

# 02 LA CONSTRUCCIÓN DE LA MUESTRA

Pedro López y Carlos Lozares (Centro de estudios sociológicos sobre la vida cotidiana y el trabajo, Departamento de Sociología, Universidad Autónoma de Barcelona)

## Introducción

### 1. Construcción de los estratos

- 1.1. El análisis factorial y la identificación de las variables—criterio de los estratos
- 1.2. El análisis de clasificación o la construcción de los estratos de la muestra estratificada

### 2. Afijación de la muestra y cuotas muestrales

- 2.1. Tamaño de la muestra y afijación
- 2.2. Cuotas muestrales
- 2.3. Distribución final de la muestra

### 3. Ponderación a posteriori de la muestra y elevación

### 4. Aplicación de los resultados de la muestra

## Introducción

Este artículo tiene por objeto exponer las principales características del diseño y del proceso de construcción de la muestra estratificada de la *Encuesta de condiciones de vida y hábitos de la población de Cataluña* en su edición de 2006, haciendo hincapié en las novedades que se han incorporado. La metodología seguida en el diseño de la muestra a lo largo de las cinco ediciones de la *Encuesta* conserva un mismo planteamiento básico, que se ha mantenido en el tiempo, si bien en cada nueva edición se han ido introduciendo algunos cambios como resultado de las exigencias de los análisis territoriales que se han ido planteando en cada momento<sup>1</sup>. La presente edición es la que incorpora mayores novedades, sobre todo por la ampliación del ámbito territorial al conjunto de Cataluña y por la introducción de cambios en el procedimiento de determinación de las cuotas de muestra.

1 Una exposición detallada del diseño y del proceso de construcción de la muestra, para la edición de la Encuesta de 1990, se puede consultar en Lozares y López (1990). Una versión reducida de este documento fue publicada en Lozares y López (1991b). Posteriormente, tomando como base esta misma muestra, se consideró el tratamiento de subpoblaciones con la finalidad de construir zonas sociales (Lozares, y Domínguez, 1996). El proceso de construcción de la edición de 1995 se presentó en forma de artículo de revista (López, Domínguez, y Lozares, 2000). Finalmente, con respecto a la edición del año 2000, se puede consultar el apartado metodológico de los textos publicados en dicha edición de la encuesta (Giner *et al.*, 2002), así como el análisis de las zonas sociales y la cartografía social que se obtiene como producto de la propia construcción de la muestra (Lozares, López, y Flores, 2003).

En primer lugar se analizan los rasgos generales de la muestra, para entrar, en los apartados siguientes, en las cuestiones específicas relativas a la construcción de los estratos, la afijación de la muestra, la asignación de cuotas y la ponderación y elevación de la muestra, para acabar con algunas consideraciones que van más allá de la simple aplicación de la muestra.

En el diseño de la muestra se establece como objetivo general la extracción de una muestra aleatoria estratificada representativa de la población de Cataluña. 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.

Por otra parte, los contenidos del estudio de la encuesta recogen una gran diversidad de temáticas sociales de interés. El objetivo consiste precisamente en poder expresar mejor la variedad de la realidad social de Cataluña respecto a 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. Esta información poblacional primaria y diversa proviene del censo de población, a través de sus indicadores. Una vez elegidas las variables más pertinentes del Censo que representan la máxima diversidad social, 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 *Encuesta*.

Así pues, el punto de partida es la información que proviene del censo de población del año 2001. Para proceder a la construcción de los estratos se agrupan los individuos en función de las secciones censales donde residen. De esta forma se dispone, para el conjunto de indicadores o variables seleccionados, del número de personas que, en cada sección, poseen la característica dada por el indicador y, en consecuencia, se dispone también del porcentaje que representa dicha variable sobre el total de la población de la sección censal. A partir de esta información se identifican, primero, las variables-criterio y, en segundo lugar, se construyen los estratos mediante métodos de análisis factorial de componentes principales y de análisis de clasificación automática.

Con los estratos construidos se procede a la afijación de un tamaño de muestra predeterminado a cada estrato según el criterio óptimo de Neyman. Realizada dicha atribución en los estratos, se asignan las cuotas de muestra a cada una de las secciones censales de cada estrato según su población. Se lleva a cabo, *a posteriori*, una subestratificación según el territorio y el hábitat, con el objetivo de conseguir una concentración territorial que facilite el trabajo de campo de la encuesta. Finalmente, las unidades muestrales se extraen de forma aleatoria en cada sección censal a partir del marco de la muestra que, en este caso, ha sido el conjunto de personas registradas en el Padrón de Habitantes actualizado a enero de 2005.

En comparación con ediciones anteriores de la *Encuesta* (Tabla 1), la construcción de la muestra presenta algunos cambios que se resumen en los siguientes aspectos:

- a) El ámbito territorial se extiende al conjunto de Cataluña. El ámbito poblacional contemplado se amplía también para considerar el de la población catalana de 16 y más años. Así, el universo poblacional se define como el conjunto de personas de 16 y más años que residen en Cataluña: un total de 5.408.264 personas según el censo de población de 2001, siendo la población total catalana registrada en aquel momento de 6.343.110<sup>2</sup>.

2 Hay que tener presente la infravaloración de este recuento, ya que se han producido importantes movimientos inmigratorios en los últimos años que motivan que, en la actualidad, la población catalana total supere los siete millones de personas. En el momento de extraer la muestra se considera el padrón de habitantes actualizado. *A posteriori*, con la ponderación de la muestra, se tienen en cuenta estas variaciones y distribuciones diferenciadas de la población de las secciones censales en el territorio.

- b) El proceso de construcción y de análisis se realiza a partir de la información contenida en el censo de población del año 2001, modificando y ampliando el conjunto de indicadores-variables utilizados para la muestra.
- c) Se ha procedido a una subestratificación de la muestra en función del territorio, con una selección aleatoria de secciones censales y una redistribución de las cuotas finales de muestra para satisfacer el requisito de concentración territorial del trabajo de campo.
- d) El marco muestral de extracción deja de ser el censo electoral para considerar la actualización del padrón de habitantes.
- e) El tamaño de muestra se amplía para facilitar los análisis independientes de dominios de estudio o subpoblaciones específicas, territoriales en particular. Para garantizar la representatividad de algunos territorios, además, se ha añadido un número de encuestas adicional que supera el tamaño de muestra fijado para datos globales.
- f) La ponderación de la muestra se ha ajustado a los nuevos requisitos del diseño y se ha aplicado por primera vez el factor de elevación para expresar los datos muestrales en términos de valores absolutos poblacionales.

Tabla 1. Comparación de las muestras de la ECVHP

Ámbito territorial	Edición de la ECVHP				
	1985	1990	1995	2000	2006
Área Metropolitana de Barcelona	Área Metropolitana de Barcelona	Región I	Región Metropolitana	Provincia de Barcelona	Cataluña
<b>Municipios</b>	27	129	162	311	946
<b>Superficie</b>	476 km <sup>2</sup>	2.459 km <sup>2</sup>	3.235 km <sup>2</sup>	7.718 km <sup>2</sup>	32.106 km <sup>2</sup>
<b>Población</b>	3.096.748	4.090.406	4.264.422	4.628.277	6.343.110
<b>% Cataluña</b>	49,2%	67,9%	70,4%	76,0%	100%
<b>Universo estadístico</b>	> 18 años	> 18 años	> 18 años	> 18 años	> 16 años
	2.174.363	2.991.131	3.275.458	3.762.462	6.049.414
<b>Muestra</b>	4.912	5.061	5.263	6.830	10.398
<b>Error relativo (n.s. 2σ)</b>	± 1,9	± 1,9	± 1,8	± 1,5	± 0,7
<b>Marco de la muestra</b>	Censo electoral 1984	Censo electoral 1989	Censo electoral 1994	Censo electoral 1999	Padrón 2005

## 1. Construcción de los estratos

El censo de población del año 2001 cumple la condición indispensable de suministrar los datos poblacionales para la construcción de la muestra estratificada. De la información que aparece en el cuestionario del censo se han seleccionado una serie de indicadores (o variables), que son los que serán utilizados para la construcción de las variables-criterio empleadas en la estratificación. Dichas variables-criterio habrán de recoger la máxima variedad o heterogeneidad social de la población de Cataluña. Dado que se han tomado las secciones censales como soporte de dichas variables, serán las secciones censales las unidades elementales de estratificación y no los individuos censados.

Esta manera de proceder se justifica por un triple motivo: por las dificultades que se derivan del tratamiento de una matriz de datos de individuos de esta magnitud; porque la agregación en secciones censales nos permite un tratamiento métrico en base a datos porcentuales derivados de

las variables seleccionadas, y porque la sección censal constituye una unidad de división territorial suficientemente pequeña para garantizar la homogeneidad de las características sociales de la población.

