

Implicaciones sociológicas en la construcción de una muestra estratificada

PEDRO LÓPEZ ROLDÁN Y CARLOS LOZARES COLINA

Departamento de Sociología
Universitat Autònoma de Barcelona
Pedro.Lopez.Roldan@uab.cat
Carlos.Lozares@uab.cat

Recibido: 05.10.2007
Aceptado: 19.12.2007

1. INTRODUCCIÓN

El objetivo de este artículo¹ no consiste en exponer de manera detallada el tipo de muestra estratificada (en adelante ME) realizada estos años para la «*Enquesta de Condicions de Vida i Hàbits de la Població de Catalunya*» (ECVHP)², incluso tampoco la de la última edición de 2006³, aunque presenta algunas variaciones importantes con relación a las precedentes. Se trata sobre todo de presentar algunos análisis y conclusiones referidos a apartados o fases específi-

¹ Una primera versión de este artículo se presentó como ponencia con el mismo título en el grupo de trabajo de Metodología en el IX Congreso Español de Sociología, celebrado en Barcelona entre el 13 y el 15 de septiembre de 2007.

² En diversas publicaciones anteriores (Dominguez, López y Lozares, 2000; López Roldán y Lozares Colina, 1999; López, Domínguez y Lozares, 2000; Lozares, López y Flores, 2003; Lozares y Dominguez 1995; Lozares y Dominguez, 1993; Lozares y Dominguez, 1996; Lozares, López, Domínguez y Borrás, 1994; Lozares y López, 1990; Lozares y López, 1991a; Lozares y López, 1991b; Lozares y López, 2000; López y Lozares, 2007) hemos dado cuenta de las diferentes muestras realizadas en las sucesivas encuestas llevadas a cabo: Área Metropolitana (1985), Región I (1990), Región Metropolitana (1995), Provincia de Barcelona (2000) y actualmente toda Cataluña (2006).

³ La encuesta, en sus cinco ediciones, ha sido un proyecto impulsado y financiado por la *Mancomunitat de Municipis de l'Àrea Metropolitana de Barcelona*, la *Diputació de Barcelona* y por el *Institut d'Estudis Regionals i Metropolitans de Barcelona* (Iermb). La edición de 2006 ha pasado a formar parte del Plan Estadístico de Cataluña mediante un acuerdo de colaboración con el *Institut d'Estadística de Catalunya* (Idescat). Para ampliar la información sobre las características de la encuesta y las diferentes publicaciones que ha generado a lo largo del tiempo se pueden consultar las webs: <http://www.enquestadecondicionsdevida.cat/> y <http://www.idescat.net/cat/societat/qualitat/ecvhp.html>.

cos de la muestra que ponen más de manifiesto la pertinencia y contenido sociológico que conlleva el diseño de ME desarrollado. No obstante, en el primer apartado, daremos cuenta, de forma resumida, de las características definitorias de su proceso de construcción.

El objetivo que nos proponemos consiste pues en mostrar que, además de la función instrumental y directa que cumple toda ME, en tanto que mecanismo de representación para una encuesta, la documentación de base empleada y el tipo y modalidad de ME que llevamos a cabo conlleva unas virtualidades y resultados que superan y desbordan su primigenio objetivo instrumental, es decir, conlleva y contiene unos productos y conclusiones intermedios de naturaleza y utilidad metodológica y sociológica con relación a la encuesta y sus resultados.

Más concretamente, dicho objetivo general se desglosa en otros más precisos como son:

- 1) la importancia de disponer de una amplia gama conceptual de indicadores y/o variables a la hora especificar y elegir las variables-criterios empleadas en la construcción de los estratos;
- 2) el rol decisivo que juegan los contenidos de las variables-criterios y de los estratos propios de la ME como un producto sociológico en sí y, además, como prefiguración y validación de los resultados de la encuesta y del censo;
- 3) la posibilidad que ofrecen todos y cada uno de los estratos, y sus territorios correspondientes, de ser objeto de un análisis sociológico autónomo, a partir de los datos que provienen del censo y de la misma encuesta en ellos; consecuentemente, por tanto, la de conjugar en cada uno de los estratos la información que proviene de los datos censales y de la encuesta así como la utilización de la tipología muestral como variable-base de comparación de los datos de la encuesta;
- 4) la ventaja que proporciona al análisis y a la interpretación social el hecho de elegir la sección censal como soporte o base de referencia de las variables usadas en la ME, en el sentido de las posibilidades que procura la territorialización y digitalización de los resultados de los estratos de la muestra y de los datos de la encuesta con relación a cualquier división territorial;
- 5) la importancia que el tipo de ME desarrollada tiene también para otras 'utilidades' sociológicas.

Por tanto, en un primer apartado, vemos, sucintamente, los rasgos y fases más importantes de la muestra de la ECVHP 2006, dando cuenta de los pasos de su construcción⁴ y, en los apartados siguientes, consideramos, de manera sucesiva, cada uno de los objetivos enunciados.

⁴ La exposición completa y detallada de la construcción de la muestra se puede consultar en López y Lozares (2007, en prensa).

2. CARACTERÍSTICAS DE UNA MUESTRA ESTRATIFICADA

En el diseño de la muestra se establece como objetivo general la extracción de una muestra aleatoria estratificada representativa de la población de Cataluña de 16 y más años. La elección del procedimiento de muestreo estratificado se justifica básicamente por criterios de precisión frente a otros métodos y por la heterogeneidad social que caracteriza a la población objeto de estudio. El proceso de construcción seguido se puede resumir en las siguientes fases.

- 1) Tratamiento previo de la información censal en vistas a la elección de las variables-criterio para la construcción e identificación de los estratos.

Los contenidos del estudio de la encuesta recogen una gran diversidad de temáticas sociales de interés⁵. El objetivo consiste precisamente en poder expresar mejor la variedad de la realidad social de Cataluña sobre las condiciones de vida y hábitos de la población. Dichos contenidos y objetivos obligan, en la construcción de la muestra estratificada, a tomar en consideración como variables-criterio de estratificación aquellas que den mejor cuenta de dicha multiplicidad social de la población. Esta información poblacional primaria y diversa proviene del Censo de Población a través de sus indicadores⁶.

Una vez elegidas las variables más pertinentes del Censo que representen la máxima diversidad social⁷ se lleva a cabo un ejercicio de reducción y estructuración de las mismas a fin de identificar y elegir las variables-criterio para construir los estratos de la muestra. Así se garantiza la presencia en la muestra de estratos que expresan las diferencias poblacionales y con unos contenidos que guardan gran correlación con los de la ECVHP. La construcción de los estratos comprende las dos fases siguientes.

- 2) Las variables-criterio para la construcción e identificación de los estratos: el análisis factorial de componentes principales.

Se pretende reducir y expresar los principales factores de diferenciación de la población y que, a su vez, se corresponden con las variables-criterio

⁵ El cuestionario recoge más de 500 variables directas con las que se pretende proporcionar datos sobre las dinámicas sociales y territoriales que se generan en Cataluña, a partir de 14 ámbitos temáticos: estructura del hogar, procedencia geográfica, lengua, nivel de estudios, situación laboral, salud, ingresos, propiedades e inversiones, consumo, vivienda, percepción del barrio y del entorno, uso e imagen del territorio, cultura y ocio, relaciones sociales, asociacionismo e ideología.

