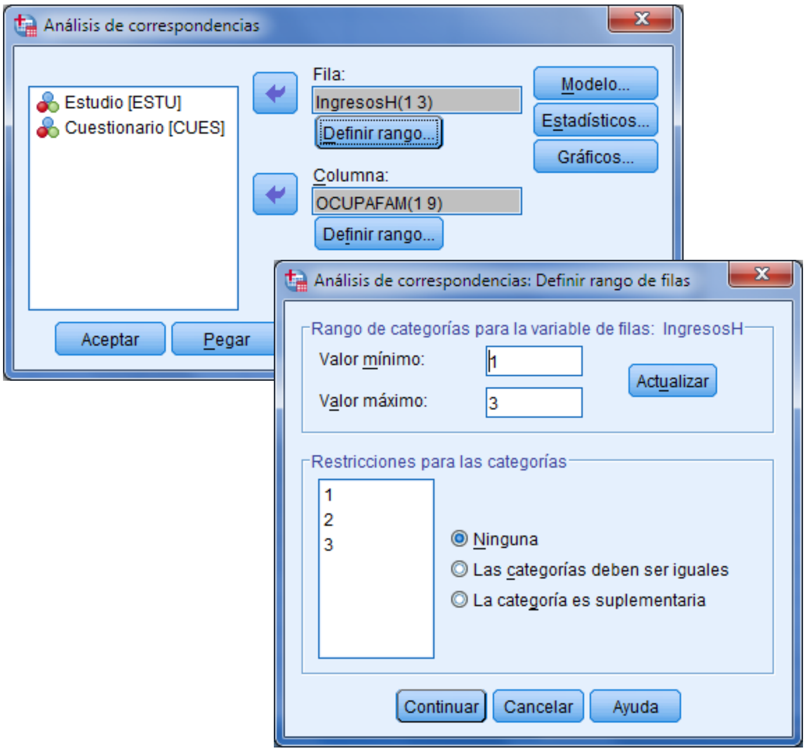


6.1. Correspondencias simples

A través del menú: Analizar/Reducción de Dimensiones/Análisis de correspondencias... llegamos al cuadro de diálogo inicial de este procedimiento:



En primer lugar se trata de especificar las dos variables del análisis, la que se considera como variable-fila, en nuestro caso IngresosH y la que se considera como variable-columna, OCUPAFAM. Una vez determinadas las variables que intervienen en el análisis hay que detallar su rango. Si hacemos clic sobre el botón Definir rango se abre un cuadro de diálogo como el siguiente:

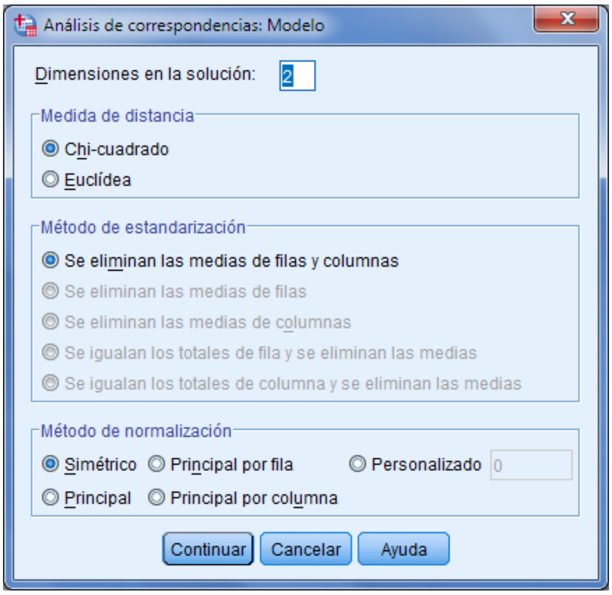


Se trata de especificar los valores mínimo y máximo que corresponden a los valores o categorías de la variable considerada. En este caso se han escrito los valores 1 como valor mínimo y 3 como valor máximo, ya que define el rango de valores de la variable IngresosH. A continuación se debe hacer clic sobre el botón de Actualizar y estos valores pasarán al recuadro inferior de Restricciones para las categorías. Las restricciones se concretan en tres posibilidades excluyentes entre sí, si bien en nuestro caso no se fijará ninguna restricción, consideraremos todas las categorías como activas.

La opción que especifica que las categorías deben ser iguales es una restricción de igualdad que se aplica si el orden obtenido por las categorías no es el deseado o si no se corresponde con el intuitivo. La opción que especifica que la categoría es suplementaria permite considerar algunas de las categorías como pasivas, para que no influyan en el análisis pero sí se representen en el espacio definido por las categorías activas, así las categorías suplementarias no juegan ningún papel en la definición de las dimensiones.

Una vez hecha la especificación del rango de las filas hay que hacer clic en el botón Continuar y se repite la operación para la variable-columna, OCUPAFAM, con los valores 1 a 9.

En el cuadro de diálogo de Modelo podemos especificar cuatro tipos de opciones:

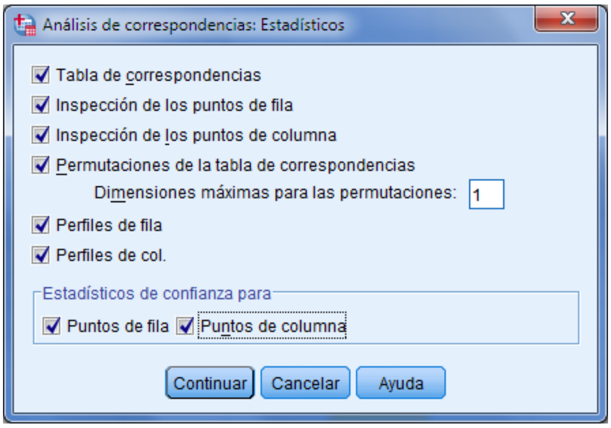


Son las siguientes:

- En el número de dimensiones consideraremos por defecto dos dimensiones con representaciones gráficas de estas dos. Si no se utilizan criterios de igualdad y todas las categorías son activas, la dimensionalidad máxima es igual al número de categorías de la variable con menos categorías menos uno.
- La medida de distancia de chi-cuadrado (una distancia ponderada donde el peso es la masa de las filas o columnas) es la especificación predeterminada para hacer un análisis de correspondencias típico y es la que utilizaremos. La medida de distancia puede ser alternativamente euclidiana (la raíz cuadrada de la suma de los cuadrados diferencias entre los valores de las dos filas o columnas).
- El método de estandarización viene determinado por la elección de la distancia, si es de chi-cuadrado implica necesariamente una media de filas y columnas. Si es euclidiana se puede optar por cualquiera de las cinco opciones que se centran sólo en las filas o las columnas, con la posibilidad de igualar previamente los marginales de fila o de columna.

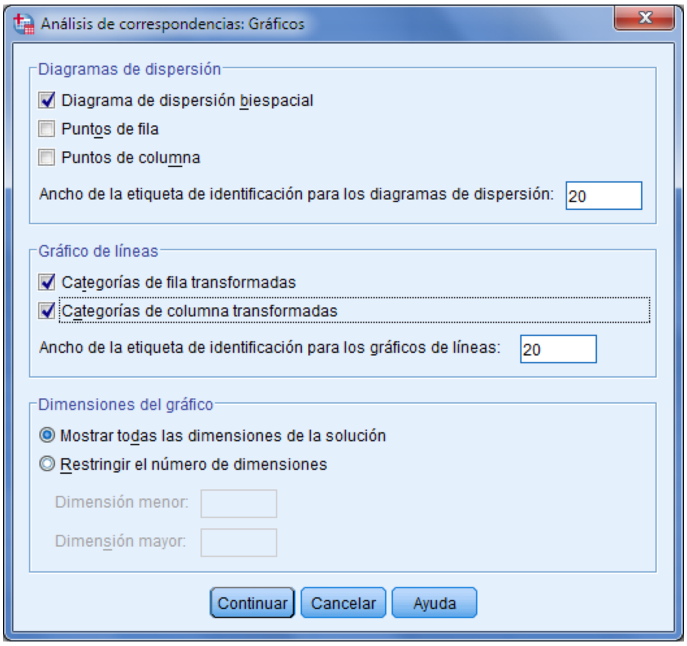
- Finalmente el método de normalización determina la forma en que se normalizan las filas y columnas. Consideraremos la normalización simétrica ya queremos comparar las correspondencias entre filas y columnas, entre las categorías de cada variable. En este caso, para cada dimensión, las puntuaciones de fila son la media ponderada de las puntuaciones de columna dividido por el valor propio coincidente y las puntuaciones de columna son la media ponderada de las puntuaciones de filas dividido por el valor propio coincidente. La normalización principal busca comparar las diferencias o similitudes entre las categorías de cada una de las variables y no entre las variables. La normalización por fila o por columna examina específicamente las categorías de la variable de las filas o de la variable de las columnas. La normalización personalizada se puede especificar un valor entre -1 y 1: el valor -1 corresponde a principal por columna, el valor 1 corresponde a principal por fila, el valor 0 corresponde a simétrico, y todos los demás valores dispersan la inercia entre las puntuaciones de columna y de fila en diferentes grados.

En el cuadro de diálogo de Estadísticos se pueden especificar los estadísticos que aparecen en un cuadro de diálogo como el siguiente:



- La Tabla de correspondencias es la tabla de contingencia con los datos absolutos entre las dos variables.
- La opción Inspección de los puntos fila o columna nos proporciona en el listado de resultados una tabla en la que para cada modalidad o categoría de la variable fila (o columna) muestra la masa, las puntuaciones, la inercia, la contribución del punto a la inercia de la dimensión y la contribución de la dimensión a la inercia del punto, y que nos sirven para describir los resultados del análisis, resultados que tienen la misma lectura en términos de la representación gráfica.
- Con las Permutaciones de la tabla de correspondencias obtenemos una permuta de la tabla original de acuerdo con el orden ascendente de las puntuaciones de las filas y columnas a partir del número máximo de dimensiones que se quiera, por defecto es considera la orden que determina la primera dimensión.
- Las opciones Perfiles de fila y Perfiles de col. proporcionan las tablas de distribuciones de porcentajes por fila o por columna.
- Los Estadísticos de confianza para dan las desviaciones típicas y las correlaciones de los puntos fila y los puntos columna activos.

En el cuadro de diálogo de Gráficos obtenemos un elemento informativo básico de la interpretación de los resultados de un ACS en forma gráfica.



- A través del Diagrama de dispersión biespacial se nos presenta por defecto la representación en dos dimensiones de las categorías de las dos variables analizadas en relación a las dimensiones o factores retenidos. Cuando el número de dimensiones es superior a 2 representa varios diagramas biespaciales. También podemos optar por pedir exclusivamente los puntos fila o columna.
- Por su parte, los Gráficos de línea representan las categorías de las variables, de fila o de columna, con las puntuaciones de estas categorías.
- Por último, con Dimensiones del gráfico podemos restringir el número de dimensiones de las representaciones.

Adicionalmente se pueden especificar otras opciones con el lenguaje de comandos que no parecen a través del menú. En particular se puede:

- Especificar datos tabulares como entrada en lugar de utilizar datos por caso (mediante el subcomando TABLE=ALL).
- Especificar el número de caracteres de etiqueta de valor que se utilizan para etiquetar los puntos para cada tipo de diagrama de dispersión matricial o diagrama de dispersión biespacial matricial (mediante el subcomandament PLOT).
- Especificar el número de caracteres de etiqueta de valor que se utilizan para etiquetar los puntos para cada tipo de gráfico de líneas (mediante el subcomando PLOT).
- Escribir una matriz de puntuaciones de fila y de columna en un archivo de datos matriciales (mediante el subcomando OUTFILE).
- Escribir una matriz de estadísticos de confianza (varianzas y covarianzas) para los valores propios y las puntuaciones en un archivo de datos matriciales (mediante el subcomando OUTFILE).

− Especificar varios conjuntos de categorías para igualar (mediante el subcomando EQUAL).

El conjunto de especificaciones que hemos detallado se corresponden con la siguiente sintaxis de SPSS[38]:

```
CORRESPONDENCE TABLE=IngresosH(1 3) BY OCUPAFAM(1 9)
 /DIMENSIONS=2
 /MEASURE=CHISQ
 /STANDARDIZE=RCMEAN
 /NORMALIZATION=SYMMETRICAL
 /PRINT=TABLE RPOINTS CPOINTS PERMUTATION(1) RPROFILES CPROFILES RCONF CCONF.
```

Como resultado de la ejecución de estas instrucciones, ya sea a través del menú o del editor de sintaxis, se obtienen los resultados que siguen.

Crédito

CORRESPONDENCE
Version 1.1
by
Leiden SPSS Group
Leiden University
The Netherlands

En primer lugar, aparece la tabla de correspondencias (de contingencia) con las frecuencias de la distribución conjunta:

Tabla de correspondencias

IngresosH Ingresos del hogar	OCUPAFAM Ocupación dominante de los padres y las madres									
	Director-Gerente	Profesional	Técnico apoyo	Administrativo	Trabajador Servicios	Cualificado Primario	Cualificado Industria	Operador-Montador	Ocupación elemental	Margen activo
Bajos	5	11	25	5	71	127	113	50	67	474
Medios	13	31	61	22	111	118	154	118	59	687
Altos	25	61	66	9	74	48	71	57	18	429
Margen activo	43	103	152	36	256	293	338	225	144	1.590

Junto con las proporciones calculadas por fila y por columna:

Perfiles de fila

IngresosH Ingresos del hogar	OCUPAFAM Ocupación dominante de los padres y las madres									
	Director-Gerente	Profesional	Técnico apoyo	Administrativo	Trabajador Servicios	Cualificado Primario	Cualificado Industria	Operador-Montador	Ocupación elemental	Margen activo
Bajos	0,011	0,023	0,053	0,011	0,150	0,268	0,238	0,105	0,141	1,000
Medios	0,019	0,045	0,089	0,032	0,162	0,172	0,224	0,172	0,086	1,000
Altos	0,058	0,142	0,154	0,021	0,172	0,112	0,166	0,133	0,042	1,000
Masa	0,027	0,065	0,096	0,023	0,161	0,184	0,213	0,142	0,091	

Considerando dos dimensiones o factores se acumula el 100% de la varianza explicada, pues dos es el número máximo de dimensiones, cuyo cálculo se determina del mínimo de categorías de fila o de columnas menos uno, es decir, $\text{mín}\{I,J\}-1=\text{mín}\{3,9\}-1=2$. En la tabla que sigue se presenta la información del valor singular (la raíz cuadrada del valor propio que aparece en la columna con el nombre de inercia), la inercia (de hecho es el valor propio) así como la proporción de varianza que suponen. Se incluye asimismo una prueba estadística que da cuenta de la significación del modelo testando la hipótesis nula de independencia entre las dos variables (el valor del chi-cuadrado

[38] En la página web se encontrará el archivo de sintaxis ACS-IngresosxOcupación.sps con estas instrucciones que se aplican sobre la matriz de de datos CIS3041+.sav. También se puede reproducir el análisis a través del menú con la matriz de datos ACS-IngresosxOcupación.sps que contiene solamente las dos variables consideradas.

dividido por el número de casos es la inercia total que se descompone en cada dimensión), que en este caso permite ser rechazada.

Resumen

Dimensión	Valor singular	Inercia	Chi cuadrado	Sig.	Proporción de inercia: Contabilizado para	Proporción de inercia: Acumulado	Valor singular de confianza: Desviación estándar	Valor singular de confianza: Correlación 2
1	0,315	0,099			0,874	0,874	0,024	0,194
2	0,120	0,014			0,126	1,000	0,024	
Total		0,114	180,568	0,000[a]	1,000	1,000		

a. 16 grados de libertad

Finalmente la información del valor singular de confianza muestra las desviaciones estándar y la correlación entre las dimensiones que nos ayudan a evaluar la precisión de las dimensiones a partir de datos muestrales. Valores bajos de las desviaciones y de la correlación, próximos a cero, nos indican un nivel de confianza aceptable para extrapolar nuestros resultados al conjunto de la población.

Como se puede observar, la primera dimensión permite sintetizar la mayor parte de la varianza o inercia explicada con un 87,4%, se trata de un patrón de asociación entre las dos variables que se puede expresar en términos de un solo factor. Como veremos seguidamente, es la asociación entre ocupación e ingresos, a medida que aumenta la categoría ocupacional los ingresos son mayores.

Los resultados anteriores se acompañan de la información de los puntos de fila y columna de confianza.

Puntos de fila de confianza

IngresosH Ingresos del hogar	Desviación estándar en la dimensión 1	Desviación estándar en la dimensión 2	Correlación 1-2
Bajos	0,057	0,052	0,576
Medios	0,068	0,042	-0,155
Altos	0,054	0,057	-0,634

Puntos de columna de confianza

OCUPAFAM Ocupación dominante de los padres y las madres	Desviación estándar en la dimensión 1	Desviación estándar en la dimensión 2	Correlación 1-2
Director-Gerente	0,094	0,086	-0,653
Profesional	0,095	0,088	-0,656
Técnico apoyo	0,033	0,044	-0,136
Administrativo	0,186	0,115	0,179
Trabajador Servicios	0,017	0,011	0,176
Cualificado Primario	0,050	0,047	0,595
Cualificado Industria	0,023	0,018	-0,438
Operador-Montador	0,097	0,059	0,155
Ocupación elemental	0,052	0,055	0,614

Así pues, los factores obtenidos expresan las correspondencias (proximidades) que se dan entre la filas, las columnas, y también entre filas y columnas por la propiedad de la equivalencia distribucional. En las tablas siguientes se presentan para cada categoría de las variables, de fila y de columna, las puntuaciones en la dimensión (las coordenadas del gráfico factorial), la masa (frecuencia o importancia de cada categoría), la inercia (la varianza explicada por cada categoría), la contribución a la inercia de la dimensión (las contribuciones absolutas) y la contribución de la dimensión a la inercia del punto (las contribuciones relativas).

Puntos de fila generales[a]

IngresosH Ingresos del hogar	Masa	Puntuación en dimensión 1	Puntuación en dimensión 2	Inercia	Contribución: Del punto en la inercia de dimensión 1	Del punto en la inercia de dimensión 2	De la dimensión en la inercia del punto 1	De la dimensión en la inercia del punto 2	Total
Bajos	0,298	0,634	-0,359	0,042	0,380	0,322	0,891	0,109	1,000
Medios	0,432	0,089	0,393	0,009	0,011	0,557	0,120	0,880	1,000
Altos	0,270	-0,843	-0,232	0,062	0,609	0,121	0,972	0,028	1,000
Total activo	1,000			0,114	1,000	1,000			

a. Normalización simétrica

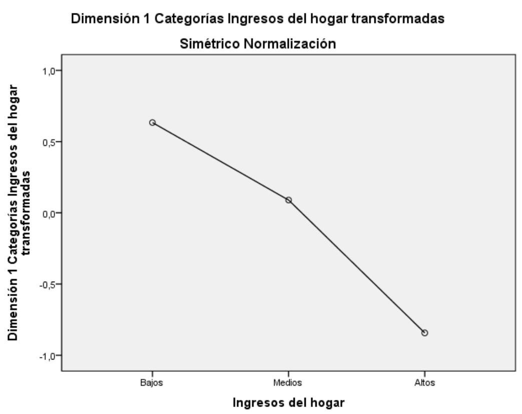
Al conservar el 100% de la inercia las contribuciones relativas suman el total 1 para cada categoría (suma por fila), mostrando la distribución de la aportación de cada categoría a cada una de las dos dimensiones. La contribución absoluta, por su parte, nos muestra qué categorías definen cada factor o dimensión (suman 1 por columna). Las categorías con mayor valor en una dimensión se corresponderán con las más alejadas del centro en la representación gráfica del gráfico factorial, es decir, tendrán una mayor coordenada.

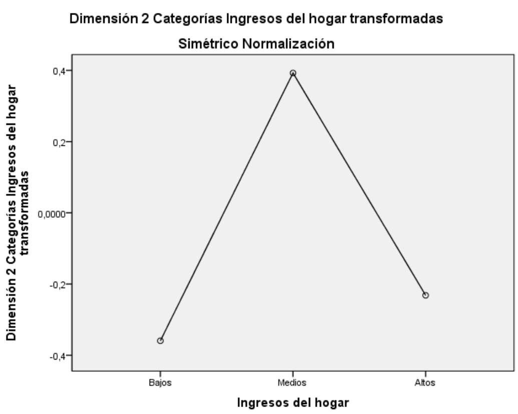
Puntos de columna generales[a]

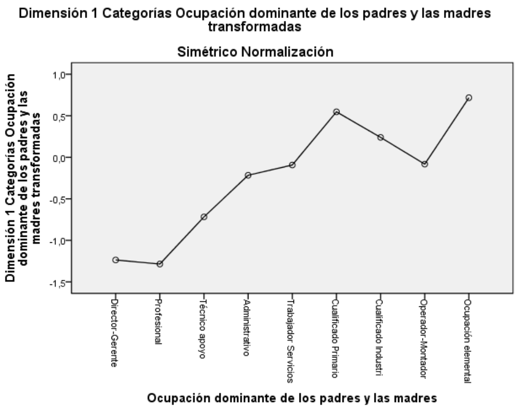
OCUPAFAM Ocupación dominante de los padres y las madres	Masa	Puntuación en dimensión 1	Puntuación en dimensión 2	Inercia	Contribución: Del punto en la inercia de dimensión 1	Del punto en la inercia de dimensión 2	De la dimensión en la inercia del punto 1	De la dimensión en la inercia del punto 2	Total
Director-Gerente	0,027	-1,237	-0,483	0,014	0,131	0,053	0,945	0,055	1,000
Profesional	0,065	-1,285	-0,480	0,035	0,340	0,125	0,950	0,050	1,000
Técnico apoyo	0,096	-0,718	-0,018	0,016	0,156	0,000	1,000	0,000	1,000
Administrativo	0,023	-0,216	1,104	0,004	0,003	0,231	0,092	0,908	1,000
Trabajador Servicios	0,161	-0,093	0,030	0,000	0,004	0,001	0,962	0,038	1,000
Cualificado Primario	0,184	0,548	-0,298	0,019	0,175	0,136	0,899	0,101	1,000
Cualificado Industria	0,213	0,239	0,084	0,004	0,039	0,013	0,955	0,045	1,000
Operador-Montador	0,142	-0,082	0,563	0,006	0,003	0,375	0,053	0,947	1,000
Ocupación elemental	0,091	0,718	-0,295	0,016	0,148	0,066	0,940	0,060	1,000
Total activo	1,000			0,114	1,000	1,000			

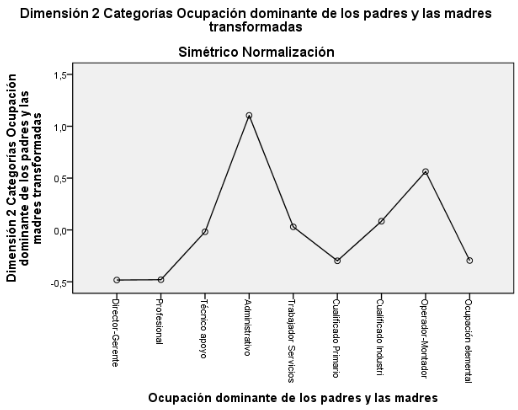
a. Normalización simétrica

Esta información, por tanto, se puede expresar gráficamente. Se pueden analizar separadamente para las categorías de cada variable en cada dimensión:

