

ESTUDIOS SOBRE EVALUACION DE LA PRECISION Y CALIDAD DE LOS RESULTADOS DE CENSOS Y ENCUESTAS (1)

Vicente Jiménez Díez de Artázcóz

INSTITUTO NACIONAL DE ESTADISTICA

1.- INTRODUCCION

Dos son los tipos generales de errores que se presentan en una encuesta por muestreo. El primero se debe a la utilización del muestreo, el segundo se presenta en cualquier investigación estadística, ya se base en la población total o en una muestra de la misma. Los del primer tipo disminuyen al aumentar el tamaño de la muestra, mientras que los del segundo son, en general, de mayor magnitud en las investigaciones exhaustivas.

Los errores debidos al muestreo han sido ampliamente tratados por la literatura estadística y existen métodos estadísticos generales para controlar este tipo de error.

Los errores del segundo tipo, denominados generalmente "errores ajenos al muestreo", tienen su origen en dos fuentes principales:

- a) errores en la fase preparatoria de la investigación: deficiencias en el marco, cuestionarios, instrucciones, etc.
- b) Errores cometidos en los trabajos de campo y tratamiento posterior de los datos.

El mayor peso relativo de los "errores ajenos al muestreo" ha motivado que

(1) Este artículo ha sido redactado por los Estadísticos Facultativos D. Vicente Jiménez Díez de Artázcóz, D.^a Rosa María Bermúdez Gómez, D.^a Isabel Muñoz Jiménez, y D. Javier de Parada Herrero.

la investigación estadística reciente se encuentre fuertemente orientada hacia la búsqueda de nuevas técnicas que permitan evaluar este tipo de errores y darles el tratamiento adecuado.

A continuación se exponen en forma sucinta algunos de los trabajos más relevantes realizados por el I.N.E. en estos últimos años para tratar de evaluar los distintos tipos de errores que menoscaban la precisión y calidad de los resultados de censos y encuestas.

2.— ERRORES DE MUESTREO

2.1.— Censo de población de 1950.

La primera aplicación del muestreo, hecha por el Instituto Nacional de Estadística, data del Censo de Población de 1950 que, aunque realizado exhaustivamente, las clasificaciones se publicaron estimando los resultados mediante una muestra del 10%.

Dados los escasos medios mecánicos que por aquella época estaban a disposición del I.N.E., el cálculo de los errores de muestreo se limitó a la clasificación de los habitantes por grupos de edad y sexo, que además de su interés y uso frecuente, ofrecía la ventaja de extenderse a todos los habitantes censados. Se calcularon para varones, mujeres, y total y en cuanto al ámbito geográfico, para el total provincial, zonas urbanas, intermedia y rural, capital de la provincia y municipios mayores de 30.000 habitantes.

En el tomo II del Censo de 1950 se publicaron los errores muestrales máximos con un coeficiente de confianza de 0,90 con lo que dicho error de muestreo máximo da la semiamplitud del intervalo de confianza, para la clasificación por grupos de edades anteriormente citada.

Para el cálculo de los errores, dentro de cada estrato se utilizaron las fórmulas del muestreo aleatorio sin reemplazamiento, y para el total provincial las del muestreo estratificado.

Para los restantes totales de clase se publicó también una tabla general de errores de muestreo y para la misma fracción de muestreo del 10%. En dicha tabla, en su parte superior, se encuentran los valores de las estimaciones; en el margen izquierdo, los tamaños de los estratos; y en la intersección de filas y columnas el error correspondiente.

2.2.— Errores de muestreo del Censo Industrial de la provincia de Barcelona

El Censo Industrial de la provincia de Barcelona se realizó con referencia al 31 de diciembre de 1957.

Decidida la enumeración completa de los establecimientos, se obtuvieron datos sobre número de obreros, potencia instalada y actividad económica, que permitió la clasificación en “grandes” con cinco o más obreros, y “pequeños” con menos de cinco. Para la obtención de los datos económicos de los primeros se procedió exhaustivamente, y se les remitieron cuestionarios especiales para cada actividad. Respecto a los segundos, se diseñó una muestra probabilística y se les envió un cuestionario en el que sólo figuraban las preguntas “Total pagado” y “Total cobrado”.

Para el diseño de la muestra se disponía de una lista de 39552 establecimientos industriales de menos de cinco obreros, localizados en la enumeración previa. Se consideró como más eficiente un muestreo probabilístico estratificado por ramas de actividad, y efectuada la estratificación se seleccionó una muestra en cada rama, excluyendo las Agrupaciones 11, 12, 30, 32 y 51 de la Clasificación Nacional de Actividades económicas que, por su pequeño número, aconsejó enumerar el 100 por 100.

Para la determinación del tamaño de la muestra con objeto de conseguir en la estimación del total un error de muestreo relativo del 5% con una probabilidad de 0,95, no se conocía la varianza poblacional, ni era posible conjeturar su valor; tampoco se disponía de estimación alguna, por no haberse realizado una encuesta piloto previa. Se estableció la hipótesis de que los datos económicos estarían fuertemente correlacionados, dentro de cada rama de actividad, con el número de obreros.

Dada la dificultad de clasificación de los pequeños establecimientos industriales por el número de obreros, dentro de cada rama de actividad, para obtener la distribución de frecuencias que permitiera calcular las varianzas poblacionales, se procedió a un estudio de la población, y de su análisis se observó que las distribuciones eran sensiblemente uniformes en cada rama de actividad, lo que permitió determinar el tamaño de la muestra en cada rama, estimando la varianza poblacional por la de la distribución uniforme.

