

López-Roldán, P. (2012). **El proceso metodológico de construcción tipológica de las clases sociales**. En Marina Subirats, *Barcelona: de la necesidad a la libertad. Las clases sociales en los albores del siglo XXI*. Barcelona: Editorial Universitat Oberta de Catalunya.

La metodología empleada en este estudio se dirige fundamentalmente a la construcción de una tipología general de grupos y clases sociales que contemple los numerosos ámbitos temáticos que los configuran según el modelo de análisis considerado. El proceso seguido de obtención y análisis de los grupos y clases sociales, a partir de los datos de la Encuesta de Condiciones y Vida y Hábitos de la Población de Cataluña 2006 (ECVHP), conlleva la aplicación de una metodología específica de tratamiento con técnicas de análisis estadístico multivariantes.

Para llegar a la obtención de la tipología general se ha procedido, en primer lugar, a la obtención de tipologías parciales de cada uno de los ámbitos considerados que posteriormente se han tratado conjuntamente para sintetizarlas en una sola de carácter global. A continuación detallamos las etapas de la metodología seguida con la descripción de las características más destacadas de las técnicas multivariantes utilizadas en su construcción.

La construcción de tipologías (López-Roldán, 1996) satisface la necesidad de clasificar o de estructurar y, en general, de resumir en un conjunto reducido y significativo de categorías o tipos a los individuos, grupos, instituciones, sociedades o cualquier otra unidad de análisis que es objeto de estudio. El objetivo final pues consiste en agrupar o clasificar los individuos en grupos homogéneos según un conjunto de variables seleccionadas que de forma conjunta y simultánea configuran los diferentes perfiles que definen unos grupos o tipos similares. En el presente estudio se desarrolló un procedimiento metodológico y de técnicas de análisis de datos destinados a estructurar la posición social de los individuos del ámbito metropolitano y catalán en términos de clases y grupos sociales.

El proceso de análisis (Domínguez y López-Roldán, 1996) de los resultados que se presentan en el estudio no ha sido apriorístico en el sentido de establecer modelos prefijados como punto de partida y como forma de teoría establecida que se debe validar. De hecho la lógica ha sido más empírica, en el sentido de seguir un procedimiento de interpretación de resultados empíricos obtenidos con el análisis de los datos para llegar a la obtención de la tipología final, lo cual no quiere decir que no se tuvieran referencias teóricas o referencias provenientes tanto de otros estudios como de la experiencia en investigaciones previas, pero no se han explicitado de forma analítica. En este sentido la expresamos como una dinámica articulada, pues los conceptos y su operativización evidencian un modelo conceptual que orienta y justifica la elección de las variables pertinentes, la formalización de las tipologías y la interpretación de los resultados. Y resultados finalmente que configuran una estructuración de la realidad social estudiada, lo que nos lleva a hablar de un proceso metodológico de construcción de una *tipología estructural y articulada*, y a dar cuenta de aquélla en términos de una *medida cualitativa* y de una *explicación estructural* (López-Roldán, 1996).

Siguiendo este modelo hemos diferenciado seis fases analíticas fundamentales desde el punto de vista metodológico:

- 1) La selección de las variables y la preparación para su tratamiento, con un análisis descriptivo simple previo, a partir de la distinción de 15 ámbitos temáticos que se establecen como estructuradores de las condiciones de vida y hábitos de la población y de las posiciones sociales de los individuos en términos de grupos sociales.
- 2) Un análisis de dimensionalización para estructurar inicialmente las variables originales y derivar los principales factores de diferenciación de los individuos en el ámbito considerado. Esta dimensionalización opera mediante el Análisis de Correspondencias Múltiples (ACM).
- 3) En función de los resultados del análisis precedente y a partir de las variables factoriales que se obtienen con el ACM, se procede a realizar un Análisis de Clasificación Automática (ACL) de las unidades en un conjunto de tipos significativos que estructuren el fenómeno estudiado: primero, desde cada ámbito, y, finalmente, desde el análisis global de la interrelación entre ámbitos.
- 4) Con las tipologías generadas se procede a la interpretación, definición y descripción de los perfiles que las caracterizan.
- 5) Finalmente, y de hecho a lo largo de todo el proceso, se procede a validar los resultados obtenidos en cada momento.
- 6) Este proceso de análisis se ha aplicado inicialmente al conjunto de la muestra del ECVHP, es decir, para toda Cataluña, y también se ha hecho un análisis que hace referencia específicamente a la submuestra del Ambito Metropolitano.

