

**UNIVERSIDAD DE LOS ANDES  
FACULTAD DE ECONOMÍA  
CENTRO DE ESTUDIOS SOBRE DESARROLLO ECONÓMICO - CEDE**



**ENCUESTA LONGITUDINAL SOBRE DINAMICA DE LOS HOGARES COLOMBIANOS  
PANEL DE HOGARES - PH**

**PH-D1-R1: DISEÑO DE LA MUESTRA**

**ENERO DE 2010**

## **CONFORMACION DEL EQUIPO TECNICO**

### **Comité Técnico Nacional**

Raquel Bernal  
Carmen Elisa Flórez  
Alejandro Gaviria  
Juliana Helo  
Ana María Ibáñez  
Christian Jaramillo  
Ximena Peña  
Álvaro Suárez

### **Componente Temático**

Paula Marcela Escobar  
Juliana Helo  
Álvaro Suárez

### **Componente Sistemas**

Martha Elena Reina  
Mercedes Tibavisco

### **Componente Operativo**

Héctor Pinilla

### **Componente Psicólogos y pruebas antropométricas**

Camila Fernández  
Paul René Ocampo

### **Elaboración de este documento**

Luis Carlos Gómez (muestrista)  
Juliana Helo  
Paula Marcela Escobar

## ÍNDICE

1. INTRODUCCIÓN .....	4
1.1. Universo de Estudio .....	4
1.2. Marco Muestral.....	4
1.3. Tipo de muestreo .....	4
1.4. Tamaño de la muestra.....	5
2. LA MUESTRA URBANA.....	7
2.1. Selección de las Unidades Primarias de Muestreo .....	7
2.1.1. Definición de Regiones .....	7
2.1.2. Definición de estratos .....	8
2.1.3. Primera fase: patrones por región .....	9
2.1.4. Segunda fase: patrones finales .....	15
2.1.5. Selección final del patrón .....	20
2.1.6. Selección de las unidades primarias dentro del patrón final .....	20
2.1.7. Distribución de la muestra en las UPM .....	23
2.2. Selección de las Unidades Secundarias de Muestreo .....	24
2.3. Selección de las Unidades Terciarias de Muestreo .....	25
3. LA MUESTRA RURAL.....	26
3.1. Selección de las Unidades Primarias de Muestreo .....	26
3.2. Selección inicial de las Unidades Secundarias de Muestreo .....	27
3.3. Selección final de Unidades Secundarias de Muestreo .....	27
3.4. Selección de las Unidades Terciarias de Muestreo .....	28
4. METODOLOGÍA PARA EL CÁLCULO DE LOS FACTORES DE EXPANSIÓN .....	29
4.1. Factores de expansión de la muestra urbana .....	29
4.1.1. Probabilidades de selección .....	29
4.1.2. Restitución del universo de hogares seleccionados no encuestados .....	30
4.1.3. Correcciones por totales y por estructura de la población .....	30
4.1.4. Factor de corrección alrededor de la unidad .....	31
4.2. Factores de expansión de la muestra rural .....	31
4.2.1. Probabilidades de selección .....	31
4.2.2. Restitución del universo de hogares seleccionados no encuestados y correcciones por totales y estructura de la población. ....	32

5. PRECISIÓN DE LAS ESTIMACIONES ..... 33  
Anexo. Modelo para Cálculo de la Precisión Observada - Fórmulas basadas en Series de Taylor ... 34

## 1. INTRODUCCIÓN

El presente documento expone las características del diseño muestral para la Encuesta Longitudinal sobre Dinámica de Hogares Colombianos, así como las metodologías implementadas para la selección de cada una de las unidades que componen la muestra en las zonas rural y urbana.

### 1.1. Universo de Estudio

El Universo de Estudio de la Encuesta en la zona urbana está constituido por los hogares particulares de estratos socioeconómicos 1 a 4; agrupados en regiones geográficas y pertenecientes a municipios con diferentes niveles de urbanización (tamaño de población). En la zona rural, el universo son todos los hogares de un conjunto de subregiones homogéneas internamente en lo agropecuario y entre ellas desde el punto de vista socioeconómico (estrato 1 básicamente).

### 1.2. Marco Muestral

El Marco Muestral está constituido por el inventario cartográfico y los archivos de viviendas, hogares y personas, agregados por manzanas para cabeceras municipales y centros poblados, obtenidos del Censo General de Población y Vivienda de 2005. En la zona rural, el marco está formado por el inventario cartográfico a nivel de veredas para las zonas rurales dispersas obtenido de los Planes de Ordenamiento Territorial o de la cartografía desarrollada en cada municipio, principalmente por las Secretarías de Planeación.

### 1.3. Tipo de muestreo

El diseño de la muestra es probabilístico, estratificado, multietápico y de conglomerados, como se presenta a continuación:

- a) Probabilístico: cada unidad de observación del universo tiene una probabilidad de selección conocida y superior a cero (0). En este tipo de muestra es factible realizar una inferencia válida del universo de estudio, determinar previamente a la recolección la precisión deseada de las estimaciones y calcular la precisión observada para cada indicador estimado<sup>1</sup>.
- b) De conglomerados: el diseño contempla diferentes tipos de conglomerados, cuya selección se hizo en etapas sucesivas. Para el universo urbano, los conglomerados son los siguientes: las cabeceras municipales (Unidades Primarias de Muestreo – UPM); manzanas (Unidades Secundarias – USM), y segmentos de viviendas contiguas (Unidades Terciarias de Muestreo - UTM). Para el universo rural: zonas rurales de los municipios (UPM); veredas (USM) y segmentos (UTM).
- c) Estratificado: las UPM para la zona urbana (municipios) fueron clasificadas o estratificadas en grupos homogéneos, de acuerdo con algunas variables independientes altamente correlacionadas con los principales indicadores de estudio: la región, el departamento, el nivel de urbanización (tamaño de la población) y el índice de Necesidades Básicas Insatisfechas (NBI). De cada estrato se seleccionó probabilísticamente una UPM.

---

<sup>1</sup> El error estándar es el indicador de precisión de las muestras probabilísticas.

d) Polietápico: se siguieron las siguientes etapas en la zona urbana:

- Primera etapa: selección de los municipios (UPM) dentro de cada estrato
- Segunda etapa: selección de las manzanas (USM) por el método de probabilidad proporcional al tamaño (PPT).
- Tercera etapa: selección de segmentos (UTM) dentro de cada manzana.

En la zona rural se siguieron las siguientes etapas:

- Primera etapa: selección de las zonas rurales de los municipios.
- Segunda etapa: selección de las veredas dentro de cada Unidad Primaria
- Tercera etapa: selección de los segmentos dentro de cada vereda.

#### 1.4. Tamaño de la muestra

La muestra urbana está compuesta por 6.000 hogares distribuidos óptimamente en las regiones Atlántica, Central, Oriental, Pacífica, Andén Pacífico y Bogotá. Sin embargo, por los costos del operativo y limitaciones financieras del proyecto, se excluyó la región del Andén Pacífico del primer levantamiento de información. Por esta razón, la muestra se redujo a 5.000 hogares. Previendo dificultades en el operativo como rechazos por parte de los hogares a las encuestadoras para responder la encuesta, se incluyó una sobremuestra del 20%, por lo cual la muestra completa fue de 6.000 hogares. En la zona rural, la muestra está compuesta por 4.000 hogares distribuidos en cuatro subregiones: Atlántica Media, Centro-oriente, Eje Cafetero y Cundiboyacense. Igualmente, se incluyó una sobremuestra del 20%, para una muestra total de 4.800 hogares.

La partición de la muestra de cada universo en fracciones iguales, se fundamenta en la relación no lineal entre tamaño de muestra y tamaño de universo, para el mismo nivel de precisión. Para universos mayores de 50.000 elementos, el tamaño de la muestra tiende a un valor constante.

La muestra urbana por nivel de urbanización o por región, y la muestra rural por región permiten, para indicadores tipo proporción, la estimación de cambios o diferencias:

- De baja magnitud (menos del 10%, es decir, buena precisión) en proporciones de 30% o más.
- De magnitud mediana (10 a 20%, precisión mediana) en proporciones inferiores a 20%.
- De mediana y alta magnitud (baja precisión), en múltiples particiones de la muestra de un NU o de una región.

Las muestras total urbana y total rural, permiten estimación de muy buena precisión de los cambios o diferencias, en proporciones, de variables de hogar. Para indicadores tipo promedio, la muestra de una región (urbana o rural), o de un nivel de urbanización permite la medición de cambios o diferencias pequeñas (menos del 10%, es decir, buena precisión), no solo en la muestra total sino en múltiples particiones de esta.

La precisión de las estimaciones de las variables de personas es mucho mayor, debido al más alto volumen de unidades de observación (cuatro veces mayor). Para subgrupos demográficos, con menos del 20% de la muestra de una región o nivel de urbanización, sólo se pueden captar cambios de mediana o alta magnitud en indicadores tipo proporción. En contraste, para indicadores tipo promedio, es factible la estimación de cambios pequeños, inclusive para múltiples particiones adicionales de la muestra en subgrupos demográficos<sup>2</sup>.

<sup>2</sup> Tomado de Gómez, LC. Diseño de la Muestra – Documento Preliminar. 2007.



## 2. LA MUESTRA URBANA

La muestra urbana de la Encuesta Longitudinal Sobre Dinámica de los Hogares Colombianos - Panel de Hogares (PH) es representativa de la población urbana de estratos uno a cuatro, en seis regiones del país y por seis niveles de urbanización. Está compuesta por 6000 hogares distribuidos óptimamente en las regiones Atlántica, Central, Oriental, Pacífica, Andén Pacífico y Bogotá.

### 2.1. Se lección de las Unidades Primarias de Muestreo<sup>3</sup>

La selección de la muestra de unidades primarias urbanas no definidas del panel de hogares (PH) se realizó con la técnica de “selección controlada” desarrollada por Goodman y Kish en 1950. Los autores definen el procedimiento como “cualquier proceso de selección en el cual, manteniendo la probabilidad asignada a cada unidad, las probabilidades de selección para algunas o todas las combinaciones preferidas de “n” unidades de un total de “N”, son más grandes que en el muestreo aleatorio estratificado. En forma correspondiente, las probabilidades de combinaciones no preferidas son más pequeñas que en el muestreo estratificado.”<sup>4</sup> El objetivo principal radica en que la muestra esté óptimamente distribuida de acuerdo con algunas variables de interés. En este caso, geográficamente, por el índice de necesidades básicas insatisfechas (NBI) y por el nivel de urbanización (tamaño de la población).

#### 2.1.1. Definición de Regiones

Se establecieron regiones de estudio para la encuesta urbana del PH: Atlántica, Oriental, Central, Pacífica, Andén Pacífico y Bogotá. Cada región a su vez, se dividió en sub-grupos con el fin de garantizar la distribución de la muestra en el mayor número de departamentos del país. La tabla 1 muestra la división en las regiones.

