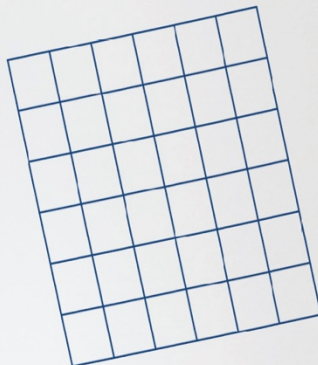


**Marco  
Maestro de  
Muestreo para  
encuestas de  
hogares**

— Enero, 2019



**Roberto Castillo**  
DIRECTOR EJECUTIVO

**Magaly Paredes**  
SUBDIRECTORA GENERAL

**Markus Nabernegg**  
COORDINADOR GENERAL TÉCNICO DE PRODUCCIÓN ESTADÍSTICA

**Christian Garcés**  
DIRECTOR DE INFRAESTRUCTURA ESTADÍSTICA Y MUESTREO

**Equipo técnico:**

Ángel Gaibor, Javier Núñez, Francisco Céspedes, Christian Garcés.

**Propiedad Institucional**

©INEC

Instituto Nacional de Estadística y Censos  
Juan Larrea N15-36 y José Riofrío. Casilla postal 135 C  
Telf: (02) 2555-701 / 2529-858

**Citar como:**

**INEC (2019).** Marco Maestro de Muestreo para encuestas de hogares. Instituto Nacional de Estadística y Censos, Quito-Ecuador.

## Contenido

Introducción .....	5
1. Marco Maestro de Muestreo (MMM) .....	5
1.1. Definición .....	5
1.2. Objetivos .....	5
1.3. Propiedades .....	6
1.4. Cobertura .....	6
1.5. Población objetivo .....	6
1.6. Unidades estadísticas muestrales de selección .....	7
2. Codificación de variables para definir la clave única de identificación para áreas geográficas y unidades estadísticas. ....	8
2.1. Áreas geográficas - División Politico Administrativa .....	9
2.1.1. Provincia .....	9
2.1.2. Codificación Provincias .....	9
2.1.3. Cantones .....	10
2.1.4. Codificación Cantones .....	11
2.1.5. Parroquias .....	11
2.1.6. Codificación Parroquias .....	12
2.1.7. Zonas no delimitadas .....	13
2.2. Unidades Estadísticas .....	14
2.2.1. Zona censal .....	14
2.2.2. Zona Censal Amanzanada .....	14
2.2.3. Zona Censal Dispersa .....	14
2.2.4. Codificación de las Zonas .....	14
2.2.5. Sector censal .....	15
2.2.6. Sector Censal Amanzanado .....	15
2.2.7. Sector Censal Disperso .....	15
2.2.8. Codificación de los Sectores .....	15
2.2.9. Conglomerado amanzanado .....	15
2.2.10. Conglomerado disperso .....	15
2.2.11. Codificación de los conglomerados .....	15
2.3. Estructura del código de identificación de la UPM .....	15
3. Construcción de la Unidades Primarias de Muestreo (Conglomerados) .....	17
3.1. Algoritmo de Generación de Conglomerados que respondan necesidades operativas .....	17
4. Estratificación del Marco Maestro de Muestreo .....	19
4.1. Objetivos de la estratificación .....	19
4.2. Dominios de estratificación .....	19
4.3. Metodología de estratificación .....	19
4.4. Aplicación de la estratificación en el Marco .....	23

4.5	Validación de la estratificación del MMM.....	23
5	Aplicación del Marco Maestro de Muestreo (caso ENEMDU) .....	26
5.1	Homogeneidad de las Unidades Primarias de Muestreo (UPM) .....	26
5.2	Optimización de la dispersión de la muestra .....	27
5.3	Mejora de la precisión de los estimadores y ampliación de la medición a nivel de jurisdicciones de relevancia nacional.....	28
6	Actualización cartográfica del Marco de Muestreo de forma continua .....	29
	<b>Referencias</b> .....	31
	<b>Anexo 1. Análisis coeficientes de variación ENEMDU enero 2017 – enero 2018</b> .....	32

## Introducción

Con el fin de mejorar la cobertura de las encuestas de hogares 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 produce 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 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 una condición necesaria, 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 Maestro de Muestreo 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.

### 1. Marco Maestro de Muestreo (MMM)

#### 1.1. Definición

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. El MMM 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 Marco 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 Marco tienen una probabilidad conocida y diferente de cero de ser seleccionados de alguna de las muestras que se puedan extraer del mismo.

#### 1.2. Objetivos

Los principales objetivos que tiene que cumplir el Marco para 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 los censos de Población y Vivienda y con otros estudios estadísticos orientados a obtener información de las condiciones de los hogares particulares.

### 1.3. Propiedades

El Marco 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, 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 objetivo 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 unidades económicas.

En este sentido, las propiedades del Marco son:

- **Exhaustividad:** Se considera exhaustivo respecto a la población objetivo, dado que todos los miembros que la constituyen (el universo) quedan cubiertos por el mismo.
- **Exactitud:** Es exacto ya que todos los miembros de la población objetivo se incluyen una sola vez.
- **Actualización:** El Marco permitirá reflejar la reciente construcción o demolición de viviendas, los movimientos migratorios en las unidades de vivienda, los nacimientos o las defunciones. Esto permitirá cumplir con el requisito de que en una muestra probabilística cada miembro de la población objetivo tenga una probabilidad conocida de ser seleccionado.

