


Departamento Administrativo Nacional de Estadística



**Dirección de Metodología y Producción
Estadística**

**Metodología Muestral
Gran Encuesta Integrada De Hogares
GEIH**

Junio de 2007

	METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH	CÓDIGO: DM-GEIH-DIM-01 VERSIÓN:02 PAGINA:2 FECHA: 23-05-07
ELABORÓ: Apoyo diseños muestrales	REVISÓ: Coordinadora de Metodología Estadística	APROBÓ : Director DIMPE

CONTENIDO

1.1. UNIVERSO	5
1.2 POBLACION OBJETIVO	5
1.3. COBERTURA GEOGRÁFICA Y REPRESENTATIVIDAD.....	5
1.4. UNIDADES ESTADÍSTICAS	6
1.4.1. UNIDADES DE MUESTREO	6
1.4.2. UNIDADES DE OBSERVACIÓN	7
1.4.3. UNIDADES DE ANÁLISIS	7
1.5. MARCO ESTADÍSTICO	7
1.6. DESAGREGACIÓN DE LOS RESULTADOS	8
1.7. PARÁMETROS A ESTIMAR.....	8
1.8. DISEÑO MUESTRAL.....	8
1.8.1. TIPO DE MUESTRA	8
1.8.2. TAMAÑO DE MUESTRA	12
1.8.3. MÉTODO DE SELECCIÓN	14
1.8.4. PROBABILIDAD FINAL DE SELECCIÓN	25
1.8.5. METODOLOGÍA DE ESTIMACIÓN	27
1.9. MANTENIMIENTO DE LA MUESTRA Y NOVEDADES	35
1.9.1. ACTUALIZACIÓN EN LOS MARCOS DE REFERENCIA	35
1.9.2. NOVEDADES	39

LISTA DE TABLAS

Tabla 1. Cobertura geográfica de la muestra para las cabeceras municipales, centros poblados y área rural dispersa	6
Tabla 2. Tamaños de muestra,	13
Tabla 3. Submuestras nacionales según el número de hogares requeridos.....	22
Tabla 4. Balance de la muestra actualizada de estratos, 2006	22
Tabla 5. Tabla para submuestreo	38

LISTA DE CUADROS

Cuadro 2. Listado totalizador del marco de muestreo.....	16
Cuadro 3. Secciones seleccionadas. Barranquilla - soledad.....	17
Cuadro 4. Total de viviendas por manzana	19
Cuadro 5. Total de viviendas por manzana.....	20

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES- GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 PÁGINA : 5 FECHA: 23-05-07</p>
---	--	---

1. DISEÑO ESTADÍSTICO

1.1. UNIVERSO

El universo para la Gran Encuesta Integrada de Hogares está conformado por la población civil no institucional residente en todo el territorio nacional.

1.2 POBLACION OBJETIVO

La población objetivo para la Encuesta está conformado por la población civil no institucional residente en todo el territorio nacional; excluyendo los nuevos departamentos, denominados Territorios Nacionales antes de la constitución de 1991, en los cuales reside aproximadamente el 4% de la población.

1.3. COBERTURA GEOGRÁFICA Y REPRESENTATIVIDAD

Tiene cobertura nacional que permite obtener resultados por zona urbana y rural, grandes regiones y total por departamento. La muestra cubre cinco regiones: Atlántica, Oriental, Central, Pacífica y Bogotá y veinticuatro ciudades y áreas metropolitanas: Bogotá, Medellín y su área metropolitana: el Valle de Aburrá conformado por los municipios de Barbosa, Bello, Caldas, Copacabana, Envigado, Girardota, Itagüí, La Estrella y Sabaneta; Cali y su área metropolitana: Yumbo; Barranquilla y su área metropolitana: Soledad; Bucaramanga y su área metropolitana: Floridablanca, Girón y a Piedecuesta, Manizales y su área metropolitana: Villamaría; Pereira y su área metropolitana: Dosquebradas y La Virginia; Cúcuta y su área metropolitana: Villa del Rosario, Los Patios y El Zulia; Pasto, Ibagué, Montería, Cartagena, Villavicencio, Riohacha, Valledupar, Sincelejo, Santa Martha, Armenia, Tunja, Neiva, Florencia, Popayán, Quibdó y San Andrés.

	METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES- GEIH	CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 PÁGINA : 6 FECHA: 23-05-07
---	--	---

Tabla 1. Cobertura Geográfica de la muestra para las cabeceras municipales, centros poblados y área rural dispersa

REGION	DEPARTAMENTOS
ATLÁNTICA	Guajira Cesar Magdalena Atlántico Bolívar Sucre Córdoba
ORIENTAL	Norte de Santander Santander Boyacá Cundinamarca Meta
CENTRAL	Caldas Risaralda Quindío Tolima Huila Caquetá Antioquia
PACÍFICA	Chocó Cauca Nariño Valle
BOGOTÁ	

1.4. UNIDADES ESTADÍSTICAS

1.4.1. Unidades de muestreo

1.4.1.1. Para la Encuesta Nacional:

- ☐ La unidad primaria de muestreo (UPM) está constituido por municipios.
- ☐ La unidad secundaria de muestreo (USM) es de dos tipos:
 - La manzana, para las cabeceras municipales y centros poblados, es decir, para los caseríos, corregimientos e inspecciones de policía.
 - La sección censal, para el área rural dispersa.
 - La unidad terciaria de muestreo (UTM) es el segmento o Medida de Tamaño, tanto en la cabecera municipal rural, caserío, corregimiento, inspección de policía y población dispersa.

1.4.1.2. Para las ciudades auto representadas:

- ☐ La unidad primaria de muestreo (UPM) es cada una de las 24 ciudades capitales con sus áreas metropolitanas.
- ☐ La unidad secundaria de muestreo (USM) es la sección censal.
- ☐ La manzana es la unidad terciaria de muestreo (UTM).

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES- GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 PÁGINA : 7 FECHA: 23-05-07</p>
---	--	---

- La unidad cuarta de muestreo (UCM) corresponde al segmento.

El segmento es la unidad final de selección de la muestra, ya que al seleccionar un segmento, que es el área compacta de aproximadamente 10 viviendas contiguas en promedio, equivale a seleccionar todas las viviendas que se encuentren dentro de sus límites. Seleccionar una vivienda es seleccionar a todos los hogares que la ocupan y seleccionar un hogar, es seleccionar a todas las personas que pertenecen a éste y que cumplen con las condiciones requeridas para formar el universo.

1.4.2. Unidades de observación

La unidad de observación está constituido por los hogares, las personas, las viviendas.

1.4.3. Unidades de análisis

Puesto que la información a considerar se obtiene de las personas y hogares, estas corresponden a las unidades de análisis.

1.5. MARCO ESTADÍSTICO

El marco estadístico está constituido por el inventario cartográfico y el listado de viviendas por ciudad, obtenidos de la información del Censo Nacional de Población y vivienda de 1993, con actualizaciones permanentes y nuevos recuentos de edificaciones y viviendas a través de la misma muestra, con el respectivo mantenimiento.

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 Página : 8 Fecha: 23-05-07</p>
---	---	--

1.6. DESAGREGACIÓN DE LOS RESULTADOS

Aunque cada vez se hace más evidente el requerimiento de los usuarios de la información de las encuestas de hogares por muestreo, de estimaciones con el máximo de desagregación geográfica posible, los recursos habitualmente disponibles, sólo permiten estimaciones directas de aceptable precisión, de fenómenos de baja frecuencia, con una periodicidad y nivel:

- ☐ Mensual: Total Nacional y Total veinticuatro ciudades capitales.
- ☐ Trimestral: Total del país, dividido por cabecera y resto y para cada una de las veinticuatro ciudades y áreas metropolitanas.
- ☐ Semestral: Por regiones.
- ☐ Anual: Para veinticuatro departamentos.

1.7. PARÁMETROS A ESTIMAR

En esta encuesta la mayor parte de los indicadores estadísticos de interés proporciones, tasas y razones.

1.8. DISEÑO MUESTRAL

1.8.1. Tipo de muestra

Teniendo en cuenta los objetivos de la Gran Encuesta Integrada de Hogares, y las premisas fundamentales, se optó por un diseño probabilístico, multietápico, estratificado, de conglomerados desiguales y autoponderado (para las veinticuatro ciudades capitales con sus áreas metropolitanas), según las siguientes definiciones:

1.8.1.1. Probabilístico:

Significa que cada unidad del universo de estudio tiene una probabilidad de selección conocida y superior a cero. Este tipo de muestra permite establecer anticipadamente la precisión deseada en los resultados principales, y calcular la precisión observada en todos los resultados obtenidos.

1.8.1.2. Multietápico:

Para las veinticuatro ciudades auto representadas, la selección de la muestra se realizó en las siguientes cuatro etapas así:

1. Selección de las unidades primarias de muestreo (UPM). Corresponden a las veinticuatro ciudades principales con sus áreas metropolitanas, se consideran

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 Página : 9 Fecha: 23-05-07</p>
---	---	---

como unidades de inclusión forzosa, es decir es obligatoria su inclusión en la muestra.

2. Selección de las unidades secundarias de muestreo (USM) o secciones censales. Dentro del universo muestral previamente estratificado y ordenado, utilizando un proceso de selección sistemática con arranque aleatorio y probabilidad proporcional a la cantidad de medidas de tamaño (M. T.).
3. Selección de las unidades terciarias de muestreo (UTM) o manzanas. Dentro de cada una de las secciones cartográficas previamente seleccionadas, se escoge una o más manzanas, con probabilidad proporcional a la cantidad de M.T.
4. Selección de las unidades cuartas de muestreo (UCM) o segmentos. Dentro de cada manzana, se selecciona un segmento aleatoriamente.

