

¿ CÓMO MEJORAR LA CALIDAD DE LA INFORMACIÓN ?
OPCIONES PARA MEJORAR LA ORGANIZACIÓN DEL TRABAJO DE
CAMPO, EL SISTEMA DE ENTRADA DE DATOS, EL ANÁLISIS DE
CONSISTENCIA Y EL MANEJO DE LA BASE DE DATOS*

JUAN MUÑOZ, CONSULTOR
(BANCO MUNDIAL)

* Documento de trabajo presentado a la Reunión de Iniciación del Programa MECOVI, celebrada en Asunción, Paraguay, el 3 y 4 de diciembre de 1996.

ÍNDICE

	<u>Página</u>
I. INTRODUCCIÓN	279
II. ORGANIZACIÓN DEL TRABAJO DE CAMPO EN LAS ENCUESTAS DE NIVEL DE VIDA	280
1. Esquema de dos rondas	280
2. Equipos móviles.....	281
III. VENTAJAS DE LA INTEGRACIÓN DE LA INFORMÁTICA AL TERRENO	282
IV. CRITERIOS DE CONTROL DE CALIDAD.....	282
1. Chequeos de rango	283
2. Verificación contra tablas de referencia.....	283
3. Controles de flujo.....	283
4. Controles de consistencia.....	284
5. Chequeos tipográficos.....	284
V. REQUERIMIENTOS PARA LA ORGANIZACIÓN CENTRAL.....	285
VI. CONSECUENCIAS PRESUPUESTARIAS.....	285
VII. OTROS ASPECTOS VINCULADOS CON LA CALIDAD DE LOS DATOS	286
1. Criterios de supervisión	286
2. Cuestionario único	286
3. Entrenamiento	287
4. Ensayo con vestuario	287
5. Substitución de hogares	287
6. Definición de identificación de las unidades estadísticas	287

¿ CÓMO MEJORAR LA CALIDAD DE LA INFORMACIÓN ? OPCIONES PARA MEJORAR LA ORGANIZACIÓN DEL TRABAJO DE CAMPO, EL SISTEMA DE ENTRADA DE DATOS, EL ANÁLISIS DE CONSISTENCIA Y EL MANEJO DE LA BASE DE DATOS

I. INTRODUCCIÓN

En este trabajo se describe el impacto que la evolución microinformática de los últimos 15 años ha tenido sobre diversos aspectos metodológicos de las encuestas de hogares, con un énfasis especial sobre aquellos que más han incidido sobre la calidad y oportunidad de los productos finales. La descripción se basa fundamentalmente en la experiencia recogida en los estudios sobre el nivel de vida de los hogares realizados por el Banco Mundial en este período, pero la mayoría de las conclusiones pueden aplicarse a otros tipos de encuestas de hogares, y particularmente a aquellas que revisten cierta complejidad e intentan observar la interacción entre varios aspectos de la realidad de los hogares.

Se tratarán, en particular, los temas relativos a la organización y logística de las operaciones de terreno, la transcripción de los datos a un soporte informático, el control de calidad de la información y la formación de los archivos destinados al análisis. Aunque el esquema de una presentación lineal obliga a analizar estos aspectos de manera separada, es bueno precisar al inicio que una de las consecuencias más importantes de la evolución informática ha sido precisamente la de integrar efectivamente todos estos elementos, haciendo cada vez más difusas las fronteras que antes los separaban.

También es bueno señalar que los cambios en la metodología de recolección de los datos han ocurrido en paralelo con otros, tanto o más importantes, en relación con lo que los usuarios esperan del proceso. El llamado “Plan de Tabulaciones”, que alguna vez fue el paradigma que debía guiar los esfuerzos de diseño e instrumentación de la encuesta al interior de una Oficina Nacional de Estadística, ya no se considera como el resultado final del proceso, sino más bien como el punto de partida de una serie de actividades analíticas, que por su naturaleza misma escapan del ámbito de la ONE. El resultado final de la encuesta ya no es un conjunto de tabulaciones sino una base de datos.

En las páginas que siguen se describe la forma en que se ha instrumentado la integración de los criterios de control de calidad apoyados por computadores a las faenas de terreno en las encuestas de nivel de vida, y las consecuencias que el proceso ha tenido sobre la organización del personal de terreno, la gestión central, la programación y el presupuesto de las encuestas. Parte del material que aquí se entrega proviene del Manual de Diseño e Instrumentación de Encuestas LSMS, editado por el Banco Mundial, y próximo a ser traducido al español con la colaboración del BID.¹

¹ Grosh, Margaret, y Juan Muñoz. 1996, *A Manual for Planning and Implementing de Living Standards Measurement Study Survey*. Living Standards Measurement Study Working Paper No. 126. World Bank, Washington, D.C.

