

La captura de datos asistida por ordenador en la Encuesta de Población Activa

por
MIGUEL ANGEL MARTÍNEZ VIDAL
Subdirector General Adjunto Recogida de Datos
INE

TERESA MORATILLA PARDO
Jefe de Área de Trabajos de Campo
INE

1. INTRODUCCIÓN

La implementación de cuestionarios complejos, con diversos apartados aplicables a distintos individuos según características determinadas en el propio cuestionario, y la necesidad de detectar errores en la captura de la información, en el momento más cercano posible a la entrevista, son algunas de las razones que llevan a plantearse la utilidad de disponer de la ayuda del ordenador en la realización de las entrevistas.

De acuerdo con la evolución sufrida por la tecnología, se desarrollaron en primer lugar las entrevistas telefónicas asistidas por ordenador (CATI, computer assisted telephone interviewing). Más tarde con el desarrollo de los ordenadores portátiles se aplicó esta técnica a las entrevistas personales en encuestas que no resulta adecuado realizar por teléfono.

Las primeras experiencias europeas llevadas a cabo en entrevistas personales asistidas por ordenador (CAPI, computer assisted personal interviewing) se realizaron en Suecia y en los Países Bajos donde se realiza la encuesta de fuerza de trabajo por este procedimiento desde enero de 1987. También en 1987 en Estados Unidos se utilizó CAPI para la recogida de una parte de la Nationwide Food Consumption Survey.

En este artículo se presentan los objetivos establecidos para la implantación de la recogida de datos de la EPA asistida por ordenador y el desarrollo y resultados de los trabajos que se han llevado a cabo.

2. VENTAJAS DE LA RECOGIDA DE DATOS ASISTIDA POR ORDENADOR

Los objetivos fundamentales que se persiguen al realizar una encuesta recogiendo los datos con ordenador con relación al sistema tradicional de recogida con cuestionario en papel, son los siguientes:

- Mejorar la calidad de la información al efectuarse entrevistas dirigidas (a cada persona se le formulan cuestiones diferentes en función de respuestas anteriores), con controles internos que impiden cometer errores o avisan de situaciones extrañas.

De esta forma se evitan los errores de flujo (se sigue correctamente el cuestionario), los de rango (sólo se admiten valores válidos) y algunas inconsistencias (no todas, porque ello ralentizaría excesivamente el ritmo de las entrevistas).

- Integrar el conjunto de documentación que se utiliza en la encuesta (cuestionario básico, posibles cuestionarios auxiliares, módulos de carácter temporal, partes de trabajo, listas de códigos, manuales, etc.) en un solo *cuestionario integrado* con lo que se consigue que el entrevistador tenga toda la información que pueda necesitar durante una entrevista de forma cómoda e inmediata.

Hay que tener en cuenta que el uso de distintos impresos durante el trabajo de campo supone que el entrevistador debe hacer su recorrido con ellos lo que acarrea soportar un peso considerable, aparte de que se pueden cometer errores en la identificación de los mismos (los cuestionarios que correspondan a una misma unidad de encuesta, vivienda, deben tener la misma identificación).

- Realizar simultáneamente a la recogida de información la grabación de los datos y permitir su codificación y depuración en el punto geográfico más próximo al de captura y en la fecha más próxima a la fecha de referencia de los datos.

De esta manera, en caso de detectarse cualquier problema con la información recogida, se incrementan considerablemente las posibilidades de subsanarlo

volviendo a contactar con el hogar que aún conservará un recuerdo reciente de la entrevista. Además se reduce así el trabajo de oficina y se favorece el tiempo de trabajo de campo.

- Reducir los plazos de obtención de resultados y de disponibilidad de los mismos al ser la codificación, depuración y grabación prácticamente simultánea a la recogida de datos.

- Permitir la fácil obtención de resúmenes de control y seguimiento de la marcha de los trabajos lo que permite planificar más adecuadamente las inspecciones a realizar.