### 1.4. Cobertura

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

### 1.5. Población objetivo

Una consideración importante a la hora de estructurar un marco maestro de muestreo, 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 de selección en las diferentes etapas del muestreo. Las unidades del marco se constituyen en un elemento básico para el cálculo, asignación y selección de la muestra de unidades poblacionales.

El Marco determina como población objetivo y unidades de última selección a las viviendas particulares<sup>1</sup> 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 Pre-censo. 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 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.

#### 1.6. Unidades estadísticas muestrales de selección

El marco 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 Marco 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 Marco.

---

<sup>1</sup> Se excluye a la población en viviendas colectivas, viviendas flotantes y sectores con población indigente

## 2. Codificación de variables para definir la clave única de identificación para áreas geográficas y unidades estadísticas.

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.

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:** Hace referencia a 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 para la identificación de la unidad estadística y 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.



## 2.1. 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 Art. 242, para la administración del Estado y la representación política el país será dividido en provincias, cantones y parroquias *"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"*.

### 2.1.1. 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.

### 2.1.2. 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 cada 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 estas 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 delimitadas que se codifican con "90".

La variable provincia se codifica con dos dígitos como se puede observar en la tabla .

**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

### 2.1.3. 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) del 2010, el país tiene 221 cantones como se señala en la Tabla 2.

**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
Sucumbios	7
Orellana	4
Santo Domingo de los Tsáchilas	2
Santa Elena	3

#### 2.1.4. Codificación Cantones

De acuerdo a la División Politico-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 para la provincia de Azuay, cuyo código de provincia es 01 (Tabla 3).

**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, donde 01 corresponde al código de la provincia de Azuay y 05 al cantón Paute.



#### 2.1.5. 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.

### 2.1.6. 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 Politico-Administrativa.

Es importante mencionar que en la División Politico-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. A continuación un ejemplo de codificación de parroquias, para la provincia de Azuay (01) y el 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 seis dígitos: dos dígitos de provincia, dos dígitos de cantón y dos dígitos de parroquia, de la siguiente manera:

**Parroquia Urbana**



**Parroquia Rural**



**2.1.7. 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
El Piedrero	90 04 51

## 2.2. 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<sup>2</sup> como Unidades Primarias de Muestreo - UPM. Los conglomerados muestrales tienen límites 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:

### 2.2.1. Zona censal

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

### 2.2.2. Zona Censal Amanzanada

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

### 2.2.3. Zona Censal Dispersa

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

### 2.2.4. Codificación de las Zonas

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

**Tabla 7:** Codificación Zonas

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

<sup>2</sup> 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.

	:
	998
AREA DISPERSA	999

### 2.2.5. Sector censal

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

### 2.2.6. Sector Censal Amanzado

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

### 2.2.7. 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.

### 2.2.8. Codificación de los Sectores

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

**Tabla 8:** Codificación Sectores

Sector	001
	002
	003
	004
	:
	999

### 2.2.9. 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.

### 2.2.10. Conglomerado disperso

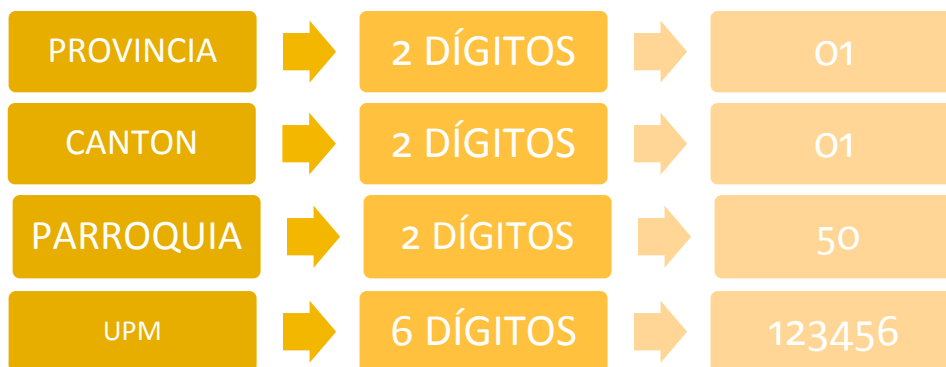
Agrupación de sectores censales integrados por un número de determinado de viviendas que comparten características en común (estrato) y que son próximas entre sí.

### 2.2.11. Codificación de los conglomerados

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

## 2.3. Estructura del código de identificación de la UPM

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



Unidades Primarias de Muestreo – UPM's homogéneas.

El nuevo diseño del Marco contempla una innovación 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's 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 dispersas en la segunda etapa, es decir, en las viviendas ocupadas.

**Tabla 9:** 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 ocho 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 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.



### 3. Construcción de la 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 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 previo al 2018, utilizaba el marco maestro de muestreo conformado por un listado de 40.610 sectores censales, los mismos que contienen un número determinado de viviendas; que teóricamente se encuentran conformados por 150 viviendas en el área amanzanada y entre 80 y 110 en el área rural.

Para generar UPM más homogéneas 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.

#### 3.1. Algoritmo de Generación de Conglomerados que respondan 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:

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

s.a.

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

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

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

$$x_{ij} = 1 \exists \{k_1, \dots, k_p\} \text{ tal que } x_{ik_1} e_{ik_1} = x_{k_1 k_2} e_{k_1 k_2} = \dots = x_{k_p j} e_{k_p j} \quad (5)$$

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 10:** Distribución provincial de conglomerados muestrales y viviendas ocupadas según información CPV 2010

<i>Provincia</i>	<i>Número de conglomerados</i>	<i>Total de viviendas ocupadas</i>
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 Rios	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
Sucumbios	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
<b>Nacional</b>	<b>107.014</b>	<b>3.718.318</b>

Durante el proceso de elaboración del Marco, cada unidad estadística está adecuadamente identificada en las diferentes fases y acompañada de información complementaria que hace 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's 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.