La selección de la muestra en las Cabeceras Municipales, Centros Poblados y Área Rural Dispersa, se realizó en las siguientes tres etapas así:

1. Selección de las unidades primarias de muestreo (UPM). Corresponde a la selección de un municipio o unión de municipios de cada estrato para cada uno de los Departamentos.
2. Selección de las unidades secundarias de muestreo (USM). Para las cabeceras municipales y los centros poblados esta constituida por la manzana y para el área rural dispersa corresponde a la sección censal.

Tanto para las cabeceras municipales, centros poblados y área rural dispersa, que forman parte del universo muestral previamente ordenado, se selecciona las unidades secundarias de muestreo, utilizando un proceso de selección sistemática con arranque aleatorio y probabilidad proporcional a la cantidad de medidas de tamaño (M. T.).

3. Selección de las unidades terciarias de muestreo (UTM). Dentro de las manzanas seleccionadas correspondientes a las cabeceras municipales y centros poblados y dentro de las secciones censales seleccionadas pertinentes al área rural dispersa, se selecciona aleatoriamente un segmento compacto de 10 viviendas en promedio.

Por lo tanto, la unidad última de muestreo (UUM), está constituida por el segmento compacto de 10 viviendas en promedio, tanto en las cabeceras municipales rurales, caseríos, corregimientos, inspecciones de policía y área rural dispersa.

Este proceso de selección por etapas sólo requiere una actualización cartográfica detallada de las manzanas seleccionadas.

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 Página : 10 Fecha: 23-05-07</p>
---	---	--

1.8.1.3. Estratificado:

Para la estratificación y selección de la muestra nacional, las áreas geográficas de los municipios se organizaron, según los principios cartográficos establecidos, en sectores, secciones y manzanas.

Para las trece principales ciudades y sus áreas metropolitanas, cada manzana del marco muestral se clasificó en un sólo estrato socioeconómico, asignándole el estrato modal, es decir, el más frecuente, de las viviendas que la componen. A su vez, cada sección cartográfica se estratificó de acuerdo con el estrato predominante en las manzanas que la componen.

Para las once ciudades se organizaron en sectores, secciones y manzanas.

Para el resto de cabeceras y área rural, los criterios de estratificación fueron:

- ❑ **Geográficos**, a nivel macro, es decir, los departamentos propiamente tales, y a nivel micro, conformado por grupos de municipios. No se incluyen aquí, las veinticuatro ciudades.
- ❑ **Socioeconómicos a nivel municipal**, esto es, dentro de cada departamento, se estratificaron las Unidades Primarias de Muestreo – UPM (municipios, básicamente), en función de macro variables altamente correlacionadas con los fenómenos demográficos, sociales y económicos que iban a ser estudiados en la muestra poblacional, así:
 1. Nivel de urbanización, en términos de la cantidad de población de las cabeceras municipales.
 2. Estructura urbano – rural de la población municipal (% de población en cabecera).
 3. Proporción de la población con necesidades básicas insatisfechas, la cual, a su vez, se clasificó en cuatro intervalos nombrados:

A = <42%; B = 42– 64%; C = 64 y más.
 4. Tamaño poblacional del estrato.

Cada municipio con 7430 o más habitantes en su población total, se constituyó en UPM. Los de menor población se agruparon con uno vecino para constituirse en UPM.

Las UPM cuyo tamaño era similar al tamaño promedio de los estratos, se constituyeron en estratos, y entraron a formar con “certeza” parte de la muestra, para efectos probabilísticos tuvieron probabilidad 1. Las demás UPM se agruparon en estratos de “no certeza” aplicando las variables de estratificación, en el orden de prioridad especificado antes.

Dentro de los estratos de “no certeza” se calculó la probabilidad de selección de las UPM agrupadas dividiendo la población total de cada UPM, por la población del estrato.

	METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH	CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 Página : 11 Fecha: 23-05-07
---	---	--

En total se conformaron 240 estratos

Como ejemplo se presenta en el cuadro 1 la estratificación del Departamento de La Guajira.

Cuadro 1. Estratificación de La Guajira

La Guajira

Código	Municipios	Estrato	2006			% CAB	% N.B.I.	Probabili dad	Probabilida d Acumulada
			TOTAL	CABECERA	RESTO				
44 001	Riohacha				7.555		57,67	1,00	
44 430	Maicao	1 2	142.783	133.879	8.904	93,8	57,96	1,00	1,00
44 847	Uribia	3	70.442	7.563	62.879	10,7	99,43	0,59	0,59
44 560	Manaure		48.407	10.145	38.262	21,0	100,00	0,41	1,00
44 650	San Juan del Cesar		36.795	27.125	9.670	73,7	49,33	0,39	0,39
44 279	Fonseca	4	26.325	23.280	3.045	88,4	49,62	0,28	0,67
44 078	Barrancas		30.175	17.563	12.612	58,2	44,27	0,33	1,00
110									
44 874	El Molino - Villanueva		25.999	24.196	1.803	93,1	51,97	0,32	0,32
44 090	Dibulla ⁴³	5	18.299	4.456	13.843	24,4	68,60	0,23	0,55
855	Urumita - La Jagua del								
44 420	Pilar ⁴⁶		10.480	9.351	1.129	89,2	65,03	0,13	0,68
44 035	Albania ⁴²		10.025	6.582	3.443	65,7	66,90	0,12	0,80
44 378	Hatonuevo ⁴⁵		8.515	7.222	1.293	84,8	48,40	0,11	0,91
44 098	Distraccion ⁴⁴		7.430	4.802	2.628	64,6	52,40	0,09	1,00

En este caso Riohacha por ser la capital del departamento es de inclusión forzosa y su probabilidad es 1, Maicao, por su tamaño en población tiene con “certeza” dos estrato, luego tiene probabilidad 1; los demás municipios, de acuerdo a las variables de estratificación, se agrupan en tres estratos de “no certeza”. Dentro de cada uno de estos tres estratos se calculó la probabilidad de selección de las UPM, como el cociente de la población total de cada UPM, entre la población del estrato. Luego se seleccionó un sólo municipio en cada estrato.

1.8.1.4. De Conglomerados:

En muestras de universos grandes, complejos y heterogéneos, como el de la Gran Encuesta Integrada de Hogares, la concentración de la muestra en conglomerados

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 Página : 12 Fecha: 23-05-07</p>
---	---	---

reduce en gran cantidad los costos, aunque genera un mayor error de muestreo, que se compensa con incremento moderado del tamaño de muestra y produce estimaciones sesgadas por la variación en el tamaño de los grandes conglomerados, la cual se reduce a niveles manejables estratificando por tamaño.

El conglomerado en la Gran Encuesta Integrada de Hogares es el segmento compacto con un promedio de 10 viviendas, en el cual se investigan todas las viviendas, todos los hogares y todas las personas.

La concentración de la muestra en conglomerados presenta grandes beneficios a nivel operativo; en efecto, el segmento representa el promedio de trabajo por recolector en un día, lo cual reduce en forma considerable, los tiempos y gastos de transporte al igual que facilita la supervisión.

Las personas que pertenecen a un mismo segmento, por lo general tienen más características comunes, que aquellas que se podrían seleccionar al azar dentro de una manzana, una sección cartográfica o un municipio.

1.8.1.5. Autoponderado:

Dentro de cada una de las veinticuatro ciudades auto representadas de la muestra, todas las unidades finales de muestreo tienen igual probabilidad de selección. Esta característica permite calcular mediante procesos manuales, los principales indicadores de fuerza de trabajo al hacer expansiones provisionales de la información, de manera que sean representativos de toda la población. Los ajustes respectivos se efectúan posteriormente para obtener los resultados definitivos

1.8.2. Tamaño de muestra

Los cálculos se hicieron con fórmulas apropiadas para el tipo de diseño muestral estratificado de conglomerados desiguales. El efecto de los conglomerados en el diseño (deff), del cual se hablará más adelante, es una relación entre la varianza real de este diseño de conglomerados y la que se obtendría con un diseño aleatorio simple de elementos, se obtuvo a partir de los resultados de la Encuesta Continua de Hogares.

Los tamaños de muestra para los diferentes dominios de estudio por trimestre son:



METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH

CÓDIGO: DM-GEIH-DIM-01

VERSIÓN : 02

Página : 13

Fecha: 23-05-07

Tabla 2. Tamaños de muestra,

DOMINIOS	Trimestral		
	Hogares	Viviendas	Segmentos
GEIH			
TOTAL NACIONAL	67.815	61.650	6.165
TOTAL CIUDADES Y AREAS	48.807	44.370	4.437
TOTAL URBANO	61.215	55.650	5.565
TOTAL RURAL	6.600	6.000	600
Bogotá	2.673	2.430	243
Medellín- AM	3.135	2.850	285
Barranquilla -AM	2.277	2.070	207
Cartagena	2.013	1.830	183
Manizales-AM	2.244	2.040	204
Montería	1.617	1.470	147
Villavicencio	1.914	1.740	174
Pasto	1.749	1.590	159
Cúcuta -AM	1.815	1.650	165
Pereira -AM	2.046	1.860	186
Bucaramanga -AM	1.749	1.590	159
Ibagué	1.947	1.770	177
Cali -AM	2.211	2.010	201
Total 13 áreas	27.390	24.900	2.490
Tunja	1.749	1.590	159
Florencia	1.881	1.710	171
Popayán	2.112	1.920	192
Valledupar	2.013	1.830	183
Quibdó	1.650	1.500	150
Neiva	2.178	1.980	198
Riohacha	1.947	1.770	177
Santa Marta	2.178	1.980	198
Armenia	2.046	1.860	186
Sincelejo	1.782	1.620	162
	19.536	17.760	1.776
San Andrés y Prov.	1.881	1.710	171
	46.926	42.660	4.266

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 Página : 14 Fecha: 23-05-07</p>
---	---	--

Para cada nueva etapa de recolección de la Encuesta a Hogares se rotan las áreas últimas de muestreo que son los segmentos, con 10 viviendas en promedio y que pueden ser una manzana o parte de esta, en los casos de manzanas muy pequeñas. La muestra de las trece principales ciudades está distribuida proporcionalmente por estrato socioeconómico, que como ya se había mencionado son seis estratos.

