

ANEXO 2. DISEÑO DE LA MUESTRA DEL ESTUDIO DE REFERENCIA

Estudio Nacional de Consumo de Sustancias Psicoactivas
en Colombia

DISEÑO DE LA MUESTRA

Luis Carlos Gómez
Consultor Muestreo Probabilístico

Bogotá, Marzo 20 de 2008

CONTENIDO

1. PRESENTACION

2. PREMISAS DEL DISEÑO MUESTRAL

2.1 Universo de estudio

2.2 Indicadores por ser estimados

2.3 Tamaño de la muestra financiable, desagregación factible de resultados y precisión esperada de estimaciones

3. ELEMENTOS DEL DISEÑO MUESTRAL

3.1 Alcance del diseño

3.2 Tipo de Muestra

3.3 Metodología de Selección de la muestra por etapas

3.3.1 Selección de UPM

3.3.2 Selección de USM

3.3.3 Selección de la UTM

3.3.4 Selección de una Persona 12 a 64 por Hogar

3.4 Procedimientos de Estimación

3.5 Modelo para Cálculo de la Precisión Observada

4. BIBLIOGRAFIA SELECCIONADA

ANEXO: Listado de Unidades Primarias de Muestreo Seleccionadas

DISEÑO DE LA MUESTRA

Luis Carlos Gómez
Consultor Muestreo Probabilístico
Marzo 20/08

1. PRESENTACIÓN.

Se describe a continuación el Diseño de la Muestra Probabilística de Hogares para el Estudio Nacional de Consumo de Sustancias Psicoactivas en Colombia, patrocinado por la Dirección Nacional de Estupefacientes (DNE) del Ministerio del Interior y de Justicia y por el Ministerio de la Protección Social, con el apoyo de la Oficina de las Naciones Unidas contra la Droga y el Delito (UNODOC) y de la Comisión Interamericana para el Control del Abuso de Drogas (CICAD) de la OEA.

El documento incluye en su orden las premisas que condicionan el diseño muestral y los elementos del diseño propiamente tal. Las premisas surgieron de los objetivos y alcances de la investigación, de los términos de referencia establecidos para este trabajo y de las puntualizaciones complementarias acordadas en las reuniones con los representantes de las instituciones promotoras del estudio. Los elementos del diseño comprenden el tipo de muestra, la metodología de selección, los procedimientos de estimación de los resultados y el modelo para el cálculo de la precisión efectivamente observada. Al final, el documento presenta las referencias bibliográficas más importantes y un Anexo con el listado de las Unidades Primarias de Muestreo seleccionadas

El trabajo fue desarrollado por el Estadístico Muestrista Luis Carlos Gómez, quien tuvo el apoyo en la implementación de la selección del Estadístico Jhon Jairo Romero, mención especial amerita la excelente colaboración del DANE mediante el suministro de las bases de datos del VIHOPE del Censo General 2005-06 y de la cartografía digital de dicho censo.

2. PREMISAS DEL DISEÑO MUESTRAL

Se refieren a la delimitación geográfica, demográfica y socioeconómica del universo de estudio; la naturaleza estadística de los indicadores por ser estimados; el tamaño de la muestra financiable; la desagregación deseada y factible de los resultados para análisis, comparación y presentación; y la precisión esperada de las estimaciones.

2.1 Universo de Estudio

Está constituido por la población de 12 a 65 años de hogares particulares, de todos los estratos socioeconómicos, residentes en las capitales departamentales y en las de municipios no capitales con 30,000 y más habitantes. También se incluyeron municipios de menor tamaño que forman parte de las grandes áreas metropolitanas

Según el Censo General 2005, los 126 municipios que configuran este universo tenían un total de 26.048.000 personas de 12 a 64 años.

2.2 Indicadores por ser estimados.

Se prevé, básicamente, la estimación rutinaria de indicadores de relación por cociente (proporciones, tasas, razones), al igual que tasas de diferencia en un indicador entre subgrupos geográficos y demográficos en un momento dado, o tasas de cambio en el tiempo de indicadores calculados en encuestas anteriores.