#### **4. Estratificación del Marco Maestro de Muestreo**

La estratificación del Marco 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.

##### **4.1 Objetivos de la estratificación**

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 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.

##### **4.2 Dominios de estratificación**

Los dominios de estratificación del Marco 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.

##### **4.3 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.
- 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:

$$\sigma(x; y_1, \dots, y_j) = B^{-1} \sum_b SSQ(x - \sum_{j \in J(b)} G_j y_j)$$

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 lleva 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 considerar 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:

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

Dónde:

$\Delta \equiv$  matriz diagonal

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

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

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 21 variables de estratificación, el cual se puede observar en la tabla 8.

**Tabla 11:** 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 ducha	
Teléfono convencional	
Teléfono celular	
Internet	
Computador	
TV cable	
Equipamiento	
Tenencia de la vivienda	
POBLACIÓN	
	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:

S	=	Valor del puntaje final
ACP_TOTAL	=	Valor del análisis de componentes principales clásico
score <sub>min</sub>	=	Valor mínimo obtenido del $\sigma$
score <sub>max</sub>	=	Valor máximo obtenido del $\sigma$

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).

#### 4.4 Aplicación de la estratificación en el Marco

El Marco Maestro 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 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. Para ello, se toma en cuenta la división o identificación de centros poblados<sup>3</sup> (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:

- 2.000 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
- 1.000.000 y 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 respetan los límites geográficos de las jurisdicciones a las cuales pertenecen, para que no existan rupturas con la asignación cartográfica preestablecida.

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

#### 4.5 Validación de la estratificación del MMM

<sup>3</sup> 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.

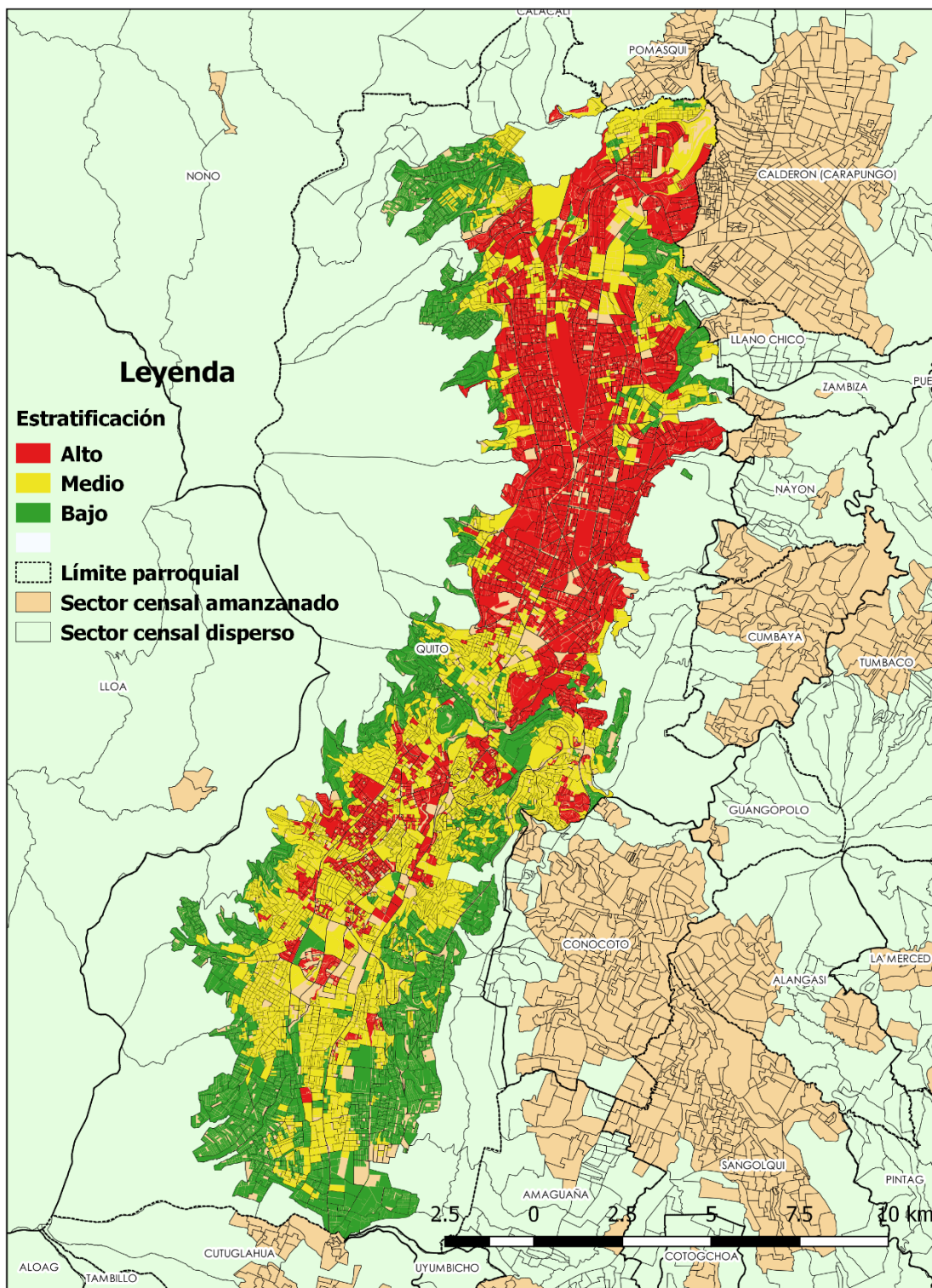
Una vez concluida la etapa de estratificación y asignación de atributos a los estratos, es necesario geo-referenciar los resultados obtenidos. Esta acción permite cargar la información generada en mapas y planos, que evidencia la distribución en el territorio nacional de los estratos establecidos.