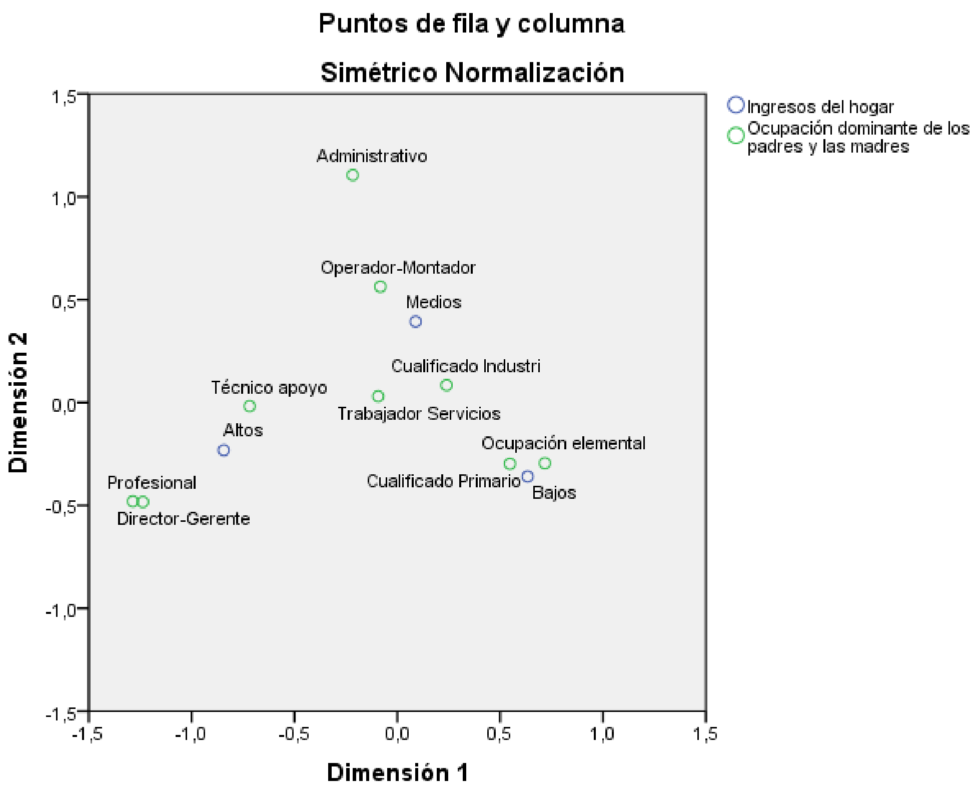




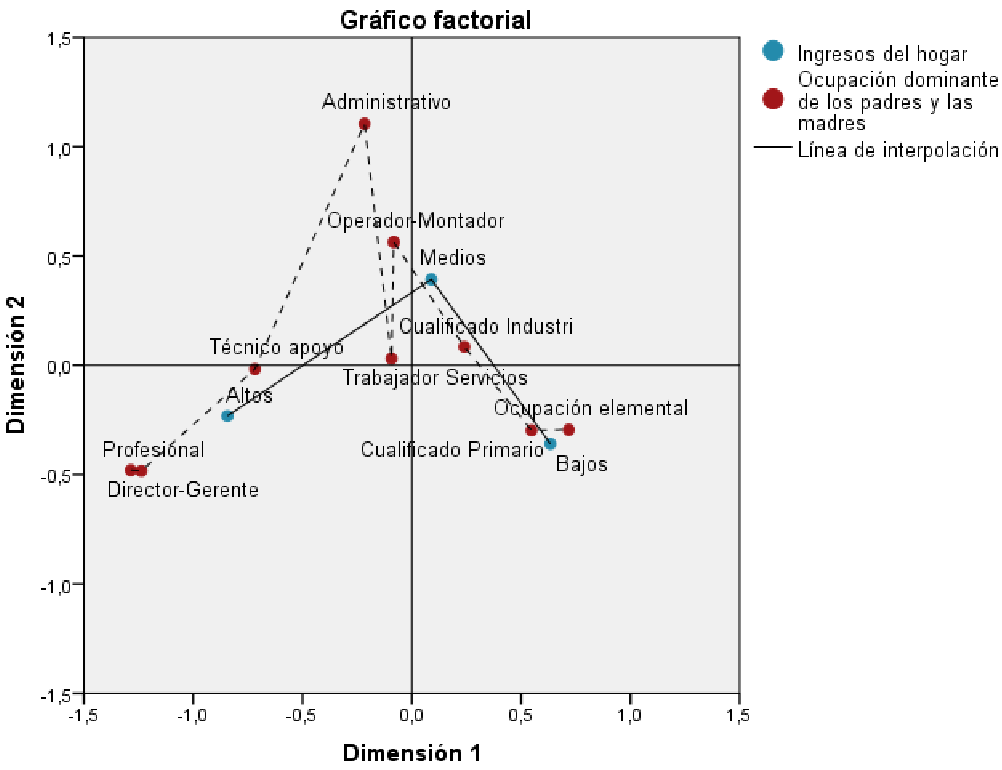




O bien directamente analizando las correspondencias entre las categorías de filas y columnas en el espacio bidimensional que consideramos en este análisis a través del gráfico factorial. El gráfico conjunto de los puntos de fila y columna que se genera es el siguiente:



El cual se puede editar para que adopte la forma siguiente donde hemos incluido la unión de los puntos a través de una línea de interpolación para destacar el recorrido de las categorías de cada variable:



Por último, se muestra una tabla interesante relativa a la permutación de la tabla de correspondencias inicial como resultado de analizar las categorías de las variables en la primera dimensión. Se procede, si es el caso, a reordenar las categorías de ambas variables a partir de las puntuaciones factoriales obtenidas. Podemos observar en este caso, comparando ambas tablas, el cambio entre Profesional y Director-Gerente así como la mejor posición alcanzada por Operador-Montador en relación a Cualificado primario y Cualificado Industria, esto es, en relación a la variable de ingresos, que conserva su orden inicial, las categorías citadas se asocian a un mayor nivel de ingresos lo que supone reordenarlas y cambiar el orden de codificación inicial de la variable. Estos cambios se pueden observar igualmente en el gráfico factorial anterior. En el caso de Profesional apenas hay diferencias con Director-Gerente, en el caso de Operador-Montador son más destacadas.

Tabla de correspondencias permutadas de acuerdo con la dimensión 1

IngresosH Ingresos del hogar	OCUPAFAM Ocupación dominante de los padres y las madres									
	Profesional	Director-Gerente	Técnico apoyo	Administrativo	Trabajador Servicios	Operador-Montador	Cualificado Industria	Cualificado Primario	Ocupación elemental	Margen activo
Altos	61	25	66	9	74	57	71	48	18	429
Medios	31	13	61	22	111	118	154	118	59	687
Bajos	11	5	25	5	71	50	113	127	67	474
Margen activo	103	43	152	36	256	225	338	293	144	1.590

Tabla de correspondencias

IngresosH Ingresos del hogar	OCUPAFAM Ocupación dominante de los padres y las madres									
	Director-Gerente	Profesional	Técnico apoyo	Administrativo	Trabajador Servicios	Cualificado Primario	Cualificado Industria	Operador-Montador	Ocupación elemental	Margen activo
Bajos	5	11	25	5	71	127	113	50	67	474
Medios	13	31	61	22	111	118	154	118	59	687
Altos	25	61	66	9	74	48	71	57	18	429
Margen activo	43	103	152	36	256	293	338	225	144	1.590

Si la variable fuera nominal, por tanto, sin un orden conceptualmente preestablecido en la variable, esta tabla de permutación y, en general, el resultado de un análisis de correspondencias cuando cuantifica las distintas categorías, proporciona una ordenación (y de cuantificación) de dichas categorías cualitativas nominales de la que carecía inicialmente, siempre como resultado de relacionarse con la otra variable.

Se propone como ejercicio reproducir los análisis de correspondencias simples presentados en el apartado 5.1.3. a partir de los datos de los archivos: España-Provincias x Educación.sav y España-Provincias x Educación-Sexo.sav, así como la matriz de datos Distancias España.sba y el archivo de sintaxis ACS-España.sps.

6.2. Correspondencias múltiples

Realizaremos en primer lugar un sencillo ejercicio para ilustrar el funcionamiento del análisis de correspondencias múltiples y sus características en SPSS. Trabajaremos con la matriz de datos Actitud.sav con 1443 individuos y 3 variables[39]. Las variables son: ACT (actitud: grado de acuerdo con la afirmación "Las mujeres deben quedarse en su casa"), EST (el nivel de estudios) y SEX (el sexo de la persona entrevistada).

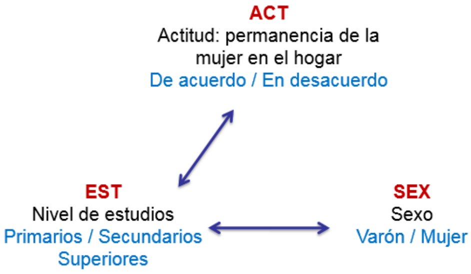
A partir de estos datos la tabla de contingencia que relaciona las tres variables se presenta a continuación:

Tabla de contingencia ACT x EST x SEX

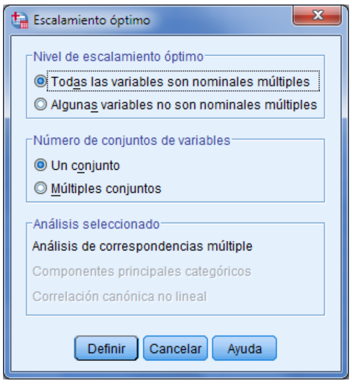
			SEX Sexo							
			1 Varón				2 Mujer			
			EST Nivel de estudios			Total	EST Nivel de estudios			Total
			1 Primarios	2 Secundarios	3 Superiores		1 Primarios	2 Secundarios	3 Superiores	
ACT Actitud: permanencia de la mujer en el hogar	1 De acuerdo	Recuento	72	110	44	226	86	173	28	287
		Frecuencia esperada	41,5	106,7	77,8	226,0	44,8	164,6	77,6	287,0
		% dentro de EST Nivel de estudios	**60,5%**	**35,9%**	**19,7%**	**35%**	**69,4%**	**37,9%**	**13,0%**	36,1%
		Residuos corregidos	6,5	,5	-5,9		8,4	1,3	-8,2	
	2 En desacuerdo	Recuento	47	196	179	422	38	283	187	508
		Frecuencia esperada	77,5	199,3	145,2	422,0	79,2	291,4	137,4	508,0
		% dentro de EST Nivel de estudios	**39,5%**	**64,1%**	**80,3%**	**65%**	**30,6%**	**62,1%**	**87,0%**	63,9%
		Residuos corregidos	-6,5	-,5	5,9		-8,4	-1,3	8,2	
Total		Recuento	119	306	223	648	124	456	215	795
		Frecuencia esperada	119,0	306,0	223,0	648,0	124,0	456,0	215,0	795,0
		% dentro de EST Nivel de estudios	100,0%	100,0%	100,0%	100%	100,0%	100,0%	100,0%	100%

El análisis de las relaciones entre las tres variables después de realizar una análisis log-lineal muestra la ausencia de interacción entre las tres variables y la configuración de un modelo de independencia condicional o doble asociación que relaciona la actitud y los estudios así como los estudios y el sexo, como se esquematiza en el gráfico siguiente.

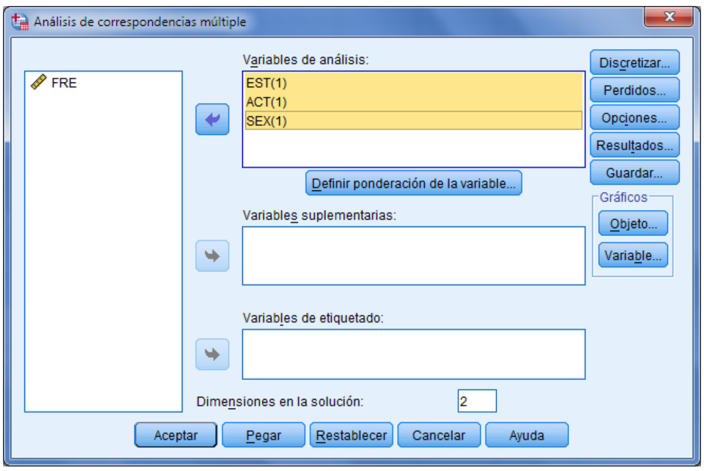
[39] A diferencia del mismo ejemplo que vimos en el tema de tablas de contingencia y log-lineal. Allí trabajamos con una matriz reducida de 12 casos correspondientes a las casillas de la tabla de contingencia. Esta alternativa se comenta en el capítulo III.2 al tratar los procedimientos de tratamiento de ficheros y en el capítulo III.6 de tablas de contingencia cuando se trata con SPSS la introducción de datos de una tabla de contingencia, donde es preciso ponderar los casos con el procedimiento del menú Ponderar casos o el comando **WEIGHT**. Ya se trabaje con los datos desagregados como aquí veremos como con los datos agregados, los resultados del análisis de correspondencias son los mismos.



Veamos cómo se expresa este resultado en términos de un análisis de correspondencias múltiples. Para obtener un ACM en SPSS se elige en el menú el Escalament óptimo: Analizar / Reducción de Dimensiones / Escalamiento óptimo. Al inicio se abre un cuadro de diálogo donde debemos escoger la modalidad de análisis de los tres procedimientos que nos ofrece, en función de si todas las variables son nominales (en general cualitativas) o bien alguna(s) son cuantitativas, y en función de si se trabaja con uno o más conjuntos de variables. En función de cada combinación disponemos de tres técnicas de análisis posibles: correspondencias múltiples (un conjunto con todas cualitativas), componentes categórico (un conjunto con alguna cuantitativa) y correlación canónica no lineal (más de un conjunto con con todas cualitativas o bien alguna cuantitativa):



La opción por defecto, que es la que presentaremos aquí, es la que corresponde al Análisis de correspondencias múltiple (comando **MULTIPLE CORRESPONDENCE**). Si hacemos clic en Definir aparece el cuadro de diálogo inicial de este procedimiento donde hemos seleccionado las tres variables activas que formarán parte del análisis y se ha trasladar el recuadro de Variables de análisis:



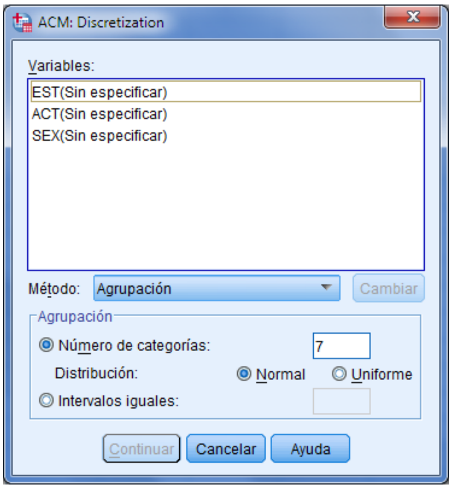
De forma optativa es posible Definir la ponderación del peso de cada variable. Por defecto, cada variable tendrá el mismo peso de 1 y así las trabjaremos habitualmente.

Una vez introducidas las variables tenemos la opción de usar Variables suplementarias, es decir, para introducir variables no activas que tendrían un papel ilustrador de las relaciones de éstas con la estructura de relaciones que emerge del conjunto de las variables activas que intervienen.

En el recuadro de Variables de etiquetado podemos incluir una variable que contenga etiquetas para los puntos-objeto (casos o individuos). Esta variable podría ser cualquiera que no forme parte del análisis (para utilizar variables del análisis como por ejemplo la edad, la ocupación, etc., se deberían crear nuevas variables que fueran una copia de éstas). En este caso no hemos especificado ninguna variable.

A continuación se plantea la decisión del número de dimensiones del análisis. El número máximo de dimensiones es el valor más pequeño de estos dos: o bien el número que sale de restar el número de categorías (12 en el ejemplo) menos el número de variables (3), o bien el número de casos (1443) menos uno. En este caso el número de dimensiones sería de 12 menos 3, es decir, 9. El número de dimensiones que considera por defecto el SPSS es de dos. Esta opción se puede considerar inicialmente pero es necesario un análisis con más dimensiones para determinar si éstas son relevantes para el estudio, sobre todo si el número de variablesy de categorías es más numeroso.

El cuadro de diálogo de Discretizar permite elegir un procedimiento para recodificar variables cuantitativas o variables definidas como cadena:



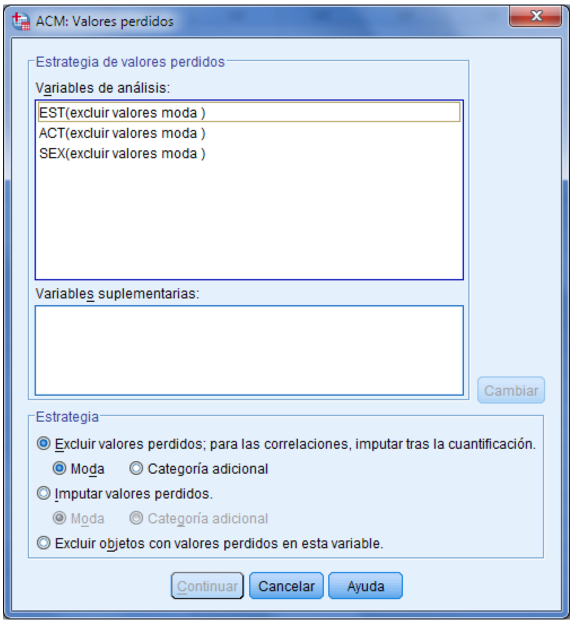
Se puede optar por tres métodos:

- Agrupación. Se recodifica en un número especificado de categorías (y eligiendo si los valores de la variable deben seguir una distribución aproximadamente normal o uniforme en estas categorías) o se recodifica por intervalos iguales. Las variables se recodifican en las categorías definidas por estos intervalos de igual tamaño especificando la longitud de los intervalos.
- Asignación de rangos. La variable se discretiza mediante la asignación de rangos a los casos.

- Multiplicación. Los valores actuales de la variable se tipifican, multiplican por 10, se redondean y se les suma una constante de manera que el menor valor discretizado sea 1.

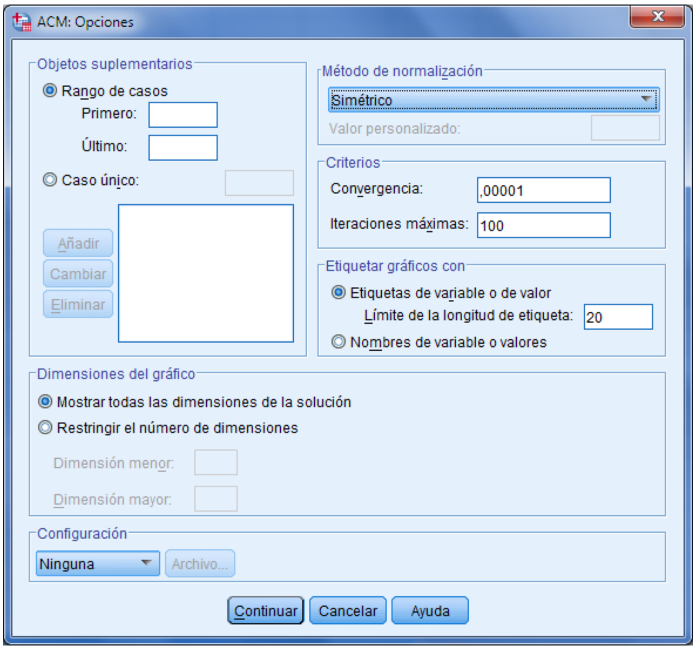
En este caso no hemos especificado ninguna discretización.

El cuadro de diálogo de Perdidos facilita el tratamiento de los valores perdidos en las variables de análisis y en las suplementarias. Se puede optar por excluir los valores perdidos (tratamiento pasivo), imputar los valores perdidos (tratamiento activo) o excluir objetos con valores perdidos (eliminación por lista):



En este ejemplo no tenemos valores perdidos, pero si existen requieren un análisis detenido para evaluar la implicaciones que pueden tener en los resultados.

En el cuadro de diálogo inicial aparece también el botón de Opciones para especificar varias opciones de esta técnica:



En primer lugar podemos escoger un rango de casos o de objetos suplementarios especificando el número de caso del objeto. Se puede especificar el método de normalización de las puntuaciones de objeto y de las variables entre cinco opciones:

- Principal por variable. Esta opción optimiza la asociación entre las variables y las categorías se sitúan en el centroide de los objetos de esa categoría. Esta opción es útil cuando el interés principal está en la correlación entre las variables.
- Principal por objeto. Esta opción optimiza las distancias entre los objetos. Esta opción es útil cuando el interés principal está en las diferencias y similitudes entre los objetos.
- Simétrico. Se utiliza esta opción de normalización si el interés principal está en la relación entre objetos y variables.
- Independiente. Se utiliza esta opción de normalización si se desea examinar separadamente las distancias entre los objetos y las correlaciones entre las variables.
- Personalizado. Se puede especificar cualquier valor real en el intervalo cerrado [-1, 1]. Un valor 1 es igual al método principal por objeto, un valor 0 es igual al método simétrico y un valor -1 es igual al método principal para variable. Si se especifica un valor mayor que -1 y menor que 1, se puede distribuir el autovalor entre los objetos y las variables.

Habitualmente emplearemos la opción de normalización simétrica, la considerada en este ejemplo.

En el apartado de Criterios se puede especificar el número máximo de iteraciones que el procedimiento puede realizar durante los cálculos. También puede seleccionar un valor para el criterio de convergencia. El algoritmo detiene la iteración si la diferencia del ajuste total entre la dos últimas iteraciones es menor que el valor de convergencia o si se ha alcanzado el número máximo de iteraciones.

En Etiquetar gráficos con se permite especificar si se utilizarán en los gráficos las etiquetas de variable y las etiquetas de valor o los nombres de variable y los valores. También se puede especificar una longitud máxima para las etiquetas.

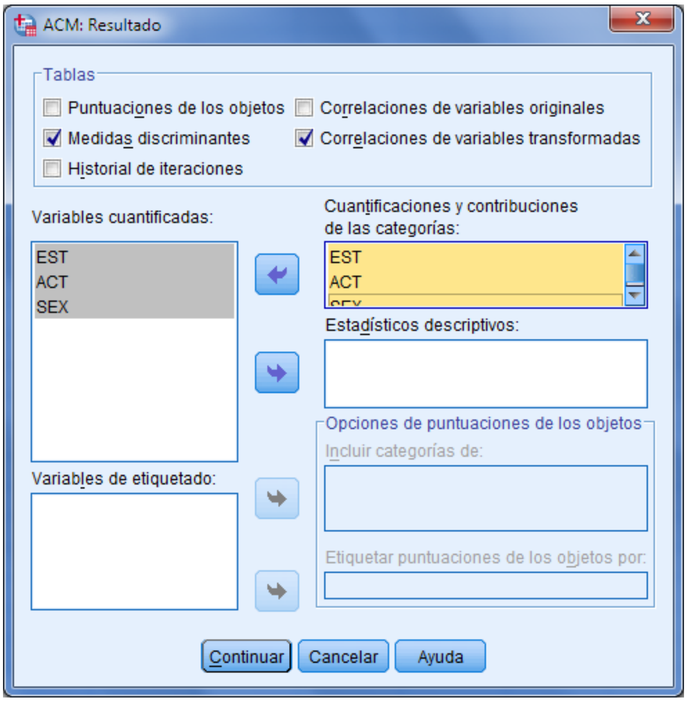
Con las Dimensiones del gráfico se controlan las dimensiones que se muestran en los resultados. O bien se muestran todas las dimensiones de la solución en un diagrama de dispersión matricial, o bien se restringe el número de dimensiones en los pares representados, seleccionando las dimensiones menor y mayor que se representarán. La dimensión menor puede variar desde 1 hasta el número de dimensiones de la solución menos 1 y se representa con respecto a las dimensiones mayores. El valor de la dimensión mayor puede oscilar desde 2 hasta el número de dimensiones de la solución e indica la dimensión mayor que se utilizará en representar los pares de dimensiones. Esta especificación se aplica a todos los gráficos multidimensionales solicitados.

Finalmente en Configuración se pueden leer datos de un archivo que contenga las coordenadas de una configuración. La primera variable del archivo deberá contener las coordenadas para la primera dimensión, la segunda variable las coordenadas para la segunda dimensión, y así sucesivamente. La opción Inicial implica que la configuración del archivo especificado se utilizará como el punto inicial del análisis. La opción Fija implica que la configuración del archivo especificado se utilizará para ajustar las variables. Las variables que se ajustan se deben seleccionar como variables de análisis, pero, al ser la configuración fija, se tratan como variables suplementarias (de forma que no es necesario seleccionarlas como variables suplementarias).

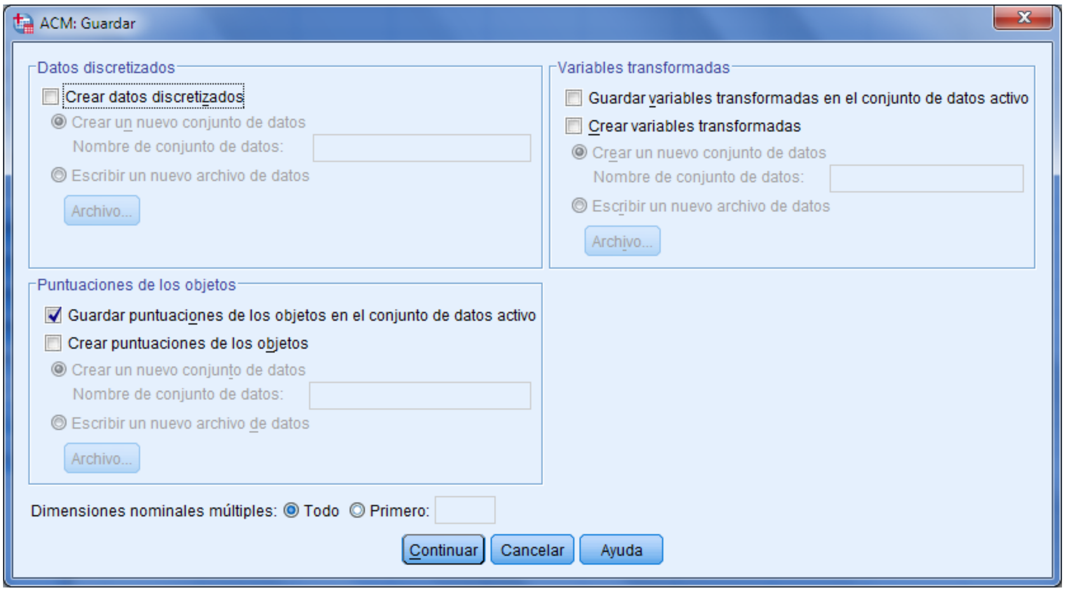
El botón de Resultados facilita mostrar varias opciones de información del análisis:

- Ver las puntuaciones de los objetos (incluidos la masa, la inercia y las contribuciones).
- Las medidas de discriminación por variable y por dimensión.
- El historial de iteraciones, y en cada iteración se muestra la varianza explicada, la pérdida y el incremento en la varianza explicada.
- La matriz de correlaciones de las variables originales y los autovalores de dicha matriz.
- La matriz de correlaciones de las variables transformadas (mediante escalamiento óptimo) y los autovalores de dicha matriz.
- Las cuantificaciones de las categorías (coordenadas, incluidas las masas, las inercias y las contribuciones) para cada dimensión de las variables seleccionadas.
- Estadísticos descriptivos. Muestra las frecuencias, el número de valores perdidos y la moda de las variables seleccionadas.

