

**ESTRATIFICACIÓN DEL MARCO DE MUESTREO PARA ENCUESTAS DIRIGIDAS A HOGARES 2024**



Diciembre, 2024

**Título separador**

Subtítulo 01

**ESTRATIFICACIÓN DEL MARCO DE MUESTREO PARA ENCUESTAS DIRIGIDAS A HOGARES 2024**

# Dominios de Estratificación

De acuerdo a un análisis demográfico y económico, se decidió incluir 10 cantones que serán tomados como dominios de estudio. Los cantones se dividen en dos grupos: cantones completos cuyos códigos empiezan en 30 y cantones urbanos cuyos códigos comienzan en 40. El orden de su codificación va en función a la División Político Administrativa del Ecuador (Tabla1). En total se consideró 34 dominios correspondientes a: 5 cantones completos, 5 cantones urbanos, 10 restos de provincias y 14 provincias completas.

|  |  |  |  |
| --- | --- | --- | --- |
| **Tabla 1: 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 |
| El Oro | - | Machala | 40 |
| Esmeraldas | - | Esmeraldas | 41 |
| Guayas | - | Guayaquil | 42 |
| Manabí | - | Manta | 43 |
| Santo Domingo | - | Santo Domingo | 44 |

Se construyeron 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.

# Estratificación

**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).

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”.

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

**DETERMINAR EN CATEGORÍAS MEDIANTE EL ALGORITMO CHAID**

A partir de la base ENEMDU acumulada 2023 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.

Para las variables techo, piso, paredes, procedencia de agua, tubería por dentro del domicilio y servicio higiénico se realizó una asignación de las categorías con bienestar utilizando el algoritmo de clasificación de árboles de decisión (CHAID) de manera particular para cada dominio de estratificación.

Para ello se realizó de forma general una clasificación de variables para 6 grupos correspondientes al cruce entre Región Natural y Área. Luego para cada dominio de estratificación se tomó como base el resultado de la región natural y área a la que pertenece, haciendo cambios en la clasificación siempre y cuando existiera un número de observaciones adecuado y difiera de la clasificación de la región natural por área.

Se decidió clasificar como bienestar a por lo menos una categoría de cada una de las variables clasificadas.

**CATEGORIZAR LAS VARIABLES DE LA BASE DEL CENSO**

# Con la categorización de bienestar obtenida del análisis CHAID se clasifica las variables del censo a nivel de personas, hogares y viviendas.

**CALCULAR LOS INDICADORES ESTABLECIDOS EN LA DIMENSIÓN DE ESTUDIO A NIVEL DE UPM**

# Se agrega la información a nivel de UPM para obtener una matriz de información que consta con las proporciones asociadas con el bienestar para los indicadores utilizados para el proceso de estratificación.

**APLICAR TÉCNICAS DE ESTRATIFICACIÓN POR DOMINIO DE ESTRATIFICACIÓN Y SELECCIONAR EL MÉTODO**

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 (y) es la primera componente del Análisis de Correspondencias principales.

Como nuestro análisis 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. (Anexos)

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:

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 presenta 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.

De acuerdo con la teoría estadística, la mejor estratificación es aquella que minimiza los errores de muestreo de los estimadores por lo cual se considera 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.

* **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.

* **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.

Donde

Es el tamaño relativo del estrato h

Es la varianza de la medida de resumen en el estrato h

* **Estratificación optima**

Estratificación para poblaciones de encuestas basada en la minimización de la siguiente expresión, ligada a la varianza de una estrategia de muestreo estratificada:

Donde

es el número de UPM en el estrato h

es el tamaño de muestra de las UPM,

es el número de UPM en el marco de muestreo y S2x

es la varianza de la medida de resumen en el estrato h.

es la regla de asignación para el tamaño de muestra

La optimización de la función objetivo puede llevarse a cabo de diferentes formas.

**Sethi:** Lavallée e Hidiroglou (1988) utilizaron el algoritmo de optimización de Sethi (1963) para encontrar los valores óptimos.

**Kozak:** Definió un algoritmo iterativo mediante arranques aleatorios para optimizar el proceso de minimización de esta técnica de estratificación.

