Encuesta de Población Activa

Informe Técnico

Madrid, mayo 2002
Área de Diseño de Muestras y Evaluación de Resultados

Índice

. Introducción

La Encuesta de Población Activa (EPA), es una encuesta de tipo continuo dirigida a investigar características socioeconómicas de la población, que viene siendo realizada por el INE desde 1964. El diseño de la encuesta se enmarca en el de la Encuesta General de Población (EGP).

Desde su implantación ha sufrido modificaciones en algunos aspectos, siempre dirigidas a una mejora en la realización de la encuesta.

El presente informe tiene por objeto recoger los aspectos metodológicos del diseño actual, así como la evaluación de la calidad de los datos de la misma.

El INE agradece de antemano cuantas sugerencias se presenten para posibles mejoras futuras de la encuesta.

II. Diseño de la Encuesta

1 Objetivos

La EPA tiene como objetivo principal el conocimiento de la actividad económica del país, en lo relativo al componente humano. Está orientada para dar información de las principales categorías poblacionales en relación con el mercado de trabajo así como obtener clasificaciones de estas categorías según distintas variables.

La experiencia ha demostrado que las diferentes fuentes estadísticas (Censo, Encuestas de Salarios, Paro registrado, etc.) que proporcionan información sobre estos temas no son adecuados para satisfacer los objetivos de la encuesta. En el caso concreto del Censo por diversas razones:

1) Su larga periodicidad impide conocer la situación en períodos intercensales.

2) Los datos del Censo son insuficientes para dar una visión detallada de la situación laboral.

3) Los datos son obtenidos por autoenumeración por lo que existen dificultades, por parte del informante, de interpretación de los conceptos utilizados.

Se justifica así la necesidad de una encuesta continua diseñada y concebida expresamente para conocer el grado de actividad económica de la población, junto a otras características estrechamente relacionadas con dicha actividad.

La encuesta está diseñada para dar resultados detallados a nivel nacional. Para las Comunidades Autónomas y las provincias se ofrece información sobre las principales características al nivel de desagregación que permiten los coeficientes de variación de los estimadores.

Como definición de población económicamente activa se ha tomado la aceptada por la Oficina Internacional de Trabajo (OIT), según la cual se considera ésta como el *conjunto de personas, que en un período de referencia dado, suministran mano de obra para la producción de bienes y servicios económicos o que están disponibles y hacen gestiones para incorporarse a dicha producción*.

La población económicamente activa está constituida por las personas de 16 y más años que en la semana de referencia satisfacen las condiciones necesarias para su inclusión entre las personas ocupadas o paradas de acuerdo con las definiciones dadas para la encuesta.

2 Ámbito de la Encuesta

El ámbito abarcado por la encuesta se desglosa en los tres apartados siguientes:

2.1 ÁMBITO POBLACIONAL

La Encuesta va dirigida a la población que reside en viviendas familiares, es decir, las utilizadas toda o la mayor parte del año como residencia habitual o permanente.

Se excluyen de la investigación los llamados *hogares colectivos*, ejemplo de los cuales son los hospitales, hoteles, cuarteles, conventos, etc.

Sí se incluyen las familias que formando un grupo independiente residan en estos establecimientos, como puede ocurrir con los directores de los centros, conserjes y porteros. En definitiva, teóricamente, sólo queda excluida de la muestra aquella población que carezca de residencia familiar.

2.2 ÁMBITO GEOGRÁFICO

La encuesta se realiza en todo el territorio nacional.

2.3 ÁMBITO TEMPORAL

La EPA es una encuesta continua con periodicidad trimestral extendiéndose las entrevistas a lo largo de las trece semanas del trimestre.

En cuanto al período de referencia hay que distinguir:

Período de referencia de los resultados de la encuesta: el trimestre.

Período de referencia de la información: se ha adoptado, como norma general, la semana anterior (de lunes a domingo) a la de la fecha en que se realiza la entrevista. Dicha semana se denomina *semana de referencia* y todos los datos deben referirse a ella, salvo las excepciones que figuran en el documento *Encuesta de Población Activa. Descripción de la encuesta, definiciones e instrucciones para la cumplimentación del cuestionario*.

3 Marco de la Encuesta

Para definir el marco de la Encuesta es necesario partir de la división administrativa de España que aparece de la forma siguiente:

Toda la Nación se encuentra dividida en 17 Comunidades Autónomas, y a su vez en 50 provincias de las cuales 47 son peninsulares y 3 insulares. Las provincias se encuentran divididas en municipios y éstos en distritos municipales.

Hasta aquí tenemos la división administrativa oficial. Después el INE juntamente con los Ayuntamientos hace una nueva subdivisión de los distritos en secciones censales.

Las secciones se utilizan para todos los trabajos encomendados al INE en los que es necesario una división inframunicipal, entre otros para fines electorales como *secciones electorales*, lo cual exige de acuerdo con la Ley Electoral que cada sección incluya un máximo de 2.000 electores y un mínimo de 500.

Por tanto, la sección censal puede considerarse como un área geográfica con límites perfectamente definidos, cuyo tamaño de población viene limitado por las condiciones antes expuestas.

El seccionado y su número varía considerablemente a lo largo del tiempo, por lo que con referencia 1 de enero de cada año coincidiendo con la revisión del Censo Electoral y en cada Censo o Padrón se realiza una actualización del mismo. Por una parte hay secciones que quedan despobladas y es necesario fusionarlas con otras y por otra también se produce el fenómeno contrario, es decir, las secciones crecen hasta superar los límites de población establecidos y es necesario dividirlas. En todos los casos se actualiza las probabilidad de selección de la sección.

4 Diseño de la muestra

4.1 TIPO DE MUESTREO. UNIDADES MUÉSTRALES

El tipo de muestreo utilizado es un muestreo bietápico con estratificación de las unidades de primera etapa.

Las unidades de primera etapa están constituidas por las secciones censales. La muestra de secciones permanece fija indefinidamente con las excepciones siguientes:

a) Cuando los resultados obtenidos en los Censos arrojen variaciones sensibles en la estructura de la población que aconsejen una afijación distinta.

b) Se agoten los hogares consultables de la sección.

c) Cuando al actualizar las probabilidades de selección le corresponda salir de la muestra.

Las unidades de segunda etapa están constituidas por las viviendas familiares principales (ocupadas permanentemente) y los alojamientos fijos (chabolas, cuevas, etc.). No se consideran las viviendas secundarias (ocupadas sólo una parte del año) y las disponibles para alquiler o venta, ya que no forman parte del ámbito poblacional definido anteriormente.

Dentro de las unidades de segunda etapa no se realiza submuestreo alguno, recogiéndose información de todas las personas que tengan su residencia habitual en las mismas.

