Buenas cifras,
mejores vidas

Marco Maestro de Muestreo

Enero 2021

Marzo, 2021

INEC

Buenas cifras,
mejores vidas

01

Instituto Nacional de Estadística y Censos (INEC)

Dirección

Dirección de Infraestructura Estadística y Muestreo

Elaborado por:

William Constante
Javier Núñez
Jorge Velásquez

Revisado por:

Francisco Céspedes

Aprobado por:

David Sánchez

Quito -Ecuador, 2021

Contenido

ENEMDU: Cálculo de errores estándar y declaración de muestras complejas

Enero 2021

Introducción

Con el fin de mejorar la cobertura de las encuestas y el proceso de inferencia estadística, se diseña y construye un Marco de Muestreo completo, actualizado y exhaustivo. El Marco Maestro de Muestreo (MMM) constituye la infraestructura para la selección de todas las muestras de las encuestas probabilísticas en hogares que levanta el Instituto Nacional de Estadística y Censos (INEC).

Está conformado por un conjunto de áreas geográficas perfectamente delimitadas, listados de viviendas ubicadas en éstas, así como los mapas y planos que permiten localizarlas en campo. De igual manera, a partir de la realización del VII Censo de Población y VI de Vivienda (CPV) en el 2010, se recopila información de las principales características sociodemográficas de la población y de las viviendas existentes en todo el territorio nacional, y una vez que se dispuso de la información definitiva del mismo, se iniciaron los trabajos de elaboración del MMM. Cabe recalcar que su actualización es sumamente importante, pues de ella depende asignar la probabilidad de selección a todas las viviendas que son objeto de estudio.

El presente documento describe el proceso de elaboración del Marco Muestral partiendo desde la construcción del identificador de las unidades muestrales hasta los conceptos, procedimientos, definiciones y la metodología empleada para el proceso de conformación de las distintas unidades de muestreo, para que sean de uso de las operaciones estadísticas por muestreo dirigidas a hogares.

Marco Maestro de Muestreo (MMM)

Definición del MMM

El MMM se presenta como una lista organizada en forma de base de datos que contiene a las viviendas registradas en el CPV 2010, que participarán en cada una de las fases de diseño, distribución y selección de la muestra de una encuesta dirigida a hogares. Principalmente se emplea para identificar y seleccionar las unidades de muestreo, y como base para realizar estimaciones basadas en los datos de la muestra, esto implica que la población a ser

seleccionada para la muestra debe estar representada de forma física, es decir, el MMM también está formado por todos los mapas y planos a diferentes escalas que permiten identificar en forma precisa y clara los límites físicos que tienen las diferentes unidades de selección, considerándose como parte principal de éste los registros y listados en los que se detalla las referencias que faciliten identificar en forma exacta las viviendas seleccionadas.

El MMM contiene información sobre la división político-administrativa y geográfica del país -subdivisiones políticas o zonificación estadística definida para efectuar la enumeración del censo de población-, como también sobre los volúmenes de viviendas y de población total, por grupos de edad y sexo, entre otras variables necesarias para clasificar a los hogares de acuerdo a determinadas características según los objetivos específicos de la encuesta. Todos y cada uno de los elementos de los que está compuesto el MMM tienen una probabilidad conocida y diferente de cero de ser seleccionados de alguna de las muestras que se puedan extraer del mismo.

Objetivos del MMM

Los principales objetivos que tiene que cumplir el MMM para que este sea de uso permanente son los siguientes:

- Abarcar todas las unidades de investigación que tiene la población de una encuesta; y,
- Permitir seleccionar muestras para varias encuestas o diferentes rondas de una encuesta dirigida a hogares.
- Identificar y proporcionar acceso a cada una de las unidades de investigación de la población encuestada, donde además tengan probabilidad conocida de selección o se pueda determinar la misma para su selección en la muestra.
- Permitir la identificación de las áreas de mayor crecimiento urbano como también las áreas rurales o zonas geográficas de mayor emigración de su población.
- Determinar en forma precisa y objetiva la cobertura geográfica y poblacional que la encuesta cubrirá.
- Permitir que los resultados sucesivos de las encuestas dirigidas a hogares sean comparables con las del Censos de Población y Vivienda y con otros estudios estadísticos orientados a obtener información de las condiciones de los hogares particulares.

Propiedades del MMM

El MMM comprende, en términos estadísticos, la población objetivo, donde todos los miembros del universo de estudio tienen una probabilidad conocida, y distinta de cero, de formar parte de alguna muestra. Además de eso, un marco muestral perfecto es aquel que es completo, exacto y actualizado; el marco debe estar exento de duplicaciones y omisiones, y actualizarse de manera permanente, con el objeto de reflejar los cambios estructurales que se van produciendo en la geografía del país, la conformación de las unidades de selección, su distribución física, el surgimiento de nuevas edificaciones y las construcciones que se generan debido al crecimiento de la población, así como la bajas registradas debido a demoliciones, fenómenos naturales o por el

hecho de que algunas viviendas se hayan transformado en negocios comerciales.

En este sentido, las propiedades que presenta el MMM se detallan a continuación:

Exhaustividad: El MMM se considera exhaustivo respecto a la población objetivo, dado que todos los miembros que la constituyen (el universo) quedan cubiertos por el mismo.

Exactitud: El MMM es exacto ya que todos los miembros de la población objetivo se incluyen una sola vez.

Actualizado: El MMM permitirá reflejar la reciente construcción o demolición de viviendas, los movimientos migratorios en las unidades de vivienda, los nacimientos o las defunciones. Lo que permitirá cumplir con el requisito de que en una muestra probabilística cada miembro de la población objetivo tenga una probabilidad conocida de ser seleccionado.

Cobertura del MMM

La cobertura geográfica del MMM es Nacional, es decir cubre las 24 provincias del país en sus áreas urbanas y rurales, e incluye las zonas no delimitadas.

Población objetivo del MMM

Una consideración importante a la hora de estructurar un Marco de Muestreo adecuado y específico, es la relación entre la población objetivo de las encuestas y las unidades de selección.

Las unidades de selección deberán estar determinadas en el Marco de Muestreo, así como también la probabilidad de selección de cada una de las unidades en las diferentes etapas del muestreo. Las unidades del Marco de Muestreo se constituyen en un elemento básico para el cálculo, asignación y selección de la muestra de unidades poblacionales.

El MMM determina como población objetivo y unidades de última selección a las viviendas particulares[1] del país, hogares y personas residentes habituales de los mismos, cuyas definiciones se detallan a continuación:

Vivienda particular: es el recinto de alojamiento separado e independiente, destinado a alojar uno o más hogares particulares o aunque no esté destinado al alojamiento de personas es ocupado como vivienda en el momento de levantarse el Precenso. Estas viviendas pueden estar ocupadas, desocupadas, en construcción o habitadas temporalmente.

