



**UNIVERSIDAD SAN FRANCISCO DE QUITO**

**Colegio de Posgrados**

**Estimadores de Áreas Pequeñas, una aplicación al cálculo de  
proporciones poblacionales**

**Víctor Roberto Morales Oñate**

**Carlos Jiménez Mosquera, Ph.D., Director de Tesis**

Tesis de grado presentada como requisito  
para la obtención del título de Magister en Matemáticas Aplicadas

Quito, marzo de 2014

**Universidad San Francisco de Quito**

**Colegio de Posgrados**

**HOJA DE APROBACIÓN DE TESIS**

**Estimación de Áreas Pequeñas, una aplicación al cálculo de proporciones  
poblacionales**

Víctor Roberto Morales Oñate

Carlos Jiménez Mosquera, Ph.D.  
Director de Tesis y  
Director de la Maestría en Matemáticas Aplicadas

\_\_\_\_\_

Julio Ibarra Fiallo, M.Sc.  
Miembro del Comité de Tesis

\_\_\_\_\_

Carlos Jiménez Mosquera, Ph.D.  
Miembro del Comité de Tesis

\_\_\_\_\_

César Zambrano, Ph.D.  
Decano de la Escuela de Ciencias

\_\_\_\_\_

Víctor Viteri Breedy, Ph.D.  
Decano del Colegio de Posgrados

\_\_\_\_\_

Quito, marzo de 2014

## © DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído la Política de Propiedad Intelectual de la Universidad San Francisco de Quito y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo de investigación quedan sujetos a lo dispuesto en la Política.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo de investigación en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Firma: \_\_\_\_\_

Nombre: Víctor Roberto Morales Oñate

C. I.: 1803590213

Fecha: Quito, marzo de 2014

## **DEDICATORIA**

A todos quienes me enviaron su fuerza

## **AGRADECIMIENTOS**

A Carlos Jiménez y Nicholas Longford, gratitud eterna.

## **RESUMEN**

Este documento evalúa la posibilidad de mejorar la estimación de la proporción de indígenas y adultos mayores en el Ecuador. Se aborda el problema haciendo uso de un conjunto de estimadores y explotando la similaridad espacial de los distritos analizados. El alcance de la aplicación abarca las provincias del Ecuador y los cantones de la provincia de Pichincha para el año 2010. Se usa la Encuesta de Empleo y Subempleo para la estimación directa, el Censo de Población y Vivienda 2001 como información auxiliar y el CPV 2010 para el contraste de resultados.

## **ABSTRACT**

This paper evaluates the possibility of improving the estimate of the proportion of indigenous and elderly people in Ecuador. Dealing with this problem is by using a set of estimators and exploiting the spatial similarity of the districts surveyed addressed. The scope of this work covers the provinces of Ecuador and the cantons of the province of Pichincha in 2010. The survey of Employment and Underemployment is used for direct estimation, the Census of Population and Housing 2001 as auxiliary information and CPV 2010 for contrasting results.

## TABLA DE CONTENIDO

<b>Resumen .....</b>	<b>07</b>
<b>Abstract .....</b>	<b>08</b>
<b>Introducción.....</b>	<b>11</b>
<b>Revisión de literatura .....</b>	<b>13</b>
<b>Metodología y diseño .....</b>	<b>16</b>
<b>3.1 Notación y nociones de estimadores de áreas pequeñas.....</b>	<b>16</b>
<b>3.2 Similaridad Espacial .....</b>	<b>18</b>
<b>3.3 Estimadores.....</b>	<b>19</b>
<b>Análisis y resultados .....</b>	<b>20</b>
<b>4.1 Contexto.....</b>	<b>21</b>
<b>4.1.1 Población Indígena en el Ecuador.....</b>	<b>21</b>
<b>4.1.2 Población Indígena en Pichincha .....</b>	<b>23</b>
<b>4.1.3 Adulto mayor en los cantones de la provincia del Guayas .....</b>	<b>24</b>
<b>4.2 Selección del estimador de área pequeña .....</b>	<b>25</b>
<b>4.2.1 Estimador para provincias del Ecuador.....</b>	<b>25</b>
<b>4.2.2 Estimador para los cantones de Pichincha.....</b>	<b>28</b>
<b>4.2.3 Estimador para los cantones de Guayas: Adulto mayor. ....</b>	<b>31</b>
<b>4.3 Resultados del estimador .....</b>	<b>33</b>
<b>Conclusiones y recomendaciones.....</b>	<b>35</b>
<b>Conclusiones.....</b>	<b>35</b>
<b>Recomendaciones.....</b>	<b>36</b>
<b>Referencias .....</b>	<b>37</b>
<b>ANEXO A: código de matriz de similaridad espacial .....</b>	<b>38</b>
<b>ANEXO B: ACRÓNIMOS DE CANTONES - GUAYAS .....</b>	<b>41</b>