4.2 ESTRATIFICACIÓN DE LAS UNIDADES DE MUESTREO

Las unidades de primera etapa se estratifican atendiendo a un doble criterio:

A.- Criterio geográfico (de estratificación)

Las secciones se agrupan en estratos de acuerdo con la provincia y tipo de municipio (según importancia demográfica) a que pertenecen.

B.- Criterio socioeconómico (de subestratificación)

Dentro de cada estrato geográfico las secciones censales se agrupan en *subestratos* atendiendo a la categoría socioeconómica de los hogares ubicados en la sección.

Estratos

Para llegar a la formación de los estratos se consideran los siguientes tipos de municipios:

1. Municipios autorrepresentados: Son aquellos que dada su categoría dentro de la provincia deben tener siempre secciones en la muestra.

Son municipios autorrepresentados:

La capital de la provincia

Municipios que tienen un número de habitantes tal, que en la afijación proporcional dentro de la provincia le corresponden al menos 12 secciones en la muestra.

Municipios que teniendo una situación demográfica destacada dentro de la provincia no hay otros similares con que agruparlos, aunque proporcionalmente le correspondan menos de 12 secciones en la muestra.

2. Municipios correpresentados: Son aquellos que dentro de la misma provincia forman parte de un grupo de municipios demográficamente similares y que son representados en común.

De acuerdo con esta clasificación, en líneas generales, los estratos teóricos considerados responden a los siguientes conceptos:

Estrato 1: Municipio capital de provincia

Estrato 2: Municipios autorrepresentados, importantes en relación con la capital.

Estrato 3: Otros municipios autorrepresentados, importantes en relación con la capital o municipios mayores de 100.000 habitantes.

Estrato 4: Municipios entre 50.000 y 100.000 habitantes.

Estrato 5: Municipios entre 20.000 y 50.000 habitantes.

Estrato 6: Municipios entre 10.000 y 20.000 habitantes.

Estrato 7: Municipios entre 5.000 y 10.000 habitantes.

Estrato 8: Municipios entre 2.000 y 5.000 habitantes.

Estrato 9: Municipios menores de 2.000 habitantes.

Hay que tener en cuenta que dada la diferente distribución de tamaños de los municipios entre las distintas provincias no se ha podido realizar una estratificación uniforme para todas ellas. Por ejemplo en la provincia de Lugo solamente hay 10

municipios con menos de 2.000 habitantes por lo que se han agrupado los estratos teóricos 8 y 9 en el estrato 8 que contiene a los municipios de menos de 5.000 habitantes. Por el contrario la provincia de Burgos tiene más de 350 municipios de menos de 2.000 habitantes, incluidos en el estrato 9, y sin embargo tiene agrupados los estratos teóricos 7 y 8 en el estrato 7, al no haber apenas municipios entre 2.000 y 5.000 habitantes. No obstante, siempre que ha sido posible, se ha procurado realizar una estratificación uniforme para todas las provincias pertenecientes a una misma Comunidad Autónoma.

Los Censos y Padrones aportan la información necesaria para actualizar la estratificación en cada provincia, en función de la distribución de la población de los municipios.

Subestratos

Para la formación de los subestratos hay que tener en cuenta la categoría socioeconómica de los hogares ubicados en la sección.

Las secciones cambian de subestrato debido a la variación de la estructura de la población, por lo que la subestratificación se revisa en cada Censo utilizando la información que éste proporciona sobre las características que intervienen en la definición de categoría socioeconómica.

Esta información permite clasificar la población económicamente activa de la sección en los siguientes grupos:

Grupo 1 (Agricultores). Comprende las siguientes categorías

01. Empresarios agrarios con asalariados

02. Empresarios agrarios sin asalariados

03. Miembros de cooperativas agrarias

04. Directores y Jefes de explotaciones agrarias

05. Resto de trabajadores agrarios

Grupo 2 (Trabajadores por cuenta propia). Comprende las siguientes categorías:

06. Profesionales por cuenta propia

07. Empresarios no agrarios con asalariados

08. Empresarios no agrarios sin asalariados

09. Miembros de cooperativas no agrarias

Grupo 3 (Directivos y profesionales por cuenta ajena y personal administrativo):

10. Directores de empresas no agrarias y altos funcionarios

11. Profesionales por cuenta ajena12. Jefes de Departamento Administrativo, Comercial y de Servicios de Empresas no agrarias o Administración Pública

13. Resto del personal administrativo y comercial

18. Profesionales de las Fuerzas Armadas

Grupo 4 (Resto de trabajadores). Comprende las siguientes categorías:

14. Resto del personal de los servicios

15. Contramaestres y capataces no agrarios

16. Obreros cualificados no agrarios

17. Obreros sin especialización no agrarios

No se incluyen en ningún grupo a los **No clasificables:**

Personas que buscan empleo por primera vez.

Personas económicamente activas que no pueden clasificarse en alguna de las rúbricas anteriores.

Existen dieciséis subestratos, quince de los cuales se obtienen en función de los porcentajes de población de los grupos 1, 2, 3, y 4 y el decimosexto (subestrato 0) está formado por aquellas secciones con un elevado porcentaje de población inactiva.

La definición de los quince primeros subestratos se establece según:1) haya un claro predominio de uno de los cuatro grupos sobre los otros tres; 2) predominen dos sobre los otros dos; 3) predominen tres grupos y 4) no hay un claro predominio de ninguno de los cuatro grupos. En alguno de los estratos pueden no existir varios de los subestratos.

El criterio adoptado para determinar si una sección pertenece a uno u otro subestrato es el siguiente:

El grupo preponderante (el de mayor porcentaje) le llamamos grupo de importancia A. Los siguientes grupos, según orden decreciente de porcentaje los denominamos B, C y D respectivamente.

Por ejemplo, si en una determinada sección los porcentajes de población activa son 40 por ciento del grupo 3, 30 por ciento del grupo 2, 20 por ciento del grupo 1 y 10 por ciento del grupo 4 entonces A=3, B=2, C=1 y D= 4.