La información censal que se ha empleado en esta edición de la *Encuesta* se ha ampliado considerablemente respecto a las ediciones anteriores, puesto que un mayor número de indicadores sociales proporciona una mayor riqueza de los estratos y de la heterogeneidad poblacional, facilitando la interpretación, sobre todo teniendo en cuenta que se ha ampliado la población objeto de estudio al conjunto de Cataluña. La relación de indicadores-variables es inicialmente muy amplia, aunque un análisis posterior de los mismos nos lleva a realizar una selección final de los que, expresando la mayor heterogeneidad social posible de Cataluña, no proporcionen información redundante.

Esta selección final de las variables-criterio obedece, en primer término, a su disponibilidad en el Censo<sup>3</sup>; responde también a criterios de pertinencia conceptual de acuerdo con los objetivos del estudio, y, finalmente, a criterios de tipo estadístico empleados en el proceso de análisis, como son la manifiesta combinación lineal entre algunas variables o el escaso valor o dispersión de otras.

En consecuencia, se dispone de una matriz de datos inicial de 234 variables-indicadores con información sobre las 5.222 secciones censales de Cataluña en términos del número de personas (o también de hogares) que poseen una determinada característica (o contenido del indicador) en el año 2001. Estas variables se expresan en términos relativos, calculando el porcentaje de la población de la sección censal que posee una determinada característica sobre el total de la población de la sección censal correspondiente.

---

Tabla 2. Variables-criterio utilizadas en la muestra estratificada

1. % Sección/Municipio	20. % personas no entienden catalán
2. % personas de 0–18 años	21. % personas entienden catalán
3. % personas de 20–29 años	22. % personas entienden y leen catalán
4. % personas de 30–39 años	23. % personas hablan, leen y escriben catalán
5. % personas de 40–49 años	24. % personas solteras
6. % personas de +65 años	25. % personas casadas
7. % personas nacidas en Cataluña	26. % personas separadas
8. % personas nacidas en España	27. % personas divorciadas
9. % personas nacidas en la UE	28. % personas de +10 años analfabetas
10. % personas nacidas en América	29. % personas de 1–3 años escolarizadas
11. % personas nacidas en África	30. % personas de 17–24 años escolarizadas
12. % personas nacionalidad española	31. % personas de +24 años universitarias
13. % personas residencia España –2 años	32. % personas con EGB
14. % personas residencia España 2–5 años	33. % personas con BUP
15. % personas residencia España 6–10 años	34. % personas con FP I
16. % personas residencia España 11–20 años	35. % personas con FP II
17. % personas residencia municipio –2 años	36. % personas estudiantes
18. % personas residencia municipio 2–5 años	37. % personas ocupadas hombres
19. % personas residencia municipio +10 años	38. % personas ocupadas mujeres

3 A partir del asesoramiento del Instituto de Estadística de Cataluña (Idescat) sobre la disponibilidad y fiabilidad de los datos del censo del 2001 descartamos una parte de las variables individuales o de hogar.

39. % personas activas hombres	61 % personas operario cualificado
40. % personas activas mujeres	62. % personas operario no cualificado
41. Tasa de paro	63. % personas trabajo en su domicilio
42. % personas con invalidez	64. % personas trabajo en su municipio
43. % personas viudas	65. % personas trabajo en otro municipio
44. % personas jubiladas	66. % personas transporte privado al trabajo
45. % personas trabajo doméstico	67. % personas transporte público al trabajo
46. % personas Agricultura	68. % personas con hipotecas
47. % personas Industria	69. % personas vivienda heredada
48. % personas Servicios	70. % personas vivienda de alquiler
49. % personas jornada <35 h	71. % personas vivienda hasta 50 m <sup>2</sup>
50. % personas jornada 35-39 h	72. % personas vivienda 51-70 m <sup>2</sup>
51. % personas jornada 40-50 h	73. % personas vivienda +100 m <sup>2</sup>
52. % personas empresarias con asalariados	74. % personas con segunda residencia
53. % personas empresarias sin asalariados	75. % personas con vehículos a motor
54. % personas asalariadas fijas	76. % hogares unipersonales
55. % personas asalariadas eventuales	77. % hogares madre con hijos
56. % personas profesionales liberales	78. % hogares pareja con hijos
57. % personas director	79. % hogares 1 núcleo con parientes
58. % personas técnico	80. % hogares más de 1 núcleo sin parientes
59. % personas administración	81. % hogares más de 1 núcleo con parientes
60. % personas contramaestre	82. % hogares sin núcleo

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, y como resultado del análisis de sus correlaciones mutuas para examinar su redundancia y significatividad, se eligen 82 indicadores o variables, que son los que aparecen en la Tabla 2.

A partir de esta selección se plantea como paso siguiente la construcción de estratos de población que sean internamente lo más homogéneos posible. Se trata de garantizar que los estratos de la muestra representen, cuanto más fidedignamente mejor, la variedad de la población. El proceso de construcción de los estratos se realiza mediante dos técnicas de análisis multivariable independientes y complementarias: el análisis factorial de componentes principales, para reducir y expresar los principales factores de diferenciación de la población y que, a su vez, se corresponden con las variables-criterio usadas para la construcción de los estratos, y el análisis de clasificación automática, para obtener grupos de población (de secciones censales) o tipos sociales que constituirán los estratos de la muestra.

## 1.1. El análisis factorial y la identificación de las variables-criterio de los estratos

Con el análisis factorial de componentes principales (ACP) se pretende reducir la información original para obtener un subespacio vectorial de menos dimensiones o factores, donde éstos sean base y, por tanto, linealmente independientes, y que, ordenados de manera jerárquica, conserven una buena parte estipulada de la varianza total. Se obtienen así las dimensiones fundamentales y reducidas de diferenciación social de la población de Cataluña. El ACP se concibe como una etapa previa y un complemento necesario para la obtención de los estratos.

Tabla 3. Varianza explicada por los componentes retenidos

Factor	Valores propios sin rotación			Valores propios de la rotación		
	Valor	% de la varianza	% acumulado	Valor	% de la varianza	% acumulado
1	15,418	18,8	18,8	13,149	<b>16,0</b>	16,0
2	14,087	17,2	36,0	9,038	<b>11,0</b>	27,1
3	12,809	15,6	51,6	8,709	<b>10,6</b>	37,8
4	6,075	7,4	59,0	7,321	<b>8,9</b>	46,6
5	3,400	4,1	63,2	7,073	<b>8,6</b>	55,2
6	2,373	2,9	66,0	5,767	<b>7,0</b>	62,3
7	2,260	2,8	68,8	5,364	<b>6,5</b>	68,8
					<b>68,8</b>	<b>100,0</b>

El ACP realizado<sup>4</sup> parte, pues, de las 82 variables de la Tabla 2. La estructura de la información permite reducir estas 82 variables a sólo 7 dimensiones, ejes o variables factoriales, que acumulan el 68,8% de la varianza total<sup>5</sup>. En la Tabla 3 se presentan los valores propios retenidos y la importancia relativa de cada componente antes y después de la rotación varimax aplicada.

A continuación se describe la identidad de los ejes factoriales que se han extraído y que estructuran y resumen la información original utilizada.

#### Primer eje: la categoría socio-profesional

Este primer eje o componente acumula, en el conjunto de las variables, el 16,0% de la varianza total, una vez efectuada la rotación, y el 23,3% de la varianza sobre el total de los ejes retenidos. Es, pues, un eje decisivo y determinante de la estructura de relaciones de las variables originales. Por lo que se refiere a su descripción, constituye el eje clásico de estructuración social que se identifica por variables vinculadas a las diferentes categorías y sectores profesionales y a los niveles educativos.

En este sentido, aparece una polaridad que opone, por un lado, la presencia de categorías profesionales y ocupaciones como directivos, profesionales y técnicos altos del sector terciario en la administración y en servicios, así como niveles educativos medios o altos (BUP, FP II, universitarios de más de 24 años y escolarización entre los 17 y 24 años) a, por otro lado, la presencia de profesiones bajas, cualificadas y no cualificadas, del sector industrial en particular, y un bajo nivel formativo (EGB o analfabetos).

Tiene un interés singular que en este eje aparezcan otras características, que acompañan a las descritas, que se extienden a lo que se puede denominar como formas de vida; de hecho, tan sólo se muestran en el primer polo enunciado, el de las categorías altas, suponiendo entonces que su negación o ausencia está vinculada al otro polo. Así, por ejemplo, los tipos de horarios de trabajo en los sectores referidos son de entre 35 y 39 horas semanales, pero también de menos de 35; se da una mayor presencia de mujeres ocupadas y activas, de situaciones de separación y divorcio y de posesión de una segunda residencia.

4 Todos los análisis se han obtenido a través del programa estadístico SPSS.

5 El examen de la matriz de correlaciones y de otras medidas nos confirma la adecuación y la validez del análisis: examen de la matriz de correlaciones, comportamiento de la matriz antiimagen de correlaciones y un índice de Kaiser-Meyer-Olkin de 0,872.

