

# NOTAS EXPLICATIVAS SOBRE LA REALIZACION DE LA EPA

Fernando FERNANDEZ MENDEZ DE ANDES  
Pilar MARCOS DOMINGUEZ

## I. INTRODUCCION (\*)

**L**OS resultados de cualquier política económica dependen, en última instancia, de la calidad de los datos con que se trabaja. De la misma manera, la verificación de cualquier hipótesis económica se apoya en una adecuada selección e identificación de la base estadística relevante, y en la calidad intrínseca de la misma.

No creemos exagerado afirmar que la fuente original de datos más utilizada en nuestro país en los últimos años ha sido, sin discusión, la Encuesta de Población Activa (EPA) del Instituto Nacional de Estadística. Y ello aunque sólo sea por la magnitud y gravedad del desempleo. La comparación de sus resultados, de origen muestral, con las cifras del paro registrado, de origen administrativo, son tan abundantes como polémicas y hasta han dado origen a interesantes artículos periodísticos. La discusión sobre aspectos técnicos de ambas fuentes de datos ha sido minuciosa y ha arrojado suficiente luz al respecto (1).

No es esa nuestra intención aquí. Son, sin embargo, numerosos los economistas, políticos y otros profesionales que utilizan habitualmente la EPA, y que a menudo se preguntan por el mecanismo concreto de recogida de datos, por el diseño de la muestra y por otras cuestiones prácticas relacionadas con la información en ella contenida.

Esta nota pretende aclarar parte de esos interrogantes y explicar de manera comprensible al público general y no especializado lo que se esconde detrás de las bambalinas de esta herramienta de trabajo tan útil y de uso tan generalizado.

El INE recoge información de una muestra y extrapola los resultados obtenidos al total de la población; los publica trimestralmente (en forma de avances) y anualmente (con resultados definitivos), junto con una evaluación de la calidad de los datos publicados trimestralmente. Además, el INE ha publicado varios trabajos sobre la meto-

dología utilizada en la elaboración de la muestra, la realización de la encuesta y los indicadores de calidad de los datos; documentos todos ellos que se recogen en la bibliografía.

En concreto, pretendemos aquí responder a las tres preguntas siguientes:

¿Cómo se obtiene la muestra EPA?

¿Cómo, y a quiénes, se realizan efectivamente las entrevistas y con qué periodicidad?

¿Cuál es el grado de fiabilidad de los datos y cómo lo cuantifica y controla el INE?

## II. LA MUESTRA

Cada diez años, el censo de la población recoge información exhaustiva (edad, sexo, nivel de educación, situación laboral, etc.) de toda la población española.

Este estudio es, evidentemente, muy costoso y exige una gran dedicación humana y de tiempo; por lo que resulta inviable repetirlo en períodos más cortos. Pero, obviamente, la información que proporciona es claramente insuficiente, por el desfase temporal entre censos, para la adecuada instrumentación y seguimiento de la política económica.

Un estudio muestral, por el contrario, permite recoger información periódica (en la EPA mensual) sobre la población estudiada a un menor coste, aunque con el error propio de todo estudio muestral, que depende del subconjunto poblacional elegido.

La muestra EPA se selecciona, en dos etapas, de la siguiente forma:

1. En una primera etapa, el territorio nacional (excepto Ceuta y Melilla) se divide en cincuenta provincias (cuarenta y siete peninsulares y tres insulares). Cada provincia consta de  $n$  municipios. Cada municipio de uno o varios distritos municipales, y cada distrito de cierto número de secciones censales (una sección censal es un subconjunto poblacional de entre 500 y 2.000 electores) (2).

Los municipios se clasifican en nueve estratos (3), en función de la probabilidad de que tengan alguna sección censal presente en la muestra (ver cuadro n.º 1). Los estratos se reagrupan en siete substratos, en función del tipo de actividad económica predominante en su sección, quedando así definidos siete substratos; los tres primeros, en los que predomina una sola actividad, otros tres en los que existen dos actividades preponderantes y un séptimo con igual importancia de las tres actividades consideradas. Es curioso señalar que las actividades económicas que sirven de clasificación no son las habituales, salvo en el caso de la agricultura, pues las otras dos combinan industria y servicios y se dividen por la categoría profesional dominante.

La selección de las secciones dentro de cada substrato se realiza con probabilidad proporcional al número de viviendas principales existentes.

Además, en cada provincia, el número de secciones por estrato es proporcional a la población de la provincia, con las siguientes restricciones:

1) El número total de secciones de la muestra en la provincia debe ser múltiplo de doce (4), y

2) Todas las provincias han de contar con un mínimo de tres y un máximo de doce bloques en la muestra.

Con esto se consigue que todas las provincias tengan una muestra representativa, que los resultados nacionales no estén sesgados y que cada agente entrevistador se haga cargo de un bloque por trimestre.

De las 32.166 secciones censales existentes (32.234 con Ceuta y Melilla) se seleccionan 3.144, que se mantienen fijas, salvo que:

a) Se agoten los hogares encuestables en alguna sección, en cuyo caso se cambia por otra.

b) Trimestralmente se observe que el aumento o disminución de población de alguna(s) sección(es) aconseja:

b.1) Partirla en varias, y seleccionar aleatoriamente una de las particiones, o

b.2) Fusionarla con otra(s) ajena(s) a la muestra, quedando automáticamente seleccionada.

c) El resultado de actualizar las probabilidades de las secciones con el nuevo censo o padrón aconsejen la sustitución de alguna sección (5).