Se utiliza la siguiente formula para el cálculo del tamaño de muestra:

$$ESrel = \frac{\sqrt{\left(1 - \frac{n}{N}\right) \frac{pq}{n} deff}}{p} \text{ es el error estándar relativo}$$

N = Tamaño del Universo

p = Probabilidad de ocurrencia del fenómeno estudiado

$$deff = \frac{Var(congl.)}{Var(MAS)} \text{ es el efecto de los conglomerados en el diseño}$$

El tamaño de muestra es:

$$n = \frac{Npqdeff}{N(p * ESrel(p))^2 + pqdeff}$$

1.8.3. Método de selección

1.8.3.1. Para las veinticuatro ciudades auto representadas

Definido el marco de muestreo por ciudad y por municipio y el tamaño de la muestra, se procede a la selección de los segmentos, según las siguientes etapas:

Primera etapa – Selección de municipios (UPM)

La primera etapa consiste en la selección de los municipios, que en este caso son las veinticuatro ciudades principales con sus áreas metropolitanas y son unidades de inclusión forzosa, es decir, su introducción en la muestra es obligatoria. Estas ciudades y áreas metropolitanas son: Bogotá, Medellín y su área metropolitana: el Valle de Aburrá conformado por Barbosa, Bello, Caldas, Copacabana, Envigado,

RutaArchivo:\GEIH\DOCUMENTACIONBASICA\METODOLOGIAS\DISEÑOMUESTRAL\METODOLOGÍADISEÑO MUESTRAL\DM-GEIH-DIM-03.doc

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 Página : 15 Fecha: 23-05-07</p>
---	---	--

Girardota, Itagüí, La Estrella y Sabaneta; Cali y su área metropolitana: Yumbo; Barranquilla y su área metropolitana: Soledad; Bucaramanga y su área metropolitana: Floridablanca, Girón y a Piedecuesta, Manizales y su área metropolitana: Villamaría; Pereira y su área metropolitana: Dosquebradas y La Virginia; Cúcuta y su área metropolitana: Villa del Rosario, Los Patios y El Zulia; Pasto, Ibagué, Montería, Cartagena y Villavicencio, Riohacha, Valledupar, Sincelejo, Santa Martha, Armenia, Tunja, Neiva, Florencia, Popayán, Quibdó y San Andrés.

Segunda etapa – Selección de secciones (USM)

La segunda etapa consiste en la selección de las secciones censales o unidades secundarias de muestreo (USM) utilizando un método de selección sistemático así.

Se hace el listado exhaustivo del marco de muestreo de cada ciudad o área metropolitana. Se clasifican las secciones por estrato y se ordenan dentro de cada estrato por su identificación, la cual incluye código del municipio, sector y sección. A cada sección se asignan los siguientes datos: cantidad de manzanas, de viviendas, de medidas de tamaño, esto es, de segmentos y el número de medidas de tamaño acumuladas hasta cada sección.

Se calcula el intervalo de selección como el cociente entre el total de medidas de tamaño de la ciudad y la cantidad de segmentos a seleccionar. Se busca el arranque aleatorio o número seleccionado al azar, mayor que cero y menor o igual al intervalo de selección. Las secciones seleccionadas corresponden a aquellas cuyo acumulado de medidas de tamaño, contiene por primera vez, el arranque aleatorio y las sumas sucesivas del intervalo de selección.

Como ejemplo se presentan los datos correspondientes a la ciudad de Barranquilla y su área metropolitana conformada por el municipio de Soledad. El cuadro 2 presenta el listado totalizador del marco de muestreo con las correspondientes medidas de tamaño por sección.

En este caso, la selección de las 206 secciones ya establecidas para la ciudad de Barranquilla y su área metropolitana se hace de acuerdo a los siguientes cálculos:

Número de secciones a seleccionar = 206

Total de medidas de tamaño acumuladas = 22697

$$\text{Intervalo de selección} = \frac{22697}{206} = 110.179 \approx 110.18$$

Arranque aleatorio = 73.768

	METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH	CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 Página : 16 Fecha: 23-05-07
---	---	--

Cuadro 2. Listado totalizador del marco de muestreo.
Medidas de tamaño a nivel de sección Barranquilla – Soledad

OBS	Dpto	Municipio	Sector	Sección	Total manzanas	de Total viviendas	Estrato modal	Medidas de tamaño	Medidas de tamaño acumuladas	de
1	08	001	101	01	11	127	1	13	13	
2	08	001	101	02	9	86	1	9	22	
3	08	001	101	03	18	141	1	14	36	
4	08	001	101	04	7	118	1	12	48	
5	08	001	101	05	10	112	1	11	59	
6	08	001	101	06	20	405	1	41	100	*
7	08	001	1201	01	15	149	1	15	115	
8	08	001	1201	05	11	101	1	10	125	
9	08	001	1203	01	21	156	1	16	141	
10	08	001	1203	02	14	275	1	28	169	
11	08	001	1301	03	17	396	1	40	209	*
12	08	001	1301	04	16	253	1	25	234	
13	08	001	1302	01	15	202	1	20	254	
14	08	001	1302	02	20	337	1	34	288	
15	08	001	1302	03	22	306	1	31	319	*
16	08	001	1303	01	5	196	1	20	339	
17	08	001	2001	09	13	585	1	59	398	
18	08	001	2001	10	23	421	1	42	440	*
19	08	001	2102	01	19	179	1	18	458	
.	
.	
.	
647	08	758	17	01	10	171	3	17	22328	
648	08	758	17	02	29	429	3	43	22371	*
649	08	758	17	03	30	504	3	50	22421	
650	08	758	17	04	26	243	3	24	22445	*
651	08	758	17	05	29	341	3	34	22479	
652	08	758	18	01	27	487	3	49	22528	
653	08	758	18	02	26	347	3	35	22563	*
654	08	758	18	03	16	291	3	29	22592	
655	08	758	18	04	25	397	3	40	22632	
656	08	758	18	05	17	235	3	24	22656	
657	08	758	4	07	16	399	4	40	22696	*
658	08	758	15	01	2	11	4	1	22697	

Municipio 001 = Barranquilla

Municipio 758 = Soledad

* = Secciones seleccionadas

Nota: Como ejemplo se presenta una parte de los 658 registros en total

	METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH	CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 Página : 17 Fecha: 23-05-07
---	---	--

El cuadro 3 muestra las secciones seleccionadas

Cuadro 3. Secciones seleccionadas. Barranquilla - Soledad

OBS	Dpto	Municipio	Sector	Sección	Estrato modal	Total viviendas	Medidas de tamaño
1	08	001	101	6	1	405	41
2	08	001	1301	3	1	396	40
3	08	001	1302	3	1	306	31
4	08	001	2001	10	1	421	42
5	08	001	2103	5	1	171	17
6	08	001	2501	2	1	351	35
7	08	001	2502	2	1	387	39
8	08	001	2503	3	1	508	51
9	08	001	2503	5	1	521	52
.
.
.
200	08	758	16	5	3	333	33
201	08	758	16	8	3	422	42
202	08	758	16	10	3	260	26
203	08	758	17	2	3	429	43
204	08	758	17	4	3	243	24
205	08	758	18	2	3	347	35
206	08	758	4	7	4	399	40

Municipio 001 = Barranquilla

Municipio 758 = Soledad

NOTA: Como ejemplo se presenta una parte de las 206 secciones en total

Este método garantiza una probabilidad de selección de secciones proporcional al tamaño, es decir a la cantidad de medidas de tamaño (M.T.) de cada sección y una representación de los diversos estratos, proporcionalmente a su tamaño, expresado como el total de las medidas de tamaño de las secciones que las conforman.

Tercera etapa – Selección de manzanas (UTM)

En las secciones seleccionadas, se procede de la siguiente forma:

Se revisa el mapa de cada sección seleccionada con el recuento de viviendas de cada manzana.

Se procede a dar límites claros e identificables a las secciones en que los límites de una o más manzanas no estén bien definidos.

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 Página : 18 Fecha: 23-05-07</p>
---	---	--

Se listan las manzanas correspondientes a las secciones seleccionadas con su total de viviendas y su equivalente en M.T., teniendo cuidado que el acumulado de M.T., de las manzanas sea igual a la cantidad de M.T., con el que se selecciona dicha sección.

Se agrupan aquellas manzanas con 4 o menos viviendas con manzanas vecinas, incluyendo aquellas que por cualquier razón: zonas verdes, parques, etc. no tenían viviendas.

Entonces una vez seleccionadas las USM, ubicadas, listadas y agrupadas sus manzanas, con al menos una M.T., se selecciona con probabilidad proporcional a la cantidad de medidas de tamaño, una o más manzanas de dicha USM, de acuerdo con su densidad de viviendas y el tamaño del intervalo de selección.

Esta operación se cumple seleccionando en forma aleatoria un número entre 1 y el total de M.T. de la sección, por cada manzana por seleccionar.

En una manzana, para las cabeceras municipales y centros poblados, al igual que en una sección censal, para las áreas rurales; que contengan más de una medida de tamaño, se hace un proceso de rotación de las mismas, cada vez que se efectúa la encuesta, hasta tomar la última medida de tamaño.

Para el ejemplo en mención se tiene en el cuadro 4 el total de viviendas de la primera sección seleccionada correspondiente al sector 101, sección 6, de la ciudad de Barranquilla. Las manzanas que tienen menos de 5 viviendas se unirán con otra cercana cartográficamente.