II. ORGANIZACIÓN DEL TRABAJO DE CAMPO EN LAS ENCUESTAS DE NIVEL DE VIDA

En las encuestas de nivel de vida, las operaciones de terreno son ejecutadas por equipos de terreno autónomos e independientes. Cada equipo está encabezado por un supervisor y está compuesto de dos o tres encuestadores, un operador de ingreso de datos y un conductor. El equipo puede además incluir a un antropometrista, en los casos en que la encuesta requiere que se registren datos de peso y estatura. Cada equipo está dotado de su propio vehículo y de un computador con impresora.

1. Esquema de dos rondas.

Las primeras encuestas LSMS, realizadas desde 1985, basaron la organización de las actividades de terreno en un esquema de entrevistas de cuatro semanas, distribuidas en un período de doce meses. En cada ciclo de cuatro semanas, el equipo realiza las entrevistas en dos localidades (Unidades Últimas de Área). El operador trabaja con su computador (un PC de escritorio) en una Oficina Regional mientras que el resto del equipo viaja entre la oficina y las dos localidades.

El cuestionario de hogares está dividido en dos partes (o *rondas*) de duración aproximadamente igual. En la primera semana del ciclo se completa la primera ronda de los cuestionarios en los hogares de la localidad A. En la segunda semana se completa la primera ronda en la localidad B, mientras que el operador registra los datos recogidos la semana precedente en A, detectando los errores e inconsistencias existentes. Durante la tercera semana, los encuestadores regresan a la unidad A para completar la segunda ronda de los cuestionarios y corregir los errores detectados en los datos de la primera, mientras el operador registra los datos de la localidad B. En la cuarta semana, los encuestadores regresan a la localidad B para completar la segunda ronda de los cuestionarios y corregir los errores detectados en la primera.

Durante la primera visita al hogar, el encuestador completa la lista de miembros y establece citas con quienes deberán informar sobre los otros módulos del cuestionario. Para las secciones individuales (educación, salud, empleo, etc.) los informantes ideales son las personas mismas, con la natural excepción de los niños pequeños, en cuyo caso se trata de entrevistar a un adulto responsable. Habitualmente, los encuestadores tratan de completar la lista de miembros en todos los hogares de la localidad durante el primer día o los primeros dos días, dejando las entrevistas individuales para el resto de la semana, a menos que los informantes prefieran hacerlo de inmediato. Para las secciones del cuestionario que se refieren al hogar como un conjunto (consumo, agricultura, actividades independientes, etc.), se identifica al informante más adecuado durante la primera ronda con el fin de entrevistarlos durante la segunda.

Esta práctica se justifica por varias razones. Por una parte, promueve que se entreviste a los informantes más calificados para cada módulo de manera individual,

reduciendo así la incidencia de las respuestas por poder. Por otra parte, permite descomponer la encuesta entera, que puede llegar a requerir de varias horas por hogar, en una serie de mini-entrevistas más soportables, de unos 30 minutos en promedio, reduciendo la fatiga de los informantes. La programación de entrevistas de acuerdo con la conveniencia de los informantes minimiza además el rechazo y otras formas de no respuesta.

2. Equipos móviles.

La organización de equipos móviles con ingreso de datos en ubicaciones fijas descrita arriba se puso en práctica por primera vez en 1985, en Costa de Marfil, y ha sido aplicada luego en Perú, Ghana, Paquistán y varios otros países. Las circunstancias particulares de ciertos países, junto con el reciente advenimiento de los computadores portátiles están trayendo consigo una variante del esquema, que posiblemente se adoptará con frecuencia cada vez mayor en las encuestas futuras.

Esta variante consiste en dotar a los equipos de terreno de computadores e impresoras portátiles, y hacer viajar al operador de ingreso de datos junto con el resto del equipo. El operador ingresa cada tarde los datos recogidos durante el día por los encuestadores, e informa permanentemente al supervisor sobre las inconsistencias detectadas, para que éstas se corrijan durante las próximas visitas a los hogares. El programa de entrevistas de cada hogar ya no necesita acomodarse al patrón de dos rondas. En la práctica el equipo visita cada localidad una sola vez, partiendo cuando la información recogida en todos los hogares se ha ingresado y se encuentra consistente. Si dos localidades que deban visitarse sucesivamente según el calendario de la encuesta se encuentran vecinas, el equipo puede de hecho viajar directamente de una a otra, sin necesidad de volver a la oficina regional.