2. Una vez seleccionadas las sec-

CUADRO N.º 1

NUMERO DE SECCIONES DE LA MUESTRA POR PROVINCIA Y ESTRATO

PROVINCIAS	1	2	5	6	7	8	9	Total
1. Alava ...	24	—	—	—	—	—	12	36
2. Albacete ...	15	—	9	—	—	—	12	36
3. Alicante ...	18	12	12	12	—	—	18	72
4. Almería ...	12	—	12	—	—	—	12	36
5. Ávila ...	12	—	—	—	—	—	24	36
6. Badajoz ...	24	—	12	12	—	—	24	72
7. Baleares ...	30	—	12	12	—	—	18	72
8. Barcelona ...	72	—	12	12	—	30	18	144
9. Burgos ...	15	—	9	—	—	—	12	36
10. Cáceres ...	18	—	18	—	—	—	36	72
11. Cádiz ...	24	12	12	12	—	—	12	72
12. Castellón ...	24	—	18	12	—	—	18	72
13. Ciudad Real ...	12	12	12	18	—	—	18	72
14. Córdoba ...	24	—	24	—	—	—	24	72
15. Coruña, La ...	24	—	12	—	—	—	36	72
16. Cuenca ...	12	—	—	—	—	—	24	36
17. Gerona ...	15	—	12	15	—	—	30	72
18. Granada ...	24	—	12	—	—	12	24	72
19. Guadalajara ...	12	—	—	—	—	—	24	36
20. Guipúzcoa ...	24	—	24	—	—	—	24	72
21. Huelva ...	12	—	12	—	—	—	12	36
22. Huesca ...	12	—	12	—	—	—	12	36
23. Jaén ...	15	9	12	12	—	—	24	72
24. León ...	24	12	12	—	—	—	24	72
25. Lérida ...	12	—	—	—	—	—	24	36
26. Logroño ...	15	—	9	—	—	—	12	36
27. Lugo ...	12	—	12	—	—	—	12	36
28. Madrid ...	120	—	—	—	—	15	9	144
29. Málaga ...	36	—	12	12	—	—	12	72
30. Murcia ...	24	12	24	—	—	—	12	72
31. Navarra ...	30	—	12	—	—	—	30	72
32. Orense ...	12	—	—	—	—	—	24	36
33. Oviedo ...	18	18	15	24	15	—	18	108
34. Palencia ...	18	—	—	—	—	—	18	36
35. Palmas, Las ...	36	—	12	12	—	—	12	72
36. Pontevedra ...	12	24	12	—	—	—	24	72
37. Salamanca ...	15	—	9	—	—	—	12	36
38. Santa Cruz de Tenerife ...	24	12	12	12	—	—	12	72
39. Santander ...	24	12	18	—	—	—	18	72
40. Segovia ...	12	—	—	—	—	—	24	36
41. Sevilla ...	48	—	18	18	—	9	15	108
42. Soria ...	12	—	—	—	—	—	24	36
43. Tarragona ...	18	—	24	12	—	—	18	72
44. Teruel ...	12	—	—	—	—	—	24	36
45. Toledo ...	18	12	12	—	—	—	30	72
46. Valencia ...	48	—	18	—	—	18	24	108
47. Valladolid ...	24	—	—	—	—	—	12	36
48. Vizcaya ...	36	9	—	—	—	15	12	72
49. Zamora ...	12	—	—	—	—	—	24	36
50. Zaragoza ...	48	—	12	—	—	—	12	72
TOTAL ...	1.194	156	501	207	15	99	960	3.132

Fuente: Encuesta de Población Activa: Informe técnico, diseño de la encuesta y evaluación de la calidad de los datos.

ciones muestrales, se procede a la elección de las viviendas que han de ser encuestadas en cada sección.

De cada sección se extraen veinte viviendas principales mediante arranque aleatorio, y con un factor de elevación aproximadamente igual al período inverso de la fracción de muestreo (6).

Las secciones se clasifican en seis grupos (o turnos de rotación) iguales; y cada trimestre se cambian todas las viviendas muestrales del turno  $i$  ( $i = 1 \dots 6$ ) (7).

De esta manera se consigue que cada vivienda sea entrevistada en seis ocasiones (durante seis trimestres consecutivos).

Aunque la encuesta es de carácter trimestral, el INE obtiene datos mensuales al dividir la muestra en tres submuestras de igual número de estratos por sección, que permiten, por extrapolación, realizar estimaciones para toda la población muestral y nacional (ver cuadro n.º 2).

El elemento muestral de la EPA es la vivienda principal. Se entrevista al grupo humano que la ha habitado durante los últimos seis meses (o que viviendo en ella menos de seis meses no tiene otra, o piensa vivir en ésta aun teniendo otra). Es decir, a todo aquel grupo humano que afirma ser esa su vivienda principal.

La encuesta recoge datos de todas las personas presentes en la vivienda (incluso cuando la hayan habitado menos de tres meses, si no ocupan otra). También se recogen datos de las personas ausentes de la vivienda durante menos de tres meses, siempre que no estén ocupando otra (8) (ver cuadro número 3).

### III. LA ENTREVISTA

El agente entrevistador recoge la información detallada en el cuestionario EPA sobre las personas de la vivienda encuestada.

Cada sección se entrevista en una semana, y el cuestionario debe responderse de acuerdo con la situación laboral de los miembros de la vivienda en la semana inmediatamente anterior a la de la entrevista.

Cada agente entrevista un bloque por trimestre y cada trimestre se sustituye 1/6 de las viviendas (siguiendo el turno de rotación preestablecido) por otras seleccionadas dentro de la misma sección.

