

Distr.
RESTRINGIDA
LC/R.1826
6 de julio de 1998
ORIGINAL: ESPAÑOL

CEPAL

Comisión Económica para América Latina y el Caribe



1^{er} TALLER REGIONAL

**PLANIFICACIÓN Y DESARROLLO DE ENCUESTAS DE HOGARES
PARA LA MEDICIÓN DE LAS CONDICIONES DE VIDA**

Aguascalientes, México, 1^o al 3 de abril de 1998



Este documento no ha sido sometido a revisión editorial.

98-7-532

PRESENTACIÓN

La reducción de los elevados niveles de pobreza e inequidad social que prevalecen en la mayoría de los países de América Latina y el Caribe constituye una de las preocupaciones prioritarias de los gobiernos de la región. De allí que revista gran importancia la necesidad de evaluar diferentes aspectos de la gestión pública en materia de política social, a fin de conocer la eficiencia y resultados de los esfuerzos que se vienen realizando orientados a mejorar las condiciones de vida de la población.

Para apoyar este objetivo se requiere contar con información relevante, confiable y oportuna, en el marco de sistemas de información estadística que permitan conocer, analizar y dar seguimiento a la evolución de los niveles de bienestar de los distintos grupos sociales, y especialmente de aquéllos que sufren situaciones de pobreza o de exclusión social.

Durante las últimas dos décadas se han desarrollado en los países latinoamericanos importantes acciones destinadas a fortalecer sus sistemas de información, y en particular aquellos que se vinculan con la generación de datos a partir de las encuestas de hogares. Sin embargo, a pesar de los avances logrados, aún se evidencian algunas limitaciones en la calidad, cobertura geográfica, amplitud temática y oportunidad de la información, lo cual dificulta su pleno aprovechamiento y utilización en el diseño, ejecución y evaluación de impacto de las medidas de política.

En este contexto, el Banco Interamericano de Desarrollo (BID), el Banco Mundial y la Comisión Económica para América Latina y el Caribe (CEPAL), con la participación directa de los países de la región, han establecido el "**Programa para el Mejoramiento de las Encuestas y la Medición de las Condiciones de Vida en América Latina y el Caribe**" (MECOVI). El objetivo central de este Programa es apoyar a los países en la tarea de generar información adecuada y de alta calidad acerca de las condiciones de vida de los habitantes de la región, en cuanto a su contenido, alcance, confiabilidad, actualidad y relevancia para el diseño de políticas.

Entre sus actividades el Programa contempla la realización -bajo la coordinación de la CEPAL- de un conjunto de talleres que buscan contribuir a mejorar la estimación y el análisis de los indicadores sociales que se obtienen de las encuestas, fortaleciendo la capacidad institucional en cada país para implementar sistemas de encuestas de hogares y utilizar la información que de ellos se deriva.

El primero de estos talleres tuvo lugar en la ciudad de Aguascalientes, México, y contó con la colaboración del Instituto Nacional de Estadística, Geografía e Informática (INEGI) de ese país. El tema abordado fue la "**Planificación y Desarrollo de Encuestas de Hogares para la Medición de las Condiciones de Vida**".

En este libro se reúnen los documentos presentados al taller, además de una síntesis de los temas tratados y de sus principales conclusiones. Con ello se persigue poner a disposición de los interesados un material que esperamos sea de gran utilidad para todos aquellos técnicos nacionales que se desenvuelven en el área del planeamiento y ejecución de las encuestas de hogares, o bien son usuarios de la información que éstas generan. Al mismo tiempo, confiamos en que este esfuerzo contribuya a incentivar la necesaria reflexión que debemos mantener y profundizar sobre esta materia, a efectos que mediante el intercambio y la cooperación horizontal entre los países se logre avanzar en el objetivo de perfeccionar el diseño metodológico, la recolección y el análisis de la información de estas encuestas, elevando así su utilidad para la formulación de políticas orientadas a mejorar las condiciones de vida de la población de América Latina.

Los documentos han sido organizados siguiendo los grandes temas en que se estructuró la agenda de la reunión. En primer lugar, figuran aquellos referidos a la **planificación y diseño de encuestas de hogares**, que comprende los aspectos relativos a la concepción del sistema integrado de encuestas, los marcos de muestreo, la elaboración de la muestra, las encuestas de panel para estudios longitudinales y el diseño de los cuestionarios. Luego, se incluyen aquellos que guardan relación con el **trabajo de campo y control de calidad**, destacando el levantamiento y control de la información y los procesos de imputación de datos. Por último, está el tema de **difusión y explotación de la información**, con especial referencia al cálculo de los errores de muestreo en encuestas complejas.

Juan Carlos Feres

Coordinador MECOVI - CEPAL

ÍNDICE

	<u>Página</u>
Agenda	5
PROGRAMA PARA EL MEJORAMIENTO DE LAS ENCUESTAS Y LA MEDICIÓN DE LAS CONDICIONES DE VIDA EN AMERICA LATINA Y EL CARIBE (MECOVI)	9
RESUMEN Y CONCLUSIONES.....	15
PLANIFICACIÓN Y DISEÑO DE ENCUESTAS DE HOGARES	27
HACIA UN SISTEMA INTEGRADO DE ENCUESTAS DE HOGARES EN LOS PAÍSES DE AMERICA LATINA. Juan Carlos Feres.....	29
ALGUNAS CONSIDERACIONES SOBRE EL DISEÑO ESTADÍSTICO DE LAS ENCUESTAS DE HOGARES. Juan Carlos Feres y Fernando Medina.....	49
PROPUESTA PARA EL DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA DE INFORMACIÓN GEOGRÁFICO PARA APOYAR LA GESTIÓN ADMINISTRATIVA Y OPERATIVA DE LAS ENCUESTAS DE HOGARES. Sergio Sommerville.....	63
TAMAÑO OPTIMO DE MUESTRA EN ENCUESTAS DE PROPÓSITOS MULTIPLES. Fernando Medina.....	117
LAS ENCUESTAS DE PANEL PARA EL ESTUDIO DE LAS CONDICIONES DE VIDA: BONDADES Y RESTRICCIONES. Fernando Medina	149
CUESTIONARIOS UTILIZADOS EN LOS PAÍSES DE AMÉRICA LATINA Y EL CARIBE PARA LA EJECUCIÓN DEL PROGRAMA PERMANENTE DE ENCUESTAS DE HOGARES. María de la Luz Avendaño y Gabriela López.....	171
LECCIONES PARA EL DISEÑO DEL CUESTIONARIO. Margaret Grosh.....	195
DISEÑO DEL CUESTIONARIO DE LA “PESQUISA NACIONAL POR AMOSTRA DE DOMICILIOS - PNAD”. Rosângela Antunes Pereira Almeida (IBGE).....	205
ENCUESTAS DE PROPÓSITOS MÚLTIPLES: LECCIONES Y DIRECCIONES PARA SU MEJORAMIENTO. Edmundo Berumen	209

TRABAJO DE CAMPO Y CONTROL DE CALIDAD.....	231
CONTROL DE CALIDAD EN EL DISEÑO Y EJECUCIÓN DE ENCUESTAS DE HOGARES. Juan Carlos Feres y Fernando Medina	233 ✓
DOCUMENTACIÓN METODOLÓGICA DE LAS ENCUESTAS DE HOGARES. Gilberto Moncada.....	253 ✓
ASPECTOS DEL DISEÑO Y LA PUESTA EN MARCHA DE LAS ENCUESTAS QUE INCIDEN EN LA CALIDAD DE LOS DATOS RECOGIDOS. Edmundo Berumen y Juan Muñoz.....	271 ✓
¿COMO 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	277 ✓
CONTROL DE CALIDAD. Kinnon Scott.....	289 ✓
ENCUESTA INTEGRADA DE HOGARES EN PARAGUAY. Dirección General de Estadística y Censos (DGEC).....	297 ✓
IMPUTACIÓN DE DATOS. Guillermo Ramírez (OCEI - Venezuela)	305 ✓
DIFUSIÓN Y EXPLOTACIÓN DE INFORMACIÓN	313
LOS ERRORES DE MUESTREO EN LAS ENCUESTAS COMPLEJAS. Fernando Medina.....	315 ✓
Lista de Participantes.....	349

PROGRAMA PARA EL MEJORAMIENTO DE LAS ENCUESTAS Y LA MEDICION DE LAS
CONDICIONES DE VIDA EN AMERICA LATINA Y EL CARIBE (MECOVI)
BID - BANCO MUNDIAL - CEPAL

**TALLER REGIONAL SOBRE PLANIFICACIÓN Y DESARROLLO DE ENCUESTAS DE
HOGARES PARA LA MEDICIÓN DE LAS CONDICIONES DE VIDA**

Aguascalientes, México, 1 al 3 de abril de 1998

INEGI

CEPAL

Agenda

Miércoles 1°	
09:00 - 09:30	Registro de participantes
09:30 - 09:45	Palabras de bienvenida y presentación del Taller (<i>INEGI, CEPAL</i>)
09:45 - 10:00	Presentación del Programa MECOVI (<i>BID-Banco Mundial-CEPAL</i>)
PLANIFICACIÓN Y DISEÑO DE ENCUESTAS DE HOGARES	
TEMA 1	Sistema integrado de encuestas de hogares
10:00 - 10:20	La conformación de los SIEHs en América Latina (<i>CEPAL</i>)
10:20 - 10:35	Las encuestas de hogares en México (<i>México</i>)
10:35 - 11:00	Discusión general Moderador: <i>Galo Arias (INEC-Ecuador)</i>
11:00 - 11:30	Café
TEMA 2	Marco de muestreo
11:30 - 11:50	El marco de muestreo y los sistemas de información geográfica (SIGs) (<i>CEPAL</i>)
11:50 - 12:10	Cambios de la muestra ante nuevos marcos muestrales (<i>Costa Rica</i>)
12:10 - 12:30	Los SIGs en el contexto de las encuestas de hogares (<i>Consultor CEPAL</i>)
12:30 - 12:50	Generación de cartografía automatizada (<i>México</i>)
12:50 - 14:00	Discusión general Moderador: <i>Guillermo Ramírez (OCEI - Venezuela)</i>
14:00 - 15:30	Almuerzo
TEMA 3	Diseño de la muestra: Tamaño y selección
15:30 - 15:50	Tamaño de muestra en la Encuesta Nacional de Hogares (<i>Perú</i>)
15:50 - 16:10	Tamaño óptimo de muestra en encuestas de propósitos múltiples (<i>CEPAL</i>)
16:10 - 16:30	Discusión general
16:30 - 16:45	Café

16:45 - 18:00	Experiencia de países Moderador: <i>Haeduck Lee (Banco Mundial)</i>
Jueves 2	
TEMA 4	Las encuestas de panel para estudios longitudinales
09:00 - 09:15	La rotación de la muestra en las Encuestas de Hogares (<i>Argentina-INDEC</i>)
09:15 - 09:30	Diseños de panel (<i>El Salvador</i>)
09:30 - 09:45	Encuestas de panel: Bondades y restricciones (<i>CEPAL</i>)
09:45 - 10:00	Discusión general Moderador: <i>Alberto Padilla (INE - Perú)</i>
TEMA 5	Diseño de cuestionarios
10:00 - 10:20	Diseño óptimo de cuestionarios (<i>Banco Mundial</i>)
10:20 - 10:40	Diseño de cuestionario en la encuesta CASEN (<i>Chile - MIDEPLAN</i>)
10:40 - 11:00	Diseño de cuestionario en la encuesta PNAD (<i>Brasil - IBGE</i>)
11:00 - 11:30	Café
11:30 - 11:50	Los cuestionarios de las Encuestas de Hogares en América Latina (<i>CEPAL</i>)
11:50 - 12:30	Discusión general Moderador: <i>Gilberto Moncada (BID)</i>
TRABAJO DE CAMPO Y CONTROL DE CALIDAD	
TEMA 6	Levantamiento y control de la información
12:30 - 12:50	Control de calidad en el diseño y ejecución de encuestas (<i>CEPAL</i>)
12:50 - 13:10	Control de calidad (<i>Banco Mundial</i>)
13:10 - 13:30	Trabajo de campo y validación de la información (<i>Paraguay</i>)
13:30 - 14:00	Discusión general Moderador: <i>Juan Carlos Feres (CEPAL)</i>
14:00 - 15:30	Almuerzo
TEMA 6	(continuación)
15:30 - 15:50	Documentación metodológica de las encuestas (<i>BID</i>)
15:50 - 16:30	Evaluación de consistencia con otras fuentes (<i>CEPAL, Chile - INE, Colombia</i>)
16:30 - 16:45	Café
16:45 - 17:15	(Continuación) (<i>Uruguay, Chile - Registro Civil</i>)
17:15 - 18:00	Discusión general Moderador: <i>Mario Jacobs (INE - Guatemala)</i>

Viernes 3	
TEMA 7	Los procesos de imputación de datos
09:00 - 09:45	Procedimientos de imputación de datos faltantes (<i>Honduras, Nicaragua - INEC, Venezuela - OCEI</i>)
09:45 - 10:00	Discusión general Moderador: <i>Saúl García (DGEC - El Salvador)</i>
DIFUSIÓN Y EXPLOTACIÓN DE INFORMACIÓN	
TEMA 8	Difusión de la información
10:00 - 11:00	Mecanismos de difusión de la información (<i>Bolivia, Cuba, Ecuador, Guatemala, Panamá, República Dominicana - Banco Central</i>)
11:00 - 11:30	Café
TEMA 9	Explotación de la información
11:30 - 11:50	Los errores de muestreo en encuestas complejas (<i>CEPAL</i>)
11:50 - 13:15	Usos de la información para el diseño y evaluación de políticas (<i>Argentina - SIEMPRO, Brasil - IPEA, Chile - MIDEPLAN, Nicaragua - MITRAB, Venezuela - MINFAM</i>)
13:15 - 14:00	Discusión general Moderador: <i>Marta Becker (DGEC - Paraguay)</i>
14:00 - 16:00	Almuerzo
LINEAMIENTOS DE ACCION	
16:00 - 17:50	Discusión general Moderadores: <i>Juan Carlos Feres (CEPAL), Haeduck Lee (Banco Mundial), Gilberto Moncada (BID)</i> ,
17:50 - 18:00	Clausura



**PROGRAMA PARA EL MEJORAMIENTO DE LAS ENCUESTAS
Y LA MEDICIÓN DE LAS CONDICIONES DE VIDA EN
AMERICA LATINA Y EL CARIBE
(MECOVI)**

ÍNDICE

	<u>Página</u>
I. ANTECEDENTES Y JUSTIFICACIÓN.....	11
II. OBJETIVO	11
III. ACTIVIDADES	12
a) Actividades a nivel de países beneficiarios.....	12
b) Actividades regionales	12
IV. ESTADO DE SITUACIÓN	13

I. ANTECEDENTES Y JUSTIFICACIÓN

La reducción de la pobreza y el incremento de la equidad social se encuentran entre los principales objetivos de los gobiernos de la región y de las instituciones multilaterales. Aunque mucho se ha aprendido en los años recientes, persiste la necesidad de conocer mejor el impacto de las políticas gubernamentales sobre el bienestar de los hogares, sobre todo de los más pobres, de América Latina y el Caribe. Para ello es imprescindible contar con información precisa sobre las características y comportamiento de los hogares: por ejemplo, fuentes de ingreso, ubicación geográfica, características socio-demográficas, hábitos alimentarios, uso de servicios sociales, etc. Esta tarea requiere disponer de información a nivel de los hogares y comunidades de cobertura nacional, comparable en el tiempo, y que sea simultáneamente confiable, oportuna y relevante.

Aunque aparentemente se dispone de un gran número de encuestas a nivel de los hogares en la región, el problema es que pocas combinan las características de ser al mismo tiempo confiables, oportunas, comparables en el tiempo y que incluyan toda la información requerida para llevar a cabo el análisis del impacto de las políticas sobre los grupos más pobres. Las deficiencias varían según el caso, pero las que se observan con mayor frecuencia son:

- *Cobertura insuficiente:* Por ejemplo en algunos países de la región no se dispone de encuestas de cobertura nacional y sólo se cubre el área urbana.
- *Cuestionarios con propósitos restringidos:* Por ejemplo, las encuestas de hogares son primariamente utilizadas para medir el empleo y el desempleo.
- *Problemas de confiabilidad de los datos.*
- *Integración insuficiente con otras fuentes de información*
- *Contenido temático limitado.*

Reconociendo la necesidad urgente de poder contar con un sistema de información confiable para medir la pobreza, la desigualdad y los indicadores sociales en todos los países de la región, el BID -conjuntamente con el Banco Mundial, la CEPAL y con la participación directa de los gobiernos de los países -está llevando a cabo una iniciativa con el doble propósito de fortalecer la capacidad institucional en cada país para *implementar* y *analizar* sistemas de encuestas de hogares con propósitos múltiples de alta calidad.

II. OBJETIVO

El objetivo del Programa MECOVI es generar en forma adecuada información sobre las condiciones de vida de la población de la región en términos de su alcance, cobertura, confiabilidad y, sobre todo, relevancia para fines de políticas dirigidas hacia la reducción de la pobreza y el aumento de la equidad social.

III. ACTIVIDADES

El Programa es patrocinado por el BID, el Banco Mundial y la CEPAL, con la participación de las instituciones especializadas en los países partícipes. MECOVI tiene dos componentes principales: a) actividades a nivel de países beneficiarios y b) actividades a nivel regional (todos los países de América Latina y el Caribe).

a) Actividades a nivel de países beneficiarios

A la fecha los países beneficiarios son Argentina, El Salvador, Paraguay y Perú. Dependiendo de la disponibilidad de recursos financieros y de los requerimientos técnicos otros países podrán incorporarse al Programa.

Las actividades en los países beneficiarios están orientadas a:

i) Mejorar el diseño y el proceso de implementación de los sistemas de encuestas de hogares para asegurar una mejor calidad de la información y una amplia cobertura (áreas urbanas y rurales a nivel nacional). De igual modo, establecer mecanismos para la interacción entre productores y usuarios de este tipo de información, y desarrollar la capacidad institucional para lograr la sostenibilidad de la ejecución de encuestas de hogares de calidad.

ii) Crear, mantener y actualizar un banco de datos con la información de las encuestas de hogares de modo oportuno, ágil y de amplio acceso a los usuarios.

iii) Fortalecer las capacidades para el uso de la información de las encuestas de hogares mediante la contratación de estudios específicos; la capacitación de los técnicos nacionales; y el uso de un Fondo de Estudios para promover la investigación de profesionales nacionales.

iv) Mejorar la calidad de las publicaciones de los resultados de las encuestas de hogares y perfeccionar los modos de difusión de las mismas.

v) Organizar cursos de capacitación a nivel nacional sobre temas relacionados con la metodología de implementación y el análisis de las encuestas de hogares.

vi) Organizar talleres a nivel nacional para discutir los resultados de las encuestas y los estudios realizados con la información de las encuestas de hogares.

b) Actividades regionales

Las actividades regionales incluyen:

- i) Talleres regionales diseñados para discutir aspectos metodológicos de las encuestas de hogares y estudios sobre las condiciones de vida en los países de la región.
- ii) Cursos de capacitación sobre la metodología de los sistemas mejorados de encuestas de hogares y el análisis para la definición de políticas.
- iii) Banco de datos con la información primaria de las encuestas de hogares disponibles en los países de la región, accesible a los usuarios.

IV. ESTADO DE SITUACIÓN

a) Talleres regionales: Primer taller (1-3 de abril, 1998), en el INEGI, Aguascalientes-México.

b) Banco de datos: Se tiene disponible un banco de datos para 17 países, que incluye 12 encuestas documentadas y revisadas por la CEPAL.

c) Países:

Paraguay: La operación de campo se inició en agosto 1997 y por el periodo de un año (5,000 viv.). Se está aplicando cuestionarios de hogares, antropometría, comunitario y de precios.

Perú: Se ha ejecutado una encuesta en el IV trim. de 1997 (7,200 viv.). La operación de campo de la encuesta de 1998 se iniciará en mayo o junio de 1998 (24,000 viviendas al año) debido a los problemas originados por El Niño. Se usará cuestionarios de hogares, comunitario y de precios.

El Salvador: La operación de campo se iniciará en abril de 1998 (18,200 viviendas al año). Se usará cuestionarios de hogares, antropometría, comunitario y precios rurales.

Argentina: Recién se ha formalizado el convenio. Se está coordinando el inicio de las actividades de MECOVI.

d) Nuevos países: Se está gestionando la incorporación de Bolivia y Nicaragua en 1998, además, se están iniciando coordinaciones para asegurar la participación de Colombia, Guatemala y Haití. Otros países pueden ser incorporados dependiendo tanto de los requerimientos de apoyo técnico como de la disponibilidad de recursos financieros.

1^{er} TALLER REGIONAL DEL MECOVI

**PLANIFICACIÓN Y DESARROLLO DE ENCUESTAS DE HOGARES
PARA LA MEDICIÓN DE LAS CONDICIONES DE VIDA**

RESUMEN Y CONCLUSIONES

RESUMEN Y CONCLUSIONES

I. Introducción

El **Primer Taller Regional del MECOVI** se llevó a cabo en la ciudad de Aguascalientes, México, entre los días 1º y 3 de abril de 1998. Tuvo por tema la **“Planificación y Desarrollo de Encuestas de Hogares para la Medición de las Condiciones de Vida”** y fue organizado en forma conjunta por la CEPAL y el Instituto Nacional de Estadística, Geografía e Informática (INEGI) de México.

El **objetivo general** de este primer taller fue analizar los aspectos conceptuales, metodológicos y operacionales vinculados al diseño e implantación de un Sistema Integrado de Encuestas de Hogares (SIEH), con especial énfasis en la definición de sus características, requerimientos, bondades y limitaciones en cuanto instrumento de diagnóstico y monitoreo de las condiciones de vida de la población.

Por su parte, los **objetivos específicos** establecidos para la realización de este evento fueron los siguientes:

- i) Conceptualizar el diseño y formulación de un Sistema Integrado de Encuestas de Hogares.
- ii) Evaluar la importancia de un marco de muestreo apropiado para la administración del SIEH, así como la pertinencia de incorporar los Sistemas de Información Geográfica (SIG's) como una herramienta fundamental para el control de calidad y actualización de la información cartográfica de apoyo para la realización del trabajo de campo.
- iii) Profundizar en el diseño de cuestionarios, en términos de la complementariedad temática y conceptual, como asimismo en su aspecto físico y de formato.
- iv) Definir estrategias y mecanismos de control de calidad de las diferentes etapas involucradas en el desarrollo de una encuesta, desde la planeación hasta la distribución y utilización de la información por parte de los usuarios, preservando la confidencialidad de los datos proporcionados por los informantes.

II. Resultado de los Debates

1. Planificación y Diseño de Encuestas de Hogares

La primera parte de la agenda se centró en discutir los elementos técnicos vinculados a la planificación y diseño de las encuestas de hogares. En primer lugar, se abordaron los aspectos metodológicos relativos al diseño de un **Sistema Integrado de Encuestas de Hogares (SIEH)**, con el propósito de precisar sus bases conceptuales, describir sus bondades y limitaciones, e identificar los obstáculos que enfrentan los diferentes países para la puesta en práctica de un sistema de este tipo. En el marco de este debate se hizo referencia también a la necesidad de esclarecer las diferencias conceptuales y operativas existentes entre una estrategia orientada a relevar en un sólo levantamiento una amplia gama de información sobre diversos temas vinculados al bienestar de las personas, y un programa permanente que considera para estos efectos un conjunto de encuestas independientes, complementarias entre sí desde el punto de vista conceptual y temático, que planean su ejecución sincronizadas en el tiempo y comparten recursos humanos y financieros.

Hubo consenso entre los participantes en que no cabe postular un *modelo único* para el diseño de un sistema integrado, ya que la composición, alcance y complejidad de un SIEH está estrechamente vinculada a la realidad específica de cada país, en términos de sus necesidades de información, la infraestructura existente en materia de encuestas, la capacidad institucional y los recursos disponibles. Sin embargo, al mismo tiempo se estableció que es posible identificar ciertos *aspectos básicos comunes* vinculados, por ejemplo, a la complementariedad temática de las encuestas, la construcción y uso de marcos compartidos, el diseño de muestras maestras y la utilización de los insumos de trabajo.

A su vez, se señaló que cada encuesta que forma parte del SIEH posee características particulares, en atención a sus objetivos específicos, diseño muestral, contenido temático, periodicidad, cobertura geográfica y costo. Normalmente los estudios se realizan por medio de *investigaciones específicas* para cada tema, o bien -teniendo como base un tema central- se incorporan al cuestionario de la encuesta las preguntas, o en ocasiones los *módulos*, relativos a los demás fenómenos que se desea investigar. Como se indicó, en todos los casos el diseño estadístico de la encuesta debiera prever esta situación, a fin de que no se produzcan sesgos en los datos y se afecte severamente la precisión de los resultados obtenidos, ya que el error de muestreo está altamente correlacionado con el tamaño de muestra y la varianza del estimador.

Entre los obstáculos y desafíos que plantea la implementación de un SIEH, se señalaron especialmente aquellos relativos al mejoramiento de los marcos muestrales, la vinculación de los diseños de las encuestas y la optimización en el uso de los recursos.

Se puso énfasis en que el diseño y construcción de marcos completos o exhaustivos es una tarea de la mayor importancia para mejorar la cobertura de las encuestas. Asimismo, la actualización permanente y la sistematización del marco que se utiliza para la selección de

la muestra es un requisito indispensable para incrementar la calidad de la información que se genera por medio de las encuestas de hogares.

En el área de utilización de los datos, se estableció que uno de los grandes retos estadísticos que se deben afrontar es el desarrollo de procedimientos teóricos y metodologías de carácter práctico que permitan *interrelacionar los resultados* obtenidos por medio de las diversas encuestas que conforman un sistema integrado. Ello es así dado que, a pesar que en un SIEH se logre la complementariedad temática y se sincronice la ejecución de las diferentes encuestas, no es clara la manera de integrar los resultados para los efectos de la caracterización de fenómenos multivariados, como es el caso del nivel de vida de los hogares.

Por otra parte, se insistió que la formación de un SIEH implica armonizar y complementar distintas encuestas, cada una de las cuales presenta, en función de sus objetivos y diseño, determinadas aptitudes y limitaciones para captar información socioeconómica de los hogares. En ese sentido, el sistema integrado difiere del concepto de encuesta única, politemática, que persigue investigar en un sólo instrumento y bajo el mismo diseño estadístico una variedad extremadamente amplia de fenómenos, pretendiendo satisfacer simultáneamente diversos objetivos analíticos, lo que por lo general dificulta la determinación de un tamaño óptimo de muestra e incide en la confiabilidad de los resultados.

Como se indicó, mientras en el esquema conceptual de un SIEH subyace la premisa que los requerimientos informativos de un cierto ámbito socioeconómico se cubren mediante un conjunto de investigaciones diferentes, pero vinculadas entre sí, en las encuestas multitemáticas de ese tipo -y que por lo general se llevan a cabo de manera esporádica- se intenta recabar en un mismo levantamiento información sobre una amplia gama de temas, normalmente muy distintos entre ellos, que difícilmente pueden ser atendidos con la misma precisión estadística.

En contraste, en el contexto del SIEH la mediciones que se efectúan sobre temas específicos tienen, por lo general, mayor precisión estadística y menor costo de operación, con lo que generan información más confiable y pueden efectuarse con mayor regularidad. Al mismo tiempo, las encuestas permanentes tienen la capacidad de captar información sobre temas adicionales al objetivo central para el que fueron diseñadas, en la medida que se hagan las modificaciones pertinentes en el tamaño de muestra y la conformación del marco de muestreo guarde relación con la temática que se desea medir. Esto sin duda incrementa los costos de operación, pero de manera marginal, ya que existe una infraestructura para el trabajo de terreno que se aprovecha con mayor intensidad. A su vez, mediante este tipo de estrategias es posible generar información para estudios longitudinales, que permitan analizar la dinámica de determinados fenómenos económicos y sociales.

Sin embargo, a pesar de las bondades señaladas, las encuestas que se realizan en el marco de un SIEH, difícilmente se pueden integrar desde un punto de vista temático, por lo

que el principal desafío -como se mencionó más arriba- consiste en diseñar métodos adecuados para complementar los datos obtenidos mediante las diferentes encuestas.

Estos y otros aspectos relativos al tema llevaron a los participantes a sugerir la necesidad de seguir profundizando en torno a los elementos conceptuales y metodológicos del SIEH y, sobre todo, en lo que dice relación con la puesta en práctica y administración de un sistema de este tipo.

En cuanto al **marco de muestreo**, el debate se concentró principalmente en lo relativo al significado de la incorporación y uso de los Sistemas de Información Geográfica (SIGs) como apoyo a los trabajos de diseño y ejecución de las encuestas de hogares. Al respecto, se insistió en que el alcance y los objetivos de un SIG suelen muchas veces rebasar no sólo las posibilidades de las oficinas de estadística sino también las necesidades de automatización de los procesos vinculados a la planeación y realización de las encuestas de hogares. Sobre esto último, se puso énfasis en reconocer que la posición geográfica de las viviendas seleccionadas es más importante que su precisión métrica.

Es claro que resulta deseable impulsar el uso de cartografía digitalizada y georeferenciada para apoyar la ubicación de las unidades de selección, y que debiera automatizarse la información que permite conformar el marco de muestreo, la selección de la muestra y el cálculo de los factores de expansión. Sin embargo, contar con estos insumos de trabajo en medios magnéticos no significa disponer de un SIG. Más bien, se trataría de un sistema de información estadística con cartografía georeferenciada, con capacidad de despliegue visual, ya que la característica fundamental de los SIG's está definida por su capacidad de sobreponer cartografía básica y temática del territorio nacional en un contexto espacial, mediante una base de datos relacional. De allí que en el taller se insistió en la conveniencia que los países hagan en cada caso particular un acucioso examen de sus necesidades (posibilidades) concretas y el análisis costo-beneficio correspondiente.

En lo que se refiere a los aspectos vinculados al **diseño de la muestra**, tanto los documentos de apoyo como las presentaciones realizadas remarcaron la necesidad de evaluar los métodos que se aplican para el cálculo del tamaño de la muestra en las encuestas de propósitos múltiples. Entre los elementos que deben ser considerados para determinar el tamaño óptimo de muestra en las encuestas de este tipo, se puso de relieve el carácter multidimensional del fenómeno de las condiciones de vida, la incidencia diferencial de los distintos temas, la distribución geográfica de la población objeto de estudio, los criterios para la estratificación del marco de muestreo, la varianza de los estimadores, el efecto de conglomeración y estratificación en el diseño de muestras complejas y el grado de precisión deseado.

Se recomendó prudencia al momento de definir los objetivos de las investigaciones, ya que con los tamaños de muestra que se utilizan actualmente en algunas encuestas es muy probable que el error estándar de los estimadores no se mantenga con un nivel de precisión razonable para todas las variables que se analizan. Al respecto, se recalcó que la manera de evaluar la confiabilidad estadística de los resultados es calculando el error de muestreo y el

efecto de diseño de los diferentes indicadores, lo cual desgraciadamente no constituye práctica habitual en las agencias nacionales productoras de la información, ni entre los usuarios que se abocan al análisis de los datos y a la formulación de políticas.

Asimismo, se sugirió que es muy importante no perder de vista las limitaciones implícitas en el diseño de una encuesta. Así, en el caso particular de las encuestas de empleo es altamente probable que los estimadores de promedios y de totales, vinculados por ejemplo a los salarios medios o al total de población desocupada, presenten un error de estimación relativamente alto, que en ocasiones llegue incluso a dificultar su uso analítico por el simple hecho que la precisión y confiabilidad de los resultados fue definida en base a una tasa y, por lo tanto, con menores exigencias en materia de tamaño de muestra.

Por otra parte, el interés por conocer la evolución de las condiciones de vida de grupos específicos de la población, como por ejemplo aquellos beneficiarios de programas sociales orientados a la erradicación de la pobreza, hace recomendable generar información que posibilite el análisis a lo largo del tiempo. En tal sentido, se puso de manifiesto la importancia de diseñar y llevar a la práctica **encuestas de panel para estudios longitudinales**, que representan un mecanismo apropiado para producir información continua acerca de una determinada cohorte poblacional, a fin de analizar los cambios experimentados en un conjunto de variables relevantes.

Al respecto, se describieron las bondades y limitaciones que presentan las opciones metodológicas conocidas para generar datos continuos, y se enfatizó la importancia de vincular las capacidades teóricas de cada alternativa con los objetivos mismos de la investigación. El propósito de introducir este tema en la agenda del taller se fundamentó en el interés expresado por los técnicos de los países en cuanto a conocer los aspectos conceptuales y estadísticos que sostienen los esquemas de rotación de las muestras actualmente en uso en las encuestas de hogares, sin pretender, por ende, obtener conclusiones definitivas acerca de los criterios que debieran adoptarse sobre la materia en los diseños de muestra de las encuestas continuas. Cada situación debe analizarse en el contexto particular de cada país, además de que es importante evaluar las experiencias pasadas a fin de conocer el nivel de correlación entre las mediciones de una misma variable, con el objeto de aprovechar esos datos para mejorar la eficiencia del diseño y tratar de estimar la proporción de traslape óptimo.

Al igual que en el caso de otros temas tratados en la reunión, se expresó que el **diseño de los cuestionarios** utilizados en las encuestas de hogares, y de manera especial en aquellas orientadas a medir las condiciones de vida, debe guardar relación directa con un conjunto de factores, entre los que cabe considerar los objetivos del estudio, la temática de la investigación, el marco conceptual, las subpoblaciones de interés, o el nivel cultural tanto del personal de campo como de los informantes.

El análisis comparativo de los cuestionarios utilizados por los países de América Latina en sus programas permanentes de encuestas de hogares reveló una gran heterogeneidad en cuanto a la forma de medir un mismo fenómeno, o la manera de captar las

diferentes variables relacionadas con éste. Así, al evaluar por ejemplo el número de preguntas que se realizan sobre cada tema y el nivel de detalle de las mismas, se confirma la variedad de situaciones que existe en lo relativo a la intensidad de los estudios, la rigurosidad con que se captan los distintos conceptos y, muy probablemente, la calidad de la información generada.

2. Trabajo de Campo y Control de Calidad

En el curso de los debates sobre este tema, se insistió de manera especial en la necesidad de reforzar los mecanismos de control durante la **ejecución del trabajo de campo**, así como durante las distintas fases de **tratamiento de la información**.

En ese sentido, se puso énfasis en la conveniencia de concebir la ejecución de una encuesta bajo la óptica de un enfoque sistémico, donde las diferentes fases del proyecto son interdependientes entre sí. Asimismo, se aludió al hecho que las herramientas de la administración por calidad total representan alternativas viables para cumplir en tiempo y forma los objetivos definidos en cada investigación y atender adecuadamente los requerimientos de los usuarios, garantizando calidad en los datos reportados y eficiencia en la operación.

Al atribuirle relevancia a todas las fases del proceso, también se reconoce que debieran redoblar los esfuerzos por **documentar** en detalle todas las acciones relacionadas con la ejecución de la encuesta, desde la planeación hasta la entrega de resultados. Así, debieran especificarse los criterios aplicados para determinar el tamaño y la distribución de la muestra; los métodos para el tratamiento de la no respuesta y la imputación de datos; si se efectuaron sustituciones y los motivos de ellas; los problemas del marco de muestreo y sus efectos en las probabilidades de selección; especificar el marco conceptual y metodológico en que se basa el diseño del cuestionario; las ventajas e inconvenientes de circunscribir la entrevista al informante directo; el procedimiento de estimación utilizado; los criterios seguidos para la crítica, codificación y análisis de consistencia de la información; y la identificación de las observaciones con las unidades de selección, para que los usuarios calificados puedan realizar el cálculo de los errores de muestreo en los dominios de estudio que les resulten de interés.

Especial énfasis se dio al tema de los **procesos de imputación de datos**, ya que es frecuente que durante la ejecución del trabajo, y pese a los múltiples esfuerzos que se realizan, ocurra la pérdida de información debido a diversas circunstancias. Se manifestó la importancia de diferenciar con claridad la no respuesta total y la parcial. En el primer caso se trata de un rechazo del informante a participar en la entrevista, o la imposibilidad de identificar la vivienda, o el hecho que al ubicar la vivienda seleccionada esta se encontraba deshabitada, había sido convertida en negocio o no fue posible ubicar a los ocupantes después de varias visitas. Por su parte, la no respuesta parcial ocurre con diferente intensidad para cada pregunta y unidad de observación, y en ocasiones es posible reconstruir la respuesta a partir de la información captada en las demás variables investigadas.

Cuando se trata de situaciones de no respuesta total y su incidencia no es muy alta, y además no se concentra en algún grupo social o área geográfica específica, la práctica común consiste en ajustar las probabilidades de selección de las unidades de vivienda que sí fueron visitadas, a efectos que el proceso de expansión de los resultados genere un total de viviendas similar al existente en el país. Se aclaró que los porcentajes de no respuesta aceptados varían entre países y según el tipo de encuesta. Asimismo, se sugirió que debieran considerarse los resultados obtenidos en experiencias anteriores como apoyo al cálculo del tamaño de muestra en las nuevas investigaciones, con el objeto de eliminar la mala práctica de efectuar sustituciones de viviendas, ya que éstas introducen sesgos que no son fáciles de identificar y mucho menos de cuantificar.

Por su parte, la no respuesta parcial debiera ser analizada cuidadosamente, y una primera recomendación que surgió es la de establecer la necesidad de conservar el archivo original de la encuesta tal y como provino del campo. Mediante esta práctica será siempre posible reconstruir la información original, e identificar los posibles sesgos que introduce en los estimadores el hecho de imputar datos faltantes.

Durante la discusión se expresaron posiciones divergentes en cuanto a la pertinencia de realizar imputaciones, o bien proporcionarle al usuario los datos originales para que sea éste quien, de acuerdo a sus propios criterios y necesidades, proceda a completar los datos faltantes. Respecto a esta última posibilidad, se argumentó en cuanto a los peligros que puede encerrar el dejar esta tarea en manos de los usuarios, a consecuencia de lo cual se sugirió la conveniencia que sea la oficina de estadística quien, con más conocimiento de las causas y efectos de la no respuesta, defina procedimientos claros y documentados de imputación que permitan un mayor control sobre la información publicada.

Al mismo tiempo, se expusieron algunos de los métodos estadísticos más comunes utilizados para completar datos faltantes, enfatizando que previo a su aplicación se debiera analizar la variable a imputar y el porcentaje y distribución de los cuestionarios sin respuesta, así como tener claros los objetivos de la imputación y sus efectos en el error estándar del estimador. Todas estas consideraciones ayudan a elegir el procedimiento más apropiado y contribuyen a no introducir sesgos innecesarios en los datos.

3. Difusión y Explotación de Información

El verdadero valor de las encuestas de hogares se mide en función de los usos analíticos que se hacen de la información y las acciones de política que se diseñan a partir de ella para mejorar las condiciones de vida de la población. En ese sentido, los participantes en el taller enfatizaron la necesidad de establecer mecanismos claros y eficientes de difusión y explotación de los datos y de intercambio técnico con los usuarios, a fin de generar procesos que contribuyan al mejoramiento de la calidad y uso de la información generada por las encuestas.

Pese a que los actuales esquemas de **difusión de información** utilizados por la mayoría de las oficinas nacionales responsables de la realización de estas investigaciones por muestreo se han adaptado al desarrollo tecnológico, y ofrecen variadas alternativas para el acceso y divulgación de los resultados, en el taller se discutió la necesidad de mejorar la presentación y contenido de los informes técnicos que acompañan la documentación de las encuestas. Ello debido a que se estima que actualmente no se proporciona información suficiente, que permita por ejemplo prevenir a los usuarios acerca de las limitaciones de los datos, las restricciones para su desagregación geográfica, su falta de capacidad para realizar inferencias sobre determinados dominios de estudio, los errores de muestreo asociados a los estimadores, la validez estadística de los resultados, la distribución de la no respuesta o los métodos de imputación utilizados.

Al respecto, se advirtió sobre las consecuencias que puede acarrear el uso inadecuado de la información, o el no considerar en el análisis las características del diseño de la muestra en una encuesta compleja. Dado que las prácticas corrientes en el análisis de la información proveniente de encuestas suele no considerar estos aspectos, es probable que algunas de las recomendaciones de política que se han formulado carezcan de la confiabilidad estadística necesaria. A su vez, es frecuente que se desee evaluar parámetros que no están acordes con la variable utilizada para determinar el tamaño de muestra, o bien construir indicadores compuestos que demandan otro tipo de diseño o un mayor número de observaciones. También está presente en las tareas cotidianas de los investigadores y diseñadores de política la tentación de analizar dominios de estudio que no fueron considerados en los objetivos de la encuesta; sin embargo, no es práctica común evaluar la confiabilidad de las estimaciones, ni analizar la distribución muestral de las variables en los subgrupos poblacionales de interés. Así, el no asumir estas precauciones puede llevar a que se invaliden los resultados y conclusiones de algunos estudios que utilizan los datos provenientes de las encuestas de hogares.

4. Lineamientos de Acción

Como en toda reunión donde se analizan temas técnicos de gran trascendencia e interés para los participantes, el tiempo fue la variable que limitó la posibilidad de profundizar en diversos aspectos conceptuales y metodológicos identificados como muy importantes a lo largo de las presentaciones y debates. Sin embargo, a pesar de este hecho, hubo consenso en que el evento llenó las expectativas de los representantes de los países y de los organizadores, sin dejar de reconocer que este esfuerzo debe continuar y algunos temas deberán ser retomados en el futuro y analizados con mayor detalle en el ámbito de reuniones especializadas.

A continuación se identifican algunos temas y líneas de acción que se desprenden de este taller y que debieran ser tenidas en cuenta para el diseño de la agenda de trabajo de las próximas reuniones regionales organizadas por la CEPAL en el marco del MECOVI, o bien en las demás actividades de este tipo previstas en el Programa a nivel regional o nacional.

i) Dedicar atención especial al tema de la **expansión** de los resultados muestrales, ya que la falta de recursos para la actualización del marco de muestreo obliga a las oficinas de estadística a recurrir a las proyecciones de población como parámetro de referencia para ajustar las probabilidades de selección y las estimaciones derivadas de la muestra. En una situación óptima, la encuesta debiera generar una estimación alternativa del total de población del país al momento del levantamiento de la encuesta, ante un marco de muestreo actualizado, y permitir identificar aspectos vinculados con el crecimiento natural de la población, la dinámica demográfica y, en particular, los cambios ocurridos en el terreno en cuanto a la edificación de nuevas construcciones para fines de vivienda y/o la destrucción de unidades contenidas en el marco.

ii) Por otra parte, y dada la sugerencia de administrar las encuestas con herramientas de calidad total, se debieran proponer metodologías para generar evidencia empírica, a nivel nacional y regional, sobre los **estándares** a utilizar para los efectos de evaluar la calidad en la ejecución de las diversas etapas consideradas en el desarrollo de las encuestas por muestreo probabilístico. Particular importancia reviste el perfeccionamiento de los métodos utilizados para el cálculo de los errores muestrales en encuestas complejas, como así también para el control de los errores no muestrales.

iii) Prestar atención al problema de aclarar las diferencias entre una encuesta única, que busca integrar en un solo relevamiento todas las dimensiones del bienestar de las personas, y un **sistema integrado de encuestas** concebidas con carácter complementario. Asimismo, se debiera profundizar en el análisis de las bondades y limitaciones técnicas de cada una de las opciones señaladas, a fin de poder evaluar su capacidad para responder a las actuales demandas de información de parte de los gobiernos y la sociedad en general que apuntan al diseño de políticas y programas orientados al mejoramiento de la situación social.

iv) Destinar algunas sesiones de trabajo al examen, con el debido rigor técnico, de los **métodos estadísticos** en los que se fundamenta la teoría del muestreo y revisar los desarrollos más actuales tanto de la estadística matemática como del análisis de información proveniente de encuestas complejas.

v) Generar información para contextos geográficos específicos obliga a analizar y evaluar métodos estadísticos de estimación sintética, como una alternativa para obtener **resultados a nivel municipal y local**. Además, esta situación lleva a la necesidad de definir estrategias para reforzar los mecanismos de recolección de información sectorial al interior de los países, a fin de reforzar los actuales esquemas de registros administrativos e integrarlos en sistemas amplios de información territorialmente desagregados. Diseñar encuestas con alta desagregación geográfica (municipal o local) puede elevar considerablemente los costos de la investigación, dificultando su realización o posterior replicación.

vi) Para avanzar en la reflexión acerca de las **encuestas de panel y los estudios longitudinales**, se requerirá seguramente generar nuevos espacios de discusión, donde se

revisen en detalle las experiencias regionales sobre la materia, así como las investigaciones teóricas y prácticas realizadas en los países de América del Norte y Europa.

vii) Establecer una **plataforma de intercambio** de experiencias y reforzar los mecanismos de cooperación horizontal entre los países, de manera de maximizar la eficiencia en el uso de los recursos regionales disponibles en el área de las encuestas de hogares.



PLANIFICACIÓN Y DISEÑO DE ENCUESTAS DE HOGARES

**HACIA UN SISTEMA INTEGRADO DE ENCUESTAS DE
HOGARES EN LOS PAÍSES DE AMÉRICA LATINA**

JUAN CARLOS FERES

CEPAL

ÍNDICE

Página

I. BASES DE UN SISTEMA INTEGRADO DE ENCUESTAS DE HOGARES (SIEH)	31
1. Antecedentes.....	31
2. Definición	32
3. Objetivos	32
4. Elementos del diseño conceptual.....	32
5. Plan operativo.....	35
6. Diferencia entre un SIEH y una encuesta multitemática “única”	36
II. OBSTÁCULOS Y DESAFIOS	38
1. Mejorar los marcos de muestreo.....	38
2. Vincular los diseños de las encuestas	39
3. Optimizar el uso de los recursos.....	41
III. LINEAMIENTOS DE ACCIÓN	41
ANEXO: NÚCLEO BÁSICO DE ENCUESTAS A INCORPORAR EN UN SIEH.....	44

HACIA UN SISTEMA INTEGRADO DE ENCUESTAS DE HOGARES EN LOS PAÍSES DE AMÉRICA LATINA

I. BASES DE UN SISTEMA INTEGRADO DE ENCUESTAS DE HOGARES (SIEH)

1. Antecedentes

Los altos índices de pobreza y desigualdad que caracterizan la situación socioeconómica de la mayoría de los países de América Latina y el Caribe, y la diversidad de las carencias a que se ven enfrentados diferentes sectores de la población, plantean la necesidad de encarar el diseño y puesta en práctica de políticas y programas cada vez más eficaces para superar los rezagos existentes y mejorar las condiciones de bienestar y equidad social.

Lo anterior se traduce, naturalmente, en una *gran demanda de información* oportuna, adecuada y de buena calidad. Sin embargo, y pese a los apreciables avances logrados en los últimos años, subsisten aún en muchos países *déficit* significativos en este campo, especialmente en cuanto a la cobertura geográfica y temática de las encuestas, su periodicidad, la consistencia de los datos generados por diferentes fuentes y la aptitud de las mismas para responder a los requerimientos impuestos por las políticas y programas sociales. Esto ha llevado en algunos casos a la *proliferación de encuestas*, o bien a *ampliar el contenido de los cuestionarios* más allá de lo que razonablemente permite la definición de cada operativo y las características y limitaciones de los instrumentos utilizados.

Por tal motivo el BID, el Banco Mundial y la CEPAL han formulado una iniciativa conjunta, expresada en el Programa para el Mejoramiento de las Encuestas y la Medición de las Condiciones de Vida en América Latina y el Caribe (MECOVI), con el fin de apoyar los esfuerzos que sobre el particular vienen desplegando los gobiernos y los organismos estadísticos de los países de la región. El propósito central del citado Programa es contribuir a establecer y/o fortalecer los *Sistemas de Encuestas de Hogares*, constituidos por un conjunto de encuestas debidamente integradas, y que representan una de las principales fuentes de información para el análisis, evaluación y seguimiento de la realidad social y las condiciones de vida de la población. Con ello se busca, al mismo tiempo, dotar al sector público y privado de los elementos necesarios que garanticen un adecuado monitoreo de las políticas en marcha y contribuyan al diseño de programas eficientes orientados a la superación de la pobreza y el mejoramiento de las condiciones de equidad.

En este contexto, la presente nota tiene por objeto formular de manera sintética y con *carácter preliminar* algunos lineamientos generales acerca del SIEH, a efectos de incentivar la discusión orientada a precisar sus bases conceptuales, describir sus bondades y limitaciones, e identificar los obstáculos que se enfrentan en los diferentes países para el diseño y puesta en práctica de un sistema de este tipo.

2. Definición

El Sistema Integrado de Encuestas de Hogares es una estrategia de producción de información sobre las características demográficas y socioeconómicas de los hogares y las personas, basada en el *diseño y ejecución coordinada* de un conjunto de encuestas que, siendo diferentes entre sí en cuanto a sus objetivos y alcances, comparten un cierto marco conceptual y metodológico, complementan sus temas de investigación, sincronizan su realización en el tiempo y optimizan el uso de los recursos humanos y financieros.

Dada las aptitudes y restricciones propias de cada tipo de encuesta, la variedad de las temáticas que abordan y las múltiples demandas que se proponen atender, el SIEH constituye un esquema de trabajo que persigue recopilar información mediante diferentes instrumentos y en distintos períodos del año, conforme a una estructura programática que refleje las prioridades nacionales, los temas de interés, la oportunidad en que se requieren los datos, los recursos disponibles, así como el uso de los métodos y procedimientos estadísticos más apropiados para generar la información, en el contexto de un marco conceptual que integre y complemente las distintas investigaciones.

3. Objetivos

La creación, perfeccionamiento y consolidación de un SIEH obedece, principalmente, a tres propósitos básicos:

a) Mejorar la calidad, cobertura, oportunidad, consistencia y complementariedad de la información que se obtiene mediante un conjunto articulado de encuestas de hogares.

b) Atender adecuadamente a los diversos requerimientos de información del área económica y social, tanto en lo relativo a la naturaleza y periodicidad de la misma, como a su pertinencia para la formulación, seguimiento y evaluación de las políticas y programas sociales.

c) Aumentar la eficiencia en el uso de los recursos.

4. Elementos del diseño conceptual

No cabe postular un *modelo único* para el diseño de un sistema integrado, ya que la composición, alcance y complejidad de un SIEH está estrechamente vinculada a la realidad específica de cada país, en términos de sus necesidades de información, la infraestructura existente en materia de encuestas, la capacidad institucional y los recursos disponibles. Sin embargo, es posible identificar ciertos *aspectos básicos comunes* vinculados, por ejemplo, a la complementariedad temática de las encuestas, la construcción y uso de marcos compartidos, el diseño de muestras maestras o la utilización de los insumos de trabajo.

En líneas generales, las diversas encuestas de hogares incorporadas en un SIEH se conciben como un *sistema integrado* porque:

- *Complementan información* de la población y de los hogares referida a distintas áreas temáticas y *programan su investigación* en el tiempo de acuerdo a la naturaleza de cada tema.

- Cada una de ellas responde a un *diseño particular* y utiliza instrumentos de captura de los datos definidos en función de los temas específicos que investiga, pero preserva la *homogeneidad conceptual* en los temas comunes.

- Las encuestas cuentan con un *marco muestral común* y las respectivas muestras se pueden extraer, de estimarse conveniente, a partir de un *marco maestro* que evita las duplicidades y reduce al máximo las visitas a los mismos hogares, incrementando las tasas de respuesta.

- El diseño de algunas de las encuestas permite el *análisis "longitudinal"*, en virtud del esquema de rotación de la muestra o porque ésta contempla expresamente una componente de "panel", con traslape parcial.

- *Comparten recursos* y permiten dar respuesta a los requerimientos de los usuarios de la información de manera eficiente y oportuna.

A su vez, la integración también hace referencia a la consistencia conceptual y metodológica de las encuestas de hogares con otras fuentes de información del sistema estadístico nacional, particularmente los censos de población y vivienda, los registros administrativos y la contabilidad macroeconómica.

Cada encuesta que forma parte del SIEH posee características particulares, en atención a sus objetivos específicos, diseño muestral, contenido temático, periodicidad, cobertura geográfica y costo. En el Anexo se presentan los lineamientos de un modelo básico que considera tres clases de relevamientos, a saber: *encuesta de propósitos múltiples, con módulo central de empleo; encuesta de condiciones de vida; y encuesta de ingresos y gastos familiares.*

No existe una regla universal para definir las encuestas que debieran ser integradas en un SIEH. Sin embargo, es frecuente encontrar que los siguientes temas suelen habitualmente ser objeto de investigación por medio de encuestas de hogares:

- Características demográficas de la población
- Migración interna e internacional
- Mortalidad y fecundidad
- Características físicas de la vivienda y disponibilidad de servicios básicos
- Tenencia de bienes electrodomésticos y de vehículos de transporte
- Niveles de escolaridad de los miembros del hogar

- Empleo, desempleo e inserción laboral
- Funcionamiento y estructura del mercado de trabajo
- Nivel, origen y composición del ingreso
- Estructura del gasto y del consumo de los hogares
- Salud y nutrición
- Acceso a los programas públicos y privados de asistencia social
- Existencia de personas discapacitadas en el hogar
- Violencia intrafamiliar y maltrato infantil
- Consumo de drogas y sustancias prohibidas

Normalmente los estudios se realizan por medio de *investigaciones específicas* para cada tema, o bien -teniendo como base un tema central- se incorporan al cuestionario de la encuesta las preguntas, o en ocasiones los *módulos*, relativos a los demás fenómenos que se desea investigar. En todos los casos el diseño estadístico de la encuesta debiera prever esta situación, a fin de que no se produzcan sesgos en los datos y se afecte severamente la precisión de los resultados obtenidos, ya que el error de muestreo está altamente correlacionado con el tamaño de muestra y la varianza del estimador.

A su vez, debe considerarse que las *unidades de análisis* para los distintos temas son diferentes. En efecto, a pesar que, por ejemplo, tanto en una encuesta de empleo como en una de presupuestos familiares las unidades de selección de última etapa son las viviendas, en las investigaciones sobre fuerza de trabajo los esfuerzos se centran en conocer las características de ocupación de la PEA, mientras que en un estudio sobre ingresos y gastos las unidades de análisis son las familias. Asimismo, es necesario considerar que el *tiempo de entrevista* y la duración del trabajo de campo se incrementan de manera considerable en las encuestas que tienen por objeto analizar las fuentes de ingreso y los hábitos de consumo de la población. Esta situación se refleja en el diseño de un *cuestionario* amplio y detallado, que permita registrar las distintas modalidades de ingreso y el destino del gasto de los hogares. No abundan los estudios analíticos que demuestren que a partir de un número reducido de preguntas sea posible estimar con precisión aceptable la estructura del gasto de los hogares o la distribución personal del ingreso.

Por otra parte, y como se señala en el siguiente documento en que se presentan algunas consideraciones sobre el diseño estadístico de las encuestas de hogares, es importante identificar y seleccionar con claridad las *variables de estratificación* que resulten adecuadas para la conformación del marco de muestreo. Sobre el particular, cabe señalar las diferencias que existen en la variable que se utiliza para la estratificación del marco de muestreo cuando se trata, por ejemplo, de una encuesta de empleo o de una orientada a investigar el presupuesto familiar. Así, mientras en el primer caso la variable adecuada puede ser la PEA, la categoría ocupacional o la rama de actividad económica de los ocupados, en el caso de las encuestas de ingresos y gastos es claro que la variable de estratificación debe estar altamente correlacionada con el nivel socioeconómico de las unidades de observación y análisis. Asimismo, en cada investigación se debe identificar la *variable de diseño* que resulta más apropiada, y calcular el tamaño de muestra óptimo

considerando el efecto de conglomeración (efd) de la variable de interés y la tasa de respuesta esperada para cada variable. Por último, hay que determinar la correlación que existe entre la variable de diseño y las características de la población que se desean estudiar con fines de inferencia estadística. Por ejemplo, las encuestas de empleo se diseñan para estimar con una determinada precisión la tasa de desocupación abierta, pero no necesariamente para realizar estimaciones de pobreza a partir del nivel de ingreso de los hogares ni tampoco para estudiar a los subempleados invisibles o al servicio doméstico.

5. Plan operativo

El diseño y puesta en práctica del SIEH implica definir, entre otras cosas, los temas a investigar, el marco conceptual general y de cada investigación, el tipo de encuestas que se levantarán, la periodicidad y cobertura geográfica y socioeconómica de cada una de ellas, y los recursos destinados a la realización de las mismas. A su vez, se deben abordar los aspectos metodológicos vinculados a la construcción del marco maestro, consultando criterios eficientes para su administración y procedimientos continuos de actualización de la cartografía. Asimismo, deben considerarse los aspectos estadísticos para el diseño y selección de las muestras, los métodos de estimación y cálculo de los errores de muestreo y los procedimientos de campo para la ejecución de las diversas encuestas.

En la etapa de diseño del SIEH es necesario establecer con claridad aspectos tales como los objetivos de cada investigación, las unidades de análisis, los dominios de estudio, las subpoblaciones de interés, los productos, las fechas para las que se espera disponer de la información y el plan de explotación y análisis de los resultados. A su vez, se debe garantizar un planteamiento conceptual y metodológico que integre a las distintas investigaciones y construir un marco de muestreo maestro que permita la selección de las diferentes muestras, ya sea para completar las rondas de una misma encuesta o para permitir la selección de muestras independientes. Por ende, se requiere además tener en cuenta la relación que existe entre las variables de estratificación -utilizadas para mejorar la eficiencia del diseño-, los parámetros de la población que se desea estimar y la(s) variable(s) considerada(s) para la determinación del tamaño de muestra.

Para la puesta en práctica del SIEH, en tanto, es menester hacer hincapié en la necesidad que el diseño del programa considere la *coordinación operativa* y la *integración sustantiva* de las diferentes encuestas que se piensa realizar. La coordinación operativa se relaciona con la utilización de los insumos comunes para la realización del trabajo de campo, en el sentido de garantizar el uso eficiente de los recursos disponibles. A su vez, la integración sustantiva se refiere a la homogeneidad conceptual y metodológica que debiera existir entre las distintas encuestas que forman parte del sistema, la que debiera basarse en las recomendaciones internacionales vigentes, de manera que la información generada revele una adecuada consistencia y permita lograr la complementariedad temática del programa. De modo que el diseño del SIEH debiera orientarse a maximizar las ventajas de coordinación e integración en el desarrollo de las diferentes etapas.

6. Diferencia entre un SIEH y una encuesta multitemática “única”

La formación de un SIEH implica armonizar y complementar distintas encuestas, cada una de las cuales presenta, en función de sus objetivos y diseño, determinadas aptitudes y limitaciones para captar información socioeconómica de los hogares. En ese sentido, el sistema integrado difiere del concepto de encuesta única, politemática, que persigue investigar en un sólo instrumento y bajo el mismo diseño estadístico una variedad extremadamente amplia de fenómenos, pretendiendo satisfacer simultáneamente diversos objetivos analíticos, lo que por lo general dificulta la determinación de un tamaño óptimo de muestra e incide en la confiabilidad de los resultados.

Como se sabe, en el hecho todas las encuestas de hogares que se llevan a cabo se pueden considerar de propósitos múltiples. Sin embargo, así como rara vez se diseña una encuesta para obtener estimaciones acerca de una sola variable, tampoco es usual -ni lógico-pretender diseñarla de manera que cubra, con precisión adecuada, todos los requerimientos de información existentes. Este punto de vista goza de general aceptación, al extremo que en la práctica las oficinas de estadística realizan normalmente distintos tipos de encuestas, con carácter complementario, para atender a las variadas necesidades de los usuarios. Tales encuestas tienen objetivos diferentes y, por ende, los diseños de muestra se elaboran considerando distintas unidades de selección, observación y análisis, con dominios de estudio específicos, asumiendo los aspectos estadísticos vinculados a la determinación de los tamaños de muestra (variabilidad del estimador, efecto de diseño y tasa de respuesta estimada), así como aquellos relativos a la cobertura geográfica y temporal, y la calendarización de los trabajos de campo y entrega de resultados.

Así, mientras en el esquema conceptual de un SIEH subyace la premisa que los requerimientos informativos de un cierto ámbito socioeconómico se cubren mediante un conjunto de investigaciones diferentes, pero vinculadas entre sí, en las encuestas multitemáticas del tipo de las que se comentan -y que por lo general se llevan a cabo de manera esporádica- se intenta recabar en un mismo levantamiento información sobre una amplia gama de temas, normalmente muy distintos entre ellos, que difícilmente pueden ser atendidos con la misma precisión estadística.

Recoger información sobre múltiples temas para la misma unidad de observación es una ventaja importante, ya que permite identificar las relaciones existentes entre diversas variables que inciden en el comportamiento de un fenómeno. Asimismo, el hecho que el levantamiento se concentre en un momento del tiempo puede llevar a reducir los costos de operación, por unidad de observación, no obstante que obliga a disponer de la totalidad de los recursos para cuando se fija el levantamiento. Sin embargo, con respecto al costo se debe tener presente, además, que si las encuestas de este tipo se propusieran investigar todos los temas con el rigor requerido, es muy probable que tanto la amplitud del cuestionario como el tiempo de entrevista y el tamaño de muestra se incrementarían en forma considerable, al punto de elevar su costo más allá que el de cualquier encuesta continua que recabe datos sobre temas específicos.

De allí que por el volumen de información que normalmente recopilan estas “encuestas integradas” y los costos asociados a su diseño y operación, es muy difícil darles un carácter permanente o aplicarlas en forma continua, dado que la mayoría de los países no disponen de recursos suficientes como para financiarlas con el presupuesto regular con que cuentan las oficinas de estadística.

A su vez, desde un punto de vista técnico la generación de información por esta vía también puede presentar algunos problemas, que no son menores. Debido a su carácter multitemático, estas encuestas enfrentan habitualmente un conflicto entre la precisión de los resultados y el tamaño de muestra utilizado. Como ha sido documentado por diferentes investigadores, en una encuesta de propósitos múltiples muy amplios es difícil lograr una adecuada compatibilidad entre los objetivos de la investigación, el tamaño de muestra óptimo y los recursos disponibles. Lo anterior significa que para lograr similar precisión para todas las variables de interés, se tendría que incrementar de manera considerable el tamaño de muestra y, por ende, el presupuesto de la encuesta. En ese sentido, la práctica común indica que necesariamente se debe sacrificar la precisión estadística de muchas de las estimaciones y ponderar los objetivos de la investigación, poniendo énfasis en los verdaderos alcances del estudio, las restricciones del análisis y la confiabilidad de la información para dominios específicos.

Para ilustrar esta situación, baste considerar el caso en que a partir de una encuesta de empleo se pretenda estimar el ingreso y el consumo promedio de los hogares, con la misma precisión estadística con que se obtiene la tasa de desocupación abierta (TDA). Una primera consideración que se debiera realizar durante la etapa de análisis de los datos, es preguntarse cuál fue la variable utilizada para la determinación del tamaño de muestra: la TDA o el coeficiente de variación del ingreso (gasto) promedio del hogar; asimismo, cuál fue el efecto de diseño y la tasa de respuesta que se utilizó para ajustar el tamaño de muestra, o sobre el error de muestreo de los estimadores.

Al respecto, es conocido el hecho que el número de selecciones que se requieren para estimar con una precisión determinada una proporción o tasa es menor que para los promedios o totales poblacionales. De modo que si el tamaño de muestra de una encuesta se calcula a partir de la proporción de población desocupada, con un cierto nivel de confianza, es muy probable que las estimaciones para los promedios de ingreso y gasto tengan una mayor variabilidad y su error relativo sea superior al requerido por el investigador para realizar inferencias y adoptar decisiones de política. Problema que se agudiza aún más cuando a partir de la muestra se desea realizar inferencias para dominios de estudio con características específicas, y no definidos entre los objetivos de la encuesta, por ejemplo los indigentes que están desempleados y que provienen del sector primario de la economía. En ese caso es muy probable que este segmento de población no tenga una correlación significativa con la variable de diseño y, por lo tanto, la baja frecuencia de observaciones con esta característica se refleje en el error de muestreo de los estimadores. Asimismo, es común que los analistas procedan a ajustar modelos econométricos para subpoblaciones con pocas observaciones, confiando en las propiedades de normalidad de los estimadores para muestras pequeñas. Sin embargo, en esas circunstancias es necesario evaluar la consistencia

del método de estimación utilizado, a efectos de garantizar que se cumplan los supuestos del modelo y se logre la precisión estadística deseada.

Por su parte, bajo el esquema del SIEH la mediciones que se efectúan sobre temas específicos tienen, por lo general, mayor precisión estadística y menor costo de operación, con lo que generan información más confiable y pueden efectuarse con mayor regularidad. Al mismo tiempo, las encuestas permanentes tienen la capacidad de captar información sobre temas adicionales al objetivo central para el que fueron diseñadas, en la medida que se hagan las modificaciones pertinentes en el tamaño de muestra y la conformación del marco de muestreo guarde relación con la temática que se desea medir. Esto sin duda incrementa los costos de operación, pero de manera marginal, ya que existe una infraestructura para el trabajo de terreno que se aprovecha con mayor intensidad. A su vez, mediante este tipo de estrategias es posible generar información para estudios longitudinales, que permitan analizar la dinámica de determinados fenómenos económicos y sociales.

Sin embargo, a pesar de las bondades señaladas, las encuestas que se realizan en el marco de un SIEH, difícilmente se pueden integrar desde un punto de vista temático, por lo que el principal desafío -como se menciona más adelante- consiste en diseñar métodos adecuados para complementar los datos obtenidos mediante las diferentes encuestas.

II. OBSTÁCULOS Y DESAFIOS

1. Mejorar los marcos de muestreo

Diseñar y construir *marcos completos o exhaustivos* es una tarea de la mayor importancia, a fin de mejorar la cobertura de las encuestas. Esta completitud no obsta para que en ocasiones puedan existir regiones del país o segmentos de la población que sean excluidos deliberadamente del marco de muestreo, por razones ligadas a los objetivos de las distintas investigaciones, falta de recursos, limitaciones de acceso, baja incidencia del fenómeno a estudiar o escasa población registrada. En el siguiente documento se examinan en detalle algunos problemas y desafíos que supone avanzar en esta etapa crucial de la instalación y desarrollo de un SIEH.

Asimismo, como se ha insistido, *mejorar y actualizar el marco* que se utiliza para la selección de la muestra es un requisito indispensable para incrementar la calidad de la información que se genera por medio de las encuestas de hogares. Sin embargo, se constata que generalmente los países no dedican recursos suficientes a esta actividad, lo que puede llevar a que en los períodos intercensales se trabaje sin considerar que las ciudades se transforman, las viviendas aumentan, las edificaciones individuales derivan en estructuras de departamentos, la población crece y se mueve libremente en el territorio o que las manzanas aumentan de tamaño y desbordan los límites establecidos para fines de muestreo. Cuando esto sucede, se alteran los pesos relativos de las distintas unidades de selección, modificando las probabilidades de selección de las viviendas sin que la estructura operativa de la encuesta

esté en capacidad de incorporar estos cambios en la base de datos y en la cartografía en uso, lo que limita las posibilidades de que la encuesta se convierta en una buena alternativa para captar la dinámica demográfica en relación al crecimiento, localización y composición de la población.

Habitualmente la expansión de los resultados de una encuesta se hace en base a una proyección de población que no siempre es apta para incorporar todos los señalamientos anteriores, y de manera especial la distribución de la población en el territorio, lo cual puede distorsionar las tasas de crecimiento poblacional de las distintas agrupaciones del país. Además, en la medida que una encuesta se realiza en una fecha muy distanciada del último evento censal, las diferencias entre las proyecciones de población y los recuentos de personas obtenidos por medio del censo se pueden agudizar, incidiendo en los resultados de la encuesta.

Por otra parte, una vez que se dispone de los resultados definitivos del censo, las estimaciones pasadas de las encuestas normalmente no se corrigen, lo que afecta la *evolución de las series históricas* e introduce sesgos de estimación en la medida que los datos se utilicen para efectuar proyecciones sobre el comportamiento de las variables.

De allí que sea muy importante generar consenso respecto de la necesidad de *actualizar el marco de muestreo por lo menos una vez al año*, aprovechando para ello la infraestructura existente para la realización del trabajo de campo. Por ejemplo, se puede adoptar la estrategia de asignar un día más en el trabajo de terreno de los enumeradores, a fin de que recorran su zona y registren las diferencias que existan entre el plano de línea utilizado para identificar su área de trabajo y la situación observada en el momento de realizar las visitas a los hogares. Del mismo modo, se puede actualizar el listado de viviendas, como una actividad rutinaria previa a la selección de la muestra.

Otra actividad a la cual se deben dedicar recursos suficientes, es la *sistematización del marco de muestreo y de la información que sirve de base para su conformación*. En ese sentido, la *digitalización de la cartografía* se presenta como una tarea necesaria que permite disponer de planos de línea de alta calidad, que apoyan de manera eficaz las labores de identificación de los segmentos de viviendas y de las unidades de observación y se actualizan rápidamente, representando un insumo fundamental para el desarrollo de cualquier investigación por muestreo. A su vez, la sistematización de los procedimientos de estratificación, selección de la muestra, determinación de las probabilidades de selección y cálculo y corrección de los factores de expansión, permiten llevar un mejor control de las actividades y reducir los errores humanos en los procesos de trabajo, evitando manipulaciones innecesarias.

2. Vincular los diseños de las encuestas

En el contexto de un SIEH es evidente que los diseños de muestra se diferencian según las características propias de cada encuesta (objetivos, contenido temático,

periodicidad, unidad de análisis, etc.). De allí que sean explícitas también las diferencias relativas a las variables utilizadas para la estratificación del marco de muestreo y la determinación del tamaño de muestra, las unidades de selección, observación y análisis y los dominios de estudio sobre los que generan información, así como la confianza de los indicadores.

La utilización de marcos maestros de muestreo contribuye a reducir el costo de diseño y operación de las encuestas, en la medida que el presupuesto requerido para su conformación se distribuye entre las distintas investigaciones que se desean realizar. A su vez, bajo una estrategia integrada, los diseños de muestra permiten definir el tamaño apropiado de las unidades de primera y segunda etapa de selección, así como el número de las mismas que debe seleccionarse, a fin de satisfacer las necesidades planteadas en un programa de mediano plazo definido para recabar información sobre diferentes temas.

Mediante esta práctica, también es posible diseñar procedimientos de administración eficientes, de tal suerte que se eviten al máximo las duplicidades de las unidades de observación y los hogares no sean visitados en forma repetida, buscando minimizar el rechazo por parte de los informantes.

Sin embargo, uno de los grandes retos estadísticos que se deben afrontar es el desarrollo de procedimientos teóricos y metodologías de carácter práctico que permitan *interrelacionar los resultados* obtenidos por medio de las diversas encuestas que conforman un sistema integrado. Ello es así dado que, a pesar que en un SIEH se logre la complementariedad temática y se sincronice la ejecución de las diferentes encuestas, no es clara la manera de integrar los resultados para los efectos de la caracterización de fenómenos multivariados, como es el caso del nivel de vida de los hogares.

En efecto, sabemos de la conveniencia de interrelacionar las bases de datos de diferentes encuestas de hogares que estudian distintos temas. No obstante, en el contexto de un SIEH normalmente las encuestas que se realizan son diseñadas con base en muestras independientes, donde los hogares visitados difieren entre las distintas investigaciones. Más aún, incluso en el caso en que se tratara de los mismos hogares pero bajo esquemas de selección distintos, subsisten las diferencias anotadas con respecto a la estratificación del marco de muestreo, las variables de diseño, los pesos relativos de las unidades de selección, los factores de expansión y la no respuesta, que constituyen -entre otras limitantes- algunos de los problemas que impiden sobreponer dos bases de datos para complementar el estudio de un determinado fenómeno.

De modo que, aun cuando el SIEH calendarize en un primer momento una encuesta de fuerza de trabajo y luego una encuesta de ingresos y gastos, no es posible utilizar, al menos en forma directa, por ejemplo la información de empleo de la primera para complementar la segunda, y evitar así que nuevamente se tenga que captar información sobre las características ocupacionales de la población. Pese a que en los dos casos se trata de una encuesta de hogares, cuyas muestras pueden haber sido seleccionadas a partir de un

mismo marco de muestreo, las distintas probabilidades de selección de las observaciones limitan la operación deseada.

3. Optimizar el uso de los recursos

Mejorar la eficiencia en el uso de los recursos técnicos, humanos y materiales, generar sinergías entre los distintos operativos que se llevan a cabo, aprovechar intensamente la infraestructura disponible e incentivar la producción de economías de escala, son algunos de los desafíos a los que se ven permanentemente enfrentadas las oficinas nacionales de estadística, y que cobran particular importancia en el contexto de instalación y desarrollo de un SIEH.

De allí que, ante la crónica insuficiencia de recursos que suele afectar a estas instituciones, se debe reforzar el empeño por desarrollar procedimientos que contribuyan a optimizar las diversas actividades, no solo en cuanto a su calidad técnica sino también en lo que se refiere al uso de los recursos disponibles. Baste señalar, tan solo a modo de ejemplo, los avances posibles de lograr en este sentido en las tareas propias de la fase de diseño y selección de la muestra; o bien, en las de actualización del marco de muestreo, que sigue siendo una de las principales limitaciones para reducir el margen de error de las estimaciones proporcionadas por las encuestas. Sobre esto último, y dado que en muchos casos es difícil que en el corto plazo se disponga de los recursos necesarios para mantener al día la información cartográfica, es conveniente aprovechar al máximo las visitas al terreno para obtener información sobre cambios en la distribución y destino de las viviendas, áreas de nuevas construcciones, desaparición de unidades habitacionales, etc..

III. LINEAMIENTOS DE ACCIÓN

En los últimos tiempos, y como fruto de una creciente demanda de información del área social, se han evidenciado con mayor fuerza en la región ciertos vacíos con relación especialmente a la cobertura temática y la periodicidad de las encuestas, como también respecto a la consistencia de los datos generados por diferentes fuentes y a la aptitud de las mismas para responder a las necesidades impuestas por las políticas y programas sociales. Esto ha llevado en algunos casos a la proliferación de encuestas, o bien a ampliar el contenido de los cuestionarios más allá de lo que razonablemente permite la definición de cada operativo y las características y limitaciones de los instrumentos utilizados. Un ejemplo de ello es el intento de abordar la investigación de los temas propios de las encuestas de ingresos y gastos mediante los procedimientos más ágiles y baratos, pero menos eficaces, de las encuestas de propósitos múltiples.

De allí que, teniendo presente que la situación en que se encuentran los distintos países de la región en materia de infraestructura de encuestas de hogares es muy diversa, para los próximos años se hace imprescindible impulsar acciones diferenciadas, acordes a la experiencia y grado de avance en que se encuentra cada país. Estas acciones, sin embargo,

debieran apuntar en último término a establecer en el mediano y largo plazo sistemas integrados de encuestas de hogares, que permitan aprovechar de manera simultánea y coherente las ventajas propias de cada relevamiento.

Desde el punto de vista regional existen dos grandes áreas de tareas técnicas asociadas a este esfuerzo. La primera comprende todas aquellas materias referidas al muestreo, en la perspectiva de la construcción de un sistema integrado de encuestas (desarrollo y perfeccionamiento del marco muestral, definición de muestras maestras, diseño de muestras panel o semipanel, etc.). La segunda, en tanto, abarca la recopilación y sistematización de experiencias acerca del diseño y contenido de los cuestionarios, técnicas de investigación, organización del trabajo de campo, métodos de evaluación de la calidad y confiabilidad de los resultados, integración de productores y usuarios y pertinencia de las diferentes encuestas en función de los objetivos perseguidos.

A su vez, el establecimiento y/o desarrollo de un sistema integrado de encuestas de hogares supone introducir mejoramientos institucionales y técnicos en los países, que en gran medida son funcionales al mejoramiento de las distintas investigaciones. Entre las tareas a asumir pueden destacarse:

- i) el fortalecimiento del nivel técnico de los funcionarios de las oficinas de estadística, tanto a nivel central como regional, incluyendo plantas permanentes o parcialmente permanentes de encuestadores;
- ii) la actualización o establecimiento de una muestra maestra;
- iii) la incorporación de una componente de panel en todos los relevamientos;
- iv) la conformación de equipos de trabajo de campo que incluyan la entrada de datos y una alta capacidad de reentrevista, junto a una solvente capacidad de supervisión;
- v) la incorporación de programas de suplementos preestablecidos y flexible;
- vi) en suma, el fortalecimiento institucional de las oficinas de estadística, con la radicación en ellas de todos los operativos.

El estado general del sistema de encuestas de cada país debiera servir de base para definir las acciones conducentes a su mejoramiento. Con el fin de ilustrar algunos cursos de acción en términos de productos que pueden constituir los primeros resultados de un programa integrado, se presentan a continuación tres ejemplos.

En un primer caso, se debieran sentar las bases de un mejoramiento del sistema integrado concentrando los esfuerzos en el fortalecimiento de la encuesta de hogares de propósitos múltiples. Esto supone, en primer lugar, aumentar su frecuencia, llegando incluso a atribuirle un carácter continuo, superando de esta manera la situación que presentan algunos de los países en cuanto a levantar la encuesta sólo una o dos veces al año.

En segundo lugar, ampliar el ámbito geográfico de la encuesta, con el fin que del área metropolitana o de algunas zonas urbanas se pase a cubrir a la población de la totalidad del territorio nacional. En tercer lugar, ampliar también la cobertura temática de la encuesta, a efectos de complementar la medición del empleo y del desempleo con información relevante para un abanico de temas sociales, que permitan establecer relaciones de causalidad y mejoren el espectro de las políticas cuyos resultados sean susceptibles de monitorearse con este instrumento. En cuarto lugar, en el caso de los países más avanzados de este grupo, la incorporación de módulos especiales, con el objeto de permitir explorar en mayor profundidad determinadas áreas de la realidad social. Por último, progresar en los aspectos metodológicos, operativos y de procesamiento de la información, con el fin de ganar en calidad y oportunidad.

En un segundo caso, los esfuerzos de nuevos levantamientos debieran concentrarse en el relevamiento de una encuesta del tipo caracterización socioeconómica, que apunte a complementar la información de la encuesta continua con un análisis más pormenorizado de las familias afectadas por situaciones de pobreza. Especial importancia tendría investigar el alcance geográfico y por estratos sociales de las políticas de gasto público, incluyendo el examen de los diversos programas que los gobiernos vienen promoviendo en los últimos años en su lucha contra los rezagos sociales. Se trataría de establecer un programa que contemple la realización de este tipo de encuestas cada dos o tres años. En este sentido, y recordando que se persigue mejorar la calidad de los datos, es imprescindible aprovechar la experiencia que ya ha acumulado la región en esta clase de encuestas.

Por último, en un tercer caso, los esfuerzos de nuevos relevamientos debieran centrarse en establecer un programa de encuestas de ingresos y gastos, en un modelo simplificado pero de mayor periodicidad, con el fin de conseguir cifras de gran calidad en torno a la evolución de los ingresos y de los gastos y de la estructura del consumo de las familias. Se trata de un objetivo ambicioso, pues requiere de mayores recursos y de una estructura institucional más consolidada. A la vez, permitiría tener un conocimiento más cabal de los efectos que los cambios estructurales en la producción, en la ocupación y en la distribución del ingreso vienen teniendo sobre la trama social, en términos de la distribución de los frutos del crecimiento económico en una nueva modalidad de desarrollo.

ANEXO

NÚCLEO BÁSICO DE ENCUESTAS A INCORPORAR EN UN SIEH

En una versión inicial y simplificada, el SIEH se puede concebir a partir de la integración de tres tipos de encuestas de hogares:

- Encuestas de Empleo (de propósitos múltiples)
- Encuestas de Condiciones de Vida
- Encuestas de Ingresos y Gastos

Cada una de estas encuestas posee características particulares, en atención a sus objetivos específicos, diseño muestral, contenido temático, periodicidad, cobertura geográfica y costo.

A continuación se describen esquemáticamente algunos de estos aspectos, con la finalidad de contribuir a avanzar en la discusión orientada a precisar los elementos e interrelaciones que pueden conformar un sistema integrado de encuestas de hogares en los países de América Latina.

I. ENCUESTAS DE EMPLEO (de propósitos múltiples)

Objetivos específicos

Medición y seguimiento del empleo, desempleo y caracterización del mercado de trabajo; e investigación de temas asociados.

Contenido temático

Módulo central: Características generales de la población, educación, medición del empleo, desempleo, caracterización del mercado de trabajo e ingresos.

Suplementos: Ampliación y profundización de los temas anteriores y complementación de nuevos temas. Los módulos especiales debieran, además:

- Formar parte de un programa con frecuencias predeterminadas.
- Aplicarse a la totalidad de la muestra o a una submuestra específica, según la naturaleza del tema.

- Incluir un número de preguntas que no represente una carga excesiva para los hogares de la muestra, al punto de afectar la confiabilidad de la información recogida en el módulo central.

Periodicidad (mínima)

Grandes ciudades o áreas metropolitanas: **Continua**, con estimaciones mensuales (mediante la técnica de promedios móviles) para las principales variables del mercado de trabajo.

Resto del área urbana : **Trimestral**, con integración de la medición del último mes del trimestre en las grandes ciudades, para generar estimaciones del área urbana nacional.

Area rural: **Semestral**, con integración de la medición urbana del segundo y cuarto trimestre, para generar estimaciones a nivel nacional.

Cobertura geográfica

Nacional, sujeta a las frecuencias señaladas en el punto anterior.

II. ENCUESTAS DE MEDICIÓN DE LAS CONDICIONES DE VIDA

Objetivos específicos

Evaluación y seguimiento de las condiciones sociales y de las situaciones de pobreza. Persiguen investigar con mayor extensión o profundidad ciertos aspectos de las condiciones de vida, como el acceso a determinados satisfactores y, muy especialmente, el impacto de políticas y programas sociales en el bienestar de diferentes sectores de la población.

Contenido temático

Además de las características generales y ocupacionales de la población, investigan con bastante detalle no sólo los recursos (ingresos) globales del hogar sino también el acceso efectivo a determinados programas asistenciales (alimentación escolar, subsidios alimentarios o de vivienda, etc.) y a los servicios públicos (salud, educación, infraestructura, etc.). Ello permite medir el impacto y evaluar el grado de focalización de esos programas, conjuntamente con el análisis de importantes atributos de los grupos beneficiarios, a base de información del hogar recogida por la encuesta y no así por los registros administrativos propios de cada programa.

Periodicidad

Cada dos años, con una duración del trabajo de campo de entre uno y tres meses.

Cobertura geográfica

Nacional, con representatividad a nivel de unidades geográficas pequeñas (comunas o distritos).

III. ENCUESTAS DE INGRESOS Y GASTOS**Objetivos específicos**

Análisis de la estructura del consumo de los hogares; cálculo de las ponderaciones del Índice de Precios al Consumidor; determinación de las elasticidades de los diferentes bienes y servicios; formación del ingreso y; estudio del nivel y distribución del ingreso y el gasto familiar.

Contenido temático

Características generales de las personas, hogares y viviendas y registro detallado del ingreso y del gasto (monetario y no monetario) de las familias. Cabe señalar que a partir de esto último es posible definir -entre otros- las canastas básicas de alimentos cuyo valor se utiliza en la estimación de las líneas de pobreza. Asimismo, estas encuestas efectúan un control (estadístico) de la estacionalidad del ingreso y de los gastos, en la medida que se entrevista a distintos hogares habitualmente a lo largo de todo un año, lo que redundará en una calidad relativamente buena de la información. A ello contribuyen también los instrumentos y procedimientos de captación, los métodos de evaluación (balance ingreso-gasto y chequeos en lugares de compra) y la selección del entrevistado (en muchos casos informante directo). Además, suelen integrarse en el marco conceptual de las Cuentas Nacionales, especialmente en lo que respecta a la estructura de consumo de los hogares, y su temática hace posible -en algún grado- el estudio del comportamiento de los gastos de consumo en respuesta al impacto de determinadas políticas económicas de coyuntura.

Periodicidad

Cada cinco años. Eventualmente, la encuesta de mitad del decenio podría realizarse con una duración del trabajo de campo inferior a doce meses (por ejemplo un trimestre), seleccionándose ese período de acuerdo a criterios de representatividad de determinados promedios anuales.

Cobertura geográfica

Nacional. De seguirse la modalidad señalada en el punto anterior, podría ser una encuesta nacional (la de duración anual) y la otra urbana.

NUCLEO BASICO DE ENCUESTAS DE UN SIEH

TIPO DE ENCUESTA	PERIODICIDAD	COBERTURA GEOGRAFICA
DE EMPLEO	Continua Trimestral Semestral	Grandes ciudades o áreas metropolitanas Resto del área urbana Area rural ↓ NACIONAL
MEDICION DE LAS CONDICIONES DE VIDA	Cada 2 años	NACIONAL (con representatividad a nivel de comunas o distritos)
INGRESOS Y GASTOS	Cada 10 años (Anual) Cada 5 años (A mitad del decenio) (Trimestral)	NACIONAL URBANA



**ALGUNAS CONSIDERACIONES SOBRE EL DISEÑO
ESTADÍSTICO DE LAS ENCUESTAS DE HOGARES**

JUAN CARLOS FERES

FERNANDO MEDINA

(CEPAL)

ÍNDICE

	<u>Página</u>
I. MARCO DE MUESTREO	51
1. Generalidades	51
2. Información del marco de muestreo	51
3. Construcción del marco de muestreo	52
4. Estratos y conglomerados	54
5. Sistematización de los marcos de muestreo	56
II. CARTOGRAFÍA	56
1. Cartografía automatizada y los Sistemas de Información Geográfica (SIGs)	56
2. Requerimientos de la digitalización	59
III. SELECCIÓN DE LAS MUESTRAS	60

ALGUNAS CONSIDERACIONES SOBRE EL DISEÑO ESTADÍSTICO DE LAS ENCUESTAS DE HOGARES

I. MARCO DE MUESTREO

1. Generalidades

Un marco de muestreo se define como una lista exhaustiva -organizada en forma de base de datos- que contiene todos y cada uno de los elementos de una población de interés que participarán en las distintas fases de selección de la muestra. A su vez, el marco está formado por un conjunto de mapas y planos a diferentes escalas, que permiten la delimitación física de las diversas unidades de selección. Asimismo, se considera parte fundamental de éste los registros físicos de las unidades de viviendas, así como los listados en que se detallan las referencias que permiten a los encuestadores la plena identificación de las viviendas seleccionadas.

El marco debiera contener información sobre la división administrativa del país - subdivisiones políticas o zonificación estadística definida para efectuar la enumeración del censo de población-, como también sobre los volúmenes de viviendas y de población total, por grupos de edad y sexo, entre otras variables necesarias para clasificar a los hogares de acuerdo a determinadas características según los objetivos específicos de la encuesta.

En un buen marco de muestreo todos los miembros del universo de estudio deben tener una probabilidad conocida, y distinta de cero, de formar parte de alguna muestra. Existen estudios en que deliberadamente se excluyen algunas zonas o regiones del país, o determinadas subpoblaciones; estos casos no violan el principio señalado, pero ello debe aclararse en los documentos que hacen referencia a la población encuestada a fin de que no se generalicen los resultados a sectores no considerados en el universo de estudio.

Por último, el marco debe estar exento de duplicaciones y omisiones, y actualizarse de manera permanente con el objeto de reflejar los cambios estructurales que se van produciendo en la geografía del país, la conformación de las unidades de selección, su distribución física, el surgimiento de nuevas edificaciones y las construcciones que se generan debido al crecimiento de la población.

2. Información del marco de muestreo

La información contenida en un marco de muestreo puede referirse a áreas geográficas o a una lista de unidades de viviendas o personas. Los estadísticos se refieren a estas alternativas como marco de áreas y de lista, respectivamente. En la mayoría de los

países de la región se utiliza habitualmente una combinación de ambos para organizar la información necesaria para elaborar los diseños de muestra de las encuestas.