Cuadro 4. Total de viviendas por manzana

Barranquilla – Sector =101 Sección =6

OBS	Segmento	Dpto	Municipio	Cart	Sector	Sección	Estrato modal	Código de Manzana	Total de viviendas
1	1	08	001	1	101	6	1	2	2
2	1	08	001	1	101	6	1	5	7
3	1	08	001	1	101	6	1	6	23
4	1	08	001	1	101	6	1	7	28
5	1	08	001	1	101	6	1	8	16
6	1	08	001	1	101	6	1	9	15
7	1	08	001	1	101	6	1	10	9
8	1	08	001	1	101	6	1	11	1
9	1	08	001	1	101	6	1	14	15
10	1	08	001	1	101	6	1	15	21
11	1	08	001	1	101	6	1	16	51
12	1	08	001	1	101	6	1	17	43
13	1	08	001	1	101	6	1	18	18
14	1	08	001	1	101	6	1	19	5
15	1	08	001	1	101	6	1	20	11
16	1	08	001	1	101	6	1	21	22
17	1	08	001	1	101	6	1	22	24
18	1	08	001	1	101	6	1	23	41
19	1	08	001	1	101	6	1	24	21
20	1	08	001	1	101	6	1	25	32

Cuarta etapa – Selección de segmentos (UCM)

Esta etapa se cumple dentro de la manzana ya seleccionada y sobre el formulario de recuento de edificaciones y viviendas así:

Si la manzana seleccionada tiene una M.T., una M.T. tiene entre 5 y 15 viviendas, ella constituye por sí misma un segmento o área a investigar, y por lo tanto, se entrevistarán todos los hogares de las viviendas de esa área.

Si la manzana tiene dos, tres o más M.T., se selecciona al azar una de ellas y allí se efectúan todas las entrevistas.

Si se trata de un área no amanzanada, se procede a dividir dicha área en las M.T. y con ayuda del recuento de viviendas se selecciona al azar una de ellas.

Para el ejemplo de Barranquilla se presenta en el cuadro 5 los segmentos seleccionados en la sección 6 del sector 101, desde la etapa 0601 de enero de 2006 hasta la etapa 0712 de diciembre de 2007.



METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH

CÓDIGO: DM-GEIH-DIM-01

VERSIÓN : 02

Página : 20

Fecha: 23-05-07

Cuadro 5. Total de viviendas por manzana.

Base geográfica – Barranquilla -A.M. Segmento = 001 Sector = 101 Sección =6
(Se inicia en la etapa 102 de Diciembre de 1998 - base precenso 1993)

Dpto	Municipio	Sector(2001)	Sección (2001)	Estrato	Manzanas (2001)	Viviendas	M.T.	M.T. Acum	EH- 0601- 0602- 0603	EH- 0604- 0605- 0606	EH- 0607- 0608- 0609	EH- 0610- 0611- 0612	EH- 0701- 0702- 0703	EH 0704- 0705- 0706	EH 0707- 0708- 0709	EH 0710- 0711- 0712
08	001	0101	06	1	5	7	1	1								
08	001	0101	06	1	6	23	2	3					*	*	A	
08	001	0101	06	1	7	28	3	6		*	*	*	A			
08	001	0101	06	1	8	16	2	8							*	*
08	001	0101	06	1	9	15	2	10								
08	001	0101	06	1	10	9	1	11								
08	001	0101	06	1	14	15	2	13								
08	001	0101	06	1	15	21	2	15								
08	001	0101	06	1	19	130	13	28								
08	001	0101	06	1	20	11	1	29								
08	001	0101	06	1	21-26-27	23	2	31								
08	001	0101	06	1	22-28	25	3	34								
08	001	0101	06	1	23	41	4	38								
08	001	0101	06	1	24	21	2	40	*	A						
08	001	0101	06	1	25	32	3	43								

NOTA: Las manzanas 21-26 y 27 se unieron para un total de 23 viviendas formando 2 M.T

Las manzanas 22 y 28 se unieron para un total de 25 viviendas formando 3 M.T.

* = Segmentos seleccionados

A = Segmentos agotados

NP= No participa

	<p>METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 PÁGINA : 21 FECHA: 23-05-07</p>
---	--	--

Así, en una sección seleccionada hay en promedio un sólo segmento seleccionado en cada trimestre.

	METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH	CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 PÁGINA : 22 FECHA: 23-05-07
---	---	--

1.8.3.2. Para las cabeceras municipales, centros poblados y área rural dispersa

Una vez definido el marco de muestreo por municipio y el tamaño de la muestra, se selecciona los segmentos, de acuerdo a las siguientes etapas:

Primera etapa – Selección de municipios (UPM)

Utilizando el método de selección controlada se seleccionan los municipios o Unidades Primarias de Muestreo UPM de cada departamento quedando finalmente seleccionado un municipio o UPM por cada estrato.

Conforme a los ajustes realizados a la muestra, se amplió a 240 estratos los cuales se distribuyeron en 12 submuestras nacionales replicadas (una por mes), y a su vez para facilitar la obtención de indicadores departamentales. Se completaron 223 UPM (con 335 municipios), con un total de 240 estratos (algunas UPM grandes tienen 2 o 3 estratos). Consiguientemente, se reestructuró toda la actual estatificación, con base en la última información disponible sobre las variables de clasificación (censo 93 y sus proyecciones). “El número 240 es modular no sólo para la conformación de las 12 submuestras replicadas mensuales de 20 estratos cada una, sino que permite la conformación de otro conjunto de clasificaciones o submuestras nacionales replicadas para los estudios coyunturales: 4 submuestras de 60 estratos cada una y 3 submuestras de 80 estratos cada una). Estas clasificaciones permiten, a su vez ofrecer un variado menú de submuestras nacionales según el número de hogares requeridos. cuya magnitud se muestra en la tabla 3.

Tabla 3. Submuestras Nacionales según el número de hogares requeridos

REQUERIMIENTO MUESTRAL DE UN ESTUDIO EN PARTICULAR		N° DE ESTRATOS	N° DE SUBMUESTRAS COMBINADAS
SEGMENTOS (10 A 20 POR ESTRATO)	HOGARES		
600 – 1200	5700 – 11200	60	1
800 – 1600	8000 – 16000	80	1
1140 – 2280	11400 - 22800	120	2 (de 60 UPM)
1600 – 3200	16000 - 32000	160	2 (de 80 UPM)
1800 – 3600	18000 - 36000	180	3 (de 60 UPM)

Se trató de mantener buena parte de las UPM de la muestra anterior, para poder seguir utilizando los materiales muestrales disponibles, haciendo el pertinente ajuste de las probabilidades de selección y de los factores básicos de expansión. De los 240 estratos de la nueva muestra, 200 son de la muestra anterior, como se muestra en la tabla 4:

Tabla 4. Balance de la muestra Actualizada de Estratos, 2006

	METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH	CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 PÁGINA : 23 FECHA: 23-05-07
---	---	--

Región	Departamento	Estratos de certeza		Estratos probabilísticos		Estratos probabilísticos	
		Antes	Actual	Antes	Actual	Antiguos	Nuevos
Atlántica	Atlántico	1	1	5	7	5	2
	Bolívar	1	1	12	13	12	1
	Cesar		1	8	7	4	3
	Córdoba	1	1	7	11	8	3
	Guajira		1	5	5	5	
	Magdalena		1	8	10	5	5
	Sucre		1	7	7	4	3
	San Andrés		1				
	ST	3	8	52	60	43	17
Oriental	Boyacá		1	19	14	13	1
	Cundinamarca			22	25	23	2
	Meta	1	1	6	4	4	0
	Norte de Sant.	1	1	7	6	6	0
	Santander	1	1	13	11	9	2
	ST	3	4	67	60	55	5
Central	Antioquia	1	1	24	32	26	6
	Caldas	1	1	8	9	8	1
	Caquetá		1	6	4	4	0
	Huila		1	9	8	6	2
	Quindío		1	6	4	2	2
	Risaralda	1	1	5	4	4	0
	Tolima	1	1	11	11	10	1
	ST	4	7	69	72	60	12
Pacífico	Cauca		1	9	11	9	2
	Chocó		1	6	4	4	0
	Nariño	1	1	14	13	13	0
	Valle	1	1	11	20	16	4
	ST	2	4	40	48	42	6
Bogotá	Bogotá	1	1				
Total		13	24	228	240	200	40

Segunda etapa – Selección de manzana (USM) en cabeceras municipales y Centros Poblados

Ruta Archivo: GEIH/DOCUMENTACION BASICA/METODOLOGIAS/DISEÑO MUESTRAL/METODOLOGÍA DISEÑO MUESTRAL/DM-GEIH-DIM-03.doc

DIMPE 01

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 PÁGINA : 24 FECHA: 23-05-07</p>
---	---	--

Selección de sección censal (USM) en área rural dispersa

La segunda etapa consiste en la selección de las unidades secundarias de muestreo (USM) que son las secciones censales en el área rural dispersa y las manzanas para el caso de las cabeceras municipales y los Centros Poblados constituidos por los caseríos, inspecciones de policía y corregimientos.

Estas unidades secundarias de muestreo, se seleccionan utilizando un método de selección sistemático así.

Se hace el listado exhaustivo del marco de muestreo del municipio. Se ordena dentro de cada municipio por su identificación, la cual incluye código del municipio, sector y sección. En el Área Rural Dispersa a cada sección se asignan los siguientes datos: cantidad de viviendas y su equivalente en medidas de tamaño, esto es, de segmentos y el número de medidas de tamaño acumuladas. Para cabeceras y Centros Poblados, a cada manzana se le asignan de la misma manera, la cantidad de viviendas, su equivalente en medidas de tamaño y el número de medidas de tamaño acumuladas.

Para Cabeceras municipales y Centros Poblados, se agrupan aquellas manzanas con 4 o menos viviendas con manzanas vecinas, incluyendo aquellas que por cualquier razón: zonas verdes, parques, etc. no tenían viviendas.