El esquema se impuso por primera vez en 1995 en Nepal, donde las dificultades de transporte impedían hacer viajar a los equipos semanalmente a una oficina regional. Se temió al principio que los niveles de supervisión central pudieran degradarse, por las dificultades para saber exactamente dónde estaba cada equipo en cada instante, pero esto se solucionó en la práctica reforzando el equipo de control central e imponiendo ciertos chequeos en el programa de actividades de los equipos de terreno. Un problema mayor fue el asegurar energía para los computadores en un país mayoritariamente no electrificado. Esto se consiguió dotando a los operadores de paneles solares y baterías de alto desempeño. Como se percibía el riesgo de robos del material informático, averías y otros desastres, se complementó la dotación de los trece equipos de terreno con tres juegos adicionales de respaldo, pero los temores en este sentido resultaron infundados: después de un año completo de faenas de terreno todas las máquinas regresaron a la capital en buenas condiciones de funcionamiento.

Más recientemente, el esquema de equipos móviles se ha aplicado con éxito a la Encuesta de Pobreza Rural en dos provincias de Argentina (Salta y Misiones, entre mayo y julio de 1996), y se piensa ponerlo también en práctica en la Encuesta de Niveles de Vida de Irán (1997) y en la Encuesta de Ingresos y Gastos de los Hogares de Honduras (en 1997-98).

III. VENTAJAS DE LA INTEGRACIÓN DE LA INFORMÁTICA AL TERRENO

La experiencia ha mostrado que el esquema organizacional basado en la integración efectiva de las operaciones de terreno y las de gestión y control de calidad de los datos ofrece enormes ventajas. La más importante y directa de ellas es que mejora significativamente la calidad de la información, porque permite corregir los errores e inconsistencias mientras los encuestadores están aún en el terreno, en lugar de confiar en procesos posteriores de “limpieza” en oficina, por medio de programas de control. Aparte de lo caros y demorosos, estos procesos conducen, en el mejor de los casos, a bases de datos internamente consistentes, pero sin garantías de reflejar la realidad observada en los hogares. La incertidumbre se debe a la multitud de decisiones, por lo general no documentadas, que tienen que tomar personas alejadas del terreno en el tiempo y el espacio.

En segundo lugar, la integración del ingreso de datos a las faenas de terreno permite disponer de una base de datos apta para el análisis de manera muy oportuna; habitualmente, pocas semanas después de las últimas entrevistas. Los beneficios de esta rapidez se reducen algo por la necesidad de distribuir las entrevistas durante un año, en vez de concentrarlas en unos pocos meses, pero es importante notar que la distribución anual evita que las conclusiones de la encuesta resulten afectadas por sesgos estacionales. La programación en doce meses también reduce la cantidad de equipos de terreno necesarios, lo que permite seleccionar al personal de manera más exigente y cuidadosa, entrenarlo en forma centralizada y uniforme, y supervisarlos efectivamente.

Otra ventaja indirecta del control de calidad de los datos en terreno es que promueve la aplicación de criterios uniformes entre todos los equipos y durante todo el período de recolección de datos; algo que es muy difícil de conseguir en la práctica con los métodos tradicionales. El computador actúa de hecho como un elemento de apoyo (incansable e incorruptible) a las misiones del supervisor del equipo.

Naturalmente, la integración de la informática al terreno no puede, por sí sola, garantizar la calidad de los datos. Los esquemas clásicos de supervisión, y la selección y capacitación cuidadosas del personal de campo juegan en las encuestas de nivel de vida un papel tanto o más importante que en las otras encuestas. La forma en que se han aplicado estos otros elementos se explica en la parte final de este documento.

IV. CRITERIOS DE CONTROL DE CALIDAD

Los datos registrados en los cuestionarios deben someterse a cinco tipos de control de calidad al momento de ser ingresados: chequeos de rango, verificación contra tablas de referencia, controles de flujos, controles de consistencia y chequeos tipográficos.