* **Estratificación geométrica**

En 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:

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

Donde

corresponde a la medición de las P variables de la matriz de información en la k-ésima UPM.

es el vector de medias de la matriz de información en el estrato h.

es una matriz diagonal de tamaño P×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, con el algoritmo se busca minimizar la siguiente función de costos:

Donde

Define un costo fijo

Es el 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).

Para seleccionar el mejor método de estratificación se considera el menor efecto de diseño generalizado.

**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.

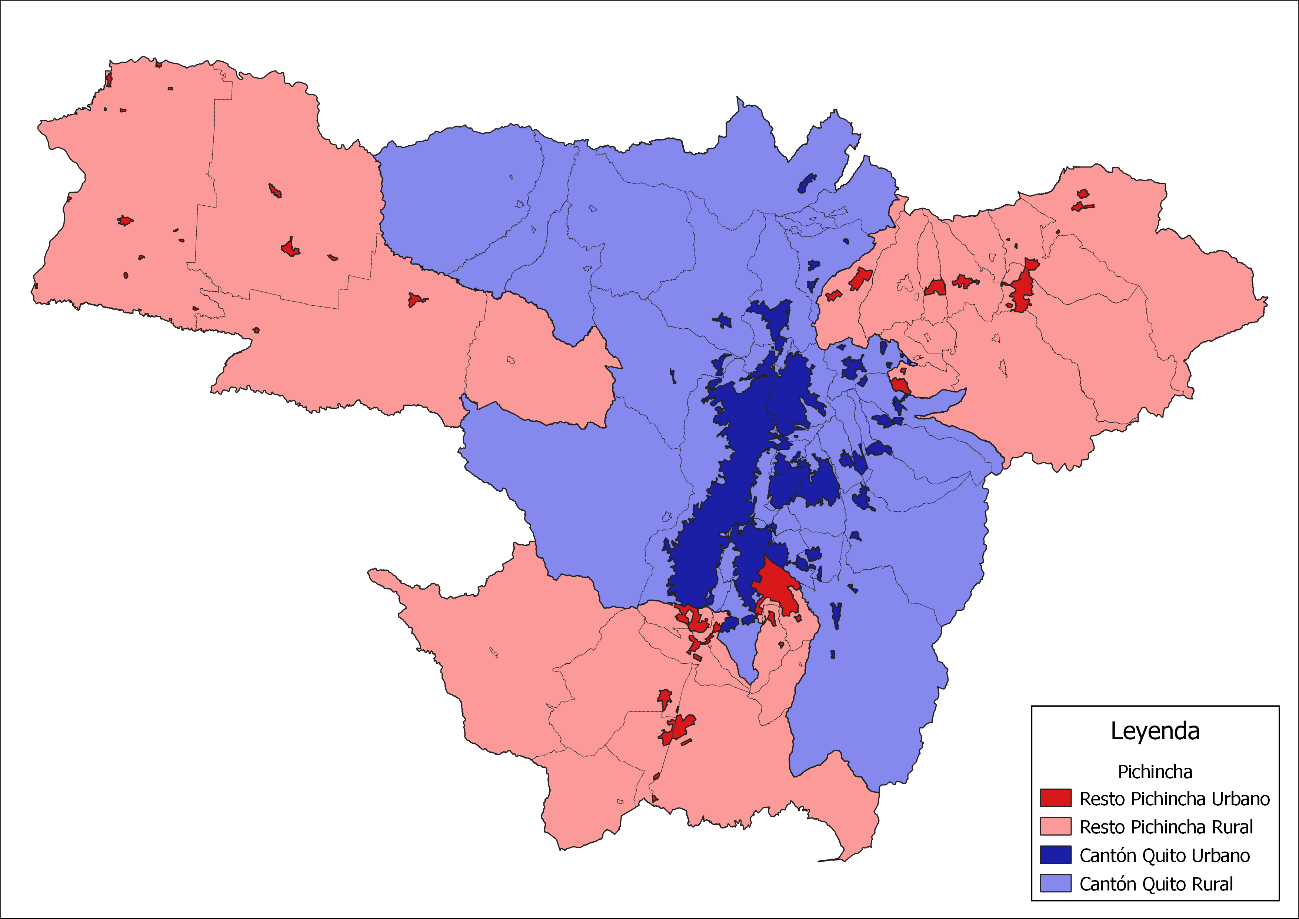
Tras realizar el proceso de estratificación se encontró que en 21 dominios de estratificación se obtuvo al menos un estrato con menos de 100 UPM, por lo que en ellos se realizó una re-estratificación esperando un estrato menos de lo que se esperaba inicialmente.