Se calcula el intervalo de selección como el cociente entre el total de medidas de tamaño del municipio y la cantidad de segmentos a seleccionar.

Se busca el arranque aleatorio o número seleccionado al azar, mayor que cero y menor o igual al intervalo de selección.

Las secciones y manzanas seleccionadas corresponden a aquellas cuyo acumulado de medidas de tamaño, contiene por primera vez, el arranque aleatorio y las sumas sucesivas del intervalo de selección.

En una manzana, para las cabeceras municipales y centros poblados, al igual que en una sección censal, para las áreas rurales; que contengan más de una medida de tamaño, se hace

un proceso de rotación de las mismas, cada vez que se efectuase la encuesta, hasta tomar la última medida de tamaño. Cuando se agotan las medidas de tamaño dentro de la manzana o dentro de la sección seleccionada, se efectúa de nuevo la selección de una medida de tamaño en la misma sección. Una manzana y una sección, permanecen en la investigación desde el momento de ser seleccionada.

Este método garantiza una probabilidad de selección de secciones y de manzanas, proporcional al tamaño, es decir a la cantidad de medidas de tamaño (M.T.) correspondientes a cada sección y a cada manzana.

Tercera etapa – Selección de segmentos (UTM)

Esta etapa se cumple dentro de la manzana y en forma correspondiente a la sección ya seleccionados y sobre el formulario de recuento de edificaciones y viviendas así:

Si la manzana o la sección seleccionada tiene una M.T., la cual tiene entre 5 y 15 viviendas, ella constituye por sí misma un segmento o área a investigar, y por lo tanto, se entrevistarán todos los hogares de las viviendas de esa área.

Si la manzana o la sección censal tiene dos, tres o más M.T., se selecciona al azar una de ellas y allí se efectúan todas las entrevistas.

Luego, en una sección seleccionada y en una manzana seleccionada, hay un sólo segmento seleccionado en cada periodo.

1.8.4. Probabilidad final de selección

Para cada municipio se tiene que:

MTT : Es el número de MT del estrato

MTmp : Es el número total de MT del municipio

MTs : Es el número de MT de la sección seleccionada

MTm : Es el número de MT de la manzana seleccionada

MTg : Es el número de MT del segmento, este número siempre es 1 ya que por definición, segmento = 1MT.

La probabilidad final de selección, si la muestra fuera de un sólo segmento (g) y por ende de todas las viviendas del segmento, para cabeceras municipales y centros poblados, está dada por:

$$P(g) = \frac{\overbrace{MTmp}^{\text{Probabilidad Municipio en el Estrato}}}{MTT} * \frac{\overbrace{MTm}^{\text{Probabilidad Manzana en el Municipio}}}{MTmp} * \frac{\overbrace{MTg}^{\text{Probabilidad Segmento en la Manzana}}}{MTm} = \frac{MTg}{MTT} = \frac{1}{MTT} = \frac{1}{N}$$

Para el caso del área rural dispersa es:

$$P(g) = \frac{\overbrace{MTmp}^{\text{Probabilidad Municipio en el Estrato}}}{MTT} * \frac{\overbrace{MTs}^{\text{Probabilidad Sección en el Municipio}}}{MTmp} * \frac{\overbrace{MTg}^{\text{Probabilidad Segmento en la Sección}}}{MTs} = \frac{MTg}{MTT} = \frac{1}{MTT} = \frac{1}{N}$$

Para cada una de las veinticuatro ciudades se tiene que:

MTT : Es el número total de MT de la ciudad
 MTs : Es el número de MT de la sección seleccionada
 MTm : Es el número de MT de la manzana seleccionada
 MTg : Es el número de MT del segmento, este número siempre es 1 ya que por definición, segmento = 1MT.

Luego la probabilidad final de selección, si la muestra fuera de un sólo segmento (g), está dada por:

$$P(g) = \underbrace{\hat{1}}_{\text{Probabilidad de selección de la Ciudad}} * \underbrace{\frac{MTs}{MTT}}_{\text{Probabilidad Sección en la ciudad}} * \underbrace{\frac{MTm}{MTs}}_{\text{Probabilidad Manzana en la Sección}} * \underbrace{\frac{MTg}{MTm}}_{\text{Probabilidad Segmento en la Manzana}} = \frac{MTg}{MTT} = \frac{1}{MTT} = \frac{1}{N}$$

Por otro lado se tiene que:

n : Es el tamaño de muestra
 N : Es el tamaño de la población
 $f_{h\alpha}$: Es la fracción global de muestreo o probabilidad de selección de un segmento α cualquiera

En general, la probabilidad de selección P de un segmento cualquiera en la ciudad o en el municipio, está dada por la probabilidad condicional:

$$\begin{aligned}
 P &= \frac{1}{MTT} + \frac{1}{(MTT-1)} \frac{(MTT-1)}{MTT} + \frac{1}{(MTT-2)} \frac{(MTT-1)}{MTT} \frac{(MTT-2)}{(MTT-1)} + \\
 &+ \frac{1}{(MTT-3)} \frac{(MTT-1)}{MTT} \frac{(MTT-2)}{(MTT-1)} \frac{(MTT-3)}{(MTT-2)} + \dots + \frac{1}{(MTT-n)} \frac{(MTT-1)}{MTT} \dots \frac{(MTT-n)}{(MTT-(n-1))} \\
 &= \underbrace{\frac{1}{MTT} + \frac{1}{MTT} + \dots + \frac{1}{MTT}}_{n-\text{veces}} = \frac{n}{MTT} = \frac{n}{N} = f_{h\alpha}
 \end{aligned}$$

Esta probabilidad (P) es igual para todos los segmentos en una ciudad o en un municipio. Además, como se investigan todos los hogares de todas las viviendas del segmento, esa probabilidad es la misma para toda vivienda, para todo hogar y para todas las personas de la misma ciudad o del mismo municipio.

Dentro de cada ciudad de la muestra urbana las probabilidades de selección de las unidades finales de muestreo son iguales; para el total urbano las probabilidades son desiguales.

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 PÁGINA : 27 FECHA: 23-05-07</p>
---	---	--

1.8.5. Metodología de estimación

Los procedimientos de estimación son aquellos que se usan para el manejo de los datos básicos y la generación de las estimaciones al universo investigado. Están orientados a minimizar los sesgos asociados con el diseño muestral, o con la cobertura incompleta de la muestra, dado por la no respuesta y a mejorar las estimaciones con base en variables exógenas de buena calidad.

1.8.5.1. Factor de expansión y ajuste por no respuesta

Los resultados de la muestra son representativos de la población total del universo investigado, por lo tanto éstos deben expandirse a partir de cada segmento seleccionado con el fin de producir estimaciones válidas de dicho universo. Los componentes del factor de expansión son los siguientes:

1.8.5.1.1. Factor básico de expansión (F_{BAS}).

Es el recíproco de la probabilidad final de selección P de un segmento cualquiera en la ciudad. Aplicado a los datos muestrales, da a cada elemento de la muestra, el peso o representación que le corresponde en el universo investigado. En consecuencia, mediante su aplicación, se estiman numéricamente, en forma aproximada, las características de dicho universo.

El factor básico es igual al “intervalo de selección” de las secciones, que es el cociente entre el total de medidas de tamaño de la ciudad y la cantidad de segmentos a seleccionar. Así:

$$F_{BAS} = \frac{N}{n}$$

Este dato es proporcionado por el muestrista y está calculado por áreas metropolitanas cuando se trata de la encuesta trimestral. Este peso básico es calculado en el momento de diseñar la muestra y se cambia en caso de efectuarse un cambio en el marco.

1.8.5.1.2. Peso de submuestreo ($P_{h\alpha}$).

El peso de submuestreo se calcula para cada uno de los segmentos. Está dado por segmento y es teóricamente igual a 1 para todos los segmentos en razón de que representan una medida de tamaño. Sin embargo el desarrollo dinámico del marco de muestreo puede hacer que este peso sea mayor que 1, de acuerdo con la densidad de viviendas del segmento en el momento de hacer la Encuesta, y por tal motivo modifica el factor básico de expansión en el segmento.

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 PÁGINA : 28 FECHA: 23-05-07</p>
---	---	--

Los pesos de submuestreo dependen de la fracción de muestreo seleccionada en la tabla de submuestreo correspondiente y son proporcionados por las oficinas regionales.

1.8.5.1.3. Ajuste de cobertura por no respuesta ($R_{h\alpha}$).

Se calcula el factor de cobertura por segmento, teniendo en cuenta las encuestas totales sobre las encuestas efectivas. Este factor debe ser máximo 1.5, en caso de que pase de este valor, se debe promediar con el factor de expansión de otros segmentos del mismo estrato; este procedimiento permite disminuir los factores de cobertura altos.

“Es importante, cuando las tasas de no respuesta varían en los subgrupos de la población de diferentes características (ejemplo, en los estratos socioeconómicos). En este caso, la no-corrección, daría lugar a estimaciones sesgadas para el conjunto de la población estudiada. El ajuste habitual es asignar a los hogares y personas no encuestadas, el promedio de las características de los encuestados en el mismo segmento. Esto se logra, corrigiendo el factor básico de expansión por un nuevo factor resultante de la razón entre número de hogares seleccionados en un segmento y el número de hogares encuestados en el mismo segmento.”

1.8.5.1.4. Ajuste por los valores poblacionales.

Las estimaciones del universo de estudio pueden ser mejoradas, si se igualan a nivel de las desagregaciones geográficas los totales de la población obtenidos de la muestra expandida, con los totales del censo de población proyectados a la fecha de la encuesta. Dicho mejoramiento se basa en la premisa de que la estructura por desagregaciones geográficas, proyectada a partir de un censo de población reciente, es más exacta que la estimada a partir de la muestra. La igualación de las cifras absolutas totales tiene ventajas prácticas para los usuarios de los datos oficiales.

