



Marco de Muestreo para Encuestas Dirigidas a Hogares 2024





Marco de muestreo para encuestas dirigidas a hogares 2024

Créditos Institucionales

Eva María Mera Intriago
Directora Ejecutiva

Marianita Granda
Subdirectora General

Gustavo Molina
Coordinador General Técnico de Producción Estadística

Christian Garcés
Director de Infraestructura Estadística y Muestreo

Revisión
Christian Garcés

Elaboración
Javier Núñez
Rafael Encalada
Angel Gaibor

Citar como: INEC (2025). Marco de muestreo para encuestas dirigidas a hogares 2024.

Propiedad Intelectual: © INEC – Instituto Nacional de Estadística y Censos. Juan Larrea N15-36 y José Riofrío. Código Postal: 170402 / Quito – Ecuador. Teléfono: (593-2) 2232303 – 2232012 - 2232151



Resumen

El marco de muestreo para encuestas dirigidas a hogares constituye la base sobre la cual se estructuran y coordinan las diferentes operaciones estadísticas por muestreo en Ecuador. Su finalidad es garantizar la representatividad y precisión de las estimaciones obtenidas a partir de encuestas dirigidas a viviendas, hogares y personas. Este marco se desarrolla utilizando los resultados definitivos y el material cartográfico del Censo de Población y Vivienda 2022, asegurando una actualización y alineación con la realidad demográfica del país. La construcción del marco de muestreo sigue tres fases fundamentales:

1. Definición de las Unidades Primarias de Muestreo: Se realiza mediante un proceso de partición de grafos con restricción de tamaño, buscando generar conglomerados más uniformes en cantidad de viviendas ocupadas. Esta estrategia permite mejorar la eficiencia operativa y garantizar una distribución equitativa de la carga de trabajo en el levantamiento de encuestas.
2. Estratificación de las Unidades Primarias de Muestreo: Se basa en características sociales, demográficas y económicas de la población, con el objetivo de minimizar la varianza de los principales indicadores a evaluar en el periodo intercensal. La estratificación permite clasificarlas en grupos homogéneos, optimizando la calidad de las estimaciones estadísticas y reduciendo los errores muestrales.
3. Coordinación de muestras a lo largo del tiempo: Se implementa a través del cálculo del Número de Pareto, una metodología que garantiza la selección eficiente de las unidades primarias de muestreo para múltiples encuestas a lo largo del periodo intercensal. Esto permite mantener la estabilidad de las muestras y reducir la carga de respuesta sobre la población encuestada.

El desarrollo del marco de muestreo contó con la asistencia técnica de la Comisión Económica para América Latina y el Caribe, a través de su División de Estadísticas Sociales, lo que permitió incorporar metodologías avanzadas y mejores prácticas internacionales en la construcción, estratificación y administración del marco.

Además, se implementaron herramientas innovadoras, como el paquete psuR en R, que facilita la generación de conglomerados homogéneos y mejora la gestión del marco de muestreo. Estas mejoras refuerzan la capacidad del INEC para llevar a cabo encuestas de hogares con mayor precisión, representatividad y comparabilidad en el tiempo, alineándose con estándares internacionales en diseño muestral.



Introducción

En el Instituto Nacional de Estadística y Censos del Ecuador (INEC) se realizan diferentes tipos de encuestas. Las mismas se podrían enmarcar en tres grupos:

- encuestas dirigidas a hogares;
- encuestas de empresas; y
- encuestas agropecuarias.

Particularmente, las encuestas dirigidas a hogares siguen un diseño muestral complejo en dos etapas. La primera etapa es un muestreo por conglomerados estratificado. Consiste en seleccionar Unidades Primarias de Muestreo (UPM), las mismas que están clasificadas según condiciones sociodemográficas en estratos alto, medio y bajo. La segunda etapa consiste en seleccionar un número fijo de viviendas dentro de cada UPM.

Hasta el año 2017, las UPM eran los sectores censales del Censo 2010, los mismos fueron usados como referencia para la construcción de los sectores utilizados para la ejecución del Censo de Población y Vivienda 2022. Por tal motivo, estos responden a necesidades operativas, ya que fueron contruidos con la finalidad de estandarizar la carga de trabajo de los censistas y generar la cartografía de apoyo para su ubicación.

Su diseño consideró un número homogéneo de viviendas totales, dando la misma consideración tanto a las viviendas ocupadas, como a las viviendas desocupadas, viviendas en construcción, viviendas convertidas en negocios, viviendas destruidas y lotes vacíos. Además, en la medida de lo posible, la estructura de los sectores censales se mantuvo entre los censos para poder realizar comparaciones entre ellos.

Sin embargo, para la gran mayoría de las diferentes operaciones estadísticas basadas en encuestas a hogares, las únicas viviendas que se deberían seleccionar son las ocupadas. Esto ha causado que el número de viviendas ocupadas en cada sector censal no responda a una distribución uniforme, es decir, hay sectores con un número pequeño de viviendas ocupadas y otros sectores que tienen un gran número de viviendas ocupadas.

El usar los sectores censales como UPM generaba inconvenientes de tipo operativo, ya que, al ser un número grande de encuestas realizadas por el INEC, la administración del marco muestral, que era el listado de sectores censales, tiene complicaciones. Esto ocurre porque se deben utilizar las mismas UPM para más de una encuesta, es decir, existe la posibilidad de que una vivienda sea elegida para diferentes investigaciones, lo cual genera malestar dentro de la población y muchas veces provoca resistencia por parte de los encuestados a dar información.

El no tener una distribución uniforme de viviendas con hogares ocupados implica que las UPM no tengan la misma probabilidad de inclusión. Al momento de inferir los resultados, cada elemento seleccionado representa a una parte de la población, esto implica que algunos de los elementos que son parte de la muestra



representen a pocos elementos de la población total y otros elementos representen a una parte más grande de la población.

Uno de los principales propósitos al construir el marco de muestreo 2022 es generar una distribución de tamaño de las UPM más uniforme, en relación con el conteo de viviendas con hogares particulares. Se presenta una herramienta que permite generar UPM más homogéneos considerando el número de viviendas ocupadas, con la restricción de que las UPM mantengan el criterio de construcción de los sectores censales que son conjuntos de manzanas o sectores que están conectados. Posteriormente, se presenta la forma en que las UPM son estratificadas. Finalmente, se dan las líneas para la correcta administración del marco muestral resultante de la generación de UPM y su respectiva estratificación.

Desde INEC, expresamos nuestro más sincero agradecimiento a la Comisión Económica para América Latina y el Caribe y a su equipo, en especial a Andrés Gutiérrez, Asesor Regional en Estadísticas Sociales, y José Zea consultor de la Unidad de Estadísticas Sociales, por guiar y acompañar en cada decisión crítica a lo largo de este proyecto. Su dedicación y arduo trabajo han sido cruciales para lograr este hito.



Revisión de literatura

El estado del arte estará estructurado en tres partes. En primer lugar, se dará una visión sugerida por Naciones Unidas de la forma en la que se debe estructurar un marco de muestreo, introduciendo las principales características que este debe tener, además se describen las limitaciones que se presentan al momento de construir los marcos de muestreo. Luego, se describirá la forma en que algunos países de América construyen sus marcos muestrales para encuestas dirigidas a hogares.

La segunda parte estará compuesta por una visión histórica de la teoría de grafos y del problema de partición. El problema de partición de grafos se divide en dos grupos dependiendo del tipo de restricciones asociadas al problema, y se explica cuál de ellos describe el problema vinculado con la construcción de conglomerados.

Finalmente, en la tercera parte se presentará una descripción general del proceso de estratificación propuesta por la Comisión Económica para América Latina y el Caribe (CEPAL), dando una visión general de los métodos utilizados al momento de realizar la estratificación.

Marco de muestreo

La Organización de las Naciones Unidas (ONU) se encarga de generar vínculos entre las naciones tomando en cuenta las políticas globales de las naciones en los ámbitos económico, social y ambiental. Por tal motivo, la ONU se involucra directamente en mejorar la calidad de los datos proporcionados por los países implantando normativas en la forma de medir los principales indicadores que permitan analizar la situación de cada país. Para normar ciertas directrices estadísticas la ONU desarrolló (UNP, 2005), el cual es un manual que expone guías prácticas para el diseño de encuestas dirigidas a hogares. Uno de los puntos principales que se mencionan son las características y guías para la construcción de marcos muestrales, los cuales se presentan a continuación.

Desde un punto de vista operativo el marco de muestreo es el conjunto de elementos del cual la muestra es seleccionada. Un marco debe proveer un medio para escoger los miembros particulares de la población objetivo a ser investigados en una encuesta. Generalmente, más de un conjunto de elementos es necesario, esto es por la naturaleza de las encuestas dirigidas a hogares. El diseño muestral de las encuestas dirigidas a hogares está compuesto por diferentes etapas, por lo que es necesario un marco para cada una de estas etapas.

Una forma de clasificar a los marcos de muestreo puede ser en marcos de área y de lista.

Los marcos de área están compuestos por unidades geográficas de un país en una disposición de orden geográfico. La forma en que se logra esto es conside-

rando las áreas geográficas que constituyen la división político-administrativa de un país. En el caso de Ecuador, está constituida por provincias, cantones, parroquias, área amanzanada y dispersa, zonas censales y sectores censales. Estas áreas deben cumplir que cubran el área del país en su totalidad, las áreas geográficas deben estar bien delimitadas, debe haber población en cada área y las mismas deben estar mapeadas.

En los marcos de muestreo dirigidos a hogares los marcos de área están conformados por los sectores censales, que son áreas geográficas conformadas por conjuntos de manzanas.

Los marcos de lista son listas de los elementos que conforman la población objetivo, en el caso de las encuestas dirigidas a hogares son listas de los hogares y deben poder ser asociados el área geográfica a la que pertenecen. Estas listas generalmente se obtienen de los censos.

De la misma forma los marcos de lista son los listados de viviendas ocupadas que conforman cada sector censal.

Así tenemos que, dada la naturaleza de las encuestas dirigidas a hogares, se tienen dos etapas, para la primera se utiliza un marco de área que es el listado de sectores censales, o los conglomerados que se van a construir a lo largo de este trabajo, para la segunda etapa es necesario un marco de lista que estará conformado por las viviendas ocupadas de cada conglomerado. Las principales características que se debe tener en cuenta para considerar a un marco perfecto es que sea completo, exacto y actualizado.

La propiedad de que un marco sea **completo** está determinada por la cobertura de la población objetivo, idealmente se espera que a partir del marco se pueda llegar a cada miembro de la población objetivo. Principalmente en las encuestas a hogares se tiene desafíos en la cobertura de las personas, puesto que en algunas ocasiones no se puede acceder a las personas que viven en instituciones, como asilos, orfanatos, hospitales, por otro lado, existen personas que no tienen un hogar fijo, esto a veces se genera por empleos que obligan a la gente a cambiar constantemente de residencia.

La **exactitud** de un marco está asociada con la duplicidad y omisión de elementos dentro del mismo. Un marco al ser un archivo que contiene el listado de viviendas puede contener omisiones, dichas listas pueden contener más de una vez al mismo hogar identificado de formas distintas o pertenecer a dos listas diferentes, además que estas listas pueden ser antiguas y omitir a las nuevas construcciones de viviendas. Esto implicaría que se generen inconsistencias en la probabilidad asignada a cada elemento de la población objetivo.

Como se ha ido describiendo anteriormente, los marcos son producidos a partir de la información obtenida en los censos, por lo que dicha información se actualiza antes de realizar el levantamiento de la información censal. Tener un marco **actualizado** es en un gasto realmente alto, por lo que una alternativa para mantener la información actualizada es realizar procesos de levantamiento de información en los conglomerados que van a ser parte de una investigación previo a la selección y entrevista de las viviendas en campo.



En los institutos de estadística, uno de los principales insumos para realizar las diferentes investigaciones es la construcción de los marcos de muestreo. Generalmente este insumo se obtiene a partir de la información levantada tras los censos realizados en cada uno de los países.

En el año 2002 se realizó el **9° Taller regional de diseño y construcción de marcos de muestreo para las encuestas de hogares** (CEPAL, 2002). El taller fue realizado por la CEPAL en coordinación con el Instituto Nacional de Estadística e Informática (INEI) de Perú, principalmente se analizaron los aspectos técnicos y operativos que intervienen en el diseño y construcción de los marcos de muestreo. A lo largo del taller se expusieron la forma en que los países de América Latina y el Caribe utilizan la información cartográfica, obtenida a partir de los censos, en la construcción de los marcos de muestreo. A continuación, se detallan los principales puntos expuestos por algunos países en el mencionado taller.

En el caso de **Bolivia** se muestra la construcción de una Muestra Maestra para las encuestas de hogares que el Instituto Nacional de Estadísticas de Bolivia (INE) realiza. Empezando por la definición de su tamaño, luego se define un tipo de muestreo, después se escoge una forma de selección y por último se establece una metodología para estimar los parámetros y sus varianzas, se selecciona una muestra y de esta se obtienen submuestras para usar en cada una de las encuestas que se tiene previsto llevar a cabo durante un periodo de tiempo establecido. En (Jimenez, 2011) se da una visión de las variables consideradas para el marco de muestreo utilizado por el INE en Bolivia, además de una visión general de las principales consideraciones para las encuestas dirigidas a hogares realizadas por el instituto.

Por su parte el equipo de **Chile** explica las ventajas del Marco de Muestreo de áreas. Las muestras que se seleccionan por áreas son más económicas y es posible actualizar la muestra para una encuesta continua sin que se incumpla la regla de la selección probabilística. Este marco se divide en 25.826 áreas con 200 viviendas urbanas y 100 viviendas rurales. El detalle de la construcción del marco muestral utilizado por el INE de Chile está descrito en (OS, 2015), aquí se da una visión detallada de cómo se estructura el marco y como ha sido posible estructurar un Sistema Integrado de Encuestas Dirigidas a Hogares.

En el caso de **Colombia**, no se trata de un marco de muestreo sino de la ampliación realizada a la muestra maestra implementada en el año 1993 para la Encuesta Nacional de Hogares. La ampliación se realizó en el año 2001 tomando en cuenta las siguientes observaciones:: definición integral de la composición y tamaño de la muestra que iba a ser usada durante los siguientes años, ampliación del número de unidades primarias de muestreo y de los estratos para así completar 12 submuestras nacionales replicadas y que se hiciera más fácil obtener indicadores por departamento, finalmente, la ampliación mantuvo gran parte de las unidades primarias de muestreo que se usaban hasta ese momento.

Los técnicos de **El Salvador** exponen que el marco muestral se basa en los datos del Censo de Población y Vivienda de 1992. El país es dividido en 5 regiones para mejorar la cobertura. Se usa la cartografía que proporciona la Dirección General de Estadísticas y Censos, así se llega a un marco final de 1'073.172 viviendas. Luego, se actualiza en 1995 y 1997, la última incluyendo visitas directas en el terreno.

Una visión más actual de la construcción de los marcos muestrales en El Salvador se presenta en (DIGESTYC, 2015).

Para la elaboración de los marcos de muestreo en **Guatemala** se usa la información cartográfica de los censos, actualizándola en los años 1992 y 1993 en miras a la realización de los Censos Nacionales de Población y Habitación de 1994. Un sector censal urbano es un área geográfica que tiene aproximadamente 200 viviendas y un sector censal rural tiene 150 viviendas aproximadamente. Con este marco, se diseñó la muestra para varias encuestas. Luego se habla de que para el IX censo de Población, VI de habitación y IV agropecuario que se realizó en 2002 se estaba elaborando otra cartografía. Para esto, un grupo de cartógrafos actualizaron la estructura de los sectores que se seleccionaron.

En el caso de **Honduras** se presenta como se elabora el Marco de Muestreo para la Encuestas Permanente de Hogares de Propósito Múltiples. Este se construyó con los datos del Censo de Población y Vivienda, sin embargo, se encontraron algunos problemas. El primero, que se creó un código 3 que se relacionaba con lo urbano/rural, y el segundo, los listados de la actualización cartográfica no constaba el número de personas en cada vivienda, por último, no se identifica la condición de la vivienda, es decir si está ocupada o no. Con esto se construye un Marco de Muestreo provisional, que será actualizado cuando se tenga la información disponible. Actualmente en las encuestas dirigidas a hogares en Honduras se utiliza muestreo por conglomerados, utilizando como tales a los segmentos censales resultado del censo de población y vivienda, tal y como se expone en (INE, 2015).

Por su parte, **Argentina** expone el Marco de Muestreo Nacional Urbano. En este la muestra de áreas que forman parte del marco se obtienen en dos etapas. En la primera, se muestrean las unidades primarias probabilísticamente, estas están definidas como aglomeraciones. En la segunda, se muestrean radios censales como unidades secundarias de muestreo en cada unidad primaria seleccionada. Mediante este proceso se ahorra recursos pues el espacio territorial se restringe. Sin embargo, es necesario extender la cobertura total a nivel nacional, pues no está disponible para encuestas que tienen como población objetivo el total de los habitantes del país. Esto ocasiona que no se pueda comparar los resultados de determinadas encuestas con otros países del MERCOSUR.