Hogar: es la unidad social conformada por una persona o grupo de personas que se asocian para compartir el alojamiento y la alimentación. Es decir, hogar es el conjunto de personas que residen habitualmente en la misma vivienda o

[1] Se excluye a la población en viviendas colectivas, viviendas flotantes y sectores con población indigente

en parte de ella (viven bajo el mismo techo), que están unidas o no por lazos de parentesco, y que cocinan en común para todos sus miembros (comen de la misma olla).

Residente habitual: es toda persona que come y duerme permanentemente en la misma vivienda en la que habita el hogar. Esta persona puede encontrarse temporalmente ausente del hogar en el momento de la encuesta por motivos de salud, estudio o vacaciones.

Unidades estadísticas muestrales de selección del MMM

El marco para el MMM se basó en los resultados definitivos y la cartografía del CPV - 2010. Las unidades de muestreo dependen de las etapas de muestreo. Para propósitos de obtener un listado actualizado eficaz en las áreas seleccionadas, es importante definir para la primera o segunda etapa de muestreo, segmentos con número de viviendas suficientes para permitir múltiples encuestas, con la finalidad de evitar en la medida de lo posible que los mismos informantes tengan que ser entrevistados repetidamente. Las unidades de selección que se utilizarán como unidades primarias de muestreo se hallan limitadas por el requisito de pertenecer únicamente a las unidades de área cartografiadas.

Una vez concluidos los trabajos de enumeración censal, surge la oportunidad de diseñar los planes y programas de las encuestas con las que se produce información para atender las demandas de los usuarios en los períodos inter-censales. En términos generales, la conformación del MMM se detalla a continuación:

La base informativa generada a partir del trabajo de campo del CPV 2010 posibilita la construcción de un marco de áreas, integrado por mapas de línea a diferentes escalas y una base de datos con la división geográfica, administrativa y estadística del país, así como información de base que permita conocer el total de viviendas particulares ocupadas en los distintos contextos geográficos, además el número de personas que habitan en cada uno de los hogares registrados en las mismas.

Posteriormente, es necesario realizar la codificación de las claves o identificadores de cada una de las áreas geográficas y unidades estadísticas que permitan iniciar con la construcción y definición del MMM.

CODIFICACIÓN DE VARIABLES PARA DEFINIR LA CLAVE ÚNICA DE IDENTIFICACIÓN PARA ÁREAS GEOGRÁFICAS Y UNIDADES ESTADÍSTICAS

1. En el Censo de Población y Vivienda y en las diversas encuestas de hogares que se realizan en el INEC, se utilizan diferentes identificadores para sus respectivos estudios de investigación, es conveniente tener conocimiento de las bases sobre las cuales se construyen estos identificadores, de esta manera se establece un identificador general que permita estandarizar la codificación de las variables de estudio.

2. Establecer un estándar en la codificación de las variables de identificación de las unidades estadísticas en los estudios que realiza el INEC, permite dar un tratamiento eficiente y eficaz a la información, además de facilitar la sistematización de información y la automatización de los procesos que realiza la institución.

Los principales conceptos estadísticos utilizados dentro de la generación de la clave de identificación muestral son los siguientes:

Unidades Estadísticas: Son requeridas por el analista de información, por una parte para saber cuál es la estrategia a seguir para la medición y, por otra, pensar en la estructura del marco de referencia de las unidades a ser estudiadas. Estas unidades son: la de investigación, análisis, observación y de muestreo.

Unidad de Investigación: Esta se refiere a la que contiene las partes que se van a analizar. Por ejemplo, en la encuesta de hogares para el estudio de la fuerza de trabajo que realiza el INEC, el motivo de la investigación es el hogar, el cual contiene las unidades a examinar, es decir, las personas. Por otro lado, en el sector industrial, la unidad de investigación está dada por el establecimiento.

Unidad de Análisis: Comprende la unidad que se analiza, es decir, de la que se busca la información y su naturaleza depende de los objetivos del estudio. Esta unidad puede ser el hogar, las personas, el establecimiento, etc. Las unidades de análisis reciben frecuentemente el nombre de "Elementos de la Población".

Unidad de Observación: Se denomina con este nombre a la unidad a través de la cual se obtiene la información, pudiendo o no coincidir con el elemento. Por ejemplo, cada uno de los miembros del hogar puede constituir un elemento de la población y sin embargo ser sólo uno de ellos, por ejemplo, el jefe del hogar, quien proporcione la información requerida. Este último, por tanto, constituirá la unidad de observación, también llamada unidad respondiente.

Unidad de Muestreo: Será un individuo o conjunto de individuos que se seleccionan en una única extracción. Como requisito se exige que el elemento o el grupo de elementos que componen el estudio reúnan las características de la población.

Identificador Unidad Estadística: es una serie numérica que agrupa diferentes variables para determinar una clave única la identificación que permite la estandarización de las variables y el número de caracteres que identifican una unidad estadística para las diferentes investigaciones que realiza el INEC, de esta manera se facilita el manejo de información de manera más eficiente.

A continuación, se indica la definición y codificación de cada una de las variables que componen el número de identificación de la Unidad Estadística, para posteriormente definir la estructura del mismo.

Áreas geográficas – División Político Administrativa

El Modelo Territorial Ecuatoriano para definición de Jurisdicciones Político Administrativas, se basa en la Ley de división Territorial del 26 de Marzo de 1897 según la cual la República del Ecuador se divide en Regiones, Provincias, Cantones y Parroquias, mediante decretos emitidos por el Congreso Nacional, se han ido creando a partir de entonces nuevas jurisdicciones.

Según la Constitución de la República Art 224, para la administración del Estado y la representación política el país será dividido en provincias, cantones y parroquias.

Art. 242.- El Estado se organiza territorialmente en regiones, provincias, cantones y parroquias rurales. Por razones de conservación ambiental, étnico-culturales o de población podrán constituirse regímenes especiales. Los distritos metropolitanos autónomos, la provincia de Galápagos y las circunscripciones territoriales indígenas y pluriculturales serán regímenes especiales.

Provincia

Son circunscripciones territoriales integradas por los cantones que legalmente les correspondan. Se encuentran agrupadas dentro de las regiones naturales.

El Ecuador está formado por 24 provincias que son las siguientes: Azuay, Bolívar, Cañar, Carchi, Cotopaxi, Chimborazo, El Oro, Esmeraldas, Guayas, Imbabura, Loja, Los Ríos, Manabí, Morona Santiago, Napo, Pastaza, Pichincha, Tungurahua, Zamora Chinchipe, Galápagos, Sucumbíos, Orellana, Santo Domingo de los Tsáchilas y Santa Elena.