Una proyección de población es un conjunto de resultados relativos a la evolución de la población futura obtenida con modelos cuyos parámetros se construyen a partir de los diagnósticos sobre la evolución reciente de las variables determinantes de su dinámica

Considerando la población total de cada ciudad o de cada municipio, como exógena, solamente los valores poblacionales de la población total (PT) se estimaban como una proyección demográfica quincenalmente hasta diciembre del año 2000.

La población en edad de trabajar (PET), la población económicamente activa (PEA) y la población desempleada (DS) se calculaban como una estimación del total a partir de la misma encuesta.

1.8.5.1.5. Factor de restitución ($F_{h\alpha}$).

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 PÁGINA : 29 FECHA: 23-05-07</p>
---	---	--

Es el producto del factor básico de expansión por el peso de submuestreo y por el ajuste de cobertura por no respuesta, aplicándose a cada segmento de la ciudad o área investigada.

Luego si:

$h=1,2,\dots,H$ son cada uno de los estratos

$\alpha=1,2,\dots,a_h$ son cada uno de los segmentos seleccionados del estrato h

Se tiene entonces:

$$F_{h\alpha} = F_{BAS} * P_{h\alpha} * R_{h\alpha}$$

Después de obtener los valores del factor de restitución se calcula el número de personas restituidas al marco muestral. Luego, si:

x_{hak} es el valor de la variable de interés para el hogar k del segmento α en el estrato h

$t_{h\alpha}$ es el total de hogares con entrevista completa en el segmento α del estrato h

$x_{h\alpha} = \sum_{k=1}^{t_{h\alpha}} x_{hak}$ es el valor de la variable de interés para el segmento α del estrato h

Se tiene que el total de la característica de interés restituida al marco de donde provino, \hat{x}' está calculado por:

$$\hat{x}' = \sum_{h=1}^H \sum_{\alpha=1}^{a_h} x_{h\alpha} F_{h\alpha}$$

Los resultados muestrales se llevan a un ajuste por variable exógena, que es un estimador independiente de población. Esto permite mejorar las estimaciones, si a nivel de las desagregaciones geográficas, se igualan los totales estimados con base en la muestra, con los totales del censo de población proyectados a la fecha de la encuesta.

Lo anterior se basa en la premisa de que la estructura por desagregaciones geográficas, proyectada a partir de un censo de población reciente, es más exacta que la estimada a partir de la muestra.

Entonces el factor de ajuste se calcula dividiendo las proyecciones de población con base en el último censo sobre el número de personas restituidas en el marco.

$L =$ Proyección de la población en edad de trabajar a la fecha de la encuesta
Población restituida de la muestra

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 PÁGINA : 30 FECHA: 23-05-07</p>
---	---	--

Este factor de ajuste se aplica al Factor de expansión de cada uno de los segmentos, luego el factor final de expansión, por segmento es

$$F_{EXP_{h\alpha}} = F_{h\alpha} * L$$

Esto es:

$$F_{EXP_{h\alpha}} = F_{BAS} * P_{h\alpha} * R_{h\alpha} * L$$

Luego el total estimado de la característica estudiada, notado por \hat{x} será finalmente:

$$\hat{x} = \sum_{h=1}^H \sum_{\alpha=1}^{a_h} x_{h\alpha} F_{EXP_{h\alpha}}$$

1.8.5.2. Estimador y estimación

Se denomina estimador a la función real $\hat{r} : S \rightarrow \Re$, con dominio sobre el conjunto S de muestras aleatorias, cada una de ellas con probabilidad $p(s_i)$. Al remplazar esta fórmula con los valores provenientes de una muestra particular, se obtiene un valor denominado estimación o valor estimado.

En el caso particular de la Encuesta a Hogares, se realizan principalmente estimación de razones.

1.8.5.2.1. Estimador de razones en una Encuesta

Las tasas, razones, proporciones y promedios, generadas a partir de este diseño muestral son de la forma de una razón, en la cual el numerador y el denominador son variables aleatorias, luego, si:

$h = 1, \dots, H$ son cada uno de los estratos
 $\alpha = 1, \dots, a_h$ son cada uno de los segmentos del estrato h
 a_h es la cantidad de conglomerados seleccionados en el estrato h
 $F_{EXP_{h\alpha}}$ es el factor final de expansión en cada uno de los segmentos α del estrato h

$\hat{y} = \sum_{h=1}^H \sum_{\alpha=1}^{a_h} y_{h\alpha} F_{EXP_{h\alpha}}$ es el valor estimado de la variable de estudio y en el universo

$\hat{x} = \sum_{h=1}^H \sum_{\alpha=1}^{a_h} x_{h\alpha} F_{EXP_{h\alpha}}$ es el valor estimado de la variable de estudio x en el universo

Entonces la razón r está dada por :

$$\hat{r} = \frac{\hat{y}}{\hat{x}} = \frac{\sum_h \sum_{\alpha} y_{h\alpha} F_{EXP_{h\alpha}}}{\sum_h \sum_{\alpha} x_{h\alpha} F_{EXP_{h\alpha}}}$$

1.8.5.2.2. Estimador de razones en dos Encuestas sucesivas

Con el objeto de medir el cambio de las variables, entre una encuesta y otra, se calcula la diferencia entre estas, así sea:

r es el valor de la razón en una encuesta en la etapa t

r' es el valor de la razón en la etapa t-1

h = 1,...,H son cada uno de los estratos

$\alpha = 1, \dots, a_h$ son cada uno de los segmentos del estrato h

$\hat{y} = \sum_{h=1}^H \sum_{\alpha=1}^{a_h} y_{h\alpha} F_{EXP_{h\alpha}}$ es el valor estimado de la variable de estudio y en el universo

$\hat{x} = \sum_{h=1}^H \sum_{\alpha=1}^{a_h} x_{h\alpha} F_{EXP_{h\alpha}}$ es el valor estimado de la variable de estudio x en el universo

Se tiene entonces:

$$\hat{r} - \hat{r}' = \frac{\hat{y}}{\hat{x}} - \frac{\hat{y}'}{\hat{x}'} = \frac{\sum_h y_h F_{EXP_{h\alpha}}}{\sum_h x_h F_{EXP_{h\alpha}}} - \frac{\sum_h y'_h F'_{EXP_{h\alpha}}}{\sum_h x'_h F'_{EXP_{h\alpha}}} = \frac{\sum_h \sum_{\alpha} y_{h\alpha} F_{EXP_{h\alpha}}}{\sum_h \sum_{\alpha} x_{h\alpha} F_{EXP_{h\alpha}}} - \frac{\sum_h \sum_{\alpha} y'_{h\alpha} F'_{EXP_{h\alpha}}}{\sum_h \sum_{\alpha} x'_{h\alpha} F'_{EXP_{h\alpha}}}$$

1.8.5.3.. Errores muestrales

Las características poblacionales son estimaciones que se calculan a partir de los elementos que incluye la muestra. Esta estimación depende del diseño de la muestra y de la combinación particular de los elementos que resultarán seleccionados. Estas estadísticas obtenidas están sujetas a dos clases de error: los errores ajenos al muestreo y los errores debidos al muestreo.

- Errores ajenos al muestreo, o errores no muestrales que se presentan porque los procedimientos de observación son imperfectos. Esta clase de errores es inherente a toda investigación estadística. Establecer su magnitud supone el uso de procesos complicados de replicación, con otros encuestadores, otras formas de preguntas, etc. que permitan comprobar la variabilidad de las respuestas frente al operativo utilizado.
- Errores debidos al muestreo, o errores muestrales, que se presentan debido a que sólo se investiga una fracción de la población total. La muestra seleccionada es una de un número de combinaciones de N elementos tomados en grupos de n posibles muestras que hubieran podido ser seleccionadas. La variación en los resultados que habrían dado estas muestras, debido al azar, forman el error de muestreo. El cálculo de estos errores se puede hacer a partir de los resultados obtenidos y permite evaluar la calidad del diseño utilizado, siempre y cuando pertenezca a la categoría de los diseños aleatorios. Para medir la magnitud de la variabilidad de la distribución muestral del estimador, denominado error muestral, se usa frecuentemente la desviación estándar y el coeficiente de variación.

La desviación estándar de una estimación se llama error estándar y está definido como la raíz cuadrada de la varianza.

$$ES(\hat{p}) = \sqrt{V(\hat{p})}$$

En la práctica no se conoce la varianza del estimador, entonces se aproxima por su estimador y su estimación; así con la muestra seleccionada se tiene la desviación estándar de la distribución de muestreo de una estimación o error estándar de la estimación:

$$\hat{ES}(\hat{p}) = \sqrt{\hat{V}(\hat{p})}$$

La varianza es la magnitud de la variabilidad de la distribución muestral debida al azar, entre menor sea la magnitud de esta variabilidad mejor será la precisión de la estimación del parámetro de interés, siempre que no haya sesgo. Esta magnitud depende del método utilizado para seleccionar la muestra, de la variabilidad natural de la variable estudiada, del tamaño de la muestra, de la frecuencia con que se presente el hecho estudiado y del parámetro a estimar, ya que para una misma muestra, es diferente si se desea estimar un total, un promedio, una razón, etc.

El coeficiente de variación se define como la relación porcentual del error estándar a la estimación central, esto es, el cociente entre el error estándar del estimador y el valor esperado del estimador, multiplicado por 100.

$$CV = \frac{\sqrt{V(\hat{r})}}{E(\hat{r})} * 100$$

El valor de este coeficiente, expresado en porcentaje, permite evaluar rápidamente la calidad de un procedimiento de estimación. Por lo general no se conoce la varianza real del parámetro $V(\hat{r})$ y tampoco su esperanza $E(\hat{r})$, estas son expresiones con las que se hacen comparaciones y se toman decisiones, sin necesidad de disponer de un valor numérico; entonces aunque no sea precisamente la misma medida, se utiliza el coeficiente de variación estimado *cve*, con la muestra particular obtenida.

$$cve = \frac{\sqrt{\hat{V}(\hat{r})}}{\hat{r}} * 100$$

En consecuencia, se suele considerar que el resultado de una estimación es bueno cuando su coeficiente de variación es menor del 5%, aceptable entre 5% y 10%; apenas regular entre 10% y 15%; y no deseable si es mayor que el 15%.