El proceso de validación se realizó junto con el personal encargado de la actualización cartográfica, 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 observar la estratificación del Marco Una descripción general, indica que el color rojo representa el estrato alto que se ubica en la parte norte, el color amarillo representa el estrato medio que se encuentra en el centro-sur de la ciudad, y el color verde representa el estrato bajo que se distribuye en las zonas periféricas de la ciudad.



Figura 1: Mapa digital de la estratificación del MMM (Quito)



## 5 Aplicación del Marco Maestro de Muestreo (caso ENEMDU)

Como se mencionó previamente, se implementa una nueva metodología de la construcción de unidades primarias de muestreo con el objetivo de generar UPMs más homogéneas. Durante el año 2017 se realizaron varias pruebas sobre los resultados y posibles efectos en la construcción del marco de muestreo a partir de la metodología de conglomerados.

Estas pruebas se las realizaron en operaciones estadísticas por muestreo que fueron parte de las asistencias técnicas brindadas por el INEC a otras instituciones, para luego incluirlas a partir de agosto de 2017 en la ENEMDU para las 5 ciudades auto-representadas en los levantamiento mensuales.

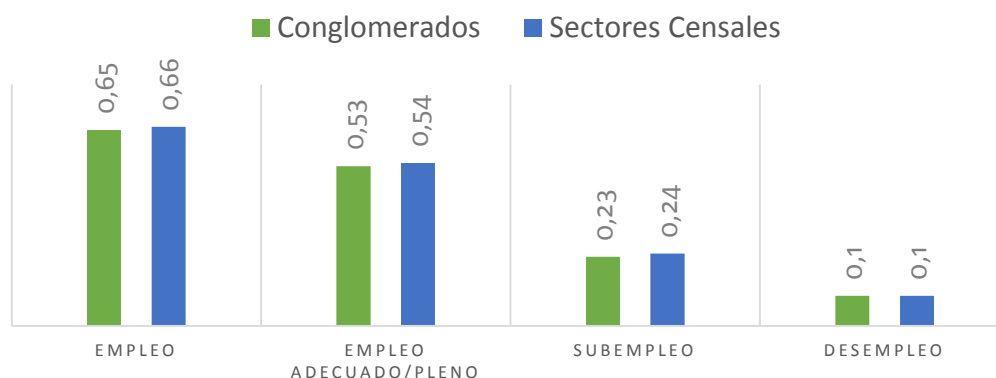
Tanto los procedimientos como los resultados obtenidos fueron discutidos en el marco de la asistencia técnica brindada por la CEPAL, la cual evaluó y avaló los procedimientos que actualmente se están aplicando para la ENEMDU, así como también emitió recomendaciones a los procesos implementados.

Las evaluaciones se centraron en los siguientes puntos: homogeneidad de las unidades primarias de muestreo, optimización de la dispersión de la muestra y mejora en la precisión de la estimación de los indicadores.

### 5.1 Homogeneidad de las Unidades Primarias de Muestreo (UPM)

Para verificar posibles efectos en las estimaciones por el cambio en la construcción de las unidades primarias de muestreo por conglomerados, se realizaron diferentes pruebas, donde se evidencia que no existen diferencias significativas en los indicadores de mercado laboral entre la muestra proveniente de sectores censales y de conglomerados. Así por ejemplo, en octubre de 2017 se levantó el 50% de la muestra de ENEMDU con el marco antiguo (sectores censales) y el 50% con el nuevo marco (conglomerados). Los dos diseños muestrales captan a la población objetivo en magnitudes y proporciones similares como se observa en la Figura 2.

**Figura 2.** Principales indicadores de mercado laboral según metodologías para la construcción de UPM, Octubre 2017  
(En porcentajes)