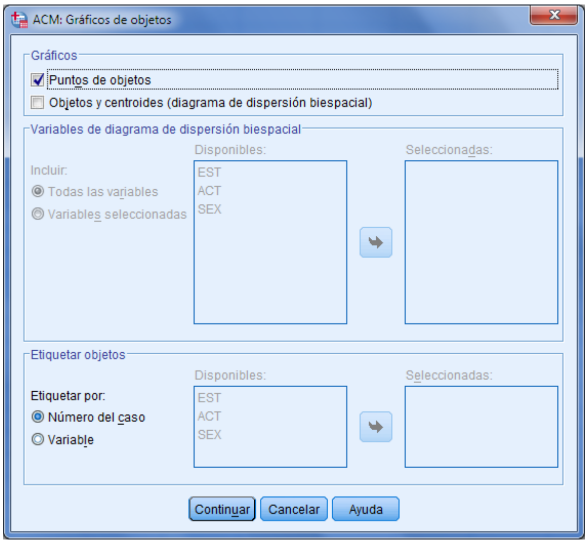
Consideraremos las opciones por defecto del procedimiento con la inclusión las variables activas en el recuadro de cuantificaciones y contribuciones de las categorías:

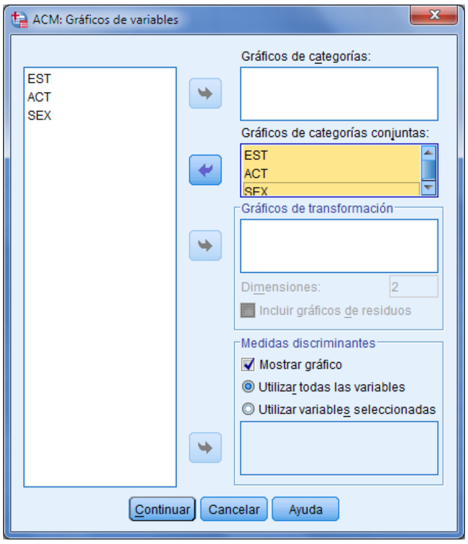


En Guardar se pueden almacenar distintas variables del análisis: los datos discretizados, las puntuaciones de los objetos y los valores transformados en un archivo de datos de SPSS, el activo u otro. En este caso se guardan las dos variables factoriales (puntuaciones) retenidas, añadiéndose a la matriz de datos activa:



Se pueden especificar finalmente en Gráficos representaciones de los resultados referidos a los objetos o las variables.





En relación a los objetos (casos) podemos pedir:

- Puntos de objetos: opción activada por defecto con los casos representados enel espacio de los factores o dimensiones.
- Objetos y centroides (diagrama de dispersión biespacial): los puntos de objetos se representan con los centroides de las variables.
- Variables de diagrama de dispersión biespacial: se pueden utilizar todas las variables para los gráficos de dispersión biespacial o seleccionar un subconjunto.
- Etiquetar objetos: se puede elegir que los objetos se etiqueten con las categorías de las variables seleccionadas (se pueden seleccionar entre los valores del indicador de

categoría o las etiquetas de valor, en el cuadro de diálogo de opciones) o con sus números de caso. Se genera un gráfico para cada variable si se selecciona Variable.

En relación a las variables podemos pedir:

- Gráficos de categorías: para cada variable seleccionada, se representa un gráfico de las coordenadas del centroide. Las categorías se encuentran en los centroides de los objetos de las categorías particulares.
- Gráficos de categorías conjuntas: este es un único gráfico de las coordenadas del centroide de cada variable seleccionada.
- Gráficos de transformación: muestra un gráfico de las cuantificaciones de las categorías óptimas en oposición a los indicadores de las categorías. Se puede especificar el número de dimensiones deseado; se generará un gráfico para cada dimensión. También se puede seleccionar si se muestran los gráficos de los residuos para cada variable seleccionada.
- Medidas de discriminación. Genera un único gráfico de las medidas de discriminación de las variables seleccionadas. Por defecto está activado.

En este caso hemos incluido todas las variables para obtener un gráfico de categorías conjuntas.

A través del menú de este procedimiento no se pueden contemplar más especificaciones, pero si usamos el lenguaje de comandos de SPSS podemos ampliar las opciones, en concreto:

- Especificar los pares de dimensiones que se van a representar, en lugar de representar todas las dimensiones extraídas, mediante la palabra clave NDIM del subcomando PLOT. En particular, para obtener diagramas de dispersión de los pares de dimensiones en lugar de un diagrama de dispersión matricial se puede utilizar la palabra clave NDIM con el subcomando PLOT: NDIM (d1, d2), que genera diagramas de dispersión de dimensión *d1* ante todas las dimensiones superiores hasta *d2*.
- Especificar nombres de raíz para las variables transformadas, puntuaciones de objetos y aproximaciones al guardarlas en el conjunto de datos activo (con el subcomando SAVE).
- Especificar una longitud máxima de las etiquetas para cada gráfico por separado (con el subcomando PLOT).
- Especificar una lista de variables distinta para los gráficos de residuos (con el subcomando PLOT).

El programa de instrucciones de SPSS que se corresponde con el ejemplo que hemos presentado través de los menús es el siguiente:

```
MULTIPLE CORRESPONDENCE VARIABLES=ACT EST SEX
 /ANALYSIS=ACT(WEIGHT=1) EST(WEIGHT=1) SEX(WEIGHT=1)
 /MISSING=ACT(PASSIVE,MODEIMPU) EST(PASSIVE,MODEIMPU) SEX(PASSIVE,MODEIMPU)
 /DIMENSION=2
 /NORMALIZATION=SYMMETRICAL
 /MAXITER=100
 /CRITITER=.00001
 /PRINT=CORR DISCRIM QUANT(ACT EST SEX)
 /PLOT=OBJECT(20) JOINTCAT(ACT EST SEX) (20) DISCRIM (20)
 /SAVE=OBJECT.
```

A continuación presentamos y comentamos los resultados más relevantes que se obtienen del análisis.

Crédito

Multiple Correspondence
Version 1.0
by
Leiden SPSS Group
Leiden University
The Netherlands

Resumen de procesamiento de casos

Casos activos válidos	1.443
Casos activos con valores perdidos	0
Casos complementarios	0
Total	1.443
Casos utilizados en análisis	1.443

Tras los créditos y la información sobre los casos procesados se informa del historial de iteraciones del algoritmo iterativo que genera los resultados del análisis:

Historial de iteraciones

	Varianza contabilizada para		
Número de iteración	Total	Aumentar	Pérdidas
31[a]	1,217873	0,000007	1,782127

a. El proceso de iteración se ha detenido porque se ha alcanzado el valor de prueba de convergencia.

Se presenta seguidamente la información de las dimensiones o factores retenidos en el análisis factorial de correspondencias múltiples. En este caso retuvimos dos factores que acumulan el 81,2% de la inercia (o varianza total explicada sin corregir), es decir, el total de los valores propios o autovalores (2,436) dividido por la inercial total (el número de variables, 3), 0,812, y multiplicado por 100. El programa proporciona los valores propios o autovalores correspondientes con el porcentaje de varianza explicada por cada dimensión.

Resumen del modelo

		Varianza contabilizada para		
Dimensión	Alfa de Cronbach	Total (autovalor)	Inercia	% de varianza
1	0,379	1,338	0,446	44,599
2	0,134	1,098	0,366	36,593
Total		2,436	0,812	
Media	0,268[a]	1,218	0,406	40,596

a. La media de alfa de Cronbach se basa en la media de autovalor.

Adjuntamos también la tabla de valores propios completa con todas las dimensiones posibles en el análisis: el número de categorías (m) menos el de variables (p), sin considerar los valores perdidos, en este caso: $m-p = 7-3 = 4$.

Resumen del modelo

		Varianza contabilizada para		
Dimensión	Alfa de Cronbach	Total (autovalor)	Inercia	% de varianza
1	0,379	1,338	0,446	44,599
2	0,134	1,098	0,366	36,594
3	-0,166	0,900	0,300	30,009
4	-0,759	0,664	0,221	22,132
Total		4,000	1,333	
Media	0,000[a]	1,000	0,333	33,333

a. La media de alfa de Cronbach se basa en la media de autovalor.

El alfa de Cronbach es un índice de fiabilidad, de consistencia interna, que se obtiene calculando la media ponderada de las correlaciones entre las variables que forman parte de cada dimensión para reflejar en qué medida un conjunto de elementos (variables) miden unidimensional una dimensión latente. Su valor oscila entre 0 y 1, y cuanto más se acerca 1 mayor es la fiabilidad de la escala y cuando los datos tienen una estructura multidimensional el alfa de Cronbach por lo general será bajo. Valores superiores a 0,6 son suficientes. En el ejemplo se generan valores bajos derivados de la baja asociación de las variables con cada eje.

Sobre el número de dimensiones factoriales el análisis de correspondencias no nos da indicaciones sobre cuántos factores retener. Recordemos que el número de factores debe ser objeto de análisis y validación como explicamos en apartados anteriores. Habitualmente se consideran tan sólo 2 dimensiones como opción por defecto. Un primer criterio fundamental es considerar los factores en función de su contenido y de la relevancia teórico-conceptual. Por este debe basado también en criterios formales adicionales. Un criterio adicional general es acumular como mínimo el 70% de la varianza. En el ejemplo siguiente aplicaremos el criterio que explicamos de Benzécri para corregir la infraestimación que supone un ACM.

Una vez seleccionado el número de dimensiones tenemos varias tablas estadísticas que nos ayudan a identificar y caracterizar las dimensiones retenidas. En primer lugar tenemos las medidas de discriminación que nos relacionan las variables originales con las dimensiones obtenidas en términos de varianza cuando es cuantificada óptimamente cada variable en cada dimensión. Cuanto mayor sea este valor mayor será la importancia de ésta en la definición del factor teniendo en cuenta que el valor máximo de esta medida sería el 1 (situación en la que todas las puntuaciones de los sujetos caen en grupos mutuamente excluyentes y que dentro de cada grupo todas las puntuaciones son idénticas). Por otra parte, y de forma similar a como sucede en el análisis de componentes, la suma de todos los valores a lo largo de una dimensión, en este caso divididos por el número de variables, es igual al valor propio asociado a la dimensión o eje factorial.

En el caso del primer valor propio (0,446) este es igual a:

$$0{,}446 = (0{,}666 + 0{,}663 + 0{,}009) \ / \ 3$$

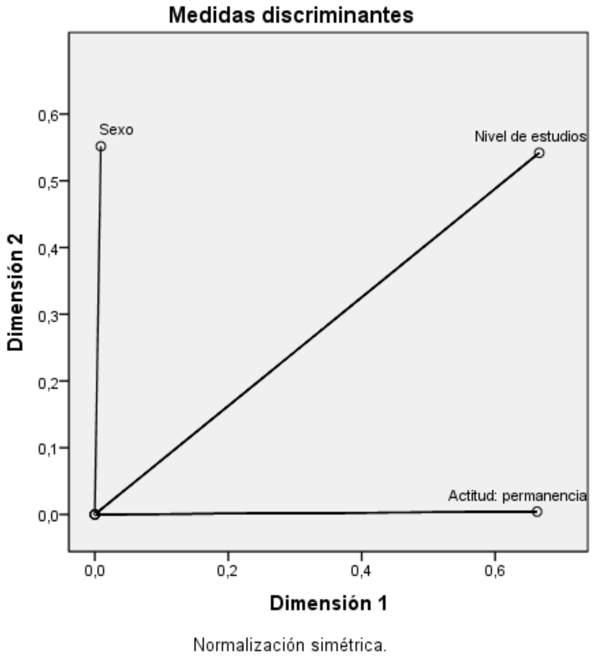
En la tabla de medidas discriminante se presta atención a las variables que tienen un valor más alto para cada dimensión: EST y ACT se asocian al factor 1 y EST y SEX al factor 2.

Medidas discriminantes

	Dimensión		
	1	2	Media
ACT Actitud: permanencia de la mujer en el hogar	0,663	0,004	0,333
EST Nivel de estudios	0,666	0,567	0,617
SEX Sexo	0,009	0,527	0,268
Total activo	1,338	1,098	1,218
% de varianza	44,599	36,593	40,596

Esta tabla también nos sirve para detectar posibles variables que tienen una baja contribución en la explicación de las diferentes dimensiones y, por tanto, puede considerarse como neutral o bien como candidata a ser eliminada del análisis si no manifiesta correspondencias con el resto de variables. En estos casos se trata de una variable con modalidades que se sitúan muy cerca del origen en la representación de los distintos ejes factoriales manifestando falta de poder de discriminación.

Estas medidas se pueden representar gráficamente. Cuando se piden más de dos dimensiones y se ejecuta el programa a través de los menús se obtiene un gráfico de dispersión múltiple o matricial y es necesario utilizar la sintaxis para generar los gráficos bidimensionles por separado. En el caso que nos ocupa tenemos solamente dos dimensiones generándose el gráfico siguiente que expresa el contenido de la tabla:



La variable relativa a la actitud se asocia al primer factor mientras que el sexo lo hace al segundo. La variable de educación muestra relaciones tanto con la primera como con la segunda dimensión por lo que se representa en diagonal entre la horizontalidad de la primera dimensión y la verticalidad de la segunda.

En el caso de soluciones factoriales de más de dos dimensiones las representaciones gráficas no suelen ser muy claras. Si se quiere ver con más precisión las representaciones entre parejas de dimensiones hay que ejecutar el procedimiento a través del lenguaje de comandos y especificar NDIM en el subcomando PLOT. En particular, si queremos obtener la representación entre la primera y la segunda dimensión y entre la primera y la tercera, de un conjunto de tres dimensones o factores retenidos, se tiene que especificar NDIM (1,3), cambiando la línea:

```
/PLOT=OBJECT(20) JOINTCAT(EST ACT SEX) (20) DISCRIM (20)
```

por:

```
/PLOT=OBJECT(20) JOINTCAT(EST ACT SEX) (20) DISCRIM (20) NDIM(1,3)
```

Pero además de la información de las variables en relación a las dimensiones, más interesante es ver cómo se relacionan las modalidades o categorías de cada variable con las dimensiones retenidas, son las llamadas cuantificaciones de las categorías. Estas cuantificaciones representan el promedio de las puntuaciones de todos los individuos (objetos) en cada dimensión y permiten precisar y mejorar la interpretación de las dimensiones. A continuación aparecen las tablas que, para cada variable, presentan, primero, la frecuencia de cada categoría, y después, las cuantificaciones los ejes o dimensiones.

ACT Actitud: permanencia de la mujer en el hogar

Puntos: Coordenadas

		Coordenadas del centroide	
		Dimensión	
Categoría	Frecuencia	1	2
1 De acuerdo	513	-1,641	0,143
2 En desacuerdo	930	0,905	-0,079

Normalización simétrica.

EST Nivel de estudios

Puntos: Coordenadas

		Coordenadas del centroide	
		Dimensión	
Categoría	Frecuencia	1	2
1 Primarios	243	-2,186	1,701
2 Secundarios	762	-0,160	-1,160
3 Superiores	438	1,491	1,075

Normalización simétrica.

SEX Sexo

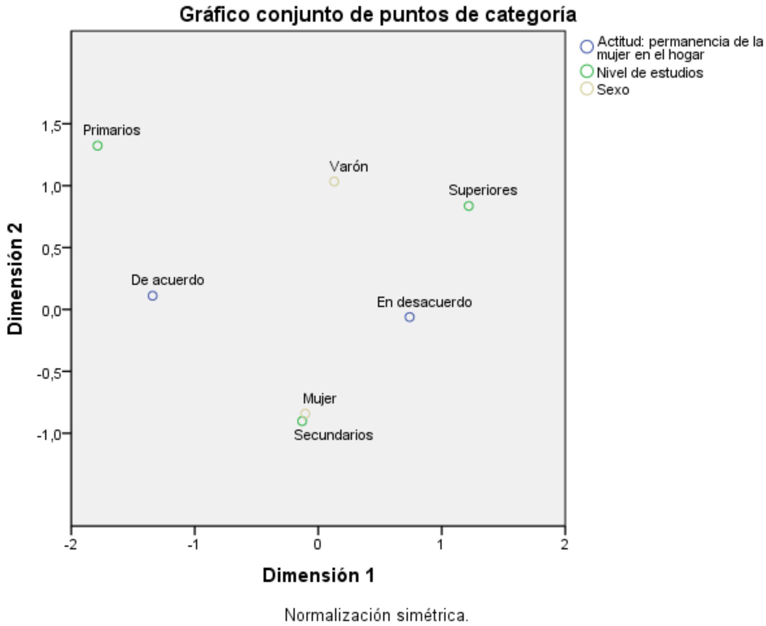
Puntos: Coordenadas

		Coordenadas del centroide	
		Dimensión	
Categoría	Frecuencia	1	2
1 Varón	648	0,158	1,329
2 Mujer	795	-0,129	-1,083

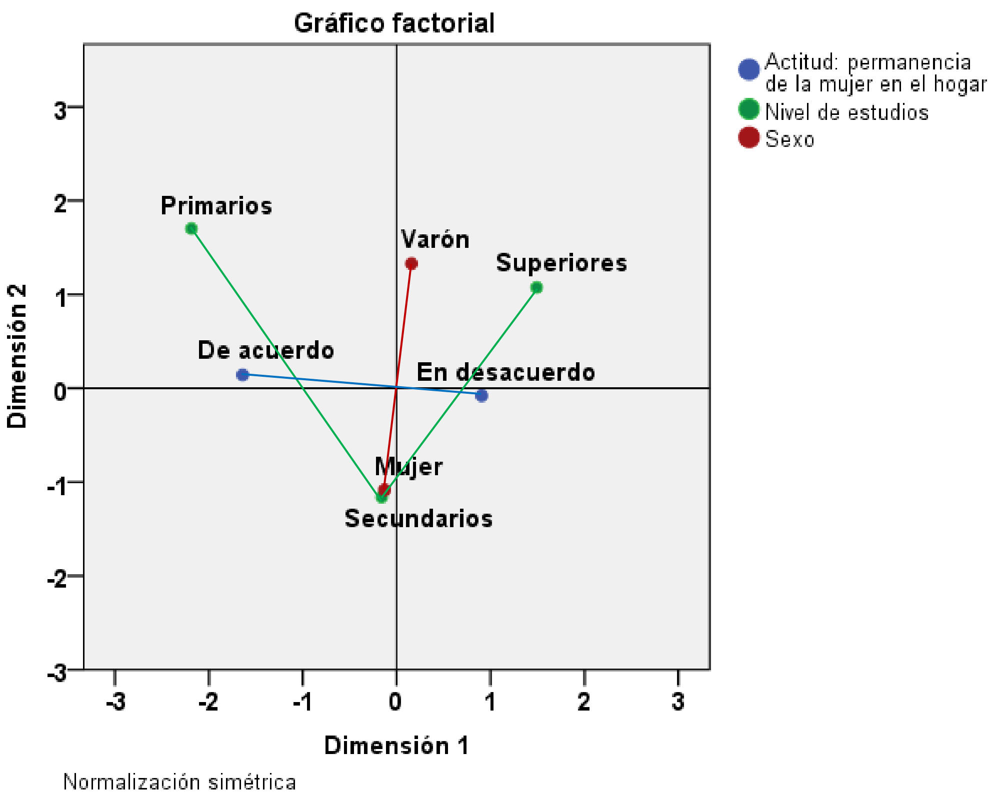
Normalización simétrica.

En el caso por ejemplo de la primera variable (ACT) tenemos que los que están de acuerdo son 513 personas en la muestra y que estos tienen un valor negativo en la primera dimensión de −1,641, pero positivo y mucho menos elevado en la segunda, próximo a cero, al centro del eje en esa dimensión. Es decir, quien manifiesta estar de acuerdo se identifica sobre todo, o contribuye sobre todo, a definir la primera dimensión en el extremo negativo. Lo mismo sucedería para los que están en desacuerdo pero en el lugar opuesto a los hombres en el extremo positivo.

Estos valores son las coordenadas de las categorías en los ejes factoriales y de nuevo, cuando más alto sea este valor más contribuirá a definir la dimensión y gráficamente más alejado estará el punto de origen. Las categorías con similares valores o coordenadas (cuantificaciones) implican que son próximas y homogéneas, que están asociadas entre sí en el mismo sentido, y gráficamente ocupan espacios próximos y opuestos a otras categorías que son heterogéneas respetos a ellas. Por lo tanto el análisis consiste en recorrer estas tablas para ver qué categorías se asocian a qué ejes y, por la acumulación de categorías, contribuir a definir los extremos de cada uno de los ejes (definir cada dimensión) y derivar así una estructura de relaciones que configuran el conjunto de las dimensiones factoriales consideradas. Para ayudar en esta lectura se dispone de la representación gráfica de las cuantificaciones, que son los valores de coordenadas. El gráfico de los puntos categoría se presenta seguidamente, tal y como se obtiene en primera instancia:



y después de ser editado cambiando la escala, las etiquetas y marcas y añadiendo las líneas entre las categorías:

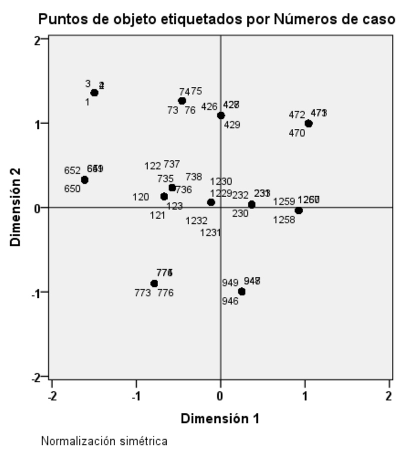


De estas informaciones derivamos la centralidad del primer factor que expresa la principal fuente de asociación entre las variables (45%): la primera dimensión, la horizontal, muestra la asociación entre estar de acuerdo o en desacuerdo con el nivel de estudios, expresa la oposición entre los que están mayoritariamente a favor, las personas con bajos niveles de estudios, frente a las personas que en mayor medida están en contra, aquéllas que tienen estudios superiores. Por su parte las personas con estudios medios aparecen en el centro de la primera dimensión pues son individuos que se reparten entre los que están a favor y en contra en la misma proporción que el conjunto de toda la muestra. De forma equivalente, la variable sexo y sus categorías, varones y mujeres, se sitúan en el centro del eje horizontal, hecho que muestra la actitud de ellos y ellas es prácticamente igual y se reparten entre los que se pronuncian a favor o en contra en la misma proporción que el conjunto o el promedio de los datos, nos indica por tanto que la variable sexo no discrimina entre estar de acuerdo o en desacuerdo, es decir, que tanto ellos como ellas se pronuncian de forma similar.

Donde se dan las diferencias por género es en el segundo factor. La segunda dimensión no tiene en consideración a la actitud como variable que acumule inercia a través de sus categorías, se sitúan en el centro con respecto del eje vertical. Este segundo factor, con una importancia menor del 37%, revela el vínculo que se produce entre el nivel de estudios y el sexo: los varones mayoritariamente tiene estudios inferiores y superiores mientras que las mujeres tienen sobre todo estudios medios.

Si relacionamos estas conclusiones con los resultados del análisis log-lineal comprobamos cómo el primer factor expresa la asociación bivariable entre actitud y estudios, mientras que el segundo la muestra la asociación entre estudios y sexo, los dos pares de relaciones significativas que emergieron del log-lineal. La relación entre actitud y sexo no apareció como significativa, es decir, ambas variables resultaron independientes. La independencia en el ACM se expresa gráficamente por la disposición perpendicular de las categorías de ambas variables; en este caso vemos como la línea punteada vertical es perpendicular a la continua horizontal, forman un ángulo de prácticamente de 90º, por lo que su correlación, el coseno de 90, es cero.

Finalmente tenemos la posibilidad de representar los individuos (objetos) en los ejes factoriales. En relación al análisis y tratamiento de los individuos insistiremos el próximo tema cuando hablamos de análisis de clasificación. Aquí se presenta el gráfico que sale considerar las 2 dimensiones consideradas. Los puntos representan simplemente la mayor o menor concentración de efectivos.



6.2.1. *Análisis de la segmentación del mercado de trabajo*

Consideramos ahora el análisis que se presentó en el apartado 5.2 sobre segmentación del mercado de trabajo para ejecutarlo con el software SPSS. Para reproducir el análisis con los datos de la Encuesta de Población Activa de 2014 a partir de la selección de variables que se presentan en la Tabla III.11.16 y en la Tabla III.11.17 se dispone de la matriz de datos EPA4T2014-Segmentación.sav junto con el archivo de sintaxis Segmentación ACM+ACL.sps que realiza el análisis de correspondencias junto con el de clasificación que se verá en el próximo capítulo.