- Permitir la introducción de distintos módulos de carácter temporal al cuestionario básico para grupos específicos de la población con características determinadas por múltiples variables.

En definitiva, se mejora la calidad de la encuesta y se consigue obtener resultados en un tiempo más corto, facilitando además el trabajo del entrevistador y el seguimiento de los trabajos.

3. REQUERIMIENTOS DEL HARDWARE UTILIZADO PARA LA REALIZACIÓN DE ENTREVISTAS PERSONALES

La experiencia piloto de 1991

El INE, de acuerdo con la Oficina de Estadística de las Comunidades Europeas (EUROSTAT) llevó a cabo un ensayo piloto de realización de entrevistas personales con ordenadores portátiles en la Encuesta de Fuerza de Trabajo de 1991 (2º trimestre 1991).

Los ordenadores portátiles utilizados, pese a ser los más adecuados en aquel momento, plantearon serios inconvenientes:

- Se componían de pantalla y teclado externo, lo que requería apoyar el teclado en una mesa para poder introducir los datos. Por tanto, era preciso convencer a los informantes de que permitieran al entrevistador la entrada a la vivienda y el uso de una mesa para poder realizar la entrevista, con el consiguiente recelo del informante.

- Su peso rondaba los tres kilos, lo que suponía un grave inconveniente para su transporte entre vivienda y vivienda.

- Las baterías eran grandes e incómodas y era preciso llevar baterías de repuesto para alcanzar 6-7 horas de trabajo autónomo, ya que mientras el entrevista-

El entrevistador efectúa las entrevistas no puede conectar con la red eléctrica por las mismas razones apuntadas anteriormente.

– Por otra parte, las dimensiones del equipo delataban claramente que se trataba de un ordenador, lo que podía suponer un riesgo de robo y peligro para el entrevistador en determinadas zonas.

Requerimientos de hardware

Cuando en 1995 se comprobó que los principales inconvenientes habidos en 1991 podían ser obviados con la evolución de la tecnología, se retomó el proyecto. Para la elección del producto final que se utilizaría se tuvieron en cuenta los siguientes condicionantes.

– El ordenador había de poder sostenerse con una única mano, no siendo necesario apoyarlo en una mesa para utilizarlo. De esta forma se trataba de que este nuevo instrumento fuese semejante al que manejaban habitualmente los entrevistadores, siendo como un cuaderno que, se apoya en una mano y con la otra se escribe en él.

Así se evita además la desconfianza de las familias ya que no es preciso que el entrevistador entre en la casa para apoyar el ordenador en la mesa.

El equipo elegido cubría este requerimiento utilizando como entrada de datos un lápiz electrónico al cual era sensible la pantalla (aunque también permitía conectarle un teclado tradicional).

– El peso y tamaño de la máquina, la comodidad para transportarlo, así como la resistencia de sus componentes a condiciones desfavorables (calor, frío, humedad, caídas...) también fue tomada en cuenta.

El equipo seleccionado pesaba 1,2 Kg. con las 2 baterías incluidas. Este es un determinante básico ya que el entrevistador debe trabajar con el portátil durante toda una jornada, y esto supone no sólo llevarlo consigo en su recorrido entre viviendas sino sostenerlo apoyado en el brazo durante todo el tiempo que dura cada una de las entrevistas escribiendo o punteando sobre él, tarea que en no pocas ocasiones puede tener lugar de pie o en la puerta de la calle.

Su tamaño es bastante reducido (algo menor que un DIN-A4) sin que ello disminuya desproporcionadamente el espacio destinado a pantalla. El menor tamaño favorece no sólo su sujeción a la hora de la realización de la entrevista, sino también un mejor acoplamiento y una mayor estabilidad.

– El tiempo de autonomía continuada del ordenador portátil utilizado con baterías sin conectar a red eléctrica.