Una vez obtenidas las estimaciones del número de obreros en cada rama, se compararon con las del Censo, obteniéndose los correspondientes errores relativos, que sólo en dos ramas de actividad superaron al 5%, siendo los restantes inferiores a dicho porcentaje. En realidad, lo que se hizo respecto al número de obreros fue una *validación* de la muestra.

Respecto a los datos económicos, se procedió a las estimaciones y cálculo de los errores de muestreo de las características “Total pagado” y “Total cobrado”. con la información que aportaron los establecimientos de la muestra. También se calcularon los intervalos de confianza con un coeficiente de confianza de 0,95.

A los valores estimados de las ramas de actividad que se investigaron por muestreo, se añadieron los valores de las que se censaron exhaustivamente, obteniéndose las estimaciones en millares de pesetas de los conceptos “Pagado” y

“Cobrado” de los establecimientos industriales con menos de cinco obreros en la provincia de Barcelona.

2.3.—Errores de muestreo del avance de las clasificaciones de la población del Censo de 1960.

La enorme tarea de los trabajos de clasificación para obtener las cifras definitivas nacionales del Censo de Población de 1960, exigía un margen de tiempo, difícil de acortar con los medios que por aquella época disponía el Instituto Nacional de Estadística.

Por ello, ante la urgente necesidad de disponer de una visión nacional de conjunto sobre las características principales de la población española, la Dirección General de Estadística ordenó se realizaran las clasificaciones fundamentales, utilizando una muestra del 1 % de los cuestionarios recogidos.

En consecuencia, se publicaron los resultados obtenidos a partir de dicha muestra, presentados en Tablas que, con bastante precisión, informaban sobre la estructura de nuestra población en el orden físico (edades), jurídico (estado civil), cultura (analfabetismo y nivel de instrucción), social (categoría socio-económica y distribución en hogares), y económico (población activa, ocupación, posición en la ocupación y ramas de actividad económica), comparando, además, las clasificaciones del Censo de 1960 con las de 1950.

El procedimiento de muestreo utilizado para seleccionar la muestra del 1 % fué el monoetápico, estratificado por municipios y selección sistemática con arranque aleatorio, tomando como unidad muestral la familia.

En la correspondiente publicación se indican las estimaciones cuyos errores de muestreo relativos son inferiores al 5%, las que superan a esta cifra pero son inferiores al 10%, y las que superan al 10 %.

2.4.—Errores de muestreo de la Encuesta de Población Activa

En el año 1964, el I.N.E. inició la serie de encuestas sobre población activa, que desde aquella época viene realizando en forma continua.

El tipo de muestreo es un bietápico estratificado, tomándose como unidades primarias las secciones censales del Censo de Población de 1960, y como últimas los hogares cuando se trata de determinar características de los mismos y, para las características referentes a los individuos, se transforma en trietápico, siendo las unidades muestrales de las etapas: sección, familia, individuos, pero no se realiza submuestreo en los últimos.

Las unidades primarias se estratificaron en 11 estratos: 5 para las secciones urbanas y 6 para las rurales. Para esta clasificación en estratos se dieron normas

detalladas que aparecen en la publicación del I.N.E. "Encuesta de Población Activa, 1964".

Las unidades primarias de una provincia, en el estrato correspondiente, se seleccionan con probabilidad proporcional al número de hogares, utilizando el método de Lahiri. De estas unidades primarias seleccionadas se conservan un 80% en el trimestre siguiente. La sustitución del 20% se hace seleccionando nuevamente las secciones, conservando para cada provincia la categoría socio-económica de las que se eliminan.

Inicialmente se seleccionaron 1000 unidades primarias y 75000 unidades secundarias cada trimestre con lo que la muestra fue del 1% aproximadamente respecto al total de familias existentes en el país.

Para el cálculo de los errores de muestreo se utilizaron las fórmulas del muestreo bietápico estratificado. Se estimaron errores de muestreo relativos, en porcentaje, para las características siguientes: activos que tienen empleo, parados, jubilados o retirados, escolares o estudiantes.

También se calcularon errores de muestreo para el total de personas activas según los siguientes grupos de edades: de 14 años, de 15 a 19, de 20 a 24, de 25 a 44, de 45 a 64, de 65 y más años.

Actualmente la Encuesta de Población Activa queda encuadrada dentro de la Encuesta General de Población, habiéndose efectuado una nueva estratificación de las unidades primarias y actualizadas las probabilidades de selección de las mismas, de acuerdo con las cifras recientes del Censo de Población de 1970. Los errores de muestreo se estimarán por el "Método de las reiteraciones".

2.5.—Errores de muestreo de la Encuesta de presupuestos familiares (marzo 1964 — marzo 1965).

Entre las fechas indicadas el I.N.E. realizó la citada encuesta. Los objetivos a alcanzar en esta investigación eran:

- 1.—Estimaciones del consumo privado.
- 2.—Establecer las ponderaciones para el cálculo del índice del coste de la vida
- 3.—Base para establecer un índice territorial del coste de la vida.
- 4.—Distribución de los gastos del consumo privado, según tipos de gasto y nivel de ingresos, tamaño del hogar, categoría socio-económica de los sustentadores principales, etc.

El marco de la investigación estaba constituido por los hogares de consumo privados fuesen o no colectivos.