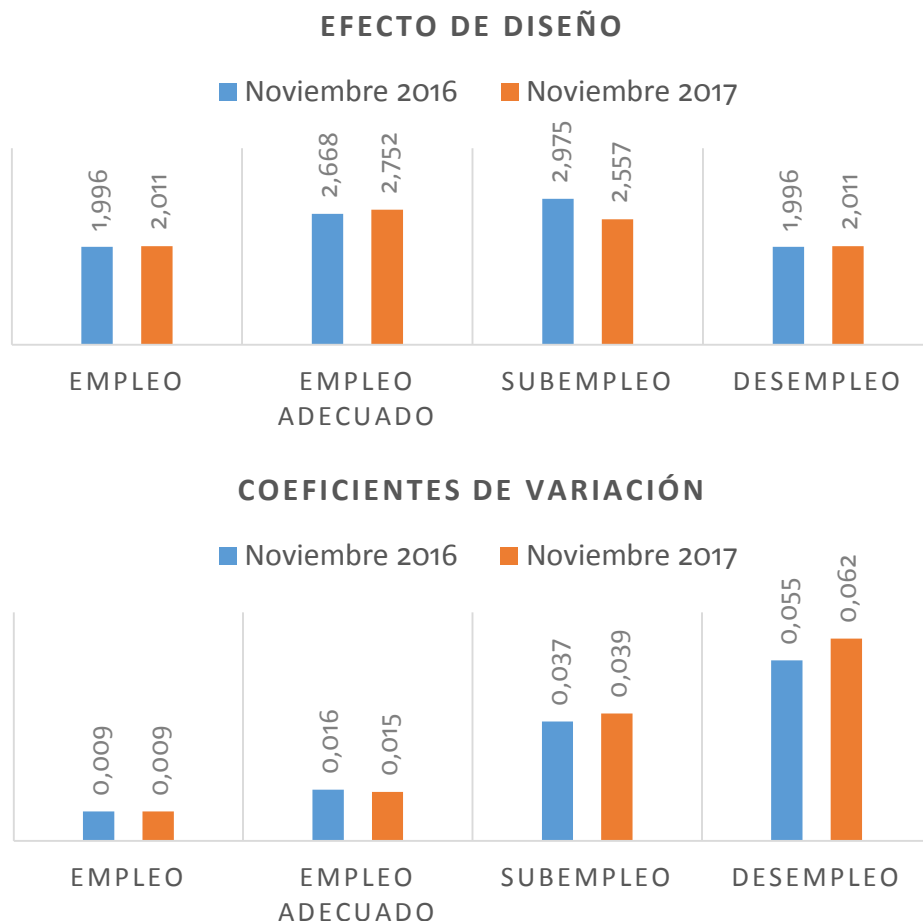


Otro ejercicio que se realizó es la comparación de noviembre 2016 y noviembre 2017. En noviembre de 2017, se incluyó el 100% de la muestra con el nuevo marco. En este caso se analizó los estimadores con respecto al diseño muestral, tales como coeficiente de variación y efectos de diseño<sup>4</sup> Con el objetivo de evaluar el nivel de

<sup>4</sup> Incluir una nota de qué es y como se calcula el efecto de diseño

precisión entre los dos diseños muestrales. La Figura 6 muestra que tanto los coeficientes de variación como los efectos de diseño de ciertos indicadores de mercado laboral son similares tanto con el diseño muestral basado en sectores censales como aquel basado en conglomerados.

**Figura 3.** Principales indicadores de mercado laboral según metodologías para la construcción de UPM, Noviembre 2016 y 2017 (En porcentajes)



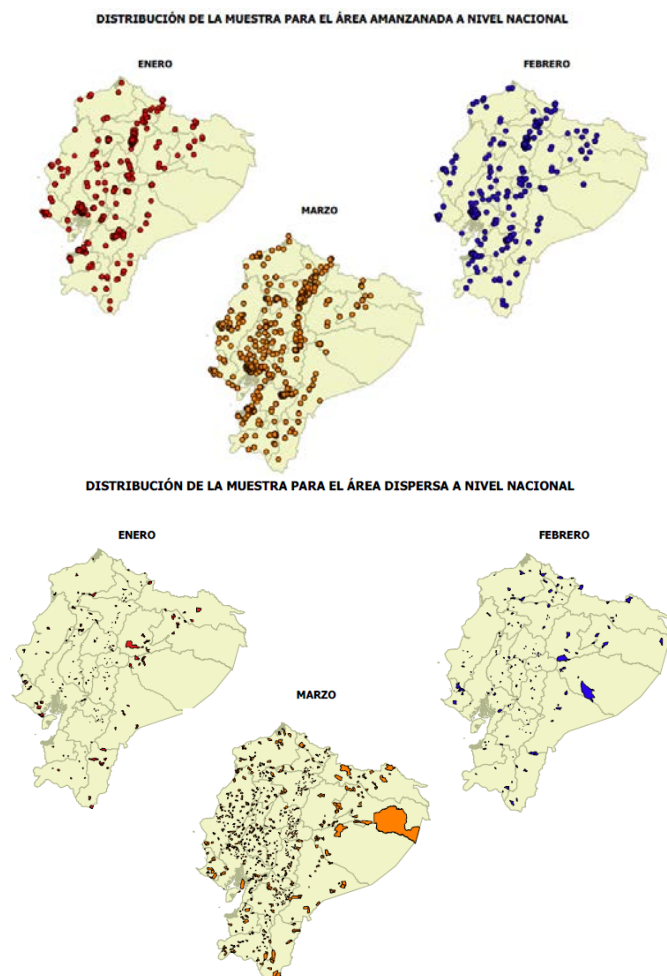
## 5.2 Optimización de la dispersión de la muestra

La dispersión de la muestra en el territorio mejora la precisión de la estimación, debido a que la muestra basada en conglomerados recoge de mejor manera la diversidad de características que de la población objetivo.

Para verificar la dispersión de la muestra se han utilizado herramientas geográficas especializadas con la finalidad de verificar espacialmente la distribución desarrollada en el plan de muestreo, cuya premisa principal es que esté presente en todo el territorio nacional.

Como se evidencia en la Figura 3, la muestra está distribuida en todo el territorio nacional con lo cual se garantiza la representatividad de la información de tanto las estimaciones en sus dominios de estudio.

**Figura 4.** Distribución de la muestra en el territorio nacional por área amanzanada y dispersa



### 5.3 Mejora de la precisión de los estimadores y ampliación de la medición a nivel de jurisdicciones de relevancia nacional

El propósito de las mejoras incluidas en la ENEMDU es optimizar la precisión de los principales indicadores de mercado laboral los cuales se pueden verificar mediante el cálculo de errores, como el coeficiente de variación e intervalos de confianza; indicadores que dan una aproximación de la eficiencia del diseño muestral ejecutado.

La actualización del Marco Maestro no afecta la comparabilidad de los indicadores de las encuestas.