La duración de los modelos estudiados estaba en torno a unas 3 horas. Puesto que la jornada de trabajo en campo de un entrevistador es mucho mayor, el tamaño y peso de las baterías así como el que éstas fuesen *standard*, también se tuvo en cuenta ya que se hace necesario llevar baterías de repuesto

El modelo elegido dispone de baterías iguales a las que se utilizan para las videocámaras, pudiendo adquirirse en cualquier establecimiento de fotografía. Estas baterías además son muy pequeñas y no pesan, pudiendo transportarse dos o incluso cuatro adicionales cómodamente en un bolsillo.

Por otra parte, el sistema de aviso del nivel de batería también es más cómodo ya que se dispone de él como un dato más en la pantalla, evitando así otros sistemas que emiten señales acústicas, lo que puede interferir en la entrevista que se esté realizando.

– La velocidad de proceso debe ser suficiente para conseguir flexibilidad y rapidez en el momento de la entrevista (paso rápido de una pregunta a la siguiente, velocidad de consulta, ...).

Características del hardware actualmente en uso

De acuerdo con los puntos anteriores los equipos con los que actualmente se trabaja (que corresponden a dos modelos diferentes) y que fueron adquiridos entre 1995 y 1997 tienen las siguientes características.

1) Fabricante: Fujitsu (modelo Stylistic 500)

Peso con las 2 baterías incluidas: 1,2 kgs.

Dimensiones: 18,2 x 0,37 x 2,72 (en cms.)

2 baterías standard tipo vídeo (por ejemplo de SONY) de ión-litio con una duración media de 3,5 horas.

Lápiz PEN activo con 3 pilas tipo reloj

Procesador INTEL 486 DX2 50 MHz.

Memoria RAM 8 Mb

Disco duro interno tipo tarjeta PCMCIA 105 MB

Posibilidad de adaptar disquetera externa de 3,5"

Pantalla LCD retroiluminada, con VGA y con resolución máxima de 640x480, y 64 escalas de grises

Tamaño de la diagonal de la pantalla 20,3 cm. (pantalla de 8")

Compatibilidad con PC's. pudiendo conectarse como periférico o principal

Posibilidad de uso como PC personal y con tablero externo accesorio (teclado)

2) Fabricante: IBM (modelo Thinkpad 730 TE)

Peso con 1 batería incluida: 1,6 Kg.

Peso con 2 baterías incluidas: 1,8 kg.

Dimensiones: 269 x 210 x 34,8 mm

Baterías tipo NiMH con adaptador AC y cargador rápido. (Estas baterías de hidruro metálico de níquel reducen drásticamente el efecto memoria de las antiguas baterías de níquel-cadmio.)

Autonomía aproximada de 1,6 horas con 1 batería y 3,3 horas con 2 baterías.

Lápiz electrónico PEN

Procesador INTEL DX4/75 MHz. con 16 Kb de caché interna

Memoria RAM base 8 Mb

Unidad de memoria en disco PCMCIA con 260 Mb

Posibilidad de unidad externa de disquete de 3,5"

Pantalla de cristal líquido con resolución VGA de 640 x 480 retroiluminada/reflejada

Pantalla de 9,5 pulgadas monocroma, sensible al lápiz PEN, con cristal grabado para un tacto como el del papel.

Replicador de puertos estándar para ampliar la conectividad: conexión a teclado, ratón, monitor, impresora y disquetera externa.

4. EL DISEÑO DE LA APLICACIÓN INFORMÁTICA

El desarrollo se realizó en entorno WINDOWS FOR PEN, en VISUAL BASIC, creando así una aplicación específica para el cuestionario electrónico. Su elaboración se encargó, en una primera versión a una empresa externa al INE, aunque con motivo del cambio de cuestionario de la EPA en 1999 para su adaptación a la encuesta comunitaria, sé reelaboró por completo la aplicación, trabajo que fue realizado íntegramente por la Subdirección General de Informática Estadística del INE.

La aplicación disponible en los portátiles tuvo las siguientes premisas que ahora constituyen las principales características de sus funcionalidades. Básicamente se trata de un cuestionario electrónico integrado, dirigido y controlado.