Los encuestadores son personal fijo del INE, con un nivel de estudios al menos equivalente al graduado escolar y que reciben un cursillo de preparación. Además, no parece existir un grupo social predominante entre los encuestadores por ninguna otra característica potencialmente relevante: sexo, edad, etc.

Cada tres entrevistadores dependen de un jefe de grupo con más experiencia, y todos ellos son supervisados por un inspector de encuestas (perteneciente al cuerpo de Estadísticos Técnicos Diplomados de su provincia).

La labor del agente entrevistador es conseguir que alguno de los miembros de la vivienda mayores de 16 años (o todos) faciliten la información solicitada en el cuestionario EPA, relativa a la situación laboral de todas las personas de la vivienda mayores de 14 años en la semana anterior a la de la entrevista, y durante los seis trimestres consecuti-

vos en los que la vivienda permanece en la muestra.

Para conseguir su objetivo, el entrevistador normalmente acude a la vivienda seleccionada en horas en las que la probabilidad de encontrar más de un miembro adulto presente, si existe, sea elevada. Así se intentan evitar las horas de concentración de la jornada laboral. A este respecto, es interesante resaltar que la persona, o personas, que contesta(n) por todo el grupo humano de una vivienda no tiene(n) que ser necesariamente la(s) misma(s) durante las seis veces que ésta permanece en la muestra (esto puede ser una fuente de los errores que se comentan más adelante).

Puede ocurrir que el entrevistador no consiga obtener información del grupo humano de la vivienda por ausencia permanente de ésta o por negativa de sus miembros a colaborar con el INE, pues el cuestionario debe conseguirse completo en un máximo de seis semanas.

Si la ausencia o negativa se produce durante la primera entrevista, la vivienda se sustituye por el primer suplente aleatorio de reserva (y así sucesivamente de repetirse el incidente), dejando constancia de la vivienda que no se entrevistó, cumplimentando un cuestionario de datos generales (9).

Si la ausencia se produce en las entrevistas segunda o siguientes, el entrevistador rellena los datos generales y el INE repite la información del trimestre anterior, durante un período.

Si la ausencia es permanente, no se repite la información, y la muestra queda disminuida en un elemento.

Si la negativa se produce en entre-

CUADRO N.º 2

#### DISTRIBUCION DEL NUMERO DE SECCIONES POR PERIODO Y ESTRATO

PERIODO	ESTRATOS							Total
	1	2	5	6	7	8	9	
1.ª submuestra ...	398	52	167	69	5	33	320	1.044
2.ª submuestra ...	398	52	167	69	5	33	320	1.044
3.ª submuestra ...	398	52	167	69	5	33	320	1.044
MUESTRA ...	1.194	156	501	207	15	99	960	3.132

Fuente: Encuesta de Población Activa: Informe técnico, diseño de la encuesta y evaluación de la calidad de los datos.

CUADRO N.º 3

CASOS POSIBLES PARA QUE UNA PERSONA SEA O NO ENCUESTADA EN LA EPA

Situación de residencia	Tiempo en meses		Ocupación de otra vivienda permanente	E o NE		EXCEPCIONES
	<	3 o más		E	NE	
P	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	SI	<input checked="" type="checkbox"/>	<input type="checkbox"/>	• Trabajadores internos en una vivienda familiar.
A	<input type="checkbox"/>	<input type="checkbox"/>	NO	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
P	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	SI	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
A	<input type="checkbox"/>	<input type="checkbox"/>	NO	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
P	<input checked="" type="checkbox"/>	<input type="checkbox"/>	SI	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
A	<input type="checkbox"/>	<input checked="" type="checkbox"/>	NO	<input type="checkbox"/>	<input type="checkbox"/>	
P	<input checked="" type="checkbox"/>	<input type="checkbox"/>	SI	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
A	<input type="checkbox"/>	<input checked="" type="checkbox"/>	NO	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
P	<input type="checkbox"/>	<input checked="" type="checkbox"/>	SI	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
A	<input checked="" type="checkbox"/>	<input type="checkbox"/>	NO	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
P	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	SI	<input type="checkbox"/>	<input checked="" type="checkbox"/>	• Trabajadores internos en una vivienda familiar.
A	<input type="checkbox"/>	<input type="checkbox"/>	NO	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
P	<input type="checkbox"/>	<input type="checkbox"/>	SI	<input type="checkbox"/>	<input checked="" type="checkbox"/>	• Hospitalizados, estudiantes, servicio militar y trabajadores en el extranjero, etc.
A	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	NO	<input checked="" type="checkbox"/>	<input type="checkbox"/>	

donde P: Presentes.  
A: Ausentes.  
E: Encuestada.  
NE: No Encuestada.

Fuente: Encuesta de Población Activa. Manual de definiciones: Instrucciones y normas de codificación.

vistas posteriores a la primera, tampoco se sustituye la vivienda, se envía el cuestionario de negativas y se sigue visitando la vivienda (por si la negativa fuera circunstancial). Caso de que la negativa se repita por segunda vez, se extrae la vivienda de la muestra (es decir, la muestra queda asimismo disminuida en un elemento) (ver gráfico 1).

Por lo que se refiere al cuestionario, éste ha sufrido dos modificaciones (en 1972 y 1976), desde su versión original en 1964, como resultado de reformas en la metodología de la encuesta. Ambas modificaciones perseguían profundizar en la situación de desempleo y actividad marginal, convirtiendo lo que era originalmente una encuesta de actividad en un instrumento que permitiese recoger de manera fiel los cambios respecto a la situación de empleo, desempleo y a las distintas situaciones

intermedias, que se estaban produciendo en el mercado de trabajo.