Es común que en las encuestas donde las unidades de observación se seleccionan en varias etapas -esquemas polietápicos-, se parta de la elección de un conjunto de áreas geográficas perfectamente delimitadas, y posteriormente se elabore un marco de lista que incluye únicamente a las unidades que pertenecen a las áreas previamente seleccionadas. Es decir, se elabora un marco de muestreo para cada una de las etapas involucradas en la selección de la muestra. Mediante este procedimiento se reducen los costos de construcción y actualización del marco, ya que para las fases subsecuentes éstos sólo incluyen aquellas unidades que fueron seleccionadas en las etapas anteriores.

Por otra parte, los marcos suelen incorporar información sobre las diferentes unidades de muestreo, las que son agrupadas conforme a criterios que permitan mejorar la eficiencia del diseño. En aquellos casos en que se trata de un esquema polietápico, a las unidades utilizadas en la primera etapa se les denomina unidades primarias de muestreo (UPM's), y a las que se seleccionan al final, unidades últimas de muestreo (UUM's). Entre esos dos extremos existen múltiples niveles, pero lo habitual es que en las aplicaciones de carácter práctico los esquemas de selección comprendan, como máximo, dos o tres etapas.

En la realidad, es usual que las organizaciones nacionales de estadística definan la construcción de un único marco con el propósito de realizar diferentes rondas de una misma encuesta o bien investigaciones puntuales de objetivos diversos. En este caso, el marco que permite la elaboración de los distintos diseños, así como la elección de las diferentes muestras, se denomina Marco Maestro de Muestreo (MMM), que representa un insumo fundamental en el diseño de un SIEH. Asimismo, es en este contexto en el que se generan las muestras maestras, definidas como toda muestra a partir de la cual pueden seleccionarse submuestras independientes -denominadas réplicas- o grupos de unidades con algún nivel de traslape que se utilizan para diferentes rondas de una encuesta o para proyectos específicos.

3. Construcción del marco de muestreo

Al finalizar la enumeración censal normalmente las divisiones de encuestas de las ONE's dedican esfuerzos y recursos a la construcción del marco maestro de muestreo. Esta etapa se asocia a la conformación de una base de datos que permite el diseño y selección de diferentes tipos de muestras, vale decir que posibilita la realización de distintas encuestas o la formación de réplicas -o submuestras interpenetrantes- necesarias para la ejecución de las encuestas periódicas o continuas que forman parte de los programas permanentes que llevan a cabo los países.

En la mayoría de los países en que se realizan encuestas de hogares, el diseño de las mismas y la selección de las unidades de observación se efectúa a partir de información obtenida de los censos de población y vivienda. Sin embargo, en ocasiones, y ante la falta de información censal actualizada, se recurre al uso de registros administrativos o mapas del

catastro elaborados por otras dependencias públicas para identificar a las unidades de vivienda. De modo que a pesar de la existencia en esos países de diversos marcos de muestreo, es necesario reconocer que deben modernizarse los métodos de trabajo utilizados para su administración y actualización, a fin de incorporar las nuevas tecnologías y las metodologías de punta que vienen siendo desarrolladas, lo que además permitiría contar al mediano plazo con cartografía automatizada y una base de datos adecuada para el manejo de la información.

La preparación y ejecución de los censos de población demanda la inversión de un monto considerable de recursos, dentro de lo cual -como se verá más adelante en el punto II- la elaboración y actualización de la cartografía es una tarea fundamental. Es así que el marco de muestreo para las encuestas se va formando de hecho en las etapas de campo previas a la ejecución del censo, debido a que los trabajos censales requieren disponer de una base cartográfica actualizada y de buena calidad.

Una vez que concluyen los trabajos de enumeración censal, surge la oportunidad de diseñar los planes y programas de las encuestas con las que se produce información para atender las demandas de los usuarios en los períodos intercensales. En términos generales, la conformación de los marcos de muestreo sigue los pasos que se detallan a continuación.

En primer lugar, la base informativa generada a partir del trabajo de campo de los censos posibilita la construcción de un marco de áreas, integrado por mapas de línea a diferentes escalas y una base de datos con la división geográfica, administrativa y estadística del país. Esta información debiera ser respaldada en un sistema automatizado, formado por una base de datos relacional, que permita la consulta directa a los usuarios, y por la cartografía digitalizada, orientada a la elaboración de mapas y a su actualización en forma automatizada.

Durante la elaboración del marco, cada unidad estadística debiera identificarse en las diferentes fases y acompañarse de información complementaria que haga posible definir su importancia relativa respecto a las demás (número de viviendas, población total, población por edad y sexo, etc.). También es importante incorporar datos sobre algunas otras características de las unidades de selección, con el objeto de contar con elementos que faciliten efectuar agrupaciones -estratificaciones- que mejoren la eficiencia del diseño y disminuyan la variabilidad de los estimadores (en caso que la variable de estratificación esté altamente correlacionada con los parámetros que se desean estimar).

En segundo lugar, las unidades censales se utilizan por lo general para formar agrupaciones (que se conocen con el nombre de conglomerados y que también forman parte del marco de muestreo por áreas) definidas comúnmente como unidades de primera etapa o unidades primarias de muestreo (UPMs). Según las características de las unidades censales, es posible que éstas puedan utilizarse como conglomerados (teniendo presente que en cualquier encuesta la primera etapa es la que genera la mayor contribución a la varianza total del diseño).

Como ha sido ampliamente documentado en textos especializados y en diversos informes técnicos elaborados por la División de Estadística de las Naciones Unidas, las unidades de observación agrupadas en conglomerados son muy distintas entre sí (heterogéneas) y tienden a incrementar la varianza de los estimadores, ya que estos conglomerados son generalmente pequeños en comparación con el universo de estudio. Sin embargo, en la práctica del muestreo es común la formación de conglomerados, ya que representan espacios geográficos que se constituyen como unidades de primera etapa para la selección de la muestra (en los documentos técnicos de Naciones Unidas a estas unidades se les denomina áreas de enumeración).

Al interior de las unidades de primera etapa existe una cierta afinidad entre los elementos que las conforman, por lo que su contribución a la varianza total (intraconglomerados) es menor que aquella que aporta las diferencias entre las agrupaciones (variación entre conglomerados). Luego, al interior de los conglomerados se definen agrupaciones más pequeñas constituidas por grupos de viviendas, que usualmente se conocen como segmentos compactos o áreas de listado y que se denominan unidades secundarias de muestreo (USMs). En esta etapa se requiere contar con un listado exhaustivo de las agrupaciones, en el que se detallen las características del sector y permitan a los entrevistadores la identificación de las viviendas seleccionadas. Finalmente, al interior de las USMs se listan todas las viviendas, las que son consideradas como unidades últimas o finales de muestreo (UUMs).

4. Estratos y conglomerados

La estratificación se refiere a la subdivisión de una población determinada en subconjuntos con características propias. Esta acción se lleva a cabo previo a la selección de la muestra y la(s) variable(s) que se utiliza(n) para ello debe(n) contener información acerca de todas las unidades de la población.

El objetivo de este procedimiento es reducir la varianza del parámetro de interés, por lo que las variables de estratificación deben estar altamente correlacionadas con aquella(s) utilizada(s) para la determinación del tamaño de muestra. De modo que los estratos son, por definición, homogéneos en su interior, lo que, de paso, establece una diferencia fundamental respecto a las características de los conglomerados.

Asimismo, los estratos no suelen utilizarse como unidades de selección debido a que es normal que todos ellos estén representados en la muestra, mientras que la naturaleza de los conglomerados los define como las típicas unidades de selección en las primeras etapas, ya que su heterogeneidad garantiza que en su interior queden incluidos diferentes elementos de la población.

Como ya se señaló, en general la formación de conglomerados tiende a incrementar la varianza muestral. Sin embargo, su utilización en el diseño de la muestra

permite flexibilidad en el número de etapas de selección y reduce substancialmente los costos asociados a la construcción del marco y a la realización del trabajo de campo.

Es común que en la mayoría de las encuestas de hogares que se llevan a cabo en los países de la región se combinen las técnicas de estratificación y conglomeración, como condición necesaria e indispensable para garantizar un diseño muestral óptimo. Sin embargo, cabe precisar que la estratificación se realiza con un objetivo específico, que está inducido por los fines de la encuesta, y guarda estrecha relación con las variables de interés. En otras palabras, es muy difícil que exista una estratificación óptima que satisfaga, con la misma precisión estadística, las necesidades de todos los temas incorporados en una encuesta y, por ende, a los distintos usuarios de la información.

Lo anterior es relevante en la medida que existen iniciativas que promueven la realización de encuestas multitemáticas que no siempre evalúan, con el rigor técnico del caso, la correlación que existe entre la estratificación del marco muestral, la variable de diseño para la determinación del tamaño de muestra y el error relativo de los diferentes indicadores que se construyen en base a las preguntas contenidas en el cuestionario y asociadas a los diferentes temas que se investigan.

Es probable que estos estudios partan de la hipótesis que existe una posible relación entre algunos de los indicadores obtenidos, la estratificación del marco y la variable de diseño. Sin embargo, también es probable que esta correlación no sea significativa desde el punto de vista estadístico, en cuyo caso se comprometen o invalidan las inferencias que se puedan hacer sobre el comportamiento de un fenómeno perteneciente a un determinado dominio de estudio. De modo que resulta indispensable estimar los errores de muestreo y calcular el efecto de diseño de las variables, con el propósito de evaluar si el tamaño de muestra utilizado resultó apropiado para los fines de la encuesta.

Por último, es importante señalar que, a diferencia de la etapa de estratificación que tiene un fin específico para cada encuesta, los trabajos que se realizan para la formación de conglomerados son de beneficio común y se pueden compartir entre las diferentes encuestas de un programa permanente. Esto significa que la definición y formación de los conglomerados, la sistematización de la información de base, la elaboración del material cartográfico y su actualización, son insumos que se pueden aprovechar para desarrollar varias rondas de una misma encuesta o para realizar investigaciones independientes sobre distintos temas, utilizando mecanismos de estratificación apropiados que mantengan una correlación significativa con la variable de diseño. Así, el costo que implica la administración de los conglomerados a objeto de mantenerlos actualizados se amortiza entre las diferentes rondas o encuestas, generando economías de escala que permiten la puesta en marcha de un programa integrado de encuestas de hogares a costos razonables y que informe de manera periódica sobre una amplia gama de temas de interés.

5. Sistematización de los marcos de muestreo

Conforme a lo expuesto, es necesario que las ONEs elaboren marcos de muestreo ad-hoc para cada una de las etapas incluidas en el diseño. Estas actividades forman parte de las labores rutinarias de las divisiones de encuestas; sin embargo, lo que aún no está muy extendido es que se incorporen en los trabajos de sistematización las metodologías y facilidades que brindan los sistemas modernos de manejo de bases de datos, que constituyen una herramienta sumamente útil para apoyar la construcción y administración de los marcos de muestreo.

En este sentido, se debieran definir y asignar prioridad al logro de algunos objetivos básicos a mediano plazo, tales como: i) establecer un vínculo estrecho entre el material cartográfico elaborado para los censos de población y vivienda y las necesidades informativas para la formación de los marcos de muestreo requeridos para el diseño de las encuestas; y ii) construir marcos de muestreo sistematizados que permitan, entre otras cosas, la actualización automática de la información, el manejo de los procedimientos estadísticos para la selección de las unidades de observación, el cálculo de los factores de expansión, la generación de cartografía automatizada y la administración y actualización de las bases de datos.

II. CARTOGRAFÍA

1. Cartografía automatizada y los Sistema de Información Geográfica (SIGs)

La identificación de las unidades de vivienda que forman parte de la muestra requiere privilegiar su posición física antes que su precisión geográfica. Esto significa que se deben establecer claramente las diferencias entre cartografía automatizada y las facilidades y alcances de un Sistema de Información Geográfica (SIG), ya que excederse en las pretensiones de sistematización implica un gasto considerable de recursos, que podría no justificarse ante la ausencia de una estrategia institucional para desarrollar un proyecto que vincule la ejecución de encuestas con el uso de la información estadística y geográfica en un mismo plano.

Debido a los avances tecnológicos ocurridos en los últimos años resulta tentador el modernizar los procedimientos actuales de manejo de información. De allí la notable relevancia que han cobrado los SIG's para la gestión de proyectos y el interés que ha surgido por incorporarlos en diversos campos de trabajo. La naturaleza de esta herramienta la hace muy apropiada para ubicar información estadística y geográfica, de diferentes temas, en el medio físico. De hecho, la posibilidad de sobreponer cartografía básica y temática representa una de las enormes ventajas de esta tecnología, y en la actualidad los SIG's se utilizan en áreas tan diversas como la gestión ambiental, el desarrollo urbano, la mercadotecnia, los servicios públicos y privados, el desarrollo de infraestructura básica, etc..

Sin embargo, las capacidades y usos de los SIG's se dan en un contexto más amplio que la generación de cartografía automatizada para el trabajo de campo de las encuestas. Es así que un SIG se puede definir como un sistema de información que permite la captura de datos, su almacenamiento, la verificación de la consistencia, la integración, la manipulación y análisis, así como el despliegue visual de la información en un contexto espacial (Chorley, 1987). En términos generales, éste consta de una plataforma mínima de información, la cual debe ser resguardada en una base de datos relacional, con atributos específicos para las observaciones, a fin de que permita el manejo de las variables y la superposición espacial de diferentes temas. Además, el hecho de disponer de cartografía georeferenciada en medios magnéticos brinda la posibilidad de ubicar físicamente la información con precisión métrica.

La idea de incorporar la herramienta de los SIG's en la gestión de proyectos estadísticos fue manifestada por J. Durbin en la reunión de la Royal Statistical Society en 1987. Durbin manifestó que "si a un mapa se le puede agregar información que permita ubicar a las oficinas de correo, los registros de electores, o cualquier otro tipo de información, se podría formular una estrategia estadística para relacionar los datos económicos y sociales provenientes de los censos y las encuestas con otros datos de utilidad para la sociedad". De modo que contar con cartografía automatizada representa una gran ventaja para los administradores de los programas de encuestas. Sin embargo, se pueden elaborar cartas y planos de línea en medios magnéticos, sin que esto signifique la disponibilidad de un SIG. Igualmente, se puede disponer de un sistema de base de datos que resguarde la información del marco de muestreo y sobreponerla en los planos de línea o cartas básicas o temáticas, sin que ello represente trabajar con un SIG.

Un aspecto que distingue a los SIG's de los actuales sistemas de manejo de información estadística y geográfica con fines de muestreo, es el hecho que para garantizar la operación eficiente de un SIG es condición indispensable que la cartografía básica se encuentre georeferenciada y se disponga de una base de datos con estructura relacional que permita el manejo de información de varios temas.

En la actualidad, la mayoría de las oficinas de estadística de la región no tienen como responsabilidad la generación de cartografía o de información sobre temas relativos al medio físico. De allí que los alcances de un SIG rebasan las atribuciones y necesidades de las divisiones de encuestas, por lo que se debe ser prudente para enfrentar el desarrollo de actividades en este campo, sin contar con un proyecto que contemple la vinculación de los trabajos de las ONEs con las tareas que realizan los Institutos Geográficos Nacionales.

Lo anterior no significa que se deba abandonar la idea de diseñar un sistema eficiente que, además de servir para los fines de ejecución de las encuestas, contribuya al estudio de diversos temas de interés en donde el posicionamiento espacial de las unidades de vivienda resulte fundamental. Por el contrario, se debiera identificar las áreas de trabajo en donde los SIG's podrían apoyar la gestión de proyectos estadísticos. En primer lugar, el uso de los SIG's puede asociarse con las etapas de diseño, recolección de información y análisis de resultados. En segundo lugar, es evidente que en relación con el diseño de

encuestas los SIG's pueden contribuir a la formación, actualización y administración de los marcos de muestreo y, en la medida que se relacionen bases de datos sobre diferentes temas, apoyar la construcción de marcos maestros. Asimismo, en el diseño estadístico de la encuesta es posible utilizar variables de la base de datos para estratificar el universo de estudio y, además, apoyar la formación de las unidades de selección, la delimitación de segmentos en las zonas urbanas y las áreas de conteo en las zonas rurales. Por su parte, la etapa de recolección de datos en campo puede ser aprovechada para captar información sobre los servicios públicos y privados disponibles en las áreas visitadas, así como su ubicación geográfica. Esta información es de gran utilidad y puede utilizarse para estudios en el ámbito del desarrollo urbano y la generación de servicios en zonas carenciadas. Por ejemplo, establecer la distancia de las viviendas a las escuelas y centros de salud y abastecimiento, o bien, determinar la distancia promedio entre las viviendas y las tomas públicas en aquellas áreas rurales que carecen de suministro de agua por tubería.

En la etapa de análisis se puede distinguir al menos dos fases: el análisis de consistencia de la información de la encuesta y la vinculación de ésta con datos disponibles sobre el medio físico (variables geográficas). En cada una de ellas los programas de análisis son, por lo general, diferentes; sin embargo, ya existen paquetes estadísticos que contemplan rutinas para el manejo de bases de datos geográficas bajo un enfoque SIG (por ejemplo, el módulo especial del Statistical Analysis System - SAS) .

En síntesis, la automatización de las actividades relativas al diseño de encuestas es posible enfrentarlas por etapas. En primera instancia, y en ausencia de un proyecto nacional de SIG, las oficinas de estadística debieran orientar sus esfuerzos a sistematizar la información disponible sobre el marco de muestreo y digitalizar la cartografía en uso, sin prestar gran atención a la precisión geográfica de la ubicación física de las viviendas. No obstante, en contacto con los institutos geográficos nacionales, se debiera cautelar que en el diseño y desarrollo del sistema se observen las normas técnicas que permitan posteriormente relacionar la información de las viviendas con otros temas.

Luego, en el contexto de un proyecto más amplio, se puede asumir la tarea de georeferenciar la cartografía. En el caso que se disponga de una red geodésica, ésta se podrá utilizar con fines de actualización. Así, las brigadas de campo, mediante el uso del GPS, pueden contribuir a mantener vigente la cartografía, sobreponiendo información sobre los rasgos físicos del terreno, la infraestructura, los servicios, las vías de comunicación, etc.. Cabe señalar que las metodologías modernas sobre desarrollo regional utilizan información del medio físico combinada con datos socioeconómicos y demográficos, a fin de lograr una mejor aproximación a las interacciones entre los fenómenos geográficos y sociales. Sin embargo, se debe estar consciente que las actividades en el ámbito geográfico suelen tener fines más amplios que rebasan por mucho los requerimientos de la elaboración y administración de un marco de muestreo.

2. Requerimientos de la digitalización

En términos generales, para disponer de un mapa digitalizado es necesario cumplir un conjunto de requisitos básicos, entre los que cabe destacar:

- Contar con los mapas a digitalizar. Estos debieran incluir, como mínimo, la división político-administrativa del país (estados, provincias, municipios, etc.). No obstante, también es posible elaborar productos cartográficos a partir de listas de coordenadas con información en cualquier sistema de referencia (cartográfico o euclidiano), ya sea que provengan de planos elaborados directamente en campo en forma manual o con ayuda de un GPS. A su vez, otra alternativa consiste en identificar determinadas zonas sobre la base de fotografías aéreas o imágenes de satélite de alta resolución.

- Definir el nivel para el cual se requiere interpolar los datos para efectos de su representación espacial y precisión estadística, dado que las encuestas normalmente recaban información sólo de una parte del universo de estudio.

- Disponer de programas para la digitalización y captura de los datos numéricos referidos a la información del marco de muestreo o de las variables de la encuesta. Para la digitalización existen distintas alternativas como: AUTOCAD, ARC/INFO, GRASS e IDRISI. Por su parte, para la información numérica es posible utilizar hojas de cálculo como EXCELL, o manejadores de bases de datos como DBASE, FOXPRO o ACCES, entre otras.

- En un mapa digitalizado cada entidad geométrica que lo compone tiene asociado un atributo, con valor numérico, que se utiliza como identificador y permite relacionar el mapa con la base de datos tabular.

- En un contexto ideal, la actualización cartográfica se puede realizar a partir de dotar a las brigadas de campo de un GPS portátil, con el fin de referir la posición geográfica del lugar en donde se efectúa la encuesta, con una exactitud tal que permita su representación espacial y su manejo cartográfico.

- Para cumplir con eficiencia estas tareas se requiere, naturalmente, capacitar al personal en el manejo de información estadística y geográfica, así como en las funciones de digitalización, manejo de bases de datos y análisis espacial.

Los equipos y sistemas recomendables dependen de la magnitud y temporalidad del proyecto. Así, si el objetivo es diseñar un sistema permanente de información, con cobertura nacional, se sugiere utilizar estaciones de trabajo con plataforma UNIX y la infraestructura necesaria para la edición e impresión de mapas. A su vez, si el propósito es generar herramientas para la elaboración de diagnósticos y trabajos de investigación que apoyen la toma de decisiones a nivel local, entonces existe la posibilidad de operar con computadoras personales con una configuración mínima que garantice capacidad de

almacenamiento y velocidad de proceso (Pentium 150 mhz., 32 mb de memoria ram, lectora de CD-Rom, monitor SVGA con 2 mb de memoria y tarjeta de gráficos para resolución de 1024 columnas por 768 renglones).

La elección de la mesa digitalizadora es independiente del tipo de plataforma que se disponga. Obviamente se debe velar porque tenga un tamaño adecuado. Existen en el mercado varias opciones (CALCOMP, ALTEK, GTCO, KURTA) y el proveedor deberá entregar los programas de enlace entre la tableta digitalizadora y el sistema de cómputo que se esté utilizando.

Finalmente, los dispositivos para la salida de los mapas pueden ser plotters de plumas para hojas de papel de 36 pulgadas e impresoras a color de inyección de tinta.

III. SELECCIÓN DE LAS MUESTRAS

Uno de los objetivos fundamentales del diseño muestral es definir el procedimiento que se aplicará para seleccionar las unidades de observación que formarán parte de la muestra. En términos generales, lo más apropiado sería seleccionar directamente las unidades de observación a partir de un listado que considere a todas y cada una de las unidades objeto de estudio. Sin embargo, proceder de esta manera incrementa apreciablemente el costo del trabajo de campo ya que, en caso que existiera un listado que incorporara todos los elementos del universo de estudio, la dispersión de la muestra implicaría grandes desplazamientos de los entrevistadores para visitar las distintas unidades seleccionadas, lo que incrementaría los costos a un nivel que haría casi inviable cualquier proyecto de encuesta.

Debido a esos altos costos asociados a la selección de una muestra aleatoria (en diseños monoetápicos con selección por muestreo aleatorio simple), es una práctica habitual el que los países utilicen diseños polietápicos; vale decir, métodos de selección en que las unidades de observación son elegidas luego de diversas fases que combinan procedimientos de conglomeración y estratificación. A su vez, en estos casos es necesario construir un marco para cada una de las etapas definidas para la selección de la muestra.

De esta manera, se busca formar unidades de marco de tamaños similares, a fin de reducir la variabilidad de las probabilidades de selección. En el caso en que las unidades formadas no difieren considerablemente en relación al número de unidades que las forman, es posible aplicar un procedimiento de selección que asigne igual probabilidad a todas las unidades. Sin embargo, es frecuente que existan diferencias considerables en relación al tamaño, por lo que en estas situaciones se sugiere que la asignación de las unidades en muestra se efectúe asignándoles una probabilidad proporcional a su tamaño, expresado por ejemplo en el total de viviendas que incluyen.

Es necesario controlar cuidadosamente la aplicación del procedimiento de selección expresado en el esquema de muestreo adoptado, y llevar un registro detallado de las unidades incluidas en la muestra y de sus probabilidades de selección, a efectos de que sea posible construir en forma clara el factor de expansión que permitirá extrapolar los resultados observados en la muestra al total del universo de estudio.

Previo a la selección de las unidades de vivienda en los listados correspondientes, se debiera depurar la lista disponible a partir de un recorrido de campo que permita conocer la situación de las viviendas que formarán parte de la investigación. Así, se debiera actualizar el listado identificando bajas y altas de unidades de vivienda, poniendo especial énfasis en las de nueva creación y en aquellas que se dedican a una actividad económica pero que a la vez operan como lugar de resguardo y convivencia para las familias.

En la fórmula aplicada para el cálculo del factor de expansión, se debieran precisar las diferencias entre los pesos asignados en gabinete a las unidades seleccionadas y aquellos determinados como fruto de los recorridos de actualización en campo. Es habitual que existan diferencias entre ambas mediciones, por lo que en el hecho resulta muy difícil preservar la característica de autoponderación en el diseño de una encuesta.

Cuando no se sigue este procedimiento, suele adoptarse el supuesto de que no han ocurrido cambios significativos entre la fecha de elaboración del marco de muestro y la de realización del trabajo de campo, lo cual obviamente es incorrecto. A fin de intentar superar esta situación, la práctica cotidiana supone que el problema se resuelve, al menos en parte, expandiendo los resultados de la encuesta a una proyección de población que tome en cuenta la dinámica demográfica del país, pero que pocas veces considera la distribución espacial de la población. Ello permite reducir los errores en los estimadores nacionales, lo que no impide que en muchas ocasiones se agudicen las diferencias regionales, generando tasas de crecimiento inconsistentes entre dos o más encuestas de un mismo programa.

Otro elemento que también es importante señalar en este apartado es el que se refiere a la práctica, muy difundida en algunas oficinas de estadística de la región, de efectuar reemplazos de viviendas en los casos de pérdida de muestra por motivos de no respuesta. De hecho, es común que las altas tasas de respuesta que reportan algunas investigaciones por muestreo escondan prácticas deliberadas de reemplazos efectuados a criterio del entrevistador y sin ningún control por parte de los encargados del diseño estadístico de la encuesta. Como se sabe estas sustituciones alteran las probabilidades de selección de las viviendas e introducen sesgos en la información que son difíciles de evaluar. De modo que para lograr la precisión deseada en las estimaciones se sugiere no fomentar estas iniciativas, sino más bien, alternativamente, ajustar el tamaño de muestra por motivos de no respuesta, con el objeto de obtener durante la ejecución del trabajo de campo el número mínimo de selecciones que garantice la confianza requerida.

**PROPUESTA PARA EL DISEÑO E IMPLEMENTACIÓN DE UN
SISTEMA DE INFORMACIÓN GEOGRÁFICO PARA APOYAR
LA GESTIÓN ADMINISTRATIVA Y OPERATIVA DE LAS
ENCUESTAS DE HOGARES**

SERGIO SOMMERVILLE, CONSULTOR

CEPAL

ÍNDICE

	<u>Página</u>
I. INTRODUCCIÓN.....	66
II. MARCO CONCEPTUAL PARA EL DESARROLLO DE INSTRUMENTOS DE APOYO.....	66
1. La importancia de la planificación del proyecto	68
2. Consideraciones claves en la implementación del SIG.....	68
3. La selección de sistemas.....	70
4. Factores adicionales de la planificación.....	71
III. LA CARTOGRAFÍA DIGITAL Y LOS SIG PARA LAS ENCUESTAS DE HOGARES	74
1. Productos para la planificación y la ejecución de trabajos de campo.....	74
2. Segmentación espacial y generación de muestras.....	74
3. El modelo urbano.....	74
4. El modelo rural	76
IV. LOS SISTEMAS DE INFORMACIÓN GEOGRÁFICOS Y SU DESARROLLO.....	77
1. Introducción.....	77
2. Los Sistemas de Información Geográficos.....	78
3. ¿Qué es un SIG?.....	80
4. Preguntas que un SIG puede contestar	84
5. La necesidad de una metodología para la implementación del SIG.....	85
6. El modelo de información SIG.....	86
7. Propuesta de diseño para la implementación del SIG	89
8. Esquema institucional de un SIG	89
9. El proyecto SIG	91
V. LA BASE DE DATOS GEOGRÁFICA	92
1. Introducción.....	92
2. Propuesta para la creación de la base de datos geográfica	93

Página

3. Uso de la cartografía nacional	96
4. Criterios para la clasificación de las cartas	97
5. Cadenas de control de calidad	97
VI. EL PROCESO DE CAPACITACIÓN	99
VII. EQUIPAMIENTO Y COSTOS DE UNA INSTALACIÓN SIG	100
1. Los equipos y programas necesarios para implementar un SIG	100
2. Consideraciones referentes a los dispositivos	101
3. Consideraciones referentes a los programas SIG	105
4. Costos generales de la implementación de un SIG	106
5. Detalle de costos	107
VIII. CONCLUSIONES	109

PROPUESTA PARA EL DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA DE INFORMACIÓN GEOGRÁFICO PARA APOYAR LA GESTIÓN ADMINISTRATIVA Y OPERATIVA DE LAS ENCUESTAS DE HOGARES

I. INTRODUCCIÓN

El tema de los Sistemas de Información Geográficos (SIG) es de suma importancia en el manejo y explotación de la información y, en particular, a aquella que se refiere a la generación de datos estadísticos vinculados con características del medio geográfico. Así, ante el avance reciente de esta metodología y de su incorporación inminente como una herramienta de apoyo para la planificación y ejecución de los procesos estadísticos, el interés de la CEPAL se centra en dar a conocer a los países metodologías de vanguardia que apoyen el fortalecimiento de sus sistemas de información, a partir de la incorporación de herramientas y técnicas modernas orientadas a hacer más eficientes sus procesos y métodos de trabajo.

La labor que llevan a cabo la mayoría de las Oficinas de Estadística de la región se suscribe a la generación de información estadística. Por este motivo, es de suma importancia que cuenten con sistemas que les permitan integrar y relacionar la cartografía básica elaborada a partir de los censos de población y vivienda, económicos y agropecuarios, con la confección del marco de muestreo que se utiliza para la realización de encuestas en hogares y establecimientos, así como con otros temas de interés público que se orientan a la definición de políticas o al desarrollo y evaluación de proyectos de inversión pública o privada.

En este sentido, es muy importante que los directores de las Oficinas de Estadística conozcan y aprecien las capacidades y posibilidades de una herramienta que permite la sistematización, manipulación y representación espacial de los datos, y ubiquen su utilidad en el contexto actual en que se desarrolla la elaboración de cartografía para la realización de censos y encuestas.

II. MARCO CONCEPTUAL PARA EL DESARROLLO DE INSTRUMENTOS DE APOYO

Los sistemas de información geográficos insertos como instrumento de apoyo global a la gestión de organización, explotación y entrega de información dentro de las Oficinas de Estadística y, en especial, como herramienta de apoyo a la gestión de las Encuestas de Hogares, poseen necesariamente un significado de reestructuración y modernización de los

procesos informáticos. La implantación institucional del SIG requiere de un proceso de transformación administrativa que incorpora dentro de la orgánica institucional cada una de las actividades y tareas que son vitales para el efectivo funcionamiento del SIG y su mantenimiento.

Los SIG, como sistemas integrados de recursos humanos, información, procedimientos, lineamientos institucionales y objetivos, programas de capacitación y perfeccionamiento, dispositivos computacionales y directrices administrativas, poseen capacidades y aplicaciones que sobrepasan el ámbito específico de acción de las Oficinas de Estadística y, en particular, a aquel de las encuestas por muestreo, las cuales para su planificación y ejecución requieren tan sólo de cartografía actualizada y una base de datos sistematizada que permita realizar ciertas operaciones para la actualización del marco de muestreo, así como para la selección de las unidades de observación. La aplicabilidad de los SIG se extiende, como instrumento de análisis espacial, a una variedad de campos donde se requiere analizar la ocurrencia de ciertos fenómenos y su relación con el medio que lo rodea. Esta aplicabilidad se extiende a las áreas económicas, agropecuarias, forestales, en la salud y educación, a los estudios de necesidades básicas insatisfechas, dentro de las instancias académicas, en el marketing y en una amplia variedad de sectores del quehacer de un país.

Hoy día, pareciera que los principales problemas de las Oficinas de Estadística se ubican en el contexto de la falta de recursos para la actualización periódica del marco de muestreo más que en restricciones para administrar los insumos de trabajo o de capacidades técnicas para diseñar procesos de automatización que les permitan facilidades en el manejo del marco de muestreo. Así, a pesar de que en algunos países se han desarrollado esfuerzos muy importantes por digitalizar la cartografía de base que se utiliza para apoyar la realización de censos y encuestas y han automatizado los procesos de selección de la muestra, no se puede afirmar que dispongan de un SIG, y sus mayores problemas se presentan por la falta de recursos asignados para la actualización sistemática del marco de muestreo. Este último aspecto posee una relación directa con la calidad de la información recolectada. En la mayoría de los casos los esfuerzos requeridos para mantener actualizada una base de datos digital-cartográfica son notablemente mayores a los esfuerzos que fueron necesarios para su creación. La representación digital de una área geográfica precisa evolucionar al mismo ritmo que ocurren las transformaciones en el terreno, en las vías, en la construcción, etc.

Existe una tentación natural, cuando la tecnología SIG se hace disponible, de improvisar las soluciones a través de su uso; esto es iniciar el proceso sin considerar donde nos llevará el proyecto. El gran peligro de realizar decisiones bajo la emoción derivada de herramientas vistosas es que deberán ser revertidas más tarde o el costo de su implementación sea muy elevado. Como consecuencia de este hecho es posible que el proyecto SIG deba ser abandonado. Para evitar estas experiencias desalentadoras, este documento propone una metodología de planificación que considera en forma objetiva el establecimiento de los objetivos, definición de las actividades a considerar y la estimación de los costos para el proyecto.

El fundamento esencial en el diseño del proyecto SIG hace referencia a la componente espacial, la cual debe ser una parte integral del proyecto desde el principio; si en cambio esta componente se asume como un elemento adicional, que se incorpora en cualquier momento del desarrollo, nos enfrentamos ineludiblemente a un resultado poco conveniente.

El diseño e implementación de un SIG es complicado y se debe asumir de esta forma. Aun cuando un SIG puede ser ejecutado desde un punto de vista técnico consistente, un diseño adecuado ejerce una influencia fuerte para que los productos de información generados puedan ser utilizados con efectividad. La evaluación de las necesidades de los usuarios, la capacitación, la recolección de datos y el proyecto piloto son vitales y críticos para alcanzar una implantación institucional consistente y garantizar el éxito del proyecto.

Más allá de los procesos técnicos propios que deben ser abordados para generar los insumos del SIG, sean estos la base de datos relacional, las base de datos digital cartográfica, los procedimientos informáticos para automatizar las actividades y tareas, los trabajos de mantenimiento y la generación de instrumentos de entrega de información, el SIG debe obedecer a un modelo que refleje los objetivos institucionales y las necesidades reales de los usuarios de la información. Este modelo de información, derivado de un análisis y diseño especializado, es la columna vertebral del SIG y debe ser el producto de un proceso de planificación de acuerdo a las necesidades reales de los trabajos de diseño de la encuesta, pruebas en terreno, ejecución de la encuesta y, finalmente, la diseminación de los resultados.

1. La importancia de la planificación del proyecto

La implementación de un SIG a nivel nacional es una actividad de alto costo en función del tiempo y dinero. De esta forma, la puesta en marcha de un SIG de esta envergadura puede tomar varios años y cientos de miles de dólares. Una planificación cuidadosa antes y durante el proyecto puede ayudar a evitar errores costosos y difíciles de resolver. Además, la planificación adecuada provee una garantía que se alcanzarán los objetivos dentro del itinerario y presupuesto.

Existe una tentación, cuando la tecnología SIG se hace disponible, improvisando las soluciones a través de su uso; esto es iniciar el proceso sin considerar donde nos llevará el proyecto. El gran peligro de realizar decisiones bajo la emoción derivada de herramientas vistosas como es la cartografía temática, es que deberán ser revertidas más tarde o el costo de su implementación es muy elevado. Como consecuencia de este hecho es posible que el proyecto SIG deba ser abandonado. Para evitar estas experiencias desalentadoras, este documento desarrolla una metodología de planificación que considera en forma objetiva el establecimiento de los objetivos, definición de las actividades a considerar y la estimación de los costos para el proyecto.

2. Consideraciones claves en la implementación del SIG

Durante el proceso de implementación del SIG existen tres aspectos claves que deben ser considerados y que merecen especial atención. Estos aspectos hacen referencia a fijar los objetivos y estimar los costos, los requerimientos funcionales y la creación de un modelo o prototipo.

2.1 Fijar los objetivos y estimar los costos

Cada etapa del proyecto SIG requiere del establecimiento de objetivos claros para la siguiente etapa y establecer los costos para alcanzar estos objetivos. Si los fondos necesarios no están disponibles, es mejor detener el proceso y ver como el proyecto falla. El proyecto puede siempre recomenzar cuando los fondos estén disponibles.

2.2 Los requerimientos funcionales

Un estudio de los requerimientos funcionales es probablemente el paso más importante durante el proceso de planificación del SIG. Durante este paso se visualizarán precisamente los requerimientos de información, como será utilizada y que productos se lograrán con ella. Dentro de un esquema institucional, este proceso es la producción de un mapa de flujos de información al interior, alrededor y fuera de cada oficina.

2.3 La creación de un modelo o prototipo

Una vez que se ingresa a la etapa de implementación existe una fuerte tendencia de alcanzar una implementación completa. Este procedimiento es altamente riesgoso pues no considera los ajustes que surgen al desarrollar un modelo prototipo. Los prototipos constituyen un paso crítico pues permiten probar y calibrar el sistema para ver si logra las expectativas y objetivos. Realizar ajustes al prototipo es lejos más sencillo que realizarlos posteriormente a la implementación total. El prototipo permite, además, a los usuarios familiarizarse con el nuevo sistema y estimar cuanto tiempo de conversión de datos y capacitación se requerirá durante las etapas de explotación. Finalmente, un prototipo exitoso puede avalar las el soporte y financiamiento de las etapas finales del proceso de planificación.

El tema del georeferenciamiento es un aspecto que ha sido el centro de discusión de aquellos que se han enfrentado a la disyuntiva de implementar un SIG. Si consideramos la naciente experiencia que se posee en estos sistemas en la región, sumado gastos considerables que se deben realizar durante las primeras etapas, resulta muy atractivo considerar la implementación de un SIG como el cumplimiento de una serie de etapas graduales que conducirán a la obtención final de este. Los costos de producir una cartografía georeferenciada son bastante más que los alcanzados cuando no se considera esta opción. Sin embargo, la cartografía digital no georeferenciada debe ser entendida como un instrumento esencialmente temático, cuyas posibilidades de análisis espacial se limitan,

generalmente, a la producción de mapas temáticos comparativos. Es importante considerar que la producción de una cartografía exenta de coordenadas no siempre resulta viable de georeferenciar y, en la mayoría de los casos, el camino para obtener una cartografía al nivel de los requerimientos de un SIG es simplemente la redigitalización sistematizada de la cartografía.

3. La selección de sistemas

Al seleccionar los programas y equipamiento, los usuarios se enfrentan a un número considerable de alternativas y combinaciones. Para un precio dado, no se puede esperar que un programa realice todas las tareas posibles. Una selección cuidadosa y bien pensada facilitará la elección de un sistema que alcanzará los objetivos del proyecto.

El proceso de selección debe tomar en consideración los siguientes aspectos:

Velocidad: La velocidad en que los sistemas pueden responder las consultas y lograr soluciones.

Las aplicaciones estadísticas, como la elaboración de la cartografía base para los trabajos de campo o la elaboración de estudios de pobreza a nivel nacional, requieren de la consolidación y unión de fuentes de información de múltiples áreas temáticas. Estos procesos manipulan, en la mayoría de los casos, millones de registros para lo cual se requiere de una velocidad de procesamiento que cumpla con las demandas de manipulación de información.

Capacidades funcionales: Las capacidades analíticas del sistema y su flexibilidad para abordar una gama amplia de procedimientos espaciales y estadísticos.

Dentro de las aplicaciones que son posibles de elaborar dentro de un proyecto SIG para apoyar los trabajos de la Encuesta de Hogares, se pueden citar aplicaciones generales como el desarrollo de cartografía temática, elaboración de cartas para los procesos de actualización cartográfica previo a la Encuesta y cartas de apoyo para los trabajos de campo. Sin embargo, el SIG provee las herramientas para desarrollar aplicaciones más complejas a través del análisis de redes. Estas aplicaciones se refieren, por ejemplo, a la automatización de la segmentación y la selección de muestras.

Tamaño de la base de datos: La habilidad de manipular grandes cantidades de información espacial y estadística.

Los grandes volúmenes de información que son administrados en las Oficinas de Estadísticas, principalmente censos nacionales y encuestas, requieren de amplios medios de almacenamiento si consideramos que estas actividades se realizan en forma periódica y, para efectos comparativos, se mantienen recopilaciones de años anteriores.

Administrar estos volúmenes de información es clave para seleccionar el SIG adecuado. Al mismo tiempo, la velocidad de respuesta puede ser menos importante si consideramos que una determinada información puede ser requerida sólo una vez al mes o, quizás, sólo una vez al año. Aún más, las capacidades funcionales son útiles, pero la mayoría de los procesos (tales como mantención y planificación) requerirán un rango limitado de capacidades analíticas.

Capacitación: El tiempo requerido para habilitar a los usuarios en el uso del sistema y la utilización de la base de datos en forma regular.

El proceso de capacitación debe ser abordado como la generación institucional de la capacidad de implementar y explotar el SIG. Este objetivo se logra introduciendo a aquellos que serán los usuarios finales del SIG, en las distintas fases de desarrollo del instrumento. Experiencias similares en países de la región han demostrado que la experiencia de implementar el SIG facilita de gran manera el proceso de implantación del SIG en las estructuras organizacionales de la institución.

Los factores expresados forman parte de algunos de escenarios que deben considerarse al seleccionar un SIG. Existen otros elementos que son más bien dependientes de las estructuras informáticas y son responsabilidad de los encargados de tales áreas.

Frecuentemente, los usuarios creen que pueden encontrar el SIG perfecto o el mejor. El mejor SIG es aquel que realiza el trabajo según los costos determinados y en los tiempos establecidos.

4. Factores adicionales de la planificación

El ciclo de planificación del proyecto SIG configuran un esquema común, pero los pasos a seguir en cada etapa del proceso variarán considerablemente de una organización a otra. Los factores adicionales que se deben considerar en proyectos SIG de gran envergadura son :

Seguridad

La seguridad de la información es uno de las mayores preocupaciones en un proyecto SIG de envergadura. Sin embargo, la seguridad es más que proteger los datos contra un uso inadecuado o malicioso. La seguridad significa, también, proteger los datos contra las caídas del sistema y catástrofes mayores. Como resultante, la seguridad debe ser considerada en varios niveles y debe anticipar la mayoría de los problemas potenciales. Esta práctica presenta desafíos complicados especialmente cuando la información debe ser accesada en forma pública (que es el caso de accesibilidad de los archivos de las Encuestas de Hogares).

Documentación

La mayoría de los proyectos SIG sobrevivirán a la gente que los implementó. Si todos los pasos involucrados en la codificación y estructuración de los archivos no están bien documentados, la información almacenada en ellos se perderá a medida que el personal se jubile o escala a posiciones más altas en la jerarquía. La documentación debe iniciarse en el mismo principio del proyecto SIG y continuar a través de su vida útil. Una práctica adecuada es asignar en forma permanente (una asignación estable de funciones a un funcionario de planta) para asegurarse que la información necesaria esta salvada y documentada.

Integridad y precisión de los datos

Cuando se descubren errores en la base de datos SIG, debe implementarse un procedimiento adecuado para su corrección (y para documentar la corrección). Aún cuando son muchos los usuarios que accederán a la información de la base de datos, no todos ellos deben tener la posibilidad de realizar cambios. Mantener la integridad de las diferentes "capas" de información en una base de datos SIG puede resultar en una tarea desafiante. La ubicación de las viviendas en las zonas rurales y la infinidad de caminos principales y secundarios es una información que debe estar disponible para la gran mayoría de los usuarios, sin embargo, el proceso de actualización debe ser realizado por funcionarios especializados en actualización cartográfica. Posiblemente, la mantención y actualización de las componentes cartográficas, propiamente tal, debería estar a cargo de instituciones especializadas que, a través de convenios multilaterales, proporcionen el servicio requerido y puedan usufructuar de los beneficios del SIG.

Sincronización en el uso

Los conjuntos de datos de un SIG utilizados a nivel de instituciones gubernamentales como las Oficinas de Estadísticas, atraerán un gran número de usuarios. Algunas partes de la base de datos puede ser requerida en forma simultánea por varios usuarios al mismo tiempo que se estén desarrollando procedimientos de mantención y actualización. Asegurándose que todos los usuarios accedan a la misma información en todo momento en una tarea complicada y desafiante para el diseño de un SIG. Si se llega a un nivel donde el uso de la base de datos sea una actividad incontrolable, los riesgos de mal uso de los datos o la duplicación de los esfuerzos de procesamiento puede resultar en un retroceso en relación a los beneficios que proporciona la herramienta.

Responsabilidad de mantención y actualización

Algunas bases de datos dentro del SIG nunca serán consideradas completas. Las ciudades, tanto a nivel urbano como rural, crecen día a día y deben ser actualizadas en forma periódica para que reflejen estos cambios. Sin embargo, estos cambios suceden en momentos indeterminados y a velocidades variables. El desarrollo de procedimientos para

registrar estos cambios debe ser imperativo. Más aún, una versión original de los bases debe mantenerse para efectos comparativos.

En los proyectos SIG de gran escala, la responsabilidad de mantener la base de datos actualizada es probablemente la tarea de un equipo estable de funcionarios.

Minimizar la redundancia

Al igual que en el párrafo anterior, dentro de los proyectos SIG a gran escala, cada "byte" de información es importante. Si en un registro de personas u hogares dentro de una base de datos a nivel nacional se sobre dimensionan los tamaños de los campos de información, el exceso de espacio utilizado puede causar serios problemas de almacenamiento. La existencia de información inutilizada no sólo causará gastos adicionales de dinero, sino que, también, degradará el tiempo de procesamiento de los archivos. Esto es el motivo por el cual en los SIG a largo plazo (como es el caso de un SIG para las Encuestas de Hogares) se debe asignar mucha atención al diseño de las estructuras de información en función de la compresión y duplicación de los datos.

Independencia de los datos

Un proyecto SIG generalmente sobrevivirá a los programas y equipos computacionales que son utilizados para implementarlo. El equipamiento computacional posee una vida útil promedio de 2 a 5 años y los programas son "upgraded" varias veces dentro de un mismo año. Si la base de datos del SIG es dependiente de una plataforma en particular o de un solo sistema computacional, se deberá actualizarla cada vez que los equipos o programas se actualicen. De esta forma, es preciso desarrollar un diseño de base de datos que sea lo más independientemente posible de los equipos y programas. A través de un diseño cuidadoso y realista, los datos pueden ser transferidos en archivos ASCII o en algún tipo de formato de intercambio de sistema a sistema.

Los diseñadores del SIG deben considerar en forma avanzada los posibles cambios y actualizaciones que deberán ser asumidos en el futuro. Es prácticamente imposible predecir lo que sucederá en el mundo de los computadores y en la tecnología de la información. Para minimizar los posibles problemas, se debe asignar una gran atención para lograr que la base de datos SIG sea lo más independiente de las herramientas utilizadas para desarrollarlas.

III. LA CARTOGRAFÍA DIGITAL Y LOS SIG PARA LAS ENCUESTAS DE HOGARES

Fundamentalmente, la utilización de un SIG para apoyar la planificación y ejecución de las Encuestas de Hogares obedece a la necesidad de implementar un modelo de información que permita un relacionamiento dinámico de la información proveniente de la encuesta con la base de datos geográfica. Siendo este su objetivo terminal, además el SIG posee una gran eficiencia en la elaboración de material para apoyar el proceso de planificación y ejecución de los trabajos de campo.

1. Productos para la planificación y la ejecución de trabajos de campo

El SIG es una herramienta que puede generar una número considerable de insumos que facilitan enormemente las etapas de planificación (en gabinete) de la encuesta así como los trabajos de campo (encuesta, supervisión y evaluación).

La generación de cartografía a escala para zonas urbanas y rurales permite proveer a los planificadores de la encuesta con material de apoyo para visualizar el área de cobertura, determinar los grupos de trabajo, encontrar las mejores rutas, etc. (ver Mapas Facsímil 1-5). Estas carta se pueden imprimir en diferentes tamaños y con toda la información geográfica y descriptiva que estos técnicos requieran.

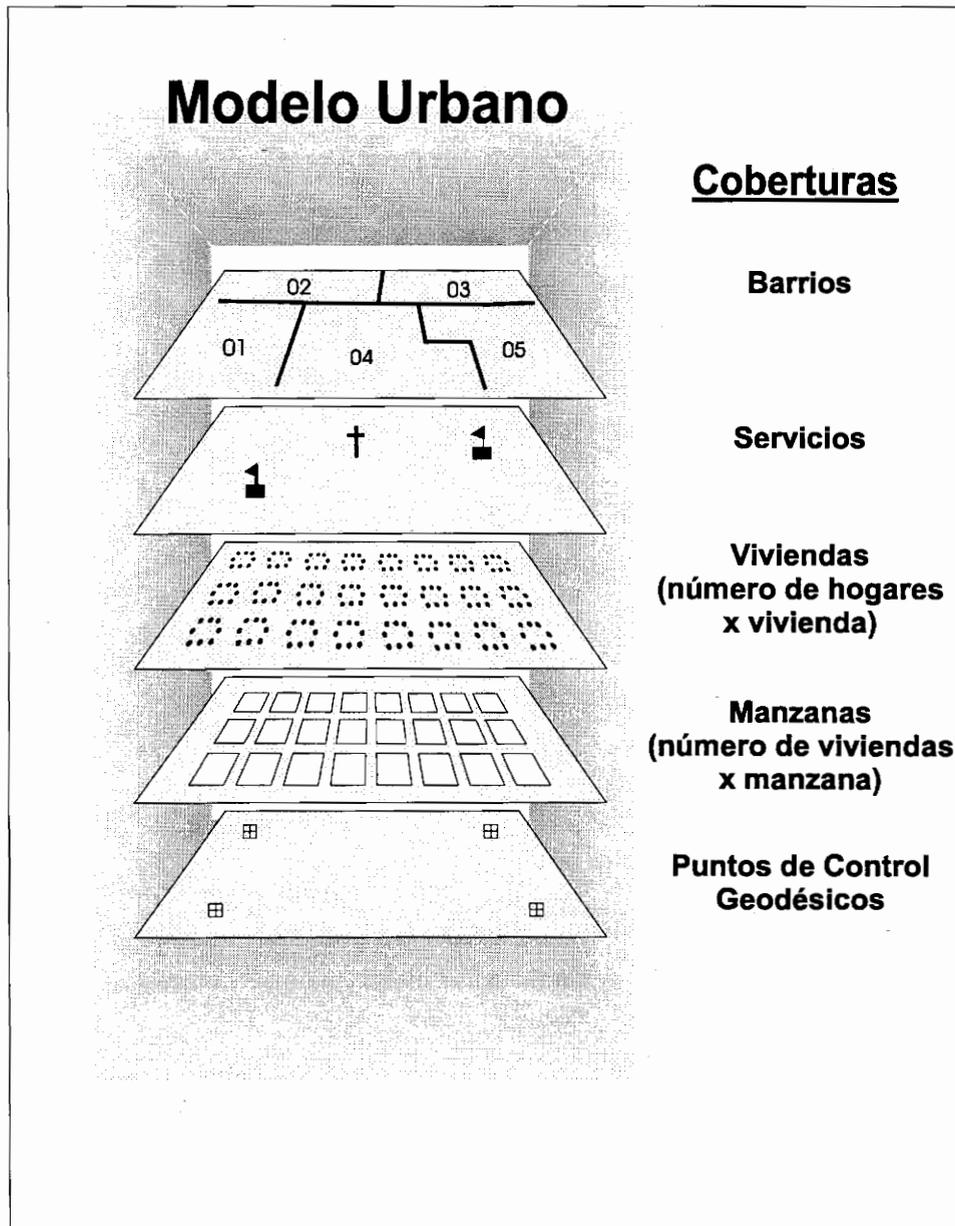
En los trabajos de campo, el SIG proporciona material de apoyo para los supervisores y censistas. Las labores de supervisión precisan contar con material que les permita visualizar la distribución de los segmentos, controlar las zonas ya censadas y mantener un registro de los controles efectuados. Por su parte, los censistas pueden contar con cartas específicas del segmento asignado para tener una visión panorámica de su área de trabajo.

2. Segmentación espacial y generación de muestras

El tratamiento de áreas urbanas y rurales en los procesos de segmentación es conceptualmente distinto. Dentro de las áreas urbanas la entidad de manzana es fundamental para el ordenamiento de las viviendas, sin embargo, a nivel rural el tratamiento de las viviendas se realiza en función de conglomerados y su relación con el entorno geográfico. Por este motivo se debe tener en consideración que la manipulación de los procedimientos de muestreo requieren de modelos diferentes y de elementos específicos para facilitar las labores de segmentación.

3. El modelo urbano

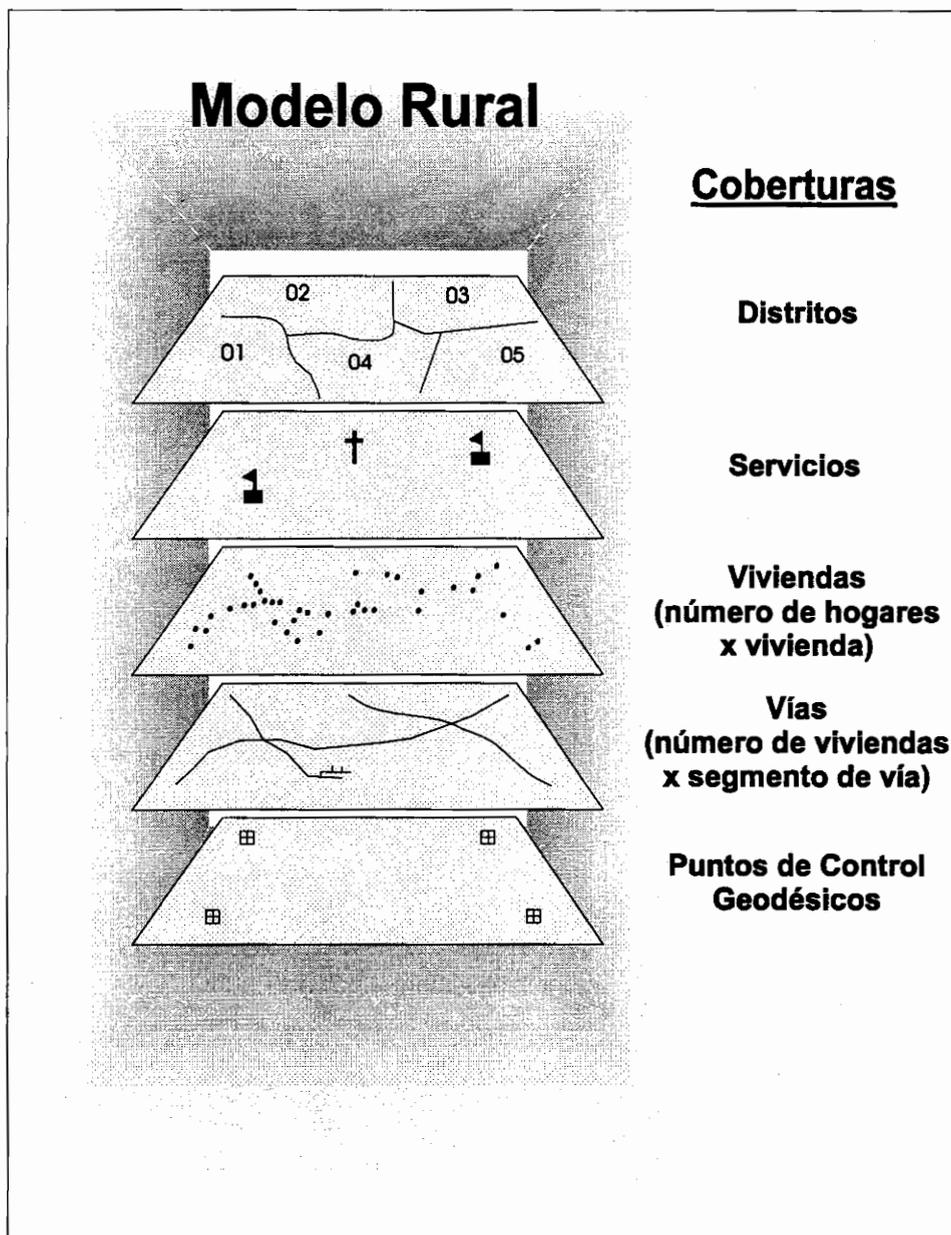
El modelo de información urbano utilizado en las aplicaciones GIS para las encuestas de hogares está basado en el concepto de relacionamiento de las manzanas. Para cada manzana de la zona geográfica se crea una tabla descriptiva con la información de los hogares existentes. A partir de esta información, se generan los totales requeridos para el procedimiento de segmentación, como por ejemplo, el total de hogares de la manzana.



El SIG proporciona, para este modelo, la relación denominada “topológica” que permite manipular el concepto de “manzanas vecinas”. A través de este concepto, el procedimiento de segmentación agrupa las manzanas hasta alcanzar el mínimo de hogares requeridos para la muestra. Por tratarse de un herramienta automatizada, este proceso se puede ejecutar repetidas veces hasta alcanzar una segmentación adecuada a la aplicación que se esté implementando. Más aún, el usuario puede especificar la manzana de origen de proceso haciendo uso del código geográfico de la manzana. La relación de vecindad entre las manzanas es generada internamente por el sistema de información geográfica. Vale destacar en este momento la importancia de contar con un SIG profesional que entregue las herramientas necesarias para generar las aplicaciones que satisfagan los requerimientos de cada aplicación.

4. El modelo rural

Al no existir un agrupamiento ordenado de viviendas como en las zonas urbanas (manzanas), en las zonas rurales es preciso definir con criterios de distancia y accesibilidad las agrupaciones de viviendas para optimizar el trabajo de los censistas.



Dado el carácter aleatorio en lo que respecta a la ubicación, en las zonas rurales se usa el concepto de “centros de gravedad” para establecer los posibles conglomerados. A partir de estos centros de gravedad y el número mínimo de viviendas que debe tener cada segmento, se implementa un procedimiento que denominaremos “espiral” que, a través de las distancias mínimas, entrega una alternativa de segmentación de la zona.

El modelo de información rural utilizado en la automatización de la generación de espacios muestrales, como se mencionó anteriormente, está basado en la formación de conglomerados de viviendas.

El algoritmo que se debe implementar para la segmentación rural es bastante complicado y requiere de asistencia especializada. La técnica utilizada se basa en el conocimiento de la operatoria espacial y sus herramientas. Su fundamento se centra básicamente en funciones recursivas y relaciones topológicas que determinan que viviendas son los candidatos más atractivos.

El procedimiento de segmentación rural puede ser enriquecido a través de la incorporación de coberturas con información de referencia. En algunas zonas muy aisladas surgen dificultades para que el censista pueda ubicar con facilidad las viviendas. Esta labor se facilita con la creación de coberturas con lugares de referencia como, por ejemplo, cruces de caminos, estancias, almacenes, etc.

IV. LOS SISTEMAS DE INFORMACIÓN GEOGRÁFICOS Y SU DESARROLLO

1. Introducción

El origen de los sistemas de información geográficos SIG se remonta a casi 30 años atrás, cuando un grupo de geógrafos desarrolló un sistema para almacenar y organizar información espacial en un computador.

Desarrollos en paralelo han permitido introducir avances notables a la tecnología original. Desde cartografía de alta calidad hasta planificación de usos de suelo, administración de recursos naturales, demanda educativa, distribución de recursos para la salud, fiscalización de impuestos e investigación demográfica, los sistemas de información geográficos se han convertido en uno de los sistemas más versátiles y adaptables a las más variadas aplicaciones.

¿En que se basa este creciente interés en los SIG?

La tecnología SIG proporciona un medio para integrar la información de un modo que permite focalizar los problemas y su relación con factores ambientales, demográficos, por nombrar sólo algunos. Los SIG nos permiten organizar el uso del computador, la información y los sistemas en torno al usuario, estableciendo la relación espacial entre los datos y por consiguiente apoyando en forma más inteligente el proceso de toma de decisiones. El manejo y análisis de los datos espaciales permite:

- *Producir datos derivados fáciles de evaluar*
- *Generar nuevos datos*
- *Contestar preguntas específicas sobre problemas complejos*
- *Analizar alternativas para un problema determinado*

La propuesta para el diseño e implementación de un sistema de información geográfico para apoyar la gestión administrativa y operativa de las encuestas de hogares requiere necesariamente establecer el marco conceptual dentro del cual se fundamenta la metodología que presenta este documento. Por este motivo realizaremos un breve recorrido por los aspectos más relevantes de la tecnología SIG, entregando los conceptos estrictamente necesarios cuando la situación así lo requiera.

2. Los Sistemas de Información Geográficos

Hoy, el número y variedad de aplicaciones SIG es impresionante. La cantidad de datos geográficos generados y acumulados supera toda expectativa. Los gobiernos locales utilizan los SIG para apoyar las labores de planificación, mantención del plan regulador, registros de loteo, seguridad pública, medio ambiente, etc. La tecnología SIG no es simplemente un medio para la creación de mapas, aunque dentro de sus aplicaciones está la producción de cartografía a diferentes escalas, proyecciones y tonalidades. Un SIG es una herramienta de análisis cuya ventaja principal radica en la capacidad para identificar las relaciones entre entidades cartográficas y sus respectivas descripciones.

Beneficios de la implementación de un SIG

Pasar de un ambiente de procesamiento manual hacia un ambiente informatizado no es solamente la adquisición y el uso de un nuevo "software"; debe ser manejado como un cambio administrativo y gerencial. El éxito en la implementación de un sistema de información geográfica se basa en asimilar el cambio como un proceso de transformación integral, donde participan todas las componentes de la institución. La implementación de un SIG institucional, como proceso informático, no es la mecanización de los procedimientos administrativos antiguos sino la introducción de nuevas herramientas tecnológicas que nos trasladan a una nueva dimensión en organización y análisis de la información. Dentro de los beneficios inmediatos podemos citar:

- ***Incremento en la calidad de la información;*** el modelo de información espacial y su metodología imponen altas metas de control de calidad sobre los datos. Estos controles de calidad permiten al usuario generar información normalizada, consistente, validada y oportuna para apoyar los procesos involucrados.
- ***Incremento en el acceso a la información;*** el acceso a la información dentro de un SIG se realiza a través de procedimientos tabulares y espaciales. Estos últimos agilizan

notablemente la accesibilidad a los datos pues relacionan las entidades geográficas con los atributos tabulares.

- **Un eficiente flujo e intercambio de la información;** la información se almacena en un formato común lo cual facilita la comunicación y el intercambio de los datos. Así mismo, se agilizan los procedimientos de captura de datos en los trabajos de campo, permitiendo bases de datos normalizadas y centralizadas.

- **Incremento de la productividad;** se reducen notablemente los procesos manuales habilitando la ejecución de procedimientos complejos con la consiguiente reducción en el tiempo de respuesta y el mejoramiento de la calidad del servicio.

- **Reducción de costos a largo plazo;** la optimización en el almacenamiento de la información implica una mayor eficiencia en las labores diarias de procesamiento de la información. El incremento en la exactitud de la información implica un apoyo consistente y confiable al proceso de toma de decisiones.

- **Personal capacitado en la tecnología;** la modernización de los recursos humanos dentro de una institución, especialmente cuando se trata de tecnología de punta, es una excelente inversión en el mediano y largo plazo. Al igual que el uso de la plataforma Windows a incrementado notablemente los niveles de productividad, el uso de un SIG genera un ambiente de satisfacción generalizado al conocer y comprender herramientas que optimizan la producción y procesamiento de la información.

Los SIG permiten la realización de análisis espacial sobre los datos.

La pregunta “Cuál es la medida de población dedicada al desarrollo de proyectos SIG” es una pregunta que no requiere de procesamiento espacial, es decir, la elaboración de su respuesta no incluye el procesamiento de información sobre latitud y longitud. Por otro lado, la pregunta “Cuántas personas viven a 100 metros de la riva de un río” incorpora la componente espacial y solamente puede ser resultado con la utilización de los datos relacionados con su localización. Un SIG proporciona las funciones necesarias para producir los resultados a este tipo de preguntas. Los objetivos de un SIG son:

- *Mejorar los procesos de administración de los datos.*
- *Disminuir la duplicidad y redundancia de datos y procedimientos.*
- *Incrementar el acceso a la información geográfica necesaria para tomar decisiones en forma objetiva.*
- *Introducir los conceptos geográficos dentro de los procesos de análisis, evaluación y presentación de la información.*

El concepto de base de datos es indispensable para la explicación de un SIG. Constituye la diferencia primordial entre este tipo de tecnología y la tecnología de dibujos y cartografía asistida por computador, cuya función es la de producir mapas de alta calidad. Cada elemento geográfico en la base de datos está directamente relacionados con sus atributos. Esto significa que a través de un atributo se puede acceder a la entidad geográfica y viceversa.

En resumen, un SIG es una herramienta para visualizar, modelar, analizar y extraer información de una base de datos con referencias geográficas.

¿Qué rol juega la componente cartográfica en la producción de resultados de un SIG?

La forma de presentar la información es esencial para la comprensión del problema que se está comunicando. Desde este punto de vista, una de las grandes ventajas de los Sistemas de Información Geográficos radica en su capacidad para producir cartografía que muestra dentro de un contexto espacial la composición de una región.

Una de las capacidades más versátiles de los SIG es su potencial en la producción de despliegues y salidas. Algunos de estos productos son:

- La producción de reportes tabulares o textuales es una función genérica de todo sistema de información. En un SIG, éstos pueden ser producidos a partir de unas condiciones geográficas determinadas.
- El despliegue y producción de mapas en medios de papel es sin duda la función de salida más importante para la distinción de un SIG de los demás sistemas de información. Las relaciones espaciales se entienden explícitamente por medio de mapas.
- Los cuadros y gráficas pueden utilizar las relaciones espaciales para ilustrar características no necesariamente de índole geográfica. Estos productos igualmente pueden ser producidos por un SIG.

3. ¿Qué es un SIG ?

Los sistemas en general permiten visualizar la información de diferentes maneras de acuerdo con los objetivos de evaluación. Un SIG permite organizar los datos de acuerdo con su localización, tamaño, dimensiones, y procesarlos a partir de coordenadas, para producir información relevante. Sus características geográficas y cualidades espaciales son las variables que permiten su categorización y organización.

Los SIG proporcionan una manera de representación instantánea y modificable de nuestro mundo.

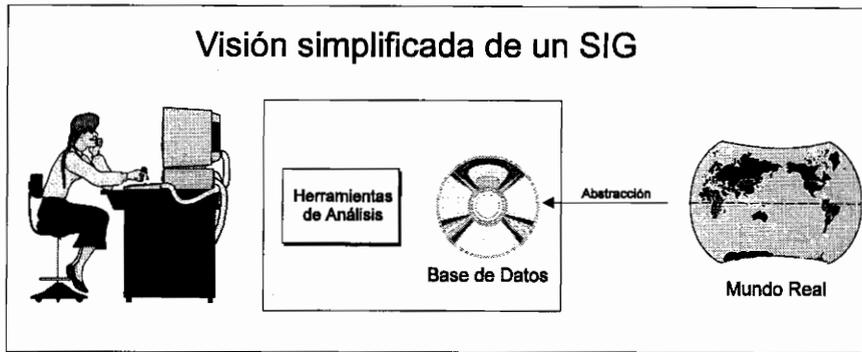
El SIG, como producto de software, posee varias componentes que permiten la conexión con otros dispositivos dentro de su ambiente. Estructuralmente posee un sistema de administración de base de datos para almacenar y procesar datos conectado con un sistema administrador gráfico para visualizar los resultados. Estos dos sistemas se relacionan con el operador a través de una interface o un interprete del lenguaje de comandos.

¿Cómo se almacena la información dentro de un SIG?

Un SIG asocia información espacial con información descriptiva sobre un objeto específico de un mapa y lo archiva internamente en forma de atributos. Por ejemplo, la infraestructura vial de una ciudad puede ser representada por un conjunto de líneas, en cuyo caso su presentación gráfica no proporciona mayor información. Los datos adicionales tales como la clase de vía, ancho de la calzada, tipo de pavimento, número de carriles, que comprenden la parte descriptiva de la base de datos y asocian información sobre los atributos del elemento gráfico. Con el SIG es posible visualizar los dos tipos de información descritos de manera integrada y reproducir la información geográfica de acuerdo a los atributos tabulares.

¿Cuáles son los dispositivos utilizados por un SIG?

Los principales dispositivos utilizados en un sistema de información geográfico son los discos para almacenar la información (alfanumérica, gráfica o imágenes), mesas digitalizadoras y scanners para ingresar los datos y ploteadoras e impresoras para presentar los resultados. En forma similar, al igual que para cualquier sistema computacional, es importante poseer una conexión a una red de comunicaciones para facilitar el intercambio de datos con otras personas e instituciones que trabajen o proporcionen datos geográficos o espaciales.



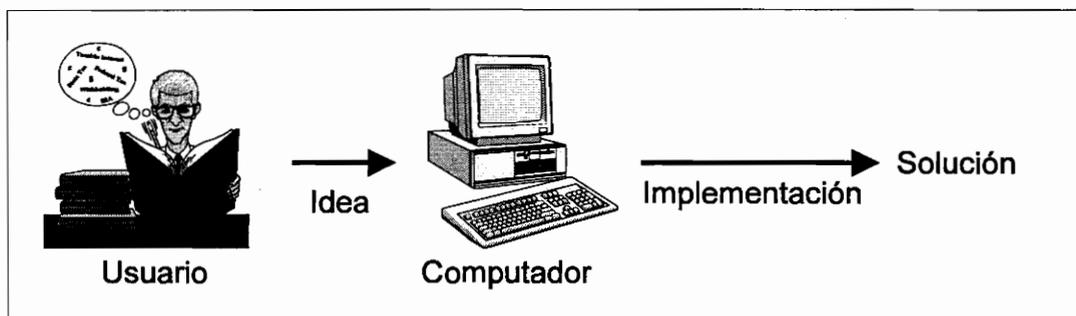
El sistema de información geográfico es un conjunto de computadores, sistemas, programas, datos geográficos y usuarios dedicados a la captura, almacenamiento, actualización, manipulación, análisis y despliegue de todo tipo de información referenciada geográficamente. Esta caracterización es bastante general, por lo cual es preciso identificar los elementos que intervienen en un SIG, y examinar la relación entre ellos.

a) El computador

El computador, como unidad de procesamiento de información, es el instrumento que permite el ingreso y almacenamiento de grandes volúmenes de datos. Estos datos, procesados a alta velocidad, permiten al usuario desarrollar modelos de información para una amplia variedad de aplicaciones. El acelerado desarrollo tecnológico que están experimentando los computadores demanda un minucioso estudio de los objetivos del proyecto para establecer que configuración es la más adecuada y más conveniente económicamente para implementar el SIG. Este informe aborda el tema de las plataformas más adelante para plantear una serie de sugerencias útiles en el momento de decidir cada una de las componentes.

b) El usuario

El centro de todo sistema de información es el usuario, quien, a partir de un problema determinado, desarrolla las herramientas que permitirán generar alternativas de solución. En el momento de diseñar un sistema de información, los analistas deben considerar que los computadores y programas no son la solución a los problemas que se intentan resolver sino el medio a través del cual el usuario implementa la solución elaborada.



c) La información

Un papel fundamental en los sistemas es la información. Para definir la información, se debe pensar en un determinado propósito o fin a cumplir. La información debe estar organizada bajo un modelo de información basado en las reales necesidades de los usuarios. De esta forma, se construyen los procedimientos que, alimentados por los datos, permiten obtener información válida, consistente, oportuna y de buena calidad. El propósito fundamental de la información es, a través de función determinada, generar una acción. Los datos son manipulados y procesados para producir información. Esto implica que los datos deben ser organizados en forma sistemática antes de que un computador pueda entrar a ser parte del sistema de información. Las actividades de diseño de la base de datos y posteriormente de administración con las responsables de convertir en realidad la solución sistematizada.

d) Los sistemas y procedimientos

Un SIG también comprende una serie de programas de aplicación que procesan y manipulan la base de datos. Esta base es solamente una abstracción del mundo real. Por medio del procesamiento de los datos y su integración de diferentes maneras se produce información derivada y es posible crear modelos sofisticados del comportamiento, por ejemplo, de una ciudad o región. Esta información se utiliza, principalmente, en la toma de decisiones e implementación de políticas.

e) Las herramientas de análisis

El eje operativo central de un SIG es el uso de un software especializado con capacidad de análisis espacial. Este software, además de administrar la base de datos y entregar elementos de análisis espacial, debe proveer los insumos básicos para construir los procedimientos especializados que requieren los usuarios.

Los procedimientos especializados, en un SIG, deben ser concebidos, diseñados, contruidos e implementados según las reales necesidades de los usuarios. Esto implica un proceso de desarrollo que en la práctica ha demostrado ser largo, difícil y poco estructurado.

La estructura de desarrollo de estas herramientas está relacionada con la diferenciación del diseño en aspectos lógicos y físicos. Se entiende por aspectos lógico aquellos orientados a establecer qué actividades debe realizar el procedimiento; sin preocuparse de detalles de implementación tales como quién lo ejecutará, con qué plataforma, en qué lugar, etc. Por el contrario, los aspectos físicos se refieren a un esquema detallado y específico, incluyendo los medios tecnológicos y de otro tipo que se utilizarán, en la realización del procedimiento.

Resulta evidente que, en primer lugar, los usuarios deben concentrarse en los aspectos lógicos, principalmente para estudiar las alternativas, para después llevar a un

detalle físico la alternativa seleccionada. En esta fase el objetivo principal es generar varias alternativas funcionalmente diferentes para la herramienta en diseño y hacer una evaluación de ellas.

4. Preguntas que un SIG puede contestar

Los SIG han sido descritos como sistemas que poseen la habilidad de ejecutar operaciones espaciales a través de la unión de diferentes conjuntos de datos utilizando la ubicación como llave común. Sin embargo, los SIG pueden ser definidos especificando el tipo de preguntas que pueden contestar. Para toda aplicación SIG existen cinco preguntas genéricas que un SIG sofisticado puede contestar:

¿Qué está ubicado en...? - La primera pregunta que un SIG puede contestar es sobre lo que existe en una ubicación determinada. Sea a través de una especificación gráfica o una ubicación por una dirección, un manzana, un distrito, una región, el SIG puede instantáneamente localizar un objeto o entidad. Esta facilidad de ubicación está dada por la característica georeferenciada que poseen estos sistemas.

¿Dónde están ...? - La segunda pregunta trata lo opuesto de la primera. Se pretende encontrar las ubicaciones donde ciertas condiciones existen. Por ejemplo, en que comunas hay más de 100.000 habitantes o que viviendas están a menos de 100 metros de un río. Esta capacidad está dada por la relación entidad geográfica - atributo que poseen los SIG. Cada elemento geográfico está vinculado a una tabla de atributos que define su posición en relación a una sistema coordenado y sus características o cualidades. Es esta facilidad la que permite filtrar a nivel tabular los elementos que cumplen ciertas condiciones.

¿Qué ha cambiado desde...? - Esta pregunta permite determinar que elementos dentro de un área han cambiado en un periodo de tiempo. El modelo de información SIG impone el la mantención de un registro histórico de la base de datos. A través de este registro se pueden comparar las mismas áreas geográficas en diferentes momentos de tiempo.

¿Qué patrones espaciales existen? - Esta pregunta pretende resolver preguntas referentes a si la ubicación de ciertos fenómenos incide en sus características. Como parte del análisis espacial que permite la implementación de un SIG, existen procedimientos para visualizar características iguales o similares adyacentes. Estas áreas pueden ser localizadas con facilidad al aplicar la capacidad inherente de cartografía temática.

¿Qué pasa si...? - Las preguntas del tipo "Que pasa si..." pretenden averiguar que pasaría si ciertas condiciones ocurren en una determinada ubicación. La relación de atributos y entidades geográficas permite con gran facilidad ingresar valores experimentales para "simular" posibles situaciones según ciertos criterios.

5. La necesidad de una metodología para la implementación del SIG

El desarrollo de los SIG se lleva a cabo en la práctica, en la mayoría de los casos, como un trabajo artesanal. Así cada analista utiliza sus propios métodos de trabajo y documentación, existiendo una mínima estandarización al respecto. Los métodos son el resultado de la experiencia práctica, no teniendo ninguna base teórica que los respalde.

La ausencia de una metodología ha traído como consecuencia una serie de problemas que se resumen a continuación:

- Al no existir una metodología que garantice una adecuada y formal definición de la función que debe cumplir el SIG dentro de la Institución, se llega, en la mayoría de los casos, a sistemas que no satisfacen las necesidades de los usuarios y que son rechazados por los mismos, con el consiguiente derroche de recursos.
- Como consecuencia del punto anterior, se produce un proceso de aproximaciones sucesivas al diseño, con múltiples iteraciones de rediseño y reprogramación, hasta alcanzar el resultado deseado, gastándose más tiempo y recursos que los estrictamente necesarios.
- El proceso de rediseño que implica el punto anterior lleva a sistemas ineficientes en cuanto al uso de los recursos computacionales y humanos involucrados. Generalmente, las herramientas resultantes de este tipo de procesos poseen numerosos “parches” que no obedecen a un patrón común.
- La falta de una buena metodología implica también mala documentación del SIG, lo que origina cierta dependencia de individuos claves que conocen el sistema. Esta dificultad se extiende además a las futuras actualizaciones que se deben realizar, limitando a los analistas a escasa documentación.
- Al carecer de la estandarización del proceso de desarrollo que proporciona una metodología, se hace difícil una buena administración de los recursos humanos que intervienen en el desarrollo de un SIG. El éxito de una instalación SIG está avalado por una buena administración del personal que lo compone.

De estos problemas, que abundan en la práctica, se desprende la necesidad de poseer una buena metodología para el diseño e implementación del SIG así como de cada una de las aplicaciones que se van a desarrollar.

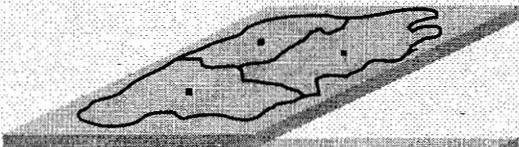
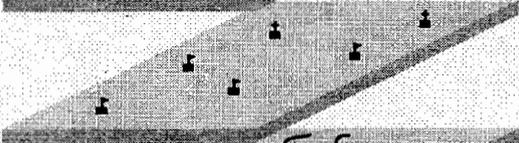
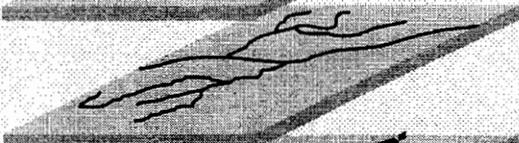
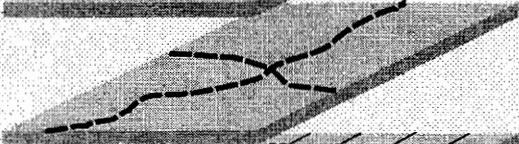
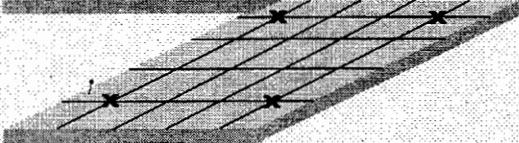
Una buena metodología no complica el proceso de desarrollo trabando en forma extrema los pasos a seguir y hacer que los encargados de la implementación usen una cantidad innecesaria de documentación. Existe la necesidad de reconocer que el diseño de un SIG es una labor creativa, que no puede acotarse dentro de límites estrechos.

6. El modelo de información SIG

Un enfoque utilizado para representar las relaciones complejas existentes en nuestro mundo es definiendo el límite del área de interés y dividiendo el conjunto total en subconjuntos o estratos temáticos. Cada estrato representa una clasificación de información semejante. Dentro del contexto natural, los estratos principales incluyen geología, vegetación, topografía e hidrología. Dentro del ambiente humano o cultural, los estratos son edificaciones, predios, vías, uso de suelo, zonificación y límites político-administrativos. Las actividades que ocurren dentro de este contexto se relacionan con: lugar de vivienda y trabajo de la población, vehículos del sistema vial, composición demográfica y densidad.

Existen tres tipos diferentes de objetos que pueden ser representados en el espacio bi-dimensional: puntos, líneas y polígonos. En un SIG sofisticado, estos objetos se catalogan como entidades y se almacenan en forma separada dentro de estructuras denominadas "coberturas". Una cobertura es una agrupación de entidades semejantes que han sido modeladas para representar el mundo real. Existen tres tipos diferentes de coberturas que corresponden a los tres tipos de representaciones espaciales: coberturas de puntos, líneas y polígonos. Cada cobertura posee la información de coordenadas que le permite posicionarse espacialmente y, además, la información alfanumérica relacionada con ese espacio.

Representación de Coberturas de Información en el Modelo SIG

<u>Estrato</u>	<u>Tipo de dato</u>
	Regiones, Distritos, División político- administrativa Comunas, etc.
	Escuelas, Iglesias, Servicios Hospitales, etc.
	Ríos, arroyos, lagos Hidrología
	Caminos, carreteras Infraestructura
	Uso de GPS Puntos de control

En la definición del modelo, el especialista en información debe considerar todos los estratos de información que serán necesarios para el procesamiento de los proyectos. Los estratos de información hacen referencia a las agrupaciones de datos que serán definidos en función de atributos similares. Dentro de cada estrato se definen los tipos de datos que conforman los límites del estrato. Por ejemplo, el estrato "Hidrología" contiene los tipos de datos "ríos", "arroyos" y "lagos".

El modelo presentado en este informe puede ser aplicada a la implementación de cualquier sistema de información geográfica. Sin embargo, los detalles específicos están orientados al apoyo de proyectos en el marco de las encuestas de hogares. Este modelo organiza los objetos dentro de coberturas o estratos temáticos. Cada cobertura representa un tipo de información en particular que debe cumplir con ciertas condiciones.

Estas condiciones reflejan la necesidad de separar la información en coberturas:

- Las componentes de la base de datos provienen de diferentes fuentes de información a diferentes escalas. Cada fuente de información debe estar reflejada y normalizada en una cobertura de tal forma que sea representada en forma espacial según sus propios atributos.
- Algunos objetos del mundo real cambian de un estrato a otro más rápidamente que otros (mayor volatilidad de la base de datos). De acuerdo con esta consideración puede ser apropiado separar aquellos estratos que tienen diferentes ciclos de actualización. Por ejemplo, los ríos y arroyos de un área determinada sufren cambios menores en relación a aquello que pueden ocurrir con la infraestructura vial de una ciudad.
- Ciertos objetos del mundo real se utilizan para la elaboración de proyectos de análisis y modelamiento de información. En este caso, la condición que determina la diferenciación de estratos está relacionada con su aplicación específica. Para un proyecto específico la creación de una cobertura de servicios (escuelas, hospitales, correos, etc.) puede satisfacer las necesidades del estudio, sin embargo, para otras aplicaciones puede ser necesario desarrollar coberturas separadas, como por ejemplo, una cobertura de escuelas.

Durante la construcción de las coberturas que almacenarán la información según cada uno de los temas, la escala de los datos y su representación gráfica debe tener suficiente detalle para cumplir con las expectativas del proceso de automatización. Esto significa que cada cobertura se debe digitalizar desde una fuente que permita la extracción en un adecuado nivel de detalle para que cada uno de los elementos que nos interesa estudiar integre la base de datos. Sin embargo, el nivel de detalle es directamente proporcional a la complejidad y al tamaño del sistema de información.

Un SIG requiere del manejo integrado de la información gráfica y descriptiva. Estos dos tipos de información cambian en forma simultánea y de manera independiente. Por ejemplo, el volumen de tráfico asociado a un segmento de vía puede cambiar sin alterar su componente gráfica y, viceversa, el segmento de vía puede ser rebuscado sin alterar su volumen de tráfico.

La administración de los datos consiste en la ejecución de los procedimientos de archivo, eliminación y extracción. Existen proyectos de una dimensión tal que requieren de un profesional de tiempo completo dedicado a las labores de administración del sistema; para proyectos pequeños esta función puede ser manejada de una manera más sencilla.

El desarrollo de modelos o representaciones permite mejorar nuestro conocimiento del mundo por medio de la clasificación, identificación y organización de los datos. Los procedimientos de despliegue y presentación gráfica de estos datos se convierten en información valiosa que posibilita la toma de decisiones y definición de políticas coherentes de manejo de los recursos.

7. Propuesta de diseño para la implementación del SIG

El presente documento propone la implementación de un sistema de información geográfico como un instrumento de apoyo al proceso de toma de decisiones, en este caso orientado a las encuestas de hogares. El usuario, como parte fundamental del SIG, interactúa con procedimientos, información y dispositivos computacionales con el objeto de desarrollar estrategias que faciliten la gestión de planificación y trabajos de campo.

El SIG, dentro de las Encuestas de Hogares, puede proporcionar apoyo en cuatro áreas principalmente:

- **planificación y desarrollo de las estrategias;** el SIG permite la producción de la cartografía base para planificar el trabajo de la encuesta. A medida que se van definiendo los segmentos muestrales, el SIG proporciona la posibilidad de crear coberturas con los segmentos y unirlos con el resto de la base de datos geográfica con el propósito de producir planos de trabajo. Estos planos permiten la visualización de la distribución de los segmentos y su ubicación.

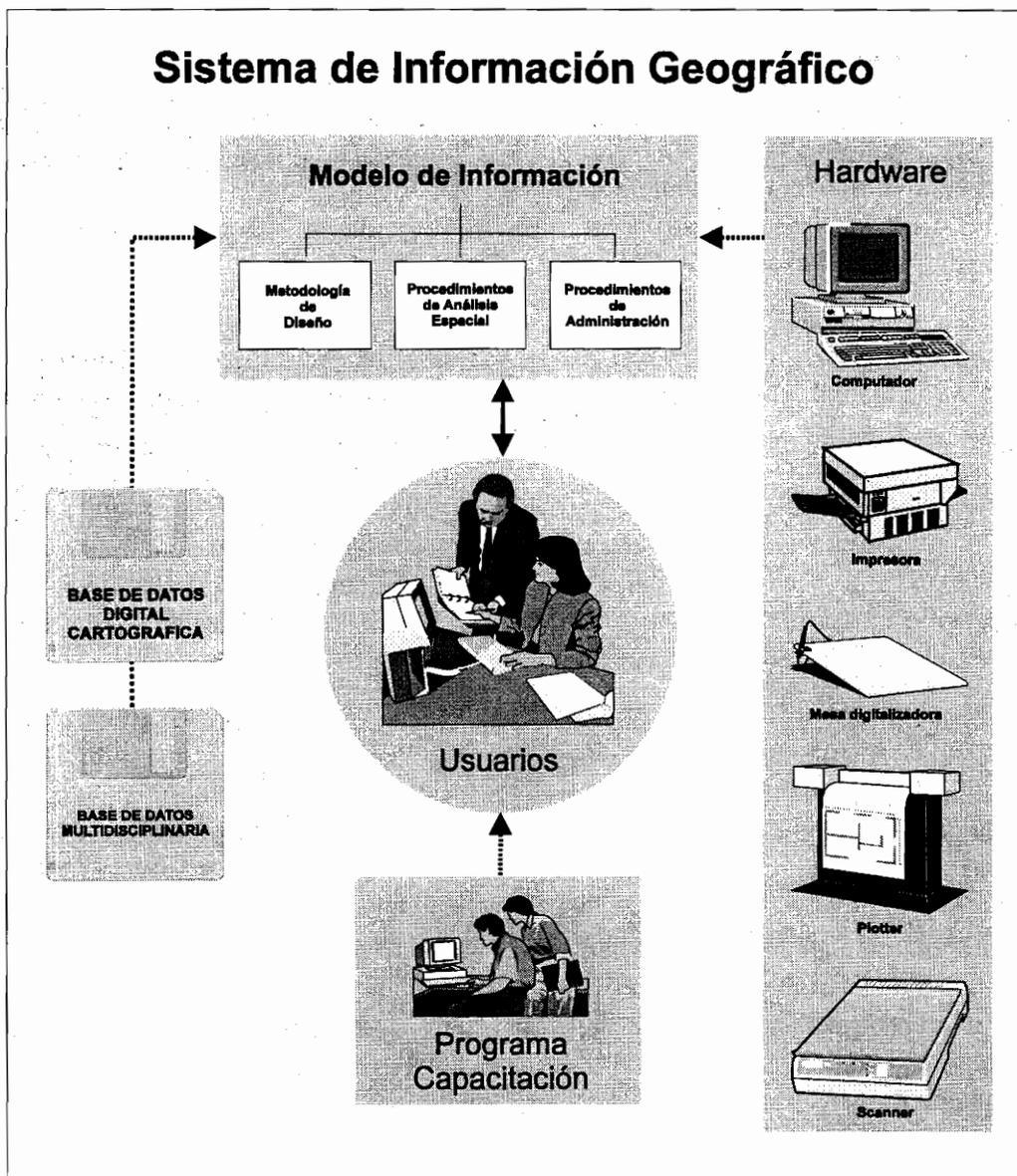
- **segmentación y muestras;** las relaciones topológicas existentes entre los elementos de las coberturas permiten automatizar el proceso de segmentación a nivel urbano y rural. Esta segmentación, según un número mínimo y máximo de viviendas por segmento, se implementa a través de la creación de nuevas coberturas que desde un punto de vista práctico implican un registro histórico. Finalmente, una vez creadas las coberturas de segmentos, un procedimiento básico permite seleccionar en forma aleatoria la muestra. Los SIG permutan la generación espacial de los segmentos muestrales. Las labores de actualización se simplifican a un grado máximo por la versatilidad de las herramientas de mantención y actualización.