En cuanto a encuestas en las provincias, también es necesario una extensión hacia las zonas rurales para que esta parte de la población esté representada. La extensión del marco propuesta tendría una identificación total de las viviendas en una muestra probabilística de áreas. Este muestreo de áreas se realizaría en dos etapas, las unidades primarias se configuran por departamentos o por conjuntos de estos y las unidades secundarias se configuran mediante radios censales. El universo de referencia del marco nuevo es la totalidad de las viviendas del país. Para este marco se eliminan zonas a las que no se puede acceder y las que tienen un alto costo de relevamiento pues tienen baja densidad poblacional. El detalle y las consideraciones actuales para la construcción de los marcos de muestreo que utiliza el instituto de estadística en Argentina se exponen en (INDEC, 2008), para conseguir conglomerados más homogéneos en el número de viviendas se propone la idea de juntar segmentos censales contiguos.

En **Estados Unidos**, según el Bureau of Labor Statistics de Julio de 2015 (Nix, 2015),

una nueva muestra de las áreas geográficas y de las direcciones seleccionadas es actualizada cada década. La población objetivo es el grupo de todas las personas civiles. Las unidades de observación son personas en una casa que se relacionan por sangre, adopción, matrimonio u otro tipo de arreglos legales. También forman parte de estas unidades aquellas personas que no están relacionadas, pero que combinan sus ingresos y hacen gastos conjuntos. Normalmente estas unidades son iguales a las viviendas en el 98 % del tiempo. Luego, los marcos muestrales son la lista de las casas de las cuales se obtiene la muestra. Este se divide en:

- Marco Unidad: viviendas regulares, el 80 %
- Marco Área: viviendas rurales, el 10 %
- Marco Permiso: nuevas construcciones , el 9 %
- Cuartos de grupo: el 1 %

Se actualiza la información del Censo 2010 dos veces por año, mediante el Servicio Postal de Estados Unidos. Para la selección de la muestra, las áreas geográficas se seleccionan aleatoriamente para representar el total de Estados Unidos. Las viviendas también se seleccionan aleatoriamente para representar las áreas geográficas.

Conglomeración

Definición de Unidad Primaria de Muestreo

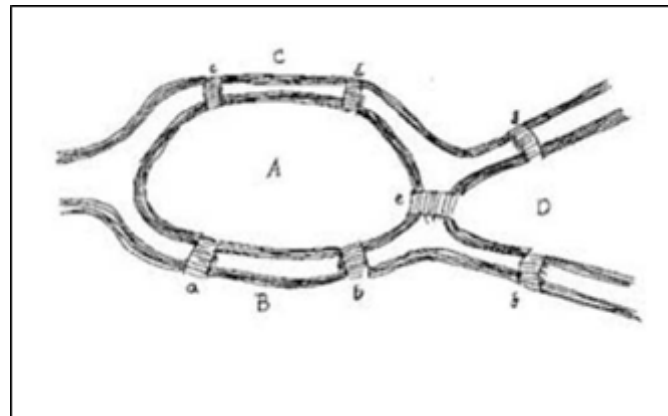
Según (Valliant et al., 2013), los conglomerados son áreas geográficas, establecimientos, escuelas, o cualquier otro tipo de agregación de unidades utilizadas en diseños muestrales multietápicas, pues son particularmente prácticos y costo-eficientes en comparación al muestreo aleatorio simple.

Generalmente, las encuestas dirigidas a hogares suelen considerar al menos dos etapas de selección, seleccionando conglomerados en la primera etapa, las cuales pasan a denominarse Unidades Primarias de Muestreo (UPM). La selección de viviendas (unidades secundarias de muestreo) dentro de cada UPM varía dependiendo de la metodología de la encuesta, pudiendo ser de enumeración completa, aleatoria, diferenciada, entre otras.

Partición de grafos

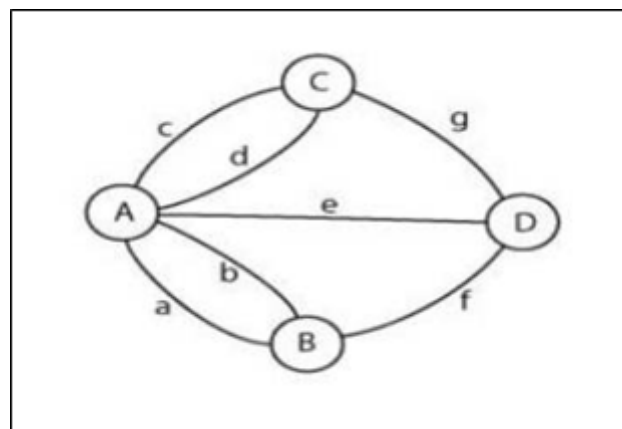
La teoría de grafos es una rama de la matemática discreta introducida en el año 1736 con el problema de los puentes de Königsberg (Euler, 1736), este consistía en encontrar el camino que transitara todos los siete puentes del río Königsberg sin pasar más de una vez por cada uno de ellos. El matemático Leonhard Euler trabajó sobre este problema y sus resultados son considerados los primeros sobre la teoría de grafos. En la Figura 1 se presenta el problema de los puentes de Königsberg en letras mayúsculas se representan los puntos que están unidos por los puentes, los mismos que están representados con letras minúsculas.

Figura 1: Puentes de Königsberg.



En la Figura 2 se representa en forma de grafo al problema con letras A, B y C se identifican a las ciudades y a, b, c, d, e, f y g representan los puentes.

Figura 2: Grafo asociado a los puentes de Königsberg .



En el año 1752 se publicó el teorema de Euler para grafos planos. Al principio estos resultados se presentaron en términos de poliedros. Después de estos avances hubo un tiempo en que no se presentaron más trabajos sobre esta área. Esto cambio en el año 1847 cuando Gustav Kirchhoff estudió un grafo especial al cual se lo llamó árbol. Este tipo de grafo es no dirigido y no tiene lazos, también es sin ciclos y conexo. Kirchhoff usó estos estudios en algunas aplicaciones de eléctrica, cuando formuló la extensión de las leyes de Ohm. Luego de diez años Arthur Cayley al contar los isómeros de hidrocarburos saturados desarrollo esta misma clase de grafos de árbol (Grimaldi, 1998).

Años más tarde, en 1852, Francis Guthrie planteó el problema de los cuatro colores (Guthrie, 1880), el mismo consiste en que dado un mapa cualquiera dividido en países, este se puede colorear únicamente con cuatro colores sin que se repita el color para dos países vecinos. El problema fue resuelto después de un siglo por los matemáticos Kenneth Appel y Wolfgang Haken (Appel y Haken, 1976). Este puede ser considerado el nacimiento de la teoría de grafos puesto que para resolverlo se definieron términos y conceptos fundamentales en esta área de la matemática.

En 1859 aparece otra idea importante en esta área, el ciclo hamiltoniano, su nombre es en honor a William Rowan Hamilton, él en 1859 formuló un acertijo usando las aristas de un dodecaedro regular el cual no es muy difícil de resolver. Sin embargo, actualmente se siguen buscando condiciones suficientes y necesarias para definir los grafos que son no dirigidos y que tienen un camino hamiltoniano. En 1930 se caracterizaron los grafos planos, esto fue realizado por el matemático Kasimir Kuratowski. En el año 1936 se publica el primer libro de teoría de grafos, este fue escrito por el matemático Dénes König (Grimaldi, 1998).

El desarrollo de este trabajo está enfocado en el particionamiento de grafos que es estudiado desde 1970. El problema de partición de grafos consiste en dividir el conjunto de nodos en subconjuntos con restricciones en el número de nodos y el peso asociado a cada partición, buscando minimizar el costo del valor de las aristas. La aplicación de este tipo de problemas aparece en varios campos como la informática, ingeniería, deportes, etc. por ejemplo en el área de los circuitos VLSI ha habido avances por parte de Kahng (Mitchell, 2001), problemas derivados del tráfico aéreo han sido estudiados por Bichot (Bichot, 2007), el desarrollo de las redes móviles ha sido estudiado en (Fairbrother et al., 2017), en el ámbito deportivo la programación de los equipos puede ser asociado con un problema de particionamiento de grafos (Mitchell, 2003).

Dentro de la partición de grafos, existen dos grupos de problemas tomando en cuenta el tipo de restricciones (Bichot y Siarry, 2013). Al primer grupo se lo conoce como particionamiento restringido, en el que el número de elementos de cada partición está fijado; el segundo grupo de problemas se denomina partición sin restricciones, en este el número de elementos que formará cada partición no está determinado previamente.

El problema con el que se asocia la construcción de conglomerados está en el segundo grupo, ya que no existe ninguna restricción en el número de manzanas que formará cada conglomerado, las restricciones están asociadas al número de viviendas que forma cada conglomerado, es decir restricciones de peso.

Estratificación

El proceso de estratificación de las UPM se realiza en función a lo expuesto en (CEPAL, 2023), en el cual se muestra la forma en la que se aborda esta temática en las oficinas nacionales de estadística de la región. La estratificación, al garantizar la homogeneidad dentro de cada estrato, disminuye la incertidumbre de las estimaciones y minimiza los errores de muestreo que se obtienen al realizar encuestas con procedimientos de muestreo probabilístico.

Este proceso se lleva a cabo utilizando información censal a nivel de personas, hogares y viviendas, en diferentes constructos o dimensiones relacionadas con la calidad de vida y el bienestar (demografía, características de la vivienda, propiedad de determinados bienes y acceso a servicios públicos, entre otras). Las variables definidas sobre estos constructos se agregan a partir de variables binarias que toman el valor 1 si el fenómeno en cuestión se asocia de forma positiva con mejores condiciones socioeconómicas y 0 en caso contrario.

La estratificación del marco de muestreo se basa en un análisis exploratorio de los datos censales para cada una de las UPM construidas. Esto significa que, cuando las UPM estén categorizadas en un estrato, todas las personas, hogares y viviendas de la misma estarán clasificadas en el mismo estrato. Preferiblemente, las variables a considerar deben estar directamente relacionadas con los fenómenos que se estudiarán en las diferentes encuestas de hogares a lo largo del período intercensal, y suelen agruparse en las dimensiones:

- Demografía y estructura de la población: sexo, edad, parentesco, origen extranjero, pertenencia a pueblos indígenas, número de hijos y número de dependientes, entre otros.
- Educación: analfabetismo, asistencia escolar, años de estudio y nivel de escolaridad, entre otros.
- Mercado de trabajo: población en edad de trabajar, participación en la fuerza de trabajo por sexo, condición de ocupación por sexo y rama de actividad, entre otros.
- Características de la vivienda: tipo de vivienda, materiales de construcción, situación de hacinamiento y equipamiento, entre otros.
- Acceso a servicios: agua potable, alcantarillado, Internet, salud y seguridad social, entre otros.
- Propiedad de determinados bienes en el hogar: en algunos países, la propiedad de bienes como televisor, microondas, aire acondicionado, automóvil o lavaplatos automático, entre otros, puede utilizarse como criterio de clasificación de los hogares.
- Necesidades básicas insatisfechas (NBI) o pobreza multidimensional: situación de hacinamiento crítico en la vivienda, servicios inadecuados, alto nivel de dependencia económica, presencia de niños en edad escolar que no asisten a la escuela y falta de acceso o acceso limitado a servicios de salud, entre otras.

Terminada la selección de variables a utilizar en la estratificación, se calculan sus agregados a nivel de UPM, sin embargo, debido a que las UPM suelen tener tamaños diferentes, la escala y el nivel en que se miden los indicadores pueden afectar los procesos de clasificación. En consecuencia, la definición de la matriz de información en términos relativos (porcentaje de aparición de un fenómeno sobre cada categoría de las variables) constituye una alternativa mejor para que el agrupamiento esté controlado por el tamaño de la UPM y supeditado únicamente a cambios estructurales asociados a la observación de las variables en cuestión. Sin embargo, debido a la forma en como se construirán las UPM, el efecto asociado a la diferencia en el tamaño de las UPM debería ser mínimo.

Con respecto a la matriz de información relativa, es importante verificar que todos los indicadores de esta matriz apunten hacia el mismo horizonte, es decir, que todos los indicadores estén expresados en términos de acceso al bienestar definido. Además, es necesario refinar esta matriz para eliminar las variables que puedan estar muy correlacionadas con el resto de las variables o que puedan expresarse como una combinación lineal de otras variables. De esta manera, se evitan los problemas de multicolinealidad y se asegura una estratificación parsimoniosa.

Finalizado este proceso, se debe contar con una matriz de información X compuesta por P columnas (variables de estratificación) y N_I filas (número de UPM en

el marco de muestreo), en la que cada fila representará la i -ésima observación de las UPM a nivel censal para cada una de las P variables.

De acuerdo con la teoría estadística, la mejor estratificación es aquella que minimiza los errores de muestreo de los estimadores, expresados en forma de varianzas o errores estándar. Además, una de las particularidades de los procesos de estratificación es que las varianzas de estos estimadores dependen, a su vez, de la variación de los microdatos observada en el censo. Sin embargo, una estratificación óptima para un indicador puede resultar ineficiente para otros. Más aún, dado que no todas las variables de interés que se examinarán en las encuestas durante el período intercensal se han medido u observado en el censo, la estratificación que se ha de utilizar debe elegirse cuidadosamente, por medio del estudio de numerosos escenarios.

En general, hay dos grandes tipos de métodos que deben evaluarse al proponer una estratificación: univariados (sobre una medida de resumen de la matriz de información) y multivariados (sobre todas las variables de la matriz de información). En ambos casos, el objetivo es encontrar la mejor división, que garantice que la varianza de los estimadores de muestreo sea mínima. En el siguiente capítulo se incluye una descripción sucinta de cada método utilizado, para más información acerca de las mismo dirigirse a (CEPAL, 2023), (Baillargeon et al., 2007) y (Barcaroli, 2014).

Implementación metodológica

En esta sección se presenta la estructura territorial del país a través del **Clasificador Geográfico Estadístico** la cual delimita el campo de acción de la **Conglomeración**, cuyo proceso se describe posteriormente. Terminada la construcción de los conglomerados se procede a la **Estratificación** de los mismos por cada dominio geográfico. Por último, se introduce la **Coordinación de Muestras** a través de la cual se administra las UPM disponibles en el Marco de Muestreo en función a las operaciones estadísticas requeridas.

Clasificador Geográfico Estadístico

Según (INEC, 2022), el Clasificador Geográfico Estadístico describe la estructura territorial del país dividiéndolo en provincias, cantones y parroquias, proporcionando los códigos numéricos y nombres de cada unidad territorial. Según la Ley de Fijación de Límites Territoriales Internos, el Comité Nacional de Límites Internos (CONALI) es el encargado de mantener actualizada la información de la División Político Administrativa, la cual adopta el INEC para todas sus operaciones estadísticas, contando con las siguientes características:

- Se asignan los dos primeros dígitos de la izquierda al código de provincia.
- El código 01 es asignado al cantón cuya cabecera es también capital provincial.
- Dentro de cada cantón se codifican las parroquias de la siguiente manera: se asigna el código 50 a la cabecera cantonal, luego, se ordenan alfabéticamente el resto de parroquias, asignándoles un código a partir del 51 en adelante (límite 99).

Las provincias quedan codificadas en orden alfabético excepto Galápagos, Sucumbíos, Orellana, Santo Domingo de los Tsáchilas y Santa Elena, creadas en fechas posteriores al establecimiento de este criterio de codificación. A estas provincias se les asigna el código inmediato superior al que tiene la última provincia, de acuerdo a la fecha de su respectiva creación.

Tabla 1: Número de cantones y parroquias por provincia según el censo 2022.

Provincia	Código provincia	Nro. cantones	Nro. parroquias
Azuay	01	15	76
Bolívar	02	7	26
Cañar	03	7	33
Carchi	04	6	32
Cotopaxi	05	7	40
Chimborazo	06	10	55



El Oro	07	14	64
Esmeraldas	08	7	64
Guayas	09	25	54
Imbabura	10	6	42
Loja	11	16	94
Los Ríos	12	13	30
Manabí	13	22	77
Morona Santiago	14	12	58
Napo	15	5	25
Pastaza	16	4	21
Pichincha	17	8	61
Tungurahua	18	9	53
Zamora Chinchipe	19	9	38
Galápagos	20	3	8
Sucumbíos	21	7	36
Orellana	22	4	32
Santo Domingo	23	2	12
Santa Elena	24	3	11
Total Nacional	-	221	1.042

Fuente: Gestión de Diseño Muestral - DINEM.