Codificación Provincias

De acuerdo a la División Político-Administrativa (DPA) actualizada por el INEC en el marco de sus funciones, la provincia será codificada de la siguiente manera:

- Se asignan dos códigos para provincia.
- Las provincias quedan codificadas en orden alfabético, a excepción de Galápagos, Sucumbíos, Orellana, Santo Domingo de los Tsáchilas y Santa Elena, creadas en fechas posteriores al establecimiento de este criterio de codificación; a éstas se les asigna el código inmediato superior al que tiene la última provincia, de acuerdo a la fecha de su respectiva creación.
- También existen las zonas no delimitas que se codifican con "90".

La variable Provincia se codifica con dos dígitos como se puede observar en la tabla 1:

Tabla 1. Codificación Provincial

PROVINCIA	CODIGO
Azuay	01
Bolívar	02
Cañar	03
Carchi	04
Cotopaxi	05
Chimborazo	06
El Oro	07
Esmeraldas	08
Guayas	09
Imbabura	10
Loja	11
Los Ríos	12
Manabí	13
Morona Santiago	14
Napo	15
Pastaza	16
Pichincha	17
Tungurahua	18
Zamora Chinchipe	19
Galápagos	20
Sucumbíos	21
Orellana	22
Santo Domingo de los Tsáchilas	23
Santa Elena	24
Zona no delimitada	90

FUENTE: División Político Administrativa DPA-2017

Cantones

Son circunscripciones territoriales conformadas por parroquias rurales y la cabecera cantonal con sus parroquias urbanas, señaladas en su respectiva ley de creación.

En base a la División Político Administrativa DPA-2010 el país tiene 221 cantones.

Tabla 2. Número de cantones por provincia

PROVINCIA	NUMERO DE CANTONES
Azuay	15
Bolívar	7
Cañar	7
Carchi	6
Cotopaxi	7
Chimborazo	10
El Oro	14
Esmeraldas	7
Guayas	25
Imbabura	6
Loja	16
Los Ríos	13
Manabí	22
Morona Santiago	12
Napo	5
Pastaza	4
Pichincha	8
Tungurahua	9
Zamora Chinchipe	9
Galápagos	3
Sucumbíos	7
Orellana	4
Santo Domingo de los Tsáchilas	2
Santa Elena	3

Codificación Cantones

De acuerdo a la División Político-Administrativa (DPA) el cantón será codificado de la siguiente manera:

El código 01 es asignado al cantón, cuya cabecera es también capital provincial. Al resto de cantones se les ordena alfabéticamente, asignándoles el código que corresponda en forma ascendente, a los cantones creados en fechas posteriores al establecimiento de este criterio de codificación se les asigna el código inmediato superior que tiene el último cantón, de acuerdo a la fecha de su respectiva creación.

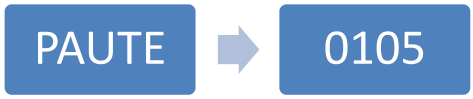
La variable cantón se codifica con dos dígitos como se indica en el siguiente ejemplo:

Provincia: Azuay **01**

Tabla 3. Codificación Cantonal de Azuay

CANTON	CÓDIGO
Cuenca (capital provincial)	01
Girón	02
Gualaceo	03
Nabón	04
Paute	05
Pucará	06
San Fernando	07
Santa Isabel	08

Un cantón se identifica con cuatro dígitos: dos dígitos de provincia y dos dígitos de cantón, de la siguiente manera:



Parroquias

- **Parroquias urbanas:** Los cantones se subdividen en parroquias, las parroquias urbanas son las que se encuentran dentro de la ciudad o área urbana.
- **Parroquias rurales:** Son aquellas que no están incluidas dentro del área urbana.

Codificación Parroquias

Para codificar la cabecera cantonal y parroquias urbanas se utiliza el código 50.

Las parroquias rurales quedan codificadas en orden alfabético desde el código 51 en adelante, a excepción de las parroquias rurales creadas en fechas posteriores al establecimiento de este criterio de codificación, a éstas se les asigna el código inmediato superior al que tiene la última parroquia rural de acuerdo a la fecha de su creación.

La asignación de códigos de parroquias está definida en la División Político-Administrativa.

IMPORTANTE: En la División Político-Administrativa a cada parroquia urbana le corresponde un código desde el 01 hasta el 49, siendo 50 la cabecera cantonal. Sin embargo, para las diferentes investigaciones realizadas, todas las parroquias urbanas son codificadas con "50".

La variable parroquia se codifica con dos dígitos.

Ejemplo de codificación de parroquias

Provincia: Azuay **01**

Cantón: Cuenca **01**

Tabla 4. Codificación parroquias urbanas del Cantón Cuenca

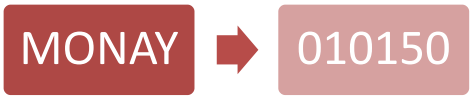
PARROQUIAS URBANAS	CODIGO
Cuenca (Cabecera Cantonal y Capital Provincial)	50
Bellavista	50 (01)
Cañaribamba	50 (02)
El Batán	50 (03)
El Sagrario	50 (04)
El Vecino	50 (05)
Gil Ramírez Dávalos	50 (06)
Huaynacápac	50 (07)
Machángara	50 (08)
Monay	50 (09)
San Blas	50 (10)
San Sebastián	50 (11)
Sucre	50 (12)
Totoracocha	50 (13)
Yanuncay	50 (14)
Hermano Miguel	50 (15)

Tabla 5. Codificación parroquias rurales del Cantón Cuenca

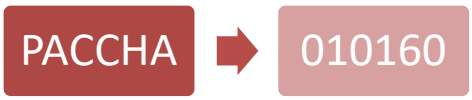
PARROQUIAS RURALES	CODIGO
Baños	51
Cumbe	52
Chaucha	53
Checa	54
Chiquintad	55
Llacao	56
Molleturo	57
Nulti	58
Octavio Cordero Palacios	59
Paccha	60
Quingeo	61
Ricaute	62
San Joaquín	63
Santa Ana	64
Sayausí	65
Sidcay	66
Sinincay	67
Tarqui	68
Turi	69
Valle	70
Victoria del Portete	71

Una parroquia se identifica con 6 dígitos: dos dígitos de provincia, dos dígitos de cantón y dos dígitos de parroquia, de la siguiente manera:

- Parroquia Urbana



- Parroquia Rural



Zonas no delimitadas

Son zonas que tienen pendiente definir sus límites o sus jurisdicciones.

En el siguiente cuadro se detalla la codificación de las zonas no delimitadas existentes en el país de acuerdo a la División Político-Administrativa vigente.