En la actualidad, parece que el INE proyecta una nueva reforma del mismo, creemos que en fase de discusión previa, para hacer frente a los cambios habidos en la naturaleza y características de la actividad laboral de la población española.

De momento, y por un período de tiempo no previsible, sigue en vigor el cuestionario de 1976. Consta de 37 preguntas, con las que se pretende clasificar la actividad de todos los miembros mayores de catorce años.

Las nueve primeras son de carácter general. La décima (ver cuadro n.º 4) hace una primera clasificación de los encuestados en relación con su actividad económica.

El encuestado debe elegir una (y sólo

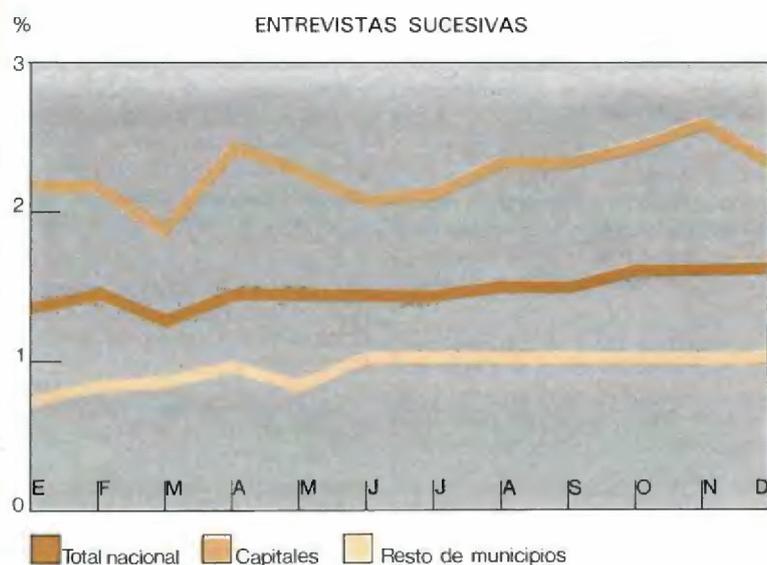
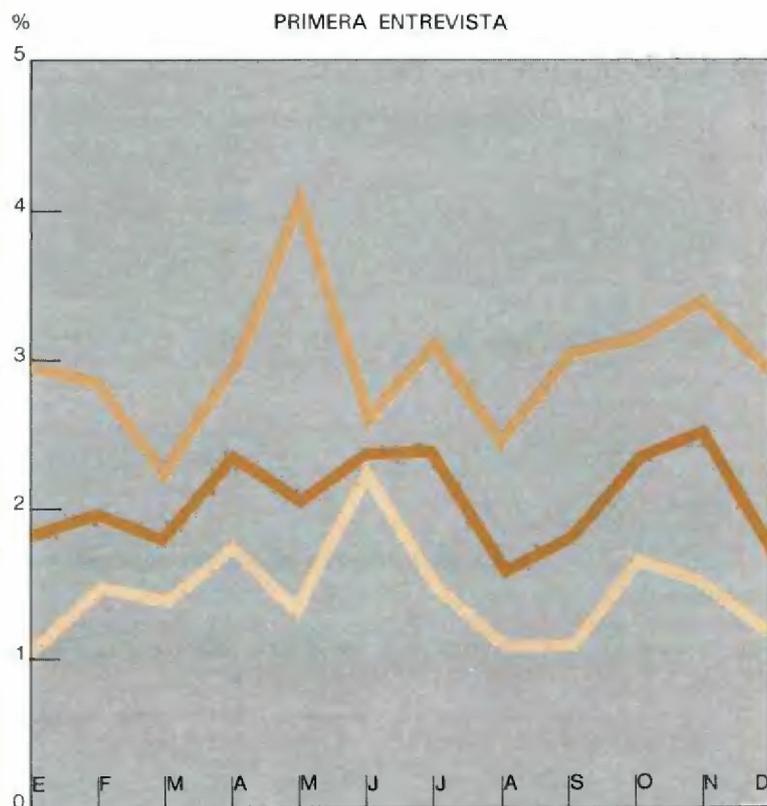
una) situación de las detalladas en la pregunta. Como quiera que no son totalmente excluyentes, el entrevistado puede dudar entre dos o más respuestas, en cuyo caso deberá optar por la enunciada en primer lugar (10).

A partir de esa clasificación inicial, las restantes preguntas intentan ampliar y concretar la información referente al tipo y grado de actividad del encuestado (11).

En lo posible, las respuestas están codificadas (o se codifican posteriormente a la entrevista).

También es labor del entrevistador evitar cualquier incoherencia aparente en las respuestas del encuestado, en cada una de las entrevistas y entre entrevistas sucesivas (12), realizando un seguimiento (en el campo) de los cambios de actividad de los encuestados.

**GRAFICO 1  
EVOLUCION DE LAS ENTREVISTAS  
NEGATIVAS**



Fuente: Evaluación de la calidad de los datos en la Encuesta de Población Activa.

Un seguimiento dinámico de los cambios de actividad de un colectivo de referencia sería muy interesante para analizar cuestiones como movilidad laboral o duración del desempleo, desagregadas por diferentes características personales. Un análisis de este tipo está en proyecto en la EPA, pero hasta ahora no se realiza.