Por tanto, es importante señalar que, si bien se trata de un factor vinculado a la posición de clase social, expresada en términos de categorías profesionales y niveles educativos, está marcado también por el tipo de profesión, la presencia activa de la mujer en el mercado de trabajo y otras características que introducen componentes vinculados a las formas de vida de la estratificación social.

### **Segundo eje: el origen geográfico, autóctonos *versus* antigua inmigración**

Este segundo eje representa el 11,0 % de la varianza original, lo que corresponde al 16,0% de la varianza conservada. Es un eje con dos vertientes que, teniendo en cuenta que se superponen sobre el mismo eje, guardan una intercorrelación que define la misma realidad de diferenciación de la población.

La primera vertiente, que es la que presenta una mayor proyección sobre este segundo eje, separa, en uno de sus polos, la población nacida en Cataluña y que manifiesta el mayor grado de conocimiento de la lengua catalana de, en el otro polo, los nacidos fuera de Cataluña que no dominan el catalán en sus diferentes niveles.

La segunda vertiente tiene un contenido más socio-profesional y socioeconómico, ligado, en uno de los extremos, a la condición de ser empresario con o sin asalariados, más la de disponer de una vivienda espaciosa de más de 100 m<sup>2</sup>. En el otro extremo, de forma opuesta, aparece la condición de ser asalariado con contrato eventual y encontrarse en situación de desempleo, con un mayor predominio de la población que se dedica a las tareas del hogar y residencia en viviendas de menor tamaño, entre 51 y 70 m<sup>2</sup> o más pequeñas.

Es preciso hacer notar igualmente la presencia de otras características que se arropan a lo largo del eje, como son el indicador del pequeño tamaño del municipio, el uso del transporte privado o la contraposición de edades de la población. Por tanto, este eje es revelador de una diferencia, que se traduce en características sociales, entre una población de origen catalán en los ámbitos más rurales y la población de origen inmigrante de primera generación llegados a Cataluña del resto de España, que se localiza en grandes núcleos urbanos.

### **Tercer eje: el ciclo vital**

Con un 10,6 % de la varianza total y el 15,4% de la varianza retenida, este tercer factor se identifica con el ciclo vital, vinculado fundamentalmente a la edad, pero también a algunas otras características concomitantes que identifican su contexto y condiciones.

Se trata, en uno de los polos del eje, de una población de menos de 49 años pero con mayor incidencia de la más joven, formada, además, por parejas con hijos y también por personas solteras, con un peso destacado de la población estudiantil. Otras dos características connotan, de alguna manera, este primer polo: la hipoteca y el uso de vehículos a motor. También merece destacarse, como rasgo distintivo, que sea en las secciones censales que corresponden a este polo donde se dé una mayor proporción de personas no autóctonas, con residencia en España desde hace un cierto tiempo.

Si las características precedentes definen uno de los polos del eje, el otro lado aparece claramente marcado por una población envejecida y por sus consecuencias con respecto al ciclo vital: mayores de 65 años, residentes en el municipio desde hace mucho tiempo, que están jubilados, viven solos en hogares sin núcleo familiar y/o en situaciones de viudedad, con un peso importante de viviendas de alquiler. Todas ellas características que, si bien denotan un perfil asociado a la edad, comportan otros indicadores familiares y formas de hogar y de vivienda.

### **Cuarto eje: el factor rural-urbano**

La parte de la varianza explicada por cada uno del resto de ejes, del cuarto al séptimo, experimenta un salto descendente muy considerable. En concreto, este cuarto eje representa el 8,9% en relación al total, y un 13,0% de la varianza retenida por los siete primeros ejes elegidos.

En este eje, sólo uno de los polos aparece claramente identificado y con unas características muy coherentes; consecuentemente, en el otro polo se sitúan los rasgos opuestos a los que describimos como definitorios del primer polo. Se trata de una población que reside en municipios pequeños, sobre todo dedicada al sector agropecuario, con una empresa familiar, con la condición de empresarios sin asalariados, y que trabajan en el mismo domicilio. Se relacionan con estos rasgos de la población otros que tienen que ver con el tipo de hogar y de vivienda. Así, se da siempre la presencia de un núcleo familiar, siendo también numerosos los hogares extensos, y con un predominio de viviendas grandes que se heredan. Se caracteriza además por una mayor proporción de personas nacidas en Cataluña.

En el otro extremo del eje se acumulan el resto de características, las de una población que vive en secciones de núcleos urbanos, de diferentes orígenes, del resto de sectores productivos, con otras formas de composición del hogar y otras viviendas.

#### **Quinto eje: la nueva inmigración**

Este eje acumula, en términos de varianza explicada, un 8,6% de la varianza total, lo que supone un 12,5% de la dispersión total atribuida a los siete ejes elegidos. Su identidad está perfectamente definida, especialmente en uno de los polos y por contraposición al otro, en el que no aparece ningún indicador posicionado de forma explícita; es decir, que dicho polo está definido por la negación de (o bien la oposición a) las características que señalamos a continuación para el primer polo.

Dicho primer polo representa la nueva inmigración, si bien se trata de personas que ya llevan un cierto tiempo en el país. Estamos hablando de secciones con un peso importante de personas que tienen la residencia en España desde hace al menos dos años, si bien no sobrepasan los diez años en el país (datos censales de 2001), y que no tienen la nacionalidad española. Su origen es sobre todo africano o latinoamericano, viven de alquiler y en viviendas pequeñas, están solteras y viven en hogares sin núcleo. Este perfil de recién llegado se contrapone a la población autóctona, enraizada y que vive en hogares con núcleos familiares.

#### **Sexto eje: el factor de actividad laboral**

Es un eje que representa el 7,0% de la varianza explicada y un 10,2% de la conservada en los siete ejes escogidos.

Los extremos de este factor contraponen la ocupación a la inactividad laboral y al paro. Un polo está claramente definido por la población activa y ocupada, tanto de hombres como de mujeres, asalariados fijos (y en menor medida eventuales), con horarios semanales exhaustivos de 40 a 50 horas, que trabaja sobre todo en la industria y en categorías laborales de la población obrera tradicional cualificada. En el polo contrapuesto se puede encontrar una mezcla de población inactiva, compuesta sobre todo de personas jubiladas, junto a perfiles de secciones con una mayor tasa de desempleo y de población estudiantil.

#### **Séptimo eje: la movilidad territorial**

Este eje representa un peso similar al precedente en la totalidad de la variabilidad, un 6,5%, y un 9,5% en relación con la varianza que representan los siete factores.

A un lado, encontramos a la población de las secciones censales con un importante peso relativo de residentes recientes en el municipio (menos de 5 años). Se trata de municipios de tamaño medio, con población joven o adulta llegada en edad reproductiva, con hipotecas y que trabaja en otro municipio, al cual se desplaza con transporte privado. Este perfil contrasta con el otro polo, donde se localiza sobre todo una población con más de 10 años de residencia en el municipio y que trabaja en él. Vemos, pues, la configuración de un factor ligado a la movilidad territorial de una población que en los últimos años se ha desplazado a ciudades intermedias y a urbanizaciones periféricas, ya sea desde la gran ciudad, ya sea desde otros núcleos.

## 1.2. El análisis de clasificación o la construcción de los estratos de la muestra estratificada

La estratificación de la muestra descansa sobre el principio según el cual la población en estudio es estadísticamente heterogénea. La homogeneidad estadística se refiere, en el caso de las variables o contenidos sociales, al grado de correlación entre las características elegidas para identificar una población; cuanto más correlacionadas estén dichas características más homogénea será la población considerada. Esta condición de máxima heterogeneidad de la población en cuanto a las variables factoriales elegidas, que pasan a ser las variables-criterio en la construcción de los estratos, se ha llenado satisfactoriamente con el análisis factorial realizado al obtener unos ejes incorrelacionados y con una buena parte de la varianza de la población incorporada.

El objetivo ahora es otro, aunque consecuente y complementario. Se trata de la estratificación, que consiste en clasificar las secciones censales en estratos compuestos por conjuntos de secciones lo más homogéneas posible dentro de cada estrato y lo más heterogéneas posible entre las secciones de los diferentes estratos, todo ello en función de las variables factoriales extraídas del ACP. Con respecto a la muestra aleatoria simple, este procedimiento asegura además, a efectos del muestreo, una mayor ganancia en precisión o disminución de error. Así pues, tomando como variables-criterio los siete ejes factoriales precedentes, o subespacio vectorial correspondiente reducido de siete dimensiones, se lleva a cabo la clasificación de las secciones censales. Este es el punto de partida y en esto consiste el análisis de clasificación automática.

Esta clasificación se realiza sin tener en cuenta ninguna restricción de contigüidad territorial, por lo que se podrá obtener como resultado un mapa de las secciones censales de Cataluña según los estratos de pertenencia y con su diferente caracterización social; es decir, en función de la pertenencia a los estratos observaremos una distribución diferenciada del territorio.