## TABLAS

Tabla 1: Algoritmo de generación de matriz de similaridad espacial .....	18
Tabla 2: Población por autoidentificación étnica – población en miles de habitantes. Excluye a Galápagos y Zonas no Delimitadas .....	21
Tabla 3: Comparaciones de conjuntos de estimadores para provincias .....	25
Tabla 4: Comparaciones de conjuntos de estimadores para cantones .....	29
Tabla 5: Matriz de distancias de los estimadores de áreas pequeñas .....	33
Tabla 6: Matriz de distancias de los estimadores de áreas pequeñas .....	33

## FIGURAS

Figura 1: Proporción de población indígena provincial .....	21
Figura 2: Proporción de población indígena cantonal .....	23
Figura 3: Suma de errores absolutos por provincias del Ecuador .....	27
Figura 4: Suma de errores absolutos por cantones de la provincia de Pichincha. ....	30

## INTRODUCCIÓN

Contar con información sociodemográfica actualizada es un recurso trascendente en el proceso de planificación y de toma de decisiones. Generalmente se cuenta con información detallada de la población a nivel de parroquias, cantones y provincias del Ecuador de manera decenal a través de los Censos de Población y Vivienda.

Una de las herramientas de recolección de datos de mayor frecuencia temporal es la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU). Esta encuesta carece de representatividad estadística tanto a nivel de las provincias de la Amazonía así como en los cantones distintos de Quito, Guayaquil, Cuenca, Ambato y Machala. En diciembre del 2010, la ENEMDU contaba con información de 82.774 personas en su muestra total. Este tamaño de muestra hace que la ENEMDU pueda ser un insumo de información importante a ser explotado en campos más allá del cálculo de indicadores de empleo.

El esfuerzo por obtener información de áreas pequeñas ha sido principalmente realizado por demógrafos. Son ellos quienes usan una gama de métodos indirectos para estimación de áreas pequeñas en años posteriores al levantamiento de datos censales. Generalmente estos métodos no incluyen muestreo y suelen enfocarse al cálculo población por área (urbano - rural), grupos de edad, sexo, migración, entre otros. Sin embargo, en la planificación de un país es necesario tener información de variables más específicas como pobreza, desnutrición, etnia, entre otros (Rao, 2003).

El presente trabajo plantea abordar el problema de representatividad estadística para las provincias de la Amazonía así como los cantones de la provincia de Pichincha y el Guayas para diciembre del año 2010 mediante el uso de una aplicación de Estimadores de Áreas Pequeñas. En primer lugar se plantea el uso de simulaciones de modo que se seleccione un estimador para el cálculo de la proporción de indígenas. Una vez elegido el estimador, se

procede a aplicar la metodología en las áreas descritas. Se usa como referencia el año 2010 debido a que se cuenta con información del Censo de Población y Vivienda para contrastar la eficiencia de los resultados obtenidos en este trabajo.

## REVISIÓN DE LITERATURA

Una manera de clasificar las formas de realizar estimaciones de áreas pequeñas en grandes líneas es a través de métodos directos, indirectos y basados en modelos (Rao, 2003).

1. Los métodos directos hacen uso de la información de la muestra tomando en cuenta el diseño muestral y sus pesos.
2. Los métodos indirectos usan información auxiliar de registros administrativos y censos para realizar estimaciones aprovechando dicha información en la estimación de áreas pequeñas. En este grupo se encuentra el estimador compuesto.
3. Los métodos basados en modelos plantean variables asociadas a la variable objetivo. Hace uso de información directa y auxiliar para sus fines.

De entre estas aproximaciones, los métodos usados en el presente trabajo se basan en el uso de estimadores indirectos a través de estimadores compuestos. Se presenta en la tabla 1 una recolección de literatura de aplicaciones de áreas pequeñas.

