



BANCO CENTRAL DE RESERVA

# Aspectos Metodológicos de la Encuesta de Hogares de Propósitos Múltiples 2022

# I. Objetivos

## A. Objetivo general

Generar información estadística actualizada, tanto cualitativa como cuantitativa relacionada con las condiciones socioeconómicas y demográficas de la población salvadoreña, para facilitar el diseño o rediseño de políticas, planes, programas y proyectos que desarrollan las instituciones públicas que contribuyan a elevar el bienestar de la población y que a la vez sea de utilidad a otros organismos nacionales e internacionales para los mismos propósitos.

## B. Objetivos específicos

- a. Proveer información estadística de las principales variables que caracterizan la situación demográfica de la población salvadoreña.
- b. Generar estadísticas actualizadas sobre las principales variables que caracterizan la situación y tendencia del empleo en el país.
- c. Estimar el monto y fuente de los ingresos en los hogares.
- d. Presentar indicadores sobre características, situación de tenencia de la vivienda en el país y de los servicios básicos que dispone.
- e. Proporcionar datos que permitan estimar indicadores sobre el nivel educativo de la población.
- f. Suministrar información sobrecifras que faciliten el cálculo de indicadores sobre la cobertura de los servicios de salud.

# II. Cobertura geográfica

La EHPM se realiza en los 14 departamentos del país, con una cobertura a nivel nacional, por área urbana y rural.

# III. Cobertura temporal

La encuesta se realiza durante el período comprendido del mes de enero a diciembre de 2023.

# IV. Variables investigadas

La Encuesta de Hogares de Propósitos Múltiples (EHPM) se diseñó como un sistema continuo de encuestas, basadas en submuestras mensuales representativas del país, con aplicación de seis módulos permanentes.

## V. Módulos de la encuesta

Se investigaron durante todo el período de la encuesta, rubros básicos sobre población (características personales de los miembros de los hogares, edad, sexo, parentesco, etc.), educación, empleo e ingreso, salud, remesas familiares y gastos del hogar.

Se incluyeron los siguientes módulos en la encuesta:

Área temática
Número de miembros fuera del hogar
Características sociodemográficas
Características de educación
Características generales de la vivienda
Empleo e ingreso
Salud
Alimentación y esparcimiento

### Sección 0: Número de miembros fuera del hogar

En esta sección se recopila la información relacionada con los miembros del hogar que se encuentran en el extranjero, los motivos del abandono de niños, niñas y adolescentes de parte de los padres y con los pupilos que habitan en el hogar.

### Sección 1: Características sociodemográficas

A través de este módulo se obtiene información estadística sobre la composición y estructura poblacional de los hogares, de tal forma que permita identificar su comportamiento demográfico.

Este módulo se aplica a la jefatura del hogar, que es la persona que los miembros del mismo reconocen como tal y en ausencia de este a un informante adecuado<sup>1</sup>. Constituye un registro básico que contiene información sobre cada uno de los miembros del hogar, tales como: parentesco con el jefe del hogar, sexo, edad y estado familiar.

### Sección 2: Características de educación

---

<sup>1</sup> Se considera un informante adecuado aquella persona en caso del sexo femenino mayor de 14 años y en el masculino mayor de 15 años, que muestre señales visibles que conozca y pueda dar la información solicitada.

Las variables captadas en este módulo permiten caracterizar las principales condiciones de educación de la población encuestada, en relación con la cobertura del sistema educativo.

Se aplica a todos los miembros del hogar, captándose información referida a la cobertura educativa, asistencia escolar, nivel y grado que estudia, tipo de centro escolar, razones de repitencia o abandono, gastos en educación, último nivel y grado aprobado y título obtenido, con dicha información el módulo suministra datos que permiten elaborar los principales indicadores educativos como: tasa de analfabetismo de la población, escolaridad promedio, tasa bruta de escolaridad.

### **Sección 3: Características generales de la vivienda**

A partir de este módulo se generan indicadores sobre las características físicas, tenencia y servicios básicos de la infraestructura utilizada como vivienda por los hogares encuestados. Se aplica de preferencia a jefes de hogar.

Los temas a investigar incluyen: tipo y tenencia de la vivienda, materiales, número de habitaciones, servicios básicos y gastos en vivienda (agua, alumbrado, tipo de servicio sanitario), equipamiento, datos financieros (cuota mensual, adquisición) y para completar las características de la vivienda, dentro de las condiciones ambientales, se incluye el tratamiento que se le da a los desechos sólidos.

### **Sección 4: Empleo e ingreso**

La información que se obtiene en este módulo permite producir estadísticas actualizadas sobre las principales variables que caracterizan la situación y tendencia del empleo en el país. Con el propósito de identificar aspectos que permitan visualizar la situación ocupacional en el período de referencia, el cual para el caso del empleo, es la semana anterior a la fecha de la encuesta.

Este módulo se aplica a cada una de las personas miembros del hogar de 5 años y más de edad. Este módulo es uno de los más importantes de la EHPM porque a partir de éste se obtiene información sobre los niveles de empleo e ingreso de la población, los cuales influyen directamente en los niveles de vida de la misma.

El módulo consta de dos partes. En la primera se capta la información sobre la condición de actividad de la población (PEA, PEI, ocupados y desocupados, etc.) En la segunda se investiga el ingreso que percibe en la actividad principal en concepto de salario, cuenta propia o patrono; así

mismo, el ingreso por empleo secundario y otros ingresos no laborales. Además, se obtiene información relacionada a las remesas recibidas, tanto en efectivo como en especie.