**Tabla 1. Regiones y departamentos**

Región	Subregión	Departamento
Atlántica	A	Guajira
		Magdalena
		Cesar
	B	Bolívar
		Atlántico
	C	Córdoba
		Sucre
Oriental	A	Santander
		Norte de Santander
	B	Meta
		Cundinamarca
Central	A	Boyacá
		Antioquia
	B	Caldas
		Quindío
Pacífica	A	Risaralda
		Valle
	B	Cauca
		Nariño
Andén Pacífico	A	Chocó

<sup>3</sup> Sección basada en el documento: Gómez, L.C y Losee, G (1968). *La Muestra de la investigación nacional de morbilidad – Estudio de recursos humanos para la salud y educación médica en Colombia* – Ministerio de Salud Pública

<sup>4</sup> Goodman, R. y Kish, L. (1950). *Controlled Selection: A Technique in Probability Sampling* Journal of the American Statistical Association, 45: 350 - 372



Región	Subregión	Departamento
	B	Nariño
		Valle
Bogotá	A	Bogotá

### Índice de Necesidades Básicas Insatisfechas

El índice de Necesidades Básicas Insatisfechas (NBI) se dividió en tres niveles para definir la riqueza de las unidades primarias de muestreo. Los límites se establecieron de forma diferencial en las regiones, como se muestra en la tabla 2.

**Tabla 2. Definición de los niveles de NBI**

Región	NBI Bajo	NBI Medio	NBI Alto
Central, Oriental y Pacífica	Menor de 20	Entre 20 y 34.99	Mayor a 35
Atlántica	Menor de 35	Entre 35 y 54.99	Mayor a 55
Andén Pacífico		Menor a 54.99	Mayor a 55

### Nivel de urbanización

Se definieron los siguientes límites de población urbana para clasificar el nivel de urbanización de una unidad primaria no definida de muestreo.

**Tabla 3. Niveles de urbanización**

Nivel	Límite Inferior	Límite Superior
1	3,000,000	
2	1,000,000	2,999,999
3	400,000	999,999
4	100,000	399,999
5	20,000	99,999
6	2000	19,999

La probabilidad de selección de las unidades primarias no definidas de muestreo con nivel de urbanización 1, 2 o 3 fue uno, lo que implicó que fueron municipios de inclusión forzosa en la muestra urbana. En total se establecieron 11 municipios en esta categoría: Bogotá, Medellín, Barranquilla, Cali, Santa Marta, Cartagena, Cúcuta, Bucaramanga, Soacha, Ibagué y Pereira. Es importante definir que dentro de la población urbana de cada municipio se incluyó el número de personas en el área metropolitana.

Los municipios con población urbana menor a 2.000 habitantes se unieron a otros aledaños para generar unidades primarias de muestreo no definidas, con el requisito de población establecido en los límites de la tabla 3.

#### 2.1.2. Definición de estratos

El primer paso en la construcción de estratos por región consistió en la organización de unidades primarias de muestreo no definidas por nivel de urbanización, con el fin de identificar aquellas de inclusión forzosa. En segundo lugar se definió el tamaño promedio de los estratos a construir en cada una de las regiones. La población urbana de las unidades con probabilidad de selección diferente a uno, constituyó el numerador para determinar este tamaño. El denominador fue el número de estratos que se debía construir y que se determinó de acuerdo al tamaño de cada región, según se muestra en la tabla 4.

**Tabla 4. Número de estratos por región**

Región	Estratos
Atlántica	7
Oriental	11
Central	10
Pacífica	8
Andén Pacífico	8
Inclusión Forzosa	11
Total	55

Por lo tanto, el tamaño promedio de estratos en la región  $i$  se definió de acuerdo a la siguiente fórmula después de suprimir las unidades primarias de inclusión forzosa:

$$Tam\_Estrato_i = \frac{Población\ Urbana_i}{Número\ de\ Estratos_i}$$

Se permitió que el tamaño promedio de los estratos en cada región variara en un 25 por ciento hacia arriba o hacia abajo. Por ejemplo, si el tamaño promedio era 500.000 personas, el intervalo de población era [375.000, 625.000]. Una unidad primaria de muestreo no definida con población superior al límite inferior constituyó un estrato y la probabilidad de selección fue 1. Con este criterio, se incluyeron directamente 9 unidades en la muestra urbana: Villavicencio, Manizales, Neiva, Pasto, Popayán, Palmira, Quibdó, Buenaventura y Tumaco.

Sin tener en cuenta las unidades de inclusión forzosa, se organizaron las restantes en cada región de acuerdo a la población requerida de cada estrato, a su nivel de urbanización, al grupo geográfico de pertenencia y a su NBI. Se agruparon de tal manera que dentro de cada estrato las unidades de muestreo no definidas que lo componían fueran similares en NBI, nivel de urbanización y ubicación geográfica, pero que cada estrato fuese muy diferente a los demás.

### 2.1.3. Primera fase: patrones por región

La primera fase de la selección controlada consistió en la construcción de un juego de patrones (o combinaciones de unidades primarias no definidas de muestreo) para cada región, con base en los criterios de control. Esto permitió asegurar la distribución adecuada de la muestra a nivel geográfico y de riqueza.

El proceso inició con el cálculo de la probabilidad de selección de cada una de las unidades primarias de muestreo no definidas en los estratos, con base en la siguiente fórmula:

$$Pr(\text{Selección UPM}) = \frac{Población\ UPM}{Población\ Estrato}$$

En la tabla 5 se muestra el estrato 1 de la región atlántica con su respectiva probabilidad de selección. Las probabilidades debían sumar 1 dentro de cada estrato.

**Tabla 5: Estrato 1 – Región Atlántica**

Municipio	Departamento	Nivel de Urbanización	Población	NBI	Nivel NBI	Probabilidad de Selección
VALLEDUPAR	CESAR (A)	4	328.891	27,16	1	0,56
MALAMBO	ATLANTICO (B)	4	100.993	31,64	1	0,17

RIOHACHA	GUAJIRA (A)	4	159.846	39,75	2	0,27
		TOTAL	589.730			1,00

Con las probabilidades individuales de cada unidad primaria de muestreo no definida se construyeron las tablas de probabilidad acumulada por estrato (Tabla 6), en donde se resumió la probabilidad de selección de cada clase. La clase se refiere a la combinación de grupo de departamentos y nivel de NBI. En el caso de la región Atlántica en donde el número de grupo de departamentos fue 3 y los niveles de NBI también fueron 3, la combinación de los dos criterios arrojó 9 posibles clases. A-NBI Bajo, A-NBI Medio, A-NBI Alto, B-NBI Bajo, B-NBI Medio, B-NBI Alto, C-NBI Bajo, C-NBI Medio, C-NBI Alto.

**Tabla 6: Probabilidades Acumuladas – Estrato 1 – Región Atlántica**

Grupo de Departamentos	NBI Bajo	NBI Medio	NBI Alto	Total
A	0.56	0.27		0.83
B	0.17			0.17
C				
Total	0.73	0.27		1.00

Se observa, por ejemplo, que la probabilidad de seleccionar una unidad del grupo A con un nivel de NBI bajo en el estrato 1 de la región Atlántica es 0.56. En este caso, esta es la probabilidad que corresponde a escoger Valledupar (Tabla 5). Si en el estrato se encontraban dos unidades pertenecientes al mismo grupo de departamento y con el mismo nivel de NBI, la probabilidad registrada en la tabla 6 era la suma de las probabilidades de cada una de las unidades y que se obtenía de la tabla 5.

Por otro lado, los totales marginales indican la probabilidad de seleccionar unidades pertenecientes a un determinado grupo de departamentos en la región o a un nivel de NBI particular. Entonces, la probabilidad de escoger una unidad del grupo A independiente del nivel de NBI fue 0.83 en el estrato 1 de la región Atlántica. Por último, el total general de la tabla (1.00) significa que en el estrato debe seleccionarse, por lo menos, una unidad primaria de muestreo no definida.

A partir de las tablas de probabilidades acumuladas por estrato, se elaboró una tabla de resumen por región (Tabla 7). En este caso, se sumó cada uno de los componentes o clases en cada uno de los estratos.

**Tabla 7. Probabilidades Acumuladas de la Región Atlántica**

Grupo de Departamentos	NBI Bajo	NBI Medio	NBI Alto	Total
A	0.77	1.55	0.36	2.68
B	0.40	0.63	0.83	1.85
C	0.57	1.69	0.22	2.47
Total	1.73	3.86	1.40	7.00

La probabilidad de selección de unidades con características grupo de departamentos A (Guajira, Magdalena, Cesar) y NBI medio en la región Atlántica es, por ejemplo, 1.55. Este número significa que se debía seleccionar por lo menos 1 unidad del grupo A y NBI medio, con la posibilidad de escoger una segunda. La probabilidad se descompone así: la probabilidad de escoger una unidad es  $0.45 \times 1 = 0.45$ ; la probabilidad de seleccionar dos unidades, cada una con igual probabilidad de selección, es  $0.55 \times 2 = 1.1$ . Esto es válido porque  $0.45 + 1.1 = 1.55$  y porque, en este caso, la probabilidad de escogencia de dos unidades es mayor que cuando se plantea la posibilidad de escoger sólo una.

De la tabla 7 también se observa que el total acumulado del grupo A de departamentos es 2.68, lo cual indica que cómo mínimo se debía escoger dos unidades que individualmente tienen una probabilidad de selección de 0.32 (complemento a 1 de la fracción decimal 0.68) o tres, con

probabilidades individuales de 0.68. El total general, 7.00, muestra que en la región atlántica se debía escoger 7 unidades. Este ejercicio se repitió para todos los estratos y todas las regiones.

La Tabla 7 constituyó la base para la construcción de una tabla de control (Tabla 8a), en donde se definieron las características de los patrones<sup>5</sup> dentro de los requisitos originados en los totales marginales de las probabilidades acumuladas. La primera columna de la Tabla 8a contiene los códigos de los grupos de departamentos (A, B, C) y del nivel de NBI (1, 2, 3). La segunda columna (X) es el máximo y mínimo de unidades primarias no definidas de muestreo que se podían seleccionar (posibilidades alternativas) en relación con cada una de las características. La tercera (Px1) son las probabilidades de selección de cada una de las alternativas. X y Px1 son el desglose de las probabilidades marginales de las tablas de resumen de cada región.

En el encabezado de la tabla, i es el número de orden de los diferentes patrones; Pi es la probabilidad asignada a cada patrón y  $\sum Pi$  son las probabilidades Pi acumuladas sucesivamente. El símbolo + señala el control. Se explicará en párrafos posteriores el mecanismo para construir las columnas restantes.