#### IV. LA CALIDAD Y FIABILIDAD DE LOS DATOS EPA

Los datos obtenidos mensual (submuestra) y trimestralmente (muestra) son una *estimación* del valor de las características estudiadas (no el verdadero valor de los parámetros) ya que la estimación está sujeta a dos tipos de errores:

1) *Error de muestreo*: Propio de este tipo de estudio y que, en el caso de los estimadores insesgados, es igual al coeficiente de variación del estimador.

Para hallar el coeficiente de variación necesitamos conocer la varianza de los estimadores EPA, que calculamos por el método de las semimuestras reiteradas.

Se forman un cierto número ( $n = 40$ ) de semimuestras aleatorias, de tal forma que:

a) Cada una contenga la mitad de la muestra, y

b) Sean representativas de la muestra completa.

En cada semimuestra (reiteración) se estima el parámetro y se calcula la varianza entre las estimaciones de las reiteraciones y de la muestra completa (13).

A partir de la varianza del estimador se calcula su coeficiente de variación, y se publica (para todas las características estudiadas) anualmente por el INE en su estudio de evaluación de calidad.

La teoría elemental de muestras determina que, a partir de la estimación y del error de muestreo (expresado como porcentaje de la estimación por el coeficiente de variación), se puede calcular un intervalo que, con un nivel de confianza del  $\alpha$  por 100 (14) contendrá al verdadero valor del parámetro. Así, cuanto menor sea el coeficiente de variación más precisa será la estimación (al ser menor la longitud del intervalo de confianza). Ver, como ejemplo, el cuadro n.º 5, en el que se especifican intervalos de confianza para

CUADRO N.º 4

CUESTIONARIO EPA: PREGUNTA NUMERO 10

10. ¿En cuál de las situaciones siguientes se encontraba la semana pasada?

- |  |    |                          |              |
|--|----|--------------------------|--------------|
| a) Está cumpliendo el servicio militar ...             | 01 | <input type="checkbox"/> | Pasar a 13   |
| b) Es ayuda familiar                                   |    |                          |              |
| - Trabajó como tal ...                                 | 02 | <input type="checkbox"/> | Pasar a 19   |
| - No trabajó como tal ...                              | 03 | <input type="checkbox"/> | Pasar a 17   |
| c) Tiene empleo continuado                             |    |                          |              |
| - Trabajó { en su empleo ...                           | 04 | <input type="checkbox"/> | } Pasar a 22 |
| fuera de su empleo ...                                 | 05 | <input type="checkbox"/> |              |
| - No trabajó ...                                       | 06 | <input type="checkbox"/> | Pasar a 11   |
| d) No tiene empleo continuado                          |    |                          |              |
| - Trabajó { ocasionalmente ...                         | 07 | <input type="checkbox"/> | } Pasar a 19 |
| de forma estacional ...                                | 08 | <input type="checkbox"/> |              |
| - No trabajó:  |    |                          |              |
| - Está disponible y busca empleo ...                   | 09 | <input type="checkbox"/> | } Pasar a 13 |
| - Está disponible y no busca empleo ...                | 10 | <input type="checkbox"/> |              |
| - Es escolar, estudiante u opositor ...                | 11 | <input type="checkbox"/> |              |
| - Es jubilado, rentista o pensionista ...              | 12 | <input type="checkbox"/> |              |
| - Se dedica exclusivamente a las labores del hogar ... | 13 | <input type="checkbox"/> |              |
| - Estuvo en otra situación (Especificar) ...           | 14 | <input type="checkbox"/> |              |
| .....  |    |                          |              |
| e) Está trabajando en el extranjero ...                | 15 | <input type="checkbox"/> | Pasar a 37   |
| f) Está incapacitado permanentemente ...               | 16 | <input type="checkbox"/> | Fin encuesta |

Fuente: Encuesta de Población Activa, Manual de definiciones: Instrucciones y normas de codificación.

las variables (agregadas) más importantes: ocupación, actividad, inactividad y paro, en los últimos trimestres disponibles.

La amplitud de estos intervalos, en el caso de elevadas probabilidades, debe invitarnos a reflexionar sobre la validez relativa (en términos probabilísticos) de las estimaciones *puntuales* que aparecen en los medios de comunicación durante el seguimiento de la evolución de estas (y otras) variables económicas y de las conclusiones que, en ocasiones, se extraen del dato. A título ilustrativo, en el segundo trimestre de 1985, con una probabilidad del 95 por 100, la población activa estaba situada

entre 13.047 y 13.345 miles de personas (la amplitud del intervalo es pues de 298 mil personas), la población ocupada se cifraba entre 9.990 y 10.456 miles de personas (diferencia de 466 mil empleos) y el paro se hallaba entre 2.783 y 3.000 miles de personas (diferencia de 217 mil parados). En tal contexto, puede quedar desvirtuado el énfasis que muchas veces se da a escasas oscilaciones de estas variables.

2) *Errores ajenos al muestreo*: Este tipo de error es propio de todos los estudios estadísticos, y puede producirse en cualquiera de sus fases: antes, durante y después de la recogida de información, por deficiencias en la la-

bor de los investigadores o declaraciones erróneas de los encuestados, deficiencias en las definiciones y/o en los cuestionarios; o por error en la codificación y elaboración de las tablas de resultados.

Con objeto de captar estos errores y evaluar su importancia, la EPA realiza una encuesta de evaluación, que consiste en repetir la encuesta a una pequeña submuestra aleatoria de las viviendas originales (en torno al 5 por 100), en un momento próximo al de realización de la encuesta original, y cotejar las posibles diferencias de respuesta.