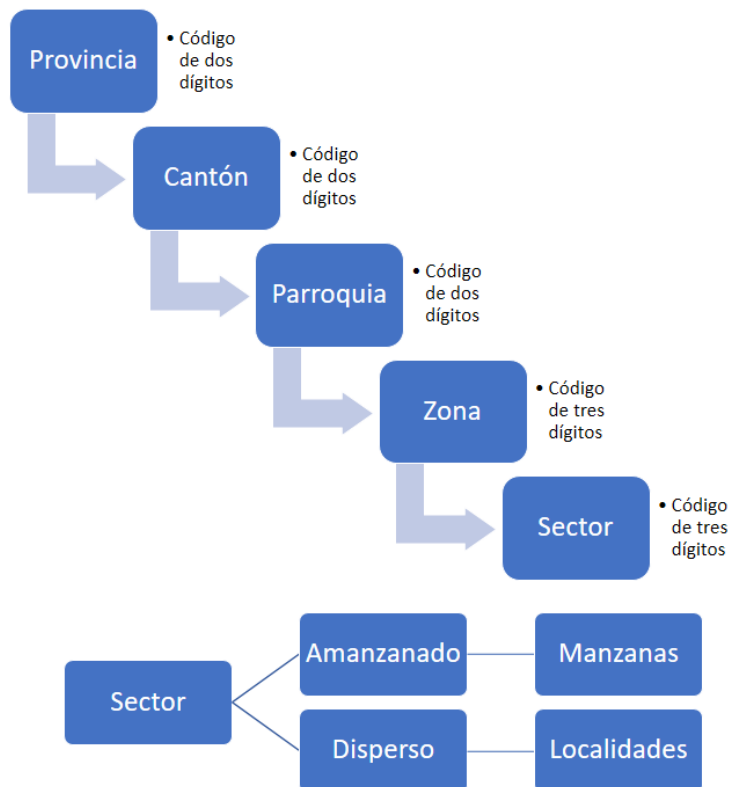
Esta codificación está integrada dentro del Marco Geoestadístico Nacional, sistema que utiliza el INEC para integrar la información estadística con la espacial en los diferentes proyectos que desarrolla, cuyos principales elementos son:

- **Zona censal amanzanada:** son superficies perfectamente delimitadas, constituidas por un promedio de 10 sectores censales amanzanados.
- **Sector censal:** es una división realizada para investigaciones estadísticas. Estas superficies son perfectamente delimitadas y geográficamente continuas, y pueden ser amanzanadas (constituidas por una o más manzanas y con un promedio de 150 viviendas) o dispersas (contienen un promedio de 80 viviendas).
- **Manzana censal:** es el elemento más característico de un plano y es el resultado de la estructura vial o hídrica que comunica las viviendas asentadas en un centro poblado.
- **Edificio censal:** es toda construcción o estructura que puede estar constituida por una o varias viviendas, establecimientos económicos, instituciones públicas o privadas, que ocupa un espacio determinado.
- **Vivienda censal:** es un recinto de alojamiento estructuralmente separado y con acceso independiente construido, edificado, transformado o dispuesto para ser habitado por personas.

En la Figura 3 se muestra la relación existente entre los elementos del Clasificador Geográfico Estadístico del INEC utilizados para la construcción del Marco de

Muestreo de encuestas dirigidas a hogares, donde se destaca la relación de contención jerárquica existente entre ellos, junto a la longitud de los códigos utilizados para su identificación única.

Figura 3: Elementos del Clasificador Geográfico Estadístico.



Fuente: Gestión de Diseño Muestral - DINEM.

Conglomeración

Formulación del problema

Las encuestas que realiza el INEC a nivel de hogares generalmente siguen un muestreo por conglomerados. Históricamente, los sectores censales han sido utilizados como conglomerados, sin embargo, el número de viviendas ocupadas por sector no sigue una distribución uniforme. Para la construcción del marco se consideró como mínimo 60 viviendas ocupadas al momento de determinar los conglomerados. En la siguiente imagen se presenta un resumen de la distribución de los sectores censales en función del número de viviendas ocupadas.

Figura 4: Distribución sectores censales por viviendas ocupadas.



Fuente: Gestión de Diseño Muestral - DINEM.

El número de investigaciones realizadas por el INEC es grande, por lo que una correcta administración de los conglomerados es indispensable, pues si se usan los mismos conglomerados para más de una investigación, se podría generar resistencia por parte de los informantes al momento de proporcionar información. Al contar con conglomerados uniformes, la administración del marco resulta más sencilla. Al problema de construcción de conglomerados se lo relacionará con un problema de partición de grafos.

Sea $G = (V, E)$ un grafo no dirigido conexo, donde $V = \{1, 2, \dots, n\}$ es el conjunto de nodos y $E = \{(i, j) : i, j \in V, i < j\}$ es el conjunto de aristas. Sea e_{ij} la variable que toma el valor de 1 si la arista $(i, j) \in E$ 0 en caso contrario.

Sea $W = (w_1, w_2, \dots, w_n) \in \mathbb{N}^n$ el vector de pesos asociado a los nodos del grafo, sea $c \in \mathbb{N}$, el peso mínimo requerido para formar un conglomerado. Sea x_{ij} la variable que toma el valor de 1 si los nodos i y j pertenecen a la misma parte, 0 en caso contrario.

La forma en la que se conectan los dos problemas será considerando a una manzana o sector censal disperso como nodo y las aristas estarán determinadas por la existencia de una frontera común entre cada par de manzanas o sectores censales dispersos. El peso de cada nodo está determinado por el número de viviendas ocupadas que existen en cada manzana o sector censal disperso.

Por lo tanto, la formulación del problema de partición de grafos con restricciones de tamaño es:



$$z = \text{mín} \sum_{i=1}^n \sum_{j=1}^n x_{ij} \quad (1)$$

Sujeto a las siguientes restricciones:

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

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

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

$$x_{ij} = x_{ji}, \quad \forall i, j \in V \quad (5)$$

$$\sum_{i=1}^n w_j x_i \geq c \quad \forall i \in V \quad (6)$$

Si $x_{ij} = 1$, entonces $\exists i = i_1, i_2, \dots, i_{m-1}, i_m = j \in V$ y

$$m \geq 2 : x_{i_1 i_2} e_{i_1 i_2} = \dots = x_{i_{m-1} i_m} e_{i_{m-1} i_m} = 1 \quad \forall i, j \in V, i \neq j \quad (7)$$

$$x_{ii} = 1 \quad \forall i \in V \quad (8)$$

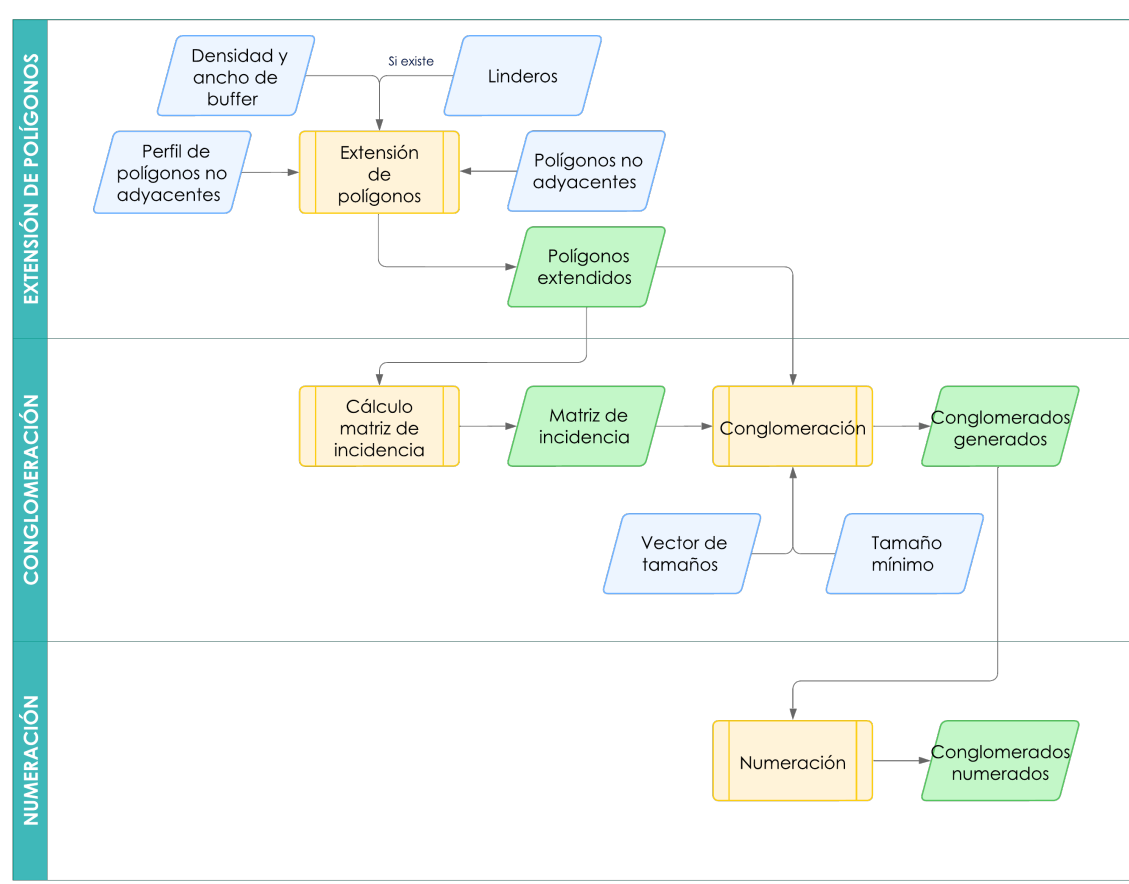
La función objetivo (1) busca minimizar el número de nodos en cada partición y de esta forma maximizar el número de conglomerados resultantes. Las restricciones (2)-(4) también denominadas desigualdades triangulares, se introducen con la finalidad de garantizar que si $x_{ij} = x_{jk} = 1$ entonces $x_{ik} = 1$. La restricción (5) es para garantizar simetría en la variable x_{ij} . La restricción de tamaño es (6), con esto se garantiza que cada conglomerado formado sume el mínimo peso requerido para formar uno. Con la restricción (7) se garantiza que, si dos nodos pertenecen a un conglomerado, entonces existe un camino de nodos en el conglomerado que los conecta. La restricción (8) es necesaria para que al momento de determinar el tamaño del conglomerado al que pertenece el nodo i se incluya el peso de dicho nodo.

Paquete psuR

La Dirección de Infraestructura Estadística y Muestreo del INEC desarrolló un paquete para el software estadístico R que permite realizar todo el proceso de conglomeración, que incluye una función asociada con una heurística que resuelve el problema de partición de grafos descrito en la sección anterior.

En el siguiente diagrama de flujo se muestra el proceso, los rectángulos amarillos representan las funciones que conforman el paquete psuR, los romboides de color celeste son los insumos necesarios para la ejecución de cada función y los romboides de color verde son los resultados de la ejecución de cada función.

Figura 5: Flujograma de psuR.



Fuente: Gestión de Diseño Muestral - DINEM.

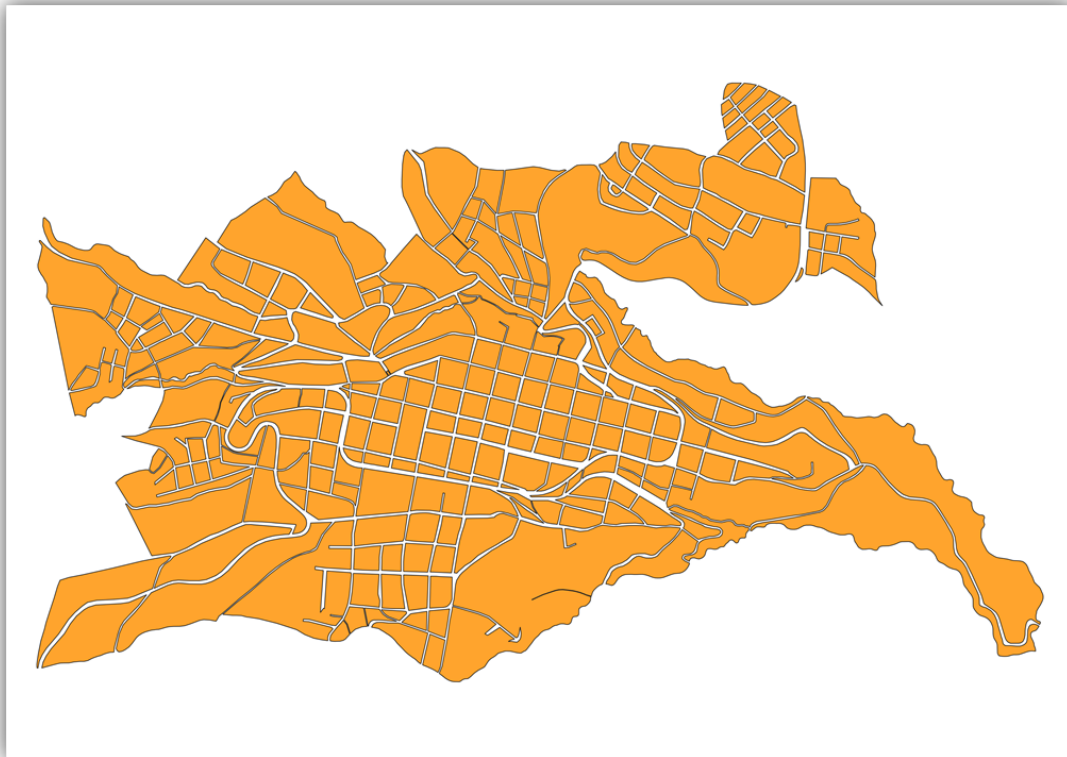
El paquete **psuR** se puede encontrar en el siguiente repositorio de GitHub <https://github.com/matjaviernunez/psuR>, el mismo fue presentado en el **Seminario Regional sobre Innovaciones y Desafíos en Encuestas de Hogares** organizado por CEPAL en el mes de noviembre de 2024.

El proceso de conglomeración se realiza de manera independiente en cada una de las parroquias que conforman el territorio ecuatoriano separando la parte amanzanada de la dispersa. El objeto geográfico que representa a los sectores dispersos está formado por polígonos adyacentes por lo que determinar la existencia de una frontera común entre un par de sectores dispersos es directa; por otro lado las manzanas no son un conjunto de polígonos adyacentes, la función *Extensión de polígonos* permite convertir el objeto geográfico de manzanas en un conjunto de polígonos adyacentes.

Extensión de polígonos

Para la ejecución de la función *Extensión de polígonos* se requiere un conjunto de polígonos no adyacentes. En la siguiente figura, se presenta un ejemplo para la parroquia San Miguel de la provincia Bolívar.

Figura 6: Manzanas de la parroquia San Miguel de Bolívar.



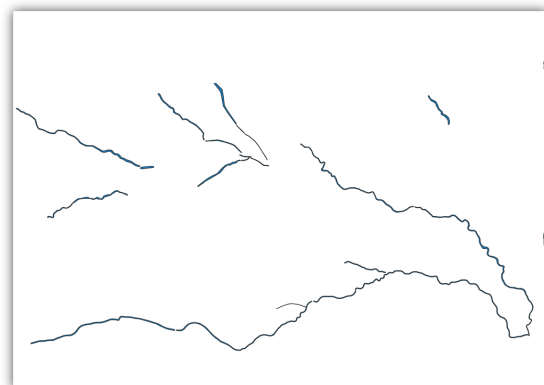
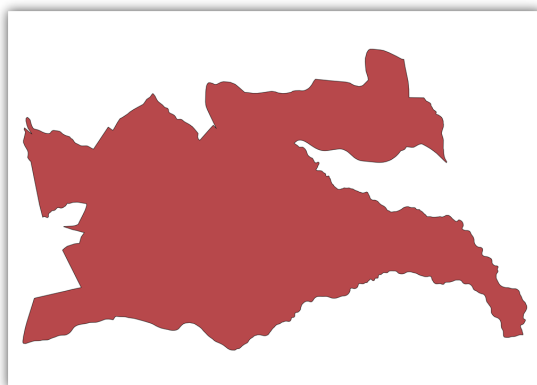
Fuente: Gestión de Diseño Muestral - DINEM.

Además, se necesita como atributo el polígono que delimita el perfil de la parte amanzanada de la parroquia que se está extendiendo. En el caso de contar con la información se puede incluir un atributo de linderos, para generar un corte en la extensión de los polígonos.

Figura 7: Insumos de la función de extensión de manzanas.

(a) Perfil de la ciudad.

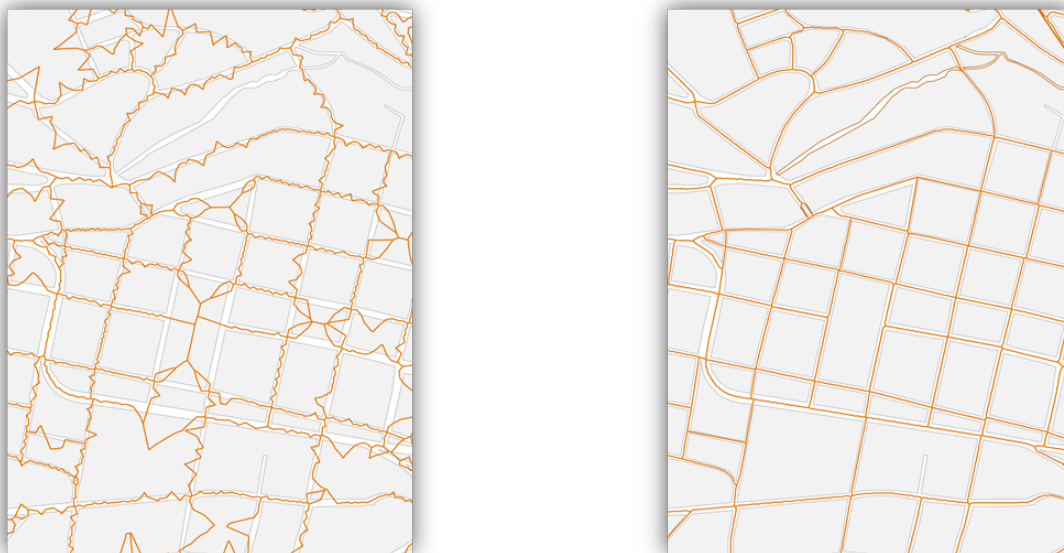
(b) Ríos de la ciudad.



Fuente: Gestión de Diseño Muestral - DINEM.