**Tabla 8a. Control de los patrones de la primera etapa – Región Atlántica**

		0.3				0.9	1.0								
	$\sum Pi$	1	0.45			9	0								
	Pi	0.3				0.0	0.0								
	i	1	0.14			2	1								
		1	2			13	14								
Características	X	Px1		Px2		Px3	.	.	.	.	Px1 3		Px1 4		
<b>A</b>	2	0.32		0.32		0.32	.	.	.	.	0.02	+	-		-
	3	0.68	+	0.36	+	0.22	.	.	.	.	0.01		0.01	+	-
<b>B</b>	1	0.15		0.15		0.15	.	.	.	.	-		-		-
	2	0.85	+	0.54	+	0.40	.	.	.	.	0.03	+	0.01	+	-
<b>C</b>	2	0.53	+	0.21	+	0.07	.	.	.	.	0.01		0.01	+	-
	3	0.47		0.47		0.47	.	.	.	.	0.02	+	-		-
<b>1</b>	1	0.27		0.27		0.27	.	.	.	.	-		-		-
	2	0.73	+	0.42	+	0.28	.	.	.	.	0.03	+	0.01	+	-
<b>2</b>	3	0.14		0.14		0.14	.	.	.	.	0.01		0.01	+	-
	4	0.86	+	0.55	+	0.41	.	.	.	.	0.02	+	-		-
<b>3</b>	1	0.60	+	0.28	+	0.14	.	.	.	.	0.02	+	-		-
	2	0.40		0.40		0.40	.	.	.	.	0.01		0.01	+	-

La Tabla 8a tuvo una extensión que se elaboró a partir de las probabilidades acumuladas por región, en donde se estableció el número máximo y mínimo de unidades a seleccionar por clase en

<sup>5</sup> Los patrones son una combinación de clases en cada uno de los estratos.

cada estrato. Por ejemplo, en la región Atlántica, de la clase A2 se podía tener 1 o 2 unidades con probabilidades respectivas de 0.45 y 0.55. La tabla 8b muestra esta extensión:

**Tabla 8b. Extensión a control de patrones de la primera etapa – región Atlántica**

$\sum P_i$	0.3	0.4	0.9	1.0
$P_i$	1	5	9	0
	0.3	0.1	0.0	0.0
	1	4	2	1
	1	2	13	14

Características	X	Px1		Px2		Px3	.	.	.	.	Px1 3		Px1 4		
<b>A1</b>	1	0.7 7	+	0.4 5	+	0.3 1	.	.	.	.	0.02		0.02	+	-
	0	0.2 3		0.2 3		0.2 3	.	.	.	.	0.02	+	-		-
<b>A2</b>	1	0.4 5		0.4 5	+	0.3 1	.	.	.	.	0.01		0.01	+	-
	2	0.5 5	+	0.2 3		0.2 3	.	.	.	.	0.02	+	-		-
<b>A3</b>	1	0.3 6		0.3 6	+	0.2 2	.	.	.	.	0.01		0.01	+	-
	0	0.6 4	+	0.3 2		0.3 2	.	.	.	.	0.02	+	-		-
<b>B1</b>	1	0.4 0		0.4 9	+	0.2 6	.	.	.	.	0.03	+	0.01	+	-
	0	0.6 0	+	0.2 9		0.2 9	.	.	.	.	-		-		-
<b>B2</b>	1	0.6 3	+	0.3 2	+	0.1 8	.	.	.	.	-		-		-
	0	0.3 7		0.3 7		0.3 7	.	.	.	.	0.03	+	0.01	+	-
<b>B3</b>	1	0.8 3	+	0.5 1		0.5 1	.	.	.	.	0.03	+	0.01	+	-
	0	0.1 7		0.1 7	+	0.0 3	.	.	.	.	-		-		-
<b>C1</b>	1	0.5 7	+	0.2 6		0.2 6	.	.	.	.	0.02	+	-		-
	0	0.4 3		0.4 3	+	0.2 9	.	.	.	.	0.02		0.02	+	-
<b>C2</b>	1	0.3 1	+	-		-	.	.	.	.	-		-		-

	2	0.6 9		0.6 9	+	0.5 5	.	.	.	.	0.03	+	0.01	+	-
<b>C3</b>	1	0.2 2		0.2 2		0.2 2	.	.	.	.	-		-		-
	0	0.7 8	+	0.4 7	+	0.3 3	.	.	.	.	0.03	+	0.01	+	-

Adicional a la tabla de control (Tabla 8) se construyó la tabla de patrones (Tabla 9), que consistió en un listado de cada uno de los estratos, las clases<sup>6</sup> que lo componen y sus respectivas probabilidades de selección. Esta tabla se construyó a partir de la tabla 6 de cada estrato. El símbolo + en la tabla indica la clase que se debía escoger en cada estrato para componer un patrón.

**Tabla 9. Patrones de Selección de la Primera Etapa**

$\Sigma P_i$	0.3	0.4	0.9	1.0
$P_i$	1	5	9	0
$i$	0.3	0.1	0.0	0.0
	1	4	2	1
	1	2	13	14

Estratos	Clases	Px1		Px2		Px3	.	.	.	.	Px1 3		Px1 4		
1	A1	0.5 6	+	0.2 5	+	0.1 1	.	.	.	.	-		-		-
	A2	0.2 7		0.2 7		0.2 7	.	.	.	.	0.03	+	0.01	+	-
	B1	0.1 7		0.1 7		0.1 7	.	.	.	.	-		-		-
2	C1	0.5 7	+	0.2 6		0.2 6	.	.	.	.	0.02	+	-		-
	C2	0.4 3		0.4 3	+	0.2 9	.	.	.	.	0.02		0.02	+	-
3	A1	0.0 7		0.0 7		0.0 7	.	.	.	.	-		-		-
	A2	0.2 6		0.2 6		0.2 6	.	.	.	.	0.02	+	-		-
	B1	0.1 9		0.1 9	+	0.0 5	.	.	.	.	0.02		0.02	+	-
	B2	0.4 9	+	0.1 7		0.1 7	.	.	.	.	-		-		-
4	A2	0.2 5		0.2 5	+	0.1 1	.	.	.	.	-		-		-
	C2	0.7 5	+	0.4 4		0.4 4	.	.	.	.	0.03	+	0.01	+	-
5	A2	0.4	+	0.1		0.1	.	.	.	.	-		-		-

<sup>6</sup> Las clases son la combinación de ubicación geográfica y nivel de NBI que se encuentra en cada estrato. Por ejemplo, en el estrato 1 de la región Atlántica se encuentran las clases A-NBI Bajo, A-NBI Medio y B-NBI Bajo.

		3		2		2								
	A3	0.0 8		0.0 8		0.0 8	.	.	.	.	-		-	-
	B3	0.3 2		0.3 2		0.3 2	.	.	.	.	0.01		0.01	+
	C2	0.1 7		0.1 7	+	0.0 3	.	.	.	.	0.02	+	-	-
<b>6</b>	A1	0.1 3		0.1 3		0.1 3	.	.	.	.	0.02		0.02	+
	A2	0.3 4	+	0.0 3		0.0 3	.	.	.	.	-		-	-
	B1	0.0 4		0.0 4		0.0 4	.	.	.	.	0.02	+	-	-
	B2	0.1 4		0.1 4	+	-	.	.	.	.	-		-	-
	C2	0.3 4		0.3 4		0.3 4	.	.	.	.	-		-	-
<b>7</b>	A3	0.2 8		0.2 8	+	0.1 4	.	.	.	.	0.01		0.01	+
	B3	0.5 0	+	0.1 9		0.1 9	.	.	.	.	0.02	+	-	-
	C3	0.2 2		0.2 2		0.2 2	.	.	.	.	-		-	-

La construcción de cada patrón inició con el establecimiento de los controles. En la región Atlántica, por ejemplo, se debían seleccionar 7 clases. De acuerdo con la Tabla 8a, existían diversas maneras de lograrlo. En cuanto a los grupos de departamentos se podían escoger 2 o 3 unidades del grupo A, 1 o 2 del grupo B y 2 o 3 del grupo C. La idea era buscar una combinación de alternativas que en total sumara 7 unidades primarias de muestreo no definidas. Tal como se estableció en la columna  $i = 1$  de la Tabla 8a, el patrón escogido (indicado por +) tendría que tener 3 unidades del grupo A, 2 del B y 2 del C. El análisis para el número de unidades por nivel de NBI se hizo de la misma manera. Se debía escoger 7 con una combinación de niveles 1, 2 o 3. En la Tabla 8b se escogieron los controles de la misma manera.

Los controles señalados en la tabla de control (Tabla 8a) con una cruz (+) fueron la base para la construcción del primer patrón:

**Tabla 10. Control para primer patrón en la región Atlántica**

Característica	X	Probabilidad
A	3	0.68
B	2	0.85
C	2	0.53
1	2	0.73
2	4	0.86
3	1	0.6
A1	1	0.77
A2	2	0.55
A3	0	0.64
B1	0	0.60
B2	1	0.63
B3	1	0.83
C1	1	0.57
C2	1	0.31

C3	0	0.78
----	---	------

Dado el número máximo de unidades que podía contener el patrón, se escogió en la Tabla 9 una combinación de clases que cumpliera con las restricciones. Se eligió una clase por estrato, con la idea de buscar aquellas con mayor probabilidad de selección. El primer patrón se señala con una cruz (+) y se resume en la siguiente tabla:

**Tabla 11. Primer patrón de la región Atlántica**

Estrato	Clases	Probabilidad
01	A1	0.56
02	C1	0.57
03	B2	0.49
04	C2	0.75
05	A2	0.43
06	A2	0.34
07	B3	0.50

Se observa que la tabla 11 cumple con los requisitos del control. La característica A se debe encontrar 3 veces y lo hace en el estrato 01, 05 y 06. Lo mismo ocurre con las clases. El control establece que se deben encontrar dos unidades A2 y en el patrón final se encuentra en los estratos 05 y 06.

Al primer patrón de la región Atlántica se asignó la probabilidad 0.31 que corresponde a la probabilidad más baja de todas las listadas en la Tabla 10. Con esta probabilidad se construyó la columna Px2, al restarla de las probabilidades de clases y controles considerados en el primer patrón. La probabilidad de las demás clases y controles se mantuvo igual. El procedimiento anterior se repitió de manera sucesiva para la construcción de los demás patrones. Este proceso concluyó cuando simultáneamente en el último patrón las probabilidades de clases y controles tomaron el valor de 0 y  $\sum P_i$  alcanzó 1.00.

Después de construir el juego de patrones para la región atlántica, se procedió de la misma manera para las restantes cuatro, Oriental, Central, Andén Pacífico y Pacífica. En la región Bogotá no se desarrolló este procedimiento por tener sólo una unidad primaria de muestreo.

#### 2.1.4. Segunda fase: patrones finales

A partir de los patrones “parciales” de cada región construidos en la primera etapa, se construyó un “patrón definitivo” que consideraba simultáneamente las 5 regiones y Bogotá. En este caso, se tomó como criterio único de control el nivel de NBI. El criterio de departamentos sólo tenía importancia a nivel regional.

El primer paso en la construcción del juego final fue la elaboración de las probabilidades acumuladas (Tabla 12), en donde se resumían las probabilidades marginales del nivel de NBI en cada una de las regiones (a partir de Tabla 7).