Tabla 6. Codificación Zonas no Delimitadas

Zona no delimitada	Código
Las golondrinas	90 01 51
Manga del cura	90 03 51
El Piedrero	90 04 51

Unidades Estadísticas

En el caso de las encuestas de hogares integrantes del Sistema Integrado de Encuesta a Hogares (SIEH), a partir del año 2018 se determinó la utilización de los conglomerados muestrales[2] como Unidades Primarias de Muestreo - UPM. Los conglomerados muestrales tienen límites bien definidos en los mapas y planos censales, que facilitan el trabajo del listado y aseguran una adecuada cobertura de las viviendas seleccionadas.

En este sentido es necesario conocer los conceptos utilizados a partir de este momento para la construcción de las unidades muestrales:

Zona censal

Es una división estadística que se define como carga de trabajo para la supervisión y control principalmente en los operativos censales.

Zona Censal Amanzanada

Es una superficie perfectamente delimitada, constituida por un promedio de 10 sectores censales amanzanados (aproximadamente 1500 viviendas)

[2] El concepto de conglomerado estará entendido como un conjunto de viviendas que cumplan cierta característica particular, principalmente la del número de viviendas. Para que las viviendas pertenezcan al mismo conjunto deben pertenecer a la misma parroquia, además que cada parroquia estará dividida en dos tipos de zonas, la amanzanada y la dispersa, para entender esto se introduce la división político administrativo del Ecuador.

Zona Censal Dispersa

Está constituida por toda el área de la Parroquia o Cabecera Cantonal, exceptuando el área amanzanada de las mismas.

Codificación de las Zonas

Las zonas se codifican con 3 dígitos tipo cadena de texto como se indica en la siguiente tabla:

ZONAS	CODIFICACION
ÁREA AMANZANADA (Cabecera Cantonal y Parroquial)	001
	002
	003
	004
	005
	:
	:
	899
NO EXISTE	900
LOCALIDAD AMANZANADA	901
	902
	903
	:
	:
	998
AREA DISPERSA	999

Sector censal

Es una división estadística que se define como carga de trabajo de los operativos de campo en investigaciones estadísticas.

Sector Censal Amanzanado

Es una superficie perfectamente delimitada y continua geográficamente, constituido por una o más manzanas. Está conformado por un promedio de 150 viviendas.

Sector Censal Disperso

La zona dispersa se subdivide en Sectores Dispersos. Estos contienen un promedio de 80 a 110 viviendas, y pueden estar constituidos por una o varias Localidades.

Codificación de los Sectores

Los sectores que se encuentran dentro de una zona determinada se codifican con 3 dígitos tipo cadena de texto como se indica en la siguiente tabla:

Sector	001
	002
	003
	004
	:
	999

Conglomerado amanzanado

Agrupación de manzanas integradas por un número de determinado de viviendas que comparten características en común (estrato) y que son próximas entre sí.

Conglomerado disperso

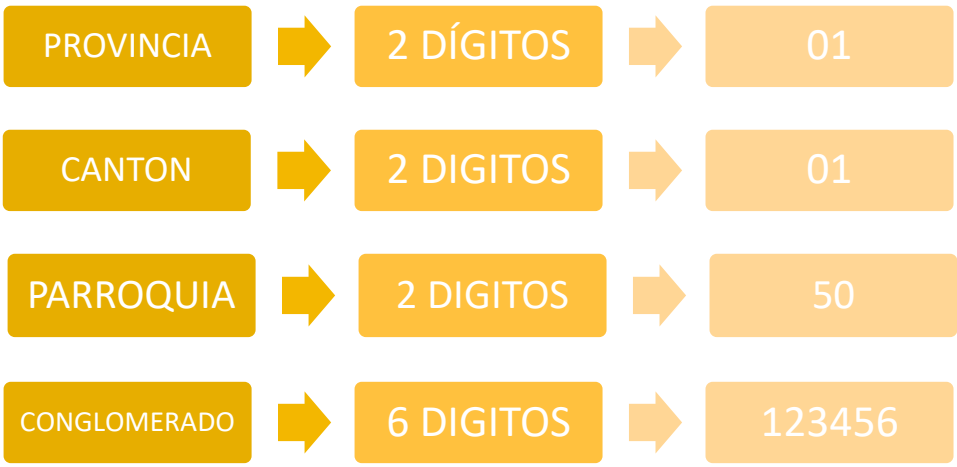
Agrupación de sectores censales integrados por un número de determinado de viviendas que comparten características en común (estrato) y que son próximas entre sí.

Codificación de los conglomerados

Es una identificación única de 6 dígitos por Unidad Primaria de Muestreo diferenciado por área urbana y rural.

Estructura del código de identificación de un conglomerado (UPM)

El número de identificación de un conglomerado (UPM) se define con 4 variables y está formado por 12 dígitos de la siguiente manera:



Unidades Primarias de Muestreo – UPM homogéneas

El nuevo diseño del MMM contempla un cambio en cuanto a la Unidad Primaria de Muestreo – (UPM). Para años anteriores se realizaba la selección de "sectores censales" en función a un criterio de operativo de campo. Sin embargo, debido al crecimiento y disminución de la población en ciertas áreas geográficas, estas UPM pasaron a ser desiguales en cuanto al número de viviendas ocupadas que tienen dentro de sus límites (originalmente, en promedio, 150 viviendas ocupadas en el área amanzanada y 80 en el área dispersa). Esta heterogeneidad ha generado probabilidades de selección inadecuadas en la segunda etapa, es decir, en las viviendas ocupadas.

Tabla 7. Ciudades auto representadas ENEMDU (viviendas totales y viviendas ocupadas por sector censal)

DOMINIO		Viv_tot	Viv_ocu
Ambato	Media	104,7	100,7
	Mínimo	27,0	23,0
	Máximo	219,0	218,0
Cuenca	Media	114,7	111,1
	Mínimo	35,0	35,0
	Máximo	226,0	225,0
Guayaquil	Media	128,5	125,6
	Mínimo	1,0	1,0
	Máximo	247,0	240,0
Machala	Media	119,2	116,0
	Mínimo	6,0	6,0
	Máximo	211,0	207,0
Quito	Media	115,3	114,1
	Mínimo	4,0	4,0
	Máximo	678,0	677,0
Total	Media	120,8	118,4
	Mínimo	1,0	1,0
	Máximo	678,0	677,0

En la tabla 7 se observa que en un período cercano a 8 años utilizando el Marco de Muestreo a partir de CPV 2010, ya se presentaban síntomas de heterogeneidad en la cantidad de viviendas por sectores. Donde los promedios esperados ya difieren de su construcción original como es el caso de los sectores amanzanados en las ciudades considerados como dominios en la ENEMDU.

Este inconveniente se solventa al reconstruir y equilibrar el tamaño de las UPM con respecto al número de viviendas ocupadas.