### **Sección 5: Salud**

A través de este módulo se identifican las condiciones de salud de la población, así como el uso de los servicios de salud a los cuales recurre la población.

Se investigan datos de asistencia a los centros de salud (a quien consultó, lugar de consulta) y las dificultades que enfrenta al buscar la atención médica.

### **Sección 6: Alimentación y esparcimiento**

En este modulo se contempla como indicador de la seguridad alimentaria todas aquellas situaciones en la que los miembros del hogar mostraron preocupación por el abastencimiento alimenticio en tiempo y cantidad adecuada, ademas se valora la priorizacion de los menores de 18 años en caso de carencia de estos. Las valoraciones obtenidas por el informante estarán basadas 3 meses previos a la entrevista.

En el caso de las preguntas de esparcimiento las variables contempladas son el acceso y uso de los espacios de recreación que posee la zona de residencia, asi como la identificación de los motivos de no uso de los mismos.

## **VI. Diseño muestral**

### **A. Cobertura de la muestra**

La población representada en la muestra correspondió a las viviendas particulares, a los hogares y personas que allí habitaban, con una cobertura nacional.

### **B. Niveles de estimación y de desagregación de resultados**

A través del nuevo diseño muestral se obtuvieron precisiones confiables para el mayor nivel de desagregación.

Nivel de desagregación	Muestra requerida
Nivel nacional - total	Trimestral
Nivel nacional - urbano	Semestral
Nivel nacional - rural	Semestral

### **C. Marco muestral y su actualización**

El marco muestral se basa en el material cartográfico del último Censo de Población y Vivienda (2007); estableciendo un nuevo marco muestral final de 1,676,002 viviendas.

El Salvador tiene una extensión territorial de 21,040.79 kilómetros cuadrados y su división política administrativa consiste en 14 departamentos, los que se subdividen en un total de 262 municipios. Cada municipio, a su vez, se divide en un área urbana (cabecera municipal) y área rural compuesta por los lugares conocidos como cantones.

### **D. Conceptualización de área urbana**

La delimitación de un área urbana se definió como el área geográfica constituida por la agrupación de los segmentos censales resultante de la aplicación de los siguientes criterios:

- El segmento censal donde está ubicada la alcaldía municipal; o sea, el segmento origen del casco urbano (OCU).
- El conglomerado de segmentos que cumplen las características de núcleo poblacional (NP), núcleo habitacional (NH) o conglomerar viviendas en manzanas (CVM) y que colindan o encierran al segmento origen del casco urbano.
- Dos o más segmentos que cumplen las características de núcleo poblacional (NP), núcleo habitacional (NH) o por contener viviendas conglomeradas en manzanas (CVM), que sean colindantes entre sí, y que suman 500 viviendas o más agrupadas continuamente.
- Los segmentos clasificados NP, NH ó CVM que colinden con el área urbana de otro municipio.
- Los segmentos que no cumplen ninguno de los criterios pero que queda circunscrito completamente dentro de una mancha urbana conformada.
- El área rural de un municipio será la conformada por el conjunto de segmentos restantes a los clasificados como área urbana.
- Aplicados los criterios anteriores, el municipio cuya población rural sea igual o menor al 5% se consideró totalmente urbano.

Definiciones utilizadas en conceptualización anterior:

## **1. Segmento Censal**

Es la unidad básica estadística establecida como área de empadronamiento, que comprende un área geográfica conformada por una o más manzanas u otras áreas dispersas, y que cumple con el criterio de estar conformadas por una cantidad específica de viviendas.

## **2. Núcleo Poblacional (NP)**

Segmento censal con densidad poblacional mayor o igual de 1,000 habitantes/km<sup>2</sup>.

## **3. Centro Poblado Urbano**

Asentamiento poblacional con un mínimo de 500 viviendas agrupadas formando manzanas contiguas. Puede estar conformado por dos o más núcleos poblacionales.

## **4. Casco Urbano (CU)**

Centro poblado urbano donde están alojadas las autoridades edilicias.

## **5. Origen del Casco Urbano (OCU)**

Segmento censal donde está ubicado el edificio de la Alcaldía Municipal.

## **6. Núcleo Habitacional (NH)**

Segmento censal con una superficie territorial menor o igual a 0.6 km<sup>2</sup>.

## **7. Conglomerado de Viviendas en Manzanas (CVM)**

Segmento que en donde el 60% o más de las viviendas están agrupadas conformando manzanas con una superficie menor o igual a 20,000 m<sup>2</sup>.

Para una mejor cobertura geográfica el país se divide en 5 regiones, estando constituidas por los departamentos que a continuación se detallan:

### Región I (Occidental)

Santa Ana, Ahuachapán y Sonsonate.

### Región II (Central I)

La Libertad y San Salvador, excluyendo los municipios que conforman el Área Metropolitana de San Salvador (AMSS); Chalatenango y Cuscatlán.

### Región III (Central II)

San Vicente, La Paz y Cabañas.

### Región IV (Oriental)

Usulután, San Miguel, Morazán y La Unión.

### Región V (Área metropolitana de San Salvador)

Constituida por los municipios de: San Salvador, Mejicanos, San Marcos, Ayutuxtepeque, Cuscatancingo, Delgado, Ilopango, Soyapango, Antiguo Cuscatlán, Nueva San Salvador, San Martín, Nejapa, Apopa y Tonacatepeque.