A continuación pasamos a detallar estas diferentes etapas.

1.1 Selección de las variables y preparación para su tratamiento

En un primer paso se seleccionan las variables que se utilizan para construir las diferentes tipologías. Las variables consideradas son de dos tipos: fundamentalmente las variables activas, aquellas que intervienen directamente en la construcción y en la definición de las tipologías de cada ámbito, y, de manera puntual, las variables ilustrativas, las que no intervienen pero se utilizan para describir y validar las clasificaciones tipológicas. El análisis ha comportado igualmente la validación del papel de estas variables y la eventual inclusión o exclusión del análisis final en función de las implicaciones interpretativas y conceptuales.

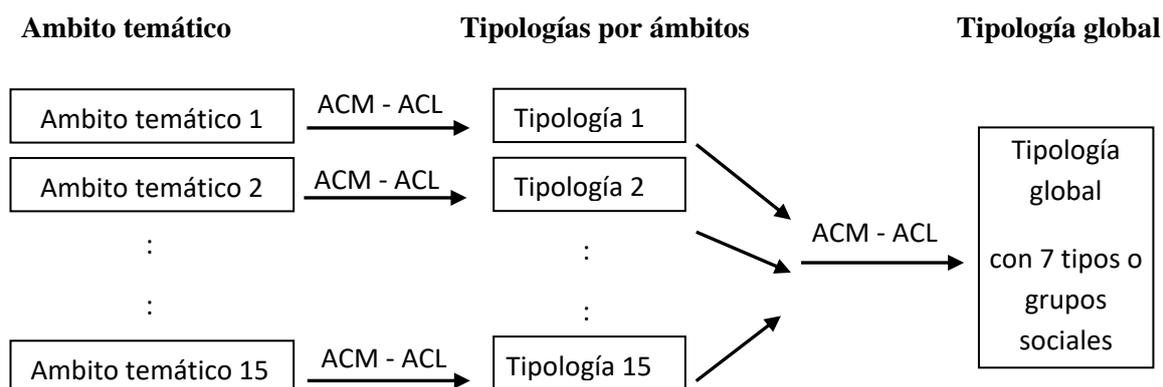
Por las técnicas estadísticas que se utilizan, el nivel de medida de las variables mínimo requerido es el nominal. De hecho todas las variables consideradas son cualitativas o categóricas, medidas a nivel nominal u ordinal. En algún caso las variables han requerido un tratamiento previo para facilitar el análisis o dar más validez a los resultados del análisis multivariable. En este sentido se han realizado dos tipos de tratamientos:

- la categorización de variables continuas para poder analizarlas estadísticamente en las técnicas utilizadas, siguiendo criterios de coherencia conceptual, y
- la recodificación de valores de variables categóricas para adecuarlas al análisis, ya sea por la baja frecuencia de algunos de sus valores o por criterios conceptuales de tratamiento.

Hemos realizado un análisis con los datos del ECVHP poco habitual en la literatura por la gran cantidad de información utilizada: finalmente se han considerado un total de 248 variables distribuidas en 15 ámbitos temáticos.