⁶ Para tratar esta información se agrupan los individuos en términos de las secciones censales donde residen (un total de 5222 para el año 2001). De esta forma se dispone, para el conjunto de indicadores o variables seleccionados (inicialmente un total de 234), del número de personas que, en cada sección, poseen la característica dada por el indicador y, en consecuencia, se dispone también del porcentaje o la tasa que representa dicha variable sobre el total de la población de la sección censal.

⁷ Se trata de variables de naturaleza socio-demográfica, cultural-educativa, de actividad laboral y profesional e inactividad, de capital económico, de composición de los hogares, de movilidad territorial y de tamaño del hábitat. De las 234 variables iniciales, y como resultado del análisis de sus correlaciones mutuas para examinar su redundancia y significatividad se eligen 82 indicadores o variables.

terio usadas para la construcción de los estratos. La estructura de la información permite reducir las 82 variables consideradas a sólo 7 dimensiones, ejes o variables factoriales, que acumulan el 69% de la varianza total.

- 3) El proceso de construcción e identificación de los estratos en la ME: el análisis de clasificación automática.

El objetivo consiste en clasificar las secciones censales en estratos que estarán compuestos por conjuntos de secciones lo más homogéneas posible dentro de cada estrato y lo más heterogéneas entre las secciones de los diferentes estratos en función de las variables factoriales extraídas del análisis factorial. Finalmente se obtienen y validan 10 estratos o grupos de secciones censales⁸.

- 4) El cálculo del error y del tamaño muestral.

El cálculo del tamaño muestral se realiza a partir de los parámetros habituales: tamaño de la población, estimación de la media y de la variabilidad, nivel de significación y error muestral⁹. El tamaño de la muestra para el conjunto de Cataluña se fijó inicialmente en 8.348 individuos que se corresponde con un error muestral relativo del 0,8%.

- 5) La afijación de la muestra en los estratos.

Este número de individuos se distribuye entre los estratos con el criterio de afijación óptima de Neyman. Con este criterio se opera el efecto de la estratificación de la muestra según el cual cuanto más grande y variable es un estrato mayor proporción de muestra se le asigna. Por tanto, no se trata de una distribución estrictamente proporcional a la población de cada estrato, sino que a efectos de optimización, de ganancia en la precisión de las estimaciones, se adopta este doble criterio.

- 6) Determinación de las cuotas muestrales.

Con el número muestral de cada estrato se procede a la asignación proporcional de las cuotas de muestra en términos de individuos correspondientes a cada sección censal del estrato. Con este reparto se garantiza el cumplimiento de la aleatoriedad de la muestra en la elección de un individuo que pertenece a una sección determinada. La determinación de las cuotas de muestra por sección sigue un proceso en dos etapas:

⁸ El análisis de clasificación (o de *cluster*) conlleva dos etapas. En una primera se realiza una clasificación jerárquica ascendente mediante el procedimiento de *Ward* (de mínima pérdida de inercia) y se fijan y clasifican las secciones en 10 grupos. En una segunda se opera una clasificación no jerárquica por la agregación alrededor de centros móviles con el objetivo de optimizar la asignación de las secciones a los estratos.

⁹ A partir de la expresión: $n = \frac{z^2 \cdot \sigma_y^2}{e^2 \cdot \bar{Y}^2}$, donde se considera un nivel de significación de 2 sigmas ($z = 2$), y como medida de variabilidad (σ_y^2) la distancia cuadrática euclidiana de cada sección al centro global de la nube de puntos compuesta de secciones censales en un espacio vectorial reducido a una base de siete vectores o variables-factores, y, como parámetro, la media de esta distancia (\bar{Y}^2).

- a) Inicialmente, se asigna a cada sección censal de un estrato la cuota de muestra que le corresponde según una distribución proporcional de acuerdo con su población de 16 o más años.
- b) La asignación inicial de cuotas para cada sección censal se distribuye en un segundo momento considerando, en cada estrato, una subestratificación a partir de un doble criterio de división territorial: según siete grandes ámbitos territoriales y según el tamaño de los municipios a los que pertenecen las secciones. Además, para favorecer la concentración territorial y reducir los elevados costes derivados de los obligados desplazamientos por todo el territorio catalán para la realización de las entrevistas personales, se ha aplicado un criterio de redistribución de las cuotas de muestra.

La necesidad de garantizar un nivel de representatividad suficiente para determinadas unidades territoriales de interés en el estudio condujo, por último, a la ampliación de la muestra añadiendo 2.094 encuestas que se repartieron de forma desigual entre algunos ámbitos territoriales y comarcas para garantizar, como submuestra, un tamaño suficiente. Así, la muestra teórica final alcanza los 10.446 individuos.

7) Ponderación a posteriori de la muestra y elevación.

El diseño muestral estratificado supone la afijación no proporcional de las cuotas de muestra de cada estrato. Según el tamaño poblacional del estrato y la variabilidad de las características sociales propias de éste hay individuos de 16 y más años que tienen una probabilidad mayor de ser elegidos a partir de la cuota que se asigna al estrato donde se ubica la sección censal a la que pertenece, es decir, no se garantiza el criterio de equiprobabilidad cuando un individuo es elegido al azar. La ponderación a posteriori de la muestra restituye la proporcionalidad poblacional de los diferentes substratos definidos por la estratificación inicial de la muestra, el ámbito territorial y el tamaño del hábitat.

Una vez realizada la ponderación de los datos muestrales, éstos se pueden expresar en magnitudes poblacionales a partir de la elevación de la muestra.

Finalmente, se obtiene información de 10.398 personas encuestadas, que una vez ponderadas, corresponden a un tamaño de muestra al que se asocia un nivel de error global del 0,7 %, para un nivel de significación de 2s.

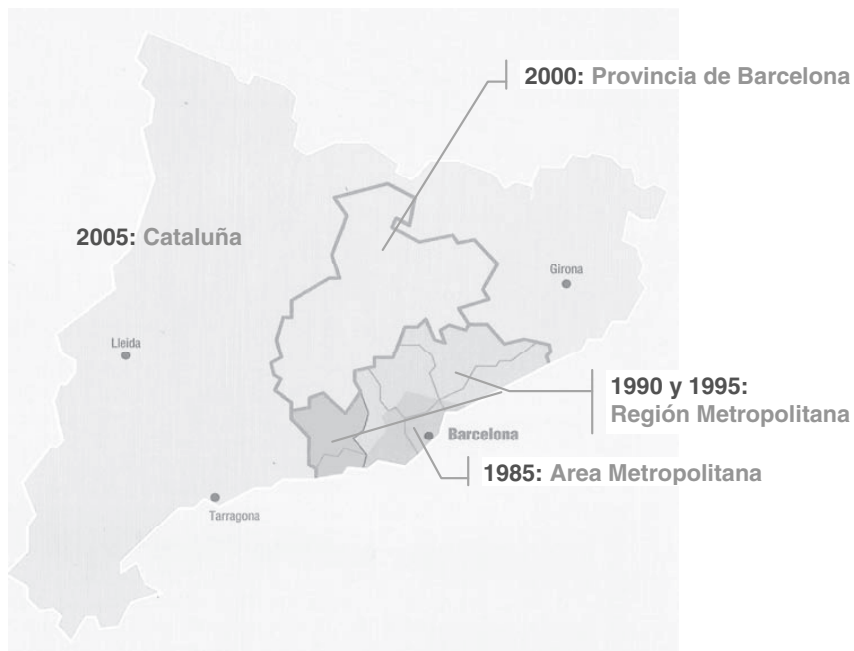
8) Validación de la muestra.