A cada sección se le asigna un código de cuatro dígitos de la siguiente manera:

a) El primer dígito es:

1 si el grupo de importancia A es el grupo 1

2 si el grupo de importancia A es el grupo 2

3 si el grupo de importancia A es el grupo 3

4 si el grupo de importancia A es el grupo 4

b) Para obtener el segundo dígito calculamos el cociente:

$$\frac{\text{Porcentaje grupo de importancia B}}{\text{Porcentaje grupo de importancia A}} \text{ que denominaremos } \frac{B}{A}$$

Pueden darse los siguientes casos:

Caso 1º: Si $\frac{B}{A} > 0{,}66$, el 2º dígito será:

1 si el grupo de importancia B es el grupo 1

2 si el grupo de importancia B es el grupo 2

3 si el grupo de importancia B es el grupo 3

4 si el grupo de importancia B es el grupo 4

Caso 2º: Si $\frac{B}{A}$ está comprendido entre 0,33 y 0,66, el 2º dígito será:

5 si el grupo de importancia B es el grupo 1

6 si el grupo de importancia B es el grupo 2

7 si el grupo de importancia B es el grupo 3

8 si el grupo de importancia B es el grupo 4

Caso 3º: Si $\frac{B}{A} < 0{,}33$ el 2º dígito será cero en cualquier caso

1

c) Para obtener el tercer dígito calculamos el cociente:

$$\frac{\text{Porcentaje grupo de importancia C}}{\text{Porcentaje grupo de importancia A}}$$

y aplicaremos el mismo criterio explicado en el apartado anterior b).

d) Para obtener el cuarto dígito calculamos el cociente:

$$\frac{\text{Porcentaje grupo de importancia D}}{\text{Porcentaje grupo de importancia A}}$$

Como fácilmente puede comprobarse el código de subestrato no puede tener dígitos repetidos (salvo el cero).

Código de subestrato cero: Como caso excepcional, si en una sección censal se verifica que:

$$\frac{\text{Poblacion de 16 y mas años Inactiva}}{\text{Poblacion de 16 y mas años Activa}} > 3$$

el código de subestrato de esta sección será cero, sin que sea de aplicación el criterio de asignación de códigos por categoría socioeconómica, expuesto anteriormente.

De acuerdo con la codificación expuesta la subestratificación se realiza de la siguiente manera:

Subestrato O

Comprende únicamente las secciones de código cero, es decir, aquellas secciones con fuerte predominio de población inactiva y, por tanto, no se ha aplicado el criterio de subestratificación por categoría socioeconómica.

Subestrato 1

Agrupa las secciones con fuerte predominio del grupo 1, es decir, aquellas cuyo código de cuatro dígitos es alguno de los siguientes:

1000 - 1600 - 1700 - 1800 - 1670 - 1760 - 1680 - 18601780 - 1870 - 1678 - 1687 - 1768 - 1786 - 1867 - 1876

Subestrato 2

Agrupa las secciones con fuerte predominio del grupo 2, códigos:

2000 - 2500 - 2700 - 2800 - 2570 - 2750 - 2580 - 2850 2780 - 2870 - 2578 - 2587 - 2758 - 2785 - 2857 - 2875

Subestrato 3

Agrupa las secciones con fuerte predominio del grupo 3, códigos:

3000 - 3500 - 3600 - 3800 - 3560 - 3650 - 3680 - 38603580 - 3850 - 3568 - 3586 - 3658 - 3685 - 3856 - 3865

Subestrato 4

Agrupa las secciones con fuerte predominio del grupo 4, códigos:

4000 - 4500 - 4600 - 4700 - 4560 - 4650 - 4670 - 47604570 - 4750 - 4567 - 4576 - 4756 - 4765 - 4657 - 4675

Subestrato 12

Agrupa las secciones con predominio de los grupos 1 y 2, códigos:

1200 - 1270 - 1280 - 1278 - 1287 - 2100 - 2170 - 2180 - 2178 - 2187

Subestrato 13

Agrupa las secciones con predominio de los grupos 1 y 3, códigos:

1300 - 1360 - 1380 - 1368 - 1386 - 3100 - 3160 - 3180 - 3168 - 3186

Subestrato 14

Agrupa las secciones con predominio de los grupos 1 y 4, códigos:

1400 - 1460 - 1470 - 1467 - 1476 - 4100 - 4160 - 4170 - 4167 - 4176

Subestrato 23

Agrupa las secciones con predominio de los grupos 2 y 3, códigos:

2300 - 2350 - 2380 - 2358 - 2385 - 3200 - 3250 - 3280 - 3258 - 3285

Subestrato 24

Agrupa las secciones predominio de los grupos 2 y 4, códigos:

2400 - 2450 - 2470 - 2457 - 2475 - 4200 - 4250 - 4270 - 4257 - 4275

Subestrato 34

Agrupa las secciones con predominio de los grupos 3 y 4, códigos:

3400 - 3450 - 3460 - 3456 - 3465 - 4300 - 4350 - 4360 - 4356 - 4365

Subestrato 123

Agrupa las secciones mixtas de los grupos 1, 2 y 3, códigos:

1230 - 1238 - 1320 - 1328 - 2130 - 2138 - 2310 - 2318 - 3120 - 3128 - 3210 - 3218

Subestrato 124

Agrupa las secciones mixtas de los grupos 1, 2 y 4, códigos:

1240 - 1247 - 1420 - 1427 - 2140 - 2147 - 2410 - 2417 - 4120 - 4127 - 4210 - 4217

Subestrato 134

Agrupa las secciones mixtas de los grupos 1, 3 y 4, códigos:

1340 - 1346 - 1430 - 1436 - 3140 - 3146 - 3410 - 4316 - 4130 - 4136 - 4310 - 3416

Subestrato 234

Agrupa las secciones mixtas de los grupos 2, 3 y 4, códigos:

2340 - 2345 - 2430 - 2435 - 3240 - 3245 - 3420 - 3425 - 4230 - 4235 - 4320 - 4325

Subestrato 1234

Agrupa las secciones mixtas de los grupos 1, 2, 3 y 4, códigos:

1234 - 1243 - 1324 - 1342 - 1423 - 1432 - 2134 - 2143 - 2314 - 2341 - 2413 - 2431 - 3124 - 3142 - 3214 - 3241 - 3412 - 3421 - 4123 - 4132 - 4213 - 4231 - 4312 - 4321

4.3 TAMAÑO DE LA MUESTRA

Para la determinación del número n de secciones y del número m de viviendas por sección de la muestra se partió de una función de coste de tipo lineal y de la expresión del coeficiente de variación para una proporción en el muestreo de conglomerados con submuestreo.

Se empleó la siguiente función de coste:

$$Q = n\,Q_S + n\,m\,Q_V \qquad \text{con} \qquad Q_S = Q_F + d\,Q_D$$

donde:

Q = Presupuesto total para el pago a los entrevistadores

Q_S = Coste por unidad primaria (sección)

Q_V = Coste por unidad última (vivienda)

n = Número de secciones

m = Número de viviendas por sección

Q_F = Coste fijo por sección

Q_D = Coste diario del trabajo de campo

d = Número de días necesarios para el trabajo de campo

Todas las variables eran conocidas excepto n y m.