**Tabla 12. Probabilidades acumuladas regiones Atlántica, Central, Oriental, Pacífica y Andén Pacífico**

Región	NBI Bajo (1)	NBI Medio (2)	NBI Alto (3)	Total
Atlántica	1.73	3.86	1.40	7.00
Oriental	6.25	2.96	0.79	10.00
Central	3.59	2.84	1.57	8.00
Pacífica	3.44	1.10	0.47	5.00



Andén	0.00	2.00	3.00	5.00
Total	15.02	12.75	7.23	35.00

La probabilidad del nivel 1 o bajo de NBI fue 15.02, lo que significa que se debía seleccionar, por lo menos, 15 unidades con esta característica. La probabilidad de escoger 15 era 0.98 mientras que la de selección de 16 era 0.02. El valor total, 35, indica el número de unidades primarias no definidas de muestreo a seleccionar. Si a 35 se le suma el total de unidades de inclusión forzosa (11 por nivel de urbanización y 9 por población superior al límite inferior del tamaño promedio del estrato) se obtiene 55, que era el número de estratos a seleccionar inicialmente.

Al igual que en la primera fase, se construyó una tabla de control (Tabla 13) a partir de las probabilidades acumuladas por región (Tabla 12). La columna X corresponde al número de unidades a seleccionar en cada uno de los niveles de NBI y Px1 a la probabilidad inicial. El resto de columnas se diligenciaron de manera simultánea con la tabla de patrones, cuyo procedimiento se explica en los párrafos siguientes.

**Tabla 13. Control de patrones de la segunda etapa – todas las regiones**

$\sum P$	0.0	0.1	0.3	0.9	0.9	1.0
i	4	8	8	8	9	0
Pi	0.0	0.1	0.2	0.0	0.0	0.0
I	4	4	0	1	1	1
	1	2	3	21	22	23

NB I	X	Px1		Px2		Px3					Px2 2		Px2 3		
1	15	0.98	+	0.94	+	0.80	+	.	.	.	0.01	+	-		-
	16	0.02		0.02		0.02		.	.	.	0.01		0.01	+	-
2	12	0.25	+	0.20		0.20		.	.	.	0.02	+	0.01	+	-
	13	0.75		0.75	+	0.61	+	.	.	.	-		-		-
3	7	0.77		0.77	+	0.63	+	.	.	.	0.01		0.01	+	-
	8	0.23	+	0.19		0.19		.	.	.	0.01	+	-		-

La tabla de patrones de la segunda etapa (Tabla 14) surgió a partir de los patrones de cada región con sus respectivas probabilidades. La primera columna indica el número del patrón parcial (i) que se encuentra en la Tabla 9 de alguna región en particular. La secuencia de clases en cada estrato corresponde al patrón particular al que se hace referencia en la columna 1. Por ejemplo, el primer patrón escogido en la región Atlántica según la Tabla 9 es A1 para el estrato 1, C1 para el 2, B2, C2, A2, A2, B3 para el resto de estratos. Esto es justamente lo que aparece debajo de cada uno de los estratos. La columna Px1 es la probabilidad de escogencia de cada patrón parcial, que también se obtiene de la Tabla 9.

**Tabla 14. Patrones de selección de la segunda etapa**

Región Atlántica										$\sum P_i$	0.042	0.183		0.988	1.00	
										$P_i$	0.042	0.141		0.007	0.01	
Patron	Estratos									i	1	2			22	23
Parcial	1	2	3	4	5	6	7			Px1	Px2	Px3		Px22	Px23	
1	A1	C1	B2	C2	A2	A2	B3			0.31	0.31	0.31	.	.	.	
2	A1	C2	B1	A2	C2	B2	A3			0.14	0.14 +	-	.	.	.	
3	A1	C1	B2	C2	A2	C2	B3			0.11	0.11	0.11	.	.	.	
4	B1	C2	A2	A2	B3	C2	C3			0.11 +	0.07	0.07	.	.	.	
5	A2	C1	A1	C2	B3	C2	A3			0.07	0.07	0.07	.	.	.	
6	A2	C2	A2	C2	B3	A1	C3			0.08	0.08	0.08	.	.	.	
7	B1	C1	A2	C2	A3	C2	B3			0.05	0.05	0.05	.	.	.	
8	A2	C2	B1	C2	B3	A2	A3			0.03	0.03	0.03	.	.	.	
9	A2	C2	B2	C2	A3	B1	C3			0.03	0.03	0.03	.	.	.	
10	B1	C2	B2	C2	A2	A1	A3			0.01	0.01	0.01	.	.	.	
11	A2	C2	B2	C2	B3	A1	A3			0.02	0.02	0.02	.	.	.	
12	A2	C1	B2	C2	C2	A1	B3			0.01	0.01	0.01	.	.	.	
13	A2	C1	A2	C2	C2	B1	B3			0.02	0.02	0.02	.	.	.	
14	A2	C2	B1	C2	B3	A1	A3			0.01	0.01	0.01	.	.	.	

**Tabla 14 (Continuación). Patrones de selección de la segunda etapa**

Región Oriental																	
Patrón Parcial	Estratos										Px1	Px2	Px3	Px22	Px23		
	1	2	3	4	5	6	7	8	9	10							
1	B1	B1	B1	B1	B2	A1	B1	A2	B2	A3	0.31	0.31	+	0.17	-	-	-
2	A2	B1	B1	B1	A2	B1	B1	A1	B2	B3	0.27	0.27		0.27	-	-	-
3	A2	A1	B1	B1	B3	B1	B1	B1	B2	A2	0.08	0.08		0.08	0.01	+	-
4	B1	A1	B1	B1	B3	A1	B1	B1	B2	A2	0.04	+	-	-	-	-	-
5	B1	A1	B1	B1	A1	B1	B1	B2	B2	A2	0.12	0.12		0.12	-	-	-
6	A2	B1	B1	B1	B2	A1	B1	B1	B2	B3	0.05	0.05		0.05	-	-	-
7	B1	B1	B1	B1	B3	A1	B1	B2	B2	A2	0.02	0.02		0.02	-	-	-
8	B1	B1	B1	B1	B2	A1	B1	A2	B2	B3	0.03	0.03		0.03	0.01	0.01	+
9	B1	B1	B1	B1	A1	B1	B1	B2	B2	A2	0.05	0.05		0.05	-	-	-
10	B1	B1	B1	B1	B2	A1	B1	B1	B2	A2	0.03	0.03		0.03	-	-	-
11	B1	A1	B1	B1	B2	B1	B1	B1	B2	A2	0.01	0.01		0.01	-	-	-

**Tabla 14 (Continuación). Patrones de selección de la segunda etapa**

Región Central																
Patrón Parcial	Estratos								Px1	Px2	Px3	Px22	Px23			
	1	2	3	4	5	6	7	8								
1	B1	A1	B1	C2	A3	A1	A2	C2	0.27	0.27	+	0.13	-	-	-	
2	A1	A2	B2	C1	A3	B1	C2	A3	0.20	0.20		0.20	-	-	-	
3	B1	A1	C2	B2	A3	B1	A2	C3	0.20	+	0.16	0.16	-	-	-	
4	B1	A1	C2	B1	A3	C1	A2	A3	0.10	0.10		0.10	-	-	-	
5	A1	A2	C1	B1	A3	A1	B2	C2	0.09	0.09		0.09	-	-	-	
6	B1	A1	B1	C2	A3	A1	A2	B2	0.02	0.02		0.02	-	-	-	
7	B1	A1	C1	B2	A3	B1	C2	A3	0.02	0.02		0.02	-	-	-	
8	B1	A1	C2	B1	A3	A1	A2	B2	0.02	0.02		0.02	-	-	-	
9	B1	A2	B1	C1	A3	A1	C2	B2	0.04	0.04		0.04	-	-	-	
10	B1	A1	B1	B2	A3	A1	C2	A3	0.02	0.02		0.02	0.01	+	-	
11	A1	A2	B1	C1	A3	A1	C2	B3	0.02	0.02		0.02	-	-	-	
12	B1	A1	C2	C1	A3	B1	A2	A3	0.01	0.01		0.01	0.01	0.01	+	

Tabla 14 (Continuación). Patrones de selección de la segunda etapa

## Región Pacífica

Patrón Parcial	Estratos					Px1	Px2	Px3	Px22	Px23
	1	2	3	4	5					
1	A1	B1	A1	A2	B3	0.17	0.17	0.17	-	-
2	A1	A1	A1	B2	B3	0.18	0.18 +	0.04	-	-
3	A1	B3	A1	A1	B2	0.07	0.07	0.07	-	-
4	A1	A1	B1	A1	B2	0.30 +	0.26	0.26	-	-
5	A1	A1	A1	B1	B2	0.09	0.09	0.09	0.02 +	0.01 +
6	A1	A2	A1	A1	B2	0.10	0.10	0.10	-	-
7	A1	B1	A1	A1	B2	0.05	0.05	0.05	-	-
8	A1	A1	A1	A2	B3	0.02	0.02	0.02	-	-
9	A1	A2	A1	A1	B3	0.02	0.02	0.02	-	-

Tabla 14 (Continuación). Patrones de selección de la segunda etapa

## Región Andén

Patrón Parcial	Estratos					Px1	Px2	Px3	Px22	Px23
	1	2	3	4	5					
1	C2	A2	B3	A3	A3	0.71 +	0.67 +	0.53	-	-
2	B2	A2	C3	A3	A3	0.24	0.24	0.24	0.01 +	-
3	B2	A2	B3	A3	A3	0.05	0.05	0.05	0.01	0.01 +

La formación de patrones finales y el diligenciamiento de las tablas se realizaron de manera similar a la primera etapa. Primero se estableció el control: 15 unidades de nivel NBI bajo, 12 de nivel medio y 8 de alto, para un total de 35 (señalado en la tabla 13 en la columna  $i = 1$  con signo +). Con este control, se estableció el patrón final con los siguientes patrones parciales:

Tabla 15. Ejemplo Primer Patrón Definitivo

Región	Patrón parcial	Probabilidad del patrón definitivo
Atlántica	4	0.04
Oriental	4	
Central	3	
Pacífica	4	
Andén Pacífico	1	

La tabla 16 muestra que el primer patrón definitivo cumplió con las restricciones impuestas en la tabla de control: 15 unidades de NBI bajo señaladas con verde, 12 de nivel 2 con azul y las restantes 8 de nivel tres en blanco.

Tabla 16. Verificación de restricciones

Region	Estrato									
	1	2	3	4	5	6	7	8	9	10
Atlántica	B1	C2	A2	A2	B3	C2	C3			
Oriental	B1	A1	B1	B1	B3	A1	B1	B1	B2	A2
Central	B1	A1	C2	B2	A3	B1	A2	C3		
Pacífica	A1	A1	B1	A1	B2					
Andén Pacífico	C2	A2	B3	A3	A3					

La menor probabilidad de los patrones parciales y de los controles fue 0.04, razón por la que ésta se le asignó al primer patrón definitivo. De las probabilidades de controles y patrones utilizadas en la elaboración del primer patrón, se restó 0.04. Las otras probabilidades permanecieron iguales. La probabilidad del patrón parcial 4 de la región oriental se eliminó tras la resta, lo que implicó que éste no fuese tomado en cuenta en la construcción del resto de patrones.