2.3 Tamaño de la muestra financiable, desagregación factible de resultados y precisión esperada de estimaciones.

Se decidió una muestra de 39,000 hogares, de los cuales se seleccionará una persona de 12 a 64 años por hogar. Se espera una muestra mínima efectiva de 30,000 hogares. Se prevén estimaciones globales para 27 particiones geográficas, las cuales incluyen los 22 departamentos más importantes, el conjunto de la Orinoquia y la Amazonia, las áreas metropolitanas de Bogotá, Medellín y Cali, y las Islas de San Andrés y Providencia; para análisis más desagregados por tipo de sustancia psicoactiva de alto consumo y por grupos del resto de sustancias, es factible una desagregación por 12 subregiones; para estimaciones por todos los tipos de sustancias, o por grupos de sustancias, y grupos demográficos o socioeconómicos, la agregación de la muestra en 6 grandes regiones permite esta posibilidad; y finalmente, para análisis detallados por sustancia, y simultáneamente por subgrupos demográficos y socioeconómicos, las estimaciones serían para el conjunto del país (ver Cuadro 1).

La precisión esperada de los resultados, en términos de Error Estándar Relativo (ES rel), para los diferentes grados de desagregación de la muestra y para prevalencias de distinta magnitud, se observa en el Cuadro 2. El error estándar es el indicador de variabilidad del azar intrínseca al muestreo probabilístico. El ESrel es el error estándar absoluto expresado como porcentaje del valor del indicador estimado puntualmente ((Esrel = ES / P)*100).

Fórmula utilizada

Es del Muestreo Aleatorio Simple de Elementos, Ajustada para Muestreo de Conglomerados.

$$ES_{rel} = \sqrt{\left(1 - \frac{n}{N}\right) \frac{PQ}{n} * deff} / P$$

En donde,

ES_{rel} = Error Estándar Relativo

n = tamaño de la muestra de una partición específica

N = Tamaño del universo de dicha partición

P = Proporción o prevalencia esperada, en valor absoluto

$Q = 1 - P$

$deff$ = Efecto de los conglomerados en el diseño. Se espera un $deff = 1.2$

$$= \frac{\text{Varianza de Muestreo de Conglomerados}}{\text{Varianza de Muestreo Aleatorio Simple de Elementos}}$$

$$= \frac{1/x^2 [\text{var}(y) + r^2 \text{var}(x) - 2r \text{cov}(y, x)]}{\left(1 - \frac{n}{N}\right) \frac{PQ}{n}} \quad (\text{ver 3.5})$$

A su vez

$$ES_{rel} = \frac{ES_p}{P} = \frac{\text{Error Estándar Absoluto}}{\text{Prevalencia o Proporción}} * 100$$

$$ES_p = ES_{rel} * P$$

Cuadro 1: Distribución muestra mínima efectiva de hogares

REGION	DEPTO	No. Munpio	Población (miles)	Muestra Hogares		
				Depto y áreas m.	Sub regiones	Regiones
Atlántico	Guajira	2	201	700	2600	7400
	Cesar	3	401	900		
	Magd	5	589	1000		
			1191			
	Atlánt.	5	1764	1300	3100	
	Bol.	5	1082	1100		
	S.Andres	2	43	700		
			2889			
	Sucre	3	292	700	1700	
	Córdoba	7	546	1000		
		838				
		32	4918			
Oriental	Nde Sant	5	826	1100	2200	6000
	Santander	6	1160	1100		
			1986			
	Boy.	5	413	900		
	Cund.	9	993	1100	2000	
			1406			
	Meta	3	448	900		
	Orin.-Amaz	10	422	900		
			870		1800	
		38	4262			
Bogotá	Bogotá	1	6763	4000	4000	4000
Central	Caldas	4	485	900	2700	4500
	Risar.	4	607	1000		
	Quindío	4	381	800		
			1473			
	Tolima	2	521	1000	1800	
	Huila	3	389	800		
			910			
		17	2383			
Antioquia	Med.Metro	10	3137	2200	3100	3100
	Resto	9	486	900		
		19	3623			
Occidental	Cali Metro	2	2120	1500	2600	5000
	Resto valle	10	1097	1100		
			3217			
	Chocó	1	100	700		
	Cauca	3	308	800	2400	
	Nariño	3	474	900		
			882			
		19	4099			
Total		126	20048			30000