El coeficiente de variación para una proporción viene dado por

$$C^2(\hat{P}) = \frac{V(\hat{P})}{\hat{P}^2} = \frac{1-\hat{P}}{\hat{P}} \cdot \frac{1+\delta(m-1)}{n\,m} = \frac{1-\hat{P}}{\hat{P}} F(\delta, m, n)$$

siendo:

$$F(\delta, m, n) = \frac{1+\delta(m-1)}{n\,m}$$

y δ el coeficiente de correlación intraclásica, que para el caso de la población activa se ha calculado y vale 0,05.

El mínimo de la expresión $C^2(\hat{P})$ respecto de las variables m y n se obtiene calculando el mínimo de la expresión $F(\delta, m, n)$ que es independiente de $\hat{P}$.

Para distintos valores de m compatibles con el trabajo de campo,

m = 4, 6, 8, 10, 11, 14, 17, 18, 19, ...91, 100

y los correspondientes valores de n dados por

$$n = \frac{Q}{Q_S + m\, Q_V}$$

se obtienen distintos valores para F (δ, m, n).

El valor mínimo de F (δ, m, n) respecto de m y n correspondió a m=20 y n=3.000.

En base a este resultado la muestra se fijó en un total de 3.000 secciones, investigándose una media de 20 viviendas por sección.

Posteriormente, con objeto de lograr una mayor representatividad en algunas Comunidades Autónomas y al mismo tiempo dar cumplimiento a las exigencias de la Unión Europea en cuanto al tamaño de la muestra en las Encuestas de Empleo la muestra ha sufrido diversas ampliaciones. A partir de 1999 se establece un tamaño de 3.484 secciones y 18 viviendas por sección, excepto en las provincias de Madrid, Barcelona, Sevilla, Valencia y Zaragoza en donde el número de entrevistas por sección es de 22.

4.4 AFIJACIÓN

Este apartado recoge los criterios seguidos para la distribución de las secciones de la muestra entre las provincias, dentro de la provincia entre estratos y dentro de éstos entre subestratos.

Para hacer la afijación entre las provincias se tuvieron en cuenta los siguientes aspectos:

a) Disponer en cada provincia de un tamaño mínimo de muestra que permita dar estimaciones de la misma.

b) Los resultados nacionales deben tener la mayor fiabilidad posible.

c) En cada provincia debe haber un número exacto de *Bloques*. Se define el bloque como el conjunto de secciones que ha de visitar un entrevistador durante un trimestre. En el caso concreto de esta encuesta, el bloque consta de 13 secciones repartidas una en cada semana del trimestre. Para compatibilizar las tres condiciones antes expuestas se ha aceptado una afijación de compromiso entre la uniforme y la proporcional, a base de agrupar provincias de importancia demográfica similar y asignarles de 3 a 12 bloques.

Dentro de cada provincia la afijación entre estratos es proporcional al tamaño de cada uno de ellos, si bien se ha potenciado los estratos donde se encuentran los municipios de mayor tamaño ya que se espera que la mayor parte de las características que se estudian estén correlacionadas con los niveles económico-sociales y culturales de los habitantes y es precisamente en estos estratos donde, en general, la dispersión debe ser mayor y donde el costo por entrevista es menor.

Dentro de los estratos, la afijación entre subestratos es estrictamente proporcional al tamaño (medido en número de viviendas familiares).

En el cuadro 1 figura la distribución de las secciones de la muestra por provincias y estratos.

Cuadro 1

Distribución de las secciones de la muestra por provincias y estratos

Provincias	1	2	3	4	5	6	7	8	9	Total
1 Alava	30				3		6			39
2 Albacete	19				7		3	6	4	39
3 Alicante	19	13		6	20	7	7	3	3	78
4 Almeria	16				7	4	3	6	3	39
5 Avila	13						4	6	16	39
6 Badajoz	20				13	6	16	13	10	78
7 Baleares	39				19	13	10	10		91
8 Barcelona	61		36	16	20	10	7	3	3	156
9 Burgos	20				7	3	9			39
10 Cáceres	19				7	4	10	16	22	78
11 Cádiz	13	13	6	20	13	7	6			78
12 Castellón	26				20	10	6	7	9	78
13 Ciudad Real	13	9			13	13	16	7	7	78
14 Córdoba	33				16	7	13	9		78
15 Coruña (La)	22			13	6	17	14	6		78
16 Cuenca	10						7	6	16	39
17 Girona	16				17	13	9	13	10	78
18 Granada	29				13	9	10	10	7	78
19 Guadalajara	20					3		3	13	39
20 Guipúzcoa	26			6	13	20	7	6		78
21 Huelva	16					10	6	7		39
22 Huesca	13					10	6		10	39
23 Jaén	16	7			13	13	13	13	3	78
24 León	23	9				10	7	10	19	78
25 Lleida	16					3	3	7	10	39
26 Logroño	26					6	6	4	10	52
27 Lugo	13					9	7	10		39
28 Madrid	98		30	10	9	3	6			156
29 Málaga	36			10	16	7	9			78
30 Murcia	30	16		6	19	13	7			91
31 Navarra	32				4	10	6	13	13	78
32 Ourense	16					3	4	13	3	39
33 Oviedo	23	29		20	10	19	7	9		117
34 Palencia	20						3	6	10	39
35 Palmas (Las)	43			9	20	7	9	3		91
36 Pontevedra	10	26			13	16	10	3		78
37 Salamanca	20					3	3		13	39
38 S.Cruz Tenerife	23	13			13	10	13	6		78
39 Santander	29	10				13	10	10	6	78
40 Segovia	16						4	4	15	39
41 Sevilla	52			10	20	16	10	9		117
42 Soria	17						4	6	12	39
43 Tarragona	19	13			7	10	9	10	10	78
44 Teruel	10					4	9		16	39
45 Toledo	13	13					17	19	16	78
46 Valencia	45			10	26	13	10	7	6	117
47 Valladolid	36					3	6		7	52
48 Vizcaya	29	7		13	9	7	6	7		78
49 Zamora	16					4			19	39
50 Zaragoza	59					4	6		9	78
51 Ceuta	13									13
52 Melilla	13									13
TOTAL	1.305	178	72	149	393	369	373	306	339	3.484

4.5 SELECCIÓN DE LA MUESTRA

La selección de la muestra se ha realizado de tal forma que dentro de cada estrato cualquier vivienda familiar tenga la misma probabilidad de ser seleccionada, es decir, se tengan **muestras autoponderadas dentro de cada estrato**.