Partimos pues de una matriz reducida con las 7 variables-factoriales que identifican las 5.222 secciones censales de Cataluña. El método de clasificación empleado, tomando como medida de proximidad la distancia cuadrática euclíadiana en el espacio vectorial de la base de los siete vectores, se desglosa en dos etapas. En ambas, se ha procedido básicamente a:

- a) En un primer momento se realiza una clasificación jerárquica ascendente, mediante el procedimiento de *Ward* (de mínima pérdida de inercia), para clasificar inicialmente las secciones censales en el número más adecuado de grupos a partir de criterios estadísticos<sup>6</sup> y del estudio conceptual-descriptivo de los estratos resultantes interpretados a partir de las variables-factoriales y de las mismas variables originales. En esta etapa se efectúa el análisis sistemático de las particiones comprendidas entre 20 y 2 grupos. Finalmente, el número de grupos, estratos a efectos muestrales, se fija en 10.
- b) En un segundo momento, con el número de grupos determinado y los centros iniciales definidos, 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. Esta reasignación óptima configura la estratificación final de población con la distribución de secciones y centros finales en los factores que se puede apreciar en la Tabla 4 y en el Gráfico 1.

Presentamos a continuación la caracterización sociológica de los diez estratos, de forma breve y resumida, descrita a partir de las dimensiones de identidad social y aludiendo también a las variables originales. Señalamos los perfiles más destacados con relación al comportamiento medio del conjunto de la población catalana.

6 Cambios en las medidas de distancias por la aglomeración de clases y análisis del cambio en la varianza explicada por cada nueva partición.

Tabla 4. Distribución de las secciones censales por estrato y centros

Estrato	Secciones censales	Centros finales						
		Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
		Categoría socioprofesional	Autóctonos	Ciclo vital	Rural-Urbano	Nueva inmigración	Actividad laboral	Movilidad territorial
1	576 11,0%	-0,780	1,379	-0,233	0,523	-0,148	0,322	-0,302
2	397 7,6%	0,366	0,153	0,387	0,053	-0,406	0,331	2,368
3	723 13,8%	-0,322	0,246	0,759	0,459	-0,282	0,523	0,097
4	504 9,7%	-0,451	0,773	-0,563	-2,290	-0,382	-0,092	-0,053
5	424 8,1%	-1,190	-0,321	0,997	0,134	0,888	-0,667	-0,196
6	425 8,1%	1,679	0,321	1,020	-0,075	-0,103	-0,914	-0,447
7	769 14,7%	-0,108	-1,424	-0,006	-0,107	-0,438	0,636	-0,348
8	407 7,8%	-0,445	-0,716	-1,064	0,628	-0,581	-1,288	0,119
9	771 14,8%	1,091	0,088	-0,701	0,431	0,329	0,027	-0,508
10	226 4,3%	-0,095	-0,459	-0,720	-0,139	2,789	0,166	0,327
Total	5.222 100%	0,000	0,000	0,000	0,000	0,000	0,000	0,000

#### El primer estrato: población catalana envejecida y de clases trabajadoras en ciudades no metropolitanas

Las secciones censales que corresponden a este estrato son 576, que representan el 11,0% de toda Cataluña. Su presencia es mayor en las Comarques Gironines y Centrals, en municipios de tamaño medio. Se proyectan, sobre todo, en el segundo eje (origen catalán) y, en menor medida, sobre el primero (categorías socio-profesionales bajas).

Su perfil característico es el de personas mayores, en buena medida jubiladas y con una destacada proporción de situaciones de invalidez. Son secciones censales con una población mayoritaria de origen catalán, casada, donde también es característica la ocupación en el sector industrial, en categorías laborales de operario cualificado, y los niveles de estudios bajos.

Comparte con el estrato 4 la mayor presencia de personas de más edad y de origen catalán, si bien en aquél se destaca la importancia del sector agropecuario de los municipios más pequeños.

#### El segundo estrato: movilidad de familias jóvenes de clase media

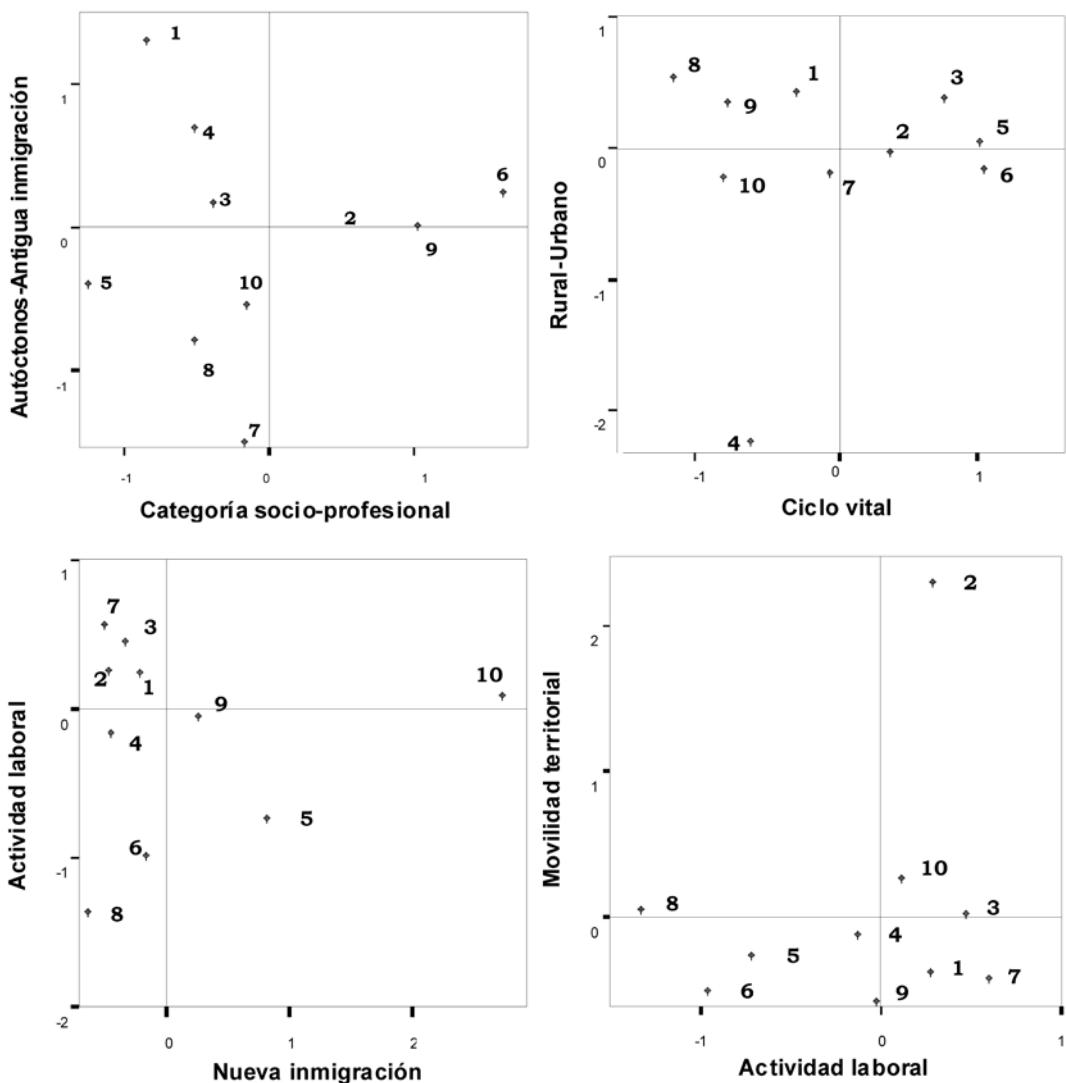
El estrato 2 lo componen 397 secciones censales, el 7,6% del total de secciones de Cataluña. Se distribuyen, sobre todo, en el ámbito metropolitano y en el Camp de Tarragona, en municipios de tamaño medio. Es un estrato con una mayor proyección sobre el factor 7 (movilidad territorial).

Se trata, por tanto, de población joven que ha llegado hace poco al municipio, en edad reproductiva, casada y en familias nucleares, que reside en viviendas grandes que financian con hipotecas. Su lugar de trabajo se localiza mayoritariamente en otros municipios, destacando la proporción de hombres ocupados en sectores industriales o de servicios, en categorías laborales medias o altas, con estabilidad laboral y jornadas extensas. Destaca también, en particular, la mayor presencia de personas procedentes de la Unión Europea con entre 6 y 10 años de residencia en España.