Cuadro 2. Precisión de la Muestra (ES rel:%) para Estimaciones Puntuales Tipo Proporción (P) (Deff=1.2)

Desa- grega- ción	Muest Efect. n	Proporción							
		0.01	0.03	0.05	0.10	0.20	0.30	0.40	0.50
		ES rel (%)							
Depto.	700	41.0	23.4	17.9	12.3	8.2	6.3	5.0	4.1
	800	38.5	22.0	16.8	11.6	7.7	5.9	4.7	3.9
	900	36.3	20.7	15.9	10.9	7.3	5.6	4.5	3.6
	1000	34.4	19.7	15.1	10.4	6.9	5.3	4.2	3.5
	1100	32.8	18.8	14.4	9.9	6.6	5.0	4.0	3.3
	1300	30.2	17.3	13.2	9.1	6.1	4.6	3.7	3.0
	1500	28.1	16.1	12.3	8.5	5.7	4.3	3.5	2.8
	2200	23.2	13.3	10.2	7.0	4.7	3.6	2.9	2.3
Sub- Región	1700	26.4	15.1	11.6	8.0	5.3	4.1	3.2	2.7
	1800	25.6	14.7	11.2	7.7	5.2	3.9	3.2	2.6
	2000	24.3	13.9	10.7	7.3	4.9	3.7	3.0	2.4
	2200	23.2	13.3	10.2	7.0	4.7	3.6	2.9	2.3
	2400	22.2	12.7	9.7	6.7	4.5	3.4	2.7	2.2
	2600	21.4	12.2	9.4	6.4	4.3	3.3	2.6	2.1
	2700	20.9	12.0	9.2	6.3	4.2	3.2	2.6	2.1
	3100	19.6	11.2	8.6	5.9	3.9	3.0	2.4	2.0
Región	4000	17.2	9.8	7.5	5.2	3.5	2.6	2.1	1.7
	4500	16.2	9.3	7.1	4.9	3.3	2.5	2.0	1.6
	5000	15.4	8.8	6.7	4.6	3.1	2.4	1.9	1.5
	6000	14.1	8.0	6.2	4.2	2.8	2.2	1.7	1.4
	7400	12.7	7.2	5.5	3.8	2.5	1.9	1.6	1.3
País	30000	6.3	3.6	2.8	1.9	1.3	1.0	0.8	0.6

De acuerdo con los resultados del Cuadro 2, el panorama de la precisión esperada de los resultados es ampliamente satisfactorio. Para las particiones departamentales y por grandes áreas metropolitanas, la precisión es buena ($ES_{rel} \leq 10\%$) para prevalencias de 0.10 o más. Y aceptables para prevalencias menores de 0.10, en el conjunto de la muestra.

Por subregión, la precisión es buena para prevalencias del 0.05 o más, en el conjunto de la muestra, y son factibles por lo menos dos particiones (ej., sexo, edad, etc.) con aceptable precisión.

Por región, se pueden estimar prevalencias de 0.030 o más con buena precisión, en la muestra total, y son factibles hasta 4 o 5 particiones por subgrupos demográficos o socioeconómicos.

Para el conjunto del país, se pueden estimar prevalencias de cualquier magnitud con excelente precisión, en la muestra total, y son posibles múltiples particiones demográficas, socioeconómicas y por nivel de urbanización, con buena precisión.

3. ELEMENTOS DEL DISEÑO MUESTRAL

3.1 Alcance del diseño

Se propone una Muestra Maestra de 117,000 hogares, constituida por tres submuestras probabilísticas idénticas de 39,000 hogares, las cuales permitirán la realización de tres encuestas nacionales de similares características, sin traslape para evitar la eventual contaminación de resultados. En el estudio que se llevará a cabo en las próximas semanas, se utilizará una de las submuestras. Los municipios serán los mismos, y los segmentos de hogares de las tres submuestras serán contiguos. Con ocasión del primero de los estudios, la empresa encuestadora deberá submuestrear y delimitar en el terreno los segmentos de viviendas (contiguas) de las tres submuestras, uno de los cuales será el encuestado.