Al no romperse la serie de las encuestas, los datos son totalmente comparables.

Este cambio es el resultado de un proceso planificado de un año de duración.

La actualización del Marco Maestro se hizo con el apoyo técnico de KOSTAT – Corea del Sur y CEPAL.

En resumen, el análisis de indicadores realizados en enero 2018 evidencia una notable mejora en las estimaciones de errores asociados al muestreo para los principales indicadores de mercado laboral para el caso puntual (Ver Anexo 1).

## 6 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 como respaldo en el momento del levantamiento de campo y presentar los resultados agregados en forma cartográfica. Sin embargo, su importancia es mayor dentro del Sistema Integrado de Encuestas de Hogares (SIEH) ya que de su adecuada implementación dependerán directamente los resultados finales de cualquier operación estadística que sea componente del SIEH.

En términos generales, un adecuado sistema de cartografía posee varios objetivos en una operación estadística, entre los cuales se destacan los siguientes:

- Asegurar la uniformidad y facilitar 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 insumo 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 así desarrollar las tareas de planificación y control a los supervisores, tanto al seguimiento de las actividades, como a la identificación de aspectos problemáticos y tomar medidas correctivas, de ser el caso

Adicionalmente, la cartografía representada en los mapas y planos facilita la presentación, el análisis y la divulgación convirtiéndose en un instrumento poderoso que permite la 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 del SIEH es un componente básico por lo que es necesario contar con esta información de forma continua, 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 maestros de muestreo pierden vigencia en el tiempo por el crecimiento de viviendas (nuevas construcciones) y el cambio en su condición de ocupación<sup>5</sup>, se hace necesario contar con una actualización cartográfica continua de las muestras utilizadas en las diferentes operaciones estadísticas, convirtiéndose en una buena práctica en el desarrollo de las investigaciones por muestreo.

Los aspectos fundamentales a tomar en cuenta para la actualización cartográfica continua son los siguientes:

- Mantenimiento de la base de datos central, que generalmente corresponden a la cartografía de los censos de población, para contar con un histórico de las actualizaciones realizadas.
- Creación de nuevas jurisdicciones legalmente constituidas que a la fecha de la actualización se hayan producido.
- Los objetivos de la operación estadística deben estar directamente relacionados con los requerimientos de actualización cartográfica.

---

<sup>5</sup> Entendiéndose como 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.

- 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

## Anexo 1. Análisis coeficientes de variación ENEMDU enero 2017 – enero 2018

Tasa	Descomposición	ene-17			ene-18			Prueba de hipótesis a dos colas						VARIACIÓN		
		Indicador	Error estándar	Coefficiente de variación	Indicador	Error estándar	Coefficiente de variación	Diferencia	Valor z	Normal	p-value	Conclusión	Significancia	COEFICIENTE DE VARIACIÓN	VARIACIÓN ABSOLUTA	VARIACIÓN PORCENTUAL
Desempleo	5 ciudades	6,5%	0,00	5,84	5,8%	0,00	5,04	-0,7%	1,3814	0,9164	0,1672	No rechazo Ho	No Significancia	MEJORÓ	-0,81	13,8%
	Quito	8,0%	0,01	9,66	8,0%	0,01	7,13	0,0%	0,0023	0,5009	0,9981	No rechazo Ho	No Significancia	MEJORÓ	-2,53	26,2%
	Guayaquil	6,1%	0,01	8,62	4,4%	0,00	9,50	-1,7%	2,5942	0,9953	0,0095	Rechazo Ho	Significancia	SUBIÓ	0,88	-10,2%
	Cuenca	3,8%	0,01	15,36	4,9%	0,01	13,47	1,1%	1,2452	0,8935	0,2131	No rechazo Ho	No Significancia	MEJORÓ	-1,90	12,3%
	Machala	4,7%	0,01	18,01	7,1%	0,01	11,20	2,4%	2,0962	0,9820	0,0361	Rechazo Ho	Significancia	MEJORÓ	-6,80	37,8%
	Ambato	5,4%	0,01	13,13	4,7%	0,01	12,42	-0,7%	0,8095	0,7909	0,4183	No rechazo Ho	No Significancia	MEJORÓ	-0,71	5,4%
Empleo Adecuado/Pleno	5 ciudades	52,8%	0,01	1,45	57,9%	0,01	1,17	5,2%	5,0466	1,0000	0,0000	Rechazo Ho	Significancia	MEJORÓ	-0,28	19,1%
	Quito	63,0%	0,01	2,15	65,5%	0,01	1,82	2,5%	1,3836	0,9168	0,1665	No rechazo Ho	No Significancia	MEJORÓ	-0,33	15,3%
	Guayaquil	45,9%	0,01	2,54	52,2%	0,01	1,98	6,3%	4,0072	1,0000	0,0001	Rechazo Ho	Significancia	MEJORÓ	-0,56	22,0%
	Cuenca	56,9%	0,02	2,99	63,6%	0,02	2,36	6,7%	2,9361	0,9983	0,0033	Rechazo Ho	Significancia	MEJORÓ	-0,63	20,9%
	Machala	47,5%	0,02	3,40	54,5%	0,02	3,00	7,0%	3,0469	0,9988	0,0023	Rechazo Ho	Significancia	MEJORÓ	-0,40	11,7%