#### El tercer estrato: población joven de clase trabajadora de municipios urbanos

El estrato 3 comparte algunos perfiles con el anterior, pero es más numeroso. Se compone de 723 secciones censales, que representan el 13,8% de las de toda Cataluña. Es igualmente población

Gráfico 1. Posición de los estratos en los factores



- Estrato 1 Población catalana envejecida y de clases trabajadoras en ciudades no metropolitanas  
 Estrato 2 Movilidad de familias jóvenes de clase media  
 Estrato 3 Población joven de clase trabajadora de municipios urbanos  
 Estrato 4 Los pequeños municipios rurales  
 Estrato 5 Población urbana joven de clase trabajadora precaria  
 Estrato 6 Clases sociales media y alta de las grandes ciudades  
 Estrato 7 Antigua inmigración en el área metropolitana  
 Estrato 8 Población envejecida urbana de la antigua inmigración  
 Estrato 9 Población de capital de clase media-alta envejecida  
 Estrato 10 La nueva inmigración

Fuente: elaboración propia

residente en el ámbito metropolitano y en el Camp de Tarragona, pero con presencia también en el Àmbit de Ponent y en las Terres de l'Ebre, en municipios de mayor tamaño. Sus secciones tienen una mayor proyección sobre el eje 3 (ciclo vital joven).

Así, contemplamos un grupo de población joven, como en el segundo estrato, casada y que convive en hogares también nucleares. Tienen igualmente hipotecas, pero de viviendas más modestas. Precisamente, la principal diferencia en relación al segundo estrato es una posición social de clase inferior; corresponden de forma destacada a hombres activos con jornadas extensas, en categorías profesionales de obreros cualificados del sector industrial, y con una mayor precariedad laboral. Comparten de nuevo con el segundo estrato el hecho de trabajar en otro municipio, si bien no obedecen al factor de movilidad territorial.

### **El cuarto estrato: los pequeños municipios rurales**

Este estrato contiene 504 secciones censales, el 9,7 % del total. Agrupa sobre todo a las secciones censales de los municipios más pequeños de fuera del ámbito metropolitano y alejados de los núcleos urbanos más poblados: se extiende por el Àmbit de Ponent, las Terres de l'Ebre, las Comarcas Gironines y Centrals, el Camp de Tarragona y el Alt Pirineu i Aran. Es el estrato que se proyecta fundamentalmente sobre el eje 4 (rural-urbano).

Se trata de una población de origen catalán, bastante envejecida. Destacan los hogares con familias extensas o sin núcleo y con residencia en viviendas grandes, principalmente heredadas. Al tratarse de un ámbito estrictamente rural, predomina la presencia de la ocupación masculina en el sector agropecuario, en las categorías de empresarios, con o sin asalariados, y con una formación básica.

### **El quinto estrato: población urbana joven de clase trabajadora precaria**

Está formado por 424 secciones censales, que corresponden al 8,1% de las secciones de Cataluña, distribuidas en los municipios de más de 20.000 habitantes del ámbito metropolitano de Barcelona, pero también en las Comarcas Gironines, el Camp de Tarragona y las Terres de l'Ebre. Se proyectan sobre todo en los ejes 1 (categoría socio-profesional baja), 3 (ciclo vital joven) y, parcialmente, en el 5 (nueva inmigración).

Estamos ante un estrato marcado por el perfil de población joven de clase trabajadora, como el estrato 3, pero en peores condiciones sociales, en categorías laborales de operarios cualificados y no cualificados, con bajo nivel de estudios y con los más altos niveles de precariedad laboral y de paro. En estas secciones predominan las personas nacidas en el resto de España, con una parte de inmigración más reciente, lo que significa hablar de hijos de una primera generación de inmigrantes que muestran unos bajos conocimientos de la lengua catalana. Son familias jóvenes, nucleares, también hipotecadas.

### **El sexto estrato: clases sociales media y alta de las grandes ciudades**

Las 425 secciones censales de este estrato, que representan el 8,1% del total de Cataluña, se concentran en las grandes ciudades catalanas, especialmente en Barcelona y en su ámbito metropolitano, pero también en Tarragona, Girona y Lleida. Su proyección se da sobre todo en el eje 1 (categoría socio-profesional alta), y también sobre el 3 (ciclo vital joven) y el 6 (actividad laboral).

En este estrato hay una importante presencia de las categorías socio-profesionales y de posición social alta. Se trata de aquellas secciones censales con el valor más elevado en cuanto al primer eje factorial: directivos, profesionales y técnicos altos, empresarios, principalmente en el sector servicios, y con jornadas laborales de menos de 40 horas semanales. Presentan unas tasas más altas de actividad y de ocupación, femenina en particular, así como elevados niveles educativos y de escolarización universitaria. Son catalanes de origen, arraigados en el municipio, en el cual también trabajan, residen en viviendas de más de 100 m<sup>2</sup> y poseen una segunda residencia.

No sobresale ninguna franja de edad como signo característico, pero se sitúan en el perfil de ciclo vital de familias jóvenes del tercer factor. Sí aparece como propio de este estrato la mayor presencia de personas divorciadas, de hogares compuestos de madres con hijos y de personas que viven en régimen de alquiler.

### **El séptimo estrato: antigua inmigración en el área metropolitana**

Este estrato, con 769 secciones censales, comporta el 14,7% del conjunto de Cataluña, siendo el más numeroso junto con el estrato 9. Sus secciones se concentran en su totalidad en Barcelona y en la primera corona metropolitana. Su proyección máxima es sobre el eje 2 (antigua inmigración).

Su rasgo característico es el de congregar a la población metropolitana que proviene de la inmigración del resto de España y que está arraigada en el municipio desde hace tiempo, más de 10 años por lo menos. A pesar de ello, presentan unos niveles bajos de conocimiento de la lengua catalana. Su estructura de edades es similar a la del conjunto de la población, con una presencia más destacada de la franja de 20 a 29 años. Es un grupo de categorías socio-profesionales bajas, sobre todo de operarios cualificados de la industria. La formación que caracteriza el estrato son la FP I y FP II. Este grupo posee un relativamente importante nivel ocupacional en su tasa de ocupación y de actividad, en particular masculina, con una cierta estabilidad contractual, si bien aparece también la contratación eventual. Trabajan en mayor proporción en otros municipios distintos al de residencia. De forma característica, son secciones de hogares nucleares y de viviendas pequeñas, de menos de 70 m<sup>2</sup>.

### **El octavo estrato: población envejecida urbana de la antigua inmigración**

Las 407 secciones censales de este estrato, el 7,8% del total, pertenecen también a los grandes núcleos urbanos del área metropolitana, pero abarcan ahora hasta la segunda corona y a una pequeña parte de los grandes núcleos urbanos del resto de Cataluña, sobre todo de Tarragona. Son secciones con una mayor proyección sobre los ejes 6 (inactividad) y 3 (ciclo vital de vejez), y también en el 2 (antigua inmigración).

Es decir, su rasgo distintivo es la elevada proporción de personas de más de 65 años, la mayor de todos los estratos, juntamente con el hecho de ser una población inmigrada del resto de España, como en el caso anterior. Comparte también el arraigo en el territorio, el desconocimiento de la lengua catalana y habitar en viviendas poco espaciosas. Al tratarse de secciones censales muy envejecidas, comporta un predominio de población inactiva en situaciones de invalidez, viudedad y jubilación.

No obstante, son zonas del territorio donde conviven igualmente personas más jóvenes que sufren en buena medida la desocupación y confluye también en ellas una mayor presencia de personas dedicadas a las tareas del hogar.

### **El noveno estrato: población de capital de clase media-alta envejecida.**

Tiene 771 secciones censales, por lo que configura el estrato más numeroso, con el 14,8% de las secciones de toda Cataluña. Están predominantemente localizadas en Barcelona ciudad; el resto se reparte entre las tres capitales de provincia catalanas. El estrato tiene su mayor proyección sobre el eje 1 (categorías socio-profesionales medias y altas) y también sobre el 3 (ciclo vital de vejez).

La población urbana de estos estratos se distribuye, no obstante, entre las diferentes edades, si bien con un peso más destacado de los que tienen más de 65 años, por lo que sobresale la proporción de personas jubiladas o en situación de viudedad. Pero es un estrato en el que se combinan otras características diversas. Aunque se trate de una población que mayoritariamente lleva mucho tiempo en el municipio, de origen catalán y con un buen conocimiento de la lengua catalana, se encuentran también en este estrato personas nacidas en el resto de España y, con una proporción destacada, de origen latinoamericano.

Son secciones con personas también activas, en particular mujeres, con niveles medios o superiores de estudios y cualificadas, con estabilidad laboral, que trabajan fundamentalmente en el sector servicios en el mismo municipio y con jornadas laborales cortas.

Este perfil se completa con un peso importante de los hogares unipersonales, sin núcleo o monoparentales. Sus viviendas son de tamaño medio, con una buena proporción de tenencia en alquiler y una proporción de segundas residencias por encima de la media global.

### **El décimo estrato: la nueva inmigración**

Finalmente, el estrato décimo, con 226 secciones censales que representan el 4,3% del total, es el de menos peso del conjunto de los estratos en número de secciones censales, si bien es especialmente significativo en cuanto a los cambios sociales más recientes. Se distribuye, sobre todo, en zonas urbanas del ámbito metropolitano, pero también en las ciudades de las Comarques Gironines, el Camp de Tarragona y las Comarques Centrales. Su proyección es mayor sobre el eje 5 (nueva inmigración), así como sobre el 3 (ciclo vital de vejez).