La muestra extendida a lo largo de un año consistió en seleccionar 50 unidades primarias por semana, una por provincia, y en cada una de ellas se eligieron

aleatoriamente 8 hogares (unidades últimas). La selección de las unidades primarias se hizo con probabilidad proporcional al número de hogares de la misma. Las unidades últimas con probabilidad constante dentro de cada unidad primaria. Los estratos fueron los once mencionados en la Encuesta de Población Activa. El tipo de muestreo fué, pues, bietápico estratificado.

Los errores de muestreo relativos, en porcentaje, se estimaron para los grandes grupos de gasto (alimentación, vestido y calzado, vivienda, gastos de casa, gastos diversos), a *nivel nacional* para el gasto total, conjuntos urbanos y suburbanos, y por separado para el conjunto urbano y para el suburbano, y a *nivel provincial* para análogas características y conjuntos.

2.6.—Errores de muestreo de la Encuesta piloto de comercio interior

Por tratarse de una encuesta piloto, mediante la cual se deseaba obtener información sobre la estructura de los establecimientos comerciales y sobre ciertas características de los mismos se eligió como ámbito geográfico de esta encuesta la provincia de Madrid.

El diseño de la muestra se desdobló en dos partes:

1.—Madrid (capital)

2.—Provincia de Madrid (excluida la capital)

Ante las dificultades de estratificación de los establecimientos (ya que se carecía de información sobre los mismos) se adoptó, para Madrid (capital) la clasificación de los establecimientos por rúbricas de la licencia fiscal, considerando cada una de éstas como un estrato distinto, con lo que el muestreo fue estratificado aleatorio.

Para la provincia (excluida la capital) no se disponía de otra información que la relación de establecimientos comerciales de cada municipio, no constando la rúbrica de la licencia fiscal. Se pensó que los establecimientos comerciales de municipios con una población aproximada, tendrían bastantes características análogas. Se formaron para la provincia cuatro estratos clasificando los establecimientos según el número de habitantes del municipio respectivo. El muestreo fue también estratificado aleatorio.

Por tratarse de una encuesta piloto, el tamaño total de la muestra fue bastante grande: 30% en la capital y 10 % en la provincia.

La encuesta se realizó en dos fases:

En la primera se trató de obtener un directorio lo más depurado posible que permitiese estratificar la muestra, así como conseguir datos provisionales sobre la estructura del comercio, personas ocupadas, horas trabajadas, etc. La primera fase permitió determinar los coeficientes de elevación dentro de cada rúbrica en la capi-

tal, tratando de subsanar las deficiencias observadas en el marco muestral, ya que un 10% de los cuestionarios contestados estaban encuadrados en rúbrica distinta a la que figuraban inicialmente.

En la segunda fase se confeccionaron dos tipos de cuestionarios: uno para *compras* y otro para *ventas*. Para la selección de la muestra en esta segunda fase se utilizó la información obtenida del total de horas trabajadas por el personal en una semana y para cada establecimiento, y se calcularon las cuasi-desviaciones típicas muestrales de esa característica, para cada rúbrica, lo que permitió afijar el tamaño de la muestra en cada una de las rúbricas, estableciendo la hipótesis, para la estimación de otra características, de que el número de horas trabajadas debería estar fuertemente correlacionada con el volumen de compras y ventas.

Los tipos de comercio se agruparon en trece clases distintas, en diez de las cuales y mediante las fórmulas del muestreo aleatorio sin reemplazamiento, se encontraron errores de muestreo relativos para el total de horas trabajadas, inferiores al 5% en diez tipos de comercio, y entre estos últimos cinco inferiores al 2,5%. De los tres errores de muestreo restantes, sólo uno alcanzó el 10,1%.

2.7.—*Errores de muestreo de la Encuesta de equipamiento y nivel cultural de la familia.*

Durante el mes de marzo de 1968, el I.N.E. realizó su primera encuesta nacional sobre el equipamiento y nivel cultural de la familia, mediante entrevista directa de más de 60.000 hogares.

El objetivo de esta encuesta era el de conseguir información sobre la distribución de la renta en el triple aspecto: funcional, personal y geográfico.

El marco de la encuesta estaba constituido por todas las familias españolas residentes en el territorio nacional en el momento de su realización.

El tipo de muestreo fue el bietápico estratificado. Los municipios fueron clasificados en 3 estratos por sus características culturales, y en 4 por sus características generales, siendo 221 el número de estratos. En la primera etapa se seleccionaron 1.842 secciones con probabilidad proporcional al número de sus habitantes; en la segunda, 35 viviendas fueron seleccionadas con la misma probabilidad. La muestra bietápica fue obtenida a partir de los estratos primarios, los cuales fueron agrupados posteriormente en las 50 provincias, 12 regiones, 3 estratos culturales y 4 generales.

Los estimadores utilizados fueron: razones, medias, proporciones y totales, y los errores de muestreo se obtuvieron por el método de los "conglomerados últimos".

La muestra fue diseñada para estimar datos nacionales; no obstante, también se publicaron estimaciones a nivel provincial, por considerar que pudieran ser úti-

les a quiénes las empleen con la suficiente cautela.

Se calcularon errores de muestreo relativos para los siguientes porcentajes a nivel nacional, regional y provincial: cabezas de familia propietarios de sus viviendas, viviendas con agua corriente, idem con inodoro, cabezas de familia con automóvil, viviendas con cocina de leña, y cabezas de familia con vacaciones inferiores a 20 días.