## **E. Muestra maestra**

El enfoque más rentable y eficiente para proporcionar muestras para las encuestas de hogares a través de un período largo de tiempo es por medio de una muestra maestra. Si continuamente se actualiza la cartografía y los datos de las áreas en muestra en la muestra maestra; la muestra para cada encuesta de hogares permanece representativa a través del tiempo.

El diseño de una muestra maestra se lleva a cabo normalmente después de haberse levantado un censo de población y vivienda. Éste involucra la selección de una muestra de Unidades Primarias de Muestreo (UPM) seguido una muestra de Unidades Secundarias de Muestreo (USM) dentro de cada UPM. La muestra tiene que ser de tal tamaño que pueda proporcionar unidades de vivienda muestrales para cada encuesta de hogares del período intercensal. Las UPM y las USM se definen más adelante. La muestra maestra utilizará un diseño multietápico estratificado. Una encuesta individual puede usar todas las UPM en la muestra maestra o sólo utilizar una submuestra. Los segmentos muestrales pueden ser parcialmente rotados cada año para reducir el sesgo de condicionamiento y las altas tasas de no entrevista que puedan deberse al hecho de mantener los mismos hogares en muestra por mucho tiempo.

## **F. Universo y unidad de análisis**

El universo para las encuestas de hogares incluye los hogares y la población que vive en la república de El Salvador. Sin embargo, la población que vive en instituciones tales como hospitales, bases militares, prisiones, conventos, etc., será excluida del marco ya que representa un porcentaje muy pequeño del total de la población resulta difícil y caro de enumerar. Si se



necesitara información para estas subpoblaciones, deberán prepararse diseños muestrales específicos para dichas subpoblaciones. Las unidades de análisis para las encuestas de hogares son los hogares y las personas que viven en los hogares. Mientras que los segmentos en el marco cubran el área completa del país (urbano y rural), el marco representará adecuadamente el universo especificado.

## G. Marco y unidades muestrales

El marco muestral para un programa intercensal de encuestas se basa generalmente en los datos censales más recientes y en la cartografía más actualizada. Se tiene disponible del censo de mayo de 2007, archivos de datos y mapas. Toda esta información se usa para construir la muestra maestra.

El cuadro 1 muestra la distribución de los hogares, usando datos del censo 2007, por departamento y por área de residencia urbano/rural, así como la distribución porcentual de los hogares. Se puede ver en el cuadro que el porcentaje de hogares urbanos varía considerablemente por departamento, desde un 1.22 por ciento en el departamento de Morazán hasta un 44.94 por ciento en el departamento de San Salvador.

**Cuadro 1: Composición del marco muestral de hogares de El Salvador: Número de hogares por departamento (urbano y rural) y distribución porcentual**

Departamento	Viviendas	%	Urbano	%	Rural	%
Ahuachapán	70,664	5.12	28,337	3.48	42,327	7.46
Santa Ana	128,362	9.29	77,235	9.49	51,127	9.01
Sonsonate	101,507	7.35	46,410	5.70	55,097	9.71
Chalatenango	44,120	3.19	13,407	1.65	30,713	5.41
La Libertad	158,931	11.51	98,438	12.09	60,493	10.66
San Salvador	402,143	29.11	365,796	44.94	36,347	6.41
Cuscatlán	51,166	3.70	14,561	1.79	36,605	6.45
La Paz	73,572	5.33	28,332	3.48	45,240	7.97
Cabañas	30,655	2.22	10,757	1.32	19,898	3.51
San Vicente	36,950	2.68	17,394	2.14	19,556	3.45
Usulután	82,939	6.00	33,656	4.14	49,283	8.69
San Miguel	105,159	7.61	54,667	6.72	50,492	8.90
Morazán	38,706	2.80	9,933	1.22	28,773	5.07
La Unión	56,430	4.09	14,983	1.84	41,447	7.30
<b>Total</b>	<b>1,381,304</b>	<b>100.00</b>	<b>813,906</b>	<b>100.00</b>	<b>567,398</b>	<b>100.00</b>

Fuente: Marco maestro del Censo de Población y Vivienda 2007.

El cuadro 2 presenta el cambio porcentual en el número de hogares a nivel nacional, urbano y rural entre el censo de septiembre de 1992 y el Censo de Población y Vivienda de mayo 2007.

**Cuadro 2: Cambio porcentual en el número de hogares entre el censo de 2007 y el de 1992**

	Hogares a nivel nacional	Hogares del área urbana	Hogares del área rural
Mayo de 2007	1,381,304	813,906	567,398
Septiembre de 1992	1,100,714	598,842	501,872
Cambio porcentual	25.49	35.91	13.06

Fuente: Censos de Población y Vivienda de 1992 y 2007.

Para el censo de 2007, El Salvador se dividió en áreas de empadronamiento (AE) con un promedio de viviendas entre 110 a 120 por AE. Sin embargo, los tamaños de las AE varían, como puede observarse en el cuadro 3. Las unidades muestrales definidas en el marco de la muestra maestra se basan en un diseño que involucra varias etapas. Las UPM dentro de cada estrato se definen normalmente en base a las subdivisiones administrativas del país y en base a consideraciones de conveniencia operacional y control. En un marco de muestra maestra es importante que las UPM sean lo suficientemente grandes para poder proporcionar hogares muestrales para cada encuesta durante el período intercensal. En encuestas de hogares anteriores, las AE fueron utilizadas para este propósito. Su tamaño es particularmente conveniente para encuestas únicas y pueden ser seleccionadas en una etapa.