A partir de estos diferentes ámbitos de análisis el proceso seguido ha consistido en la obtención de una tipología parcial en cada uno de los ámbitos, y un análisis conjunto de las diferentes tipologías obtenidas para construir la tipología global de estructuración de las clases y grupos sociales (Gráfico 1)

Gráfico 1. Esquema del proceso de construcción tipológica



No presentamos la relación detallada de las variables que se han incluido en el análisis, ni los resultados tipológicos parciales obtenidos. Destacaremos algunos contenidos descriptivos básicos de estos ámbitos y, por tanto, parte del contenido del modelo de análisis que se deriva.

Los ámbitos son los siguientes:

- 1) **Nivel de estudios y hábitos culturales** (15 variables). Incluye el nivel de estudios acabados del entrevistado e indicadores culturales como el número de libros en el hogar o la frecuencia de lectura.
- 2) **Trabajo productivo** (18 variables). Relación con la actividad y características del empleo: categoría profesional, sector, tipo de empresa, tipo de relación laboral, tipo de contrato, jornada, salario, estabilidad,...
- 3) **Tipos familiares** (12 variables). En relación al número de miembros del hogar y composición, y distribución del trabajo doméstico y de cuidado.
- 4) **Consumo** (11 variables). Formas de pago en las compras, tipo de establecimiento, periodicidad de las compras, compra del hogar de productos.
- 5) **Recursos económicos** (21 variables). Contempla diversas variables referidas a los ingresos anuales del hogar, a las dificultades económicas, a las posibilidades de ahorro, a las inversiones realizadas, a la posesión de propiedades, a los seguros contratados, a los gastos en vivienda y a la valoración del nivel de vida.
- 6) **Vacaciones** (6 variables). Pernoctaciones fuera del domicilio en varios momentos del año y lugar de vacaciones.
- 7) **Equipamiento del hogar** (18 variables). Posesión de varios artículos en el hogar: lavavajillas, secadora, congelador independiente, vitrocerámica, cámara de vídeo, ordenador, internet, móviles,... así como el número de coches y motos.
- 8) **Vivienda** (17 variables). Régimen de tenencia, superficie, localización, equipamiento de la vivienda (calefacción, aire acondicionado, ascensor, inodoros, duchas, zonas de jardín,

piscina), inconvenientes de la vivienda y posesión de segunda residencia.

- 9) **Formas de relación** (12 variables). Frecuencias de relaciones con familiares, vecinos, amigos,... personas a las que acude en caso de enfermedad, problemas,... y actividades de voluntariado.
- 10) **Tiempo libre** (34 variables). Actividades de tiempo libre en casa: leer, ver televisión, escuchar radio, bricolaje, ordenador, familia,... y fuera de casa: cine, teatro, bares, pasear, excursiones, compras, familia, deporte,...
- 11) **Origen geográfico** (6 variables). Lugar de nacimiento, año de llegada a Cataluña y al municipio, lengua del entrevistado, nivel de conocimiento del catalán y situación legal de residencia.
- 12) **Salud** (6 variables). Valoración de la salud personal y manifestación de dificultades para realizar diversas actividades: salir de casa, subir escaleras, moverse, vestirse y lavarse o comer solo.
- 13) **Entorno** (29 variables). Valoración del estado en el barrio o el entorno más inmediato en relación al alumbrado, la limpieza, las zonas verdes, de diversos servicios y equipamientos,...., de la contaminación, del tráfico, de la inseguridad, de la calidad de vida,...
- 14) **Expectativas** (8 variables). Valoración del nivel de vida actual y futuro de la familia y general, del nivel de ingresos y del trabajo, de las relaciones personales y de la evolución social y política de Cataluña.
- 15) **Voto y asociacionismo** (6 variables). Número de asociaciones a las que pertenece, posicionamiento político, voto en las elecciones municipales, autonómicas y generales.

En cada ámbito, y después también en el análisis global final, se ha aplicado un mismo procedimiento que combina dos técnicas principales: el análisis factorial de correspondencias múltiples con el análisis de clasificación, para obtener las tipologías. Las comentamos en los apartados siguientes.