El ejercicio de ponderación se acompañó de la necesaria validación de los resultados a partir de contrastar los datos muestrales derivados de la encuesta con la información actualizada del Padrón de Habitantes del año 2006.

Para acabar esta descripción de las fases de la muestra y su caracterización se presenta (Tabla 1 y Mapa 1) un resumen comparativo de las diferentes ediciones.

Tabla 1. Comparación de las muestras de la ECVHP

Edición de la ECVHP					
	1985	1990	1995	2000	2006
Ambito Territorial	Área Metropolitana de Barcelona	Región I	Región Metropolitana	Provincia de Barcelona	Cataluña
Municipios	27	129	162	311	946
Superficie	476 km ²	2459 km ²	3235 km ²	7718 km ²	32106 km ²
Población % Cataluña	3.096.748 49,2%	4.090.406 67,9%	4.264.422 70,4%	4.628.277 76,0%	6.343.110 100%
Universo estadístico	> 18 años 2.174.363	> 18 años 2.991.131	> 18 años 3.275.458	> 18 años 3.762.462	> 16 años 5.408.264
Muestra	4.912	5.061	5.263	6.830	10.398
Error relativo (n.s. 2σ)	± 1,9	± 1,9	± 1,8	± 1,5	± 0,7
Marco de la muestra	Censo electoral 1984	Censo electoral 1989	Censo electoral 1994	Censo electoral 1999	Padrón 2005

**Mapa 1.** Territorios de las diferentes ediciones de la ECVHP.

3. EL CONJUNTO DE VARIABLES (INDICADORES CENSALES) QUE FORMAN UNA BASE AMPLIA SOCIO-CONCEPTUAL PARA LA ELECCIÓN POSTERIOR DE LAS VARIABLES-CRITERIO DE ESTRATIFICACIÓN

Una muestra aleatoria, estratificada o no, será menos sesgada cuanto mejor y más precisamente represente la realidad social del territorio en cuestión y no sólo, aunque también, la distribución numérica de la población. La muestra estratificada, persigue, entre otros, alcanzar los dos objetivos. No sólo ha de tener en cuenta la distribución aleatoria y proporcional del número de muestra sino que tiene también como finalidad que dicha distribución se haga según criterios de contenido social, es decir, criterios que toman en cuenta las proporciones o distribución de los valores de algunas variables (muchas o pocas) en el territorio. Cuanto más den cuenta dichas variables elegidas¹⁰ la variación de contenidos sociales de la población tanto más se acercarán y respetarán el objetivo anunciado. Además, dado que sobre dichas variables se lleva a cabo la factorialización y que los factores elegidos contienen una buena parte de la variabilidad o varianza de dichas variables aseguramos también por ellas en la muestra las diferencias sociales del territorio. Por tanto, la necesidad de la estratificación descansa sobre dichas exigencias, es decir, del supuesto fundamentado de que la muestra es tanto más precisa, además de por el error estadístico, cuanto mejor representan las variables elegidas de origen la realidad y la variedad social del territorio.

Así pues estas variables iniciales, de las que la proporción y distribución de sus valores se asegura en la muestra, juegan un rol decisivo. La tipificación del territorio en los estratos de la muestra estratificada, y a partir de ellas más directamente de las variables criterio, ha de ser la mejor imagen posible de la realidad del territorio. Cuando se procede a obtener muestras de cada estrato se asegura la presencia en la muestra final de la proporción adecuada de esa o esas variables-criterio.

Es habitual limitar estas variables originales a unas pocas, y generar los estratos por las combinaciones de sus valores (por ejemplo, la edad, género, categoría socio-profesional, ...), de las que se puede tener una cierta seguridad previa de su funcionamiento en el territorio. Evidentemente con dicho número de dos o tres variables de partida no tiene sentido llevar a cabo un análisis factorial para elegir las variables-criterio. Estas variables acostumbran pues a ser directamente las variables-criterio, que son sobre las que se construyen los estratos.

En la muestra que nos ocupa, la cuestión de la elección de las variables-criterio tiene que resolver la necesidad de analizar múltiples aspectos de la vida social, siendo difícil justificar la elección de unas pocas para satisfacer el trata-

¹⁰ En el apartado siguiente se comenta cómo se procede a la factorialización de las variables que aquí mencionamos. Dichos factores resultantes serán denominados allí variables-criterio, que serán las definitivamente elegidas para formar los estratos.

miento de esa multiplicidad de fenómenos. Una razón adicional para llevar a cabo el paso de una estratificación bivariada o trivariada a otra multivariada, es la disponibilidad de los indicadores censales con un grado de validez suficiente, dado que se extienden a toda la población.

La solución adoptada, pues, consiste en analizar un conjunto extenso de variables censales, las proporcionadas por el Censo de Población de 2001, con el objetivo ya explicitado de que den cuenta lo más precisamente posible de la diversidad social de Cataluña. Su elección, evidentemente, está condicionada por la disponibilidad en el cuestionario censal, y por su fiabilidad, pero también por la pertinencia conceptual de acuerdo con los objetivos del estudio. Y aquí la mirada sociológica adquiere un protagonismo insustituible para orientar y justificar la elección final de las variables.

De un total de 234 indicadores iniciales extraídos del Censo, y expresados como variables mediante el cálculo del porcentaje de la población de una sección censal que posee una determinada característica sobre el total de la población de dicha sección (en total 5.222), se procede a una selección de 82, resultado de un análisis estadístico-correlacional (que muestra o la manifiesta combinación lineal entre algunas variables o el escaso valor o dispersión de otras) y de su justificación conceptual por el sentido que adquieren en la estructuración social de la realidad.

El valor añadido de este proceder, dentro de la idea de los efectos benéficos de la modalidad de la construcción de la muestra que proponemos, consiste en disponer de y utilizar una panoplia considerable de variables para la búsqueda, por su reducción en la factorialización, de las variables-criterio. La variedad de contenidos que representan, se acercan lo más posible, en la medida que existen en el censo, a las que se enuncian y se obtienen en la muestra. Esto quiere decir, que como resultado parcial de la construcción de la muestra, disponemos de una apreciable cantidad y diversidad conceptual de variables cuyos resultados se pueden cotejar con los obtenidos en la encuesta. Por tanto, tenemos un medio de validar, con relativa extensión de contenidos, los resultados de la encuesta.