Construcción de las unidades primarias de Muestreo (Conglomerados)

El objetivo principal de la aplicación de esta nueva metodología es desarrollar un algoritmo eficiente que permita generar conglomerados de viviendas que respondan las necesidades de las investigaciones realizadas por el Instituto Nacional de Estadística y Censos – INEC; desde el punto de vista estadístico y operativo. Se abordará el problema de tal manera que se generará un algoritmo que modifique en pequeña medida la organización logística del operativo, pero a la vez garantice resultados estadísticos más robustos.

Para las investigaciones socio-demográficas el INEC utilizaba el marco maestro de muestreo, conformado por un listado de 40.610 sectores censales (Unidades Primarias de Muestreo - UPM), los mismos que contienen un número determinado de viviendas; que teóricamente se encuentran conformados por 150 viviendas en área amanzanada y entre 80 y 110 en el área rural.

Para resolver este problema se ha generado un "Algoritmo de Generación de Conglomerados para Necesidades Operativas-AGCNO". Este algoritmo garantiza que el número de viviendas por conglomerado se encuentre entre un número t de viviendas ocupadas hasta un número 2t; de esta manera, se asegura que los factores de expansión de diseño pertenezcan a un intervalo que va desde s hasta 2s. El número de viviendas ocupadas que tienen actualmente los conglomerados son de 30 a 60 viviendas ocupadas, tanto para el área amanzanada como para el área *dispersa*.

Algoritmo de Generación de Conglomerados que respondan a necesidades operativas

El problema que se abordará está asociado a un problema de partición de grafos, en el que las manzanas o las grillas son los nodos, las aristas son las fronteras que existen entre las manzanas o grillas y los pesos de los nodos están dados por el número de viviendas ocupadas dentro de cada manzana o grilla.

Sea $z \in \{1, \dots, n\}^n$ tal que $z_i = min_{x_{ij}} j$. Sea $y \in \{0,1\}^n$ tal que $y_i = 1$ si $z_i = i$ si no. La formulación general del problema de partición de grafos con restricciones de tamaño, peso y selección es:

$$z = \min \sum_{i=1}^{n} a_i \qquad (1)$$

s.a.

$$x_{ij} + x_{jk} - x_{ik} \leq 1, \qquad \forall 1 \leq i < j \leq k \qquad (2)$$

$$x_{ij} - x_{jk} + x_{ik} \leq 1, \qquad \forall 1 \leq i < j \leq k \qquad (3)$$

$$-x_{ij} + x_{jk} + x_{ik} \leq 1, \qquad \forall 1 \leq i < j \leq k \qquad (4)$$

$$x_{ij} = 1 \exists \{k_1, \dots, k_p\} \; tal \; que \; x_{ik1} e_{ik1} = x_{k_1 k_2} e_{k_1 k_2} = \cdots = x_{k_p j} e_{k_p j} \qquad (5)$$

La función objetivo (1) busca maximizar el número de conglomerados que se van a formar. Las restricciones (2)-(4) son también llamadas desigualdades triangulares y garantizan que, si tres nodos de V están unidos por dos aristas en un conglomerado, entonces el tercer nodo también pertenece al mismo conglomerado. La restricción (5) garantiza que dos nodos pertenecen a un mismo conglomerado siempre y cuando exista un camino que los junte con nodos que pertenecen al mismo conglomerado.

A nivel Nacional existen 107.014 conglomerados o Unidades Primarias de Muestreo, distribuidos en las 24 provincias del país y a nivel de áreas urbano/rurales. El 74,5% de los sectores censales están ubicados en áreas urbanas, mientras que el 25,5% restante en áreas rurales del país.

Tabla 8. Distribución provincial de conglomerados muestrales y viviendas ocupadas según información CPV 2010

Provincia	Número de conglomerados	Total de viviendas ocupadas
Azuay	5.066	181.359
Bolívar	1.279	46.122
Cañar	1.595	56.694
Carchi	1.288	43.923
Cotopaxi	2.807	100.202
Chimborazo	3.196	114.597
El Oro	4.659	160.519
Esmeraldas	3.733	127.564
Guayas	26.974	936.826
Imbabura	3.040	103.361
Loja	3.194	110.928
Los Ríos	5.459	189.694
Manabí	9.071	315.439
Morona	975	34.563
Napo	701	24.091
Pastaza	617	21.361
Pichincha	21.169	732.845
Tungurahua	3.901	138.560
Zamora Chinchipe	669	23.108
Galápagos	348	8.520
Sucumbíos	1.250	43.234
Orellana	982	33.317
Santo Domingo De Los Tsáchilas	2.807	96.373
Santa Elena	2.025	67.667
Zonas no Delimitadas	209	7.451
Nacional	107.014	3.718.318

Durante la elaboración del MMM, cada unidad estadística está adecuadamente identificada en las diferentes fases y acompañada de

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

Según el informe técnico elaborado por la División de Estadística de las Naciones Unidas (Naciones Unidas, 2007) las unidades de observación agrupadas en conglomerados son muy distintas entre sí (heterogéneas) y tienden a incrementar la varianza de los estimadores, ya que estos conglomerados son generalmente pequeños en comparación con el universo de estudio. Sin embargo, en la práctica del muestreo es común la formación de conglomerados, ya que representan espacios geográficos que se constituyen como unidades de primera etapa para la selección de la muestra.
Al interior de las unidades de primera etapa existe una cierta afinidad entre los elementos que las conforman, por lo que su contribución a la varianza total (intra conglomerados) es menor que aquella que aporta las diferencias entre las agrupaciones (variación entre conglomerados).

Finalmente, al interior de las UPM se requiere contar con un listado exhaustivo de todas las viviendas, en el que se detallen las características de las mismas y permitan a los entrevistadores la identificación de las viviendas seleccionadas, las que son consideradas como unidades últimas o finales de muestreo.

Estratificación del Marco Maestro de Muestreo

La estratificación del MMM consiste en agrupar de acuerdo a ciertas similitudes, las Unidades Primarias de Muestreo (UPM) creadas previamente en base a la Información del Censo de Población y Vivienda 2010.

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

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

Objetivos de la estratificación del MMM

Es muy importante que las unidades de marco se agrupen de acuerdo a características homogéneas, ya que eso reduce el número de selecciones y contribuye a minimizar la varianza. Los objetivos del proceso de estratificación del MMM son los siguientes:

- Agrupar a las unidades de marco en grupos de acuerdo a un conjunto de características socioeconómicas.
- Formar grupos de conglomerados homogéneos en su interior y heterogéneos entre ellos.
- Mejorar el diseño muestral incrementando la eficiencia del mismo y controlando la varianza de los estimadores.

Dominios de estratificación del MMM

Los dominios de estratificación del MMM considerados son cada una de las provincias continentales, divididas en sus componentes urbanas y rurales.