Se consideran 8 variables activas en el análisis (CONTRATO, JORNADA DURACION, ANTIGÜEDAD, OCUPACION, ACTIVIDAD, EMPRESA y EDUCACION) y 3 variables sumplementarias o ilustrativas que complementan la interpretación de los resultados sin formar parte del cálculo de las dimensiones o factores (EDAD, SEXO y NACIONALIDAD).

Para determinar el número de dimensiones a retener utilizamos como criterio general, junto con la interpretación sustantiva de los resultados, alcanzar al menos un 70% de varianza o inercia explicada. En el caso del SPSS la información completa de todos los valores propios correspondientes a todos los factores no aparece como información, tan solo se presentan los datos en relación a las dimensiones retenidas. Por defecto, además, considera solamente dos factores.

Por otra parte, siguiente el razonamiento expuesto en el apartado 5.2.2, ante la infraestimación de la importancia relativa de los factores que supone la técnica en su versión múltiple, nosotros proponemos un procedimiento de determinación de número de factores transformando los valores propios iniciales en unos nuevos valores propios corregidos que expresen el peso real que tienen esos factores (transformación de Benzécri). Para ello necesitamos, en primer lugar, disponer de la información completa relativa al Resumen del modelo, tabla donde se lista la información sobre la varianza explicada por cada factor o dimensión retenida. Sabemos que el número máximo de dimensiones en un ACM es el número de categorías (m) menos el de variables (p), sin considerar los valores perdidos, en este caso: $m-p = 40-8 = 32$. Así pues, ejecutaremos primeramente un análisis de correspondencias múltiples solicitando 32 dimensiones.

La tabla que se obtiene de resumen del modelo se presenta seguidamente. A partir de ella copiaremos la información de la columna Inercia y la pegaremos en la columna destacada en rojo Valor propio del documento en Excel VPCorregidos.xlsx: como se muestra a continuación en la tabla.

De los 32 valores propios iniciales seleccionamos aquellos que son superiores a $1/p$, es decir, $1/8=0{,}125$, un total de 13, y aplicamos la transformación de Benzécri de la Ecuación 22. En la tabla siguiente vemos la distribución del nuevo porcentaje de varianza explicada por los 13 primeros factores.

Resumen del modelo

Dimensión	Alfa de Cronbach	Varianza contabilizada para Total (autovalor)	Inercia
1	0,765	3,022	0,378
2	0,641	2,277	0,285
3	0,472	1,702	0,213
4	0,399	1,537	0,192
5	0,324	1,396	0,175
6	0,273	1,314	0,164
7	0,208	1,223	0,153
8	0,178	1,185	0,148
9	0,090	1,086	0,136
10	0,072	1,067	0,133
11	0,035	1,031	0,129
12	0,031	1,028	0,128
13	0,008	1,007	0,126
14	-0,003	0,997	0,125
15	-0,007	0,994	0,124
16	-0,016	0,987	0,123
17	-0,040	0,966	0,121
18	-0,062	0,948	0,119
19	-0,078	0,936	0,117
20	-0,102	0,918	0,115
21	-0,141	0,890	0,111
22	-0,278	0,804	0,101
23	-0,340	0,771	0,096
24	-0,472	0,708	0,088
25	-0,541	0,679	0,085
26	-0,584	0,662	0,083
27	-0,947	0,547	0,068
28	-1,473	0,437	0,055
29	-1,755	0,394	0,049
30	-2,372	0,325	0,041
31	-4,240	0,212	0,027
32	0,000	0,000	0,000
Total		32,051	4,006
Media	0,002[a]	1,002	0,125

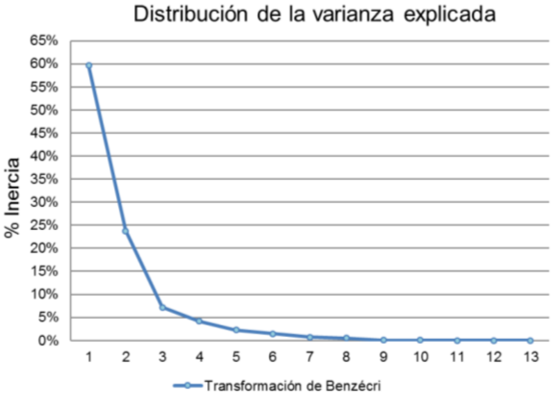
a. La media de alfa de Cronbach se basa en la media de autovalor

Factor	Valor propio	% Inercia
1	0,377752	9,43%
2	0,284649	7,10%
3	0,212794	5,31%
4	0,192172	4,80%
5	0,174505	4,36%
6	0,164235	4,10%
7	0,152826	3,81%
8	0,148095	3,70%
9	0,135712	3,39%
10	0,133431	3,33%
11	0,128916	3,22%
12	0,128474	3,21%
13	0,125919	3,14%
14	0,124679	3,11%
15	0,124241	3,10%
16	0,123324	3,08%
17	0,120760	3,01%
18	0,118524	2,96%
19	0,116992	2,92%
20	0,114801	2,87%
21	0,111232	2,78%
22	0,100547	2,51%
23	0,096314	2,40%
24	0,088451	2,21%
25	0,084848	2,12%
26	0,082703	2,06%
27	0,068368	1,71%
28	0,054622	1,36%
29	0,049305	1,23%
30	0,040640	1,01%
31	0,026539	0,66%
32	0,000000	0,00%
Suma	**4,006371**	**100%**

De los 32 valores propios iniciales seleccionamos aquellos que son superiores a *1/p*, es decir, *1/8=0,125*, un total de 13, y aplicamos la transformación de Benzécri de la *Ecuación 22*. En la tabla siguiente vemos la distribución del nuevo porcentaje de varianza explicada por los 13 primeros factores.

Factor	Valor propio	% Inercia	Valor propio corregido (*)	% Inercia (1)	% Acumulado
1	0,377752	9,43%	0,083440	**59,65%**	59,65%
2	0,284649	7,10%	0,033290	**23,80%**	**83,44%**
3	0,212794	5,31%	0,010067	**7,20%**	90,64%
4	0,192172	4,80%	0,005893	**4,21%**	94,85%
5	0,174505	4,36%	0,003201	**2,29%**	97,14%
6	0,164235	4,10%	0,002011	**1,44%**	98,58%
7	0,152826	3,81%	0,001011	**0,72%**	99,30%
8	0,148095	3,70%	0,000697	**0,50%**	99,80%
9	0,135712	3,39%	0,000150	**0,11%**	99,91%
10	0,133431	3,33%	0,000093	**0,07%**	99,97%
11	0,128916	3,22%	0,000020	**0,01%**	99,99%
12	0,128474	3,21%	0,000016	**0,01%**	100,00%
13	0,125919	3,14%	0,000001	**0,00%**	100,00%

Si consideraremos el criterio cuantitativo de alcanzar el 70% de la varianza explicada retendremos los dos primeros factores que acumulan el 83,44% de la varianza total. Esta conclusión se ve reforzada por la inspección visual del gráfico de sedimentación que podemos elaborar los valores propios transformados:



Con la decisión del número de dimensiones volvemos a ejecutar el ACM y obtenemos los resultados siguientes. En primer lugar, observamos la varianza explicada por los dos primeros factores conservados y los indicadores aceptables que proporciona el Alfa de Cronbach para cada dimensión:

Resumen del modelo

Dimensión	Alfa de Cronbach	Varianza contabilizada para	
		Total (autovalor)	Inercia
1	0,765	3,022	0,378
2	0,641	2,277	0,285
Total		5,299	0,662
Media	0,712[a]	2,650	0,331

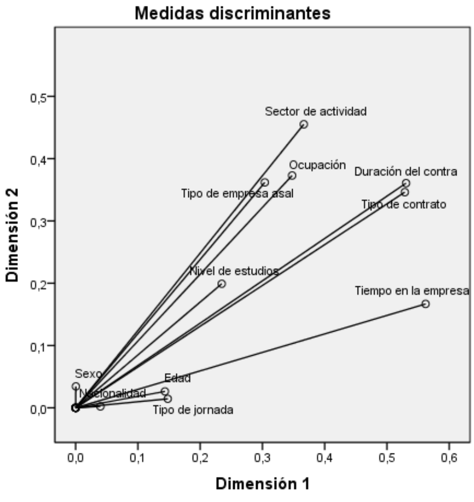
a. La media de alfa de Cronbach se basa en la media de autovalor.

Las medidas discriminantes nos muestran a qué dimensión se asocia cada variable original, información que se representa también gráficamente:

Medidas discriminantes

	Dimensión 1	Dimensión 2	Media
Tipo de contrato	**0,529**	0,346	**0,437**
Tipo de jornada	**0,148**	0,015	0,081
Duración del contrato	**0,531**	0,361	**0,446**
Tiempo en la empresa	**0,562**	0,167	**0,364**
Ocupación	0,348	**0,373**	**0,360**
Sector de actividad	0,366	**0,455**	**0,411**
Tipo de empresa asalariados	0,304	**0,362**	**0,333**
Nivel de estudios	**0,235**	0,199	0,217
Edad[a]	**0,143**	0,026	0,085
Sexo[a]	0,000	**0,034**	0,017
Nacionalidad[a]	**0,040**	0,002	0,021
Total activo	3,022	2,277	2,650

a. Variable complementaria.



Normalización simétrica.

Como se puede observar, la duración del contrato, el tipo de contrato y el sector de actividad son las variables con mayor capacidad de discriminación, sus valores de media total (0,446, 0,437 y 0,411, respectivamente) son los más elevados. El tiempo en la empresa y la duración del contrato se asocian en mayor medida al primer factor, mientras que el sector de actividad lo hace con el segundo. El tiempo en la empresa, el tipo de empresa y la categoría ocupacional también muestran valores elevados. La primera variable marcando el carácter del primer factor y las otras dos algo más relacionadas con el segundo. El resto de variables expresan contribuciones menos importantes como se puede observar en el gráfico por su proximidad al origen (0,0).

La información detallada de las categorías de estas variables se puede analizar a partir de las Cuantificaciones. Las tablas siguientes, con las variables activas en el análisis y la variables suplementarias, que son de ayuda en la interpretación de los factores, se aprecia la posición de cada categoría en cada dimensión, sus coordenadas.

Variables activas del análisis

Categoría	Frecuencia	Coordenadas del centroide Dimensión 1	 2
Tipo de contrato			
Indefinido	36.228	0,669	0,622
Temporal	11.608	-2,091	-1,951
Tipo de jornada			
Completa	39.472	0,287	0,101
Parcial	8.364	-1,360	-0,493
Duración del contrato			
Indefinido	36.228	0,669	0,622
Hasta 1 mes	5.622	-2,260	-1,659
Hasta 6 meses	2.921	-2,429	-1,862
Más de 6 meses	2.602	-1,391	-2,842
Perdidos	463		
Tiempo en la empresa			
Hasta 1 año	8.083	-2,251	-1,539
2-3 años	5.199	-0,979	0,049
4-10 años	13.326	0,116	0,549
11-20 años	10.862	0,666	0,637
Más de 20 años	10.366	1,397	-0,209
Ocupación			
Directivos	1.188	1,061	0,752
Profesionales	8.590	1,455	-2,190
Técnicos apoyo	5.081	0,518	0,205
Administrativos	5.587	0,598	0,220
Trabajador servicios	10.524	-0,420	0,074
Cualificado primario	525	-1,306	0,419
Cualificado industria	5.069	-0,667	1,268
Operadores	3.854	-0,346	1,724
Elementales	7.080	-1,489	0,277
Perdidos	338		
Sector de actividad			
Primario	1.464	-2,277	-0,234
Industria 1	2.518	-0,421	1,601
Industria 2	2.862	0,056	1,471
Industria 3	2.089	0,049	1,269
Construcción	2.291	-1,112	0,722
Comercio	10.034	-0,782	0,745
Transporte-Comunicaciones	3.283	0,166	0,697
Financiero-Profesionales	5.588	-0,108	0,541
Administración	14.061	1,325	-1,884
Otros servicios	3.646	-1,119	0,378
Tipo de empresa asalariados			
Pública	11.368	1,606	-2,022
Privada	36.468	-0,501	0,627
Nivel de estudios			
Primaria incompleta	557	-1,469	0,622
Primaria	2.596	-1,115	0,852
Secundaria 1a	13.279	-0,872	0,912
Secundaria 2a General	6.176	0,052	0,396
Secundaria 2a Profesional	4.775	-0,442	0,450
Superior	20.453	0,834	-0,948

Normalización simétrica.

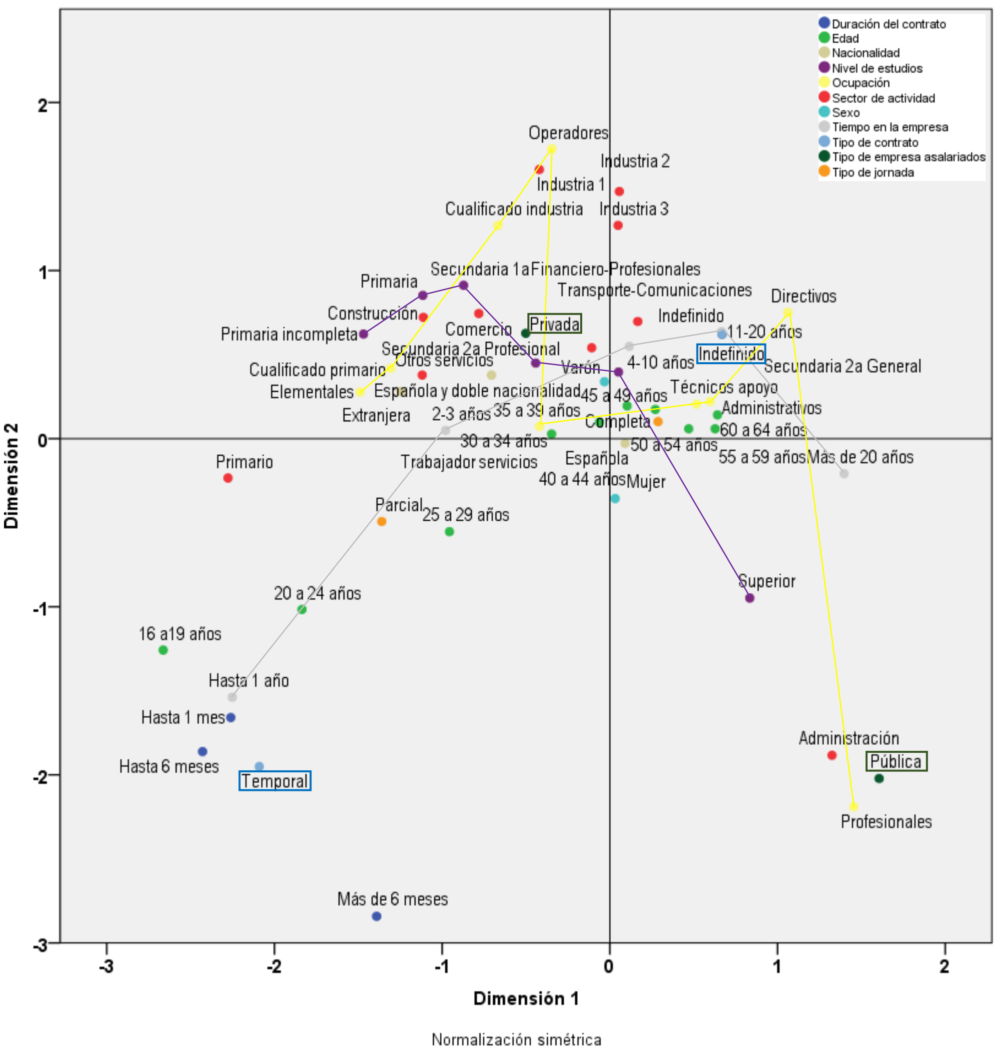
Variables suplementarias

Categoría	Frecuencia	Coordenadas del centroide Dimensión 1	Coordenadas del centroide Dimensión 2
Edad[a]			
16 a19 años	229	-2,664	-1,258
20 a 24 años	2.045	-1,836	-1,016
25 a 29 años	4.232	-0,957	-0,553
30 a 34 años	5.250	-0,348	0,028
35 a 39 años	7.224	-0,063	0,098
40 a 44 años	7.274	0,103	0,195
45 a 49 años	7.314	0,271	0,173
50 a 54 años	6.766	0,470	0,059
55 a 59 años	5.001	0,628	0,059
60 a 64 años	2.501	0,642	0,141
Sexo[a]			
Varón	24.356	-0,032	0,339
Mujer	23.480	0,032	-0,357
Nacionalidad[a]			
Española	44.166	0,091	-0,028
Española y doble	981	-0,706	0,378
Extranjera	2.689	-1,247	0,281

Normalización simétrica.
a. Variable complementaria.

Esta información de la tabla se representa mediante un gráfico factorial (Gráfico conjunto de puntos de categoría):

Gráfico conjunto de puntos de categoría



La información que se obtiene mediante el procedimiento en SPSS difiere de la obtenida con SPAD, tal y como comentamos en el apartado 5.2 y se comenta posteriormente en el apartado 7. No obstante, la lectura de la información conduce a los misma interpretación.

En el análisis de la segmentación laboral observamos cómo la primera dimensión se configura por las contribuciones de la contratación temporal en el polo positivo frente a la contratación indefinida en el polo negativo. La eventualidad de la contratación vemos que se asocia con los contratos de duración determinada y con llevar poco tiempo en la empresa, así como con jornadas de trabajo parciales. Si nos fijamos en las variables de cualificación, la polaridad de contratación indefinida se corresponde en mayor medida con las ocupaciones de profesionales y niveles de estudios universitarios, mientras la contratación temporal se asocia principalmente con las categorías ocupacionales más elementales. El tipo de empresa aparece en este primer factor para destacar la contratación indefinida que en mayor medida se da en el sector público. Podemos hablar por tanto de una primera dimensión que expresa la segmentación laboral asociada doblemente a la estabilidad y a la cualificación.[40]

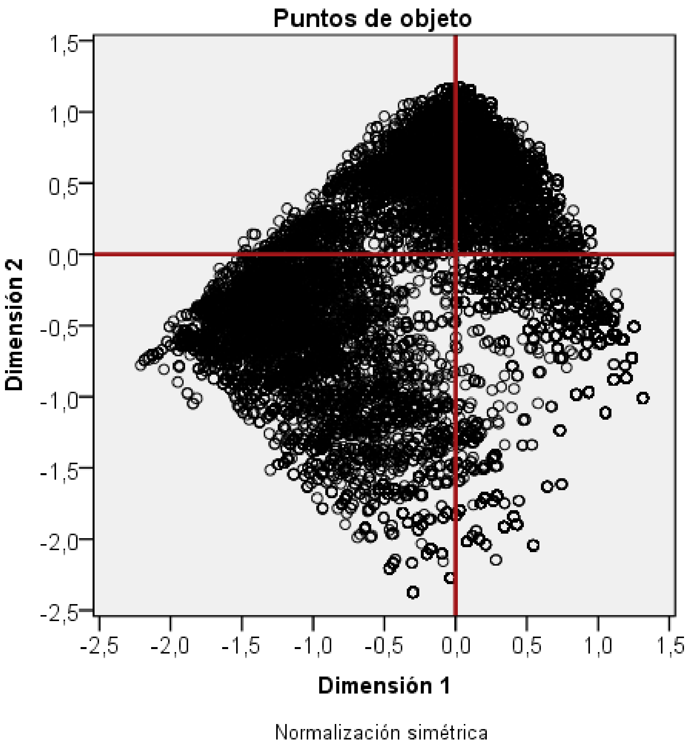
La segunda dimensión está protagonizada específicamente por el perfil de trabajador/a de la administración pública con altos niveles de cualificación, son técnicos y profesionales con niveles de estudios superiores. Junto a este perfil que ocupa el polo negativo de la dimensión se ubican también las categorías de la ocupación temporal configurando así un doble perfil en este extremo. El resto de las categorías ocupa de forma generalizada valores moderados en la polaridad positiva, o próximos a cero, en general del sector privado, y donde podemos destacar el perfil de trabajador estable de la industria con cualificaciones medias-bajas. Se dibuja pues un segunda dimensión de sector público y privado.

Si observamos las categorías de las variables suplementarias podemos apreciar cómo de forma sistemática las distintas edades se ubican a lo largo del primer eje desde las coordenadas positivas más alejadas del centro, las que corresponden a los más jóvenes, hasta los valores negativos más alejados de las edades más maduras. El origen inmigrante diferencia también a los autóctonos en posiciones centrales frente a las personas de nacionalidad extranjera o doble nacionalidad que se ubican primordialmente en el lado positivo de la precariedad del empleo. El sexo no marca diferencias en esta primera dimensión, por lo que cabe concluir que la estabilidad-inestabilidad del primer afecta por igual a ambos sexos. De forma modesta en la segunda dimensión las mujeres tienden a situarse en valores negativos por su mayor presencia en la administración pública mientras que los varones tienden a alejarse hacia valores positivos por su mayor presencia en los sectores industriales.

Finalmente, con las puntuaciones factoriales asignadas a los individuos, se obtiene la siguiente representación gráfica donde se posiciona a los individuos en el espacio factorial, lo que implica cuantificar o asignar un valor a cada individuo según cada una de las dos variables factoriales, caracterizadas por tener ser linealmente independientes

[40] Cabe hacer notar la particularidad anecdótica del gráfico factorial proporcionado por cada software: en SPSS la primera dimensión se representa, de menos a más, entre la inestabilidad y la estabilidad; en el caso de SPAD al revés. Justamente este hecho revela que no es importante el signo de las puntuaciones sino las posiciones relativas y polares de las categorías en el espacio, por ello es lo mismo leerlos hacia una lado que hacia el otro, ya sea mirando el gráfico "por delante" o "por detrás".

entre sí, tener una métrica continua y estar tipificadas (tienen media es cero y la desviación típica uno). Estas variables factoriales se pueden utilizar para realizar análisis complementarios, gráficos como en este caso o utilizando estas variables con otras técnicas. En particular, este será el caso de la metodología que seguiremos al combinar el análisis factorial con el análisis de clasificación con el objetivo de la construcción de tipologías: las variables factoriales se considerarán como las variables-criterio con las que clasificar las unidades en grupos homogéneos.



Si al ejecutar el procedimiento de **MULTIPLE CORRESPONDENCE** especificamos que guarde las dos variables factoriales del análisis las encontraremos al final de la matriz de datos. Por defecto se asigna un nombre a estas variables, en nuestro casos las dos variables reciben el nombre de OBSCO1_1 y OBSCO2_1, es decir, "*object scores*" o puntuaciones factoriales para la primera y segunda dimensión del análisis 1 que acabamos de realizar. Si se ejecuta a través del menú de instrucciones este es el nombre asignado, y cada vez que ejecutemos de nuevo se guardarán con el número de orden del análisis (2, 3, 4, etc.). A través de la sintaxis, es posible guardarlas con el nombre raíz que el usuario/a elija, por ejemplo, Factor, generando las variables Factor1_1 y Factor2_1:

	Factor1_1	Factor2_1
1	0,28	0,78
2	1,20	-0,87
3	0,37	0,55
4	0,23	0,31
5	0,32	0,81
6	0,59	0,01
7	0,32	0,75
8	0,77	-0,43
9	-1,08	-0,59
10	0,22	0,97

6.2.2. *Análisis de conciliación de la vida laboral y familiar*

Con los datos de la matriz Conciliación.sav y el archivo de sintaxis Conciliación ACM+ACL.sps se pueden replicar los análisis sobre concililiación de la vida laboral y familiar publicados en el artículo de López-Roldán y Lozares (2007) con el título "La conciliación entre las exigencias del ámbito productivo y las condiciones socio-familiares: estudio de caso de una empresa".[41]

7. Análisis de correspondencias con SPAD

Ilustraremos el uso del software SPAD con el objetivo de realizar un análisis de correspondencias presentaremos dos ejemplos. Uno de correspondencias simples donde analizaremos los datos de la Tabla III.11.10, la tabla de correspondencias entre la ocupación dominante del hogar y el nivel de ingresos, y otro de correspondencias múltiples con un ejemplo sencillo de tres variables que vimos en el capítulo III.5 y III.6 de tablas de contingencia y log-lineal donde se relacionan la actitud, los estudios y el sexo.