4. LAS VARIABLES-CRITERIO DE ESTRATIFICACIÓN Y LOS ESTRATOS COMO PRE-IMAGEN SOCIOLOGICA DE LA ESTRUCTURA SOCIAL DE CATALUÑA

Dado el gran número de variables inicialmente elegidas, 82, y de la imposibilidad de usar todas ellas como variables-criterio de estratificación, se requiere un procedimiento que permita reducir dicha cantidad e información original para utilizarla en la construcción de los estratos. Además, mantener dicho número no sólo no permite utilizar todas las variables para realizar los estratos sino que además es parcialmente inútil porque mucha de la información que comportan dichas variables es redundante dada la correlación existente entre ellas.

El proceso seguido de construcción de los estratos combina un análisis factorial de componentes principales que nos permite reducir y expresar en términos de siete variables factoriales las dimensiones fundamentales de diferenciación de la población de Cataluña (variables-criterio de estratificación), con un análisis de clasificación automática que conglo mer a las secciones censales en 10 grupos (los estratos de la muestra).

Las implicaciones sociológicas de este proceder se expresan en un doble sentido. En primer lugar, como producto substantivo de naturaleza sociológica, autónomo, que supera la necesidad primera e instrumental de constitución de los estratos de la muestra. Tanto los factores como los grupos obtenidos más adelante con la estratificación nos permiten describir, dimensionalizar y estructurar la realidad social catalana actual. Es decir, descubrir, cuáles son los contenidos sociales que más explican, diferencian o discriminan más a la población; es decir una buena pre-imagen de los factores sociales.

El simple enunciado de la identidad de los factores (Tabla 2) nos ofrece una amplia panoplia y complejidad de contenidos en cuanto a los factores explicativos del conjunto de los datos, que va más allá de la simplicidad (o reducción de su número) con que se abordan los factores o variables más explicativas en las encuestas. La identidad de estos factores, de claro contenido sociológico, se equipara a las variables-criterio de estratificación y servirán, además, de guía, validación y complemento en la interpretación de los datos de la encuesta.

Tabla 2. Factores del análisis de componentes principales (*)

El primer factor:	<i>La categoría socio-profesional (23%)</i>
El segundo factor:	<i>El origen geográfico: autóctonos vs. antigua inmigración (16%)</i>
El tercer factor:	<i>El ciclo vital (15%)</i>
El cuarto factor:	<i>Rural-urbano (13%)</i>
El quinto factor:	<i>La nueva inmigración (13%)</i>
El sexto factor:	<i>La actividad laboral (10%)</i>
El séptimo factor:	<i>La movilidad territorial (10%)</i>

(*) Entre paréntesis se incluye el porcentaje de varianza explicada por cada factor.

De manera complementaria, el análisis de clasificación genera y da identidad (Tabla 3) a grupos sociales en una tipología de estructuración social, equiparándose a los estratos poblacionales que expresan su homogeneidad interna y la heterogeneidad entre ellos.

De esta forma la construcción de los estratos va más allá de un mero instrumento muestral para convertirse en una análisis de estructura social y en una guía sustantiva de análisis de la encuesta y de su validación posterior. Esta función es impensable en la forma habitual de aplicar las ME.

Tabla 3. Grupos del análisis de clasificación (*)

Estrato 1	<i>Población catalana envejecida, clases trabajadoras en ciudades no metropolitanas (11%)</i>
Estrato 2	<i>Movilidad de familias jóvenes de clase media (8%)</i>
Estrato 3	<i>Población joven de clase trabajadora de municipios urbanos (14%)</i>
Estrato 4	<i>Los pequeños municipios rurales (10%)</i>
Estrato 5	<i>Población urbana más joven de clase trabajadora precaria (8%)</i>
Estrato 6	<i>Clases sociales medias y alta de las grandes ciudades (8%)</i>
Estrato 7	<i>Antigua inmigración en el área metropolitana (15%)</i>
Estrato 8	<i>Población envejecida urbana de la antigua inmigración (7%)</i>
Estrato 9	<i>Población de núcleos urbanos de clase media-alta algo envejecida (15%)</i>
Estrato 10	<i>La nueva inmigración (4%)</i>

(*) Entre paréntesis se incluye el porcentaje de secciones censales de cada estrato sobre el total.

En un segundo sentido, cabe señalar también la dimensión metodológica del procedimiento de construcción. Como hemos sugerido en diversas publicaciones (Lozares, 1990; López, 1994, 1996), el uso complementario de estas técnicas constituye un diseño de análisis que satisface la necesidad habitual de la investigación sociológica de construir tipologías estructuradoras de lo social. Es lo que hemos denominado como tipologías estructurales y articuladas, un instrumento de operativización conceptual, construido de forma articulada entre la teoría y la realidad empírica, y destinado a definir, estructurar y medir la complejidad multidimensional de los fenómenos sociales¹¹.

¹¹ En este sentido hemos propuesto el concepto de «explicación estructural» (López, 1996), en la medida en que una tipología sistematiza una regularidad relacional entre tipos. La explicación del fenómeno proviene de comparar y de organizar la relación entre una pluralidad de situaciones o hechos concretos que conceptualizamos buscando en ellos un orden o una regularidad que los reúne, relaciona e identifica en forma de tipos que expresan la estructura no necesariamente manifiesta del fenómeno. La diversidad del fenómeno así identificada pone de manifiesto las diferencias y las semejanzas que caracterizan un todo y sus partes y donde, si bien no existe una relación de causalidad en el sentido que habitualmente se le atribuye a la explicación, sí se obtiene un esquema de interrelación que explica cómo se ordena y regulariza la realidad estudiada, cómo subyace un modelo de comportamiento de la realidad social que nos permite hablar de una explicación estructural. Como de forma sugerente nos señala J. Ibáñez (1985: 97) el significado del término «explicar» es el de desplegar, llevar la madeja o la maraña a un plano. En inglés *explain* significa explicar, proyectar sobre un plano. Por su parte, la palabra «estructura» tiene un significado vinculado a la noción de espacio y a la disposición o colocación de partes o elementos de una cosa, de un todo.

5. EL TERRITORIO Y LA POBLACIÓN DE CADA ESTRATO COMO OBJETOS DE ANÁLISIS SOCIOLOGICOS AUTÓNOMOS: DE LOS ESTRATOS A LAS ZONAS SOCIALES

Los resultados sustantivos de la ME se pueden conjugar posteriormente con los de la encuesta. Lo que denominamos aquí Zonas Sociales¹² es el resultado de implementar en los estratos construidos los datos y resultados correspondientes que provienen de la encuesta.

Dado que las unidades de referencia son secciones censales, el conjunto de las Zonas Sociales llenan toda Cataluña cual regueros o «manchas» sociales que se cruzan e interpenetran en el territorio. Cada Zona Social puede ser objeto de un análisis sociológico con los datos de la encuesta y cada territorio geográfico o administrativo puede estar atravesado por varias Zonas Sociales. En cada territorio, en particular, se puede efectuar un análisis específico de diferencias sociales internas, de dimensionalización y configuración de grupos sociales.