**Cuadro 3: Composición del marco muestral de hogares de El Salvador. Número de AE y promedio de hogares por AE por departamento y área**

Departamento	Total		Urbano		Rural	
	# AE	# HH/AE	# AE	# HH/AE	# AE	# HH/AE
Ahuachapán	668	106	228	125	440	97
Santa Ana	1,124	114	594	130	530	97
Sonsonate	901	113	347	134	554	100
Chalatenango	514	86	115	117	399	77
La Libertad	1,419	112	801	123	618	98
San Salvador	3,148	128	2,812	130	336	108
Cuscatlán	430	119	103	142	327	112
La Paz	747	99	230	124	517	88
Cabañas	302	102	88	123	214	93
San Vicente	412	90	152	115	260	76
Usulután	778	107	270	125	508	97
San Miguel	958	110	468	117	490	103
Morazán	389	100	85	117	304	95
La Unión	633	90	140	107	493	84
Total	12,423	112	6,433	127	5,990	95

Fuente: Marco muestral del Censo de Población y Vivienda 2007.

Como existen varias AE que contienen pocas viviendas, se recomienda que las mismas se consoliden con AE aledañas en la misma área geográfica (incluyendo zona urbana/rural) antes de la creación de las UPM. Esta operación redujo el número de AE de 12,423 a 12,065.

Seguidamente, crear UPM que contengan, en promedio, tres AE o segmentos contiguos. En algunos casos, podrá ser necesario que ciertas UPM contengan cuatro AE contiguas y en otros casos dos áreas contiguas, dependiendo de las condiciones en el terreno y del tamaño de las AE.

Las AE que tienen menos de 51 hogares deberán consolidarse con AE adyacentes antes de crear las UPM para que, de esta manera, no haya ninguna AE que tenga menos de 51 hogares. Todas las UPM deben encontrarse completamente dentro de la misma Municipalidad. Ninguna UPM puede contener AE de Municipalidades diferentes o de un área de residencia diferente (urbano/rural). El país estará dividido en aproximadamente 4,020 UPM de tamaño promedio igual a 3 AE. Esta metodología proporcionará UPM de tamaño suficiente para garantizar suficiente muestra para cualquier encuesta en el periodo intercensal. Los tamaños de las AE son también apropiados porque muchas encuestas requieren un número grande de hogares por AE en el estrato rural. Otra consideración es que las encuestas demográficas y de migraciones normalmente requieren muestras más grandes dentro de cada USM (el AE o un segmento de la AE), ya que las características demográficas tienen una correlación intraclase más baja que las variables económicas. Por lo tanto, se recomienda definir USM con un tamaño mínimo de 51 viviendas.

Se eligió una muestra de 1,664 UPM. Este tamaño de muestra satisfará los requisitos de muestra de la mayoría de las encuestas que se levanten durante el periodo intercensal. La institución puede usar todas las 1,664 UPM para las encuestas de gran envergadura o puede usar una submuestra de las UPM para encuestas más pequeñas. La submuestra de UPM para las encuestas pequeñas deberá elegirse en el mismo orden en que se eligió la muestra original de las 1,664 UPM. La distribución de las UPM muestrales por área urbano/rural se presenta en el cuadro 4.

**Cuadro 4: Composición de los hogares en el marco muestral: Estimación del número de UPM y USM por departamento y área**

Departamento	# AE en el Marco			# UPM en la Muestra		
	Urbano	Rural	Total	Urbano	Rural	Total
Ahuachapán	228	440	668	43	50	93
Santa Ana	594	530	1,124	68	85	153
Sonsonate	347	554	901	65	52	117
Chalatenango	115	399	514	32	34	66
La Libertad	801	618	1,419	132	54	186
San Salvador	2,812	336	3,148	282	51	333

Departamento	# AE en el Marco			# UPM en la Muestra		
	Urbano	Rural	Total	Urbano	Rural	Total
Cuscatlán	103	327	430	41	31	72
La Paz	230	517	747	51	47	98
Cabañas	88	214	302	39	67	106
San Vicente	152	260	412	32	31	63
Usulután	270	508	778	38	35	73
San Miguel	468	490	958	43	53	96
Morazán	85	304	389	38	62	100
La Unión	140	493	633	44	64	108
<b>Total</b>	<b>6,433</b>	<b>5,990</b>	<b>12,423</b>	<b>948</b>	<b>716</b>	<b>1,664</b>

Fuente: Marco muestral del Censo de Población y Vivienda 2007.

Una vez que las UPM se seleccionan en la muestra, se debe seleccionar las USM. Para hacer esto, de cada UPM en muestra, se seleccionará una USM con probabilidad proporcional al tamaño (PPT). En algunos casos, puede ser necesario tener que segmentar la USM si la misma contiene un gran número de hogares. Si una USM contiene 150 hogares o menos, no se debe segmentar. Sin embargo, si la USM contiene más de 150 hogares, debe segmentarse en partes que no excedan las 100-150 viviendas por parte. Una vez que la USM se ha dividido en segmentos, se debe seleccionar uno de los segmentos con PPT. Se hace esto para tener segmentos de tamaño 150 viviendas o menos antes de llevar a cabo la operación de listado en los segmentos seleccionados. Si la USM es muy grande, la operación de listado puede encarecer los costos de la encuesta.

Se debe notar que la selección de la USM con PPT seguido de la selección de una AE con PPT dentro de cada USM es equivalente a seleccionar el AE con PPT en una sola etapa. Sin embargo, las AE adicionales en las USM muestrales estarán disponibles para otras encuestas de hogares en el período intercensal. Desde el punto de vista operacional, esta etapa adicional de selección es bien fácil de implementar.