Confluye en este estrato una doble realidad social definida: por una parte, a partir de la edad y, por otra, de la nueva inmigración. De un lado, se trata de una población con una proporción destacada de personas de más de 65 años, circunstancia que confluye con el predominio de las situaciones de jubilación y viudedad. Pero, a la vez, otro de sus rasgos destacables es la concentración de la población recientemente inmigrada y llegada al municipio con orígenes diversos, africanos y latinoamericanos sobre todo, no precisamente mayor, y que convive con autóctonos nacidos en Cataluña o en el resto de España.

En consecuencia, se dan diferentes perfiles característicos: la alta proporción de hogares sin núcleo y unipersonales, de personas solteras, de viviendas pequeñas y de alquiler, o el bajo conocimiento de la lengua. La población activa está ocupada especialmente en el sector servicios, con baja cualificación y con la más alta tasa de eventualidad, realidad que se acompaña con altas tasas de desempleo.

## **2. Afijación de la muestra y cuotas muestrales**

Una vez determinados los estratos, con la caracterización social descrita y como expresión de conjuntos homogéneos de población agregada en secciones censales, corresponde ahora determinar la distribución de los individuos o unidades de la muestra en estos estratos, partiendo de un número total preestablecido. 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 muestral.

### **2.1. Tamaño de la muestra y afijación**

Como medida de variabilidad se considera la distancia cuadrática euclíadiana de cada sección al centro global de la nube de puntos compuesto de secciones censales en un espacio vectorial reducido a una base de siete vectores o variables-factores, componentes del ACP, y, como parámetro, la media de esta distancia. De esta manera se consigue estimar y caracterizar cada unidad o sección censal no con una variable que exprese una sola característica aislada de las secciones censales, sino con una variable densa, compuesta de combinaciones de los ejes factoriales que concentran una gran parte de la varianza. De dicha variable sintética se calcula la distancia a cada sección, su valor medio y su desviación, datos fundamentales para el cálculo de la muestra.

El tamaño de la muestra para el conjunto de Cataluña se fijó inicialmente en 8.348 individuos. Si consideramos un nivel de significación de  $2\sigma$ , con este tamaño muestral  $n$  para datos globales de Cataluña, el error muestral relativo que se obtiene es del 0,78%, aplicando la fórmula siguiente:

$$n = \frac{z^2 \cdot \sigma_y^2}{e^2 \cdot \bar{Y}^2} \quad \text{de donde} \quad e = \sqrt{\frac{z^2 \cdot \sigma_y^2}{n \cdot \bar{Y}^2}}$$

con:

- $z$  el número de sigmas de nivel de significación,
- $\sigma_y^2$  la varianza de la distancia cuadrática euclíadiana de las secciones censales al centro de la totalidad de la nube (valor obtenido 0,783),
- $e$  el error muestral,
- $\bar{Y}$  la media de las distancias cuadráticas euclidianas (valor obtenido de 2,493).

Este número de individuos se distribuye entre los estratos definidos por la muestra 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 que se expresa en la fórmula:

$$n_h = \frac{N_h \cdot \sigma_h}{\sum_{h=1}^K N_h \cdot \sigma_h} \cdot n$$

donde:

- $n_h$  es el tamaño muestral del estrato  $h$  ( $h=1 \dots 10$ ),
- $N_h$  es la población de 16 o más años del estrato  $h$ ,
- $\sigma_h$  es la desviación de la distancia cuadrática euclíadiana de las secciones censales del estrato  $h$  al centro de su estrato.

Los valores que se obtienen de la afijación se presentan en la Tabla 5.

Tabla 5. Afijación muestral óptima de Neyman para cada estrato

Estrato	$N_h$	$\sigma_h$	$N_h \cdot \sigma_h$	Coeficiente de afijación	$n_h$
1	640.561	0,47324	303.138,8	0,10173	849,27
2	556.508	0,83655	465.545,6	0,15624	1.304,27
3	977.070	0,40956	400.172,8	0,13430	1.121,12
4	248.814	0,78853	196.197,4	0,06584	549,66
5	467.962	0,67811	317.328,4	0,10650	889,02
6	435.320	0,60354	262.731,6	0,08817	736,07
7	845.423	0,47077	397.996,4	0,13357	1.115,02
8	354.574	0,56753	201.231,2	0,06753	563,77
9	686.607	0,42044	288.678,8	0,09688	808,76
10	195.425	0,75077	146.719,3	0,04924	411,05
<b>Total</b>	<b>5.408.264</b>	<b>0,62386</b>	<b>2.979.740,3</b>	<b>1,00000</b>	<b>8.348,00</b>

## 2.2. 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 cuatro etapas:

### Etapa 1: distribución proporcional de cuotas

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 mediante la fórmula:

$$n_{sh} = \frac{N_{sh}}{N_h} \cdot n_h$$

donde:

- $n_{sh}$  es la cuota de muestra de la sección  $s$  del estrato  $h$ ,  
 $N_{sh}$  es la población de 16 o más años de la sección  $s$  del estrato  $h$ ,  
 $N_h$  es la población de 16 o más años del estrato  $h$ ,  
 $n_h$  es el tamaño muestral del estrato  $h$ .

### Etapa 2: subestratificación

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<sup>7</sup> y según el tamaño de los municipios (hábitat) a los que pertenecen las secciones (Tabla 6).

Tabla 6. Variables territoriales de subestratificación

Ámbito territorial	Población 16 y más	%	Hábitat	Población 16 y más	%
Àmbit Metropolità	3.749.174	69,3	Hasta 500 hab.	78.739	1,5
Comarques Gironines	468.550	8,7	501–2.000 hab.	253.618	4,7
Camp de Tarragona	380.708	7,0	2.001–5.000 hab.	332.814	6,2
Terres de l'Ebre	135.700	2,5	5.001–20.000 hab.	952.450	17,6
Àmbit de Ponent	255.825	4,7	Más de 20.000 hab.	3.790.643	70,1
Comarques Centrals	364.167	6,7	<b>Total</b>	<b>5.408.264</b>	<b>100,0</b>
Alt Pirineu i Aran	54.140	1,0			
<b>Total</b>	<b>5.408.264</b>	<b>100,0</b>			

De esta forma, se realiza la distribución de las 8.348 encuestas determinadas inicialmente entre 350 subestratos  $h_j$  teóricos, resultado de combinar los 10 estratos con los 7 ámbitos territoriales y con los 5 tamaños municipales, que generan 187 subestratos reales no nulos.

<sup>7</sup> Ámbitos territoriales definidos por el Plan Territorial de Cataluña.

Esta distribución implica contemplar todas y cada una de las secciones censales de Cataluña<sup>8</sup>, lo que significa, desde el punto de vista del trabajo de campo, unos costes muy elevados, derivados de los obligados desplazamientos por todo el territorio catalán para la realización de las entrevistas personales. En este sentido, y para favorecer la concentración territorial, se ha aplicado un criterio de redistribución de las cuotas de muestra a partir de la diferenciación de las secciones que pertenecen a los municipios más pequeños (hasta 20.000 habitantes) de los más grandes.

En el caso de los subestratos con más de 20.000 habitantes, se consideran todas las secciones censales con la cuota de muestra asignada inicialmente.

Si la sección censal es de un subestrato de municipios con una población inferior o igual a 20.000 habitantes, en total 137 subestratos, se efectúa la concentración territorial mediante la atribución de la cuota total de muestra del subestrato a una selección aleatoria de secciones del mismo de la forma siguiente:

- a) En primer lugar se realiza la extracción aleatoria de secciones. Para ello se divide la cuota de muestra del subestrato entre 3 para fijar el número de secciones del subestrato que serán seleccionadas de forma aleatoria, es decir:

$$ns_{h_j} = \frac{n_{h_j}}{3}$$

donde:

$ns_{h_j}$  es el número de secciones a seleccionar del subestrato  $h_j$  (con  $j=1\dots 137$ ) de municipios hasta 20.000 habitantes,

$n_{h_j}$  es la cuota muestral del subestrato  $h_j$ .

- b) Se procede a la concentración de la cuota de muestra del subestrato en las secciones censales seleccionadas ( $\bar{s}$ ) de forma proporcional al peso poblacional de la sección sobre el total de las secciones seleccionadas del subestrato. La asignación de cuotas se expresa en la fórmula:

$$n_{\bar{s}h_j} = \frac{N_{\bar{s}h_j}}{N_{\bar{h}_j}} \cdot n_{h_j}$$

donde:

$n_{\bar{s}h_j}$  es la cuota de muestra de la sección  $s$  seleccionada ( $\bar{s}$ ) del subestrato  $h_j$  (con  $j=1..137$ ) de municipios hasta 20.000 habitantes,

$N_{\bar{s}h_j}$  es la población de 16 y más años de la sección  $s$  seleccionada ( $\bar{s}$ ) del subestrato  $h_j$ ,

$N_{\bar{h}_j}$  es la población de 16 y más años total del subestrato  $h_j$  con las secciones seleccionadas,

$n_{h_j}$  es la cuota muestral del subestrato  $h_j$ .