3.2 Tipo de Muestra

El diseño es probabilístico, de conglomerados y tetraetápico

Probabilístico, porque cada persona del universo estudiado tendrá una probabilidad de selección conocida y superior a cero. En este tipo de muestra es factible una inferencia válida del universo de estudio, su nivel de precisión puede fijarse anticipadamente en forma aproximada, y puede calcularse la precisión observada para cada indicador estimado. El error estándar es el indicador de precisión de las muestras probabilísticas.

El diseño es de probabilidades desiguales, lo cual obliga a la aplicación de un factor de corrección que permita generar estimaciones insesgadas.

De conglomerados. Los municipios, que son los conglomerados de primer nivel, no serán muestreados, y todos serán visitados. Dentro de los municipios los conglomerados que son objeto de muestreo son las manzanas de 20 o más hogares. Las manzanas con menos hogares se anexaron a una manzana vecina para constituir una unidad de muestreo. Cada conglomerado seleccionado será dividido en partes con aproximadamente el mismo número de viviendas (8-10). Tres de estas partes serán los segmentos de la muestra maestra.

El muestreo por conglomerados permite la concentración de la muestra en unos cuantos de estos, con enorme ganancia en los costos y precisión aceptable. La cantidad y tamaño de los conglomerados de la muestra es función del tamaño de muestra de unidades de observación, de la varianza interna (correlación intraclásica) de los fenómenos estudiados y de la varianza interconglomerado. A mayor correlación intraclásica, menor la cantidad de conglomerados de tamaño grande que pueden ser seleccionados, y viceversa, si la correlación intraclásica es baja.

Tetraetápica.

Cuatro etapas tiene el proceso de selección dentro de un municipio. a) Selección de las manzanas y grupos de manzanas (Unidades Primarias de Muestreo, UPM) de 20 o más hogares; b) si la UPM tiene 45 o más hogares, selección de un área (Unidad Secundaria de Muestreo, USM) con un tamaño promedio de 30 hogares (26 viviendas aproximadamente); c) selección de un segmento (Unidad Terciaria de Muestreo, UTM), con un promedio de 10 hogares (8 viviendas aproximadamente); y d) selección de una persona (elemento) de 12 a 64 años, por hogar.

Las manzanas de menos de 20 hogares debieron ser anexadas a una manzana vecina, para completar en conjunto 20 o más hogares, y participar en la selección como unidad primaria.

Cada etapa tiene su probabilidad de selección. El producto de las probabilidades de las varias etapas, es la probabilidad final de cada persona de la muestra.

3.3 Metodología de Selección de la muestra por etapas.:

3.3.1 Selección de UPM

La selección de las UPM en cada municipio se hizo sobre la base de datos VIHOPE (viviendas, hogares, personas) del Censo General 2005-06, la cual tiene un ordenamiento geográfico continuo. En total, se seleccionaron 3900 UPM, con la distribución señalada en el Cuadro 1 por departamento, subregión y región.

La distribución de la muestra de UPM en las ciudades por ser estudiadas en cada departamento, se realizó proporcional a la población de 12 a 64 años.

El procedimiento de selección dentro de un municipio fue el siguiente:

- Cálculo para cada manzana registrada en el VIHOPE del número de medidas de tamaño (MT), de 30 hogares, existentes, así:

N° de Hogares	N° de MT
20 – 44	1
45 - 74	2
75 – 104	3
105 – 134	4
etc.	etc.

Las manzanas con menos de 20 hogares, se anexaron a una contigua.

- Cálculo de Frecuencia Acumulada de MT para el municipio.