También se calcularon errores de muestreo para el número medio de personas por vivienda, para el porcentaje de cabezas de familia que superaron el nivel cultural de sus padres, y para aquéllos cuyo nivel cultural ha sido superado por alguno de sus hijos.

En la Encuesta de Equipamiento y nivel cultural de las familias (muestreo por conglomerado) se calcularon estimaciones y errores de muestreo del total de clase de personas activas en cada provincia, lo que permitió una vez estimados los errores de muestreo mediante el muestreo aleatorio simple, calcular los coeficientes de variación y evaluar para cada provincia *el efecto de conglomerado*. (J. Parada "Análisis comparativo de los resultados de población activa, a nivel nacional, del Padrón de 1965 y de la Encuesta de Población Activa", revista Estadística Española, n° 47 abril-junio de 1970).

La estimación del efecto de conglomerado ha sido útil en el cálculo de los intervalos de confianza de las estimaciones de la "Encuesta de Economías domésticas de Barcelona".

2.8.—Errores de muestreo de la Encuesta de Turismo receptivo, 1970.

El Instituto Nacional de Estadística realizó en el mes de agosto de 1970 una encuesta de turismo receptivo.

El propósito de dicha encuesta fue el estudio de diversas características de los visitantes temporales principales, en alojamientos hoteleros y acampamentos turísticos, que se encontraban en nuestro país en agosto de 1970.

La encuesta se realizó por muestreo polietápico estratificado, y abarcó 4.127 alojamientos hoteleros y 183 acampamentos turísticos, habiendo sido 39.556 el número de entrevistas.

Las tablas de datos se refieren a un día medio del mes de agosto de 1970.

Los errores de muestreo para las características consideradas, se estimaron por el *método de las reiteraciones*.

Se calcularon los siguientes errores de muestreo relativos, en porcentaje, a nivel nacional, para totales de clase y medias, según:

- 1.—Nacionalidad de los visitantes temporales principales.
- 2.—Importe medio de los presupuestos de los visitantes por nacionalidad.

3.—Días de estancia media en España por categoría socio-económica de los visitantes.

3.—TRABAJOS SOBRE LA EVALUACION DE CALIDAD DE LOS DATOS ESTADISTICOS.

La calidad de los datos estadísticos es tema de permanente atención por parte de Institutos y Centros que se dedican a la elaboración de estadísticas en virtud de la información recogida mediante Censos y Encuestas. Las numerosas fuentes de error que amenazan a las distintas fases del desarrollo de una encuesta obligan a una minuciosa vigilancia y escurpulosidad tanto en el planteamiento de la investigación como en la recogida, elaboración y proceso de datos.

El I.N.E., productos de estadísticas, consciente de la necesidad de la lucha diaria por alcanzar baremos óptimos de calidad en su producto, arbitra medios cada vez más importantes para el control de calidad y crea, en 1966, un Servicio exclusivamente dedicado al estudio y puesta en práctica de nuevas técnicas estadísticas para medir y contrastar la calidad de los datos recogidos.

A continuación se exponen someramente algunos trabajos sobre evaluación de calidad realizados por el I.N.E. en estos últimos años.

3.1.—*Estudios previos a la iniciación de un programa de control de calidad en la Encuesta de Población Activa.*

Como paso previo al establecimiento de un programa general de control de calidad en la Encuesta General de Población se realizó un estudio piloto en una submuestra de la Encuesta de Población Activa. Comprendió este estudio dos fases:

- 1) Un análisis de inconsistencias en los cuestionarios de hogar y
- 2) un examen del nivel de instrucción de inspectores y agentes de dicha encuesta.

1) *Análisis de los cuestionarios de hogar.* Se analizaron los cuestionarios de 286 hogares que figuraban en la mecánica de la encuesta como unidades de observación repetida en dos trimestres sucesivos. El análisis fue realizado en dos etapas. en una primera etapa se procedió al examen de todos los cuestionarios para determinar las inconsistencias internas en cada uno de ellos; posteriormente, se compararon los cuestionarios de dos trimestres consecutivos pertenecientes a un mismo hogar con objeto de detectar cambios e inconsistencias entre los datos de los dos trimestres. La principal dificultad que surgió en esta 2ª fase radicaba en el desconocimiento de los valores verdaderos y en la imposibilidad de contrastar los datos con otras informaciones. Se pudo detectar la existencia de inconsistencias y discrepancias importantes pero no las causas a las cuales eran imputables. Como conse-

cuencia de los resultados obtenidos se consideró conveniente establecer en esta encuesta un programa de control de calidad, generalizado a toda España, que permitiera evaluar la calidad de los datos y determinar las posibles fuentes de errores.

2) *Nivel de instrucción de inspectores y agentes.* La fase de recogida de información en la Encuesta de Población Activa se realiza por medio de agentes entrevistadores. Por este motivo, al considerar el I.N.E. la necesidad de implantar un programa de control de calidad, se quiso determinar previamente tanto los conocimientos que poseían los agentes entrevistadores como el grado de preparación de sus supervisores. Se decidió, entonces, someter a inspectores y agentes a unas pruebas procurando que realizaran en ellas una labor similar a la que desarrollaban en la encuesta.

Para los inspectores se confeccionaron dos juegos de cuestionarios que correspondían a dos hogares imaginarios y en los que se habían introducido una serie de errores e incompatibilidades. Cada inspector tenía que examinar los cuestionarios y determinar si estaban formalizados adecuadamente, procediendo a su depuración en caso contrario.