La asignación que se obtiene da lugar a cuotas de muestra no enteras para cada sección censal, por lo que, a efectos de elección de los individuos, se computa el redondeo generando un tamaño muestral final de 8.352 encuestas. La distribución de frecuencias de las cuotas de muestra finales se presenta en la Tabla 7.

El proceso seguido implica finalmente tomar 4.476 de las 5.222 secciones censales; se dejan de considerar, por tanto, 746, aquellas que tienen un peso poblacional más reducido y que, siendo del mismo subestrato, aleatoriamente se han descartado para favorecer la concentración en el territorio.

8 Teniendo en cuenta que algunas secciones censales inicialmente obtienen una cuota de muestra redondeada de 0 (Tabla 7).

Tabla 7. Cuotas de muestra finales

Cuota de muestra	Secciones	%	Cuota de muestra	Secciones	%
0	746	14,3	6	26	0,5
1	2.101	40,2	7	7	0,1
2	1.443	27,6	8	3	0,1
3	562	10,8	9	1	0,0
4	230	4,4	11	1	0,0
5	102	2,0			
			<b>Total</b>	<b>5.222</b>	<b>100,0</b>

### Etapa 3: ampliación de la muestra

La necesidad de garantizar un nivel de representatividad suficiente para determinadas unidades territoriales de interés en el estudio condujo a la ampliación de la muestra, por lo que se añadieron 2.094 encuestas que se repartieron de forma desigual entre algunos ámbitos territoriales y comarcas para garantizar, como submuestra, un tamaño suficiente, según se indica en la tabla siguiente:

Tabla 8. Muestra ampliada

Territorio	Muestra Adicional	Territorio	Muestra Adicional
Camp de Tarragona	164	Bages	111
Terres de l'Ebre	361	Berguedà	15
Àmbit de Ponent	348	Garraf	124
Alt Pirineu i Aran	504	Osona	109
Alt Penedès	173	Solsonès	19
Anoia	166	Total	2.094

Para cada uno de estos 11 territorios, el tamaño de muestra adicional se repartió de forma proporcional al peso de la cuotas de muestra de cada territorio. El reparto se expresa en la fórmula:

$$na_{\bar{s}t} = \frac{n_{\bar{s}t}}{n_t} \cdot na_t$$

donde:

- $na_{\bar{s}t}$  es la cuota de muestra adicional de la sección  $s$  seleccionada ( $\bar{s}$ ) del territorio  $t$  ( $t=1\dots11$ ),
- $n_{\bar{s}t}$  es la cuota de muestra de la sección  $s$  seleccionada ( $\bar{s}$ ) en el territorio  $t$ ,
- $n_t$  es la cuota de muestra del territorio  $t$ ,
- $na_t$  es la muestra adicional del territorio  $t$  (Tabla 8).

La cuota final de las secciones censales de estos territorios son el resultado del redondeo al número entero de la suma de la cuota inicial más la adicional.

### Etapa 4: concentración de cuotas en Alt Pirineu i Aran

Finalmente, en el caso particular del subestrato de los municipios de hasta 500 habitantes del territorio del Alt Pirineu i Aran, que pertenecen al estrato 4, se operó una segunda extracción aleatoria de secciones con el objetivo de reiterar la concentración territorial de las entrevistas, dadas las características particulares de escasa población de esta zona. Del total de 46 secciones censales de este subestrato se seleccionaron aleatoriamente 13, y se les asignó la cuota de muestra del subestrato, 132 encuestas, de forma proporcional al peso muestral de éstas.

### 2.3. Distribución final de la muestra

Teniendo en cuenta las redistribuciones de cuotas y la muestra adicional, se obtiene la distribución final de encuestas a realizar en cada una de las secciones censales seleccionadas. En las Tablas 9, 10 y 11 se presenta la distribución de las cuotas de muestra que resulta del proceso seguido: la distribución inicial, la redistribución con el criterio de concentración territorial en los municipios de hasta 20.000 habitantes y la ampliación de la muestra. Estas tres distribuciones se presentan según el estrato, el ámbito territorial y el hábitat.

Tabla 9. Distribución de las cuotas de muestra según el estrato

Estrato	Cuotas iniciales		Redistribución de cuotas		Muestra adicional	Muestra total	
Estrato 1	849	10,2%	850	10,2%	446	1.296	12,4%
Estrato 2	1.304	15,6%	1.300	15,6%	196	1.496	14,3%
Estrato 3	1.121	13,4%	1.113	13,3%	352	1.465	14,0%
Estrato 4	550	6,6%	549	6,6%	691	1.240	11,9%
Estrato 5	889	10,6%	884	10,6%	182	1.066	10,2%
Estrato 6	736	8,8%	735	8,8%	67	802	7,7%

Tabla 10. Distribución de las cuotas de muestra según el ámbito territorial

Ámbito territorial	Cuotas iniciales		Redistribución de cuotas		Muestra adicional	Muestra total	
Àmbit Metropolità	5.661	67,8%	5.669	67,9%	297	5.966	57,1%
Comarques Gironines	773	9,3%	774	9,3%	0	774	7,4%
Camp de Tarragona	620	7,4%	618	7,4%	164	782	7,5%
Terres de l'Ebre	239	2,9%	237	2,8%	361	598	5,7%
Àmbit de Ponent	435	5,2%	434	5,2%	348	782	7,5%
Comarques Centrals	525	6,3%	530	6,3%	420	950	9,1%
Alt Pirineu i Aran	95	1,1%	90	1,1%	504	594	5,7%
<b>Total</b>	<b>8.348</b>	<b>100,0%</b>	<b>8.352</b>	<b>100,0%</b>	<b>2.094</b>	<b>10.446</b>	<b>100,0%</b>

Tabla 11. Distribución de las cuotas de muestra según el hábitat

Hábitat	Cuotas iniciales		Redistribución de cuotas		Muestra adicional	Muestra total	
Hasta 500 hab.	169	2,0%	162	1,9%	181	343	3,3%
501–2.000 hab.	484	5,8%	487	5,8%	462	949	9,1%
2.001–5.000 hab.	561	6,7%	557	6,7%	306	863	8,3%
5.001–20.000 hab.	1.565	18,7%	1.563	18,7%	619	2.182	20,9%
Más de 20.000 hab.	5.570	66,7%	5.583	66,8%	526	6.109	58,5%
<b>Total</b>	<b>8.348</b>	<b>100,0%</b>	<b>8.352</b>	<b>100,0%</b>	<b>2094</b>	<b>10.446</b>	<b>100,0%</b>

### 3. Ponderación *a posteriori* de la muestra y elevación

El diseño muestral estratificado de la *Encuesta de condiciones de vida y hábitos de la población de Cataluña, 2006* 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, habrá individuos de 16 y más años que tendrán 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 pertenecen, es decir, no se garantiza el criterio de equiprobabilidad cuando un individuo es elegido al azar.

La afijación óptima que se ha aplicado tiene un sentido instrumental, ya que nos asegura la presencia en la muestra de aquellas características menos frecuentes en la población, pero, a la vez, sobredimensiona la presencia de los individuos de los estratos más heterogéneos que las poseen. Por otra parte, por el procedimiento seguido en la distribución y posterior redistribución de las cuotas muestrales, la ponderación *a posteriori* de la muestra restituye igualmente la proporcionalidad poblacional de los diferentes subestratos definidos por la estratificación inicial de la muestra, el ámbito territorial y el tamaño del hábitat. En consecuencia, una vez obtenida la muestra, es preciso restituir el valor real de las frecuencias ponderando su peso en el conjunto y garantizar así una muestra aleatoria proporcionalmente dimensionada a su población.

La ponderación  $w_s$  que se aplica es una magnitud que transforma la probabilidad real de que un individuo, perteneciente a una sección censal, sea elegido en la probabilidad teórica bajo la hipótesis de estricta aleatoriedad, y se expresa con la relación siguiente:

$$w_s = \frac{\text{Probabilidad teórica}}{\text{Probabilidad real}} = \frac{N_s / N}{n_s / n}$$

donde:

- $w_s$  es la ponderación de cada individuo de la muestra de la sección  $s$ ,  
 $N_s$  es la población de 16 y más años de la sección  $s$ , del año 2006,  
 $n_s$  es el número de individuos de la muestra de la sección  $s$ ,  
 $N$  es la población total de 16 y más años, del año 2006,  
 $n$  es el número de casos de la muestra.

Es decir, se otorga un menor peso a aquellos individuos que tienen una probabilidad mayor de ser elegidos, y un mayor peso a aquellos otros con una probabilidad menor.