Tabla 1: Literatura de aplicaciones de áreas pequeñas

<b>Año</b>	<b>Autor</b>	<b>Método</b>
2010	Sakshaug and Raghunathan	I
2000	Citro y Kalton	M
2009	Herrador	M
2011	Goerndt	I,M
2011	Coelho	M
2007	Longford	I,M
2010	Datta	M
2012	Longford	M
2008	Longford	M
2013	Benedetti	M
2013	Schmid	M
2009	Salvati	M
2011	Longford	I
2010	Longford	M

I: Indirecto

M: Basado en modelos

En la literatura se especifica las aplicaciones que usan el método indirecto y el basado en modelos para estimadores de áreas pequeñas. Las aplicaciones presentadas en la tabla 1 abordan temáticas de diferente índole. Se tiene estimaciones de áreas pequeñas de ingresos, pobreza, empleo, forestales, entre otros. Varios trabajos se enfocan en la propuesta mejoras a los métodos tradicionales planteados principalmente en (Rao, 2003).

De entre las aplicaciones recopiladas, “*Small –area estimation with spatial similarity*” (Longford, 2009) es la que provee una serie de estimadores más apropiados para los fines de este estudio.

El Banco Mundial es una de las instituciones internacionales que ha realizado esfuerzos por tener mediciones de áreas pequeñas en cuanto a pobreza. Se han enfocado en el mapeo de indicadores de pobreza y han generado un software para estos fines, PovMap.

Una de las experiencias más notorias en el Ecuador en cuanto a estimaciones de áreas pequeñas es el estudio de desnutrición realizado por el Ministerio Coordinador de Desarrollo Social (MCDS) (MCDS, 2010). En este estudio se realiza una estimación de la desnutrición crónica para niños menores a cinco años. La fuente principal de información fue la Encuesta de Condiciones de Vida 2005-2006 y se usó información auxiliar de registros administrativos y el CPV 2001.

Tanto el Banco Mundial como el Ministerio de Desarrollo Social usaron estimaciones de áreas pequeñas basadas en modelos. El uso de modelos en la estimación de áreas pequeñas puede ser de gran ayuda si se cuenta con estado del arte o evidencia empírica que minimice problemas de especificación. Sin duda la pobreza y la desnutrición crónica cumplen con este requerimiento sobremanera. Sin embargo, cuando se trata de la estimación de indicadores que

no cuentan con los mencionados requisitos, la estimación basada en modelos podría no ser la mejor alternativa.

En diciembre de cada año la ENEMDU tiene cobertura nacional con representatividad estadística para el área urbana y rural de las provincias de la Sierra, Costa y Amazonía. La Amazonía es tomada como un solo dominio de estudio y se excluye a la Región Insular (InecPedia, 2014). Los indicadores generados a partir de la ENEMDU suelen ser obtenidos mediante estimación directa o haciendo uso del diseño muestral de la encuesta. Esto funciona exclusivamente para los dominios en los cuales la encuesta es representativa. No obstante, los requerimientos de información para la planificación nacional y local suelen ir más allá de esta limitante.

Este trabajo se centra en la estimación de áreas pequeñas mediante la comparación de la estimación directa y compuesta para las áreas en la que la ENEMDU no es representativa. Para ello, se hace uso del CPV 2001 como información auxiliar, la ENEMDU 2010 como estimación directa y el CPV 2010 para marco referencial de comparación de los resultados. La estimación se centra en el cálculo de la proporción de indígenas para todas las provincias del Ecuador y evaluar sus resultados. Asimismo se realiza una aplicación para los 9 cantones de la Provincia de Pichincha y la proporción de adultos mayores en los cantones de la provincia del Guayas.

## METODOLOGÍA Y DISEÑO

### 3.1 Notación y nociones de estimadores de áreas pequeñas

Se tiene una población  $\mathcal{P}$  conformada por  $D$  áreas pequeñas y estas forman una partición  $\mathcal{P}_d$ ,  $d = 1, \dots, D$  y la variable  $Y$  está definida para todas las subpoblaciones (distritos) de  $\mathcal{P}$ . El objetivo es calcular el estadístico  $\theta_d$  de  $Y$  para cada distrito  $d$  (Longford, 2005).

Se asume que  $\theta_d$  se puede calcular por una función  $\Theta$  que puede ser usada para evaluar cualquier distrito de  $\mathcal{P}$ ,  $\theta_d = \Theta \mathcal{P}_d$ .