Para ello, las unidades de primera etapa (secciones censales) se seleccionan con probabilidad proporcional al número de viviendas familiares principales, según los datos del último Censo o Padrón. Dentro de cada sección seleccionada en primera etapa, se selecciona un número fijo de viviendas familiares con igual probabilidad mediante la aplicación de un muestreo sistemático con arranque aleatorio. Para esta encuesta se ha determinado seleccionar 18 viviendas por sección (ver apartado 4.3)

Por tanto, la probabilidad de selección de la vivienda i, perteneciente a la sección j del estrato h, donde se han afijado K_h secciones sería

$$P(V_{ijh}) = P(S_{jh}) \times P(V_{ijh} / S_{jh}) = K_h \times \frac{V_{jh}}{V_h} \times \frac{18}{V_{jh}} = K_h \times \frac{18}{V_h}$$

siendo

$P(S_{jh})$ = Probabilidad de selección de la sección j del estrato h

$P(V_{ijh}/S_{jh})$ = Probabilidad de selección de la vivienda i condicionada a la selección de la sección j.

V_{jh} = Total de viviendas de la sección j.

V_h = Total de viviendas del estrato h.

Como se ve, esta probabilidad no depende de i ni de j.

4.6 DISTRIBUCIÓN DE LA MUESTRA EN EL TIEMPO

La distribución de la muestra es uniforme en el tiempo.

Cada período de la encuesta es de un trimestre siendo cada una de las secciones de la muestra visitada en una de las 13 semanas del mismo.

La totalidad de la muestra está dividida en tres submuestras independientes representativas, cada una de ellas, de toda la población.

4.7 TURNOS DE ROTACIÓN

Como hemos dicho en el párrafo anterior cada período de la encuesta es de un trimestre, repitiéndose ésta sucesivamente.

Las secciones censales permanecen fijas en la muestra indefinidamente (salvo las excepciones que figuran en el apartado 4.1), sin embargo las viviendas familiares son renovadas parcialmente cada trimestre de encuesta, a fin de evitar el cansancio de las familias. Esta renovación se efectúa en una sexta parte de las secciones.

A estos efectos, la muestra total se halla dividida en seis submuestras que denominamos *Turnos de rotación*. Cada sección viene identificada por un código de cinco dígitos. El último dígito nos expresa el turno de rotación a que pertenece, estando numerado del 1 al 6.

Cada trimestre se renuevan las viviendas que pertenecen a las secciones de un determinado turno de rotación. Por tanto cada vivienda pertenece a la muestra durante seis trimestres consecutivos, al cabo de los cuales sale de la misma para ser reemplazada por otra de la misma sección.

Cada trimestre las secciones del turno de rotación cuyas viviendas son entrevistadas por última vez, se actualizan con objeto de poder incorporar a la muestra en el siguiente período aquellas viviendas, tanto de nueva construcción como las que se han transformado en viviendas familiares, las cuales, cuando se realizó el último Censo o Padrón no existían o se encontraban desocupadas o destinadas a otras finalidades diferentes a la de vivienda principal.

Estas viviendas se incorporan a la muestra con una probabilidad igual a la original de las viviendas de la sección.

Se considera *Ciclo* en la encuesta a cada uno de los períodos de la misma. Éstos están numerados correlativamente desde el período de implantación. En función del número de Ciclo se puede determinar el Turno de Rotación correspondiente a las secciones cuyas viviendas se renuevan en la encuesta, mediante la siguiente ecuación:

Turno de rotación = (ȯ - número de ciclo) + 1

ȯ expresa el número múltiplo de 6 más próximo por exceso al número de ciclo

Así por ejemplo, en el primer trimestre de 2002 se realizó el ciclo número 118 de la encuesta. El múltiplo de 6 más próximo por exceso es 120, de donde

TR = (120 - 118) + 1 = 3

Quiere decir por tanto que en ese trimestre se renovaron las viviendas de las secciones que pertenecen al Turno de Rotación 3.

4.8 ESTIMADORES

Hasta el año 2001, se han utilizado **estimadores de razón** tomando como variable auxiliar las Proyecciones Demográficas de población elaboradas por el INE, siendo la expresión del estimador de una determinada característica Y en un trimestre de encuesta la siguiente:

$$\hat{Y} = \sum_{h} \frac{P_h}{p_h} \sum_{i=1}^{n_{th}} y_{hi} \qquad (1)$$

extendiéndose el sumatorio h a los estratos de una provincia, una comunidad autónoma o al total nacional, y donde:

P_h : es la proyección de la población, que reside en viviendas familiares, en el estrato h, referida a la mitad del trimestre.

p_h : es el número de personas que habitan en las viviendas de la muestra, en el estrato h, en el momento de la entrevista.

n_h : es el número de viviendas en las secciones de la muestra en el estrato h.

y_{hi} : es el valor de la característica investigada en la vivienda i-ésima, del estrato h.

A partir del primer trimestre de 2002, se aplican **Técnicas de reponderación** a los estimadores con objeto de ajustar las estimaciones de la encuesta a la información procedente de fuentes externas.

La técnica de reponderación consiste en lo siguiente:

$s = \{u_1 .... u_k ..... u_n\}$

Se considera una población $U = \{u_1 ..... u_N\}$ de la cual se extrae una muestra

La expresión (1) puede escribirse de la siguiente forma:

$$\hat{Y} = \sum_{k \in s} d_k \, y_k$$

donde:

y_k : Valor de la característica investigada en la unidad muestral k.

d_k : Factor de elevación de la unidad k obtenido mediante la expresión $\frac{P_h}{p_h}$, siendo h el estrato al que pertenece la unidad.

$\sum_{k \in s}$: Sumatorio extendido a todas las unidades de la muestra s.

Se dispone de J variables auxiliares cuyos valores son conocidos para la muestra y cuyos totales son conocidos para la población

$$X_j = \sum_{k \in U} x_{jk}$$

Se trata de encontrar un nuevo estimador

$$\hat{Y}_w = \sum_{k \in s} w_k \, y_k$$

donde los nuevos pesos w_k cumplan las siguientes condiciones:

$\forall \, j = 1 ..... J$

- Sean próximos a los pesos iniciales d_k
- Verifiquen la ecuación de equilibrado

  $$\sum_{k \in s} w_k \, x_{jk} = X_j$$

El planteamiento del problema es encontrar unos valores w_k que hagan mínima la expresión:

$$\sum_{k \in s} d_k \, G\left(\frac{w_k}{d_k}\right) \qquad \text{con la condición} \qquad \sum_{k \in s} w_k \, X_k = X$$

siendo:

G = Función de distancia.

X = Vector de dimensión (J,1) con los totales de las variables auxiliares.

X_k = Vector de dimensión (J,1) con los valores de las variables auxiliares en la unidad muestral k.

La solución del problema depende de la función de distancia G que se utilice.