El proceso seguido se puede esquematizar de la forma siguiente (Gráfico 1):

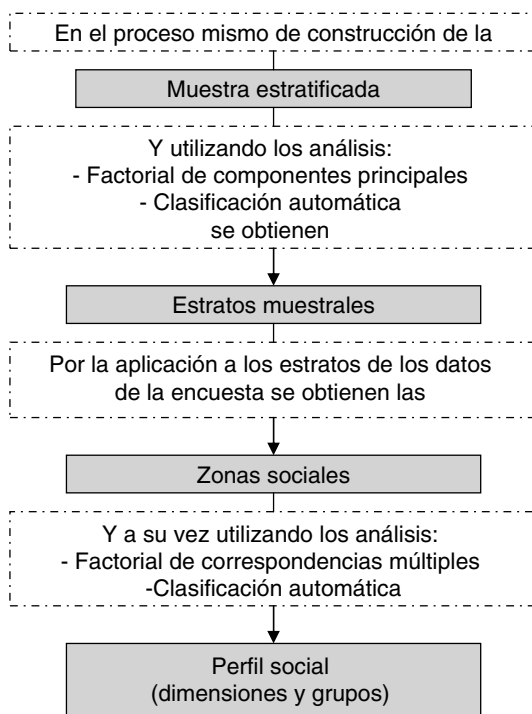


Gráfico 1. Esquema del análisis de la Zonas Sociales.

¹² Para una aplicación de estos supuestos ver Lozares y Domínguez (1996); Lozares, López y Flores (2003).

Tomando como base las secciones censales, se pueden hacer también análisis de territorios administrativos en función de las Zonas Sociales que los cruzan, o en los que se inmergen, obteniendo así, primero, una visión sociológica diferenciada del territorio y, segundo, una posibilidad comparativa con otros territorios, dado que los estratos se refieren a toda Cataluña. Además, dados los niveles de error es posible generar una variable categórica con los estratos como valores y cruzarla por las variables de la encuesta al tener las secciones censales como referencia común de ambos tipo de datos.

6. UNA CARTOGRAFÍA SOCIAL DE CATALUÑA

Dada la base territorial, al tomar las secciones censales como localización de los estratos (y de las Zonas Sociales) y la posibilidad de digitalizarlas, es factible, entonces, obtener una cartografía social o construcción de mapas sociales. La visión que se adquiere del territorio (Cataluña u otros de menor rango) es la de unos conglomerados de realidades sociales plurales que los cruzan. Una visión unificada de un territorio: Cataluña, provincias, comarcas, ámbitos territoriales, municipios, etc. puede ocultar su diversidad territorial bajo la homogeneidad que impone un estudio general de los mismos. La vuelta a una realidad, localizable geográficamente, más desmembrable como las Zonas Sociales, nos permite ver un mapa más variado en sus calles y secciones censales y descubrir, además, primero que en el interior de cada territorio, existen otros micromundos sociales y, segundo, que en el interior de cada Zona Social se puede dar una nueva variedad social si la analizamos en sí misma como es factible.

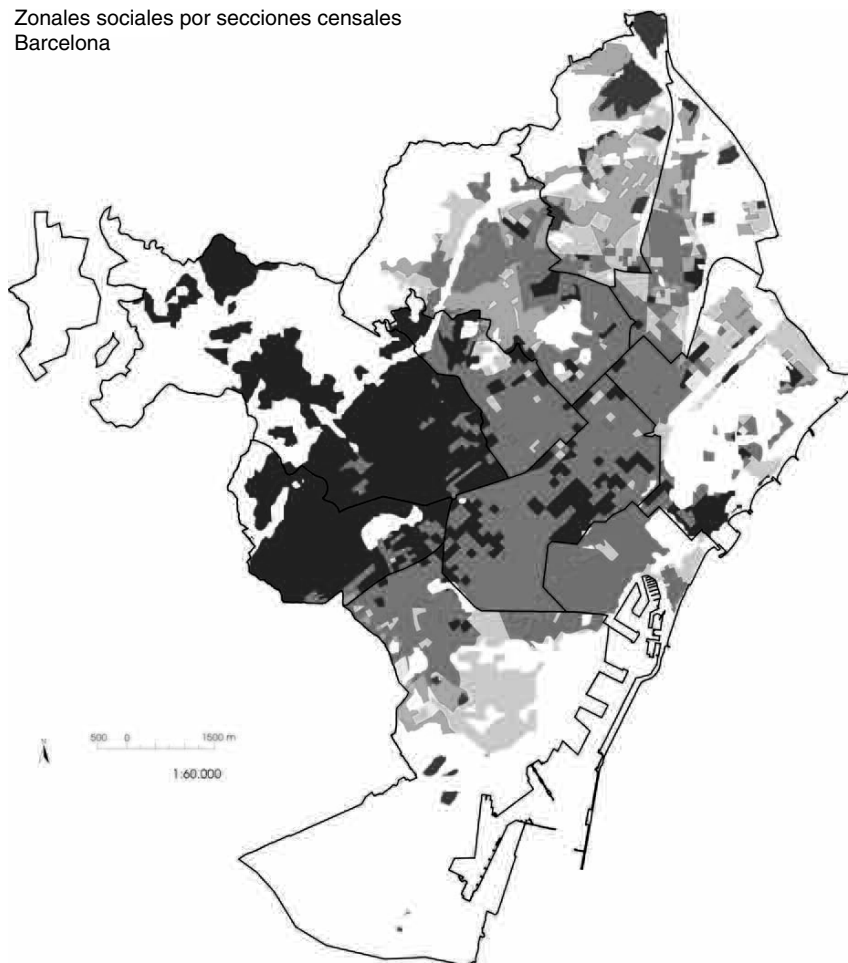
A modo de ejemplo de aplicación se presenta el mapa social de la ciudad de Barcelona (Mapa 2), aunque el contenido es el de los estratos (con el contenido de los datos del censo) no el de las Zonas Sociales (con los contenidos de los datos de la encuesta pues éstos se encuentran en fase de elaboración y análisis).

7. OTRAS «UTILIDADES» INSTRUMENTALES Y SOCIOLÓGICAS DE LA ME PROPUESTA

Genéricamente la ME tiene una serie de ventajas instrumentales adicionales de naturaleza técnica, además de algunas de las estadísticas señaladas aunque no desarrolladas, que se acrecientan con la modalidad de ME presentada, por ejemplo, entre otras:

- la reducción del sesgo en el trabajo de campo pues se puede controlar la proporción muestral en cada estrato por las variables-criterio siguiendo las encuestas entrantes;
- la posibilidad de llevar a cabo sustituciones en el caso de no respuesta tomando la encuesta sustituyente en el mismo estrato, dadas las características homogéneas de su población aparte de conocer previamente su localización;

Zonales sociales por secciones censales
Barcelona



Zonas sociales

- Población urbana más joven de clase trabajadora precaria
- Población catalana envejecida, clases trabajadoras en ciudades no metropolitanas
- Los pequeños municipios rurales
- Población envejecida urbana de la antigua inmigración
- Población joven de clase trabajadora de municipios urbanos
- Antigua inmigración en el área metropolitana
- La nueva inmigración
- Movilidad de familias jóvenes de clase media
- Población de núcleos urbanos de clase media-alta algo envejecida
- Clases sociales medias y alta de las grandes ciudades

Mapa 2. Cartografía social de la ciudad de Barcelona.