Una muestra  $\mathcal{S}$  de  $\mathcal{P}$  tiene una partición conforme con  $\mathcal{P}_d$  en las subpoblaciones tal que  $\mathcal{S}_d = \mathcal{S} \cap \mathcal{P}_d$ . El estimador directo  $\hat{\theta}_d$  de  $\theta_d$  depende solamente de los valores de  $Y$  y del diseño muestral. Se asume que  $\theta_d$ ,  $d = 1, \dots, D$  se relacionan de modo que  $\theta_d = \Theta \mathcal{S}_d$ . Un ejemplo de  $\Theta$  y  $\Theta$  es la media poblacional y la media muestral respectivamente.

Los tamaños de las poblaciones de los distritos se denotan por  $N_d$  y  $n_d$  correspondientemente (siendo  $N$  y  $n$  sus totales). Las fracciones dentro de las subpoblaciones  $f_d = \frac{n_d}{N_d}$  no necesariamente son idénticas. Se asume que  $\theta_d$  es un estimador insesgado, entonces  $v_d = \text{var } \theta_d$ ; se asume que  $v_d$  es conocida.

Se define la media y la varianza para poblaciones finitas de un conjunto  $\theta_1, \dots, \theta_D$ :

$$\theta = \frac{1}{D} \theta_1 + \dots + \theta_D$$

$$\sigma_0^2 = \frac{1}{D} \sum_{d=1}^D \theta_d - \theta^2$$

Se denota con subíndices  $\mathcal{S}$  de  $\mathcal{D}$  a los resultados de muestrales y por distritos respectivamente,  $v_d = \text{var}_{\mathcal{S}} \theta_d$  y  $\sigma_0^2 = \text{var}_{\mathcal{D}} \theta_d$ .

Se considera un conjunto de estimadores  $\theta_d^h$ ,  $h = 1, \dots, H$  para  $\theta_d$  y su combinación convexa:

$$\theta_d = \sum_{h=0}^H b_d^h \theta_d^h \quad (1)$$

En (1) se estiman los coeficientes  $b_d^1, \dots, b_d^H$  y  $b_d^0 = 1 - b_d^1 - \dots - b_d^H$  para los cuales el error medio cuadrático  $ECM \theta_d; \theta_d = E_{\mathcal{S}} (\theta_d - \theta_d^2) \theta_d$  se minimiza. Se asume que el estimador  $\theta_d = \theta_d^0$  es insesgado.

Si  $H = 1$  en (1) se tiene el estimador compuesto básico:

$$\theta_d = (1 - b_d) \theta_d + b_d \theta \quad (2)$$

Donde se supone que los distritos son fruto de un muestreo aleatorio simple estratificado  $SSRS_d$ ,  $b_d = \frac{1}{1+n_d\omega}$  y  $\omega$  es el ratio de las varianzas entre y dentro de los distritos.

### 3.2 Similaridad Espacial

El supuesto de similaridad entre áreas pequeñas cercanas es bastante intuitivo y ha sido abarcado en aplicaciones de varias índoles como estadística espacial, geoinformación y estimadores de áreas pequeñas.

Se definen entonces un conjunto de estimadores de  $\theta_d$  libres de supuestos en cuanto a distribución y se asume una estructura natural de correlación relacionando la distancia. Se asume que una función de distancia  $\xi d, d'$  está definida para cualquier pareja de distritos  $d$  y  $d'$ . Esta función es simétrica, no negativa e igual a cero cuando  $d = d'$  (Longford, 2009).

A continuación se presenta el algoritmo para generar la matriz de similaridad espacial

Tabla 1: Algoritmo de generación de matriz de similaridad espacial

<p>Entrada: Matriz indicatriz simétrica <math>M_{d \times d}</math> de vecinos (1 si es vecino de <math>d</math>, 0 caso contrario)</p> <p>Salida: Matriz de similaridad espacial</p>
<p>1) Seleccionar un vector fila <math>m_{i \times d}</math> de la matriz de entrada y generar un vector <math>vei</math> con las posiciones de los vecinos de <math>m_{i \times d}</math>.</p>
<p>2) Seleccionar una sub-matriz con las posiciones de las filas calculadas en uno con todas las columnas de 1)</p>
<p>3) Si la longitud del vector calculado en 1) es mayor a uno, generar un vector indicatriz de la sub-matriz calculada en 2) (<math>vei2</math>: vecinos de los vecinos)</p>
<p>4) Calcular: <math>m_{i \times d} + (\max M_{d \times d} + vei2_{1 \times d}) * n_{i \times d} * vei2_{1 \times d}</math>, donde <math>n_{i \times d}</math> es el vector indicatriz de los NO vecinos de <math>m_{i \times d}</math></p>