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, 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 un estrato por lo que unicamente para este dominio de estratificación se decidió colapsar el estrato 1 y 2 de la estratificación original. En la siguiente tabla se presenta el número de estratos por provincia.

|  |  |  |
| --- | --- | --- |
| **Número de estratos por provincia** | | |
| **Provincia** | **Número de Estratos** | **Número de UPM** |
| AZUAY | 10 | 2704 |
| BOLIVAR | 4 | 651 |
| CAÑAR | 5 | 820 |
| CARCHI | 3 | 606 |
| COTOPAXI | 6 | 1477 |
| CHIMBORAZO | 7 | 1638 |
| EL ORO | 8 | 2459 |
| ESMERALDAS | 7 | 1834 |
| GUAYAS | 9 | 14397 |
| IMBABURA | 5 | 1582 |
| LOJA | 8 | 1578 |
| LOS RIOS | 6 | 2987 |
| MANABI | 8 | 5170 |
| MORONA SANTIAGO | 4 | 586 |
| NAPO | 2 | 363 |
| PASTAZA | 2 | 346 |
| PICHINCHA | 11 | 9935 |
| TUNGURAHUA | 9 | 1999 |
| ZAMORA CHINCHIPE | 2 | 342 |
| GALAPAGOS | 1 | 112 |
| SUCUMBIOS | 4 | 670 |
| ORELLANA | 3 | 560 |
| SANTO DOMINGO DE LOS TSACHILAS | 5 | 1667 |
| SANTA ELENA | 4 | 1138 |
| **Total** | **133** | **55621** |

En el dominio de estratificación 242 correspondiente a Provincia 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**



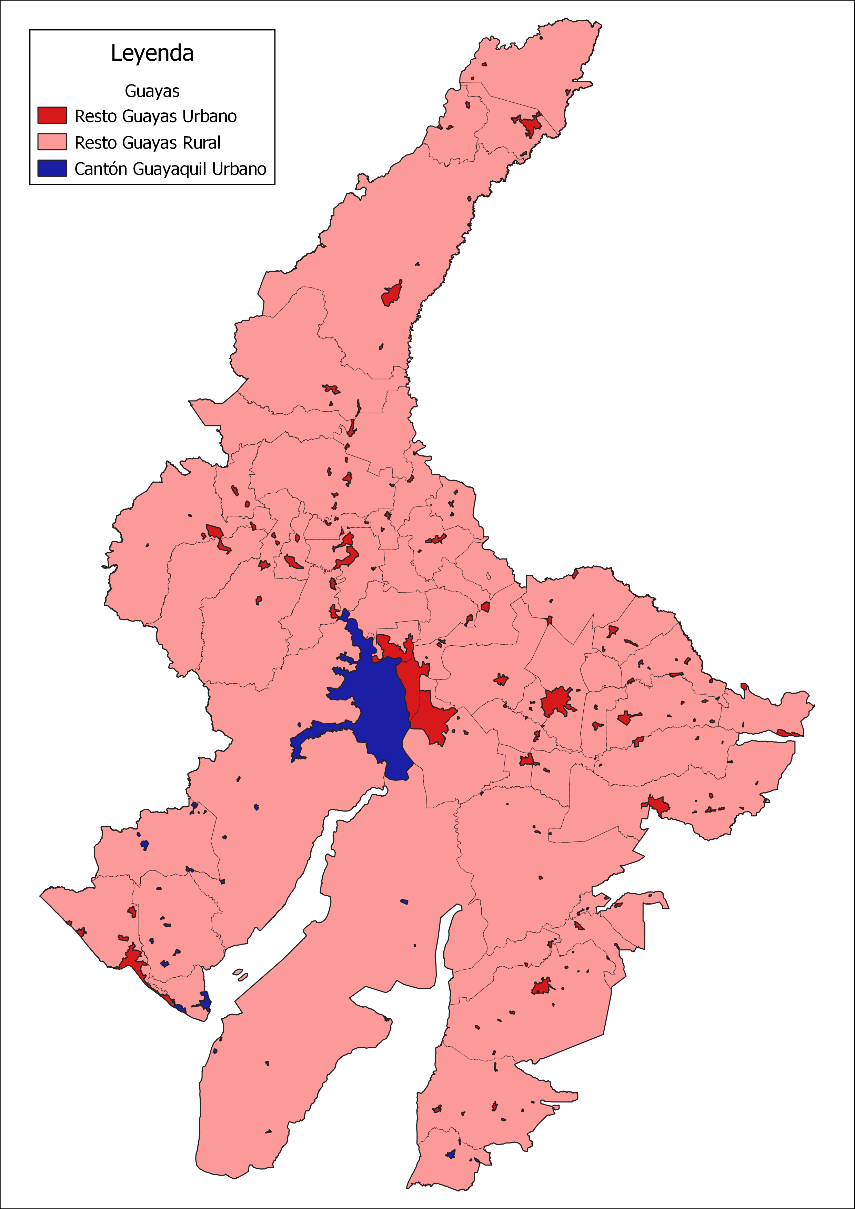
La provincia de Pichincha consta de cuatro dominios de estratificación. (Cantón Quito Urbano, Cantón Quito Rural, Resto Pichincha Urbano, Resto Pichincha Rural)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tabla 3: Resultado estratificación Pichincha** | | | | | | |
| **Provincia** | **Dominio** | **Área** | **Estrato** | **N° UPM Dominio** | **N° UPM Estrato** | **Proporción por dominio** |
| 17 | 171 | 1 | 1711 | 943 | 319 | 33,83 |
| 17 | 171 | 1 | 1712 | 943 | 428 | 45,39 |
| 17 | 171 | 1 | 1713 | 943 | 196 | 20,78 |
| 17 | 172 | 2 | 1721 | 408 | 133 | 32,6 |
| 17 | 172 | 2 | 1722 | 408 | 146 | 35,78 |
| 17 | 172 | 2 | 1723 | 408 | 129 | 31,62 |
| 17 | 331 | 1 | 3311 | 8.082 | 2.664 | 32,96 |
| 17 | 331 | 1 | 3312 | 8.082 | 3.337 | 41,29 |
| 17 | 331 | 1 | 3313 | 8.082 | 2.081 | 25,75 |
| 17 | 332 | 2 | 3321 | 502 | 185 | 36,85 |
| 17 | 332 | 2 | 3322 | 502 | 245 | 48,8 |
| 17 | 332 | 2 | 3323 | 502 | 72 | 14,34 |