El procedimiento se repitió de manera sucesiva hasta completar los 23 patrones definitivos. Este número se alcanzó tras eliminar las probabilidades de cada uno de los patrones parciales y acumular 1.00 en  $\sum P_i$ .

### 2.1.5. Selección final del patrón

De los 23 patrones definitivos construidos en la segunda etapa se seleccionó el patrón final, con base en la probabilidad asignada a cada uno de ellos (probabilidad que se encuentra al inicio de la Tabla 14). Por ser la última probabilidad acumulada 1.000 se escogió un número aleatorio de tres decimales entre 0.000 y 1.000. El valor fue 0.756, que correspondió al patrón 8 (Tabla 17).

Tabla 17. Patrón Final seleccionado

Region	Estrato									
	1	2	3	4	5	6	7	8	9	10
Atlántica	A2	C1	A1	C2	B3	C2	A3			
Oriental	A2	B1	B1	B1	A2	B1	B1	A1	B2	B3
Central	B1	A1	B1	C2	A3	A1	A2	C2		
Pacífica	A1	B1	A1	A2	B3					
Andén Pacífico	B2	A2	C3	A3	A3					

### 2.1.6. Selección de las unidades primarias dentro del patrón final

El número de unidades primarias de muestreo (UPM) en cada clase seleccionada en los patrones parciales podía ser mayor a uno. Por ejemplo, en el estrato 3 de la región Atlántica se encontraban dos UPM pertenecientes a la clase A1. En caso que en la clase se encontrara sólo una unidad, esa UPM por definición formaba parte de la muestra. En contraste, si se encontraba más de una UPM en esa clase, se debía hacer una selección aleatoria de acuerdo a la probabilidad individual de selección de cada una de ellas. Tras acumular la probabilidad de selección de cada UPM en la

clase correspondiente, se debía escoger un número aleatorio entre 0 y el valor total de la probabilidad acumulada en la clase escogida para definir la UPM que entraría a la muestra.

La tabla 18 muestra un ejemplo para el estrato 3 de la región Atlántica. Con el patrón definitivo se estableció que en el estrato 3 se debía escoger aleatoriamente una UPM de clase A1. En este caso, dos unidades cumplían con ese requisito: San Juan del Cesar (Guajira) y Fonseca (Guajira). Las probabilidades de selección respectivas eran 0.04 y 0.03, que acumulaban 0.07. Se escogió un número aleatorio entre 0.00 y 0.07. En este caso, el valor fue 0.05, por lo que se escogió Fonseca – Guajira como representante del estrato 3 de la región Atlántica.

**Tabla 18. Ejemplo selección UPM en el Estrato 3 – Región Atlántica**

Codmpio	Municipio	Departamento	Nivel de Urbanización	Población	NBI	Nivel NBI	Grupo Departamento	Probabilidad de Selección	Probabilidad Acumulada
44650	SAN JUAN DEL CESAR	GUAJIRA	5	23090	24.58	1	A	0.04	0.04
44279	FONSECA	GUAJIRA	5	20443	27.95	1	A	0.03	0.07
8078	BARANOA	ATLANTICO	5	44634	25.46	1	B	0.07	0.15
8573	PUERTO COLOMBIA	ATLANTICO	5	21499	25.53	1	B	0.04	0.18
8685	SANTO TOMAS	ATLANTICO	5	23110	33.07	1	B	0.04	0.22
13468	MOMPOS	BOLIVAR	5	23548	32.84	1	B	0.04	0.26
8638	SABANALARGA	ATLANTICO	5	71278	36.33	2	B	0.12	0.38
8520	PALMAR DE VARELA	ATLANTICO	5	23373	37.35	2	B	0.04	0.41
8296	GALAPA	ATLANTICO	5	31540	37.78	2	B	0.05	0.47
8634	SABANAGRANDE	ATLANTICO	5	26028	40.11	2	B	0.04	0.51
13430	MAGANGUE	BOLIVAR	5	81721	49.09	2	B	0.14	0.64
13836	TURBACO	BOLIVAR	5	60288	49.65	2	B	0.10	0.74
20060	BOSCONIA	CESAR	5	29732	49.31	2	A	0.05	0.79
44430	MAICAO	GUAJIRA	5	92797	51.58	2	A	0.15	0.95
44560	MANAURE	GUAJIRA	5	31986	52.63	2	A	0.05	1.00
			TOTAL	605067					

Tras sucesivas repeticiones de este procedimiento se obtuvo la lista de UPM para la muestra urbana, la cual se resume en la tabla 19.

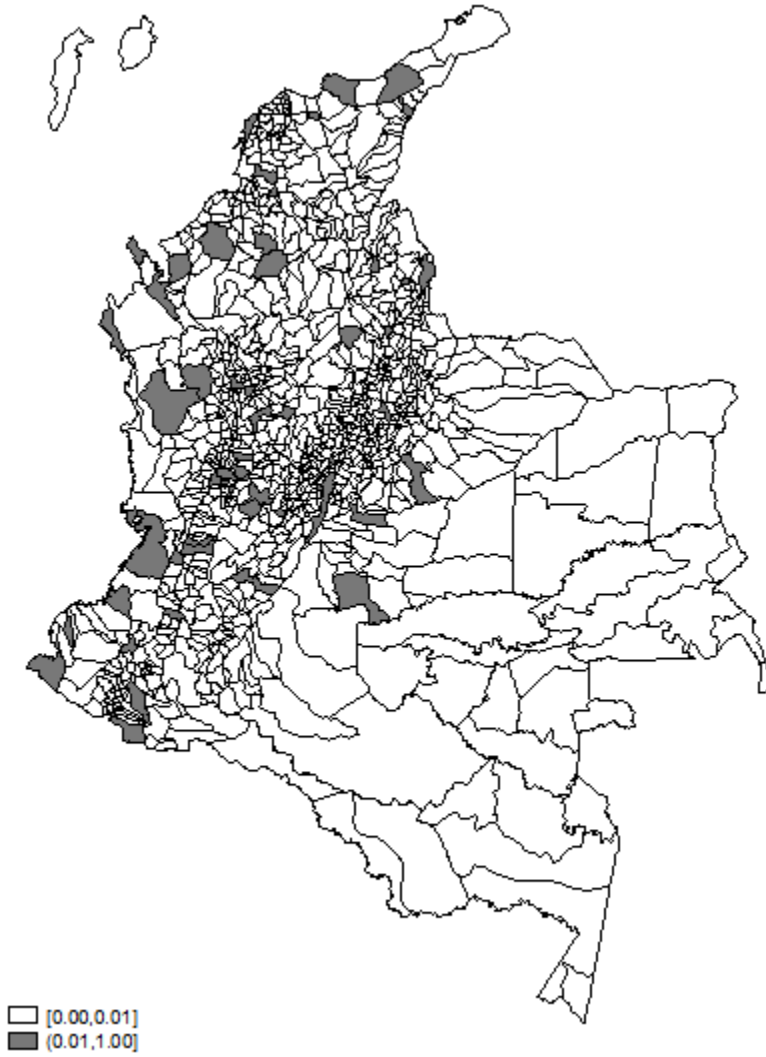
**Tabla 19. Muestra Urbana de Unidades Primarias de Muestreo (UPM)**

Región	Estrato	Clase	Codmpio	Municipio	Departamento
Bogotá	UPM de inclusión forzosa		11001	Bogotá	Bogotá
Atlántica			8001	Barranquilla	Atlántico
Atlántica			47001	Santa Marta	Magdalena
Atlántica			13001	Cartagena	Bolívar
Oriental			54001	Cúcuta	Norte de Santander
Oriental			68001	Bucaramanga	Santander
Oriental			25754	Soacha	Cundinamarca
Oriental			50001	Villavicencio	Meta
Central			5001	Medellín	Antioquia
Central			73001	Ibagué	Tolima
Central			66001	Pereira	Risaralda
Central			17001	Manizales	Caldas
Central			41001	Neiva	Huila
Andén pacífico			27001	Quibdó	Chocó
Andén pacífico			76109	Buenaventura	Valle
Andén pacífico			52835	Tumaco	Nariño
Pacífica			76001	Cali	Valle
Pacífica			52001	Pasto	Nariño
Pacífica			19001	Popayán	Cauca

Pacífica			76520	Palmira	Valle
Atlántica	1	A2	44001	Riohacha	Guajira
	2	C1	23001	Montería	Córdoba
	3	A1	44279	Fonseca	Guajira
	4	C2	23068	Ayapel	Córdoba
	5	B3	13244	El Carmen de Bolívar	Bolívar
	6	C2	70717	San Pedro	Sucre
	7	A3	47268	El Retén	Magdalena
Oriental	1	A2	68081	Barrancabermeja	Santander
	2	B1	15176	Chiquinquirá	Boyacá
	3	B1	25899	Zipaquirá	Cundinamarca
	4	B1	25430	Madrid	Cundinamarca
	5	A2	54498	Ocaña	Norte de Santander
	6	B1	15516	Paipa	Boyacá
	7	B1	25785	Tabio	Cundinamarca
	8	A1	54418	Lourdes	Norte de Santander
			54313	Gramalote	Norte de Santander
	9	B2	25260	El Rosal	Cundinamarca
10	B3	50711	Vistahermosa	Meta	
Central	1	B1	63001	Armenia	Quindío
	2	A1	5615	Rionegro	Antioquia
	3	B1	66682	Santa Rosa de Cabal	Risaralda
	4	C2	73268	Espinal	Tolima
	5	A3	5837	Turbo	Antioquia
	6	A1	5756	Sonsón	Antioquia
	7	A2	5847	Urrao	Antioquia
	8	C2	73624	Rovira	Tolima
Pacífica	1	A1	76147	Cartago	Valle
	2	B1	52356	Ipiales	Nariño
	3	A1	76248	El cerrito	Valle
	4	A2	76020	Alcalá	Valle
	5	B3	52405	Leiva	Nariño
Andén Pacífico	1	B2	52490	Olaya Herrera	Nariño
	2	A2	27006	Acandí	Chocó
	3	C3	19809	Timbiquí	Cauca
	4	A3	27600	Rio Quito	Choco
	5		27372	Jurado	Choco

De las UPM seleccionadas, 53 se constituyeron con un sólo municipio y dos estuvieron compuestos por dos municipios. Estos casos son el estrato 8 de la región Oriental y el estrato 5 de la región Andén Pacífico. El primero lo compone Lourdes y Gramalote en Norte de Santander y el segundo Jurado y Riosucio en Chocó.

El mapa a continuación muestra la distribución geográfica de las unidades primarias de muestreo urbanas seleccionadas resaltadas en gris.

**Mapa 1. Municipios de la muestra urbana**

### 2.1.7. Distribución de la muestra en las UPM

El siguiente paso a la selección de las unidades primarias de muestreo (UPM) fue la definición del número de hogares a encuestar en cada uno de estos municipios. En total, se debían distribuir 6.000 hogares más una sobre muestra del 20 por ciento en las seis regiones de estudio<sup>7</sup>. Por lo tanto, a cada región le correspondieron 1.200 hogares. La distribución de éstos se realizó proporcionalmente a la población del estrato que cada UPM representaba. Se definió que una medida de tamaño (MT) contendría aproximadamente 10 hogares. Este número podía variar entre 5 y 14 hogares<sup>8</sup>.