1. Chequeos de rango.

Para cada variable de la encuesta debe establecerse el conjunto de valores posibles. Las variables cualitativas sólo pueden tomar uno de ciertos valores predefinidos, de acuerdo con una codificación preestablecida, por ejemplo, el sexo sólo puede codificarse como “1” (masculino) o “2” (femenino). Las variables cronológicas sólo pueden contener fechas válidas; por ejemplo el 29 de febrero sólo puede aparecer en años bisiestos. Las variables cuantitativas sólo pueden tomar valores en un intervalo; por ejemplo, la edad de una persona debe ser entre 0 y 99 años.

El programa de ingreso debe acusar al operador el ingreso de datos anómalos por medio de señales audibles y visibles. para que éste pueda corregirlo de inmediato si se trata de un error tipográfico. Sin embargo, el operador debe ser capaz de forzar el ingreso de un dato erróneo cuando éste aparece de esa forma en el cuestionario. En ese caso el programa debe generar un mensaje impreso que permita al supervisor ordenar que se verifique la situación durante una nueva visita al hogar. En otras palabras, el operador no debe tratar de adivinar cuál debería haber sido el valor correcto: esa es la misión del encuestador.

2. Verificación contra tablas de referencia.

La consistencia de las medidas antropométricas (edad, peso y estatura) debe controlarse por medio de las tablas de referencia de la Organización Mundial de la Salud. Los valores que se escapen más de tres desviaciones típicas de los valores medios de la tabla son probables errores de registro y deben acusarse como tales para que se repitan las medidas.

Una verificación similar permite controlar la verosimilitud de los datos de consumo de alimentos. Usando una tabla de composición de alimentos se debe calcular que el consumo de calorías por persona en el hogar esté dentro de márgenes admisibles. También debe controlarse que el aporte de cada alimento específico, o de ciertos grupos de alimentos, no exceda cierta proporción de la ingesta total del hogar.

También deben imponerse límites mínimos y máximos, y de manera específica para cada rúbrica, a los precios unitarios que puedan aparecer, implícita o explícitamente, en las partes del cuestionario destinadas al registro de gastos.

3. Controles de flujo.

Los cuestionarios contienen habitualmente instrucciones para que el encuestador formule o no ciertas preguntas dependiendo de la respuesta a una pregunta precedente. Por ejemplo, el monto de la factura de teléfono sólo debe registrarse en los hogares que disponen de este servicio.

El programa de ingreso de datos debe verificar que el encuestador siguió correctamente las instrucciones de flujo, pero sin guiar al operador en el proceso: la misión

es registrar fielmente lo que el encuestador registró en el cuestionario. Por ejemplo, después de ingresar el dato sobre la existencia de teléfono en el hogar, el programa debe presentarle al operador el campo destinado al monto de la factura, aunque la respuesta anterior sea que no hay teléfono. Si en el cuestionario aparece un monto a pesar de ello, el programa deberá señalar un error de flujo y el encuestador tendrá que verificar en terreno la naturaleza del problema. Si el programa siguiera automáticamente el flujo, el problema podría pasar desapercibido.

4. Controles de consistencia.

Los controles de consistencia verifican que las respuestas a una pregunta sean coherentes con las respuestas a otras preguntas. Los casos más simples y numerosos son los controles entre dos o más variables relativas a la misma unidad estadística (por ejemplo, la edad y la fecha de nacimiento de una persona). Sin embargo, son los controles más complejos los que otorgan más oportunidades para verificar globalmente la calidad de la información registrada en el cuestionario como un todo. Entre los más importantes están:

La consistencia demográfica del hogar. -Consiste en verificar la coherencia entre las edades, sexos y vínculos de parentesco entre los miembros del hogar. Por ejemplo, los padres deben ser al menos 15 años mayores que sus hijos, los cónyuges deben ser de diferentes sexos, etc.

La consistencia de las ocupaciones. La presencia o ausencia de ciertos módulos debe ser consistente con las ocupaciones declaradas por los individuos. Por ejemplo, la sección de agricultura debe estar presente si y sólo si algún miembro se declara como agricultor en la sección de actividades.

La consistencia entre la edad y otras características individuales, tales como el estado civil, la relación de parentesco con el jefe de hogar o el nivel educativo alcanzado.

Gastos. La presencia de ciertos gastos debe ser consistente con la información registrada en otros módulos del cuestionario. Por ejemplo, el gasto en artículos escolares o matrículas debería ser consistente con la presencia de escolares en el hogar.