7.1. Correspondencias simples

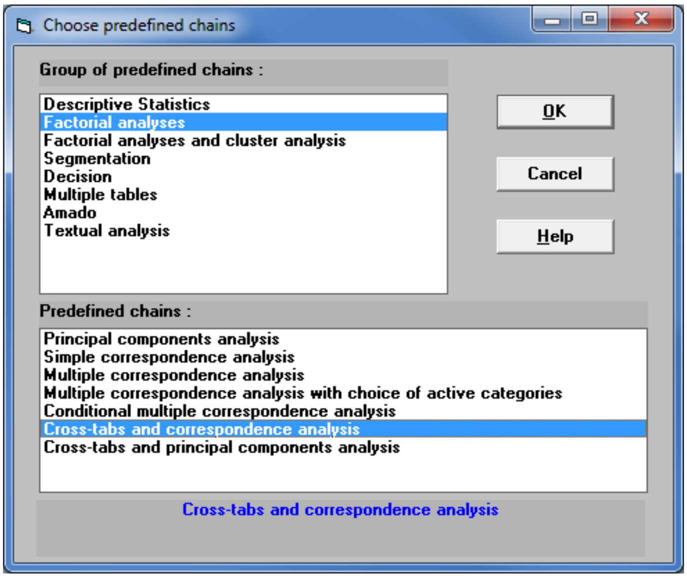
Para realizar un ACS recurriremos al asistente de programas predefinidos de SPAD en el menú Template. Elegiremos en el grupo de cadenas predefinidas la opción Factorial Analysis que da lugar a las diferentes alternativas de procedimientos factoriales. En el caso del ACS disponemos de dos alternativas:

a) Simple Correspondence Analysis (método `CORBI`) que aplica la técnica de análisis a partir de una tabla de correspondencias ya construida y almacenada con ese formato de filas y columnas en una base de datos del sistema con la extensión sba. Este es el procedimiento seguido en el caso del ejemplo de análisis de la relación en la ocupación dominante del hogar y el nivel de ingresos.

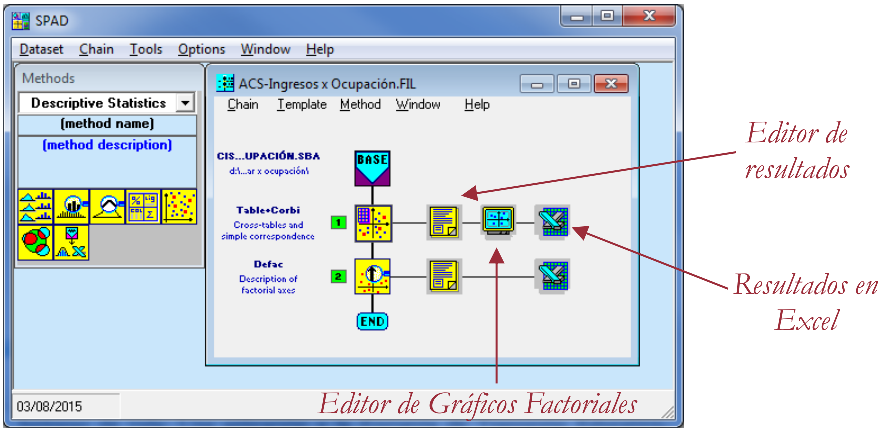
b) Cross-tabs and Correspondence Analysis (método `TABLE+CORBI`) donde se realiza el ACS a partir de una matriz de datos de individuos por variables que requiere previamente que se construya la tabla de contingencia entre dos variables las matriz. Este es el procedimiento que se ha aplicado en el caso de los ejemplos la relación entre provincias y educación de la Tabla III.11.14, y luego también con el sexo, así como en el caso de la Tabla III.11.15 de distancias por carretera.

El cuadro de diálogo que permite esta elección se presenta a continuación. Tanto en una como en otra opción la cadena o programa de instrucciones que se genera prevé un método adicional, de nombre `DEFAC`, para la descripción de los ejes factoriales.

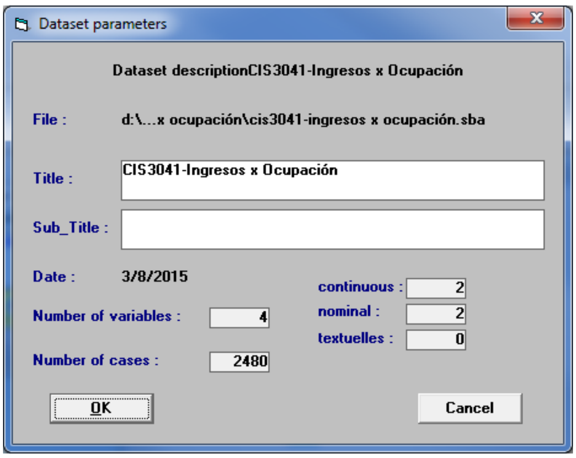
[41] Publicación accesible en http://ddd.uab.cat/pub/papers/02102862n83p123.pdf.



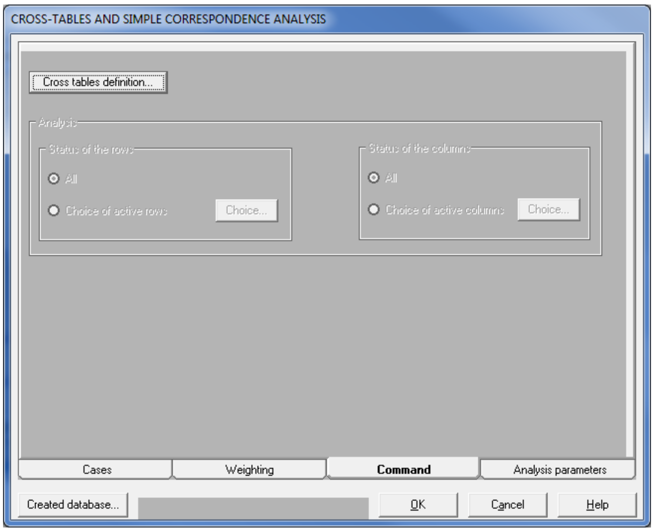
Si optamos por realizar un cruce previo al ACS la ventana de SPAD con los procedimientos ejecutados es la siguiente:



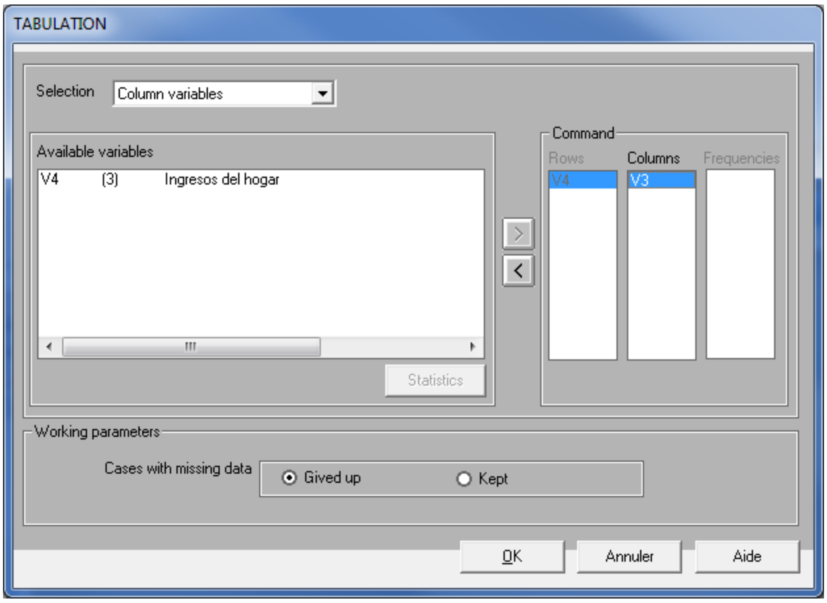
Trabajaremos con la matriz de datos CIS3041-Ingresos x Ocupación.sba proveniente de los datos del CIS donde se han conservado únicamente las dos variables del ejemplo.



Si parametrizamos el procedimiento TABLE+CORBI, al hacer doble clic sobre el icono nos aparece el siguiente cuadro de diálogo donde se solicita definir la tabla cruzada desde la pestaña Command:

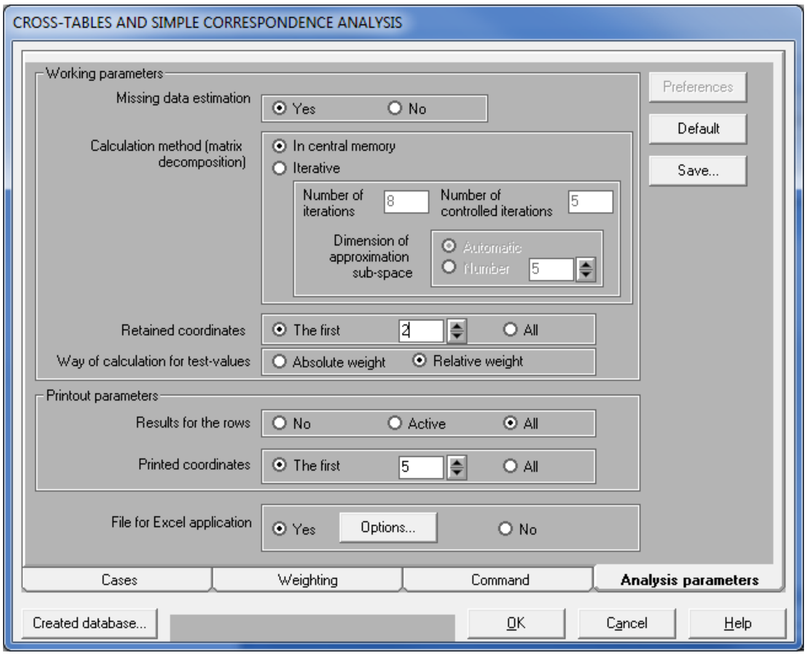


El cuadro de diálogo para precisar el cruce es el siguiente:



Elegiremos la variable que va en las filas, la V4 que corresponde a los ingresos. Las filas se consideran como "casos". En el desplegable elegiremos a continuación la opción de la variable en columnas y pasaremos la variable V3, la ocupación dominante del hogar. Las columnas se consideran como "frecuencias".

Seguidamente tenemos la opción de seleccionar casos si fuera necesario o de ponderar los datos. En nuestro caso iremos directamente a definir los parámetros del análisis:

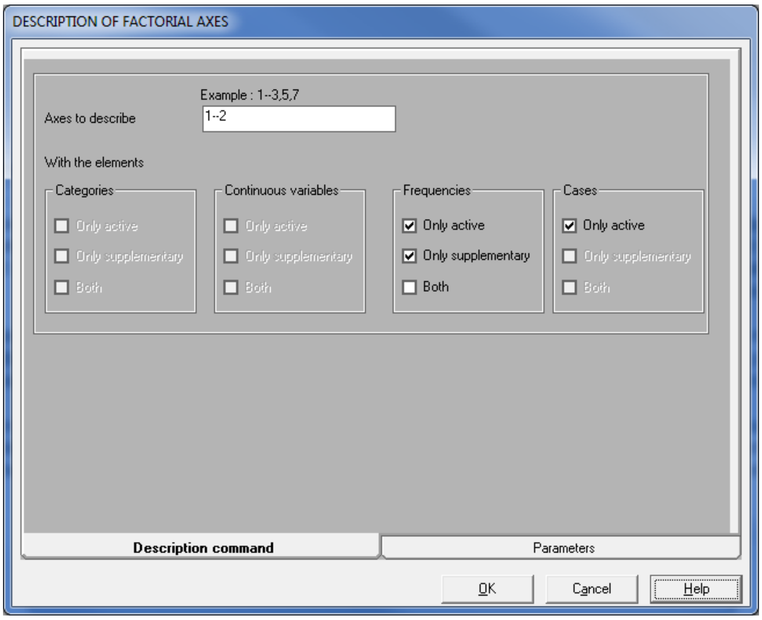


El análisis se puede realizar directamente con las opciones por defecto, solamente es necesario precisar las variables de fila y columna. En los parámetros del análisis la especificación más relevante es el número de factores retenidos. Siendo el máximo número el mínimo de filas y columnas menos 1, el número lo podemos fijar después de ver los resultados. Por defecto considera 10. En el ejemplo el número máximo es de 2 y es el valor que hemos introducido.
Adicionalmente existe la opción, si fuera el caso, de estimar valores perdidos según un algoritmo creado por Benzécri. También el método de cálculo de los vectores y valores propios por defecto, In central memory, se puede cambiar por un método iterativo recomendado para las tablas que incluyen un gran número de categorías (varios cientos) donde el usuario controla los cálculos a partir del número de iteraciones, de casillas y la dimensión del subespacio. Como en otros procedimientos disponemos igualmente de opciones de control de los resultados que mantendremos sin cambios.

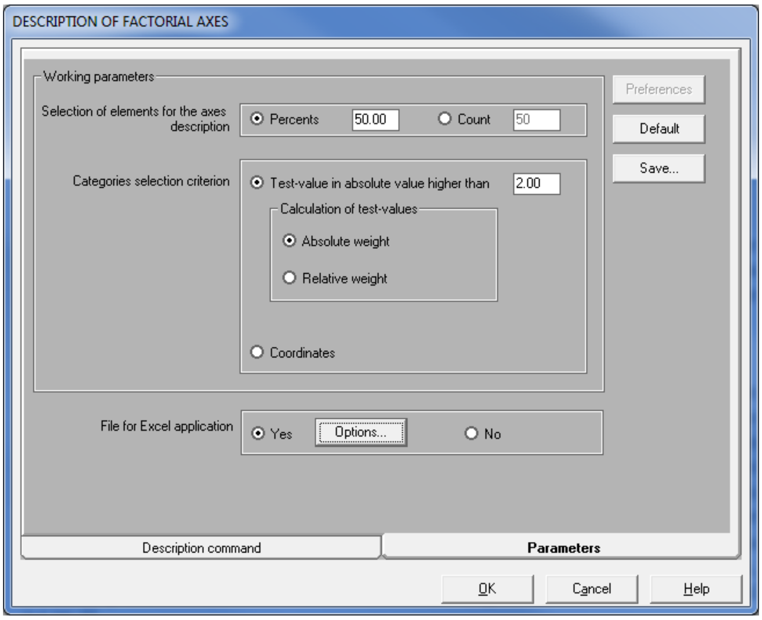
Finalmente el botón Created database abre un cuadro de diálogo que permite guardar la tabla de contingencia construida en el formato de un archivo de datos del sistema SPAD. A partir de esta base de datos, a continuación, se pueden aplicar otros métodos de SPAD.

Por su parte el procedimiento `DEFAC` complementa los resultados del análisis de correspondencias con la descripción de los factores para ayudar a la interpretación a partir de los elementos más significativos. Estos elementos pueden ser los casos, las variables categóricas o las variables o frecuencias continuas, y se pueden utilizar como elementos activos o ilustrativos. Los elementos característicos se clasifican en función de sus coordenadas. En el caso de las variables categóricas se pueden clasificar de acuerdo con el criterio estadístico de un valor-test.

En el cuadro de diálogo inicial, Command Description, se precisan los factores que serán descritos. Por defecto aparece la opción 1 a 3, pero si en el análisis sólo se generan 2 o se quieren más de 3 hay que cambiarlo al valor correspondiente.



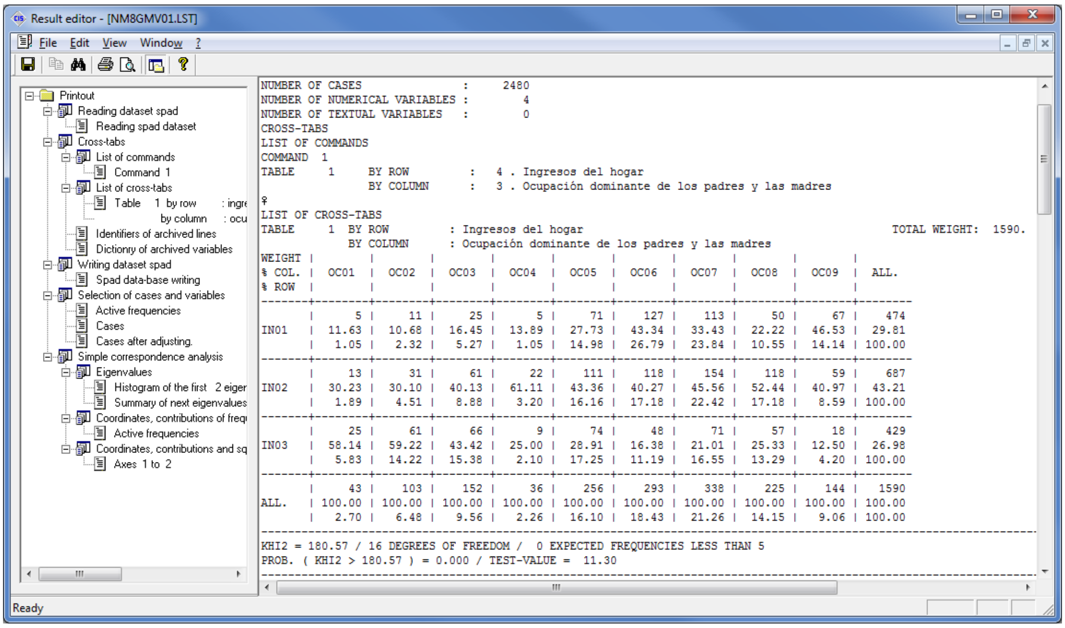
La pestaña de parámetros permite especificar dos aspectos. Por un lado el porcentaje o el número de elementos más significativos. Con el 50%, opción por defecto, la salida se produce para todas las descripciones con el 25% más significativo en cada uno de los extremos de los ejes.



Por otro, el criterio de selección de las categorías se basa en una prueba estadística de utilidad para tablas de datos de gran tamaño y para la lectura de los análisis multidimensionales complejos. Con la disposición de los elementos (variables o categorías) de forma descendente se ponen de manifiesto los rasgos más sobresalientes de caracterización de los factores. Para evaluar la magnitud de las diferencias más significativas entre las proporciones (si son categorías de variables cualitativas) o entre las medias (si son variables cuantitativas), la prueba estadística que se aplica da lugar a un valor del estadístico z de la normal. Si el valor, el número de desviaciones estándar, es mayor que 2, en valor absoluto, una desviación es significativa en el umbral habitual del 5%[42].

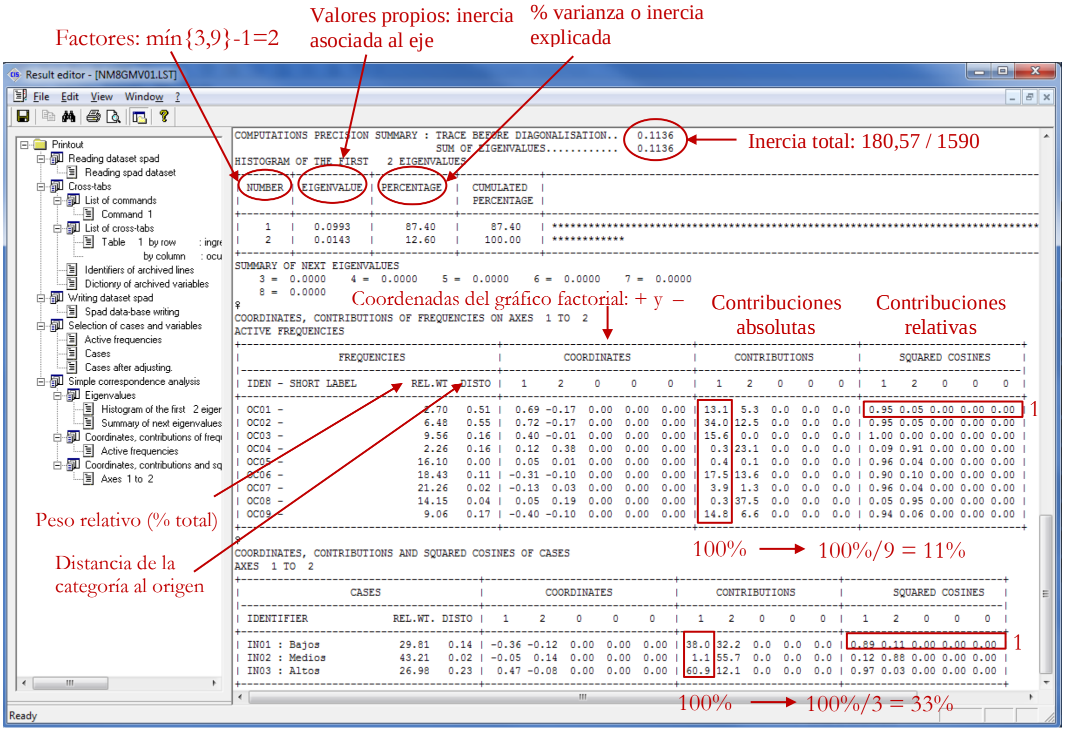
De la ejecución de estos procedimientos comentaremos los aspectos más relevantes de las salidas de tablas y gráficos, y remitimos al lector/a a la interpretación realizada anteriormente.

En las dos imágenes que siguen se recogen resultados parciales de la salida del procedimiento del ACS con SPAD. En la primera se puede ver la tabla de Burt y en la siguiente la información de los valores propios y las tablas con las coordenadas, las contribuciones absolutas y las contribuciones relativas en relación a los dos factores considerados.

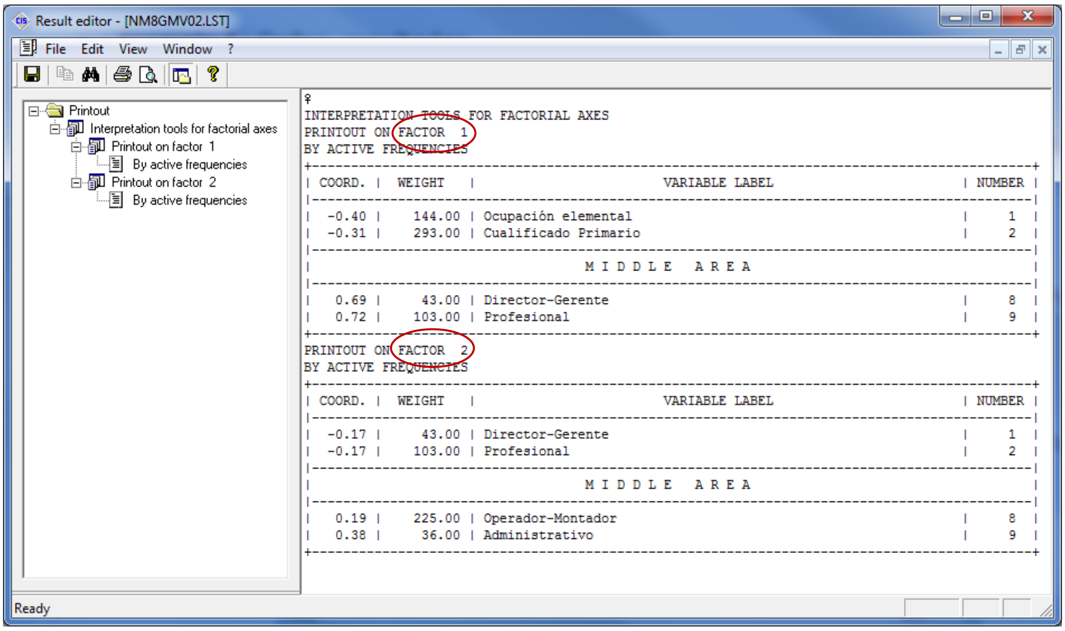



```
NUMBER OF CASES              :     2480
NUMBER OF NUMERICAL VARIABLES :       4
NUMBER OF TEXTUAL VARIABLES   :       0
CROSS-TABS
LIST OF COMMANDS
COMMAND  1
TABLE     1     BY ROW        :   4 . Ingresos del hogar
                BY COLUMN     :   3 . Ocupación dominante de los padres y las madres
♀
LIST OF CROSS-TABS
TABLE     1  BY ROW         : Ingresos del hogar                             TOTAL WEIGHT:  1590.
             BY COLUMN      : Ocupación dominante de los padres y las madres
WEIGHT |        |        |        |        |        |        |        |        |        |
% COL. |  OC01  |  OC02  |  OC03  |  OC04  |  OC05  |  OC06  |  OC07  |  OC08  |  OC09  |  ALL.
% ROW  |        |        |        |        |        |        |        |        |        |
-------+--------+--------+--------+--------+--------+--------+--------+--------+--------+--------
       |      5 |     11 |     25 |      5 |     71 |    127 |    113 |     50 |     67 |    474
IN01   |  11.63 |  10.68 |  16.45 |  13.89 |  27.73 |  43.34 |  33.43 |  22.22 |  46.53 |  29.81
       |   1.05 |   2.32 |   5.27 |   1.05 |  14.98 |  26.79 |  23.84 |  10.55 |  14.14 | 100.00
-------+--------+--------+--------+--------+--------+--------+--------+--------+--------+--------
       |     13 |     31 |     61 |     22 |    111 |    118 |    154 |    118 |     59 |    687
IN02   |  30.23 |  30.10 |  40.13 |  61.11 |  43.36 |  40.27 |  45.56 |  52.44 |  40.97 |  43.21
       |   1.89 |   4.51 |   8.88 |   3.20 |  16.16 |  17.18 |  22.42 |  17.18 |   8.59 | 100.00
-------+--------+--------+--------+--------+--------+--------+--------+--------+--------+--------
       |     25 |     61 |     66 |      9 |     74 |     48 |     71 |     57 |     18 |    429
IN03   |  58.14 |  59.22 |  43.42 |  25.00 |  28.91 |  16.38 |  21.01 |  25.33 |  12.50 |  26.98
       |   5.83 |  14.22 |  15.38 |   2.10 |  17.25 |  11.19 |  16.55 |  13.29 |   4.20 | 100.00
-------+--------+--------+--------+--------+--------+--------+--------+--------+--------+--------
       |     43 |    103 |    152 |     36 |    256 |    293 |    338 |    225 |    144 |   1590
ALL.   | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00
       |   2.70 |   6.48 |   9.56 |   2.26 |  16.10 |  18.43 |  21.26 |  14.15 |   9.06 | 100.00
--------------------------------------------------------------------------------------------------
KHI2 = 180.57 / 16 DEGREES OF FREEDOM /  0 EXPECTED FREQUENCIES LESS THAN 5
PROB. ( KHI2 > 180.57 ) = 0.000 / TEST-VALUE =  11.30
--------------------------------------------------------------------------------------------------
```

42 Cuando se trata de la comparación de dos proporciones se utiliza la ley hipergeométrica para evaluar las diferencias. Para comparar dos medias se utiliza la t de Student. Una presentación de estos conceptos se encuentra en Morineau (1984) y Lebart, Morineau y Piron (1997).