Destinados a las pruebas de los agentes se redactaron unos supuestos en cada uno de los cuales se describía un hogar con las características individuales de sus miembros y las actividades que éstos ejercían. El trabajo de los agentes consistía en cumplimentar los cuestionarios correspondientes y, a continuación, codificar los datos.

La revisión de las pruebas demostró que las diferencias de preparación entre los agentes eran mucho más acusadas que entre los inspectores. En la distribución de errores por agente existía algunos valores extremos que eran del todo inadmisibles ya que sobrepasaban cualquier nivel de aceptación y demostraban su desconocimiento de las normas a seguir en la encuesta. Se consideró necesario someter a estos agentes a un periodo de adiestramiento finalizado el cual serían sometidos a observación durante algún tiempo para determinar si su trabajo se encontraba dentro de los límites de aceptación y, en otro caso, tomar las medidas que se juzgasen oportunas. Asimismo, se abrió una ficha para cada uno de los examinados constatando el número y tipo de errores que habían cometido. La finalidad de estas fichas era mantener al día un registro de errores, ampliándolo con experiencias posteriores, que permitiera determinar tendencias y establecer unos niveles de aceptación del trabajo realizado por agentes e inspectores.

3.2.—*Encuesta de Equipamiento y Nivel cultural de las familias.*

Los resultados obtenidos en el estudio piloto de la Encuesta de Población Activa demostraron la conveniencia de introducir programas de evaluación de resultados en los censos y encuestas que realiza el I.N.E.

En marzo de 1968 se realizó la primera encuesta nacional sobre el equipamiento y nivel cultural de las familias españolas aplicándose, en ella, un programa general de evaluación de calidad de los datos. Se siguió el modelo matemático aplicado por la Oficina de Censos de los Estados Unidos debido a Hansen, Hurwitz y Bershad, basado en la "entrevista repetida". El procedimiento operativo, muy simple, consiste en repetir las entrevistas en una muestra de los hogares seleccionados para la encuesta. Posteriormente se cotejan los datos en ambas ocasiones con objeto de investigar las inconsistencias y cuantificar los errores mediante la aplicación de una serie de índices. El modelo de Hansen, Hurwitz y Bershad presupone que se obtienen en las segundas entrevistas o "entrevista repetida" los valores verdaderos de las características en estudio. Aunque en la práctica es difícil comprobar si se ha alcanzado este objetivo, los datos de las entrevistas repetidas, realizadas con más medios y mejores agentes, serán de superior calidad que los primitivos y permitirán basar en ellos todos los cálculos de errores y sesgos.

En la Encuesta de Equipamiento se utilizaron, para la evaluación de calidad, entrevistadores de una categoría y formación superiores. Por otra parte, se pudo efectuar con mayor detenimiento y minuciosidad el trabajo de campo por ser más reducido el número de familias entrevistadas. Todo esto permitió suponer que se habían alcanzado las condiciones que para su aplicación exige el modelo anteriormente citado.

El programa de evaluación se extendía al estudio de las inconsistencias en la cobertura (omisiones o inclusiones erróneas de las unidades muestrales) y en el contenido (asignación incorrecta de las unidades elementales a las distintas clases que se establecen en las clasificaciones por características de la encuesta). Se analizaron ambos tipos de errores figurando en la publicación de la citada encuesta los resultados obtenidos. También se estudió la labor individual desarrollada por los agentes comparando la calidad alcanzada por cada uno de ellos con su actuación en las pruebas de la Encuesta de Población Activa.

La experiencia obtenida en esta encuesta fue de gran utilidad para los programas de evaluación que se establecerían posteriormente para encuestas de tipo continuo y censos.

3.3.—Encuesta sobre bienes de consumo duradero en las familias (marzo de 1969).

Un año después de la Encuesta de Equipamiento e intenciones de compra de 1968 se realizó, esta vez por correo, una 2ª encuesta a los mismos hogares de la primera con objeto de estudiar la evolución experimentada en el equipamiento de los hogares y conocer hasta que punto se habían cumplido las intenciones de compra reflejadas en la 1ª encuesta. Las contestaciones útiles recibidas, después de

reiterados avisos, alcanzaron aproximadamente los 2/3 del total de cuestionarios enviados. El importante volumen del "no consta" (33%) debido a múltiples causas (cuestionarios incompletos o mal cumplimentados, devoluciones de la oficina de Correos por "desconocido", dirección insuficiente, cambio de familia, ausencia o fallecimiento y la no-respuesta inherente a toda encuesta por correo) podía producir una significativa pérdida de representatividad de la muestra. Por ello el estudio de la calidad de esta encuesta fue orientado hacia la contrastación de la representatividad de la muestra y, en su caso, la realización de los ajustes necesarios para aminorar el sesgo de la no-respuesta.

El equipamiento de la familia está fuertemente relacionado con la categoría socio-económica por lo que la no-respuesta podría introducir fuertes sesgos en los datos si los no-respondientes se distribuyesen en forma sesgada entre las distintas categorías socio-económicas. En tal caso la 1ª y 2ª encuesta no corresponderían a muestras aleatorias de la misma población y, por tanto, todo intento de comparación sería inútil. Con objeto de conocer si la muestra de la encuesta por correo podía considerarse, a pesar del volumen de no-respuesta, como una submuestra aleatoria de la muestra de la 1ª encuesta se aplicó un contraste de homogeneidad mediante la χ^2 en base a las frecuencias teóricas y observadas de hogares por estratos y categorías socioeconómicas. El contraste resultó significativo (es decir, hubo que rechazar la hipótesis que la 2ª muestra fuese homogénea con la 1ª) por lo que se procedió a homogeneizar la muestra mediante el ajuste de la no-respuesta, causa activa de distorsión. Se estudió a qué clases pertenecían los hogares no respondientes procediéndose, a la vista de esto, a una subestratificación de la 2ª muestra dando una ponderación adecuada a cada substrato según el volumen de no-respuesta habido en el mismo.