5. Chequeos tipográficos.

Uno de los puntos débiles de la digitación integrada a las faenas de terreno es la dificultad para instrumentar el concepto de “verificación” (o doble digitación) de los datos, debida a la necesidad de trabajar con un sólo operador en una sola máquina. Esta condición entraña el riesgo de introducir errores de difícil detección, particularmente en los módulos destinados al registro de gastos. Por ejemplo, es probable que un gasto en carne de \$14 pase desapercibido si se digita incorrectamente como \$41 (mientras que la misma transposición de dígitos en la edad de una persona de 41 a 14 años, es susceptible de provocar más de alguna incoherencia reveladora).

Es posible reproducir el efecto de la verificación de antaño, programando el computador para que obligue al operador a repetir el ingreso de algunos datos, pero esta opción sólo puede aplicarse de manera excepcional, para ciertas variables o módulos muy críticos. En la práctica es preferible instrumentar soluciones alternativas, al nivel del diseño del cuestionarios. La mejor y la más simple de estas soluciones es la introducción de líneas de totales y subtotales, en todas las páginas del cuestionario que se presenten de manera tabular, y para todas las columnas que sea sensato sumar verticalmente. Los encuestadores pueden calcular esos totales con una calculadora de bolsillo y registrarlos a mano en las líneas destinadas a ese efecto; el programa de ingreso de datos puede luego verificar que los totales cuadren y advertir al operador de los posibles errores de digitación. Otra solución simple es dotar a los códigos de las rúbricas de gasto de un dígito verificador.

V. REQUERIMIENTOS PARA LA ORGANIZACIÓN CENTRAL

La integración del control de calidad informático a las faenas de terreno impone ciertas condiciones a la organización y funcionamiento del grupo responsable de la concepción y la supervisión central de la encuesta.

El grupo central debe estar dirigido por un Director de Proyecto, secundado por un Jefe de Operaciones de Terreno y un Jefe de Operaciones Informáticas. Dependiendo de las condiciones de cada país, puede ser necesario reforzar este grupo por dos jefes asistentes y, eventualmente, por un responsable administrativo y una secretaria. Este grupo debe organizarse para funcionar *como un equipo*, dedicado a la encuesta de manera exclusiva y bajo el liderazgo único del Director de Proyecto durante todas las etapas de preparación y operación. Esto es particularmente importante -y a veces difícil de conseguir- en las Oficinas Nacionales de Estadística que aún se encuentran estructuradas según los esquemas departamentales tradicionales. Si en lugar de responder ante el Director de Proyecto los miembros del grupo central de la encuesta se ven forzados a responder separadamente ante los directores de estos departamentos (por ejemplo, el Jefe de Operaciones Informáticas al director del departamento de procesamiento de datos, el Jefe de Operaciones de Terreno al director del departamento de encuestas, etc.), será muy difícil asegurar que todos los detalles del diseño se completen a tiempo y ensamblen como es debido.

VI. CONSECUENCIAS PRESUPUESTARIAS

La integración de la informática al terreno tiene incidencias en el presupuesto de la encuesta, aunque -contrariamente a lo que pudiera suponerse- el costo de los computadores y el resto del material informático no es lo más importante.

En efecto, la dotación requerida por cada equipo de terreno (consistente de un computador, una impresora y, algunos otros suministros) puede conseguirse hoy día por

menos de \$2.500; cifra relativamente modesta si se la compara, por ejemplo, con los costos anuales de transporte del equipo (combustibles, mantenimiento, depreciación del vehículo, etc.).

Más importantes son los costos generados por la elevada tasa de supervisión implícita en el esquema (un supervisor cada dos o tres encuestadores, en lugar de uno por cinco o más como es habitual), por la necesidad de movilizar en el terreno a los operadores de ingreso de datos junto con los encuestadores, en lugar de mantenerlos en estaciones fijas, y por la necesidad de nominar a un Jefe de Operaciones Informáticas (y tal vez, también a un asistente), durante todas las etapas de diseño de la encuesta, levantamiento de la información, y consolidación y documentación de la base de datos.

Dependiendo de las condiciones locales, el diseño y depuración del programa de ingreso de datos podría también requerir asistencia técnica específica.

VII. OTROS ASPECTOS VINCULADOS CON LA CALIDAD DE LOS DATOS

Entre los muchos otros aspectos metodológicos de las encuestas de nivel de vida que tienen incidencia sobre la calidad de los datos recogidos, pueden mencionarse los siguientes:

1. Criterios de supervisión.

La integración del control de calidad informático al terreno no es un reemplazo sino un apoyo a las funciones clásicas del supervisor. Si bien esas funciones y sus efectos sobre la calidad de los datos son conocidas de antaño, las encuestas de nivel de vida han sido tal vez más cuidadosas que lo habitual en instrumentarlas de manera efectiva. Merece particular mención en este sentido la ejecución de dobles visitas de control a una proporción relativamente importante de los hogares (20 a 30 por ciento), y la puesta en operación de formularios y procedimientos para permitir que esta misión de los supervisores pueda ser a la vez supervisada por el equipo central de gestión de la encuesta.