Finalmente, se necesita determinar la densidad y el ancho de buffer para determinar la calidad de la extensión de las manzanas. En las siguientes figuras se muestra el resultado de la extensión de polígonos para diferentes anchos de buffer y densidad.

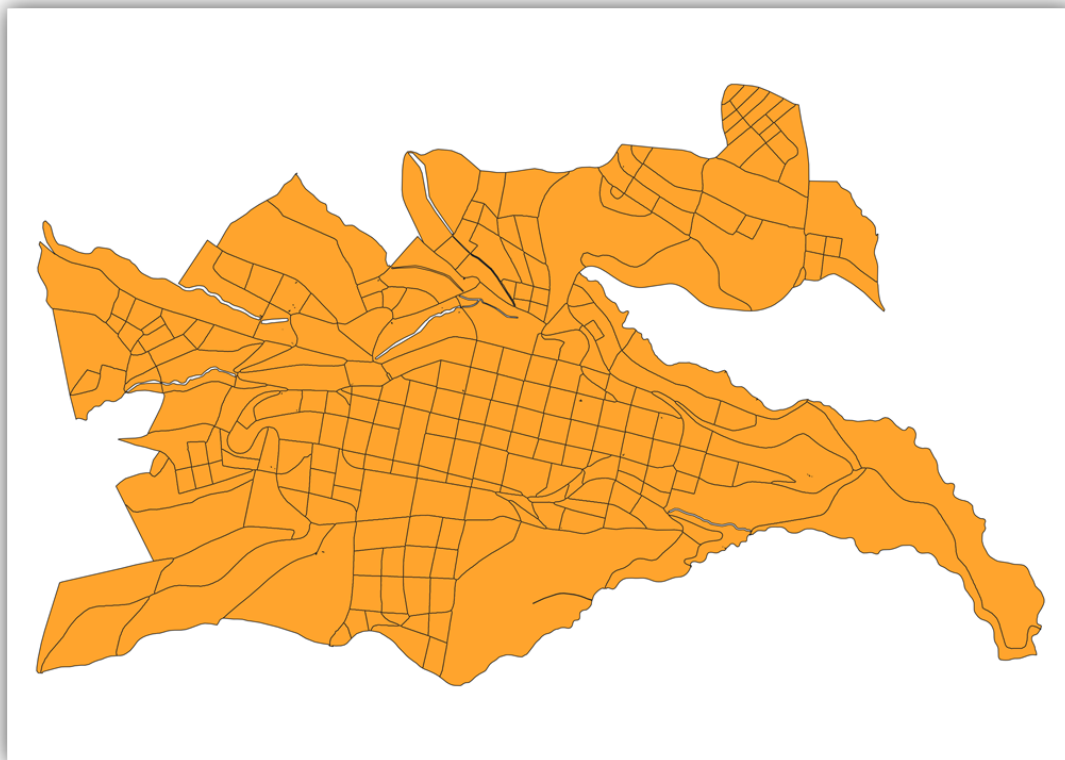
Figura 8: Resultado de extensión para diferentes valores de densidad y buffer.



Fuente: Gestión de Diseño Muestral - DINEM.

En la siguiente imagen, se presenta el resultado de la extensión de polígonos, como se utilizó información de "linderos", se evitará la unión de manzanas que se encuentren separadas por ríos.

Figura 9: Extensión de manzanas en la parroquia San Miguel de Bolívar.



Fuente: Gestión de Diseño Muestral - DINEM.

Cálculo matriz de incidencia

La matriz de incidencia es una matriz cuadrada que tiene como dimensión el número de polígonos considerados en la conglomeración, cuyas entradas toman valores de 1 o 0 en función de la existencia o no de una frontera en común entre cada par de manzanas o sectores dispersos respectivamente. La función requiere como insumos un archivo geográfico de polígonos adyacentes y el nivel de tolerancia, este valor determina la existencia de una frontera en común en función de la extensión de la intersección entre cada par de polígonos, si dicha intersección supera la tolerancia existe frontera en común.

Figura 10: Transformación de frontera común en matriz de incidencia.



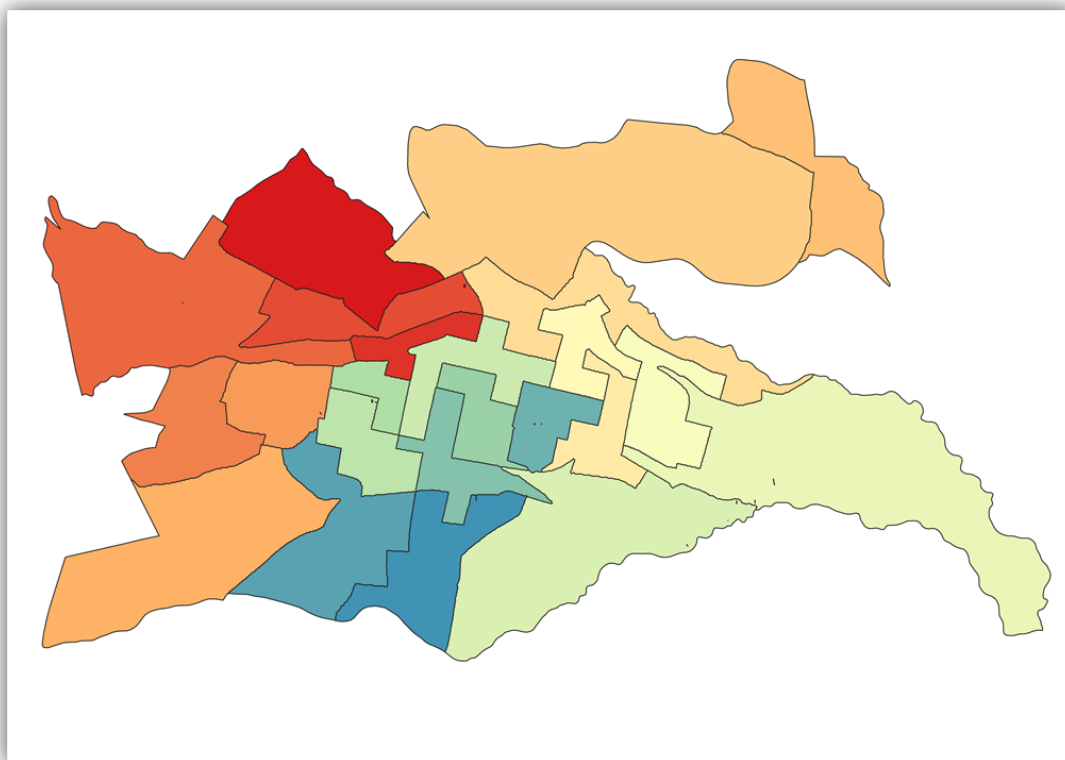
	6	7	8	9	10	14	20
6	0	0	1	0	0	1	0
7	0	0	1	0	0	1	0
8	1	1	0	0	1	0	0
9	0	0	0	0	1	0	0
10	0	0	1	1	0	0	0
14	1	1	0	0	0	0	1
20	0	0	0	0	0	1	0

Fuente: Gestión de Diseño Muestral - DINEM.

Conglomeración

La función que permite realizar la conglomeración requiere como insumos la matriz de incidencia y una base de datos con los identificadores de los polígonos y la variable de peso (número de viviendas ocupadas), es importante que tanto la matriz de incidencia como la base de datos de pesos estén ordenadas de la misma manera. El resultado de la ejecución de la función *Conglomeración* es la base de datos de pesos junto con una variable que indica a que conglomerado pertenece cada polígono.

Figura 11: Resultado de la ejecución de la función *Conglomeración*.

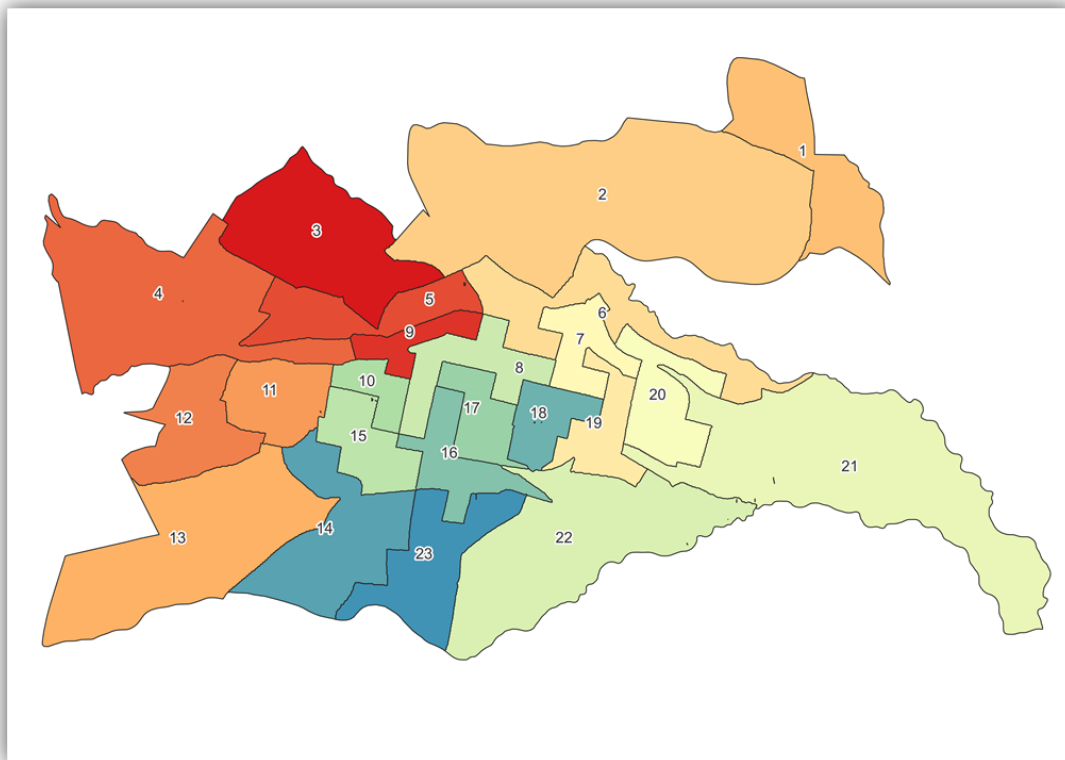


Fuente: Gestión de Diseño Muestral - DINEM.

Numeración

La última función que está presente en el paquete *psuR* es la de *Numeración*, la misma requiere el objeto geográfico de los conglomerados, cuyo resultado son los conglomerados numerados desde 1 hasta n , donde n es el número de conglomerados resultante, siguiendo un orden serpenteante iniciando en aquel conglomerado cuyo centroide se encuentra más al norte.

Figura 12: Numeración serpenteante para los conglomerados de San Miguel.

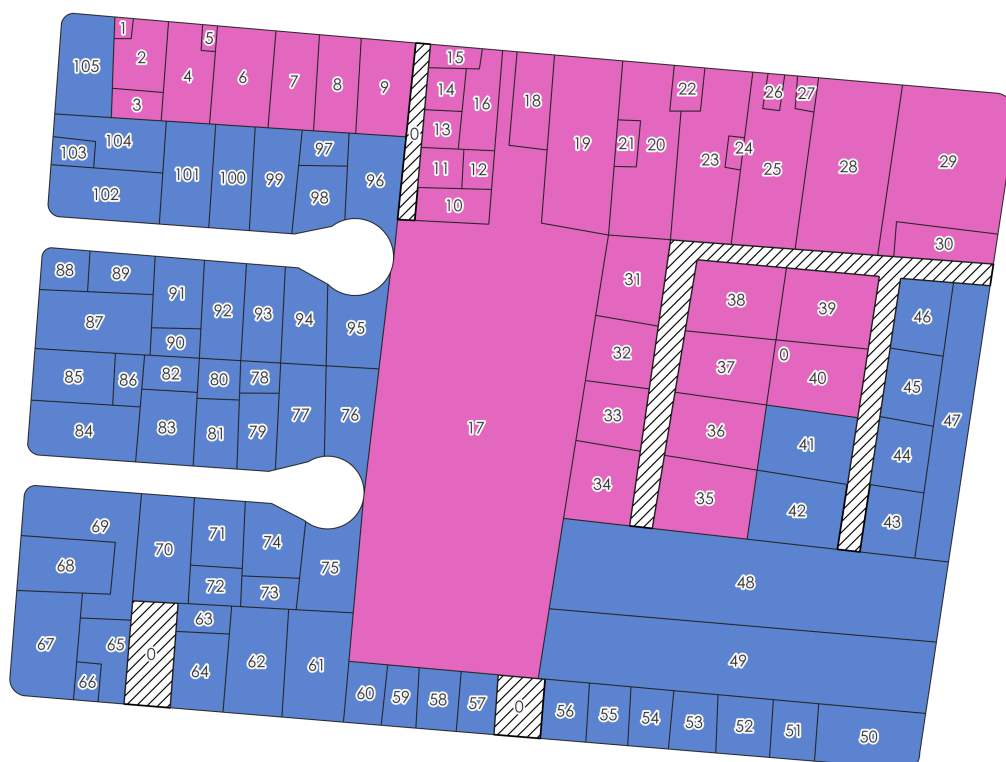


Fuente: Gestión de Diseño Muestral - DINEM.

Tras realizar el proceso de conglomeración, se identificaron conglomerados que no alcanzaban el número mínimo de viviendas, esto se debe a que hay partes amanzanadas o dispersas en parroquias que al juntar en su totalidad todas las manzanas o sectores dispersos que la conforman no llegan al límite establecido.

Además, existen conglomerados que son considerablemente grandes, esto se debe a que existen manzanas con un número muy grande de viviendas. Dado que la cartografía con la que cuenta el INEC llega a nivel de edificio censal se puede generar una partición de las manzanas teniendo así un número más homogéneo de viviendas. En la siguiente imagen se presenta un ejemplo de partición de manzanas.

Figura 13: Resultado de partición de manzanas grandes.



Fuente: Gestión de Diseño Muestral - DINEM.

Con estas consideraciones finales se llega a un total de 56.239 UPM distribuidas de la siguiente forma.

Figura 14: Distribución de las UPM por viviendas ocupadas.



Fuente: Gestión de Diseño Muestral - DINEM.

Definición del peso mínimo requerido para formar un conglomerado

En la sección anterior se presentaron los resultados de la conglomeración considerando un peso mínimo de 60 viviendas para formar un conglomerado, sin embargo, se consideraron tres escenarios de conglomeración teniendo un peso mínimo de 60, 80 y 100 viviendas en la parte amanzanada, en la parte dispersa siempre se consideró un mínimo de 60 viviendas. En la siguiente tabla se presenta el número de UPM resultante para cada escenario.

Tabla 2: Resultado escenarios de conglomeración.

Escenario	Número de UPM amanzanado	Número de UPM disperso	Número de UPM total
Mínimo 60	44.623	11.616	56.239
Mínimo 80	34.009	11.616	45.625
Mínimo 100	27.630	11.616	39.246

Fuente: Gestión de Diseño Muestral - DINEM.

Para determinar cual de los escenarios es el ideal se procede con el cálculo del coeficiente de correlación intraclase para cada uno de los indicadores que fueron considerados al momento de la estratificación. En la siguiente tabla se presentan los resultados para los indicadores de la dimensión de vivienda¹, se incluye una columna de la razón del efecto de diseño generalizado obtenidos en los escenarios de 60 y 100 viviendas.

Tabla 3: Coeficiente de correlación intraclase para los indicadores de la dimensión vivienda.

Indicador	ρ_{60}	ρ_{80}	ρ_{100}	Razón deff
techo	0,374	0,370	0,368	1,01
paredes	0,287	0,284	0,281	1,01
piso	0,338	0,334	0,332	1,01
r_agua	0,491	0,488	0,485	1,01
p_agua	0,804	0,802	0,800	1,00
cx_sh	0,625	0,621	0,618	1,01
luz	0,116	0,116	0,116	1,00
basura	0,577	0,576	0,576	1,00

Fuente: Gestión de Diseño Muestral - DINEM.

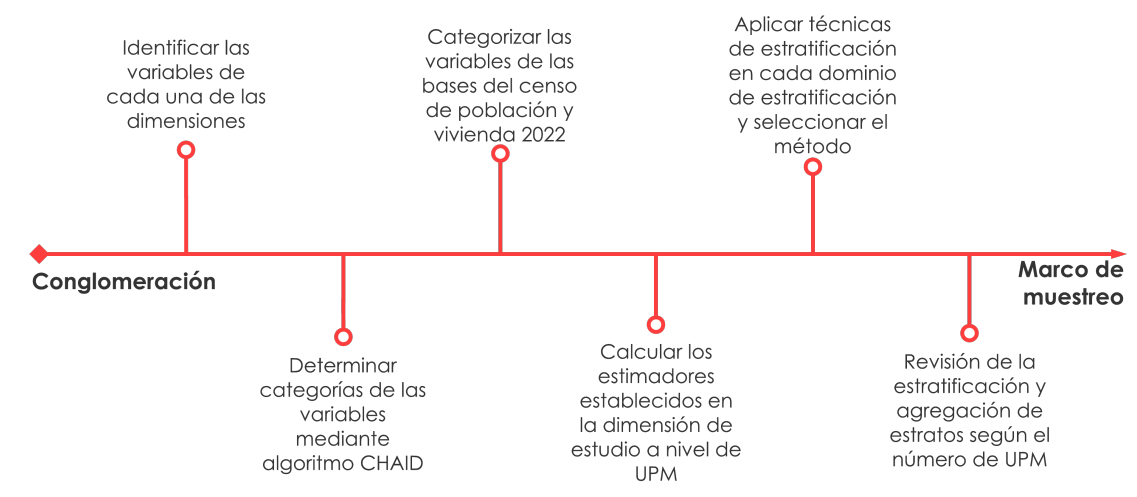
Al analizar los resultados, se obtuvo que el coeficiente de correlación intraclase es similar para los tres escenarios. Al calcular la razón del efecto de diseño generalizado para los escenarios de mínimo 60 y 100 viviendas, se obtuvieron valores muy cercanos a 1, lo que indica que, al momento de calcular un tamaño de muestra con un marco con mínimo 60 viviendas se tendría un incremento mínimo con respecto al escenario de 100 viviendas. Además, el escenario de mínimo 60 viviendas cuenta con un 68,8 % más de UPM que en el escenario de mínimo 100 viviendas, razones por las cuales se escogió el primer escenario.