1.2 Análisis de Correspondencias Múltiples (ACM)

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 o categóricas. El resultado de un ACM es la obtención de unas nuevas variables, dimensiones o factores de diferenciación de los individuos. Estos factores se generan por combinaciones (lineales) de un conjunto original de variables, que originalmente son cualitativas, pero que pueden tratarse con una métrica cuantitativa a partir de la consideración de la definición de una medida de distancia entre los valores o perfiles de las variables que posibilitan el tratamiento estadístico factorial.

Los factores obtenidos son, en número, menores que el de dimensiones inicial configurado por las variables originales; perdiendo una parte de la información inicial (expresada en términos de varianza o de inercia) pero ganando en significación y parsimonia. Además, estos factores se caracterizan por estar ordenados jerárquicamente por su importancia (valores propios) como factores de diferenciación de los individuos, y entre ellos son (linealmente) independientes.¹

Por otra parte, estos factores se pueden representar gráficamente, lo que supone un importante elemento descriptivo y de ayuda inestimable para la interpretación de los resultados al destacar geoméricamente las interacciones entre las variables y, por tanto, ilustra cómo se estructura la información. El objetivo de este tipo de técnica factorial es la ordenación escalar tanto los individuos como de las categorías de las variables analizadas.

¹ Los datos del estudio se han tratado principalmente con el programa informático SPAD (*Système Portable pour l'Analyse des Données*), versión 6. Los procedimientos de ACM y, posteriormente de ACL, son los implementados en este software.

Mediante el ACM, y a partir de las dimensiones o factores que se retienen², podemos asignar a cada individuo una coordenada (puntuación factorial) en un espacio multidimensional continuo. Estos factores tienen la ventaja adicional, de interés para el análisis de clasificación posterior, que se expresan en las mismas unidades de medida (son variables estandarizadas).

1.3 Análisis de Clasificación Automática (ACL)

La técnica del análisis de clasificación automática o de conglomerados (*cluster analysis*) está destinada propiamente a obtener las tipologías al agrupar los individuos según sus similitudes. En nuestro caso las similitudes se calculan a partir de las distancias entre individuos según las puntuaciones factoriales resultantes del ACM considerando como medida la distancia euclídea.

De esta manera (Lozares y López-Roldán, 2000: 147) la técnica estadística multivariable de la ACM nos proporciona las condiciones de aplicación deseables del proceso clasificatorio. Además de reducir la información, nos proporciona, a partir de las variables originales tratadas, un conjunto nuevo de variables de dimensión significativamente menor mediante la acumulación de la mayor parte de la varianza, y, al ser variables que forman base, engendran el subespacio vectorial, resultan incorrelacionadas o linealmente independientes.

Como el número de componentes que se utilizan es de menor dimensión que la definida por las variables originales, las distancias entre los puntos o individuos evaluadas a partir de la ACL diferirán de las distancias definidas con las variables originales, pero precisamente en el mejor sentido a efectos del análisis y objetivos de la ACL, pues lo que obtenemos es una nube de puntos donde los individuos se disponen en función de aquellas características que más los discriminan y los hacen diferentes, con las ventajas adicionales mencionadas de reducción e incorrelación estadística.

Finalmente nos facilita que las unidades de medida de las variables clasificatorias (los factores en este caso) sean las mismas, así se consigue que a la hora de hacer las comparaciones de las unidades sus diferencias vengan expresadas estrictamente por la medida de similitud utilizada y no por el efecto del cambio de unidad de medida, se logra por tanto ponderar su importancia relativa y evita que afecte a los resultados de la clasificación.