- la facilidad para agrupar encuestas dentro del mismo estrato cuando hay dificultad o es costoso el acceso a alguna localidad o sección en las que se han de realizar muy pocas encuestas;
- la seguridad de la presencia en la muestra de grupos minoritarios de especial interés para los objetivos del estudio;
- la ausencia de incompatibilidad con respecto a otros métodos o diseños de muestreo, por ejemplo, el polietápico y el de conglomerados.

La modalidad de ME presentada tienen además otros efectos benéficos y útiles para otro tipo de estudios que necesiten un conocimiento del territorio similar al representado por lo que significan las Zonas Sociales.

En primer lugar, facilita la elección de otros tipos de muestras con una entidad de base distinta a los individuos, en el interior de territorios con diferentes Zonas Sociales: por ejemplo, selección de centros de enseñanza, de comercios, de asociaciones, hoteles, servicios sociales, etc.

Así, por ejemplo, los resultados de la muestra estratificada se han utilizado en la construcción de la muestra del *Panel de Famílies i Infància*¹³ donde el objetivo era extraer una muestra representativa de la población del alumnado de los centros educativos de secundaria públicos y privados de Cataluña. El diseño comporta dos etapas hasta la elección de las unidades finales. En una primera etapa se selecciona una muestra aleatoria de centros de acuerdo con su estratificación social según el territorio donde están localizados. En una segunda etapa se procede a la selección del alumnado de cada centro.

El diseño de esta muestra bietápica tiene la finalidad de concentrar territorialmente la selección del alumnado de secundaria a partir de la selección de un número reducido de centros y constituir el tamaño de muestra necesario con el alumnado de estos centros que actúan de conglomerados. La cuestión que se suscita, en consecuencia, es cómo efectuar la selección de estos centros para garantizar la presencia en la muestra de la diversidad de características sociales de la población estudiantil de secundaria de acuerdo con los objetivos del estudio. La solución adoptada es relativamente novedosa.

Primeramente, el diseño de una muestra estratificada es la estrategia más adecuada para garantizar la presencia de aquellos rasgos poblacionales conocidos y correlacionados la información que se quiere obtener en la investigación. En este sentido una primera alternativa era considerar variables criterio de estratificación como la titularidad del centro, su ubicación territorial, su tamaño o la distribución del alumnado por cursos (desde primero a cuarto de ESO), información poblacional que se disponía de cada centro.

¹³ Estudio realizado por el *Institut d'Infància i Món Urbà* (CIIMU) por encargo del *Departament d'Acció Social i Ciutadania de la Generalitat de Catalunya*, bajo la dirección científica Lluís Flaquer, y destinado al análisis de la realidad de la infancia y la adolescencia en Cataluña a partir de una muestra de 3000 individuos. Para ampliar la información se puede consultar el web: <http://www.ciimu.org/webs/panel/marc1.htm>.

Pero se adopta una segunda alternativa de especial interés para el análisis sociológico de la realidad social del estudiantado de secundaria y de sus familias. Consideramos como hipótesis que fundamenta esta alternativa de estratificación que las características sociales del alumnado y de sus familias en un centro determinado coinciden fundamentalmente con las características sociales de la población donde se ubica el centro educativo. Con la información precisa sobre el perfil social de la población referida a unidades homogéneas y reducidas, esta relación es plausible y nos permite considerar estratos poblacionales de diferenciación social en base a los cuales establecer la selección de los diferentes centros con alumnado que también estará socialmente diferenciado según aquellos rasgos poblacionales. Se procede así a identificar la sección censal de cada centro de acuerdo con la estratificación social de Cataluña, y se pondera la importancia de cada centro en función de la importancia de la distribución de los estratos sociales. Finalmente se extrae una muestra aleatoria de centros escolares a partir de la afijación proporcional del tamaño de la muestra entre los estratos sociales a los cuales pertenecen aquéllos y según su peso poblacional.

En segundo lugar, otra ventaja de utilidad de la muestra estratificada va en el sentido de la contextualización social necesaria para los estudios de casos que puedan darse en el interior de una de las Zonas Sociales en varios casos comparativos en función de las Zonas de pertenencia. Nos referimos al estudio de un centro educativo o de varios comparados y del medido de procedencia de los alumnos, al de un mercado, al de una empresa, al recorrido 'social' de un medio de transporte, etc.

Finalmente, la utilidad de la muestra se extiende a la elección de entrevistados o de grupos de discusión según las características sociales de un territorio o comparativamente de diversos sea en vistas a la homogeneidad de los participantes, sea a su heterogeneidad. Así lo empleamos en un estudio sobre la realidad juvenil en Cataluña¹⁴, donde se trataba de identificar comportamientos, representaciones y actitudes de la población catalana entre 15 y 29 años, con un énfasis en las condiciones de vida, especialmente aquellas relacionadas con la dimensión territorial considerando las 41 comarcas catalanas. El estudio se basó en un diseño multimétodo, combinado entrevistas, grupos de discusión y datos secundarios estadísticos¹⁵. En particular, se realizaron 35 grupos de discusión, cuya composición en términos sociales y territoriales se guió y concretó con la ayuda del análisis de los perfiles obtenidos en el proceso de estratificación que venimos comentando.

¹⁴ López, P.; Valls, N.; Verd, J. M.; Vidal, P. (coords.) (2006): *La realitat juvenil a Catalunya* [documento en línea]. Observatori Català de la Joventut, Secretaria General de Joventut de la Generalitat de Catalunya. Disponible en <http://www6.gencat.net/joventut/catala/sgi/observatori/ reavjovcat.htm>

¹⁵ La estrategia metodológica que se siguió ejemplifica la ponencia presentada con el título «La eficiencia teórica y metodológica de los diseños multimétodo» por J. M. Verd y P. López Roldán en el grupo de trabajo de Metodología en el IX Congreso Español de Sociología.

8. CONCLUSIONES

1. Los datos censales constituyen una adecuada base informativa que nos dan una previa y amplia imagen del campo sociológico que corresponde a la encuesta que se realizará sobre la base de la muestra estratificada. Dicho beneficio es más eficaz cuando se trata de una encuesta de amplio espectro en sus contenidos como son las que se refieren a los análisis de estructura social.

2. Con el objetivo de identificar las variables-criterio que forman la base de la estratificación de la ME se ha llevado a cabo una doble operación de factorialización y estratificación de las variables originales previamente escogidas y seleccionadas del censo. Este cálculo y su análisis e interpretación correspondientes constituyen ya un pre-análisis de la estructura social de Cataluña, es decir, de sus factores y grupos sociales. Por ello, el contenido de las variables-criterio y de los estratos, además de prefigurar una estructura social, puede ser un medio de validar los resultados de la encuesta y de dar pistas para el análisis. La propuesta mantenida a lo largo de estas páginas va en esta dirección: la ME es algo más que un mero instrumento muestral-estadístico para ser también, en la modalidad propuesta, guía sustantiva de análisis de la encuesta y un gran potencial de su validación, algo impensable en la manera habitual de realizarse las ME. Evidentemente las condiciones ideales para ello son las presentadas, que las variables-criterio sean conocidas en la población real y que tengan una manifiesta asociación con variables claves del objeto o hipótesis de estudio.