- Cálculo de Intervalo de Muestreo (IM), dividiendo el total de MT existentes en el municipio (cabecera), por el número de UPM a ser seleccionadas (cálculo con un decimal).
- Selección de un número al azar (número de arranque) entre 1.0 y el valor del intervalo de muestreo (IM). El número seleccionado indica la primera UPM seleccionada.
- Selección del resto de UPM del municipio sumando sucesivamente el IM al número de arranque.

La distribución de la muestra de UPM según número de MT es la siguiente:

N° de MT	N° de UPM	%
1	1883	48.3
2	1075	27.5
3	533	13.7
4	245	6.3
5	68	1.7
6 y +	96	2.5
	3900	100.0

Casi la mitad de la muestra (49%) está conformada por UPM con 30 hogares en promedio (rango entre 20 y 44). En estas UPM se procede directamente a la selección de las Unidades Terciarias (UTM), listando las viviendas de la UPM directamente en el terreno, dividiéndola en 3 segmentos de aproximadamente el mismo número de viviendas, y seleccionando al azar uno de ellos (UTM).

La distribución anterior señala una característica particular de la conformación urbanística de las grandes ciudades del país y de algunos de los municipios anexos de las áreas metropolitanas: la existencia de grandes conjuntos urbanos, con edificaciones multifamiliares, que quedaron nombrados en el censo de población como un solo conglomerado.

3.3.2 Selección de USM

En el resto de UPM (52%), conformadas con 2 o más MT, es necesario un paso adicional, la selección de una USM. Para el efecto, se evaluarán en cada caso, dos opciones.

La primera, realizar un submuestreo “proxi” de áreas a nivel de oficina, utilizando las divisiones visibles en la USM geo-referenciada con base en la cartografía digital del DANE. Al revisarse las UPM con 2 o más MT delimitadas automáticamente por el software de la cartografía digital, se observa que muchas de ellas corresponden a conjuntos residenciales de gran tamaño, en los cuales están señalizados los edificios de apartamentos y ciertas divisiones internas, lo que permitiría aproximar una división preliminar y un submuestreo a priori. Lo anterior simplificaría el trabajo de campo de recuento de viviendas, división y submuestreo, pues lo restringiría a un área más reducida.

La segunda opción consiste simplemente en realizar en el terreno el recuento de viviendas, y la subdivisión y submuestreo de la UPM. Esta tarea también sería necesaria en el caso de que el submuestreo "a priori" de áreas a nivel de oficina haya sido parcial. Ello implica:

- Hacer un croquis de la topografía periférica de la UPM según como aparezca en la plancha cartográfica que la contiene (ver Forma 1).
- Realizar un rápido recuento (no listado) de las viviendas existentes en cada lado o sección identificable físicamente. Anotar sobre el croquis dichas cantidades de viviendas.
- Dividir el área de la UPM en tantas partes como el número de medidas de tamaño (MT) preestablecido (dado a nivel central). Las partes deben tener un promedio de 25-30 viviendas, pero el número podría fluctuar entre 17 y 39 viviendas. La división debe ser fácilmente identificable en el terreno. Por lo menos, deberían anotarse sobre el mapa las direcciones de las viviendas que están al inicio y al final del recorrido de cada parte.

Excepcionalmente, por deficiencias del VIHOPE, el número esperado de viviendas de una UPM podría ser muy diferente al existente en la realidad. En este caso particular, el número de partes debería acomodarse a tal realidad. En el caso extremo de inexistencia de viviendas en una UPM seleccionada, se suministraría por fax la UPM de reemplazo en un plazo no mayor de 24 horas después de comunicado el evento.

- Numerar las partes de 1 en adelante, y seleccionar una aleatoriamente (con Tabla de números aleatorios).

Estudio Nacional de Sustancias Psicoactivas en Colombia
- Selección de la Muestra –

Forma 1: CROQUIS DE LA UNIDAD PRIMARIA DE MUESTREO – UPM, O
DE LA USM SUBMUESTRADA A PRIORI

Nombre del Municipio _____ Código _____
Código de la UPM o USM _____ Dirección: Calles _____ Carreras _____
Nombre del Encuestador _____ Código _____
Nombre del Supervisor _____ Código _____

(La figura delimitada debe coincidir con la señalada en la plancha de la cartografía digital)

Definiciones de vivienda y hogar

Vivienda: Es una unidad o espacio independiente y separado, habitado o disponible para ser habitado por una o más personas, que pueden conformar uno o más hogares.