2. Cuestionario único.

Las encuestas de nivel de vida siempre han usado un cuestionario único para cada hogar, organizando el registro de informaciones individuales en páginas de formato uniforme, con una línea para cada persona. Esto ha evitado los problemas de compaginación y de identificación, habituales cuando se usan cuestionarios separados para hogares e individuos.

Tampoco se han usado cuestionarios distintos para áreas urbanas y rurales, sino un cuestionario uniforme, con preguntas filtro para detectar la necesidad de aplicar los módulos agrícolas. Esto ha evitado el problema logístico de manejar dos stocks de cuestionarios

distintos y la necesidad de arbitrar en el caso de los hogares ubicados en localidades calificadas como urbanas pero que desempeñan actividades agrícolas.

3. Entrenamiento.

La formación de los encuestadores y operadores de ingreso de datos requiere de un período de entrenamiento de cuatro semanas: dos semanas de formación teórica intercaladas con dos semanas de trabajo práctico supervisado. Habitualmente se entrena a más encuestadores de los necesarios para la encuesta, reservando la selección final para después de observar el desempeño de los candidatos durante el entrenamiento.

4. Ensayo con vestuario.

El calendario de la recolección de datos considera siempre una breve interrupción de las actividades, luego de las primeras dos a cuatro semanas de trabajo en terreno. Durante esta interrupción, se reúne a los supervisores de los equipos y, cuando las condiciones lo permiten, a los equipos completos, para evaluar los problemas encontrados y hacer los ajustes necesarios en las instrucciones y procedimientos de campo (incluyendo depuraciones en el programa de ingreso de datos). Aunque por razones tácticas no se ha considerado aconsejable hacerlo explícito, la intención, ha sido siempre la de considerar las primeras semanas como un ensayo de hecho, reservándose la opción de desechar los datos recogidos en ese intervalo y prolongar al encuesta por un tiempo equivalente después del período programado de doce meses, en el caso que los problemas encontrados lo ameriten. Hasta ahora no ha sido necesario hacerlo.

5. Substitución de hogares.

Aunque está claro que la substitución de hogares no soluciona el problema de los rechazos y otras formas de no-respuesta, las encuestas nivel de vida han optado habitualmente por entregar a los supervisores una breve lista de hogares de reserva (de dos a cuatro por localidad), para substituir . en los casos mencionados. La idea es suprimir el incentivo perverso de que los encuestadores vean premiados sus fracasos con menos trabajo. Las substituciones se registran en todo caso como tales en las bases de datos, para dejar a los analistas la opción de considerarlas o no en sus cálculos.

6. Definición de identificación de las unidades estadísticas.

Las encuestas de niveles de vida recogen información sobre varias unidades estadísticas diferentes (hogares, viviendas, personas, cultivos, parcelas, rúbricas de gasto, etc.). Dos elementos claves del buen diseño de un cuestionario son: (1) hacer explícito en cada sección o módulo cuál es la unidad estadística sobre la que se recoge información y (2) asegurar que cada unidad estadística quede identificada de manera única y sin ambigüedades. La segunda condición puede descomponerse en (2a) asegurar la identificación unívoca de los hogares (por medio de un número único para cada hogar, idealmente no muy largo) y (2b) incluir en el diseño del cuestionario identificadores

explícitos (idealmente preimpresos) para cada unidad observada; por ejemplo, códigos de identificación para los miembros del hogar, códigos para los ítems presupuestarios, etc.

Es fundamental que el programa de ingreso de datos -y todos los procesos informáticos posteriores que conduzcan a la formación de las bases de datos finales- conserven en todo momento la identificación única de cada unidad estadística. En las encuestas de nivel de vida esto se ha conseguido asignando a cada unidad observada un identificador en tres partes. La primera es la “clase de registro”, que aparece al principio de cada registro y dice de que parte del cuestionario proviene la información. La clase de registro viene seguida -en todos los registros- por el número del hogar. En la mayoría de los registros sigue un tercer identificador para diferenciar las distintas unidades estadísticas del mismo tipo que puede contener un hogar (por ejemplo, el código del individuo).