Para encuestas de hogares individuales, es deseable tener una muestra de áreas pequeñas en las que se pueda llevar a cabo la operación de listado de hogares de manera oportuna y económica antes de pasar a la selección de los hogares que se incluirán en la muestra. Se ha encontrado en encuestas de hogares anteriores que muchas AE eran demasiado grandes para permitir un listado que sea económico. Por lo tanto, las AE grandes se dividieron en segmentos más pequeños con límites bien definidos. Es muy importante asegurarse de que cada segmento tenga límites bien definidos en el terreno, sin importar el tamaño del segmento. En caso de no poder tener un límite bien definido e identificable en el terreno, es mejor usar segmentos más grandes que arriesgar tener problemas de cobertura con segmentos que no están bien delineados en el campo.

Se llevará a cabo un listado de hogares dentro de cada segmento muestral antes de cada encuesta, o al menos se hará el listado una vez al año cuando se tenga que compartir el mismo listado en varias encuestas. En muchos países, las viviendas se seleccionan en la última etapa ya que son estructuras permanentes con mucha menor variabilidad en el tiempo que los hogares. Sin embargo, en El Salvador, una estructura o una unidad de vivienda puede tener más de un hogar (esto es más pronunciado en la parte rural del país) y esta situación puede presentar problemas para controlar la carga de trabajo de los gestores de información.

Por lo tanto, sería operacionalmente más conveniente seleccionar hogares directamente a partir del listado. Esta es una opción viable si el listado se lleva a cabo en un período no mayor de tres meses anteriores al levantamiento de la encuesta. Si no, se tendrá errores de cobertura debido a cambios en los hogares a través del tiempo.

Cuando se seleccionan UPM, con PPT en las etapas respectivas, y luego se selecciona un número constante de hogares por segmento en la última etapa, la muestra será auto ponderada dentro del estrato. La variabilidad en las ponderaciones depende en la exactitud de las medidas de tamaño (número de hogares) en el marco.

## **H. Estratificación**

Uno de los rasgos más importantes de un diseño muestral eficiente es la estratificación del marco muestran en áreas homogéneas. La selección de la muestra se lleva a cabo independientemente dentro de cada estrato, aunque se desea ordenar las UPM bajo ciertos criterios dentro de cada estrato para proporcionar, de esa manera, una estratificación adicional implícita cuando se usa selección sistemática. La naturaleza de la estratificación depende de las características más importantes que se quieran medir en la encuesta, así como de los dominios de estimación o de análisis.

La estratificación corresponde a los dominios geográficos más importantes que se definen en las encuestas de hogares: Departamento y 50 Municipalidades dentro de los Departamentos. Las unidades muestrales se estratificarán aún más por área urbano/rural para propósitos de muestreo únicamente. No habrá una estratificación explícita a nivel urbano/rural.

## **I. Tamaño de muestra y asignación**

El tamaño de muestra para una determinada encuesta depende de la exactitud requerida para las estimaciones en cada dominio, así como de los recursos y de las limitaciones operacionales. La

exactitud de los resultados de una encuesta depende tanto del error muestral como del error no muestral. El error muestral puede medirse por medio de la varianza mientras que el no muestral sólo puede parcialmente medirse a través de re-entrevistas muy caras o de estudios de validación. El error muestral es inversamente proporcional al tamaño de muestra. Por otro lado, los errores no muestrales pueden aumentar con un incremento en el tamaño de muestra ya que es más difícil controlar la calidad de una operación de gran envergadura. Por lo tanto, es importante que el tamaño de muestra global sea manejable desde el punto de vista de control de calidad y control operacional.

El diseño de la muestra maestra será flexible en términos de satisfacer requisitos de tamaños de muestra diferentes para las distintas encuestas durante el período intercensal. El número de USM será suficiente para las encuestas de hogares que requieren el tamaño de muestra más grande. Las encuestas más pequeñas usarán una submuestra de UPM de la muestra maestra. El número de hogares que se elegirá por segmento muestral también variará de encuesta a encuesta, dependiendo de los objetivos de la encuesta y de la correlación intraclase de las principales variables que se midan. Por ejemplo, las características socioeconómicas tienen generalmente una correlación intraclase más alta que las características demográficas. Como la correlación intraclase aumenta el efecto del diseño (la varianza obtenida con el diseño actual dividido por la varianza correspondiente proveniente de una muestra aleatoria simple del mismo tamaño), es generalmente más eficiente usar tamaños de segmentos más pequeños para encuestas socioeconómicas (aunque también es necesario tener en cuenta las limitaciones de recursos y logísticas).

En base a los objetivos de las diversas encuestas de hogares que se tienen planificadas y dada la mayor variabilidad y menor costo de recolección de datos en las áreas urbanas, se recomienda elegir una muestra de 1,664 UPM para la muestra maestra, que proporcionará 948 UPM urbanas y 716 UPM rurales en base a una asignación más o menos proporcional. Esta asignación asegura estimaciones eficientes a nivel nacional y un tamaño de muestra adecuado en cada departamento del país.

La eficiencia muestral para una encuesta particular depende en parte del número de hogares seleccionados dentro de cada área muestral (como lo es el segmento o el AE) en la última etapa.

Para encuestas socioeconómicas en otros países en vías de desarrollo se ha determinado que el límite superior del número óptimo de hogares por segmento es de 10 para áreas urbanas y 15 para las rurales. Aunque sea necesario usar un número más grande de hogares por segmento para

algunas de las encuestas debido a limitaciones logísticas o de recursos, es deseable más adelante en el período intercensal tratar de reducir el tamaño del segmento para encuestas socioeconómicas.