Hogar: Es el conjunto de personas (familiares la mayoría de ellos, habitualmente) que conviven en la misma vivienda en un espacio específico, y que comparten los alimentos (comen de la misma olla). Una persona residente en una vivienda en un espacio específico, y que no comparte los alimentos, constituye un hogar.

Las personas residentes transitoria o permanentemente en hogares institucionales tales como cuarteles, cárceles, ancianatos, orfanatos, hospitales, no forman parte del universo de investigación. Los hogares familiares ubicados en las viviendas institucionales (del administrador, del cuidandero, etc) sí son parte del universo.

3.3.3 Selección de la UTM

En cada UPM de solamente una MT, y en cada USM, se debe realizar el listado de las viviendas existentes, en la Forma 2, dividir el área sobre el croquis de la UPM, o de la USM, en tres partes con aproximadamente el mismo número de viviendas, señalando para cada parte la dirección de la vivienda inicial y final del segmento, y seleccionar al azar (con tabla de números aleatorios) uno de los segmentos.

Proceda a realizar la encuesta en todos los hogares de las viviendas existentes en el segmento seleccionado.

Diligencie la Forma 3, Listado de Hogares y personas 12 a 64 del segmento seleccionado.

Estudio Nacional de Sustancias Psicoactivas en Colombia

- Selección de la Muestra -

Forma 2: LISTADO DE VIVIENDAS DE LA USM Y CONFORMACIÓN DE LOS SEGMENTOS (UTM)

Nombre del Municipio _____ Código _____
Código de USM _____ Dirección: Calles _____ Carreras _____
Nombre del Encuestador _____ Código _____
Nombre del Supervisor _____ Código _____
Fecha: Día ____ Mes ____ Año ____

N° de orden Vivienda	Dirección de Vivienda	Observaciones
_____	_____	_____

Nota: Señalizar los cortes de los segmentos. Deben coincidir con lo señalado en el Croquis de USM. Enmarcar el segmento seleccionado

Estudio Nacional de Sustancias Psicoactivas en Colombia

- Selección de la Muestra –

- TABLA DE NÚMEROS ALEATORIOS –

(Reproducir la tabla del texto “Muestreo de Encuestas, Leslie Kish, Editorial Trillas, México, 1972, Págs. 717, 718”)

Instructivo: el número de arranque de un dígito está marcado. Fue seleccionado aleatoriamente. Si necesita un número de dos dígitos, considere el vecino de la columna de la derecha.

- a) Identifique el rango de los números, de los cuales debe seleccionar uno aleatoriamente.
- b) Revise verticalmente hacia la parte inferior de la tabla, cada uno de los números hasta encontrar uno que esté dentro del rango requerido. Este es el número seleccionado. Márquelo. No puede saltar números. No se puede cambiar el número útil.
- c) Si necesita más números, continúe revisando los siguientes de la columna.

Al terminar de verificar los números de la columna inicial, continúe con la columna siguiente, de la derecha, empezando por el primer número, parte superior. Si se han verificado todos los números de la última columna, continúe con la primera columna de la tabla.

- Selección de la Muestra -

Nombre del Municipio _____ Código _____
 Código de segmento _____ Dirección: Calles _____ Carreras _____
 Nombre del Encuestador _____ Código _____
 Nombre del Supervisor _____ Código _____
 Fecha: Día _____ Mes _____ Año _____

[illegible]

17

3.3.4 Selección de una Persona 12 a 64 por Hogar

(Se utilizará procedimiento de Kish)

3.4 Procedimientos de Estimación

Se denominan así los esquemas para garantizar estimaciones insesgadas para el universo de estudio. Son fuentes de sesgo las desiguales probabilidades de selección, la cobertura incompleta de la muestra seleccionada, los problemas del marco de muestreo, los errores en los instrumentos de recolección, la deficiente capacitación de los encuestadores, los problemas de control de calidad de las mediciones y los errores del procesamiento de los datos. Los problemas de diseño, capacitación, recolección y procesamiento, se manejan con un buen plan de encuesta y con un equipo de recolección bien escogido, motivado y capacitado.