5) Repetir el proceso para todos las filas de $M_{d \times d}$ para obtener $M'_{d \times d}$ .
6) Mientras $\min M'_{d \times d} = 0$ , repetir los pasos del 1 al 5.
7) Llenar con ceros la diagonal de la matriz generada en 6)

Para cada distrito, se genera un *anillo*  $- h$  como la población circundante a  $d$ ,

$$\mathcal{P}_d^h = \sum_{d'; \xi_{d,d'}=h} \mathcal{P}_{d'}$$

que en el caso muestral es  $\mathcal{S}_d^h$ .

### 3.3 Estimadores

El aporte más importante de (Longford, 2009) es un conjunto de estimadores que toman en cuenta la matriz de similaridad espacial. Es decir que el estimador compuesto para cada distrito toma en cuenta la información de los  $h - \text{anillos}$  alrededor de  $d$ . El estimador de  $\theta_d^h$  es,

$$\theta_d^h = \frac{1}{n_d} \sum_{d' \in \mathbb{d}_d^h} n_{d'} \theta_{d'}$$

donde  $\mathbb{d}_d^h$  es el conjunto de distritos que forman  $\mathcal{P}_d^h$ . De esta manera (1) queda definida completamente. Reescribiendo (1) tenemos:

$$\theta_d = \sum_{h=0}^H b_d^h \theta_d^h = 1 - \mathbf{b}_d^T \mathbf{1} \theta_d + \mathbf{b}_d^T \boldsymbol{\psi}_d$$

donde  $\mathbf{b}_d = b_d^1, \dots, b_d^H$ ,  $\boldsymbol{\psi}_d = \theta_d^1, \dots, \theta_d^H$ ,  $\theta_d^1 = \theta_d$  y los coeficientes  $b_d^0 + \dots + b_d^H = 1$ . Tanto  $\mathbf{b}_d$  como  $\boldsymbol{\psi}_d$  tienen longitud  $H$ .

En cuanto a la estimación de  $\sigma_0^2$ ,  $\sigma_h^2$  y la covarianza con los  $h$  – anillos ( $\gamma_h$ ) se tienen los siguientes estimadores:

$$\sigma_0^2 = \frac{S_0}{D} - \frac{1}{D} \sum_{d=1}^D \left( 1 - 2 \frac{q_d}{q_+} \right) v_d - v$$

donde  $S_0 = \sum_{d=1}^D \theta_d - \theta^2$ ,  $q_d$  son los coeficientes de  $\theta = \frac{q_1 \theta_1 + \dots + q_D \theta_D}{q_+}$  y  $q_+$  es su total.

Sea  $m_d^h$  es número de distritos en el  $h$  – anillo del distrito  $d$  y  $m_+^h = m_1^h + \dots + m_D^h$ , entonces:

$$\sigma_h^2 = \frac{1}{m_+^h} \sum_{d=1}^D \sum_{d' \in \mathbb{d}_d^h} \left( \theta_d - \theta_{d'} \right)^2 - 2 \sum_{d=1}^D m_d^h v_d$$

de modo que el estimador insesgado de la varianza es  $\sigma_h^2 = 2 \sigma_0^2 - \gamma_h$ , entonces,

$$\gamma_h = \sigma_0^2 - \frac{1}{2} \sigma_h^2$$

De este modo se incorpora el criterio de los  $h$  – anillos a la estimación de áreas pequeñas.