Si se considera la función de distancia lineal de argumento $z = \frac{w_k}{d_k}$:

$$G(z) = \frac{1}{2}(z-1)^2, \quad z \in R$$

el problema se resuelve mediante la utilización de los multiplicadores de Lagrange que conducen a la obtención de un conjunto de factores w_k que verifican las

condiciones de equilibrado y proporcionan las mismas estimaciones que el estimador de regresión generalizada.

En el caso particular de la EPA se ha optado por utilizar la función de distancia lineal pero truncada (para evitar las soluciones negativas del sistema de ecuaciones), con objeto de aprovechar las propiedades del estimador de regresión, de pequeña varianza y mínimo sesgo.

Como variables auxiliares se han utilizado:

- Población de 16 y más años por grupos de edad y sexo a nivel de Comunidad Autónoma.
- Población de 16 y más años por provincia.

De esta forma con los estimadores actuales utilizados en la EPA se estima correctamente la población por grupo de edad y sexo.

Para la solución práctica de este problema se ha utilizado el software CALMAR (CALage sur MARges) programado por el INSEE (Institut National de la Statistique et des Études Économiques) de Francia.

5 Actualizaciones en el marco de la encuesta

Las continuas variaciones de población bien en sus características, bien en su distribución espacial exigen realizar actualizaciones en el marco que necesariamente repercuten en la estructura muestral.

En el marco de la EPA se consideran tres tipos de actualizaciones:

Actualizaciones en el marco de secciones, consecuencia de las modificaciones producidas por diversas incidencias como particiones, fusiones o variaciones de límites en las secciones seleccionadas. En cada uno de estos casos es necesario determinar la probabilidad de selección de las nuevas secciones así como el número de entrevistas a realizar en las mismas.

Actualizaciones en el marco de viviendas con carácter restringido y exclusivo para las secciones de la muestra. Esta actualización como ya se dijo en el apartado 4.7, tiene por objeto incorporar las viviendas principales *altas* de la sección en la relación de viviendas de la misma.

Actualización con carácter general relativa a todas las secciones y viviendas de la población, en la cual se actualiza la probabilidad de selección de la sección y que se realiza periódicamente cuando se dispone de la información necesaria.

5.1 INCIDENCIAS EN LAS SECCIONES DE LA MUESTRA

Se consideran los siguientes casos:

5.1.1 Partición de secciones

En el caso de una sección S en la que el crecimiento del número de viviendas principales exige que se escinda en diversas partes S_1, S_2 … S_K, bien para formar nuevas secciones o para incorporarse a otras ya existentes.

Se plantea el problema de determinar las probabilidades de selección de las nuevas secciones para conocer cual es la que va a permanecer en la muestra, así como el número de viviendas a entrevistar en la misma para que la muestra siga siendo autoponderada.

Se distinguen dos casos:

A) La sección S se fragmenta para formar dos o más secciones completas.

En este caso se opera como sigue:

1) Llamamos

V_S = Número de viviendas de la sección S según el último Censo

V'_S = Número de viviendas de la sección S después de actualizada.

V_{Sj} = Número de viviendas de la parte j de la sección S según datos del último Censo.

V'_{Sj} = Número de viviendas de la parte j de la sección S después de actualizada.

2) Se selecciona una de las nuevas secciones S_j con probabilidad proporcional a su tamaño actualizado V'_{Sj} / V'_S

3) El número de viviendas que deben ser objeto de entrevista es

$$m_j = 18 \frac{V'_s}{V_s}$$

las cuales son seleccionadas sistemáticamente.

De esta manera la muestra continúa siendo autoponderada.

B) La sección S se fragmenta para anexionarse a una o más secciones existentes.

En este caso:

1) Se selecciona uno de los fragmentos con probabilidad proporcional a su tamaño según el último Censo V_{Sj} / V_S y la nueva sección S'_j a donde se haya incorporado dicha parte quedará automáticamente seleccionada.

2) El número de viviendas que han de ser entrevistadas viene dado por

$$m_j = 18 \frac{V'_{s'_j}}{V_{s'_j}}$$

siendo

$V'_{s'j}$ = Número de viviendas principales en la actualidad en la nueva sección S'_j.

$V_{s'j}$ = Número de viviendas principales que existían en el último Censo o Padrón dentro de los límites de la nueva sección S'_j.

5.1.2 Fusión de secciones

Debido a que algunas secciones por los movimientos migratorios y naturales de la población van quedando vacías se procede a su fusión con otra u otras, de forma que en caso de ser seleccionada tengan unidades que investigar.

El caso de fusión de secciones no es sino un caso particular de la partición estudiada en el apartado 5.1.1.B.

Por tanto si la sección S_j seleccionada se fusiona con otra para formar la nueva sección S, ésta queda incorporada automáticamente a la muestra y el número de viviendas a entrevistar es:

$$m = 18 \frac{V'_S}{V_S}$$

siendo

V'_S = Número de viviendas principales en la actualidad en la nueva sección S

V_S = Número de viviendas principales, según último Censo o Padrón, dentro de los límites de la nueva sección S.

5.1.3 Variación de límites

Este es el caso de una sección que se forma con fragmentos de dos o más secciones por reajuste en sus límites.

Para el cálculo de la probabilidad de selección, este caso puede considerarse como un proceso en dos etapas: la primera de partición de cada sección y la segunda de fusión adecuada de las secciones resultantes de la partición.

En todos los casos antes expuestos, las nuevas secciones se incorporan a la muestra cuando por *Turno de rotación* corresponde renovar las familias en las secciones afectadas por dichas incidencias.

5.2 RENOVACIÓN DE LA MUESTRA COMO CONSECUENCIA DE LA ACTUALIZACIÓN DE LAS PROBABILIDADES DE SELECCIÓN

Cuando se dispone de información, bien procedente de los ficheros electorales, Censos de Población ó Padrón se procede a actualizar las probabilidades de selección de las secciones y a ajustar a 18 el número de entrevistas por sección.

Este procedimiento se realiza de tal forma que las probabilidades de selección de las secciones sean proporcionales al número de viviendas que en ese momento tenga cada una. En principio esto podría lograrse partiendo de cero y seleccionando una muestra nueva, pero ello provocaría una ruptura total con la muestra antigua, lo cual es arriesgado en el caso de encuestas continuas como es la EPA. Por ello se arbitra un procedimiento que sin distorsionar las probabilidades de selección que realmente corresponden a cada sección mantenga la muestra con las mínimas variaciones.