<sup>7</sup> La sobre muestra tenía por objeto cubrir las tasas de no respuesta, de rechazo o cualquier inconveniente que se presentara en el operativo de campo.

<sup>8</sup> Debido a la alta tasa de rechazos en los hogares de estrato 4, se decidió ampliar las Medidas de Tamaño a 13 hogares en promedio, variando entre 7 y 19, para las cuatro ciudades principales (Bogotá, Medellín, Cali, Barranquilla).



La tabla 20 muestra este procedimiento. En la primera columna (¿?) se listan las UPM de la región y en la segunda, la población urbana del estrato correspondiente. La tercera es la distribución porcentual que se obtiene al dividir la población del estrato por la población total de la región. Por último se calcula el número de MT que se debe escoger en cada UPM, por medio de la multiplicación de la distribución porcentual y el número de MT totales que se requiere en cada región. Dado que cada MT contiene en promedio 10 hogares, el total de hogares a escoger en cada región sería  $120 \times 10 = 1200$ .

**Tabla 20. Procedimiento para la distribución del número de hogares**

Municipio	Población Estrato	Distribución Porcentual	No. MT en la Muestra
X	1.000.000	$1.000.000/2.000.000 = 50\%$	$0,50 \times 120 = 60$
Y	700.000	$700.000/2.000.000 = 35\%$	$0,35 \times 120 = 42$
Z	300.000	$300.000/2.000.000 = 15\%$	$0,15 \times 120 = 18$
Total	2.000.000	100%	120

A manera de ejemplo, en la tabla 21 se muestra la distribución de la muestra en la región Atlántica.

**Tabla 21. Distribución de la muestra – Región Atlántica**

Codmpio	Municipio	Departamento	Población Estrato	Distribución %	No. MT en la Muestra
8001	Barranquilla	Atlántico	1.671.819	0,24	29
13001	Cartagena	Bolívar	876.334	0,13	16
47001	Santa Marta	Magdalena	409.480	0,06	7
44001	Riohacha	Guajira	589.730	0,09	10
23001	Montería	Córdoba	533.796	0,08	9
44279	Fonseca	Guajira	605.067	0,09	11
23068	Ayapel	Córdoba	450.110	0,07	8
13244	El Carmen de Bolívar	Bolívar	466.003	0,07	8
70717	San Pedro	Sucre	669.517	0,10	12
47268	El Retén	Magdalena	571.118	0,08	10
<b>Total</b>			6.842.974	1	120

## 2.2. Selección de las Unidades Secundarias de Muestreo

La distribución de la muestra en las diferentes unidades de muestreo (UPM) permitió la escogencia de las unidades secundarias de muestreo (USM), en este caso, manzanas estratos 1 a 4 de los municipios urbanos. La selección se basó en el listado de viviendas, hogares y personas (VIHOPE), construido a partir del Censo 2005 llevado a cabo por el Departamento Administrativo Nacional de Estadística (DANE). El VIHOPE contiene, para cada municipio, un listado de cada una de las manzanas urbanas (clase 1) con su identificador, el sector y la sección a la que pertenece y el número de vivienda y hogares en ella. A esta información se unió la información de estratificación del DANE y se eliminaron aquellas con estrato 5, 6 o 0.

A partir de esta información se determinó el número de MT en cada manzana, con la aproximación de la tabla 22. Si se encontraba una manzana con menos de 5 hogares, se debía unir a una manzana aledaña para cumplir con el requisito de tamaño. En consecuencia, entraban juntas en la selección final.

**Tabla 22. Equivalencia de hogares y medidas de tamaño (MT)**

Hogares	Número de MT
5 – 14	1
15 – 24	2
25 – 34	3
35 – 44	4
...	

Posteriormente se acumuló el número de MT en la UPM. La tabla 23 muestra un ejemplo del procedimiento para la UPM A de la tabla 20.

**Tabla 23. Procedimiento para determinar las MT en una UPM**

Sector	Sección	Manzana	Viviendas	Hogares	MT	MT Acumulada
01	01	01	100	110	11	11
		02	30	32	3	14
		...				...
	02	01	4	4	Se debe sumar con la siguiente manzana = 9	
		02	5	5	1	...
		...				
					Total	120

La tabla 23 se construyó para cada una de las UPM de la muestra urbana. A partir de ésta se determinó un intervalo de muestreo (IM) que permitió la selección específica de las manzanas. El IM se calculó por medio de la división del número de MT en la cabecera (Total en la tabla 23) entre el número de MT requeridas en cada UPM (Tabla 22).

$$IM = \frac{\text{No. de MT existentes en la cabecera}}{\text{No. de MT en la muestra}}$$

Se escogió un número aleatorio entre 1 y el IM y éste se ubicó en la columna MT Acumulada de la tabla 23. Esta posición determinó la primera manzana seleccionada de esa UPM. Por ejemplo, si el IM es 15, el número aleatorio se debe encontrar entre 1 y 15. Si se obtuvo el número 12, este número se ubica entre la fila 1 y la 2 de la tabla 23. Por lo tanto, se escoge la manzana correspondiente a la fila 2. La siguiente manzana del municipio se determinó al sumarle el IM al número de arranque. Este proceso se repitió de manera sucesiva hasta completar las manzanas correspondientes de cada UPM.

### 2.3. Selección de las Unidades Terciarias de Muestreo

Por cada USM seleccionada se debían encuestar en promedio 10 hogares. En los casos en que una manzana o una combinación de manzanas tuviesen una sola medida de tamaño (MT), se debía encuestar todos los hogares en esa USM. Sin embargo, era posible encontrar que una USM estuviese compuesta por más de una MT. En este caso, se desarrolló el siguiente procedimiento para la selección de los segmentos durante el operativo de campo. Primero, se realizó el recuento de todos los hogares y se dividió la manzana en pedazos iguales que contuviesen aproximadamente 10 hogares cada uno. Posteriormente se seleccionó de manera aleatoria uno de estos pedazos, que fue encuestado en su totalidad.

### 3. LA MUESTRA RURAL

La muestra rural es representativa de cuatro subregiones internamente homogéneas en su economía campesina, pero diferentes en su modelo económico: Atlántica, Cundiboyacense, Eje Cafetero y Centro-oriental. Cada subregión en cuestión puede estar ubicada integral o parcialmente en un conjunto de municipios aledaños, y fueron seleccionadas de acuerdo con los objetivos propios de la Encuesta.

#### 3.1. Selección de las Unidades Primarias de Muestreo

Al interior de cada subregión se seleccionó un conjunto de municipios de acuerdo con el tamaño de su población y las características de su economía en zonas rurales dispersas (no incluye centros poblados). Esta selección no fue probabilística, sino orientada a contar con un conjunto de municipios que dentro de cada subregión permitiera recolectar información sobre las características propias de la economía campesina en cada una.

Cada subregión tiene una muestra de 1.200 hogares, distribuidos en 240 segmentos de un promedio de veinte (20) hogares. Al interior de cada subregión, la muestra de 1.200 hogares se distribuyó entre los municipios seleccionados, proporcional al tamaño de la población. :

**Tabla 24: Municipios, segmentos y hogares**

Subregión	Departamento	Municipio	Encuestas	Segmentos
Atlántica	Córdoba	Cereté	300	15
		Chinú	180	9
		Ciénaga de Oro	260	13
		Sahagún	320	16
	Sucre	Sampués	140	7
Cundiboyacense	Boyacá	Saboya	456	23
	Cundinamarca	Simijaca	192	10
		Susa	216	11
	Santander	Puente Nacional	336	17
Eje Cafetero	Quindío	Circasia	288	14
		Córdoba	96	5
		Filandia	240	12
	Risaralda	Belén de Umbría	576	29
Centro-oriental	Cundinamarca	Tocaima	168	8
	Tolima	Natagaima	192	10
		Ortega	576	29
		Purificación	264	13
<b>Total</b>		17	4.800	240

### 3.2. Selección inicial de las Unidades Secundarias de Muestreo

Para cada municipio se realizó una selección de las Unidades Secundarias de Muestreo (veredas) proporcional al tamaño a partir del número de hogares en cada una. Con este fin:

- Se calculó, a partir de la información del Sisbén, el número de hogares por vereda.
- Se construyó una frecuencia acumulada de hogares para todas las veredas del municipio y se obtuvo el intervalo de muestreo (IM) dividiendo el número total de hogares en las veredas por el número de segmentos que se querían seleccionar en el municipio.
- Aleatoriamente se obtuvo un número entre 1 y el valor del intervalo, el número de arranque, y con éste se identificó en la tabla de frecuencia acumulada por veredas la primera vereda de la muestra.
- Se obtuvo el resto de USM de la muestra, sumando sucesivamente el IM al número de arranque, y ubicando estos valores en la frecuencia acumulada calculada.

### 3.3. Selección final de Unidades Secundarias de Muestreo

Al realizar la selección de las veredas se presentaron dificultades relacionadas con la desactualización de la información del Sisben y la cartografía. Estas se resumen básicamente en:

- Las veredas disponibles en las bases del Sisben no coincidían con las que aparecían en los mapas o en las páginas web de los municipios.
- Aún cuando las veredas coincidían entre las bases del Sisben y los mapas, el número de hogares en cada una de ellas no coincidía con la información de la Alcaldía del municipio, que en la mayoría de los casos estaba más actualizada.
- No existía cartografía a nivel de vereda (solo corregimientos) para algunos municipios.

Para los municipios en los cuales se presentaron estos problemas se solicitó a la Secretaría de Planeación de cada uno de ellos el mapa con límites veredales y el listado de veredas y hogares en cada uno según el Sisben para poder realizar la selección con información actualizada.

En los casos en que la Alcaldía contaba con dicha información, se realizó la selección siguiendo la metodología descrita anteriormente. Debido a que en muchos casos los mapas enviados por los municipios diferían de la cartografía que estaba disponible en Bogotá, se ajustaron los mapas y se seleccionaron los puntos de arranque.

Hubo algunos casos particulares donde no fue posible implementar la metodología definida:

Para los municipios de Cereté y Sampués, no fue posible contar con el listado de veredas y hogares ni con los mapas con límites veredales (estaban disponibles solo con corregimientos). En el caso de Cereté, la información del Sisben correspondía a corregimientos, por lo que el equipo en terreno debió verificar las veredas que pertenecían a cada uno y seleccionar en campo las veredas que se trabajarían.

Para Sampués se entregó al equipo en campo el listado de veredas del municipio para que, mediante verificación del número de hogares de cada vereda con los líderes locales, se seleccionara aquellas con el mayor número de hogares hasta completar el número de segmentos

necesarios. Tanto en Sampués como en Cereté los equipos debieron dibujar las veredas en los mapas para poderse ubicar y para poder contar con la ubicación de las veredas en rondas futuras.