Al realizar la estratificación socioeconómica se obtuvieron 12 estratos socioeconómicos, sin embargo, el estrato 3323 presenta menos de 100 UPM por lo cual se consideró re-estratificar el dominio 332 considerando dos estratos.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tabla 4: Resultado re-estratificación Pichincha** | | | | | | |
| **Provincia** | **Dominio** | **Área** | **Estrato** | **N° UPM Dominio** | **N° UPM Estrato** | **Proporción por dominio** |
| 17 | 171 | 1 | 1711 | 943 | 319 | 33,83 |
| 17 | 171 | 1 | 1712 | 943 | 428 | 45,39 |
| 17 | 171 | 1 | 1713 | 943 | 196 | 20,78 |
| 17 | 172 | 2 | 1721 | 408 | 133 | 32,6 |
| 17 | 172 | 2 | 1722 | 408 | 146 | 35,78 |
| 17 | 172 | 2 | 1723 | 408 | 129 | 31,62 |
| 17 | 331 | 1 | 3311 | 8.082 | 2.664 | 32,96 |
| 17 | 331 | 1 | 3312 | 8.082 | 3.337 | 41,29 |
| 17 | 331 | 1 | 3313 | 8.082 | 2.081 | 25,75 |
| 17 | 332 | 2 | 3321 | 502 | 381 | 75,9 |
| 17 | 332 | 2 | 3322 | 502 | 121 | 24,1 |