- **labores de campo y fiscalización;** la producción de planos a diferentes escalas y tamaños permite la generación de material de apoyo para las labores de campo. Cada censista puede llevar un plano de ubicación con una perspectiva global de la zona de trabajo y, además, un plano a escala reducida del segmento muestral con información detallada de cada una de las viviendas que debe visitar. Estos mismos planos pueden ser utilizados para retornar modificaciones a la cartografía.

- **cartografía temática;** a partir de los procesamientos iniciales de las fichas de la encuesta, la producción de cartografía temática se facilita enormemente al disponer de mecanismos de ingreso, almacenamiento y clasificación de la información. Dada la relación directa entre los datos recabados y las coberturas a través de la codificación, el enlace para la producción de cartas temáticas se pone al alcance de todos los usuarios.

8. Esquema institucional de un SIG

El sistema de información geográfico debe considerarse como la interacción entre información espacial y descriptiva, un sistema informático capaz de realizar análisis espacial, los equipos computacionales dimensionados a la medida de los requerimientos, los usuarios, un programa de capacitación y los procedimientos administrativos para garantizar un relacionamiento armónico. El modelo de información es un reflejo de los objetivos del proyecto en términos de organización y manipulación de los datos. Este modelo debe definir claramente la metodología de procesamiento y sus limitaciones analíticas.



Cada uno de estos elementos sumados a la creatividad de los usuarios conforma el concepto de “sistema de información”. Por este motivo es importante tener en cuenta que cuando se habla de un sistema de información geográfico no se está hablando solamente de un programa, sino de la interacción de un conjunto de componentes que obedecen a un objetivo común.

El concepto de SIG institucional significa, necesariamente, un proceso de modernización de los recursos humanos y procedimientos administrativos. Su implementación requiere de la capacidad de los usuarios para asimilar la nueva tecnología e introducir técnicas eficientes de procesamiento de datos y evitar la mecanización de los procesos antiguos.

Esquemáticamente, el SIG institucional posee información, equipamiento, procedimientos de análisis, procedimientos administrativos y, lo más importante, usuarios capacitados para explotar las herramientas elaboradas. Como lo muestra la figura, debe existir una coordinación de todos estos elementos en función de las necesidades de los usuarios. En este punto es relevante destacar la importancia del soporte informático como factor de “lubricación” de cada uno de las componentes.

9. El proyecto SIG

El proceso de implementación de un proyecto SIG puede ser organizado en una serie de cuatro pasos. Aún cuando la mayoría de los proyectos que se desarrollen seguirán una secuencia similar, se debe tomar especial cuidado al diseñar cada proyecto en forma específica. Esto significa que los analistas encargados del diseño deben extraer de los usuarios las necesidades reales y evitar desarrollar lo que ellos consideren arbitrariamente como la solución. Según lo afirmado, los pasos a seguir son:

a) Determinar los objetivos; los elementos más importantes a considerar cuando se esté determinando los objetivos del proyecto son:

- ¿Cual es el problema por resolver? ¿Como se está resolviendo ahora? ¿Existen métodos alternativos de solución utilizando un SIG?
- ¿Cuales son los productos finales del proyecto - informes, mapas de trabajo, mapas con calidad de presentación? ¿Cuan frecuentes serán estos productos generados?
- ¿Cual es el auditorio objetivo del producto final - técnicos, planificadores o el público en general?
- ¿Existen o existirán otros usuarios de la misma información? Si es así, ¿Cuales son los requerimientos específicos?

b) Construcción de la base de datos; esta es la etapa más crítica del proyecto. El nivel de cobertura y la exactitud de la base de datos determina la calidad del análisis y productos finales

- Diseño de la base de datos - se debe determinar los límites del área de estudio, que sistema de coordenadas a utilizar, que coberturas de información geográfica se necesitan, cuales son los elementos presentes en cada cobertura, que atributos requiere cada elemento y como codificar y organizar los elementos.

- Automatizar los datos - ingresar la información espacial a la base de datos, hacer que la datos espaciales sean utilizables verificando y editando los errores y, por último, ingresar los atributos a la base de datos.

- Administrar la base de datos - ejecutar procedimientos de actualización y validación permanentes sobre la base de datos.

c) Analizar la información; en el análisis es donde ocurre el verdadero valor de un sistema de información geográfico. Un SIG puede ejecutar tareas analíticas de gran consumo de tiempo que muchas veces son hasta imposibles de realizar manualmente. En este ambiente se puede probar diferentes escenarios para una misma situación realizando modificaciones menores en los atributos.

d) Presentar los resultados del análisis; Un SIG ofrece distintas alternativas para presentar la información en forma de mapas o informes. El producto final debe guardar siempre estrecha relación con los objetivos del proyecto y con el auditorio objetivo.

V. LA BASE DE DATOS GEOGRÁFICA

1. Introducción

La implementación de un SIG a nivel institucional requiere del cumplimiento de un número significativo de objetivos. Uno de estos objetivos guarda relación con la producción de la base de datos geográfica. Como elemento fundamental del SIG, la base de datos geográfica provee la componente espacial al sistema tanto a nivel de análisis como en la presentación de resultados. Por este motivo, el desarrollo de esta base de datos, que es una de las componentes mas difíciles de implementar, conlleva, generalmente, el mayor gasto de recursos, tanto humanos como económicos.

El primer paso para la creación de una base de datos geográfica consiste en convertir la cartografía tradicional en formato digital. Este proceso llamado *digitalización*, se desarrolla utilizando dispositivos como un scanner o una mesa digitalizadora.

- El **scanner** requiere de un esfuerzo menor en la etapa inicial pues las cartas cartográficas son “leídas” automáticamente y transformadas a formato digital. Sin embargo, el scanner no hace diferencia entre los elementos que está leyendo y produce un archivo con todo lo encontrado. A partir de este punto se deben ejecutar procedimientos de limpieza para aislar cada componente en una cobertura separada. Estos procedimientos, que poseen límites poco adecuados de resolución, aumentan el riesgo de producir cartografía con bajos niveles de precisión.

- La **mesa digitalizadora** implica un procedimiento absolutamente manual, donde el operador o “digitalizador” coloca una carta sobre la mesa y, a través del uso de un puntero electrónico, recorre cada uno de los trazos a digitalizar marcando puntos (se debe presionar un botón en el puntero) lo más próximo posible de tal forma que cuando estos puntos se unan imaginariamente representen con precisión el trazo original. Este procedimiento, que resulta bastante más lento que el uso del scanner, permite altos niveles de precisión y requiere de menores esfuerzos de limpieza.

En la actualidad, el uso de la mesa digitalizadora se ha popularizado por haber demostrado altos índices de calidad. Este documento, dentro de su desarrollo, sólo trata el uso de este dispositivo como instrumento de creación de la base de datos geográfica. Básicamente, este proceso de automatización de los datos espaciales en un mapa, a través de la mesa digitalizadora, requiere del cumplimiento esencialmente de tres etapas. Estas etapas son:

- *Preparar la cartografía para la automatización*
- *Capturar los elementos cartográficos del mapa*
- *Visualmente evaluar la calidad de la captura de datos*

Durante este capítulo discutiremos los diferentes tópicos que se deben abordar para construir una base de datos digital cartográfica como parte del proceso de implementación de un sistema de información geográfica. Estos tópicos abarcan desde la preparación de la cartografía base, la aplicación de la tecnología SIG en la ejecución de proyectos, el proceso de capacitación y modernización de los recursos humanos y, por último, la producción de cartografía temática.

2. Propuesta para la creación de la base de datos geográfica

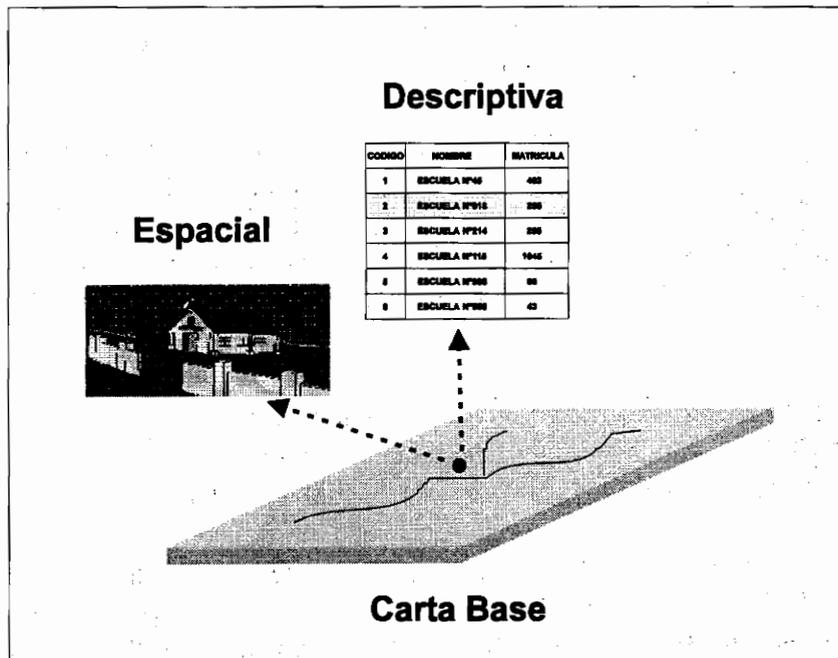
El primer paso sustantivo en la implementación de un SIG institucional es la creación de la base de datos geográfica. Para automatizar la geografía, se debe especificar explícitamente que información se va a almacenar, que estructura informática se va a utilizar y que uso se le dará a la base de datos.

Un SIG provee un modelo de datos que permite representar mapas en el computador. Una vez que se comprende como estos mapas digitales son creados y almacenados, las

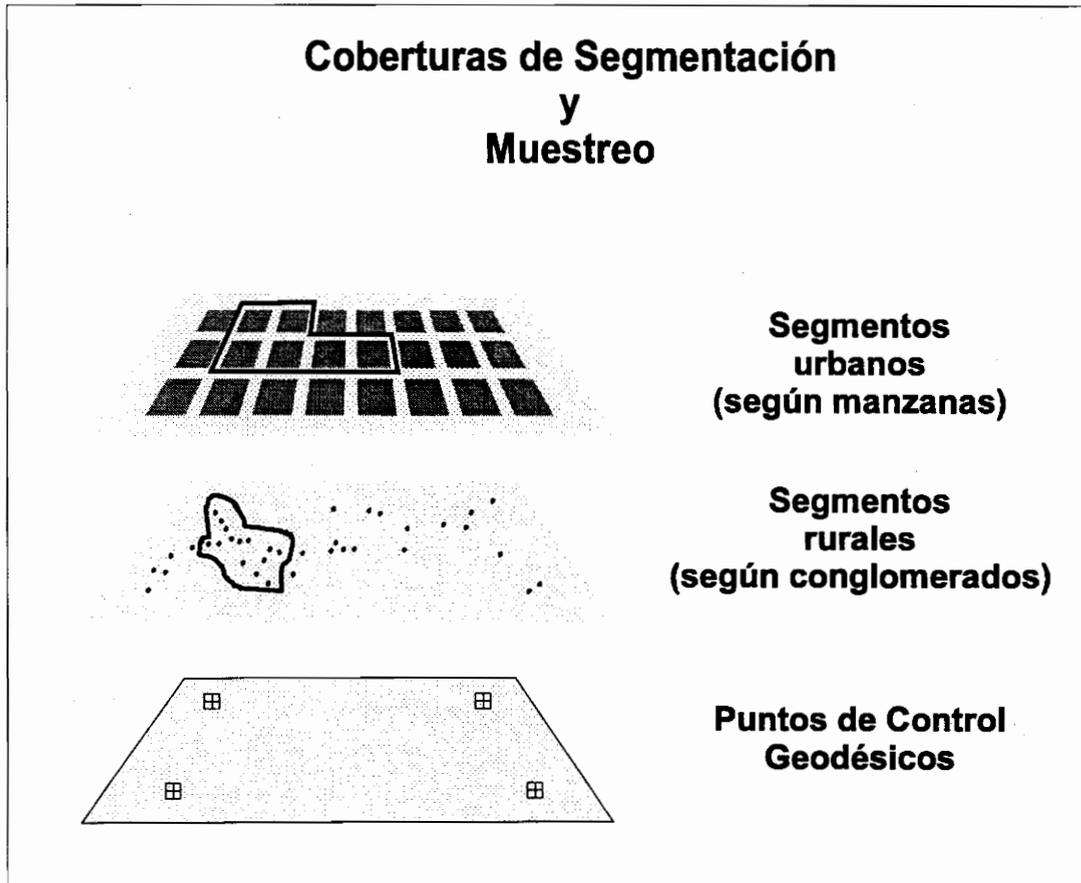
personas que implementarán el SIG pueden comenzar a construir la base de datos para apoyar el o los proyectos en cuestión.

Los tópicos fundamentales que deben ser considerados dentro de la implementación de esta base de datos son:

Tópico 1: Los dos tipos básicos de información: espacial y descriptiva. Dentro de un sistema de información geográfica se pueden manipular los objetos que existen dentro de una connotación espacial. Estos objetos poseen información referente a su tamaño, ubicación y relación con los demás objetos. Además, existe la información descriptiva que, en formato tabular, se almacena en forma de atributos.



Tópico 2: Los elementos geográficos necesarios para responder eficientemente a los requerimientos del proyecto. La información geográfica considerada en la implementación de un SIG orientado a la planificación y ejecución de trabajos relacionados con las encuestas de hogares requiere la presencia de elementos que faciliten la localización y referencia de las viviendas. En este sentido se debe considerar tanto los elementos geográficos naturales tales como vías, ríos y arroyos y los límites político administrativos del área de trabajo tales como límites comunales, distritales, provinciales, etc.



Tópico 3: El almacenamiento de los elementos geográficos según un sistema de coordenadas (georeferenciación). Antes de iniciar las labores de digitalización se debe tomar la decisión referente a que sistema coordinado se va a utilizar. Si una base de datos se va a utilizar para análisis espacial, todas las componentes deben estar registradas bajo un sistema coordinado común.

Un sistema coordinado esta compuesto por un esferoide (una descripción matemática de la forma de la Tierra) y una proyección de mapa (una conversión matemática de coordenadas esféricas a planas). El sistema UTM (Universal Transverse Mercator) es un ejemplo de sistema que utiliza proyecciones de mapas para representar los elementos tridimensionales de la Tierra en una superficie plana.

Tópico 4: Relacionamiento entre todos los objetos espaciales (topología). La capacidad de los sistemas de información geográfica de relacionar diferentes elementos en el espacio se logra a través del concepto de topología. Como componente del SIG, la estructura topológica de la base de datos la genera automáticamente el sistema, construyendo los atributos de relación entre los puntos, líneas y polígonos.

Tópico 5: Almacenamiento de la información descriptiva en formato tabular. La estandarización del formato de datos DBF ha facilitado la integración de diversas fuentes de información. La información descriptiva en un SIG puede ser administrada por una amplia gama de sistema de base de datos que soporte el formato DBF. Este hecho permite flexibilizar la construcción de aplicaciones y facilita la mantención de la base de datos.

Tópico 6: Los procedimientos de agrupación de la información espacial con la descriptiva. El poder de un SIG yace en su conexión entre la información gráfica (espacial) y la información tabular (descriptiva). El SIG que se adopte en una implementación debe generar y mantener automáticamente esta conexión.

3. Uso de la cartografía nacional

La construcción de la base de datos geográfica constituye una de las etapas más complejas e importantes en la implementación de un SIG. Por este motivo, se debe tener especial cuidado su elaboración y en los insumos (cartas base) que se adopten para desarrollar las coberturas.

Existen dentro de los países de la región existen varias alternativas para proveerse de la cartografía base. Esta fase de desarrollo debe estar respaldada por una actividad de recopilación y comparación de las cartas que se pueden obtener en instituciones como aquellas que se citan a continuación:

Instituto Nacional de Estadísticas; generalmente, y con propósitos censales, los INE poseen cartografía nacional urbana y rural. Esta cartografía es comúnmente bastante completa y posee buenos niveles de actualización derivados de las actividades de precenso que se realizan, en la mayoría de los casos, aproximadamente cada diez años. A pesar de esto, la cartografía generalmente no está georeferenciada y su índice de distorsión es bastante elevado. Para tal efecto, las cartas del INE pueden ser utilizadas como material de referencia y para extraer los límites político-administrativos. Sin embargo, sin mayores garantías de precisión, el uso de un GPS puede aportar las coordenadas para la georeferenciación.

En algunos países de la región, los INE ya poseen parte de su cartografía en formato digital. En estos casos se debe considerar si el formato de digitalización es compatible con el sistema que se utilizará y los posibles procesos de transformación que serán requeridos.

Instituto Geográfico Militar; dentro de los IGM se puede encontrar posiblemente al mejor cartografía en términos de calidad y precisión. La cartografía del IGM se puede adquirir en diferentes escalas y coberturas de acuerdo al nivel de detalle que se desee utilizar en la digitalización. Sin embargo, esta cartografía no siempre posee todos los límites político-administrativo por lo cual puede ser enriquecida por la cartografía censal.

Municipios e Intendencias; en el tema urbano, los municipios e intendencias poseen, generalmente, cartografía a nivel de manzanas con altos niveles de actualización. Este punto se debe, en la mayoría de los casos, a la necesidad de cartografía actualizada para fines de recaudación de impuestos. En algunos casos estas cartas poseen coordenadas, sin embargo, el uso de un GPS a nivel urbano permite optimizar la precisión cuando esto sea necesario.

Como insumo adicional y material de apoyo, las fotos aéreas son un buen medio para la extracción de información geográfica cuando no existe cartografía para el área en cuestión. Aún cuando se puede digitalizar directamente de la fotografía, estos insumos resultan altamente convenientes para verificar y aclarar zonas que resulten poco nítidas en una carta base. Se debe tener en cuenta que las fotos aéreas se distorsionan a medida que la distancia al centro de la foto aumenta.

4. Criterios para la clasificación de las cartas

La implementación de un sistema de información geográfica impone a su vez una nueva clasificación de las cartas. Tradicionalmente éstas fueron clasificadas según los temas representados: cartas físicas, políticas, económicas, etc.; dentro de la nueva visión “digital” de la cartografía, el amplio número de aplicaciones que se pueden desarrollar implica necesariamente una generalización de esta clasificación pues dentro de la temática actual de los SIG, la cartografía es básicamente un insumo que puede ser tratado en función de diferentes niveles de profundidad y precisión.

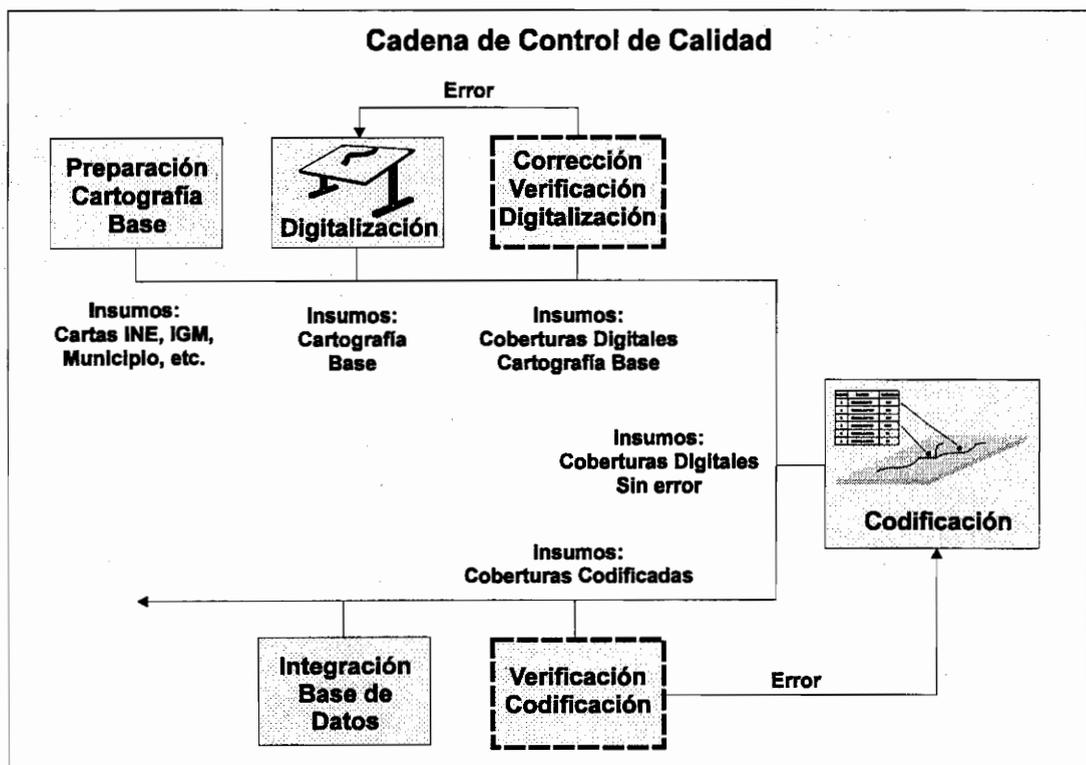
Dentro de la operatoria de un SIG, es práctico clasificar las cartas según su escala, clasificación que se mantiene siempre vigente, ya que la representación cartográfica conlleva una reducción de la realidad; por tanto la escala expresará la capacidad de la carta relativa a su contenido, es decir, la cantidad de información que puede comunicar, su grado de generalización y su precisión respecto a la localización. También dependerá de la escala la extensión del área que puede representarse en relación a un tamaño de hoja que sea manuable. En este sentido, Chile podría representarse en una sola hoja de escala 1:250.000, pero esto resultaría en un tamaño que no es práctico, perdiendo su visión global.

Actualmente, un manejo adecuado de las cartas respecto a la clasificación es aquel que contempla la manera de entregar la información. De acuerdo a este aspecto existen dos grandes tipos de cartas: las cartas topográficas y las cartas temáticas. La carta topográfica se puede definir como aquella confeccionada a base de un levantamiento original y que incluye la descripción de un lugar representando todos aquellos fenómenos u objetos visibles de la superficie terrestre en una dimensión altimétrica y planimétrica con una localización precisa lograda a través de la Geodesia. La carta temática es aquella confeccionada sobre una carta base, que trata uno o más temas en forma destacada. Tiene a su vez la capacidad de expresar fenómenos visibles o invisibles, concretos o abstractos.

5. Cadenas de control de calidad

En la construcción de la base de datos geográfica se deben cumplir una serie etapas desde la recopilación de información base hasta la formalización de la base de datos. Estas etapas se desarrollan por personal capacitado en cada una de ellas y verificadas por un técnico especializado. Este proceso de construcción de las coberturas que integrarán la base de datos debe administrarse a través de un proceso o cadena de control de calidad.

Preparación de la Cartografía Base; El control de calidad se desarrolla partir de la preparación de la cartografía base seleccionando las cartas más adecuadas para los objetivos del proyecto. En la selección de estas cartas se debe otorgar la debida importancia a la escala y al estado de conservación de las mismas. Según estas variables se crea una base de cartas insumo que se utilizarán para las labores de digitalización.



Digitalización; El proceso de digitalización requiere del uso de cartografía base apta para los objetivos del proyecto. Los datos en un mapa son ingresados al computador digitando punto por punto cada uno de los elementos geográficos, sean estos puntos que representen escuelas u hospitales o líneas que representen límites político-administrativos, ríos o vías.

Verificación de la Digitalización; esta verificación constituye el primer control de calidad de la cadena de producción. En esta etapa se ejecutan los procesos de limpieza y topología de la cobertura. A través de este control, el técnico verifica que todos los

elementos señalados en la cartografía base se hayan ingresado al computador. Este proceso se efectúa imprimiendo las coberturas y realizando con control visual con la carta base. Si se detectan errores u omisiones, esta cobertura se retorna a la mesa de digitalización.

Codificación; durante esta etapa se ingresan los códigos geográficos y atributos principales de las coberturas. Este proceso se realiza en utilizando las cartas base que fueron debidamente codificadas manualmente durante la etapa de preparación. Considerando el alto costo de tiempo que implica esta labor se sugiere la construcción de procedimientos que automaticen la codificación.

Verificación de la Codificación; este control de calidad tiene por objetivo verificar que todos los elementos que fueron incluidos en la digitalización posean su respectivo código. Este procedimiento puede ser automatizado a través de la creación de procedimientos a nivel tabular que verifiquen directamente en las tablas de atributos (información descriptiva) la existencia de los códigos. Cualquier error detectado en esta etapa debe ser remitido de vuelta a la codificación.

Integración a la base de datos; el procedimiento final durante la construcción de la base de datos consiste en integrar a la base de datos las nuevas coberturas. Este procedimiento debe considerar la estructura de la base de datos para la ubicación adecuada de los nuevos elementos.

VI. EL PROCESO DE CAPACITACIÓN

La metodología de implementación de un sistema de información geográfica propuesta en este documento incluye al usuario como factor central. Esta importancia dada al usuario está basada en el hecho de que son ellos los encargados de elaborar las soluciones y aportar la componente de creatividad. Además, esta metodología propone la implementación del SIG como parte de un proceso de modernización de los recursos humanos. Es imprescindible que los usuarios participen en cada componente de esta implementación, asimilando y aplicando las nuevas herramientas en sus tareas cotidianas.

El proceso de modernización debe incluir a todos los usuarios a través de la incorporación a un programa de capacitación especialmente diseñado para atender los nuevos desafíos que impone esta tecnología. Por este motivo es preciso distinguir a los usuarios en términos del papel que jugarán dentro del SIG.

Dentro del proceso de implementación, implantación, mantención y explotación de un SIG existen un número considerable de tareas por ejecutar. Cada una de estas tareas debe ser realizada por el personal indicado y con la capacitación adecuada. Por este motivo el diseño del programa de capacitación debe considerar a los usuarios según las funciones que van a realizar y según el nivel de responsabilidad que tendrán.

- En la etapa de implementación se debe considerar un proceso de capacitación orientado a la preparación de la cartografía base, digitalización, corrección, codificación e integración de la base de datos. Este grupo de producción puede incluir a cualquier usuario que haya recibido la adecuada capacitación.
- La etapa de implantación requiere que se capaciten a usuarios que un nivel intermedio de responsabilidad (jefes de unidad o departamento) para que coordinen y supervisen el armado generar de la base de datos. Estos usuarios deben controlar los ingresos y egresos de la información así como los cambios administrativos necesarios para facilitar la asimilación de la nueva tecnología.
- La etapa de mantención agrupa a los usuarios con mayor capacidad que han participado del grupo de implementación. Estos usuarios deben coordinar, orientar y ejecutar todos los procedimientos de actualización e ingreso de nuevos datos.
- La etapa de explotación incluye al mayor número de usuarios a nivel institucional pues su objetivo es habilitar el uso de las nuevas herramientas en el mayor número posible de aplicaciones. Estos usuarios deben ser capacitados en los procedimientos de análisis espacial, elaboración de cartografía aplicada, cartografía temática, etc.

VII. EQUIPAMIENTO Y COSTOS DE UNA INSTALACIÓN SIG

1. Los equipos y programas necesarios para implementar un SIG

El proceso de selección de equipos para una plataforma SIG debe evaluarse según los objetivos planteados en el diseño, el tamaño y volumen de información que se va a manejar y la modalidad operatividad del sistema. En la mayoría de los casos las inversiones iniciales en equipamiento se transforman en el largo plazo en el factor de menor costo dentro del proceso de implementación del SIG. Por este motivo se debe tener especial cuidado al elegir la plataforma y la calidad de los equipos para asegurar un buen servicio en el largo plazo.

Las diferencias más importantes entre las plataformas computacionales existentes en el mercado, radican en su capacidad de almacenamiento y velocidad de procesamiento. Estos parámetros aumentan proporcionalmente según el número de usuarios y aplicaciones, la resolución de los datos, el número de entidades y la complejidad de las aplicaciones. Para tal efecto se puede implementar un SIG a nivel de PC o estaciones de trabajo (workstation).

Los computadores personales o PC's están catalogados como el nivel inferior dentro de las plataformas computacionales. Este hecho se debe principalmente a su tradicional capacidad de almacenamiento y velocidad limitadas. Sin embargo, en la actualidad este factor está cambiando aceleradamente hasta el punto de acercarse al poder de

una estación de trabajo. En general estos equipos poseen una orientación mono-usuario pero los avances que han introducido los sistemas operativos multi-usuarios han añadido una nueva dimensión al trabajo en grupo. Si a este punto le agregamos los bajos costos de las unidades, la plataforma PC se transforma en una alternativa muy atractiva.

Las estaciones de trabajo, que poseen una arquitectura similar a los equipos mainframe y mini-computadores, introducen una nueva dimensión al trabajo con tecnología SIG. Su alta capacidad de procesamiento y almacenamiento las ubican en una excelente posición para albergar un ambiente SIG. Sin embargo, las estaciones de trabajo aún se cotizan a niveles de costo, situándose las más simples por sobre los 30.000 dólares.

2. Consideraciones referentes a los dispositivos

El la implementación de un SIG, la elección de los dispositivos que van a configurar la componente de hardware debe cumplir con una serie de exigencias en términos de producción y costos para garantizar el buen desarrollo del sistema. Los dispositivos a considerar son:

a) Procesadores; la experiencia a nivel de la región lleva a la conclusión de que la implementación de un SIG a nivel institucional se realiza con menor trauma como un proceso participativo donde los usuarios asimilan la nueva tecnología a medida que se integran al uso de las herramientas. Esto implica que la implementación es un proceso gradual donde cada etapa es ejecutada en función de las capacidades de los usuarios.

El uso masivo y la familiaridad con los microcomputadores y su bajo costo son factores relevantes en el momento de decidir el uso de una plataforma. En la actualidad los procesadores Pentium con velocidades de hasta 266Mhz son ampliamente capaces de ofrecer una alternativa eficiente para introducir a los usuarios en la implementación y explotación de un SIG. Estos procesadores, cuyos precios oscilan entre los 1.000 y 4.000 dólares, permiten configurar una pequeña red de no más de tres equipos que satisface los requerimientos iniciales de un SIG.

b) Medios de almacenamiento; a nivel nacional, los requerimientos de una base de datos geográfica puede oscilar, para un país del tamaño de Ecuador o Paraguay, en aproximadamente 200Mb. Este requerimiento de espacio resulta ínfimo se consideramos que el costo promedio de un disco IDE de 2.5Gb es de 300 dólares.

Las unidades lecto-grabadoras ópticas de 600Mb con costos aproximados de 1.000 dólares más 15 dólares el disco, las grabadoras de CDRom con capacidad de 650MB y costos aproximados de 800 dólares más 10 dólares el disco y, finalmente, las unidades de disquetes de 100Mb y 230Mb ofrecen alternativas tentadoras para almacenar grandes bases de datos a bajo costo.

c) **Equipos de entrada de datos;** en el proceso de creación de la base de datos geográfica, el uso de las mesas digitalizadoras por sobre los scanners es prácticamente un estándar. Su amplio control y precisión durante el proceso de digitalización, permite a los usuarios generar las coberturas con facilidad y eficiencia.

El proceso de digitalización varía en función del nivel de detalles y de la calidad de la cartografía base. Estos factores sumados a los tiempos esperados de implementación perfilan la cantidad de mesas necesarias para ejecutar el trabajo. Sin embargo, se debe tener en consideración que este posible uso masivo de las mesas es sólo en una etapa inicial y su uso se reduce solamente a actividades de actualización y al ingreso de nuevas fuentes de información geográfica. Una mesa tamaño A1 y A0 oscilan entre los 2.500 y 3.500 dólares.

En el proceso de digitalización de Paraguay se utilizaron cuatro mesas digitalizadoras en un periodo de tiempo real de un año con turnos de 12 horas diarias.

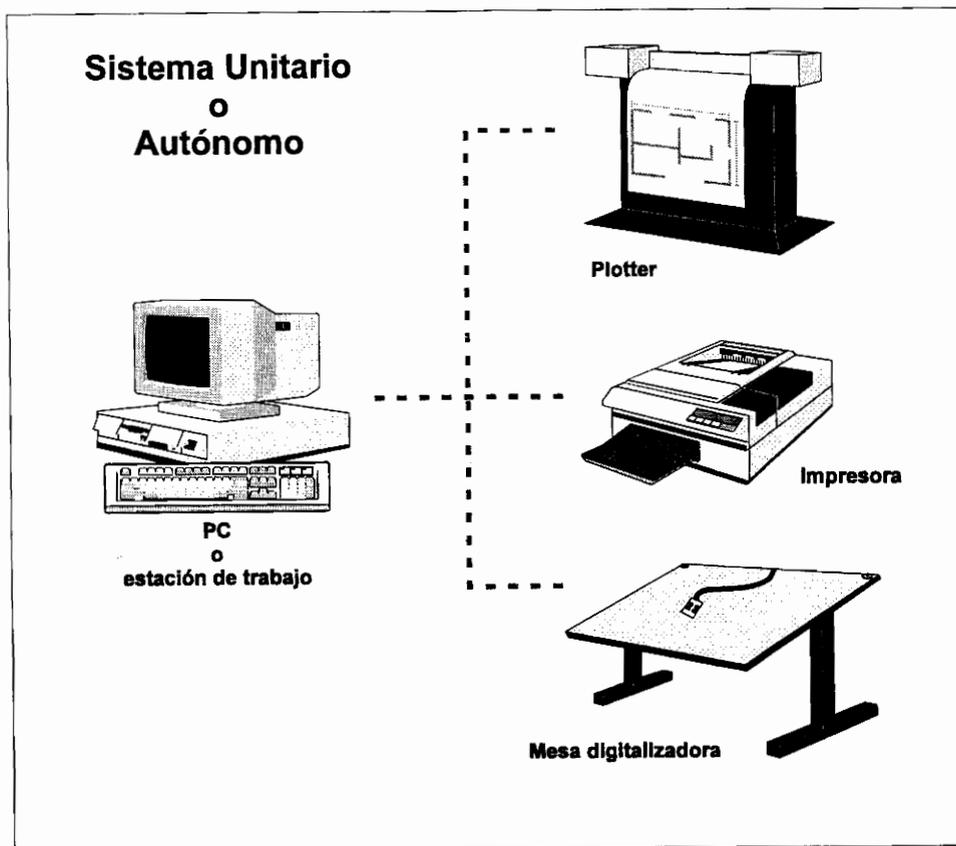
d) **Equipos de salida de datos;** existen variadas formas de presentar los productos resultantes del análisis espacial. Dentro de los más utilizados podemos citar las impresoras a color y los plotters.

Las impresoras a color con tecnología "inkjet" se han popularizado por su alto rendimiento, excelente resolución (300dpi) y bajos costos. Estas impresoras pueden producir cartografía temática con buen nivel de presentación y sus costos oscilan entre los 400 y 800 dólares aproximadamente.

Para trabajos más exigentes, el uso de plotters tamaño A1 y A0 proveen excelentes niveles de presentación con 300 dpi para imágenes en colores y hasta 720 dpi para imágenes en blanco y negro. Sin embargo, a pesar que sus costos promedian los 9.000 dólares, los insumos como el papel y la tinta continúan siendo un factor de cierta complejidad. Un rollo de papel normal para plotter tamaño A0 fluctúa en los 50 dólares.

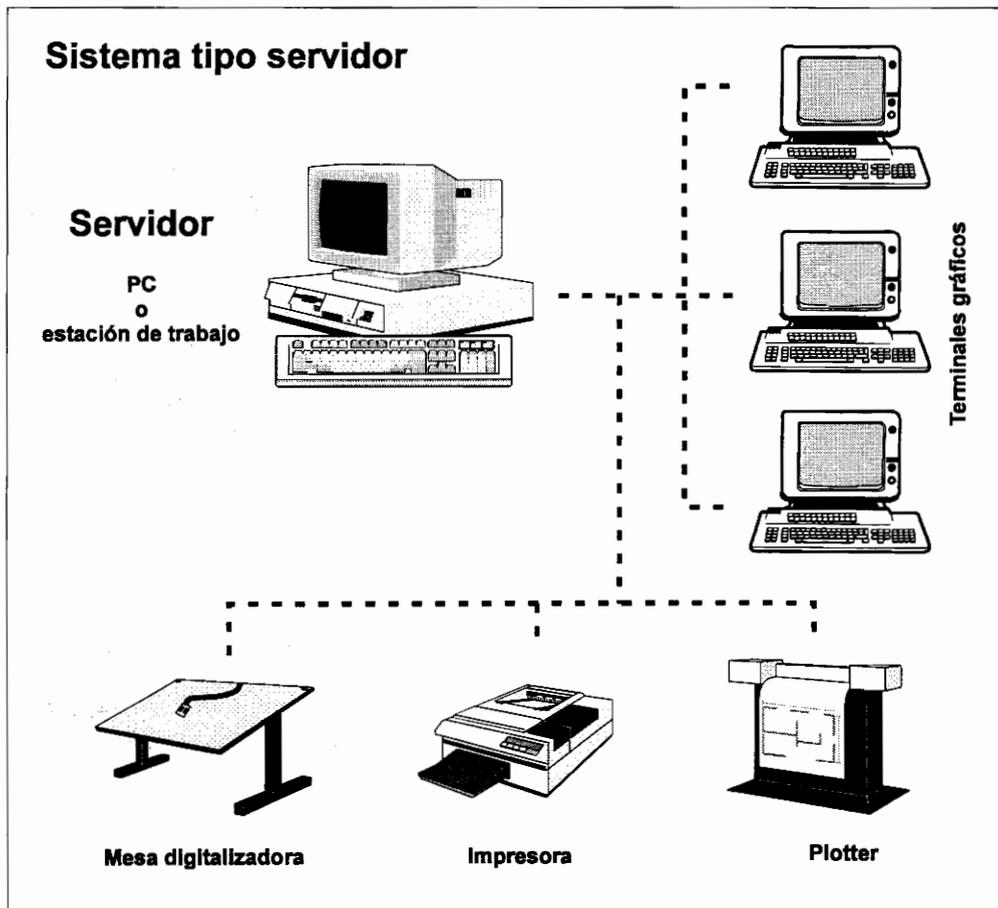
e) **Configuración**

- **Sistemas unitarios** o más conocido como "stand-alone" está constituido por un computador que funciona en forma aislada. Este tipo de sistemas posee un alto grado de control sobre las aplicaciones, se pueden adaptar fácilmente a las necesidades específicas de cada sector, sus costos son módicos y poseen un alto potencial para integrarse a redes en forma gradual. Este tipo de sistemas es utilizado con frecuencia en instalaciones SIG en las etapas iniciales experimentales especialmente por su fácil manejo y reducido costo.



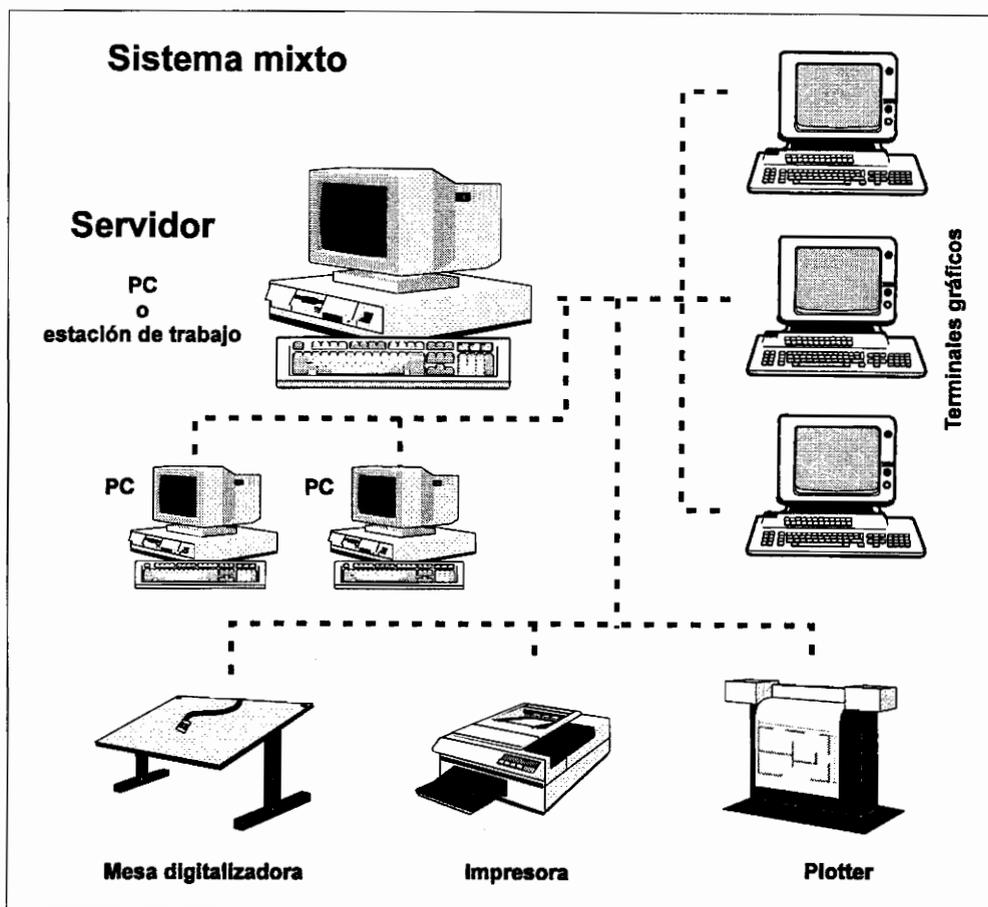
Esta configuración presenta una alternativa muy atractiva para iniciar el proceso de implementación de un SIG. A pesar de su limitado alcance sumado al hecho que puede ser operado simultáneamente por un usuario, con esta configuración se puede crear una base de datos geográfica e implementar procedimientos de análisis espacial con cartografía temática. Aún así, bajo una coordinación apropiada, esta configuración puede ser utilizada por un grupo de usuarios.

- **Sistemas tipo servidor** están basados en una operación y administración centralizada. Las aplicaciones se desarrollan en las unidades terminales y su almacenamiento y ejecución son encargadas al servidor. Sus ventajas más relevantes en la implementación y explotación de un SIG son referentes al intercambio sencillo de información, flexibilidad al compartir los periféricos, almacenamiento estandarizado de las aplicaciones y la información y un soporte técnico centralizado. Sin embargo este tipo de configuraciones tiene algunas desventajas tales como el alto costo y la necesidad de terminales con alta capacidad gráfica.



- **Sistemas mixtos** son aquellos que permiten la conexión de computadores de diferente tamaño en red. Su configuración permite formar grupos en forma de ambientes de trabajo para aplicaciones específicas. Por tratarse de una configuración flexible y versátil, los sistemas mixtos son los más utilizados en las instalaciones SIG.

Las características principales de los sistemas mixtos se refieren a su capacidad de integrar procesadores independientes, coordinación operativa y administrativa y la posibilidad de desarrollar integradamente las aplicaciones. Sus ventajas más relevantes incluyen procesamiento gráfico integrado a la red, flexibilidad y autonomía departamental, costos razonables y posibilidades amplias de crecimiento. Entre sus desventajas esta la complejidad operativa y dificultad de estandarización.



3. Consideraciones referentes a los programas SIG

El proceso de selección del software para la implementación del SIG es recomendable que se efectúe tomando en consideración la experiencia de otras instituciones como, por ejemplo, los institutos de estadística.

La variable costos del sistema es un factor de gran importancia si consideramos que los sistemas más competitivos trabajan en base a llaves de protección (licencias). Cada copia del sistema se acompaña de una llave que evita su uso en más de una máquina. Cada una de estas llaves o licencias tiene un costo generalmente significativo en relación al costo inicial del sistema. Estos costos varían notablemente de una plataforma PC a una plataforma workstation.

Los altos costos que implican estos programas requieren que su selección tome en cuenta aspectos tales como el alcances de sus herramientas de análisis espacial, su potencial en la creación de las coberturas geográficas, su adaptabilidad a diversos proyectos en diferentes ambientes, interface amigable para los usuarios y un excelente soporte y documentación.

4. Costos generales de la implementación de un SIG

Los aspectos financieros juegan un rol de gran importancia en el momento de planificar e implementar un SIG. Aún cuando los procesos involucrados en la implementación generan en la mayoría de los casos altos costos, existen alternativas a nivel de microcomputador que reducen los costos notablemente. Sin embargo, antes de clasificar los costos en altos o bajos debemos tener en cuenta cuáles son los beneficios que brindará la nueva tecnología.

La cuantificación de los costos y los beneficios es la mejor estrategia para evaluar el impacto del SIG. No obstante, los cambios inherentes a todo proceso informático producen nuevos elementos que son muchas veces impredecibles y difíciles de cuantificar. Una metodología adecuada de estimación de costos implica abordar cada una de las etapas de la implementación para verificar los costos y beneficios según su origen.

En el proceso de evaluación de la factibilidad de un proyecto ciertos costos son relativamente fáciles de predecir. El tiempo y costo estimado de entrenamiento y capacitación, o la digitalización de los mapas que hacen parte del proyecto se pueden cuantificar por medio de propuestas a entidades externas.

Las actividades de planificación, asignación de personal, capacitación y entrenamiento son tan importantes dentro de un proyecto SIG como los computadores, programas y procesos de digitalización.

Costos

- Realización de un estudio de factibilidad para implementación del SIG a nivel institucional. El estudio debe ser realizado con una perspectiva corporativa y por un especialista en análisis de sistemas.
- Diseño del sistema considerando las variables informáticas y administrativas. El diseño debe considerar los aspectos de implantación del sistema a nivel institucional.
- Adquisición del equipamiento computacional de acuerdo al diseño. Se debe tener en consideración la plataforma computacional.
- El sistema de información geográfica y la producción de las aplicaciones específicas para satisfacer los requerimientos planteados.
- El proceso de capacitación del personal y la transición hacia el uso de los nuevos procedimientos.
- La producción de la base de datos geográfica.

- La operación del sistema y las labores de mantención y actualización de la base de datos.

Beneficios directos

- El uso de la tecnología SIG implica una mayor eficiencia en el procesamiento de la información.
- La componente geográfica introduce la variable espacial en el análisis de la información.
- El alto nivel de codificación que requieren las aplicaciones SIG optimizan los procedimientos y tiempos de acceso a toda la información.
- El mejor uso de la información y la introducción de nuevas herramientas de análisis espacial proveen una mayor objetividad en la toma de decisiones.

Fundamentalmente, el uso de la tecnología SIG impone un incremento de la productividad a través de la reducción de los procesos manuales, la habilidad para ejecutar procesos complejos y la reducción en el tiempo de respuesta. Además, el ordenamiento informático implícito en la implementación del sistema reduce los costos a largo plazo al optimizar el proceso de almacenamiento de la información, mejorando la manipulación inter-departamental de la información e incrementando su exactitud.

5. Detalle de costos

Para realizar una cálculo detallado de los costos que implica la implementación e implantación de un sistema de información geográfica en una institución, se deben considerar en forma detallada un número bastante considerable de factores. Cada uno de estos factores se listan a continuación:

Costos de equipos y programas

- **Compra de los equipos;** equipo central o servidor, terminales, impresoras, mesas digitalizadoras, plotters, pantallas de alta calidad gráfica, medios de almacenamiento masivo, unidades de respaldo, etc.
- **Compra de los programas;** un sistemas de información geográfico (el programa propiamente tal), un sistema de administración de base de datos, un programa para georeferenciar fotografías aéreas, etc.
- **Mantenimiento de los equipos;** un contrato de mantención permanente de los equipos.

- **Soporte para los programas;** soporte telefónico o a través de INTERNET para realizar las consultas necesarias.
- **Adecuación de las instalaciones;** habilitación de un espacio físico apropiado para implantar el sistema de información geográfica.
- **Actualización de los equipos;** un programa permanente de actualización y renovación de equipos.

Costos de desarrollo de la base de datos

- **Control Geodésico;** el control geodésico a través del uso de GPS debe asimilarse como una actividad permanente con el objeto de realizar las actualizaciones necesarias a la base de datos. Cada elemento que se va a ingresar a la base debe ser controlado eficazmente para evitar degradar el nivel de precisión.
- **Obtención de fotografías aéreas;** como material de apoyo al proceso de creación de la base de datos geográfica se debe considerar la adquisición de fotografías aéreas. Estas fotos pueden ser, además, utilizadas como base para digitalización cuando no existe cartografía base adecuada.
- **Preparación de la cartografía fuente;** el proceso de recopilación de la cartografía base puede resultar largo y complicado. Las diferencias de escala y temática influyen para aumentar los tiempos de obtención de una cartografía apropiada para los fines del proyecto. Este proceso puede demandar costos considerables.
- **Digitalización y traducción de formatos;** dependiendo del tamaño del área requerida para atender en forma adecuada los objetivos del proyecto, el proceso de digitalización puede requerir extensos turnos de un número elevado de digitalizadores. A estos costos se pueden sumar la incorporación de datos que provengan de otros formatos.
- **Procesos de control de calidad;** la implementación de un SIG debe estar, en forma permanente, monitoreada por un proceso de control de calidad. Este proceso garantizará información válida y oportuna para satisfacer los requerimientos del proyecto. Sin embargo, este proceso debe estar dirigido por un técnico especializado en técnicas de validación y consistencia.

Costos de consultoría externa especializada

- Elaboración de requisitos de sistematización
- Elaboración de análisis, planificación del proceso de sistematización
- Diseño de la configuración de equipos y programas
- Diseño de la base de datos

- Diseño de procedimientos y aplicaciones
- Entrenamiento y capacitación
- Evaluación del contexto legal y políticas de manejo de la información

Costos de personal

- Personal directivo y coordinador
- Soporte técnico
- Encargados de la capacitación
- Seminarios

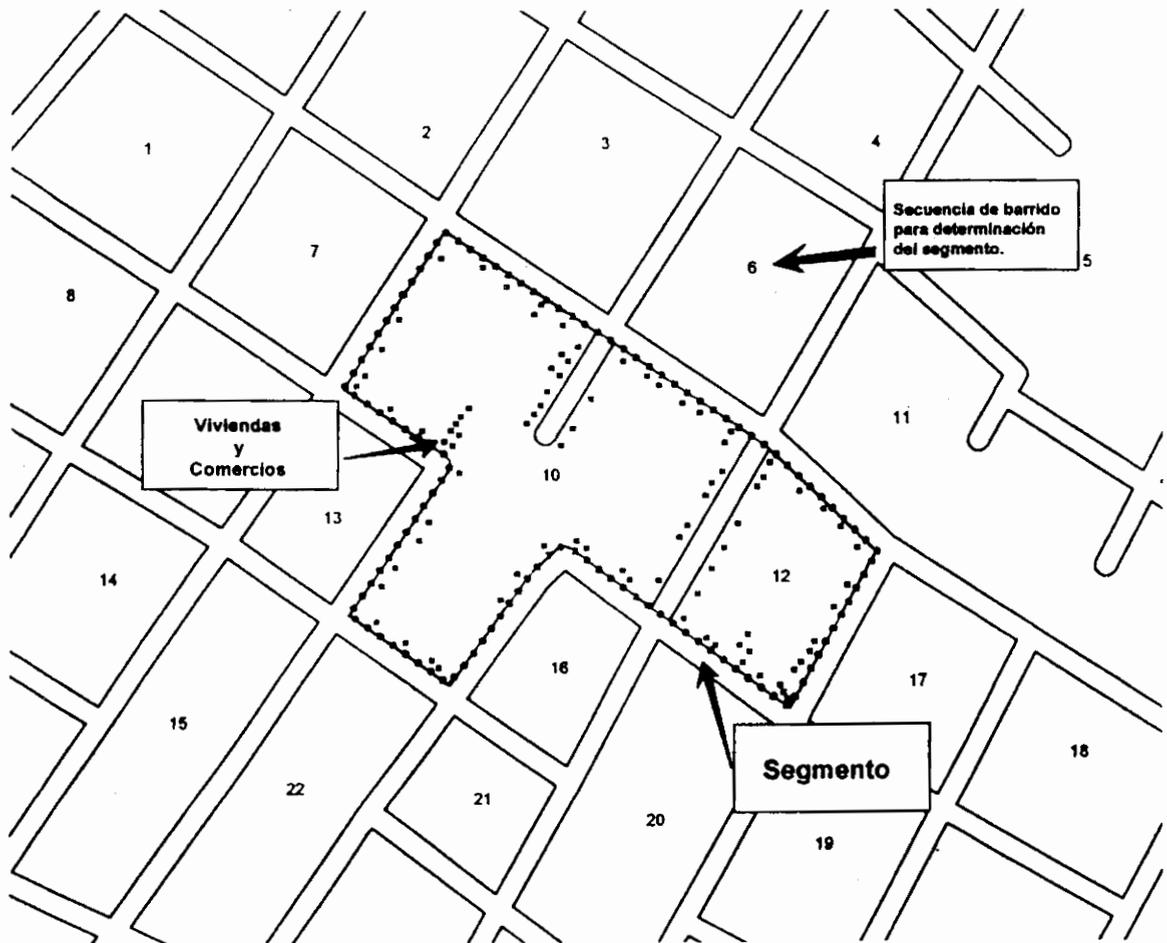
VIII. CONCLUSIONES

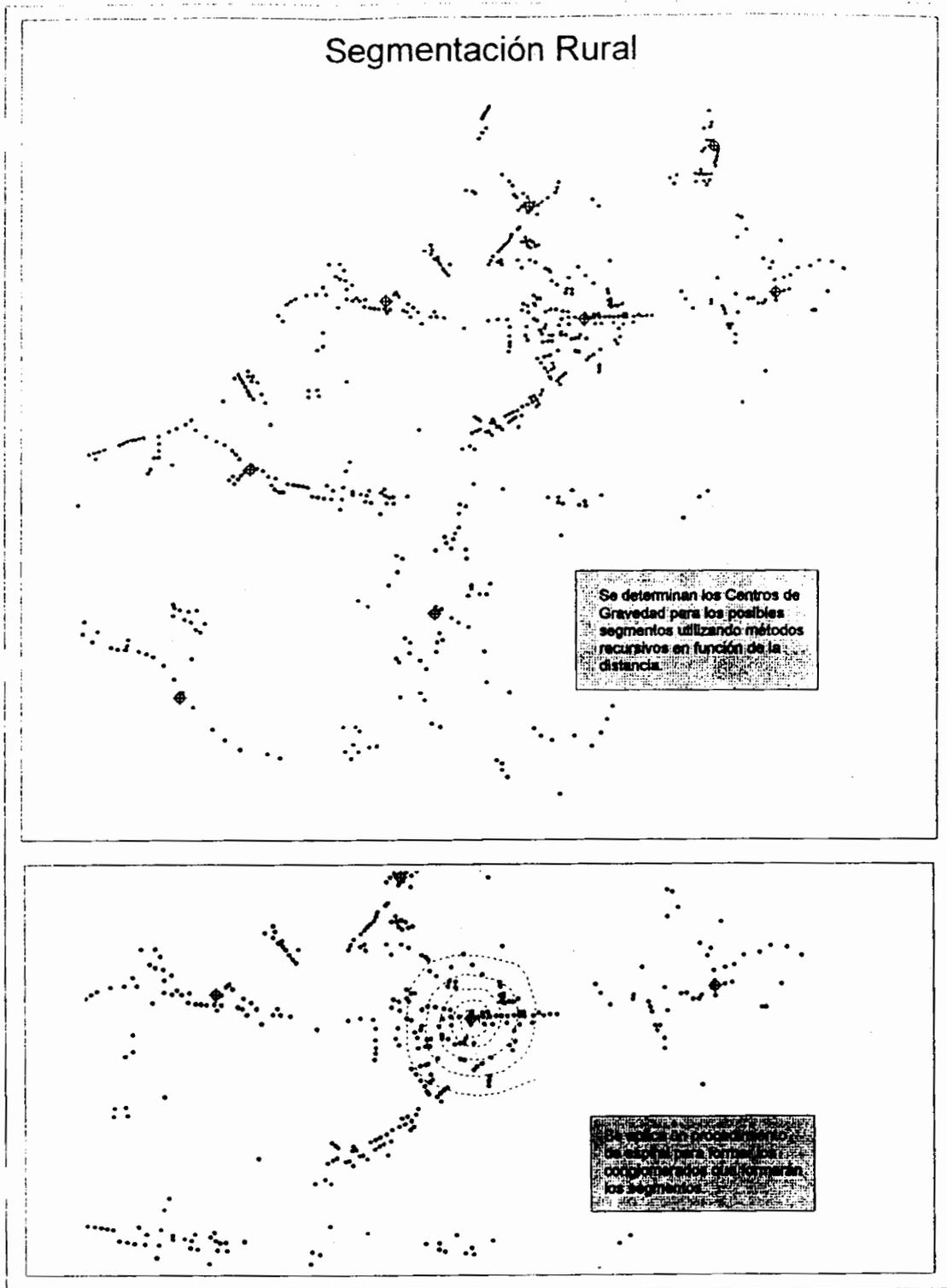
El diseño e implementación de procedimientos de automatización de las actividades y tareas en torno a la ejecución de los trabajos en una encuesta de hogares debe visualizarse necesariamente en el contexto de un proceso de transformación y cambio cuyo objetivo final sea el diseño y puesta en marcha de un SIG. Como primera etapa es aconsejable orientar los objetivos al proceso de digitalización de la cartografía. Dentro de esta etapa es altamente recomendable establecer convenios de intercambio de información con el o los organismos oficiales encargados de la generación de la cartografía nacional y con cualquier otra institución que haya desarrollado esfuerzos similares de calidad comprobada. Paralelamente, se puede iniciar el diseño de la base de datos relacional que permita integrar la información estadística de las unidades básicas definidas con fines de muestreo (por ejemplo estados, municipios, localidades, manzanas, segmentos y viviendas), con diversos aspectos del medio geográfico y su distribución espacial.

La creación de la base de datos digital-cartográfica es, generalmente, un de los procesos que presenta las mayores dificultades en la producción del SIG. La cartografía proveniente de las Oficinas de Estadística se puede encontrar en variados niveles de calidad, detalle, escalas y cobertura. Generalmente, es necesario recurrir a fotografías aéreas para obtener información de zonas poco documentadas y faltas de procesos de actualización. En los países de la región, se ha adoptado una política muy recomendable la cual consiste es establecer convenios de apoyo técnico entre las Oficinas de Estadística y los Institutos Geográficos. Estos convenios poseen una utilidad recíproca al permitir el intercambio de experiencias, información y técnicas que permiten perfeccionar a los recursos humanos. Además, esta práctica enriquece la generación de una cartografía común para el establecimiento de una base de datos normalizada, válida y consistente.

Elementos Cartográficos Producción de Segmentos Muestrales

Manzanas, Viviendas y Comercios

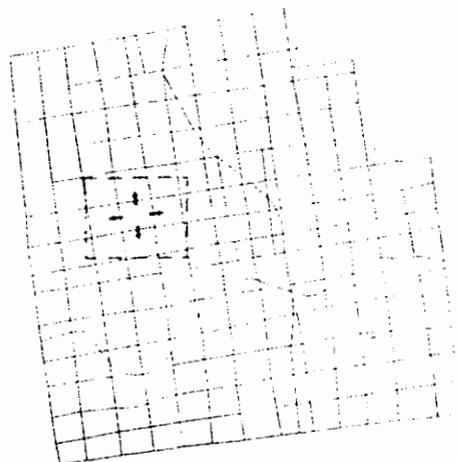




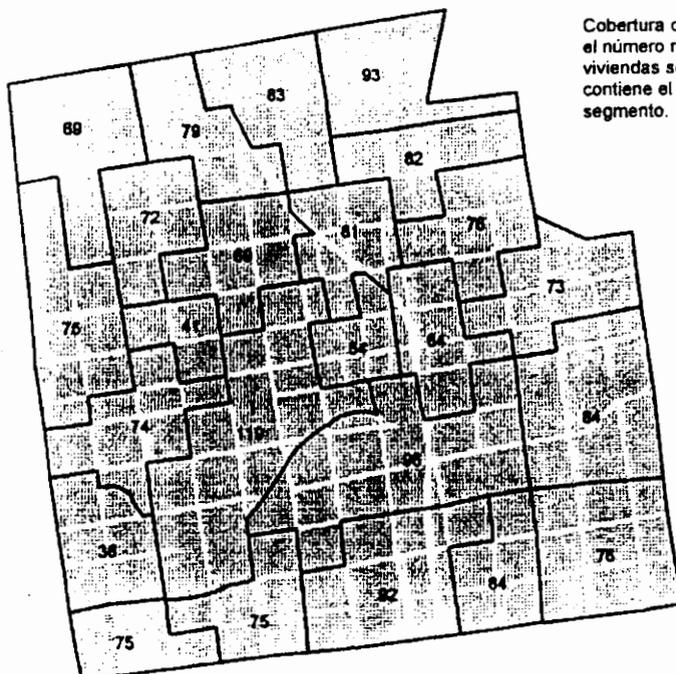
Segmentación Urbana



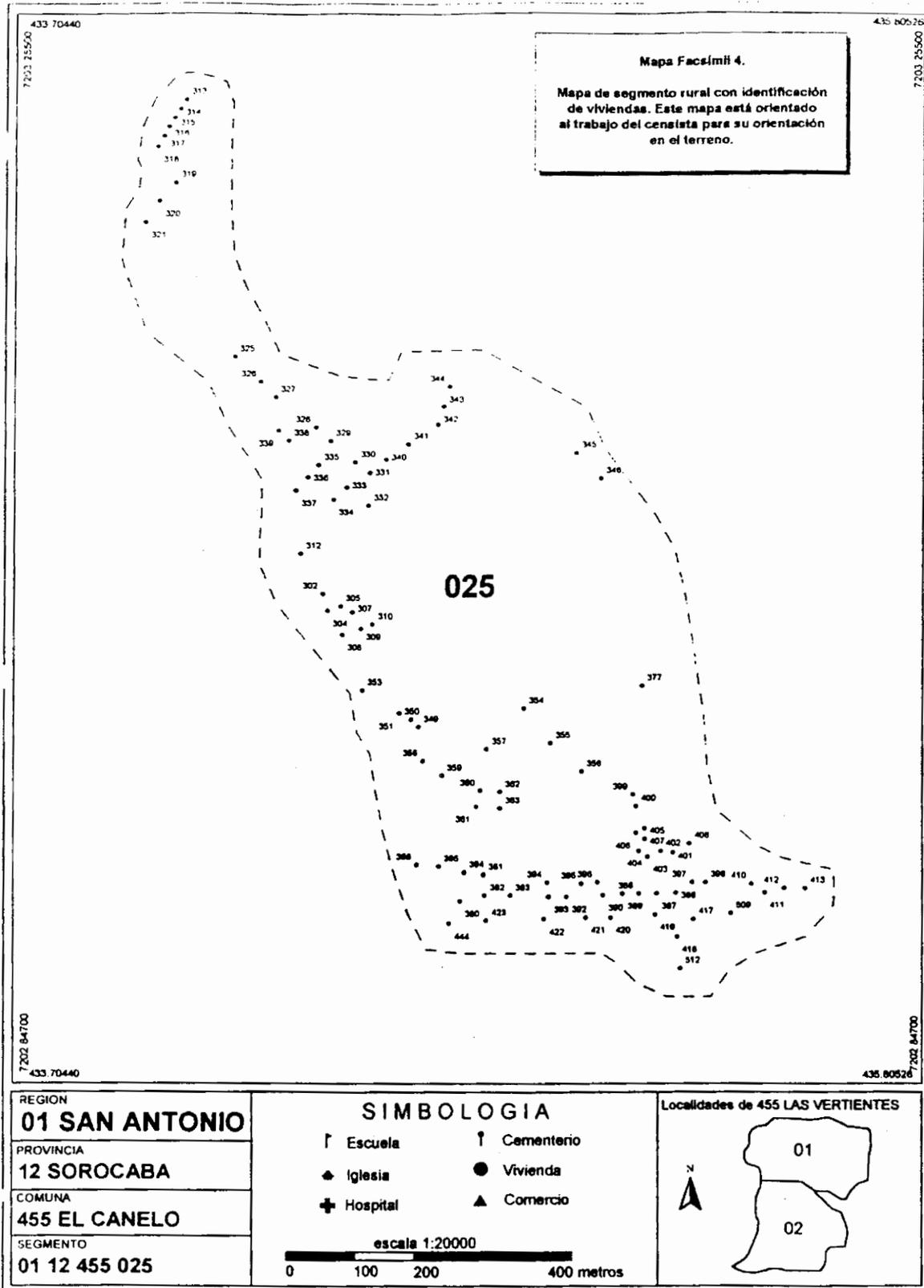
Cobertura de manzanas con el número de viviendas almacenado como atributo.

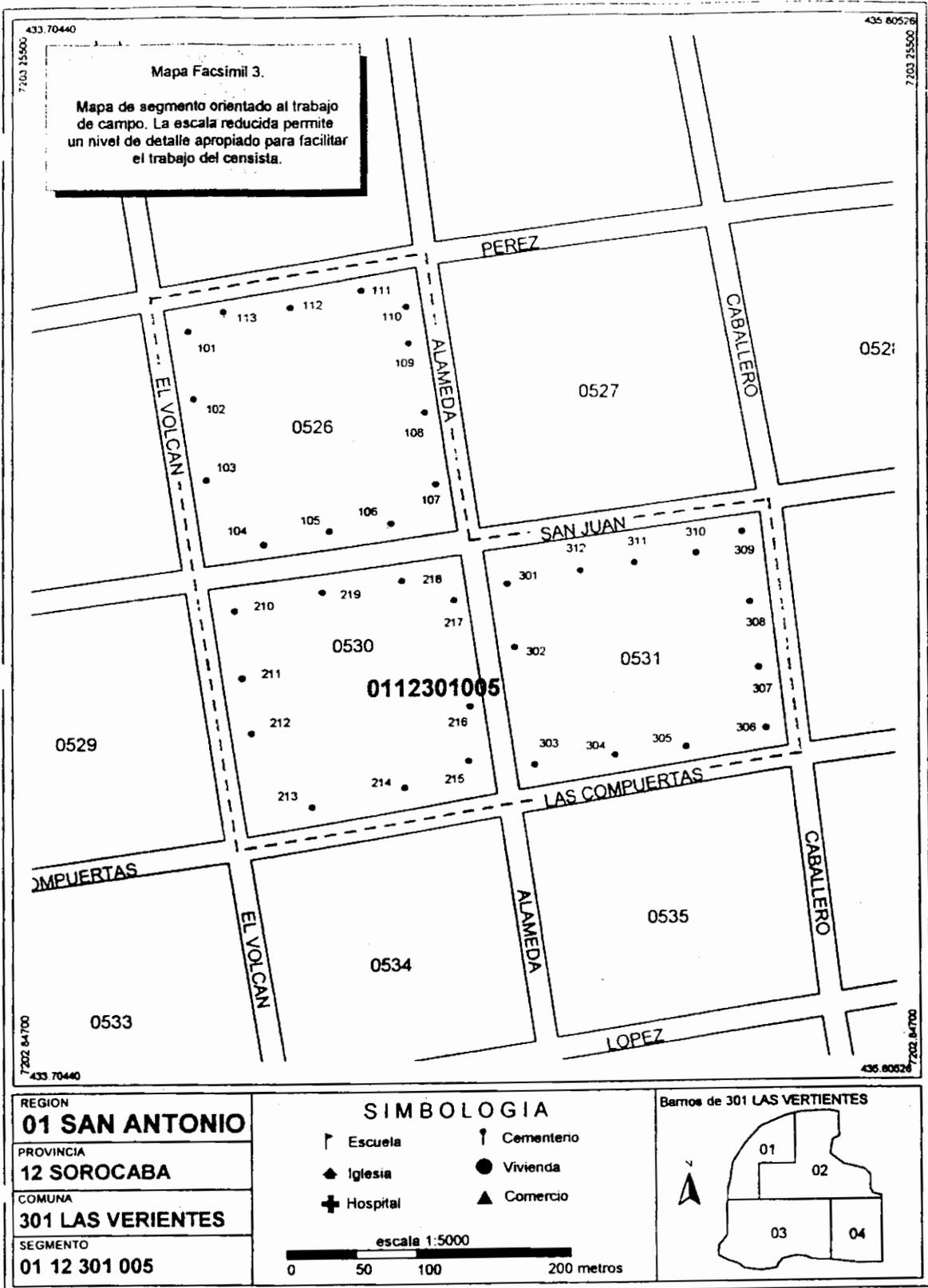


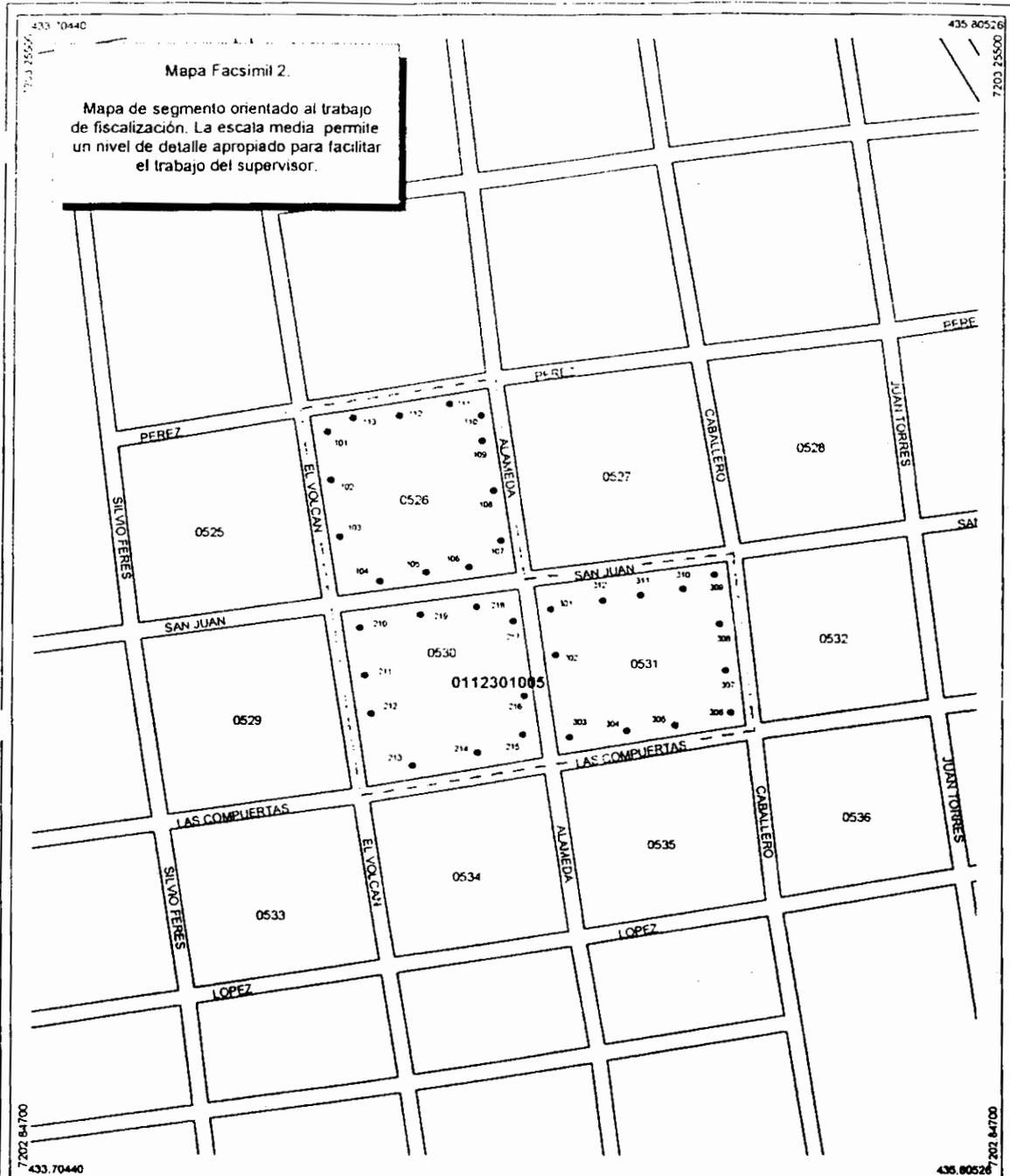
Cobertura de ejes centrales con la relación topológica entre cada manzana y su vecino



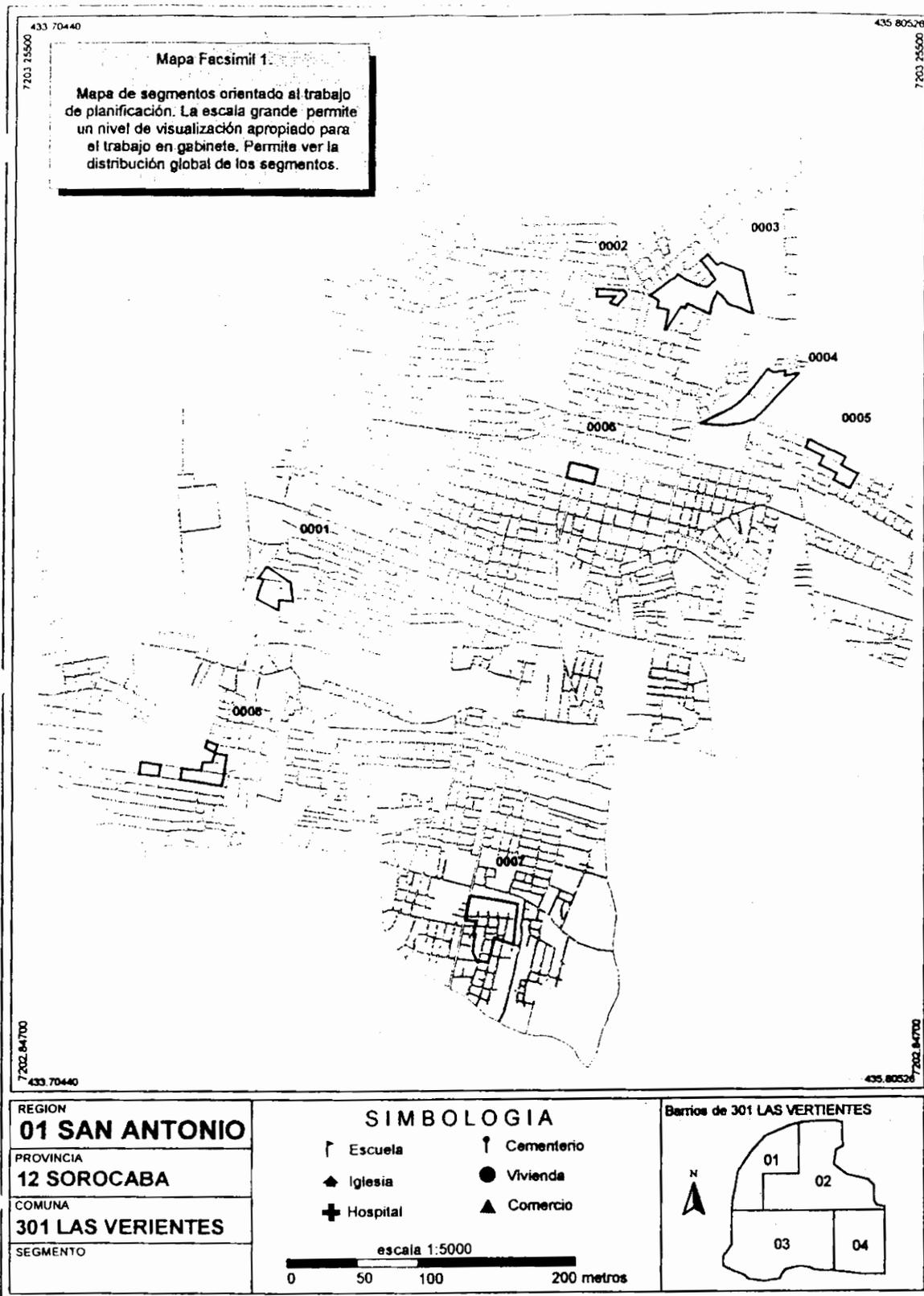
Cobertura de segmentos según el número mínimo y máximo de viviendas solicitado. cada segmento contiene el número de viviendas del segmento.







<p>REGION 01 SAN ANTONIO</p> <p>PROVINCIA 12 SOROCABA</p> <p>COMUNA 301 LAS VERIENTES</p> <p>SEGMENTO 01 12 301 005</p>	<p style="text-align: center;">SIMBOLOGIA</p> <table border="0"> <tr> <td>Escuela</td> <td>Cementerio</td> </tr> <tr> <td>Iglesia</td> <td>Vivienda</td> </tr> <tr> <td>Hospital</td> <td>Comercio</td> </tr> </table> <p style="text-align: center;">escala 1:5000</p> <p style="text-align: center;">0 50 100 200 metros</p>	Escuela	Cementerio	Iglesia	Vivienda	Hospital	Comercio	<p style="text-align: center;">Barrios de 301 LAS VERTIENTES</p>
Escuela	Cementerio							
Iglesia	Vivienda							
Hospital	Comercio							



**TAMAÑO ÓPTIMO DE MUESTRA EN ENCUESTAS
DE PROPÓSITOS MÚLTIPLES**

FERNANDO MEDINA

CEPAL

ÍNDICE

	<u>Página</u>
I. INTRODUCCIÓN	119
II. TAMAÑO DE MUESTRA ÓPTIMO PARA DISTRIBUCIONES MULTINOMIALES: EL CASO UNIVARIADO PARA DISEÑOS MONOETÁPICOS	120
1. El cálculo del tamaño de muestra: Aspectos prácticos.....	123
2. Ejemplo de cálculo	126
III. LA DETERMINACIÓN DEL TAMAÑO DE MUESTRA EN DISEÑOS POLIETÁPICOS Y MULTITEMÁTICOS	133
1. El planteamiento matemático	135
2. El criterio de optimalidad	136
3. Formulación específica: Muestreo aleatorio simple estratificado	137
4. Aplicación del método	138
5. Muestreo polietápico	140
6. Aplicación del método	141
IV. CONCLUSIONES	144
REFERENCIAS BIBLIOGRÁFICAS	146

TAMAÑO ÓPTIMO DE MUESTRA EN ENCUESTAS DE PROPÓSITOS MÚLTIPLES

I. INTRODUCCIÓN

La determinación del tamaño de muestra en el diseño de una encuesta por muestreo probabilístico es una de las etapas más importantes, por lo que debe afrontarse con estricto apego a las consideraciones de carácter técnico de la teoría estadística, así como a los objetivos de la investigación y a los usos futuros de la información.

Resulta muy frecuente escuchar entre los especialistas en el tema que para calcular el número de unidades que formarán parte de la muestra hay que considerar la varianza de la variable de interés, así como la precisión con la que se desean obtener las estimaciones y la confianza requerida. Sin embargo, se pone muy poco énfasis en la importancia de tener presentes los objetivos de la encuesta, el tipo de variables e indicadores que se desean estimar, los dominios de estudio que se quieren analizar, así como el esquema de muestreo que se utilizará para seleccionar a las unidades de observación. Es cierto que el tamaño de muestra tiene un efecto sobre la varianza de las estimaciones (S^2/n); sin embargo, también se debe recordar que ésta, a su vez, depende en gran medida del esquema que se utilice para la selección de la muestra (Kish, 1965).

Es habitual que las encuestas que se realizan en la práctica sean de propósitos múltiples y por lo tanto se requiera la estimación de diversas estadísticas, las cuales, de manera individual, pueden conducir a tamaños de muestra y esquemas de selección diferentes. En este caso, la habilidad del especialista en muestreo conlleva a determinar el número de observaciones necesarias para cubrir los objetivos de un estudio multitemático, logrando generar un balance apropiado entre los costos de ejecución y la precisión deseada para los diferentes parámetros que se desean estimar. Sin embargo, se debe enfatizar que es muy difícil encontrar soluciones óptimas cuando las características que se desean investigar son demasiado dispares o las frecuencias observadas son muy bajas en los dominios de estudio de interés. Es obvio que las encuestas que se diseñan para estimar totales a niveles agregados y aquellas en donde se desean analizar subpoblaciones con características muy particulares requieren fracciones de muestreo muy diferentes.

En los textos que estudian la teoría del muestreo la determinación del tamaño de muestra para investigaciones multitemáticas se analiza, en primera instancia, como si se tratara de un problema univariado (Cochran, 1953 y Kish, 1965) sin considerar que prácticamente todas las encuestas son de propósitos múltiples. Así, en el caso particular de la determinación del tamaño de muestra en un estudio multipropósitos, cuando se desea estimar una proporción o porcentaje de una población -con función de densidad binomial- que posee ciertos atributos, se asume normalidad en la distribución de probabilidad del parámetro (p) y no se considera la corrección por población finita (cpf) (Cochran, 1953). Si

al aplicar este procedimiento se ignora que otras características de la población también serán estimadas a partir de los datos recabados, no será posible determinar la precisión para las variables analizadas en forma simultánea y muy probablemente se incrementará el error de muestreo.

Conforme a lo anterior, el objetivo de este trabajo es llamar la atención sobre la necesidad de asumir con solvencia técnica la etapa del cálculo del tamaño de muestra en las encuestas de propósitos múltiples y alertar sobre las limitaciones que se presentan en la confiabilidad de las estimaciones para dominios de estudio que no fueron considerados en los objetivos de la investigación. De manera particular, se señala la necesidad de evaluar la precisión de los indicadores que se obtienen y asumir con responsabilidad la formulación de hipótesis sobre el comportamiento de subpoblaciones muy específicas sobre las que no se puede garantizar que la inferencia estadística tenga validez.

II. TAMAÑO DE MUESTRA ÓPTIMO PARA DISTRIBUCIONES MULTINOMIALES: EL CASO UNIVARIADO PARA DISEÑOS MONOETÁPICOS

El cálculo del tamaño de muestra para estimar parámetros de proporciones con distribución multinomial es una situación cotidiana en el diseño de las encuestas que realizan las Oficinas Nacionales de Estadística¹.

De esta manera, el problema de determinar el número de observaciones necesarias para la estimación simultánea de proporciones multinomiales es equivalente a la construcción de intervalos de confianza simultáneos para variables con distribución multinomial, con la diferencia de que en el cálculo del tamaño de muestra los límites son fijados a priori por el investigador, a fin de controlar la probabilidad de que el intervalo contenga al verdadero valor del parámetro.

Cuando se realiza una encuesta para estimar parámetros de distribuciones multinomiales, el objetivo se debe centrar en calcular intervalos de confianza para cada una de las categorías de la variable. Sin embargo, en el trabajo cotidiano este hecho frecuentemente se pasa por alto, ya que en la práctica el procedimiento utilizado consiste en considerar a cada categoría - versus el resto-, como si se tratara de una variable binomial y se utiliza este hecho para determinar un conjunto de intervalos de confianza para cada una de las proporciones observadas en las celdas de manera independiente (Cochran, 1953)².

Esta manera de actuar no es del todo apropiada ya que este procedimiento no permite el cálculo del coeficiente de confianza para el conjunto de intervalos, puesto que

¹ Por ejemplo, la proporción de desempleados por categoría ocupacional en las encuestas de fuerza de trabajo o la distribución de mujeres en edad fértil por grupos de edad en las investigaciones demográficas.

² En el procedimiento propuesto por Cochran incluso se supone distribución normal para el parámetro p y se ignora la corrección por finitud (cpf) (Cochran, 1953 pág. 50).

únicamente establece consideraciones sobre los valores de las proporciones analizadas de manera individual³.

En Quesenberry y Hurts (QH, 1964) se propone un procedimiento para construir intervalos de confianza simultáneos para distribuciones multinomiales basados en una aproximación a la distribución chi-cuadrada (X^2) mediante el siguiente razonamiento.

Sean n_1, n_2, \dots, n_k las frecuencias observadas en una muestra de tamaño n con distribución de probabilidad multinomial y sean $\pi_1, \pi_2, \dots, \pi_k$ los parámetros asociados a la distribución. En este caso, n_i denota el número de observaciones clasificadas en la i -ésima celda de la distribución, mientras que π_i representa la probabilidad de que alguna observación caiga en la i -ésima celda ($i=1, 2, \dots, k$), y es evidente que $\pi_i \geq 0$ y $\sum \pi_i = 1$.

Conforme a lo anterior, QH propusieron los siguientes intervalos de confianza simultáneos para los k parámetros de las $\pi_1, \pi_2, \dots, \pi_k$:

$$\pi_i^i \leq \pi_i \leq \pi_i^s \quad (i=1, 2, \dots, k) \quad (1)$$

donde:

$$\pi_i^i = \{ A + 2n_i - \{ A [A + 4n_i (n - n_i) / n] \}^{1/2} \} / [2(n + A)]$$

representa el límite inferior y

$$\pi_i^s = \{ A + 2n_i + \{ A [A + 4n_i (n - n_i) / n] \}^{1/2} \} / [2(n + A)]$$

permite el cálculo del límite superior del intervalo.

En las expresiones anteriores A representa el porcentaje superior ($\alpha * 100$) de una distribución chi-cuadrada (X^2) con $(k - 1)$ grados de libertad y n el tamaño total de la muestra ($A \sim X^2_{(k-1, \alpha)}$). Así, cuando n tiende a infinito ($n \rightarrow \infty$), la probabilidad de que los k intervalos de confianza contengan el verdadero valor del parámetro será menor a $(1 - \alpha)$.

Las expresiones para los intervalos de confianza definidas en (1) fueron recomendadas para el caso en que la probabilidad de que las aseveraciones acerca de los verdaderos valores π_i son correctas y deben ser mayores o iguales a $(1 - \alpha)$; es decir, cuando la probabilidad de que la afirmación sea incorrecta es menor o igual que α .

En un trabajo posterior, Goodman (1965) mejoró el método de cálculo de QH y propuso otra metodología que genera intervalos de confianza más pequeños, basados en la aproximación de la densidad binomial a la distribución normal, utilizando la desigualdad de

³ Esta manera de proceder no es apropiada para distribuciones multinomiales con más de dos categorías ya que no es realista considerar siempre que todos los parámetros son iguales a $1/2$ ("worst case") y que la suma de los mismos es igual a la unidad (Thompson, 1987).

Bonferroni y el teorema del límite central para determinar la probabilidad de que los intervalos sean correctos en forma simultánea⁴.

Goodman demostró que los estimadores propuestos por QH se podrían obtener a partir de la solución de una ecuación cuadrática en π_i :

$$(p_i - \pi_i)^2 = A \pi_i (1 - \pi_i) / n \quad (i = 1, 2, \dots, k) \quad (2)$$

en donde $p_i = n_i / n$ y A tiene la misma interpretación que la sugerida por QH.

Cuando $k=2$ y $(n \rightarrow \infty)$, la probabilidad de que los intervalos de confianza definidos por (1) y (2) contengan al verdadero valor del parámetro es $(1 - \alpha)$; y corresponden a los que se construyen para estimar los parámetros de una distribución binomial.

Cuando se trata de una variable multinomial, y en particular cuando $k > 2$, Goodman(*op. cit.*) demuestra que las expresiones (1) y (2) se pueden mejorar reemplazando A por B , en donde B corresponde al valor máximo del $(\alpha / 2) * 100$ percentil de una distribución chi-cuadrada con un grado de libertad $(B \sim X^2_{(1, \alpha/2)})$ ⁵.