Análisis y resultados



Provincia	Autoidentificación Étnica								Total
	Indígena	Afro	Negrx	Mulatx	Montubix	Mestizx	Blancx	Otrx	
AZ AZUAY	17.6	10.8	0.9	3.9	2.9	637.9	36.7	1.3	712.1
BO BOLIVAR	46.7	1.2	0.2	0.6	2.1	127.8	4.9	0.2	183.6
CÑ CAÑAR	34.2	4.2	0.5	1.3	2.4	172.6	9.6	0.4	225.2
CA CARCHI	5.6	6.8	1.7	2.1	0.4	142.9	4.7	0.2	164.5
CO COTOPAXI	90.4	4.8	0.4	1.6	7.3	294.8	9.3	0.5	409.2
CH CHIMBORAZO	174.2	3.6	0.2	1.2	1.2	267.9	10.0	0.4	458.6
EL EL ORO	4.1	24.2	4.7	12.6	16.9	489.8	46.8	1.7	600.7
ES ESMERALDAS	15.0	123.1	56.6	54.9	13.0	238.6	31.3	1.6	534.1
GU GUAYAS	46.2	204.3	36.4	111.4	411.0	2461.7	355.3	19.1	3645.5
IM IMBABURA	102.6	12.2	4.1	5.2	1.2	261.7	10.8	0.5	398.2
LO LOJA	16.5	8.3	0.6	1.8	3.2	404.9	13.2	0.5	449.0
LR LOS RIOS	5.0	30.3	7.1	10.7	272.7	411.9	38.5	2.0	778.1
MA MANABI MORONA	2.5	62.2	8.6	11.4	262.7	954.2	64.3	3.9	1369.8
MO SANTIAGO	71.5	1.1	0.2	0.6	0.3	68.9	4.6	0.8	147.9
NA NAPO	58.8	0.8	0.2	0.7	0.6	39.5	2.8	0.2	103.7
PA PASTAZA	33.4	0.6	0.2	0.5	0.3	46.4	2.4	0.1	83.9
PI PICHINCHA	137.6	65.4	12.8	38.4	34.6	2115.0	163.2	9.4	2576.3
TU TUNGURAHUA ZAMORA	62.6	4.7	0.3	2.2	2.3	414.5	17.4	0.7	504.6
ZA CHINCHIPE	14.2	0.8	0.2	0.3	0.2	73.4	1.9	0.3	91.4
SU SUCUMBIOS	23.7	4.3	2.3	3.8	1.7	132.4	8.0	0.4	176.5
OR ORELLANA SANTO	43.3	2.6	1.7	2.4	1.6	78.4	6.0	0.3	136.4
SD DOMINGO	6.3	16.4	3.8	8.2	9.0	298.2	25.1	1.0	368.0
SE SANTA ELENA	4.2	20.6	1.5	4.2	15.2	244.3	11.4	7.4	308.7

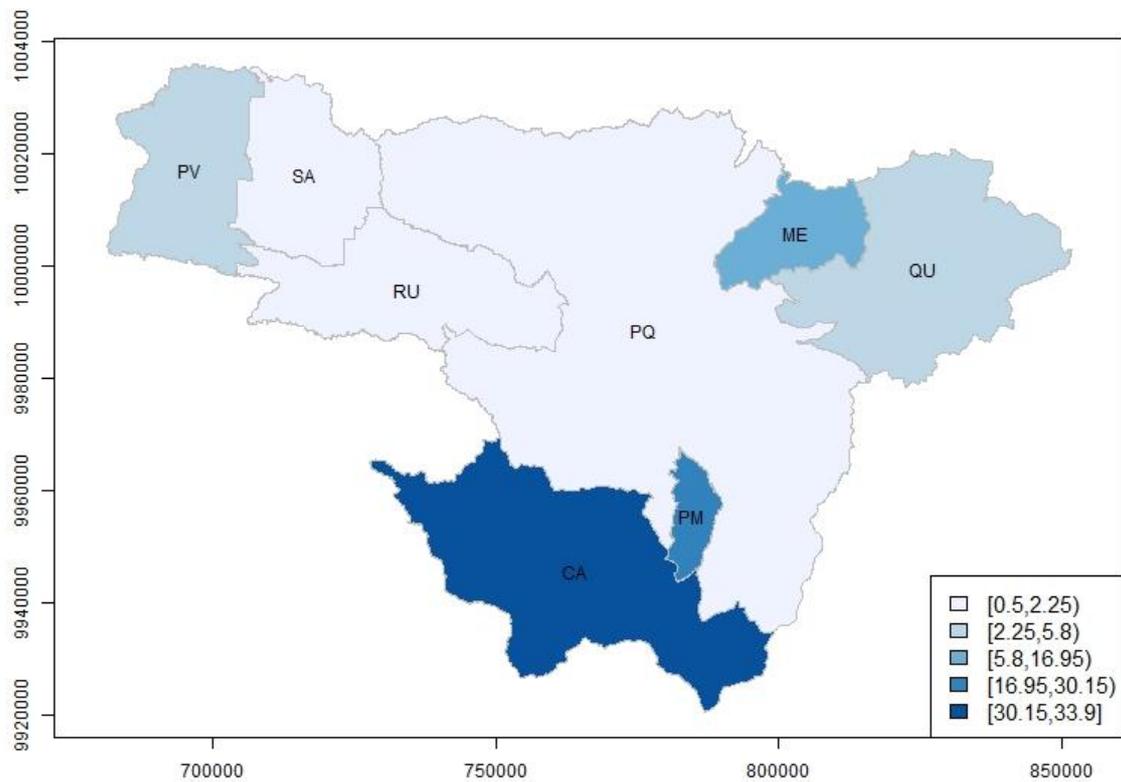
Cabe mencionarse la diferencia que se establece al analizarse los datos en valores absolutos y en proporciones. Por ejemplo, la provincia de Pichincha tiene aproximadamente 138 mil habitantes que se autoidentifican como indígenas pero en términos relativos solo muestra un 5,3% de su composición por etnias. Napo por su parte es la provincia que tiene mayor presencia indígena a nivel Nacional (56,7%).