El conocimiento de las variables-criterio, dado por el análisis factorial, nos permite conocer la varianza explicada y acumulada por cada una y por todas las variables-criterio con relación a la varianza total, lo que nos habla de su poder explicativo y de su capacidad para generar estructura social, ya que concitan la máxima heterogeneidad.

Por otro lado, de los estratos se puede conocer su varianza o desviación, es decir, su nivel de homogeneidad versus heterogeneidad interna y entre ellos. Eso nos permite, desde un punto de vista técnico, saber de antemano si la ME que se construye de esta manera puede aumentar su eficacia en precisión (menor error) con relación a la muestra aleatoria simple.

Dado que es posible construir los estratos con la máxima flexibilidad, esta forma de elaborar la ME posibilita no dejar de lado ningún grupo o parte de la población en el análisis de la encuesta. Esto es importante a la hora de distribuir la muestra y sus encuestas si se pretende reflejar en ella la variedad social de la población.

3. Se ha visto también que cada estrato, sus territorios y población correspondiente, pueden ser objeto de análisis sociológicos autónomos utilizando tanto los datos de censo como los de la encuesta, lo que permite tener una información no sólo complementaria y redundante informativamente en vistas a su validación mutua sino una información que conjuga resultados de las dos bases

de datos, del censo y de la encuesta. Precisamente, a los estratos que contienen la información que viene del censo, los denominamos Zonas Sociales cuando acumulan también la información que proviene de la encuesta. Ello es posible porque en cada estrato, las muestras de la encuesta realizadas en ellos, tienen un error satisfactorio debido al procedimiento del cálculo del error, como hemos mostrado, y a la poca varianza de los estratos debido a su gran homogeneidad. Varias consecuencias se derivan de todo ello.

La primera, que se puedan realizar análisis de territorios administrativos y/o geográficos cruzados por las diferencias sociales de las Zonas Sociales que se proyectan sobre ellos y que se refieren a toda Cataluña. Esta posibilidad se da porque la base de referencia en la construcción de los estratos son las secciones censales.

La segunda, que es factible generar una variable nominal con los estratos como valores y emplearla como variable de cruce con las variables de la encuesta.

La tercera, que se puede llevar a cabo una mapificación o cartografía social de Cataluña al digitalizar en un mapa sus secciones censales.

Hemos visto también en el apartado 6 otras utilidades sociológicas que provienen de la manera de construir estas muestras estratificadas.

4. Todo el ejercicio y la argumentación mostrada clarifica, y quizás identifica mejor, la función del sociólogo con un perfil de metodólogo aplicado, o al menos de conocedor de los diseños, métodos, fases, aplicación de procedimientos de recogida de información, de transformación en datos y de análisis en la investigación social. En efecto, no se trata de ser un especialista o un experto en profundidad y extensión como lo son los que a ello se dedican exclusivamente, aunque también pudiera serlo, si bien los tiempos para las especializaciones múltiples son escasos, de las teorías sociológicas o sectoriales o, como en el caso, de la estructura social de una población. Sin embargo, tal y como se ha mostrado, este personaje ha de tener un conocimiento suficiente de referentes sustantivos sociológicos. Lo mismo se puede afirmar con relación la metodología de investigación de corte epistemológico; en este caso, dentro de una perspectiva cuantitativa. Ídem con respecto al conocimiento de la estadística y/o del álgebra y sus métodos, como es el caso: no será un estadístico especializado y experto, aunque, de nuevo, pudiera serlo. Un especializado exclusivamente en estadística

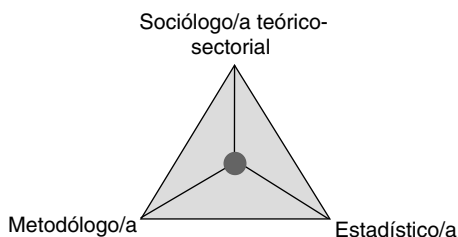


Gráfico 2. *Metodólogo sociológico.*

puede tener los mismos problemas, en otro sentido, con relación al conocimiento necesario para orientar la ME en el sentido y pertinencia sociológica que aquí se ha mostrado.

Por tanto, el perfil que se reivindica es el de un metodólogo sociológico que se encuentra en medio de un triángulo cuyos vértices son tres especialidades: la del teórico general o sectorial en sociología, la del metodólogo, vertiente epistemológica, y la del estadístico-matemático: tres personajes distintos para un solo perfil verdadero. Evidentemente, la necesidad y complicación del personaje se acrecienta cuando se ve abocado a investigaciones multimétodo o multiestratégicas o bajo una perspectiva compleja, o simplemente, de combinación entre perspectivas cuantitativas y cualitativas. La panoplia del 'semiexperto' global se extiende, además de a las citadas, a la lingüística, la antropología, la psicología, la comunicación, etc., por no ir más lejos e introducir además la simulación social y otras matemáticas de nuevo cuño, al menos dentro del campo de la sociología, como los sistemas dinámicos, la teoría de la complejidad, las redes sociales, etc.

BIBLIOGRAFÍA

- AZORIN, F.; SÁNCHEZ CRESPO, J. L. (1986): *Métodos y aplicaciones del muestreo*. Madrid: Alianza.
- BABBIE, E. (1990): *Survey Research Methods*. Belmont, CA: Wadsworth Publishing Company.
- BAILEY, K. (1987): «Survey Sampling». En: *Methods of Social Research*. Nueva York: Free Press. P. 79-103.
- BETHEL, J. (1989): «Sample allocation in multivariate surveys». *Survey Methodology*, 15: 47-57.
- BETHEL, J. (1989): «Minimum variance estimation in stratified sampling». *Journal of the American Statistical Association* 84: 260-265.
- COCHRAN, W. G. (1978): *Técnicas de muestreo*. México: CECSA.
- DOMÍNGUEZ, M.; LÓPEZ, P.; LOZARES, C. (2000): *Informe sobre la construcción de la muestra de la Enquesta de la Regió de Barcelona 2000*. Barcelona: Institut d'Estudis Regionals i Metropolitans de Barcelona.
- GROSBAS, J. M. (1987): *Méthodes statistiques des sondages*. Paris: Economica.
- IBÁÑEZ, Jesús (1985): «Las medidas de la sociedad». *Revista Española de Investigaciones Sociológicas* 29: 85-127.
- KHAN, M. G. M.; NAJMUSSEHAR and AHSAN, M. J. (2005): «Optimum Stratification for Exponential Study Variable under Neyman Allocation». *Journal of the Indian Society of Agricultural Statistics* 59 (2): 146-150.
- KISH, L. (1972): *Muestreo de encuesta*. México: Trillas.
- LÓPEZ ROLDÁN, P. (1994): *La construcción de tipologías en Sociología: propuesta metodológica de construcción, análisis y validación. Aplicación al estudio de la segmentación del mercado de trabajo en la Regió Metropolitana de Barcelona*. Tesis Doctoral. Bellaterra: Universitat Autònoma de Barcelona.
- LÓPEZ ROLDÁN, P. (1996): «La construcción de tipologías: metodología de análisis». *Papers. Revista de Sociologia* 48: 9-29.