¹En la Tabla 5 se detalla cada indicador

Estratificación

El proceso de estratificación se realiza luego de finalizada la conglomeración, el cual transforma el conjunto de conglomerados uniformes en tamaño en el Marco de Muestreo de encuestas dirigidas a hogares. En la siguiente figura se muestran los subprocesos involucrados en la estratificación.

Figura 15: Flujograma de la estratificación.



Fuente: Gestión de Diseño Muestral - DINEM.

Dominios de estratificación

De acuerdo a un análisis demográfico y económico, se decidió incluir 10 cantones que serán tomados como dominios independientes. Los cantones se dividen en dos grupos: cantones completos cuya numeración empieza en 30 y cantones urbanos cuya numeración comienza en 40. Esta distinción se debe a que la parte rural del segundo grupo es muy pequeña con respecto al número de UPM.

El orden de su codificación va en función a la División Político Administrativa del Ecuador (ver Tabla 4). En total, se dividió al Ecuador en 34 dominios geográficos correspondientes a: 5 cantones completos, 5 cantones urbanos, 10 restos de provincias y 14 provincias completas.

Tabla 4: Códigos de los dominios cantonales.

Provincia	Cantón	Cantón Urbano	Código
Azuay	Cuenca	-	30
Chimborazo	Riobamba	-	31
Loja	Loja	-	32
Pichincha	Quito	-	33
Tungurahua	Ambato	-	34

Provincia	Cantón	Cantón Urbano	Código
El Oro	-	Machala	40
Esmeraldas	-	Esmeraldas	41
Guayas	-	Guayaquil	42
Manabí	-	Manta	43
Santo Domingo	-	Santo Domingo	44

Fuente: Gestión de Diseño Muestral - DINEM.

Sobre los 34 dominios geográficos se determinaron un total de 62 dominios de estratificación que corresponden a:

- 10 dominios de estratificación de cinco cantones por área.
- 5 dominios de estratificación de cinco cantones considerando únicamente su parte urbana.
- 1 dominio de estratificación correspondiente a la provincia de Galápagos.
- 10 dominios de estratificación de diez restos de provincias en su área urbana.
- 5 dominios de estratificación de cinco restos de provincias en su área rural.
- 13 dominios correspondientes a provincias en su área urbana.
- 18 dominios correspondientes a provincias en su área rural.

Identificar variables en cada una de las dimensiones

El Marco de Muestreo debe permitir clasificar las UPM de acuerdo a su estructura socioeconómica con el fin de aumentar la calidad de inferencia de las encuestas y minimizar los errores de muestreo, para lo cual se utilizó la información censal a nivel de personas, hogares y viviendas, en las siguientes dimensiones:

- características de la vivienda,
- acceso a los servicios básicos,
- condiciones del hogar,
- patrimonio de los hogares,
- educación,
- mercado laboral y seguros,
- maternidad temprana y
- demografía.

En total se consideraron 46 variables para la estratificación del Marco de Muestreo (Tabla 5).

Tabla 5: Variables de estratificación.

Dimensión	Indicadores
Características de la vivienda	Porcentaje de viviendas con material de techo adecuado. (techo)
	Porcentaje de viviendas con material de paredes exteriores adecuado. (paredes)
	Porcentaje de viviendas con material de piso adecuado. (piso)
Acceso a los servicios básicos	Porcentaje de viviendas con procedencia de agua adecuado. (p_agua)
	Porcentaje de viviendas que reciben el agua de forma adecuada. (r_agua)



Dimensión	Indicadores
Acceso a los servicios básicos	Porcentaje de viviendas que poseen servicio higiénico adecuado. (cx_sh)
	Porcentaje de viviendas que poseen servicio de luz eléctrica adecuado. (luz)
	Porcentaje de viviendas que eliminan la basura de forma adecuada. (basura)
Condiciones del hogar	Porcentaje de hogares no hacinados.
	Porcentaje de hogares con fuente de agua mejorada para beber.
	Porcentaje de hogares no monoparentales.
	Porcentaje de hogares con tasa de dependencia baja.
Patrimonio de los hogares	Porcentaje de hogares que poseen de una fuente adecuada para cocinar.
	Porcentaje de hogares que disponen de servicio de teléfono convencional.
	Porcentaje de hogares que disponen de servicio de teléfono celular.
	Porcentaje de hogares que disponen de televisión por cable.
	Porcentaje de hogares que disponen de servicio de internet.
	Porcentaje de hogares que disponen de computadora.
	Porcentaje de hogares que disponen de refrigeradora.
	Porcentaje de hogares que disponen de lavadora.
	Porcentaje de hogares que disponen de secadora.
	Porcentaje de hogares que disponen de microondas.
	Porcentaje de hogares que disponen de extractores de olores.
	Porcentaje de hogares que disponen de auto de uso exclusivo.
	Porcentaje de hogares que disponen de moto de uso exclusivo.
	Porcentaje de hogares que disponen de cuarto o espacio exclusivo para cocinar.
	Porcentaje de hogares que disponen de cuarto o espacio exclusivo para el servicio higiénico.
	Porcentaje de hogares que disponen de cuarto o espacio exclusivo para ducharse.
Educación	Porcentaje de hogares cuyo jefe de hogar posee más de 13 años de escolaridad.
	Tasa neta de asistencia de educación general básica (5 a 14 años).
	Tasa neta de asistencia ajustada en bachillerato (15 a 17 años).
	Alfabetismo digital.
Mercado laboral y Seguros	Tasa de participación bruta.
	Tasa de participación global.
	Tasa de ocupados.
	Porcentaje de personas que están cubiertos por algún tipo de seguro.
	Tasa de no trabajo infantil (5 a 17 años).
Maternidad Temprana	Tasa de jóvenes que trabajan o estudian (18 a 24 años).
	Porcentaje de mujeres adolescentes sin hijos (12 a 19 años).
Demografía	Porcentaje de mujeres.
	Porcentaje de hombres.
	Tasa de juventud (15 a 29 años).
	Tasa de envejecimiento (+ 65 años).
	Porcentaje de personas que se autoidentificaron como Indígenas y Montubios.
	Porcentaje de personas que se autoidentificaron como Afroecuatorianos y Mulatos.
	Porcentaje de personas que se autoidentificaron como Blancos, Mestizos, Otros.

Fuente: Gestión de Diseño Muestral - DINEM.



Los estratos deben jerarquizar las UPM del país de acuerdo a un nivel cuantificable de ausencia o presencia de los fenómenos. Por lo tanto, las variables utilizadas en el proceso de estratificación permiten agrupar las UPM de acuerdo a cierto grado de bienestar de la población, es decir, mientras existan más condiciones relacionadas con el bienestar en las UPM, tenderán a ser consideradas de estrato “alto”; por el contrario, mientras las UPM tiendan a la ausencia de estas características (en conjunto) serán consideradas de estrato “bajo”.

Determinar categorías mediante algoritmo CHAID

La clasificación de bienestar de las variables pertenecientes a las dimensiones **Condiciones del hogar, Patrimonio de los hogares, Educación, Mercado laboral y Seguros, Maternidad Temprana y Demografía** se realizó de manera general para todos los dominios de estratificación, mientras que es independiente por dominio para las variables de las dimensiones **Características de la vivienda y Acceso a los servicios básicos**. Para esta clasificación se llevó a cabo un análisis CHAID, el cual consiste en generar árboles de decisión mediante estadísticos de chi-cuadrado, tomando como variable dependiente la pobreza por ingresos, calculada desde la Encuesta Nacional de Empleo y Desempleo (ENEMDU) Acumulada 2023.

No todos los dominios de estratificación tienen un número de casos (Unidades primarias de muestreo) suficiente para realizar la clasificación; en estos se utilizaron los resultados de la clasificación correspondiente al cruce entre Región Natural y Área a la que pertenece dicho dominio. Se decidió clasificar como bienestar a por lo menos una categoría de cada una de las variables clasificadas. En la siguiente tabla se presenta el resultado de la categorización de las opciones de respuesta de la pregunta “V03: ¿El material predominante del techo de la vivienda es:” para todos los dominios de estratificación, el valor 1 aparece cuando la opción de respuesta está relacionada con condiciones de bienestar 0 en caso contrario.

Tabla 6: Categorización material techo vivienda con bienestar por dominio de estratificación.

Dominio estratificación	¿El material predominante del techo de la vivienda es:					
	Opción 1	Opción 2	Opción 3	Opción 4	Opción 5	Opción 6
011	1	1	0	1	0	0
012	1	0	0	0	0	0
021	1	1	0	1	0	0
022	1	0	0	1	0	0
031	1	1	0	1	0	0
032	1	1	0	1	0	0
041	1	0	0	0	0	0
042	1	0	0	0	0	0
051	1	0	0	0	0	0
052	1	0	0	0	0	0
062	1	0	0	0	0	0
071	1	1	1	1	0	0
072	1	1	0	0	0	0
081	1	1	0	1	0	0
082	1	1	0	0	0	0
091	1	1	0	0	0	0
092	1	0	0	0	0	0
101	1	0	0	0	0	0
102	1	0	0	0	0	0



Dominio estratificación	¿El material predominante del techo de la vivienda es:					
	Opción 1	Opción 2	Opción 3	Opción 4	Opción 5	Opción 6
111	1	1	0	1	0	0
112	1	0	0	0	0	0
121	1	1	0	1	0	0
122	1	0	0	1	0	0
131	1	1	0	1	0	0
132	1	0	0	1	0	0
141	1	1	0	0	0	0
142	1	1	0	1	0	0
152	1	1	0	0	0	0
171	1	1	0	1	0	0
172	1	0	0	0	0	0
182	1	0	0	0	0	0
211	1	1	0	0	0	0
212	1	1	0	0	0	0
221	1	1	1	0	0	0
222	1	1	0	0	0	0
241	1	0	0	1	0	0
301	1	1	1	1	0	0
302	1	0	0	1	0	0
311	1	0	0	1	0	0
312	1	0	0	1	0	0
321	1	1	0	1	0	0
331	1	1	0	1	0	0
332	1	0	0	1	0	0
341	1	1	0	1	0	0
342	1	0	0	1	0	0
441	1	1	0	1	0	0
232	1	0	0	1	0	0
401	1	1	0	1	0	0
411	1	1	0	1	0	0
421	1	1	0	1	0	0
431	1	1	0	1	0	0

Fuente: Gestión de Diseño Muestral - DINEM.

Las opciones de respuesta para la pregunta son las siguientes:

- Opción 1: Hormigón (losa, cemento)?
- Opción 2: Fibrocemento, asbesto (eternit, eurolit)?
- Opción 3: Zinc, aluminio (lámina o plancha metálica)?
- Opción 4: Teja?
- Opción 5: Palma, paja u hoja?
- Opción 6: Otro material?

Al analizar el resultado de la clasificación se puede concluir que la opción 1 está relacionada con bienestar en todos los dominios de estratificación mientras que las opciones 5 y 6 se relacionan con no bienestar en todos los dominios.

Cálculo de los indicadores a nivel de UPM y reducción de dimensiones

Con la categorización de bienestar obtenida del análisis CHAID, se clasifican las variables del censo a nivel de personas, hogares y viviendas, para luego agregar la información a nivel de UPM para obtener una matriz de información con las proporciones asociadas con el bienestar para todos los indicadores utilizados para el proceso de estratificación.

Con el fin de garantizar el carácter multiuso de las encuestas de hogares, se utilizó una técnica multivariada de reducción de datos que permita resumir la variación y las correlaciones entre variables, donde la medida de resumen de la matriz de información asociada al bienestar, definida por:

$$y = f(X_1, X_2, \dots, X_p)$$

es la primera componente del Análisis de Correspondencias Principales².

Como el análisis realizado se centra en la primera componente, se empieza por analizar la calidad de representación de las variables, el criterio nos dice que a mayor valor del coseno cuadrado se tiene una mayor vinculación o representación de la variable en el respectivo componente.

Luego se analiza la contribución de las variables en la contabilidad de la variabilidad de la primera componente. La teoría sugiere excluir aquellas variables que contribuyen menos del:

$$\frac{1}{\# \text{ de variables usadas en el ACP}} = \frac{1}{46} = 2,17\%.$$

Para evaluar la mejor representación de nuestra medida de resumen se analizaron 2 escenarios de forma independiente para cada uno de los dominios de estratificación:

- Escenario 1: Exclusión de las variables cuya calidad de representación es menor que el 25%.
- Escenario 2: Exclusión de las variables cuya calidad de representación es menor que el 30% y cuya contribución es menor a 2.17%.

Una vez obtenida nuestra medida de resumen (y), se presentan los efectos de diseño y efecto de diseño generalizado $G(S)$, considerando el número de estratos de acuerdo al número de UPM por dominio de estratificación, para todas las 46 variables iniciales. Dado que el efecto de diseño generalizado total del Escenario 1 es mejor que el del Escenario 2 se elige el primer escenario para la estratificación.

Aplicar técnicas de estratificación y seleccionar el método

Tal y como se expone en (CEPAL, 2023), la mejor estratificación es aquella que minimiza los errores de muestreo de los estimadores, por lo cual se consideran múltiples métodos de estratificación tanto univariados (Cuantiles, Dalenius, Estratificación Óptima, Estratificación Geométrica) como multivariados (K-medias, Genéticos) que buscan encontrar la mejor división para cada dominio de estratificación de forma independiente. A continuación se presenta una breve descripción de cada uno.

²De ahora en adelante, se utilizará la notación (y) para referirse a la medida de resumen de la matriz de información asociada al bienestar.

División en cuantiles.

Divide la población de las UPM en grupos creados a partir de la división de la distribución de la medida de resumen en intervalos regulares. Suele ser ideal si la estratificación se realiza alrededor de una variable.

Método de raíz de frecuencia acumulada.

Técnica de estratificación basada en la raíz cuadrada de las frecuencias acumuladas de la medida de resumen sobre las UPM, cuyo objetivo principal consiste en encontrar grupos que minimicen la siguiente función:

$$D = \sum_{h=1}^H W_h \sqrt{S_{y_h}^2}$$

donde:

- W_h = tamaño relativo del estrato h ,
- $S_{y_h}^2$ = varianza de la medida de resumen en el estrato h .

Estratificación óptima.

Es un método de estratificación basado en la minimización de la siguiente expresión, ligada a la varianza de una estrategia de muestreo estratificada:

$$\sum_{h=1}^{H-1} \left(\frac{N_h}{H} \right)^2 \left(\frac{1}{(n - N_H)a_H} - \frac{1}{N_h} \right) S_{y_h}^2$$

donde:

- N_h = número de UPM en el estrato h ,
- n = tamaño de muestra de las UPM,
- N = número de UPM en el marco de muestreo y $S_{y_h}^2$,
- $S_{y_h}^2$ = varianza de la medida de resumen en el estrato h ,
- a_h = regla de asignación para el tamaño de muestra.

La optimización de la función objetivo puede llevarse a cabo de diferentes formas, mientras que (Lavallée y Hidiroglou, 1988) utilizaron el algoritmo de optimización de (Sethi, 1963) para encontrar los valores óptimos, (Kozak, 2004) define un algoritmo iterativo mediante arranques aleatorios para optimizar el proceso de minimización de esta técnica de estratificación.

Estratificación geométrica.

Para el caso de poblaciones de UPM con medidas de resumen sesgadas, las varianzas relativas (coeficientes de variación) de la medida de resumen en cada estrato eran similares, es decir:

$$\frac{S_{x_1}}{\bar{X}_1} \cong \frac{S_{x_2}}{\bar{X}_2} \cong \dots \cong \frac{S_{x_H}}{\bar{X}_H}$$

(Gunning y Horgan, 2004) desarrollaron un método con el objetivo de que los coeficientes de variación de la medida de resumen tendieran a ser iguales dentro de los estratos y, de esta forma, encontraron que los límites que definían esos grupos estaban conformados en progresión geométrica.

K-medias de Jarque.

(Jarque, 1981) propone utilizar una versión modificada del algoritmo de k-medias (Macqueen, 1967), cuyo objetivo es la minimización de la siguiente función de distancia:

$$\sum_{h=1}^H \sum (X_k - \bar{X}_h) \Lambda^{-1} (X_k - \bar{X}_h)$$

donde:

- X_k = medición de las P variables de la matriz de información en la k -ésima UPM,
- X_h = vector de medias de la matriz de información en el estrato h .,
- Λ = matriz diagonal de tamaño $P \times P$ cuyas entradas se definen como la varianza de las P variables de la matriz X .