Posteriormente, durante la ejecución del operativo se presentaron dificultades en las veredas seleccionadas por:

- Problemas de orden público
- Población indígena principalmente (en resguardos o en cabildos)
- Veredas muy lejanas y de difícil acceso
- Veredas correspondientes a grandes productores

Para resolver estas restricciones, se solicitó a las supervisoras de los grupos afectados verificar con las autoridades y líderes locales, las veredas en las cuales no se presentarían dichos inconvenientes. Sobre este listado se seleccionaron los reemplazos necesarios, eliminando de listados las veredas seleccionadas inicialmente y haciendo nuevamente un muestreo por PPT, en caso de contar con información sobre el número de hogares. En aquellos municipios en los cuales no se contaba con información precisa sobre el número de hogares, la selección se hizo con base en las recomendaciones de los líderes y las autoridades sobre las veredas que cumplían de mejor manera las condiciones de población objetivo y seguridad necesarias.

Por otra parte, uno de los objetivos en zonas rurales consistió en analizar las formas de tenencia de tierra, por lo cual era necesario que todos los hogares seleccionados tuvieran acceso a tierras en cualquier forma de tenencia (propia, arriendo, usufructo, etc.). Sin embargo, durante los dos primeros meses del operativo se incluyeron hogares con tierras, por lo que fue necesario ajustar la población objetivo, de acuerdo con la zona: en los municipios del Eje Cafetero se aceptaron hogares con y sin tierras, con el fin de estudiar la forma particular de tenencia en esta subregión, y se aceptaron además predios pequeños sin límite inferior de tamaño. En las demás subregiones se solicitó a las supervisoras no incluir hogares sin tierras y se les dio un mínimo de 900 m<sup>2</sup> como límite para la inclusión de predios.

Las restricciones en términos de población objetivo (no indígena), propiedad de la tierra (sólo hogares con acceso a tierra por cualquier forma de tenencia) y tamaño de los predios, dificultaron que se alcanzara el número de hogares por segmento (20 hogares). En estos casos, se autorizó a las supervisoras a tomar veredas vecinas para completar el tamaño de los mismos.

### **3.4. Selección de las Unidades Terciarias de Muestreo**

Con base en los mapas de cada municipio<sup>9</sup> se seleccionaron los puntos de arranque para cada uno de los segmentos, de acuerdo con el número de éstos que correspondieron a cada vereda en el muestreo por PPT. Estos puntos se escogieron de acuerdo con las vías, fuentes de agua y edificaciones cercanas.

Para cada punto de arranque el recorrido se realizó en zigzag y en dirección norte, haciendo un barrido del área hasta completar el tamaño del segmento (20 hogares). El recorrido se hizo de acuerdo con las instrucciones establecidas en el Manual de Supervisión y en el Manual de Recuento.

---

<sup>9</sup> Mapas con límites veredales cuando estaban disponibles, vías principales y secundarias, hidrografía, toponimia y centros poblados.

#### 4. METODOLOGÍA PARA EL CÁLCULO DE LOS FACTORES DE EXPANSIÓN

La probabilidad de selección de un hogar está dada por el producto de las probabilidades de selección de las unidades en las tres etapas del muestreo: unidades primarias (UPM), secundarias (USM) y terciarias (UTM):

$$P(H_{ijkl}) = P(UPM_{jk}) * P(USM_{ijk}) * P(UTM_{ijkl})$$

Donde

$P(H_{ijkl})$  es la probabilidad de selección del un hogar de la i-ésima UTM (MT), de la j-ésima USM (manzana) de la k-ésima UPM (municipio) del l-ésimo estrato (.).

$P(UPM_{kl})$  es la probabilidad de selección del k-ésimo municipio del l-ésimo estrato (probabilidad de selección de la UPM).

$P(USM_{jkl})$  es la probabilidad de selección de la j-ésima manzana, del k-ésimo municipio, del l-ésimo estrato (probabilidad de selección de la USM).

$P(UTM_{ijkl})$  es la probabilidad de selección de la i-ésima Medida de Tamaño (MT), de la j-ésima manzana, del k-ésimo municipio del l-ésimo estrato (probabilidad de selección de la UTM).

El factor básico de expansión para el hogar se calcula como el recíproco de su probabilidad de selección:

$$FBE(H_{ijkl}) = \frac{1}{P(H_{ijkl})}$$

El cálculo de cada una de probabilidades varía entre la zona urbana y rural debido a las diferencias metodológicas y operativas implementadas para la selección de las USM y las UTM en cada una.

#### 4.1. Factores de expansión de la muestra urbana

##### 4.1.1. Probabilidades de selección

La probabilidad de las UPM de la muestra urbana es igual a 1 para aquellas unidades de inclusión forzosa, y para las demás es igual a:

$$P(UPM_{kl}) = \frac{\text{Población de } k - \text{ésima UPM}}{\text{Población total del } l - \text{ésimo estrato}}$$

La población que aquí se contempla hace referencia a la población objetivo de la Encuesta, es decir, a la población en estratos socioeconómicos 1 a 4, excluyendo hogares unipersonales, temporales y hogares conformados solamente por personas mayores de 65 años. En el denominador, el estrato se refiere al l-ésimo conjunto de municipios construidos dentro de cada región a partir de su indicador de NBI y el nivel de urbanización, al que pertenece el municipio k.

En el caso de las USM, la probabilidad de selección es

$$P(USM_{jkl}) = \frac{\text{No. de MT seleccionadas del } k - \text{ésimo municipio del } l - \text{ésimo estrato}}{\text{No. total de MT del } k - \text{ésimo municipio del } l - \text{ésimo estrato}}$$

La probabilidad de selección de las UTM está dada por:

$$\left\{ \begin{array}{l} P(UTM_{ijkl}) = \frac{\text{No. de MT en } j - \text{ésima manzana del } k - \text{ésimo municipio del } l - \text{ésimo estrato según VIHOPE}}{\text{No. de MT en } j - \text{ésima manzana del } k - \text{ésimo municipio del } l - \text{ésimo estrato según recuento}} \\ \quad \text{si la relación de MT del VIHOPE al recuento era mayor que cero y menor que 3} \\ 1 \quad \text{si la relación de MT del VIHOPE al recuento era mayor o igual que 3} \end{array} \right.$$

Esta forma de cálculo de la tercera probabilidad se debe a las diferencias en el número de hogares entre el VIHOPE y el recuento, que surgen de los cambios en las viviendas y hogares de las manzanas a lo largo del tiempo (el VIHOPE es de 2005) y de la metodología de distribución de la muestra por manzanas empleada por el Censo de Población de 2005.

En aquellos casos en los cuales el número de hogares de la manzana era muy pequeño para conformar una sola MT y se unieron manzanas para obtener áreas geográficas de tamaño suficiente, se considera el número total de MT del área completa.

#### 4.1.2. Restitución del universo de hogares seleccionados no encuestados

Debido a que dentro de los hogares de cada MT seleccionada, algunos rehusaron responder la encuesta, se implementó un factor de ajuste de cobertura a nivel de segmento de la forma

$$FAC_{jkl} = \frac{\text{No. de hogares seleccionados en la } j - \text{ésima manzana}}{\text{no. de hogares encuestados en la } j - \text{ésima manzana}}$$

para cada segmento en la  $j$ -ésima manzana, del  $k$ -ésimo municipio del  $l$ -ésimo estrato.

En aquellos casos en que el factor de ajuste fue superior a 1.5 se unieron manzanas del mismo estrato socioeconómico<sup>10</sup> con el fin de recuperar los hogares seleccionados no encuestados a partir de un número más amplio de hogares encuestados. Por ejemplo, si en una manzana los hogares seleccionados fueron 10 pero sólo 5 respondieron ( $FAC = 2$ ), se unió la manzana con otra (con 9 encuestados de 10 seleccionados por ejemplo) del mismo estrato calculando el factor de ajuste como si se tratara de una sola manzana. En este caso el factor se reduciría ( $FAC = \frac{20}{9} = 1.4$ ), al contar con más encuestados respecto a los que no respondieron la encuesta.

Este factor multiplica al factor básico de expansión, siendo el mismo para cada hogar dentro de cada manzana o conglomerado geográfico (en caso de áreas geográficas ampliadas). Se obtiene el factor de expansión ajustado por cobertura.

#### 4.1.3. Correcciones por totales y por estructura de la población

<sup>10</sup> Para los estratos 1, 2 y 3, se unieron manzanas dentro del mismo municipio o dentro de la misma región. Para el estrato 4 se unieron todas las manzanas de la muestra.

Se calculó un factor de ajuste por totales y estructura de la población para reconstituir correctamente el universo, el cual se obtiene de la relación

$$FAte = \frac{\text{poblacion proyectada}}{\text{población expandida}}$$

para cada uno de los campos de la siguiente matriz:

Estrato Socioeconómico	Sexo									
	Hombre					Mujer				
	0-9 años	10-19	20-39	40-64	65 y más	0-9 años	10-19	20-39	40-64	65 y más
1										
2										
3										
4										

Esta matriz fue calculada por regiones tanto con la población proyectada como con la población expandida para obtener el factor de ajuste. La población proyectada se obtuvo teniendo en cuenta:

- Las proyecciones de población del DANE a 2010 por rangos de edad para cada municipio.
- La población por estrato socioeconómico, la cual se obtiene aplicando los porcentajes de usuarios del servicio de energía eléctrica (que es cercana al 100% en el país y en la mayoría de los municipios), a la población total proyectada en cada municipio.
- Los porcentajes de población por sexo y por cada uno de los grupos de edad según el DANE, que de nuevo se aplican a la proyección de población.

El factor de ajuste por totales y estructura multiplica al factor de expansión ajustado por cobertura para cada hogar dentro de cada grupo poblacional de la matriz, por región. En este punto, el factor de expansión para un hogar está dado por:

$$F(H_{ijkl}) = FBE(H_{ijkl}) * FAC_{jkl} * FAte$$

#### 4.1.4. Factor de corrección alrededor de la unidad

Finalmente, se calculó para cada hogar, el factor de expansión ajustado a la unidad. Este se hace para que los factores de expansión no expandan la muestra al tamaño del universo, sino para que, al usarlos, se mantenga el tamaño de la muestra obtenida pero se ajusten los totales y estructuras para que sean iguales a los que se observan en el universo. Este factor se obtuvo mediante la fórmula:

$$Fu(H_{ijkl}) = \frac{F(H_{ijkl})}{\frac{\sum_{i=1}^n F(H_{ijkl})}{n}}$$

donde el denominador es el promedio aritmético de los factores de expansión ajustados de los n hogares de la muestra.

## 4.2. Factores de expansión de la muestra rural

### 4.2.1. Probabilidades de selección



Debido a que la selección de los municipios de la muestra rural no fue probabilística, la probabilidad de selección de cada UPM rural es igual a 1.