Conforme a lo anterior, la probabilidad de que el intervalo de confianza obtenido sea incorrecto para π_i es igual a α/k ; mientras que la probabilidad de que al menos uno de los k intervalos de confianza sea incorrecto es igual a α .

Mediante este procedimiento se obtienen intervalos más pequeños que los propuestos por QH, y además según Goodman (*op. cit.*) este es el procedimiento apropiado para garantizar que la probabilidad de las afirmaciones sobre la precisión del intervalo sean correctas.

Las expresiones propuestas por QH para determinar intervalos de confianza en una muestra aleatoria de tamaño n proveniente de una distribución multinomial con parámetros p_1, p_2, \dots, p_n son:

$$p_i = p_i \pm Z_{(\alpha_i/2)} [p_i (1 - p_i) / n]^{1/2}; \quad i = 1, 2, \dots, k. \quad (3)$$

los cuales contendrán al verdadero valor del parámetro en forma simultánea con probabilidad $(1 - \sum \alpha_i)$, y donde p_i representa la proporción observada en la i -ésima celda y $Z_{(\alpha_i/2)}$ es el $(1 - \alpha_i / 2) * 100$ percentil de una distribución normal estandarizada.

⁴ A pesar de que Goodman no trata de manera explícita la determinación del tamaño de muestra su procedimiento sentó las bases para el desarrollo de otras aplicaciones que se mencionarán más adelante.

⁵ De hecho, Goodman asegura que para $k > 2$ se demuestra que $B < A$ a los niveles de probabilidad usuales ($\alpha = .01, .05$ ó $.10$) y los valores para B pueden ser obtenidos usando las tablas de la distribución chi-cuadrada $(B \sim X^2_{(1, \alpha/2)})$. Es importante señalar que B también se puede obtener de una tabla con distribución normal ya que su valor corresponde al cuadrado del $(\alpha/2) * 100$ percentil de una distribución normal con media cero y varianza unitaria.

Posteriormente, Angers(1974) teniendo como referencia el trabajo de Goodman propuso un método gráfico para determinar tamaños de muestra para distribuciones multinomiales. Angers propuso que los puntos medios de los intervalos de confianza se pueden calcular utilizando la siguiente expresión:

$$d_i = Z_{(\alpha_i/2)} [p_i (1-p_i) / n]^{1/2} ; i= 1, 2, \dots , k \quad (4)$$

A partir de (4) es posible calcular los valores de las d_i 's para determinados niveles de confianza α_i 's. Sin embargo, Angers (*op. cit.*) sugiere que es muy tedioso determinar el valor óptimo de n para las d_i 's, si el objetivo es que los k intervalos $p_i = p_i \pm d_i$ contengan al verdadero valor del parámetro en forma simultánea con un nivel de significancia $\alpha = \sum \alpha_i$. Así, la propuesta consiste en sugerir que el tamaño de muestra se determine de manera gráfica sustituyendo $\sum \alpha_i$ por $\sum \alpha_i / 2$, por medio del siguiente procedimiento que se apoya en un razonamiento gráfico⁶: elija de manera arbitraria un tamaño de muestra n y calcule los k cocientes por medio de la expresión:

$$n d_i^2 / p_i (1 - p_i) ; i = 1, 2, \dots , k \text{ (categorías de la variable de diseño)} \quad (5)$$

que representan los valores de las abscisas, mientras que en el eje de las ordenadas se ubican los niveles de confianza ($0 \leq \alpha \leq .10$ y $0 \leq \alpha \leq .01$).

Posteriormente, se busca en las gráficas propuestas los valores obtenidos en el eje de las abscisas a fin de identificar los correspondientes niveles de confianza (α_i 's); y se compara la $\sum \alpha_i$ con el valor de α definido por el investigador, el criterio que se utiliza para decidir es que si la sumatoria ($\sum \alpha_i$) es mayor(menor) que α , entonces el tamaño de muestra propuesto es muy pequeño (grande), por lo que se deberá modificar el tamaño de muestra en múltiplos de n y continuar con el procedimiento descrito hasta encontrar un intervalo ($n_1 \leq n \leq n_2$) que contenga al valor buscado. Cuando se logre ubicar el intervalo, el número final de observaciones se obtiene por medio de interpolación lineal simple⁷.

1. El Cálculo del Tamaño de Muestra: Aspectos Prácticos

La relación de los procedimientos anteriores con la práctica cotidiana se puede ilustrar a partir del trabajo de Tortora (1978). Haciendo referencia a la manera en que Cochran(1953, pág. 50) aborda en su libro clásico la determinación del tamaño de muestra para estimar una proporción aplicando un esquema de selección aleatorio simple y para el

⁶ Los detalles de la aplicación del método pueden ser consultados en el trabajo de Angers(1974) en donde se proponen dos gráficas que en el eje de las abscisas ubican los valores de la expresión $n d_i^2 / p_i (1 - p_i)$ y en el de las ordenadas los correspondientes niveles de confianza α_i 's.

⁷ En el ejemplo presentado en el artículo de Angers (1974) se aplicó el procedimiento descrito y se determinó una muestra de 1,750 observaciones cuando se consideró una variable trinomial ($p_1 = .4$ y $p_2 = p_3 = .3$) y se definió un valor de $\alpha = .05$ y $d_1 = .05$ y $d_2 = d_3 = .025$.

caso univariado⁸, Tortora retoma la propuesta de Goodman(*op. cit.*), y plantea el cálculo del tamaño de muestra de la manera siguiente.

Suponga que una antropóloga desea investigar el tipo de sangre de los habitantes de una isla y quiere seleccionar una muestra de tamaño n para estimar el porcentaje de los que tienen sangre de tipo O. Así, según el procedimiento descrito por Cochran(*op. cit.*), el experto en muestreo debiera considerar a la variable(tipo de sangre O) como si se tratara de eventos Bernoulli y el efecto del tamaño de muestra sobre el valor del estimador se debe medir en términos del error estándar o del intervalo de confianza⁹. Sin embargo, de acuerdo al razonamiento de Tortora (*op. cit.*) suponga ahora que la antropóloga desea estudiar de manera simultánea la distribución de los distintos tipo de sangre de todos los habitantes de la isla.

Esto significa para el muestrista responder a la siguiente pregunta, ***qué tan grande debe ser la muestra para garantizar con suficiente precisión afirmaciones sobre la distribución de los porcentajes de los distintos tipos de sangre (A, B ,O y AB) de los habitantes de la isla.***

La pregunta anterior debiera ser la que los muestristas de las Oficinas Nacionales de Estadística se plantearan y respondieran durante la etapa de elaboración de los diferentes diseños de muestra que involucran distribuciones multinomiales, ya que es muy común que en las encuestas se definan objetivos como los que se plantea la antropóloga hipotética del ejemplo de Cochran¹⁰. Sin embargo, los hechos indican otra cosa, y claramente se enfrenta una inconsistencia entre el deber ser y la realidad, por lo que es posible que en algunas encuestas no se tengan observaciones suficientes para estimar con precisión parámetros de interés en subpoblaciones, con el agravante de que además en la divulgación, interpretación y utilización de los resultados existe poca preocupación de los usuarios por la precisión estadística de los datos y por supuesto las inferencias y decisiones de política que se plantean nunca hacen referencia al error de muestreo asociado a las estimaciones¹¹.

⁸ En caso de que se tratara de un esquema de selección polietápico el tamaño de muestra obtenido mediante muestreo aleatorio simple tendría que corregirse utilizando el factor denominado efecto de diseño (efd).

⁹ Esta es la manera en que se procede en la mayor parte de los diseños de las encuestas que se realizan en los países de la región, y sobre todo en aquellas investigaciones interesadas en medir el comportamiento de la fuerza de trabajo y en particular la estimación de la tasa de desocupación abierta.

¹⁰ De hecho, si a partir de los datos que generan las encuestas de hogares que estudian el comportamiento de la fuerza de trabajo se desean realizar estimaciones sobre la distribución de la población en diferentes niveles de pobreza por categoría ocupacional (indigentes, pobres y no pobres), en la definición del tamaño de muestra se debiera utilizar una variable multinomial y no la tasa de desocupación abierta como variable de diseño como ocurre en la práctica. Sin embargo, en este caso es importante señalar que para evaluar la magnitud de la pobreza se debiera utilizar como variable de diseño el ingreso medio de la población y no la proporción p (de pobres), ya que la estratificación de la población en los distintos niveles de pobreza se logra de manera indirecta toda vez que se determinó el monto de ingresos de los hogares.

¹¹ Se trabaja definiendo subclases y calculando indicadores como si la muestra hubiera estado diseñada para los fines específicos de la investigación. Existe abuso y una gran inconciencia por parte de los investigadores y diseñadores de política y un uso irracional de los resultados de las encuestas, ya que frecuentemente se hacen aseveraciones que probablemente carezcan de confiabilidad estadística.

El procedimiento propuesto por Tortora(*op. cit.*) parte de dividir a una muestra de tamaño n en k categorías exhaustivas y mutuamente excluyentes. Así, sea π_i la proporción de la población ubicada en la i -ésima categoría ($\sum \pi_i = 1$) y n_i la frecuencia observada en la i -ésima categoría de una muestra aleatoria. Para una precisión α deseada, se quiere determinar el conjunto de intervalos I_i ($i = 1, 2, \dots, k$) tal que:

$$\Pr \{ \bigcap_{i=1}^k (\pi_i \in I_i) \} \geq (1 - \alpha) \quad (6)$$

Es decir, se desea que la probabilidad de que todo intervalo I_i contenga al verdadero valor del parámetro π_i sea menor a $(1 - \alpha)$.

A partir del trabajo de Goodman(*op. cit.*) se presentan expresiones que permiten aproximar el cálculo de intervalos de confianza cuando $n \rightarrow \infty$:

$$\pi_i^i \leq \pi_i \leq \pi_i^s \quad (i=1, 2, \dots, k) \quad (7)$$

en donde:

$$\pi_i^i = \pi_i - [B \pi_i (1 - \pi_i) / n]^{1/2}$$

$$\pi_i^s = \pi_i + [B \pi_i (1 - \pi_i) / n]^{1/2}$$

representan el límite inferior y superior y B es el percentil superior $(\alpha / 2 * 100)$ de una distribución X^2 con 1 grado de libertad ($B \sim X^2_{(1, \alpha/2)}$).

El análisis de las ecuaciones dadas en (7) revelan que $[\pi_i (1 - \pi_i) / n]^{1/2}$ es la desviación estándar de la i -ésima celda de una población multinomial, y además se debe recordar que las funciones marginales de probabilidad corresponden a una densidad binomial. Así, si N es el tamaño total de la población y se incorpora el factor de corrección por población finita (cpf) y la varianza de cada π_i , se tiene una aproximación a los intervalos de confianza (Cochran, 1953).

$$\pi_i^i = \pi_i - [B (N - n) \pi_i (1 - \pi_i) / (N - 1) n]^{1/2}$$

$$\pi_i^s = \pi_i + [B (N - n) \pi_i (1 - \pi_i) / (N - 1) n]^{1/2} \quad (8)$$

Considerando que cuando $N \rightarrow \infty$ las expresiones anteriores convergen a (7).

Para la determinación del tamaño de muestra se requiere definir la precisión de cada parámetro de la distribución multinomial¹². De esta manera, suponga que se desea una precisión absoluta b_i para cada celda; entonces, a partir de (7) se tiene que:

¹² Esta situación representa una diferencia sustantiva respecto al procedimiento tradicional en donde generalmente se elige una variable de diseño y sobre ella se determina el número de observaciones necesarias para realizar la investigación.

$$\begin{aligned}\pi_i - b_i &= \pi_i - [B \pi_i (1 - \pi_i) / n]^{1/2} \\ \pi_i + b_i &= \pi_i + [B \pi_i (1 - \pi_i) / n]^{1/2}\end{aligned}\quad (9)$$

Despejando el valor de b_i en las ecuaciones (9):

$$b_i = [B \pi_i (1 - \pi_i) / n]^{1/2} \quad (10)$$

y resolviendo para n se obtiene que el tamaño de muestra necesario para estimar cada celda con una precisión b_i es igual a:

$$n = \text{máx}_i \{ B \pi_i (1 - \pi_i) / b_i^2 \} \quad (11)$$

En caso que se desee incorporar la corrección por finitud (cpf) (11) se convierte en:

$$n = \text{máx}_i \{ B N \pi_i (1 - \pi_i) / [b_i^2 (N - 1) + B \pi_i (1 - \pi_i)] \} \quad (12)$$

Conforme a lo anterior, para decidir sobre qué tamaño de muestra resulta más apropiado se deben calcular los k pares (b_i, π_i) y seleccionar el valor de n que resulte mayor dejando el valor de B fijo en cada caso.

Es importante observar que en (11) y (12) el tamaño de muestra es una función de π_i y de b_i , por lo que n decrece cuando $\pi_i \rightarrow 1/2$ y también lo hace en el caso en que $b_i \rightarrow 0$. En la situación en que $b_i = b \forall i$, sólo se requiere hacer un cálculo y se sugiere que este sea con la π_i más cercana a $1/2$.

Es frecuente que en la práctica se desconozca un valor aproximado de la proporción π_i que se desea estimar; por lo que se puede suponer -como frecuentemente se hace- que $\pi_i = 1/2$ y $b_i = b \forall i = 1, \dots, k$, por lo que $n = B / 4 b^2$.

2. Ejemplo de Cálculo

A fin de ejemplificar el procedimiento propuesto por Tortora, continuemos desarrollando el ejemplo hipotético de la antropóloga que desea estimar las proporciones de habitantes con tipos de sangre A, O, B y AB. Para este fin, supóngase que de un estudio previo se sabe que la distribución de los tipos de sangre en una población similar fue la que se presenta en el cuadro 1.

Cuadro 1

DISTRIBUCIÓN DEL TIPO DE SANGRE

TIPO DE SANGRE	PORCENTAJE
A	27
B	19
O	43
AB	11

Además, suponga que se desea obtener una precisión absoluta de $\pm 5\%$ para cada proporción y un coeficiente de confianza de .95 ($\alpha = 5\%$). Según la notación utilizada por Tortora, en este caso $b_i = .05$ ($i = 1, 2, 3$ y 4) ya que se requiere la misma precisión para cada celda y $\alpha = .05$ ($(1 - \alpha) = .95$), y como supuesto adicional se asume que la población es suficientemente grande por lo que se ignora la cpf.

Debido a que $b_i = b \forall i$, entonces sólo debemos efectuar un cálculo utilizando la proporción que se encuentre más cerca de .5¹³, y que en este caso corresponde a la proporción de la población con tipo de sangre O. Así, utilizando la expresión (11) y sustituyendo de forma conveniente:

$$n = B \pi_i (1 - \pi_i) / b_i^2 = 6.5381 * 0.43 (1 - 0.43) / (.05)^2 = 641$$

se obtiene que el tamaño de muestra requerido es de 641 personas. En este caso, $k=4$ y B es igual a 6.5381, que corresponde al valor en tablas de una distribución $X^2_{(1, \alpha/k=0.0125)}$ con un grado de libertad para un valor de $\alpha/k = 0.0125$

(1.25%); ya que $B \sim X^2_{(1, \alpha/k)}$ ¹⁴.

Por otra parte, si el problema se resolviera de la manera tradicional se tendrían que calcular tres tamaños de muestra diferentes -uno para cada valor de p - suponiendo en cada caso una distribución binomial (normal) del parámetro de interés¹⁵.

¹³ En la literatura a este valor se le conoce como "el peor de los casos" (worst case) y corresponde a un valor de $p=.5$ que representa el punto en el que se maximiza la varianza (pq) de una distribución binomial con parámetro p .

¹⁴ En el ejemplo presentado en Cochran se asume un valor de $t^2 \cong 4$ y un error absoluto de $\pm 5\%$. De esta forma, para un valor de $.30 \leq p \leq .60$ el tamaño de muestra necesario ($n = t^2 p q / d^2 = 4 p q / 25$) se encontraría entre $336 \leq n \leq 400$ y según la práctica habría que optar por seleccionar 400 personas para garantizar de manera conservadora la precisión buscada.

¹⁵ Ya se comentó que Cochran(1953) propone como aproximación a la solución del problema considerar cada celda como una distribución binomial y obtener un tamaño de muestra para cada una de las proporciones de interés. El valor más grande que resulte de estos cálculos será el que deberá elegirse para los fines de la encuesta.

Así, considere la popular expresión para el cálculo del tamaño de muestra para una proporción binomial:

$$n' = t^2 p q / d^2 \quad (13)$$

en donde t representa la abcisa de la curva normal para una confianza α determinada, mientras que d es el error máximo absoluto con el que se desean obtener las estimaciones. De esta manera, con $\alpha=.05$ los tamaños de muestra calculados para cada una de las proporciones requeridas son:

$$n_A = 303 ; n_B = 236 ; n_O = 377 \text{ y } n_{AB} = 150$$

Según el criterio definido por Cochran se debiera escoger el valor máximo y seleccionar 377 personas; sin embargo, observe que este valor representa tan sólo el 58.8% del total de observaciones calculadas mediante el método propuesto por Tortora. Es decir, se estaría seleccionado una muestra 41.2% menor de lo que realmente se necesita. Ante esta situación, es evidente que se incrementa el error de muestreo cuando se intentan desagregar las observaciones para analizar características de interés en dominios de estudio específicos con baja frecuencia de aparición.

La comparación de ambas fórmulas se logra por medio del cociente de las expresiones (11) y (12) el cual indica que:

$$n / n' = B / t^2 \quad (14)$$

Observe que la relación varía de acuerdo al valor que asume $B \sim X^2_{(1,\alpha/k)}$, el cual a su vez depende de la precisión requerida (α) y del número de categorías en que se distribuye la variable de interés (k). De esta manera, es posible construir una tabla en donde se establezca la relación que existe entre el tamaño de muestra de una distribución binomial y aquel que se obtiene por medio de intervalos de confianza simultáneos para distribuciones multinomiales.

¹⁶ En esta situación se asume que se trata de un esquema de muestreo aleatorio simple de una sola etapa, por lo que el tamaño de muestra no se afecta por el efecto de diseño (efd) y además no se involucra el ajuste por no respuesta. Sin embargo, el tamaño de muestra siempre se debe ajustar en el caso de una encuesta real que suponga un esquema de selección polietápico en el cual se prevé la pérdida de observaciones por diferentes causas.

Cuadro 2

RELACIÓN ENTRE EL TAMAÑO DE MUESTRA BINOMIAL Y DE INTERVALOS DE CONFIANZA SIMULTÁNEOS MULTINOMIALES¹

CATEGORÍAS	3 ²	4	5	6	7	8	9	10
CONFIANZA α								
.010	1.30	1.38	1.44	1.48	1.54	1.56	1.60	1.63
.025	1.39	1.49	1.57	1.63	1.69	1.74	1.78	1.82
.050	1.49	1.62	1.73	1.81	1.89	1.94	2.00	2.05
.075	1.58	1.74	1.87	1.97	2.05	2.13	2.20	2.26
.100	1.68	1.86	2.00	2.12	2.22	2.31	2.38	2.45

¹ En las celdas se reportan los valores del cociente $n / n' = B / t^2$ y no se considera la corrección por población finita (cpf), además de que se asume la misma precisión absoluta para cada una de las categorías.

² Cuando $k=2$, se trata de una distribución binomial y el tamaño de muestra se determina de la manera tradicional propuesta por Cochran.

La interpretación del cuadro 2 se debe hacer de la manera siguiente. Supóngase que se desean estimar las proporciones de una cierta variable con distribución de probabilidad multinomial formada por 3 categorías y se quiere garantizar una confianza de $(1 - \alpha) = 95\%$ y la misma precisión para todas las celdas. En este caso, la fórmula propuesta por Tortora indicaría que se tendría que seleccionar una muestra 49% mayor a la que se determinaría si se aplicara la expresión tradicional que se presenta en Cochran(1953). Esto significa que (13) subestimaría el verdadero tamaño de la muestra necesario para obtener la precisión deseada en cada una de las tres categorías en que se clasifica la variable de interés a una confianza del 95%.

El método de Tortora representa la manera correcta de proceder para determinar el tamaño de muestra de una proporción multinomial bajo un esquema de selección aleatorio simple y cuando se utiliza sólo una variable de diseño; sin embargo, en Angers (1979) se mejora esta propuesta ya que el autor considera que en ocasiones los resultados generados por la expresión (11) son muy conservadores y estiman tamaños de muestra innecesariamente grandes.

Angers (*op. cit.*), basa su procedimiento usando para el cálculo un valor muy cercano a .5 para cada parámetro mediante la siguiente fórmula:

$$n = \min_{\alpha_i} \max_i \{ B_i \pi_i (1 - \pi_i) / b_i^2 ; i = 1, \dots, k \} \quad (15)$$

tal que $\alpha_i \leq a_i \leq \alpha$ y $\sum_{i=1}^k \alpha_i \leq \alpha$. En esta propuesta, $B \sim X^2_{(1-\alpha/k)}$ tiene la misma interpretación que en el método de Tortora y las a_i 's = α / k y el valor de α son

propuestos por el investigador. Así, la probabilidad de que el i -ésimo intervalo de confianza contenga al verdadero valor del parámetro t_i es al menos $(1 - \alpha_i)$, mientras que $(1 - \alpha)$ representa la probabilidad de que los k intervalos sean correctos de manera simultánea¹⁷.

Según los resultados de las aplicaciones que presenta Angers en su investigación, esta forma de calcular el tamaño de muestra es más robusta y puede llegar a significar ahorros de casi un 12% en el número de observaciones a seleccionar en relación al método de Tortora.

Thompson (1987) hace una revisión de las investigaciones realizadas y concluye que el método propuesto por Angers(1984) supera las propuestas anteriores pero resulta muy tedioso en su aplicación. En este sentido, el autor propone una manera de determinar el "peor de los casos" (worst case) para un vector de parámetros multinomiales cuando se desean obtener intervalos de confianza simultáneos para cada uno de los componentes del vector π .

Thompson plantea que el objetivo consiste en determinar el tamaño de muestra n para una variable aleatoria de una distribución multinomial, de tal forma que la probabilidad de que todas las proporciones estimadas de manera simultánea estén contenidas en el intervalo sea menor que $(1 - \alpha)$; esto es,

$$\Pr \{ \bigcap_{i=1}^k | p_i - \pi_i | \leq d_i \} \geq 1 - \alpha \quad (16)$$

en donde π_i es la proporción de observaciones en la i -ésima categoría en la población, p_i la proporción observada en la muestra y k el número de categorías y

$$\alpha_i = \Pr \{ |Z_i| \geq d_i \sqrt{n} / \sqrt{\pi_i (1 - \pi_i)} \} = 2(1 - \phi(Z_i)) \quad (17)$$

en donde Z_i es la variable normal estandarizada, ϕ la función acumulativa de probabilidad y $Z_i = d_i \sqrt{n} / \sqrt{\pi_i (1 - \pi_i)}$ ¹⁸.

A fin de obviar la corrección por finitud y lograr la normalidad del estimador, se asume que la población es lo suficientemente grande y que el muestreo se hace sin reemplazo. En la expresión (16) y siguiendo a Angers (1974)¹⁹, se define las d_i como:

¹⁷ Bajo esta óptica el procedimiento de Tortora se puede considerar un caso particular del propuesto por Angers cuando $\alpha_i = \alpha / k$.

¹⁸ Cuando $k=2$ y $d_1=d_2=d$ se trata de una distribución binomial y el tamaño de muestra se determina de la manera tradicional ($n = z^2 \pi_i (1 - \pi_i) / d_i$). Si la proporción π_i es desconocida, se utiliza el criterio de máxima varianza (worst case) con $\pi_i = .5$.

¹⁹ Del análisis realizado al artículo de Thompson(1987) asumimos que esa debe ser la definición de las d_i , ya que el autor no hace explícito en su investigación la expresión para calcularlas.

$$(18) \quad d_i = Z_{(\alpha_i / 2)} [p_i (1 - p_i) / n]^{1/2}$$

en donde $Z_{(\alpha_i / 2)}$ es el percentil $(1 - \alpha_i/2)*100$ de una distribución normal estandarizada²⁰.

Como resultado de esta investigación, Thompson demuestra que para un determinado nivel de significancia α_i de las d_i 's (excepto cuando $\sum_i^k = \alpha$), el vector de parámetros de una distribución multinomial para determinar "el peor de los casos" se maximiza para $\sum \alpha_i = \alpha$, cuando $\pi_i = 1/m$ para un número determinado de categorías y $\pi_i = 0$ para el resto; es decir,

$$p \pm Z_{(\alpha/2)} [(1/m)(1 - 1/m) / n]^{1/2}, i = 1, \dots, k \quad (19)$$

en donde los valores de m varían conforme a k y α ²¹.

Conforme a lo anterior, y suponiendo que $d_i = d \forall i=1, \dots, k$ y sin tener ninguna información acerca de la forma en que se distribuye la variable en las distintas categorías, Thompson (*op. cit.*) propone la siguiente expresión para el cálculo del tamaño de muestra:

$$n = \max_m Z^2 (1/m) (1 - 1/m) / d^2 \quad (20)$$

En donde $Z = \phi(1 - \alpha/2m)$ es el límite superior del $(\alpha/2m)$ percentil de una distribución normal y $m \in Z^+$.

Thompson sugiere que el tamaño de muestra se puede determinar a partir del cuadro 3 dividiendo $d^2 n / d^2$, cuando el investigador especifica un nivel de confianza α y una distancia (error de estimación) d .

Para ilustrar el uso del cuadro 3 suponga que se desea determinar el tamaño de muestra para realizar una encuesta que pretende estudiar el comportamiento de las mujeres en edad fértil por grupos de edad en función al número de hijos nacidos vivos. Así, si se desea una confianza de 95% para todas las categorías, entonces se fija $\alpha=5\%$ y en el cuadro 3 se observa que se deberían seleccionar un total de 510 mujeres ($d^2 n=1.27359$), mientras que en la última columna se estaría indicando el número de parámetros distintos de cero. Es

²⁰ Este resultado se justifica a partir de la aplicación del teorema del límite central y la desigualdad de Bonferroni (Angers, 1989).

²¹ Durante la presente década se han realizado otros desarrollos que han mejorado los procedimientos de cálculo aquí presentados. Así, Sison y Glaz(1995) propusieron un algoritmo de cálculo para intervalos de confianza y determinación del tamaño de muestra para proporciones multinomiales que supera en precisión a los anteriores. Sin embargo, su aplicación práctica es muy compleja ya que requiere el cálculo de momentos de tercero y cuarto orden para distribuciones Poisson truncadas y por este motivo no se analizó en este trabajo.

importante notar que no se requiere conocer a priori el número de grupos de edades que se formarán, lo cual significa que bajo esta manera de proceder no es necesario saber el número de categorías(k) en que se divide la población.

Cuadro 3

**Tamaño de Muestra para Estimar en forma Simultánea
los parámetros de una distribución Multinomial
con una distancia "d" del verdadero valor
con un nivel α de significancia**

α	d^2n	ncon $d=.05$	m
.50	.44129	177	4
.40	.50729	203	4
.30	.60123	241	3
.20	.74739	299	3
.10	1.00635	403	3
.05	1.27359	510	3
.025	1.55963	624	2
.02	1.65872	664	2
.01	1.96986	788	2
.005	2.28514	915	2
.001	3.02892	1212	2
.0005	3.33530	1342	2
.0001	4.11209	1645	2

Reproducida de Thompson (*op. cit.*)

A partir del análisis realizado se puede concluir que cuando se desean estimar parámetros de distribuciones multinomiales y garantizar la precisión simultánea de todos ellos, la expresión propuesta por Cochran no resulta apropiada ya que subestima de manera importante el número de observaciones necesarias para garantizar la precisión deseada en todas las categorías en las que se distribuye la variable de interés.

Asimismo, es importante enfatizar que en la práctica es habitual que se deseen estimar proporciones multinomiales por lo que se aconseja aplicar procedimientos robustos que están disponibles y que superan a los que se utilizan de manera tradicional para determinar el tamaño de muestra ($n=t^2pq/d^2$).

III. LA DETERMINACIÓN DEL TAMAÑO DE MUESTRA EN DISEÑOS POLIETÁPICOS Y MULTITEMÁTICOS

Hasta ahora, en los desarrollos presentados además de los supuestos en los que se basan los métodos se asume que la selección de las unidades de observación se realiza en una sola etapa y se trabaja como si sólo se deseara estimar una característica de la población. Es decir, se supone un esquema de muestreo monoetápico y aleatorio simple (MAS), en donde únicamente existe una variable de interés (proporción, media o total). Sin embargo, en la práctica no es común que se presenten estas situaciones ya que en la mayoría de los diseños de muestra al menos se requieren dos etapas para seleccionar a las unidades de observación, y prácticamente todas las encuestas se pueden considerar de propósitos múltiples.

Lo anterior significa que a los problemas derivados de la falta de aplicación de un procedimiento apropiado para la determinación del tamaño de muestra se agrega el hecho de que el número de observaciones a seleccionar se debe incrementar debido a la necesidad de formar estratos y conglomerados para la selección de las unidades primarias de muestreo (UPM's), lo cual aumenta la varianza de los estimadores.

En la práctica, el problema de seleccionar la muestra en varias etapas y formar estratos y conglomerados para facilitar la identificación de las unidades de observación, se resuelve corrigiendo el tamaño de muestra obtenido mediante muestreo aleatorio simple por un factor denominado efecto de diseño (**efd**) que relaciona el coeficiente de correlación intraconglomerados con el tamaño promedio de las unidades de segunda etapa²².

En Cochran (*op. cit.*), se señala que es habitual que en la planeación de las encuestas las organizaciones ejecutoras estén interesadas en estudiar múltiples características de la población, por lo que propone que para determinar el tamaño de muestra se fije el error máximo aceptable para cada una de las características que se consideren más importantes y se calcule de manera independiente el número de observaciones requeridas para cada caso.

Posteriormente, se sugiere comparar los diversos valores obtenidos y decidir al respecto: si éstos son similares y el tamaño de muestra más grande está acorde con el presupuesto destinado para la realización de la encuesta, se opta por este número de observaciones el cual garantizará suficiente precisión para todas las variables de interés. Sin embargo, la práctica cotidiana revela que esta situación es poco frecuente y en muchas

²² El efecto de diseño fue propuesto por Kish(1965) y se define por $efd = [1 + \rho (m^* - 1)]$ en donde ρ representa el coeficiente de correlación intraconglomerados y m^* el tamaño medio del conglomerado, los cuales se espera que sean de aproximadamente iguales. Es habitual que el

$efd > 1$ por lo que el tamaño de muestra para un diseño estratificado y polietápico (n_c) en general es mayor que el que se obtiene por muestreo aleatorio simple (n_{mas}) y que comúnmente se denomina tamaño efectivo de muestra ($n_e = n_c / efd$). Así, se tiene que $[n_c = n_{mas} (1 + \rho (m-1))]$.

ocasiones los valores son tan discordantes que es necesario renunciar a investigar algunas características de la población²³.

Es difícil que los diseñadores de encuestas con poca experiencia -y sobre todo los usuarios de la información- comprendan esta situación, y por lo tanto es frecuente que se decida investigar características de la población en donde el tamaño de muestra utilizado no está en condiciones de estimar con la precisión adecuada²⁴, sin que exista preocupación por evaluar la magnitud del error de estimación.

Asimismo, cuando se desean analizar subclases con características poco frecuentes es difícil estimar el tamaño de muestra ya que la distribución de la variable se conoce a posteriori y es muy probable que se incremente el error de muestreo de manera considerable (Yates, 1953). Esta situación es cotidiana en las encuestas de hogares (empleo, niveles de vida y presupuestos familiares) sin que los usuarios de la información intuyan que la calidad de los resultados que analizan y las recomendaciones de política que formulan pueden carecer de validez estadística²⁵ (CEPAL, 1998).

De esta manera, los problemas de subestimación del tamaño de muestra se agravan con la consiguiente pérdida de precisión e incremento en el error de muestreo y sesgo de los estimadores²⁶.

En los textos sobre la teoría del muestreo se trata ampliamente el tema de la determinación del tamaño de muestra para diseños que requieren obtener el número óptimo de selecciones de primera y segunda etapa para un presupuesto determinado o una varianza deseada, y además el interés del investigador se centra en estudiar una sola variable (Cochran, 1953, Snedecor y Cochran, 1980). Sin embargo, la mayoría de las

²³ No sólo hay que analizar el problema de las encuestas de propósitos múltiples desde la perspectiva del tamaño de muestra, también es importante considerar el esquema de muestreo y la definición del marco muestral. Asimismo, cuando se desean obtener estimaciones por dominios de estudio con una varianza (V) para cada subdivisión (k), el tamaño de muestra total puede aproximarse a k veces el necesario para una estimación global con la misma precisión, lo cual frecuentemente se pasa por alto en las unidades ejecutoras de encuestas.

²⁴ En Kish (1965) se señala que existen factores que limitan el desarrollo de encuestas con objetivos múltiples y en ocasiones es imposible llegar a soluciones que satisfagan todas las pretensiones de los usuarios, ya que una encuesta nacional puede requerir de una fracción de muestreo diferente a la que sería necesario aplicar cuando se desea investigar una subpoblación de interés en un nivel geográfico menor.

²⁵ En el análisis de datos es frecuente que se plantee el problema del tamaño de muestra mínimo para poder garantizar que se cumpla la aproximación a la distribución X^2 de un estadístico que se forma a partir de los datos observados y esperados en una tabla de contingencia. En este sentido, se habla de un mínimo de observaciones por celda (5, 10, 20 o 30) que se debe determinar y para lo cual no existe consenso. Un criterio analítico para determinar los mínimos necesarios que considera las proporciones en que se distribuye la variable se puede encontrar en Eaton (1978) en donde se hace una propuesta y se establece una regla de carácter práctico.

²⁶ De hecho, la precisión de la información se mide a partir del cálculo del error de muestreo; sin embargo, habitualmente se ignora la magnitud del sesgo de los resultados ya que muy pocas veces éste se puede evaluar de modo que los datos obtenidos se asumen como correctos.

encuestas requieren la estimación simultánea de varias variables con una precisión adecuada y en este caso los procedimientos para calcular el tamaño de muestra óptimo no están bien definidos.

1. El Planteamiento Matemático

Ya se señaló que la manera tradicional de proceder consiste en estimar el tamaño óptimo para cada variable -utilizando la expresión propuesta por Cochran y posteriormente decidirse por un valor que de alguna manera signifique una solución que satisfaga los requerimientos en forma conjunta (denominada solución de compromiso); sin embargo, esta no es la forma más adecuada de resolver el problema ya que frecuentemente no se logra obtener la solución óptima. De hecho, cuando se trata de estimar de manera simultánea un determinado número de características ($p \geq 2$) no existe un criterio único para la determinación del tamaño de muestra óptimo (Kokan, 1963), por lo que actuar de la manera tradicional no garantiza la confiabilidad requerida en las diversas variables de interés.

No existen muchas investigaciones que traten este tema y es preciso señalar que la solución matemática para determinar el número óptimo de selecciones de primera y segunda etapa cuando el interés se centra en la estimación simultánea de varias variables, significa minimizar una función de costos no lineal sujeta a restricciones no lineales y varianza constante lo cual puede resultar bastante complejo.

Si el análisis se centra en una variable el procedimiento para determinar el tamaño óptimo de muestra para diseños de una etapa está bien documentado en la literatura. Es decir, cuando se desea determinar el número óptimo de selecciones que minimice el costo de la encuesta para una varianza deseada ($V = a_0 + \sum_i^k a_i/m_i$) o minimizar la varianza del estimador para un presupuesto determinado ($C = c_0 + \sum_i^k c_i m_i$); en donde mas m_i 's son función del número de selecciones y $k \in \mathbf{Z}^+$ depende del procedimiento de muestreo seleccionado; c_i es el costo por muestrear la unidad m_i y a_i no depende de m_i pero si de su variabilidad.

Así, Kokan (*op. cit.*) demuestra que para un costo C fijo el valor mínimo de la varianza se obtiene por medio de:

$$V_{\min} = a_0 + (\sum_i^k \sqrt{a_i c_i})^2 / (C - c_0) \quad (21)$$

Por otra parte, cuando la varianza V es fija el valor mínimo de C se obtiene por medio de:

$$C_{\min} = c_0 + (\sum_i^k \sqrt{a_i c_i})^2 / (V - a_0) \quad (22)$$

y los tamaños de muestra óptimos son:

$$m_i = \sqrt{a_i/c_i}(C - c_0) / \sum_i^k \sqrt{a_i c_i} ; (i=1, \dots, k) \text{ (para } C \text{ fijo)} \quad (23)$$

$$m_i = \sqrt{(a_i/c_i) (\sum_i^k \sqrt{(a_i c_i)} / (V - a_0))} ; (i=1, \dots, k) \text{ (para } V \text{ deseada)} \quad (24)$$

En la literatura no existe una solución tan simple cuando se trata de una encuesta que estudia varias características de manera simultánea, ya que las a_i 's cambian y es probable que los costos c_i también dependiendo de la variable de estudio, y por lo tanto no es posible encontrar un método único que minimice la varianza de todas las variables en forma simultánea para un presupuesto determinado.

Existen algunas propuestas para solucionar esta situación (Chakravarthy, 1955 y Ghosh, 1958) que han planteadas por Kokan (*op. cit.*) como un problema de optimización de una función sujeta a restricciones mediante técnicas de programación no lineal de la manera siguiente²⁷.

2. El Criterio de Optimalidad

Suponga que se dispone de un determinado presupuesto C para la realización de una encuesta en la cual se desean investigar p características de una población de interés. Es bien sabido que no se puede obtener un valor exacto en las estimaciones por lo que se está dispuesto a aceptar un cierto margen de error el cual dependerá de la característica de estudio, y para el cual se puede fijar una cota superior a la varianza (V) con algún nivel de confianza α .

Así, un vector $\mathbf{m}=(m_1, m_2, \dots, m_k)$ se dice que es óptimo si minimiza el costo de la investigación sujeto a la condición de que la varianza V_j de la j -ésima característica de interés no sea mayor a una cantidad v_j prefijada, para un nivel de confianza $1-\alpha_j$ ($0 < \alpha_j < 1$), $j=1, \dots, p$, y en donde p representa el número de variables a estudiar.

Siguiendo a Kokan (*op. cit.*) la solución general se puede plantear de la manera siguiente. Considere una población de tamaño N (U_1, U_2, \dots, U_N) en donde se desean medir p características variables. Se determina un esquema de selección y se extrae una muestra de tamaño n y se fija la varianza V_j para la j -ésima variable. De esta manera, el problema de asignación óptima de la muestra consiste en minimizar una función de costos C que depende del vector \mathbf{m} sujeta a las restricciones de varianza (V_j), lo cual se plantea de la manera siguiente:

$$\text{Mín } C = \phi(\mathbf{m}) \quad (25)$$

$$\text{s.a. } V_j \leq v_j \quad ; (j=1, \dots, p) \quad (26)$$

$$0 \leq m_i \leq M_i \quad ; (i=1, \dots, k) \quad (27)$$

²⁷ Existen referencias del uso de programación lineal para resolver el problema en los trabajos de Dalenius(1953) y Nordbotten(1956) y de métodos no lineales en Dalenius(1957).

Si existe un vector \mathbf{m} que satisface las $(p+2k)$ restricciones impuestas por (26) y (27) se dice que existe una solución factible. Además, si esa solución minimiza la función objetivo (25) se obtiene un mínimo local óptimo²⁸.

3. Formulación Específica: Muestreo Aleatorio Simple Estratificado

Suponga que se forman L estratos y que n_h unidades son seleccionadas sin reemplazo del estrato h , y sin pérdida de generalidad suponga que se está interesado en estudiar la media de la población \bar{Y}_j . Así, un estimador insesgado de la media poblacional \bar{Y}_j es $\bar{y}_{j\text{est}}$ el cual tiene una varianza muestral igual a :

$$V_j = \text{var}(\bar{y}_{j\text{est}}) = \sum_h^L a_{hj} x_h \quad (28)$$

en donde $a_{hj} = W_h^2 S_{hj}^2 \quad (h=1, \dots, L; j=1, \dots, p) \quad (29)$

y $x_h = (1/n_h) - (1/N_h) ; (h=1, \dots, L) \quad (30)$

Si c_h es el costo de evaluar las p características de interés en una unidad cualquiera del h -ésimo estrato, entonces el costo total de la encuesta se determina por medio de la siguiente expresión :

$$C = c_0 + \sum_h^L c_h n_h \quad (31)$$

en donde c_0 es un costo fijo. Si se aplican (29) y (30) en (28), la función de costo se transforma en:

$$C = c_0 + \sum_h^L N_h c_h / (1 + N_h x_h) \quad (32)$$

Por consideraciones de carácter práctico en cada uno de los estratos el tamaño de muestra debe ser mayor o igual que uno (si fuera igual a cero no tendría sentido formar el estrato) y menor al tamaño del estrato, lo cual implica que:

$$0 \leq x_h \leq 1 - (1/N_h) \quad (33)$$

La situación ahora consiste en determinar los valores n_1, \dots, n_L con objeto de minimizar (31) sujeto a $V_j \leq v_j$ y a (33), para un costo c_0 fijo que puede no ser considerado en la solución del problema. **Minimizar C es lo mismo que maximizar $-C$** por lo que se formula el siguiente planteamiento:

²⁸ Kokan(op.cit.) señala que se puede transformar la función objetivo en una función convexa y las restricciones se pueden expresar en forma lineal, lo cual conlleva a que sea posible determinar una solución factible y que esta sea la óptima.

$$\text{Max } \phi = - \sum_h^L N_h c_h / (1 + N_h x_h) \quad (34)$$

$$\text{s.a. } \sum_h^L a_{hj} x_h \leq v_j \quad (j=1, \dots, p) \quad (35)$$

$$0 \leq x_h \leq 1 - 1/N_h \quad (h=1, \dots, L) \quad (36)$$

Observe que (34) es una función no lineal y convexa, mientras que (35) y (36) son lineales en las x_h 's, y los valores V_j deben ser fijados por el investigador.

4. Aplicación del Método

Suponga que se desea realizar una encuesta para estimar el volumen de la producción (X_1) y del personal ocupado (X_2) de la industria manufacturera en un país determinado. Se dispone de un directorio de establecimientos elaborado a partir del Censo Económico más reciente y se cuenta con información sobre el tamaño del establecimiento lo cual, a partir de la aplicación de un procedimiento multivariado apropiado, permite la formación de tres estratos como se muestra a continuación:

Cuadro 4

**Clasificación de los Establecimientos de la Industria
Manufacturera por Tamaño**

Tamaño del Establec.	h	No. Estab. N_h	Varianza de la producción $(S_{h1}^2) X_1$	Varianza del no. de Empleados $(S_{h2}^2) X_2$	W_h	W_h^2
Chicos	1	800	4,000	12	.2857	.0816
Medianos	2	1,500	46,500	35	.5357	.2870
Grandes	3	500	500,000	250	.1786	.0319
Total		2,300			1.0000	

Las varianzas para los estimadores del total X_1 y X_2 se obtienen por medio de:

$$V_1 = \text{var}(x_1) = \sum_{h=1}^3 N_h^2 S_{h1}^2 x_h \quad \text{y} \quad V_2 = \text{var}(x_2) = \sum_{h=1}^3 N_h^2 S_{h2}^2 x_h$$

$$x_1 = 1/n_1 - (1/N_1) = 1/n_1 - (1/800) = 1/n_1 - .00125$$

$$x_2 = 1/n_2 - (1/N_2) = 1/n_2 - (1/1,500) = 1/n_2 - .000666$$

$$x_3 = 1/n_3 - (1/N_3) = 1/n_3 - (1/500) = 1/n_3 - .0020$$

Se desea que la diferencia entre el valor real y el estimador de la producción total (x_1) no sea mayor a 500,000, y que la diferencia entre el total de empleados estimados (x_2) y el total real se ubique alrededor de las 1,200 unidades. Además, se quiere que el **error de estimación** no sea mayor al 5% con un nivel de probabilidad $p=0.95$. De esta forma, suponiendo normalidad de la variable se tiene que;

$$\Pr \{ |y_j - Y_j| \leq .05 Y_j \} = \Pr \{ |y_j - Y_j| \leq S.E(y_j) * 1.96 \} = .95 \quad j = 1, 2.$$

Además, para los fines del estudio se requiere que las varianzas de las estimaciones sean: $V_1 \leq 150'000,000$ ($D.S(y_1)=12,247.45$) y $V_2 \leq 80,000$ ($D.E(y_2)=282.84$), y sin incorporar restricciones adicionales se asume que no existen diferencias en los costos de entrevista en los diferentes estratos. De esta forma, se tiene que $C = c_0 + c \sum_{h=1}^3 n_h$ ($c_1=c_2=c_3=c$) y como los costos no afectan la solución del problema se define la siguiente función a maximizar:

$$\begin{aligned} \text{Max} \quad \phi &= \sum_{h=1}^3 n_h = - [(800/1+800x_1) + (1,500/1+1,500x_2) + (500/1+500x_3)] \\ \text{s.a} \quad &2.56 X_1 + 104.625 X_2 + 125 X_3 \leq .15 \\ &7.68 X_1 + 78.75 X_2 + 6.25 X_3 \leq .08 \\ &0 \leq X_1 \leq .99875; \quad 0 \leq X_2 \leq .999334; \quad 0 \leq X_3 \leq .998000 \end{aligned}$$

Los resultados obtenidos mediante la aplicación del algoritmo de optimización indican que $x_1=.003153$, $x_2=.000708$ y $x_3=0$, por lo que se deben seleccionar un total de $n=1,454$ (63.2% del total) empresas distribuidas como se muestra a continuación: $n_1=227$, $n_2=727$ y $n_3=500$. Obsérvese que la solución obtenida señala que hay que censar el estrato de las empresas mayores que es donde se observa mayor variabilidad lo cual encarece el costo de la encuesta²⁹.

En una investigación posterior realizada por Kokan y Khan (1967) el problema presentado se replantea de la manera siguiente: si se asume que $x_i= 1/n_i$, entonces las ecuaciones (33), (34) y (35) se pueden reescribir como se muestra a continuación:

²⁹ La optimización de la función objetivo se realizó utilizando el algoritmo del "gradiente conjugado" (Powell, 1977 y Beale, 1972) para funciones no lineales que contiene el sistema de análisis estadístico SAS. Sin embargo, las pruebas realizadas mostraron que ante cambios en la varianza en el estrato de empresas mayores la solución obtenida no se modifica y se mantiene la propuesta de seleccionar a todas las empresas grandes. Asimismo, ante la hipótesis de cambios (incrementos en $V_1 \leq 180'000,000$ y $V_2 \leq 90,000$) en la varianza de las variables, el tamaño de muestra disminuye en los dos primeros estratos ($n_1=212$, y $n_2=679$ y $n_3=500$) pero se mantiene la necesidad de censar el estrato de las empresas grandes. Una situación similar se observó al variar el tamaño del estrato de las empresas mayores a 250, disminuir los valores de las varianzas al interior del estrato e incrementar la variación esperada. Esta situación no resulta consistente, lo cual parece indicar la necesidad de analizar las capacidades prácticas de la solución propuesta por Kokan(1963). De hecho, siempre que $x_h \rightarrow 0 \Rightarrow n_h \rightarrow N_h$ y en el límite son iguales, situación que se presenta en el ejemplo analizado en donde el valor estimado para $x_3 \rightarrow 0$.

$$\text{Max } \phi = - \sum_h^L N_h c_h / (1+N_h x_h) \quad (37)$$

$$\text{s.a. } \sum_h^L a_{hj} x_h \leq v_j + \sum_h^L a_{hj} / N_h ; (j=1,\dots,p) \quad (38)$$

$$1/N_i \leq x_h \leq 1 ; \quad (h=1, \dots, L) \quad (39)$$

Conforme a lo anterior, se reestimaron los tamaños de muestra del ejemplo que estamos analizando a partir de la optimización del siguiente sistema de ecuaciones:

$$\begin{aligned} \text{Max } \phi &= \sum_{h=1}^3 n_h = - [(800/1+800x_1) + (1,500/1+1,500x_2) + (500/1+500x_3)] \\ \text{s.a } &2.56 X_1 + 104.625 X_2 + 125 X_3 \leq .15+.0032+.06975+.25 \\ &7.68 X_1 + 78.75 X_2 + 6.25 X_3 \leq .08+.0096+.0525+.125 \\ &.00125 \leq X_1 \leq 1; .0000666 \leq X_2 \leq 1 \text{ y } .002 \leq X_3 \leq 1 \end{aligned}$$

A partir de la aplicación del algoritmo de optimización no lineal se tiene que $x_1=0.014385$, $x_2=0.001779$ y $x_3=0.002$, por lo que los tamaños de muestra estimados para cada estrato son $n_1=64$, $n_2=409$ y $n_3=250$. De esta forma, la solución encontrada indica que se deben seleccionar un total de $n=723$ empresas (25.8% del total) de la industria manufacturera, con objeto de obtener estimaciones del volumen de la producción y del total de personal ocupado con el nivel de confianza requerido.

Es importante señalar que existen diferencias importantes respecto a la manera en que se procede comúnmente. Así, es habitual que primero se decida el número total de observaciones que se van a seleccionar (n) y posteriormente se distribuya la muestra con algún criterio de asignación entre los diferentes estratos³⁰. En el procedimiento aplicado, el total de unidades que formarán parte de la investigación se determina a partir de agregar los tamaños de muestra calculados para cada uno de los estratos considerando la variabilidad de la variable de interés y las diferencias de tamaño.

5. Muestreo Polietápico

Suponga que se decide un esquema de selección en dos etapas en donde las unidades primarias de muestreo (UPM's) son de igual tamaño y que se está interesado en estimar p características de las unidades de observación. Se asume, sin pérdida de

³⁰ Es conocido que existen diversos procedimientos para asignar la muestra en los estratos y comúnmente ante la falta de información sobre la varianza de cada estrato, lo cual limita la posibilidad de lograr la asignación óptima, se utiliza el criterio de asignación proporcional con el objeto de lograr muestras autoponderadas. Sin embargo, se puede demostrar que con esta manera de proceder no siempre se logra una buena distribución de la muestra por lo que investigaciones recientes sugieren la aplicación del procedimiento denominado "asignación de potencia" (Power Allocation. Bankier, M., 1988) que se considera más robusto y conduce a mejores resultados.

generalidad, el interés por estimar la media \bar{y}_j cuya varianza muestral tiene la siguiente expresión:

$$V_j = \text{var}(\bar{y}_j) = (1/n - 1/N) S_{uj}^2 + (1/nm - 1/NM) S_{wj}^2 \tag{40}$$

Se define $x_1 = (1/n - 1/N)$ y $x_2 = (1/nm - 1/NM)$, de tal manera que $n = N/(1+Nx_1)$ y $nm = NM/(1+NMx_2)$. Como $1 \leq n \leq N$ y $1 \leq m \leq M$, se tiene que $0 \leq x_1 \leq 1-1/N$ y $0 \leq x_2 \leq 1-1/NM$. Así, si definimos a FC como la función de costos:

$$FC = c_0 + c_1 n + c_2 nm + c_3 n^{1/2} \tag{41}$$

En donde c_0 es un costo fijo y c_1, c_2 representan los costos de entrevista en las unidades de primera y segunda etapa respectivamente, mientras que c_3 es el costo de traslado entre las UPM's y USM's.

Debido a que c_0 es una cantidad constante y fijando el límite superior de la varianza esperada $V_j=v_j$, se plantea el problema de maximizar:

$$\text{Max } \phi = - [(c_1 N / (1+Nx_1)) + (c_2 NM / (1+NMx_2)) + (c_3 N^{1/2} / (1+Nx_1))] \tag{42}$$

$$\text{s.a. } S_{u1}^2 x_1 + S_{w1}^2 x_2 \leq v_1 \tag{43}$$

.

$$S_{up}^2 x_1 + S_{wp}^2 x_2 \leq v_p$$

$$0 \leq x_1 \leq 1-1/N, \quad 0 \leq x_2 \leq 1-1/NM \tag{44}$$

6. Aplicación del Método

Suponga que se realizó una encuesta para estudiar el tema de la desocupación en donde se seleccionaron 25 unidades de primera etapa (n) y 10 de segunda etapa (m) con igual probabilidad, y se estimaron las siguientes varianzas para las variables x_1 y x_2 que representan la tasa de desocupación abierta y la de subempleados en una determinada ciudad.

$$S_{i1}^2 = .0087, S_{i2}^2 = .0432, S_{e1}^2 = .0492 \text{ y } S_{e2}^2 = .0850$$

en donde los subíndices i y e indican que se trata de la varianza intraconglomerados y entre conglomerados respectivamente. En este caso, $n=25$ y $m=10$ por lo que $nm=250$, mientras que el costo de entrevista en las UPM's es $c_1=\$7.8$ y en las USM's se estimó en $c_2=\$3$, mientras que $c_3=0$.

Suponga ahora que la Oficina Nacional de Estadística (ONE) desea realizar una encuesta y para la elaboración del nuevo diseño de muestra requiere **determinar los valores óptimos de n y m** de tal forma que se minimice el costo total de la investigación, ya que se cuenta con un presupuesto limitado y se desea utilizar como insumos las varianzas estimadas de la investigación anterior. Observe que no se conocen los valores de N y M los cuales se pueden suponer como infinitos lo cual no impone ninguna restricción adicional a la formulación y solución del problema.

Conforme al procedimiento descrito en la sección anterior se plantea el siguiente problema de optimización.

$$\begin{aligned} \text{Max } \phi &= - [7.8/x_1 + 3/x_2] \\ \text{s.a. } S_{u1}^2 + S_{w1}^2 &\leq v_1 \\ S_{u2}^2 + S_{w2}^2 &\leq v_2 \\ 0 < x_1 = 1/n &\leq 1 \\ 0 < x_2 = 1/nm &\leq 1 \end{aligned} \quad (45)$$

Para determinar el límite superior de V_j ($j=1,2$) se utiliza la siguiente expresión propuesta por Kokan(1963):

$$\Pr \{ V \leq (n-1)s_e^2 / (nmX_0^2) \} = 1 - \alpha \quad (46)$$

En este caso, como se desea un $\alpha=.05$ entonces $X_{0(n-1, \alpha)}^2 = X_{0(24, .05)}^2 = 13.85$, por lo que los valores de v_1 y v_2 se obtienen como se indica a continuación:

$$\begin{aligned} v_1 &= (24)(.0492)/(250)(13.85) = .000341 \\ v_2 &= (24)(.0850)/(250)(13.85) = .000589 \end{aligned}$$

Asimismo, el límite superior para S_{wj}^2 se obtiene por medio de³¹:

$$\Pr \{ S_w^2 \leq 2n(m-1)s_w^2 / [(2nm-2n-1)]^{1/2} - t_{\alpha} \} = 1 - \alpha \quad (47)$$

Para el problema que nos ocupa se tiene que:

$$\begin{aligned} s_{w1}^2 &= (50)(9)(.0087) / [(500-50-1)^{.5} - 1.64]^2 = .010244 \\ s_{w2}^2 &= (50)(9)(.0432) / [(500-50-1)^{.5} - 1.64]^2 = .050865 \end{aligned}$$

Por su parte, los límites inferior y superior de s_{ui}^2 se determinan por medio de las siguientes expresiones:

³¹ Si se utilizan dos colas el valor de α se debe dividir entre 2 por lo que $Z(\alpha/2)=1.64$, y Z tiene una distribución normal estandarizada.

$$\frac{[(n-1)s_e^2/mX_1^2] - [2n(m-1) s_w^2/m[(2nm-2n-1)]^{1/2} - t_{1/2\alpha}]^2}{[(n-1)s_e^2/mX_0^2] - [2n(m-1) s_w^2/m[(2nm-2n-1)]^{1/2} + t_{1/2\alpha}]^2} \quad (48)$$

Siguiendo con nuestro ejemplo, en donde sólo nos interesa conocer el límite superior tenemos que;

$$s_{u1}^2 = [(24)(.0492)/(10)(13.85)] - [(50)(9)(.0087)] / \{9[(500-50-1)^{-5} + 1.96]^2\} = .007714$$

$$s_{u2}^2 = [(24)(.0850)/(10)(13.85)] - [(50)(9)(.0432)] / \{9[(500-50-1)^{-5} + 1.96]^2\} = .010699$$

Conforme a los cálculos anteriores, se plantea el siguiente problema de optimización no lineal en la función objetivo y lineal en las restricciones impuestas a los valores de las x_i 's.

$$\text{Max } \phi = - [7.8/x_1 + 3/x_2]$$

$$\text{s.a. } .007714 x_1 + .010244 x_2 \leq .000341$$

$$.010699 x_1 + .050865 x_2 \leq .000589 \quad (49)$$

$$0 < x_1 = 1/n \leq 1, \quad 0 < x_2 = 1/nm \leq 1$$

Se observa que de la última relación de las ecuaciones (44) se deduce el número óptimo de selecciones de primera y segunda etapa (n y m), y por lo tanto el total de observaciones (nm) que se deben seleccionar para la realización de la encuesta.

La solución óptima del problema indica que $x_1 = .023404$ y $x_2 = .006657$ por lo que se deben seleccionar un total de $n=43$ unidades primarias de muestreo (UPM's) y dentro de estas $m=4$ selecciones de segunda etapa (USM's), lo cual genera un total de $nm=172$ observaciones elementales (hogares en este caso) a seleccionar para la realización de la encuesta. Si se asume un promedio de 3 personas mayores de 12 años por hogar se tendría un total de 516 personas en edad de trabajar que formarían parte de la investigación.

Otras investigaciones posteriores han estudiado el problema de asignación óptima de la muestra a partir de métodos de programación no lineal a lo cual le han denominado programación convexa por las características de la función objetivo (Kokan y Khan(1967) y Huddleston, Claypool y Hocking(1970). En todas las investigaciones se concluye que no existe solución única y por lo tanto se proponen algoritmos particulares que generan soluciones óptimas bajo ciertas suposiciones que se describen en las propuestas elaboradas³².

³² En los trabajos de Kokan(1963 y 1967) que han servido de base para el desarrollo aquí presentado, se analizan las conclusiones para el caso de muestreo doble y para los errores de respuesta y además se menciona que los métodos propuestos pueden ser extendidos a diseños más complejos. De hecho, Kokan y Khan(1967) proponen una solución analítica la cual genera un algoritmo de aplicación iterativo que permite encontrar la solución óptima y analizan un ejemplo de una encuesta con 6 estratos y cinco categorías de interés. Por su parte, en de Huddleston et.al.(1970), se utiliza un algoritmo propuesto por Hartley y Hocking (1963) para analizar una encuesta con 15 estratos y 7 variables de interés.

Los desarrollos analizados muestran que existen alternativas para determinar el número óptimo de selecciones de primera y segunda etapa en las encuestas de propósitos múltiples que deben ser consideradas a fin de mejorar la eficiencia de los diseños de muestra. Asimismo, se demuestra que es posible mejorar los métodos de trabajo actuales aprovechando las capacidades de cómputo y se evidencia la necesidad de generar controles durante la planeación de la encuesta y la ejecución del trabajo de campo que permitan la estimación de los costos que involucra la realización de una encuesta, a fin de poder aplicar algoritmos de cálculo como los que se han analizado.

Además, se hace énfasis en la necesidad de calcular errores de muestreo de los estimadores, así como el efecto de diseño, a fin de disponer de información valiosa para mejorar el cálculo del tamaño de muestra. Esta práctica, sin duda, permitirá mejorar los futuros diseños incrementando la precisión de los estimadores y reduciendo los costos de operación.

IV. CONCLUSIONES

1.- Es claro que una de las preocupaciones principales en la planeación de una encuesta por muestreo probabilístico se orienta a determinar con certeza el número de observaciones necesarias para poder garantizar la precisión y confianza requerida en los estimadores calculados. En este sentido, es importante que las organizaciones ejecutoras estén conscientes que una mala decisión puede limitar la validez de los resultados y en ocasiones llegar a invalidar la investigación completa.

2.- Las encuestas que se realizan en todos los países se pueden considerar de propósitos múltiples. Sin embargo, es práctica común que la determinación del tamaño de muestra se efectúe como si sólo se desearan obtener estimaciones para una sola variable. Posteriormente, se forman dominios de estudios y se analiza la información sin evaluar si las restricciones impuestas por el tamaño de muestra conducen a la pérdida de confiabilidad en las relaciones de causalidad observadas en subpoblaciones de interés y se formulan recomendaciones de política que pueden llegar a carecer de validez estadística.

3.- En muchas ocasiones, se procede como si el único interés de la investigación se centrara en estimar proporciones (p) y posteriormente con los datos generados se forman totales y promedios sin estar conscientes de que el error de muestreo de los estimadores se incrementa por el simple hecho de que para lograr la misma precisión que en el caso de una proporción se requeriría aumentar el tamaño de muestra.

4.- Para el caso de una sola variable la literatura documenta cómo determinar el tamaño de muestra cuando se desea estimar una proporción binomial y no se dispone de un valor que aproxime su varianza. Sin embargo, en la práctica sucede frecuentemente que las variables que se desean estimar tienen una distribución multinomial, por los procedimientos

que habitualmente se aplican no resultan adecuado ya que subestima el número de observaciones que se deben seleccionar y no es posible controlar la precisión simultánea de los estimadores.

5.- Para el caso multivariado se han desarrollado procedimientos matemáticos que permiten explorar alternativas y generan soluciones al problema de determinar el tamaño óptimo de muestra y su asignación en los diferentes estratos en que se divide la población.

6.- Cuando se trata de encuestas de propósitos múltiples en donde la muestra se selecciona en diversas fases y se desea obtener el número óptimo de selecciones de primera (n) y segunda etapa (m) es muy difícil obtener una solución única. Sin embargo, existen algunos algoritmos que permiten soluciones bajo ciertas restricciones.

7.- La única manera de saber si se cumplieron las expectativas de confiabilidad de los estimadores definidas en el diseño de la muestra se logra a partir del cálculo de los errores de muestreo y del efecto de diseño, tanto para la variable principal como para todos aquellos indicadores que se analizarán a partir de la información generada.

8.- Para lograr la elaboración de diseños de muestra óptimos es necesario disponer de información sobre la varianza de los estimadores y sobre los costos de operación de las diferentes etapas de una encuesta. En este sentido, los organismos ejecutores debieran aplicar mayores controles en las diversas etapas de planeación y ejecución del trabajo de campo. Esta práctica permitirá mejorar la eficiencia de los diseños, con el consiguiente ahorro de recursos.

REFERENCIAS BIBLIOGRÁFICAS

- Angers, C. (1974). "A Graphical Method to Evaluate Sample Sizes for the Multinomial Distribution". *Technometrics*, vol. 16, No. 3, 469-471.
- Angers, C. (1979). "Sample Size Estimation for Multinomial Populations". *The American Statistician*. Vol. 33, No. 3, 163-164.
- Angers, C. (1989). "Note on Quick Simultaneous Confidence Intervals for Multinomial Proportions". *American Statistical Association*, vol. 43, No.2, 91.
- Bankier, M. D. (1988). "Power Allocation: Determining Sample Size for Subnational Areas". *American Statistician*, vol. 42, No.3, 174-177.
- CEPAL, División de Estadística y Proyecciones Económicas (1988). "Los Errores de Muestreo en las Encuestas Complejas" (documento de trabajo por publicar).
- Cochran, W. G. (1953). "Sampling Technics". John Wiley & Sons, Inc.
- Cowan, C. D. and Malec, D.J. (1988). "Sample Allocation for A Multistage, Multilevel, Multivariate Survey". Bureau of the Census. Fourth Annual Research Conference.
- Chakravarthy, I. M. (1955). "On the Problem of Planing a Multistage Survey for Multiple Correlated Characters", *Sankhyā*, vol. 14, 211-216.
- Fitzpatrick, S. and Scott, A. (1987). "Quick Simultaneous Confidence Intervals for Multinomial Proportions". *American Statistical Association*, vol. 82, No. 399, 875-878.
- Ghosh, S. P. (1958). "A note on stratified random sampling with multiple characters", *Calcuta Statisc. Ass. Bull.*, 8, 81-89.
- Goodman, A. Leo. (1965). "On Simultaneous Confidence Intervals for Multinomial Proportions". *Technometrics*, vol. 7, No. 2, 247-254.
- Huddleston, H. F., Claypool, P.L. and Hocking, R.R. (1970). "Optimal Sample Allocation to Strata Using Convex Programming". *Journal of the Royal Statistical Society (Series C)*, vol. 19, 273-278.
- Kish, L. (1965). "Survey Samplig". John Wiley & Sons, Inc.
- Kokan, A. R. (1963). "Optimum Allocation in Multivariate Surveys". *Journal of the Royal Statistical Society (Series A)*, vol. 126, 557-565.
- Kokan, A. R. and Khan, S. (1967). "Optimum Allocation in Multivariate Surveys: An Analytical Solution". *Journal of the Royal Statistical Society (Series B)*, vol. 29. 115-125.

- Quesenberry, C. P. and Hurst, D. C. (1984). "Large Sample Simultaneous Confidence Intervals for Multinomial Proportions". *Technometrics*, vol. 6, No. 2., 191-195.
- Sedransk, J. (1965). "A Double Sampling Scheme for Analytical Surveys" *Journal of the American Statistical Association*. 985-1004.
- Sison, C. P. and Glaz, J. (1995). "Simultaneous Confidence Intervals and sample Size determination for Multinomial Proportions". *Journal of the American Statistical Association*, vol. 90, No. 429, 366-369.
- Thompson, S. K. (1987). "Sample Size for Estimating Proportions". *The American Statistician*, vol. 41, No.1, 42-46.
- Tortora, R. D. (1978). "A Note on Sample Size estimation for Multinomial Populations". *The American Statistician*, vol. 32, No. 3, 100-102.
- Waters, J. R. and Chester, A. J. (1987). "Optimal Allocation in Mulivariate, Two Stage Sampling Desings". *The American Statistician*, vol. 41, No. 1, 46-50.
- Yates, F. (1953). "Sampling Methods for Census and Surveys". 2nd. Editon, Charles Griffin & Co., London.



**LAS ENCUESTAS DE PANEL PARA EL ESTUDIO
DE LAS CONDICIONES DE VIDA:
*BONDADES Y RESTRICCIONES***

FERNANDO MEDINA

CEPAL

ÍNDICE**Página**

I. INTRODUCCIÓN.....	151
II. ESTIMACIÓN DE UN PARÁMETRO EN OCASIONES SUCESIVAS	152
III. ALGUNOS ESQUEMAS DE ROTACIÓN DE LA MUESTRA.....	159
IV. LAS MUESTRAS REPETIDAS Y LAS ENCUESTAS CONTINUAS	161
V. MÉTODOS DE ESTIMACIÓN EN LAS ENCUESTAS DE PANEL.....	163
VI. ALGUNAS CONCLUSIONES	166
REFERENCIAS BIBLIOGRÁFICAS	168

LAS ENCUESTAS DE PANEL PARA EL ESTUDIO DE LAS CONDICIONES DE VIDA: BONDADES Y RESTRICCIONES

I. INTRODUCCIÓN

Ante lo cambiante de la realidad económica y social es muy común que los investigadores, diseñadores y evaluadores de política se interesen por conocer y medir los cambios en el comportamiento de un determinado conjunto de variables que guardan estrecha relación con las acciones realizadas por los gobiernos para mejorar los niveles de vida de la población y por ende con las variaciones del ciclo económico.

En este sentido, las agencias estadísticas se ven en la necesidad de hacer mediciones, de manera regular, sobre determinados parámetros de la población que resulten de interés. De hecho, para evaluar el efecto de las acciones de política en el nivel de vida de las familias o para estar en capacidad de conocer la movilidad social de una determinada cohorte de población, es necesario desarrollar estudios longitudinales y contar con instrumentos estadísticos que permitan seguir en el tiempo a un grupo de personas con el fin de observar las transformaciones ocurridas en su nivel educativo, su inserción laboral, sus percepciones económicas, su acceso a los bienes y servicios básicos, a las transferencias públicas y privadas que reciben los miembros del hogar, así como a los cambios ocurridos en la magnitud de la pobreza y la concentración del ingreso.

Para cumplir este objetivo, se requiere contar con instrumentos estadísticos que permitan generar mediciones sucesivas sobre un mismo conjunto de observaciones. La alternativa que surge para dar respuesta a las interrogantes planteadas son las encuestas de panel. Desde la óptica del muestreo, por panel debe entenderse una muestra que observa y hace mediciones en un mismo grupo de elementos en dos o más ocasiones sucesivas, con objeto de evaluar los cambios que están ocurriendo en el nivel de una variable o estimar el comportamiento promedio de alguna característica de interés. En este sentido, se debe aclarar que solamente por medio de los denominados estudios longitudinales es posible medir los cambios que se experimentan en una población respecto de un determinado parámetro, así como explicar las causas que lo originan y las relaciones que existen entre las variables que se analizan. De hecho, la información sobre el cambio bruto que está detrás de un cambio neto en el valor de un indicador, solo se puede explicar por medio de las muestras de panel (Kish, 1979).

Considere el caso de que a partir de dos investigaciones independientes se concluye que la incidencia de la pobreza pasó de 6% a 12%. Ante esta evidencia, el investigador no está en condiciones de afirmar si la tasa estimada a partir de la muestra que está analizando es del 6%, del 12% o de algún otro valor intermedio. De hecho, ante las dos mediciones, solo se estaría en posibilidad de medir el cambio bruto en el valor del indicador. Solamente por medio de los estudios de panel será posible conocer los cambios individuales en la

condición de pobreza de los hogares, así como para evaluar el efecto de programas en los niveles de vida y determinar la dinámica de las interrelaciones entre las variables que explican la formación del ingreso de los hogares y sus posibilidades de acceder a un conjunto de satisfactores básicos (Ashenfelter, Deaton y Solon, 1986).

Por otra parte, algunos expertos (Raj, 1979) opinan que las encuestas continuas son más baratas que las que se realizan en solo una ocasión, además de que permiten proveer información actualizada sobre las tendencias en la evolución de algún fenómeno de interés. Sin embargo, adoptar una decisión para elegir entre una encuesta de panel o una continua con traslape parcial, no es fácil debido principalmente a que normalmente las encuestas persiguen varios objetivos lo cual conducirá necesariamente a diferentes diseños óptimos que difícilmente pueden ser armonizados en una sola encuesta. Además, las correlaciones que normalmente se observan entre los valores de una variable en distintos momentos del tiempo producen resultados contradictorios en la varianza de los estimadores. Así, y como se verá más adelante, en el caso de que se desee estimar la diferencia (d) entre dos muestras sucesivas, la opción de traslape total tiende a reducir la varianza de d , pero incrementa la de la suma de dos parámetros o el promedio.

La manera de resolver el conflicto se debe basar en las conjeturas que se puedan hacer acerca del comportamiento de las variables de interés intentando ofrecer soluciones a partir de diseños de muestra con traslapes parciales. Así, la pregunta que se debe responder es cuál es el diseño muestral que resulta más adecuado para satisfacer un objetivo específico con un presupuesto determinado y una precisión requerida. Esta es la disyuntiva para elegir entre una encuesta de panel o un levantamiento de corte transversal.

II. ESTIMACIÓN DE UN PARÁMETRO EN OCASIONES SUCESIVAS

En los fenómenos asociados a la coyuntura económica es muy común que las agencias productoras de estadísticas estén interesadas en conocer cómo cambia el nivel de un estimador en el tiempo, a consecuencia de acciones de política planeadas o ante variaciones inesperadas en la evolución del ciclo económico.

De acuerdo a lo anterior, y ante el interés por estudiar, por ejemplo, el cambio de la tasa de desocupación abierta entre el tiempo s_1 y s_2 , surgen las siguientes alternativas para el diseño y selección de una muestra de n unidades, que tiene como objetivo medir el nivel de desocupación en cada levantamiento y/o el cambio del valor del estimador entre las dos mediciones.

i) Utilizar la misma muestra en ambas ocasiones. En este caso, se estaría utilizando un diseño de panel en el cual las observaciones seleccionadas en el período s_1 son las mismas que se estudian el momento s_2 . Esta opción también se conoce con el nombre de muestras completamente traslapadas o diseño de muestra fija.

ii) Otra opción implicaría seleccionar dos muestras independientes de tamaño n_1 y n_2 respectivamente. Es decir, no existiría ningún porcentaje de traslape entre las observaciones.

iii) Seleccionar en el tiempo s_1 una muestra de tamaño n y en s_2 mantener m unidades de la primera muestra; es decir, eliminar $n-m$ observaciones y por lo tanto añadir la misma cantidad a fin de que se conserve el mismo tamaño de muestra (n). En esta situación, y conforme a los objetivos específicos de la investigación, se trataría de dos muestras con traslape parcial. Esta técnica también se conoce como rotación de la muestra, muestreo en ocasiones sucesivas con reemplazo parcial de unidades (Yates, 1949) y como muestreo de series de tiempo (Hansen *et. al.*, 1953).

En general, puede decirse que existen ventajas y desventajas entre las opciones señaladas por lo que el investigador debe asumir un posición clara respecto a los objetivos de la encuesta; es decir, se debe definir si se desea garantizar una mayor precisión al estimador que mide el cambio sacrificando la confiabilidad de las estimaciones individuales, o por el contrario se desea privilegiar las mediciones puntuales sin importar que disminuya la confianza de la diferencia entre las observaciones sucesivas.

En todos los casos, las implicaciones estadísticas para estimar la diferencia en el nivel de la ocupación ($t_1 - t_2$) son diferentes. En caso de decidirse por la opción i), se tendría la posibilidad de conocer los cambios individuales en la condición de ocupación de la población activa, además de la diferencia bruta en el nivel del estimador. Sin embargo, mucho se ha escrito sobre algunos de los inconvenientes que se pueden presentar al trabajar por períodos prolongados con muestras de panel en encuestas continuas. De hecho, entre los argumentos en contra de esta práctica se comentan los altos costos que acarrea un programa de esta naturaleza debido a la movilidad física de las familias, así como a la posible contaminación de las respuestas ante la fatiga y predisposición de los informantes para participar en diferentes períodos de entrevista. Esta situación, evidentemente incrementa los errores de estimación y además en la medida de que el tiempo que transcurra entre las mediciones sea considerable, la tasa de respuesta tenderá a decrecer lo cual sin duda incidirá en la menor confiabilidad de los resultados de la investigación.

Cuando el porcentaje de traslape es total; es decir, cuando $n_1=n_2=n$ y $\Pi_1=\Pi_2=1$, la varianza de la diferencia entre dos muestras idénticas también se puede conceptualizar como la diferencia de dos mediciones sobre una misma muestra efectuadas en distintos momentos.

Si definimos como $d=(\bar{x} - \bar{y})$ el estimador de cambio de la media en los momentos s_1 y s_2 entonces, bajo la hipótesis de que las muestras son las mismas, la expresión para obtener el estimador de la varianza estaría dado por:

$$\text{Var}(d) = (\bar{x}/n - \bar{y}/n) = 1/n (S_x^2 + S_y^2 - 2S_{xy}) = 1/n (S_x^2 + S_y^2 - \rho_{xy} S_x S_y) \quad (1)$$

en donde ρ_{xy} es el coeficiente de correlación entre las mediciones efectuadas sobre una misma variable en dos momentos de tiempo.

Reescribiendo la expresión(1) y cuando las varianzas son iguales $S_x^2 = S_y^2$ tenemos que:

$$\text{Var}(d) = 2/n S^2(1 - \rho_{xy}) \quad (2)$$

De acuerdo con la expresión (2) se deduce que la ventaja estadística de estimar d a partir de una muestra de panel, respecto a mediciones independientes, está dada por la presencia del factor $(1 - \rho_{xy})$. De hecho, se debe observar que el coeficiente de correlación mide la reducción en la varianza del estimador debido al traslape de las observaciones. Así, en la medida de que $\rho_{xy} \rightarrow 1$ la variabilidad de la diferencia se mantendrá en el nivel $2/n S^2$.

Cuando el objetivo de la investigación se centra en estudiar el cambio de "status" de una observación respecto a una característica particular que interese al investigador, entonces se aconseja utilizar la misma muestra en las diferentes mediciones que se deseen realizar, ya que la correlación positiva en las observaciones suele contribuir a reducir la varianza del estimador (Kish, *op.cit.*). Esta situación es particularmente importante en los estudios de movilidad social y los programas orientados a la evaluación de políticas económicas y sociales, cuando el objetivo del estudio se centra en conocer el efecto de alguna medida en el nivel de vida de las familias.

Para los estudios de movilidad laboral, social o económica(vinculados al bienestar y a la distribución del ingreso), la encuestas de panel tienen ventajas considerables respecto a los estudios de corte transversal y a los datos de series de tiempo, sobre todo en la etapa de análisis de los resultados cuando se emplean métodos econométricos. Una primera cualidad de las muestras de panel se presenta al observar que cuando se dispone de varias mediciones sobre una misma unidad o individuo, se incrementan los grados de libertad y se reduce el nivel de multicolinealidad entre las variables explicatorias, con lo que se obtienen mejoras considerables en los estimadores de los parámetros obtenidos a partir de modelos econométricos. Asimismo, diversos autores sostienen que un aspecto más importante es el hecho de que a partir de estudios longitudinales se pueden analizar diversos tipos de fenómenos económicos que no pueden ser evaluados a partir de datos de series de tiempo y aquellos que se obtienen de estudios de corte transversal, y sostienen que por medio de las muestras de panel también es posible reducir los sesgos de estimación (Hsiao,1986). En Ashenfelter, *et.al (op.cit)* se señala que existen algunos fenómenos económicos que solo pueden ser evaluados a partir de muestras de panel y como ejemplo se mencionan los estudios de niveles de vida, la dinámica del ingreso y la evaluación de programas sociales.

De manera particular, en un estudio realizado por Freeman y Medoff (1981) se señalan los problemas que se enfrentan para analizar la dinámica del mercado laboral a partir de estudios de corte transversal, y abundan las investigaciones que identifican enormes virtudes a las muestras de panel, aunque no desconocen las restricciones que se enfrentan en el análisis econométrico sobre todo en el caso de los modelos no lineales (Hsiao, *op. cit.*).

Desde el punto de vista de la teoría del muestreo es importante recordar que a pesar de que en un estudio de panel se mantengan las mismas unidades de observación las actitudes y respuestas de los informantes pueden haber cambiado lo cual sin duda introducirá sesgos adicionales en la estimación de la diferencia (d) que no siempre son posibles identificar y por tanto evaluarlos.

Con el fin de reducir los costos de operación de las encuestas debido a la movilidad de los informantes, algunas investigaciones de corte longitudinal han optado por mantener en la muestra paneles de viviendas o de segmentos compactos que incluso son aún más estable que las viviendas. A pesar de las bondades de esta opción, se reconoce que las viviendas y los segmentos compactos muestran niveles de correlación menores que en las familias, sin perder de vista que esta situación solo afecta a la proporción de unidades que son renovadas entre dos levantamientos consecutivos, por lo que las implicaciones en el error del estimador no son considerables. Además, se ha estudiado que es muy probable que los nuevos ocupantes de las viviendas muestren una correlación significativa con la familia anterior lo cual contribuirá a que los errores de estimación no se vean muy afectados por esta circunstancia.

En caso de que se desee estimar el comportamiento promedio de un fenómeno a lo largo de un período de tiempo, es evidente que la mejor opción será seleccionar muestras independientes ya que la correlación positiva que se observa entre las variables a partir de un estudio de panel con traslape parcial o total estaría afectando el nivel y la varianza del estimador.

Así, en este caso el coeficiente de correlación en (2) sería igual con cero ya que la covarianza entre ambas mediciones no existiría.

En este sentido, la expresión para la varianza de la diferencia (d) está dada por:

$$\text{Var}(d) = S_x^2 / n_x + S_y^2 / n_y \quad (3)$$

y cuando las varianzas son iguales en ambas mediciones $S = S_x^2 = S_y^2$ se tiene que;

$$\text{Var}(d) = S^2 (1/n_x + 1/n_y) \quad (4)$$

Asimismo, en el caso de que los tamaños de muestra se mantengan constantes $n = n_x = n_y$, (4) se convierte en:

$$\text{Var}(d) = 2/n S^2 \quad (5)$$

En el caso de la opción iii), que señala la existencia de un cierto porcentaje de traslape entre dos muestras consecutivas, la expresión para la varianza dada en (2) se transforma en:

$$\text{Var}(d) = 2/n S^2 (1 - \rho_{xy} \Pi_m) \quad (6)$$

En donde $\Pi_m > 0$ representa la proporción de la muestra que se traslapa. De esta expresión se deduce que en el caso que $\rho_{xy} > 0$ la ganancia en precisión será proporcional al producto que se forma al considerar la correlación entre las observaciones y la proporción de traslape ($\rho_{xy} \Pi_m$), y como es evidente la máxima ganancia se obtiene en el caso en que $\Pi_m=1$ y $\rho_{xy}=+1$, que es equivalente a la situación de trabajar con un panel de observaciones; es decir, con traslape total.

El efecto de un traslapamiento parcial ($1-\rho_{xy}\Pi_m$) sobre la varianza del estimador de diferencias será proporcional a la proporción de traslape observado Π_m . Sin embargo, en caso que se desee estimar la suma de dos medias se tiene un efecto exactamente opuesto al señalado: es decir, el factor ($1+\rho_{xy}$) se obtiene para un traslape total y ($1+\rho_{xy}\Pi_m$) en caso de un traslapamiento parcial.

En la literatura sobre el tema se dice que trabajar con muestras traslapadas es adecuado en aquellos casos en los que se desea obtener una estimación de la media relativa de una determinada variable en el momento en que se efectúa la medición; es decir, el cambio de una característica en una población determinada. Sin embargo, también pueden utilizarse para comparar las medias de dos características diferentes. De hecho, en las encuestas periódicas que generan series de tiempo es adecuado trabajar con muestras de tamaño constante y traslapamientos parciales, ya que en esta situación las ganancias debidas a la correlación entre las observaciones se asocian a la proporción de traslape observada y son directamente proporcionales a Π_m .

En el caso particular de las encuestas que estudian los niveles de empleo y la utilización de la fuerza de trabajo, traslapar la muestra también tiene efectos sobre el valor del estimador y su varianza. En efecto, si se deseara conocer la evolución de la tasa de desocupación abierta (tda) y su análisis se realizara a partir de muestras independientes, los cambios en el nivel del estimador estarían influidos, tanto por las variaciones de la variable como por el efecto del marco y la utilización de una nueva muestra.

Cuando se trabaja con diseños de panel la pregunta obvia que surge es cuál es el porcentaje de traslape óptimo que se desea mantener en el caso que se esté interesado en que la encuesta permita medir las diferencias entre dos o más mediciones consecutivas. En ese sentido, es importante señalar que la fracción de la muestra que se sustituya en cada ocasión dependerá de los objetivos de la investigación, así como de la intensidad en la correlación existente entre las observaciones en las mediciones sucesivas, la cual por lo general es positiva y no podrá ser determinado en forma arbitraria sin haber examinado la relación entre las variables y haber fijado con claridad los objetivos de la investigación. Si la asociación entre las mediciones es muy alta, se podrá conservar solo una pequeña fracción sin que esto tenga una incidencia negativa en la calidad de los resultados. En general, algunos expertos recomiendan que deberá reponerse más del 50% de las unidades de la primera encuesta para ser incluidas en la segunda ocasión (Raj, *op.cit.*).

Está claro que las alternativas de utilizar un diseño con traslape total o decidirse por un esquema de muestra independientes responden a objetivos contrapuestos, por lo que se debe buscar un equilibrio entre las alternativas señaladas. Así, una encuesta que utiliza muestras traslapadas debiera tratar de:

i) Mantener una proporción de traslape Π_m lo más alto posible a fin de maximizar la precisión de los estimadores de cambio en ocasiones sucesivas, si es que eso se define como el objetivo central de la encuesta.

ii) Tratar de que Π_m sea lo menor posible a fin de evitar que se contamine la muestra por observaciones aberrantes o falta de interés de los entrevistados a participar en la encuesta.

Antes de entrar a discutir las diferentes alternativas en uso para el diseño de encuestas con muestras traslapadas, se hace referencia a las variantes que existen en la aplicación de la opción definida como traslape parcial.

Es evidente que la opción más común es definir una proporción de traslape $\Pi_m > 0$ entre dos muestras consecutivas. Sin embargo, también es posible que la muestra utilizada en la segunda ocasión (n_y) sea un subconjunto de la primera (n_x); es decir, que $n_x > n_y = n_m$. En esta situación, observe $\Pi_y = 1$ pero $\Pi_x < 1$. Conforme a lo anterior, la varianza de la diferencia d se puede expresar como:

$$\text{Var}(d) = S^2/n_x + S_y^2/n_m - 2 S_{xy}/n_x = S_y^2/n_m + (S_x^2 - 2\rho_{xy} S_x S_y)/n_x \quad (7)$$

Cuando $S_y^2 = S_x^2 = S^2$, entonces (7) se transforma en;

$$\text{Var}(d) = S^2 [1/n_m + (1 - 2\rho_{xy})/ n_x] \quad (8)$$

La situación planteada tiene consecuencias prácticas ya que en el caso que $(S_x^2 - 2\rho_{xy} S_x S_y) < 0$, las observaciones que no forman parte del traslape ($n_x - n_m$) tenderán a incrementar la varianza. Asimismo, cuando $S_x^2 = S_y^2$ y $\rho_{xy} > 0.5$, entonces el incremento de la muestra n_x , más allá del traslapamiento, reducirá la ganancia que pudo obtenerse debido a la correlación positiva entre las unidades que forman parte del traslapamiento.

Adicionalmente a lo planteado, puede darse el caso de que la media de una muestra se compare con la media de su subconjunto lo cual ocurre frecuentemente cuando se desea comparar una población con una de sus subclases (por ejemplo, los desocupados encubiertos con los desocupados). Esta situación se presenta al observar que cuando dos variables son iguales ($X_i = Y$) y muestran una correlación perfecta ($\rho_{xy}=1$), entonces;

$$\text{Var}(d) = S^2(1/n_m - 1/n_x) = S^2/n_m (1 - n_m/n_x) = S^2/n_m (1 - \Pi_x) \quad (9)$$

Ante una situación como la planteada muchos investigadores y usuarios de la información se confunden frecuentemente ya que proceden a calcular la varianza de la diferencia como si se tratara de dos estimaciones independientes, sin darse cuenta de que existe una alta correlación positiva entre las variables analizadas ya que una de el dominio de estudio analizado en un subconjunto de la población de referencia. Es evidente que la varianza del estimador de la diferencia se reduce en una proporción Π_x que identifica el porcentaje de traslape entre ambas muestras.

De hecho, se debe señalar que el número de observaciones traslapadas nm no contribuye al incremento de la varianza, ya que la diferencia entre $(n_x - n_m)$ explica una proporción importante de la misma.

En la práctica, es habitual que se observen reducciones en la varianza de las diferencias debido a la correlación que existe entre las unidades traslapadas, ya que el signo de la covarianza entre las variables es negativo, mientras que las correlaciones son positivas (Kish, *op.cit.*). Así, por ejemplo, si se está interesado en medir la condición de desocupación de la población activa a partir de dos encuestas continuas en donde la correlación en relación con la condición de ocupación de la fuerza de trabajo es de $\rho=.75$, entonces la diferencia entre $(tda_1 - tda_2)$ a partir de una encuesta con traslapamiento completo se afectaría por el factor $(1-.75)=0.25$, mientras que en el caso de un traslape parcial $\Pi_m=.50$ el factor de ajuste para la varianza sería $(1-.5*.75)=0.625$.

Se han desarrollado algunos procedimientos estadísticos a fin de encontrar métodos eficientes para estimar el cambio en el nivel del estimador más allá que la simple comparación entre las dos medias. Debido a la correlación que existe entre las observaciones, cada elemento de la muestra que pertenece a la proporción de traslape Π_m contribuye en una menor proporción a la formación de la varianza del estimador en $(1-\rho)$ en relación a los elementos que se ubican fuera de proporción traslapada; es decir, en el subconjunto $(1-\Pi_m)=Q$. En términos algebraicos, la varianza mínima del estimador de diferencia se obtiene cuando decrecen las ponderaciones de la proporción $(1-\Pi_m)$ en el factor $(1-\rho)$. De esta manera, la expresión para obtener el estimador de diferencia con varianza mínima es:

$$D(\bar{x} - \bar{y}) = [\Pi_m / (1 - Q\rho)] (\bar{x}_{nx} - \bar{y}_{ny}) + \{[Q(1-\rho)] / (1 - Q\rho)\} (\bar{x}_{Qx} - \bar{y}_{Qy}) \quad (10)$$

En la expresión (10), $(\bar{x}_{Qx} - \bar{y}_{Qy})$ representa la diferencia entre las medias en proporción de la muestra Q y la primera parte se pondera por Π_m y la segunda por $Q(1-\rho)$ y ambas se relativizan al dividir las entre $(1 - Q\rho)$.