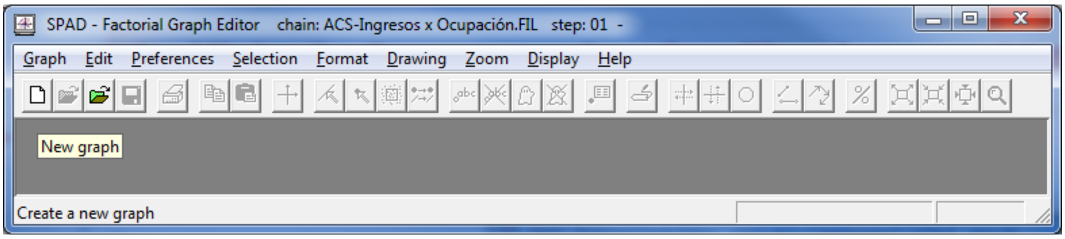


Los resultados de la descripción de los factores se pueden ver en la imagen siguiente. Este es un tipo de resultado que gana en interés cuando el número de elementos descriptivos es importante. A partir de la zona central (Middle area) ubica las categorías de filas (casos) y columnas (frecuencias activas) en función de su significación en cuanto al alejamiento hacia la polaridad positiva o negativa de cada eje.

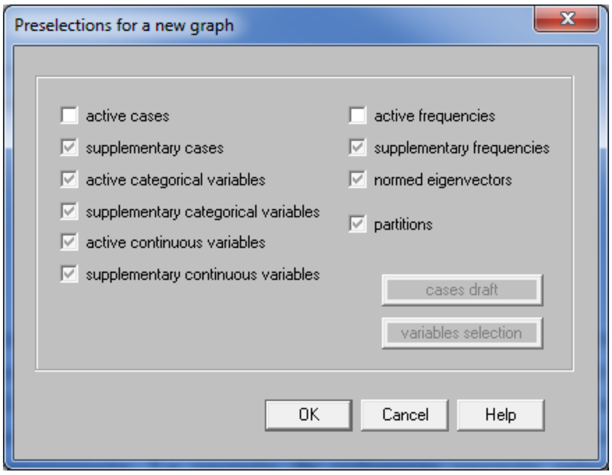


Algunos de los resultados que acabamos de comentar se pueden exportar a Excel con una plantilla de presentación de las tablas que genera el propio software. Haciendo doble clic sobre el icono se generan automáticamente.

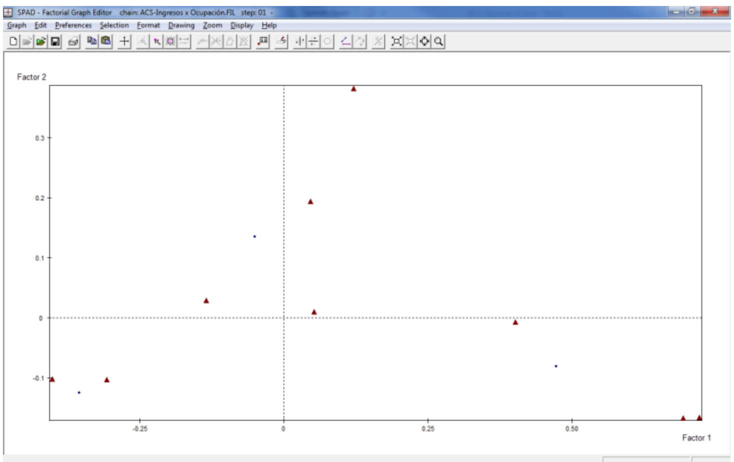
Nos queda finalmente comentar el editor gráfico para reproducir el gráfico factorial del análisis. Una vez abierto el editor haciendo doble clic sobre el icono nos aparece esta ventana inicial donde podemos abrir o crear un nuevo gráfico:



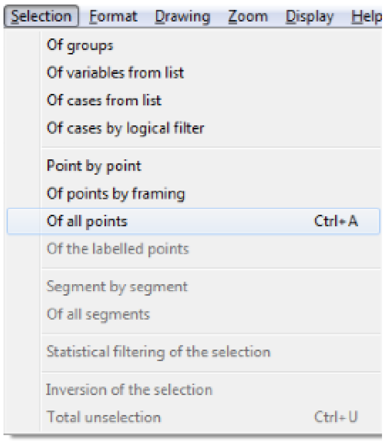
Clicaremos sobre New graph y accederemos a un cuadro de diálogo donde nos facilita la elección de qué elementos aparecerán en el gráfico:



De la relación de elementos, los que aparecen en blanco son los elegibles mientras que los que aparecen marcado en gris no están disponibles. En este caso marcaremos tanto la opción actives cases (categorías en fila de la tabla de correspondencias) como actives frequencies (categorías en columna) para ver la representación simultánea y analizar sus correspondencias. Cuando se activa la selección los botones cases draft y variables selection ofrecen la posibilidad de seleccionar individuos y seleccionar categorías. Al apretar sobre OK nos aparece una ventana como esta:

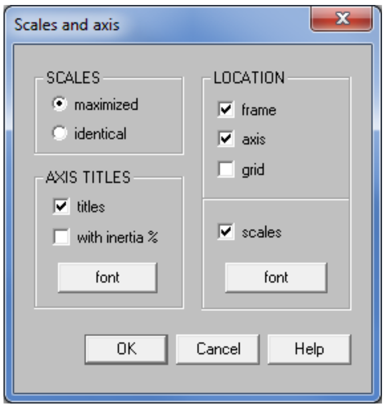


Aparecen los puntos-fila y puntos-columna con marcas diferentes, puntos azules y triángulos granates, respectivamente. Para etiquetarlos es necesario seleccionarlos a través del menú de selección:



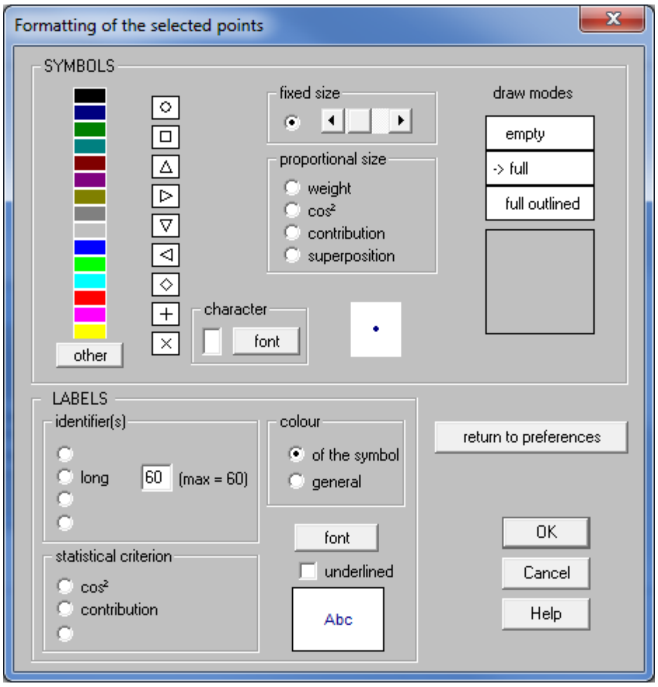
o directamente con <CRTL><A>. Al seleccionarlos cambian el color a fucsia. Sobre esta selección es posible realizar diversos cambios, por ejemplo, de formato, o etiquetarlos como haremos a continuación clicando sobre el botón . A continuación para deseleccionar podemos clicar sobre .

El gráfico aparece por defecto con un tamaño que maximiza su presentación en la pantalla del ordenador. Este tamaño se puede ajustar a través del menú Display y la opción Scales and axis:



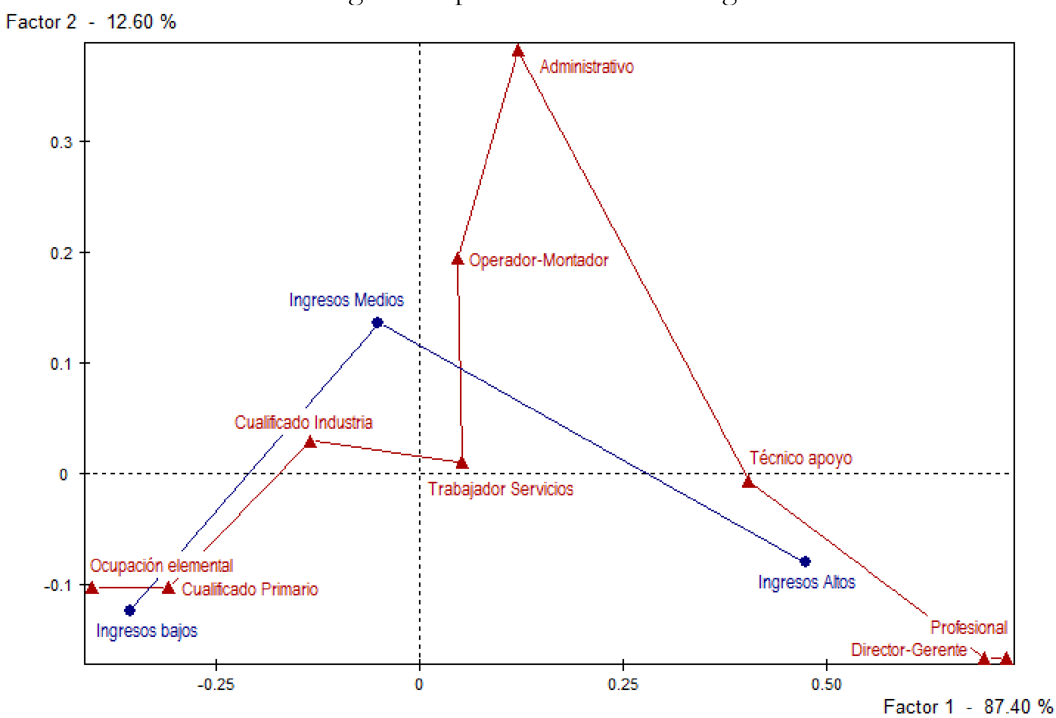
Podemos clicar sobre identical en la opción SCALES con lo que obtendremos una imagen proporcional entre el eje horizontal y vertical del gráfico. Adicionalmente podemos clicar sobre with inertia % para visualizar sobre el gráfico el porcentaje de inercia explicada por cada factor.

Las etiquetas de los puntos se pueden mover para ubicarlas en un espacio distinto. Esto será especialmente últil cunado el número de categorías sea numeroso. Para cambiar el formato de los puntos y las etiquetas se pueden seleccionar y a través del menú Format / Colours, symbols,… cambiarlos de tamaño, color, tipo de letra, etc.:



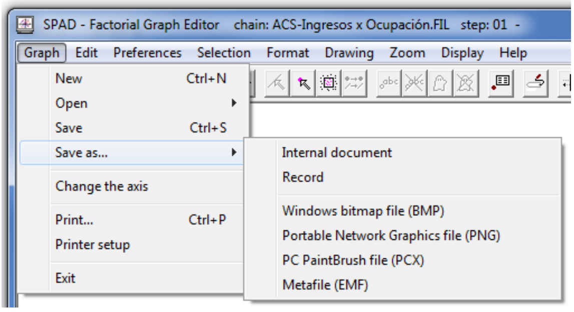
A través del botón de segments drawing podemos también trazar líneas que unan los diferentes puntos para así obtener una imagen del recorrido de las categorías como ayuda visual para dar también identidad a los ejes factoriales.

Con todos estos cambios el gráfico aparece como esta imagen:



Al operar estos distintos cambios es habitual que se desdibujen los diferentes elementos del gráfico, por ello es necesario refrescar la pantalla a través del botón .

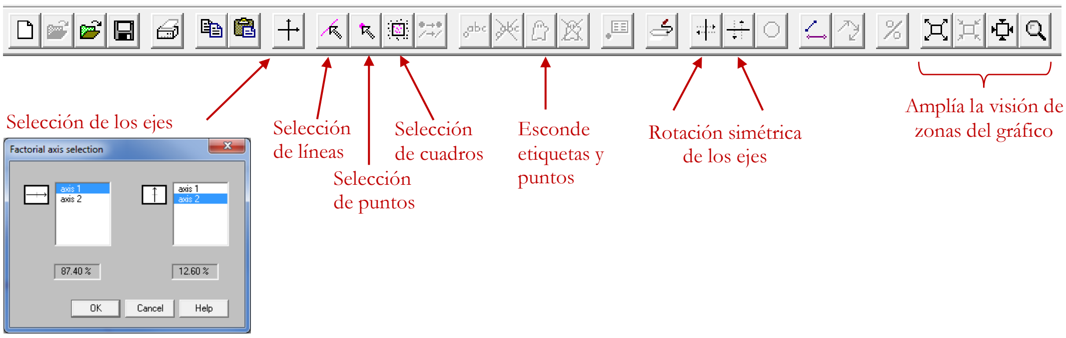
Para guardar un gráfico podemos optar por diversas alternativas:



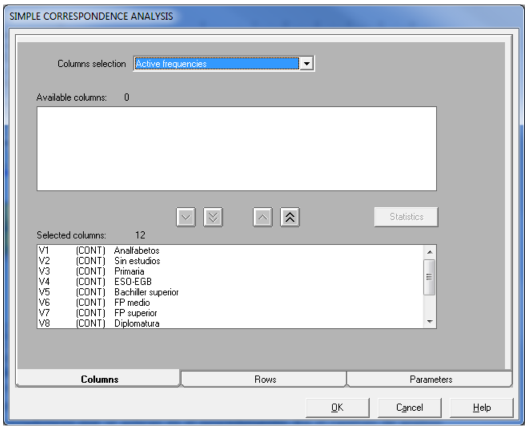
Si se guarda un documento interno el gráfico depende del programa de instrucciones por lo que si se cambian o se borran afectará igualmente al gráfico. La utilidad de esta opción es que todas las funciones de las anotaciones y las propiedades de los planos factoriales permanecen disponibles.

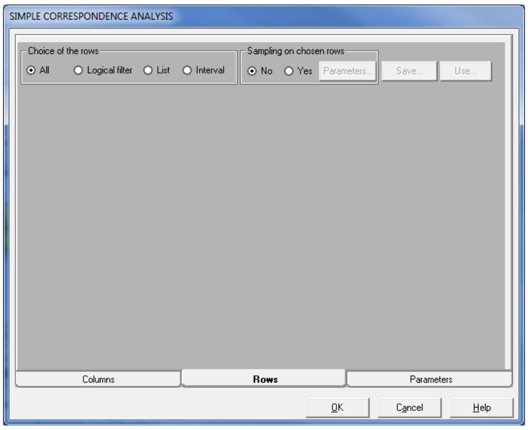
Si se guarda como registro en un archivo el gráfico es independiente del programa de instrucciones. Los archivos se guardan con la extensión GFA. En este caso algún formato no está disponible, pero en ambas opciones se pueden volver a cargar los gráficos para trabajar de nuevo la edición.

Alternativamente el editor de planos factoriales permite guardar los gráficos en formato de imagen BMP, PCX, PNG o EMF. El formato EMF Metafile ofrece la mejor calidad de imagen.
Otras opciones disponibles para la edición del gráfico factorial se presentan en la imagen siguiente:



Por último, en relación al ACS, comentaremos los cuadros de diálogo específicos del procedimiento `CORBI` cuando se realiza un análisis de correspondencias simples con una matriz de datos que corresponde a una tabla de correspondencias ya construida. La novedad son las dos pestañas referidas a las columnas y a las filas, mientras que la de parámetros coincide con la comentada anteriormente. La pestaña de columnas permite seleccionar las categorías en columna de la tabla, denominadas frecuencias, y seleccionar las que se considerarán como frecuencias activas (opción obligatoria) o frecuencias suplementarias. La pestaña de filas permite la selección de las categorías de las filas consideradas como casos. Se dispone de diversas alternativas de selección en un cuadro de diálogo similar en diversos procedimientos. Esta modalidad es la que hemos aplicado en el caso del análisis de la relación entre las provincias y el nivel educativo.

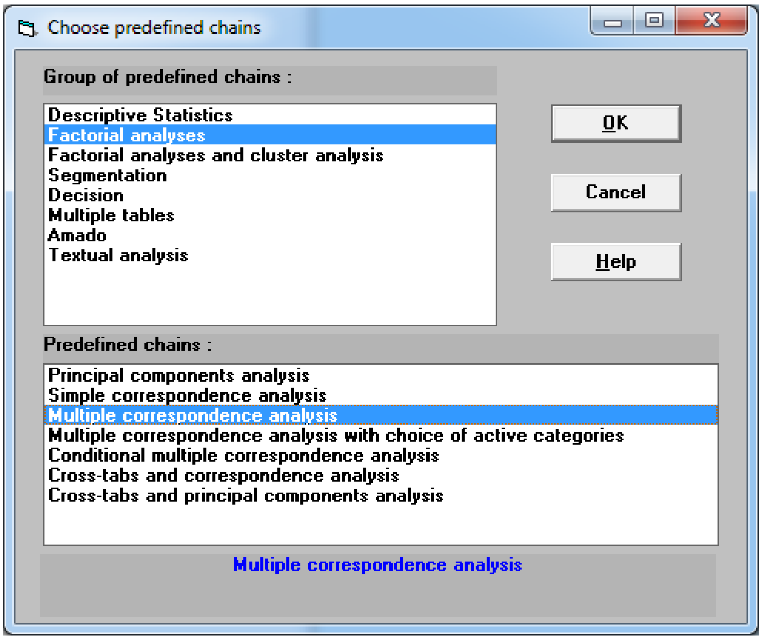




Con los datos de los archivos: España-Provincias x Educación.sba y España-Provincias x Educación-Sexo.sba, así como la matriz de datos Distancias España.sba se pueden reproducir los análisis de correspondencias simples presentados en el apartado 5.1.3.

7.2. Correspondencias múltiples

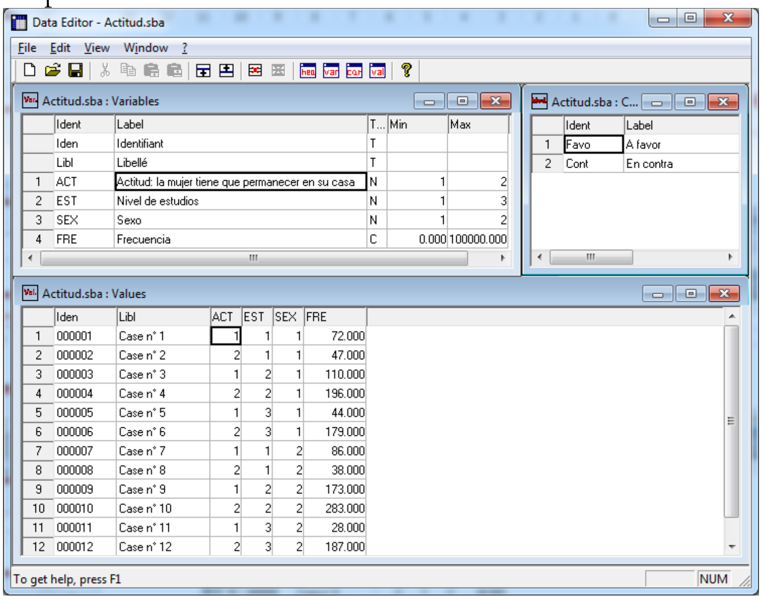
Si realizamos un análisis factorial de correspondencias múltiples podemos construir la cadena o programa de instrucciones con la ayuda de los programas predefinidos y optando por esta combinación[43]:



43 Veremos en el capítulo siguiente la opción de combinar un análisis de correspondencias con un análisis de clasificación con el objetivo de construir tipologías. En ese caso elegiremos Factorial analyses and cluster analysis.

Junto a la opción de correspondencias múltiples habitual, el método CORMU en terminología del comando de SPAD, existe la posibilidad de realizar un ACM con una selección de categorías activas, método COREM, que elimina las categorías seleccionadas, o bien de realizar un análisis de correspondencias condicional (Escofier, 1990; Lebart, Morineau y Piron, 1997), método CORCO, donde una variable categórica actúa restricción condicional de los resultados.

Ejecutaremos un sencillo ejercicio para ilustrar el funcionamiento del ACM y sus características en SPAD. Trabajaremos con la matriz de datos Actitud.sba que posee la mismas disposición de los datos que vimos en los temas de tablas de contingencia y log-lineal. La variables son tres: ACT (actitud: grado de acuerdo con la afirmación "Las mujeres deben quedarse en su casa"), EST (el nivel de estudios) y SEX (el sexo de la persona entrevistada), y aparecen en forma de matriz de datos de individuos por variables, pero con un solo caso por combinación de categorías de las tres variables, por lo que habrá que ponderar por una variable de frecuencia. La matriz de datos presenta este aspecto en el editor:



	Ident	Label	T...	Min	Max
	Iden	Identifiant	T		
	Libl	Libellé	T		
1	ACT	Actitud: la mujer tiene que permanecer en su casa	N	1	2
2	EST	Nivel de estudios	N	1	3
3	SEX	Sexo	N	1	2
4	FRE	Frecuencia	C	0.000	100000.000

	Ident	Label
1	Favo	A favor
2	Cont	En contra

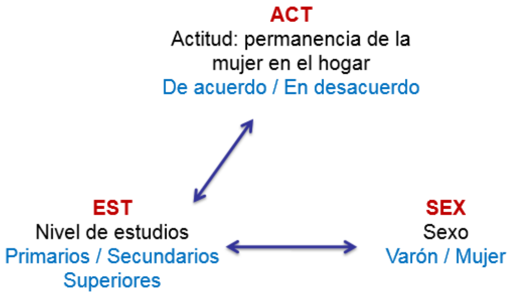
	Iden	Libl	ACT	EST	SEX	FRE
1	000001	Case n° 1	1	1	1	72.000
2	000002	Case n° 2	2	1	1	47.000
3	000003	Case n° 3	1	2	1	110.000
4	000004	Case n° 4	2	2	1	196.000
5	000005	Case n° 5	1	3	1	44.000
6	000006	Case n° 6	2	3	1	179.000
7	000007	Case n° 7	1	1	2	86.000
8	000008	Case n° 8	2	1	2	38.000
9	000009	Case n° 9	1	2	2	173.000
10	000010	Case n° 10	2	2	2	283.000
11	000011	Case n° 11	1	3	2	28.000
12	000012	Case n° 12	2	3	2	187.000

La tabla de contigencia que relaciona las tres variables es esta:

Tabla de contingencia ACT x EST x SEX

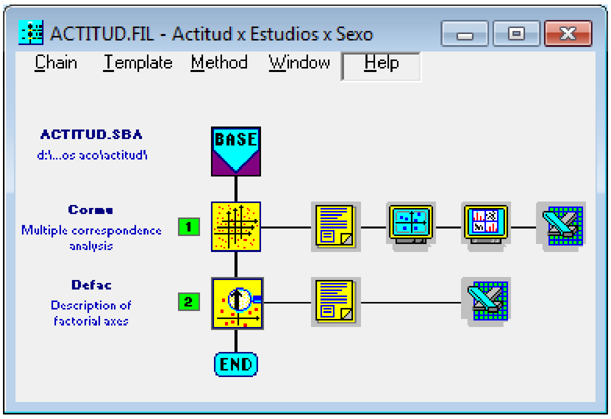
			SEX Sexo							
			1 Varón				2 Mujer			
			EST Nivel de estudios				EST Nivel de estudios			
			1 Primarios	2 Secundarios	3 Superiores	Total	1 Primarios	2 Secundarios	3 Superiores	Total
ACT Actitud: permanencia de la mujer en el hogar	1 De acuerdo	Recuento	72	110	44	226	86	173	28	287
		Frecuencia esperada	41,5	106,7	77,8	226,0	44,8	164,6	77,6	287,0
		% dentro de EST Nivel de estudios	60,5%	35,9%	19,7%	35%	69,4%	37,9%	13,0%	36,1%
		Residuos corregidos	6,5	,5	-5,9		8,4	1,3	-8,2	
	2 En desacuerdo	Recuento	47	196	179	422	38	283	187	508
		Frecuencia esperada	77,5	199,3	145,2	422,0	79,2	291,4	137,4	508,0
		% dentro de EST Nivel de estudios	39,5%	64,1%	80,3%	65%	30,6%	62,1%	87,0%	63,9%
		Residuos corregidos	-6,5	-,5	5,9		-8,4	-1,3	8,2	
Total		Recuento	119	306	223	648	124	456	215	795
		Frecuencia esperada	119,0	306,0	223,0	648,0	124,0	456,0	215,0	795,0
		% dentro de EST Nivel de estudios	100,0%	100,0%	100,0%	100%	100,0%	100,0%	100,0%	100%

El análisis de las relaciones entre las tres variables después de realizar una análisis log-lineal muestra la ausencia de interacción entre las tres variables y la configuración de un modelo de independencia condicional o doble asociación que relaciona la actitud y los estudios así como los estudios y el sexo, como se esquematiza en este gráfico:

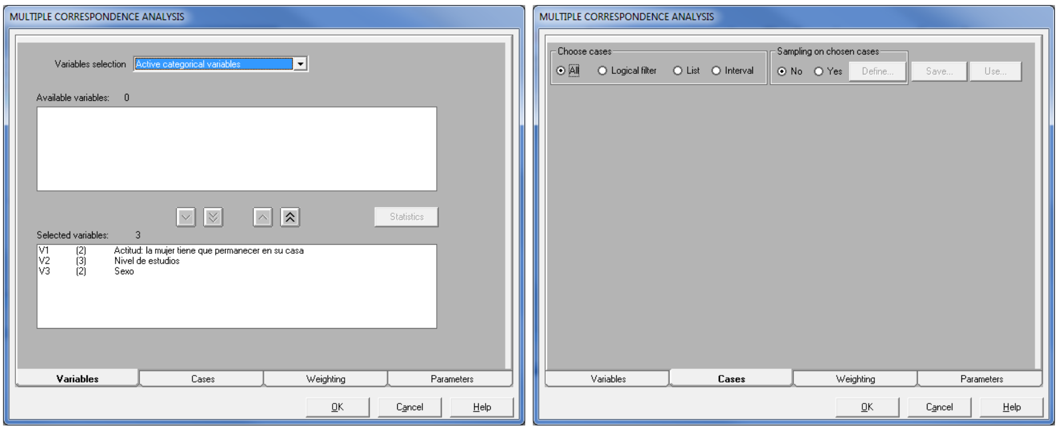


Veamos cómo se expresa este resultado en términos de un análisis de correspondencias múltiples.