Empleo Adecuado/Pleno	Ambato	47,7%	0,02	3,65	55,8%	0,02	2,78	8,2%	3,5106	0,9998	0,0004	Rechazo Ho	Significancia	MEJORÓ	-0,87	23,7%
Subempleo	5 ciudades	18,9%	0,01	3,34	14,9%	0,01	3,43	-4,0%	4,9241	1,0000	0,0000	Rechazo Ho	Significancia	SUBIÓ	0,09	-2,7%
Subempleo	Quito	11,1%	0,01	9,56	9,4%	0,01	8,98	-1,8%	1,2920	0,9018	0,1964	No rechazo Ho	No Significancia	MEJORÓ	-0,58	6,1%
Subempleo	Guayaquil	25,7%	0,01	3,86	19,8%	0,01	4,11	-5,8%	4,5584	1,0000	0,0000	Rechazo Ho	Significancia	SUBIÓ	0,25	-6,4%
Subempleo	Cuenca	9,8%	0,01	10,18	10,6%	0,01	10,31	0,8%	0,5413	0,7058	0,5883	No rechazo Ho	No Significancia	SUBIÓ	0,12	-1,2%
Subempleo	Machala	17,6%	0,02	11,23	12,7%	0,01	9,65	-4,9%	2,1179	0,9829	0,0342	Rechazo Ho	Significancia	MEJORÓ	-1,58	14,0%
Subempleo	Ambato	19,8%	0,02	7,95	12,7%	0,01	7,74	-7,1%	3,8452	0,9999	0,0001	Rechazo Ho	Significancia	MEJORÓ	-0,21	2,6%
Otro empleo no pleno	5 ciudades	16,5%	0,01	3,33	16,7%	0,01	3,18	0,3%	0,3552	0,6388	0,7225	No rechazo Ho	No Significancia	MEJORÓ	-0,15	4,5%
Otro empleo no pleno	Quito	13,3%	0,01	7,11	14,0%	0,01	6,31	0,7%	0,5592	0,7120	0,5760	No rechazo Ho	No Significancia	MEJORÓ	-0,80	11,2%
Otro empleo no pleno	Guayaquil	17,3%	0,01	4,81	18,3%	0,01	4,60	1,0%	0,8412	0,7999	0,4002	No rechazo Ho	No Significancia	MEJORÓ	-0,21	4,3%
Otro empleo no pleno	Cuenca	20,8%	0,01	6,54	15,5%	0,01	7,31	-5,3%	2,9977	0,9986	0,0027	Rechazo Ho	Significancia	SUBIÓ	0,78	-11,9%
Otro empleo no pleno	Machala	23,8%	0,02	9,89	20,3%	0,01	5,72	-3,5%	1,3334	0,9088	0,1824	No rechazo Ho	No Significancia	MEJORÓ	-4,17	42,1%
Otro empleo no pleno	Ambato	15,7%	0,01	8,05	18,7%	0,01	6,23	3,0%	1,7407	0,9591	0,0817	No rechazo	No Significancia	MEJORÓ	-1,82	22,6%


										Ho	a					
No remunerado	5 ciudades	4,7%	0,00	6,72	4,1%	0,00	7,02	-0,6%	1,3699	0,9146	0,1707	No rechazo Ho	No Significancia	SUBIÓ	0,30	-4,5%
No remunerado	Quito	3,7%	0,01	15,68	2,8%	0,00	15,29	-0,9%	1,3022	0,9036	0,1929	No rechazo Ho	No Significancia	MEJORÓ	-0,39	2,5%
No remunerado	Guayaquil	4,1%	0,00	10,87	4,4%	0,00	10,97	0,3%	0,4645	0,6788	0,6423	No rechazo Ho	No Significancia	SUBIÓ	0,10	-0,9%
No remunerado	Cuenca	8,7%	0,01	11,29	5,4%	0,01	12,18	-3,3%	2,7535	0,9971	0,0059	Rechazo Ho	Significancia	SUBIÓ	0,89	-7,9%
No remunerado	Machala	6,2%	0,01	13,66	5,3%	0,01	12,26	-0,9%	0,8217	0,7944	0,4113	No rechazo Ho	No Significancia	MEJORÓ	-1,40	10,3%
No remunerado	Ambato	10,7%	0,01	9,05	8,0%	0,01	9,28	-2,7%	2,2100	0,9864	0,0271	Rechazo Ho	Significancia	SUBIÓ	0,23	-2,5%
Participación Bruta	5 ciudades	47,2%	0,00	1,02	47,1%	0,00	0,84	-0,1%	0,1955	0,5775	0,8450	No rechazo Ho	No Significancia	MEJORÓ	-0,18	17,8%
Participación Bruta	Quito	45,3%	0,01	1,68	46,1%	0,01	1,57	0,8%	0,7508	0,7736	0,4528	No rechazo Ho	No Significancia	MEJORÓ	-0,12	7,1%
Participación Bruta	Guayaquil	48,3%	0,01	1,63	47,9%	0,01	1,23	-0,4%	0,3720	0,6451	0,7099	No rechazo Ho	No Significancia	MEJORÓ	-0,40	24,7%
Participación Bruta	Cuenca	48,4%	0,01	1,72	45,6%	0,01	1,82	-2,8%	2,3851	0,9915	0,0171	Rechazo Ho	Significancia	SUBIÓ	0,10	-6,1%
Participación Bruta	Machala	44,8%	0,01	3,01	45,8%	0,01	1,96	1,0%	0,6345	0,7371	0,5258	No rechazo Ho	No Significancia	MEJORÓ	-1,05	34,9%
Participación Bruta	Ambato	51,9%	0,01	1,78	50,0%	0,01	1,46	-1,9%	1,6177	0,9471	0,1057	No rechazo Ho	No Significancia	MEJORÓ	-0,32	18,0%
Participación Global	5 ciudades	64,6%	0,01	0,86	62,6%	0,00	0,74	-2,0%	2,6968	0,9965	0,0070	Rechazo Ho	Significancia	MEJORÓ	-0,13	14,6%