Es un cuestionario **integrado** porque engloba los siguientes impresos y libros que se requieren para la realización de la EPA:

– Impresos:

- Parte de trabajo o documento donde se indican las direcciones de las viviendas a visitar en cada sección para realizar las entrevistas y en el que se reflejan las incidencias que surgen durante el trabajo de campo (negativas, ausencias, vacías, sustituciones realizadas...)

- El cuestionario de la EPA propiamente dicho, que se utiliza cuando un grupo acepta a colaborar (grupo encuestado).

- El cuestionario de negativas, que se utiliza cuando un grupo se niega a colaborar por primera vez.

- El cuestionario Ficha de Alquiler, se trata de un cuestionario auxiliar que recoge datos de las viviendas en alquiler y que sólo se realiza en los cuartos trimestres (módulo temporal) a los grupos encuestados.

- El módulo que se incluye en los segundos trimestres de cada año (en 1999 módulo de accidentes y enfermedades profesionales, en 2000 módulo de transición de la educación inicial al mercado laboral).

- La ficha de vivienda, que recoge la historia de las distintas entrevistas realizadas a lo largo de los 6 trimestres de encuesta mientras no cambie el grupo que ocupa la vivienda. En este caso se abre una nueva ficha de vivienda.

Esta ficha se completa además con una Tabla de Composición del Grupo que ocupa la vivienda y una Ficha de persona para cada uno de los componentes del grupo.

– Libros:

- Manual del Entrevistador relativo al trabajo de campo

- Clasificaciones o codificaciones de:

- Estudios terminados (CNED adaptada a la CINE-97)

- Estudios en curso (CNED adaptada a la CINE-97)

- Ocupaciones (CNO-94)

- Actividad del establecimiento (CNAE-93)

- Municipios por provincias al 1 de enero

- Países por continentes

Todas ellas incluyen los literales correspondientes junto con su código asociado.

- Listados de consulta:

- Empresas en franquicia
- Organismos públicos

Se trata de un cuestionario **dirigido** porque la propia aplicación dirige el trabajo del entrevistador, indicándole el cuestionario que debe hacer en cada caso y avisando de cuándo hay que efectuar sustituciones y cómo.

Se trata de un cuestionario **controlado** porque incluye controles en el ámbito de sección, de vivienda y persona, y además hace los controles tanto durante el trabajo de campo como después de campo.

Los datos en el portátil van encriptados de manera que en caso de pérdida o sustracción del mismo queda preservada la intimidad de los informantes.

Esta aplicación se ejecuta de igual forma en PC's (cambiando el lápiz electrónico por ratón y/o teclado) de manera que las fases de depuración posteriores a la recogida de datos se pueden realizar en el mismo entorno. Esta opción también se utiliza para realizar entrevistas telefónicas desde la oficina de la Delegación Provincial del INE.

El sistema de almacenamiento y envío a los servicios centrales del INE de la información para entrar en los procesos centralizados de obtención de resultados también están informatizados con aplicaciones que enlazan con la fase de trabajo ya descrita.

Existen varias formas de introducir los datos según el tipo de pregunta:

- Elegir una opción precodificada, marcando con el lápiz la respuesta válida, en preguntas en las que aparecen las posibles opciones de respuesta (por ejemplo, en la Clave de vivienda: encuestable o en el sexo: varón).

- Por reconocimiento de caracteres numéricos o alfanuméricos, escribiendo con el lápiz en los peines correspondientes el texto o cifra que proceda, en preguntas en las que aparecen peines para indicar la respuesta (por ejemplo, observaciones, el nombre y apellidos de una persona o el teléfono).

El ordenador reconoce lo escrito y si no entendiera alguna letra o número lo indica con una interrogación para que se escriba de nuevo.

- Por teclado numérico o alfanumérico, en preguntas en que aparece un icono o símbolo de teclado en la parte superior de la pantalla. Al pulsar este símbolo con el lápiz aparece el dibujo de un teclado en la pantalla, y se puntea con el lápiz sobre dicho teclado la respuesta (por ejemplo, la dirección, el número habitual de horas de formación a la semana).