Las probabilidades desiguales implican la ponderación de los datos muestrales por el recíproco de la probabilidad final de selección de cada domicilio encuestado, o por un factor equivalente. Tal recíproco, denominado también “factor básico de expansión, no sólo corrige el sesgo originado en las probabilidades desiguales, sino que genera una estimación de los valores absolutos del universo estudiado. Un factor equivalente puede ser calculado alrededor de la unidad, dividiendo cada factor de expansión por el promedio de todos los factores. En este caso, la ponderación de la muestra no expande los resultados pero sí corrige el sesgo por probabilidades desiguales.

La cobertura incompleta de la muestra afecta la estimación de las características del universo de estudio, por las variables tasas de cobertura que se suelen observar entre subgrupos de población (Ej. por estrato socioeconómico) que tienen comportamientos diferentes. Es conveniente, por lo tanto, evaluar la posibilidad de aplicar un factor de ajuste que busque recuperar la información de los no cubiertos, y ello es factible cuando existe evidencia de que, en promedio, son similares las características de los domicilios cubiertos y los no cubiertos. El factor es una razón entre la muestra seleccionada y la muestra encuestada, y se debe obtener para la menor desagregación posible (segmento en este caso, o pequeños grupos de segmentos). Cuando no existe la evidencia mencionada, la inferencia de los resultados debe restringirse solamente al universo representando por los domicilios encuestados.

Los problemas del marco de muestreo También podrían generar sesgo en los resultados. En nuestro caso, el marco es el Censo General 2005 del DANE, específicamente con respecto al universo de viviendas y hogares, por sector, sección y manzana, y la cartografía digital. Desde el punto de vista de su

calidad, podemos considerarlo muy aceptable y es de lejos, la mejor opción disponible.

Por otra parte, la reconstrucción de los totales y de la estructura por edad y sexo de los universos departamentales y subregionales, mediante el factor básico de expansión, puede tener imprecisiones. En consecuencia, es conveniente la implementación de un factor de ajuste calculado en función de la proyección de la población 12 a 64 del Censo 2005 por municipio, edad y sexo. El factor se calcula ajustando para cada departamento la muestra expandida con el factor básico de expansión, al universo proyectado por edad y sexo.

Probabilidades y factores

Probabilidades:

P₁ = De la UPM en el municipio

$$= \frac{N^{\circ} \text{Medidas de Tamaño (MT) de la UPM}}{N^{\circ} \text{Medidas de Tamaño del Municipio}} * N^{\circ} \text{de UPM seleccionadas}$$

P₂ = De la USM en la UPM

$$= \frac{1}{N^{\circ} \text{de USM en la UPM}}$$

P₃ = De la UTM en la USM

$$= \frac{1}{3}$$

P₄ = De la persona 12- 64 seleccionada en el hogar

$$= \frac{1}{N^{\circ} \text{de personas 12 – 64 en el hogar}}$$

P = Probabilidad Final

$$= P_1 * P_2 * P_3 * P_4$$

Factores:

F_B = Factor Básico de Expansión

$$= 1 / P$$

F_{CH} = Factor de Ajuste por No Cobertura (Respuesta) de Hogares

$$\frac{N^{\circ} \text{ de Hogares Seleccionados (existentes) en el segmento}}{N^{\circ} \text{ de Hogares Encuestados (permitieron listado y seleccion de persona para encuesta)}}$$

Si el factor es mayor de 1.5, este factor se debe calcular para el conjunto de 2 o más segmentos del mismo estrato socioeconómico, del mismo municipio.