**Guayas**

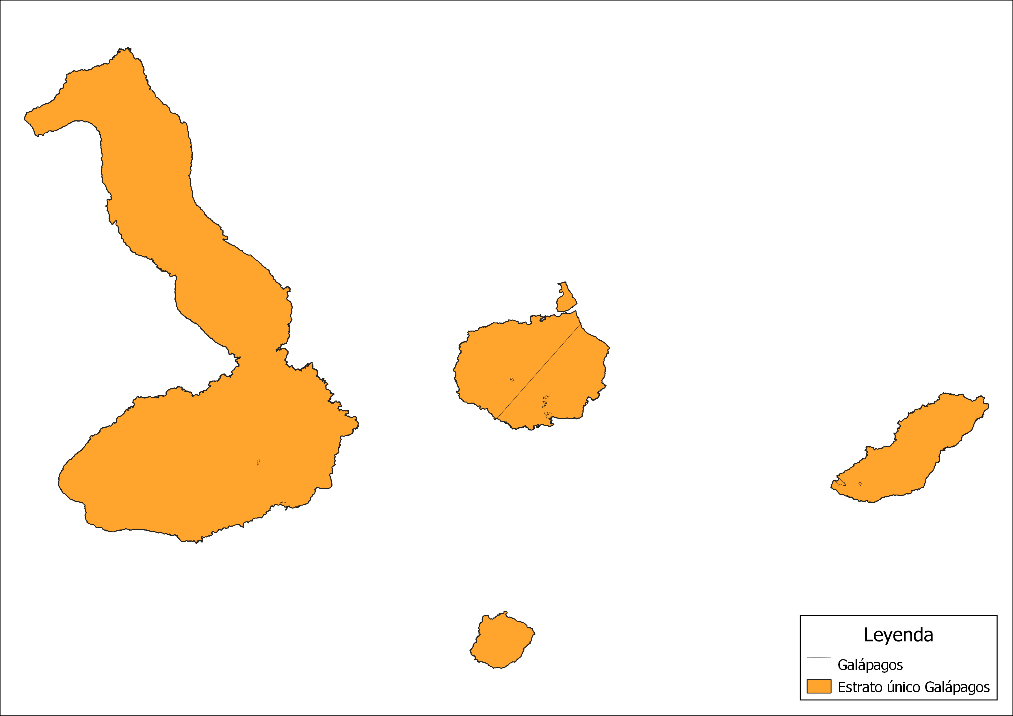


La provincia de Guayas consta de tres dominios de estratificación. (Cantón Guayaquil Urbano, Resto Guayas Urbano, Resto Guayas Rural)

La parte rural del cantón Guayaquil es muy pequeña, por lo cual se estratifica como Guayas rural.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tabla 5: Resultado estratificación Guayas** | | | | | | |
| **Provincia** | **Dominio** | **Área** | **Estrato** | **N° UPM Dominio** | **N° UPM Estrato** | **Proporción por dominio** |
| 9 | 091 | 1 | 0911 | 4292 | 2076 | 48,37 |
| 9 | 091 | 1 | 0912 | 4.292 | 1.609 | 37,49 |
| 9 | 091 | 1 | 0913 | 4.292 | 607 | 14,14 |
| 9 | 092 | 2 | 0921 | 1.236 | 450 | 36,41 |
| 9 | 092 | 2 | 0922 | 1.236 | 388 | 31,39 |
| 9 | 092 | 2 | 0923 | 1.236 | 398 | 32,2 |
| 9 | 421 | 1 | 4211 | 8.869 | 2.852 | 32,16 |
| 9 | 421 | 1 | 4212 | 8.869 | 3.801 | 42,86 |
| 9 | 421 | 1 | 4213 | 8.869 | 2.216 | 24,99 |