Este estudio demostró la necesidad de conceder una primerísima atención, especialmente en encuestas por correo, al tratamiento de la no-respuesta arbitrando los medios técnicos y económicos necesarios para reducirla a un porcentaje mínimo.

3.4.—Contraste entre el Padrón Municipal de habitantes y la Encuesta de Población Activa.

Con ocasión de la renovación total del Padrón de Habitantes en 1965, operación que en virtud de la Ley de Régimen Local están obligados a realizar los Ayuntamientos todos los años terminados en 0 y en 5, fue llevado a cabo por el I.N.E. en colaboración con los Ayuntamientos del país un ensayo para la explotación con fines estadísticos de la información que se obtiene sobre las características de los habitantes (sexo, edad, estado civil, actividad, etc.).

La explotación de resultados se realizó en base a un muestreo sistemático de las Hojas Familiares de Empadronamiento. Se consideraba que de resultar prác-

tico el ensayo podría disponerse de una gran cantidad de información estadística que pudiese al día la obtenida mediante el Censo General de Población a realizar por el I.N.E. todos los años terminados en cero.

Dado que para el año 1965 se disponía de los datos de la encuesta habitual y periódica del I.N.E. sobre Población Activa se creyó oportuno contrastar los resultados de la encuesta con los obtenidos mediante la explotación muestral de las hojas del Padrón. Las discrepancias, en lo referente a población activa, fueron bastante notables. Si bien la cifra total de activos eran muy similar en ambos documentos, el desglose por sexos presentaba marcadas diferencias que por ser de signo contrario se compensaban en la cifra total (las cifras de población activa del Padrón eran superiores en casi un millón de varones e inferiores en más de un millón de mujeres que las correspondientes de la Encuesta de Población Activa). Estas discrepancias podían obedecer, dada la utilización de diferentes muestras para la obtención de resultados, a dos grandes tipos de errores: errores de muestreo y errores ajenos al muestreo. Fueron analizados los errores de muestreo en ambos tipos de investigaciones (Padrón y Encuesta de Población) construyéndose intervalos de confianza del 95 % para ambas estimaciones y se llegó a la conclusión de que los errores de muestreo podían justificar el 15 % de la diferencia en población activa varones y tan sólo el 8 % de la diferencia obtenida en la cifra de mujeres. El hecho de que los errores de muestreo sólo justifiquen un pequeño porcentaje de la diferencia habida en los resultados de ambas muestras probó que la causa principal de estas diferencias se encuentran en fuentes ajenas al muestreo.

Del análisis de los resultados obtenidos destacaron tres hechos:

1º) Respecto a los varones, figuraban con profesiones activas en el cuestionario del Padrón individuos clasificados en la categoría de jubilados, pensionistas o rentistas (categorías inactivas) según la Encuesta de Población Activa. La falta de definiciones concretas en los cuestionarios del Padrón respecto a la profesión, oficio u ocupación y la ausencia general de instrucciones sobre la forma de autocumplimentar el cuestionario en esta importante rúbrica fue causa de la imprecisión en la enumeración de la actividad en los cuestionarios del Padrón.

2º) Respecto a las mujeres, se observó en la rúbrica "sus labores" una diferencia a favor del Padrón de cerca de 2 millones de mujeres. Esto fue debido a que la rúbrica "sus labores" del Padrón recogía la casi totalidad de mujeres que se encontraban en régimen de "Ayuda Familiar" (ya que el cuestionario del Padrón no recoge ni define este aspecto) así como una buena parte de las que se encontraban como "jubiladas, rentistas o pensionistas" que cumplimentaron el cuestionario del Padrón poniendo como profesión "sus labores".

3º) En la Agricultura se observó una fuerte diferencia, a favor de la Encues-

ta de Población Activa, en las mujeres activas empleadas. Esto parece confirmar lo referente a “Ayuda Familiar” expuesto anteriormente, pues en la Agricultura se da en alto grado el régimen femenino de “Ayuda Familiar” que, al no ser recogido por el Padrón, infravalora en éste la población activa femenina dedicada a labores agrícolas.

En conclusión y con vistas a una mejor explotación de los datos del Padrón, en lo referente a Población Activa, se hace necesario intensificar esfuerzos tendentes a definir y recoger en el cuestionario:

- A) La situación de Ayuda Familiar.
- B) La mención expresa de categorías inactivas como jubilados, retirados, pensionistas, asilados, etc.
- C) Normas precisas que permitan el deslinde entre profesiones que por sus características peculiares pueden ser clasificadas en idéntico o distinto grupo.

3.5.— Evaluación de calidad en la Encuesta General de Población.

El proyecto de evaluación de calidad en la Encuesta General de Población perseguía un doble objetivo: controlar el trabajo de los agentes que intervienen en la encuesta y evaluar la calidad de los resultados.

Con este fin se debía repetir parte del trabajo asignado a cada agente, para así comparar los resultados obtenidos en la entrevista original (E.O.) con los de la entrevista repetida (E.R.). Esto permite evaluar de alguna manera dos grandes tipos de errores ajenos al muestreo:

- I) Errores de cobertura, que afectan a las unidades omitidas o erróneamente incluídas en la encuesta.
- II) Errores de contenido, que afectan a las características o ítems observados dentro de las unidades correctamente incluídas.