Algoritmos genéticos.

(Ballin y Barcaroli, 2013) argumentan que la mejor estratificación es la división del marco de muestreo que asegura el mínimo costo muestral que satisfaga algunas restricciones de precisión o maximice la precisión de los indicadores de interés bajo las restricciones. De esta forma, este algoritmo busca minimizar la siguiente función de costos:

$$C_0 + \sum_{h=1}^H c_h n_h$$

donde:

- C_0 = costo fijo.
- c_h = costo promedio de observar un hogar en el estrato h .

Las estratificaciones resultantes deben ser evaluadas mediante un criterio de estratificación multivariante que contemple cada una de las variables de la matriz de información mediante el efecto de diseño generalizado $G(S)$.

$$G(S) = DEFF_p = \sum_{p=1}^P \frac{1}{S_{X_p}^2} \sum_{h=1}^H W_h S_{X_p}^2$$

donde:

$$W_h = \frac{N_h}{N}, \text{ tamaño relativo del estrato } h.$$

$$S_{X_p}^2 = \text{varianza de la variable } X_p \text{ en la población.}$$

Para seleccionar el mejor método de estratificación se considera el menor efecto de diseño generalizado en cada dominio de estratificación. En la Tabla 15 se presenta el método utilizado para cada dominio de estratificación.

Revisión de la estratificación y agregación de estratos según el número de UPM

Según recomendaciones de CEPAL, un estrato socioeconómico debe constar de mínimo 100 UPM, para facilitar la administración del marco de muestreo y evitar desgaste en los informantes. Considerando esta recomendación, se construyen tres, dos o un estrato socioeconómico en cada dominio de estratificación, por este motivo se tiene que 11 dominios de estratificación tendrán un estrato socioeconómico único, en 10 dominios de estratificación se tiene un número de UPM que va desde 200 hasta 299 por lo que se realizará la estratificación a dos estratos socioeconómicos y en 41 dominios de estratificación se tiene un número de UPM mayor o igual a 300 por lo que se realizará la estratificación a tres estratos socioeconómicos.

Tras realizar el proceso de estratificación se encontró que en 21 dominios de estratificación se obtuvo al menos un estrato con menos de 100 UPM, por lo que en ellos se realizó una re-estratificación esperando un estrato menos de lo que se esperaba inicialmente.

En el dominio de estratificación 041, al realizar la estratificación esperando tres estratos, se obtuvo el siguiente número de UPM por estrato: 1-36, 2-159 y 3-187. Al realizar la re-estratificación a dos estratos, se obtuvo el siguiente número de UPM por estrato: 1-84 y 2-298. Teniendo aún un número de UPM menor a 100 en el estrato 1, únicamente para este dominio de estratificación se decidió colapsar los estratos 1 y 2 del primer resultado para formar dos estratos que cuenten con el número mínimo de UPM. En la siguiente tabla se presenta el número de estratos por provincia.



Tabla 7: Número de estratos por provincia.

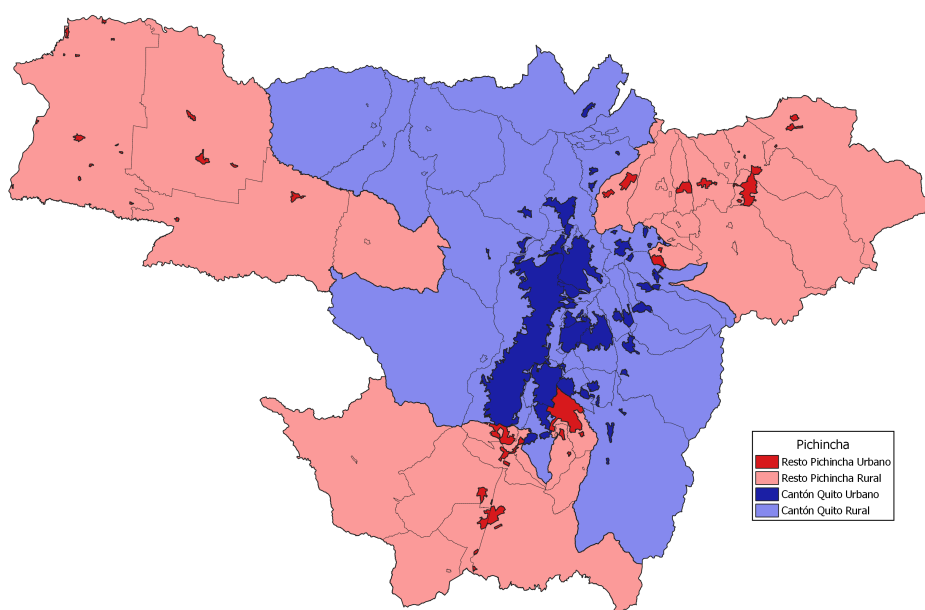
Provincia	Número de estratos	Número de UPM
Azuay	10	2.709
Bolívar	4	651
Cañar	5	820
Carchi	3	606
Cotopaxi	6	1.478
Chimborazo	7	1.639
El Oro	8	2.461
Esmeraldas	7	1.841
Guayas	9	14.414
Imbabura	5	1.590
Loja	8	1.582
Los Ríos	6	2.989
Manabí	8	5.176
Morona Santiago	4	586
Napo	2	363
Pastaza	2	346
Pichincha	11	10.494
Tungurahua	9	2.002
Zamora Chinchipe	2	342
Galápagos	1	112
Sucumbíos	4	672
Orellana	3	560
Sto. Domingo de los Tsáchilas	5	1.668
Santa Elena	4	1.138
Total Nacional	133	56.239

Fuente: Gestión de Diseño Muestral - DINEM.

En el dominio de estratificación 242 correspondiente a Santa Elena Rural se tiene un total de 37 UPM, siendo el único estrato en el que se tiene menos de 100 UPM. A continuación, se presentan algunos ejemplos.

Pichincha.

Figura 16: Dominios de estratificación - Pichincha.



Fuente: Gestión de Diseño Muestral - DINEM.

La provincia de Pichincha consta de cuatro dominios de estratificación. (Cantón Quito Urbano, Cantón Quito Rural, Resto Pichincha Urbano, Resto Pichincha Rural), al estratificar considerando tres estratos meta se obtuvo la siguiente distribución:

Tabla 8: Resultados preliminares estratificación Pichincha.

Provincia	Dominio	Área	Estrato	Nro. UPM dominio	Nro. UPM estrato	Proporción por dominio
17	171	1	1711	955	319	33,40
	171		1712		434	45,45
	171		1713		202	21,15
	172	2	1721	408	133	32,60
	172		1722		146	35,78
	172		1723		129	31,62
	331	1	3311	8.629	2734	31,68
	331		3312		3491	40,46
	331		3313		2404	27,86
	332	2	3321	502	185	36,85
	332		3322		245	48,8
	332		3323		72	14,34

Fuente: Gestión de Diseño Muestral - DINEM.

Sin embargo, el estrato 3323 presenta menos de 100 UPM por lo cual se consideró re-estratificar el dominio 332 considerando dos estratos, obteniendo los siguientes resultados:

Tabla 9: Resultados estratificación Pichincha.

Provincia	Dominio	Área	Estrato	Nro. UPM dominio	Nro. UPM estrato	Proporción por dominio
17	171	1	1711	955	319	33,40
	171		1712		434	45,45
	171		1713		202	21,15
	172	2	1721	408	133	32,60
	172		1722		146	35,78
	172		1723		129	31,62
	331	1	3311	8.629	2734	31,68
	331		3312		3491	40,46
	331		3313		2404	27,86
	332	2	3321	502	381	75,90
	332		3322		121	24,10

Fuente: Gestión de Diseño Muestral - DINEM.

Galápagos.

Figura 17: Dominios de estratificación - Galápagos.



Fuente: Gestión de Diseño Muestral - DINEM.

La Provincia de Galápagos cuenta con muy pocas UPM por lo cual se decidió colapsar la provincia en un solo dominio sin área.

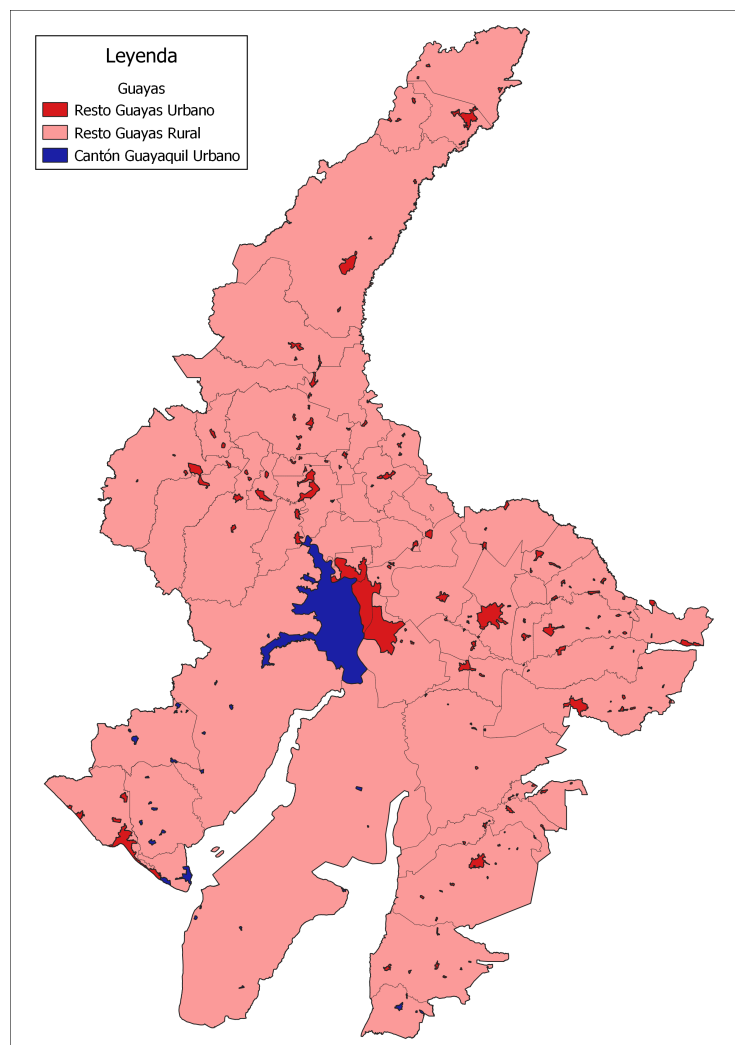
Tabla 10: Resultados estratificación Galápagos.

Provincia	Dominio	Área	Estrato	Nro. UPM dominio	Nro. UPM estrato	Proporción por dominio
20	209	9	2099	112	112	100,00

Fuente: Gestión de Diseño Muestral - DINEM.

Guayas.

Figura 18: Dominios de estratificación - Guayas.



Fuente: Gestión de Diseño Muestral - DINEM.

La provincia de Guayas consta de tres dominios de estratificación. (Cantón Guayaquil Urbano, Resto Guayas Urbano, Resto Guayas Rural). La parte rural del cantón Guayaquil es muy pequeña, por lo cual se estratifica como Guayas rural.

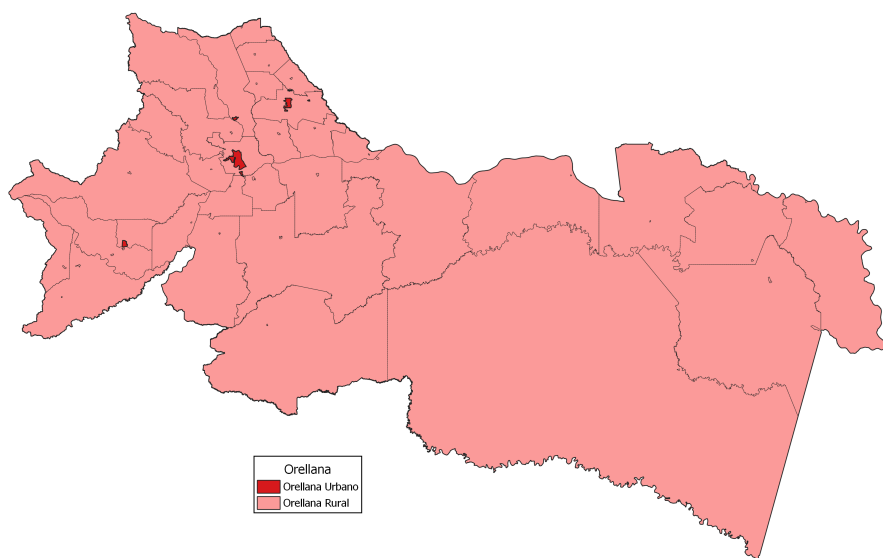
Tabla 11: Resultados estratificación Guayas.

Provincia	Dominio	Área	Estrato	Nro. UPM dominio	Nro. UPM estrato	Proporción por dominio
09	091	1	0911	4293	2076	48.36
		1	0912	4293	1609	37.48
		1	0913	4293	608	14.16
	092	2	0921	1236	450	36.41
		2	0922	1236	388	31.39
		2	0923	1236	398	32.2
	421	1	4211	8885	2856	32.14
		1	4212	8885	3812	42.9
		1	4213	8885	2217	24.95

Fuente: Gestión de Diseño Muestral - DINEM.

Orellana.

Figura 19: Dominios de estratificación - Orellana.



Fuente: Gestión de Diseño Muestral - DINEM.

La provincia de Orellana consta de dos dominios de estratificación. (Orellana Urbano, Orellana Rural), los resultados preliminares de su estratificación fueron:

Tabla 12: Resultados preliminares estratificación Orellana.

Provincia	Dominio	Área	Estrato	Nro. UPM dominio	Nro. UPM estrato	Proporción por dominio
22	221	1	2211	292	64	21,92
			2212		228	78,08
	222	2	2221	268	110	41,04
			2222		158	58,96

Fuente: Gestión de Diseño Muestral - DINEM.

Al realizar la estratificación socioeconómica se obtuvieron 4 estratos socioeconómicos, sin embargo, el estrato 2211 presenta menos de 100 UPM por lo cual se consideró re-estratificar el dominio 221 considerando un estrato.

Tabla 13: Resultados estratificación Orellana.

Provincia	Dominio	Área	Estrato	Nro. UPM dominio	Nro. UPM estrato	Proporción por dominio
22	221	1	2219	292	292	100,00
			2221		110	41,04
	222	2	2222	268	158	58,96

Fuente: Gestión de Diseño Muestral - DINEM.

Coordinación de Muestras

Debido a que el INEC realiza varias encuestas dirigidas a hogares en el periodo intercensal, la administración del marco de muestreo se vuelve crucial para poder responder a los diferentes requerimientos de información, manteniendo un ecosistema donde puedan coexistir las diferentes operaciones estadísticas. (CEPAL, 2023) plantea que la coordinación de muestras se realice a partir de una planificación rigurosa que permita conocer de antemano qué UPM se seleccionarán a lo largo del siguiente período intercensal. Esta relación debe estar supeditada a todas las operaciones estadísticas basadas en encuestas de hogares.

Puesto que las UPM a participar en las encuestas dirigidas a hogares del INEC serán seleccionadas con probabilidad proporcional al tamaño, se utilizará una coordinación de muestras proporcionales utilizando el algoritmo de Pareto (Rosen, 1997), las cuales se adaptan a la medida de tamaño de las UPM, la cual corresponde al total de viviendas particulares ocupadas en las mismas. Este algoritmo se basa en los principios de la función de distribución de Pareto con parámetros (I, J) y crea los siguientes números aleatorios permanentes:

$$\xi_i^{par} = \frac{\frac{\xi_i^P}{1 - \xi_i^P}}{\frac{\pi_i}{1 - \pi_i}}$$

donde:

- ξ_i^P = número aleatorio permanente de la UPM_i .
 π_i = $n_I \cdot p_i$, probabilidad de inclusión de la i -ésima UPM (mayor o igual a 1).

Por consiguiente, al ordenar el marco mediante los números ξ_i^{par} y seleccionar los primeros elementos, se obtendrá una muestra secuencial de Poisson. En la Tabla 14 se ejemplifica la selección de dos muestra coordinadas negativamente³. Cabe mencionar que el número aleatorio permanente ξ_i^P es fijo para cada UPM en el periodo intercensal, mientras que el número de Pareto ξ_i^{par} puede variar, puesto que también depende de la probabilidad de selección de cada UPM.

Tabla 14: Ejemplo de selección de dos muestras de Pareto coordinadas negativamente.

Unidad Primaria de Muestreo	ξ_i^P	ξ_i^{par}	s^1	s^2
006	0,00054	0,00191	1	0
004	0,10390	0,31927	1	0
005	0,13682	0,74385	1	0
010	0,27018	1,05055	0	0
007	0,27574	1,68336	0	1
011	0,35584	2,39573	0	1
001	0,49264	3,97909	0	1
002	0,58316	6,18585	0	0
008	0,77467	10,05965	0	0
003	0,82851	19,43617	0	0
009	0,94986	87,13004	0	0

Fuente: Gestión de Diseño Muestral - DINEM.