Este procedimiento, debido a KISH (1971), es el siguiente:

Sea S una sección perteneciente al estrato h, seleccionada en un Censo o Padrón, C, con probabilidad

$$P_S = \frac{V_S^C}{V_h^C} = \frac{\text{Viviendas en S segun Censo C}}{\text{Viviendas en el estrato h segun Censo C}}$$

y supongamos que en el siguiente Censo o Padrón, C', le corresponde una probabilidad de selección dada por

$$P'_S = \frac{V_S^{C'}}{V_h^{C'}} = \frac{\text{Viviendas en S segun Censo C'}}{\text{Viviendas en el estrato h segun Censo C'}}$$

Se compara P_S con P'_S pudiendo ocurrir uno de los dos siguientes casos:

1) Si $P'_S > P_S$ la sección S permanece en la muestra con probabilidad P'_S, ya que si fue seleccionada con una probabilidad P_S inferior a la que actualmente le corresponde, con mayor motivo hubiera salido seleccionada aplicándole su probabilidad actual P'_S.

2) Si $P'_S < P_S$ la sección permanece en la muestra con probabilidad P'_S/P_S y sale de la muestra con probabilidad $1 - P'_S / P_S$.

Este criterio motivará la salida de la muestra de un cierto número de secciones. Estas serán sustituidas por otras secciones del mismo estrato pero seleccionadas de **entre las que no perteneciendo a la muestra hayan aumentado de probabilidad**.

Con este criterio se mantiene el esquema de que la probabilidad que tiene una sección de pertenecer a la muestra es la que realmente le corresponde, es decir, proporcional al número de viviendas actuales.

III. Evaluación de la calidad de los datos

1 Introducción

Los errores que afectan a toda encuesta pueden agruparse en dos grandes grupos:

Errores de muestreo, que se originan por la obtención de resultados sobre las características de una población, a partir de la información recogida en una muestra de la misma.

Errores ajenos al muestreo, que son comunes a toda investigación estadística, tanto si la información es recogida por muestreo como si se realiza un Censo. Estos errores se presentan en cualquier fase del proceso estadístico:

- Antes de la toma de datos: por deficiencias del marco e insuficiencias en las definiciones y cuestionarios.

- Durante la toma de datos: por defectos en la labor de los entrevistadores e incorrecta declaración por parte de los informantes.

- Tras la recogida de los datos: errores en la depuración, codificación, grabación, tabulación e impresión de los resultados.

2 Errores de muestreo

Trimestralmente se calculan los errores de muestreo de las estimaciones de algunas de las principales características investigadas.

Para la obtención de los errores de muestreo se utiliza el método de las *semimuestras reiteradas*.

Este procedimiento consiste en obtener sucesivas semimuestras de la muestra inicial. A partir de cada semimuestra se calcula la estimación de la característica de la que queremos obtener el error de muestreo. Una vez calculadas todas las estimaciones con cada una de las semimuestras, así como la estimación con la muestra completa, el estimador de la varianza viene dado por:

$$\hat{V}(\hat{Y}) = \frac{1}{r}\sum_{i=1}^{r}(\hat{Y}_i - \hat{Y})^2$$

donde:

r : es el número de semimuestras obtenidas, esto es el número de reiteraciones

$\hat{Y}_i$: es la estimación obtenida con la i-ésima reiteración

Para cada reiteración se repite el proceso de estimación general, es decir, se aplica la técnica de reponderación utilizando el software CALMAR.

$\hat{Y}$: es la estimación basada en la muestra completa

En el caso de la EPA el número de reiteraciones que se utiliza es de 40. Para formarlas se procedió de la siguiente forma:

a) Se agruparon todas las secciones de cada estrato por pares, procurando que las dos secciones de cada par pertenecieran al mismo turno de rotación de la EPA.

b) Se asignó aleatoriamente la primera sección de cada par a 20 reiteraciones y la otra sección a las otras 20.

De esta forma cada reiteración queda constituida por un número de secciones equivalente al 50 por ciento de la muestra (semimuestra) y cada sección aparece en la mitad de las reiteraciones.

En las tablas se publica el error de muestreo relativo en porcentaje (coeficiente de variación), obtenido de la forma:

$$CV(\hat{Y}) = \frac{\sqrt{\hat{V}(\hat{Y})}}{\hat{Y}} . 100$$

3 Errores ajenos al muestreo

El estudio de los errores ajenos al muestreo presenta numerosas dificultades debido a la gran variedad de causas que los originan, así como a las hipótesis en que se basan los modelos teóricos que, en general, no se cumplen en la realidad, lo que lleva a obtener resultados aproximados.

En la EPA el análisis de los errores ajenos al muestreo se basa en el modelo matemático elaborado por la Oficina de Censos de los Estados Unidos debido a Hansen, Hurwitz y Bershad, y que, operativamente, consiste en repetir las entrevistas de la encuesta en una submuestra de la muestra de viviendas originalmente seleccionada. Posteriormente se cotejan los datos obtenidos en ambas ocasiones con objeto de investigar las inconsistencias y cuantificar los errores mediante la aplicación de diversos índices de calidad.

Aparte de la *entrevista repetida* se realiza un estudio específico de aquellas unidades seleccionadas que son encuestables pero que se negaron a facilitar los datos solicitados.

Para estas unidades que se niegan a contestar se cumplimenta un *cuestionario de negativas* en el que se recogen una serie de características básicas como son el sexo, la edad y la relación con la persona principal de la persona que rehusa a ser entrevistada, así como la edad, el sexo, los estudios terminados, la relación con la actividad y la rama de actividad de la persona principal.

3.1 ENCUESTA DE EVALUACIÓN

Esta encuesta se realiza con periodicidad semestral, es decir, la muestra seleccionada se extiende a lo largo de dos trimestres consecutivos.

La comparación de los resultados obtenidos en la encuesta de evaluación (entrevista repetida, ER) con los obtenidos en la entrevista original (EO) permite evaluar dos grandes tipos de errores ajenos al muestreo.

a) Errores de cobertura, producidos por la omisión o por la inclusión errónea de unidades en la encuesta original.

b) Errores de contenido, que afectan a las características investigadas en las personas encuestables.

El trabajo de campo se lleva a cabo por agentes especializados, los cuales realizan la entrevista repetida a lo sumo 15 días después de la original, refiriéndose los datos de ambas entrevistas al mismo período de tiempo.

Teniendo en cuenta el doble objetivo perseguido con la encuesta de evaluación, evaluar la calidad de los resultados y controlar el trabajo de los agentes que intervienen en la EPA, la muestra de secciones en las que se realiza la segunda entrevista se subdivide en dos submuestras:

Submuestra A: Para la selección de esta submuestra se forman con todos los bloques de la encuesta, salvo los correspondientes a Ceuta y Melilla, 26 itinerarios de 10 bloques aproximadamente cada uno. En cada semana, de las veintiséis en que se considera dividido el semestre, se selecciona aleatoriamente uno de estos itinerarios visitándose la sección correspondiente de cada bloque.