En la selección de la muestra para la entrevista repetida hubo que tener en cuenta el diseño de la Encuesta General de Población. En él, la muestra total se reparte de un modo uniforme a lo largo de un semestre, encontrándose dividida en tres submuestras que se extienden en el tiempo a un bimestre, representando cada una de ellas a toda la población. Por otra parte se hizo necesario considerar el bloque, unidad de trabajo de la E.G.P., que corresponde a la cantidad de secciones censales que puede visitar un agente entrevistador durante el tiempo de duración de la encuesta.

Según ésto, una vez determinado el tamaño de la muestra para la encuesta de evaluación, y siendo uno de los objetivos de la misma controlar el trabajo de cada agente se plantearon dos procedimientos de afijación de la muestra:

- 1º) Distribuir la bimestralmente, de manera que cada agente sea controlado una vez al bimestre o, lo que es lo mismo, tres veces durante el tiempo de duración de la E.G.P.
- 2º) Distribuir la semestralmente, esto es controlando a cada agente una vez durante el tiempo total de duración de la E.G.P.

El primer caso exigiría respecto al segundo, y para un mismo tamaño de muestra, unos costes de desplazamiento tres veces mayores y un número de entrevistas en cada desplazamiento tres veces menor. Razones de coste y presupuesto de la encuesta llevaron a adoptar el segundo procedimiento de afijación.

El trabajo de campo se lleva a cabo por agentes especializados, los cuales realizan la entrevista repetida 15 días después de la original, refiriéndose los datos de ambas entrevistas al mismo período de tiempo. A su vez, estos agentes están a las órdenes de unos inspectores de campo que atienden especialmente los casos de falseamiento de resultados en la entrevista original, bien por no haberse realizado las entrevistas o por haberse entrevistado a un hogar distinto del seleccionado.

Finalizado el trabajo de campo se lleva a cabo en oficina el tratamiento de los resultados. Teniendo en cuenta que los índices de calidad de la encuesta se basan en la comparación de los cuestionarios obtenidos en E.O. y E.R. para un mismo individuo es fundamental que la identificación de los dos boletines que correspondan a una misma persona sea correcta.

Una vez identificados, los cuestionarios son comparados mediante ordenador obteniéndose una serie de resultados que permiten realizar los siguientes análisis:

- Semanalmente y para cada sección censal:
 - 1) *Errores de cobertura en las viviendas*
 - 2) *Errores de cobertura en las personas*
- Semestralmente:
 - 3) *Errores de contenido.* Para cada una de las características investigadas se confecciona una tabla de doble entrada con la clasificación de las unidades en la E.O. y en la E.R. Las características que se estudian de la Encuesta General de Población son las siguientes:
 - Categorías activa o inactiva
 - Edad
 - Estado Civil
 - Nivel Cultural
 - Activos marginales
 - Ocupación
 - Situación profesional
 - Categoría socioeconómica
 - Rama de actividad

Estas tablas permiten analizar entre que clases se produce el mayor volumen de confusiones o donde se producen los mayores errores.

4) *Indice de la variabilidad de respuesta para cada una de las características estudiadas.*

Estos índices son:

- a) *Porcentaje de idénticamente clasificados.* Es un índice que mide la estabilidad de las respuestas.
- b) *Tasa de diferencia bruta.* Representa el porcentaje de unidades distintamente clasificadas en ambas entrevistas respecto al total de unidades entrevistadas. Es un indicador de las “desviaciones de respuesta”.
- c) *Tasa de diferencia neta.* Representa el porcentaje de la diferencia de unidades que poseen una determinada característica, según cada entrevista, respecto al total de unidades entrevistadas. La diferencia neta es una estimación del “sesgo de respuesta”.
- d) *Indice de cambio bruto.* Representa el porcentaje de la diferencia bruta respecto al total de unidades con la característica, según la encuesta de evaluación.
- e) *Indice de cambio neto.* Representa el porcentaje de la diferencia neta respecto al total de unidades con la característica, según la encuesta de evaluación.
- f) *Indice de inconsistencia.* Representa la contribución de la variabilidad de respuesta a la variabilidad total. Su valor oscila entre 0 y 1 siendo tanto más consistentes los resultados cuanto más próximo a cero es el índice.

5) *Indices de errores por rúbricas y agentes.*

Con el fin de observar los errores más frecuentes en cada uno de los agentes se obtiene una tabla con los porcentajes de errores en cada rúbrica por agente.

3.6.—*Encuesta de Evaluación de Calidad de los Censos Generales 1970.*

Posteriormente a la realización del Censo de Población, Viviendas, Edificios y Locales se llevó a cabo un programa de evaluación previsto para servir dos principales propósitos:

- a) Proveer a los utilizadores del Censo de medidas de su acuracidad.
- b) Orientar al I.N.E. en las mejoras de futuros Censos y Encuestas.

Operaciones esenciales en la aplicación de la encuesta fueron las siguientes:

A) Respecto al diseño:

- 1) Selección probabilística de una muestra de 700 secciones censales.
- 2) Recorrido exhaustivo de la sección para determinar los errores de cobertura en edificios, locales y viviendas.
- 3) Selección de una submuestra de viviendas dentro de las secciones seleccionadas.
- 4) Entrevista repetida en las viviendas seleccionadas para determinar la cobertura de población y los errores de contenido.