La varianza de la diferencias ponderada se obtiene por medio de:

$$\text{Var}(D) = [(1-\rho)/(1-Q\rho)] 2S^2/n \quad (11)$$

El factor entre paréntesis indica la reducción en la varianza debido al traslapamiento de la muestra, y se interpreta como el efecto de la correlación en la varianza de la diferencia ponderada, mientras que $(1 - \Pi\rho)$ lo es en la diferencia simple.

Siguiendo con el ejemplo sobre los cambios en la condición de ocupación de la fuerza de trabajo considere $\rho=.75$ y $\Pi=.5$; entonces, la media simple sin ponderar reduce la varianza en $(1 - .5*.75)=0.625$, mientras que el efecto de la estimación ponderada es $(1 - .75)/(1 - .5*.75)= 0.25/0.625=0.4$. En este caso las ponderaciones serían $0.5/0.625=0.8$ en la proporción traslapada $\Pi=0.5$, y $0.5(1-.75)/0.625=0.125/0.625=0.2$ en la proporción no traslapada Q.

Finalmente se señala que la diferencia entre la varianza de los estimadores ponderados "versus" los simples es:

$$[(1-\rho)/(1-Q\rho)]/(1-\Pi\rho) = (1-\rho)/(1-\rho + \Pi Q\rho^2) \quad (12)$$

De la expresión (12) se deduce que cuando la correlación ρ entre las observaciones es grande y las proporciones de traslape Π y Q son moderadas, el término $\Pi Q\rho^2$ es muy diferente a $(1 - \rho)$. En el caso que se está analizando, con los valores definidos para $\rho=.75$ y $\Pi=0.5$, se tiene que $(1-.75)/(1-.75 + .25*.5625)=0.64$, que es igual al cociente entre los valores de las varianzas estimadas anteriormente $(.40/.625=0.64)$.

III. ALGUNOS ESQUEMAS DE ROTACIÓN DE LA MUESTRA

Al decidir trabajar con un esquema de muestras traslapadas, una pregunta que frecuentemente surge durante el trabajo práctico es qué porcentaje de la muestra se debe conservar para un levantamiento futuro. Esta pregunta no tiene una respuesta universal, ya que la proporción de la muestra que se debe preservar dependerá de los objetivos de la encuesta.

La rotación de la muestra puede ser efectuada de diversas formas. En Binder y Hidiroglou (1988) se mencionan las siguientes alternativas:

- i) Con un nivel de rotación muestral
- ii) Con un seminivel de rotación muestral; y
- iii) Con multiniveles de rotación

La manera de efectuar la rotación de la muestra con un nivel se interpreta de la manera siguiente. Suponga que la muestra consiste de n elementos que deben ser encuestados en diferentes ocasiones; entonces, $(1-\pi)n$ elementos ($0 \leq \pi \leq 1$) pueden estar presentes en la muestra del período s_1 y además podrían conservarse para ser incluidos en el momento s_2 . El resto de las observaciones (πn) serán reemplazadas por un número igual a fin de complementar el tamaño de muestra (n) requerido.

Bajo este esquema, se pueden citar, como ejemplos, las encuestas de Fuerza de Trabajo realizadas por Statistics Canada, las cuales emplean un esquema de rotación que significa reemplazar un sexto de los hogares en muestra mensualmente. De esta forma, la muestra se compone de seis paneles cada uno de los cuales permanece en operación durante seis meses consecutivos y posteriormente es reemplazado por otro del mismo tamaño.

Como ejemplo de las encuestas con un seminivel de rotación de la muestra, se encuentra la experiencia desarrollada por Hansen *et.al.* (1955) e introducida en la Encuesta Corriente de Población(CPS) realizada en forma mensual por la Oficina del Censo de los Estados Unidos de Norteamérica. Bajo este esquema de rotación, las unidades permanecen en la muestra en un número consecutivo de ocasiones y no son utilizadas por algunos períodos y posteriormente vuelven a encuestarse durante un determinado número de ocasiones. Una generalización de este procedimiento fue desarrollada por Rao y Graham(1964) de la manera siguiente. Sean N y n el tamaño de la población y de la muestra respectivamente, entonces un total de $n_2 \geq 1$ unidades de observación permanecen en la muestra durante r períodos($n=n_2r$), salen por m ocasiones y regresan para estar vigentes durante r repeticiones, posteriormente vuelven a salir m veces y el ciclo se repite un determinado número de rondas.

En el caso particular de la experiencia de la Oficina del Censo de los Estados Unidos la muestra está compuesta por ocho paneles de rotación. Así, un determinado grupo de rotación permanece en la muestra durante cuatro meses consecutivos, sale por ocho meses y regresa para participar durante cuatro veces más. Este esquema de rotación es conocido como **4-8-4** y tiene la característica que durante un mes cualquiera, el 75% de la muestra es común respecto al mes anterior, de modo que existe un 50% de traslape entre dos años consecutivos, respecto al mismo mes del año anterior.

Evidentemente que el procedimiento descrito admite variantes. Por ejemplo, para una encuesta trimestral, puede suceder que cada unidad elemental permanezca en la muestra durante seis trimestres de modo el traslape entre dos mediciones consecutivas sea de un 83%.

En párrafos anteriores se ha indicado que una permanencia prolongada de las unidades elementales en la muestra puede inducir sesgos de respuesta por parte de los informantes. En este sentido, los trabajos de Bailar (1975, 1978 y 1979) para las encuestas de población activa de los Estados Unidos, así como las investigaciones de Pearl (1979) en relación con las encuestas de consumo en ese mismo país, son ejemplos de los análisis que se han realizado para evaluar el sesgo en el valor de los estimadores en los diseños de muestras con traslape.

El hecho de que las unidades de la muestra tengan diferentes valores según el número de veces que han sido encuestadas introduce el denominado sesgo del grupo de rotación, en el entendido de que en caso de no existir sesgo en cada uno de los levantamientos cada panel de rotación tendría el mismo valor esperado en el nivel de la variable analizada.

En Sánchez C. (1984) se da una expresión para calcular un índice para cada grupo de rotación a partir de la siguiente expresión:

$$I_{\text{grot}} = (A_{\text{grot}} / \bar{A}) * 100 \quad (13)$$

en donde:

I_{grot} \approx Índice del grupo de rotación

A_{grot} \approx Número de personas con el atributo de interés en un grupo de rotación determinado.

\bar{A} \approx Número de personas con el atributo de interés en todos los períodos de la muestra.

A fin de ejemplificar el uso de este indicador, Sánchez C. (*op.cit.*) utiliza los datos de Bailar (1979) para mostrar el uso del índice en el análisis de los varones en situación de desempleo, utilizando la Encuesta Permanente de Empleo de los Estados Unidos de Norteamérica.

Así, $I_{\text{grot1}}=105.9$ y $I_{\text{grot7}}=95.5$ estaría indicando que la estimación de los varones estimados en la primera oportunidad, es un 5.9% mayor que la media estimada sobre todos los períodos de la encuesta. Asimismo, la medición efectuada en el grupo de rotación 7, está por debajo de la media en un 4.5%. Según las conclusiones obtenidas por Bailar (*op.cit.*), a partir de este índice se demuestra la manera en que el sesgo de rotación afecta el nivel de los estimadores, el cual tal vez se pueda suavizar mediante técnicas de promedios móviles que permitan hacer un ajuste estacional de las observaciones pero que nunca logran eliminarlo por completo.

Los resultados de la investigación no evidenciaron con claridad los problemas que originaron las diferencias y se citan como probables causas los sesgos de entrevista, el hecho de que la entrevista se hace por teléfono, los cambios en el cuestionario y los grupos de rotación de la muestra, entre otros.

IV. LAS MUESTRAS REPETIDAS Y LAS ENCUESTAS CONTINUAS

La alternativa de trabajar con muestras repetidas se presenta ante los estudios que deciden realizar el trabajo de campo en un momento determinado. De hecho, existen opiniones que sostienen las ventajas al realizar varias encuestas menores en ocasiones sucesivas, a pesar de que se mantiene la divergencia de opiniones y no es posible formular una recomendación universal sobre este tema y las soluciones particulares deben aplicarse según los objetivos de la investigación.

Entre las bondades que se asocian a las encuestas continua se encuentra la posibilidad de mejorar la precisión de los estimadores y/o reducir costos de operación en la ejecución del trabajo de campo, ya que el financiamiento se puede obtener en la medida que se vaya requiriendo para darle continuidad a las labores de obtención de la información. En contraparte, la realización de una encuesta grande en un momento reducido de tiempo obliga a afinar muy bien toda la maquinaria de operación y sobre todo la etapa de diseño de cuestionarios, elaboración de manuales y capacitación de los encuestadores, a fin de reducir los sesgos en la captación de la información. Asimismo, los costos de operación pueden incrementarse de manera importante y se requiere disponer del presupuesto en su totalidad para garantizar la conclusión de la investigación.

Tal vez una de las mayores ventajas que se asocian a las muestras repetidas se refieran a la posibilidad de generar información sobre las variaciones temporales en el valor de los estimadores, además de que permiten identificar y estimar las tendencias estacionales y seculares de una serie de tiempo y brindan la posibilidad de captar las variaciones irregulares en la tendencia del indicador debido a un comportamiento inesperado del ciclo económico.

Este tipo de encuestas resultan apropiadas para el diseño y evaluación de políticas en la medida de que las autoridades y los agentes económicos tengan la capacidad de reaccionar ante los cambios en el nivel y la tendencia del indicador. Esto significa que antes de diseñar una encuesta con estas características, es necesario responderse a la pregunta de para qué se quieren los datos y si los diseñadores de política tienen la capacidad de reaccionar e influir, por ejemplo, ante las variaciones mensuales del nivel de desempleo, con acciones que incidan en el nivel de la tasa de desocupación abierta en el próximo período.

La evidencia indica que esto último no es muy factible y que habitualmente la información recogida mensualmente alimenta bancos de información de series de tiempo que permiten la elaboración y estimación de modelos para el pronóstico del comportamiento futuro del indicador. En este sentido, esta situación puede llegar a ocasionar altos costos de operación sin que la información sea utilizada para la toma de decisiones en el corto plazo.

Por otra parte, también se afirma que la agregación de la muestra en las encuestas repetidas puede conducir a mejores estimaciones que en el caso de una sola encuesta. En este sentido Kish (*op.cit.*) opina que "*la selección probabilística de segmentos de tiempo a partir de la totalidad de un intervalo permite realizar inferencia estadística, a partir de la media de una condición promedio del intervalo*". En caso contrario, la inferencia hecha a partir de un segmento "tipo" implica hacer supuestos en relación con la estabilidad del fenómeno de estudio lo cual puede significar ignorar las variaciones estacionales, seculares e irregulares de las variables que se están midiendo. En este sentido, las encuestas repetidas, a partir de promediar en forma conveniente las distintas mediciones, puede agregar observaciones con fines de inferencia mejorando la precisión del estimador.

V. MÉTODOS DE ESTIMACIÓN EN LAS ENCUESTAS DE PANEL

Los procedimientos de estimación de los parámetros de interés a partir de los datos que generan las encuestas de panel son de gran importancia para obtener una visión realista sobre la evolución de las variables que se analizan.

En este sentido, es importante señalar que tanto las mediciones obtenidas por las encuestas de panel como las que se generan con datos de corte transversal y aquellas que provienen de diseños con muestras traslapadas, tienen errores de medición y errores de no muestreo que afectan la confiabilidad de los estimadores. Sin embargo, un mismo tipo de error tiene diferentes implicaciones en la precisión de los resultados.

Conforme a lo anterior, se identifican dos grandes métodos que se usan comúnmente (Binder y Hidiroglou *op.cit.*): El procedimiento clásico y el de series de tiempo. En el método clásico (Jesenn, 1942, Tikkiwal, 1979 y Wolter, 1979) la sucesión de observaciones y_{it} que se forma a partir de la medición de ciertos parámetros de interés $\{\theta\}$ como la media o el total de la variables, se consideran fijos en el tiempo y se supone que de alguna manera están relacionados con los valores obtenidos en las mediciones anteriores $y_{t-1,i}$ en la misma unidad bajo una cierta estructura de correlación que se determina a partir del análisis de las observaciones.

El problema de muestreo en ocasiones sucesivas fue considerado por primera vez por Jesenn (1942) en el análisis de datos provenientes de una encuesta de predios agrícolas. En esta investigación, se trataba de que de una muestra de $n=900$ unidades de observación usada en una encuesta efectuada en 1938, 450 se retuvieron para ser empleadas en la ronda de 1939 y una cantidad similar se seleccionó de manera independiente a fin de conservar el tamaño de muestra original, con lo cual la muestra de 1939 puede considerarse una submuestra de la muestra original.

Por otra parte, los procedimientos de estimación basados en la filosofía de las series de tiempo consideran al parámetro θ_t como una variable aleatoria cuyo valor cambia en el tiempo como resultado de un proceso estocástico (choques aleatorios) (Blight and Scott, 1973 y Jones, 1980). En este sentido, los métodos de estimación deben considerar la autocorrelación entre las observaciones, así como la existencia de diferentes componentes de la serie de tiempo (tendencia, efecto secular, componente estacional e irregular).

Siguiendo a Hansen *et.al*(1953) y el trabajo de Ashenfelter *et.al.*(*op.cit.*), a continuación se exponen algunos resultados relacionados con los errores de medición y la proporción óptima de traslape en las encuestas que consideran en su diseño paneles de rotación de la muestra.

Considere la situación en que se desea hacer dos encuestas sucesivas para medir la evolución de un determinado parámetro de interés. En este caso, sea n el tamaño total de la

muestra en ambos levantamientos y π la proporción de la muestra que se mantendrá entre las investigaciones. Así,

$x_1 \approx$ es la media, en el período 1, de los $n\pi$ hogares que pertenecen a la proporción de hogares trasladada;

$x_2 \approx$ es la media, en el período 2, de los $n\pi$ hogares que pertenecen a la proporción de hogares trasladada;

$y_1 \approx$ es la media, en el período 1, de los $n(1-\pi)$ hogares que pertenecen a la muestra vieja;

$y_2 \approx$ es la media, en el período 2, de los $n(1-\pi)$ hogares que pertenecen a la muestra nueva;

S_1 y $S_2 \approx$ la desviación estándar de la muestra en los períodos 1 y 2 para la variable de interés;

$\rho \approx$ el coeficiente de correlación entre dos mediciones consecutivas para la variable de interés.

Conforme a lo anterior, si el interés de la investigación se centra en estimar las medias poblacionales μ_1, μ_2 y la diferencia entre estos valores ($\mu_1 - \mu_2$), entonces los errores de muestreo se pueden expresar como combinaciones lineales de los parámetros estimados como se muestra a continuación.

$$\hat{\mu}_1 = \alpha_{11} x_1 + \alpha_{12} x_2 + \beta_{11} y_1 + \beta_{12} y_2 \quad (14)$$

$$\hat{\mu}_2 = \alpha_{21} x_1 + \alpha_{22} x_2 + \beta_{21} y_1 + \beta_{22} y_2 \quad (15)$$

$$\hat{\Delta}_\mu = (\mu_1 - \mu_2) = \alpha_{31} x_1 + \alpha_{32} x_2 + \beta_{31} y_1 + \beta_{32} y_2 \quad (16)$$

Para garantizar la confiabilidad de los resultados, se requiere que los estimadores sean insesgados y que el error de muestreo sea lo más pequeños posible. Observe que bajo el supuesto de que $\alpha_{12} = \beta_{12} = 0$ implicaría de que el estimador $\hat{\mu}_1$ tendría que ser calculado posterior al levantamiento de la segunda encuesta. Así, dado un cierto porcentaje de traslape $\pi > 0$, los valores de las α 's y β 's no dependen de los ponderadores seleccionados.

Con base a lo anterior, las expresiones para estimar μ_2 y Δ son:

$$\begin{aligned} \hat{\mu}_2 = \{ & [\rho\pi(1-\pi)] / [1 - (1-\pi)^2 \rho^2] \} [(S_2/S_1) (y_1 - y_2)] + [\pi / 1 - (1-\pi)^2 \rho^2] + \\ & + \{ (1-\pi)[1 - (1-\pi)^2 \rho^2] / [1 - (1-\pi)^2 \rho^2] \} y_2 \end{aligned} \quad (17)$$

$$\begin{aligned} \Delta_{\mu} = & \{(1-\pi)[1-(1-\pi)^2\rho^2] / [(1-(1-\pi)^2\rho^2)(y_2-y_1) + \{\pi/[1-(1-\pi)^2\rho^2]\}(x_2-x_1)] \\ & + [\rho\pi(1-\pi)] / [1-(1-\pi)^2\rho^2] \{(y_2-y_1)(S_2/S_1) - (y_2-x_2)(S_2/S_1)\} \end{aligned} \quad (18)$$

En las expresiones (16) y (17) es fundamental conocer cuál es el valor de π que resulta más apropiado y que minimiza el error de estimación; asimismo, se debe conocer el coeficiente de correlación (ρ) entre las mediciones efectuadas y los estimadores puntuales de la variable de interés. En este sentido, se recuerda que el valor de ρ y el cociente de los errores estándar (S_1/S_2) cambia para cada variable, por que es muy difícil encontrar un valor óptimo de π que satisfaga los diversos objetivos de una encuesta de propósitos múltiples.

En el caso en que se dese estimar el valor de Δ y en el caso de que las varianzas $S_1=S_2$ sean iguales, se tiene que la varianza del estimador de diferencia se obtiene por medio de:

$$\sigma_{\Delta_{\mu}}^2 = 2(1-\rho) s_2 / n[1-(1-\pi)\rho] \quad (20)$$

A partir de (20) se comprueba que cuando $\rho > 0$, el diseño óptimo se logra al traslapar toda la muestra; es decir, cuando $\pi=1$ lo cual implica que se visitan en la segunda ronda los mismo hogares que en la primera, por lo que el valor mínimo que puede obtenerse en (20) es;

$$\sigma_{\Delta_{\mu}}^2 = [2(1-\rho) s_2 / n] \quad (21)$$

la cual solo depende del tamaño de muestra, el coeficiente de correlación entre las mediciones sucesivas y la varianza muestral del estimador de la variable de interés. Es evidente, que cuando se trata de un diseño de muestra complejo el efecto de diseño (deff) deberá ser considerado, por que el valor mínimo de la expresión (21), que corresponde a un esquema de selección aleatorio simple, se incrementa en el caso de que el $deff > 1$, lo cual regularmente sucede en la práctica.

Es evidente que si se trabaja con dos muestras independientes, la correlación se estará sacrificando la eficiencia del estimador debido a que en esta situación no se considera la correlación entre las mediciones efectuadas sobre una misma variable. Por su parte, para los diseño con traslape parcial la pregunta importante es determinar cuál es la pérdida de eficiencia en el valor del estimador del error de muestreo $\sigma_{\Delta_{\mu}}^2$, en la medida que cambia la proporción de traslape considerada π . En el trabajo de Ashenfelter *et.al.(op.cit.)*, se incluye una tabla que muestra cómo cambia el error estándar del estimador en función a diferentes combinaciones en los valores de π y de ρ . Así, se afirma que en el caso de que se retenga para un segundo levantamiento la mitad de la muestra seleccionada ($\pi= 50\%$), la pérdida de eficiencia en el valor de (21), nunca excederá del 30%.

Asimismo, se presenta otra tabla en donde se varía la precisión asociada con la segunda encuesta para estimar el parámetro μ_2 y se presenta la siguiente expresión para el cálculo de la varianza:

$$\sigma_{\mu_2}^2 = S^2/n \{ [1 - \rho^2(1-\pi)] / [1 - \rho^2(1-\pi)^2] \} \quad (22)$$

En el caso en que las varianzas de la variable de interés en las mediciones sucesivas sean iguales, se tiene que el diseño óptimo se logra cuando la proporción de traslape π es igual a:

$$\pi = 1 - [1 - (1-\rho^2)^{1/2}] / \rho^2 \quad (23)$$

VI. ALGUNAS CONCLUSIONES

i) Las encuestas de panel o de corte longitudinal resultan de gran utilidad para el diseño y definición de políticas, así como para la evaluación de programas que inciden en los niveles de vida de la población. De hecho, para aquellas investigaciones cuyo objetivo es evaluar el cambio en la situación de un indicador en un grupo social entre dos instantes consecutivos de tiempo, los estudios longitudinales resultan ser el mecanismo más apropiado, a pesar de que no están exentos de error.

ii) Los datos obtenidos a partir de estudios de panel tienen algunas ventajas importantes en la identificación y estimación de modelos econométricos, pero también presentan ciertas limitaciones en su tratamiento sobre todo en el caso de modelos no lineales. De cualquier manera, lo más importante es reconocer que los datos provenientes de las encuestas de panel requieren un tratamiento especial y para su análisis se deben utilizar métodos estadísticos que consideren la estructura de correlación entre las variables estudiadas y la colinearidad entre los indicadores que se utilizarán para la especificación y estimación de los modelos econométricos.

iii) Para los estudios económicos asociados a características dinámicas de las personas la mejor opción es estudiar la evolución de los fenómenos por medio de estudios longitudinales.

iv) Desde el punto de vista práctico las encuestas de panel pueden llegar a ser muy costosas sobre todo en las situaciones en donde se presenta una gran movilidad de las unidades de observación consideradas en la muestra. En este sentido, las tasas de no respuesta tienden a incrementarse incidiendo en el error de estimación y por supuesto la calidad de los resultados. Las ventajas y desventajas de los estudios de panel no deben considerarse en forma absoluta y para su aplicación se deben equilibrar unas con otras.

v) Cuando en los objetivos de la investigación se define el interés por estimar el cambio en el nivel de una variable, los diseños de muestras traslapadas representan una alternativa a los estudios longitudinales. En este caso, la pregunta crucial que se debe responder en cuál es la proporción óptima de traslape a utilizar, y la respuesta debe guardar estrecha relación con los objetivos de la encuesta, la correlación entre las mediciones efectuadas y la utilización que se quiera hacer de los datos recabados.

vi) La varianza de un estimador de diferencia para medir cambio es menor en los estudios de panel pero esto a su vez incrementa la variabilidad de un estimador de suma o promedios, por lo que es muy importante relacionar los objetivos de la investigación con los indicadores que se desean analizar y los métodos de estimación que se utilicen.

vii) Las encuestas traslapadas o con paneles de rotación parcial evidencian algunas bondades importantes respecto a los estudios de panel total y su utilización en los programas nacionales de encuestas que realizan las oficinas de estadística de la región denota que desde el punto de vista práctico tienen mayor aceptación para aportar datos sobre la evolución de algunos fenómenos económicos y de manera particular en relación al nivel de ocupación y el desempleo; sin embargo, presentan limitaciones para el estudio de fenómenos dinámicos.

REFERENCIAS BIBLIOGRÁFICAS

- Ashenfelter, O., Deaton, A. and Solon, G. (1986). "Collecting Panel Data in Developing Countries". World Bank, Living Standards Measurement Study LSMS Working Papers, No. 23, Washington, D.C..
- Bailar, B.A. (1975). "The effects of rotation group bias on estimates from panel surveys", *Journal of the American Statistical Association (JASA)*, vol. 70, No. 349, 23-30.
- (1979). "Rotation samples biases and their effects on estimates of changes". 42 Sesión del Instituto Internacional de Estadística, Manila, Filipinas.
- Bailey, L. and Corby, C. (1978). "A comparison of some adjustment and weighting procedures for survey data". *Survey sampling and measurement*. Namboodiri, New York.
- Binder, D.A. and Hidiroglou, M.A. (1988). "Sampling in Time". *Handbook of Statistics*, Krishnaiah, P.R. and Rao, C.R. vol. 6, 187-211.
- Blight, B.J. and Scott, A.J. (1973). "A stochastic model for repeated survey". *Journal of the Royal Statistical Society, Series B* 35, 61-68.
- Freeman, R.B., and J.L. Medoff (1981). "The Impact of Collective Bargaining: Illusion or Reality? Mimeograph. Harvard University.
- Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953). "Sample Surveys Methods and Theory". Volume 2. Wiley, New York.
- Hansen, M.H., Hurwitz, W.N., Nisselson, H. and Steinberg, J. (1955). "The redesign of the Census Current Population Survey". *Journal of the American Statistical Association* 50, 701-719.
- Hsiao, C. (1986). "Analysis of Panel Data". *Econometric Society Monographs*. Cambridge University Press.
- Jesenn, R.J. (1942). "Statistical investigation of a farm survey for obtaining farm facts". *Iowa Agricultural Station Research Bulletin* 304, 54-59.
- Jones, R.G. (1980). "Best linear unbiased estimators for repeat surveys". *Journal of the Royal Statistical Society, Series B* 42, 221-226.
- Kish, L. (1979). "Muestreo de Encuestas". Editorial Trillas. México, D.F.
- Pearl, R.B. (1979). "Revaluation of the 1972-1973". U.S. Consumer Expenditure Survey, Technical Paper No. 46, U.S. Bureau of Census, 46-50.

Raj,D.(1979)."La Estructura de las Encuestas por Muestreo", Fondo de Cultura Económica, México, D.F..

Sánchez C.,J.L.(1984)."Curso Intensivo de Muestreo en Poblaciones Finitas". Instituto Nacional de Estadística de España(INE), Madrid, España.

Tikkiwal,B.D.(1979)."Successive sampling: A review". Proceedings of the 42nd session of the International Statistical Institute held in Manila.Book 2,367-384.

Wolter,K.M.(1979)."Composite estimation in finite populations".Journal of the American Statistical Association 74,604-613.

Yates,F.(1949)."Sampling Methods for Censuses and Surveys".First Edition.Charles Griffin,London.



**CUESTIONARIOS UTILIZADOS EN LOS PAÍSES DE AMÉRICA
LATINA Y EL CARIBE PARA LA EJECUCIÓN DEL PROGRAMA
PERMANENTE DE ENCUESTAS DE HOGARES**

MARÍA DE LA LUZ AVENDAÑO

GABRIELA LÓPEZ

(CEPAL)

ÍNDICE

	<u>Página</u>
I. PRESENTACIÓN.....	173
II. PROGRAMAS DE ENCUESTAS DE HOGARES EN AMÉRICA LATINA Y EL CARIBE.....	174
1. Antecedentes	174
2. Cuestionarios utilizados	176
 BIBLIOGRAFÍA	 183
 CUADROS.....	 185
Cuadro 1: América Latina: Programa Permanente de Encuestas de Hogares. 1995.....	186
Cuadro 2: América Latina: Temas investigados en el cuestionario central de las encuestas de hogares de propósitos múltiples	187
Cuadro 3: América Latina: Características de la vivienda investigadas en la encuesta de hogares de propósitos múltiples	188
Cuadro 4: América Latina: Características generales de la población investigadas en las encuestas de hogares de propósitos múltiples	189
Cuadro 5: América Latina: Variables de migración investigadas en las encuestas de hogares de propósitos múltiples. 1995.....	190
Cuadro 6: América Latina: Características educacionales investigadas en las encuestas de hogares de propósitos múltiples	191
Cuadro 7: América Latina: Características ocupacionales investigadas en las encuestas de hogares de propósitos múltiples. 1995.....	192
Cuadro 8: América Latina: Concepto investigado en las encuestas de hogares de propósitos múltiples	194

CUESTIONARIOS UTILIZADOS EN LOS PAÍSES DE AMÉRICA LATINA Y EL CARIBE PARA LA EJECUCIÓN DEL PROGRAMA PERMANENTE DE ENCUESTAS DE HOGARES

I. PRESENTACIÓN

La necesidad de contar con información confiable y oportuna para apoyar el diagnóstico y la toma de decisiones en el ámbito económico y social, ha llevado a los países de América Latina y el Caribe a impulsar el fortalecimiento de los sistemas estadísticos nacionales, incluido el desarrollo de programas para la implementación de encuestas de hogares por muestreo.

En consonancia con este objetivo, la Comisión Económica para América Latina y el Caribe (CEPAL), al igual que otros organismos internacionales como el BID y el Banco Mundial, ha prestado asistencia técnica a los países de la región con el fin de mejorar y consolidar la capacidad de gestión y análisis de la información por parte del personal de las Oficinas Nacionales de Estadística (ONEs) involucrado en la planeación, diseño, ejecución y administración de los programas permanentes de encuestas de hogares.

Como se sabe, atendiendo a la historia y características propias del sistema estadístico de cada país y a la naturaleza de sus requerimientos de información, tales programas suelen presentar rasgos específicos. Pese a que en general se inscriben en un marco conceptual y metodológico relativamente homogéneo, derivado básicamente de las recomendaciones internacionales existentes sobre la materia, estos tienden a diferir, por ejemplo, en cuanto a la periodicidad con que se realizan las encuestas, su ámbito temático, cobertura geográfica, calidad y disponibilidad de la información y los instrumentos utilizados para la recolección de los datos. Sobre esto último, hay consenso que, sin pretender estandarizar el contenido y formato de dichos instrumentos, resulta del todo deseable propender hacia el establecimiento de bases comunes para la elaboración de los cuestionarios, que permitan mejorar la calidad y pertinencia de los antecedentes que se relevan, a la par con hacer posible una más amplia y adecuada comparabilidad de la información entre los distintos países.

En este sentido, la presente publicación busca contribuir a la formación de una base de datos documental a nivel regional, que ayude a conocer mejor la forma en que los países de América Latina vienen resolviendo sus necesidades en materia de generación de información por medio de encuestas a los hogares, así como la manera en que ésta se refleja en el diseño de los cuestionarios utilizados. A partir de la revisión de la documentación oficial, se describe el contenido del cuestionario aplicado por cada país en la encuesta del programa permanente realizada en 1995. De este modo, se aprecian los diferentes temas investigados, los conceptos que se indagan y la estructura y formulación de las preguntas. Asimismo, se corrobora que, pese a que en muchos casos se aborda una misma temática, la

amplitud y profundidad de la investigación es diferente, lo que queda de manifiesto en la naturaleza y número de las preguntas que se efectúan a los informantes. A su vez, también se constata que tanto el tema central investigado en cada encuesta como los temas complementarios son diversos, dependiendo de las particularidades del país y sus necesidades específicas al momento en que la información fue levantada.

Cabe agradecer a los funcionarios de las Oficinas Nacionales de Estadística por el apoyo brindado para la elaboración de este estudio, al tiempo que los invitamos a continuar aportando sus valiosos comentarios y sugerencias a los efectos de actualizar y enriquecer periódicamente el conocimiento de la documentación relativa a las encuestas de hogares que se llevan a cabo en los países de América Latina y el Caribe. Compendios de información como el que aquí se presenta pueden contribuir a fortalecer la cooperación horizontal, a la vez que orientan los esfuerzos que los organismos internacionales vienen realizando en la región en materia de asistencia técnica en el ámbito de las encuestas de hogares.

II. PROGRAMAS DE ENCUESTAS DE HOGARES EN AMÉRICA LATINA Y EL CARIBE

1. Antecedentes

Las acciones llevadas a cabo en los países de América Latina y el Caribe para desarrollar un sistema de información socioeconómica derivado de las encuestas de hogares se remontan a principios de la década de los sesenta. En efecto, los trabajos realizados en el marco del programa de "Estudios Conjuntos sobre Integración Latinoamericana" (ECIEL) definieron el inicio de los esfuerzos regionales con miras a generar información estadística confiable sobre el nivel y composición del consumo y el ingreso de las familias en 18 de los principales centros urbanos de 10 países sudamericanos. Durante los años comprendidos entre 1966 y 1972 se efectuó el diseño de las encuestas y los correspondientes operativos de campo para la recolección de la información, a partir de lo cual se elaboraron diversos estudios que dieron sustento a importantes análisis sobre el consumo y la distribución del ingreso en los países incorporados al programa.

En estas encuestas se utilizó el denominado modelo Atlántida, elaborado por la Oficina del Censo de los Estados Unidos de América, con el auspicio de la Agencia Internacional para el Desarrollo (AID). El material desarrollado para implementar las encuestas (referido a un país ficticio llamado Atlántida) fue presentado en un seminario regional realizado en Ciudad de México en 1965, en el que participaron más de 50 estadísticos de 15 países de América Latina y el Caribe, especialistas en encuestas por muestreo. Este modelo sirvió de base para las encuestas aplicadas en un número considerable de países de la región, dando origen a programas regulares de ejecución de encuestas demográficas y socioeconómicas que en varios casos siguen vigentes incluso hasta nuestros días.

Posteriormente, en 1979, se puso en marcha el "Programa para Desarrollar la Capacidad Nacional para Efectuar Encuestas de Hogares" (NHSCP), impulsado por las Naciones Unidas y concebido como un esfuerzo interagencial que contó con el apoyo de organismos especializados como el Banco Mundial y el Programa de las Naciones Unidas para el Desarrollo (PNUD). Los primeros informes elaborados en el marco de este programa, evidenciaron en diversos países la existencia de encuestas específicas, que analizaban de manera aislada y parcial temas tales como el empleo, la migración, la fecundidad y los gastos de consumo. Se trataba de encuestas especiales que mostraban una gran heterogeneidad y falta de integración en cuanto a su diseño conceptual y metodológico y al tratamiento estadístico de los datos, así como una escasa relación entre los distintos levantamientos realizados.

Paralelamente a esta iniciativa de carácter mundial, los países de América Latina y el Caribe dieron origen al "Programa Interamericano de Encuestas de Hogares" (PIDEH). Este programa fue propuesto en la VII Conferencia Interamericana de Estadística, convocada por la Organización de Estados Americanos y celebrada en la República Dominicana en 1977, en tanto que los objetivos y metas fueron establecidos luego en Santiago de Chile en 1981.

Sin embargo, a fin de aunar los esfuerzos regionales en esta área y optimizar el uso de los recursos y capacidades técnicas de los diferentes organismos multilaterales que participaban en los distintos proyectos, en 1982, en Washington D.C., se decretó la fusión de ambos programas. Al mismo tiempo, se acordó integrar las actividades de diseño conceptual y estadístico con las acciones de capacitación que se consideraban necesarias para fortalecer la capacidad técnica de los profesionales encargados de la coordinación e implementación de las encuestas. A su vez, se dispuso la utilización de un marco conceptual y metodológico coherente, acorde con las recomendaciones emanadas de las reuniones y consultas técnicas realizadas por los organismos especializados del Sistema de las Naciones Unidas y de otras instancias regionales. No obstante, y más allá que en la mayoría de los países el diseño de las encuestas se ajustó en líneas generales a dichas recomendaciones, los cuestionarios tendieron a reflejar de manera específica la realidad y preocupaciones de cada país, en ocasiones en desmedro de una adecuada comparabilidad internacional.

Por otra parte, con posterioridad a la conclusión de los referidos programas, los países han mantenido el interés en reforzar su capacidad técnica y analítica en el área de las encuestas de hogares, así como en mejorar la calidad de la información que se obtiene. Sin embargo, la crisis que afectó a la región durante la década pasada limitó fuertemente las posibilidades de seguir avanzando en los países que ya contaban con un programa regular de encuestas de propósitos múltiples, y dificultó su implantación en aquellos otros que se encontraban relativamente más rezagados.

En este contexto surge el "Programa para el Mejoramiento de las Encuestas de Condiciones de Vida en América Latina y el Caribe" (MECOVI), coauspiciado por el Banco

Interamericano de Desarrollo (BID), el Banco Mundial y la Comisión Económica para América Latina y el Caribe (CEPAL), que con la participación directa de los países de la región pretende contribuir a la implantación y fortalecimiento de sistemas integrados de encuestas de hogares, que permitan generar información adecuada y de alta calidad acerca de la situación social y niveles de vida de las familias.

En suma, a pesar de los importantes esfuerzos que se han realizado en el pasado, resulta necesario actualizar los sistemas existentes y reforzar la capacidad teórica y analítica de los técnicos nacionales, para lograr un mejor aprovechamiento de la información disponible. Asimismo, se requiere apoyar la realización de encuestas en aquellos países que carecen de información confiable sobre las condiciones de vida de la población, bajo el concepto de un Sistema Integrado de Encuestas en Hogares (SIEH) que permita armonizar y homogeneizar, conceptual y metodológicamente, las diversas encuestas que se realizan, incorporando de manera adecuada aquellos temas de interés que contribuyen al conocimiento de la situación económica y social de los hogares.

2. Cuestionarios utilizados

Los cuestionarios de las encuestas de los 16 países de América Latina analizados en esta nota (véase el cuadro 1), presentan una gran diversidad respecto a su cobertura temática, diseño y formato, como también en lo que se refiere a la población objeto de estudio en las distintas áreas. En lo que sigue no se pretende calificar la pertinencia de la inclusión o no de cada tema por parte de los países, ni tampoco emitir juicios de valor sobre la calidad de la información recabada. Simplemente se desea describir el contenido de los cuestionarios y, de paso, llamar la atención acerca de las notables diferencias que existen y la virtual imposibilidad de pretender estandarizar los instrumentos de captación de información o establecer patrones uniformes para el estudio de cada tema. Se debe estar consciente que cada país presenta sus propios problemas, y en función de ellos establece prioridades y estrategias diferentes para analizar y evaluar un determinado fenómeno.

De modo que se aborda una descripción comparativa desde el punto de vista de la cobertura temática de los cuestionarios, donde la importancia e intensidad de la investigación de cada tema se manifiesta en el número y nivel de detalle de las preguntas incorporadas en ellos. En ese sentido, la revisión de los cuestionarios que se utilizan actualmente en el trabajo de campo de las distintas encuestas abarcó la identificación de los temas más recurrentes, así como los conceptos que se investigan y el número de preguntas que se efectúan.

Esta revisión incluyó un conjunto de 16 países (Argentina, Bolivia, Brasil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Honduras, México, Panamá, Paraguay, Perú, República Dominicana, Uruguay y Venezuela) y los cuestionarios analizados corresponden, en la mayoría de los casos, a las encuestas de hogares levantadas el año 1995.

Los temas que con mayor frecuencia se investigan en las encuestas se pueden agrupar como sigue: vivienda, características sociodemográficas de la población, migración, educación, ocupación e ingresos de los miembros del hogar. Los objetivos de las encuestas no siempre son los mismos, lo cual explica de cierta forma las diferencias observadas en cuanto al interés e intensidad con que se investiga cada tema. Sin embargo, como se indicó anteriormente, el examen de los antecedentes pone de manifiesto que gran parte de las encuestas se basan en un marco metodológico común, el cual se fundamenta en las recomendaciones emanadas de los foros de consulta y en los lineamientos formulados por los organismos internacionales especializados.

2.1 Visión general

De la revisión de los contenidos de los cuestionarios se desprende que la gran mayoría de los países investiga regularmente las características de la vivienda. Sólo 5 de ellos (si en el caso de Chile se considera la encuesta de empleo del INE) no incluyeron este tema en los formularios de 1995. (Véase el cuadro 2) A su vez, aquellos países que captan esta información lo hacen con una intensidad variable: Argentina es el país que menor énfasis pone en el tema (con sólo 6 preguntas), mientras que El Salvador, con 25 preguntas, y Brasil con 30, son los países en donde los aspectos relacionados con la vivienda y sus servicios se investigan con mayor amplitud y profundidad.

La totalidad de los países incorporan preguntas sobre características sociodemográficas de la población residente en los hogares. La moda se ubica entre 4 y 5 preguntas, destacando Ecuador y Panamá cuyos cuestionarios incluyen únicamente 3 referencias sobre las características generales de la población, y Brasil y Chile con 7 y 8, respectivamente.

La migración es otro tema que también es investigado por prácticamente todos los países, con excepción de Chile (INE) y El Salvador, aun cuando en general se destinan a este fin nada más que 2 ó 3 preguntas. Brasil es el que mayor importancia asigna a este tema, con 12 preguntas, seguido de Argentina con 8, en tanto que en varios casos se indaga exclusivamente sobre el lugar de nacimiento de los miembros del hogar.

Dada la importancia que reviste la educación en muchos campos de la vida de las personas y en particular en sus oportunidades de bienestar, todas las encuestas examinadas incorporan explícitamente preguntas relativas a esta dimensión. El promedio de preguntas consideradas en los cuestionarios gira alrededor de 7, siendo Brasil el país que incluye un mayor número de ellas (22), mientras que Colombia, Ecuador, Honduras y Panamá registran sólo 3.

La ocupación es el tema central de las encuestas del programa permanente en los 16 países analizados, con la excepción de la CASEN de Chile cuyo propósito es más amplio dado que se orienta a proporcionar información para la caracterización socioeconómica de la

población y el seguimiento y evaluación de las políticas y programas sociales. Al igual que para los demás temas mencionados, la profundidad en el estudio del empleo y del funcionamiento del mercado de trabajo en general, se manifiesta en el número de preguntas incorporadas en los cuestionarios. República Dominicana es el país que registra menos preguntas en esta área (sólo 16). En cambio, Brasil y Paraguay asignan gran relevancia al análisis de la ocupación, por lo que sus encuestas incluyen 133 y 124 preguntas, respectivamente. Los restantes países destinan a este objeto un número variable de preguntas, que oscila entre 24 y 49, correspondiendo el valor más bajo a Argentina y el más alto a México.

Por último, en todas las encuestas examinadas se consigna la investigación del valor de ciertos tipos de ingresos (monetarios y en especie) para determinadas categorías de personas. El número de preguntas sobre este tema es también muy variable y se ubica entre 2 y 67. México es el país cuya investigación del ingreso es más limitada, referida sólo a los asalariados, en tanto que Uruguay inquiriere sobre un concepto amplio que incluye el ingreso total (monetario y no monetario) de las personas, más una imputación por el uso de la vivienda propia y el valor de los bienes producidos para autoconsumo.

2.2 Análisis por tema

a) Vivienda

Las variables más comúnmente investigadas son: tipo de vivienda; materiales de construcción de techos, paredes y pisos; tenencia de la vivienda; número de cuartos y dormitorios; tipo de combustible utilizado para cocinar; origen del agua y tipo de abastecimiento; disponibilidad de servicio sanitario y sistema de eliminación de excretas; disponibilidad de energía eléctrica y tenencia de bienes durables en el hogar. (Véase el cuadro 3)

Los únicos países que incluyen en sus cuestionarios preguntas sobre prácticamente todos estos temas son Brasil, Chile (CASEN), El Salvador, Paraguay, Perú y República Dominicana. Por su parte Venezuela no investiga el combustible utilizado para cocinar, ni la disponibilidad del agua, mientras que Chile no indaga sobre el combustible ni la existencia de bienes durables en el hogar. En el caso de México, en tanto, además de que también están ausentes los conceptos anteriores, tampoco se incluye una pregunta sobre el servicio sanitario. Colombia y Honduras presentan una situación similar entre ellos, ya que sus encuestas no captan información sobre el material de construcción de los techos de la vivienda, el combustible utilizado para cocinar ni la tenencia de bienes durables. Finalmente, la menor cobertura de los temas planteados se presenta en Argentina, ya que la Encuesta Permanente de Hogares (EPH) incluye preguntas sobre únicamente el 50% de las variables consideradas.

b) Características sociodemográficas

Todas las encuestas registran preguntas sobre algunas variables demográficas de interés. Así, la relación de parentesco con el jefe del hogar, el género de los miembros del hogar, y en particular el de los integrantes de la fuerza de trabajo, la edad y estado civil de las personas, son los conceptos incluidos con mayor frecuencia en los cuestionarios analizados (véase el cuadro 4). En ningún país se deja de considerar al menos una pregunta sobre cada uno de estos tópicos, resaltando el hecho que en prácticamente todos ellos ésta se limita exclusivamente a sólo una.

c) Migración

La investigación de este fenómeno se aborda por medio de preguntas tales como el lugar de nacimiento de las personas, la localidad en que residían anteriormente, las razones del traslado y el tiempo de residencia en el lugar en que se efectúa la entrevista (véase el cuadro 5). El examen de los cuestionarios revela que no hay ningún país que cubra simultáneamente los cuatro conceptos señalados, como asimismo que en Chile (INE) y El Salvador el tema está totalmente ausente de la encuesta. En México la indagación se limita al lugar de nacimiento de los miembros del hogar, en tanto que en Chile (CASEN) y Costa Rica al lugar de residencia anterior. Por su parte, Argentina, Brasil, Ecuador, Paraguay, República Dominicana y Uruguay inquieran sobre tres de los cuatro conceptos planteados. La pregunta que con menor frecuencia se incluye en los cuestionarios analizados se refiere a la razón del cambio del lugar de residencia, ya que solamente es formulada por Bolivia, Paraguay y República Dominicana. A su vez, el tiempo de residencia en el lugar de la entrevista se investiga en 9 de los 16 países, y 13 de las encuestas comparadas incorporan la pregunta sobre el lugar de residencia anterior.

d) Educación

Pese a la importancia que reviste este tema, no todos los países lo investigan con la misma intensidad. Como era de esperar, el concepto relacionado con el nivel de estudios de los miembros del hogar se incluye en todos los cuestionarios. Asimismo, la condición de analfabetismo se investiga en la mayor parte de los países, siendo las excepciones nada más las encuestas de Costa Rica, México, Panamá, Paraguay, República Dominicana y Uruguay. (Véase el cuadro 6)

Los cuestionarios que contemplan un mayor número de preguntas sobre el tema educacional son los de Chile (CASEN), El Salvador, Perú, República Dominicana y Uruguay. En el primer caso, este hecho se explica debido a que los objetivos de la encuesta CASEN exigen profundizar en esta área y por tanto incorpora, por ejemplo, preguntas relativas a la asistencia preescolar y escolar de los menores, las razones de no asistencia, el último año aprobado, el nivel y grado al que asiste y la jornada escolar. A su vez, países como Costa Rica, Perú y Uruguay investigan, además, algunos antecedentes descriptivos del tipo de educación no formal recibida por los miembros del hogar.

e) Ocupación

La investigación del mercado de trabajo, empleo y desempleo, constituye el objetivo central de la gran mayoría de las encuestas analizadas. En este tema, un primer aspecto a destacar se refiere a la edad mínima de los miembros del hogar para ser considerados parte de la población económicamente activa (PEA). Mientras Argentina no establece un límite de edad para este efecto, Perú lo fija en 6 años y Bolivia y Paraguay en 7 años. A su vez, hay un grupo de 8 países en los cuales esta edad mínima está definida a partir de los 10 años (Brasil, Colombia, Costa Rica, Ecuador, El Salvador, Honduras, República Dominicana y Venezuela), en tanto que Chile (CASEN), Colombia y México la fijan en 12 años. Por último, Uruguay y Panamá establecen los límites más altos, ya que la información sobre condición de actividad es captada para las personas de 14 y 15 y más años de edad, respectivamente (véase el cuadro 7). Este aspecto es relevante no sólo para la comparabilidad de las cifras entre países sino también debido a que, aunque los menores de esas edades debieran asistir de manera regular a los centros educativos, es probable que según el mínimo establecido se pueda incurrir en subestimaciones de la población que desempeña una actividad económica, ya que la realidad confirma que en algunos países existe un número significativo de menores que trabajan.

Por su parte, la profundidad con que se investiga la ocupación muestra diferencias considerables entre los países. Para determinar la condición de actividad de la población se utilizan desde un mínimo de 4 preguntas (Brasil y Chile-Ine), hasta un máximo de 15 (México). Dentro de este rango, en los cuestionarios de Chile (Casen), Perú, República Dominicana y Venezuela se incorporan 5 preguntas, en tanto que en el resto de los países la media se ubica en torno a 8 preguntas.

En cuanto a las características del empleo, en todos los países el interés se centra fundamentalmente en investigar el tipo de ocupación, la categoría del empleo y la rama de actividad económica a que pertenece la empresa donde la persona desarrolla su actividad principal. En sólo siete de las encuestas examinadas (Bolivia, Brasil, El Salvador, México, Paraguay, Perú y República Dominicana) es posible, a su vez, conocer el tipo de establecimiento (público o privado) en que trabajan los ocupados. En prácticamente todos los países, con excepción de Colombia y República Dominicana, se capta información sobre el tamaño del establecimiento donde el entrevistado desempeña su ocupación principal. Finalmente, sólo Brasil, Chile, Colombia, El Salvador, Paraguay, Perú y Venezuela informan sobre las prestaciones sociales recibidas por los trabajadores, y en particular sobre la cobertura de la seguridad social de los ocupados.

Por otra parte, debido a los frecuentes vaivenes que experimentan los niveles de actividad económica, en algunos países de la región es cada vez más frecuente que las persona busquen mejorar sus ingresos a partir de tener más de una ocupación remunerada. En ese sentido, es importante que las investigaciones sobre la utilización de la fuerza de trabajo consideren captar información sobre el pluriempleo, a fin de evaluar el esfuerzo que realiza la población por mantener un adecuado nivel de vida. En Argentina, Bolivia, Chile,

Colombia, Honduras, Panamá, República Dominicana y Uruguay no se investigan las características de la ocupación secundaria de las personas que manifiestan tener más de un empleo, ni tampoco se registra información sobre los ingresos obtenidos o las horas dedicadas a ese empleo. En cambio, en Brasil, El Salvador y Paraguay la investigación de la ocupación secundaria es exhaustiva, indagándose acerca del tipo de ocupación, la rama de actividad y la categoría en el segundo empleo, además de la clase y el tamaño del establecimiento.

Sobre la calidad y protección del empleo, la mayoría de las encuestas incluyen preguntas para conocer si el tipo de contratación de los ocupados es permanente o tiene carácter temporal. Asimismo, en todos los países existe interés por determinar el número de horas trabajadas, lo que permite estimar diversas modalidades de subocupación de la fuerza de trabajo.

Para determinar la condición de búsqueda de trabajo por parte de los desocupados, las diversas encuestas establecen un período de referencia específico para la captación de la información. Las variantes van de aquellas que utilizan una semana fija o bien la semana anterior a la realización de la entrevista (Argentina, Bolivia, Colombia-urbana, Costa Rica, Ecuador, Honduras, México, Paraguay y Uruguay), a otros que utilizan un período más amplio, como un mes e incluso un año (Brasil, Chile, Colombia-rural, El Salvador, Panamá, Perú y Venezuela). A su vez, Honduras es el único país que establece tanto la semana como el mes anterior a la fecha de la entrevista.

En este mismo ámbito, en la mayoría de los países los cuestionarios incluyen preguntas que indagan acerca de la experiencia laboral de los desocupados, además de la ocupación, rama y categoría ocupacional que tuvieron en su último empleo. Colombia y Paraguay investigan, además, las pretensiones de quienes buscan trabajo en cuanto al tipo de ocupación buscada y la rama de actividad y categoría en que desean insertarse.

Los inactivos también son objeto de análisis, debido a que entre ellos se encuentran personas que desean incorporarse al mercado de trabajo o que cuentan con experiencia laboral. Las encuestas de Costa Rica, El Salvador, México, Panamá, Paraguay y Venezuela investigan determinadas características de este segmento de la población.

f) Ingresos

En la mayoría de los países cuyas encuestas fueron examinadas se hacen esfuerzos por captar, con distinto grado de desglose, los ingresos corrientes totales de las personas y los hogares. Uruguay es uno de los casos en que mediante un gran número de preguntas (67) se intenta aproximarse a la medición del concepto de ingreso total disponible, más una estimación del valor imputado por el uso de la vivienda propia y de los bienes producidos o suministrados por el hogar. En una situación similar se encuentra Chile (CASEN) con 28 preguntas, Perú con 27 y El Salvador con 22. (Véase el cuadro 8)

Luego, hay un grupo de 3 países (Brasil, Panamá y Paraguay) que captan los ingresos totales, monetarios y en especie, y otros 3 (Argentina, Bolivia y Ecuador) que sólo lo hacen respecto de los ingresos monetarios. Las encuestas de los demás países se limitan a intentar medir un concepto de ingreso más restringido, entre los que destaca México con la menor cobertura conceptual, dado que investiga únicamente los sueldos y salarios monetarios de la ocupación principal y lo hace mediante la utilización de tan sólo 2 preguntas.

Sin embargo, en todos los casos se debe estar conscientes que la medición de los ingresos de las personas que se obtiene mediante las encuestas de hogares suele adolecer, por lo general, de problemas de falta de respuesta y de subestimación de los valores reportados. Estos sesgos tienden a afectar a algunas corrientes de ingreso con mayor intensidad que a otras. De manera que es de particular importancia efectuar siempre, previo al uso de la información, un análisis de la confiabilidad de los datos de ingreso y, si procede, incorporar los ajustes correspondientes.

BIBLIOGRAFÍA

A fin de complementar los antecedentes presentados respecto al contenido de los cuestionarios utilizados por los países en sus programas permanentes de encuestas de hogares de propósitos múltiples, a continuación se incluye una lista de textos referidos principalmente a las recomendaciones técnicas que han sido formuladas en este campo y que las oficinas nacionales de estadística utilizan, de manera regular, en los diversos proyectos que llevan a cabo sobre la materia. En ese sentido, se consideran una fuente de consulta necesaria para propender hacia una mayor coherencia y uniformidad conceptual entre los distintos países.

1. Naciones Unidas (1975), "Hacia un Sistema de Estadísticas Sociales y Demográficas". Serie F, No. 18, Nueva York.
2. Naciones Unidas (1978), "Estadísticas de Migración Interna". Informe Técnico, Serie F, No. 23, Nueva York.
3. Naciones Unidas (1979), "Principios y Métodos para el Mejoramiento de las Estadísticas Sociales en los Países en Desarrollo". Serie F, No. 25, Nueva York.
4. Naciones Unidas (1980), "Principios y Recomendaciones para los Censos de Población y Habitación". Nueva York.
5. Naciones Unidas (1982), "Medición del Empleo y de los Ingresos Rurales". Estudios e Informes de la CEPAL, No 19, Santiago.
6. Naciones Unidas (1983), "Papel que desempeña el programa para desarrollar la Capacidad Nacional de Efectuar Encuestas de Hogares en el suministro de información sobre la salud en los países en desarrollo". Estudio Técnico Nro.3, Departamento de Cooperación Técnica para el Desarrollo y Oficina de Estadística, Nueva York.
7. Naciones Unidas (1986), "Encuestas de Ingresos y Gastos: Conceptos y Métodos en la Experiencia Latinoamericana". Cuadernos de la CEPAL, No 53, Santiago.
8. Naciones Unidas (1987), "Manual de Encuestas sobre Hogares". Serie F, No. 31, Nueva York.
9. Naciones Unidas (1988), "Cómo Pesar y Medir Niños. Determinación del Estado Nutricional de niños pequeños mediante Encuestas de Hogares". Departamento de Cooperación Técnica para el Desarrollo y Oficina de Estadística, Nueva York.
10. Naciones Unidas (1992), "Manual de Censos de Población y Habitación. Características Demográficas y Sociales". Serie F, No. 23, Nueva York.
11. Naciones Unidas (1995), "Guidelines for Household Surveys on Health". National Household Survey Capability programme, New York.
12. Organización Internacional del Trabajo (1983), "Decimocuarta Conferencia Internacional de Estadígrafos del Trabajo, 1982". Informe de la Conferencia, Ginebra.
13. Comisión Económica para América Latina y el Caribe (1995), "Recopilación de las Recomendaciones Internacionales sobre las Estadísticas de Empleo, Desempleo y Subempleo". Documento Interno, División de Estadística y Proyecciones Económicas.

14. Instituto Nacional de Estadística y Censos (INDEC) (1989), "Segunda Reunión del Comité de Expertos en Estadísticas Sociodemográficas". Buenos Aires, Argentina.
15. International Labour Office (1990), "Survey of Economically Active population, Employment, Unemployment and Underemployment". Geneva.
16. Ferreira, J. y E. Klein (1988), "Empleo Rural: Metodologías de Medición". Oficina Internacional del Trabajo, PREALC.

CUADROS

Cuadro 1
AMÉRICA LATINA: PROGRAMA PERMANENTE DE ENCUESTAS DE HOGARES, 1995.

PAÍS	ENCUESTA	ORGANISMO EJECUTOR	FECHA Y PERIODICIDAD	COBERTURA GEOGRÁFICA	TAMAÑO DE LA MUESTRA
Argentina	Encuesta Permanente de Hogares	Instituto Nacional de Estadística y Censos	Mayo y Octubre Dos al año	A.M. y 20 Aglomerados Urbanos	Oct. 1994 21771 Hogares
Bolivia	Encuesta Integrada de Hogares	Instituto Nacional de Estadística	Junio Anual	8 Cap. de Departamento más El Alto	5569 Hogares
Brasil	Pesquisa Nacional por Amostragem de Domicílios	Fundacao Instituto Brasileiro de Geografia e Estadística	Septiembre Anual	Nacional	85270 Hogares
Chile	Encuesta Nacional de Empleo	Instituto Nacional de Estadística	Continua Anual	Nacional	IV Trimestre 31004 Hogares
	Encuesta de Caracterización Socioeconómica	Ministerio de Planificación y Cooperación	Cada dos Años	Nacional	1996 33636 Hogares
Colombia	Encuesta Nacional de Hogares-Fuerza de Trabajo	Departamento Administrativo Nacional de Estadística	Septiembre Variable Cuatro al año	Nacional Urbana Rural	18255 Hogares 7339 Hogares
Costa Rica	Encuesta de Hogares de Propósitos Múltiples	Dirección General de Estadística y Censos	Julio Anual	Nacional	9631 Hogares
Ecuador	Encuesta Periódica de Empleo y Desempleo en el Área Urbana	Instituto Nacional de Estadísticas y Censos	Junio y Noviembre Dos al año	Urbana	Noviembre 8203 Hogares
El Salvador	Encuesta de Hogares de Propósitos Múltiples	Dirección de Información del Ministerio de Relaciones Exteriores	Todo el Año	Nacional	8482 Hogares
Honduras	Encuesta Permanente de Hogares de Propósitos Múltiples	Dirección General de Estadística y Censos	Octubre Dos al año	Nacional	6331 Hogares
México	Encuesta Nacional de Empleo Urbano	Instituto Nacional de Estadística, Geografía e Informática	Continua Anual	Urbana	III Trimestre 1992 38388 Hogares
	Encuesta de Hogares	Dirección de Estadísticas y Censos	Agosto Anual	Nacional	9875 Hogares
Paraguay	Encuesta de Hogares	Dirección General de Estadística, Encuestas y Censos	Agosto a Noviembre Anual	Nacional	4667 Hogares
Perú	Enc. Nac. de Hogares sobre Niveles de Vida y Pobreza	Instituto Nacional de Estadísticas e Informática	Variable	Nacional	No disponible
República Dominicana	Encuesta Nacional de Fuerza de Trabajo	Banco Central de la República Dominicana	Julio Anual	Nacional	5546 Hogares
Uruguay	Encuesta Continua de Hogares	Instituto Nacional de Estadística	Semestre Dos al año	Urbana	Año 1995 20057 Hogares
Venezuela	Encuesta de Hogares por Muestreo	Oficina Central de Estadística e Informática	Semestre Dos al año	Nacional	18702 Hogares

Fuente: CEPAL, División de Estadística y Proyecciones Económicas.

Cuadro 2
 AMÉRICA LATINA: TEMAS INVESTIGADOS EN EL CUESTIONARIO CENTRAL DE LAS ENCUESTAS
 DE HOGARES DE PROPOSITOS MÚLTIPLES, 1995.
 (Número de preguntas incluidas en cada tema)

	Argentina	Bolivia	Brasil a/b/	Chile		Colombia		Costa Rica	Ecuador	El Salvador	Honduras	México	Panamá	Paraguay	Perú	República Dominicana	Uruguay	Venezuela
				INE	CASEN c/d/	Urbana	Rural											
Vivienda	6	.	30	.	22	17	17	.	.	25	12	11	.	20 ^{e/}	23	21	15	21
Características generales	5	4	7	4	8	4	4	4	3	5	4	5	3	5	5	4	5	5
Migración	8	3	12	.	1	3	2	1	4	.	2	1	1	4	2	3	3	2
Educación	6	5	22 ^{f/}	2	10	3	3	7	3	14	3	6	3	6	12 ^{g/}	6	6	5
Ocupación	24	46	133 ^{h/}	29	25	35	35 ^{h/}	31	28	28	25	49	25	124	32	16	41	38
Ingreso:																		
Nº de preguntas	15	8	16	11	28	6	5	5	7	22 ^{h/}	16	2	9	9	33	2	67	5
Concepto de ingreso y invalides	IPM+T	IPM+T	IT	IT+AI+ A	IT+AI+ A	IP+T	IPM+T	IP+T+A	ITM	IT+AI+GA	IPM+T	SYSM	IT	IT	IT+AI+A	IPM+AI	IT+AI+A	IPM+T+AI

Fuente: CEPAL, División de Estadística y Proyecciones Económicas, sobre la base de los cuestionarios de las encuestas levantadas en 1995.

- a/ Además investiga raza, nupcialidad, fecundidad y tenencia de bienes.
- b/ Incluye trabajo de los niños.
- c/ La encuesta CASEN se realiza cada dos años, en este caso nos referimos a la encuesta levantada en 1996.
- d/ Además investiga salud, raza y seguridad ciudadana.
- e/ Además investiga tenencia de animales y equipo agrícola.
- f/ Incluye educación superior.
- g/ Además investiga acerca de la capacitación técnica para el trabajo.
- h/ Incluye los ingresos del sector agrícola.
- i/ ITM: Ingreso total monetario.
- IPM: Ingreso primario monetario.
- IT: Ingreso total.
- T: Transferencias.
- IP+T: Ingreso primario más transferencias.
- A: Autoconsumo.
- AI: Arrendo imputado.
- SYSM: Sueldos y salarios monetarios.
- GA: Ganancia agrícola.

Cuadro 3

AMÉRICA LATINA: CARACTERÍSTICAS DE LA VIVIENDA INVESTIGADAS EN LAS ENCUESTAS DE
HOGARES DE PROPÓSITOS MÚLTIPLES, 1995.

	Argentina	Bolivia	Brasil	Chile		Colombia		Costa Rica	Ecuador	El Salvador	Honduras	México	Panamá	Paraguay	Perú	República Dominicana	Uruguay	Venezuela
				INE	CASEN a/	Urbano	Rural											
Tipo de vivienda	X	-	X	-	X	X	X	-	-	X	X	X	-	X	X	X	X	X
Materiales:																		
Techo	-	-	X	-	X	-	-	-	-	X	-	X	-	X	X	X	-	X
Paredes	X b/	-	X	-	X	X b/	X b/	-	-	X	X	X	-	X	X	X	-	X
Piso	-	-	-	-	X	X	X	-	-	X	X	X	-	X	X	X	-	X
Tenencia de la vivienda	X	-	X c/	-	X d/	X c/	X c/	-	-	X c/d/	X	X	-	X	X c/d/	X	X c/d/	X c/d/
Cuartos:																		
Total	X	-	X	-	X	X	X	-	-	X	X	X	-	X	X	X	X	X
Dormitorios	-	-	X	-	X	X	X	-	-	X	X	X	-	X	X	X	X	X
Relativo a cocina o combustible para cocinar	-	-	X	-	-	-	X	-	-	X	-	-	-	X	X e/	X	X	-
Agua:																		
Origen	-	-	X	-	X	X	X	-	-	X	X	X f/	-	X	X	X	X	X
Abastecimiento	X	-	X g/	-	X	-	-	-	-	X	X	-	-	X	X	X	X	-
Servicio sanitario:																		
Servicio higiénico	X	-	X	-	X	X	X	-	-	X h/	X	-	-	X i/	X	X	X	X i/
Sist. de elim. de excretas	-	-	X	-	X	X	X	-	-	X	X	X	-	X j/	X	X	X	X
Electricidad	X	-	X	-	X	-	-	-	-	X	X	X	-	X	X	X	X	X
Tenencia de bienes durables	-	-	X	-	-	-	-	-	-	X	-	-	-	X y/	X	X	X	X

Fuente: CEPAL, División de Estadística y Proyecciones Económicas, sobre la base de los cuestionarios de las encuestas levantadas en 1995.
X= indica que la característica o variable es investigada en la encuesta.

a/ La encuesta CASEN se realiza cada dos años, en este caso nos referimos a la encuesta levantada en 1996.

b/ Paredes exteriores.

c/ Además investiga el monto del alquiler.

d/ Además investiga el monto del alquiler imputado.

e/ También investiga acerca del gasto en este combustible.

f/ Sólo investiga si tiene agua potable.

g/ Además investiga si tiene filtro de agua.

h/ Además investiga si tiene ducha.

i/ También investiga acerca de la eliminación de la basura.

j/ Investiga acerca de la tenencia de transporte propio, animales y equipo agrícola.

Cuadro 4

AMÉRICA LATINA: CARACTERÍSTICAS GENERALES DE LA POBLACIÓN INVESTIGADAS EN LAS
ENCUESTAS DE HOGARES DE PROPOSITOS MÚLTIPLES, 1995.

(Número de preguntas incluidas en cada tema)

	Argentina	Bolivia	Brasil a/	Chile		Colombia		Costa Rica	Ecuador	El Salvador	Honduras	México	Panamá	Paraguay c/	Perú	República Dominicana	Uruguay	Venezuela
				INE	CASEN a/b/	Urbano	Rural											
Parentesco	1	2d/	2d/	1	3d/	1	1	1	1	1	1	1	1	1	1	1	1	2d/
Sexo	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Edad	2	1	1	1	1	1	1	1	1	2	1	2	1	2	1	1	1	1
Estado civil (límite de edad)	Sin límite	Sin límite	-	15 y más años	12 y más años	Sin límite	Sin límite	10 y más años	-	12 y más años	Sin límite	12 y más años	-	Sin límite	12 y más años	Sin límite	Sin límite	12 y más años

Fuente: CEPAL, División de Estadística y Proyecciones Económicas, sobre la base de los cuestionarios de las encuestas levantadas en 1995.

a. Además investiga raza.

b/ La encuesta CASEN se realiza cada dos años, en este caso nos referimos a la encuesta levantada en 1996.

c/ Además investiga idioma.

d/ Además investiga las relaciones familiares.

Cuadro 5

AMÉRICA LATINA: VARIABLES DE MIGRACIÓN INVESTIGADAS EN LAS ENCUESTAS
DE HOGARES DE PROPÓSITOS MÚLTIPLES, 1995.

	Argentina	Bolivia	Brasil	Chile		Colombia		Costa Rica	Ecuador	El Salvador	Honduras	México	Panamá	Paraguay	Perú	República Dominicana	Uruguay	Venezuela
				INE	CASEN a/	Urbano	Rural											
Lugar de nacimiento	X	-	X	-	X	-	-	-	X	-	-	X	-	X	X	X	X	-
Dónde vivía antes	X	Xb/	Xc/	-	Xe/	Xd/	Xe/	Xd/	Xe/	-	Xe/	-	Xb/	Xb/c/	Xc/	-	Xe/	Xe/
Razón del cambio	-	X	-	-	-	-	-	-	-	-	-	-	-	X	-	X	-	-
Tiempo de residencia	X	-	X	-	X	-	X	-	X	-	X	-	-	-	-	X	X	X

Fuente: CEPAL, División de Estadística y Proyecciones Económicas, sobre la base de los cuestionarios de las encuestas levantadas en 1995.

X= indica que la característica o variable es investigada en la encuesta.

a/ La encuesta CASEN se realiza cada dos años, en este caso nos referimos a la encuesta levantada en 1996.

b/ Donde vivía hace un año.

c/ Donde vivía hace cinco años.

d/ Donde vivía hace dos años.

e/ Sin referencia de tiempo.

Cuadro 5
 AMÉRICA LATINA. CARACTERÍSTICAS EDUCACIONALES INVESTIGADAS EN LAS ENCUESTAS DE
 HOGARES DE PROPÓSITOS MÚLTIPLES. 1995.

	Argentina	Bolivia	Brasil a/	Chile		Colombia		Ecuador	El Salvador	Honduras	México	Panamá	Paraguay	Perú	República Dominicana	Uruguay	Venezuela
				INE	CASEN b/	Urbano	Rural										
Límite de edad de la pob.	Sin límite	5 y +	Sin límite	5 y +	Variable	5 y +	5 y +	10 y + c/	4 y +	Sin límite	6 y +	5 a 18	5 y + d/	Sin límite	Sin límite	Sin límite	3 y + e/
Afabetismo	X	X	X	X f/	X	X	X	X	X	X	-	-	-	X	-	-	X
Asistencia preescolar	X	-	X	X	X	-	X	-	-	X	-	-	-	X	X	X	-
Asistencia escolar	X	X	Xg/	X	X	X	X	X	X	X	-	X	X	X	X	X	X
Razón de no asistencia	-	X	-	-	X	-	Xiv/	-	X	-	-	X	X	-	-	-	X
Último nivel aprobado	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Último año aprobado	X	-	X	X	X	-	-	-	X	X	X	X	-	X	X	X	X
Nivel al que asiste	X	-	X	X	X	-	-	-	X	X	-	-	-	X	X	X	-
Grado (año) al que asiste	X	-	X	X	X	-	-	-	X	X	-	-	-	X	X	X	-
Tipo de establecimiento	-	-	-	-	X	-	-	-	X	-	-	-	X	-	X	X	-
Jornada	-	-	-	-	X	-	-	-	-	-	-	-	-	-	X	-	-
Gasto en educación	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-
Recibe merienda	-	-	-	-	X	-	-	-	X	-	-	-	-	-	-	-	-
Instrucción no formal	-	-	-	-	-	-	X	-	-	-	-	-	-	X	-	-	-

Fuente: CEPAL, División de Estadística y Proyecciones Económicas, sobre la base de los cuestionarios de las encuestas levantadas en 1995.

X= indica que la característica o variable es investigada en la encuesta.

a/ Además se investiga acerca de la educación supletiva, para los mayores de 15 años.

b/ La encuesta CASEN se realiza cada dos años, en este caso nos referimos a la encuesta levantada en 1996.

c/ Investiga acerca del grado y nivel al que se matriculó en 1995.

d/ Las preguntas de educación se realizan a todos los mayores de 5 años, y la no asistencia a los menores de 35 años.

e/ La asistencia escolar se investiga para la población entre 3 y 30 años y el nivel educativo para los de 3 años y más.

f/ Se puede obtener en forma indirecta.

g/ Además se investiga si el curso es seriado (anual).

h/ Se pregunta sólo a personas hasta 35 años y el nivel educativo para los de 5 y más años.

i/ Se investiga alfabetismo para mayores de 10 años y asistencia para mayores de 6 años.

j/ Incluye un módulo de evaluación de programas sociales (educación) para amas de casa y jefes de hogar.

k/ Investiga acerca de cursos de capacitación para el trabajo.

Cuadro 7
AMÉRICA LATINA. CARACTERÍSTICAS OCUPACIONALES INVESTIGADAS EN LAS ENCUESTAS DE HOGARES DE PROPOSITOS MÚLTIPLES. 1995.

	Argentina	Bolivia	Brasil	Chile		Colombia		Costa Rica	Ecuador	El Salvador	Honduras	México	Panamá	Paraguay	Perú	República Dominicana	Uruguay	Venezuela	
				INE	CASEN a/	Urbano	Rural												
Condición de actividad																			
Límite de edad	Sin límite	7 y +	10 y + b/	15 y +	12 y +	10 y + b/	10 y +	10 y +	10 y +	10 y +	10 y +	12 y +	15 y +	7 y +	6 y +	10 y +	14 y +	10 y +	
Nº de preguntas	11	6	4	4	5	8	8	8	8	8	4	15	7	8	5	5	6	5	
Periodo de referencia	Semana fija	Semana pasada	Semana fija	Semana pasada	Semana pasada	Semana pasada	Semana fija	Semana fija	Semana pasada	Semana pasada	Semana pasada								
OCUPADOS																			
Ocupación principal																			
Categoría ocupacional	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Rama de actividad	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Ocupación	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Tipo de establecimiento	-	X	X	-	-	-	-	-	-	-	-	X	-	X	X	X	-	-	-
Tamaño de establecimiento	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Cobertura de seguro social	-	-	X	-	X	-	-	-	-	X	-	-	-	X	X	-	-	X	X
Ocupación secundaria	-	-	X	-	-	-	-	-	-	-	-	-	-	X	X	-	-	-	-
Categoría ocupacional	-	-	X	-	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Rama de actividad	-	-	X	-	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Ocupación	-	-	X	-	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Tipo de establecimiento	-	-	X	-	-	-	-	-	-	-	-	-	-	X	X	X	X	X	X
Tamaño de establecimiento	-	-	X	-	-	-	-	-	-	-	-	-	-	X	X	X	X	X	X
Trab. permanente o temporal	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Hrs. efectivas O.P.	X	-	-	X	-	-	-	-	-	X	X	X	X	X	X	X	X	X	X
Hrs. efectivas O.S.	X	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hrs. extras	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Total horas (suma)	X	-	-	X	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-
Hrs. normales O.P.	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Hrs. normales O.S.	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Total horas (suma)	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-
Busca otra ocupación	X	-	X	-	-	X	-	-	-	-	-	-	-	-	X	-	-	-	-

Cuadro 7

(continuación)

AMÉRICA LATINA: CARACTERÍSTICAS OCUPACIONALES INVESTIGADAS EN LAS ENCUESTAS DE HOGARES DE PROPOSITOS MÚLTIPLES, 1995.

	Argentina	Bolivia	Brasil	Chile		Colombia		Costa Rica	Ecuador	El Salvador	Honduras	México	Panamá	Paraguay	Perú	República Dominicana	Uruguay	Venezuela	
	Semana fija	Semana pasada	Año anterior	2 meses	Semana pasada	Último año	Semana pasada	Semana pasada	Semana fija	1 mes	1 mes y 1 semana	Semana pasada	3 meses	Semana pasada	2 meses	Semana pasada	Semana pasada	2 meses	
DESOCUPADOS																			
Búsqueda de trabajo																			
Tipo de búsqueda			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Tiempo de desempleo						X	X	X											
Tiempo de búsqueda	X	X		X	X		X	X	X	X	X	X	X	X	X	X	X	X	X
Experiencia laboral anterior	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X
Categoría ocup. anterior	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X
Rama de activ. anterior	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X
Ocupación anterior	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X
Tipo de establec. anterior			X																
Tamaño de establec. anterior	X			X															
Categoría ocup. buscada				Xc/		X													
Rama de activ. buscada					X	X	X												
Ocupación buscada					X	X	X												
INACTIVOS																			
Tipo de inactividad	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Deseos de trabajar		X																	
Experiencia laboral anterior		X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X
Categoría ocup. anterior			X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X
Rama de activ. anterior			X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X
Ocupación anterior			X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X
Tamaño de establec. anterior				X															
Actividades marginales			X			X													

Fuente: CEPAL, División de Estadística y Proyecciones Económicas, sobre la base de los cuestionarios de las encuestas levantadas en 1995.

X= indica que la característica o variable es investigada en la encuesta.

a/ La encuesta CASEN se realiza cada dos años, en este caso nos referimos a la encuesta levantada en 1996.

b/ Además se investiga el trabajo de los niños del hogar.

c/ Se investiga como clase de trabajo.

d/ Se investiga sólo para los inactivos con deseos de trabajar.

Cuadro 8

AMÉRICA LATINA: CONCEPTO DE INGRESO INVESTIGADO EN LAS ENCUESTAS DE HOGARES DE PROPÓSITOS MÚLTIPLES, 1995.

	Argentina		Bolivia	Brasil	Chile		Colombia		Costa Rica	Ecuador	El Salvador	Honduras	México	Panamá	Paraguay	Perú	República Dominicana	Uruguay	Venezuela
					INE	CASEN a/	Urbano	Rural											
Ocupación principal	X		X	X	Xc/	X	Xc/	X	X	X	X	X	X	X	Xb/	X	X	X	X
Sueldos y salarios monetarios	d/			X	Xc/	X	Xc/	X	d/		Xc/	d/				X		X	d/
Sueldos y salarios en especie	X		X	X	Xc/	X	Xc/	X	X	X	X	X	X			X		X	
Ganancias monetarias	d/			X	Xc/	X	Xc/		X		Xc/	d/						X	d/
Ganancias en especie	X		X	X	Xc/	X	Xc/		X		Xc/	X						X	
Ocupación secundaria					Xc/	X	Xc/	X			Xc/	X		X	Xb/			X	
Sueldos y salarios monetarios			X	X	Xc/	X	Xc/	X	X	X	X	X				X		X	
Sueldos y salarios en especie	d/			X	Xc/	X	Xc/				Xc/	d/						X	d/
Ganancias monetarias	X		X	X	Xc/	X	Xc/	X	X	X	X	X				X		X	
Ganancias en especie	d/			X	Xc/	X	Xc/				Xc/	d/						X	d/
Transferencias					Xc/	X	Xc/	X			Xc/	X						X	
Jubilaciones y pensiones	X		X	X	X	X			X	X	X	X		X	X	X		X	
Otras				X	X	X			X		X	Xe/		X	X	X		X	
Ingresos de capital	X		Xb/				h/	h/	X	Xb/	Xb/			X					h/
Arrendos				X	X	X						X			X	X		X	
Intereses					X	X						X			X	X		X	
Dividendos						X						X				X		X	
Otros ingresos	X		X	X	X	X				X	Xg/			X	X	X		X	
Arriendo imputado casa propia					X														
Concepto de Ingreso h/	ITM		ITM	IT	IT+A+AI	IT+A+AI	IP+T	IPM+T	IP+T+A	ITM	IT+AI+GA	IPM+A	YSM	IT	IT	IT+AI+A	IPM+AI	IT+AI+A	IPM+T
Nro. de preguntas	15		8	16	11	28	6	5	5	7	22	15	2	9	9	27	3	67	5

Fuente: CEPAL, División de Estadística y Proyecciones Económicas, sobre la base de los cuestionarios de las encuestas levantadas en 1995.

X= indica que la característica o variable es investigada en la encuesta.

a/ La encuesta CASEN se realiza cada dos años, en este caso nos referimos a la encuesta levantada en 1996.

b/ Se investigan estos rubros en forma agregada.

c/ Se investiga en conjunto la actividad principal y secundaria.

d/ No valorados.

e/ Pregunta por ingresos en concepto de subsidios.

f/ Se investiga en conjunto con las transferencias.

g/ Pregunta por las remesas del exterior, ayuda de familiares y ayuda particulares.

h/ ITM: Ingreso total monetario.

IT: Ingreso total.

T: Transferencias.

IP+T: Ingreso primario más transferencias.

A: Autoconsumo.

AI: Arriendo imputado.

YSM: Sueldos y salarios monetarios.

GA: Ganancia agrícola.

**LECCIONES PARA EL DISEÑO DEL CUESTIONARIO
DERIVADAS DE LA EXPERIENCIA DE LA ENNIV/LSMS***

**MARGARET GROSH
(BANCO MUNDIAL)**

* Transparencias utilizadas por la autora durante su presentación de este tema en el Taller.

LECCIONES PARA EL DISEÑO DEL CUESTIONARIO

Diseño “óptimo”

- No hay un solo diseño óptimo
 - diferencias en objetivos, circunstancias, etc.
- Pero hay pautas para llegar a una encuesta exitosa

Tres Temas

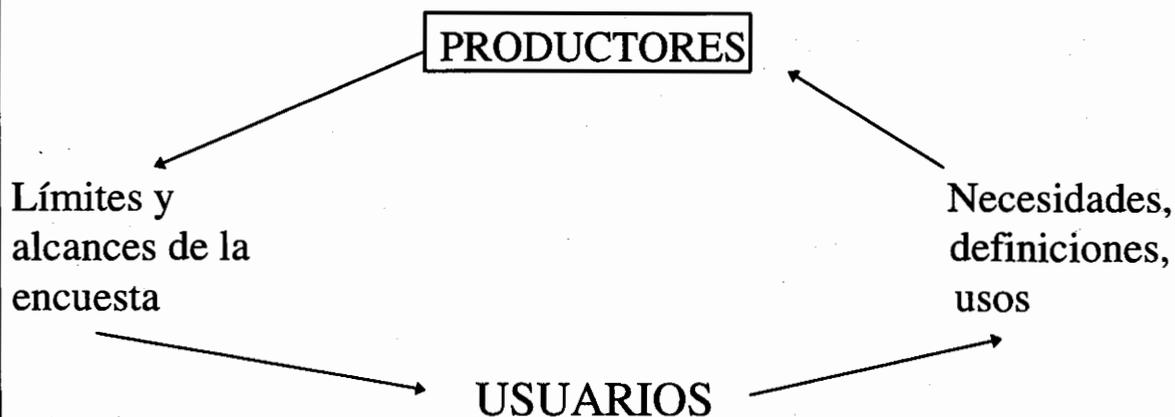
1. Vínculos entre productores y usuarios de datos
2. Contenido básico de la encuesta
3. Control de calidad

1. Vínculos con los Usuarios

- Fundamentales para el éxito
- Deben incluirse en todas las etapas
 - definición de propósitos y parámetros de la encuesta
 - formulación de cuestionarios
 - preparación del resumen estadístico, documentación, archivos de datos
 - análisis más sofisticado
 - evaluación de la encuesta

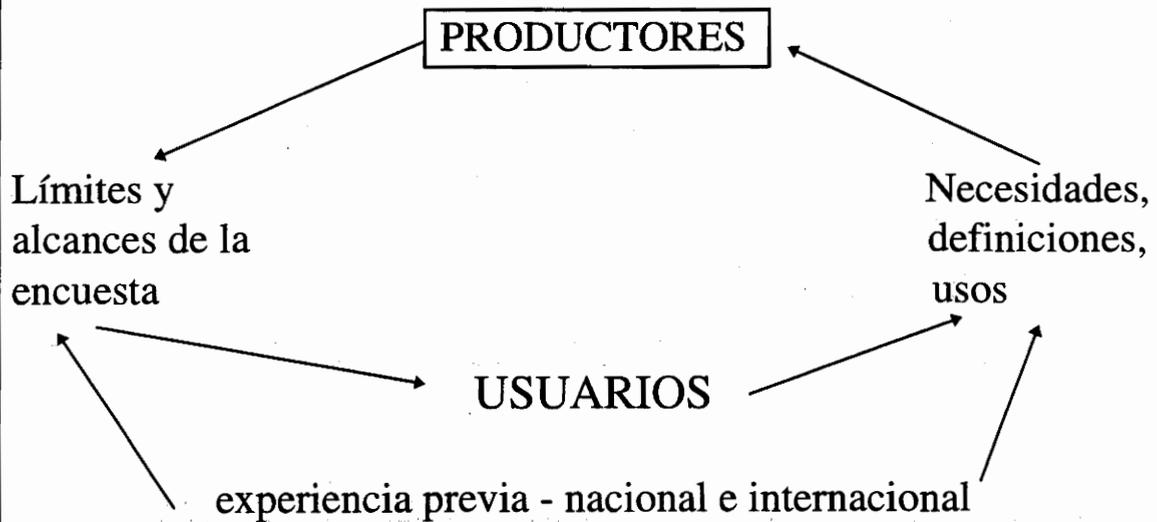
1. Vínculos con los Usuarios

Elementos de educación mutua



1. Vínculos con los Usuarios

Elementos de educación mutua requerida



Aportes de la experiencia LSMS

- Manual de implementación
- Manual de diseño de cuestionarios
- Ejemplares
 - cuestionarios, manuales de encuestadores, etc.
- Contactos
 - <http://www.worldbank.org/lsms/lsmshome.html>
 - lsms@worldbank.org

1. Vínculos con Usuarios

- **Medios formales**

- comité asesor de la encuesta
- servicio de difusión de datos de la ONE
- seminarios, foros, charlas, etc.
- publicaciones

- **Medios informales**

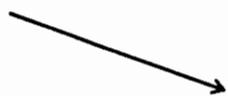
- consultas ad hoc
- contactos informales

Tema 2: Contenido de los cuestionarios

MARCO CONCEPTUAL

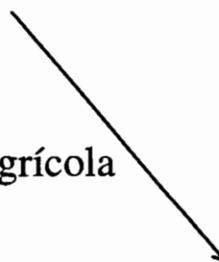
- **Objetivos Sociales**

Reducir pobreza
Reducir mortalidad infantil
Elevar tasas de enseñanza
Mejorar eficiencia agrícola
...etc...



- **Instrumentos**

Clínicas
Escuelas
Agentes de extensión agrícola
...etc...



- **Impacto final**

Personas
Hogares
Fincas
...etc...

Implica una encuesta integrada

- entender como lograr progreso
 - ¿cuáles programas/servicios deben ofrecerse?
 - ¿a quiénes? ¿a dónde?
- para medir progreso hacia los objetivos
 - objetivos finales, tales como la desnutrición, el analfabetismo, la pobreza
 - objetivos intermedios, tales como el recibo de suplementos, la asistencia escolar, el empleo

El contenido detallado varía según los objetivos

- Medir bienestar
- Describir políticas
- Monitoreo de los mismos
- Modelación y simulación

Variación - ejemplo de antropometría

- no incluir nada
- medir los niños menores de 5 años
 - describir bienestar, focalización de programas
- medición repetida
 - monitoreo
- medir todos los miembros
 - impacto de programas de nutrición en niños
 - impacto de nutrición en aprendizaje
 - efectos de nutrición en productividad de adultos

Se selecciona el diseño de la encuesta según:

- Objetivo analítico
 - puede diferir según el tema
- Capacidad institucional
 - de recolección de datos
 - capacidad analítica
- Otros datos disponibles
- Condicionantes financieros

Opciones comunes *:

<u>Cuestionario</u>	<u>Periodicidad</u>	<u>Muestra</u>
ENNIV Completo	cada 4 - 5 años	2000 - 5000
Módulos básicos y rotativos	anual	2000 - 8000
ENNIV abreviada	anual	2000 - 12000

* Nueva encuesta o reformulación de una encuesta existente

Paquete básico de una encuesta integrada (módulos básicos)

- Cuestionario del hogar
 - valorización del consumo como medición de bienestar monetario
 - indicadores de bienestar no monetario
 - uso de servicios sociales y recibo de subsidios
 - temas de interés puntual
- Cuestionario de comunidad
- Cuestionario de precios
- Detalles varían según los objetivos

3. Control de calidad

- importante
- posible
- barreras son burocráticas y financieras, no técnicas

3. Herramientas para la calidad

- hay muchas técnicas
- detalle adecuado
- proceso institucional importa
 - consulta adecuada con varios puntos de vista
 - usuarios
 - productores
 - experiencia previa
 - tiempo adecuado

**DISEÑO DEL CUESTIONARIO DE LA “PESQUISA NACIONAL
POR AMOSTRA DE DOMICILIOS - PNAD”**

**ROSÂNGELA ANTUNEZ PEREIRA ALMEIDA
(IBGE)**

DISEÑO DEL CUESTIONARIO DE LA "PESQUISA NACIONAL POR AMOSTRA DE DOMICILIOS - PNAD"

El sistema de encuestas domiciliarias, implantado progresivamente en Brasil a partir de 1967 con la creación de la Pesquisa Nacional por Amostra de Domicilios - PNAD, tiene la finalidad de producir informaciones básicas para el estudio del desarrollo socioeconómico del país.

Se trata de una encuesta por muestreo de domicilios, de periodicidad anual, que - por tener propósitos múltiples- investiga diversos temas; unos de carácter permanente, como las características generales de la población, educación, trabajo, ingresos y vivienda, y otros con periodicidad variable, como las características sobre migración, fecundidad, nupcialidad, salud, nutrición y otros, que son incluidos en el sistema con periodicidad no definida, de acuerdo con las necesidades de información del país.

Desde mediados de la década de 1980 y hasta 1990 la PNAD utilizó dos instrumentos de investigación: el "Cuestionario Básico" (PNAD 1.01), para la investigación de los temas permanentes de la encuesta, y un "Cuestionario Suplementario" (PNAD 1.02), para la investigación de otros temas. La ventaja de la utilización del cuestionario básico es que, manteniéndose inalterado de un año para otro, el conjunto de preguntas a ser investigadas, los sistemas de entrada de datos, la crítica, la expansión y la tabulación de la información de los temas básicos no requieren ser modificados, a no ser que digan relación con los períodos de referencia de las variables que se modifican anualmente (por ejemplo, el valor del salario mínimo usado para la tabulación). Para los temas suplementarios es necesaria la definición de todo un sistema de depuración.