- Mediante calendario, en las preguntas tipo fecha en las que aparece una tecla que indica Calendario, al pulsar esta tecla con el lápiz se despliega un calendario en el que se señala el día, mes y año (por ejemplo, en la fecha de visita o en la fecha de nacimiento).

- Mediante un spín de incremento (▲) o de decremento (▼), en las preguntas en que aparecen estos símbolos, pulsando con el lápiz sobre el spín para aumentar o disminuir la cifra (por ejemplo, en las horas de visita o en las horas habituales de trabajo).

- Acudiendo a una lista de códigos, en las preguntas en las que aparece una tecla que indica Seleccionar. Al pulsarla se despliega la lista de códigos correspondiente junto con sus literales (por ejemplo en países, municipio-provincia, estudios terminados y en curso, ocupación, actividad del establecimiento).

- Haciendo anotaciones (grafismos) con el lápiz sobre la pantalla en los lugares destinados al efecto en las preguntas en que aparece el icono correspondiente. Esto permite que el Entrevistador escriba sobre la pantalla con su propia letra para no perder tiempo en buscar en la lista de códigos y así no interrumpir la entrevista. Admiten grafismos las preguntas de ocupación, actividad, estudios, nacionalidad y lugar de nacimiento. Posteriormente, en la oficina, deben codificarse.

- Mediante el despliegue de una lista de números, señalando los que procedan (por ejemplo, el precio del alquiler o la superficie de la vivienda).

En algunas preguntas son posibles varias formas de cumplimentación, mientras que en otras sólo es posible utilizar una de ellas. Cuando para una pregunta son válidas distintas formas de cumplimentación aparecen en pantalla los iconos o símbolos correspondientes. El Entrevistador puede utilizar la que le resulte más cómoda.

Cuando una persona haya sido entrevistada previamente (en trimestres anteriores) las contestaciones facilitadas en la entrevista anterior relativas a variables cuya modificación no se espera, aparecen ya incluidas en el cuestionario, aunque como mínimo siempre se confirma la respuesta. Esto facilita enormemente la realización de las entrevistas posteriores a la primera.

5. EFICIENCIA DEL SISTEMA

La implantación del sistema de captura de datos para la EPA con ordenadores portátiles ha supuesto una serie de ventajas importantes como consecuencia de que la detección de errores durante la propia entrevista aumenta considerable-

mente la calidad de los datos, ya que el entrevistador puede consultar en ese mismo momento al entrevistado. Entre las principales mejoras se puede citar:

- Se han reducido significativamente los errores que se producían con el sistema anterior (cuestionarios en papel y grabación centralizada), con el consiguiente impacto en la calidad de las estimaciones:

Se han disminuido en aproximadamente un 70% el número de registros de personas con algún error en la fase de imputación automática (pasando de un 7,43% en el cuarto trimestre de 1995 a un 2,14% en el cuarto trimestre de 1999).

Además de reducirse el número de registros con error, se ha reducido el número de errores de cada uno de los registros con error. En el cuarto trimestre de 1999 los errores detectados en personas con error son de 1,06 frente al 1,15 del cuarto trimestre de 1996.

Estos últimos, errores referidos a inconsistencias entre diferentes variables de una misma persona, no se eliminan completamente (ya que no se han incluido en la aplicación todas las inconsistencias para no ralentizar la entrevista en exceso) pero son mínimos.

La media de errores detectados por registro según tipo de error es:

- Errores de flujo o contenido: 0,0052
- Errores en estudios terminados o en curso: 0,0111
- Errores en la relación entre situación profesional, estudios terminados y ocupación: 0,0028
- Errores de ocupación y actividad: 0,0030

Los errores en números de identificación de viviendas o personas, códigos de sección, municipio, provincia, país, actividad, ocupación y nivel de estudios y los relativos a estructura del hogar, así como los flujos del cuestionario, se eliminan prácticamente en su totalidad.