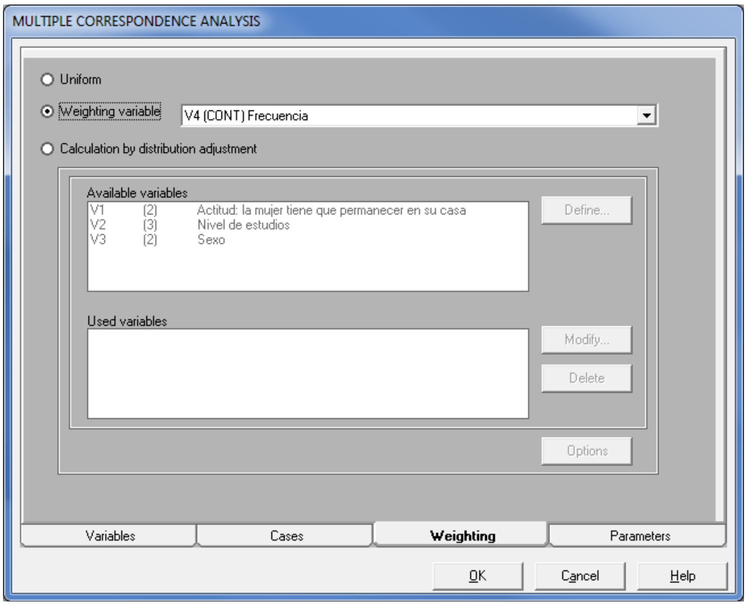
Para realizarlo en SPAD, de forma similar al ACS, el programa de instrucciones incluye el procedimiento del ACM junto a la descripción de los factores:



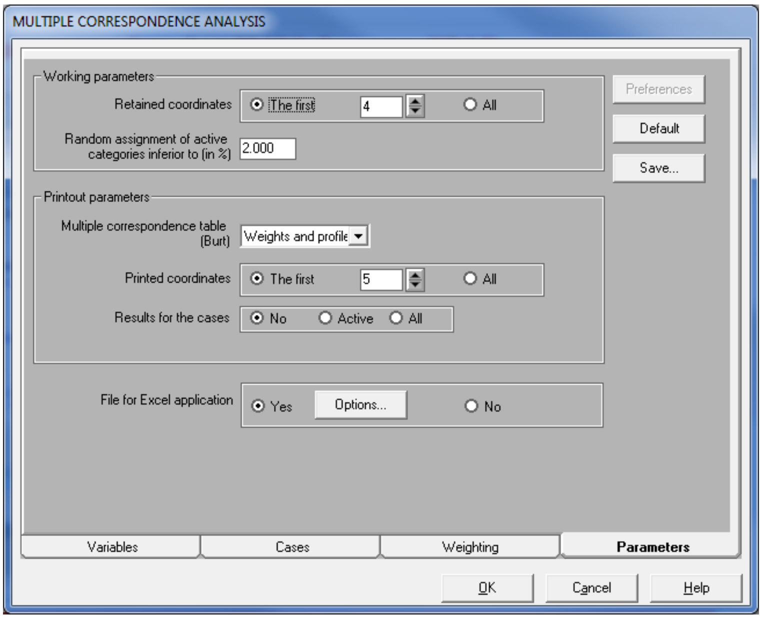
El cuadro de diálogo de `CORMU` incluye cuatro pestañas. En la primera se eligen las variables categóricas activas y las suplementarias categóricas o continuas. En nuestro caso elegiremos las tres únicas variables consideradas como activas. En la pestaña de casos no es necesario precisar nada pues consideraremos todos los casos.



La pestaña Weighting permite ponderar nuestros datos. Utilizaremos esta opción de forma instrumental para multiplicar cada caso de nuestra matriz por la frecuencia de la casilla a la que corresponde. Para ello elegiremos la opción Weighting variable y seleccionaremos la variable V4:



Finalmente nos queda la pestaña de parámetros donde en primer lugar se decide el número de factores a conservar. El total de ejes factoriales es el número de categorías activas menos el de variables, 7-3, por ello aparece el valor 4 en cuadro de diálogo. Como consecuencia del análisis posterior podemos reducir el número de dimensiones para tener un resultado más parsimonioso y suficiente para dar cuenta de los patrones de asociación existentes entre las variables. Para ello volveremos a ejecutar el procedimiento con el número de factores decidido.



Un segundo aspecto de interés en la parametrización del ACM es el procedimiento automático que le permite generar resultados robustos del análisis al tener en cuenta las categorías más débiles de las variables a través de la distribución aleatoria de los casos con muy baja frecuencia sobre el resto de categorías de las variables. El criterio que se maneja por defecto es el 2% como aparece en el recuadro correspondiente.

Esta opción nos permite además manejar las categorías que son valores perdidos y tienen frecuencias alrededor del 2%, de manera que su asignación aleatoria permite conservar esos casos. Cuando el número de valores perdidos es elevado la eliminación de los casos seguramente distorsionará los resultados del análisis, por ejemplo, con valores del 15% de no sabe y no contesta en los ingresos. En esa situación es conveniente conservar todos los casos y contemplar los valores perdidos como una categoría más, opción que nos permitirá también descubrir el perfil de los mismos: si se sitúan en el centro del gráfico factorial serán "neutrales" y si se ubican en un espacio alejado del centro nos mostrará a qué características pertenecen los individuos que no contestaron. De hecho este aspecto abre la posibilidad a poner en marcha un procedimiento de imputación de acuerdo con el principio del vecino más cercano, es decir, asignar un valor de ingresos al que no los tiene por una cuantía similar a la de las personas que ocupan el mismo espacio social, con quienes comparte unas mismas características, y de quienes sí sabemos su nivel de ingresos.

Como criterio habitual, si se quieren conservar todos los datos, es conveniente trabajar con las variables sin valores perdidos codificando todas sus categorías y sabiendo que las que tengan una frecuencia del 2% o menos serán "neutralizadas" y el resto se tratarás como una categoría más. Si se trabaja con valores perdidos, todos los casos perdidos en una variable se suprimirán inmediatamente en el resto de variables, por lo que es necesario estudiar los efectos de esta decisión. Alternativamente existe la posibilidad de ejecutar el ACM con una selección específica de categorías como comentamos al inicio.

Como resultado de la aplicación de este criterio de asignación aleatoria el procedimiento nos proporciona información de a qué categorías con baja frecuencia se le ha aplicado el criterio de "limpieza" a través de una tabla de frecuencias.

Si queremos visualizar la tabla de correspondencias de Burt podemos optar por presentarla en los resultados con las frecuencias absolutas, relativas o ambas. También se configura el número de coordenadas presentadas en los resultados, que dependerá de los factores retenidos, así como los resultados para los casos y la opción de resultados en Excel.

Las opciones del procedimiento de descripción de los factores, `DEFAC`, son similares a las que comentamos en el caso del ACS.

Presentamos a continuación los resultados más relevantes del análisis.

En primer lugar aparece la relación de variables y el número de categorías asociadas, junto con las frecuencias y una representación gráfica de las mismas:

```
SELECTION OF CASES AND VARIABLES
ACTIVE CATEGORICAL VARIABLES
     3 VARIABLES       7 ASSOCIATED CATEGORIES
----------------------------------------------------------------------------------------------------------------------------------
   1 . Actitud: la mujer tiene que permanecer en su casa            (   2 CATEGORIES )
   2 . Nivel de estudios                                            (   3 CATEGORIES )
   3 . Sexo                                                         (   2 CATEGORIES )
----------------------------------------------------------------------------------------------------------------------------------
CASES
------------------------------ NUMBER -------------WEIGHT ---------------
 WEIGHT OF CASES      : Frecuencia
 KEPT .............. NITOT =     12      PITOT =           1443.000
 ACTIVE ............ NIACT =     12      PIACT =           1443.000
 SUPPLEMENTARY ...... NISUP =     0      PISUP =              0.000
-------------------------------------------------------------------------
♀
MULTIPLE CORRESPONDENCE ANALYSIS
ELIMINATION OF ACTIVE CATEGORIES WITH SMALL WEIGHTS
THRESHOLD (PCMIN)  :      2.00 %       WEIGHT:     28.86
BEFORE CLEANING    :      3 ACTIVE QUESTIONS        7 ASSOCIATE CATEGORIES
AFTER CLEANING     :      3 ACTIVE QUESTIONS        7 ASSOCIATE CATEGORIES
TOTAL WEIGHT OF ACTIVE CASES :   1443.00
MARGINAL DISTRIBUTIONS OF ACTIVE QUESTIONS
----------------------------+------------------+----------------------------------------------------------------------------------
        CATEGORIES          | BEFORE CLEANING |         AFTER CLEANING
IDENT      LABEL            | COUNT    WEIGHT | COUNT    WEIGHT       HISTOGRAM OF RELATIVE WEIGHTS,
----------------------------+------------------+----------------------------------------------------------------------------------
    1 . Actitud: la mujer tiene que permanecer en su casa
Favo - A favor              |    6     513.00 |   6     513.00  **********************
Cont - En contra            |    6     930.00 |   6     930.00  ****************************************
----------------------------+------------------+----------------------------------------------------------------------------------
    2 . Nivel de estudios
Infe - Inferiores           |    4     243.00 |   4     243.00  ***********
Medi - Medios               |    4     762.00 |   4     762.00  *********************************
Supe - Superiores           |    4     438.00 |   4     438.00  *******************
----------------------------+------------------+----------------------------------------------------------------------------------
    3 . Sexo
Homb - Hombre               |    6     648.00 |   6     648.00  ****************************
Muje - Mujer                |    6     795.00 |   6     795.00  ***********************************
----------------------------+------------------+----------------------------------------------------------------------------------
```

Sigue la tabla de correspondencias múltiples:

```
MULTIPLE CORRESPONDENCE TABLE
       | Favo Cont | Infe Medi Supe | Homb Muje |
 -----+-----------+----------------+-----------+
 Favo |  513    0 |
 Cont |    0  930 |
 -----+-----------+----------------+
 Infe |  158   85 |  243    0    0 |
 Medi |  283  479 |    0  762    0 |
 Supe |   72  366 |    0    0  438 |
 -----+-----------+----------------+-----------+
 Homb |  226  422 |  119  306  223 |  648    0 |
 Muje |  287  508 |  124  456  215 |    0  795 |
 -----+-----------+----------------+-----------+
       | Favo Cont | Infe Medi Supe | Homb Muje |
♀
MULTIPLE CORRESPONDENCE TABLE (ROW PROFILES)
MARGINAL DISTRIBUTION OF EACH QUESTION IS ON THE CORRESPONDING DIAGONAL
ALL NUMBERS ARE GIVEN AS PERCENTAGES
       | Favo Cont | Infe Medi Supe | Homb Muje |
 -----+-----------+----------------+-----------+
 Favo | 35.6  0.0 | 30.8 55.2 14.0 | 44.1 55.9 |
 Cont |  0.0 64.4 |  9.1 51.5 39.4 | 45.4 54.6 |
 -----+-----------+----------------+-----------+
 Infe | 65.0 35.0 | 16.8  0.0  0.0 | 49.0 51.0 |
 Medi | 37.1 62.9 |  0.0 52.8  0.0 | 40.2 59.8 |
 Supe | 16.4 83.6 |  0.0  0.0 30.4 | 50.9 49.1 |
 -----+-----------+----------------+-----------+
 Homb | 34.9 65.1 | 18.4 47.2 34.4 | 44.9  0.0 |
 Muje | 36.1 63.9 | 15.6 57.4 27.0 |  0.0 55.1 |
 -----+-----------+----------------+-----------+
       | Favo Cont | Infe Medi Supe | Homb Muje |
-
```

Los valores propios, el porcentaje de varianza explicada y el histograma se presentan seguidamente:

```
EIGENVALUES
COMPUTATIONS PRECISION SUMMARY : TRACE BEFORE DIAGONALISATION..   1.3333
                                 SUM OF EIGENVALUES............   1.3333
HISTOGRAM OF THE FIRST   4 EIGENVALUES
+--------+------------+-------------+-------------+----------------------------------------------------------------------------------+
| NUMBER | EIGENVALUE | PERCENTAGE  |  CUMULATED  |                                                                                  |
|        |            |             |  PERCENTAGE |                                                                                  |
+--------+------------+-------------+-------------+----------------------------------------------------------------------------------+
|    1   |   0.4460   |     33.45   |     33.45   | ******************************************************************************** |
|    2   |   0.3659   |     27.45   |     60.89   | ******************************************************************               |
|    3   |   0.3001   |     22.51   |     83.40   | ******************************************************                           |
|    4   |   0.2213   |     16.60   |    100.00   | ****************************************                                         |
+--------+------------+-------------+-------------+----------------------------------------------------------------------------------+
```

En el caso del ACM vimos cómo la importancia cuantitativa de los factores está infravalorada y es necesario recalcular el porcentaje de inercia explicada a partir de la transformación de los valores propios en valores propios corregidos. En la tabla siguiente se presentan estos cálculos siguiendo el criterio de Benzécri (1979):

Factor	Valor propio	% Inercia
1	0,4460	33,4%
2	0,3659	27,4%
3	0,3001	22,5%
4	0,2213	16,6%
Suma	**1,333334**	**100%**

Factor	Valor propio	% Inercia	Valor propio corregido	% Inercia	% Acumulado
1	0,4460	33,4%	0,0286	92,3%	92,3%
2	0,3659	27,4%	0,0024	7,7%	100,0%
3	-	-	-	-	-
4	-	-	-	-	-
Suma	**0,811925**	**61%**	**0,030947**	**100%**	

Así, a partir del número total de factores, cuatro, sobre los que se reparte la inercia total, 1,33, consideramos solamente los dos primeros pues son los valores propios superiores a *1/p*, es decir, *1/3=0,33*, y sobre los que aplicamos la transformación. En este caso vemos como el primer factor acumula de forma casi total la inercia, con el 92%, mientras que el segundo factor expresa el 8% de la variabilidad total.
Para interpretar los factores obtenidos disponemos de las tablas de coordenadas y contribuciones, además del gráfico factorial.

Coordenadas factoriales

Contribuciones absolutas

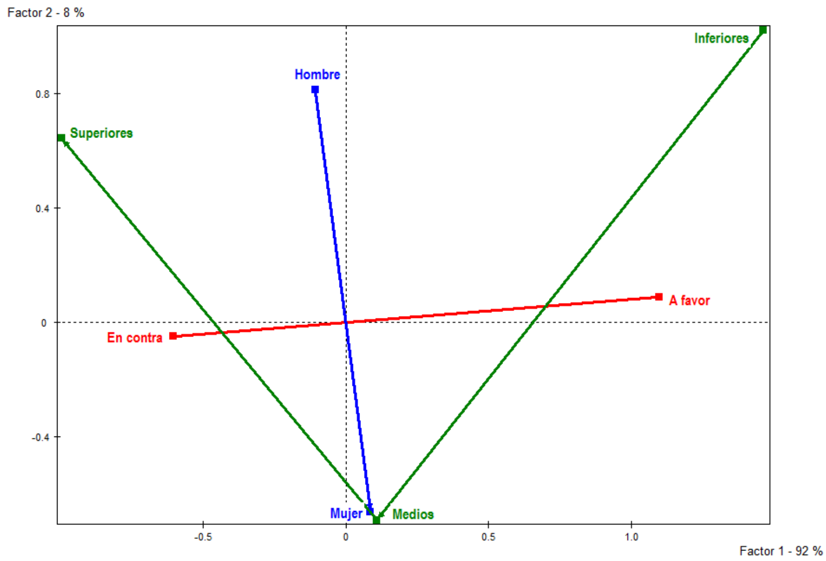
Contribuciones relativas

```
LOADINGS, CONTRIBUTIONS AND SQUARED COSINES OF ACTIVE CATEGORIES
AXES  1 TO  4
+------------------------------------------+-------------------------------+--------------------------+-------------------------+
|                CATEGORIES                |            LOADINGS           |      CONTRIBUTIONS       |     SQUARED COSINES     |
|------------------------------------------+-------------------------------+--------------------------+-------------------------|
| IDEN - LABEL              REL. WT. DISTO |  1     2     3     4     0    |  1    2    3    4    0   |  1    2    3    4    0  |
+------------------------------------------+-------------------------------+--------------------------+-------------------------+
|   1 . Actitud: la mujer tiene que permanecer en su casa                                                                     |
| Favo - A favor             11.85    1.81 |  1.10  0.09 -0.06 -0.77  0.00 | 31.9  0.2  0.1 32.1  0.0 | 0.66 0.00 0.00 0.33 0.00 | 1
| Cont - En contra           21.48    0.55 | -0.60 -0.05  0.03  0.43  0.00 | 17.6  0.1  0.1 17.7  0.0 | 0.66 0.00 0.00 0.33 0.00 |
+------------------------------------------+------- CUMULATED CONTRIBUTION = 49.5  0.4  0.2 49.9  0.0 +-------------------------+
|   2 . Nivel de estudios                                                                                                     |
| Infe - Inferiores           5.61    4.94 |  1.46  1.02  0.82  1.05  0.00 | 26.8 15.9 12.5 27.9  0.0 | 0.43 0.21 0.13 0.22 0.00 |
| Medi - Medios              17.60    0.89 |  0.11 -0.69 -0.63  0.06  0.00 |  0.5 23.2 23.2  0.3  0.0 | 0.01 0.54 0.44 0.00 0.00 |
| Supe - Superiores          10.12    2.29 | -1.00  0.64  0.64 -0.69  0.00 | 22.5 11.4 13.9 21.8  0.0 | 0.43 0.18 0.18 0.21 0.00 |
+------------------------------------------+------- CUMULATED CONTRIBUTION = 49.8 50.6 49.6 50.1  0.0 +-------------------------+
|   3 . Sexo                                                                                                                  |
| Homb - Hombre              14.97    1.23 | -0.11  0.81 -0.74  0.02  0.00 |  0.4 27.0 27.7  0.0  0.0 | 0.01 0.54 0.45 0.00 0.00 |
| Muje - Mujer               18.36    0.82 |  0.09 -0.66  0.61 -0.02  0.00 |  0.3 22.0 22.5  0.0  0.0 | 0.01 0.54 0.45 0.00 0.00 |
+------------------------------------------+------- CUMULATED CONTRIBUTION =  0.7 49.1 50.2  0.1  0.0 +-------------------------+
♀
LOADINGS AND TEST-VALUES OF CATEGORIES
AXES  1 TO  4
+---------------------------------------------+-------------------------------+---------------------------------------+---------+
|                CATEGORIES                   |          TEST-VALUES          |               LOADINGS                |         |
|---------------------------------------------|-------------------------------|---------------------------------------|---------|
| IDEN - LABEL                  COUNT  ABS.WT |  1     2     3     4     0    |   1      2      3      4      0     |  DISTO. |
+---------------------------------------------+-------------------------------+---------------------------------------+---------+
|    1 . Actitud: la mujer tiene que permanecer en su casa                                                                      |
| Favo - A favor                    6  513.00 |  30.9   2.5  -1.6 -21.9   0.0 |   1.10   0.09  -0.06  -0.77   0.00 |    1.81 |
| Cont - En contra                  6  930.00 | -30.9  -2.5   1.6  21.9   0.0 |  -0.60  -0.05   0.03   0.43   0.00 |    0.55 |
+---------------------------------------------+-------------------------------+---------------------------------------+---------+
|    2 . Nivel de estudios                                                                                                      |
| Infe - Inferiores                 4  243.00 |  25.0  17.4  13.9  17.9   0.0 |   1.46   1.02   0.82   1.05   0.00 |    4.94 |
| Medi - Medios                     4  762.00 |   4.3 -27.9 -25.3   2.5   0.0 |   0.11  -0.69  -0.63   0.06   0.00 |    0.89 |
| Supe - Superiores                 4  438.00 | -25.0  16.1  16.1 -17.3   0.0 |  -1.00   0.64   0.64  -0.69   0.00 |    2.29 |
+---------------------------------------------+-------------------------------+---------------------------------------+---------+
|    3 . Sexo                                                                                                                   |
| Homb - Hombre                     6  648.00 |  -3.6  27.9 -25.5   0.8   0.0 |  -0.11   0.81  -0.74   0.02   0.00 |    1.23 |
| Muje - Mujer                      6  795.00 |   3.6 -27.9  25.5  -0.8   0.0 |   0.09  -0.66   0.61  -0.02   0.00 |    0.82 |
+---------------------------------------------+-------------------------------+---------------------------------------+---------+
```

100%

De estas informaciones derivamos la centralidad del primer factor que expresa la principal fuente de asociación entre las variables: este primer eje horizontal expresa la oposición entre los que están mayoritariamente a favor, las personas con bajos niveles de estudios, frente a las personas que en mayor medida están en contra, aquéllas que tienen estudios superiores. Por su parte las personas con estudios medios aparecen en el centro de la primera dimensión pues son individuos que se reparten entre los que están a favor y en contra en la misma proporción que el conjunto de toda la muestra.

De forma equivalente, la variable sexo y sus categorías, varones y mujeres, se sitúan en el centro, hecho que muestra la actitud de ellos y ellas es prácticamente igual y se reparten entre los que están a favor o en contra en la misma proporción que el conjunto o promedio de los datos.



La segunda dimensión por su parte no tiene en consideración a la actitud como variable que acumule inercia a través de sus categorías, se sitúan en el centro con respecto del eje vertical. Este segundo factor, con una importancia menor del 12%, revela el vínculo que se produce entre el nivel de estudios y el sexo: los varones mayoritariamente tiene estudios inferiores y superiores mientras que las mujeres tienen sobre todo estudios medios.

Si relacionamos estas conclusiones con los resultados del análisis log-lineal comprobamos cómo el primer factor expresa la asociación bivariable entre actitud y estudios, mientras que el segundo la muestra la asociación entre estudios y sexo, los dos pares de relaciones significativas que emergieron del log-lineal. La relación entre actitud y sexo no apareció como significativa, es decir, ambas variables resultaron independientes. La independencia en el ACM se expresa gráficamente por la disposición perpendicular de las categorías de ambas variables; en este caso vemos como la línea roja es perpendicular a la azul, forman un ángulo de prácticamente de 90º, por lo que su correlación, el coseno de 90, es cero.

7.2.1. *Análisis de la segmentación del mercado de trabajo*

Para reproducir el análisis de segmentación del mercado de trabajo con los datos de la Encuesta de Población Activa de 2014 a partir de la selección de variables que se presentan en la Tabla III.11.16 se dispone de la matriz de datos EPA4T2014-Segmentación.sba que reproduce el análisis de correspondencias junto con el de clasificación que se verá en el próximo capítulo.

7.2.2. *Análisis de conciliación de la vida laboral y familiar*

Con los datos de la matriz Conciliación.sba se pueden replicar los análisis sobre conciliación de la vida laboral y familiar publicados en el artículo de López-Roldán y Lozares (2007) con el título "La conciliación entre las exigencias del ámbito productivo y las condiciones socio-familiares: estudio de caso de una empresa".[44]

8. Análisis de correspondencias con R

El análisis de correspondencia en R se puede realizar a través de diversos paquetes: el paquete ca de Nenadic y Greenacre (2007), el paquete FactoMineR (Husson, Lê, y Pagès, 2011), el paquete ade4, el paquete MASS, el paquete homals, el paquetes anacor (de Leeuw y Mair, 2009) y el paquete de visualización factoextra.