Las unidades que se desean estratificar son los conglomerados que serán considerados como las unidades primarias de muestreo; las unidades de observación son las viviendas particulares ocupadas, los hogares, los residentes habituales de las mismas y las características de las viviendas u hogares.

Metodología de estratificación

Existen varias metodologías que permiten, mediante técnicas matemáticas, la creación de subconjuntos aplicando el análisis de variables cuantitativas como cualitativas. Se debe considerar que la fuente principal de información es el Censo de Población y Vivienda 2010 (CPV 2010), dependiendo de las características de sus variables, se ha tomado como alternativa las siguientes técnicas:

- Para las variables cualitativas, se emplea el *análisis de componentes principales no lineales*. En este caso en particular, se utilizó la técnica del *escalamiento óptimo, mediante mínimos cuadrados alternantes*; el cual permite transformar las variables originales asignando valores a las categorías de cada una de las variables y luego correlacionarlas para caracterizar o analizar la estructura de los datos, es decir, se transforma las variables cualitativas en cuantitativas con el fin de mejorar la combinación lineal de las variables tratadas y minimizando problemas como subjetividad y manipulación.

- Mientras que, para variables cuantitativas, se parte de un *análisis de componentes principales,* el cual tiene como principal objetivo reducir el grupo de variables seleccionadas que representan la naturaleza observada, por un grupo más pequeño de variables no correlacionadas que presenten la mayor parte de información que se encuentra en las variables originales; en este caso con las variables que pertenecen a los módulos de vivienda, hogar y personas del CPV 2010.

a) Análisis de componentes principales mediante mínimos cuadrados alternantes

La metodología se basa en la aplicación de componentes principales mediante mínimos cuadrados alternantes (escalamiento óptimo). Esta metodología se desprende del modelo de componentes principales clásico y un doble algoritmo de escalamiento óptimo, el mismo que, cuantifica las categorías de las variables para maximizar la correlación entre las variables seleccionadas de manera inicial. Posteriormente, con las cuantificaciones obtenidas se estima los parámetros del análisis de componentes principales lineal.

Por otro lado, se denominan *estimaciones de mínimos cuadrados condicionales*, debido a que las estimaciones de mínimos cuadrados de los parámetros en un subconjunto asumen que los parámetros en todos los demás subconjuntos son constantes, puesto que la naturaleza de los mínimos cuadrados es condicional a los valores de los otros parámetros en otros subconjuntos.

Se debe considerar que el primer componente principal es la suma ponderada de las variables originales con mayor varianza, es decir, que las variables nuevas generadas por los componentes principales son combinaciones lineales (sumas ponderadas) de las variables originales.

Para generalizar un conjunto de múltiples variables al grupo de índices J de las *j* variables, se realiza una partición en un subconjunto B, explicado de la siguiente manera:

$$\sigma(X; Y_1, \dots, Y_j) = \mathcal{B}^{-1} \sum_{b} SSQ(X - \sum_{j \in J(b)} G_j Y_j)$$

Como parte del escalamiento óptimo, es importante destacar que las categorías de las variables originales deben pasar por un proceso de recodificación de mayor a menor en su numeración, es decir, si existe en una variable con cinco categorías, éstas según su frecuencia, deben ser recodificadas, considerando que la categoría con mayor frecuencia deberá tener el máximo valor es decir cinco. Este procedimiento se lo debe llevar a cabo con cada una de las variables seleccionadas.

b) Análisis de componentes principales clásico

Esta metodología señala las relaciones que se presentan entre n variables correlacionadas, que se transforman en un nuevo conjunto de variables sin correlación entre sí (que no redundan en información). Este conjunto se denomina conjunto de componentes principales. A su vez son combinaciones lineales de las anteriores variables, construidas por su orden de importancia o según la dimensión a la cual pertenezcan.

Para el análisis de componentes principales, no se requiere el supuesto de normalidad multivariante, no obstante, se debe considerar que las variables fueron escaladas, para evitar subjetividad con sus categorías. Se debe tomar en consideración una serie de variables sobre un grupo de individuos (vivienda, hogar y personas), con lo cual se calcula un nuevo grupo de variables cuya varianza decrezca de una manera progresiva.

Las variables escaladas son clasificadas por dimensiones entre vivienda, hogar y personas, para obtener un mejor resultado y minimizar la varianza en el

ACP_TOTAL, el cual está compuesto por los resultantes del ACP_VIVIENDA, ACP_HOGAR y ACP_PERSONAS. De manera general, se obtiene un vector resultante para posteriormente evaluarlo y clasificarlo:

$$\sum_{i=1}^{p} Var(y_i) = \sum_{i=1}^{p} \lambda_i = traza(\Delta)$$

Dónde:

$\Delta \equiv matriz\ diagonal$

Con el fin de contar con el porcentaje de la varianza total, se recoge un componente principal:

$$\frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i} = \frac{\lambda_i}{\sum_{i=1}^{p} var(x_i)}$$

Cabe recalcar que no se suele seleccionar más de tres componentes principales, con el fin de facilitar el análisis, es decir, se debe seleccionar aquellas variables que explican una proporción aceptable de la varianza global.

c) Dimensiones y variables de estratificación

Tras realizar el cálculo del análisis de componentes principales y obteniendo los resultados de la aplicación se obtienen tres dimensiones y un total de veinte y uno variables de estratificación, el cual se puede observar en la tabla 9.

Tabla 9. Dimensiones y variables de estratificación

VIVIENDA	Tipo de viviendas
	Vías de acceso
	Materiales del techo
	Materiales del piso
	Materiales de las paredes
	Procedencia del agua
	Agua que recibe la vivienda
	Servicio higiénico (eliminación de excretas)
	Eliminación de Basura
HOGAR	Disponibilidad de servicio higiénico del hogar
	Disponibilidad de ducha
	Teléfono convencional
	Teléfono celular

	Internet
	Computador
	TV cable
	Equipamiento del Hogar
	Tenencia de la vivienda
POBLACIÓN	Nivel de instrucción
	Grado
	Características económicas

d) Asignación de puntajes para las dimensiones y variables de estratificación

La metodología de componentes principales clásico proporciona una asignación numérica del 1 al 1.000 para observar de mejor manera la posible agrupación que pertenece por sus características, es por ello que se utiliza el siguiente algoritmo de clasificación:

$$S = \frac{ACP\_TOTAL - score_{min}}{score_{max} - score_{min}}$$

Dónde:

$$\begin{array}{rcl} \text{S} & = & \text{Valor del puntaje final} \\ \text{ACP\_TOTAL} & = & \text{Valor del análisis de componentes principales clásico} \\ score_{min} & = & \text{Valor mínimo obtenido del } \sigma \\ score_{max} & = & \text{Valor máximo obtenido del } \sigma \end{array}$$

A partir de esta puntuación se identifica cada una de las viviendas con sus hogares para realizar la clasificación final, se toman en consideración la presencia de cada una de las dimensiones y variables de estratificación en cada uno de los dominios, los cuales dan como resultado una agrupación única.

e) Clasificación K-medias

Debido a las diferencias en la métrica de las variables, se las transforman antes del análisis, de manera que todas ellas tengan variabilidades similares, con el fin de eliminar del cálculo las distancias.