La técnica de clasificación automática utilizada es un algoritmo mixto implementado en el software SPAD y que, partiendo de las puntuaciones factoriales del conjunto de individuos, aplica un triple proceso clasificatorio (Lebart, Morineau y Piron, 2004: 177 a 184):

- 1) Una primera clasificación se obtiene por el cruce de varias particiones de base construidas alrededor de centros móviles.
- 2) Las clases estables que se obtienen de este primer procedimiento se agregan a continuación por un método de clasificación jerárquica ascendente según el criterio de *Ward* o de mínima pérdida de inercia.
- 3) Finalmente, las diferentes particiones de los individuos que se pueden obtener a partir del árbol de agregación del procedimiento *Ward* se optimizan o se consolidan mediante una reasignación a los diferentes grupos creados en cada partición con un nuevo proceso de clasificación por centros móviles que mejora la inercia entre los grupos.

A partir del árbol de agregación se trata de determinar el corte que corresponde a la mejor entre las mejores particiones. Esta decisión se toma teniendo en cuenta fundamentalmente un criterio sustantivo de configuración de los grupos, y un criterio más formal o matemático derivado de los cambios que se producen en el árbol de agregación. En el caso del análisis global realizado en

² En el caso del análisis global se han considerado un total de seis factores que conservan el 94,7% de la inercia total según el cálculo transformado de los valores propios (Benzécri, 1979; Greenacre, 2008).

este estudio se decidió considerar una tipología con 7 grupos o tipos cuyos contenidos fundamentan el análisis realizado en este libro.

1.4 Validación de los resultados

Por último, como en todo proceso de investigación, se ha procedido de diversas maneras a validar los resultados del estudio para comprobar su consistencia y la estabilidad de las tipologías obtenidas. En este sentido se han realizado diversos análisis de ACM y de ACL a partir de la inclusión o exclusión de varias variables y de su consideración con codificaciones diferentes (con más o menos modalidades). Hemos tratado los datos también contrastando el conjunto de Cataluña con la submuestra del ámbito metropolitano. Asimismo se han contrastado los resultados de la actual edición de la encuesta con los de la de la anterior y se han validado desde el punto de vista sustantivo.

Por último hay que señalar dos particularidades de los datos del ECVHP para dar sentido a la configuración de los grupos sociales. Por un lado, los datos hacen referencia a estimaciones sobre una población definida en términos de individuos de 16 y más años, por lo que las características sociales de su hogar no dejan de ser atributos individualizados, y la distribución del peso de los grupos sociales se hace, en consecuencia, en relación a individuos que también pertenecen a hogares.

Por otro lado se ha procedido a construir la tipología de grupos sociales a partir de un análisis que consideró el conjunto de toda la muestra, es decir, la totalidad de Cataluña y, a continuación, se ha considerado la distribución resultante con la submuestra específica del ámbito metropolitano. Esta forma de proceder nos permite tener una visión global de la estructura social catalana y comparar las dinámicas de estructuración social metropolitanas con las del conjunto catalán. Hemos constatado, sin embargo, dado el importante peso poblacional metropolitano y la actual estructuración de la realidad social, que los perfiles sociales de los grupos no generan grandes diferencias como se han dado en el pasado.

Referencias bibliográficas

Benzécri, J.-P. (1979). «Sur le calcul de taux d'inertie dans l'analyse d'un questionnaire». *Les Cahiers de l'Analyse des Données* 4, 377-378.

Domínguez, M.; López-Roldán, P. (1996) «La construcció de tipologies: procés i tècniques d'anàlisi de dades». *Papers. Revista de Sociologia* 48. P. 31-39.

Greenacre, M. (2008). *La práctica del análisis de correspondencias*. Madrid: Fundación BBVA.

Lebart, L.; Morineau, A.; Piron, M. (2004). *Statistique exploratoire multidimensionnelle*. Paris: DUNOD.

López-Roldán, P. (1996). «La construcción de tipologías: metodología de análisis». *Papers. Revista de Sociologia* 48. P. 9-29.

Lozares, C.; López-Roldán, P. (2000). *Anàlisi multivariable de dades estadístiques*. Bellaterra: Universitat Autònoma de Barcelona. Colección Materials, 93.