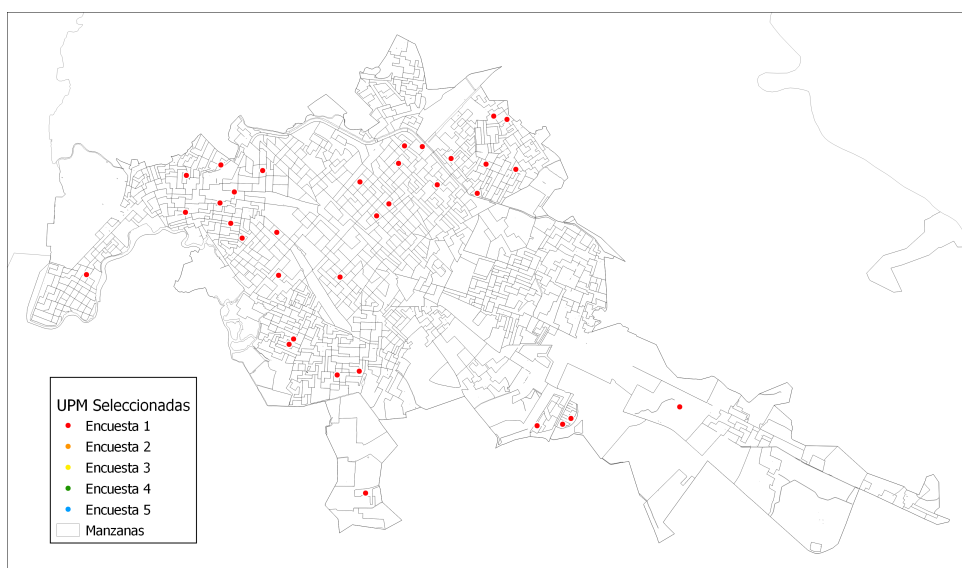
A continuación se presenta una serie de mapas que ilustran un esquema de coordinación de muestras negativo entre 5 encuestas para el Cantón Urbano Machala.

Encuesta Mensual de Mercado Laboral

Las primeras tres operaciones estadísticas a coordinar están dirigidas a reportar indicadores asociados a las condiciones laborales de la Población Económicamente Activa. Dado su diseño muestral, son encuestas con periodicidad mensual, compartiendo tanto tamaño como distribución a lo largo de todo su periodo de ejecución.

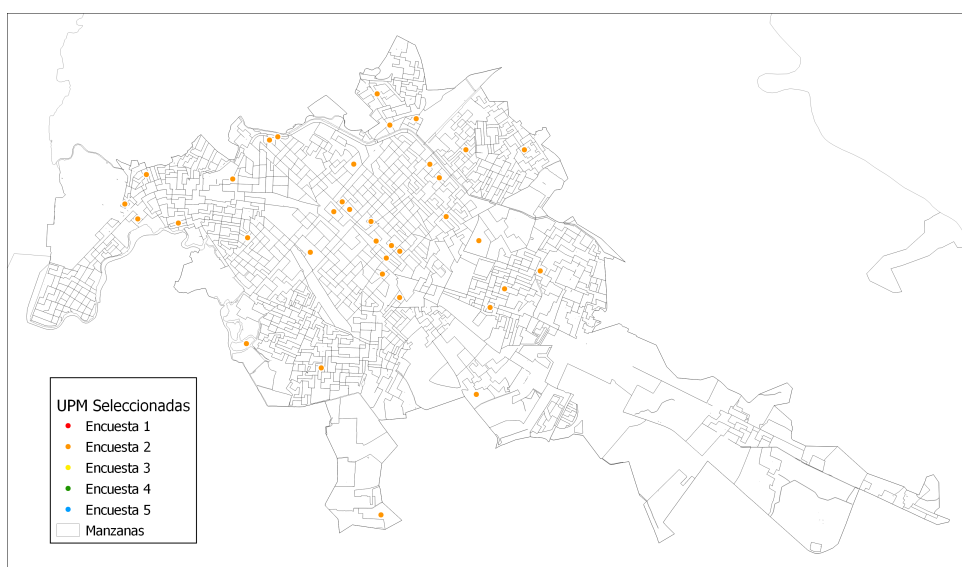
³Todas las encuestas del periodo intercensal serán coordinadas negativamente a excepción que se especifique lo contrario en su requerimiento de diseño, o el total de UPM disponibles por estrato sea menor al requerido para mantenerla.

Figura 20: Ejemplo de selección encuesta 1 Machala.



Fuente: Gestión de Diseño Muestral - DINEM.

Figura 21: Ejemplo de selección encuesta 2 Machala.



Fuente: Gestión de Diseño Muestral - DINEM.

Figura 22: Ejemplo de selección encuesta 3 Machala.

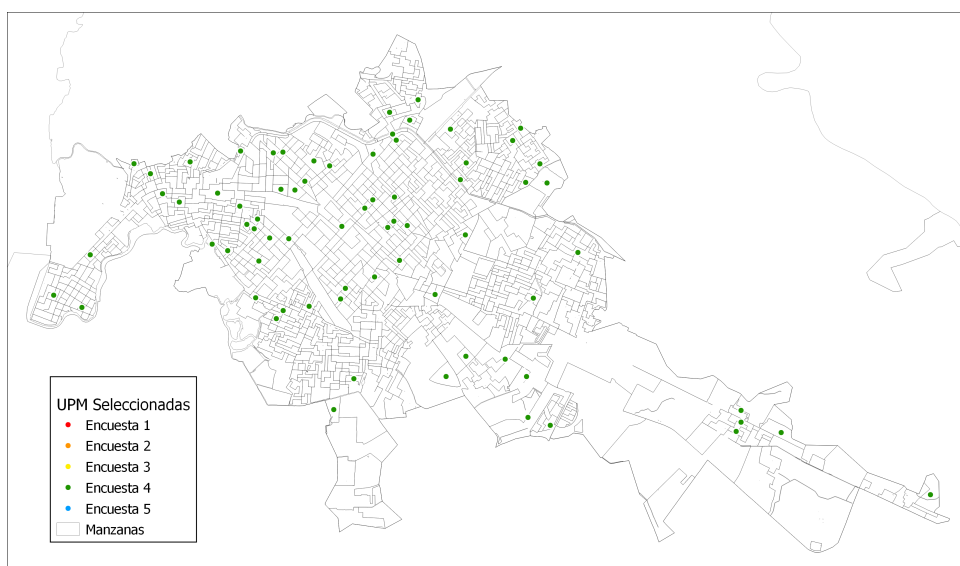


Fuente: Gestión de Diseño Muestral - DINEM.

Encuesta Decenal de Ingresos y Gastos

La cuarta operación estadística a coordinar es una encuesta de corte económico, cuyo principal atención está en recabar información acerca de los patrones de consumo de los hogares con la finalidad de generar índices asociados a la Canasta Básica Familiar y la Pobreza por Ingresos. En una encuesta de carácter provincial que, idealmente, se realiza una vez en cada periodo intercensal.

Figura 23: Ejemplo de selección encuesta 4 Machala.

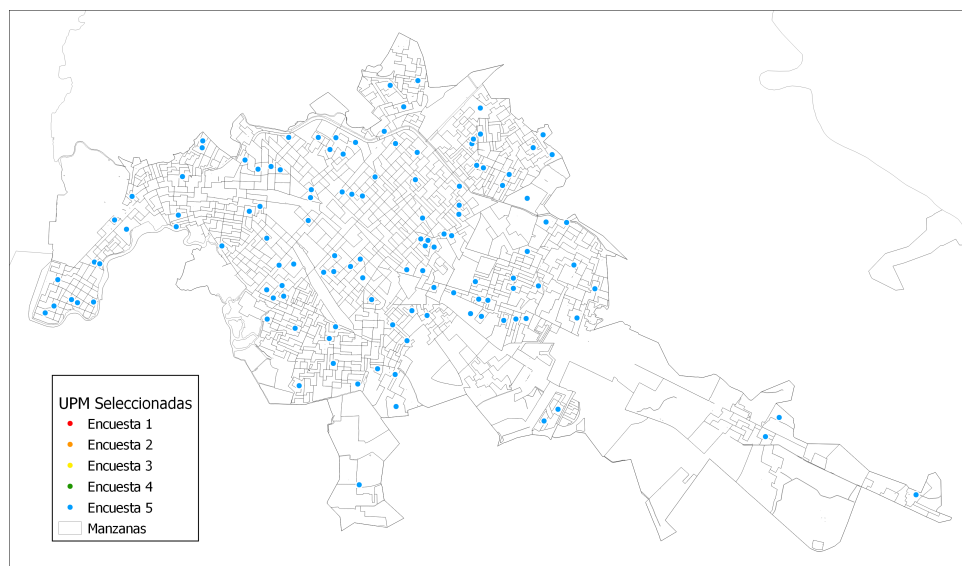


Fuente: Gestión de Diseño Muestral - DINEM.

Encuesta Quinquenal de Salud y Nutrición

La quinta operación estadística a coordinar es una encuesta demográfica provincial, cuyo principal interés es generar indicadores de salud y nutrición, tanto para la primera infancia como para las mujeres en edad fértil, entre otras poblaciones de interés. Idealmente, se realiza al menos una vez en cada periodo intercensal.

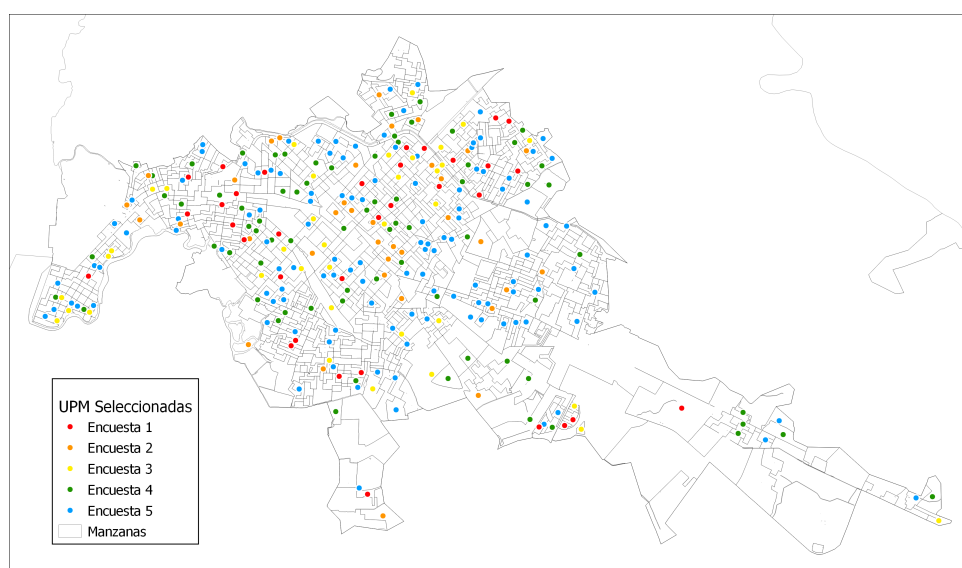
Figura 24: Ejemplo de selección encuesta 5 Machala.



Fuente: Gestión de Diseño Muestral - DINEM.

Por último, en el siguiente mapa se muestra cómo quedaría la distribución espacial de las cinco encuestas coordinadas de manera negativa. Nótese la dispersión de cada una de las mismas, estando presentes sobre toda la extensión del dominio de estudio, junto al no solapamiento de las mismas.

Figura 25: Ejemplo de Coordinación de Muestras sobre Machala.



Fuente: Gestión de Diseño Muestral - DINEM.

Resultados

Estratificacion

En la Tabla 15 se presenta el total de UPM disponibles por estrato. La codificación de los dominios geográficos se realiza en función al Clasificador Geográfico Estadístico y los dominios de estratificación cantonales presentados en la Tabla 4.

La codificación del estrato resulta de la concatenación del dominio geográfico, área y un secuencial del total de subconjuntos resultantes de la estratificación, donde los estratos terminados en 1 tienen mejores condiciones de bienestar. Se utiliza el código 9 si no se estratificó, como en el caso de la provincia de Galápagos.

Tabla 15: Resultados estratificación por provincia.

Provincia	Dominio geográfico	Área	Método	Estrato	Nro. UPM
Azuay	01	1	Ninguno	0119	215
		2	Dalenius	0121	144
				0122	233
				0123	136
	30	1	LH Sethi	3011	403
				3012	647
				3013	330
		2	LH Kozak	3021	139
				3022	252
				3023	210
Bolívar	02	1	Kmeans	0211	102
				0212	151
		2	Kmeans	0221	152
				0222	246
Cañar	03	1	LH Sethi	0311	140
				0312	129
				0313	153
		2	Kmeans	0321	112
				0322	286
Carchi	04	1	Geometric	0411	195
				0412	187
		2	Ninguno	0429	224
Cotopaxi	05	1	Dalenius	0511	183



Provincia	Dominio geográfico	Área	Método	Estrato	Nro. UPM
Cotopaxi	05	1	Dalenius	0512	238
				0513	178
		2	Dalenius	0521	252
				0522	374
				0523	253
Chimborazo	06	1	Ninguno	0619	193
		2	Kmeans	0621	301
				0622	239
	31	1	LH Sethi	3111	126
				3112	310
				3113	270
		2	Ninguno	3129	200
El Oro	07	1	LH Sethi	0711	325
				0712	466
				0713	340
		2	Dalenius	0721	211
				0722	150
	40	1	Dalenius	4011	295
				4012	463
				4013	211
Esmeraldas	08	1	Dalenius	0811	178
				0812	243
				0813	147
		2	Kmeans	0821	296
				0822	313
	41	1	Kmeans	4111	290
				4112	374
Guayas	09	1	LH Kozak	0911	2.076
				0912	1.609
				0913	608
		2	Kmeans	0921	450
				0922	388
				0923	398
	42	1	Dalenius	4211	2.856
				4212	3.812
				4213	2.217
Imbabura	10	1	LH Sethi	1011	238



Provincia	Dominio geográfico	Área	Método	Estrato	Nro. UPM
Imbabura	10	1	LH Sethi	1012	467
				1013	510
		2	Kmeans	1021	205
				1022	170
Loja	11	1	Kmeans	1111	128
				1112	145
		2	Kmeans	1121	372
				1122	134
	32	1	LH Kozak	3211	164
				3212	250
				3213	258
		2	Ninguno	3229	131
Los Ríos	12	1	LH Kozak	1211	795
				1212	813
				1213	438
		2	LH Kozak	1221	298
				1222	465
				1223	180
Manabí	13	1	Dalenius	1311	973
				1312	1.195
				1313	577
		2	Kmeans	1321	499
				1322	609
				1323	409
	43	1	Percentiles	4311	457
				4312	457
Morona Santiago	14	1	Ninguno	1419	218
		2	LH Kozak	1421	107
				1422	137
				1423	124
Napo	15	1	Ninguno	1519	163
		2	Ninguno	1529	200
Pastaza	16	1	Ninguno	1619	194
		2	Ninguno	1629	152
Pichincha	17	1	LH Sethi	1711	319
				1712	434
				1713	202



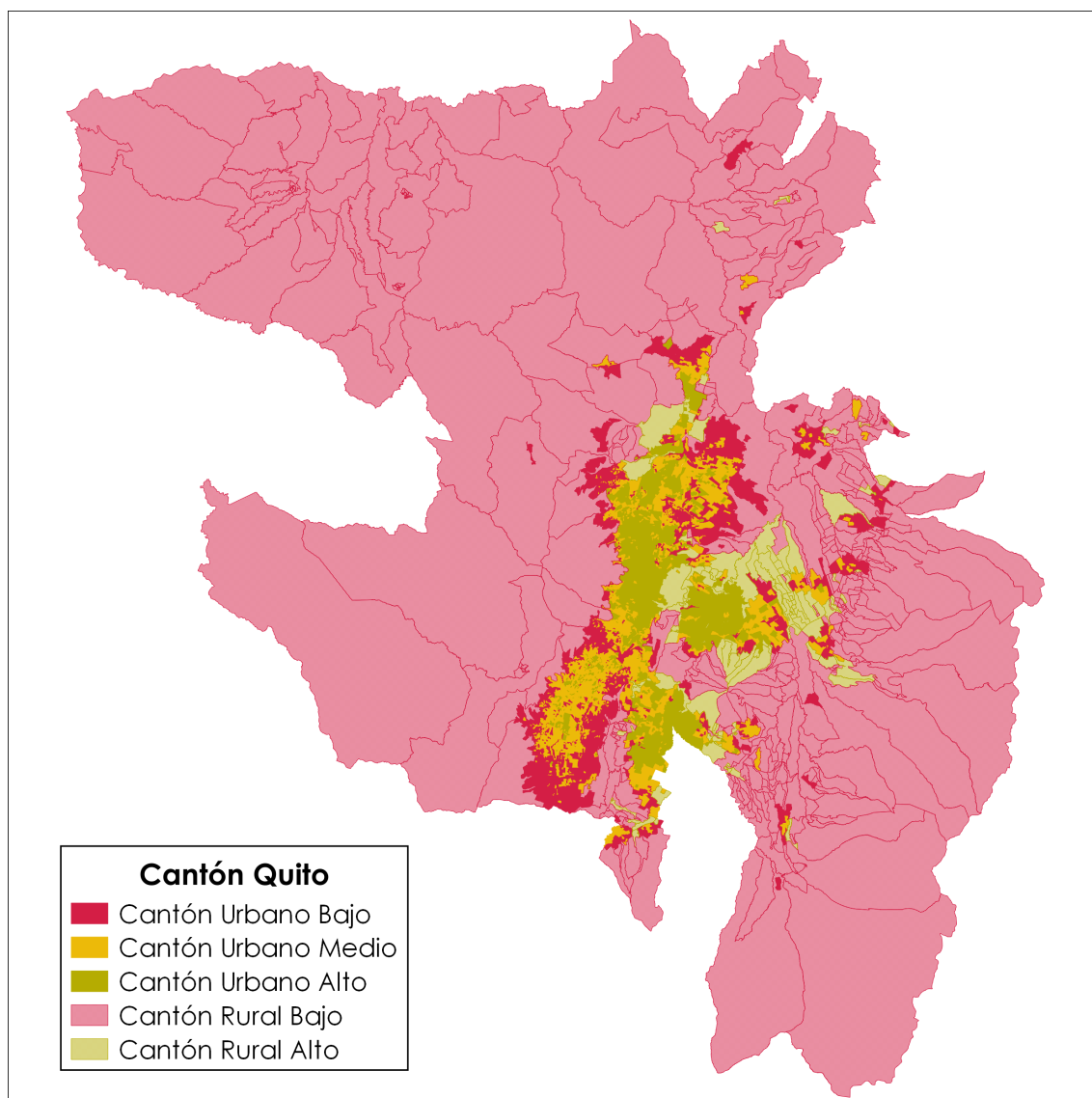
Provincia	Dominio geográfico	Área	Método	Estrato	Nro. UPM
Pichincha	17	2	LH Sethi	1721	133
				1722	146
				1723	129
	33	1	LH Kozak	3311	2.734
				3312	3.491
				3313	2.404
		2	Jarque	3321	381
				3322	121
Tungurahua	18	1	Ninguno	1819	178
		2	LH Sethi	1821	175
				1822	197
				1823	112
	34	1	Dalenius	3411	194
				3412	478
				3413	284
		2	Kmeans	3421	156
				3422	228
Zamora Chinchipe	19	1	Ninguno	1919	154
		2	Ninguno	1929	188
Galápagos	20	9	Ninguno	2099	112
Sucumbíos	21	1	Kmeans	2111	213
				2112	136
		2	Kmeans	2121	101
				2122	222
Orellana	22	1	Ninguno	2219	292
		2	Kmeans	2221	110
				2222	158
Santo Domingo	23	1	Ninguno	2319	154
		2	Ninguno	2329	259
	44	1	Dalenius	4411	435
				4412	554
				4413	266
Santa Elena	24	1	LH Kozak	2411	442
				2412	451
				2413	208
		2	Ninguno	2429	37

Fuente: Gestión de Diseño Muestral - DINEM.