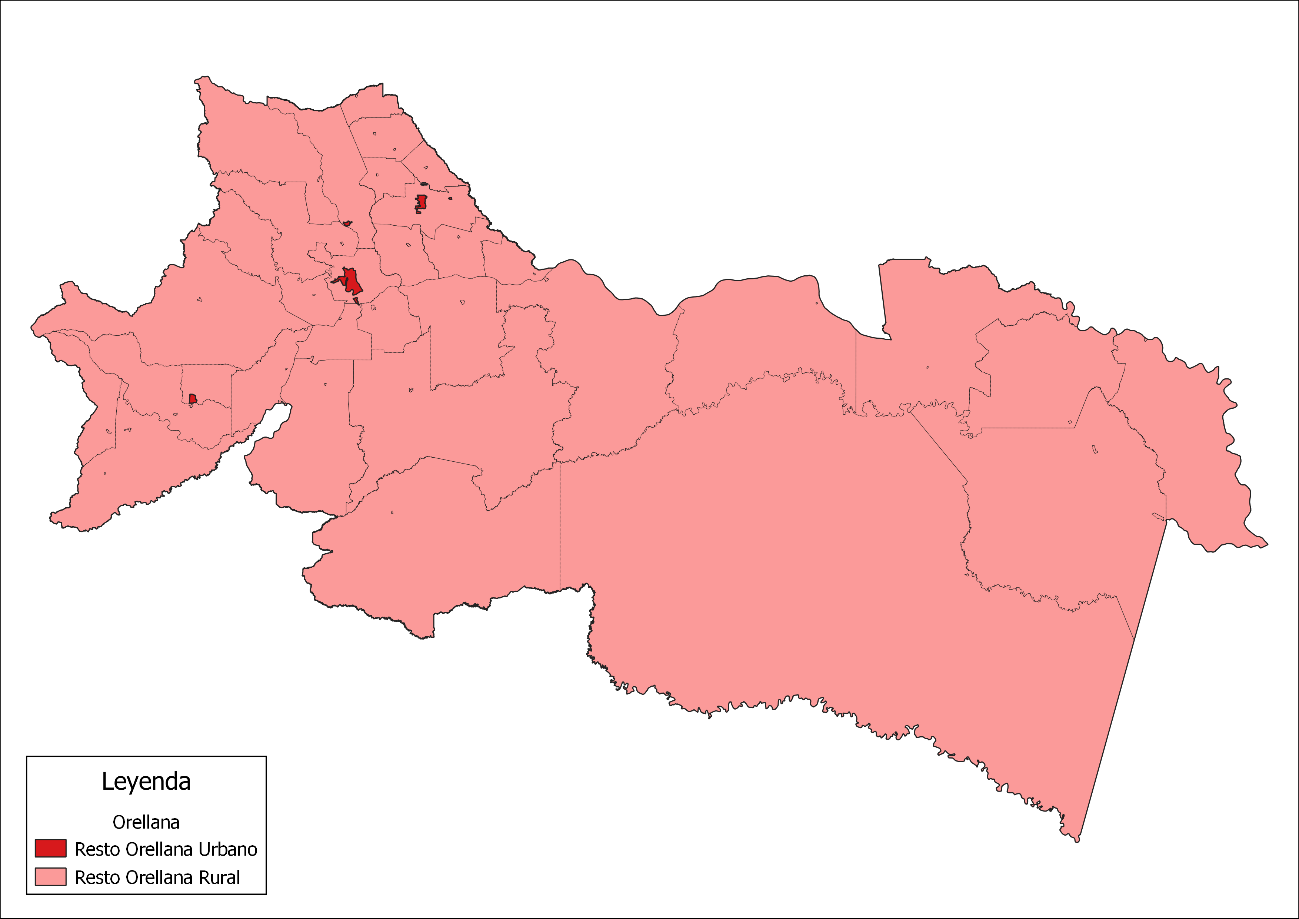
**Galápagos**

****

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

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tabla 6: Resultado estratificación Galápagos** | | | | | | |
| **Provincia** | **Dominio** | **Área** | **Estrato** | **N° UPM Dominio** | **N° UPM Estrato** | **Proporción por dominio** |
| 20 | 209 | 9 | 2099 | 112 | 112 | 100 |

**Orellana**



La provincia de Orellana consta de dos dominios de estratificación. (Orellana Urbano, Orellana Rural)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tabla 7: Resultado estratificación Orellana** | | | | | | |
| **Provincia** | **Dominio** | **Área** | **Estrato** | **N° UPM Dominio** | **N° UPM Estrato** | **Proporción por dominio** |
| 22 | 221 | 1 | 2211 | 292 | 64 | 21,92 |
| 22 | 221 | 1 | 2212 | 292 | 228 | 78,08 |
| 22 | 222 | 2 | 2221 | 268 | 110 | 41,04 |
| 22 | 222 | 2 | 2222 | 268 | 158 | 58,96 |

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 8: Resultado re-estratificación Orellana** | | | | | | |
| **Provincia** | **Dominio** | **Área** | **Estrato** | **N° UPM Dominio** | **N° UPM Estrato** | **Proporción por dominio** |
| 22 | 221 | 1 | 2219 | 292 | 292 | 100 |
| 22 | 222 | 2 | 2221 | 268 | 110 | 41,04 |
| 22 | 222 | 2 | 2222 | 268 | 158 | 58,96 |