Aunque la magnitud de la variabilidad de la distribución muestral del estimador es medida por la varianza, el error estándar y el coeficiente de variación, este último tiene la ventaja de dar dicha medida en términos porcentuales, por lo cual es una medida común para las estimaciones.

Los datos obtenidos permiten evaluar además la eficiencia del diseño de la muestra utilizada. Se trata del efecto de diseño (*deff*). Este se define como el cociente entre la varianza del estimador \hat{r} de r en un diseño $p(s)$ cualquiera y la varianza de una muestra aleatoria simple de la misma cantidad de elementos, esto es:

$$deff(p, \hat{r}) = \frac{V_p(\hat{r})}{V_{MAS}(\hat{r}_\pi)}$$

1.8.5.3.1. Precisión de la razón en una Encuesta

Los errores estándar de la ENH, pueden ser calculados con fórmulas desarrolladas para el Muestreo Estratificado de Conglomerados Desiguales aquí utilizado. Un método apropiado para calcular o aproximar los valores de los errores estándar es el de “Propagación de Varianzas” que “se refiere a un método asintótico de varianzas aproximadas de funciones aleatorias en muestras grandes” (Kish, 1979, 674).

Para cada municipio se tiene:

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 PÁGINA : 34 FECHA: 23-05-07</p>
---	---	--

$h = 1, \dots, H$ son cada uno de los estratos

$\alpha = 1, \dots, a_h$ son cada uno de los segmentos del estrato h

$f_{h\alpha}$ es la fracción global de muestreo o probabilidad de selección del segmento α en el estrato h

$\hat{y}_h = \sum_{\alpha=1}^{a_h} y_{\alpha} F_{EXP_{h\alpha}}$ es el valor estimado de la variable de estudio y en el estrato h

$\hat{x}_h = \sum_{\alpha=1}^{a_h} x_{\alpha} F_{EXP_{h\alpha}}$ es el valor estimado de la variable de estudio x en el estrato h

$$\hat{\text{var}}(\hat{y}_h) = \text{varianza de } \hat{y}_h = \frac{1-f_{h\alpha}}{a_h-1} \left[a_h \sum_{\alpha} (y_{h\alpha} F_{EXP_{h\alpha}})^2 - \left(\sum_h \sum_{\alpha} y_{h\alpha} F_{EXP_{h\alpha}} \right)^2 \right]$$

$$\hat{\text{var}}(\hat{x}_h) = \text{varianza de } \hat{x}_h = \frac{1-f_{h\alpha}}{a_h-1} \left[a_h \sum_{\alpha} (x_{h\alpha} F_{EXP_{h\alpha}})^2 - \left(\sum_h \sum_{\alpha} x_{h\alpha} F_{EXP_{h\alpha}} \right)^2 \right]$$

$$\hat{\text{cov}}(\hat{y}_h, \hat{x}_h) = \text{covarianza de } \hat{y}_h, \hat{x}_h = \frac{1-f_{h\alpha}}{a_h-1} \left[a_h \sum_{\alpha} (y_{h\alpha} x_{h\alpha} F_{EXP_{h\alpha}}) - \sum_h \sum_{\alpha} x_{h\alpha} y_{h\alpha} F_{EXP_{h\alpha}} \right]$$

$$\hat{ES}(\hat{r}_h) = \sqrt{\frac{1}{(\hat{x}_h)^2} [\hat{\text{var}}(\hat{y}_h) + \hat{r}_h^2 \hat{\text{var}}(\hat{x}_h) - 2\hat{r}_h \hat{\text{cov}}(\hat{y}_h, \hat{x}_h)]}$$

Entonces el error estándar de una razón es:

$$\hat{ES}(\hat{r}) = \sqrt{\frac{1}{(\hat{x})^2} \left[\sum_h \hat{\text{var}}(\hat{y}_h) + \hat{r}^2 \sum_h \hat{\text{var}}(\hat{x}_h) - 2\hat{r} \sum_h \hat{\text{cov}}(\hat{y}_h, \hat{x}_h) \right]}$$

1.8.5.3.2. Precisión de la razón en dos Encuestas sucesivas

Para calcular el error estándar del estimador de la diferencia, en las etapas en que no se hizo el traslape, se usa la siguiente fórmula

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

En las etapas donde se hace traslape se utiliza la siguiente fórmula:

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 PÁGINA : 35 FECHA: 23-05-07</p>
---	---	--

$$\text{var}(r - r') = \sqrt{\text{var}(r) + \text{var}(r') - 2\text{cov}(r, r')}$$

1.9. MANTENIMIENTO DE LA MUESTRA Y NOVEDADES

El sistema de selección de la muestra permite tomar en forma sucesiva varias muestras probabilísticas a partir de un listado de manzanas, dado en el marco de muestreo. Lo anterior se explica porque al seleccionar en forma aleatoria, los segmentos con probabilidades iguales en la última etapa de la encuesta, los segmentos restantes constituyen también una muestra aleatoria.

Por esta razón se pueden efectuar nuevas selecciones, así:

- Si la manzana seleccionada tiene una medida de tamaño, en la primera etapa del proyecto en que se use, se encuestarán los hogares de todas las viviendas de esa manzana, y para la siguiente etapa, esa manzana se remplazará, aleatoriamente con probabilidad proporcional a la cantidad de medidas de tamaño, por otra de la misma sección.
- Si la manzana seleccionada se puede partir en dos o más segmentos de acuerdo con sus MT, se tomará una para la primera etapa, otra para la segunda, y así sucesivamente, hasta agotar la manzana. Una vez agotadas las MT de la manzana, se seleccionará al azar con probabilidad proporcional a la cantidad de medidas de tamaño, una nueva manzana de la misma sección, para remplazar la manzana agotada.

De igual manera se puede mantener la misma sección cartográfica en las muestras mientras no se agoten todos los segmentos o medidas de tamaño de las manzanas y todas las manzanas dentro de la sección.

1.9.1. Actualización en los marcos de referencia

La actualización permanente de los marcos de referencia de las muestras es necesaria para que se logre acercarlos de nuevo a los universos que están representando. Ya que por la dinámica de crecimiento, los municipios incorporan nuevas manzanas, las cuales deben ser incluidas en los marcos y tengan oportunidad de ser seleccionadas. Estas actualizaciones se refieren a los siguientes aspectos:

- Actualización de la cartografía con base en la información obtenida en las diferentes instituciones que desarrollan programas de vivienda y planifican el crecimiento de las ciudades, mediante verificación en el terreno de los límites cartográficos de las ciudades, levantamiento de mapas con nuevos desarrollos y revisión de las áreas de antiguo

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 PÁGINA : 36 FECHA: 23-05-07</p>
---	---	--

desarrollo a fin de registrar la creación o eliminación de manzanas y la variación de límites con referencia a la antigua cartografía.

- Recuento de edificaciones y viviendas, elaborando un listado con las direcciones e identificaciones de las viviendas por manzanas, para todas las secciones cartográficas de la ciudad; por creación de nuevas manzanas, cambios en las sectorizaciones cartográficas y actualización del número de viviendas por manzana. La información se registra en el formulario de recuento de edificaciones y viviendas. El recuento para la encuesta rural, tiene diferencias con respecto a la encuesta urbana, sobre todo en los segmentos que pertenecen a caseríos, corregimientos y área rural dispersa. En la encuesta rural este proceso también recibe el nombre de segmentación, por el cual se localiza, delimita y se realiza el recuento de las áreas seleccionadas en la muestra, así mismo, se realiza una identificación detallada de acuerdo con límites naturales o artificiales de los segmentos seleccionados. El proceso de recuento en la encuesta se realiza con base en el documento “Manual de recuento” de la Encuesta de Hogares.
- Cambios en la estratificación de las áreas existentes como efecto de la extensión de los servicios públicos, la pavimentación de las vías y el mejoramiento o empeoramiento de las viviendas de las áreas. La estratificación de las manzanas y de las secciones cartográficas se hace según la ubicación y características de las viviendas que las conforman; asignando a la manzana y a las secciones, el estrato modal observado, de acuerdo al estrato predominante en las manzanas que la componen.

Por su propia naturaleza, los cambios en los marcos muestrales se asocian con operaciones cartográficas de gran amplitud, que solamente tienen lugar de manera esporádica y se derivan usualmente de operaciones censales o de recuentos de viviendas que requieren de un aporte económico bastante grande.

El marco original de la muestra, es objeto de una actualización permanente a través de la misma muestra, cada trimestre. Esta actualización recoge los cambios operados en las áreas seleccionadas, tales como:

- Alteraciones físicas en la manzana seleccionada, ocasionadas por el continuo proceso de renovación urbana, con el consiguiente aumento o disminución en la cantidad de viviendas. Estas alteraciones pueden darse por modificación de parte de la manzana, la fusión de ella con otra, e inclusive, su misma desaparición.
- Crecimiento por nuevos desarrollos urbanos de aquellas manzanas o áreas con menos de 4 viviendas, que por lo mismo, fueron agregadas con otras de mayor tamaño para conformar una medida de tamaño.
- Creación de nuevas manzanas a partir de las unidades seleccionadas.
- Cambios en la nomenclatura de la sección seleccionada.
- Corrección de los defectos del recuento de viviendas.

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 PÁGINA : 37 FECHA: 23-05-07</p>
---	---	--

f) Modificación de los límites de las secciones seleccionadas.