B) Respecto al trabajo de campo:

- 1) Mejor calidad y entrenamiento en los entrevistadores de la encuesta de evaluación.
- 2) Obtención de la información del *mejor respondiente*. A tal fin las características personales de habitantes mayores de 18 años se debían requerir de esas mismas personas. Los datos de menores de 14 años se requerían del cabeza de familia o responsable principal del menor. Se dejaría a elección del entrevistador la especificación del respondiente para personas entre 14 y 18 años.
- 3) Utilización de cartografía y material auxiliar. Los entrevistadores de la encuesta de evaluación fueron provistos de croquis y callejero de la sección que tenían que recorrer y del cuaderno del agente de entrevista original donde figuraban todas las unidades censadas de edificios, locales y viviendas dentro de la sección. El agente de la encuesta de evaluación en su recorrido de la sección anotaba en un nuevo cuaderno las discrepancias entre lo observado por él y lo registrado en el cuaderno del agente censal y cumplimentaba nuevos cuestionarios de las unidades seleccionadas en la submuestra.

El diseño de la encuesta de evaluación descansó sobre las siguientes bases:

- a) Muestreo trietápico estratificado, siendo:
 - Unidad de primera etapa: Municipio
 - Unidad de segunda etapa: Sección censal
 - Unidad de tercera etapa: Segmento (Conglomerado compacto de unas 10 viviendas aproximadamente).
- b) Selección de una muestra autoponderada de unidades últimas con una fracción de muestreo global de 1/1500.

La decisión de escoger el segmento como conglomerado compacto de unidades elementales se tomó a raíz de los resultados obtenidos en el ensayo de la encuesta de evaluación durante el censo piloto donde se probaron dos procedimientos alternativos de selección de las unidades elementales:

- I) División de la sección en segmentos y elección al azar de 1 segmento por sección.
- II) Listado de viviendas de la sección y selección sistemática de viviendas con arranque aleatorio.

Ambos procedimientos tienen sus ventajas e inconvenientes. La principal ventaja de la selección de segmentos es la gran economía de tiempo que se produce en el trabajo de campo. Esto pudo ser comprobado en el ensayo piloto donde se contabilizó el tiempo real empleado en el trabajo de campo que resultó, en media, 3 veces superior en las secciones con selección sistemática que en las secciones con selección de segmentos. El principal inconveniente de la selección de segmentos es la formación de éstos que requiere una documentación cartográfica básica de gran precisión. Por otra parte, la selección sistemática presenta la ventaja de extender la muestra a lo largo de la sección lo que, en general, producirá una mayor precisión de las estimaciones al disminuir el efecto del conglomerado en el error de muestreo.

La ponderación de ventajas e inconvenientes de cada método mostró la conveniencia del empleo del diseño por segmentos que, aparte las consideraciones sobre costes y tiempos anteriormente expuestas, presentaba un favorable impacto psicológico sobre el entrevistado, menos reacio a admitir una segunda entrevista si su vecino también la sufría.

El trabajo de campo de la encuesta comprendió el recorrido exhaustivo de 700 secciones censales con más de 400.000 viviendas visitadas donde se preguntaba al cabeza de familia si había sido censado y el número de miembros que componían el hogar. Para el estudio de los errores de contenido se cumplimentaron cuestionarios censales para todas las unidades (edificios, locales, viviendas y población) ubicada en los segmentos seleccionados.

Actualmente se está procediendo a los trabajos de oficina y a las tabulaciones correspondientes para obtener las conclusiones y resultados definitivos de la encuesta.

B I B L I O G R A F I A

- Censo de Población de 1950 (Tomo II).— I.N.E.
- Censo Industrial de la Provincia de Barcelona, 1957.—Establecimientos de cinco o más empleados.— I.N.E.
- Censo de la Población y de las Viviendas, 1960.— Avance de las clasificaciones de la población obtenido mediante una muestra del 1 por 100.— I.N.E.
- Baylos Corroza, Victoria.— Encuesta de la población activa española.— Revista “Estadística Española”, n° 24.—julio-septiembre 1964.— I.N.E.
- Pena Traperó, J.B.— Encuesta de presupuestos familiares. Revista “Estadística Española”, n° 24.— julio — septiembre.— I.N.E.
- Ginard Ferrer, Juan.— Encuesta de Comercio Interior.— Revista “Estadística Española”, n° 24.— julio — septiembre 1964.— I.N.E.
- El diseño de la muestra en las encuestas de Población Activa, Presupuestos familiares y Comercio Interior, por E. García España, V. Jiménez Díez de Artazcoz, L. Cabrero Ledesma, E. Ibarrola Muñoz, J. Leal Antón, M. Hernández Vara, y E. García de Robles. “Estadística Española” n° 24.— I.N.E.
- Encuesta de Población activa (1964).— I.N.E.
- Encuesta de Presupuestos familiares (marzo 1964 — marzo 1965).— I.N.E.
- Encuesta de Equipamiento y nivel cultural de la familia (1968).— I.N.E.
- Parada, J.—Análisis comparativo de los resultados de población activa, a nivel nacional, del Padrón de 1965 y de la Encuesta de Población Activa.
Revista “Estadística Española,” n° 47, abril — junio, 1970.
- Encuesta de Turismo receptivo, 1970.— I.N.E.
- Encuesta de Turismo receptivo, 1970.—Evaluación de resultados.—I.N.E.