F_{CP} = Factor de Ajuste por No Cobertura (Respuesta) de Personas a la Encuesta.

$$= \frac{N^{\circ} \text{ de Personas Seleccionadas para encuesta en segmento}}{N^{\circ} \text{ de Personas encuestadas en segmento}}$$

F_{BC} = Factor Básico de Expansión ajustado por no cobertura de hogares y personas

$$= F_B * F_{CH} * F_{CP}$$

F_{ES} = Factor de Ajuste por Estructura de Edad y Sexo

$$= \frac{N^{\circ} \text{ de personas proyectadas a partir Censo 05 en grupo específico de edad y sexo}}{N^{\circ} \text{ de personas expandidas a partir de } F_{BC} \text{ en el mismo grupo de edad y sexo}}$$

F_f = Factor Final de Expansión

$$= F_{BC} * F_{ES}$$

3.5 Modelo para Cálculo de la Precisión Observada

Se sugieren modelos estadísticos basados en las series de Taylor. Los cálculos deben hacerse con valores ponderados por el factor final de expansión.

Para estimaciones a nivel departamental, subregional, regional y nacional.

Muestreo de conglomerados desiguales:

$$\text{var}(r) = \frac{1}{x^2} [\text{var}(y) + r^2 \text{var}(x) - 2r \text{cov}(y.x)] \quad (4)$$

En donde:

$$r = \text{Estimador de razón} = \frac{y}{x} = \frac{\sum_{\alpha=1}^a y_{\alpha}}{\sum_{\alpha=1}^a x_{\alpha}}$$

$$\alpha = 1, 2, \dots, a$$

= cada uno de los segmentos.

a = total de segmentos de la muestra en el departamento, subregión, región o país

y, x son variables aleatorias.

y = valor acumulado de la variable "y" en el departamento, subregión, región o país

$$= \sum_{\alpha=1}^a y_{\alpha}$$

y_{α} = En variables discretas: número de personas que tienen una determinada característica en el segmento α

En variables continuas: acumulado del valor de la variable en el segmento α

x_{α} = Tamaño de la muestra de personas encuestadas en el conglomerado α

$$x = \sum_{\alpha=1}^a x_{\alpha}$$

= Tamaño de la muestra de personas encuestadas en el departamento, subregión, región o país

$$\text{var}(y) = \frac{a}{a-1} \left(\sum_{\alpha} y_{\alpha}^2 - \frac{y^2}{a} \right)$$

$$\text{var}(x) = \frac{a}{a-1} \left(\sum_{\alpha} x_{\alpha}^2 - \frac{x^2}{a} \right)$$

$$\text{cov}(y.x) = \frac{a}{a-1} \left(\sum_{\alpha} y_{\alpha} x_{\alpha} - \frac{yx}{a} \right), \text{ covarianza de } y, x.$$

Error estándar de r

$$es(r) = \sqrt{\text{var}(r)}$$

Error estándar relativo

$$esrel = \frac{es(r)}{r} * 100$$

Varianza de la diferencia del indicador entre dos mediciones.

Conglomerados desiguales

$$Var(r-r') = var(r) + var(r') - 2cov(r,r')$$

En donde:

$Cov(r,r') = covarianza$

$$= \frac{1}{xx'} [\text{cov}(y.y') + rr' \text{cov}(x.x') - r \text{cov}(y'x) - r' \text{cov}(y.x')]$$

Siendo

$$y,y' = \left(\sum_{\alpha} \sum_j y_{\alpha j} \right) \left(\sum_{\alpha} \sum_j y'_{\alpha j} \right)$$

en forma similar, $x.x'$; $y'x$; $y.x'$

Error estándar de la diferencia

$$Es(r-r') = \sqrt{\text{var}(r-r')}$$

BIBLIOGRAFIA SELECCIONADA

- Cochran, W. , Sampling Techniques, John Wiley and Sons, Inc., New York, 1963
- Gómez, L.C., Ampliación de la Muestra Maestra de Hogares. Documento integral, DANE, Bogotá, Diciembre 10 de 2001 (Informe Técnico)
- Gómez, L.C., Adecuación de la Muestra Maestra de Hogares para la Encuesta Continua, Modernización de la Encuesta Nacional de Hogares, DANE, Bogotá, Febrero 28 / 2000 (Informe Técnico)
- Kish, L., Survey Sampling, John Wiley and Sons, Inc, New York, 1965