Participación Global	Quito	60,7%	0,01	1,66	60,3%	0,01	1,41	-0,4%	0,2997	0,6178	0,7644	No rechazo Ho	No Significancia	MEJORÓ	-0,25	15,0%
Participación Global	Guayaquil	67,7%	0,01	1,27	64,6%	0,01	1,05	-3,1%	2,8189	0,9976	0,0048	Rechazo Ho	Significancia	MEJORÓ	-0,21	16,8%
Participación Global	Cuenca	63,5%	0,01	1,71	60,3%	0,01	1,75	-3,2%	2,1307	0,9834	0,0331	Rechazo Ho	Significancia	SUBIÓ	0,04	-2,6%
Participación Global	Machala	62,5%	0,01	2,22	62,6%	0,01	1,64	0,1%	0,0500	0,5199	0,9601	No rechazo Ho	No Significancia	MEJORÓ	-0,58	26,0%
Participación Global	Ambato	68,6%	0,01	1,88	65,3%	0,01	1,44	-3,2%	2,0372	0,9792	0,0416	Rechazo Ho	Significancia	MEJORÓ	-0,43	23,2%
Empleo Global	5 ciudades	93,5%	0,00	0,41	94,2%	0,00	0,31	0,7%	1,3814	0,9164	0,1672	No rechazo Ho	No Significancia	MEJORÓ	-0,09	23,1%
Empleo Global	Quito	92,0%	0,01	0,84	92,0%	0,01	0,62	0,0%	0,0023	0,5009	0,9981	No rechazo Ho	No Significancia	MEJORÓ	-0,22	26,2%
Empleo Global	Guayaquil	93,9%	0,01	0,56	95,6%	0,00	0,43	1,7%	2,5942	0,9953	0,0095	Rechazo Ho	Significancia	MEJORÓ	-0,13	22,6%
Empleo Global	Cuenca	96,2%	0,01	0,61	95,1%	0,01	0,69	-1,1%	1,2452	0,8935	0,2131	No rechazo Ho	No Significancia	SUBIÓ	0,09	-14,3%
Empleo Global	Machala	95,3%	0,01	0,88	92,9%	0,01	0,86	-2,4%	2,0962	0,9820	0,0361	Rechazo Ho	Significancia	MEJORÓ	-0,03	3,0%
Empleo Global	Ambato	94,6%	0,01	0,75	95,3%	0,01	0,61	0,7%	0,8095	0,7909	0,4183	No rechazo Ho	No Significancia	MEJORÓ	-0,14	19,0%
Empleo Bruto	5 ciudades	60,4%	0,01	0,95	59,0%	0,00	0,80	-1,4%	1,8938	0,9709	0,0583	No rechazo Ho	No Significancia	MEJORÓ	-0,15	15,6%
Empleo Bruto	Quito	55,8%	0,01	1,91	55,5%	0,01	1,57	-0,4%	0,2632	0,6038	0,7924	No rechazo Ho	No Significancia	MEJORÓ	-0,35	18,2%
Empleo Bruto	Guayaquil	63,5%	0,01	1,34	61,8%	0,01	1,14	-1,8%				No	No	MEJORÓ	-0,20	15,0%

Empleo Bruto	Cuenca	61,1%	0,01	1,92	57,3%	0,01	1,74	-3,8%	1,6069	0,9460	0,1081	rechazo Ho	Significancia	MEJORÓ	-0,18	9,3%
Empleo Bruto	Machala	59,6%	0,02	2,73	58,2%	0,01	1,68	-1,4%	2,4420	0,9927	0,0146	rechazo Ho	Significancia	MEJORÓ	-1,05	38,5%
Empleo Bruto	Ambato	64,9%	0,01	2,23	62,3%	0,01	1,58	-2,6%	0,7579	0,7758	0,4485	rechazo Ho	Significancia	MEJORÓ	-0,65	29,3%
Sector informal	5 ciudades	32,5%	0,01	2,51	31,4%	0,01	2,43	-1,1%	1,4791	0,9304	0,1391	rechazo Ho	Significancia	MEJORÓ	-0,65	29,3%
Sector informal	Quito	18,7%	0,01	6,35	21,1%	0,01	5,69	2,4%	0,9730	0,8347	0,3305	No rechazo Ho	No Significancia	MEJORÓ	-0,08	3,1%
Sector informal	Guayaquil	44,3%	0,01	2,92	40,7%	0,01	3,01	-3,6%	1,4427	0,9254	0,1491	Rechazo Ho	Significancia	SUBIÓ	0,09	-2,9%
Sector informal	Cuenca	23,8%	0,02	8,39	20,5%	0,02	7,70	-3,3%	2,0334	0,9790	0,0420	No rechazo Ho	No Significancia	MEJORÓ	-0,69	8,2%
Sector informal	Machala	29,1%	0,02	8,29	32,7%	0,02	5,85	3,6%	1,2919	0,9018	0,1964	No rechazo Ho	No Significancia	MEJORÓ	-2,43	29,3%
Sector informal	Ambato	22,6%	0,02	8,11	21,2%	0,01	5,58	-1,4%	1,1716	0,8793	0,2414	No rechazo Ho	No Significancia	MEJORÓ	-2,52	31,1%

**CADA  
HECHO  
DE TU  
VIDA**  
*Cuenta*

 @ecuadorencifras

 INEC/Ecuador