La selección de las USM (veredas) se hizo por probabilidad proporcional al tamaño (PPT) de acuerdo con el número de hogares en cada una según el SISBEN, y durante el operativo fue necesario reemplazar algunas de las veredas seleccionadas por problemas de orden público, siguiendo el mismo procedimiento. Sin embargo, en algunos municipios se encontró un amplio número de población indígena organizada en cabildos o resguardos en las veredas escogidas, al punto que fue necesario buscar a la población objetivo de la Encuesta en todas las veredas del municipio y no solo en aquellas que habían sido seleccionadas. Por lo anterior, la probabilidad de selección para las USM en la muestra rural no se basa en las medidas de tamaño calculadas para la selección por PPT sino directamente en el número de hogares en cada vereda, de la siguiente forma:

$$P(USM_{jkl}) = \left( \frac{\text{No. de hogares de la } j - \text{ésima vereda}}{\text{No. de hogares en todas las veredas}} \right) * \text{No. de veredas seleccionadas}$$

en el k-ésimo municipio de la j-ésima subregión.

La probabilidad de selección de las UTM es

$$P(UTM_{jkl}) = \frac{\text{No. de Hogares encuestados en la } j - \text{ésima vereda}}{\text{No. de total de Hogares en la } j - \text{ésima vereda}}$$

en el k-ésimo municipio de la j-ésima subregión.

#### **4.2.2. Restitución del universo de hogares seleccionados no encuestados y correcciones por totales y estructura de la población.**

No se realizarán ajustes por hogares no encuestados debido a que el número máximo de rechazos por municipio es de tres, siendo cero en la mayoría de municipios. Por otra parte, debido a que la población rural dispersa proyectada (del DANE) incluye población indígena que no fue objeto del estudio, no es posible hacer una comparación adecuada entre las proyecciones de población y su estructura respecto a la estructura del universo expandido. En consecuencia, no se realiza ajuste por totales y estructura en la muestra rural.

## 5. PRECISIÓN DE LAS ESTIMACIONES

La precisión de las estimaciones puntuales se expresa en términos de error estándar, que es el indicador de variabilidad del azar intrínseca al muestreo probabilístico. El error estándar es función del tamaño de la muestra, del método de selección, de la naturaleza estadística de las estimaciones, de la frecuencia de las características cualitativas y de la variabilidad de las cuantitativas.

El error estándar relativo (ES rel) es el error estándar absoluto expresado como proporción (o porcentaje) del valor del indicador estimado. Para las estimaciones puntuales se desea, en general, un ES rel del 10% o menos. Para indicadores de relación por cociente, se acepta cierta variación en la magnitud del ES rel, como se muestra en la tabla 25.

**Tabla 25. Variaciones del ES rel para indicadores de cociente**

Valores del indicador de cociente	Valores del ES rel
$P < 10\%$	50% a 10%
$10 \leq P \leq 30\%$	10% a 7%
$P > 30\%$	7% a 5%

Para los indicadores de diferencia o cambio en el tiempo, la precisión se determina en función de la magnitud de tales diferencias o cambios que desean ser captados, con significancia estadística, para un determinado nivel de confianza (por ejemplo, 95%). Mientras más pequeños sean los cambios que se desee identificar, con significación estadística, más grande debe ser la muestra y más precisas sus estimaciones. Por otra parte, la velocidad de cambio en el tiempo de un indicador, varía con la naturaleza del indicador, y con el comportamiento de todos los factores que lo afectan, incluyendo por supuesto, las intervenciones a que a veces son sometidos.

Para un 95% de confianza, el error estándar de la diferencia o cambio ( $ES_{p_1-p_2}$  o  $ES_{\bar{y}_1 - \bar{y}_2}$ ) es igual a la diferencia mínima que se desea captar con significancia estadística, dividida por 1.96 (valor de z).

Los errores estándar de los resultados, o errores de muestreo, deben calcularse y usarse sistemáticamente, para evitar conclusiones equivocadas en la inferencia al universo de estudio. El método de cálculo se desprende directamente del método de selección de la muestra y de los procedimientos de estimación. El Anexo presenta el modelo basado en las Series de Taylor, para un Muestreo Estratificado de Conglomerados Desiguales.

### Anexo. Modelo para Cálculo de la Precisión Observada - Fórmulas basadas en Series de Taylor

#### A.3.1. Estimación de promedios y proporciones $\bar{y}$ , $p = r$

En el muestreo de conglomerados desiguales, los estimadores de cociente (promedios, proporciones) se los denomina estimadores de razón ( $r$ ), entre dos variables aleatorias: la variable de estudio “ $y$ ” y el tamaño de la muestra “ $x$ ” o sea  $r = y/x$ . En el caso de los promedios, “ $y$ ” es simplemente el acumulado del valor cuantitativo de la variable continua  $y$  en las proporciones, tasas, etc., es el número de unidades finales de observación que tienen la característica de interés.

##### A.3.1.1. Estimador

$$r = \frac{\sum_{h=1}^H \sum_{\alpha=1}^{a_h} \sum_{\delta=1}^{d_{h\alpha}} \sum_{\gamma=1}^{c_{h\alpha\delta}} y_{h\alpha\delta\gamma} F_{h\alpha\delta\gamma}}{\sum_{h=1}^H \sum_{\alpha=1}^{a_h} \sum_{\delta=1}^{d_{h\alpha}} \sum_{\gamma=1}^{c_{h\alpha\delta}} x_{h\alpha\delta\gamma} F_{h\alpha\delta\gamma}}$$

donde:

$h =$  estrato “ $h$ -ésimo”, que varía desde 1 hasta  $H$  (total de estratos)

$\alpha =$  1, 2, ...,  $a_h$ , es cada una de las UPM encuestadas en el estrato  $h$ .

$\delta =$  1, 2, ...,  $d_{h\alpha}$ , es cada uno de las USM encuestadas en la UPM  $\alpha$ , del estrato  $h$

$\gamma =$  1, 2, ...,  $c_{h\alpha\delta}$ , es cada una de las UTM encuestadas, en la USM  $\delta$

$a_h =$  El número de conglomerados seleccionados en el estrato “ $h$ ”. En nuestro caso, es el número de UPM seleccionadas en el estrato  $h$ .

$d_{h\alpha} =$  Número de USM encuestadas en la UPM  $\alpha$

$c_{h\alpha\delta} =$  Número de UTM encuestadas en la USM  $\delta$

$y_{h\alpha\delta\gamma} =$  El valor acumulado de la variable, en UTM  $\gamma$ , de la USM  $\delta$ , de la UPM  $\alpha$ , del estrato  $h$

$x_{h\alpha\delta\gamma} =$  El número de unidades de observación (hogares o personas) en la UTM  $\delta$ , de la USM  $\delta$ , de la UPM  $\alpha$ , del estrato  $h$

$F_{h\alpha\delta\gamma} =$  factor final de expansión en la UTM  $\gamma$ , de la USM  $\delta$ , de la UPM  $\alpha$ , del estrato  $h$

## A.3.1.2. Varianza de r

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

En donde, despreciando 1-f que es el factor de corrección para poblaciones finitas:

$$\text{var}(\hat{Y}_h) = \frac{a_h}{a_{h-1}} \left( \sum_{\alpha} \sum_{\delta} \sum_{\gamma} (y_{h\alpha\delta\gamma} F_{h\alpha\delta\gamma})^2 - \frac{\left( \sum_{\alpha} \sum_{\delta} \sum_{\gamma} (y_{h\alpha\delta\gamma} F_{h\alpha\delta\gamma}) \right)^2}{a_h} \right)$$

$$\text{var}(\hat{X}_h) = \frac{a_h}{a_{h-1}} \left( \sum_{\alpha} \sum_{\delta} \sum_{\gamma} (x_{h\alpha\delta\gamma} F_{h\alpha\delta\gamma})^2 - \frac{\left( \sum_{\alpha} \sum_{\delta} \sum_{\gamma} (x_{h\alpha\delta\gamma} F_{h\alpha\delta\gamma}) \right)^2}{a_h} \right)$$

$$\text{cov}(\hat{Y}_h, \hat{X}_h) = \frac{a_h}{a_{h-1}} \left( \sum_{\alpha} \sum_{\delta} \sum_{\gamma} y_{h\alpha\delta\gamma} x_{h\alpha\delta\gamma} F_{h\alpha\delta\gamma} - \frac{\left( \sum_{\alpha} y_{h\alpha\delta\gamma} F_{h\alpha\delta\gamma} \right) \left( \sum_{\alpha} x_{h\alpha\delta\gamma} F_{h\alpha\delta\gamma} \right)}{a_h} \right)$$

## A.3.1.2. Error Estándar de r

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

A.3.2. Estimadores de cambio en el tiempo, o de diferencia entre dos regiones o subgrupos  $r_1 - r_2$ 

## A.3.2.1. Varianza para Muestras Independientes

En un estrato:

$$\text{var}(r_{h1} - r_{h2}) = \text{var}(r_{h1}) + \text{var}(r_{h2})$$

En el conjunto de estratos:

$$\text{var}(r_1 - r_2) = \sum_h \text{var}(r_{h1}) + \sum_h \text{var}(r_{h2})$$

A.3.2.2. Varianza para muestras dependientes o muestras panel

En un estrato:

$$\text{var}(r_1 - r_2) = \text{var}(r_1) + \text{var}(r_2) - 2 \text{cov}(r_1, r_2)$$

En el conjunto de estratos:

$$\text{var}(r_1 - r_2) = \sum_h \text{var}(r_{h1}) + \sum_h \text{var}(r_{h2}) - 2 \sum_h \text{cov}(r_{h1}, r_{h2})$$

Siendo,

$$\begin{aligned} \text{cov}(\hat{Y}_{h1}, \hat{Y}_{h2}) &= \\ &= (1 - f_h) \frac{a_h}{a_{h-1}} \left( \sum_{\alpha} \sum_{\delta} \sum_{\gamma} y_{h\alpha\delta\gamma(1)} y_{h\alpha\delta\gamma(2)} F_{h\alpha\delta\gamma} - \frac{\left( \sum_{\alpha} \sum_{\delta} \sum_{\gamma} y_{h\alpha\delta\gamma(1)} F_{h\alpha\delta\gamma} \right) \left( \sum_{\alpha} y_{h\alpha\delta\gamma(2)} F_{h\alpha\delta\gamma} \right)}{a_h} \right) \end{aligned}$$

$$\text{cov}(r_{h1}, r_{h2}) = \frac{1}{\hat{X}_1 \hat{X}_2} \left[ \text{cov}(\hat{Y}_{h1}, \hat{Y}_{h2}) + r_{h1} r_{h2} \text{cov}(\hat{X}_{h1}, \hat{X}_{h2}) - r_{h1} \text{cov}(\hat{Y}_{h2}, \hat{X}_{h1}) - r_{h2} \text{cov}(\hat{Y}_{h1}, \hat{X}_{h2}) \right]$$

donde cada una de las covarianzas listadas es de la forma de la arriba desarrollada  $(\text{cov} \hat{Y}_{h1}, \hat{Y}_{h2})$

A.3.2.3. Error estándar de cambio o diferencia

$$es(r_1 - r_2) = \sqrt{\text{var}(r_1 - r_2)}$$