La práctica eliminación del proceso posterior de tratamiento de los datos que se necesita para los sistemas sobre cuestionarios en papel (Baket (1995)) se concreta en:

- El equipo de grabación existente en la Unidad de Proceso de Datos se ha destinado a otras operaciones distintas de la EPA ya que la grabación de la EPA se realiza en el mismo momento de la captura de los datos.

- En cuanto al equipo de depuración existente en Servicios Centrales que, antes de pasar a grabación y después de recogida la información en cada provincia, efectuaba una depuración exhaustiva de todos los cuestionarios recibidos, se ha

podido reducir de 17 auxiliares de estadística a sólo 4. Este equipo reducido efectúa una fase de revisión posterior a la detección automática de errores con corrección manual de los mismos.

– Se ha adelantado el calendario de publicación de los datos de manera que a la séptima semana después del trimestre se publican los resultados del mismo. En otros países que realizan encuestas continuas con referencia trimestral el tiempo que media entre la finalización del periodo de referencia y la publicación de los datos son de 10 semanas en el caso del Reino Unido, 8 en el caso de Dinamarca y 12 en el caso de Bélgica. En otros países donde la referencia es mensual este plazo es de 3 semanas en el caso de Finlandia y de dos en el caso de Francia.

6. CONCLUSIÓN

La recogida de datos asistida por ordenador tiene un objetivo directo como es el de mejorar la calidad de los datos recogidos en el mismo momento que se está realizando la entrevista. Sin embargo su efecto se extiende prácticamente a todas las fases posteriores de tratamiento de los datos consiguiendo eliminar los procesos de grabación, agilizar los de depuración e imputación automáticas y acortar los plazos de disponibilidad de la información.

Otra ventaja que posee el uso de ordenadores para la recogida de información es que aumenta la potencialidad de los cuestionarios autoadministrados. El control total sobre el flujo del cuestionario y de los errores de rango y contenido permite diseñar cuestionarios electrónicos complejos, orientados a ser cumplimentados por los entrevistados sin que medie la presencia del entrevistador. De esta forma se proporciona al entrevistado unas mayores garantías de privacidad que reducen uno de los principales riesgos de las entrevistas personales: la *presión* que la mera presencia del entrevistador, aún cuando su actitud sea totalmente correcta, ejerce sobre las respuestas que proporciona el entrevistado y que le puede llevar a inducir errores de medida adecuándolas a patrones de comportamientos socialmente aceptados.

Esta utilidad hace especialmente útil este instrumento para encuestas sobre temas sensibles para la intimidad de los entrevistados y ha sido ya utilizado por el INE como una posibilidad que se les ha ofrecido a las mujeres que formaron parte de la muestra de la Encuesta de Fecundidad (1998).

Las ventajas que esta técnica proporciona hace que el INE se plantee su aplicación a otras operaciones estadísticas en un futuro inmediato.

REFERENCIAS

- BAKER REGINALD P., BRADBURN NORMAN M. AND JOHNSON A. (1995). «Computer-Assisted Personal Interviewing: An experimental evaluation of data quality and costs». *Journal of Official Statistics*, 11, 4, 415-434.
- COUPER MICK ET AL (1998). «Computer Assisted Survey Information Collection». New York: Wiley.
- MARTIN JEAN, O'MUIRCHEARTAIGH COLIN AND CURTICE JOHN. «The use of CAPI for attitude surveys: an experimental comparison with traditional methods». *Journal of Official Statistics*, 9, 3, 641-661.
- VAN BASTELAER, R.A., KERSSEMAKERS, F. AND SIKKEL, D. (1988). «A test of the Netherlands's continuous labour force survey with hand-held computers: contributions to questionnaire design». *Journal of Official Statistics*, 4, 141-154.
- WEEKS, M.F. (1992). «Computer assisted Survey information collection: a review of CASIC methods and their implications for survey operations». *Journal of Official Statistics*, 8, 445-465.