Se supone (15) que en la entrevista

repetida se obtienen los verdaderos valores de las características estudiadas.

La comparación de los resultados de ambas encuestas (encuesta original y encuesta repetida) permite evaluar dos tipos de errores ajenos al muestreo: de cobertura y de contenido.

1) *Errores de cobertura:* Se producen por la inclusión u omisión errónea de unidades (viviendas y/o personas) en la encuesta original.

2) *Errores de contenido:* Consisten en clasificar de forma distinta a una persona en las entrevistas original y repetida. Se estudia mediante las tablas de concordancia (ver cuadro n.º 6), y mediante una serie de indicadores de calidad como los siguientes:

— Porcentaje de idénticamente clasificados

$$PIC = \frac{a}{a + b} \cdot 100, \quad 0 \leq PIC \leq 100$$

— Tasa de diferencia bruta

$$tdb = \frac{c + b}{n} \cdot 100, \quad tdb \geq 0$$

es un indicador de la varianza de la respuesta, o

— Índice de consistencia global

$$ICG = \frac{\sum a_i}{n} \cdot 100, \quad 0 \leq ICG \leq 100,$$

donde  $a_i$  es el número de personas con igual clasificación en ambas entrevistas; si  $ICG = 100$  todas las características han sido medidas correctamente.

Los resultados de la Encuesta de Evaluación, junto con los indicadores de calidad, se publican paralelamente (con el lógico desfase temporal) a los resultados EPA, pero, en la práctica, suelen ser obviados (¿olvidados?) por ser propios de todos los estudios estadísticos. Quizá convendría recordarlos en un momento en el que se está modifican-

do el diseño de la muestra, e intentar aplicar todas las mejoras técnicas que contribuyan a minimizarlos. Y así conseguir mejorar, día a día, la calidad de los datos utilizados en la investigación económica (y, por ende, la calidad de esa investigación).

## V. NOTA FINAL

Sería pretencioso, en un estudio como el que hemos realizado, hablar de las conclusiones obtenidas. En realidad, se trataba de explicar los pasos necesarios para que el público en general pueda disponer de una encuesta de situación laboral y contribuir a su mejor conocimiento. Hemos estudiado el proceso de obtención de la información contenida en la EPA, sin analizar las necesidades adicionales de la misma o las mejoras posibles en su presentación y aprovechamiento.

CUADRO N.º 5 (\*)

(5.1)	POBLACION ACTIVA (MILES)			
	3.º trimestre 1984 (J.A.S.)	4.º trimestre 1984 (O.N.D.)	1.º trimestre 1985 (E.F.M.)	2.º trimestre 1985 (A.M.J.)
$\alpha = 67\%$ $\hat{X} \pm CV$	(13.148,20; 13.240,48)	(13.183,23; 13.273,17)	(13.203,62; 13.296,37)	(13.047,58; 13.345,82)
$\alpha = 95\%$ $\hat{X} \pm 2 CV$	(13.101,94; 13.286,66)	(13.138,26; 13.318,14)	(13.157,25; 13.342,75)	(12.898,46; 13.492,24)
$\alpha = 99\%$ $\hat{X} \pm 2,5 CV$	(13.078,85; 13.309,75)	(13.115,77; 13.363,11)	(13.134,06; 13.365,93)	(12.823,90; 13.569,50)
$\alpha = 99,7\%$ $\hat{X} \pm 3 CV$	(13.055,76; 13.332,84)	(13.093,29; 13.363,11)	(13.110,87; 13.389,12)	(12.749,34; 13.644,06)
	OCUPADOS (MILES)			
(5.2)	3.º trimestre 1984 (J.A.S.)	4.º trimestre 1984 (O.N.D.)	1.º trimestre 1985 (E.F.M.)	2.º trimestre 1985 (A.M.J.)
$\alpha = 67\%$ $\hat{X} \pm CV$	(10.360,14; 10.441,26)	(10.223,89; 10.316,31)	(10.177,59; 10.277,81)	(10.106,56; 10.339,64)
$\alpha = 95\%$ $\hat{X} \pm 2 CV$	(10.319,58; 10.481,82)	(10.177,67; 10.362,53)	(10.127,48; 10.327,92)	(9.990,02; 10.456,18)
$\alpha = 99\%$ $\hat{X} \pm 2,5 CV$	(10.299,30; 10.502,10)	(10.154,57; 10.385,63)	(10.102,43; 10.352,97)	(9.931,75; 10.514,45)
$\alpha = 99,7\%$ $\hat{X} \pm 3 CV$	(10.279,02; 10.522,38)	(10.131,46; 10.408,74)	(10.077,37; 10.378,03)	(9.873,47; 10.572,73)

CUADRO N.º 5 (\*) (continuación)

( 5. 3 )	PARADOS (MILES)			
	3.º trimestre 1984 (J.A.S.)	4.º trimestre 1984 (O.N.D.)	1.º trimestre 1985 (E.F.M.)	2.º trimestre 1985 (A.M.J.)
$\alpha = 67\%$ ... $\hat{X} \pm CV$ ...	(2.675,71; 2.745,09)	(2.831,62; 2.906,78)	(2.887,95; 2.961,05)	(2.837,83; 2.946,57)
$\alpha = 95\%$ ... $\hat{X} \pm 2 CV$ ...	(2.641,02; 2.779,78)	(2.794,04; 2.944,36)	(2.851,40; 2.997,60)	(2.783,46; 3.000,94)
$\alpha = 99\%$ ... $\hat{X} \pm 2,5 CV$ ...	(2.623,67; 2.797,12)	(2.775,25; 2.963,15)	(2.833,12; 3.015,87)	(2.756,27; 3.028,12)
$\alpha = 99,7\%$ ... $\hat{X} \pm 3 CV$ ...	(2.606,33; 2.814,47)	(2.756,46; 2.981,94)	(2.814,85; 3.034,15)	(2.729,09; 3.055,31)