## **J. Procedimientos de selección de muestra**

En base a la metodología descrita anteriormente para el diseño de la muestra maestra, se describe a continuación los pasos que se han seguido para desarrollar el marco de la muestra maestra y la subsiguiente selección de unidades muestrales en cada etapa de selección.

- a. Definición de los estratos (50 municipios autorrepresentados y los restos departamentales divididos por área).
- b. Selección de las UPM utilizando selección aleatoria sistemática con probabilidad proporcional al tamaño.
- c. Selección de tres USM por UPM utilizando selección aleatoria sistemática.

Si la UPM contiene más de 150 estructuras, deberá segmentarse en partes que no excedan las 100 – 150 estructuras por partición. Una vez que se haya dividido la UPM, se procederá a seleccionar una de las particiones de forma aleatoria. Esto se hace para tener segmentos de unas 150 viviendas o menos antes de llevar a cabo el listado en los segmentos muestrales. Es importante tener en cuenta que se deben obtener divisiones de segmentos que tengan límites bien definidos e identificables en el terreno.

Antes de cada encuesta se debe hacer el listado de estructuras correspondientes. Si esto no fuera posible, hay que asegurarse de que el listado de estructuras en los segmentos muestrales se haga al menos una vez al año.

Para la última etapa de selección, se actualizarán los bloques seleccionados en el mapa del segmento. Una vez actualizados, se procederá a la selección de las estructuras reportadas como ocupadas por hogares para su posterior levantamiento. Esta cantidad no debe ser menor de 4 ni mayor de 6 estructuras por bloque.

## **K. Cálculo de las probabilidades de selección de los hogares muestrales**

La probabilidad de selección de un hogar que se selecciona en dos etapas es:

$$(1) \quad p_{hijk} = \frac{n_h M_{hi}}{M_h} p_{2hij} \frac{m_{hij}}{M'_{hij}}$$

Donde:

- $p_{hijk}$  = Probabilidad de selección del k-ésimo hogar en el j-ésimo segmento dentro de la i-ésima UPM del estrato h.
- $n_h$  = Número de UPM seleccionadas en el estrato h.
- $M_h$  = Número de hogares en el marco dentro del estrato h.
- $M_{hi}$  = Número de hogares en el marco dentro de la i-ésima UPM del estrato h.
- $p_{2hij}$  = Probabilidad de selección del j-ésimo segmento dentro de la i-ésima UPM del estrato h.
- $m_{hijk}$  = Número de hogares seleccionados en el segmento j-ésimo de la i-ésima UPM del estrato h.
- $M'_{hij}$  = Número de hogares listados en el j-ésimo segmento dentro de la i-ésima UPM del estrato h.

El componente  $p_{2hij}$  en la fórmula se calcula:

$$(2) \quad p_{2hij} = \frac{M_{hij}}{M_{hi}}$$

Donde:

- $M_{hij}$  = Número de hogares en el j-ésimo segmento dentro de la i-ésima UPM del estrato h.
- $M_{hi}$  = Número de hogares dentro de la i-ésima UPM del estrato h.

Si los segmentos dentro de cada UPM contienen más o menos el mismo número de hogares, se puede entonces seleccionar el segmento con probabilidad igual y, en ese caso, la fórmula para  $p_{2hij}$  viene dada por:

$$(3) \quad p_{2hij} = \frac{1}{s_{hi}}$$

Donde:

- $s_{hi}$  = Número total de segmentos en la i-ésima UPM del estrato h.

El factor de ponderación básico o inicial es la inversa de la probabilidad inicial de selección del hogar y viene dado por:

$$(4) \quad W_{hijk} = \frac{M_h M_{hi} M'_{hij}}{n_h M_{hi} M_{hij} m_{hijk}}$$

Donde:

- $W_{hijk}$  = Factor de ponderación básico para los hogares en el j-ésimo segmento muestral dentro de la i-ésima UPM del estrato h.



Si  $m_{hijk}$  es constante dentro de cada estrato (por ejemplo, 10 hogares muestrales) y  $p_{2hij}$  se basa en una selección con probabilidad proporcional al tamaño, la muestra será aproximadamente auto-ponderada dentro de cada estrato. Estos factores de expansión variarán un poco debido a la diferencia entre el número de hogares en el marco y el número de hogares listados dentro del segmento.

Es importante también ajustar los factores de expansión para tomar en cuenta la tasa de no entrevista. Como los factores de expansión serán calculados a nivel de segmento muestral, es conveniente ajustar los factores a este nivel. El factor de expansión final  $W'_{hijk}$  para los hogares muestrales dentro del j-ésimo segmento de la i-ésima UPM en el estrato h se puede expresar de la siguiente manera:

$$(5) W'_{hijk} = W_{hijk} \frac{m_{hijk}}{m'_{hijk}}$$

Donde:

$m_{hijk}$  = Número total de estructuras seleccionadas en el j-ésimo segmento muestral de la i-ésima UPM del estrato h.

$m'_{hijk}$  = Número total de hogares entrevistados dentro del j-ésimo segmento de la i-ésima UPM en el estrato h.

Posteriormente, los factores de expansión de calibran de acuerdo con las proyecciones de población, tomando en cuenta los totales departamentales y la distribución nacional por sexo.

## L. Estimaciones

Las estimaciones más comunes que se calculan en las encuestas son totales y razones. La estimación de un total viene dada por:

$$(6) \hat{Y} = \sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{k=1}^{m'_{hijk}} W'_{hijk} y_{hijk}$$

Donde:

$L$  = Número de estratos.