En una próxima edición de este capítulo se realizará una presentación del análisis de correspondencias con el software R.

9. Bibliografía

Abad, J.; Blanco, P.; García, A. (2008). Análisis de Correspondencias y estudio de historias de vida: Una aplicación a la Encuesta de Transición Educativo-Formativa e Inserción Laboral. *Pecvnia*, 6, 1-27. http://pecvnia.unileon.es/pecvnia06/06_001_028.pdf

Abad, J.; Muñiz, N.; Cervantes, M. (2003). Análisis de correspondencias simples y múltiples. En J.-P. Lévy y J. Varela, *Análisis Multivariable para las Ciencias Sociales*. Madrid: Pearson Prentice Hall, 361-416.

Abascal, E. et al. (2012). A comparison of two modes of data collection. Using multidimensional analysis. *Revista Internacional de Sociología*, 70, 3, septiembre-diciembre, 511-532. http://revintsociologia.revistas.csic.es/index.php/revintsociologia/article/view/474/495

Abascal, E.; Rada, V. D. D. (2014). Analysis of 0 to 10-point response scales using factorial methods: a new perspective. *International Journal of Social Research Methodology*, 17, 5, 569-584. DOI: 10.1080/13645579.2013.799736.

Aluja, T., Morineau, A. (1999). *Aprender de los datos: el análisis de componentes principales. Una aproximación desde el Data Mining*. Barcelona: Ediciones Universitarias de Barcelona.

Aparicio, F. (1988). La díficil realización de un Análisis de Componentes Principales mediante los programas estadísticos más difundidos en el mercado. *Estadística Española*, 30, 117, enero-abril, 99-114.

Barton, A. H. (1985). Concepto de espacio de atributos en Sociología. En R. Boudon y P. F. Lazarsfeld, *Metodología de las Ciencias Sociales. I. Conceptos e índices*. Barcelona: Laia, 195-219.

[44] Publicación accesible en http://ddd.uab.cat/pub/papers/02102862n83p123.pdf.

Batista; J. M. (1983). *Introducción al análisis factorial confirmatorio*. Barcelona: Edicions Universitat de Barcelona.

Batista; J. M. (1984). Componentes Principales y Análisis Factorial (Exploratorio y Confirmatorio). En J. J. Sánchez Carrión, *Introducción a las técnicas de análisis multivariable aplicadas a las ciencias sociales*. Madrid: Centro de Investigaciones Sociológicas, 17-74.

Batista; J. M.; Martínez, M. R. (1989). *Análisis multivariante. Análisis en componentes principales*. Barcelona: Hispano Europea.

Bécue, M.; Valls, J. *Manual de introducción a los métodos factoriales y clasificación con SPAD*. Bellaterra: Servei d'Estadística de la UAB. http://sct.uab.cat/estadistica/sites/sct.uab.cat.estadistica/files/manualSPAD.pdf

Benzécri, J. P. (1973). *L'Analyse des données. II. L'analyse des correspondances*. Paris: Dunod.

Benzécri, J. P. (1979). Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire. *Les Cahiers De l'Analyse Des Données*, 3, IV, 377-388. http://archive.numdam.org/ARCHIVE/CAD/CAD_1979__4_3/CAD_1979__4_3_377_0/CAD_1979__4_3_377_0.pdf

Benzécri, J. P.; Benzécri, F. (1980). *La pratique de l'analyse des données. I Analyse des correspondances. Exposé élémentaire*. Paris: Dunod.

Bertier, P.; Bourouche, J. M. (1983). *Analyse des données multidimensionnelles*. Paris: PUF.

Borràs, V. (1996). L'estructuració del consum a través de l'anàlisi de correspondències. *Papers. Revista de Sociologia*, 48, 89-102. http://ddd.uab.cat/pub/papers/02102862n48/02102862n48p89.pdf

Bourdieu, P. (1998). *La distinción. Criterios y bases sociales del gusto*. Madrid:Taurus.

Bourdieu, P. (2009). Pierre Bourdieu - La Distinción - Parte 2 HD. Video de Youtube. https://www.youtube.com/watch?v=Grvbu-AlBYo&playnext=1&list=PL952F7F5645297DF5&feature=results_main

Bourouche, J.-M.; Saporta, G. (1980). *L´Analyse des données*. Paris: PUF.

Brändle, G. (2007). Consumo y cambio social en España: evolución en el equipamiento doméstico (1983-2005). *Revista Española de Investigaciones Sociológicas*, 120, octubre-diciembre, 75-114. http://www.reis.cis.es/REIS/PDF/REIS_120_0031196339676936.pdf

Burt, C. L. (1940). *The factors of the mind: An introduction to factor analysis in psychology*. Londres: University of London.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245–276.

Cea d'Ancona, M. A. (2002). Análisis factorial. En M. A. Cea d'Ancona, *Análisis multivariable. Teoría y práctica en la investigación social*. Madrid: Síntesis, 428-515.

Cea d'Ancona, M. A. (2002). La medición de las actitudes ante la inmigración: evaluación de los indicadores tradicionales de "racismo". *Revista Española de Investigaciones Sociológicas*, 99, julio-septiembre, 87-111. http://www.reis.cis.es/REIS/PDF/REIS_099_06.pdf http://pagines.uab.cat/plopez/sites/pagines.uab.cat.plopez/files/ManualSPAD5-1.pdf

Chávez Molina, E.; Molina Derteano, P. (2009). La movilidad socio-ocupacional en la mira. Un estudio de caso exploratorio para debatir viejas y nuevas cuestiones. Ponencia *9º Congreso Nacional de Estudios del Trabajo*. Buenos Aires, Asociación Argentina de Especialistas en Estudios del Trabajo (ASET).

Cibois, Ph. (1983). *L'analyse factorielle* . Paris: PUF.

CISIA-CERESTA (2001). *Introduction à SPAD Version 5.0. Manuel de Prise en Main*. Montreuil: CISIA-CERESTA

CISIA-CERESTA (2001). *Système SPAD pour Windows Version 5.0. SPAD-Base. Aide à l'interprétation*. Montreuil: CISIA-CERESTA. http://pagines.uab.cat/plopez/sites/pagines.uab.cat.plopez/files/ManualSPAD5-2.pdf

Clausen, S.-E. (1998). *Applied Correspondence Analysis*. Thousand Oaks (California): Sage. http://www.relmis.com.ar/ojs/index.php/relmis/article/view/9/12

COHERIS-SPAD (2007). *SPAD7.0. Introduction à SPAD*. Guide de l'utilisateur. Courbevoie: SPAD. http://tic-recherche.crifpe.ca/docs/guides/fr/SPAD7_guide.pdf

Comrey, A. L. (1985). *Manual de análisis factorial*. Madrid: Cátedra.

Comrey, A. L.; Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale: Erlbaum.

Cornejo, J. M. (1988). *Técnicas de investigación social: el análisis de correspondencias (Teoría y Práctica)*. Barcelona: Promociones y Publicaciones Universitarias.

Correa, G. (2008). *Contribuciones al analisis multivariante no lineal*. Tesis doctoral. Departamento de Estadística de la Universidad de Salamanca. http://gredos.usal.es/jspui/handle/10366/19182

Crivisqui, E. M. (1993). *Análisis factorial de Correspondencias, un instrumento de investigación en las Ciencias Sociales*. Asunción: Universidad Católica de Asunción.

Cuadras, C. M. (2012). *Nuevos métodos de análisis multivariante*. Barcelona: CMC Editions. http://www.ub.edu/stat/personal/cuadras/metodos.pdf

de Leeuw, J.; Mair, P. (2009). Simple and canonical Correspondence Analysis using the R package anacor. *Journal of Statistical Software*, 31, 5.

Domínguez, M.; López-Roldán, P. (1996). La construcció de tipologies: procés i tècniques d'anàlisi de dades. *Papers. Revista de Sociologia*, 48, 31-39. http://ddd.uab.cat/pub/papers/02102862n48p31.pdf

Dunteman, G. H. (1989). *Principal Components Analysis*. Newbury Park: Sage

Elzo, J. (2010). Una tipología de los españoles de 2008, atendiendo a sus sistemas de valores. En J. Elzo, *Un individualismo placentero y protegido: cuarta encuesta europea de valores en su aplicación a España*. Deusto: Universidad de Deusto, 251-298.

Elzo, J. et al. (2010). *Valores sociales y drogas 2010*. Madrid: Fundación de Ayuda contra la Drogadicción. http://www.fad.es/sala_lectura/valores2010.pdf

Escobar, M. (2011). La calidad democrática. Una propuesta para su medición por expertos. *Revista Española de Investigaciones Sociológicas*, 133, enero-marzo, 59-80. http://www.reis.cis.es/REIS/PDF/Reis_133_041295261682795.pdf

Escofier, B.; Pagès, J. (1992). *Análisis factoriales simples y múltiples: objetivos, métodos e interpretación*. Bilbao: Universidad del País Vasco.

Escofier, B. (1990). L'analyse des correspondances conditionnelle. *La revue de Modulad*, 5, 13-27. https://www.rocq.inria.fr/axis/modulad/archives/numero-5/Escofier-5/Analysedescorrespondances.pdf

Etxeberría, J.; García, E.; Gil, J.; Rodríguez, G. (1995). *Análisis de datos y textos*. Madrid: RA-MA.

Fachelli, S. (2010) *Nuevo modelo de estratificación social y nuevo instrumento para su medición. El caso argentino*. Tesis doctoral, Bellaterra, Barcelona. http://tdx.cat/handle/10803/5149

Fachelli, S. (2010). Trayectorias de los hogares argentinos según estrato social entre 1997 y 2006. *Revista Latinoamericana de Estudios del Trabajo*, 23-24, 81-112. http://relet.iesp.uerj.br/Relet_23-24/art5.pdf

Fachelli, S.; López-Roldán, P. (2010). An attempt to measure social stratification and social changes in terms of distances. *XVII ISA World Congress of Sociology*, 11-17 de Julio, Göteborg (Suecia).

Fachelli, S.; López, N.; López-Roldán, P.; Sourrouille, F. (2012). *Desigualdad y diversidad en América Latina: hacia un análisis tipológico comparado*. Buenos Aires: SITEAL, Instituto Internacional de Planeamiento de la Educación (UNESCO-OEI). Libros digitales, 2.
http://www.siteal.org/sites/default/files/siteal_libro_digital_desigualdad_y_diversidad.pdf
http://pagines.uab.cat/plopez/sites/pagines.uab.cat.plopez/files/SITEAL-UBA.pdf

Fachelli, S. (2013). ¿La crisis aumenta las diferencias entre estratos sociales? La medición del cambio social en Argentina. *EMPIRIA. Revista de Metodología de Ciencias Sociales*, 22, enero-junio, 13-46.
http://dialnet.unirioja.es/descarga/articulo/4118170.pdf
http://pagines.uab.cat/plopez/sites/pagines.uab.cat.plopez/files/Estratos-UBA.pdf

Fernández Santana, J. O. (1988). Comprensión y manejo del análisis factorial. *Revista Internacional de Sociología*, 46, 1, 7-35.

Filgueira, C. (2007). Actualidad de las Viejas temáticas: clase, estratificación y movilidad social en América Latina. En R. Franco, A. León y R. Atria (Coordinadores) *Estratificación y movilidad social en América Latina. Transformaciones estructurales de un cuarto de siglo.* Santiago de Chile, LOM-CEPAL-GTZ.

García, E.; Gil, J.; Rodríguez, G. (2000). *Análisis factorial.* Madrid: La Muralla.

García Santesmases, J. M. (1984). Análisis Factorial de Correspondencias. En J. J. Sánchez Carrión, *Introducción a las técnicas de análisis multivariable aplicadas a las ciencias sociales*. Madrid: Centro de Investigaciones Sociológicas, 75-106.

Gifi, A. (1981). *Nonlinear multivariate analysis*. Leiden: University of Leiden.

Gómez Bezares, F. (1988). Análisis factorial por componentes principales (Algunos aspectos interesantes). *Estadística Española, vol. 30, 118, mayo-agosto*. P. 215-232.

Greenacre, M. J. (1984). *Theory and Application of Correspondence Analysis*. London: Academic Press.

Greenacre, M. J. et al. (1994). *Correspondence Analysis in the Social Sciences*. London: Academic Press.

Greenacre, M. J.; Blasius, J. (2006). *Multiple correspondence analysis and related methods*. London: Chapman & Hall/CRC.

Greenacre, M. J. (2008). *La práctica del análisis de correspondencias*. Madrid: Fundación BBVA.
http://www.fbbva.es/TLFU/tlfu/esp/publicaciones/libros/fichalibro/index.jsp?codigo=300

Greenacre, M. J. (2008). *Biplots in Practice*. Madrid: Fundación BBVA.
http://www.fbbva.es/TLFU/tlfu/esp/publicaciones/libros/fichalibro/index.jsp?codigo=571

Guttman, L. (1941). The quantification of a class of attributes: A theory and method of scale construction. En P. Horst, *The Prediction of Personal Adjustment*. New York: Social Science Research Council.

Hair, J. F. et al. (2011). *Multivariate Data Analysis*. Upper Saddle River: Prentice Hall.

Herrera-Usagre, M. (2011). El consumo cultural en España. Una aproximación al análisis de la estratificación social de los consumos culturales y sus dificultades metodológicas. *EMPIRIA. Revista de Metodología de Ciencias Sociales*, 22, julio-diciembre, 141-172. http://e-spacio.uned.es/fez/eserv.php?pid=bibliuned:Empiria-2011-22-5060&dsID=Documento.pdf

Hotelling, H. (1933). Analysis of a Complex of Statistical Variables Into Principal Components. *Journal of Educational Psychology*, 24, 417-441 y 498-520.

Husson, F.; Lê, S.; Pagès, J. (2011). *Exploratory Multivariate Analysis by Example using R*. London: Chapman & Hall. http://factominer.free.fr/book

Itzcovich, G.; Sourrouille, F. (2012). *Condiciones sociales, configuraciones familiares y vínculos de escolarización en adolescentes de 15 a 17 años. Aproximación desde una perspectiva relacional*. SITEAL, Instituto Internacional de Planeamiento de la Educación (UNESCO-OEI). Cuaderno 13. http://www.siteal.org/sites/default/files/cuaderno13_20121002.pdf

Joaristi, L.; Lizasoain, L. (1999). *Análisis de correspondencias*. Madrid: La Muralla.

Jolliffe, I. T. (2002). *Principal Component Analysis*. New York: Springer.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187-200.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.

Kaiser, H. F. (1974). An index of factor simplicity. *Psychometrika*, 39, 31–36.

Kessler, G.; Espinoza, V. (2003). *Movilidad y trayectorias ocupacionales en Argentina: rupturas y algunas paradojas del caso de Buenos Aires*. Santiago de Chile, CEPAL-Naciones Unidas.

Kim, J.-O.; Mueller, Ch.M. (1978). *Introduction To Factor Analysis. What It Is And How To Do It*. Beverly Hills: Sage.

Kim, J.-O.; Mueller, Ch.M. (1978). *Factor analysis. Statistical Methods and Practical Issues*. Beverly Hills: Sage.

Le Roux, B.; Börjesson, M., Bonnet, Ph. (2006). Performing Multiple Correspondence Analysis (MCA) using SPAD (version 6.5) http://www.skeptron.uu.se/broady/sec/p-gda-0609-spadguide-mca.pdf

Le Roux, B.; Rouanet, H. (2004). *Geometric Data Analysis, From Correspondence Analysis to Structured Data Analysis*. Dordrecht: Kluwer.

Le Roux, B.; Rouanet, H. (2010). *Multiple Correspondence Analysis*. Thousand Oaks (California): Sage.

Lebaron, F (1991). Geometric Data Analysis in a Social Science Research Program: The case of Bourdieu's Sociology. CURAPP, Université de Picardie Jules Verne-CNRS. Prefacio a El oficio del Sociólogo (Bourdieu y Passeron). https://www.u-picardie.fr/curapp/IMG/pdf/SLDS.pdf

Lebart, L.; Morineau, A.; Fenelon, J. P. (1985). *Tratamiento estadístico de datos. Métodos y programas*, Marcombo, Barcelona.

Lebart, L.; Morineau, A.; Piron, M. (1997). *Statistique exploratoire multidimensionnelle*. Paris: Dunod.

Lefebvre, J. (1983). *Introduction aux analyses statistiques multidimensionnelles*, Masson, Paris.

Long, J. S. (1983). *Confirmatory Factor Analysis* . Beverly Hills: Sage Publications.

López-Roldán, P. (1996a). La construcción de tipologías: metodología de análisis. *Papers. Revista de Sociologia*, 48, 9-29. http://ddd.uab.cat/pub/papers/02102862n48p9.pdf

López-Roldán, P. (1996b). La construcción de una tipología de segmentación del mercado de trabajo. *Papers. Revista de Sociologia*, 48, 41-58. http://ddd.uab.cat/pub/papers/02102862n48p41.pdf

López-Roldán, P. (2014). *Recursos per a la investigació social*. Bellaterra: Universitat Autònoma de Barcelona. http://ddd.uab.cat/record/89349 | http://pagines.uab.cat/plopez/

López-Roldán, P.; Fachelli, S. (2015). *Metodología de construcción de tipologías para el análisis de la realidad social*. Bellaterra: Universitat Autònoma de Barcelona. https://ddd.uab.cat/record/118082

López-Roldán, P.; Lozares, C. (2007). La conciliación entre las exigencias del ámbito productivo y las condiciones socio-familiares: estudio de caso de una empresa. *Papers. Revista de Sociologia*, 83, 123-144. http://ddd.uab.cat/pub/papers/02102862n83p123.pdf

López-Roldán, P.; Lozares, C. (2007). Implicaciones sociológicas en la construcción de una muestra estratificada. *EMPIRIA*, 14, 87-108. http://e-spacio.uned.es/fez/eserv.php?pid=bibliuned:Empiria-2007-14-0001&dsID=Documento.pdf

López-Roldán, P.; Lozares, C.; Domínguez, M. (2000). Disseny i construcció d'una mostra estratificada a partir de dades censals. *Qüestiió*, 24, 1, 111-136. http://www.raco.cat/index.php/Questiio/article/viewFile/143995/195695

Lozares, C.; López-Roldán, P.; (1991a). El análisis de componentes principales. Aplicación al análisis de datos secundarios. *Papers. Revista de Sociologia*, 37, 31-63. http://ddd.uab.cat/pub/papers/02102862n37/02102862n37p31.pdf

Lozares, C.; López-Roldán, P.; (1991b). El muestreo estratificado por análisis multivariado. En M. Latiesa, *El pluralismo metodológico en la investigación social: ensayos típicos*. Granada: Universidad de Granada, 107-160.

Lozares, C.; López-Roldán, P. (2000a). L'anàlisi factorial de components principals. En C. Lozares y P. López-Roldán, *Anàlisi multivariable de dades estadístiques*. Bellaterra (Barcelona): Universitat Autònoma de Barcelona, 69-142.

Lozares, C.; López-Roldán, P. (2000b). Elements matemàtics per a l'anàlisi de dades. En C. Lozares y P. López-Roldán, *Anàlisi multivariable de dades estadístiques*. Bellaterra (Barcelona): Universitat Autònoma de Barcelona, 23-68.

Lozares, C.; López-Roldán, P.; Borràs, V. (1998). La complemetariedad del log-lineal y el análisis de correspondencias en la elaboración y análisis de tipologías. *Papers. Revista de Sociologia*, 55, 79-93. http://ddd.uab.cat/pub/papers/02102862n55p79.pdf

Mallo Fernández, Fernando (1985). *Análisis de componentes principales y técnicas factoriales relacionadas: teoría, computación, aplicaciones*. León: Secretariado de Publicaciones de la Universidad de León.

Meulman, J. J. (1982). *Homogeneity analysis of incomplete data*. Leiden: DSWO Press.

Meulman, J. J. (1996). Fitting a distance model to homogeneous subsets of variables: Points of view analysis of categorical data. *Journal of Classification,* 13, 249-266.

Meulman, J. J.; Heiser, W. J. (1997). Graphical display of interaction in multiway contingency tables by use of homogeneity analysis. En M. Greenacre y J. Blasius, *Visual Display of Categorical Data*. New York: Academic Press.

Miguélez, F.; Martín, A.; de Alós-Moner, R.; Esteban, F.; López-Roldán, P.; Molina, Ó.; Moreno, S. (2012). *Trayectorias laborales de los inmigrantes en España*. Barcelona: Obra Social "la Caixa".

http://multimedia.lacaixa.es/lacaixa/ondemand/obrasocial/pdf/Trayectorias_laborales_de_los_inmigrantes_en_Espana.pdf

Morineau, A. (1984). Note sur la Caractérisation Statistique d'une Classe et les Valeurs-tests. *Bulletin Technique Centre Statistique Informatique Appliquées*, 2, 1-2, 20-27.

Murtagh, F. (2005). *Correspondence Analysis and Data Coding with JAVA and R*. Boca Raton (FL): Chapman & Hall/CRC.

Nenadic, O.; Greenacre, M. (2007). Correspondence Analysis in R, with Two- and Three-dimensional Graphics: The ca Package. *Journal of Statistical Software*, 20, 3.

Narvaiza, J. L. (1981). El Análisis Factorial. Exposición Gráfica e Intuitiva». *Boletín de Estudios Económicos*, 113, 259-285.

Navarro Gómez, M. L. (1983). Aspectos teóricos y una aplicación práctica del análisis factorial de correspondencias. *Estadística Española*, 99, 33-59.

Nishisato, S. (1984). Forced classification: A simple application of a quantification method. *Psychometrika,* 49, 25-36.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 6, 2, 11, 559-572.
http://www.stats.org.uk/pca/Pearson1901.pdf

Rouanet, H.; Ackennann, W.; Le Roux, B. (2001). El análisis geométrico de encuestas: La lección de La distinción de Bourdieu. *Revista Colombiana de Sociología*, VI, 1, 139-145.
http://www.revista.unal.edu.co/index.php/recs/article/view/11063/11729

Salvia, A.; Vera, J. (2010): Heterogeneidad Estructural, Mercado Laboral y Desigualdad Social: El patrón de distribución de los ingresos y los factores subyacentes durante dos fases de distintas reglas macroeconómicas. Ponencia *IV Congreso de la Asociación Latinoamericana de Población* (Cuba), Asociación Latinoamericana de Población (ALAP).

Salvia, A.; Quartulli, D. (2010): *La movilidad y la estratificación socio-ocupacional en la Argentina*. Observatorio de la Deuda Social en Argentina, Buenos Aires, Pontificia Universidad Católica Argentina.

Salvia, A.; Comas, G.; Gutiérrez Ageitos, P.; Quartulli, D.; Stefani, F. (2008): *Cambios en la estructura social del trabajo bajo los regímenes de convertibilidad y post-devaluación. Una mirada desde la perspectiva de la heterogeneidad estructural,* Observatorio de la Deuda Social en Argentina, Buenos Aires, Pontificia Universidad Católica Argentina.

Sánchez, C.; Domínguez, M. (2001). Anàlisi de l'estructura social de les comarques catalanes a partir de dades censals. Metodologia i primera aproximació als resultats. *Revista Catalana de Sociologia*, 14, 193-213.
http://publicacions.iec.cat/repository/pdf/00000024/00000067.pdf

Spearman, C. (1904). "General Intelligence", Objectively Determined and Measured. *The American Journal of Psychology*, 15, 2, 201–292.
http://www.jstor.org/stable/pdf/1412107.pdf

Subirats, M.; López-Roldán, P.; Sánchez, C. (2010). Clases y grupos sociales en la Región Metropolitana de Barcelona. *Papers. Regió Metropolitana de Barcelona*, 52, 105-120.
http://www.iermb.uab.es/htm/descargaBinaria.asp?idRevArt=276

Subirats, M.; López-Roldán, P.; Sánchez, C. (2010). Classes i grups socials a la Regió Metropolitana de Barcelona. *Papers. Regió Metropolitana de Barcelona*, 52, 8-37.
http://www.iermb.uab.cat/htm/descargaBinaria.asp?idRevArt=270

Tabachnick, B.; Fidell0, L. (2007). Principal Components and Factor Analysis. En B. Tabachnick y L. Fidell0, *Using Multivariate Statistics*. Boston: Pearson, 607-675.

Tenenhaus, M.; Young, F. W. (1985). An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis, and other methods for quantifying categorical multivariate data. *Psychometrika,* 50, 91-119.

Thurstone, L. L. (1947). *Multiple-factor analysis*. Chicago: University of Chicago Press

Vicente, M. A. ; Manera0, J. (2003). El análisis factorial y por componentes principales. En J.-P. Lévy y J. Varela, *Análisis Multivariable para las Ciencias Sociales.* Madrid: Pearson Prentice Hall, 327-360.

Van Rijckevorsel, J. 1987. *The application of fuzzy coding and horseshoes in multiple correspondence analysis*. Leiden: DSWO Press.

Volle, M. (1978). *Analyse des données*. Paris: Economica.

Winch, R. F. (1947). Heuristic and Empirical Typologies: A Job for Factor Analysis. *American Sociological Review*, 12, 1, febrero, 68-75.