Cada semana se investigan pues, dentro de la submuestra A unas 10 secciones, correspondientes a bloques diferentes. Como la selección de itinerarios es sin reemplazamiento y estos contienen todos los bloques de que consta la EPA, esto es, todos los agentes de EO, al finalizar el semestre se ha investigado una sección de cada entrevistador.

Submuestra B: Para la selección de la submuestra B se forman con los bloques que cubren el ámbito peninsular de la encuesta, es decir, eliminando los correspondientes a las provincias insulares y a Ceuta y Melilla, 81 zonas de tres bloques aproximadamente cada una. Cada semana del semestre se selecciona aleatoriamente con reemplazamiento una de ellas, exceptuándose para la selección de cada semana aquellas que tengan algún bloque seleccionado en la submuestra A, visitándose igualmente una sección de cada uno de los tres bloques de la zona seleccionada.

En total se investigan aproximadamente 340 secciones cada semestre.

En las secciones seleccionadas se repite la entrevista en la mitad de las viviendas encuestadas en la EO.

3.2 ERRORES DE COBERTURA

Con la comparación de los resultados obtenidos en ambas entrevistas se obtienen indicadores sobre la cobertura de viviendas, la de personas así como indicadores sobre los errores de contenido.

Cobertura de viviendas: se obtienen las viviendas que son encuestables en ambas entrevistas, las encuestables en ER y no en EO y viceversa.

Cobertura de personas: para estudiar los errores en la cobertura de personas, estas se clasifican en:

- Personas cotejables, son aquellas que ambos agentes han considerado encuestables.

- Personas omitidas, son aquellas cuyos datos ha recogido el agente de ER por considerarlas encuestables, pero de las que no existe información en la EO.

- Personas erróneamente incluidas, que figuran en la EO pero no en la ER por considerar el agente de entrevista repetida que no eran encuestables.

3.3 ERRORES DE CONTENIDO

Los datos sobre errores de contenido se basan en la información suministrada en las dos entrevistas por las personas clasificadas como cotejables.

Así para una característica C con las modalidades M_1, ..., M_k, una persona cotejable puede incluirse en una tabla con el siguiente formato:

E.O. / E.R.	Total	M_1	M_2	. . .	M_j	. . .	M_k
Total personas	**n**	**$n_{.1}$**	**$n_{.2}$**	**. . .**	**$n_{.j}$**	**. . .**	**$n_{.k}$**
M_1	**$n_{1.}$**	n_{11}	n_{12}	. . .	n_{1j}	. . .	n_{1k}
M_2	**$n_{2.}$**	n_{21}	n_{22}	. . .	n_{2j}	. . .	n_{2k}
. .	**. . .**	. . .	. . .	. . .	. . .	. . .	. . .
. .	**. . .**	. . .	. . .	. . .	. . .	. . .	. . .
. .	**. . .**	. . .	. . .	. . .	. . .	. . .	. . .
M_i	**$n_{i.}$**	n_{i1}	n_{i2}	. . .	n_{ij}	. . .	n_{ik}
. .	**. . .**	. . .	. . .	. . .	. . .	. . .	. . .
. .	**. . .**	. . .	. . .	. . .	. . .	. . .	. . .
. .	**. . .**	. . .	. . .	. . .	. . .	. . .	. . .
M_k	**$n_{k.}$**	n_{k1}	n_{k2}	. . .	n_{kj}	. . .	n_{kk}

n_{ij} representa el número de personas clasificadas en la modalidad M_i según la ER y en la M_j según la EO.

La diagonal principal (n_{ii}) representa el número de personas que han sido idénticamente clasificadas en ambas entrevistas.

Para cada modalidad M_i de la característica C se puede obtener la siguiente tabla:

E.R. \ E.O.	Con la Modalidad M_i	Sin la Modalidad M_i	Total
Con la Modalidad M_i	a	b	a + b
Sin la Modalidad M_i	c	d	c + d
TOTAL	a + c	b + d	n

Comparando con la tabla anterior tenemos las siguientes equivalencias:

$a = n_{ii}$ número de personas clasificadas con la modalidad M_i en ambas entrevistas.

$b = n_{i.} - n_{ii}$ número de personas clasificadas con la modalidad M_i en ER y con otra diferente en EO.

$c = n_{.i} - n_{ii}$ número de personas clasificadas con la modalidad M_i en EO y con otra distinta en ER.

$d = n - n_{i.} - n_{.i} + n_{ii}$ número de personas que se han clasificado en modalidad diferente a la de M_i en ambas entrevistas.

$n = a + b + c + d$ es el total de personas que se han clasificado en ambas entrevistas respecto de la característica C estudiada.

En base a estas tablas reducidas se definen los siguientes indicadores de calidad para la modalidad M_i:

a) Porcentaje de idénticamente clasificados

$$P.I.C.(M_i) = \frac{a}{a+b} \times 100 = \frac{n_{ii}}{n_{i.}} \times 100$$

Varía entre cero y cien. Es un indicador de la estabilidad de respuesta. Su valor óptimo (100) expresa que todas las personas pertenecientes según la ER a la modalidad M_i se clasificaron de igual forma en la EO.

b) Indice de cambio neto

$$I.C.N.(M_i) = \frac{c-b}{a+b} \times 100 = \frac{n_{.i} - n_{i.}}{n_{i.}} \times 100$$

Puede ser positivo ($c > b$ o $n_{.i} > n_{i.}$) o negativo ($b > c$ o $n_{i.} > n_{.i}$). Es un indicador del sesgo de respuesta, expresado como porcentaje del número de personas pertenecientes a M_i según la ER.

c) Tasa de diferencia neta

$$T.D.N.(M_i) = \frac{c-b}{n} \times 100 = \frac{n_{.i} - n_{i.}}{n} \times 100$$

Similar al anterior, pero en este caso es un porcentaje respecto al total de personas que se han clasificado en ambas entrevistas respecto a la característica de referencia.

d) Indice de cambio bruto

$$I.C.B.(M_i) = \frac{c + b}{a + b} \times 100 = \frac{n_{.i} + n_{i.} - 2n_{ii}}{n_{i.}} \times 100$$

Puede ser nulo o positivo. Es un indicador de la varianza de respuesta.

e) Tasa de diferencia bruta

$$T.D.B.(M_i) = \frac{c + b}{n} \times 100 = \frac{n_{.i} + n_{i.} - 2n_{ii}}{n} \times 100$$

Para comparar la calidad general de las distintas características evaluadas se utiliza el **índice de consistencia global**, obtenido a partir de la tabla en la que aparecen todas las modalidades de la característica C. Se define como

$$I.C.G.(C) = \frac{\sum_{i=1}^{k} n_{ii}}{n} \times 100$$

El valor que expresa la inexistencia de error es el 100.