De esta manera, se aplica el análisis de conglomerados de K-medias, el cual se basa en las distancias euclídeas existentes entre un conjunto de variables, asignando de los K primeros casos a los K conglomerados (centroídes).

$$d_{ii'} = \sqrt{\sum_j (x_{ij} - x_{i'j})^2}$$

Dónde:

X se refiere a las puntuaciones obtenidas por el caso i y el caso i' (i =/ i') en cada una de las j = 1, 2,..., p variables incluidas en el análisis (la sumatoria de la

expresión incluirá p términos, es decir, tantos p como variables).

Aplicación de la estratificación en el Marco de muestreo

El Marco de Muestreo aplicado para la ENEMDU, se lo ha dividido por dominios de estudio, y dentro de ellos sus correspondientes conglomerados, a los cuales se asignó un estrato tomando principalmente en entre otras características geográficas, socio-económicas y socio-demográficas, con la finalidad de mejorar la precisión y exactitud de los estimadores, minimizando su varianza.

Diferenciar entre área urbana y rural es necesario, lo cual se considera como estrato implícito dentro del Marco de Muestreo. Para ello, se toma en cuenta la división o identificación de centros poblados[3] (criterio poblacional) sugerido por la Comunidad Andina de Naciones (CAN); en el – "Seminario: Censos 2000 de Población y Vivienda de los Países Andinos", clasificándolos de la siguiente manera:

Rural:

- Población dispersa y centros poblados con menos de 2000 habitantes.

Urbano:

- 0 a 4.999 habitantes
- 5.000 a 9.999 habitantes
- 10.000 a 19.999 habitantes
- 20.000 a 49.999 habitantes
- 50.000 a 99.999 habitantes
- 100.000 a 199.999 habitantes
- 200.000 a 499.999 habitantes
- 500.000 a 999.999 habitantes
- 1.000.000y más habitantes

Para las encuestas a hogares se considera esta clasificación de área urbana y rural, por lo tanto, la estratificación también usa este criterio para la asignación de los estratos.

La identificación de estrato para cada conglomerado a nivel nacional respeta los límites geográficos de las jurisdicciones a las cuales pertenecen, con la finalidad de que no existan rupturas con la asignación cartográfica preestablecida.

Los estratos creados para los dominios del Marco de Muestreo que se presentan son: (1) Estrato A, (2) Estrato B, (3) Estrato C.

[3] Es todo lugar del territorio nacional rural o urbano, identificado mediante un nombre y habitado con ánimo de permanencia. Sus habitantes se encuentran vinculados por intereses comunes de carácter económico, social, cultural e histórico. Dichos centros poblados pueden acceder, según sus atributos, a categorías como: caserío, pueblo, villa, ciudad y metrópoli.

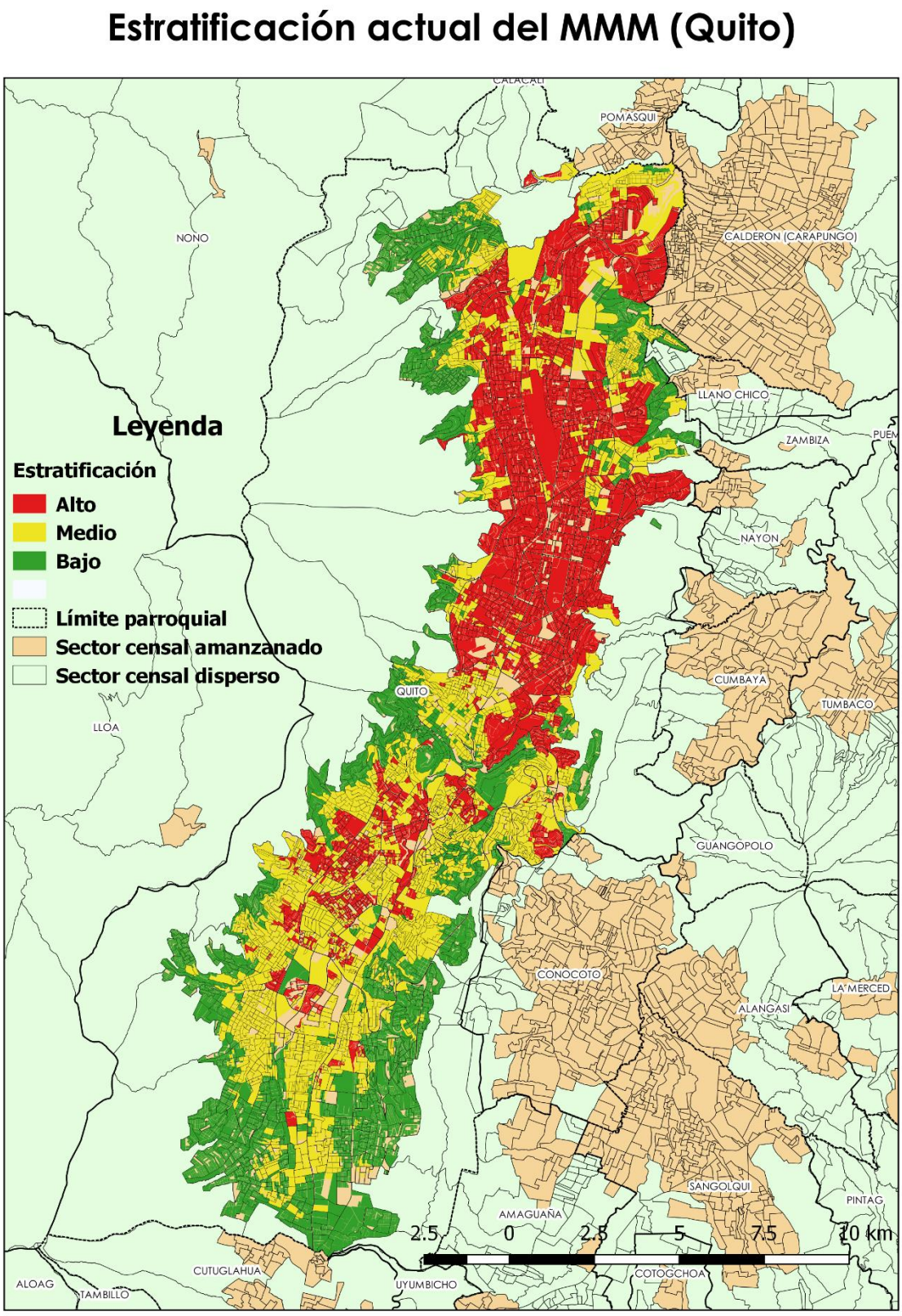
Validación de la estratificación del MMM

Una vez concluida la etapa de estratificación y asignación de atributos a los estratos, es necesario georeferenciar los resultados obtenidos. Esta acción permite cargar la información generada en mapas y planos, que evidencia la distribución en el territorio nacional de los estratos establecidos.