Como se mencionó anteriormente, las provincias de la Amazonía tienen representación alta de indígenas. Esto aumenta a la relevancia de contar con indicadores periódicos a nivel provincial en lo referente a la proporción de indígenas.

#### 4.1.2 Población Indígena en Pichincha

Según el CPV 2010, el 5% de la población se auto identifica como indígena en Pichincha. El siguiente mapa muestra la distribución cantonal de este valor provincial.

Figura 2: Proporción de población indígena cantonal



En el caso de la proporción de indígenas en los cantones de la provincia de Pichincha, este indicador tiene un rango de 0,5% a 34%. Su rango es menor que el caso analizado a través de provincias pero sigue habiendo diferencias marcadas.

La presencia de la población indígena cantonal en orden descendente es: Cayambe (CA), Pedro Moncayo (PM), Mejía (ME), Quito (QU), Pedro Vicente Maldonado (PV), Rumiñahui (RU), San Miguel de los Bancos (SA) y Puerto Quito (PQ).

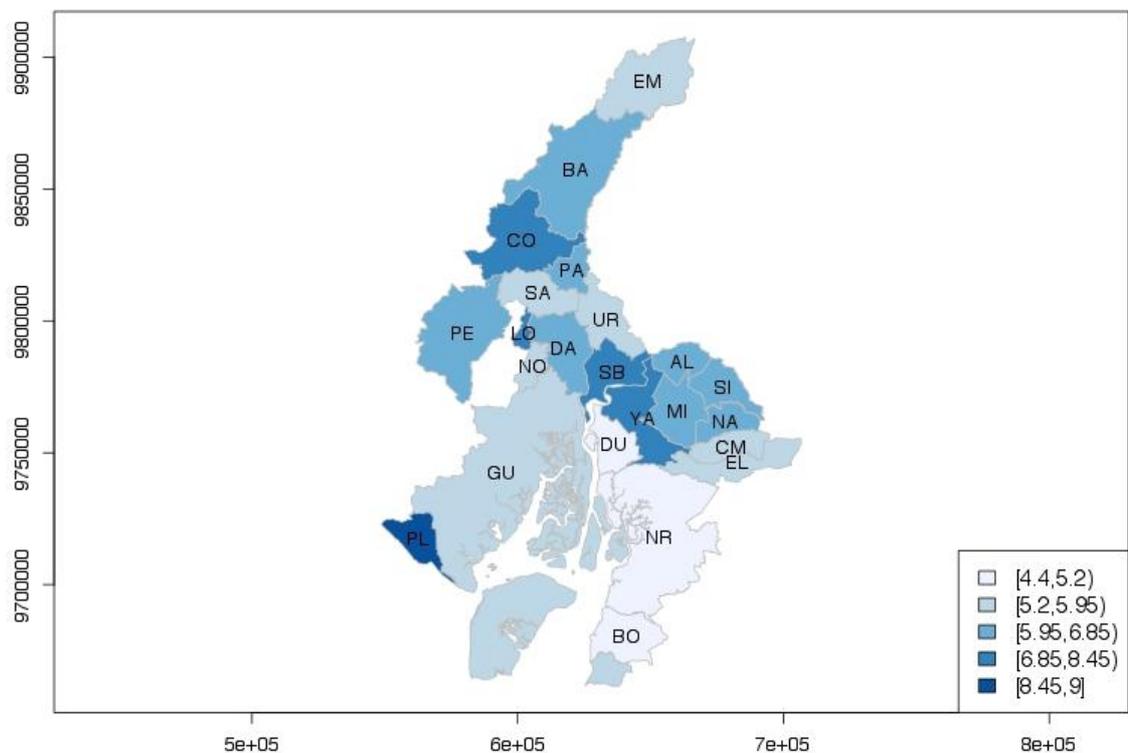
Similar al caso anterior, la figura 2 sugiere que la estimación del indicador analizado puede ser mejorada al tomar en cuenta la similaridad espacial de los cantones objetivo.