- LÓPEZ ROLDÁN, P.; VALLS, N.; VERD, J. M.; VIDAL, P. (coords.) (2006): *La realitat juvenil a Catalunya* [documento en línea]. Observatori Català de la Joventut, Secretaria General de Joventut de la Generalitat de Catalunya. Disponible en <http://www6.gencat.net/joventut/catala/sgi/observatori/reaajovcat.htm>
- LÓPEZ ROLDÁN, P.; LOZARES COLINA, C. (2007): «La construcción de la muestra de la “Enquesta de Condicions de Vida i Hábits de la Població de Catalunya 2006”». *Recerques i Metodologies*, 1. Barcelona: Institut d’Estudis Regionals i Metropolitans de Barcelona (en prensa).
- LÓPEZ, P., DOMÍNGUEZ, M., LOZARES, C. (2000): «Disseny i construcció d’una mostra estratificada a partir de dades censals». *Qüestió* 24, 1: 111-136.
- LOZARES, C. (1990): «La tipología en Sociología: más allá de la taxonomía». *Papers. Revista de Sociologia* 34: 139-164.
- LOZARES, C.; DOMÍNGUEZ, M. (1993): *Enquesta de la Regió Metropolitana de Barcelona 1990. Territori i realitat social: les zones socio-demogràfiques de la Regió Metropolitana de Barcelona*. Barcelona: Mancomunitat de Municipis de l’Area Metropolitana de Barcelona y Diputació de Barcelona.
- LOZARES, C.; DOMÍNGUEZ, M. (1995): *Diseño de la muestra y supervisión metodológica de la Enquesta de Salut de Barcelona 1992-93*. Barcelona: Ajuntament de Barcelona.
- LOZARES, C.; DOMÍNGUEZ, M. (1996): «Tratamiento multivariado de subpoblaciones en una gran encuesta social: la construcción de zonas sociales». *Papers. Revista de Sociologia* 48: 71-87.
- LOZARES, C.; LÓPEZ, P. (1990): *Enquesta Metropolitana de la Regió Metropolitana de Barcelona. Construcció de la mostra estratificada*. Sèrie Documents de Treball 90/1. Bellaterra: Institut d’Estudis Metropolitans de Barcelona.
- LOZARES, C.; LÓPEZ, P. (1991a): «El análisis de componentes principales. Aplicación al análisis de datos secundarios». *Papers. Revista de Sociologia* 37. P. 31-63.
- LOZARES, C.; LÓPEZ, P. (1991b): «El muestreo estratificado por análisis multivariado». En: *El pluralismo metodológico en la investigación social: ensayos típicos*, editado por M. Latiesa. Granada: Universidad de Granada. P. 107-160.
- LOZARES, C.; LÓPEZ, P. (2000): *Anàlisi multivariable de dades estadístiques*. Bellaterra (Barcelona): Universitat Autònoma de Barcelona. Colección Materials 93.
- LOZARES, C.; LÓPEZ, P.; DOMÍNGUEZ, M.; BORRAS, V. (1994): *Tratamiento multivariado un una gran encuesta aleatoria: La muestra estratificada, la Estructura Social y el análisis de Zonas sociales. Metodología y complementariedad mutuas* Bellaterra: Institut d’Estudis Metropolitans de Barcelona.
- LOZARES, C.; LÓPEZ, P.; FLORES, J. L. (2003): *Zones socials per al territori de la província de Barcelona: una cartografia social* Barcelona: Institut d’Estudis Regionals i Metropolitans.
- MILLER, T. J.; SKALSKI, J. R.; IANELLI, J. N. (2007): «Optimizing a stratified sampling design when faced with multiple objectives». *Journal of Marine Science* 64: 97-109.
- MONROE, G. S. (1972): «Stratified Sample Surveys with Multiplicity». *Journal of the American Statistical Association* 67, 337: 224-227.
- MULVEY, J. M. (1983): «Multivariate Stratified Sampling by Optimization». *Management Science* 29, 6: 715-724.
- NEYMAN, J. (1934): «On two different aspects of the representative method: the method of stratified sampling and the method of purposive selection». *Journal of the Royal Statistical Society* 97: 558-606.

- PODGURSKI, A.; YANG, C. (1993): *Partition testing, stratified sampling, and cluster analysis*. New York: ACM Press.
- PULIDO SAN ROMÁN, A. (1978): «Selección, método y determinación de la muestra». En *Estadística y técnicas de investigación social*. Madrid: Pirámide. P. 161-186.
- RODRÍGUEZ OSUNA, J. (1991): *Métodos de muestreo*. Madrid: CIS.
- SÁNCHEZ CRESPO, J. C. (1971): *Principios elementales del muestreo estimación de proporciones*. Madrid: INE.

RESUMEN

El artículo, tras exponer concisamente las principales características del diseño y del proceso de construcción de la muestra estratificada de la «Enquesta de Condicions de Vida i Hàbits de la Població de Catalunya», en particular del año 2006, se centra y tiene como objetivo específico la presentación de algunos resultados que, proviniendo directa o indirectamente de la construcción de la muestra, tienen un gran interés por su implicación sociológica. Por el procedimiento seguido en el diseño, por los criterios utilizados, por las decisiones tomadas y por su (re)utilización a posteriori, el tipo muestra que se presenta proporciona unos resultados, tanto metodológicos como sociológicos, para el conocimiento del territorio que van más allá de los esperados de naturaleza puramente instrumental. Nos referimos a la identificación de los estratos que permiten un análisis específico de la estructura social. A la cartografía social que se deriva de dichos estratos al considerar la sección censal. A la información adicional que se obtiene por la aplicación de los datos de las encuestas a los estratos en términos de lo que denominamos como «zonas sociales». A la reutilización para la construcción de otras muestras, de otros análisis sociales, para validar los propios resultados de la encuesta y otros aspectos técnicos.

PALABRAS CLAVE

Muestreo estratificado, encuesta, cartografía social, análisis de datos censales, análisis multivariable, tipología.

ABSTRACT

The article firstly presents concisely the main characteristics of the design and the process of construction of the stratified sample of the «Enquesta de Condicions de Vida i Hàbits de la Població de Catalunya», particularly for the year 2006. Then, the article is centered, and has as specific objective, the presentation of some results that —coming directly or indirectly from the construction of the sample— have a great interest by their sociological implication. The type of sample that is used provides results of high interest for the understanding of the territory, both in the methodological and sociological aspects. This is achieved by means of the procedure followed in the design, the used criteria, the taken decisions and their subsequent use. These results are, namely, the identification of the stratum —that allows a specific analysis of the social structure—, the production of a social cartography that derives from these stratum when considering the census section, the identification of «social zones» using the additional information that it is obtained by the application of the data of the

surveys to the stratum, and the reusability for the construction of other samples or other social analyses, or to validate our own results of the survey and other technical aspects.

KEY WORD

Stratified sample, survey, social cartography, census data analysis, multivariate analysis, typology.