( 5. 4 )	INACTIVOS (MILES)			
	3.º trimestre 1984 (J.A.S.)	4.º trimestre 1984 (O.N.D.)	1.º trimestre 1985 (E.F.M.)	2.º trimestre 1985 (A.M.J.)
$\alpha = 67\%$ ... $\hat{X} \pm CV$ ...	(14.079,72; 14.167,28)	(14.137,82; 14.228,58)	(14.203,51; 14.294,79)	(14.254,38; 14.557,62)
$\alpha = 95\%$ ... $\hat{X} \pm 2 CV$ ...	(14.025,94; 14.211,06)	(14.092,44; 14.273,96)	(14.158,02; 14.340,38)	(14.102,76; 14.709,24)
$\alpha = 99\%$ ... $\hat{X} \pm 2,5 CV$ ...	(14.014,05; 14.232,95)	(14.069,75; 14.296,65)	(14.135,23; 14.363,17)	(14.026,95; 14.785,05)
$\alpha = 99,7\%$ ... $\hat{X} \pm 3 CV$ ...	(13.992,16; 14.254,84)	(14.047,06; 14.319,34)	(14.112,43; 14.385,97)	(13.951,14; 14.785,05)

(\*) En el cuadro n.º 5 se muestran ejemplos de intervalos de confianza que, con una probabilidad  $\alpha\%$ , contienen el verdadero valor de las siguientes características globales:

- 5.1. Población activa.
- 5.2. Población ocupada.
- 5.3. Parados.
- 5.4. Inactivos.

En los últimos trimestres disponibles.

A partir del segundo trimestre de 1985, se inicia el proceso de implantación de un nuevo diseño muestral, con modelación de ciertos estratos, consecuencia de estudios posteriores al censo de 1981 en cuanto al diseño de la EPA. La implantación gradual de este diseño a lo largo de seis trimestres consecutivos, en los que coexistirán estratos antiguos y nuevos, en principio, provocará un ligero incremento de los errores de muestreo que, es de esperar, vaya reduciéndose conforme avance el proceso de implantación.

CUADRO N.º 6  
TABLAS DE CONCORDANCIA

CLASIFICACION SEGUN E.O. CLASIFICACION SEGUN E.R.	Modalidad que se estudia	Restantes modalidades	Total personas
Modalidad que se estudia ...	a	b	a + b
Restantes modalidades ...	c	d	c + d
TOTAL PERSONAS ...	a + c	b + d	n

Fuente: Encuesta de Población Activa: Informe técnico, diseño de la encuesta y evaluación de la calidad de los datos.

Si al finalizar hemos conseguido responder a alguno de los interrogantes que a menudo nos planteamos los usuarios de la EPA nos damos por satisfechos. En cualquier caso, creemos que un mejor conocimiento del diseño de la muestra y del proceso de la entrevista nunca es ocioso, y permite apreciar mejor las dificultades de la estadística laboral vía encuestas, particularmente compleja por hacer referencia a actitudes y comportamientos humanos difícilmente medibles y a veces hasta contradictorios.

**NOTAS**

(\*) Agradecemos las facilidades proporcionadas por los centros oficiales de documentación estadística y las valiosas sugerencias del personal que realiza la EPA. Cualquier error que subsista es, por supuesto, de nuestra exclusiva responsabilidad.

(1) Para una discusión técnica sobre estos aspectos, pueden verse, entre otros, PAPELES DE ECONOMÍA ESPAÑOLA n.º 8 y *Análisis de las Estadísticas del Mercado de Trabajo*, del Ministerio de Economía y Comercio.

(2) La amplitud de este intervalo puede dar lugar a alguna crítica sobre la posibilidad de sobreestimación de las actividades urbanas (en particular del sector servicios, ya que las secciones de menor número de electores se ubican geográficamente en zonas poco pobladas). No obstante, creemos poder afirmar que esto no es así, puesto que el INE tuvo en cuenta este dato, junto con el hecho de ser mayor la variabilidad de respuesta en las zonas urbanas de mayor tamaño, y dio mayor peso relativo al estrato 1 (de municipios autorrepresentados de mayor tamaño en cada provincia) para evitar este conflicto.

Claro que esto no quiere decir que las fluctuaciones estacionales que se producen en determinadas ocupaciones del sector servicios queden perfectamente recogidas, por su complejidad intrínseca.

(3) Los estratos 1, 2, 3 y 4 se componen de municipios autorrepresentados, es decir, de municipios que siempre tienen al menos una sección presente en la muestra. Los estratos 5, 6, 7, 8 y 9 de municipios demográficamente similares de menor tamaño (en la actualidad no existen municipios en los estratos 3 y 4).

(4) Para que haya un número exacto de bloques. Un bloque es el conjunto de secciones que ha de visitar un entrevistador en un trimestre. Como entrevista una sección por semana, cada bloque contiene doce secciones.