El diseño de la muestra de la *Encuesta* implica la selección de personas a encuestar pertenecientes a unidades agregadas homogéneas como son las secciones censales. Como hemos indicado, este diseño conlleva una selección aleatoria de secciones censales, de forma que el conjunto de encuestas no se ha realizado en todas ellas; la concentración de cuotas de muestra realizada en algunos subestratos implica que una parte de las encuestas sean expresión de un número de secciones y de población más amplio. La muestra final obtenida ha sido de 10.398 individuos que pertenecen a 4.415 secciones censales, elegidas aleatoriamente, del total de 5.219<sup>9</sup>. Estas secciones censales están distribuidas en 166 subestratos, que abarcan una población de 5.356.760 personas de 16 y más años, según los datos recogidos en el Padrón Continuo a 1 de enero de 2006, con relación a una población catalana total de 16 y más años de 6.049.414, que se distribuye en 187 subestratos.

Es decir, en relación con el total poblacional, el diseño muestral, en su objetivo de favorecer la concentración territorial del trabajo de campo, no recoge unidades muestrales con respecto a 692.654 personas de 16 y más años de la población del año 2006: de 21 subestratos (cuya cuota muestral

<sup>9</sup> Inicialmente las secciones eran 5.222, pero tres de ellas desaparecieron en el Padrón de Habitantes del año 2005. Inicialmente, también el tamaño de la muestra era de 10.446, pero la muestra final se cerró con 10.398 personas encuestadas.

era mínima e inferior a 0,5, y comprendía a 8.855 personas de la población) y de 804 secciones que no son elegidas en la muestra final (y que afectan a una población de 683.799 personas).

La ponderación de la muestra, que actúa en relación con las secciones elegidas y a su peso poblacional, restituye la representación del conjunto de la población de Cataluña de 16 y más años. A efectos de ponderación, y con el objetivo de abarcar la magnitud total de la población, se ha procedido a la redistribución de la población excluida entre las secciones con cuota muestral del mismo subestrato.

$$w_s = \frac{\frac{N_{s_{h_j}} + \sum N_{s_{h_j}} \cdot (N_{h_j} - \sum N_{s_{h_j}})}{\sum N_{s_{h_j}} / n} / n}{n_{s_{h_j}} / n} = \frac{N_{s_{h_j}} + \bar{N}_{s_{h_j}} / N}{n_{s_{h_j}} / n}$$

donde:

- $w_s$  es la ponderación de cada individuo de la muestra de la sección  $s$ ,
- $N_{s_{h_j}}$  es la población de 16 y más años de la sección  $s$  elegida en la muestra del subestrato  $h_j$ ,
- $\sum N_{s_{h_j}}$  es la población de 16 y más años de todas las secciones  $s$  elegidas en la muestra del subestrato  $h_j$ ,
- $\bar{N}_{s_{h_j}}$  es la población de 16 y más años de todas las secciones  $s$  no elegidas en la muestra del subestrato  $h_j$ ,
- $N$  es la población total de 16 y más años,
- $n_{s_{h_j}}$  es el número de casos de la muestra pertenecientes a la sección  $s$  del subestrato  $h_j$ ,
- $n$  es el número de casos de la muestra.

Esta redistribución se ha efectuado de forma proporcional al peso poblacional de las secciones del subestrato elegidas para la muestra. De esta forma, la ponderación  $w_s$  adopta la expresión:

Una vez realizada la ponderación de los datos muestrales, éstos se expresaran en magnitudes poblacionales a partir de la elevación de la muestra introduciendo un factor de ponderación  $f$  para cada individuo de la muestra, que se obtiene simplemente poniendo en relación el tamaño de la población y el tamaño de la muestra, a partir de la relación:

$$\text{Factor de elevación} = f = \frac{N}{n}$$

Finalmente, pues, se ha obtenido información de 10.398 personas encuestadas, que, una vez ponderadas, corresponden a un tamaño de muestra al que se le asocia un nivel de error global del 0,72%, para un nivel de significación de  $2\sigma$ .

## 4. Aplicación de los resultados de la muestra

Para acabar, queremos apuntar brevemente algunos resultados o productos que se derivan del diseño y de la construcción de la muestra estratificada que hemos presentado. Más allá de los beneficios de eficacia estadística que el diseño comporta, el procedimiento seguido proporciona conclusiones de interés para el análisis sociológico y para el conocimiento social del territorio.

- a) En primer lugar, la modalidad del diseño de estratificación propuesta es de por sí un estudio sociológico, basado en datos censales de toda la población, sin error aleatorio salvo el de la medición, que nos proporciona un contenido social de gran valor como imagen de la estructura social de Cataluña.
- b) Este análisis e interpretación de la realidad social que supone la confección de la muestra estratificada tiene, a su vez, una base y un soporte territorial geográfico, la sección censal, que nos proporciona la posibilidad de elaborar una cartografía social de gran interés con una gran flexibilidad en la elección de los territorios y en su comparación.

- c) En este sentido, es posible reutilizar la información de la muestra con el objetivo de construir lo que hemos denominado zonas sociales (Lozares, y Domínguez, 1993, 1996; Lozares, López, y Flores, 2003), que no son otra cosa que los estratos y los resultados del análisis e interpretación de los datos censales en ellos, añadidos a los datos de la encuesta, puesto que la homogeneidad de los estratos permite un análisis autónomo de los mismos a partir de las encuestas realizadas en ellos. Así pues, el contenido social de estas zonas se enriquece con la información obtenida en la propia *Encuesta*, que facilita la validación de los resultados y proporciona una nueva imagen social de la realidad.
- d) La estratificación, a partir de los datos censales, tal y como se ha desarrollado, sirve además para ser aplicada en la realización de otras muestras y estudios que asuman las diferencias sociales expresadas en el territorio.

## Bibliografía

- COCHRAN, W. G. *Técnicas de muestreo*. México: CECSA, 1971.
- GINER, S. [et al.]. *Enquesta de la Regió de Barcelona 2000. Informe General*. Barcelona: Mancomunitat de Municipis de l'Àrea Metropolitana de Barcelona i Diputació de Barcelona, 2002.
- GROSBAS, J. M. *Méthodes statistiques des sondages*. París: Ed. Economica, 1987.
- LÉBART, L. [et al.]. *Tratamiento estadístico de datos*. Barcelona: Ed. Marcombo, 1985.
- LÓPEZ ROLDAN, P.; LOZARES COLINA, C. *Anàlisi bivariante de dades estadístiques*. Bellaterra: Universitat Autònoma de Barcelona, 1999. (Col·lecció Materials; 79)
- LÓPEZ, P. [et al.]. «Disseny i construcció d'una mostra estratificada a partir de dades censals». *Qüestió*, vol. 24, núm. 1 (2000), p. 111-136.
- LOZARES, C.; DOMÍNGUEZ, M. *Enquesta de la Regió Metropolitana de Barcelona 1990. Territori i realitat social: les zones sociodemogràfiques de la Regió Metropolitana de Barcelona*. Barcelona: Mancomunitat de Municipis de l'Àrea Metropolitana de Barcelona y Diputació de Barcelona, 1993.
- LOZARES, C.; DOMÍNGUEZ, M. «Tratamiento multivariado de subpoblaciones en una gran encuesta social: la construcción de zonas sociales». *Papers. Revista de Sociología*, núm. 48 (1996), p. 71-87.
- LOZARES, C.; LÓPEZ, P. *Enquesta Metropolitana de la Regió Metropolitana de Barcelona. Construcció de la mostra estratificada*. Sèrie Documents de Treball 90/1. Barcelona: Institut d'Estudis Metropolitans de Barcelona, 1990.
- LOZARES, C.; LÓPEZ, P. «El análisis de componentes principales. Aplicación al análisis de datos secundarios». *Papers. Revista de Sociología*, 37, (1991a), p. 31-63.
- LOZARES, C.; LÓPEZ, P. «El muestreo estratificado por análisis multivariado». En: LATIESA, M. *El pluralismo metodológico en la investigación social: ensayos típicos*, Granada: Universidad de Granada, 1991b, p. 107-160.
- LOZARES, C.; LÓPEZ, P. *Anàlisi multivariante de dades estadístiques*. Bellaterra: Universitat Autònoma de Barcelona, 2000. (Col·lecció Materials; 93)
- LOZARES, C. [et al.]. *Zones socials per al territori de la província de Barcelona: una cartografia social*, Barcelona: Institut d'Estudis Regionals i Metropolitans de Barcelona, 2003.
- NEL·LO, O. [et al.]. *Enquesta Metropolitana de Barcelona. Condicions de Vida i Hàbits de la Població. La transformació de la societat metropolitana*. Barcelona: Mancomunitat de Municipis de l'Àrea Metropolitana de Barcelona i Diputació de Barcelona, 1998.
- SÁNCHEZ CARRIÓN, J. J. *Introducción a las técnicas de análisis multivariante aplicadas a las ciencias sociales*. Madrid: Centro de Investigaciones Sociológicas, 1984.