Las variaciones anteriores causan algunas veces pérdida o exceso de la muestra, al aumentar o disminuir la cantidad de viviendas en el segmento seleccionado. Cuando se presentan cambios en el marco muestral, se aplican los siguientes procedimientos:

1.9.1.1. Actualización del segmento o área seleccionada

Una vez seleccionada una manzana, el segmento seleccionado aleatoriamente, se recorre completamente, delimitado por las direcciones inicial y final del mismo, de acuerdo con lo señalado en el recuento original. Esta operación se realiza en la misma manzana, durante el número de períodos, equivalentes al número de segmentos que hubiere en la manzana. Lo anterior significa que no puede omitirse en el recorrido ninguna fracción del segmento, bajo el supuesto de que las unidades que la componen no son viviendas.

1.9.1.2. Aumento en la cantidad de viviendas del segmento

Si el segmento ha crecido hasta 20 viviendas, se aplica la encuesta a todos los hogares de todas las viviendas del segmento. Si el segmento ha crecido de manera que tuviera más de 20 viviendas, antes de aplicar las encuestas, se procede a realizar un nuevo recuento de toda la manzana o manzanas que originaron el segmento. Con el nuevo recuento, el segmento se divide en tantas partes como Medidas de Tamaño Teóricas (MTT) tenía originalmente, con el fin de ajustar el factor de expansión de tal manera que reflejara este aumento al hacer las expansiones.

Se aplica la tabla de submuestreo para escoger la fracción indicada, esta se utiliza de la siguiente forma:

Se toma la fila que contiene el cociente entre el total de viviendas del nuevo recuento y el total de MT de la manzana, cruzada con la columna correspondiente a la cantidad de segmentos que se toman en cada periodo de recolección.

Se fragmenta la primera parte del nuevo recuento en tantas divisiones como indique la fracción de submuestreo y se selecciona aleatoriamente una de estas subdivisiones para encuestar todos los hogares que allí residen.

Se respeta siempre el número de MTT de las manzanas o manzanas (en caso de manzanas pequeñas) a que pertenece el segmento. Si quedasen aún MT por investigar en etapas posteriores, se procederá con ellas de idéntica manera como se hizo en la primera parte. Si la manzana tiene una sola M.T.T. se aplica el mismo procedimiento, pero esa manzana queda agotada para la próxima etapa.

**Tabla 5. Tabla para submuestreo
Medidas de tamaño teóricas de la manzana**

Tabla para submuestreo		
Número de unidades de viviendas listadas	Fracción de muestreo	Peso de submuestreo
0-20	1	1
21-40	$\frac{1}{2}$	2
41-60	$\frac{1}{3}$	3
61-80	$\frac{1}{4}$	4
81-100	$\frac{1}{5}$	4*
101-120	$\frac{1}{6}$	4*
121 y más	**	**
* El máximo peso de submuestreo es 4. Esto puede introducir un pequeño sesgo en las estimaciones, pero reduce el posible efecto de grandes varianzas muestrales.		
** Si una manzana u otra área asignada para listar, se encuentra que tiene más de 120 unidades de vivienda se debe consultar con el Grupo de Metodología Estadística del DANE Central.		

Por ejemplo, si se ha seleccionado una nueva manzana que tenía en el recuento original 46 viviendas, lo cual es equivalente a 4 MT aproximadamente y en el momento de hacer el recorrido en uno de estos segmentos delimitados por las direcciones inicial y final, se encuentran 25 viviendas, se hace necesario realizar un nuevo recuento a toda la manzana.

Si después de tal recuento se encuentran 120 viviendas, entonces se divide 120 en tantas partes como Medidas de Tamaño Teóricas (MTT) tenía originalmente, es decir 4 MTT, de donde se obtiene 30 viviendas en promedio cada una.

Se busca entonces en la tabla para submuestreo, que se presenta en la Tabla 4, la fracción de muestreo que corresponde al cruce de la segunda fila, que contiene a las 30 viviendas en promedio, con la primera columna, correspondiente a la fracción de muestreo que se toma en la primera etapa de la manzana.

La fracción resultante es $\frac{1}{2}$, por lo cual la primera parte del nuevo recuento es decir las 30 viviendas, debe subdividirse en dos partes aproximadamente iguales, en este caso cada una de 15 viviendas, de las cuales se selecciona una de ellas en forma aleatoria, para aplicar la encuesta. Las otras 15 viviendas se descartaban para aplicar la encuesta, pero se recuperan con el peso de submuestreo que para este caso sería 2.

1.9.1.3. Disminución en la cantidad de viviendas del segmento

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 PÁGINA : 39 FECHA: 23-05-07</p>
---	---	--

Si las viviendas del segmento seleccionado desaparecieron total o parcialmente, el faltante se considera pérdida de muestra para la etapa en estudio. No obstante, debe observarse toda la manzana para verificar lo que acontece en el resto de MT de la manzana.

1.9.2. Novedades

Algunas veces se pueden encontrar algunas novedades, entre ellas están:

1.9.2.1. Las direcciones seleccionadas no correspondían al número de la manzana según la cartografía

Pueden presentarse dos casos, en los cuales se aplican los siguientes procedimientos:

- a) Que las direcciones pertenezcan a otra manzana de la misma sección. En este caso, primero se verifican las direcciones, luego se procede a realizar las encuestas en la manzana a la que corresponden dichas direcciones y se corrige el número de la manzana en el recuento, en el marco de la muestra y en la cartografía.
- b) Que las direcciones no correspondan a ninguna manzana de la sección. Este caso indica que muy probablemente el recuento de la sección esta mal elaborado. Por consiguiente, se realiza un nuevo recuento a la sección, y se procede a seleccionar la manzana y el segmento correspondiente.

1.9.2.2. El segmento no se logra ubicar en el terreno

Pueden presentarse dos casos, en los cuales se aplican los siguientes procedimientos:

- a) Por demolición comprobada de las viviendas. En este caso se realiza una verificación de la cartografía de toda la sección en el terreno, haciendo las correcciones necesarias e introduciendo los cambios, desaparición o creación de manzanas en la cartografía y en el marco y se pasa el respectivo reporte a muestras, de donde solamente se selecciona un nuevo segmento en el caso que fuese un problema de marco, esto es, que el terreno sea un parque, una zona comercial o un área sin desarrollar; en caso contrario se da como pérdida de muestra, es decir se pierden las encuestas que se debían realizar en esta área, ya que reemplazar los segmentos inicialmente seleccionados alteraría las probabilidades de selección.
- b) Por cambios de nomenclatura en las direcciones. Se procede a establecer las equivalencias entre la nomenclatura antigua y nueva, en especial las concernientes al segmento: Si no es posible, entonces hay que reelaborar el recuento, actualizar las direcciones y seleccionar a partir de éste, la manzana y el segmento correspondientes, actualizando la cartografía y el marco de la muestra.

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 PÁGINA : 40 FECHA: 23-05-07</p>
---	---	--

BIBLIOGRAFÍA

KISH, Leslie. Muestreo de encuestas. México : Editorial Trillas S. A. 1979

CANAVOS, C. George. Probabilidad y Estadística. Aplicaciones y Métodos. México: Mc Graw-Hill, 1996.

COCHRAN, W., Sampling Techniques. New York : John Wiley and Sons, Inc., 1963.

DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADÍSTICA. Metodología de la Encuesta Nacional de Hogares. Bogotá. D.C.: DANE. 1985.

_____ Cartilla de cartografía censal. Santafé de Bogotá. D.C. : DANE. 1999.

_____ Encuesta Continua de Hogares. Ciudades y áreas metropolitanas. Manual de conceptos básicos y de recolección. Etapa 001. Santafé de Bogotá. D.C. : DANE. 2000.

_____ Guía Metodológica: Encuesta Nacional de Hogares. Santafé de Bogotá. D.C. : DANE. 1999.

_____ Guía para documentar la actividad estadística. Santafé de Bogotá. D.C. : DANE. 1998.

_____ Manual de Recuento. Encuesta de Hogares.

_____ Mejoras a la Encuesta Nacional de Hogares. Santafé de Bogotá. D.C. : DANE. 1996.

_____ Metodología de la Encuesta Nacional de Hogares. Santafé de Bogotá. D.C.: DANE. 1997.

_____ Metodología de la Estratificación de viviendas urbanas en los municipios. Santafé de Bogotá. D.C.: DANE. 1992.

_____ Resumen Metodológico Encuesta Nacional de Hogares. Santafé de Bogotá. D.C. : DANE. 1999.

DUQUE, Nelly; HINCAPIÉ Cecilia y DE LIBREROS Marión. 20 años de la Encuesta de Hogares en Colombia. 1970-1990. Santafé de Bogotá. D.C. : DANE. 1991.

GÓMEZ, Luis Carlos, et al. Ampliación de la Muestra Maestra de Hogares. Primer informe. Bogotá : DANE. 2001.

GÓMEZ, Luis Carlos. Modernización de la Encuesta Nacional de Hogares (ENH). Adecuación de la muestra maestra de hogares para la encuesta continua. Bogotá : DANE. 2000.

	<p align="center">METODOLOGÍA MUESTRAL GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH</p>	<p>CÓDIGO: DM-GEIH-DIM-01 VERSIÓN : 02 PÁGINA : 41 FECHA: 23-05-07</p>
---	---	--

GÓMEZ, Luis Carlos, et al. La muestra de la Encuesta Nacional de Hogares (ENH) y elementos de la muestra maestra nacional de población. Bogotá : DANE. 2000.

OSPINA, David. Introducción al muestreo. Bogotá, D.C : Universidad Nacional de Colombia. Facultad de Ciencias. Departamento de Matemáticas y Estadística. 2001.

SARNDAL, Carl., Swensson, Bengt. Model Assisted Survey Sampling. New York: Springer-Verlag. 1992.