$y_{hijk}$  = Valor de la variable y para el k-ésimo hogar muestral en el j-ésimo segmento dentro de la i-ésima UPM del estrato h.

La estimación de una razón viene dada por:

$$(7) \hat{R} = \frac{\hat{Y}}{\hat{X}}$$

Donde  $\hat{Y}$  y  $\hat{X}$  son estimaciones de totales para las variables  $Y$  y  $X$ , respectivamente, que se calculan como se muestra en la fórmula anterior.

Cuando se utilizan diseños con conglomerados, las medias y las proporciones son casos particulares de razones. En el caso de la media, la variable  $X$  en el denominador de la razón es igual a 1 para cada unidad, de tal manera que el denominador es la suma de los factores de expansión. En el caso de una proporción, la variable  $X$  en el denominador también es igual a 1 para todas las unidades, pero la variable  $Y$  en el numerador es binomial y es igual a 0 ó 1, dependiendo de la ausencia o presencia, respectivamente, del atributo especificado para la unidad observada.

## M. Procedimientos de estimación de varianzas

Además de presentar cuadros que muestren errores muestrales, es importante también incluir algunas fuentes de errores no muestrales para que el usuario de datos pueda tener una idea clara de la validez de los datos.

El error muestral se usa para determinar el nivel de confiabilidad de una estimación y contiene una parte del error no muestral. El estimador de la varianza debe tomar en cuenta los diferentes aspectos del diseño muestral, como la estratificación y el uso de conglomerados. Hay muchos programas que se pueden utilizar para calcular errores muestrales (o varianzas). Algunos de estos son gratuitos (CENVAR) y otros, como SPSS o STATA, contienen módulos que permiten el cálculo de varianzas para diseños complejos. Todos estos paquetes usan el estimador que en inglés se conoce como "ultimate cluster variance estimator" para calcular varianzas para estimadores lineales. La fórmula para estimar la varianza de un valor agregado o total viene dada por:

$$(8) V(\hat{Y}) = \sum_{h=1}^L \left[ \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left( \hat{Y}_{hi} - \frac{\hat{Y}_h}{n_h} \right)^2 \right]$$

Donde:

$$(9) \hat{Y}_{hi} = \sum_{k=1}^{m/hijk} W'_{hijk} Y_{hijk}$$

$$(10) \hat{Y}_h = \sum_{i=1}^{n_h} \hat{Y}_{hi}$$

La fórmula para estimar una razón viene dada por:

$$(11) V(\hat{R}) = \frac{1}{\hat{X}^2} [V(\hat{Y}) + \hat{R}^2 V(\hat{X}) - 2\hat{R} \text{Cov}(\hat{X}, \hat{Y})]$$

Donde  $V(\hat{X})$  y  $V(\hat{Y})$  se calculan como se muestra en las fórmulas anteriores.

## **N. Información que se debe mantener para la muestra maestra**

Es importante mantener un archivo que contenga un registro por cada UPM indicando las medidas de tamaño usadas en cada etapa de selección para poder calcular las probabilidades de selección y las ponderaciones correspondientes. Este mismo archivo puede usarse para introducir el número de viviendas y hogares listados en cada segmento muestral durante cada encuesta o período de encuesta, para facilitar la verificación de la actualización de las áreas muestrales e identificar problemas potenciales.

## **O. Consideraciones metodológicas en una encuesta de hogares continua**

En las encuestas de hogares que miden características afectadas por la estacionalidad, es ideal tener submuestras representativas distribuidas a través del tiempo para representar a nivel nacional la estacionalidad a través del año.

Algunos países están usando este tipo de encuesta continua para medir el desempleo a nivel mensual o trimestral. Las submuestras representativas se distribuyen a través del tiempo y del espacio. Con este tipo de metodología, se puede mantener un conjunto de gestores de información permanentes que entrevistan una submuestra de segmentos cada mes. Una ventaja de este enfoque es que es posible usar personal permanente y bien entrenado. Otra ventaja es que la recolección y el procesamiento de los datos se distribuyen equitativamente a través del año. Esto facilita también la implementación de controles de calidad y operacionales.

La asignación propuesta de la muestra en el cuadro 3 permite la posibilidad de definir submuestras urbanas y rurales, ya sean trimestrales o mensuales, que pueden distribuirse equitativamente a través del año usando todas las USM disponibles en la muestra maestra.

Una manera simple de distribuir la muestra de manera representativa a través del año es asignar un código de mes (del 1 al 12) a las USM de la muestra maestra dentro de cada estrato en el mismo orden en que las mismas fueron elegidas, comenzando con un número aleatorio entre 1 y 12. Todas las USM muestrales (y segmentos muestrales correspondientes) con código de mes 1 serán asignadas al mes de enero, con código de mes 2, al mes de febrero y así sucesivamente.

## **P. Consideraciones para la rotación de la muestra**

No es deseable mantener los mismos hogares en la muestra durante muchos años porque la tasa de no respuesta aumentará a través del tiempo, y las respuestas pueden estar sujetas a un sesgo de condicionamiento después de varias entrevistas. Por esta razón la muestra deberá rotarse periódicamente en encuestas continuas.

Otra consideración es que un traslapo en la muestra entre dos rondas consecutivas mejora la precisión de las estimaciones de las diferencias (como por ejemplo la diferencia entre dos mediciones anuales del desempleo). En el caso de una encuesta continua se recomienda que la mitad de los segmentos sea rotada cada año para poder tener un traslapo en la muestra de un 50 por ciento entre dos rondas anuales.