A partir de 1992, cuando fue implementada la encuesta revisada, la recolección de información pasó a ser hecha en un único instrumento, manteniéndose la misma idea del cuestionario básico y suplementario. Esto es, los temas permanentes son investigados mediante el mismo conjunto de preguntas en cada año.

La revisión de la encuesta provocó un acentuado crecimiento en la investigación de los temas básicos, en relación al utilizado en la década de 1980. Así, fue necesario cambiar el layout del instrumento, que pasó a ser presentado en la forma de cuaderno y estructurado en partes. De esa manera, cada tema de la encuesta pasó a ser investigado como una parte del cuestionario, permitiendo que en cada año de realización de la encuesta se pudiese excluir una o más partes/temas, incluir nuevas partes/temas, o bien volver con una excluida anteriormente.

En 1992, la composición del cuestionario estaba distribuida así:

1. Identificación y control (15 preguntas).
2. Características de la Unidad de Domicilio (30 preguntas).

3. Identificación de los Moradores (3 preguntas).
4. Características Generales de los Moradores (7 preguntas).
5. Características de Migración de los Moradores (12 preguntas).
6. Características de Educación (11 preguntas).
7. Características del Trabajo de los Niños de 5 a 9 años de edad (13 preguntas).
8. Características de la Enseñanza "Supletiva" de los Moradores de 14 años o más de edad (11 preguntas).
9. Características del Trabajo y el Ingreso de los Moradores de 10 años o más de edad (125 preguntas).
10. Características de Nupcialidad de los Moradores de 15 años o más de edad (4 preguntas).
11. Características de Fecundidad de las Mujeres Moradoras de 15 años o más de edad (11 preguntas).

En 1993 y 1995 el cuestionario se mantuvo inalterado. En 1996 fueron excluidas las partes 7, 8 y 10 e incluido el tema "Movilidad Social" en la parte 12. En 1997 se excluyó la parte 12, además de las partes ya retiradas en 1996. Así, en 1998 se volverá a incluir en la encuesta la parte 7 y serán investigados nuevos temas como "Características de Salud de los Moradores", parte 13, y "Características de Movilidad Física", parte 14.

A pesar que el cuestionario impresiona por su tamaño, debemos considerar que la función de este instrumento es traducir en forma simple y encadenada toda la complejidad de la investigación, haciendo que la entrevista fluya más fácilmente y cuidando de obtener la mejor calidad de la información. Esto significa transformar conceptos y definiciones en una serie de preguntas simples, facilitando la comprensión del entrevistado y retirando del entrevistador la responsabilidad de transmitir al entrevistado toda la conceptualización existente detrás de la investigación, evitando el riesgo que las indagaciones que no están explicitadas en el cuestionario sean omitidas o enunciadas en forma incorrecta y, consecuentemente, las respuestas no reflejen lo que se pretendía investigar. El cuestionario está estructurado de forma de investigar las características de hasta 4 moradores. Los recortes de las páginas del instrumento permiten que se identifique, en lo alto de la página, a la persona que está siendo entrevistada, el sexo, la fecha de nacimiento y el número de trabajos. En el caso de los domicilios con más de 4 moradores (tamaño medio del domicilio en Brasil) se utiliza otro formulario. Para las entrevistas no realizadas, el cuestionario es reducido a solamente la carátula.

Las preguntas, en su mayoría, son precodificadas; otras, que no se encuadran en esta forma, sufren posteriormente un proceso de codificación automática. El flujo de la entrevista está bien definido a través de los comandos de salto existentes en cada pregunta, lo que permite el encadenamiento de informaciones facilitando la entrevista.

En cada año el cuestionario es actualizado en lo que dice relación con las fechas, períodos de referencia y fecha de nacimiento límite para la inclusión de un morador en una parte determinada.

Es claro que para cada nuevo tema incluido en la investigación, continúa siendo necesario el desarrollo del sistema de depuración, que todavía no es suficientemente ágil para permitir trabajar con todos los temas simultáneamente. Así, como en la década del ochenta, la divulgación de los datos de los temas inéditos se hace en una fecha posterior a la del cuerpo básico.

Cabe resaltar que en las partes mantenidas o retornadas no se deben hacer alteraciones internas, esto es, incluir, excluir o modificar sustancialmente las preguntas que las componen, para no retardar la divulgación de los resultados, dado que son necesarias alteraciones en los programas de entrada de datos, crítica (la PNAD tiene hoy un proceso de depuración de los datos que envuelve cerca de 15.000 reglas de crítica) y tabulación.

**ENCUESTAS DE PROPÓSITOS MÚLTIPLES:
LECCIONES Y DIRECCIONES PARA SU MEJORAMIENTO***

**EDMUNDO BERUMEN, CONSULTOR
(BID)**

* 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.....	211
II. PRÁCTICAS DE CAMPO CUESTIONABLES	
v/s BUENAS PRÁCTICAS.....	212
A. Sustitución.....	212
B. Más de un informante individual por hogar.....	212
C. Selección directa de hogares.....	213
D. Retroalimentación al personal de campo.....	213
E. Pruebas piloto v/s ensayos.....	213
F. Capacitación del personal de campo.....	214
G. Relación supervisor-encuestadores.....	214
H. Controles de campo.....	215
I. Retroalimentación a la unidad de diseño.....	215
J. Devolver información a la unidad observada.....	215
III. PRÁCTICAS EN LA UNIDAD DE DISEÑO.....	216
A. Objetivos; diseño conceptual.....	216
B. Cambios al cuestionario.....	217
C. Documentación de la encuesta.....	218
D. Diseño muestral.....	218
E. Acceso, difusión y análisis.....	220
ANEXO 1: Cuadros.....	221
ANEXO 2: Documento elaborado para el Taller del Programa MECOVI de 1995.....	224

ENCUESTAS DE PROPÓSITOS MÚLTIPLES LECCIONES Y DIRECCIONES PARA SU MEJORAMIENTO

I. INTRODUCCIÓN

En agosto de 1995 el consultor preparó un documento que narra la situación general que guarda la infraestructura para levantar encuestas de hogares en los países de América Latina y el Caribe, tratando entre otros temas los siguientes:

- * tamaños de muestra y diseños de muestreo,
- * desagregación de resultados y periodicidad de las encuestas,
- * trabajo de campo,
- * aprovechamiento de datos existentes,
- * propósitos múltiples vs tiempo de entrevista.

Con el trabajo anterior como antecedente, y en el contexto del programa MECOVI, trataremos algunos temas complementarios que apuntan en distintas direcciones para el mejoramiento de los actuales programas de encuestas, y prácticas metodológicas asociados a estos.

Como referencia general, mencionamos algunos de los parámetros metodológicos interesantes, en varios de los programas de encuestas de la región:

- a) Un tiempo promedio de entrevista que oscila entre 10 minutos y 8 horas a lo largo de 7 días de contacto con la unidad informante.
- b) Una relación supervisor: encuestador que va desde 1:1 hasta 1:10.
- c) Esquemas de muestreo que cumplen con rigor los lineamientos para obtener una muestra probabilística, y otros que relajan el rigor en algunas etapas, de manera consciente o sin saberlo.
- d) Distintas revisitas a la unidad seleccionada para lograr la entrevista, y diferentes prácticas de sustitución de unidades difíciles de medir.
- e) Altas tasas de respuesta, que oscilan entre 80%-95%, que en ocasiones esconden prácticas de sustitución.

f) Unidades Primarias de Muestreo (UPM) cuyo tamaño oscila entre 20-600 viviendas.

g) Toma de muestra de UPM que varía entre 3-25 viviendas.

h) Tamaños de muestra (a nivel diseño) por ronda que oscilan entre 250-50,000 viviendas.

i) Coberturas que van desde una sola ciudad o área metropolitana hasta encuestas nacionales con varios dominios de estudio.

II. PRÁCTICAS DE CAMPO CUESTIONABLES v/s BUENAS PRÁCTICAS

A. Sustitución.

La sustitución de unidades de la población de interés (sean las finales o la última de "área" que contiene a las finales), por la razón que sea (dificultad para localizarla, rechazo directo, rechazo disfrazado, fallas del marco de muestreo, etc.) no es una práctica recomendable. Por más controles que se instalen para su "buena aplicación", el personal de campo eventualmente encuentra la manera de eludirlos, introduciendo vicios que facilitan el que "cumpla" con cargas de trabajo asignadas en los tiempos permisibles.

Lo recomendable es prever (y luego controlar) el nivel aceptable de los diferentes componentes de la "no-respuesta", y compensar el tamaño de muestra para tomar en cuenta estas mermas. Ya que es común en los programas de encuesta el tener un "cuestionario de hogar" donde se recoge su composición, características básicas de sus miembros y su vivienda, y otro "individual" (ya sea como módulo o cuestionario aparte), es necesario el tomar en cuenta la tasa de no-respuesta a nivel hogar y cuestionario individual, y compensar adecuadamente ambas.

B. Más de un informante individual por hogar.

El aplicar cuestionarios individuales en más de un miembro por hogar del mismo grupo de interés, aumenta el riesgo de contaminación de entrevistas posteriores a la primera, y en general eleva la tasa de no-respuesta individual. Desde otra perspectiva, el tener más de una medición individual por hogar no aumenta en la misma proporción la información obtenida por el efecto de la correlación intraclase que generalmente es positiva.

Es fácil contar con formatos de campo que identifiquen y enumeren a los individuos de la población de interés, agrupados por hogar, aplicando sobre estas listas un sencillo esquema de muestreo sistemático que disminuya considerablemente la probabilidad de seleccionar más de uno por hogar, y a la vez mantener bajo control el número de viviendas a seleccionar.

C. Selección directa de hogares.

En la práctica, la identificación y ubicación en campo de la muestra seleccionada a nivel gabinete de acuerdo al diseño particular utilizado no es una operación sencilla. En particular, la enumeración e identificación de hogares ("conjunto de personas que") en algunos marcos de muestreo, induce a diseños que se basan en la selección "directa" de estos a nivel gabinete. Las dificultades se presentan en campo, donde es complicado identificar inequívocamente al hogar seleccionado, lo que aumenta el riesgo de introducir sesgos de selección.

El uso de la vivienda en las últimas etapas de selección, facilita su ubicación correcta por ser una unidad espacial, "de área". Ya ubicada ésta, es fácil aplicar una regla ya sea de inclusión de todos los hogares que habitan en la misma vivienda, o de selección de uno de ellos mediante un mecanismo aleatorio/insegado al momento de hacer la entrevista. Si bien los encuestadores pueden tener dificultad en encontrar un domicilio/vivienda seleccionado, esto es más fácil que identificar directamente un hogar particular.

D. Retroalimentación al personal de campo.

El personal de campo da la cara ante el informante tanto a nivel institucional como sobre el tema sustantivo bajo investigación. El que no está cabalmente informado y convencido de lo que hace, se traduce en inseguridades que a su vez externa ante el informante, creando un ambiente incómodo que impacta en forma negativa la calidad de los datos que recoge.

En programas continuos de encuesta, no implica mayor esfuerzo el dedicarle tiempo suficiente para periódicamente enterar al personal de campo sobre los usos pasados y previstos de la información que recogen, en un lenguaje accesible a ellos y al de los informantes con los que tratan. Esto se puede lograr mediante la entrega de un "reporte ligero" que reseñe algunos resultados/usos del estudio, mezclado con "noticias" sobre anécdotas interesantes del trabajo de campo mismo.

E. Pruebas piloto v/s ensayos.

Las pruebas piloto por lo general se constriñen a ensayos sencillos donde lo fundamental es mejorar el lenguaje y mecánica del cuestionario. Usualmente queda pendiente un verdadero "ensayo con vestuario", que cubra aspectos de logística de campo, estimación de la productividad del personal de campo, estimar tiempos de traslado entre zonas en muestra, "semi-cerrar" preguntas abiertas, probar los mecanismos de "envío-recepción" de materiales de campo, ensayar la crítica-codificación, captura-validación, y producción de tabulados prioritarios.

La inversión que se realice en tiempo de preparación, ejecución y análisis; recursos humanos y materiales; y soporte administrativo necesario para llevar a cabo verdaderas pruebas piloto que permitan detectar a tiempo deficiencias graves en todos los componentes

del operativo de una encuesta, redituar en una mejor calidad de la encuesta, con claros mecanismos identificados apriori para el manejo de contingencias y eventualidades, evitando que éstas "sorprendan" al operativo de la encuesta sin preparación alguna para resolverlas.

F. Capacitación del personal de campo.

Relacionado al punto anterior, en particular la capacitación del personal de campo con frecuencia se realiza sin contar con los manuales e instructivos detallados para el correcto llenado de formatos y cuestionarios de campo; y la capacitación se reduce a una instrucción verbal, en un ambiente de gabinete, quizá con algún juego de "roles" para practicar frente al grupo lo instruido. Si además la capacitación es en "cascada", los malentendidos se multiplican y se pierde la homogeneidad de criterios y procedimientos para la correcta aplicación de la encuesta.

El aprendizaje de gabinete se ve enriquecido considerablemente cuando se cuenta con la debida anticipación con los manuales e instructivos correspondientes, y se complementa la capacitación con un ejercicio de campo, observado de cerca por los diseñadores de la encuesta, y vuelta a gabinete para resaltar lo observado, comentar aciertos y fallas, reforzar la instrucción en temas difíciles, y aplicar medidas correctivas donde se requiera. Cuando necesariamente se requiere una capacitación en "cascada", se reducen los riesgos de fomentar diversidad de criterios si se apoya a los instructores con material audiovisual (videograbado) que se filme durante la capacitación central, y se edite y duplique para su uso en las subsecuentes.

G. Relación supervisor-encuestadores.

En papel, la relación del número de encuestadores por supervisor, oscila entre 1-10. En la práctica, con frecuencia el "supervisor" es o un encuestador más (y las cargas de trabajo así lo toman en cuenta) o sólo un "vínculo administrativo" con el encuestador, a efectos de entregar/recoger cuestionarios y pagar sus servicios.

Las tareas de supervisión y revisión sustantiva "in situ", del trabajo del encuestador/equipo, es crucial para mantener un nivel de calidad aceptable en el trabajo de campo. La supervisión es, y así se debe reconocer, una tarea intrínsecamente distinta y complementaria a la del encuestador. En particular, es recomendable que el inicio de toda encuesta considere explícitamente un programa intenso de supervisión al arranque, con un alto (breve) al operativo de campo para examinar de cerca el resultado, corregir desviaciones y reforzar instrucciones cruciales. La ausencia de una real supervisión abre de inmediato la puerta al establecimiento de vicios en el personal de campo difíciles de erradicar después. Otras prácticas recomendables son: las re-entrevistas del supervisor con cuestionarios diseñados *ex professo* para ello; entrevistas del encuestador "observadas" (y luego comentadas) por el supervisor; grabación (con permiso del informante) de algunas entrevistas, para su posterior análisis en campo y en gabinete; verificación al 100% de todo tipo de "no-respuesta".

H. Controles de campo.

Si bien hay excepciones, lo común es una ausencia de controles individuales del desempeño de los encuestadores, supervisores, crítico-codificadores, y coordinadores regionales, respecto a parámetros como "tasas de no-respuesta" según componente; productividad; porcentaje de errores en aplicación correcta del cuestionario (global y por secciones); inconsistencias en la codificación; etc. El contar sólo con cifras globales de estos parámetros impide el detectar oportunamente fallas graves en algunos elementos de campo/gabinete.

El sistematizar la identificación y registro individual de estadísticas sencillas que den cuenta del desempeño del personal de campo, sólo requiere el asignar una clave de identificación del encuestador-supervisor-crítico-coordinador que se pueda ligar a cada cuestionario individual, y el diseño de formatos sencillos en los puntos de control adecuados. De esta suerte se pueden contrastar los resultados de un encuestador particular con los de su equipo, región, y global; y así para los demás elementos del operativo de campo.

I. Retroalimentación a la unidad de diseño.

Si bien la tarea principal del personal de campo es hacer un buen levantamiento de la encuesta particular, apegándose a los procedimientos definidos para la identificación de la unidad a observar, y la aplicación del cuestionario, en general se "desperdicia" el potencial que tiene la fuerza de campo para retroalimentar a la unidad de diseño, para mejoras futuras de la encuesta.

Sin que implique cargas adicionales de trabajo en detrimento de su tarea fundamental, se puede aprovechar cada ronda de la encuesta para detectar algunos elementos valiosos para la unidad de diseño, como por ejemplo: áreas importantes de nuevo crecimiento, tanto dentro de las UPM en muestra, como fuera de ellas; actualizaciones de la cartografía, incluyendo cambios en las calles (modificación en su extensión o contorno, desaparición de algunas, incorporación de nuevas, cambios de nombre, etc.); registro sistemático de términos coloquiales/regionales para identificar algunos conceptos del cuestionario; unidades y medidas particulares a distintas zonas; etc.

J. Devolver información a la unidad observada.

En encuestas que revisitan en varias rondas a la misma unidad de observación, no se acostumbra retroalimentarla de manera directa con información elaborada en base a su aportación y la de otros informantes en muestra. El no apreciar directamente cómo contribuye al estudio es una de las razones de experimentar tasas crecientes de no-respuesta en este tipo de encuesta.

No requiere un gran esfuerzo el producir brevísimos reportes con ilustraciones gráficas de indicadores interesantes a nivel general, que se distribuyan en rondas

subsecuentes de encuestas periódicas, o por correo en mediciones "ad-hoc". Lo anterior cumple con el derecho a la información que en particular tienen las unidades informantes, y sirve para establecer un mejor "rapport" entre entrevistado y entrevistador.

III. PRÁCTICAS EN LA UNIDAD DE DISEÑO

A. Objetivos; diseño conceptual.

En general, quienes diseñan y levantan encuestas (productores), ya sean organismos públicos o privados, usualmente lo hacen para uno o varios clientes (usuarios), no para consumo propio. La relación productor-usuario para determinar el objetivo y alcance de la encuesta, siempre es difícil. El usuario tiene dificultad en expresar en términos operativos lo que necesita medir, y el grado de precisión y confianza que le es suficiente. Recae entonces la responsabilidad en el productor para ayudar al usuario a definir sus requerimientos, y hacerle explícito desde el inicio las limitaciones y resultados posibles por medio de encuestas. En la ausencia de esta fase, ahí es donde nacen fallas serias de las encuestas, malas interpretaciones futuras de sus resultados, y conflictos productor-usuario por insatisfacción con los resultados. Entre las preguntas relevantes que los productores pueden hacer a los usuarios para esclarecer los objetivos están las siguientes.

¿Sobre quiénes requiere el usuario los datos? La definición explícita y amplia de la población de interés al usuario es fundamental. La mayoría de las encuestas no pueden cubrir (hacer inferencia válida) con sus procedimientos a todos los elementos de la población de interés, por problemas de acceso (zonas de riesgo, climatológicas), costo, imperfecciones en los marcos de muestreo, o problemas de comunicación en la etapa de campo (grupos monolingües, por ejemplo). El productor debe informarle al usuario de éstas limitaciones y estimar la parte de la población de interés que no quedar cubierta por la encuesta. El precisar estas limitaciones evitar el hacer generalizaciones no válidas de los resultados.

¿Qué precisión y confianza requiere el usuario de los resultados, a la luz de su costo? Ninguna encuesta es de una sola variable, de una sola pregunta. Las consideraciones de control para las distintas fuentes de error que debe contemplar el diseño de una encuesta, se deben hacer para cada parámetro de interés. Declaraciones generales sobre la precisión y confianza que los resultados de una encuesta tienen, dicen todo y nada. ¿Para qué variables o subpoblaciones de interés? (por ejemplo, mujeres en edad fértil, ancianos, discapacitados, menores de 5 años). Será necesario hacer compromisos y concentrarse en las variables principales, para los dominios de estudio prioritarios. Al final de la encuesta, es importante efectivamente calcular y presentar las estimaciones de error y confianza para las variables principales, en una forma sencilla y comprensible para el usuario. Desafortunadamente, una práctica poco común en la mayoría de las encuestas.

Por otro lado, el cálculo rutinario de errores de muestreo para variables prioritarias en todas las rondas de los programas continuos de encuesta, permite contar con la información necesaria para ratificar o rectificar los tamaños de muestra en uso, y hacer una mejor asignación de los recursos de la encuesta, ya sea para lograr mayores desagregaciones de los resultados, o fortalecer otras áreas (por ejemplo supervisión) con los recursos "ahorrados" en los casos donde se pueda disminuir el tamaño de la muestra sin sacrificar la precisión y confianza requerida en las estimaciones.

¿La encuesta es puntual o se requieren levantamientos periódicos? Por un lado, los parámetros de interés en las distintas encuestas tienen dinamismo distinto. Hay variables que cambian significativamente con más rapidez que otras. Pero, justifica esto el que las estemos midiendo con una periodicidad afín a su dinamismo? .O es más prudente el medirlas con una periodicidad acorde a la capacidad del usuario para reaccionar y tomar medidas que incidan sobre el fenómeno investigado? De qué sirve medir de nuevo el parámetro si aún no se actúa con respecto a la medición anterior? Sólo para observar su evolución? Ciertamente, se pueden construir series de datos que permitirán modelar y analizar mejor la estadística, pero esto poco se hace, y a qué costo/beneficio?

Vale la pena idear estrategias en donde la periodicidad de la encuesta se acorta conforme entendemos la medición y aprendemos a reaccionar ante ella. Esto permitirá un aprovechamiento más intensivo de cada encuesta, al poder dedicar los recursos especializados al análisis profundo, en lugar de continuamente estar preocupados sólo por la medición siguiente. Dar también tiempo a quienes tienen que tomar acción a tomar ésta con mayor comprensión del fenómeno, y dar tiempo a las acciones tomadas de poder incidir en la variable de interés antes de volver a medirla. Si a futuro se llega a un escenario donde la capacidad de acción es pareja al dinamismo de la variable, enhorabuena, midamos con más frecuencia, ahora sí se justifica.

Con la respuesta a estas y otras preguntas, el productor debe entonces traducir las necesidades/objetivos del usuario a un plan de análisis, usualmente los tabulados que se desea obtener de la encuesta, y acordar con el usuario la relevancia del plan. Después de concertado el plan de análisis, el productor debe desarrollar un cuestionario sencillo, ágil, de corta duración en tiempo de entrevista, y con preguntas que no hieran ni la sensibilidad del informante ni su inteligencia. Las preguntas a su vez deben concentrarse en el tema y ser suficientes para obtener las estimaciones de las variables de interés contenidas en el plan de análisis.

B. Cambios al cuestionario.

Encuestas continuas. Cuando una encuesta se repite en el tiempo para seguir la evolución de un tema sustantivo, surge la inquietud de mantener o no el mismo cuestionario, el mismo diseño de muestra, la misma estrategia del trabajo de campo, el mismo proceso de estimación. El alterar cualquiera de estos elementos puede inducir diferencias en los resultados que se confundirán con cambios naturales en el nivel de las variables medidas. Esto complica considerablemente el análisis y empaña las conclusiones.

Por esta razón, algunas Oficinas Nacionales de Estadística en distintos países, mantienen los mismos conceptos, cuestionario, diseño y logística de campo de una encuesta durante largos periodos, a pesar de conocer y tener manera de mejorar alguno o varios aspectos de la metodología. Cuando el cambio viene, es porque el beneficio de incorporar las mejoras excede el costo de 'romper' la serie. En estos casos, hay distintas estrategias para incorporar los cambios: se pueden introducir todos de golpe; se pueden programar para introducirlos en forma paulatina en levantamientos sucesivos (primero el cuestionario, luego el diseño de muestra, etc); cuando el interés y el presupuesto lo permiten, se puede tener un periodo de levantamientos paralelos con la metodología 'vieja' y la 'nueva', y computar los resultados como un promedio ponderado que va desfasando la primera y reforzando la segunda hasta que solo queda ésta; etc. En este campo no hay receta general, cada situación deber considerarse de manera particular, entre productores y usuarios, y actuar en consecuencia.

C. Documentación de la encuesta.

La documentación detallada de la metodología y procedimientos utilizados en los programas de encuesta en general es escasa, si no es que ausente. Esta situación se recrudece en particular en lo que se refiere a los sistemas y programas de procesamiento, donde en ocasiones son verdaderas "cajas negras", algunas de origen desconocido, sin nadie que se atreva a tocarlas.

Falta disciplina para registrar, archivar, ordenar y finalmente documentar: compromisos entre definiciones conceptuales de las variables prioritarias de interés y las definiciones operativas utilizadas en campo; la metodología del diseño y selección de muestra (cobertura, estratificación, etapas y unidades de muestreo de cada una, probabilidades de selección de cada etapa y final, variables prioritarias utilizadas para definir el tamaño de muestra); reseña de la capacitación, incluyendo manuales del cuestionario; procedimientos de supervisión y manuales respectivos; los resultados finales del trabajo de campo; los procedimientos y manuales de crítica-codificación; los criterios de validación y consistencia utilizados en la captura de datos; los paquetes o programas utilizados para el cálculo de errores de muestreo y parámetros asociados, y las estimaciones mismas que se lograron para las variables prioritarias.

D. Diseño muestral.

El diseño muestral de la encuesta, determinación del tamaño de muestra y esquema de muestreo, tiene un historial variado en la región. Retomando algunos de los puntos del documento elaborado para el taller del programa MECOVI realizado en agosto de 1995:

" La razón de ser del tamaño de muestra en la mayoría de las encuestas periódicas de la región es hoy en día un misterio. En algunos casos, el tamaño se remonta al origen de la encuesta, varios lustros atrás, asumiendo las razones y recomendación de algún consultor, que probablemente no contaba con el apoyo de parámetros relevantes locales, y usó los de otros países y otras encuestas. Situación justificable en el origen. ¿Pero y ahora?

En cuanto a los diseños de muestreo, falta formación en la mayoría de los cuadros técnicos para que tengan la capacidad de renovar los diseños en uso, mejorándolos desde una visión integral de mediano y largo plazo, con muestras maestras que atiendan las demandas de varias encuestas de distintos productores (ONE y otros), con actividades explícitas para el mantenimiento y actualización de los marcos de muestreo.

Por último, también sucede que al paso del tiempo, el marco de muestreo se proporciona a distintos usuarios, sin que el área de muestreo de la ONE lleve un control sistemático y exhaustivo de su uso. Cuando se tienen muestras grandes en áreas urbanas, se corren entonces riesgos innecesarios de que distintas encuestas "toquen a la misma puerta", con los consiguientes conflictos de entrevistas paralelas y abusos al informante."

En base a los puntos anteriores y otros, en el trabajo se hacían recomendaciones que continúan vigentes, y es pertinente el rescatarlas:

"a) Calculen en próximas rondas de sus programas de encuesta, de manera rutinaria, los errores de muestreo y par metros asociados, para las principales variables de la(s) encuesta(s).

b) Calculen los errores de muestreo para varias rondas recientes de las encuestas actuales, para tener la información necesaria para ratificar o rectificar los tamaños de muestra en uso.

c) Capaciten y formen a personal técnico en el diseño y selección de muestras, con un programa de actualización periódica.

d) Cuestionen la periodicidad y desagregación actual, en consulta con los usuarios prioritarios de cada encuesta (razones relevantes y fundamentadas).

e) Elaboren marcos de muestreo apegados a las necesidades del programa de encuestas y no al revés. Para ello, almacenar en bases de datos la información relevante del marco que se use para cada unidad de muestreo, en sus distintas etapas, para así facilitar el uso de diferentes diseños. Cuando se llegue al listado de viviendas de las "últimas unidades de área" utilizadas en el diseño, igualmente almacenar estos en archivos apropiados para su seguimiento y control.

f) Conserven en los archivos de datos de cada ronda de la encuesta: los ponderadores correspondientes a cada registro; y los identificadores necesarios de cada registro para poder reconstruir (si así se desea) el ponderador que le corresponde, y poder calcular errores de muestreo.

g) Reinstalen las actividades de mantenimiento y actualización de los marcos de muestreo, en particular el seguimiento de la "nueva construcción" y procedimientos para su incorporación a la muestra de rondas sucesivas, y la actualización cartográfica de reas en muestra.

h) Reinstalen programas permanentes de supervisión de las actividades de campo, incluyendo: re-entrevistas del supervisor con cuestionarios ad-hoc, entrevistas "observadas" por el supervisor, grabación (con permiso del informante) de entrevistas, verificación de la "no-respuesta" (se recomienda se haga al 100%), y la elaboración de formatos y gráficas que faciliten el seguimiento y control de encuestadores individuales, supervisores (y su grupo de encuestadores), estratos, regiones y dominios de estudio.

i) Elaboren ensayos y artículos sustantivos con la información ya disponible de rondas recientes de la encuesta, relacionada a temas de niveles de vida y afines a estos. Subcontratando para ello a profesionistas locales, con seminarios/talleres programados para difundir y discutir los resultados.

j) Facilitar a las instituciones/organismos interesados, los archivos básicos de datos de la encuesta (salvaguardando el secreto estadístico, con medidas apropiadas), para que promuevan la elaboración de otros estudios y artículos que se discutan en los seminarios/talleres programados en el punto anterior.

k) Elaboren explícitamente un programa integrado de encuestas, que se difunda con los usuarios prioritarios para que sepan que temas y cuando se investigarán en éste."

E. Acceso, difusión y análisis.

Por último, y pidiendo disculpas por volver a citar el trabajo anterior, en el tema del aprovechamiento cabal de los datos ratificamos que:

"... si en algo se va a la zaga en el tema de encuestas de hogares en la región, es en el análisis de los datos ya disponibles. Levantada una encuesta y publicados sus resultados primarios, es raro encontrar un aprovechamiento ulterior de estos. Ante "nuevas" demandas de información, la respuesta fácil es "realicemos una nueva encuesta". El reto de intentar una primera respuesta aprovechando con otra óptica datos de encuestas recientes o no, es por lo general evitado, concentrando el esfuerzo en explicar por qué, los datos disponibles "no sirven", y a partir de ahí justificar la nueva encuesta.

La crónica falta de recursos dentro de las ONE, aunado al recelo de dar acceso a los usuarios a los archivos primarios de datos en aras de proteger el "secreto estadístico", han propiciado en el pasado ésta situación.

Cuando sí se desarrollan trabajos ulteriores, la mayoría son realizados no sólo fuera de las ONE y organismos nacionales afines al tema, si no incluso fuera del país, con poca o nula participación sustantiva de profesionistas locales. En no pocas ocasiones el producto final es del hábito de un muy reducido círculo de expertos locales, si acaso."

ANEXO 1

INFORMACION RECOLECTADA EN 1991					
TAMAÑOS DE MUESTRA (1991)					
PAIS	ENCUESTA	UPM EN MUESTRA	TAMAÑO DE UPM (VIV)	VIV EN MUESTRA POR UPM	TOTAL DE VIV EN DISEÑO
ARGENTINA	PERMANENTE DE HOGARES	2575	300	10-15	30000
BOLIVIA	INTEGRADA DE HOGARES	1764	20	5	8820
BRASIL	MENSUAL DE EMPLEO	1542	250	25	35550
CHILE	NACIONAL DE EMPLEO	2394	300	15	35910
COLOMBIA	NACIONAL DE HOGARES	1989	400	10	19890
COSTA RICA	DE HOGARES DE PROPOSITOS MULTIPLES	719	60	13	9347
CUBA	NACIONAL DE OCUPACION	1620	150	8	12960
GUATEMALA	NACIONAL SOCIO-DEMOGRAFICA	611	166	15-20	10934
HONDURAS	PERMANENTE DE HOGARES DE PROPOSITOS MULTIPLES	875	340	10U-14R	10250
MEXICO	NACIONAL DE EMPLEO URBANO	1370	>480	30	41100
PANAMA	DE HOGARES	1600	50	6	9600
URUGUAY	CONTINUA DE HOGARES	7925	20	3	23775
VENEZUELA	DE HOGARES POR MUESTREO	3960	200	10	39600

TRABAJO DE CAMPO (1991)					
PAIS	DURACION ENTREVISTA (MINUTOS)	ENCUESTADORES POR SUPERVISOR		TASA DE RESPUESTA (%)	
ARGENTINA	10-15	10		80	
BOLIVIA	8	4		90	
BRASIL	40	6		82	
CHILE	15	5		90	
COLOMBIA	25-30	3		90	
COSTA RICA	20-45	3		89	
CUBA	15	2-3		95	
GUATEMALA	20	4-5		85	
HONDURAS	30-40	5		91	
MEXICO	20-40	4-5		90	
PANAMA	25	4-5		96	
URUGUAY	30	3-5		90	
VENEZUELA	15	3		95	
PAISES SIN PROGRAMAS CONTINUOS DE ENCUESTAS (1991)					
PAIS	ENCUESTA	AÑO	COBERTURA	TAMAÑO DE MUESTRA	DOMINIOS DE PUBLICACION
ECUADOR	PERIODICA SOBRE EMPLEO Y DESEMPLEO	1989	URBANA	11800	NACIONAL, COSTA, SIERRA Y ORIENTE
EL SALVADOR	DE HOGARES DE PROPOSITOS MULTIPLES	1990	ND	6400	ND
HAITI	DE PRESUPUESTO FAMILIAR Y CONSUMO	1986-1987	NACIONAL	3100	NACIONAL Y REGIONAL
NICARAGUA	PARA ESTABLECIDOS Y REINSERTADOS	1991	TRES MUNICIPIOS	2400	MUNICIPAL
PARAGUAY	DE HOGARES (MANO DE OBRA)	1990	AREA METROPOLITANA DE ASUNCION	1000	ASUNCION Y RESTO
PERU	NACIONAL SOBRE MEDICIONES DEL NIVEL DE VIDA	1986	NACIONAL	5000	COSTA, SIERRA, SELVA Y LIMA
REPUBLICA DOMINICANA	DE FUERZA DE TRABAJO	1991	ND	1700	ND

PAISES DEL CARIBE DE HABLA INGLESA (1991)						
PAIS	POBLACION EN MILES (1990)	ULTIMO CENSO	ENCUESTA MAS RECIENTE	AÑO	COBERTUR A	TAMAÑO
ARUBA	60	7/10/91	ND	ND	ND	ND
BAHAMAS	253	1/5/90	MANO DE OBRA	1989	NACIONAL	2440
BARBADOS	255	1/5/90	NACIONAL SOBRE SIDA	1990	NACIONAL	3600
BELICE	187	12/5/91	GASTOS DE LOS HOGARES	1989	NACIONAL Y URBANO-RURAL 6 DISTRITOS	2100
ISLAS CAYMAN	26	15/10/91	INGRESOS Y GASTOS DE LOS HOGARES	1990	NACIONAL	750
DOMINICA	82	12/5/91	MANO DE OBRA	1989	NACIONAL	500
GRANADA	98	12/5/91	MANO DE OBRA	1988	NACIONAL	1000
JAMAICA	2456	7/4/91	MANO DE OBRA	1990	NACIONAL	7800
ANTILLAS HOLANDESES	188	27/1/92	MANO DE OBRA	1989	CURAZAO	2250
SAINT KITTS	44	12/5/91	FUERZA DE TRABAJO	1987	NACIONAL	ND
SANTA LUCIA	150	12/5/91	INGRESOS Y CUIDADOS DE SALUD	1989	NACIONAL	600
SURINAME	422	1992	FUERZA DE TRABAJO Y DEMOG.	1989	PARAMARI BO Y WANICA	600
TRINIDAD Y TOBAGO	1234	15/5/90	CONTINUA DE POB.	1989	NACIONAL	9700

ANEXO 2

DOCUMENTO ELABORADO PARA EL TALLER DEL PROGRAMA MECOVI DE 1995*

I. INTRODUCCIÓN

1. El papel fundamental que juega el sector de los hogares en el desarrollo socio-económico de los países en desarrollo, es un hecho ampliamente reconocido. Los hogares dan cuenta de una buena parte de la actividad productiva y son a su vez afectados por los cambios económicos y sociales. El desarrollo humano, nivel de vida y bienestar general de la población de un país, se refleja finalmente en los hogares, y los grupos familiares y miembros individuales que los integran. Conocer las características de los hogares para distintos estratos socio-económicos y áreas geográficas de interés, así como el perfil de sus integrantes agrupados en distintas clases poblacionales, es ya una demanda cotidiana en distintas esferas de la gestión pública, organismos no de gobierno, agencias internacionales y aún el sector privado.

2. Por esta razón y como complemento a (o ante la ausencia de) un buen sistema de estadísticas continuas o periódicas basadas en registros administrativos, una de las herramientas estadísticas apropiadas para medir y dar seguimiento oportuno, a costo razonable, de indicadores demográficos, sociales y económicos, es la encuesta de hogares en base a muestras probabilísticas.

3. En América Latina y el Caribe, varios de los países de la región tienen desde hace tiempo programas periódicos de encuestas de hogares, o han llevado a cabo varias encuestas en forma ad-hoc de acuerdo a las demandas de usuarios. Esto ha conformado una infraestructura a distintas escalas según la experiencia y necesidad de cada país (nacional, regional, áreas metropolitanas, capital del país, etc.) que se puede aprovechar a efectos de dar apoyo al seguimiento de distintos indicadores de interés en el campo socio-económico, demográfico y de desarrollo.

4. Este trabajo se ha realizado en la región tanto dentro de las Oficinas Nacionales de Estadística (ONE) como fuera de ellas, con o sin el apoyo de éstas (por ejemplo para el diseño y selección de muestras), y cubren un espectro bastante extenso y rico en datos en la región. Si bien no se reconoce de manera explícita, aún por los productores, todas las encuestas levantadas han sido de propósitos múltiples, al menos al considerar los datos levantados por sus operativos de campo. ^ De ahí que en algunos casos varios de los indicadores que interesan en un país particular, son ya parte integrante de los resultados

* Realizado en la sede del BID en Washington el 4 de agosto de 1995.

rutinarios de su programa de encuestas, en otros será posible construirlos a partir de la información recolectada, y en otros sólo será necesario el agregar un módulo especial a la encuesta para captar la información requerida. En pocos países se tendría que partir de cero y diseñar un programa exclusivo para obtener varios de los indicadores necesarios a través de encuestas de hogares para el seguimiento de las metas de interés.

5. De hecho, aunque tampoco se reconoce, varios de los programas de encuestas en la región cuentan con información longitudinal, pues están basados en muestras con algún esquema de rotación, donde parte de los hogares son revisitados dos o más veces, captando además de las variables sustantivas de la encuesta cambios en la estructura del hogar (entradas y salidas de miembros).

6. ¿Quiere decir esto que todo está bien en la región en el campo de encuestas de hogares? No. En las siguientes secciones se tratan áreas problemáticas que requieren atención.

II. TAMAÑOS DE MUESTRA Y DISEÑOS DE MUESTREO

7. La razón de ser del tamaño de muestra en la mayoría de las encuestas periódicas de la región es hoy en día un misterio. En algunos casos, el tamaño se remonta al origen de la encuesta, varios lustros atrás, asumiendo las razones y recomendación de algún consultor, que probablemente no contaba con el apoyo de parámetros relevantes locales, y usó los de otros países y otras encuestas. Situación justificable en el origen. ¿Pero y ahora?

8. Aún en las ONE más avanzadas de la región lo común es que no se calculen $\hat{\sigma}$ para las principales variables de la encuesta sus errores de muestreo (errores estándar), intervalos de confianza; coeficientes de variación, correlación intraclase; efectos de diseño; etcétera. Situación no justificable con la disponibilidad desde hace más de una década de programas de distribución gratuita para el cálculo de estos parámetros, por complejo que sea el diseño de muestra utilizado. Su ausencia hace que se perpetúen los tamaños iniciales, sin mayor justificación.

9. La promoción del sano hábito de su cómputo rutinario para las principales variables de la encuesta permitiría ratificar o rectificar (lo más seguro) los tamaños en uso. No sería sorprendente que un simple análisis de los errores de muestreo evidencie que algunas encuestas están sobredimensionadas, y que con una reducción sustancial de muestra no hay pérdida significativa en la confianza y precisión de las estimaciones de los parámetros prioritarios, con el consecuente ahorro, pues el principal componente de las encuestas son los costos de campo.

10. Para los tabulados complejos que requieren más casos, simplemente se "acumula" muestra a lo largo de varias rondas de la encuesta para producirlo.

11. En cuanto a los diseños de muestreo, falta formación en la mayoría de los cuadros técnicos para que tengan la capacidad de renovar los diseños en uso, mejorándolos desde una visión integral de mediano y largo plazo, con muestras maestras que atiendan las demandas de varias encuestas de distintos productores (ONE y otros), con actividades explícitas para el mantenimiento y actualización de los marcos de muestreo.

12. En ocasiones, la manipulación del marco y muestra se complica innecesariamente por restricciones del área de cómputo. Conviene revisar y adecuar procedimientos para que estos sean de acuerdo a la conveniencia metodológica del programa de encuesta, y no a las limitaciones de cómputo, hoy en día no justificables.

13. Por otro lado, se dan varios casos donde los marcos de muestreo incluyen "muchas" variables para usar en diseños estratificados, y luego sean poco usadas. Conviene que las ONE las revisen y sólo se queden con pocas, generales, de apoyo a sus programas periódicos de encuesta fundamentalmente. Las principales serían: medida de tamaño de los segmentos censales, su clasificación urbana-rural, su identificación político administrativa (geográfica), y su número de identificación final que contiene una estratificación implícita al haber sido éste adjudicado de manera sistemática en forma de "serpentina" o "serpiente".

14. Por último, también sucede que al paso del tiempo, el marco de muestreo se proporciona a distintos usuarios, sin que el área de muestreo de la ONE lleve un control sistemático y exhaustivo de su uso. Cuando se tienen muestras grandes en áreas urbanas, se corren entonces riesgos innecesarios de que distintas encuestas "toquen a la misma puerta", con los consiguientes conflictos de entrevistas paralelas y abusos al informante.

III. DESAGREGACIÓN Y PERIODICIDAD DE LAS ENCUESTAS

15. Si bien las encuestas son un medio para solventar la carencia de información de otras fuentes regulares (registros administrativos y censos), no se les debe confundir con sustitutos de éstas. En particular, no se les debe exigir la misma capacidad en cuanto a la desagregación de sus estimaciones; no son el instrumento para tener resultados locales, para un gran número de localidades.

16. No obstante lo anterior, la creciente demanda y persistente ausencia de datos locales presiona cada vez más a los programas de encuesta para su generación. Así por ejemplo, algunos programas de encuestas de empleo continúan creciendo en número de ciudades cubiertas, para producir estimaciones separadas de cada ciudad, aún cuando la instrumentación de políticas de empleo no se haga de manera diferencial para cada ciudad. Lo paradójico es que se crece a una encuesta con un tamaño global varias veces superior al necesario para tener una estimación nacional o nacional urbana de los parámetros de empleo, sin poder hacerlo.

17. Continuando con el ejemplo de las encuestas de empleo, tomemos el tema de periodicidad. ¿Es necesario, como lo tienen algunos programas de la región, conocer mensualmente los indicadores y tasas de empleo-desempleo, o sería suficiente el estimarlo de manera trimestral? Pregunta más relevante aún ante la actual crisis regional. Más allá de los análisis que los expertos pueden realizar con los datos mensuales, ¿qué con los que "instrumentan las políticas de empleo"? Después de todo, "ellos" son los usuarios prioritarios de la encuesta. ¿Reaccionan "ellos" ante la evidencia mensual del cúmulo de datos que ahora reciben, "instrumentando" ajustes o nuevas políticas de empleo cada mes? Si no hay esa capacidad administrativa de reacción, ¿en qué les afecta el que el flujo fuera cada trimestre o semestre? La muestra anual de estos programas de encuesta se podría reducir aún más con el consecuente ahorro adicional.

18. En tanto, las investigaciones a profundidad se pueden realizar con la serie ya disponible de rondas pasadas. Si en algo se va a la zaga, no es en la producción de información, es en el aprovechamiento cabal de las estadísticas ya disponibles. Programas de encuestas en otras áreas requieren un cuestionamiento similar.

IV. TRABAJO DE CAMPO

19. La supervisión de campo tanto para la infraestructura de soporte (listados de viviendas, nueva construcción, viviendas omitidas, y cartografía asociada), como en el levantamiento mismo de las encuestas, ya no se realiza en la mayoría de los programas de encuesta. Urge restablecerla. Para ello es necesario que las ONE preparen los formatos y procedimientos a usar, pero sobre todo tener ya probada una mecánica ágil de qué hacer con la información que se recoja, en particular en lo que toca a la toma de decisiones y acciones correctivas (incluyendo parar momentáneamente la encuesta para recapacitar, o incluso cesar y cambiar personal con graves fallas). Es imperativo que el personal de campo se sepa supervisado y sujeto a cuentas, en particular obligado a repetir trabajo deficiente, sin pago adicional por ello.

20. El uso de personal de planta, fijo, como encuestadores y supervisores, tiene ventajas indudables, pero precisamente ante la ausencia de un programa de supervisión permanente, conlleva a vicios y abusos. Un abuso común, es la noción de que el personal fijo fue contratado para levantar "n" entrevistas por semana/mes, y que cualquier tarea distinta que se le solicite es una carga adicional, que le debe ser compensada. Se dan entonces situaciones donde la carga usual de levantamiento es ridículamente baja, y sin embargo no se pueda usar el resto del tiempo del personal para actividades de actualización y mejora metodológica del programa de encuesta. Este personal usa el tiempo libre para "contratarse" con otras encuestas ad-hoc incluso internas a la misma ONE.

V. APROVECHAMIENTO DE DATOS DE PROGRAMAS EXISTENTES

21. Como ya se señaló, si en algo se va a la zaga en el tema de encuestas de hogares en la región, es en el análisis de los datos ya disponibles. Levantada una encuesta y publicados sus resultados primarios, es raro encontrar un aprovechamiento ulterior de estos. Ante "nuevas" demandas de información, la respuesta fácil es "realicemos una nueva encuesta". El reto de intentar una primera respuesta aprovechando con otra óptica datos de encuestas recientes o no, es por lo general evitado, concentrando el esfuerzo en explicar por qué los datos disponibles "no sirven", y a partir de ahí justificar la nueva encuesta.

22. La crónica falta de recursos dentro de las ONE, aunado al recelo de dar acceso a los usuarios a los archivos primarios de datos en aras de proteger el "secreto estadístico", han propiciado en el pasado ésta situación.

23. Cuando sí se desarrollan trabajos ulteriores, la mayoría son realizados no sólo fuera de las ONE y organismos nacionales afines al tema, si no incluso fuera del país, con poca o nula participación sustantiva de profesionistas locales. En no pocas ocasiones el producto final es del ámbito de un muy reducido círculo de expertos locales, si acaso.

VI. PROPÓSITOS MÚLTIPLES VS "POLITEMÁTICA"

24. Como ya se mencionó, al menos desde el punto de vista de información recogida, todo programa de encuesta de la región ha sido de "propósitos múltiples". Los temas múltiples pueden crecer, han crecido, y la encuesta puede recoger respuestas a todos ellos, la pregunta es, ¿dónde está el balance entre agregar temas y todavía obtener respuestas "fiables" del informante?

25. La recomendación casi universal de conservar el tiempo de entrevista con cualquier unidad informante a lo más entre 20-40 minutos, se basa en que la fatiga que provoca la entrevista motiva al informante que hacia el final conteste más con el objetivo de terminar ésta que de pensar su respuesta. No obstante, 20-40 minutos son suficientes para cubrir una gran cantidad de temas cuando el cuestionario es ágil y bien estructurado.

26. Visitas múltiples a la unidad informante, incluso el cambio de informante, ayudan a aliviar la fatiga, pero aún ésta estrategia tiene un límite. En todo caso hace falta investigación seria para contestar las dudas sobre la calidad de información que se obtiene cuando la encuesta es politemática, con varios informantes, y tiempos de entrevista superiores a los 40 minutos.

27. Si bien para efectos de ajustar modelos de análisis lo ideal es tener medidas las distintas variables de interés en el mismo hogar, hace falta investigación que explore la posibilidad de elevar el análisis a nivel "unidad primaria de muestreo", donde

sería posible sumar información de distintas rondas (incluso distintas encuestas con el mismo diseño) con distintos módulos temáticos, evitando así los problemas de fatiga de informantes individuales. De ser factible, los temas de interés se pueden distribuir en rondas sucesivas del programa de encuestas.

VII. SUGERENCIAS GENERALES

28. A manera de resumen, se recomienda que el "Programa para el mejoramiento de las encuestas para medir las condiciones de vida en América Latina y el Caribe" tome en cuenta el asignar recursos para promover que las ONE:

a. Calculen en próximas rondas de sus programas de encuesta, de manera rutinaria, los errores de muestreo y parámetros asociados, para las principales variables de la(s) encuesta(s).

b. Calculen los errores de muestreo para varias rondas recientes de las encuestas actuales, para tener la información necesaria para ratificar o rectificar los tamaños de muestra en uso.

c. Capaciten y formen a personal técnico en el diseño y selección de muestras, con un programa de actualización periódica.

d. Cuestionen la periodicidad y desagregación actual, en consulta con los usuarios prioritarios de cada encuesta (razones relevantes y fundamentadas).

e. Elaboren marcos de muestreo apegados a las necesidades del programa de encuestas y no al revés. Para ello, almacenar en bases de datos la información relevante del marco que se use para cada unidad de muestreo, en sus distintas etapas, para así facilitar el uso de diferentes diseños. Cuando se llegue al listado de viviendas de las "últimas unidades de área" utilizadas en el diseño, igualmente almacenar estos en archivos apropiados para su seguimiento y control.

f. Conserven en los archivos de datos de cada ronda de la encuesta: los ponderadores correspondientes a cada registro; y los identificadores necesarios de cada registro para poder reconstruir (si así se desea) el ponderador que le corresponde, y poder calcular errores de muestreo.

g. Reinstalen las actividades de mantenimiento y actualización de los marcos de muestreo, en particular el seguimiento de la "nueva construcción" y procedimientos para su incorporación a la muestra de rondas sucesivas, y la actualización cartográfica de áreas en muestra.

h. Reinstalen programas permanentes de supervisión de las actividades de campo, incluyendo: re-entrevistas del supervisor con cuestionarios ad-hoc, entrevistas "observadas" por el supervisor, grabación (con permiso del informante) de entrevistas, verificación de la "no-respuesta" (se recomienda se haga al 100%), y la elaboración de formatos y gráficas que faciliten el seguimiento y control de encuestadores individuales, supervisores (y su grupo de encuestadores), estratos, regiones y dominios de estudio.

i. Elaboren ensayos y artículos sustantivos con la información ya disponible de rondas recientes de la encuesta, relacionada a temas de niveles de vida y afines a estos. Subcontratando para ello a profesionistas locales, con seminarios/talleres programados para difundir y discutir los resultados.

j. Facilitar a las instituciones/organismos interesados, los archivos básicos de datos de la encuesta (salvaguardando el secreto estadístico, con medidas apropiadas), para que promuevan la elaboración de otros estudios y artículos que se discutan en los seminarios/talleres programados en el punto anterior.

k. Elaboren explícitamente un programa integrado de encuestas, que se difunda con los usuarios prioritarios para que sepan que temas y cuando se investigarán en éste.

^ Por ejemplo, además de algunas características de la vivienda y su equipamiento, es común encontrar que recogen los datos básicos de composición por sexo y edad de los miembros del hogar, parentesco con el jefe de hogar, estado civil, nivel educativo, y algunos indicadores sobre ocupación y/o nivel de ingresos.

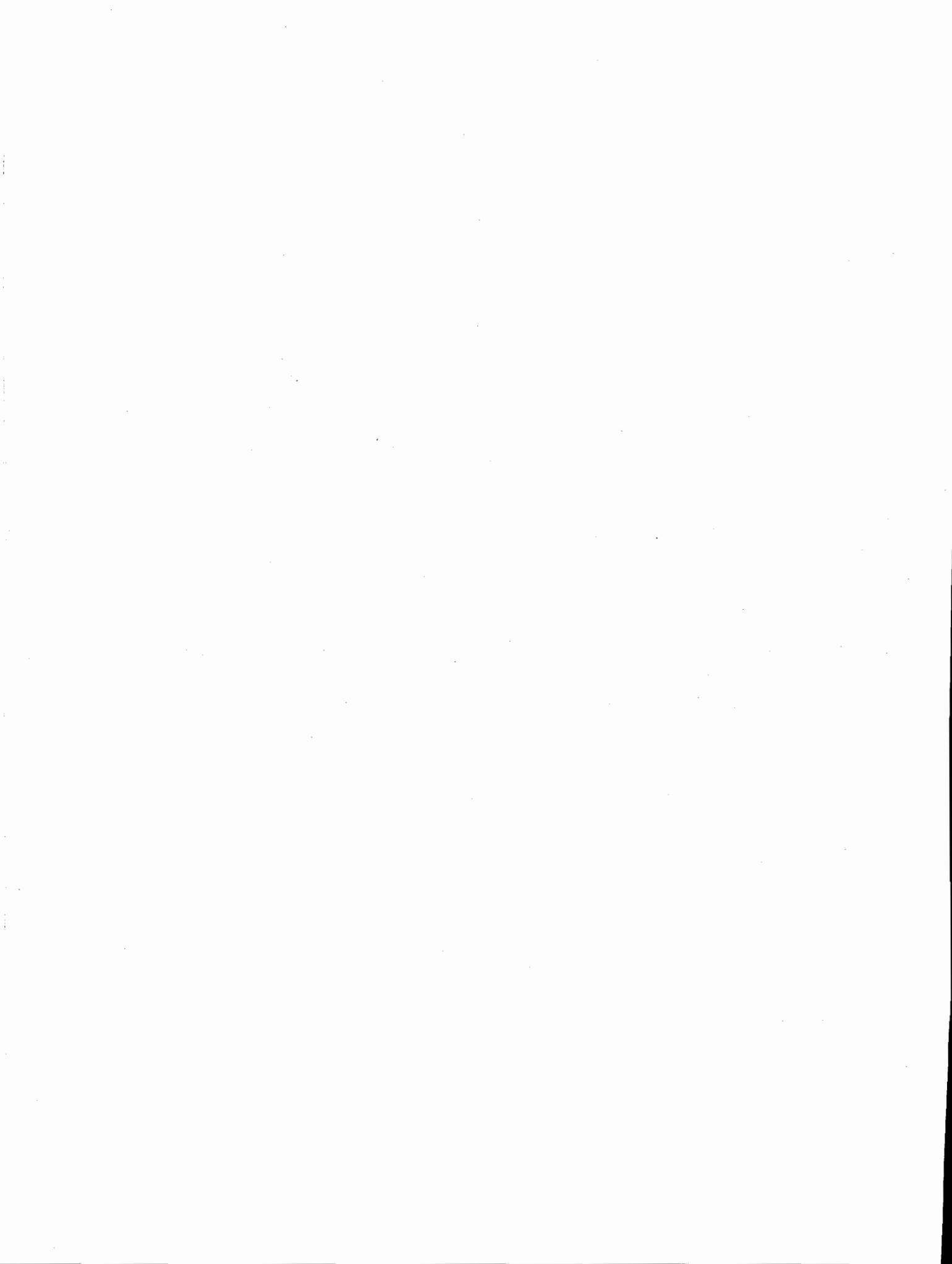
^ O se calculen de manera errónea, utilizando las fórmulas de un muestreo aleatorio simple, cuando el diseño de muestra utilizado está bastante alejado del modelo básico de texto.

^ Quizá por recomendación de algún consultor, y donde posteriormente la ONE no tuvo el personal adecuado para aprovecharlas, pero las mantienen en el marco.

^ Al inicio de los programas sí se hacían, pero muy pronto se abandonaron. Aún cuando la figura de "supervisor" existe, en general se constriñe a un "acompañante" del grupo de entrevistadores, que rescata algunas entrevistas, sirve de "chofer", y lleva algún control de recursos y materiales.



TRABAJO DE CAMPO Y CONTROL DE CALIDAD



**CONTROL DE CALIDAD EN EL DISEÑO Y EJECUCIÓN DE
ENCUESTAS DE HOGARES**

JUAN CARLOS FERES

FERNANDO MEDINA

(CEPAL)

ÍNDICE

	<u>Página</u>
I. INTRODUCCIÓN	235
II. ÁMBITO METODOLÓGICO	235
1. Diseño conceptual	235
2. Diseño estadístico.....	240
III. LEVANTAMIENTO DE LA INFORMACIÓN	243
1. Organización del trabajo de campo	243
2. Capacidad del personal de campo de oficina	244
3. Distribución de la muestra y de las cargas de trabajo.....	244
4. Supervisión del trabajo de campo	245
IV. TRATAMIENTO DE LA INFORMACIÓN	247
1. Crítica y codificación	247
2. Falta de respuesta	248
3. Captura y congruencia de los datos	249
4. Consistencia de la información	250
5. Precisión estadística de los estimadores.....	250
6. Documentación de los procesos de la encuesta.....	251
7. Difusión de resultados y relación con los usuarios	251

CONTROL DE CALIDAD EN EL DISEÑO Y EJECUCIÓN DE ENCUESTAS DE HOGARES

I. Introducción

El diseño y ejecución de encuestas de hogares por muestreo exige un sólido sustento técnico, así como la aplicación de sistemas modernos de gestión y administración que aseguren la producción de información estadística de calidad. El cumplimiento de los objetivos de la investigación se traduce en el logro de una plena satisfacción de los requerimientos de los usuarios, donde se minimicen los errores y sesgos de la información y se alcance eficiencia en el uso de los recursos disponibles mediante la generación de resultados confiables, oportunos y a bajo costo.

En ese sentido, todas las etapas involucradas en el diseño y ejecución de una encuesta son esenciales para garantizar la calidad de la información. De modo que en cada una de ellas se debieran contemplar y poner en práctica mecanismos de control que permitan identificar y corregir las distorsiones y posibles errores en los datos recabados.

Bajo esta perspectiva, de orden más bien sistémico, a continuación se analizan sucintamente las diversas etapas que caracterizan la ejecución de una encuesta de hogares, a la vez que se sugieren algunos criterios, procedimientos y medidas de control que se consideran de utilidad para reducir los errores y sesgos de los datos y, en consecuencia, elevar la calidad de la información.

II. Ambito metodológico

1. Diseño conceptual

Bajo la óptica de la calidad total, los objetivos de las encuestas se deben orientar preferentemente a satisfacer, en tiempo y forma, las necesidades de los usuarios de la información: analistas, diseñadores de políticas y tomadores de decisiones. De modo que en la etapa de definición de los métodos y procedimientos para llevar a cabo la investigación, se debe dar respuesta a preguntas tales como qué tipo de información se desea obtener; de qué manera se alcanzarán los objetivos de la encuesta y cumplirán los plazos establecidos para la entrega de resultados; cómo se relacionarán los datos generados con aquellos disponibles de otras fuentes y cuál es la precisión deseada para las estimaciones de interés.

a) Requerimientos de información

Un primer requisito indispensable es definir con claridad el fenómeno a estudiar, con el fin de identificar los objetivos generales y específicos de la encuesta. Son pocas las ocasiones en que no existe ninguna información acerca del tema que se desea analizar, por lo que es aconsejable hacer primeramente una revisión bibliográfica y de los antecedentes

disponibles de estudios similares realizados en el pasado, en el país y en el extranjero, así como efectuar consultas con especialistas y usuarios de la información a efectos de precisar los alcances de la investigación y los indicadores que se desean obtener.

Asimismo, es necesario tener claro los usos que se harán de la información recopilada, lo que permitirá identificar los dominios de estudio, establecer las hipótesis de comportamiento, formular posibles relaciones de causalidad entre las variables estudiadas y definir los niveles de desagregación y precisión requeridos para los datos. Esta última información constituye, a su vez, un insumo fundamental para determinar el tamaño de la muestra y definir los procedimientos para la selección de las observaciones que formarán parte de la investigación.

b) Marco conceptual

Una vez identificados los requerimientos de información, la elaboración del marco conceptual y metodológico debiera orientarse a cumplir con los objetivos de la encuesta, considerando las recomendaciones internacionales vigentes. A su vez, si entre las distintas encuestas que forman parte del sistema integrado existiese traslape de algunos temas, las definiciones y conceptos debieran homologarse a fin de evitar inconsistencias en los resultados proporcionados por las diferentes fuentes de datos.

Es habitual que algunas encuestas investiguen determinados temas con mayor o menor profundidad que otras, situación que también puede redundar en que los resultados obtenidos no sean estrictamente comparables. Las encuestas de empleo, por ejemplo, normalmente profundizan en la identificación de las personas que manifiestan estar desocupadas (cesantes o buscando empleo por primera vez), o bien que están afectadas por distintas formas de subempleo. Para ello, y en atención a las recomendaciones internacionales sobre la materia, se formulan varias preguntas, algunas de las cuales actúan como filtro, con el objeto de distinguir a estas subpoblaciones de interés, pero que a la vez determinan la medición de la fuerza de trabajo. Asimismo, existen también algunas encuestas orientadas a evaluar los niveles de vida de las familias que estudian el empleo como una variable importante para explicar la formación del ingreso de los hogares; sin embargo, es frecuente que las preguntas que se incluyen para captar la condición de ocupación de la población en edad de trabajar adolezcan de cierta superficialidad, lo que genera discrepancias en los resultados respecto a otras estadísticas oficiales sobre empleo y desempleo.

Por su parte, la sistematización del marco conceptual y metodológico de la encuesta debe traducirse en la elaboración de un conjunto de manuales que sirvan para apoyar las actividades de capacitación de los supervisores, entrevistadores, críticos y codificadores, y del personal de todo nivel involucrado en la ejecución del proyecto. En estos manuales se debieran incorporar los conceptos y definiciones de las preguntas contenidas en el cuestionario, ejemplificando con gran detalle las variantes que pueden enfrentarse durante la realización del trabajo de campo, especificando lo que se considera y lo que se excluye en cada caso y la manera de decidir ante distintas situaciones prácticas.

Asimismo, dichos manuales debieran adoptarse como documentos de consulta que ayuden a aclarar dudas y resolver situaciones complejas que puedan surgir durante las visitas a los hogares.

c) Cuestionario y plan de tabulados

El logro de los objetivos de una encuesta de propósitos múltiples se basa en gran medida en la disposición de los hogares a colaborar, así como en la calidad de los datos que proporcionan los informantes. De manera que la etapa del diseño del cuestionario se puede considerar como una de las más críticas en la planeación de una encuesta por muestreo, ya que su aceptación por parte de los informantes está condicionada a disponer de un instrumento de fácil entendimiento y estructurado en términos lógicos. Un cuestionario bien redactado y ordenado atrae la atención del informante y facilita su relación con el entrevistador.

El cuestionario es el instrumento que permite conducir la entrevista y evita que el empadronador elija el orden y redacción de las preguntas, o las categorías de respuesta. A su vez, establece el flujo adecuado de las preguntas e identifica a las subpoblaciones de interés en cada tema. Por otra parte, para facilitar su manejo en campo el cuestionario debe tener un formato adecuado y ser resistente a las inclemencias climáticas y al constante manoseo de que son objeto.

Dado que el cuestionario representa la expresión operativa de los objetivos de la investigación, su diseño debe guardar estrecha relación con los demás componentes de la encuesta. Por ende, es muy recomendable que su confección esté a cargo de un equipo interdisciplinario integrado por especialistas en los temas de la encuesta, diseño estadístico, enfoque conceptual, personas con experiencia en el trabajo de campo y los encargados del procesamiento de datos, entre otros. En la actualidad es común encontrar que tiendan a repetirse los estudios sobre diversos temas en un mismo país e incluso entre países, por lo que la experiencia en el diseño de cuestionarios debe entenderse como un proceso acumulativo. Sin embargo, esto no significa que un formato elaborado con objetivos específicos pueda aplicarse en cualquier lugar o circunstancia sin que se efectúen las modificaciones pertinentes en atención a las características del entorno cultural, el nivel de instrucción de la población objetivo, la calificación del personal de campo o el grado de desarrollo del sistema estadístico nacional.

Como se ha señalado en diversos documentos técnicos, es importante tener consciencia que las personas que redactan el cuestionario no son las mismas que van a responder la entrevista, por lo que se debe pensar siempre en los informantes al momento de formular las preguntas. En este sentido, deben omitirse al máximo los tecnicismos y las palabras de uso poco frecuente ya que su inclusión seguramente contribuirá a generar confusión y a desestimular la participación de los entrevistados.

El buen conocimiento del tema a investigar, así como las consultas a los usuarios, facilitan una apropiada redacción de las preguntas. Asimismo, las pruebas de campo en

zonas geográficas diversas que presenten situaciones poco comunes y para diferentes estratos socioeconómicos, permiten afinar la redacción y ayudan a percibir la aceptación de los informantes y su disposición a colaborar. No siempre una redacción detallada genera mejor información que las preguntas breves y directas. Sin embargo, por lo general un desglose amplio de las opciones de respuesta constituye un mecanismo adecuado para ganar en precisión y reducir las omisiones, al tiempo que en muchas variables económicas permite la comparabilidad con datos a nivel macro y con registros administrativos.

El diseño físico del formulario puede basarse en un formato que registre en una primera parte información de carácter general como el nombre de la encuesta, el organismo que la ejecuta, las disposiciones legales que amparan su aplicación, el número de folio e información de control que permita determinar con facilidad la ubicación geográfica de la unidad de observación, el tiempo de duración de la visita al hogar y el resultado final de la entrevista. A continuación, es aconsejable incorporar preguntas que informen sobre las características básicas de las unidades de observación. Así, en el caso de las encuestas de hogares se puede registrar, entre otras, las características de la vivienda y el número de hogares que la componen, sus materiales de construcción, la disponibilidad de servicios básicos y el equipamiento doméstico.

La segunda parte del formato se puede destinar a consignar información general sobre los miembros del hogar que permita una primera clasificación de la población objeto de estudio para iniciar el registro de datos sobre los temas centrales de la investigación.

Finalmente, las preguntas relacionadas con el o los temas específicos de la encuesta deberán estructurarse conforme a una lógica simple, donde los aspectos de carácter general antecedan a las preguntas más puntuales sobre los tópicos de interés. Los primeros cuestionamientos deben ser breves y fáciles de responder, a fin de que el entrevistado adquiera confianza e incremente su interés por participar. Asimismo, las preguntas comprometedoras o íntimas se debieran ubicar de la mitad del cuestionario hacia adelante, para aplicarlas una vez que el entrevistador haya establecido una relación de confianza con el informante. En todos los casos el cuestionario debiera disponer de suficiente espacio en blanco para que el entrevistador pueda hacer las aclaraciones pertinentes que ayuden a los responsables de la crítica y codificación a entender el porqué se registró esa información.

En relación al plan de tabulados, éste suele verse simplemente como la acción que permite decidir a los responsables de la difusión sobre el esquema básico de publicación de resultados. Sin embargo, hay que enfatizar que esta fase constituye una oportunidad para identificar, junto con los usuarios, los dominios analíticos en los que se desea desagregar la información, así como para conocer cuáles serán los indicadores que se desean calcular y las respectivas subpoblaciones de interés. Asimismo, el análisis preliminar de los cruces que se quieren realizar contribuye a estimar el número mínimo de observaciones necesarias para garantizar la confiabilidad estadística de las relaciones que se establezcan entre categorías de las variables estudiadas, lo cual ayudará a la determinación del tamaño de muestra necesario.

Por otra parte, el plan básico de tabulados debe ser interpretado como un primer intento por sintetizar los resultados que se generarán, y en ningún caso debiera considerarse como un procedimiento orientado a la explotación total de la información. Por mas tabulados y gráficas que se incluyan en una publicación, nunca se podrá llegar a satisfacer todas las demandas de los usuarios o lograr una completa y exhaustiva explotación de la base de datos.

El verdadero valor de los resultados generados por medio de una encuesta por muestreo está determinado, en última instancia, por el uso analítico que se haga de la información, con el fin de establecer un sistema de datos que describa y/o contribuya a explicar la ocurrencia de un fenómeno de interés.

Una práctica que se viene desarrollando frecuente es poner a disposición de los usuarios discos magnéticos o compactos con los tabulados de la encuesta. Esto constituye una interesante y eficaz forma de incorporar el desarrollo tecnológico a los procedimientos de difusión de información estadística que facilita al usuario la consulta y almacenamiento de los datos. No obstante, en ningún caso releva de la necesidad de agregar mayor valor a la información de la encuesta mediante los estudios analíticos que normalmente se elaboran en los centros de investigación y de diseño de políticas, tanto públicos como privados.

En ese sentido, cabe enfatizar que los usuarios especializados requieren disponer esencialmente de las bases de microdatos, con el fin de ampliar y profundizan sus posibilidades de análisis. De modo que las unidades ejecutoras debieran asumir que las actividades de la encuesta continúan más allá de la generación de tabulados y la publicación de resultados, al proporcionar a los usuarios los archivos con los registros individuales, así como asesoría estadística para garantizar el buen uso de los datos y preservar la confidencialidad de la información.

d) Estructura funcional

Se requiere establecer con claridad los diferentes niveles de mando que intervienen en el desarrollo de la encuesta, así como las funciones y responsabilidades asociadas a cada uno de ellos. Para esto es recomendable elaborar un organigrama que represente la estructura jerárquica y las interrelaciones de las distintas áreas que intervienen en el diseño y ejecución de la encuesta. Asimismo, debiera elaborarse un manual de operación que defina los procedimientos administrativos que corresponde cumplir para la solicitud y comprobación de recursos, el envío y recepción de materiales, el uso de vehículos y combustible, los diversos gastos que se requieren para la elaboración y actualización del marco de muestreo, la impresión de materiales, la ejecución del trabajo de campo, etc.

De paso, esta información resulta esencial para futuras investigaciones en la medida que los diseños óptimos de encuestas requieren disponer de datos sobre los costos de ejecución en las diferentes fases, para determinar con exactitud el número óptimo de unidades de selección de primera y segunda etapa. Y, en otro orden, permite también elaborar controles para la distribución de los materiales de trabajo (útiles de oficina,

cuestionarios, manuales, etc.) y diseñar diversos formatos para el manejo de las actividades de la encuesta.

2. Diseño estadístico

El diseño estadístico de una encuesta comprende el desarrollo de las siguientes actividades: i) Determinación del tamaño de muestra; ii) Elaboración del marco de muestreo; iii) Selección de las unidades que formarán parte de la muestra; y iv) Definición de los procedimientos de estimación y cálculo de los errores de muestreo. Cada una de estas actividades debe ser planeada y ejecutada con rigor y eficiencia, con el fin de contribuir a minimizar los errores de muestreo y los sesgos de estimación.

a) Tamaño de la muestra

Uno de los primeros requerimientos del diseño estadístico de una encuesta es identificar, dentro del conjunto de variables para las que se propone generar información, cuál(es) de ella(s) se utilizará(n) para determinar el tamaño de la muestra. Al mismo tiempo, se debe decidir si se obtendrán estimaciones para proporciones, promedios o totales, así como la precisión y confianza deseada para los resultados.

Conocer la varianza de las variables de diseño es fundamental para calcular el número óptimo de selecciones que permitan lograr estimaciones con la precisión establecida. No obstante, es común que en muchas encuestas (sobre todo las que estudian el mercado laboral y la fuerza de trabajo y también las que estudian las condiciones de vida) el tamaño de muestra se calcule como si el objetivo de la encuesta se centrara exclusivamente en estimar proporciones, aunque posteriormente la información sea utilizada para calcular promedios y totales. En este sentido, se debe enfatizar que esta práctica tiene consecuencias en el error de muestreo, ya que éste se incrementa necesariamente por el simple hecho que la muestra calculada no resulta suficiente para garantizar la precisión esperada en estimaciones de promedios y totales. A su vez, también es habitual que no se disponga de estimaciones acerca del error de muestreo de la variable de interés para los diseños estratificados y de conglomerados. En este caso, comúnmente el tamaño de muestra se determina como si se fuera a efectuar una selección con igual probabilidad, y posteriormente el valor obtenido se incrementa por un factor de ajuste denominado efecto de diseño (efd), así como por las posibles pérdidas de observaciones que se puedan originar durante la ejecución del trabajo de campo, o tasa de no respuesta esperada (tnr).

Esta fase suele desarrollarse en forma rutinaria y muchas veces no se le presta la debida atención. Como se sabe, un menor número de observaciones que las necesarias incrementa los errores de muestreo e introduce sesgos importantes en los datos.

Cuando no se dispone de valores actualizados sobre el efecto de conglomeración del marco (efd) y se utiliza alguno obtenido de una investigación "similar", es muy probable que se subestime el tamaño de muestra requerido para obtener resultados con la precisión deseada. Asimismo, si se sobreestima el valor de este factor de ajuste el número de

selecciones se incrementa innecesariamente, lo cual redundaría en un desperdicio de recursos. Es importante recordar que cada encuesta tiene características propias y que éstas varían conforme a las circunstancias de la investigación, la participación de los informantes, la calidad de los encuestadores, etc.. En tal sentido, se señala que deben actualizarse los coeficientes de variación y efectos de diseño de las variables, ya que el no hacerlo puede significar pérdida de precisión por malos cálculos en el tamaño de muestra.

Por último, para prever la pérdida de observaciones en campo, por motivos de rechazo de los informantes y defectos del marco, y preservar el grado de precisión, es recomendable introducir ajustes en el tamaño de la muestra. De modo que se considera una mala práctica incorporar reemplazos de manera deliberada, ya que además de alterar las probabilidades de selección de las unidades de observación, induce a un mal uso por parte de los entrevistadores e introduce sesgos de estimación difíciles de evaluar.

b) Marco de muestreo

Para seleccionar las unidades que formarán parte de la encuesta se requiere disponer de una relación que identifique de manera inequívoca a cada una de ellas. Sin embargo, por lo general no se dispone de un listado que permita la ubicación y selección directa de las unidades de observación, lo que hace necesario elaborar diseños polietápicos que suponen la confección de marcos de muestreo para cada fase de selección.

A su vez, para minimizar los errores de selección e identificación de las unidades se requiere que el marco de muestreo esté actualizado y no adolezca de duplicaciones u omisiones. En tal sentido, y a pesar que el marco puede haberse elaborado con información reciente, es necesario efectuar conteos que permitan actualizar el número de viviendas existentes en las unidades seleccionadas, y luego listarlas a fin de identificar aquellas que son susceptibles de ser entrevistadas. En el caso en que se trate de una encuesta de hogares, se eliminan de este proceso los establecimientos económicos y las viviendas de uso colectivo, por no formar parte de la población objeto de estudio.

El listado de viviendas debiera contar con un croquis que permita ubicar físicamente la manzana o área geográfica seleccionada, así como el segmento donde se encuentran las viviendas que se desea visitar. A fin de que los entrevistadores puedan identificarlas con precisión, se deben describir los rasgos físicos perdurables de la zona y de la vivienda, de tal suerte que sean fácilmente identificables en el terreno.

Asimismo, para no incorporar distorsiones en la selección de la muestra, durante la etapa de elaboración del marco se deben asignar a las unidades definidas medidas de tamaño (pesos) actualizadas, con el objeto de diferenciar su importancia relativa tanto en el proceso de selección como en la fase de generación de resultados.

En todos los casos es deseable que los procedimientos utilizados tengan la capacidad de ser replicables, en el sentido que si otra persona siguiera las instrucciones definidas necesariamente debiera aproximarse al mismo resultado.

A fin de apoyar la consolidación y resguardo de la información, es importante recurrir a métodos de automatización que permitan clasificar e indexar la base de datos en función a los atributos de interés definidos por los usuarios, a fin de facilitar los procedimientos de estratificación y conglomeración en el caso de las encuestas polietápicas.

c) Selección de la muestra

La etapa de selección de la muestra es fundamental para elegir de manera inequívoca a las unidades que formarán parte de la investigación. Así, a partir del esquema de selección definido se deberán seleccionar las unidades en las diferentes etapas, identificando el estrato de procedencia, así como el área geográfica a la que pertenece el conglomerado seleccionado. En todos los casos, se deberá registrar el "peso" de la unidad seleccionada con el fin de poder determinar la probabilidad de selección de las viviendas y los factores de expansión asociados.

La incorporación de los Sistemas de Información Geográfica (SIGs) a la ejecución de proyectos de generación de información estadística es una realidad en los países de la región. En este sentido, es cada vez más común que los marcos de muestreo estén resguardados en medios magnéticos y los procedimientos de conglomeración y selección de la muestra se hagan en forma automatizada. Sin embargo, se deberá tener cuidado en la ejecución mecánica de los procedimientos ya que para efectos del diseño y la estrategia de control y supervisión del trabajo de campo, no siempre las soluciones matemáticas obtenidas mediante algún algoritmo de optimización resultan ser las más eficientes desde el punto de vista práctico. Asimismo, se debe recordar que para efectos de muestreo se debe dar prioridad a la identificación física de la vivienda y no a la precisión geográfica de su ubicación, por lo que la falta de referencia métrica de la cartografía no debe ser impedimento para automatizar estas fases del proceso y disponer de planos de línea que se puedan actualizar e imprimir en forma automática.

d) Procedimiento de estimación y cálculo de los errores de muestreo

La manera de evaluar si una encuesta cumplió con las especificaciones previstas en el diseño, en relación a la confiabilidad de los estimadores, es mediante el cálculo de los errores de muestreo para las variables de interés. De modo que una vez definidas las expresiones que se utilizarán para construir los estimadores, se debe especificar claramente en el diseño de la muestra el procedimiento que se aplicará para el cálculo de los errores de muestreo de las variables de interés.

Pese a que en los documentos elaborados por los países muchas veces se indica que para la estimación de la varianza se pueden utilizar algunos de los distintos métodos conocidos (Muestras Replicadas, Replicaciones Repetidas y Balanceadas, Replicaciones Repetidas por el método Jackknife y el método de Series de Taylor), en la mayoría de los casos la información contenida en la ficha técnica de la encuesta no permite evaluar la factibilidad de aplicar estas alternativas. Por ejemplo, el método de muestreo replicado requiere que se hayan realizado al menos dos selecciones (un número par) al interior de cada

unidad primaria de muestreo (UPM). Así, si esta información no se consigna en el reporte de la encuesta y la base de datos no permite identificar las selecciones por UPM, el analista no estará en condiciones de calcular el error de muestreo.

Para efectuar las estimaciones es necesario asociar a cada una de las observaciones el inverso de su probabilidad de selección, llamado también factor de expansión. En esta tarea se deben tener en cuenta las diferentes etapas de selección de la muestra, así como los pesos relativos de las unidades de selección. Asimismo, en el cálculo del factor de expansión se requiere considerar los ajustes necesarios por no respuesta y por la post-estratificación de los datos, toda vez que los valores muestrales de las estructuras por edad y sexo se contrasten con los resultados obtenidos a partir de un censo reciente o de una proyección de población confiable.

III. Levantamiento de la información

1. Organización del trabajo de campo

Una de las etapas más importante en el desarrollo de una encuesta es la ejecución del trabajo de campo. Es posible elaborar un buen cuestionario, o efectuar un óptimo diseño de muestra e impartir una excelente capacitación a los encuestadores; sin embargo, la falta de una adecuada organización y supervisión del trabajo de campo puede llegar a invalidar los resultados de la investigación.

Por ende, es imprescindible que la unidad ejecutora confeccione un manual en el que se especifiquen detalladamente los aspectos organizativos del levantamiento y las facultades, atribuciones y responsabilidades del personal, así como las relaciones entre las diferentes instancias que participan en la planeación y ejecución del trabajo de campo.

A su vez, se deben definir los controles necesarios para garantizar que se minimicen los errores de muestreo y ajenos al muestreo, así como los sesgos involuntarios que se pueden introducir producto de decisiones inadecuadas adoptadas por supervisores y/o entrevistadores. Al respecto, es recomendable que el equipo de campo evite adoptar iniciativas sustantivas sin el previo consentimiento de la coordinación central del proyecto, y al mismo tiempo asegure que los criterios aplicados para la clasificación de las respuestas estén de acuerdo con los lineamientos entregados en la etapa de capacitación y consignados en el manual correspondiente.

Por otra parte, es conveniente elaborar bitácoras de problemas surgidos en las diferentes zonas donde se esté llevando a cabo la encuesta y de sus respectivas soluciones, a fin de dar a conocer al personal de campo las decisiones adoptadas ante cada situación. Asimismo, y valiéndose de las facilidades que ofrecen los medios de comunicación modernos, como el correo electrónico, las oficinas regionales debieran estar permanentemente informadas sobre los criterios asumidos por la coordinación del proyecto,

respecto a las dudas surgidas durante el trabajo del campo en las distintas zonas geográficas en donde se realiza la investigación.

2. Capacitación del personal de campo y de oficina

La puesta en marcha de un adecuado programa de capacitación del personal involucrado en el levantamiento de la información es otra tarea fundamental. Este programa debiera abarcar todos los niveles de la unidad ejecutora, a la vez que es necesario que contenga prácticas en campo que permitan conocer el desempeño y rendimiento de la estructura operativa que participará en la fase de ejecución: supervisores, entrevistadores, codificadores, choferes, etc.. Es aconsejable que tales actividades comprendan la capacitación en sala, con ejercicios prácticos, a partir de juego de roles, en donde se ponga a prueba la claridad conceptual y de redacción del cuestionario, así como la capacidad de los manuales para responder a dudas concretas planteadas en situaciones reales. Asimismo, se sugiere realizar evaluaciones permanentes que fomenten el intercambio de puntos de vista entre los responsables de la elaboración del marco conceptual y metodológico, por una parte, y los entrevistadores y el personal encargado de hacer la supervisión y la crítica-codificación de la información, por la otra. Al mismo tiempo, se debieran planear sesiones de prácticas de campo en distintos ámbitos geográficos, seguidas del correspondiente análisis de evaluación que permita perfeccionar los procedimientos, así como precisar conceptos o incorporar eventuales correcciones que afecten incluso la redacción de algunas preguntas. Respecto a esto último, es necesario recalcar que la correcta comprensión por parte de los empadronadores tanto de los conceptos a investigar como de las preguntas que serán formuladas a los entrevistados, son una condición esencial para disminuir los errores de no muestreo y los probables sesgos de la información.

3. Distribución de la muestra y de las cargas de trabajo

Existen otras actividades relativas al levantamiento de los datos a las que también se les debe prestar especial atención. Una de ellas se refiere a lograr una adecuada distribución de la muestra de conformidad a los plazos estimados para la realización de la encuesta. Esta debe ser congruente con el calendario en términos del tiempo asignado para el levantamiento y la entrega de resultados, buscando el equilibrio entre la disponibilidad de recursos y las necesidades del trabajo de campo. A su vez, se debiera apuntar a que los diferentes períodos proyectados para la recopilación de la información (semana, mes, trimestre, etc.) mantengan un balance adecuado, con el objeto de no saturar las actividades en alguna fecha determinada que pudiera incrementar desproporcionadamente la carga de trabajo de los encuestadores y afectar la confiabilidad de la información. Asimismo, es importante distribuir la muestra en función de las condiciones climáticas, privilegiando aquellas zonas o regiones de difícil acceso o en las que sólo se puede transitar en ciertas épocas del año, a fin de garantizar que las entrevistas se realicen en los plazos señalados.

Por otra parte, la asignación de las cargas de trabajo a los encuestadores y supervisores debiera estar acorde con la capacidad física del personal, las condiciones del terreno y el tiempo estimado para llevar a cabo la entrevista. En este sentido, el número de

visitas diarias que razonablemente puede realizar un entrevistador dependerá de la extensión del cuestionario y la complejidad de las preguntas, la distancia entre las viviendas (tanto en las zonas urbanas como en las rurales), la duración promedio de la entrevista y el número de visitas necesarias para lograr su cooperación. Se debe velar por no sobrecargar al personal de campo, ya que ello podría llevar a que las entrevistas se realicen sin el debido cuidado o, peor aún, se estimule el autollenado de cuestionarios por parte de los enumeradores. En ambos casos se incrementan los sesgos de estimación, con el agravante que, como se ha dicho, es muy difícil evaluar su efecto en la calidad de los resultados.

4. Supervisión del trabajo de campo

La supervisión del trabajo de campo es otra de las tareas fundamentales para garantizar buenos resultados. En primera instancia, se debiera conformar un equipo encargado de la supervisión de los entrevistadores sin que sus integrantes tengan la responsabilidad directa de realizar entrevistas. Ello implica destinar recursos para que este grupo de personas, con mayor nivel de capacitación y entrenamiento, evalúen la labor de los enumeradores y además dediquen tiempo para realizar reentrevistas que permitan corroborar la veracidad de la información proporcionada por los miembros del hogar. De este modo, la supervisión permite aclarar dudas que se presentan en los datos, verificar si el encuestador está cumpliendo con su trabajo y, en general, evaluar la calidad de la información recopilada.

Actualmente las oficinas de estadística están preocupadas por reducir el tiempo que media entre la ejecución del trabajo de campo y la entrega de resultados, lo cual sin duda es muy importante para mejorar la oportunidad de la información de cara al diseño de políticas y la toma de decisiones. En esta dirección se han realizado ejercicios que incorporan el uso de computadoras portátiles para ingresar en campo la información proporcionada por los hogares, desarrollando en forma paralela la fase de crítica y análisis de consistencia.

A primera vista esta situación podría considerarse ideal, ya que la información se ingresa y valida en el momento mismo en que se captura, a partir del diseño de sistemas en línea que permiten definir criterios de consistencia de los datos. Sin embargo, se debe tener presente que es muy difícil que los diseños de gabinete prevean todas las contingencias que surgen en la realidad, y por lo tanto cubran suficientemente los diversos aspectos relativos a la congruencia de la información. De allí que, especialmente en encuestas complejas, la experiencia hace aconsejable mantener y reforzar la etapa de crítica y codificación, como una tarea fundamental para detectar errores e inconsistencias no identificados en la fase de supervisión, lo que contribuye a mejorar la calidad de la información.

Por otra parte, en aquellas encuestas que contienen preguntas cuyas respuestas deben codificarse después de realizar el análisis de un conjunto de variables y la consulta de nomencladores específicos, es difícil realizar esta actividad en forma automática. Un ejemplo de ello es la codificación del tipo de ocupación o la rama de actividad económica en que se desempeña la fuerza de trabajo, que en la medida que se registre con un amplio nivel

de desagregación requiere de gran experiencia por parte de los codificadores, así como el uso de catálogos especializados.

La captura y consistencia de la información en campo no implica que la congruencia interna de los datos sea suficiente para minimizar los errores de muestreo y garantizar la validez de los resultados. Como se sabe, el error cuadrático medio de los estimadores se compone del error de muestreo y el sesgo, donde éste último componente no puede evaluarse estadísticamente. A su vez, los errores ajenos al muestreo generan los mayores sesgos y distorsiones respecto de la realidad y es difícil que la incorporación de procedimientos automáticos puedan disminuirlos de manera significativa. En caso que existan errores en los datos, tales procedimientos automatizados tenderán más bien a estandarizarlos pero no a eliminarlos. De modo que el propósito de reducir el tiempo empleado en el procesamiento de la información no debe llevar a descuidar el resto de la etapas. Así, no sirve de mucho obtener resultados rápidos si los tamaños de muestra utilizados no son apropiados para cumplir con los objetivos planteados y no se conocen los errores de muestreo asociados a las estimaciones, ni la consistencia de la información con otras fuentes de datos disponibles. Más aún, la pronta disponibilidad de resultados no elimina de por sí las malas prácticas en el uso de la información en el caso en que las investigaciones que se efectúen apliquen, por ejemplo, métodos de análisis que no consideren como es debido la estructura del diseño de la muestra.

En la práctica, la mayoría de las oficinas de estadística de la región asignan importantes recursos a las tareas de supervisión y crítica-codificación. Muchas de ellas incluso han diseñado sistemas de captura y consistencia de la información que les permiten depurar las bases de datos en tiempo real, a medida que los datos se ingresan, con lo cual han reducido de manera apreciable el tiempo que transcurre entre la realización del trabajo de campo y la entrega de resultados. Sin embargo, corresponde insistir en que debe prestarse una mayor atención a la congruencia interna de los datos, la consistencia de la información generada por la encuesta con aquella otra disponible en el sistema nacional de información, así como a su comparación con los resultados de encuestas similares realizadas en el pasado.