El proceso de validación se realizó junto con personal encargado de actualización cartográfica del INEC, mediante la observación en mapas digitales de la distribución y ubicación de los estratos en cada uno de los dominios. Esto permitió concluir que geográficamente el método empleado representa de gran manera la realidad del territorio en campo.

Por ejemplo, la figura 4 corresponde al mapa digital de la ciudad de Quito. En el mapa se puede ver representada gráficamente la estratificación del MMM. Una descripción general, permite decir que el color rojo que representa el estrato alto de la ciudad que se ubica en la parte norte, el color amarillo que representa el estrato medio que se encuentra en el centro-sur de la ciudad, y el color verde que representa el estrato bajo que se halla distribuido en las zonas periféricas de la ciudad.

Figura 4. Mapa digital de la estratificación del MMM (Quito)



Actualización cartográfica del Marco de Muestreo de forma continua

La primera noción sobre la función de la cartografía en las operaciones estadísticas por muestreo es la de servir de respaldo en el momento del levantamiento de campo y presentar los resultados agregados en forma cartográfica; sin embargo, su importancia es mayor dentro del SIEH ya que de su adecuada implementación dependerá directamente en los resultados finales de cualquier operación estadística que sea componente del SIEH.

En términos generales, un adecuado sistema de Cartografía cumple varias finalidades en una operación estadística, entre los cuales se destacan los siguientes:

- Asegurar la uniformidad y facilitan las actividades propias de una operación estadística.
- Garantizar la cobertura de las muestras generadas y, al mismo tiempo, asegurar que no existan duplicaciones.
- Facilitar la reunión de datos y servir como insumos para los procesos de supervisión de las actividades de levantamiento de información.

Durante el operativo de campo, la cartografía asegura que los investigadores o encuestadores puedan identificar con facilidad el conjunto de unidades muestrales asignadas. Permitiendo desarrollar las tareas de planificación y control asignadas a los supervisores, tanto al seguimiento de las actividades, como a la identificación de aspectos problemáticos y tomar con prontitud medidas correctivas.

Además, la cartografía representada en los mapas y planos facilita la presentación, el análisis y la divulgación de los resultados convirtiéndose en un instrumento poderoso que permite su visualización de resultados, lo que ayuda a identificar modalidades locales de importantes indicadores demográficos y sociales. Por lo tanto, la cartografía es parte integral del análisis de las políticas en los sectores público y privado.

En base a este antecedente la cartografía dentro de la ejecución del SIEH juega un papel importante, por lo que es imperante contar con esta información de forma permanente, actualizada y oportuna; para poder garantizar calidad, coherencia, complementariedad y confiabilidad de los resultados obtenidos en las operaciones estadísticas integrantes del SIEH.

Debido a que los Marcos de Muestreo pierden vigencia en el tiempo por el crecimiento de viviendas (nuevas construcciones) y el cambio en su condición de ocupación[4], se hace necesario contar con una actualización cartográfica

4 Entendiéndose cómo condición de ocupación a como se encontraba la vivienda en el momento de la actualización cartográfica de acuerdo a este criterio, las viviendas se clasifican en viviendas ocupadas, desocupadas, en construcción, demolidas, destruidas y temporales.

permanente de las muestras utilizadas en las diferentes operaciones estadísticas, convirtiéndose en una buena práctica en el desarrollo de las investigaciones por muestreo.

Esta actualización permanente debe estar totalmente desligada de los aspectos presupuestarios ya que se debe propender a contar un financiamiento corriente para su ejecución.

Los aspectos fundamentales que se deben tomar en cuenta dentro de la actualización cartográfica permanente se resumen en lo siguiente:

- Mantenimiento de la base de datos central, que generalmente corresponden la cartografía de los censos de población, para contar con un histórico de las actualizaciones realizadas.
- Creaciones de nuevas jurisdicciones legalmente constituidas que a la fecha de la actualización se hayan producido.
- Los objetivos de la operación estadística deben están directamente relacionados con los requerimientos de actualización cartográfica.
- Coordinación con las unidades de muestreo para la planificación de la actualización cartográfica.
- Coordinación con el operativo de campo para la ejecución de la actualización cartográfica.
- Diseño de herramientas informáticas para el ingreso de la información cartográfica actualizada.
- Controles de calidad de la información cartográfica actualizada.
- Validación de información cartográfica actualizada.
- Producción de materiales de apoyo para levantamiento de campo.

Referencias

Ambrosio L., Villa A. e Iglesias L (1996). Estratificación multivariante - Criterios de evaluación. MAdrid: Estadística Española

Cochran, W.G. (1977). Survey Techniques. Nueva York: John Wiley & Son

Comunidad Andina de Naciones (2010). Decisión 730. Lima: Comunidad Andina de Naciones

Cox, D.R. (2006). Principles of Statistical Inference. Oxford: Nuffield College

Feres, J. y Medina, F. (2001) Hacia un sistema integrado de encuestas de hogares en los países de América Latina. Santiago de Chile: Naciones Unidas.

Haggard, E.A. (1958). Intraclass correlation and the analysis of variance. New York: Dryden.

INEC (2011) Memorias del VII Censo de Población y VI de Vivienda. Quito: Instituto Nacional de Estadística y Censos.

Kish, L. (1965). Survey Sampling. Nueva York: John Wiley & Son.

Knight, Keith (1999). Mathematical Statistics. Ontario: Chapman and Hall / CRC.

Moncada G. y Lee H.. (2005). MECOVI: Mejora de las encuestas y medición de las condiciones de vida en América Latina y el Caribe. Banco Mundial.

Naciones Unidas (2007). Encuestas de hogares en los países en desarrollo y en transición. New York: Naciones Unidas

Sampha, S. (2001). Sampling Theory and Methods. New Delhi: Narosa Publishing House

Särndal, C.E., Swenson, B., & Wretman, H.J (1991). Model assisted survey sampling. Nueva York: Springer-Verlag

INEC

Buenas cifras,
mejores vidas

@ecuadorencifras @ecuadorencifras @InecEcuador t.me/ecuadorencifras INEC/Ecuador INECEcuador INEC Ecuador