En los primeros dos años de la encuesta, se deben rotar hogares en los mismos segmentos muestrales. Al principio del tercer año es una buena idea empezar a rotar un cierto porcentaje de segmentos (25 por ciento). De esta manera se renueva el stock de segmentos y mejora la actualización del marco muestral.

## **Q. Procedimiento para el trabajo de campo**

Una vez seleccionados los bloques de una UPM, su identificación (mapa y listado) se entrega al gestor de brigada. Los gestores de información se dirigirán no sólo a las viviendas que figuran como ocupadas, sino que investigan en campo la situación de todas las construcciones del bloque, y realizan las encuestas en todas las viviendas que se encuentren ocupadas en el área delimitada por el bloque. Esto significa que, al momento de la Encuesta, algunas viviendas que figuraban como desocupadas hayan pasado a ser ocupadas, y en ese caso, el gestor de información realizará una actualización al respecto y podrá hacer más de cuatro entrevistas por bloque.

También podría ocurrir el caso contrario que una vivienda ocupada se hubiera desocupada entre el momento del listado y el momento de la Encuesta. Este procedimiento no es arbitrario, sino que responde a la necesidad de calcular los coeficientes de expansión corrigiendo las cifras del marco mediante la información que los propios gestores de información obtienen en campo en ocasión de la realización de las encuestas.

El trabajo de cada equipo de gestores de información y gestores de brigada se resume en las dos tareas siguientes:

- a. Entrevistar a todos los hogares que residen en las viviendas ocupadas en los bloques seleccionados en la muestra.

- b. Registrar el número de viviendas encontradas en cada lote, cuántas de ellas son viviendas ocupadas, y cuántas de éstas pudieron ser entrevistadas (algunas podrían no ser entrevistadas por rechazo, moradores ausentes u otras formas de no contacto).

## **R. Distribución de la muestra en el tiempo**

Considerando que el período de la encuesta es de enero a diciembre, los 1,664 segmentos seleccionados fueron distribuidos a lo largo del año en cada uno de los meses del año, conteniendo cada uno 139 segmentos aproximadamente a ser entrevistadas en campo. Los municipios y los cantones se distribuyeron a lo largo de los 12 meses, de forma que, en lo posible, un número similar de segmentos se visite para un departamento en particular a lo largo de todo el país urbano y rural, y así poder determinar la carga de trabajo a los gestores de información y gestores de brigada.

La asignación de un cantón a un mes en particular estuvo determinada por la cercanía a los municipios que se visitarían en la semana de trabajo.

## **S. Consideraciones sobre la no respuesta**

En toda investigación de naturaleza estadística se presenta el fenómeno que la información obtenida no corresponda exactamente a todas las unidades muestrales seleccionadas, conociendo este comportamiento como “no respuesta”.

Una definición precisa por tipo se presenta a continuación.

### **Tipo A**

Se presenta en aquellos casos cuando en las viviendas ocupadas no se obtiene respuesta, ya sea porque al visitar la vivienda en varias ocasiones nunca se encontró a nadie o los miembros del hogar están ausentes temporalmente o porque éstos se negaron a proporcionar la información solicitada.

### **Tipo B**

Este tipo de casos está referido cuando la vivienda seleccionada se encuentra desocupada, ya sea en forma regular o permanente, así mismo cuando la vivienda es utilizada temporalmente para negocio.

### Tipo C

Se refiere a los casos cuando al momento de la entrevista la vivienda seleccionada ya no existe, debido a que la han demolido o funciona como tienda, almacén, taller o cualquier otro tipo de negocio, etc. También se incluyen en esta tipología aquellos casos en los que no se tuvo acceso a la estructura por razones de seguridad o delincuencia.

Las no respuestas de los tipos B y C se deben a problemas en el marco muestral.

## VII. Consistencia y procesamiento de datos

### A. Entrada de datos

El software que permite capturar la información (CSPPro) permite crear los siguientes documentos:

- a. Diccionario, en el cual se definen todas las secciones, variables, rangos de variables, nombres y títulos de variables, entre otras características.
- b. Diseño del formulario. Permite diseñar el formulario en formato digital para facilitar al gestor de información la captura de datos de todas las secciones y variables.
- c. Filtros de consistencias. Este es un lenguaje de programación basado en el flujo de la información capturada y las diferentes inconsistencias que el equipo de metodología especifique. Prácticamente existe un control total del flujo de la información que se programa según los módulos de la encuesta.

### B. Inconsistencias en la entrada de datos

Las inconsistencias de entrada de datos se controlan por medio del diccionario y programas que le dan el seguimiento a los saltos y consistencias entre preguntas.

### C. Inconsistencias en la salida de datos

Terminado el proceso de captura, se genera un proceso de revisión de consistencia de datos, esto permite ejecutar los programas y corregir las inconsistencias. Este proceso se realiza a través de CSPPro y SPSS, que permiten ejecutar los procedimientos almacenados para la ejecución de estos chequeos.

#### **D. Generación de variables auxiliares que no están en la boleta**

Esta generación de variables facilita la interpretación de los indicadores. Este proceso es calculado partiendo de otras variables y respetando los algoritmos para el cálculo de estas.

Una vez concluida la digitación de los datos la integración de las nuevas variables y la depuración de la base de datos se realizan los siguientes pasos:

- a. Se procede a la actualización de los reportes para la tabulación de los datos.
- b. La exportación de la base de datos de SQL a SPSS y otros formatos.