A continuación, se presenta una colección de mapas de la estratificación de varios dominios geográficos.

Cantón Quito

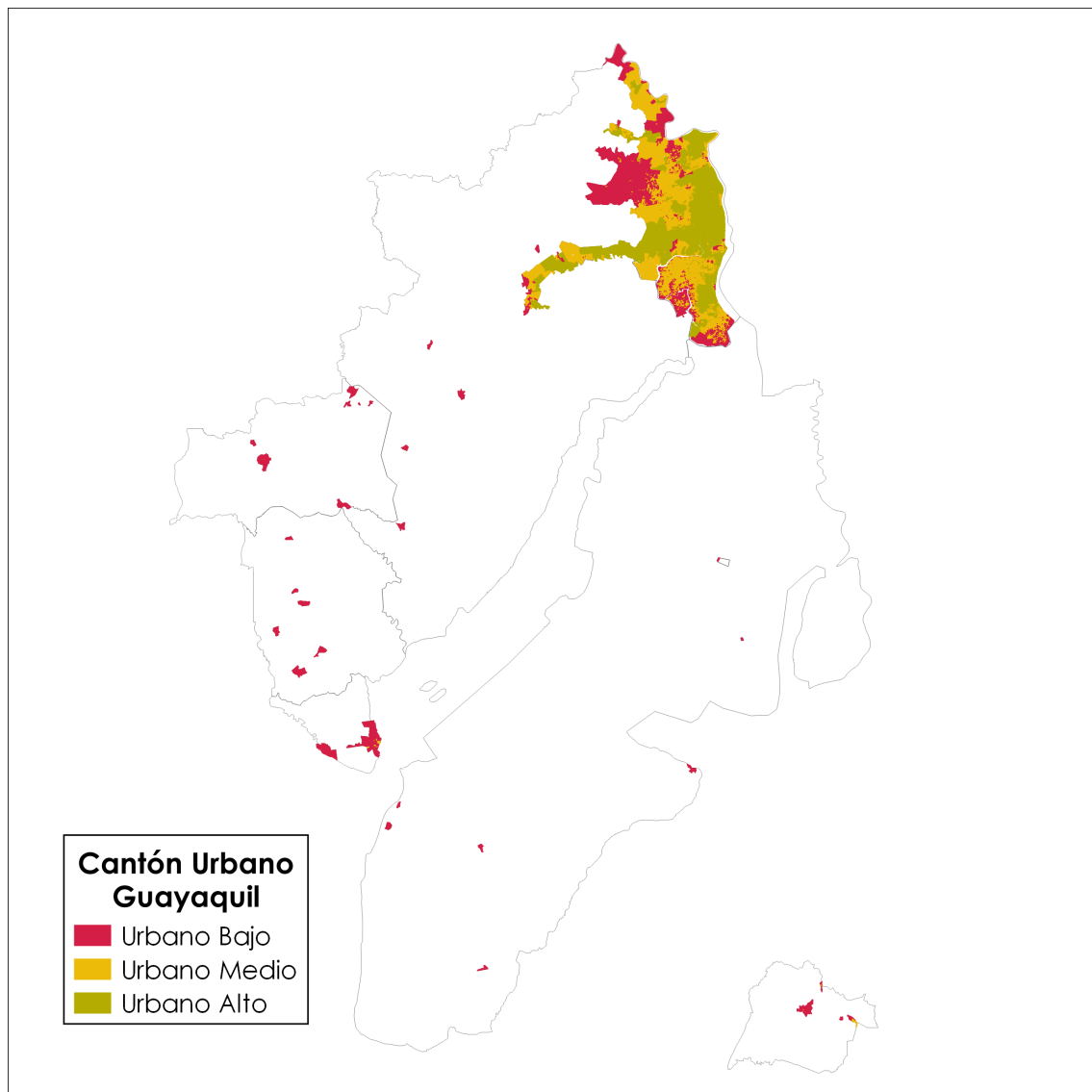
Figura 26: Resultado de la estratificación Cantón Quito.



Fuente: Gestión de Diseño Muestral - DINEM.

Cantón Urbano Guayaquil.

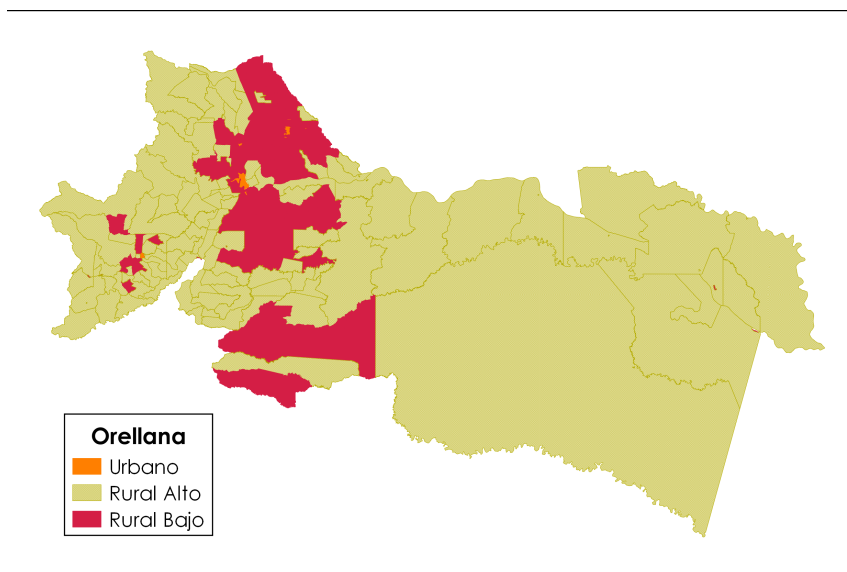
Figura 27: Resultado de la estratificación Cantón Urbano Guayaquil.



Fuente: Gestión de Diseño Muestral - DINEM.

Provincia de Orellana.

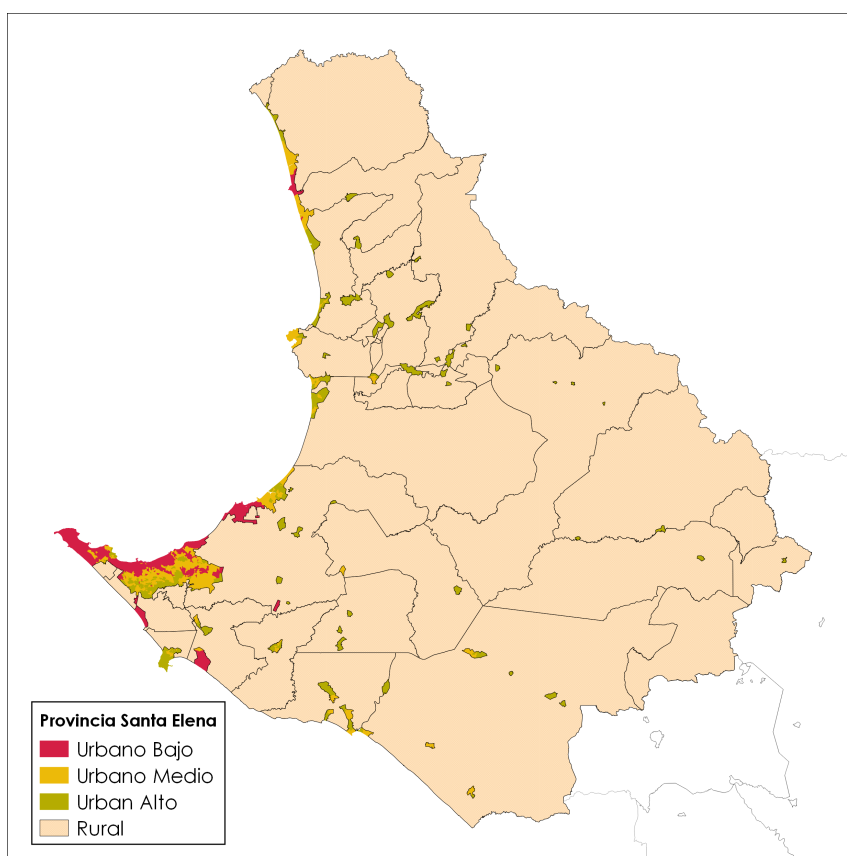
Figura 28: Resultado de la estratificación Orellana.



Fuente: Gestión de Diseño Muestral - DINEM.

Provincia de Santa Elena.

Figura 29: Resultado de la estratificación Santa Elena.

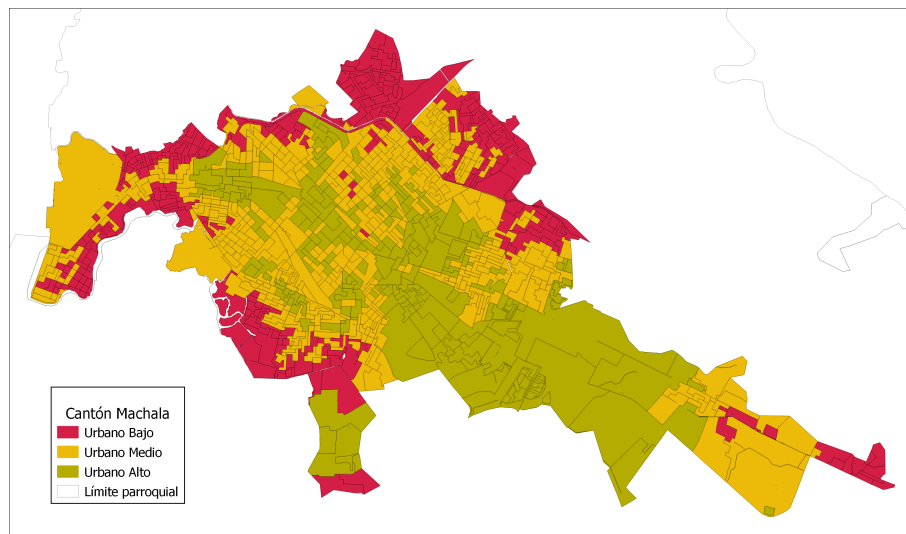


Fuente: Gestión de Diseño Muestral - DINEM.



Cantón Urbano Machala

Figura 30: Resultado de la estratificación en Machala.



Fuente: Gestión de Diseño Muestral - DINEM.



Bibliografía

- Apple, K., & Haken, W. (1976). Every planar map is four colorable. *Bulletin of the American Mathematical Society*, 82(5).
- Baillargeon, S., Rivest, L., & Ferland, M. (2007). *Stratification en enquêtes entreprises: une revue et quelques avancées* (Assemblée annuelle de la SSC) (https://ssc.ca/sites/default/files/survey/documents/SSC2007_S_Baillargeon.pdf).
- Ballin, M., & Barcaroli, G. (2013). Joint determination of optimal stratification and sample allocation using genetic algorithm. *Survey Methodology*, 39(2), 369-393.
- Barcaroli, G. (2014). *SamplingStrata: an R package for the optimization of stratified sampling* (Vol. 61). Journal of Statistical Software.
- Bichot, C. (2007). *Élaboration d'une nouvelle métaheuristique pour le partitionnement de graphe: la méthode de fusion-fission. Application au découpage de l'espace aérien* (Toulouse. INPT 2007).
- Bichot, C., & Siarry, P. (2013). *Graph Partitioning*. Wiley.
- Comisión Económica para América Latina y el Caribe. (2002). *9º Taller Regional: Diseño y construcción de los marcos de muestreo para las encuestas de hogares*. CEPAL.
- Comisión Económica para América Latina y el Caribe. (2023). *Diseño y análisis estadístico de las encuestas de hogares de América Latina*. Metodologías de la CEPAL.
- Department of Economic and Social Affairs. (2005). *Designing household survey samples: Practical guidelines*. United Nations Publications.
- Dirección General de Estadística y Censos. (2015). *Encuesta de Hogares de Propósitos Múltiples*. DIGESTYC.
- Dirección Provincial de Estadística y Censos - Provincia del Neuquén. (2008). *Metodología para la construcción del Marco de Muestreo Provincial Urbano*. INDEC.
- Euler, L. (1736). Solutio problematis and geometrian situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, 8, 128-140.
- Fairbrother, J., Letchford, A., & Briggs, K. (2017). A Two-Level Graph Partitioning Problem Arising in Mobile Wireless Communications.
- Grimaldi, R. (1998). *Matemáticas discreta y combinatoria. Una introducción con aplicaciones*. Wesley.
- Gunning, P., & Horgan, J. M. (2004). A new algorithm for the construction of stratum boundaries in skewed populations. *Survey Methodology*, 30(2), 20-33.
- Guthrie, F. (1880). Note on the coloring of maps. *Proceedings of the Royal Society of Edinburgh*, 10, 727-728.
- Instituto Nacional de Estadística de Honduras. (2015). *Encuesta permanente de hogares y propósitos múltiples*. INE.



- Instituto Nacional de Estadística y Censos. (2022). Clasificador Geográfico Estadístico (https://www.ecuadorencifras.gob.ec/documentos/web-inec/Geografia_Estadistica/Micrositio_geoportal/index.html, Visitado por Última vez en 29.01.2025).
- Jarque, C. M. (1981). A solution to the problem of optimum stratification in multivariate sampling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 30(2), 163-169.
- Jimenez, M. (2011). *Encuestas de hogares*. INE.
- Kozak, M. (2004). Optimal stratification using random search method in agricultural surveys. *Statistics in Transition*, 6(5), 797-8206.
- Lavallée, P., & Hidiroglou, M. A. (1988). On the stratification of skewed populations. *Survey Methodology*, 14(1), 33-43.
- Macqueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. <https://api.semanticscholar.org/CorpusID:6278891>.
- Mitchell, J. (2001). *Branch-and-cut for the k-way equipartition problem* (ISMP 2000. Technical report, Rensselaer Polytechnic Institute) (<https://optimization-online.org/wp-content/uploads/2001/02/288.pdf>).
- Mitchell, J. (2003). Realignment in the National Football League: Did they do it right? *Naval Research Logistics (NRL)*, 50(7), 683-701.
- Nix, B. (2015). *Sample Design and Weights* (2015 CE Survey Microdata Users' Workshop) (<https://www.bls.gov/cex/sampling-methods.pdf>).
- Observatorio Social. (2015). *Metodología de Diseño Muestral de la Encuesta de Caracterización Socioeconómica Nacional*. Ministerio de Desarrollo Social.
- Rosén, B. (1997). On sampling with probability proportional to size. *Journal of Statistical Planning and Inference*, 62(2), 159-191. [https://doi.org/https://doi.org/10.1016/S0378-3758\(96\)00186-3](https://doi.org/https://doi.org/10.1016/S0378-3758(96)00186-3)
- Sethi, V. K. (1963). A note on the optimum stratification of populations for estimating the population means. *Australian and New Zealand Journal of Statistics*, 5(1), 20-33.
- Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical Tools for Designing and Weighting Survey Samples*. Springer.



@InecEcuador



@ecuadorencifras



@ecuadorencifras



INECEcuador

Marco de Muestreo para Encuestas Dirigidas a Hogares 2024