Esto no desestima, sino muy por el contrario refuerza, los intentos por entregar resultados oportunos pero a la vez de la mayor confiabilidad. El éxito de una investigación por muestreo no se basa en disponer de computadores en el terreno, si acaso se olvida que los errores de los datos se generan de hecho en diferentes instancias y que un proceso automatizado no es suficiente por sí sólo para eliminar los sesgos. De cualquier manera, se debe seguir propugnando que se reduzcan los plazos para la entrega de resultados, ya que en ocasiones la información es conocida varios años después de concluido el trabajo de campo, con lo que ésta se incorpora al acervo histórico del país pero sin representar una contribución efectiva al estudio de la coyuntura y, por ende, a la toma de decisiones.

Por último, es preciso señalar que la fase de supervisión y reentrevista a los hogares no debe entenderse como una tarea que signifique volver a estudiar toda la muestra o seguir a los entrevistadores por toda su área de trabajo. Para cubrir satisfactoriamente esta

etapa basta con el uso del muestreo de lotes, lo cual permite llevar un registro y evaluar el trabajo efectuado por los entrevistadores, generar una calificación de su desempeño en el terreno y controlar la calidad de la información. De hecho, esta práctica se sugiere para las encuestas de panel donde los hogares son visitados en más de una ocasión. Además, es recomendable rotar a los entrevistadores a fin de minimizar las posibilidades de repetición de información en forma deliberada o el autollenado de cuestionarios. Esta práctica es frecuente sobretodo cuando los entrevistadores visitan las mismas viviendas durante diversas rondas de una misma encuesta (por ejemplo, en las encuestas continuas de hogares que investigan el tema del empleo).

IV. Tratamiento de la información

1. Crítica y codificación

En esta etapa se definen los códigos válidos para cada categoría de respuesta y sus correspondientes rangos, las subpoblaciones de interés para cada tema y las relaciones permitidas entre las variables.

Las oficinas encargadas de administrar las encuestas debieran disponer de un grupo de técnicos con gran experiencia y profundo conocimiento de los temas investigados y de los alcances del cuestionario, encargados exclusivamente de evaluar la calidad de la información recolectada y la asignación de códigos a las preguntas abiertas que requieren clasificarse.

A pesar de que resulta aconsejable que la mayor parte de las preguntas están precodificadas en el formato del cuestionario, existen ciertos temas en los cuales -como se indicó más arriba- no es conveniente que se efectúe una asignación automática ya que habitualmente se cometen errores de interpretación y clasificación. Por ejemplo, y como ya se mencionó, cuando se investiga el tema del empleo y se pretende codificar el tipo de ocupación (principal y secundaria) que realizan los miembros del hogar que manifestaron estar ocupados durante el período de referencia de la encuesta, es habitual que se presenten errores en la codificación efectuada en campo. De allí que para lograr información de calidad, se solicita a los entrevistadores que obtengan de los informantes ocupados una amplia descripción de sus actividades, a los efectos que en la oficina se pueda asignar posteriormente un código apropiado a partir de las alternativas consignadas en la clasificación de ocupaciones elaborada por los organismos internacionales especializados en el tema del empleo. Asimismo, hay situaciones en las que tampoco se puede inferir en forma directa la codificación de la rama de actividad de la empresa por lo que es necesario revisar catálogos específicos elaborados para este fin.

Se debe estar conscientes, sin embargo, que es normal que se presenten problemas de campo no previstos en los manuales y formatos de llenado. De hecho es prácticamente imposible prever todas las situaciones y reflejarlas en la modalidad de captura, crítica y

validación de los datos en campo, sin quedar expuesto a que se generen una cantidad considerable de errores de clasificación y sin que los programas de verificación sean capaces de detectar tales inconsistencias. Del mismo modo, es preciso reiterar una vez más que los sesgos en los datos son difíciles de evaluar y en la mayor parte de los casos son más dañinos que los errores de muestreo.

Por otra parte, la etapa de crítica y codificación también permite evaluar el trabajo de los supervisores y representa otra instancia de control de la actuación de los encuestadores. Llevar registros apropiados y desarrollar actividades de actualización del personal son mecanismos que ayudan a reducir los errores ajenos al muestreo sobre todo en las encuestas permanentes y/o continuas.

2. Falta de respuesta

La pérdida de información es una situación que comúnmente se presenta durante la ejecución del trabajo de campo. De allí que llevar un control adecuado de sus causas y su distribución entre grupos de hogares y regiones del país sea visto como un elemento fundamental para reducir sesgos de estimación y preservar al máximo la calidad de la información. La no respuesta siempre va a existir, pero lo importante es tener una idea precisa de su magnitud y localización a fin de realizar los ajustes pertinentes en el tamaño de muestra y reducir las causas que la originan.

La falta de actualización del marco de muestreo es una de las situaciones que provocan que no se identifiquen los segmentos seleccionados o se subregistre el número de viviendas existentes en alguna zona determinada. Esto puede preverse y controlarse, efectuando recorridos en las zonas seleccionadas y elaborando conteos y listados de viviendas en las fechas cercanas a la realización del trabajo de campo. En muchas ocasiones la desactualización del marco ocurre a gran velocidad, por lo que es recomendable que no exista mucha diferencia entre la fecha de actualización y la correspondiente a la realización del trabajo de campo.

A su vez, la no respuesta por rechazo del informante se puede minimizar mediante visitas a los hogares por parte de los supervisores, o informando previamente sobre la realización de la encuesta por medio de cartas dirigidas a los jefes de los hogares donde se expongan sus objetivos y la fecha en que serán visitados por personal de la oficina de estadística.

A partir de experiencias previas, es posible estimar la tasa de respuesta esperada e incrementar el tamaño de muestra a fin de lograr el número mínimo de entrevistas para obtener la precisión requerida en las estimaciones. Como se indicó, no se considera recomendable realizar sustituciones de las viviendas no entrevistadas por otras que hayan sido preseleccionadas y definidas como reemplazos. Este procedimiento introduce sesgos y altera las probabilidades de selección de las observaciones. En todo caso, como una actividad posterior a la conclusión del trabajo de campo y antes de ajustar los factores de expansión, se sugiere efectuar un análisis de la no respuesta total, por causas, por

entrevistador, por grupos de hogares y por zonas geográficas, a fin de evaluar posibles errores de omisión o subespecificación. Mediante este procedimiento se pueden identificar posibles sesgos u omisiones, determinar la calidad del marco y evaluar la eficiencia de la labor realizada por los enumeradores. La concentración de la no respuesta en alguna zona o grupo social específico puede distorsionar la calidad de la información y en ocasiones incluso invalidar los resultados de la investigación.

Cuando el número de entrevistas no logradas es poco significativo o se encuentra dentro del rango estimado, la eliminación de éstas observaciones no afecta la precisión de los datos. Sin embargo, será necesario ajustar los factores de expansión con el objeto que los resultados generados por medio de la muestra se aproximen a los valores verdaderos del universo de estudio (post-estratificación). Dependiendo de la magnitud y distribución de la no respuesta los factores de expansión originales tenderán a modificarse, por lo que es casi un hecho que los diseños autoponderados perderán esa característica al finalizar la etapa del trabajo de campo. Así, un buen registro y control de la falta de respuesta es un requisito para garantizar la confiabilidad de los datos reportados.

Por otra parte, es también necesario prestar atención a la no respuesta parcial, la cual ocurre cuando algunas preguntas o secciones del cuestionario no son respondidas por el hogar o alguno de sus miembros. Cuando la frecuencia de ocurrencia es baja, las observaciones que presentan esta condición pueden ser eliminadas sin que ello afecte la precisión de los resultados. De lo contrario, es posible aplicar procedimientos automáticos para imputar valores faltantes, lo que no necesariamente contribuye a mejorar la calidad de las estimaciones y puede llevar a incrementar el error del estimador.

En efecto, esta práctica permite reducir el número de preguntas sin respuesta pero la imputación no siempre es el método apropiado para reducir los errores de estimación. A pesar de que las observaciones ubicadas en una misma unidad muestral tienen una alta correlación en relación con la variable de estratificación, esto no necesariamente significa que ésta se preservará para cualquier característica de la población que se desee estimar.

3. Captura y congruencia de los datos

La fase de captura de datos representa el primer paso para la sistematización de la información. El desarrollo tecnológico y los programas disponibles permiten que esta etapa sea ahora más simple que las prácticas adoptadas en el pasado. Actualmente, la capacidad de las computadoras personales y las facilidades para compartir recursos que brindan las redes locales han desplazado a los antiguos "main frames", que requerían de amplios espacios, ambientación especial y complicados lenguajes de programación. A su vez, los modernos sistemas de digitación permiten el diseño de pantallas a imagen del cuestionario, la definición de criterios de validación, la verificación de información durante la captura y la grabación en línea de los datos. Los tiempos de respuesta se han acortado y la mayor velocidad y capacidad de memoria permiten reducir los plazos para la depuración y procesamiento de la información. Por último, también se sugiere aquí aplicar métodos de

muestreo por lotes para verificar la calidad de la información capturada, así como para evaluar el desempeño del personal encargado de la digitación.

4. Consistencia de la información

Esta etapa se realiza toda vez que se ha concluido con las tareas de digitación y comprende al menos las dos etapas siguientes: i) análisis de consistencia interna y ii) validación de los resultados con encuestas sobre el mismo tema realizadas en el pasado, con otras encuestas que generen datos sobre temas comunes, datos macroeconómicos, estudios internacionales y registros administrativos. En este contexto, esta actividad es fundamental ya que la calidad de la información no se puede calificar de manera exclusiva por la tasa de respuesta, por la baja incidencia de preguntas sin respuesta o la rapidez para la obtención de resultados.

Los datos generados se considerarán de buena calidad y útiles para la toma de decisiones en la medida que sus órdenes de magnitud guarden relación con el resto de cifras micro y macro del sistema nacional de información. Asimismo, se podrán considerar adecuados desde el punto de vista estadístico en la medida que el orden de su error absoluto y relativo esté en un rango aceptable, el cual fue definido al momento de establecer los objetivos de la investigación.

5. Precisión estadística de los estimadores

La evaluación de la calidad de la información desde la perspectiva estadística se efectúa a partir del cálculo de los errores de muestreo y del efecto de diseño. En primer lugar, los responsables de la administración del proyecto deben verificar si el margen de error utilizado en la determinación del tamaño de muestra está acorde con los resultados observados en la encuesta para la(s) variable(s) de diseño y algunas otras que estén altamente correlacionadas con ésta(s). Asimismo, se debe calcular el efecto de diseño y verificar si su valor es similar al que se utilizó para ajustar el tamaño de muestra debido a la conglomeración del marco.

Estas comparaciones mínimas, permitirán determinar si el esquema de muestreo aplicado fue adecuado y si el número de observaciones seleccionadas fueron suficientes para generar resultados con la precisión requerida. En caso que el error estimado se encuentre en un rango aceptable, será posible afirmar que los resultados obtenidos se pueden utilizar para establecer relaciones de causalidad y realizar inferencia al universo de estudio con la confianza asumida en los objetivos de la investigación. En ese sentido, es preciso recordar que el error de estimación y el efecto de diseño son diferentes para cada variable y que su valor cambia entre las diferentes encuestas que se realizan, por lo que en la medida de lo posible se debe evitar utilizar resultados de otras investigaciones para estimar tamaños de muestra, calcular coeficientes de variación, límites de intervalos de confianza o probar la significancia estadística de algún parámetro.

6. Documentación de los procesos de la encuesta

Las actividades posteriores a la generación de resultados también deben considerarse como fases importantes en el control de calidad de la información. La documentación de las diferentes fases del proceso de la encuesta, así como de las decisiones que se asumieron ante problemas concretos, resultan fundamentales para evaluar la consistencia de los resultados y explicar distorsiones o sesgos en los datos en caso de que estos existan. Esta afirmación es válida para todas las etapas de la encuesta, pero sobre todo en la que se vincula en forma directa con el tratamiento de la información y los procedimientos de ajuste e imputación que se aplican para suplir los datos faltantes.

En la medida que no se documenten las decisiones adoptadas sobre, por ejemplo, la concentración de la falta de respuesta total y parcial, los motivos y criterios para el ajuste de los factores de expansión, los procedimientos utilizados para la corrección de los ingresos y la imputación de la condición de actividad de la población ocupada, los usuarios de la información carecerán de elementos para determinar si sus análisis y las hipótesis en las que se basan los modelos de comportamiento se cumplen en función de la naturaleza de los datos.

Adicionalmente, se debe enfatizar la conveniencia de proveer a los usuarios de todos los elementos necesarios para que juzguen la calidad de la información y evalúen la posibilidad de utilizarla para la toma de decisiones. Así, se debe mejorar el contenido de la ficha técnica que acompaña la documentación de la encuestas. En el mejor de los casos, los documentos que se ponen a disposición de los usuarios contienen un resumen muy apretado del diseño estadístico de la encuesta y de algunos de los parámetros que se utilizaron para determinar el tamaño de la muestra. Sin embargo, son muy pocas las ocasiones en que se divulga el esquema de selección de la muestra y se identifican las unidades primarias y secundarias de selección, la distribución geográfica y por estrato social de la no respuesta, el miembro del hogar que proporcionó la información, el número de preguntas que se tuvieron que imputar, etc..

Se debe recordar que una buena manera de sensibilizar al usuario sobre los malos usos de la información es indicarle con claridad las limitaciones de los datos y las restricciones impuestas por el tamaño de muestra y los objetivos de la investigación en aquellos casos en que la información no puede desagregarse para ciertos dominios de estudio. Además, deben proporcionársele los insumos informativos que le permitan estimar el error de muestreo de las variables que le interesen.

7. Difusión de resultados y relación con los usuarios

Las responsabilidades de las oficinas de estadística no se acaban con la divulgación del informe que contiene los resultados básicos de la encuesta. La relación con los usuarios debe ser permanente y cubrir un amplio espectro de actividades, que van desde la definición conjunta de los objetivos de la investigación a la retroalimentación mutua a partir de la entrega de las bases de datos con los registros individuales de la encuesta.

Es recomendable llevar un control de las personas e instituciones a las que se les facilita la información, y establecer algún tipo de convenio en que la oficina de estadística se comprometa a brindar asesoría sobre aspectos específicos de la encuesta y el archivo de datos y los usuarios asuman el compromiso de consultar sobre la confiabilidad de la información antes de su publicación y la responsabilidad por los errores y/o imprecisiones que se cometan en la interpretación de los resultados.

**DOCUMENTACIÓN METODOLÓGICA
DE LA ENCUESTA DE HOGARES**

GILBERTO MONCADA*
(BID)

* El autor agradece los comentarios de Rob Vos y Margaret Grosh.

ÍNDICE

	<u>Página</u>
I. ANTECEDENTES.....	255
II. OBJETIVOS.....	255
III. PRINCIPALES DOCUMENTOS PROPUESTOS DEL SISTEMA INTEGRADO DE ENCUESTAS DE HOGARES	256
1. Programa del Sistema Integrado de Encuestas de Hogares.....	257
2. Cuestionarios.....	259
3. Manual del encuestador.....	259
4. Manual del supervisor	261
5. Muestra.....	262
6. Metodología de la operación de campo.....	262
7. Manual del codificador.....	262
8. Manual de entrada de datos	264
9. Manual de códigos utilizados.....	264
10. Información metodológica básica para usuarios	265
11. Variables agregadas y errores muestrales.....	267
12. Informe ex-post de la operación de campo.....	268
13. Manual de uso de la base de datos y códigos de las variables.....	268
14. Perfil de condiciones de vida	268
15. Información sobre principales resultados.....	269
16. Publicación de divulgación popular	270

DOCUMENTACIÓN METODOLÓGICA DE LA ENCUESTA DE HOGARES

I. ANTECEDENTES

Una de las lecciones más importantes de la experiencia y la buena práctica en la ejecución de encuestas de hogares es la necesidad de documentar todos los procedimientos y métodos utilizados en cada una de las etapas de la encuesta: planeamiento y diseño, implementación, control de calidad, procesamiento, análisis y divulgación de resultados.

Lamentablemente, ésta es una práctica muy poco realizada por las diferentes instituciones que llevan a cabo encuestas de hogares en América Latina y El Caribe, lo cual tiene varias consecuencias. De una parte, limita el uso potencial de las bases de datos al no permitir conocer el alcance y debilidades de la información; además, dificulta la etapa del procesamiento e interpretación plena de los resultados. Asimismo, la no documentación de las encuestas de hogares genera un costo enorme en el proceso de aprendizaje y acumulación de conocimientos y buenas prácticas en las instituciones productoras de este tipo de información, dado que se desaprovechan las lecciones de las experiencias anteriores.

Uno de los principales propósitos del Programa MECOVI, a la luz de la experiencia señalada, es justamente estimular y fortalecer la capacidad de las oficinas de estadística para que se documente todos los procesos llevados a cabo durante la ejecución de las encuestas de hogares.

II. OBJETIVOS

El objetivo del presente documento es proveer a las oficinas de estadística un esquema básico de la información metodológica que debe ser organizada de manera sistemática para las encuestas de hogares que se desarrollarán dentro del marco del Programa MECOVI. La disponibilidad y sistematización de esta información será muy beneficiosa para las instituciones productoras de estadísticas, y para el país, por las razones expuestas anteriormente.

Se busca también generar una interacción entre las oficinas de estadística y las instituciones co-patrocinadoras del Programa MECOVI, a fin de llegar a un consenso en relación a la información de soporte que debe ser preparada a nivel de los países integrantes del Programa, y, de ser posible, a nivel regional, lo que facilitará lograr un objetivo de comparabilidad en el uso de las bases de datos.

Generalmente las oficinas de estadística producen alguna documentación metodológica básica, pero no siempre de manera sistemática y completa de todos los procesos de las encuestas, por ello se debe dar por entendido que esta es una tarea crucial que deben cumplir todas las entidades ejecutoras de encuestas de hogares en el marco del Programa MECOVI.

III. PRINCIPALES DOCUMENTOS PROPUESTOS DEL SISTEMA INTEGRADO DE ENCUESTAS DE HOGARES

Los documentos propuestos tienen carácter multipropósito, la mayoría de ellos se pueden usar para un mismo fin, por ejemplo, durante la etapa de análisis de la información; o también de modo independiente en alguna otra etapa de la encuesta de acuerdo al énfasis de cada uno de ellos. Algunos están diseñados principalmente para implementar la encuesta, otros particularmente para ser usados durante el procesamiento y análisis; y otros para proveer los resultados del análisis de los datos.

En general, los documentos se pueden organizar del siguiente modo:

Sobre información global del sistema de encuestas

1. Programa del Sistema Integrado de Encuestas de Hogares

Documentos utilizados principalmente para implementar y procesar la encuesta:

2. Cuestionarios
3. Manual del encuestador
4. Manual del supervisor
5. Muestra
6. Metodología de la operación de campo
7. Manual del codificador
8. Manual de entrada de datos
9. Manual de códigos utilizados

Documentos utilizados principalmente para el uso y análisis de las encuestas:

10. Información metodológica básica
11. Variables agregadas y errores muestrales
12. Informe ex-post de la operación de campo
13. Manual de uso de la base de datos y códigos de variables

Documentos que proveen los resultados del procesamiento y análisis de los datos recolectados:

14. Perfil de pobreza
15. Información sobre resultados principales
16. Publicación de divulgación popular

Algunos de los documentos propuestos también se podrían organizar como parte de un documento más grande, de manera temática, pasando cada uno de ellos a constituir un capítulo.

1. Programa del Sistema Integrado de Encuestas de Hogares

Este es uno de los primeros documentos que deben ser elaborados durante la etapa de planeamiento del sistema de encuestas a implementar. Su propósito es definir un perfil general del tipo de investigación a realizar, y que sirva de guía para las acciones a tomar. El documento resume de manera general los objetivos y propósitos de las encuestas, los principales temas a ser investigados, las características básicas de la investigación, y las definiciones y conceptos básicos.

El marco de referencia son los objetivos principales del Programa MECOVI: i) mejorar la calidad, cobertura, consistencia y complementariedad de la información a través de un conjunto articulado de encuestas de hogares o de módulos de una encuesta de propósitos múltiples; ii) aumentar la eficiencia en el uso de los recursos; iii) atender adecuadamente a diversos requerimientos de información del área social tanto en lo relativo a la naturaleza, periodicidad y alcance de la misma, como a su pertinencia para la formulación, seguimiento y evaluación de las políticas y programas sociales.

Asimismo, las condiciones necesarias para la existencia de un Sistema Integrado de Encuestas de Hogares (SIEH): i) Tener un marco muestral maestro; ii) uniformidad en conceptos, definiciones y formas de captar la información; iii) tratamiento complementario de la temática a través de diferentes instrumentos y/o en encuestas de tipo modular; iv) una clara programación de la temática en una estrategia de mediano plazo.

Si bien los objetivos y las condiciones necesarias para tener un SIEH mencionados anteriormente constituyen el marco de referencia para la preparación de este documento, cada oficina de estadística deberá organizarlo en función a los planes de acción específicos dentro del Programa MECOVI.

El contenido de este documento podría estructurarse de la siguiente manera:

A. *Objetivos generales y específicos.*

Indicar lo que se busca con las encuestas de hogares en términos de su significancia para la institución y para el país. Por ejemplo, satisfacer la demanda de información del área social, tanto para fines de investigación y análisis, así como para la formulación, seguimiento y evaluación de las políticas y programas sociales por las autoridades correspondientes. Asimismo, reforzar la capacidad institucional en cuanto a la realización de este tipo de investigaciones; buscar información específica sobre los programas sociales del gobierno a efectos de monitorear sus resultados, etc.

B. *Características básicas de la encuesta*

Describir el tamaño y las características generales de la organización, los niveles jerárquicos, la logística que será empleada, el diseño y la muestra seleccionada, los planes

para la ejecución de la operación de campo, el método a utilizar para la entrevista, los tipos de cuestionarios que serán usados, el control de calidad (crítica y consistencia), el procesamiento, y los resultados esperados.

C. Conceptos y definiciones básicas

Es necesario definir los principales conceptos y variables que se utilizarán en la investigación para generar un idioma común entre los principales involucrados. Por ejemplo, unidades de muestreo, vivienda particular, unidad de análisis, hogar, jefe de hogar, residente habitual, presupuesto común, receptor de ingresos, aportante principal, unidad de gasto, autoconsumo, autosuministro, pago en especie, etc.

D. Principales temas a ser investigados

Es necesario describir los temas o módulos a investigar, y dentro de cada uno de ellos definir las variables que potencialmente serían tratadas por la encuesta. Asimismo, estos temas deberían ser precisados en cuanto a la programación para su ejecución, de acuerdo a las rondas de encuestas a realizar.

Los módulos y variables podrían ser, entre otros:

- Características demográficas de los miembros del hogar (edad, sexo, parentesco, estado civil, lengua materna, etc.)
- Vivienda (valor de la vivienda, tipo de propiedad, servicios básicos, materiales de construcción, etc.)
- Educación y capacitación (alfabetismo, nivel de educación alcanzado, asistencia escolar, razones de no asistencia, capacitación para el trabajo, educación infantil, etc.)
- Salud y seguridad social (consulta médica, personal de salud, salud reproductiva de la mujer, salud infantil, vacunas, etc.)
- Antropometría (peso, talla, perímetro braquial)
- Actividad económica de los miembros del hogar (empleo, ocupación, sector de actividad, ingresos, etc.)
- Negocios del hogar e ingreso por trabajo independiente agropecuario y no agropecuario (tamaño de la parcela, producción, ingresos, ventas, autoconsumo, gastos en insumos, pago de planillas, etc.)
- Migración (interna, externa, causas, etc.)
- Gastos (alimentos, bienes y servicios no alimenticios, autoconsumo, autosuministro, etc.)
- Ahorro y crédito (fuentes, inversiones, etc.)
- Activos y bienes del hogar
- Uso de tiempo
- Otros ingresos (transferencias, remesas, pensiones, otros ingresos de capital, intereses, dividendos, etc.)
- Acceso a programas y servicios sociales.

2. Cuestionarios

Normalmente los cuestionarios pueden ser tres:

- i) Cuestionario de hogares
- ii) Cuestionario comunitario
- iii) Cuestionario de precios

Sin embargo, en caso de haber uno o más instrumentos de recolección, es igualmente recomendable incluirlos en la lista de documentos de soporte.

Sería conveniente describir, en un anexo a los cuestionarios, recomendaciones para su lectura adecuada, por ejemplo, el flujograma de las preguntas, la simbología utilizada, significado del tipo de letra utilizada, etc.

Una ventaja adicional para el procesamiento de la información es escribir en cada uno de los cuestionarios a distribuir, en las preguntas correspondientes, el código utilizado para su procesamiento. Para el uso de la base de datos de la encuesta es muy importante disponer de una copia de los cuestionarios utilizados.

3. Manual del encuestador

Este manual tiene por finalidad poner a disposición del encuestador las recomendaciones de carácter general respecto a la técnica de la entrevista y las instrucciones específicas el llenado del cuestionario.

En el encuestador recae la importante misión de recopilar información de buena calidad. Del buen desempeño de su trabajo dependerá, en gran medida, el éxito de la encuesta. Por lo tanto, es muy necesario insistir para que cada encuestador estudie detenidamente este manual y lo lleve siempre consigo para hacer consultas cuando realice las entrevistas durante la operación de campo.

El manual de encuestador podría incluir la siguiente información:

A. *Organización para la operación de campo*

Este punto está referido a establecer la dependencia jerárquica, funciones y responsabilidades del personal que tendrá a su cargo el recojo de información. Se debe especificar las responsabilidades del personal involucrado en la operación de campo (jefe de proyecto, supervisor, encuestador, etc.). De igual modo, la asignación de la carga de trabajo, los aspectos logísticos para la recepción y entrega de documentos y materiales, y las indicaciones para la redacción del informe de operación de campo que debe realizar cada uno de los encuestadores.

B. Tareas del encuestador

Es la persona que tiene la importante misión de solicitar y obtener la información verídica y confiable sobre el hogar y cada uno de sus miembros respecto a diferentes temas, y anotarlos en el cuestionario.

Se debe especificar de manera detallada cuales son las funciones y tareas que deben cumplir, así como los tipos de materiales que utilizarán (manuales, cuestionarios, etc.).

C. La entrevista

La entrevista es un modo de obtener información a través de preguntas efectuadas a personas para su respuesta inmediata y directa. Completar una entrevista con éxito es un arte y como tal no debe tratarse como un proceso mecánico. Debe ser ejecutada como una conversación normal entre dos personas; sin embargo, es necesario observar varias reglas básicas para el éxito de una entrevista.

Es necesario describir los aspectos más importantes de la técnica de la entrevista, como son el acceso a la persona entrevistada, la importancia del carácter privado de la entrevista, la confidencialidad de las respuestas, la neutralidad del encuestador respecto al contenido de la entrevista, el control de la entrevista o conducción de la misma, el modo de actuar frente a personas indecisas, y la importancia de la entrevista directa con la persona indicada.

De igual modo, es necesario que el manual contenga instrucciones específicas sobre el arte de hacer preguntas. Por ejemplo, las preguntas se deben hacer exactamente como están escritas en el cuestionario, se debe indagar sobre respuestas incompletas o no satisfactorias, no se debe asumir cosas por adelantado, no apresurar la entrevista, y enseñar como finalizar una entrevista.

D. Conceptos y definiciones básicas

Es necesario suministrar al encuestador, en el manual, los conceptos y definiciones básicas de las variables y temas que son tratados en la encuesta, por ejemplo, unidades de muestreo, vivienda particular, hogar, residente habitual, miembros del hogar, jefe de hogar, etc.

E. Instrucciones para el diligenciamiento del cuestionario

En este punto se debe señalar de manera muy específica las instrucciones que debe seguir el encuestador para resolver el cuestionario durante la entrevista. En primer lugar, es necesario indicar de manera general las secciones o módulos que contiene el cuestionario, su contenido, objetivo y los miembros del hogar que deben responderla; asimismo, las instrucciones generales para el llenado del cuestionario (según las características propias de

cada cuestionario, por ejemplo, códigos, pases, etc.), como escribir las respuestas, instrucciones para la carátula del cuestionario, etc.

En segundo lugar, es necesario describir de manera específica las instrucciones para el llenado de cada una de las secciones o módulos del cuestionario, sus objetivos, los informantes, y las instrucciones para el llenado pregunta por pregunta, definiendo cada una de las variables.

4. Manual del supervisor

Este manual tiene como propósito proveer al supervisor las instrucciones y procedimientos necesarios para el buen desarrollo de la encuesta, y además constituye el documento de consulta para la solución correcta de los problemas que se presenten en la operación de campo.

El supervisor es el funcionario responsable de controlar y supervisar el trabajo de un número determinado de encuestadores, así como de apoyarlos en el cumplimiento de sus funciones a fin de lograr los objetivos de la encuesta, de acuerdo a los procedimientos detallados en este manual.

A. Responsabilidades

El manual debe detallar de manera explícita las responsabilidades del supervisor, entre las que se encuentran: la actualización cartográfica, para lo cual es necesario describir las pautas o método a seguir; preparar y distribuir los documentos y materiales de los encuestadores; coordinar con las autoridades locales de ser necesario; preparar el plan de trabajo para la operación de campo (ubicación de los segmentos, revisión de planos, asignar cargas de trabajo); establecer reuniones periódicas de trabajo y coordinación con los encuestadores para evaluar el avance del trabajo y buscar soluciones a posibles problemas; controlar la muestra a fin de evitar reemplazos; recepcionar los documentos y materiales de los encuestadores; ejecutar la entrevista a personas notables de los centros poblados para resolver el cuestionario comunitario; resolver el cuestionario de precios; elaborar un informe final de trabajo de campo de su equipo de trabajo.

B. Pre-crítica de los cuestionarios diligenciados

El supervisor también deberá proceder a realizar una pre-crítica de la información recopilada, para ello el manual debe estipular las instrucciones pertinentes para cada una de las secciones o módulos que contiene el cuestionario de hogares. Por ejemplo, preguntas que siempre deben tener respuestas, relación entre preguntas afines, preguntas filtro, unidad de tiempo, etc.

De existir inconsistencias el supervisor deberá llamar al encuestador para que corrija en el campo dicha información.

5. Muestra

El objetivo de este manual es describir el tamaño y el método de selección de la muestra de la encuesta de hogares. En general, se debería explicitar la población y la cobertura del marco muestral; la procedencia de la información básica del marco muestral y sus principales características, así como los métodos de actualización; la forma, uso y disponibilidad de la cartografía; la definición de la muestra (probabilística, multietápica, etc.); las unidades muestrales (UPM, etc.); la determinación del tamaño muestral; la distribución de la muestra (por áreas geográficas- urbana/rural- dominios, conglomerados, etc.); los factores de expansión.

De igual modo, es importante documentar los hechos observados durante el uso de la muestra, por ejemplo, se debe incluir información sobre la no-respuesta prevista y la no-respuesta efectivamente lograda, el porcentaje de rechazo respecto a la no respuesta y a la muestra, las razones explicativas y las sugerencias de como solucionar el problema, la existencia de algunos patrones para lo no respuesta, información sobre cuales fueron los hogares reemplazados si se da el caso. Asimismo, información sobre las diferencias entre las probabilidades de selección proporcionales al tamaño usado para extraer la muestra y las probabilidades del tamaño descubierto luego de relistar las viviendas.

6. Metodología de la operación de campo

Este documento debe contener de manera detallada los métodos y procedimientos que se deben seguir para la ejecución de la operación de campo. El informe debe mostrar todas las etapas que implica realizar el levantamiento de la información en el campo, desde su organización, el aspecto logístico, la capacitación de los encuestadores, supervisores, codificadores y digitadores; asimismo, la prueba piloto para probar el cuestionario y la prueba piloto para la práctica de campo y la selección del personal encuestador.

Además, el informe debe documentar todos los procedimientos para las acciones complementarias a realizar durante el trabajo de campo, como son la identificación de las viviendas seleccionadas, la recuperación de la información, la observación de las entrevistas y la realización de reentrevistas a cargo de los supervisores, los modos de evitar los reemplazos, la consistencia de la información en el campo, etc.

7. Manual del codificador

El objetivo del manual es contener las definiciones y procedimientos que deberían seguirse en la ejecución de la codificación y conversión de la información contenida en los cuestionarios de las encuestas de hogares.

En la elaboración de este manual se debe tener en cuenta si la digitación se realizará en lugares fijos (oficina central, oficinas regionales), o de manera integrada con el trabajo de campo utilizando computadores portátiles e impresoras.

Si la digitación se realizará en algún o algunos lugares fijos, entonces se deberá organizar un equipo de codificación y conversión que realice esta labor, previa a la digitación. Si la entrada de datos se hará de modo integrado al trabajo de campo, entonces los responsables de la codificación y conversión serán los encuestadores y el supervisor, en cada equipo de trabajo. En ambos casos, para simplificar la tarea de codificación es importante que el cuestionario esté precodificado.

En cualquiera de las formas que se realice la digitación es muy importante que el manual incluya las siguientes partes:

A. Definiciones básicas

Revisión: Consiste en verificar que las respuestas anotadas estén en forma legible, detectar toda aquella información innecesaria y ubicar en el lugar correspondiente la información registrada en cada pregunta. En general, la revisión consiste en examinar las anotaciones hechas en cada pregunta para asegurar la correcta entrada de datos a la computadora.

Codificación: Consiste en la asignación de claves numéricas o códigos a la información registrada y revisada de los cuestionarios (por ejemplo, códigos CIU, CUIO, etc.), lo que facilitará el posterior procesamiento electrónico de la información.

Conversión: Es la tarea que consiste en homogeneizar diferentes unidades de medida, tales como cantidades anotadas en fracciones a enteros y decimales, información obtenida en dólares a su equivalencia en moneda nacional, o también el caso de diferentes unidades de medida para productos alimenticios a una sola unidad.

Las tres definiciones señaladas, es decir la revisión, codificación y conversión se efectuarán en una sola operación, pregunta por pregunta en cada cuestionario diligenciado.

B. Funciones y responsabilidades del codificador

Es necesario tener muy bien definidas las responsabilidades de las persona que harán las labores de codificador, a fin de que este proceso sea muy estricto. Entre estas responsabilidades podrían estar, por ejemplo, tener un perfecto conocimiento de las normas e instrucciones contenidas en el manual, tener estrecha coordinación con el jefe de equipo, ceñirse estrictamente a las reglas técnicas establecidas, mantener absoluta reserva sobre la información contenida en los cuestionarios, etc.

C. Funciones y responsabilidades del responsable de la codificación-conversión

Entre ellas, tener perfecto conocimiento de las normas e instrucciones contenidas en el manual a fin de resolver las consultas del personal a su cargo, recepcionar los documentos a codificar, administrar la provisión del material y equipo necesarios para el desarrollo de la tarea, llevar el control de avance diario de cuestionarios codificados, informar

oportunamente cualquier situación no prevista que pueda presentarse durante el trabajo, coordinar con la unidad de digitación de la información, etc.

D. Instrucciones generales y específicas para la ejecución de la codificación-conversión

Esta sección debe contemplar todas las instrucciones necesarias para que los codificadores lleven a cabo su trabajo de una manera estricta y completa. Pueden haber instrucciones generales, como el tipo de bolígrafo a utilizar y el color de la tinta, así como instrucciones específicas relacionadas a secciones particulares del cuestionario, por ejemplo, respecto a la ubicación muestral en la carátula, o al uso de los códigos de ocupaciones, de actividades económicas o de cultivos, etc. Las instrucciones de codificación se deben elaborar tomando en consideración las características particulares de los cuestionarios utilizados.

8. Manual de entrada de datos

Documento destinado a otorgar las instrucciones necesarias para la digitación de la información registrada en los cuestionarios. Este documento debe contener el diccionario de variables o códigos utilizados para cada una de las preguntas del cuestionario, la descripción del programa o software a utilizar; y las instrucciones generales y específicas para que los digitadores puedan procesar de manera eficiente la información, tomando en cuenta para ello si la digitación se realizará en lugares fijos (oficina central, oficinas regionales) o estará integrada con el trabajo de campo.

También debería incluir los criterios adoptados para la consistencia mecanizada de los datos, el tipo de software elegido y el modo de uso. Las normas deberían indicar las responsabilidades específicas y las acciones que debería tomar el equipo encargado de la entrada de datos, el jefe de proyecto y los analistas responsables de la calidad de información, cuando se detecten inconsistencias en los datos o la información está incompleta. Nuevamente, el tipo de normas va depender si la digitación se realiza en lugares fijos o de manera integrada al trabajo de campo. Por ejemplo, si la entrada de datos está integrada al trabajo de campo y se detectan inconsistencias, entonces una recomendación será realizar una reentrevista al hogar. Esta es una diferencia muy importante con el tipo de digitación y chequeo primario de los datos en lugares fijos, que no permite, en la mayoría de los casos, volver a reentrevistar al hogar.

9. Manual de códigos utilizados

Es imprescindible adjuntar toda la información relacionada a los códigos utilizados para la encuesta. Entre los principales están:

- Código de ocupaciones
- Código de áreas geográficas
- Clasificador de actividades económicas

- Códigos de productos (agrícolas, pecuarios, etc.)
- Códigos para unidades de medida (peso, superficie, capacidad)

10. Información metodológica básica para usuarios

El propósito de este documento es mostrar de una manera general los principales lineamientos metodológicos utilizados en el planeamiento, diseño, implementación, control de calidad, procesamiento, análisis y difusión de resultados, de las encuestas de hogares realizadas.

En general, este documento es un resumen de los demás informes metodológicos que serán preparados y que de manera independiente se adjuntarán a las bases de datos al momento de su distribución. Sin embargo, su utilidad práctica radica en que de una manera rápida y resumida puede documentar a los usuarios sobre la metodología utilizada. Además, este tipo de informe se presta de manera ideal para adjuntarlo a boletines informativos sobre la encuesta de hogares, también para distribuirlo conjuntamente con los tabulados básicos e informe descriptivo general de resultados, y con el documento sobre el perfil de la pobreza.

Contenido:

1. Introducción (*remarcar si hay cambios respecto a encuestas anteriores o la medida en la cual es comparable con las anteriores*). Sería recomendable incluir un cuadro con la programación de las encuestas integrantes del SIEH.

2. Cuestionario de la encuesta

- 2.1 Cobertura del cuestionario al hogar (*módulos, variables, periodo de referencia, informante*)
- 2.2 Informantes del cuestionario al hogar y los miembros del hogar
- 2.3 Funcionalidad del cuestionario al hogar (*ventajas y desventajas del formato, tipo de letra, forma de organización de los módulos, etc.*)
- 2.4 Cuestionario de la comunidad
- 2.5 Cuestionario de precios
- 2.6 Diferencias con cuestionarios de las encuestas anteriores
- 2.7 Flujo del cuestionario (*indicaciones sobre rutas, saltos, simbología, etc.*)

3. Muestra de viviendas

- 3.1 Tamaño y cobertura de la muestra
- 3.2 Diseño y selección de la muestra
- 3.3 Muestra implementada (*no respuesta prevista, no respuesta lograda, causas, etc.*)
- 3.4 Muestra para el cuestionario comunitario
- 3.5 Comparabilidad con el diseño de la muestra de encuestas anteriores e implicancias

4. Organización y ejecución de la encuesta

- 4.1 Dirección
- 4.2 Cronograma de todas las etapas de la encuesta
- 4.3 Capacitación y selección de personal
 - 4.3.1 Curso para selección de personal
 - 4.3.2 Curso para supervisores, jefes regionales y otros
 - 4.3.3 Curso para operadores de microcomputadores
- 4.4 Encuesta piloto
- 4.5 Comunicación pública (si aplica)
- 4.6 Operación de campo
 - 4.6.1 Organización
 - 4.6.2 Aspectos administrativos (*manejo de recursos, logística, etc.*)
 - 4.6.3 Supervisión
 - 4.6.4 Ejecución (*visitas, duración de la entrevista, modos utilizados, contingencias, rendimientos individuales y por equipos, etc.*)
 - 4.6.5 Cuestionario comunitario
 - 4.6.6 Labores complementarias (*recuperación de información, observación a las entrevistas, revisión de cuestionarios diligenciados, reentrevistas*)
- 4.7 Transporte
- 4.8 Consistencia de la información
 - 4.8.1 Crítica-codificación
 - 4.8.2 Consistencia mecanizada

5. Calidad y comparabilidad de la información

- 5.1 Factores limitantes
 - 5.2 Errores muestrales de todas las variables principales
 - 5.3 Calidad comparada con otras fuentes y encuestas anteriores
 - 5.4 Periodo de referencia de los datos recolectados y estacionalidad de la información (*por ejemplo, ingresos, año escolar, etc.*)
- 6. Información del consumo agregado
 - 7. Información del ingreso agregado
 - 8. Deflatación de valores monetarios y precios relativos entre regiones
 - 9. Archivos de la base de datos

10. Canastas básicas de consumo
11. Líneas de pobreza

ANEXOS

- A. Modo de obtener la base de datos de la encuesta de hogares
- B. Lista de instituciones de contraparte y contactos
- C. Lista de documentos de soporte (documentos que deben adjuntarse durante la entrega de la base de datos).
- D. Lista de publicaciones en base a la encuesta de hogares, medios y formas de difusión
- E. Distribución de la muestra urbana y rural por dominios, áreas geográficas, segmentos, viviendas, etc.
- F. Índices de precios al consumidor (nacional, capital y/o grandes ciudades), del periodo de referencia de la encuesta
- G. Códigos, a nivel de 1 dígito, de actividades económicas y ocupaciones.

11. Variables agregadas y errores muestrales

La construcción o elaboración de variables agregadas que resulten de la investigación, tales como el ingreso, gasto, años de escolaridad, etc., debería ser explícitamente definida en este manual. Estas variables son elaboradas por las oficinas de estadística a partir de la información original de la encuesta, y pueden formar parte de la base de datos de la encuesta como información adicional. Por su parte, los investigadores y usuarios en general pueden desear construir dichas variables de acuerdo a sus propios criterios metodológicos usando la información original, o desear usar la definición estructurada por la oficina de estadística, para lo cual es necesario conocer los criterios utilizados para su elaboración.

Es recomendable detallar las preguntas y las secciones involucradas en la definición de las variables. De igual manera, es importante describir los supuestos utilizados o las imputaciones realizadas, el método de corrección de la inflación de las variables monetarias, el periodo base utilizado, el modo de cálculo de los valores per cápita, la definición de los grupos de consumo (alimentos, alquileres, etc.), la definición de perceptor, etc.

Además, este documento también tiene el objetivo de mostrar las bondades y limitaciones de la información de la encuesta, mediante la estimación de los errores muestrales de todas las variables principales, entre ellas, empleo, ingreso, gasto, y otras que pueden ser de gran uso para el análisis de los datos.

El manual, a efectos de agilizar su uso, debería incluir ejemplos sobre casos concretos de como interpretar los datos cuando la información se presenta en manera de

rangos, su significado en términos de las potencialidades y las limitaciones de los datos. También se podría incluir aspectos conceptuales y sus aplicaciones interpretativas con los resultados de la investigación, información sobre el tipo de software utilizado, etc.

12. Informe ex-post de la operación de campo

Este informe es muy valioso para conocer los acontecimientos sucedidos durante la operación de campo, y además, para tener un mejor conocimiento y entendimiento de la calidad de la información recolectada. Además de los errores muestrales, la calidad de la información de las encuestas de hogares es afectada por diversas causas, tales como rechazos, cansancio del informante u olvido de fechas o gastos, equivocaciones de los encuestadores, etc. Estos factores generan lo que se conoce como errores no muestrales. Para reducirlos al mínimo es necesario realizar una buena planificación de todas las etapas de la encuesta, y una estricta supervisión.

El documento se debe elaborar sobre la base de los informes de campo que deben realizar los encuestadores y supervisores, y debe contener las experiencias y acontecimientos ocurridos durante el trabajo de campo (contingencias, rechazos, las actitudes de los entrevistados frente a la encuesta, duración de la entrevista, etc.). De igual modo, este documento debe resumir las sugerencias aparecidas en el campo sobre la funcionalidad de los cuestionarios, los manuales, la muestra, la cartografía, etc.

13. Manual de uso de la base de datos y códigos de las variables

Este documento debería incluir todas las instrucciones necesarias para el uso adecuado de las bases de datos de las encuestas de hogares. Por ejemplo, debe indicar como están organizadas las bases de datos, los códigos de los archivos correspondientes, y el lenguaje en el cual están organizados. De igual manera, la identificación de los archivos de todas las variables agregadas construidas por la oficina de estadística.

Asimismo, la codificación que ha sido utilizada para identificar a cada una de las variables contenidas en las secciones y preguntas de los cuestionarios utilizados. Se recomienda que la codificación sea la más sencilla posible y esté claramente señalada para agilizar el procesamiento de la base de datos.

Una forma adicional de presentar los códigos de las variables es mediante una serie de tablas de registro, para cada sección o módulo, que tengan las siguientes columnas: código de identificación (segmento, vivienda, hogar); sección y pregunta del cuestionario; código de las variables; tamaño de muestra; promedio; valor mínimo; valor máximo.

14. Perfil de condiciones de vida

Este es uno de los productos o resultados de la investigación que será realizado durante la etapa de análisis de la encuesta. Este documento será preparado por la oficina de estadística con el soporte técnico del Programa MECOVI.

El informe contendrá una cuantificación de la pobreza y una caracterización general de las condiciones de vida de la población bajo estudio. Su distribución conjuntamente con los demás documentos de soporte servirá de marco de referencia a otros investigadores que hagan uso de la información.

15. Información sobre principales resultados

Este documento debería enfatizar en un esquema de presentación ágil y amigable para los usuarios. Su diseño debe considerar el uso de manera complementaria de texto, gráficos y tabulados o cuadros estadísticos.

De manera similar que el documento sobre el perfil de la pobreza, la distribución de los resultados principales de la información de la encuesta servirán como marco de referencia a los diversos usuarios de la bases de datos.

Una breve descripción de los principales hallazgos de la investigación, referidos a un determinado tema, informa y ayuda al usuario a interpretar mejor los cuadros presentados. De igual modo, la presentación de gráficos sencillos y claros resaltando los principales resultados ayudan a fijar las ideas sobre las conclusiones del estudio.

El contenido del informe debería estar conformado solamente por los aspectos más sobresalientes de la encuesta. La priorización de estos resultados debe ser materia de un trabajo en equipo de los técnicos de la oficina de estadística, y contar, además, con el soporte de un consultor y de los co-patrocinadores del Programa MECOVI. La información presentada debería incluir los márgenes de error de las variables para dimensionar su uso.

La forma del documento debería contemplar un diseño gráfico ágil. Algunos aspectos que se podrían considerar para lograrlo son los siguientes:

- Carátula con diseño adecuado e impresa en cartulina de un gramaje que permita su uso sin dificultad.
- El tipo y tamaño de letra debe permitir una fácil lectura.

La impresión se debería realizar en papel blanco alisado, por ejem. de 80 gramos, y si los costos lo permiten, sería recomendable que la impresión se realice a dos colores.

Esto ayudaría a realzar los gráficos, títulos, subtítulos, etc.

- Los diferentes tipos de letra y accesorios ayudan a graficar mejor la presentación de los resultados, por ejemplo, letra itálica, subrayados, letra en "negrita", se pueden usar para diferenciar o resaltar algunos aspectos de la información, tales como subtítulos, totales y subtítulos en los cuadros, etc.
- El tamaño del formato debe permitir un buen diseño gráfico, por Ejem. 20.5 por 28 cm.

- El índice del documento debe estar estructurado de acuerdo a una secuencia lógica de los temas presentados. Este índice debe incluir las áreas temáticas, gráficos, cuadros estadísticos, metodología, etc.
- El diseño de los cuadros estadísticos debe ser muy cuidadoso. El título del cuadro es clave para entender su contenido, por lo tanto debe ser escrito considerando todas las pautas para ello. Las tablas deben tener todas las notas y llamadas a pie de cuadro que sean necesarias para clarificar su contenido e interpretación.
- Los valores numéricos en los cuadros deben expresarse en valores redondeados a miles o millones, si aplican, a fin de que éste sea más claro y entendible. Las cifras en decimales, en lo posible, redondear a un sólo decimal. De igual manera, incluir los rangos de confianza para las principales variables.
- Capacitar a un funcionario de la oficina en el manejo de programas editores que existen en el mercado.

16. Publicación de divulgación popular

El objetivo de esta publicación es divulgar los resultados de la encuesta a través de un documento elaborado en términos muy simples y directos, con un énfasis netamente comunicador a la población no especializada en leer e interpretar estadísticas.

El diseño de esta publicación debe combinar el uso de un texto muy breve que resuma las ideas o mensajes principales, los datos más importantes de los resultados de la investigación y, especialmente, dibujos muy gráficos que expresen el mensaje o el contexto de la información. Al respecto, se recomienda revisar el documento "Divulgación Popular de la Encuesta de Hogares de 1995", publicado por la Dirección General de Estadística, Encuestas y Censos del Paraguay (DGEEC).

**ASPECTOS DEL DISEÑO Y LA PUESTA EN MARCHA DE LAS
ENCUESTAS QUE INCIDEN EN LA CALIDAD
DE LOS DATOS RECOGIDOS***

EDUMUNDO BERUMEN

JUAN MUÑOZ

* Esta nota es parte de las conclusiones de la Reunión de Iniciación del Programa MECOVI, celebrada en Asunción, Paraguay, el 3 y 4 de diciembre de 1996.

ÍNDICE

	<u>Página</u>
Actualización del marco muestral	273
Listado de hogares	273
Documentación de la muestra	273
Control de la no respuesta	274
Entrada y validación de datos	274
Precisión de los resultados publicados	274
Formación de los equipos técnicos	274
Formación de los equipos de terreno	275
Supervisión	275
Vehículos de difusión	275
Normas de buena práctica en el uso de los datos	275
Períodos de referencia	275

ASPECTOS DEL DISEÑO Y LA PUESTA EN MARCHA DE LAS ENCUESTAS QUE INCIDEN EN LA CALIDAD DE LOS DATOS RECOGIDOS

La siguiente es una lista preliminar de los tópicos que se mencionaron durante la reunión como relevantes en relación con la calidad de las informaciones recogidas en las encuestas de nivel de vida. Aunque no es exhaustiva ni se encuentra organizada por temas o prioridad, se espera que sirva como un primer ayuda-memoria de los aspectos tratados. Agradeceremos los comentarios destinados a completarla.

Actualización del marco muestral

Las ONEs deberían contemplar la actualización del marco muestral (particularmente en las áreas periurbanas y otras de alto crecimiento esperado) como una actividad prioritaria, permanente y básica de un programa integrado de encuestas de hogares.

Listado de hogares

La operación de listado de hogares es fundamental y debe tener un lugar explícito en el calendario y en el presupuesto de las encuestas. La operación debe completarse y supervisarse antes de la encuesta propiamente tal, con el objeto de asegurar su exhaustividad y precisión. Los procedimientos de campo de la encuesta deben considerar el rescate de viviendas omitidas por la operación de listado.

Documentación de la muestra

La documentación de la muestra debe contener al menos los siguientes elementos:

Definición explícita de la población objetivo, incluyendo si es preciso la presencia de estratos de inclusión o exclusión forzada.

Definición explícita de todas las etapas del proceso de selección de las Unidades Últimas de Área. Esta definición debe materializarse en la constitución de una lista de todas las unidades en un archivo informático (por ejemplo, en una hoja de cálculo), con las siguientes variables asociadas a cada UUA:

Identificación y pertenencia a estratos y dominios de análisis relevantes (urbano/rural, regiones, etc.)

Tamaño esperado (según el marco muestral):

- Tamaño efectivamente observado durante la operación de listado.
- Tasa de no respuesta esperada.

- Tasas de no respuesta efectiva.
- Fechas programadas de levantamiento de datos.
- Fechas efectivas de levantamiento de datos.
- Equipo responsable del listado.
- Equipo responsable del levantamiento de datos.

El archivo de documentación debería irse completando a medida que se realizan las etapas relevantes de la encuesta (no al final del proceso) y estar disponible en todos los momentos como testigo de su ejecución.

Control de la no respuesta

La sustitución no resuelve el problema de la no respuesta y puede entrañar riesgos de difícil prevención. Se deberían supervisar en terreno la totalidad de los casos registrados de no respuesta y monitorear la evolución de esa tasa, por equipo y por encuestador, a lo largo de todo el período de recolección de datos. Deberían además hacerse esfuerzos por estimar el sesgo producido por la no respuesta.

Entrada y validación de datos

Debe completarse la integración de procedimientos informáticos de control de calidad a las operaciones de campo de las encuestas. La instrumentación efectiva de esta práctica tiene implicaciones organizacionales, logísticas y de recursos humanos, que deben tenerse en cuenta al presupuestar la encuesta y programar las actividades de diseño y puesta en marcha. La incorporación de la informática al terreno no presume el abandono del cuestionario de papel.

Precisión de los resultados publicados

Todas las publicaciones generadas por la ONE a partir de la encuesta deberían presentar los errores muestrales y los intervalos de confianza para las variables más significativas. La ONE debería además promover la aplicación de esta práctica en las publicaciones generadas por usuarios externos, junto con otras formas de prevenir la publicación de resultados de significación dudosa a partir de los datos de la encuesta (por ejemplo, desagregaciones más finas que las permitidas por el diseño).

Los resultados publicados no deberían contener más cifras significativas que las que razonablemente pueda proveer la muestra (por ejemplo, los porcentajes publicados no deberían tener más de un decimal). Se recomienda establecer esta práctica, subrayando en la publicación las razones en que se fundamenta, con el fin de educar a los usuarios sobre los alcances y limitaciones de las encuestas por muestreo.

Formación de los equipos técnicos

La formación del personal técnico responsable del diseño y la puesta en marcha de las encuestas debería considerarse como una actividad prioritaria. La avidez del mercado laboral por el personal entrenado en el proceso es inevitable y debe considerarse como una situación de hecho, con consecuencias sobre el carácter permanente que debe tener el entrenamiento.

Formación de los equipos de terreno

El correcto entrenamiento del personal de terreno es uno de los elementos claves para asegurar la calidad de los datos. Debería otorgarse a esta actividad la importancia debida, y cautelarse que contenga los componentes necesarios de formación teórica y trabajo práctico efectivo.

Supervisión

Los procedimientos de control de encuestadores deben instrumentarse de manera efectiva durante todo el período de levantamiento de datos y estar apoyados en elementos formales (formularios, criterios de selección de hogares supervisados, etc.) que permitan que la tarea pueda a su vez ser supervisada por el equipo central de la encuesta. El dimensionamiento de los recursos humanos necesarios para la operación de campo debería asegurar que se mantengan tasas de supervisión adecuadas durante todo el trabajo de campo.

Vehículos de difusión

Deberían instrumentarse mecanismos que aseguren el intercambio de experiencias entre los países. Se han mencionado al respecto la edición de un boletín periódico y la animación de una página web.

Normas de buena práctica en el uso de los datos

La política de libre acceso a los datos podría complementarse con el establecimiento de recomendaciones de buena práctica en el uso de los mismos. Las sugerencias podrían referirse, en particular, a las formas de cautelar la confidencialidad de la información individual, las normas para exhibir errores muestrales e intervalos de confianza, las menciones a la fuente de origen de los datos, y las formas de mantener la consistencia de conceptos entre publicaciones.

Períodos de referencia

Debería evitarse la confusión entre tres tipos de intervalos mencionados al definir la cobertura temporal y la periodicidad de una encuesta; a saber (a) el período de referencia de la información, (b) el período de levantamiento de los datos y © los cortes efectuados para la presentación de los resultados.



¿ 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).

CONTROL DE CALIDAD*

KINNON SCOTT
(BANCO MUNDIAL)

* Transparencias utilizadas por la autora durante su presentación de este tema en el Taller.

CONTROL DE CALIDAD

Porque?

- Impacto
 - Descriptivo v/s Prescriptivo
 - Valor
- Uso futuro de datos e información
 - Cultura de demanda
- Credibilidad institucional

Como?

Diseño del Cuestionario

- Preguntas Cerradas
- Preguntas Precodificadas
- Completas
- Flujos o saltos explícitos
- Traducciones

Prueba Piloto

- Probar todo instrumento en todas las situaciones
- Revisar - ajustar
- Reprobar

Capacitación

- Todo el equipo
- Propósito
- Contenido
- Capacitación Específica
- Que sea rigurosa
- Manual detallado

Supervisión

- Razón supervisor/encuestador
- Tareas específicas
- Herramientas necesarias

Entrada de datos en paralelo

- Estilo tradicional
- CAPI o concurrente
- Complejidad
- Mejor uso de recursos de supervisión
- Oportuna

Informante Directo

- Mejor Informado
- Menor cansancio

Tamaño de la muestra

- Errores muestrales
- Errores no muestrales
 - Sesgos
 - Sin remedio

Acceso Libre

- Múltiples revisiones
- Comparación con otra información
- Especialización

Cómo?

- Diseño del cuestionario
- Prueba Piloto
- Capacitación
- Supervisión
- Entrada de datos en paralelo
- Informante Directo
- Tamaño de la muestra
- Acceso libre

Control de Calidad

- Urgente
- Permanente
- Costo - Beneficio
- Desafío



ENCUESTA INTEGRADA DE HOGARES EN PARAGUAY

DIRECCIÓN GENERAL DE ESTADÍSTICA Y CENSOS (DGEC)

ÍNDICE**Página**

INTRODUCCIÓN.....	299
CUESTIONARIOS DE LA EIH	300
ORGANIZACIÓN DEL TRABAJO DE CAMPO	301
PROCESAMIENTO	302
CONTROL DE CALIDAD DE LA ENCUESTA	302

ENCUESTA INTEGRADA DE HOGARES EN PARAGUAY

INTRODUCCIÓN

La Encuesta de Hogares en Paraguay se inicia en el año 1976 con un proyecto experimental de encuestas sobre empleo del Programa de Empleo para América Latina y el Caribe (PREALC). El tamaño de la muestra era de 500 viviendas y el área de cobertura cubría a Asunción, la capital del país.

La instauración de una encuesta de hogares con carácter permanente recién se inicia en el año 1983 fecha a partir de la cual se dispone de una serie ininterrumpida con cobertura ampliada al Area Metropolitana de Asunción (unas 10 ciudades circunvecinas) y un tamaño de muestra de 1000 viviendas.

Hasta el año 1992 la encuesta fue ejecutada con la misma cobertura geográfica, el mismo instrumento de recolección de la información, el mismo marco muestral y tamaño de muestra.

Al año siguiente se inicia el proceso de cambios ya que el marco muestral es actualizado utilizando los datos del Censo Nacional de Población y Viviendas 1992 y ampliado a todo el Departamento Central urbano con un tamaño de 1400 viviendas distribuidas en igual proporción entre ambos estratos. Asimismo el cuestionario incorpora unas pocas nuevas preguntas que amplía la información existente sobre empleo.

El año 1994 marca el inicio de la encuesta urbana a nivel nacional. Con el mismo cuestionario se expande la encuesta a todo el país y se genera información para nuevos estratos (total país, ciudades fronterizas, resto urbano, además de los ya citados).

Como paso inicial hacia una Encuesta Integrada de Hogares (EIH), en 1995 se implementan modificaciones en el diseño y formato del cuestionario, se amplía la información sobre empleo y se agregan nuevas preguntas relacionadas con la vivienda, el confort del hogar, educación y a modo experimental se incorporan preguntas sobre salud. La encuesta toma carácter urbano-rural, llegando la muestra a un tamaño de 4500 hogares.

Atendiendo al plan de implementar una encuesta integrada, 1996 marca el momento de incorporar un conjunto de módulos que hasta esa fecha no habían sido abordados. Esta actividad se realiza ya en el marco del apoyo del Programa de Mejoramiento de las Encuestas y Medición de las Condiciones de Vida en América Latina y el Caribe (MECOVI). Aunque la encuesta es aplicada sólo a las áreas urbanas del país la misma incorpora módulos independientes sobre gastos alimentarios y no alimentarios del hogar, sobre salud y educación, aunque centrado en la temática de empleo.

Se prevé que en un período de 4 años habrá de implementarse dos Encuestas Integradas de Hogares en el marco del programa MECOVI, intercaladas con encuestas con módulos reducidos denominada Encuesta Permanente de Hogares (EPH). Actualmente se está implementado la EIH que dio inicio en agosto de 1997 y se tiene prevista su culminación para julio del corriente año. Asimismo la segunda EIH tendrá lugar en el año 2000.

CUESTIONARIOS DE LA EIH

Para reunir datos consistentes con los objetivos de la Encuesta Integrada de Hogares se utilizan 3 tipos de cuestionarios:

Cuestionario a Hogares: a través del cual se pregunta a los miembros del hogar diversos aspectos relacionados con su bienestar

Cuestionario Comunitario: el cual se formula a los líderes comunales sobre los servicios e infraestructura disponibles en las áreas rurales del país

Cuestionario de Precios: en el cual se pregunta a los oferentes de bienes y servicios sobre los precios de venta al consumidor.

Cuestionario a Hogares: La Encuesta Integrada de Hogares 1997/8 recoge información sobre las características y condiciones de vida de la población. Está conformado por un conjunto de módulos íntimamente vinculados que permiten tener una visión a nivel individual y de hogares.

Debido a que el objetivo principal de la EIH 97/8 es la medición del bienestar, el cuestionario fue diseñado incorporando preguntas detalladas sobre gastos en efectivo, valor del autoconsumo, la forma de obtención de los diversos productos, valor de la vivienda y propiedad de bienes duraderos.

Atendiendo al objetivo de caracterizar a la población según los niveles de pobreza, la EIH recoge una amplia variedad de información sobre características de la vivienda, educación de las personas de 5 y más años de edad, salud de todos los miembros del hogar, fertilidad, antropometría, empleo e ingresos laborales, los cuales permiten conocer, entre otros, los retornos e impactos de la educación, la nutrición de los niños, la prevalencia de vacunas, la incidencia de la diarrea en los niños, uso de los servicios de salud, empleo principal y secundario y sus características, las fuentes de ingresos en efectivo y en especie.

En el ámbito de los hogares se recaba información de empresas agropecuarias y no agropecuarias para producir una estimativa de los ingresos netos de las unidades productivas que funcionan en el hogar. El cuestionario consta de 9 secciones con sus respectivas partes, abarcando aproximadamente 600 preguntas.

Cuestionario Comunitario: aplicado en esta edición sólo a las comunidades rurales, a un total de 261 localidades del interior del país que conforman la muestra. Consta de 6 partes con un total de 64 preguntas que investigan sobre demografía, actividades económicas y migraciones laborales, infraestructura y servicios comunitarios, educación, salud, organización comunal y actividad agropecuaria de la localidad visitada.

Las preguntas van dirigidas a los líderes comunitarios, personas representativas o autoridades del lugar y es aplicada por el supervisor mediante contactos y coordinación de horarios con los entrevistados. El número de personas entrevistadas no supera a 3 ni tampoco es aplicada a una sola persona.

Cuestionario de Precios: Se obtiene información de los precios de los principales productos alimenticios y otros productos básicos del consumo familiar, tanto en áreas urbanas como en áreas rurales del país conociendo de esta forma su nivel, su variación relativa en el tiempo y por regiones. Al momento de la visita al segmento de trabajo, los supervisores realizan tres observaciones, tomando información en primer lugar en las despensas/almacenes, luego en los supermercados y por último en los mercados y ferias. El cuestionario toma precios de 9 rubros principales de productos que forman parte de las informaciones captadas en la sección de gastos del cuestionario a los hogares. Se recoge información sobre 78 productos referidos a pan; harina, pastas y arroz; carnes y huevos; embutidos; aceites y grasas; lácteos; frutas frescas y productos varios.

ORGANIZACIÓN DEL TRABAJO DE CAMPO

La Encuesta Integrada de Hogares desarrollada actualmente en el país presenta innovaciones importantes respecto a las versiones anteriores de la encuesta de hogares. Un aspecto sustantivo de los cambios es la incorporación de nuevos contenidos temáticos, seccionados en diversos módulos y la adición de los cuestionarios comunitarios para localidades rurales y de precios para las áreas urbanas y rurales de todo el país.

Una innovación trascendente en la encuesta es que la misma se realiza durante todo el año en lugar de sólo tres meses como se venía desarrollando en lo que hoy se denomina la Encuesta Permanente de Hogares (EPH).

El trabajo de campo y el ingreso de los datos se encuentran descentralizados y su supervisión es estricta. Cada equipo está encabezado por un supervisor y compuesto por dos encuestadores, un operador de ingreso de datos y un conductor.

En cada segmento de trabajo cada equipo permanece aproximadamente una semana visitando en promedio 10 viviendas. Cada hogar es visitado 2 o más veces y en cada visita se realizan entrevistas con los diferentes miembros del hogar, puesto que cada adulto debe responder por sí mismo, lo cual permite minimizar errores que pueden cometer informantes sustitutos.

La estructura organizacional de la encuesta está conformada por un jefe de encuestas, dos coordinadores de campo, un encargado de procesamiento, un equipo técnico compuesto por cuatro integrantes y el personal de campo. Se cuenta con un total de 10 equipos de trabajo distribuidos según la carga de trabajo estipulada para cada zona creada.

Como ya se mencionó la EIH ha exigido la necesidad de establecer 4 bases de trabajo en el interior del país ubicadas en puntos geográficos estratégicos del país (dos en zonas fronterizas de mayor concentración de población, una en el Área Metropolitana de Asunción y el último la zona central de del país).

A efectos de dar cuenta de la estacionalidad y permitir que se analicen los datos recogidos a partir del primer cuatrimestre generando estimadores insesgados a nivel nacional, la muestra fue distribuida aleatoriamente en cuatro partes iguales atendiendo la carga de trabajo por equipos, por área urbana y rural.

PROCESAMIENTO

De acuerdo a la organización actual del operativo de campo, el tratamiento informático de la encuesta es realizado en el campo, simultáneamente con la recolección de los datos y generalmente en las oficinas locales de la encuesta. Una vez recogido los datos por parte de los encuestadores en su primera visita, el supervisor verifica rápidamente el cuestionario para detectar errores o inconsistencias fácilmente detectables. Evaluado el cuestionario es entregado al digitador quien realiza la codificación de las preguntas abiertas e ingresa los datos al computador portátil. Cuando el trabajo de campo permite que la encuesta se realice sin pernocte, la carga de datos se realiza en las oficinas de las distintas bases, sin necesidad de desplazamiento de los equipos de informática. Sin embargo, una proporción significativa de la muestra se carga trasladando los equipos a las localidades visitadas.

Inmediatamente ingresados los datos, el programa utilizado permite obtener listado de inconsistencias y errores cometidos en el ingreso o en el momento de la recolección de la información. Esto permite corregir durante la segunda visita al hogar los errores o inconsistencias encontrados en la primera parte del cuestionario.

CONTROL DE CALIDAD DE LA ENCUESTA

Tradicionalmente el control de calidad de la encuesta estaba fuertemente relacionado con la conformación de equipos de codificadores-verificadores que revisaban minuciosamente cada cuestionario una vez acumulado una cantidad significativa que justifique el funcionamiento de esta área. Consistía en un largo proceso de revisión de cuestionario sobre la base de un listado de inconsistencias, que era evaluado según el juicio del crítico para dar como correcta o incorrecta la información, ya que el personal de campo,

supervisores y entrevistadores, era contratado sólo durante el operativo de campo y en fechas en que la información era revisada ya no se encontraban como personal de planta. Ocasionalmente se les solicitaba presentarse para realizar aclaraciones respecto al cuestionario diligenciado.

Con el proceso de descentralización que exige la EIH, lo que implica una apertura de los trabajos en 4 frentes, la calidad de la encuesta está sujeta principalmente a un buen programa de ingreso de datos que detecta en terreno todos aquellos errores que pueden ser cometidos en la toma y digitación de los datos mediante reportes que genera. Existen diversos tipos de control de calidad de la información al momento de ser ingresados los datos como ser: control de rango, verificación con respecto a tablas de referencia, control de flujo y controles de consistencia.

Revisión del cuestionario por el Supervisor: Los cuestionarios son diligenciados en el campo por los encuestadores quienes en dos visitas como mínimo obtienen respuestas a las secciones respectivas del cuestionario, entregando posteriormente al supervisor quien realiza un chequeo de las partes completadas, constatando el seguimiento de las recomendaciones establecidas para el correcto llenado del cuestionario. El supervisor analiza mínimamente la coherencia de la información en su conjunto, revisando que la información esté en las líneas correspondientes, los cálculos hayan sido realizados, los saltos atendidos según edades, etc.

Observación de las entrevistas: Fundamentalmente en las etapas iniciales del trabajo de campo se ha puesto mucho énfasis en que los supervisores acompañen a sus encuestadores y observen las entrevistas, en algunos casos las encuestas eran hechas solamente con la presencia del supervisor. Recién en el momento en que el supervisor está convencido de que el encuestador reúne condiciones adecuadas para realizar una entrevista, es autorizado a aplicar sólo un cuestionario. Esta actividad es repetida periódicamente por los supervisores y acompañado por el equipo de supervisión central.

Equipo Central de Supervisión: Además de las tareas propias de aspectos administrativos, logísticos, de verificación de la muestra, etc. el equipo central se ocupa constantemente del control de calidad de la encuesta. En la mayor parte de las visitas realizadas se observa por lo menos una entrevista, se evacuan dudas conceptuales, se uniforman criterios, se transmiten decisiones adoptadas en otras bases de trabajo, se realiza un chequeo de los listados ya revisados por los supervisores de campo y una revisión global de los cuestionarios. Para la evacuación de dudas no es estrictamente necesario la presencia física de los supervisores centrales, ya que es asiduamente utilizada la radio que dispone cada uno de los móviles, o bien la comunicación telefónica. Los errores e inconsistencias que se presentan tienen correspondencia con las posibilidades de seguimiento y apoyo que pueda brindarse desde el equipo central, ya que las encuestas de la base central tienen menos inconvenientes por la posibilidad de evacuación de dudas en forma permanente.

Revisión de cuestionarios en gabinete: Hasta el momento de iniciar los trabajos para la publicación de los resultados correspondientes al primer cuatrimestre de la EIH, el control y chequeo de la información era principalmente realizado por los equipos de campo (supervisor, digitador, encuestador) sobre la base de los reportes generados en la carga de datos. Una vez que se ha reunido en la oficina central la base de datos de las distintas regiones del país se ha procedido a una revisión de los listados de inconsistencias reportados por los distintos equipos para observar la uniformidad de criterios y conceptos utilizados por los distintos supervisores y digitadores. Paralelamente se han preparado frecuencias y tablas que permitieron detectar incoherencias en la distribución o en los cruces de variables realizados. Esta primera actividad del control de calidad de la información ha arrojado la necesidad de devolver cuestionarios a las distintas bases del país para que el personal de campo regrese a los hogares en caso de ser necesario. La devolución de los cuestionarios al personal de campo se realizó acompañada de un personal de la oficina central, quien en taller con los mismos, intenta despejar las dudas e inconsistencias, con la obligación de regresar al terreno en caso de ser necesario. De este chequeo practicado a la base de datos ha surgido un conjunto de consistencias nuevas que se adosan al programa de entrada y consistencia.

Reentrevistas: una proporción relativamente importante de los hogares visitados (10 por ciento) es nuevamente entrevistada por los supervisores mediante unas pocas preguntas de cada sección del cuestionario de hogar. Las preguntas formuladas se refieren a definición de condición de las personas, saltos para pasar a otras secciones o partes o bien una información determinante. Con los datos obtenidos los supervisores controlan los cuestionarios aplicados por los encuestadores, estableciendo un diálogo sobre los datos obtenidos en el hogar en caso de encontrar discrepancias en alguna de las respuestas. Asimismo, a través de las reentrevistas, se calculan los errores de contenido utilizando la variable edad.

Informes periódicos: Independientemente de la comunicación que se establece con cada equipo de trabajo en forma diaria por medio de la comunicación por radio o vía telefónica, los supervisores llenan un formulario en el que dan cuenta todos aquellos aspectos relativos a sus labores en el período, como ser los segmentos visitados, la situación de las entrevistas, el desempeño de los encuestadores, digitadores y choferes, los requerimientos de materiales e insumos que utilizarán y todos aquellos inconvenientes que ameriten su comunicación. De esta forma se está alerta a cualquier problema que requiera la participación del personal de la oficina central.

IMPUTACIÓN DE DATOS*

GUILLERMO RAMÍREZ
(OCEI - VENEZUELA)

* Transparencias utilizadas por el autor durante su presentación de este tema en el Taller.

IMPUTACIÓN DE DATOS

INTRODUCCIÓN

En el análisis de grandes volúmenes de información es muy frecuente encontrar la existencia de datos con valores perdidos.

De hecho, los datos completos constituyen más una excepción que la regla.

Esta situación constituye una severa limitación, puesto que los métodos estadísticos tradicionales están diseñados para ser aplicados sobre conjuntos de datos completos.

Los continuos avances de la informática han hecho posible el surgimiento y puesta en práctica de nuevas metodologías para el tratamiento de información con datos faltantes, los cuales, en su mayoría, producen resultados aceptables cuando hay pocos valores perdidos.

A pesar de la variedad de métodos existentes, el problema permanece abierto, sin que hasta ahora parezca haberse hallado una solución definitiva.

ALTERNATIVAS DE SOLUCIÓN

1.- Descartar los individuos o variables con valores perdidos

Esta modalidad resulta atractiva debido a su simplicidad, ya que es posible analizar los datos sin necesidad de modificar el método estadístico utilizado.

Solamente es recomendable en situaciones con muy pocos valores perdidos, diseminados a su vez sobre muy pocas unidades, puesto que de otra manera la pérdida de información parcial útil puede llegar a ser considerable.

2.- Estimar los valores perdidos utilizando algún método de imputación

Una segunda opción consiste en utilizar la información contenida en los datos observados, para construir estimaciones “razonables” de los valores perdidos. Al ser imputadas tales estimaciones en las observaciones perdidas, se obtiene un juego de datos completo, susceptible de ser analizado con los métodos estadísticos estándar.

EVALUACIÓN DE LOS MÉTODOS DE IMPUTACIÓN

Una de las estrategias usualmente utilizadas requiere una matriz de datos con propiedades conocidas, de la cual se elimina aleatoriamente una cierta proporción de valores asumidos como perdidos, y que son posteriormente estimados e imputados.

Luego se construyen índices, tanto de la bondad de las estimaciones como del grado de fidelidad con que son reconstruidas las propiedades de la matriz original a partir de la matriz imputada.

ALGUNAS TÉCNICAS DE IMPUTACIÓN

Imputación por Medias

Este método, propuesto por primera vez por Wilks (1932), es posiblemente uno de los procedimientos de imputación más antiguo y más sencillo.

Consiste en estimar los valores perdidos de la j -ésima variable mediante la media de sus valores observados, la cual ha sido llamada por Little y Rubin (1987), media de los valores disponibles.

Imputación por Regresión

Este método, propuesto por Buck (1960), supone que las filas de la matriz de datos constituyen una muestra aleatoria de una población normal multivariante. El vector de medias y la matriz de varianzas y covarianzas de los datos completos son utilizados como estimaciones de los parámetros poblacionales, con los cuales se ajustan ecuaciones en regresión para cada una de las variables con datos perdidos, en término de las restantes.

Imputación por Componentes Principales

Este método, sugerido por Dear (1959) y descrito en detalle por Beale y Little (1974), tiene como punto de partida una matriz de datos estandarizada con datos completados mediante el método de Imputación por Medias.

El proceso consiste en:

- Aplicar un ACP sobre la matriz imputada.
- Determinar las distancias entre cada unidad con datos perdidos y las restantes, basándose en sus puntuaciones sobre la primera componente.
- Imputar como valor perdido, el valor observado correspondiente al individuo más cercano al que contiene la información perdida.

Existen métodos basados en procedimientos máximo-verosímiles, y otros más sofisticados basados en la descomposición de una matriz en valores singulares, como es el caso del método de Krzanowsky (1988) y el de Gleason y Staelin (1975).

La mayoría de todos estos métodos parten de una premisa, según la cual los valores perdidos se producen por mecanismos completamente aleatorios.

TÉCNICAS DE IMPUTACIÓN EN LA ENCUESTA DE HOGARES

Imputación Manual

En el proceso de revisión y codificación de la EHM se aplican diversos métodos de imputación en el caso de no respuesta. Por ejemplo, no se permiten blancos en variables claves como edad, sexo y aquellas relacionadas con la situación en el mercado laboral. Por lo tanto, se imputan valores de acuerdo con ciertas especificaciones contenidas en el Manual del Crítico-Codificador.

Ajustes en los Factores de Expansión

En la EHM pueden producirse dos tipos de no respuesta que afectan los factores de expansión:

- Pérdida completa de un segmento
- Pérdida de viviendas en un segmento

En ambos casos se ajustan los FE, multiplicándolos convenientemente de modo que se mantengan las representatividades. Este proceso equivale a duplicar registros con iguales características que los registros perdidos.

Detección de zonas de no respuesta

En Venezuela, y probablemente en otros países, el mayor porcentaje de no respuesta se presenta en las clases altas (alrededor del 30%). Además ocurre que hay mayor no respuesta en las grandes ciudades que en el resto del país. Los encuestadores realizan hasta tres intentos para obtener la información, más una última visita del supervisor. Luego de esta cuarta visita no se insiste en la consecución del dato. Cuando se detectan zonas importantes de no respuesta, se llevan a cabo campañas promocionales que tienen como objetivo disuadir a los hogares a suministrar la información.

MÉTODO DE LAS ASIGNACIONES DINÁMICAS

Los métodos de imputación automática mediante la técnica de asignación dinámica (“Hot Deck”) tradicionalmente utilizados en los Censos de Población y Vivienda, no son actualmente aplicados en la Encuesta de Hogares por Muestreo en Venezuela.

COMENTARIOS FINALES

- A diferencia de un Censo, en el caso de una encuesta por muestreo la modificación de un registro implica modificaciones en todos los registros que él representa.
- Existe una tendencia a no modificar la no respuesta, codificarla como tal y dejarle al investigador la potestad de utilizar o no un método de imputación.



DIFUSIÓN Y EXPLOTACIÓN DE INFORMACIÓN



LOS ERRORES DE MUESTREO EN LAS ENCUESTAS COMPLEJAS

FERNANDO MEDINA
(CEPAL)

ÍNDICE

Página

I. INTRODUCCIÓN	317
II. DETERMINACIÓN DEL TAMAÑO DE MUESTRA ÓPTIMO	317
1. El efecto de diseño y el error estándar del estimador	319
2. El tamaño de muestra efectivo	322
3. Un ejemplo real	325
III. EL MODELO DE REGRESIÓN Y EL DISEÑO DE LA MUESTRA.....	336
IV. LA ESTIMACIÓN DEL EFECTO DE DISEÑO Y EL USO DE LOS PONDERADORES.....	340
V. CONCLUSIONES.....	345
REFERENCIAS BIBLIOGRÁFICAS	347

LOS ERRORES DE MUESTREO EN LAS ENCUESTAS COMPLEJAS

I. INTRODUCCIÓN

La utilización de las encuestas por muestreo como una herramienta útil para generar información ha cobrado gran relevancia en las últimas décadas. De hecho, parecería que la credibilidad de las afirmaciones que se hacen sobre el comportamiento de un determinado fenómeno aumentan en la medida en que al público se le hace saber que la información se generó por medio de una investigación por muestreo la cual se argumenta que es "representativa y confiable al 95%".

De esta manera, a pesar de ser práctica común que diversas organizaciones nacionales e internacionales participen en la planeación y apoyen el desarrollo de encuestas, se considera pertinente evaluar la manera en que éstas se proyectan y analizan, planteando las siguientes preguntas:

- i) **Se diseñan las encuestas en forma adecuada?**
(se determina el tamaño de muestra óptimo, se aplican procedimientos robustos de estratificación, se asigna la muestra de manera adecuada);
- ii) **Los resultados obtenidos se interpretan de manera correcta?**
(la inferencia estadística está basada en el diseño muestral); y
- iii) **Las relaciones de causalidad que se prueban y las recomendaciones de política que se formulan tienen validez estadística?**
(es estadísticamente válida la inferencia basada en modelos de comportamiento).

Tratando de contribuir a dar respuesta a estas preguntas, el objetivo de este trabajo se centra en evaluar los aspectos técnicos que se consideran más importantes y que debieran ser el foco de atención de los estadísticos de encuestas, pero sobre todo para los usuarios de la información en la fase de interpretación y utilización de los resultados para la formulación de hipótesis de comportamiento y el diseño de políticas en el ámbito social y económico.

II. DETERMINACIÓN DEL TAMAÑO DE MUESTRA ÓPTIMO

Para determinar el número de unidades que formarán parte de la muestra las expresiones propuestas suponen que la unidad ejecutora de la encuesta cuenta con antecedentes acerca de la(s) varianza(s) de la(s) variable(s) que se desean estudiar¹. Así, cuando interesa estimar una proporción binomial bajo un esquema de selección aleatorio simple (MAS) Cochran (1953), suponiendo normalidad en la distribución del parámetro,

¹ Esta forma de proceder llevaría a suponer que la encuesta sólo desea obtener estimaciones de una variable, sin considerar que prácticamente todas las encuestas son de propósitos múltiples.

propuso una expresión para calcular el tamaño de muestra ($n = t^2 pq/d^2$) que se utiliza con mucha frecuencia y también con alguna ligereza². Así, en el caso de que no se tenga ninguna información acerca de la varianza del parámetro de interés -lo cual sucede frecuentemente- la práctica aconseja asumir un valor de $p=.5$ ("worse case") el cual maximiza la variabilidad del estimador ($pq=.25$) y genera un tamaño de muestra que en apariencia garantiza la precisión deseada³.

Como dan cuenta diversas investigaciones (Tortota, 1978; Angers, 1979; Thompson, 1987 y CEPAL, 1998) esta manera de proceder no es del todo apropiada, ya que en la realidad es habitual que la variable de interés tenga una distribución multinomial. En este sentido, además de que el procedimiento sugerido por Cochran puede subestimar en forma importante el tamaño de la muestra, tampoco permite fijar en forma simultánea un coeficiente de confianza para todas las categorías en que se distribuye la variable, lo cual inhibe al investigador sobre la posibilidad de controlar la precisión deseada de las estimaciones.

Es evidente que en la práctica cotidiana las encuestas que se realizan son de propósitos múltiples. De esta forma, para determinar el tamaño de muestra óptimo se deben considerar todas las variables de interés, de tal suerte que el número de observaciones propuestas garantice que se minimice la varianza de los estimadores en forma simultánea, o se pueda conocer el nivel de la variabilidad de las estimaciones para un monto de presupuesto asignado. Para resolver este problema, en la literatura se han propuesto algunos algoritmos de optimización no lineal que permiten calcular el tamaño de muestra óptimo (Kokan, 1963, Kokan y Khan, 1967 y CEPAL, *op. cit.* para ejemplificar su uso).

Por otra parte, cuando la selección de las unidades de observación y análisis se realiza en varias etapas, es necesario aplicar técnicas de estratificación y conglomeración adecuadas que permitan identificar a las unidades de estudio, lo cual incrementa la varianza del estimador, así como el número de selecciones que se deben realizar para garantizar la eficiencia del diseño en términos de varianza mínima⁴.

En este sentido, decidir sobre el total de observaciones a seleccionar se debe apoyar en criterios objetivos que fijen prioridades sobre las variables de interés, la precisión y confiabilidad deseada, los dominios analíticos que se formarán, el método para la selección de la muestra, los indicadores derivados que se desean calcular, así como los

² En los diseños poliéticos y multipropósitos, el tamaño de muestra obtenido se multiplica por un factor que se analizará más adelante y que se denomina comúnmente como efecto de diseño (efd).

³ Para los casos en que se desean obtener estimadores de medias y totales es necesario conocer la variabilidad asociada a la variable de interés, por lo que proceder como si se tratara de estimar una proporción es incorrecto. Así, si se decide estimar el tamaño de muestra utilizando la expresión para proporciones, se debe estar consciente que el error de muestreo se incrementará sin que exista la posibilidad de evaluar su repercusión en la varianza de los estimadores.

⁴ Es importante recordar que la muestra no depende del tamaño de la población, a excepción de su efecto en el factor de corrección por finitud (CPF) en el caso de que este se considere. Además de que la muestra necesaria, es inversamente proporcional al cuadrado del error estándar, por lo que para lograr una reducción de k veces su nivel, se requiere incrementar la muestra en un factor k^2 .

costos que involucra la identificación de las unidades de marco. Así, para encuestas de propósitos múltiples se sugiere utilizar algoritmos que permitan encontrar el tamaño óptimo de la muestra por estrato que minimice la variabilidad de los estimadores en forma simultánea (Kokan y Khan, *op. cit.*).

Es muy usual que no se disponga de información sobre la varianza de las variables de estudio, ya que la mayoría de las Oficinas Nacionales de Estadística no la calculan ni publican de forma rutinaria. Asimismo, tampoco es habitual que se realicen estudios piloto a gran escala que generen información confiable que pueda ser utilizada como insumo para determinar el tamaño de muestra óptimo. De esta manera, se deben buscar alternativas de solución que permitan planear encuestas con información limitada.

1. El Efecto de Diseño y el Error Estándar del Estimador

Para tratar de resolver la situación que se presenta en el cálculo del tamaño de muestra en encuestas complejas, en donde no se cuenta con información sobre la varianza de la variable de interés, Kish(1979 pp. 302-309) propuso una manera de solucionar el problema definiendo un factor de ajuste que a partir de una muestra aleatoria simple MAS permite aproximarse al número de selecciones necesarias para que un diseño de conglomerados proporcione la misma varianza. El factor se conoce como el efecto de diseño (efd_k) y se define por medio de :

$$efd_k = \text{Var}(\bar{y}) / (1-f) S^2/n \quad (1)$$

en donde $\text{Var}(\bar{y})$ representa la varianza de la variable de interés en el diseño de conglomerados y $(1-f) S^2/n$ la del esquema de selección aleatoria, de tal forma que :

$$\text{Var}(\bar{y}) = (1-f) S^2/n * efd_k \quad (2)$$

Según Kish (*op. cit.*), “este extenso factor apunta a resumir las diversas complejidades en el diseño de una muestra, sobre todo los de conglomeración y estratificación”. Asimismo, este autor señala que una mejor aproximación a la varianza S^2 del estimador se logra por medio de:

$$S^2 = s^2 [1+(efd_k-1)/n] \quad (3)$$

De (3) se observa que s^2 será una buena aproximación a S^2 cuando $efd_k=1$. De esta forma, es común que en los diseños estratificados el $efd_k < 1$ ya que los procedimientos de estratificación tienden a reducir la varianza del estimador debido a la homogeneidad que se logra al interior de los grupos formados. Lo contrario sucede en el caso de diseños de conglomerados en donde generalmente se observa que $efd_k > 1$, lo cual estaría indicando que el efecto de conglomeración tiende a incrementar la varianza del estimador y por lo tanto es

necesario aumentar el tamaño de muestra obtenido bajo un esquema aleatorio simple, a fin de lograr la precisión deseada⁵.

En el caso de que los conglomerados sean en promedio de igual tamaño, el efecto de diseño se puede expresar por medio de:

$$efd_k = [1 + \rho(M-1)] \quad (4)$$

$$efd_k = [1 + \rho(M^*-1)] \quad (5)$$

en donde M y M^* representan el tamaño del conglomerado y el tamaño promedio, y ρ se interpreta como el coeficiente de correlación intraconglomerados y se puede calcular por medio de⁶:

$$\rho = \left[\sum_i^M \sum_{j,l}^{M^*} (X_{ij} - \bar{X})(X_{il} - \bar{X}) \right] / \left[M(M^* - 1) M^* \sigma^2 \right] \quad (6)$$

y M y M^* representan el número de conglomerados en la población y el tamaño medio del conglomerado respectivamente y $\sigma^2 = (MM^* - 1)S^2 / MM^*$.

La varianza del estimador se puede escribir en términos del efd_k como se muestra a continuación:

$$V(\bar{x}) = (1-f) S^2 [1 + (M^* - 1) \rho] nM \quad (7)$$

lo cual permite comparar la varianza de un diseño aleatorio simple sin reemplazo (**MASSR**), con la que se obtiene de un esquema de selección por conglomerados (**DC**), lo cual se puede interpretar como una medida de la eficiencia relativa del diseño.