(5) Sea  $P(S, t) = \frac{M(S, t)}{N(E, t)}$  la probabilidad de selección de una sección S, donde

$M(S, t)$  = al número de viviendas principales de la muestra, y

$N(E, t)$  = al número de viviendas principales del estrato en el momento t. Si:

a)  $P(S, t+5) = P(S, t)$ : la sección S permanece en la muestra.

b)  $P(S, t+5) < P(S, t)$ : la sección S permanece en la muestra con probabilidad de  $\frac{P(S, t+5)}{P(S, t)}$ , y desaparece (y es sustituida por otra) con probabilidad  $1 - \frac{P(S, t+5)}{P(S, t)}$ .

(6) El método de arranque aleatorio consiste en extraer aleatoriamente un elemen-

to de la muestra y, a partir de él, seleccionar los elementos que determinan el factor de elevación. Por ejemplo, si la muestra es de 100 elementos y el factor de elevación igual a 4 y, por arranque aleatorio, obtenemos el elemento 57, a partir de él los elementos: 61 (igual a 57 + 4), 65 (61 + 4), 69 ... y así hasta seleccionar 25 elementos.

Si el muestreo fuese aleatorio simple, el factor de elevación sería exactamente igual al período inverso de la fracción de muestreo x, tal que

$$\frac{1}{x} = \frac{\text{número de viviendas principales de la muestra (= 20)}}{\text{número de viviendas principales totales en la sección (= n)}}$$

Esto no es exactamente así porque la muestra es autoponderada en viviendas. En la EPA el factor de elevación oscila entre 200 y 300 unidades.

(7) El turno de rotación (el grupo que debe cambiarse en cada trimestre) se calcula mediante la fórmula siguiente:

$$T.R. = (6 - n.º \text{ de ciclos}) + 1$$

(8) Y, en cualquier caso, se incluyen siempre los datos referidos a: estudiantes, varones en el servicio militar, enfermos en sanatorios, hospitales, etc., y personas de la vivienda que trabajen temporalmente en el extranjero o estén en viaje de negocios, sean pescadores, navegantes del mar o de las líneas aéreas. Además, el servicio doméstico interno se encuesta siempre donde trabaja (independientemente del número de meses que lleve en ese domicilio).

(9) EGP-C1 para las ausencias, y CN para las negativas.

El cuestionario EGP-C1 recoge datos generales de los miembros de la vivienda en base a, por ejemplo, el padrón municipal.

El cuestionario de negativas (CN) contiene información sobre:

- el modo de obtención del cuestionario de negativas (en forma directa o vía el padrón);
- los motivos para rehusar la entrevista, y
- el número de entrevista (primera, segunda ... sexta) que se ha rehusado contestar).

(10) Nótese que una simple reordenación de las situaciones propuestas en la pregunta diez podría dar lugar a resultados diferentes. Por ejemplo, en lo que se refiere a las situaciones de ayuda familiar y no trabajo: está disponible y busca empleo o es escolar, estudiante y opositor.

(11) Para comprender en toda su dimensión la magnitud del problema de definición existente y los criterios utilizados para la identificación de una determinada situación laboral, nos remitimos al cuadro de C. de Miguel aparecido en el número 8 de PAPELES DE ECONOMÍA ESPAÑOLA, págs. 90-93.

(12) Por ejemplo, se trata de observar y evitar que se declaren distintos niveles de estudios en entrevistas sucesivas, que el tipo de actividad y el nivel de estudios declarados sean inconsistentes, etc. Parte de estos problemas pueden surgir al poder

contestar distintas personas por el mismo grupo humano en entrevistas sucesivas.

(13) Se demuestra que los estimadores de la muestra y sus reiteraciones son insesgados, por lo que el error cuadrático medio coincide con la varianza del estimador, donde la varianza así calculada será:

$$V(X) = \frac{\sum_{r=1}^{r=40} (X_r - X)^2}{40}$$

siendo:

- $x$ : el estimador de la muestra completa,
- $x_r$ : el estimador de la r.ª reiteración.

(14)

INTERVALO	Grado de confianza (%)
X (1 ± CV) ... .. .	67
X (1 ± 2 CV) ... .. .	95
X (1 ± 2,5 CV) ... .. .	99
X (1 ± 3 CV) ... .. .	99,7

(15) Basándose en el modelo de Hansen, Hurwitz y Bershad utilizado en la elaboración del censo de EE.UU.

**BIBLIOGRAFIA**

DE MIGUEL, C., «La medición del paro y la Encuesta de Población Activa», PAPELES DE ECONOMÍA ESPAÑOLA, n.º 8, págs. 86-104.

- «Reformas necesarias en la Encuesta de Población Activa» (Comunicación 1.6, Ministerio de Trabajo).

INE, *Encuesta de Población Activa*. Publicaciones anuales de los años 1964, 1972, 1976 y 1985. Publicaciones trimestrales de los años 1984 y 1985.

- *Encuesta de Población Activa*. Informe Técnico. Diseño de la Encuesta y Evaluación de la calidad de los datos, Madrid, 1978.

- *La Encuesta de Población Activa*, Madrid, 1979.

- *Encuesta de Población Activa*. Manual de definiciones: Instrucciones y normas de codificación, Madrid, 1979.

- *La Encuesta de Población Activa*. Manual Técnico: Corrección de errores mediante proceso electrónico, clasificación de la población encuestada y tablas de resultados, Madrid, 1978.

- Evaluación de la calidad de los datos de la Encuesta de Población Activa (varios años).

MINISTERIO DE ECONOMÍA Y COMERCIO, *Análisis de las Estadísticas del Mercado de Trabajo*.

PAPELES DE ECONOMÍA ESPAÑOLA, n.º 8, 1981.

SANCHEZ CRESPO, J. L., *Curso Intensivo de muestreo en poblaciones finitas*, Madrid, 1980.