Es evidente que se desea que los conglomerados mantengan una baja correlación de sus elementos, lo cual se logra cuando $\rho = -1/(M^* - 1)$ por lo que en este caso, tanto la varianza del estimador como el efd_k son iguales a cero. Por otra parte, el caso más desfavorable se presenta cuando $\rho \rightarrow +1$, lo cual induce a un incremento sistemático en la varianza del estimador, y en esta situación todos los elementos son iguales por lo que el efecto de diseño asumiría un valor igual al tamaño promedio del conglomerado, de modo que la varianza será tan grande como la de las unidades elementales ($efd_k = M^*$).

⁵ Una interpretación adecuada del factor propuesto por Kish sería la siguiente. Cuando no se conoce la varianza de la variable de interés en un diseño complejo, pero si se dispone de una aproximación bajo un esquema aleatorio simple, entonces se calcula el tamaño de muestra bajo este esquema y luego se ajusta multiplicándolo por el efecto de diseño. De esta forma, el nuevo número de observaciones garantiza la misma varianza que un esquema de selección MAS.

⁶ La expresión (5) es la que se utiliza con mayor frecuencia, ya que en la práctica los conglomerados se forman cuidando que sean aproximadamente del mismo tamaño.

En el caso de que $\rho=0$, significaría que la variable está completamente distribuida al azar, por lo que el efecto de diseño es igual a uno ($\text{efd}_k=1$), y tanto el esquema de selección MAS como el de conglomerados DC generarían la misma varianza⁷.

El término $(M^* - 1)\rho$ debe interpretarse como el aumento que se genera en la varianza del estimador por haber seleccionado n conglomerados de tamaño M^* en lugar de nM^* unidades elementales⁸, y con esta información se está en condiciones de analizar **cuál es el efecto de un diseño de muestra complejo en el error estándar del estimador $\hat{\theta}$** .

Una manera de responder esta pregunta se logra al comparar las varianzas obtenidas por los dos diseños: el verdadero, que para una encuesta habitual corresponde a un esquema de selección estratificado y de conglomerados (DC) y uno hipotético generado por una selección MASSR (Skinner, 1989).

Como fue señalado, Kish (*op. cit.*) propuso comparar la eficiencia relativa del diseño complejo DC respecto a la varianza del MASSR, definiendo el factor conocido como efecto de diseño:

$$\text{efd}_k(\hat{\theta}) = \text{Var}_{\text{DC}}(\hat{\theta}) / \text{Var}_{\text{MAS}}(\hat{\theta}) \quad (8)$$

Skinner (*op. cit.*), señala que esta medida es apropiada cuando se desean comparar dos diseños alternativos; sin embargo, toda vez que se realizó la encuesta y si se aplicó un esquema diferente a un MASSR, resulta irrelevante hacer la comparación propuesta por Kish(efd_k), ya que los datos fueron generados por un mecanismo de selección compleja y poco o nada tienen que ver con un proceso de selección aleatorio y sin reemplazo.

La intención de Kish al proponer el efd_k fue medir el efecto de la varianza sobre el valor del estimador del parámetro $\hat{\theta}$; sin embargo, para los fines de una encuesta compleja es más importante conocer cuál es el efecto del diseño de muestra sobre el estimador de la varianza. Así, siguiendo a Skinner (*op. cit.*) sea $v_0 = \hat{\text{var}}_{\text{MAS}}(\hat{\theta})$ un estimador de la varianza del estimador del parámetro, bajo un esquema de selección MAS (asumiendo que las observaciones son independientes e idénticamente distribuidas iid)⁹. Así, el efecto de un diseño complejo sobre el estimador de la varianza se puede evaluar al comparar la distribución de v_0 con la varianza del diseño v_{DC} , lo cual se interpreta como el sesgo de estimación.

$$\text{sesgo}(v_0) = E_{\text{DC}}(v_0) - \text{Var}_{\text{DC}}(\hat{\theta}) \quad (9)$$

⁷ A pesar de que los conglomerados estén bien mezclados, en las situaciones prácticas el coeficiente de correlación intraconglomerado (ρ) tiende a ser mayor que uno, por lo que el efd_k puede ser considerable en la medida de que el tamaño del conglomerado sea grande.

⁸ Es normal que al interior de los conglomerados las unidades tengan un cierto parecido ($\rho > 0$), por lo que se espera que el diseño de conglomerados genere menor precisión. Sin embargo, es posible encontrar situaciones en que la correlación entre los elementos sea negativa ($\rho < 0$) lo cual significaría que el muestreo por conglomerados fue más preciso que un MAS.

⁹ Por ejemplo, para el caso de una proporción $v_0 = p(1-p)$.

A fin de lograr congruencia con la propuesta de Kish, (9) se puede expresar en términos relativos lo cual da lugar a la definición del factor **efme** que se refiere al efecto en la varianza del estimador debido a una mala especificación en el diseño de la muestra (“misespecification effect, **meff**”) propuesto en Skinner (*op. cit.*):

$$\text{efme}(\hat{\theta}, v_0) = \text{Var}_{\text{DC}}(\hat{\theta}) / E_{\text{DC}}(v_0) \quad (10)$$

En este caso, $\text{Var}_{\text{DC}}(\hat{\theta})$ representa la varianza observada en el diseño mientras que $E_{\text{DC}}(v_0)$ es el valor esperado de la varianza de un diseño MASSR y el $\text{efme}(\hat{\theta}, v_0)$ se interpreta como un factor que evalúa qué tanto v_0 tiende a subestimar o sobrestimar la varianza verdadera $\text{var}(\hat{\theta})$, tal que;

$$\begin{aligned} &< 1 \text{ si el sesgo}(v_0) > 0 \\ \text{efme}(\hat{\theta}, v_0) &= 1 \text{ si el sesgo}(v_0) = 0 \\ &> 1 \text{ si el sesgo}(v_0) < 0 \end{aligned} \quad (11)$$

En los diseños estratificados y de conglomerados con tamaños de muestra apropiados, $E_{\text{DC}}(v_0) \approx \text{Var}_{\text{MAS}}(\hat{\theta})$ por lo que se puede establecer una relación entre los factores de ajuste mencionados¹⁰;

$$\text{efd}_k(\hat{\theta}) = \text{efme}(\hat{\theta}, v_0) = \text{var}_{\text{DC}}(\hat{\theta}) / E_{\text{DC}}(v_0) \quad (12)$$

2. El Tamaño de Muestra Efectivo

Determinar el error de muestreo de un estimador es importante por las siguientes razones: i) Es necesario conocer la precisión y confiabilidad estadística de los datos generados; ii) Se debe evaluar la eficiencia estadística del diseño de muestra aplicado; y iii) Para generar información útil que pueda ser aprovechada para planear futuras investigaciones por muestreo.

A pesar de que la mayor parte de los estadísticos de encuestas ponen un mayor énfasis en minimizar los errores de muestreo, no hay que olvidar que también existen errores de no muestreo que afectan la calidad de los datos los cuales deben ser minimizados a partir de la supervisión y control de las actividades de diseño conceptual y del trabajo de campo.

En este sentido, a partir del efecto de diseño es posible derivar otras expresiones que son de gran utilidad práctica: el denominado factor de diseño **fdd** y el tamaño efectivo de muestra n_e . El factor de diseño se define como;

$$\text{fdd} = \text{efd}_k^{1/2} \quad (13)$$

¹⁰ En Skinner(*op.cit.*) se presentan expresiones para el cálculo del efd para diseños polietápicos y multivariados que llevan a la definición de una matriz V_0 de efectos de diseño de un vector de parámetros θ .

y es muy importante para la interpretación de la precisión de los estimadores y se utiliza fundamentalmente como un factor de ajuste para corregir el error estándar de estimador y la longitud de los intervalos de confianza (Verma et. al., 1980).

Por su parte, Kish(1965) definió el tamaño efectivo de muestra por medio de;

$$n_e = n_0 / efd_k \quad (14)$$

el cual puede ser interpretado como la cantidad de información contenida en una muestra. Por ejemplo, para un $efd_k=1.30$ en una muestra de conglomerados de tamaño $n_{DC}=10,000$ hogares, el tamaño efectivo indica que sólo se requeriría de $n_e=7,692$ hogares para estimar el parámetro θ con la misma precisión deseada a partir de un esquema aleatorio simple¹¹.

Es bien sabido que los errores de estimación y las pruebas de significancia juegan un papel fundamental en el análisis de los resultados de una encuesta ya que permiten, entre otras cosas, contrastar las hipótesis del comportamiento de una población ante un determinado fenómeno de interés.

De hecho, los errores de muestreo son muy diferentes cuando se trata de estimar estadísticas descriptivas a nivel poblacional, que cuando éstas se calculan para dominios de estudio específicos. Asimismo, la pruebas de significancia permiten establecer relaciones de causalidad y formular recomendaciones de política, por lo que su validación estadística es fundamental para determinar la confiabilidad y precisión de las aseveraciones que se realizan sobre un tema específico.

A pesar de que esto es muy importante, es muy frecuente que investigadores y analistas pasen por alto este hecho y utilicen la información sin preguntarse si los supuestos en que se basan los modelos que aplican se cumplen con la muestra observada. De hecho, los paquetes de cómputo más utilizados y convencionales (SAS, SPSS, entre otros) no consideran esta situación, y las rutinas de cálculo que tienen disponibles para efectuar inferencia estadística y estimar modelos de causalidad, asumen que las observaciones provienen de una muestra aleatoria simple extraída sin reemplazo, situación que no se cumple en el caso de los diseños de muestra complejos.

En este sentido, el factor de diseño **fd** debe tenerse presente al momento de interpretar los resultados de las encuestas, y de manera especial cuando se calculan los errores de muestreo de los estimadores. De esta manera, a continuación se presentan dos fases tradicionales de la inferencia estadística en donde el diseño de una muestra compleja tiene incidencia en la precisión de los resultados de la encuesta.

¹¹ El concepto de tamaño efectivo de muestra no debe ser confundido con el mismo término que se utiliza para señalar al número de unidades diferentes que se eligen en una muestra con reemplazo.

Suponga que se han procesado los resultados de la investigación y se desea calcular un intervalo de confianza para un conjunto de estimadores de interés. Se utiliza un paquete de cómputo convencional el cual asume que la muestra se extrajo en forma aleatoria y sin reemplazo, por lo que se considera una muestra de n observaciones independientes e idénticamente distribuidas (iid).

La expresión tradicional para calcular intervalos de confianza para un parámetro $\hat{\theta}$ parte de suponer que;

$$t_0 = (\hat{\theta} - \theta)/v_0^{1/2} \sim N(0,1) \quad (15)$$

de tal suerte que un intervalo al 95% de confianza se obtiene por medio de;

$$C_0 = \{ \theta : |\hat{\theta} - \theta| < 1.96 v_0^{1/2} \} = (\hat{\theta} - 1.96 v_0^{1/2}, \hat{\theta} + 1.96 v_0^{1/2}) \quad (16)$$

Ahora, la pregunta que se debe plantear y responder es **cuál es el efecto del diseño de muestra sobre los supuestos en que se basa la construcción del intervalo de confianza** propuesto en (16). Bajo un diseño de conglomerados se puede asumir que $\hat{\theta}$ es insesgado y tiene una distribución normal; es decir

$$\hat{\theta}_{DC} \sim N[\theta, \text{var}_{DC}(\hat{\theta})] \quad (17)$$

Además, para muestras grandes el valor de v_0 se aproxima a $E_{DC}(v_0)$ de tal forma que la distribución de $t_0 \approx (\hat{\theta} - \theta) / E_{DC}(v_0)^{1/2}$, por lo que de la ecuación (16) se tiene que:

$$t_{0DC} \sim N[0, \text{var}_{DC}(\hat{\theta}) / E_{DC}(v_0)] = N[0, \text{efd}_k(\hat{\theta}, v_0)] \quad (18)$$

Conforme a lo anterior, el efecto de diseño se puede interpretar como un factor que mide los cambios en la varianza (aumento o disminución) del estimador basados en el valor del estadístico t_0 . De esta manera, cuando

$$\begin{aligned} &= 1 \quad C_0 \text{ es correcto} \\ \text{efd}_k &> 1 \quad C_0 \text{ es muy pequeño} \\ &< 1 \quad C_0 \text{ es muy amplio} \end{aligned} \quad (19)$$

En términos prácticos, cuando se dispone de un estimador del efd_k un intervalo de confianza al 95% se puede calcular por medio de¹²:

$$C_0^* = (\hat{\theta} - 1.96 (v_0 \text{efd}_k)^{1/2}, \hat{\theta} + 1.96 (v_0 \text{efd}_k)^{1/2}) \quad (20)$$

¹² En Kish (1979) se ejemplifica que calcular un intervalo de confianza conforme a la expresión (16) equivale a utilizar la siguiente expresión: $\bar{y} \pm (t' / \sqrt{\text{efd}_k}) (s \sqrt{\text{efd}_k} / n)$. Así, para una $t=2$ y un $\sqrt{\text{efd}_k}=1.5$, entonces $t'=2/1.5=1.34$, y en consecuencia se tendría un incremento en la razón del error del 5% al 9%.

Por otra parte, en caso de que se desee comprobar la significancia de algún parámetro de interés ($\hat{\theta}$), la teoría estadística convencional establece que bajo el supuesto de normalidad la hipótesis simple $H_0: \theta=0$ se rechaza si $T_0 = (\hat{\theta} - \theta)/v_0^{1/2}$ se ubica en la región crítica (de rechazo) de una distribución normal estandarizada; así, para una prueba de dos colas a un 95% de confianza se tiene que:

$$H_0 \text{ se rechaza si } |T_0| > 1.96 \text{ bajo } H_0 \quad (21)$$

Sin embargo, bajo un diseño de muestra complejo;

$$T'_0 = T_0 / \text{efd}_k^{1/2} = (\hat{\theta} - \theta) / (v_0 \text{efd}_k)^{1/2} \quad (22)$$

3. Un Ejemplo Real ¹³

A continuación, se evalúan los resultados de una encuesta a fin de tratar de cumplir los siguientes objetivos: i) Evidenciar la manera en que se altera el error de estimación cuando se considera el diseño de la muestra; ii) Mostrar como se incrementan los errores de muestreo cuando se ajustan por el efecto de diseño (efd_k); iii) Indicar como crecen los errores en los estimadores a consecuencia de que en el diseño de muestra no se tuvieron las precauciones para determinar el número de observaciones necesarias. iv) Evidenciar los riesgos de mala interpretación que se enfrentan por el hecho de que los usuarios no validan la confiabilidad estadística de la información en los dominios de estudio que se forman.

En términos generales, se puede señalar que la información utilizada corresponde a una encuesta realizada por algún país de la región y se planeó con el objetivo central de estimar la tasa de desocupación abierta (TDA), pero que tiene como objetivo secundario -y tal vez esto es lo más importante- caracterizar el ámbito laboral de diferentes regiones del país, utilizando un modelo teórico que supone la existencia de un mercado laboral segmentado. Para el cálculo del tamaño de muestra se utilizó la expresión propuesta por Cochran (*op. cit.*) suponiendo una distribución binomial del parámetro de interés y se fijó una precisión y un error máximo esperado.

De esta manera, se deseaban generar estimaciones para el área urbana de la capital del país, así como para cada una de las zonas urbanas de las cabeceras municipales de los departamentos más importantes; y por agregación se podrían obtener estimaciones para el conjunto urbano. Además de la clasificación tradicional que permite identificar a los desocupados abiertos, el marco conceptual utilizado establece y define diversas modalidades de desocupación y subutilización de la fuerza de trabajo, por lo que se consideró de interés

¹³ Los resultados presentados corresponden a una encuesta real efectuada en algún país de Hispanoamérica, y el único objetivo de presentar resultados es con el ánimo de ejemplificar los errores que se pueden cometer por el mal uso de la información y la utilización de los datos para efectos distintos a los que motivaron la realización de la investigación.

disponer de una caracterización de la población económicamente activa ubicada en las diversas categorías de estudio¹⁴, así como de las variables sociodemográficas y económicas que condicionan la utilización de la mano de obra¹⁵.

Para la elaboración del diseño de la muestra se definieron como unidades primarias de muestreo (UPM's) las cabeceras municipales de los departamentos del país, mientras que las unidades de segunda etapa (USM's) se formaron a partir de segmentos de viviendas. En ambos casos, la selección de las observaciones que formaron parte de la muestra se hizo asignando igual probabilidad a todas las unidades ya que aparentemente no había diferencias considerables en su tamaño

Se trata de una encuesta que se realiza en forma continua y que permite generar información para estudiar la evolución del mercado de trabajo en la zona urbana del país, y de manera particular sobre los niveles de desocupación y subutilización de la fuerza de trabajo.

En la tabla 1 se presentan resultados de los totales expandidos para distintas variables interés, los cuales fueron estimados utilizando el paquete PC CARP el cual considera en sus rutinas de cálculo la estructura del diseño de la muestra (efecto de estratificación y conglomeración) y las diferentes probabilidades de selección de las observaciones¹⁶.

¹⁴ La clasificación utilizada corresponde a la que recomienda la Organización Internacional del Trabajo(OIT) para el estudio de la desocupación y el comportamiento del mercado laboral.

¹⁵ Además, y ante falta de información actualizada, es frecuente que estas encuestas se utilicen para realizar estimaciones de pobreza lo cual le impone restricciones adicionales a los objetivos de la investigación, tamaños de muestra y precisión de los estimadores.

¹⁶ Es importante señalar que el PC CARP utiliza para los cálculos de la varianza el método del conglomerado último propuesto por Hansen *et. al.*(1963), por lo que es necesario identificar la unidades primarias de muestreo y definir un ponderador para cada una de las observaciones. Por las características del diseño muestral, el factor de expansión se calculó a partir del cociente entre una estimación de población proyectada para un año determinado, entre el total de observaciones en muestra.

Cuadro 1

**Coefficiente de Variación y Efecto de Diseño para Estimadores de Totales
Area Urbana del País**

VARIABLE	ESTIMADO R DEL TOTAL	ERROR ESTÁNDAR	COEF. DE VARIACIÓN	EFFECTO DE DISEÑO	COEF. DE VARIAC.* AJUSTADO
Población Total	1'352,960	16,733.20	1.2368	****	****
Activos	639,703	9,271.82	1.4494	3.3852	2.6667
<i>Ocupados</i>	<i>572,802</i>	<i>8,486.60</i>	<i>1.4816</i>	<i>2.8957</i>	<i>2.5212</i>
Plenos	375,236	6,274.02	1.6720	1.9277	2.3214
Subocupados	198,706				
<i>Invisibles</i>	<i>108,531</i>	<i>3,858.63</i>	<i>3.5553</i>	<i>1.9807</i>	<i>5.0036</i>
<i>Visibles</i>	<i>90,175</i>	<i>3,831.52</i>	<i>4.2490</i>	<i>2.3163</i>	<i>6.4667</i>
<i>Desocupados</i>	<i>66,901</i>	<i>3,059.73</i>	<i>4.5735</i>	<i>1.9550</i>	<i>6.3947</i>
Cesantes	51,896	2,621.14	5.0508	1.8282	6.8292
Aspirantes	15,005	1,246.60	8.3079	1.3908	9.7978
Inactivos	713,256	11,085.70	1.5542	4.8393	3.4190
<i>Con Exper.</i>	<i>36,765</i>	<i>2,155.11</i>	<i>5.8619</i>	<i>1.7245</i>	<i>7.6979</i>
<i>Sin Exper.</i>	<i>676,491</i>	<i>10,737.50</i>	<i>1.5872</i>	<i>4.5266</i>	<i>3.3769</i>

* Corresponde al valor que se obtiene al multiplicar el error estándar del estimador por la raíz cuadrada del efecto de diseño.

En todos los casos, los coeficientes de variación(CV) estimados son menores al 10%. Sin embargo, se aprecia que en la medida que las categorías en que se distribuye la población ocupada presentan frecuencias más bajas, el error del estimador se incrementa pasando el CV de 1.24% para el total de población, al 8.31% que corresponde al caso de los aspirantes a obtener un empleo.

Asimismo, se debe observar que los mayores valores en el efecto de diseño se presentaron en las subclases en donde la homogeneidad de las observaciones es mayor. Así, por ejemplo, correspondió un $efd_k = 9.8$ en la categoría de aspirantes a conseguir un empleo, que indicaría un valor alto en el coeficiente de correlación intraconglomerados, lo cual conlleva a que se incremente la varianza del diseño de muestra respecto a una selección aleatoria. En este caso, el valor del coeficiente de correlación intraconglomerados estimado fue $\rho = 0.1086$, lo cual evidencia la relación que existe entre las observaciones¹⁷.

¹⁷ Es importante recordar que para efectos del cálculo de la varianza, se sugiere que se mantengan correlaciones intrasconglomerados muy bajas, con objeto de que el valor del efd_k sea cercano a la unidad. Es decir, se busca que el diseño de conglomerados no incremente de manera artificial la varianza del estimador.

De la misma manera, en las categoría de activos, ocupados (con valores en el efd_k mayores de 2) y en la de los ocupados plenos, se percibe una mayor heterogeneidad intraconglomerados, y en este caso los valores de ρ fueron **0.0225**, **0.0188** y **0.0163** respectivamente, lo cual contribuyó a disminuir la varianza del estimador y se reflejó en los valores bajos observados en el coeficiente de variación (CV).

Por otra parte, también se debe señalar que el verdadero error del estimador se determina al considerar en el cálculo de la varianza el factor de ajuste que se obtiene por medio de la raíz cuadrada del efecto de diseño (\sqrt{efd}). Así, es evidente que debido a que para todas las variables el $efd_k > 1$, se observa que el coeficiente de variación ajustado por el diseño de muestra ($CV^* = CV * \sqrt{efd}$) se incrementa, y para el caso de las personas que se clasificaron en la categoría de aspirantes su valor crece en el entorno del 10% y para los inactivos con experiencia se ubicó en 7.7%.

Asimismo, a fin de ilustrar cómo cambian los límites del intervalo de confianza al considerar el efecto del diseño sobre la varianza del estimador, a continuación se presentan los resultados obtenidos para algunas de las variables analizadas.

Total de Activos: **693,703 ± 18,173 sin ajuste y**
 693,703 ± 33,436 ajustados por efd_k ;

Total de Inactivos: **713,256 ± 21,728 sin ajuste y**
 713,256 ± 47,798 ajustados por efd_k .

En el caso de los activos existe una diferencia absoluta de 15,263 personas en los límites calculados por ambos procedimientos, lo cual podría representar una diferencia poco importante. Por otra parte, en el caso de los inactivos los resultados ajustados evidencia una diferencia de más de dos veces el valor de los límites obtenido sin ajustar el error estándar del estimador, lo cual permite apreciar la relevancia de estimar en forma apropiada el error de muestreo.

Para el caso de los **desocupados**, se tiene que la encuesta estima que 66,901 personas se consideraron en situación de desocupación abierta con una diferencia de $\pm 5,997$ personas según el error estándar del estimador. Sin embargo, cuando en los cálculos se incorpora el efecto de diseño se tiene que los límites cambian y se observa una diferencia de $\pm 8,385$ personas. Si se decide diseñar acciones de política a partir de la puesta en marcha de un programa emergente de empleo que involucre a toda la población desocupada (en este caso se debiera considerar como techo presupuestario el límite superior del intervalo de confianza), es evidente que los recursos que se requieren invertir son diferentes en los casos señalados. De hecho, la diferencia absoluta de 2,388 personas desempleadas establece una diferencia significativa en el presupuesto que se debiera asignar para la operación del programa.

Siguiendo a Kish (1979) se enfatiza que el trabajar de la manera tradicional (expresión (16)) supone que tanto el valor de tablas (t) obtenido de la distribución normal estandarizada como el error estándar del estimador, se afectan por la raíz cuadrada del efecto de diseño ($\bar{y} \pm t' [s \sqrt{\text{efd}_k / n}]$, en donde $t' = t / \sqrt{\text{efd}_k}$). Así, en el caso que nos ocupa, para obtener un intervalo al 95% de confianza ($\alpha=5\%$) se sabe que $t=1.96$; por lo que para el total de activos se tiene que $\sqrt{\text{efd}_k}=1.8399$ y $t'=1.0653$ lo cual significa que el error de estimación se incrementa del 5% al 14.5%; es decir, aumenta más de nueve puntos porcentuales, sin que el analista se percate de esta situación y en sus conclusiones continúe afirmando que el intervalo de confianza que contiene el verdadero valor del parámetro corresponde a un $\alpha=5\%$.

En este análisis los valores del error de muestreo con y sin efd_k pueden parecer menores para aquellos que trabajan los métodos convencionales de cálculo; sin embargo, en este caso el hecho de que las diferencias no sean tan grandes se debe, en parte, a que se trabaja a un nivel de agregación en donde el tamaño de muestra es suficiente para garantizar una precisión adecuada¹⁸. Sin embargo, un análisis detallado de los datos permite observar que en la categoría de personas que se clasificaron como aspirantes a ocupar un puesto en el mercado de trabajo, sólo se ubicaron 212 observaciones lo cual denota que cualquier intento por desagregar este tamaño de muestra podría conducir a resultados erróneos de interpretación. Es decir, no es posible siquiera intentar conocer a nivel desagregado la rama de actividad a la cual desearía insertarse esta subpoblación.

Asimismo, se señala que 60 de las 212 observaciones registradas (28.3%) se concentraron en la capital del país, mientras que en 10 municipios se reportaron menos de 10 casos y en uno de ellos no se obtuvo ninguna observación, lo cual rompe con la lógica de selección de la muestra ya que los municipios se definieron como UPM's y se espera que todos contribuyan en la formación del estimador del área urbana del país.

Un comportamiento similar se observó al analizar la distribución muestral de la categoría de subempleados visibles. En este caso, se dispone de 1,241 personas que se ubicaron en esta situación, de las cuales el 31.7% se concentró en la capital del país. Asimismo, existen 6 municipios en donde el tamaño de muestra reportado fue menor a treinta observaciones y en uno de ellos sólo se registraron dos casos.

Situaciones como la descrita se presentaron en algunas otras de las categorías de análisis, lo cual debe obligar a los usuarios a reflexionar sobre la confiabilidad estadística que puede tenerla información desagregada cuando ésta se pretende utilizar para formular hipótesis de comportamiento o establecer acciones de política. De hecho, hay que enfatizar que la encuesta cumple sus propósitos para estimar con precisión la tasa de desocupación abierta (TDA) a nivel departamental y para la zona urbana del país, pero presenta limitaciones para caracterizar el mercado de trabajo a nivel municipal, y evidentemente tiene

¹⁸ En la encuesta que se utilizó para ejemplificar los conceptos aquí presentados, existen departamentos en donde los errores de estimación para algunas variables de interés son mayores y por lo tanto las diferencias en los límites de los intervalos de confianza son más notables.

escasa capacidad para establecer modelos de comportamiento y formular relaciones de causalidad entre las variables que explican la conformación del mercado laboral.

Sin embargo, a pesar de la evidencia este tipo de situaciones frecuentemente se pasan por alto y muchos analistas actúan como si la información recabada fuera lo suficientemente confiable para inferir sobre el comportamiento de la población en su conjunto y abusan de la bondad de los datos al momento de elaborar sus análisis y diseñar alternativas de políticas.

Es obvio que el no considerar la estructura del diseño de la muestra incrementa el error cuadrático medio de los estimadores, sin que los analistas tengan capacidad de controlar el efecto que esto puede tener en las conclusiones derivadas de sus análisis.

En la tabla 2 se presentan para el área urbana del país los resultados de algunas tasas que son de gran utilidad para analizar el comportamiento del mercado de trabajo. Así, se tiene que toda vez que el error de muestreo se ajustó por el efecto de diseño, la mayor variabilidad observada es menor al 6% la cual corresponde al estimador de la tasa de desocupación abierta (TDA). Esta situación confirma la confiabilidad de la información a este nivel de agregación y corrobora que el total de observaciones logradas es suficiente para satisfacer la precisión deseada.

Cuadro 2

Coeficiente de Variación y Efecto de Diseño para Estimadores de Tasas Área Urbana del País

Variable	Estimador Tasa en %	Error Estándar	Coef. de Variación %	Efecto de Diseño	Coef. de Variación Ajustado
PEA/PO10	47.2818	.004286	0.9066	1.3244	1.0433
PEI/PO10	52.7182	.004286	0.8131	1.3244	0.9357
PEA/PEI	89.6877	.015423	1.7197	1.3244	1.9791
OCUP/PEA	89.5419	.004384	0.4896	1.7435	0.6464
PLE/OCUP	65.5088	.007646	1.1672	1.9683	1.6375
DESO/PEA	10.4581	.004384	4.1919	1.7435	5.5350
INAS/PEI	94.8455	.002921	0.3080	1.6528	0.5090

PEA ≈ Población Económicamente Activa

PO10 ≈ Población en edad de trabajar de 10 y más años

PEI ≈ Población Económicamente Inactiva

OCUP ≈ Ocupados

PLE ≈ Ocupados Plenos

DESO ≈ Desocupados

INAS ≈ Inactivos sin experiencia

Los intervalos de confianza para los estimadores de algunas tasas de interés se presentan a continuación:

Tasa Neta de Participación : 47.28 ± 0.84 sin ajuste y ± 0.97 con ajuste;
Tasa de Ocupación : 65.51 ± 0.86 sin ajuste y ± 1.13 con ajuste;
Tasa de Desocupación Abierta: 10.46 ± 0.86 sin ajuste y ± 1.13 con ajuste.

De esta forma, cuando no se considera el factor de ajuste la tasa de desocupación abierta puede tener un valor máximo de 11.3%, mientras que cuando se tiene en cuenta el efecto de diseño el valor de este indicador se incrementa y puede involucrar hasta 11.6% de la fuerza de trabajo, lo cual para todo efecto práctico no resulta significativo.

A fin de conocer el error en que se incurre por no considerar en el proceso de estimación las características del diseño de la muestra, en la tabla 3 se presentan resultados obtenidos por diversas alternativas de cálculo.

Cuadro 3

Error Estándar para Estimadores de Tasas Area Urbana del País

Variable	Estimad or PCCAR P ¹	Error Est. PCCAR P	Error Est. ² PCCAR P	Estimad or SASP ³	Error Est. SASP	Estimad or SASSP ⁴	Error Est. SASSP ⁵
Tasa Neta de Partic.	47.28	.004286	.004932	47.41	.003725	47.28	.003725
Tasa de Ocupac.	89.54	.004384	.005789	89.54	.003317	89.54	.003316
Tasa de Desocup Abierta	10.46	.004900	.006470	10.46	.003317	10.46	.003316

¹ Corresponde a los resultados considerando la estructura del diseño de muestra y los ponderadores de las observaciones.

² Ajustado por el efecto de diseño que calcula la rutina incluida en el PC CARP.

³ Corresponde a los resultados ponderados calculados con el SAS.

⁴ Corresponde a los resultados sin ponderar calculados con el SAS.

⁵ El error estándar calculado por el SAS (pq/\sqrt{n}) se considera una aproximación a la varianza verdadera basada en el teorema del límite central para muestras grandes (Kakwani,

1990). Sin embargo, a pesar de este hecho se nota que al no considerar la estructura del diseño se subestima el error de muestreo del estimador.

En primera instancia, es importante señalar que no existen diferencias significativas en los valores de las tasas estimadas con los datos ponderados y sin ponderar, debido a que a partir de una proyección de población (y utilizando un estimador de razón) se definieron factores de expansión constantes por departamento; es decir, se tuvieron muestras autoponderadas a ese nivel geográfico¹⁹.

Por otra parte, se observa que más allá de los valores absolutos de los errores calculados por las diferentes metodologías existe la tendencia general a subestimar el verdadero error de muestreo cuando en la rutina de estimación no se consideran las restricciones impuestas por el diseño de la muestra ni las diferentes probabilidades de selección de las observaciones.

Los problemas de tamaño de muestra que se señalaron en párrafos anteriores se agravan y complican las capacidades de explotación de la información, en la medida que se desagregan los datos y el análisis se efectúa para dominios de estudio más específicos, en donde los tamaño de muestra resultan insuficientes para garantizar una adecuada precisión de los estimadores, así como de los estadísticos utilizados para la contrastación de hipótesis.

Así, en la tabla 4 se presentan estimaciones de algunas variables de interés para el área urbana de la capital del país. De esta forma, a pesar de que se consignan totales de subocupados visibles, aspirantes a conseguir un empleo y de cesantes, en unidades de millar, cualquier intento por desagregar esta información y formular hipótesis de comportamiento debiera obligar al usuario a observar que los valores muestrales para esas categorías fueron 394, 279 y 60 casos respectivamente.

¹⁹ A pesar de que esta es una práctica común, es claro que esta forma de actuar limita la capacidad de uso de las encuestas. Esto significa que en aquellas organizaciones que asignan recursos permanentes para la actualización del marco tienen la oportunidad, en cada levantamiento, de poner al día la cartografía y los pesos relativos de las unidades de selección lo cual permite generar estimaciones de población alternativas en períodos intercensales. Sin embargo, las que por razones de falta de presupuesto no realizan esta actividad se deben conformar con trabajar con estimadores de razón lo cual puede llegar a ocasionar severas inconsistencias entre las estimaciones derivadas de las muestras y los censos.

Cuadro 4

**Coefficiente de Variación y Efecto de Diseño para Estimadores de Totales
Area Urbana de la Capital del país**

Variable	Estimador de Totales	Error Estándar	Coef. De Variación %	Efecto de Diseño	Coef. De Variación Ajustado
PLENOS	230,620	5,155.2	2.2353	1.4363	2.6789
INVISIBLE	52,326	2,998.94	5.7313	1.6162	7.2862
VISIBLE	44,916	2,865.62	6.3799	1.7018	8.3227
CESANTE	31,806	2,309.88	7.2624	1.5340	8.9948
ASPIRAN.	6,840	950.33	13.894	1.1683	15.0177
INAC/EXP.	18,924	17,833.5	9.4237	1.5107	11.5827
INAS/EXP	393,528	83,795.0	2.1239	3.1674	3.7799
ACTIVOS	365,370	6,453.89	1.7664	1.8856	2.4256
INACTIVO	412,452	8,473.15	2.0543	3.2501	3.7035
OCUPADO	326,724	5,975.91	1.8290	1.6530	2.3515
DESOCUP.	38,646	2,635.75	6.8202	1.6591	8.7848

PLENOS	≈ Activos Plenos	INAC	≈ Inactivo con
Experiencia			
INVISIBLE	≈ Subempleados Invisibles	INAC SIN	≈ Inactivos sin
Experiencia			
VISIBLE	≈ Subempleados Visibles	INACTIVO	≈ Inactivos
CESANTE	≈ Desocupados Abiertos	OCUPADO	≈ Ocupados
ASPIRANTE	≈ Aspirante a ocupado	DESOCUP	≈ Desocupados

Esto significa que las particiones que se quieren formar con los datos deben considerar este hecho, a fin de prever que los tamaños de muestra observados garanticen que los estadísticos de prueba converjan a las distribuciones teóricas que se utilizan para calcular intervalos de confianza y contrastar hipótesis.

Los resultados obtenidos al comparar algunos procedimientos de cálculo para la parte urbana de la capital del país se presentan en la tabla 5. Así, al igual que en el análisis realizado para el área urbana en su conjunto, se confirma la tendencia de los paquetes de uso convencional a subestimar el verdadero error de muestreo.

Cuadro 5

Error Estándar para Estimadores de Tasas. Area Urbana de la Capital del País

Variable	Estimad or PCCAR P ¹	Error Est. PCCAR P	Error Est. ² PCCARP	Estimad or SASP ³	Error Est. SASP	Estimad or SASSP ⁴	Error Est. SASSP ⁵
Tasa Neta de Partic.	46.97	.006313	.008116	46.97	.006043	46.97	.006043
Tasa de Ocupac.	89.42	.006732	.010264	89.42	.005433	89.42	.005433
Tasa de Des. Ab.	10.58	.006732	.010264	10.57	.005433	10.58	.005433

¹ Corresponde a los resultados considerando la estructura del diseño de muestra y los ponderadores.

² Ajustado por el efecto de diseño que calcula la rutina incluida en el PC CARP.

³ Corresponde a los resultados ponderados calculados con el SAS.

⁴ Corresponde a los resultados sin ponderar calculados con el SAS.

⁵ El error estándar calculado por el SAS (pq/\sqrt{n}) se considera una aproximación a la varianza verdadera basada en el teorema del límite central para muestras grandes (Kakwani, 1990). Sin embargo, a pesar de este hecho se nota que al no considerar la estructura del diseño se subestima el error de muestreo del estimador.

A fin de ejemplificar qué pasa en la situación de aquellos dominios de estudio en los que se enfrentan mayores limitaciones en relación al tamaño de muestra, la tabla 6 presenta información sobre los resultados obtenidos para una de las provincias del país.

En este caso los valores muestrales para las variables subempleados visibles, cesantes y aspirantes fueron 94, 24 y 16 respectivamente. Asimismo, las frecuencias absolutas observadas para los inactivos, con y sin experiencia, se ubicaron en 13 y 496.

De esta manera, nuevamente se pone de manifiesto las restricciones que se enfrentan al tratar de inferir sobre el comportamiento de subpoblaciones con características muy específicas y tamaños de muestra insuficientes.

Cuadro 6

**Coefficiente de Variación y Efecto de Diseño para Estimadores de Totales
Area Urbana de una Provincia del país**

Variable	Estimador de Totales	Error Estándar	Coef. de Variación %	Efecto de Diseño	Coef. De Variación Ajustado
PLENOS	15,345	1,018.67	6.6385	1.6883	8.6257
INVISIBLE	5,610	647.73	11.5460	1.5087	14.1818
VISIBLE	5,170	931.79	18.0230	3.3587	33.0303
CESANTE	1,320	370.53	28.0710	1.9347	39.0449
ASPIRAN.	880	308.74	35.0840	1.9987	49.6001
INAC/EXP.	715	233.21	32.6170	1.3995	38.5860
INAS/EXP	27,280	1,374.08	5.0369	2.4381	8.0362
ACTIVOS	28,325	1,457.81	5.1467	2.7418	8.5221
INACTIVO	27,995	1,420.14	5.0729	2.6019	8.1828
OCUPADO	26,125	1,317.68	5.0438	2.2517	7.5686
DESOCUP.	2,200	451.36	20.5160	1.7504	27.1432

La tabla 7 confirma el comportamiento observado en relación a que se subestima el error de muestreo a partir de rutinas de cálculo que suponen que la muestra corresponde a un diseño monoetápico en donde las unidades se seleccionaron con muestreo aleatorio simple sin reemplazo (MASSR).

Cuadro 7

Error Estándar para Estimadores de Tasas. Area Urbana para una Provincia del País

Variable	Estimador PCCAR P ¹	Error Est. PCCAR P	Error Est. ² PCCAR P	Estimador SASP ³	Error Est. SASP	Estimador SASSP ⁴	Error Est. SASSP ⁵
Tasa Neta de Partic.	5.29	.015574	.041363	50.29	.015632	59.29	.015632
Tasa de Ocupac.	92.84	.014578	.015399	92.23	.0118056	92.23	.0118056
Tasa de Des. Ab.	7.77	.014578	.015399	7.77	.0118056	7.77	.0118056

¹ Corresponde a los resultados considerando la estructura del diseño de muestra y los ponderadores.

² Ajustado por el efecto de diseño que calcula la rutina incluida en el PC CARP.

- ³ Corresponde a los resultados ponderados calculados con el SAS.
⁴ Corresponde a los resultados sin ponderar calculados con el SAS.
⁵ El error estándar calculado por el SAS (pq/\sqrt{n}) se considera una aproximación a la varianza verdadera basada en el teorema del límite central para muestras grandes (Kakwani, 1990). Sin embargo, a pesar de este hecho se nota que al no considerar la estructura del diseño se subestima el error de muestreo del estimador.

III. EL MODELO DE REGRESIÓN Y EL DISEÑO DE LA MUESTRA

La utilización de información proveniente de encuestas por muestreo para formular y validar relaciones de causalidad y recomendar opciones de política es una actividad que se realiza de manera habitual en diversas organizaciones públicas y privadas, así como en organismos internacionales, universidades y centros de investigación, utilizando como insumos los resultados generados por las encuestas de hogares (empleo, ingresos y gastos y niveles de vida, entre otras).

En este sentido, es muy importante conocer cuáles son las limitaciones impuestas a los datos por el diseño de muestra que generó la información, a fin de evaluar si las hipótesis sobre las que se basan los métodos de ajuste y análisis se cumplen, y al mismo tiempo determinar si las conclusiones tienen validez estadística.

Tal vez una de las técnicas estadísticas a las que se recurre con mayor frecuencia para explicar la relación que existe entre un conjunto de variables sea el análisis de regresión.

De esta manera, en esta sección se realiza una evaluación de los supuestos en que se basa la aplicación de esta técnica así como de las repercusiones en su uso, por el hecho de que la mayor parte de las muestras que se analizan en el ámbito socioeconómico provienen de encuestas efectuadas a partir de diseños de muestra multietápicos y estratificados.

Es bien sabido que a partir del modelo de regresión es posible corroborar si existe dependencia lineal entre un vector de variables explicativas o independientes ($\sim x$) y una variable y denominada dependiente la cual se asume con distribución normal.

La expresión que relaciona la variable y con el vector $\sim x$ se expresa por medio de una combinación lineal de parámetros desconocidos y las variables independientes;

$$y_i = x_0 \beta_0 + x_{1i} \beta_1 + x_{2i} \beta_2 + \dots + x_{pi} \beta_i + \varepsilon_i; \quad i = 1, \dots, n \quad (23)$$

en donde las β_i 's son los parámetros que se desean estimar, y se establece como hipótesis que el término de error ε_i tiene una distribución normal con media cero y varianza constante ($\varepsilon_i \sim N[0, \sigma^2]$). Asimismo, se asume que los errores no están correlacionados y que las observaciones son independientes e idénticamente distribuidas (iid) y que provienen de una muestra aleatoria con una cierta distribución de probabilidad.

En términos matriciales (23) se puede expresar como;

$$E(Y/X=x) = \alpha + X^t \beta + \varepsilon \quad (24)$$

La estimación de la ecuación (24) por medio del método de mínimos cuadrados ordinarios(MCO) genera la solución:

$$\hat{\beta} = (X^t X)^{-1} X^t Y \quad (25)$$

a partir de suponer que los datos provienen de una muestra aleatoria seleccionada sin reemplazo (MASSR) y que no están relacionados con el diseño de la muestra. Sin embargo, este supuesto no se cumple en la realidad ya que el efecto de conglomeración genera correlación entre los errores (Kish y Frankel, 1974), por lo que es necesario establecer algunas consideraciones cuando los datos provienen de encuestas complejas.

De hecho, el método de estimación MCO supone que $V(\varepsilon / X) = \sigma^2 I$; sin embargo, la realidad indica que $V(\varepsilon / X) = \sigma^2 V$ por lo que se debe buscar un método alternativo y adecuado de solución.

A partir de lo anterior se puede afirmar que no se verifica la hipótesis de independencia del término de error (Sul, E. *et. al.* 1994) por lo que los estimadores y sus varianzas son sesgados, lo cual tiene consecuencias en el cálculo de intervalos de confianza y en las expresiones para obtener los estadísticos de prueba que se utilizan para la contrastación de hipótesis.

Una expresión del estimador $\sim\beta$ para el caso de probabilidades desiguales de selección se obtiene por medio de:

$$\hat{\beta}^* = (X^t D^{-1} X)^{-1} X^t D^{-1} Y \quad (26)$$

en donde la matriz $X^t D^{-1} X$ tiene un elemento $\sum_s X_{ij} X_{ik} / \pi_i$, y la sumatoria se extiende sobre las n unidades que forman parte de la muestra y D representa una matriz con las probabilidades de selección incluidas en la diagonal²⁰.

De igual forma se afirma que la varianza del vector ($\sim\beta$):

$$V_{MCO}(\beta) = [n_0(n_0 - k)]^{-1} \sum e^2_t V^{-1}_{xx} \quad (27)$$

puede ser inconsistente aún en el caso de que la muestra sea aleatoria simple debido a la presencia de heterocedasticidad en los datos, mientras que un estimador de la varianza de $\hat{\beta}^*$ se calcula como:

²⁰ De acuerdo con Kish y Frankel (1974), la estimación de (26) se puede lograr por los siguientes tres métodos para estimar la varianza del estimador para muestras repetidas de tamaño n provenientes de encuestas complejas: Expansión de serie de Taylor ST, Replicaciones Repetidas y Balanceadas RRB y Replicaciones con Jackknife RJ.

$$\hat{V}(\hat{\beta}^*) = (\mathbf{X}' \mathbf{D}^{-1} \mathbf{X})^{-1} \sigma^2 \quad (28)$$

Por la forma de la expresión (23) una alternativa para resolver el problema de estimar de manera apropiada (24) consiste en utilizar el método de mínimos cuadrados ponderados (MCP) (Fuller, 1975; Holt *et. al.*, 1980 y Shah *et. al.*, 1977) y como alternativas posibles se proponen el método de mínimos cuadrados generalizados (MCG) y el de máxima verosimilitud (MV) (Holt *et. al. op. cit.*). En este sentido, el ponderador es el inverso de la probabilidad de selección de la observación (factor de expansión) o un factor ajustado por no respuesta y postestratificación (Sul *et. al., op. cit.*)²².

Esta forma de proceder no considera la posible correlación que exista entre el término de error y su efecto sobre la varianza del estimador. De hecho, la correlación entre los errores aleatorios tiene un efecto mínimo sobre el estimador de los parámetros el cual puede ser ignorado.

EJEMPLO REAL

Suponga que a partir de los datos de una encuesta de empleo se quiere estudiar la relación que existe entre la condición de ocupación de un individuo y algunas variables sociodemográficas como la edad, los años de estudio y el sexo, la cual se desea modelar a partir de definir una relación lineal entre las variables.

$$\text{cond. de ocupación} = f(\text{edad, años de estudio, sexo}) \quad (29)$$

De esta forma, a partir del modelo de regresión lineal simple se desea estimar el valor de los parámetros β_i , así como su error estándar y corroborar su significancia estadística.

$$\text{condocu} = \beta_1 + \beta_2 \text{ edad} + \beta_3 \text{ edu} + \beta_4 \text{ sexo} + \varepsilon \quad (30)$$

Los resultados para distintos métodos de estimación se presentan en la tabla 8²³.

Como se puede observar, existen pequeñas diferencias entre los valores de los parámetros estimados por el método de MCO y los obtenidos cuando las observaciones se

²¹ Esta expresión no corresponde a la propuesta por Kish y Frankel (*op. cit.*); sin embargo, en el caso de que no se detecte la presencia de heterocedasticidad en los datos, la estimación propuesta para estimar la varianza de los estimadores es una buena aproximación (Holt *et. al. op. cit.*)

²² Los procedimientos de postestratificación han sido analizados por Holt y Smith (1989) quienes han concluido que se pueden considerar como técnicas robustas de estimación. En su trabajo ellos demuestran que, contrario a lo que se supone habitualmente, las muestras autoponderadas generan estimadores sesgados para la media y la postestratificación permite solucionar este problema.

²³ Los datos utilizados para la estimación del modelo son reales y se obtuvieron de una encuesta de empleo realizada por una Oficina Nacional de Estadística de la región.

ponderan utilizando el factor de expansión y el método de MCP. De hecho, se presenta una total coincidencia entre los resultados generados utilizando el PC CARP, considerando el factor de expansión y la estructura del diseño de muestra, y aquellos que se producen a partir del SAS y un método de estimación que considera únicamente el factor de expansión²⁴.

Cuadro 8

Estimadores y su Error Estándar por diferentes Métodos de Estimación

Parámetro	Estimador	Error ¹ Estándar	Estimador	Error ² Estándar	Estimador	Error ³ Estándar
β_1	.869483	.016124	.862701	.012760	.869484	.012450
β_2	.002139	.000272	.002348	.000224	.002139	.000221
β_3	-.003405	.001298	-.003333	.001256	-.003405	.001233
β_4	-.000814	.006116	-.004796	.005969	-.000814	.005799

¹ Corresponde al estimado considerando el factor de expansión y la estructura del diseño de la muestra y utilizando para la estimación el PC CARP.

² Corresponde al estimado por Mínimos Cuadrados Ordinarios con SAS.

³ Corresponde al estimado sin considerar el diseño de la muestra, pero estimado por Mínimos Cuadrados ponderados con el SAS.

Lo que se debe señalar de los resultados obtenidos es el hecho que el error estándar del estimador se subestima siempre que no se considera el diseño de la muestra y los ponderadores, y aún en la situación en que se ponderan los resultados pero el esquema de selección de la muestra no se tiene en cuenta²⁵.

Esto significa que, a pesar de que se aplique un método de estimación que involucre las distintas probabilidades de selección de las observaciones, es posible que persista una subestimación del error del estimador.

Es evidente que en el caso ilustrado no se generarían mayores problemas de inferencia debido a las ligeras discrepancias observadas en las estimaciones del error; sin embargo, en la literatura abundan los casos en donde las diferencias pueden ser considerables y conllevan a asumir conclusiones erróneas por el mal manejo de los datos (Skinner *op. cit.* y Holt *et. al. op. cit.*).

²⁴ Los parámetros β_1 , β_2 y β_3 resultaron estadísticamente distintos de cero, mientras que β_4 no tiene significancia estadística lo cual indica que para este conjunto de datos el sexo de la persona no ayuda a explicar la condición de actividad de la población en edad de trabajar.

²⁵ Resultados similares fueron obtenidos por Sue, *et al.* (1989), para una investigación que relacionaba el índice de masa corporal con distintas variables sociodemográficas.

Una aproximación al efecto de conglomeración sobre la varianza de los parámetros estimados se presenta en la tabla 9, el cual se puede interpretar como el porcentaje de subestimación que existe en el estimador del error estándar al considerar a las observaciones como una muestra aleatoria seleccionada sin reemplazo.

Cuadro 9

**EFFECTO DE DISEÑO PARA LOS PARÁMETROS
DEL MODELO ESTIMADO**

Parámetro	Efecto de Diseño
β_1	1.2636
β_2	1.2143
β_3	1.0334
β_4	1.0246

**IV. LA ESTIMACIÓN DEL EFECTO DE DISEÑO Y EL USO
DE LOS PONDERADORES**

Se ha evidenciado que para hacer un uso adecuado de la información es necesario que los usuarios conozcan los detalles del método de estratificación utilizado y el procedimiento que se aplicó para la selección de la muestra, a fin de estar en condiciones de calcular los factores de expansión así como para estimar el efecto del diseño sobre la varianza de los estimadores.

Sin embargo, la realidad parece indicar que son pocas las organizaciones que calculan errores de muestreo y el efecto de diseño para las distintas variables de interés casi nunca se conoce. En este sentido, los usuarios -las más de las veces con poco conocimiento sobre métodos de análisis de encuestas complejas- no le dan importancia a este hecho lo cual genera un círculo vicioso de recriminaciones y culpas por los posibles malos usos de la información.

De hecho, la única manera de sensibilizar a los usuarios sobre los riesgos que representa hacer caso omiso de las especificaciones del diseño de la muestra y las limitaciones de desagregación de los datos, es que las Oficinas Nacionales de Estadística publiquen de manera regular estimaciones de los errores de muestreo y del efecto de diseño, y además acompañen las bases de datos con la información necesaria para que los analistas especializados tengan la posibilidad de calcular la magnitud del error de estimación para

dominios de estudio específicos que no se consideraron entre los objetivos del diseño muestral.

Como es muy difícil que esta situación ideal se logre concretar en el corto plazo, tal vez sería suficiente incorporar en las bases de datos la mínima información necesaria para lograr aproximaciones al efecto de diseño.

Para el cálculo del efd_k se requiere al menos identificar a las unidades de primera etapa de selección (UPM). Es decir, para una encuesta estratificada se debiera disponer como mínimo de la información que permita ubicar el estrato y la UPM de procedencia de las observaciones. Así, en el caso de que sólo una UPM se haya seleccionado -como sucede en algunos diseños- es evidente que en este caso su identificación coincide con la del estrato.

Si no se conoce el estrato de procedencia el efd_k se puede calcular como si se tratara de un diseño de conglomerados en la medida que sea posible identificar el segmento al cual pertenecen las observaciones.

Asimismo, para un diseño de muestra estratificado se debe asegurar que al menos existan dos UPM's seleccionadas por estrato ya que en caso contrario no es posible estimar la varianza. En la situación de que sólo un conglomerado haya sido seleccionado por estrato, se sugiere formar pseudoestratos para lo cual el analista requiere contar con un buen conocimiento del diseño de muestra.

A fin de ejemplificar los procedimientos sugeridos, a continuación se presentan los resultados obtenidos al estimar errores y efectos de diseño para algunas variables de interés.

Los datos utilizados corresponden a una encuesta de empleo realizada por un país de la región, que tiene por objetivo central estimar la tasa de desocupación abierta (TDA) respecto a la población económicamente activa (PEA). Las unidades de observación se seleccionaron a partir de un diseño estratificado (cuatro estratos) en tres fases. Para las dos primeras etapas, la probabilidad de selección fue proporcional a una medida de tamaño, mientras que las unidades de última etapa (segmentos de viviendas) se eligieron con muestreo aleatorio simple.

Para evaluar el efecto que tiene la selección multietápica y la estratificación de la unidades primarias en el error estándar del estimador se probaron las siguientes alternativas de diseño: i) En el primer caso (**D1**), se utilizaron las observaciones muestrales simulando que la información se obtuvo por muestreo aleatorio simple de elementos. ii) Otra opción (**D2**), fue utilizar las agrupaciones de viviendas de última etapa para estimar el error y efecto de diseño asumiendo un esquema aleatorio simple de conglomerados. iii) Se suprimió la información de la primera etapa de selección y sólo se consideraron las unidades de segunda y tercera fase (**D3**). iv) Por último, se trabajó con el diseño de muestra trietápico aplicado para la realización de la encuesta (**D4**).

En todos los casos, los cálculos se realizaron utilizando el PC CARP y se estimaron totales y proporciones (promedios en el caso del ingreso) a fin de analizar los cambios observados en el efecto de diseño como consecuencia de incorporar etapas adicionales en la selección de la muestra.

Un primer comentario que se debe realizar es que al igual que lo que se observa en muchas de las encuestas de empleo que se realizan en la región, el ingreso de las personas está subestimado ya que sólo corresponde a lo que declaran percibir los ocupados por concepto de sueldos y salarios del trabajo remunerado. Asimismo, se señala que para los cálculos se eliminaron aquellos registros en donde no se consignó ninguna respuesta(10.7%) (código de no especificado), pero si se consideraron los casos que reportaron cero ingresos(5.6%), lo cual es posible que incida en la calidad de los resultados.

En la tabla 10 se presentan los resultados obtenidos para los totales estimados de algunas variables de interés. En teoría, se esperaba que el efecto de diseño se incrementara para todas las variables en la medida que se incorporan subsecuentes etapas de selección de la muestra según lo obtenido en otras investigaciones similares (Verma *et. al. op. cit.*). En este sentido, al pasar de la opción D1 a D2 se registró un aumento en el efd_k , comportamiento que se mantuvo entre las opciones D2 y D3(a excepción de lo observado en el ingreso total). Por ejemplo, en la categoría ocupados el aumento fue del 32%, mientras que para el total de desocupados abiertos el incremento fue del 13.9%.

Sin embargo, esta tendencia creciente se revirtió al comparar los resultados de las opciones D3 y D4. Por ejemplo, para la categoría ocupados el efd_k disminuyó 9.3%, mientras que en el caso de los desempleados se redujo 6.4%.

En el caso del ingreso continuó el comportamiento errático que ya había sido identificado, a pesar de que el coeficiente de variación estimado para esta variable (2.48%) fue menor al observado para el total de desocupados el cual se ubicó en 5.0%.

Por su parte, los resultados obtenidos para las tasas y promedios evidencia un comportamiento más acorde con lo esperado. En efecto, en el caso de la tasa neta de participación (**TNP**) y la de desocupación abierta (**TDA**), el valor del efecto de diseño muestra una tendencia creciente en la medida que se incrementa en número de etapas para la selección de la muestra.

Cuadro 10

EFECTO DE DISEÑO PARA TOTALES DE VARIABLES DE INTERÉS

Variable	Estimador ¹ del Total	Efecto de Diseño D1.	Efecto de Diseño D2	Efecto de Diseño D3	Efecto de Diseño D4
Ocupados	5'363,210	1.0000	2.2292	2.9452	2.6715
Desocup. Abiertos	423,655	1.0000	1.3361	1.5213	1.4244
Ingreso	877,562 ²	1.0000	2.1590	1.6098	1.6509

¹ Los estimadores de totales fueron los mismos para ambos procedimientos.

² En millones de unidades.

De esta manera, y conforme a los datos que se presentan en la tabla 12, se observó un incremento del 45% en la varianza de la TNP por el hecho de estimar los datos a partir de un diseño de muestra bietápico en lugar de un esquema de conglomerados. Asimismo, se corrobora que una etapa más de selección no tiene una incidencia importante en la varianza del estimador, ya que el aumento registrado en el efd_k no alcanza ni un punto porcentual(0.16%).

Cuadro 11

EFECTO DE DISEÑO PARA TASAS DE INTERÉS

Variable	Estimador ¹ del Total	Efecto de Diseño D1.	Efecto de Diseño D2	Efecto de Diseño D3	Efecto de Diseño D4
Tasa Neta de Ocupación	48.61	1.0000	1.0126	1.4695	1.4719
Tasa de Desocupación Abierta	3.84	1.0000	1.2151	1.3274	1.3307
Ingreso Promedio	795.47 ²	1.0000	2.1921	1.4008	1.4022

¹ Los estimadores de totales fueron los mismos para ambos procedimientos.

² En unidades.

Por su parte, los resultados observados en la TDA muestran un comportamiento similar; sin embargo, en este caso el aumento en varianza por pasar de un diseño de conglomerados D2 a un esquema de selección bietápico D3 es del 9.2%, mientras que incorporar una etapa más de muestreo en el proceso de estimación (pasar de D3 a D4) sólo incrementa el error en 0.25%.

En el caso del ingreso promedio de los ocupados se mantuvo el comportamiento errático que se había evidenciado para el estimador del total, y en esta situación también se observó un error de muestreo inferior al registrado para la tasa de desocupación abierta; es decir, los coeficientes de variación reportados fueron 2.2% y 4.8% respectivamente.

Los efectos de diseño verdaderos, que corresponden al esquema que se utilizó para la selección de la muestra D4, son muy parecidos entre sí y muestran valores bastante aceptables lo cual haría suponer un nivel muy bajo de correlación de las observaciones entre conglomerados. De hecho, los valores del coeficiente de correlación intracase ρ fueron **0.1179**, **0.0827** y **0.1005** para la tasa neta de participación, tasa de desocupación abierta y el ingreso medio, respectivamente.

Cuadro 12

**INCREMENTO PORCENTUAL EN EL EFECTO DE DISEÑO
PARA TASAS DE INTERÉS**

Variable	Estimador ¹ del Total	D1	D3/D2	D4/D3	D4/D2
Tasa Neta de Ocupación	48.61	1.0000	45.12	0.16	45.35
Tasa de Desocupación Abierta	3.84	1.0000	9.24	0.25	9.51
Ingreso Promedio	795.47 ²	1.0000	(-36.1)	0.09	(-36.0)

¹ Los estimadores de totales fueron los mismos para ambos procedimientos.

² En unidades.

V. CONCLUSIONES

i) Cualquier resultado generado por medio de una encuesta está sujeto a un error de estimación, el cual se requiere conocer para determinar la confiabilidad estadística de los datos. Además, en la medida que se relajan las restricciones impuestas por el diseño y se desagreguen los datos, la probabilidad de que el número de observaciones resulte insuficiente aumenta y por lo tanto disminuye la precisión de los resultados.

ii) La estructura del diseño de muestra afecta la varianza de los estimadores. Esta situación se afirma debido a que mientras que el proceso de estratificación ayuda a disminuir la varianza, la conglomeración y las diferentes probabilidades de selección de las observaciones incrementa la variabilidad de los estimadores. De hecho, también contribuye a incrementar la varianza el tamaño promedio del conglomerado ya que guarda estrecha relación con el error estándar, por lo que todos estos elementos se deben controlar durante la etapa del diseño muestral.

iii) Prácticamente todas las encuestas que realizan las Oficinas Nacionales de Estadística de los países de la región corresponden a diseños de muestra complejos. En este sentido, se sugiere incorporar en las publicaciones de resultados estimaciones de los errores de muestreo y del efecto de diseño, a fin de que los usuarios conozcan la magnitud del error y confiabilidad de las estimaciones. Asimismo, se debe ser más enfático y advertir sobre los riesgos que se enfrentan por la utilización de los resultados de la encuesta, para fines diferentes a los que se definieron en los objetivos centrales de la investigación.

iv) Las encuestas de empleo, ingresos y gastos, demográficas y de niveles de vida, entre otras, tienen restricciones naturales en su capacidad explicativa las cuales están impuestas por los objetivos de las mismas, así como por la variable de diseño y los tamaños de muestra utilizados. En este sentido, se recomienda a los analistas de información realizar una evaluación estadística de los resultados y elegir métodos de análisis que consideren entre sus rutinas de cálculo las limitaciones que impone el diseño de la muestra, como una etapa previa a la formulación de modelos de causalidad, conclusiones y recomendaciones de acciones de política.

v) Si una encuesta de propósitos múltiples no se diseñó de manera adecuada en relación con la aplicación de un procedimiento de estratificación y conglomeración robusto, así como la determinación del tamaño óptimo de muestra, es muy probable que la confiabilidad de los datos no sea la adecuada por lo que en el análisis de los resultados se debieran asumir las precauciones necesarias.

vi) Los objetivos de la investigación, el tamaño de muestra y el esquema de selección le imponen restricciones a los datos que deben ser consideradas durante la etapa de análisis de la información. Así, en caso de que no se respeten los aspectos señalados existe el riesgo de que la inferencia que se realice carezca de significancia estadística.

vii) Los usuarios de los datos deben enfrentar la etapa de análisis con prudencia y hacerse asesorar por personas que conozcan el diseño de la muestra y que manejen los conceptos de la inferencia estadística en el caso de encuestas complejas. De lo contrario, es probable que algunas de las formulaciones de política que recomienden carezcan de confiabilidad estadística.

viii) El no considerar el diseño de la muestra en el análisis de los resultados, significa que se subestima el error de muestreo y se obtiene intervalos de confianza imprecisos que pueden llevar a asumir conclusiones erróneas respecto a los estadísticos de prueba que se utilizan para contrastar la significancia estadística de parámetros de interés, así como de los que se obtienen a partir del ajuste de un modelo econométrico.

REFERENCIAS BIBLIOGRÁFICAS

- Angers, C. (1979). "Sample Size Estimation for Multinomial Populations". *The American Statistician*. Vol. 33, No. 3, 163-164.
- Bankier, M. D. (1988). "Power Allocation: Determining Sample Size for Subnational Areas". *American Statistician*, vol. 42, No. 3, 174-177.
- CEPAL, División de Estadística y Proyecciones Económicas (1988). "Tamaño Óptimo de Muestra en Encuestas de Propósitos Múltiples". (documento de trabajo por publicar).
- Cochran, W. G. (1953). "Sampling Technics". John Wiley & Sons, Inc.
- Fuller, W. A. (1975). "Regression Analysis for Sample Surveys". *Sankhya*, 37 (Series C), 117-132.
- Holt, D. and T. M. F. Smith (1979). "Post Stratification". *Journal of the Royal Statistical Society* 142 (Series A), 33-46.
- Holt, D., T. M. F. Smith and P. D. Winter (1980). "Regression Analysis of Data from Complex Surveys". *Journal of the Royal Statistical Society*, 143 (Series A), 174-487.
- Kish, L. and Frankel, M. R. (1974). "Inference from Complex Samples" (with Discussion). *Journal of the Royal Statistical Society*. (Series B), vol. 36, 1-37.
- Kish, L. (1979). "Muestreo de Encuestas". Editorial Trillas, México, D.F..
- Kakwani, N. (1990). "Testing for the Significance of Poverty Differences". World Bank. Living Standard Measurement Study, Working Papers No. 62.
- Kokan, A. R. (1963). "Optimum Allocation in Multivariate Surveys". *Journal of the Royal Statistical Society* (Series A), vol. 126, 557-565.
- Kokan, A. R. and Khan, S. (1967). "Optimum Allocation in Multivariate Surveys: An Analytical Solution". *Journal of the Royal Statistical Society* (Series B), vol. 29. 115-125.
- Skinner, C. J. (1989). "Analysis of Complex Surveys". Chapter 2, Introduction to Part A., Edited by C. J. Skinner, D. Holt and T. M. F. Smith. John Wiley and Sons Ltd., 23-58.
- Sul, E., N. Forthofer, R. and J. Loromier, R. (1994). "Analyzing Complex Survey Data". Sage University Papers on Quantitative Applications in the Social Science, Num. 71, Beverly Hills: Sage Pubns.

- Thompson, S. K. (1987). "Sample Size for Estimating Proportions". *The American Statistician*, vol. 41, No. 1, 42-46.
- Tortora, R. D. (1978). "A Note on Sample Size Estimation for Multinomial Populations". *The American Statistician*, vol. 32, No. 3, 100-102.
- Verma, V., Scott, C. and O'Muircheartaigh (1980). "Sample Designs and Sampling Errors for the World Fertility Survey". (With Discussion). *Journal of the Royal Statistical Society (Series A)*, vol. 143, 431-473.

LISTA DE PARTICIPANTES

PROGRAMA PARA EL MEJORAMIENTO DE LAS ENCUESTAS Y LA MEDICION DE LAS
CONDICIONES DE VIDA EN AMERICA LATINA Y EL CARIBE (MECOVI)
BID - BANCO MUNDIAL - CEPAL

**TALLER REGIONAL SOBRE PLANIFICACIÓN Y DESARROLLO DE ENCUESTAS DE
HOGARES PARA LA MEDICIÓN DE LAS CONDICIONES DE VIDA**

Aguascalientes, México, 1 al 3 de abril de 1998

INEGI

CEPAL

Lista de Participantes

ARGENTINA

CLYDE ELISA CHARRE DE TRABUCHI

Directora de Encuestas a Hogares
Instituto Nacional de Estadística y Censos (INDEC)
Pte. Julio A. Roca 609, Buenos Aires - Argentina
fono: 541-3499214 / 54-1-3499212
fax: 541-3499218
E-mail: ctrib@indec.mecon.ar
Dirección Part.: French 2740
fono: 54-1-8242315

IRENE NOVACOVSKY

Coordinadora del SIEMPRO
Secretaría de Desarrollo Social de la Nación
Av. 9 de julio 1925 piso 13 (1332), Buenos Aires - Argentina
fono: 541-3846155
fax: 541-3832079 / 3836764
E-mail: siempro@datamarkets.com.ar
Dirección Part.: Enrique Martínez 205
Buenos Aires

PABLO PERELMAN

Gerente de Información Social del SIEMPRO (Economista)
Secretaría de Desarrollo Social de la Nación
Av. 9 de julio 1925 piso 13 (1332), Buenos Aires - Argentina
fono: 541-3846155
fax: 541-3832079 / 3836764
E-mail: siempro@datamarkets.com.ar
Dirección Part.: Humboldt 2481 6°22
Buenos Aires

BOLIVIA**WALTER CASTILLO GUERRA**

Jefe de Departamento de Indicadores Sociales

Instituto Nacional de Estadística

Plaza Mario Guzmán Aspiazu N°1, La Paz - Bolivia

fono: 591-2-392205 / 392693

fax: 591-2-342923

E-mail: wcastil@ine.gov.bo

Dirección Part.: Avenida Tejada Sorzano N° 913, 5° piso, La Paz

fono: 591-2-241051

BRASIL**ROSÂNGELA ANTUNES FERREIRA ALMEIDA**

Chefe da Divisão de Estudos e Análises do depto, de emprego e rendimento/Estadística

Instituto Brasileiro de Geografia e Estadística (IBGE)

Av. Chile 500 6° andar, Ríó de Janeiro - Brasil

fono: 021- 5144525

fax: 021-2626570

E-mail: rosangel@ibge.gov.br

Dirección Part.: R. Antonio Basilio 201/601 Tijuca

Ríó de Janeiro

CEP. 20511-190

MARCIA MARÍA QUINTSLR M.

Chefe do Departamento de Índices de Preços - DESIP

Instituto Brasileiro de Geografia e Estadística IBGE

Av. Chile, 500 6° andar, Ríó de Janeiro - Brasil,

fono: 021-5144526

fax: 021-2626670

E-mail: marciaq@ibge.gov.br

Dirección Part.: Rua Brão de Mesquita 600/301 - Tijuca - Cep 20540000, Ríó de Janeiro**ALFONSO EDUARDO RODRÍGUEZ ARIAS**

Consultor, Economista

Directoria de Política Social

Instituto de Pesquisa Econômica Aplicada (IPEA-Brasilia)

Edif. BNDES 14° andar. Setor Bancário Sul, Brasília, D.F.

fono: 061-3155275

fax: 061-3155400

E-mail: alicem@ipea.gov.br

Dirección Part.: SQN 116 Bloco Dapt. 502

Asa Norte. CEP 70 773-040 Brasília, DF, Brasil

CHILE**JÓSE LUIS LOBOS AGUIRRE**

Subdirector de Operaciones
Instituto Nacional de Estadísticas
Avda. Presidente Bulnes 418

Santiago

fono: 3667730

fax: 3690056

E-mail: inesdop@ctcreuna.cl

Dirección Part.: Mayflower N° 2386

Santiago

fono: 2775558

TEOBALDO ROMÁN URTUBIA

Jefe Departamento Estadísticas Laborales Máster en Estadística
Instituto Nacional de Estadísticas
Avda. Presidente Bulnes 418

Santiago

fono: 3667600

fax: 3690056

E-mail: inelab@ctcreuna.cl

Dirección Part.: Huérfanos 2308 Santiago

fono: 6713512

BERTA RIVSCA TEITELBOIM GRINBLATT

Jefa Departamento Información Social
Ministerio de Planificación y Cooperación MIDEPLAN
Ahumada N° 48, Piso 5

fono: 6984771 - 6751540

fax: 6729027

E-mail: bteitelb@mideplan.cl

Dirección Part.: Vicente Pérez Rosales N° 1356-I, La Reina,

Santiago

Fono: 6729027

ROBERTO URMENETA

Subdirector de Operaciones
Servicio de Registro Civil e Identificación
Huérfanos 1570 3er piso, Santiago

fono: 56-2-6984766

fax: 56-2-2236292

E-mail: rurmenet@netline.cl

Dirección Part.: Hindenburg 660 Providencia

Santiago

COLOMBIA**CIRO MARTÍNEZ GÓMEZ**

Director Técnico de Censos, Economista - Demógrafo

Departamento Nacional de Estadística,

DANE - CAN- Avda. El Dorado

Bogotá - Colombia

fono: 2223178

fax: 2221845

E-mail: dane6@impsat.net.co

Dirección Part.: Cra 52 N° 88-73 Barrio Entrerrios

Bogotá

COSTA RICA**MARÍA ELENA GONZÁLEZ QUESADA**

Coordinadora Unidad Técnica de Programación, Operaciones Censales y Encuestas, Licenciado en Estadística

Ministerio de Economía, Industria y Comercio - Área de Estadística y Censos

Avenida 4, Calle 0 y 2 Edificio Rex, tercer piso

San José - Costa Rica - C.A.

fono: 506-2219656

fax: 506-2230813

E-mail Directora: dgec@meic.gov.cr

Dirección Part.: 50 este y 100 norte de la Facultad de Agronomía

Universidad de Costa Rica, San Pedro de Montes de Oca

San José

fono: 2245159

CUBA**MARÍA DEL CARMEN ZAMORA HERNÁNDEZ**

Estadística Global Principal, Economista

Oficina Nacional de Estadísticas (ONE)

Paseo N° 60 Esquina A 5^{ta} Vedado

Ciudad de la Habana - Cuba

fono: 309041

fax: 537-333083

E-mail: mzamora@onered.one.gov.cu

Dirección Part.: Calle 35 N° 18 Apto. 1 entre 2 y paseo

Plaza de la Revolución

fono: 706071

Ciudad de la Habana

ECUADOR**GALO ARIAS VELOZ**

Subdirector Técnico del INEC, Economista

Instituto Nacional de Estadísticas y Censos

Juan Larrea N° 534 y Riofrio

Quito - Ecuador

fono: 529858 / 544326 / 544561/231602

fax: 509836

Dirección Part.: Magnolias N° 46-165

Quito - Ecuador

GERMÁN EFRAÍN FLORES ESCOBAR

Director de Estadística Sociodemográficas y Jefe - Ecv. III Ronda - 1998/ Ingeniero

Instituto Nacional de Estadísticas y Censos

Juan Larrea N° 534 y Riofrio

Quito - Ecuador

fono: 529858 / 544326 / 544561/231602

fax: 509836

Dirección Part.: Quito/Urb. Condado # 425-c

fono: 494751

EL SALVADOR**SAÚL FRANCISCO GARCÍA CARBALLO**

Jefe de División de Información Social

Dirección General de Estadísticas y Censos

Av. Juan Bertis N° 79, Ciudad Delgado

San Salvador - El Salvador

fono: 2765900

fax: 2765900 ext. 129

Dirección Part.: Urbanización Metropolis,

Paseo Las Margaritas N° 59

San Salvador

DANIEL FLORES DE PAZ

Técnico Asesor IV

Dirección General de Estadísticas y Censos

Av. Juan Bertis N° 79, Ciudad Delgado

San Salvador - El Salvador

fono: 2765900

fax: 2765900 ext. 129

Dirección Part.: Final 3ª C.O. Urb. Trinidad Pje. 2 N°75

Cuscatancingo, San Salvador

GUATEMALA**MARIO JACOBS**

Gerente, Ingeniero Industrial

Instituto Nacional de Estadística (INE)

8va Calle 9-55 zona 1 Edificio América

Ciudad de Guatemala - Guatemala

fono: 502-2- 2326136 /502-2- 2382587

fax: 502-2-2324790

E-mail: ine_g@rds.org.gt

Dirección Part.: 38 Avenida 11-69 Zona 7, El Rodeo La Floresta,
Ciudad de Guatemala

FRANCISCO GUILLERMO FLORES RÁMILA

Director de Producción y Difusión Estadística, Economista

Instituto Nacional de Estadística

8va Calle 9-55 zona 1 Edificio América

Ciudad de Guatemala - Guatemala

fono: 502-2- 2535687 /502-2- 2382587

fax: 502-2-2324790

E-mail: ine_g@rds.org.gt

Dirección Part.: 11 av. "A" 12-68 zona 2,
Guatemala- Guatemala

HONDURAS**ROSA IDALIA MENDOZA**

Jefe de Programa de Encuesta Permanente de Hogares de Propósitos Múltiples, Ms. En
Población y Desarrollo

Dirección General de Estadística y Censos

6° avenida 8° calle

Comayagua - Honduras

fono: 2228450

Dirección Part.: Tegucigalpa Colonia Hato de Enmedio Sector N° 10
Bloque N° 7 casa N°1

DOMINGO ALBERTO ORDOÑEZ GUZMÁN

Analista de Datos, Encuesta Permanente de Hogares de Propósitos Múltiples

Dirección General de Estadística y Censos

6° avenida 8° calle

Comayagua - Honduras

fono: 2228450

Dirección Part.: Colonia Hato de Enmedio Sector 7, Bloque 99
Peatonal 9, casa N° 5116,
Tegucigalpa

MÉXICO

MIGUEL CERVERA, Director General
Dirección General de Estadística, INEGI
Edificio Sede, Puerta 3, 1er nivel, Prol. Héroe de Nacozari N° 2301 Sur, Aguascalientes
fono: 491-81550/81735/81302/81546/81680
fax: 491-81487
E-mail: mcervera@dge.inegi.gob.mx

MARIO PALMA ROJO, Coordinador Administrativo
INEGI, México, Coordinación Administrativa
Edificio Sede, Pta 4 Niv. 1 Héroe de Nacozari N° 2301 Fracc. Jardines del Parque,
Aguascalientes, México
fono: 49-181342
fax: 49-181302

EDUARDO AMAYA ROJAS
Director de Planeación y Capacitación
INEGI
Centro de Capacitación del INEGI. Av. Héroe de Nacozari 2301 Sur, Fracc. Jardines del Parque
Aguascalientes, México
fono: 49-105461 / 166637
fax: 49-181280
E-mail: eamaya@cap.inegi.gob.mx

MARÍA GUADALUPE AGUILAR FRÍAS
Subdirectora de Características Sociodemográficas
INEGI, DGE Dirección del Censo de Población y Vivienda 2000
Prol. Héroe de Nacozari 2301 Pta. 1 1er nivel., Aguascalientes - México
fono: 149-181416
fax: 149-180396
Dirección Part.: Plan de Ayutla N° 308-3 colonia del Carmen, Aguascalientes
C.P. 20070
fono: 183808

TATIANA BARCENAS SANDOVAL
Profesional Dictaminador de Servicios Especializados
Instituto Nacional de Estadística Geografía e Informática (INEGI)
Torre el Dorado Av. Las Américas s/n
Fracc. El Dorado C.P. 20235, Aguascalientes - México
fono: 787912 / 105711
fax: 787959
E-mail: enighplifags.podernet.com.mx
Dirección Part.: Vivero del Valle N° 202
Fracc. Casa Blanca C.P. 20270, Aguascalientes

MARÍA ENRIQUETA CRUZ LARA
Prof. Dic. De Servicios Especializados
Instituto Nacional de Estadística Geografía e Informática (INEGI)
República Mexicana esq. República del Salvador S/N
Fracc. El Dorado C.P. 20235
Aguascalientes - México
fono: 787912 / 105711
fax: 787959
E-mail: enighplifags.podernet.com.mx
Dirección Part.: Av. Ojo caliente N° 1502
Fracc. Ojo caliente C.P. 20190
fono: 702854
Aguascalientes

TERESA GONZÁLEZ MELÉNDEZ
Prof. Dic. De Servicios Especializados
Instituto Nacional de Estadística Geografía e Informática (INEGI)
República Mexicana esq. República del Salvador S/N
Fracc. El Dorado C.P. 20235
Aguascalientes - México
fono: 787912 / 105711
fax: 787959
E-mail: enighplifags.podernet.com.mx
Dirección Part.: El Refugio N° 217
Fracc. Ojo caliente C.P. 20190
fono: 700378
Aguascalientes

SUSANA MEJIA MURILLO
Prof. Dic. De Servicios Especializados
Instituto Nacional de Estadística Geografía e Informática (INEGI)
República Mexicana esq. República del Salvador S/N
Fracc. El Dorado C.P. 20235
Aguascalientes - México
fono: 787912 / 105711
fax: 787959
E-mail: enighplifags.podernet.com.mx
Dirección Part.: Las Norias N° 234
Fracc. Ojocaliente C.P. 20190
fono: 700030
Aguascalientes

MIGUEL ANGEL ROJAS MINGUER

Prof. Ejec. De Servicios Especializados

Instituto Nacional de Estadística Geografía e Informática (INEGI)

Av. Héroe de Nacozari N° 2301

Aguascalientes - México

fono: 181678

fax : 181409

E-mail: mrojas@economi.inegi.gob.mx

Dirección Part.: El Novillo N° 101 E-3-304 Fracc. Ojo caliente, Aguascalientes

fono: 700776

NICARAGUA**ALVARO AGURTO ROBELO**

Analista de Programas, Ingeniero Agrónomo

Instituto Nacional de Estadísticas y Censos INEC-Nicaragua

Fte. Hosp. Lenín Fonseca Apado. 2031, Managua Nicaragua

fono: 2662825 / 2666178

fax: 2662031

Dirección Part.: Colonia Ciudad Xolotlaio, Casa N° 63, Managua

fono: 2444754

fax: 2897566

MARBEL GAMBOA DE ESPINOZA

Directora General de Empleo y Salarios, Economista

Ministerio del Trabajo (MITRAB.)

Estadio Nacional, 2 cuadras al Norte, Managua - Nicaragua

fono: 2223628

fax: 2223628

E-mail: mitrab@ibw.com.ni

Dirección Part.: Residencial "Las Brisas" Casa N° Q-32, Managua

fono: 2662080

PANAMÁ**LOIRA ELNISA AGRAZAL DE RUÍZ**

Coordinadora de Encuestas de Hogares Estadísticos I

Contraloría General de la República

Dirección de Estadísticas y Censos

Av. Balboa Apartado 5213 Panamá 5, Panamá

fono: 2643734 / 2640777 ext. 129

fax: 2697294

E-mail: cgrdec@keops.utp.ac.pa

Dirección Part.: Pedregal, San Joaquín casa E-1

Panamá

PARAGUAY

MARTA GRACIELA BECKER DUARTE, Directora General
Dirección General de Estadística, Encuestas y Censos
Naciones Unidas y Zenteno - Zona Norte (Fernando de la Mora)
Fernando de la Mora - Paraguay
fono: 511016 / 510867 / 506988
fax: 508493
E-mail: dgec@sce.cnc.una.py
Dirección Part.: Tte. Fretes N°762 c/Bruno Guggiari
fono: 551331

OSCAR SALVADOR BARRIOS LEIVA
Jefe Departamento de Encuestas, Economista
Dirección General de Estadística, Encuestas y Censos
Naciones Unidas y Zenteno - Zona Norte (Fernando de la Mora)
Fernando de la Mora - Paraguay
fono: 511016 / 510867
fax: 508493
E-mail: dgec@sce.cnc.una.py
Dirección Part.: San José c/María F. González, Fernando de la Mora, Zona Sur

ALBERTO RAMÓN DUARTE ARIAS
Coordinador de Campo, Departamento de Encuestas
Dirección General de Estadística, Encuestas y Censos
Naciones Unidas y Zenteno - Zona Norte (Fernando de la Mora)
Fernando de la Mora - Paraguay
fono: 511016 / 510867
fax: 508493
E-mail: dgec@sce.cnc.una.py
Dirección Part.: Artillería N°356 c/14 de mayo, Fernando de la Mora
fono: 515545

PERÚ

ALBERTO PADILLA TREJO
Coordinador Ejecutivo del Programa MECOVI - PERÚ, Licenciado en Estadística
Instituto Nacional de Estadística e Informática (INEI-PERÚ)
Av. General Garzón N°658 Jesús María, Lima - Perú
fono: 51-1-4334223 anexo 316
fax: 51-1-4333159
E-mail: apadilla@inei.gob.pe
Dirección Part.: José Urdanivia 409 San Luis, Lima - Perú
E-mail: padilla@telematic.edu.pe
fax: 51-1-4379470

REPÚBLICA DOMINICANA**SUSANA GAMEZ SEOANE**Consultor Económico, Departamento Cuentas Nacionales y Estadísticas Económicas/Economista
Banco Central de la República Dominicana

Pedro Henríquez Ureña esq. Leopoldo Navarro, Santo Domingo - República Dominicana

fono: 2219111 ext. 2038 / 6850783

fax: 6827666

E-mail: BANCEN3@CODETEL.NET.DO

Dirección Part.: Fantino Falco N° 77 Res Remi. Apart 3A Piatini**URUGUAY****MARTHA RODRÍGUEZ GÓMEZ**

Administrativa - Especializada

Instituto Nacional de Estadística

Río Negro 1520, Montevideo - Uruguay

fono: 9027303

fax: 9014653

E-mail: martrodr@ine.gub.uy

Dirección Part.: Constitución 2365 Apto. 5

fono: 2008026

VENEZUELA**GUILLERMO JOSÉ RAMIREZ**

Director General / Estadístico

Oficina Central de Estadística e Informática

Avda. Boyaca, Edif. "Fundación la Salle", 4^{to} piso, Mariperez, Caracas - Venezuela

fono: 582-7937191

fax: 582-7829755

E-mail: gramirez@platino.gov.ve

Dirección Part.: Calle Géminis, Res. "Mayurupi", piso 7, apto. 7C

Santa Paula, El Cafetal, Caracas

RAMÓN GERARDO MALDONADO

Jefe División de Desarrollo Social / Economista

Ministerio de la Familia

Av. Lecuma. Torre Oeste, Parque Central Núcleo 4 piso 41

Caracas - Venezuela

fono: 5740576 / 5766067

fax: 5746879

E-mail: análisis@eldish.net

Dirección Part.: Av. Sucre Residencias Naiguata. Torre C. Piso 16, Apto. 16-5

Caracas

CEPAL

JUAN CARLOS FERES

Coordinador MECOVI - CEPAL

División de Estadística y Proyecciones Económicas

CEPAL

Av. Dag Hammarskjold S/N

Vitacura, Santiago, Chile

fono: 56-2-2102408

fax: 56-2-2102472

E-mail: jferes@eclac.cl

FERNANDO MEDINA

Asesor Regional

División de Estadística y Proyecciones Económicas

CEPAL

Av. Dag Hammarskjold S/N

Vitacura, Santiago, Chile

fono: 56-2-2102417

fax: 56-2-2102472

E-mail: fmedina@eclac.cl

MARÍA DE LA LUZ AVENDAÑO

Coordinadora Banco de Datos de Encuestas de Hogares

División de Estadística y Proyecciones Económicas

CEPAL

Av. Dag Hammarskjold S/N

Vitacura, Santiago, Chile

fono: 56-2-2102441

fax: 56-2-2102472

E-mail: mavendaño@eclac.cl

SERGIO SOMERVILLE

Asesor Gobierno de Chile

Alameda 1351

fono: 6953351- 1829

CEPAL

E-mail: somerville@interactiva.cl

Dirección Part.: Las Compuertas 0236

fono: 8711917

Santiago

BID

GILBERTO MONCADA

Coordinador MECOVI - BID

1300 New York Avenue, N.W. Washington, D.C. 20577

fono: 202-6232716

fax: 202-6233299

E-mail: gilbertom@iadb.org

Dirección Part.: 10701 Muirfield DR. Potomac, Maryland 20854 USA

BANCO MUNDIAL

HAEDUCK LEE

Coordinador del Proyecto MECOVI del Banco Mundial / Economista

1818 H St., N.W., Washington, D.C. 20433, USA

fono: 202-4736687

fax: 202-5222119

E-mail: hlee1@worldbank.org

KINNON SCOTT, Especialista en Encuestas, Medición de Pobreza

1818 H Street, N.W., Washington D.C. 20433, USA

fono: 202-4738124

fax: 202-5221153

E-mail: kscott1@worldbank.org

MARGARET GROSH, Economista

1818 H Street, N.W., Washington D.C. 20433, USA

fono: 202-4739249

fax: 202-5221153

E-mail: mgrosh@worldbank.org

PNUD

ANABELLE GUERRERO, Consultor en Evaluación

Programa de Naciones Unidas para el Desarrollo PNUD

UNPD. One United Nations Plaza Room N° DC1-2278

Nueva York, NY 10017

U.S.A.

fono: 212-9065416

fax: 212-9065892

E-mail: guerrero@undp.org.

Dirección Part.: 215 West 105 St. N° 4d

NY 1002

INEGI (CAPACITACIÓN)-ORGANIZACIÓN DEL EVENTO

GABRIEL ALCOLEA HERRERA

Subdirector de Capacitación Técnica en Estadística

Instituto Nacional de Estadística, Geografía e Informática (INEGI)

Centro de Capacitación INEGI. Av. Héroe de Nacozari 2301 Sur, Fracc. Jardines del Parque,
C.P. 20270, Aguascalientes, México

fono: 49-105462 / 166637

fax: 49-181280

E-mail: galcolea@cap.inegi.gob.mx

JACKELINE MARTÍNEZ PONCE

Subdirectora de Aplicaciones Estadísticas

Instituto Nacional de Estadística, Geografía e Informática (INEGI)

Centro de Capacitación INEGI. Av. Héroe de Nacozari 2301 Sur, Fracc. Jardines del Parque,
C.P. 20270, Aguascalientes, México

fono: 49-105462 / 166637

fax: 49-181280

E-mail: jmartinez@cap.inegi.gob.mx

GASPAR HERNÁNDEZ NOYOLA

Jefe del Departamento de Capacitación en Estadística

Instituto Nacional de Estadística, Geografía e Informática (INEGI)

Centro de Capacitación INEGI. Av. Héroe de Nacozari 2301 Sur, Fracc. Jardines del Parque,
C.P. 20270, Aguascalientes, México

fono: 49-105462 / 166637

fax: 49-181280

E-mail: ghernandez@cap.inegi.gob.mx