#### 4.1.3 Adulto mayor en los cantones de la provincia del Guayas

La provincia del Guayas cuenta con 8435 individuos en su muestra para diciembre de 2010. Es la provincia que tiene la mayor muestra a nivel Nacional<sup>1</sup> en los datos que contiene la ENEMDU. Está dividida en 23 cantones cuyos acrónimos se muestran en el anexo B.

Según el CPV 2010, las personas de 65 años y más representan el 5.8% (205128 personas) del total provincial. La figura 3 muestra este valor distribuido a nivel cantonal.

Figura 3: Proporción de población de 65 años y más – cantonal Guayas



<sup>1</sup> Pichincha: 5846 individuos en su muestra a diciembre de 2010.

El rango de la proporción de adultos mayores se encuentra entre 4,37 y 8,98. Es mucho menor al rango de los puntos 4.1.1 y 4.1.2. Es posible que la menor volatilidad y el número de cantones de esta Provincia ayude a mejorar la estimación.

#### **4.2 Selección del estimador de área pequeña**

En primer lugar se generan 500 muestras con muestreo aleatorio simple estratificado donde los estratos son los distritos  $d$  ( $SSRS_d$ , por sus siglas en inglés). Estas muestras son generadas a partir de las medias poblacionales de las provincias del Ecuador y de los cantones de Pichincha y Guayas según el caso. Finalmente se calculan las proporciones y tamaños muestrales de cada distrito y cada muestra para ser evaluadas por un conjunto de estimadores.

De la familia de estimadores posibles, el presente trabajo se orienta a la estimación directa, compuesta univariada y compuesta bivariada. Para estos dos últimos tipos de estimadores se tiene los casos en los que se puede tomar en cuenta la matriz de similaridad espacial. Las etiquetas de las estimaciones compuestas vienen dadas por  $K - Comp - H$ , donde  $K$  indica el uso de información auxiliar ( $K = 1$  si no usa y  $K = 2$  caso contrario) y  $H$  el valor de la matriz de similaridad espacial ( $H = 1$  si no la usa y  $H > 1$  con el valor de la matriz de similaridad espacial truncada en  $H$ ).

##### **4.2.1 Estimador para provincias del Ecuador**

En la tabla siguiente se resumen los resultados de las simulaciones para evaluar los estimadores  $K - Comp - H$  para las provincias del Ecuador.

Tabla 3: Comparaciones de conjuntos de estimadores para provincias

	<i>Directo</i>	<i>1-Comp-1</i>	<i>1-Comp-2</i>	<i>1-Comp-3</i>	<i>2-Comp-1</i>	<i>2-Comp-2</i>
<i>Directo</i>		0 (0)	0 (0)	0 (0,001)	0,001 (0,001)	0,001 (0,001)
<i>1-Comp-1</i>	0,003 [16]		0 (0)	0 (0,001)	0,001 (0,001)	0,001 (0,001)
<i>1-Comp-2</i>	0,005 [17]	0,003 [15]		0 (0)	0,001 (0,001)	0,001 (0,001)
<i>1-Comp-3</i>	0,008 [15]	0,005 [16]	0,004 [12]		0,001 (0,001)	0,001 (0,001)
<i>2-Comp-1</i>	0,016 [23]	0,014 [22]	0,014 [22]	0,016 [22]		0 (0)
<i>2-Comp-2</i>	0,019 [21]	0,016 [23]	0,017 [21]	0,019 [22]	0,006 [17]	

La tabla 3 está dividida en dos paneles. En la parte inferior de la diagonal se muestra la suma del valor absoluto del error de estimación. En corchetes se expresa el número de provincias en que el estimador de la fila es superior al estimador de la columna. En la parte superior de la diagonal se muestra la media del valor absoluto del error de estimación y su desviación estándar encerrada en paréntesis.

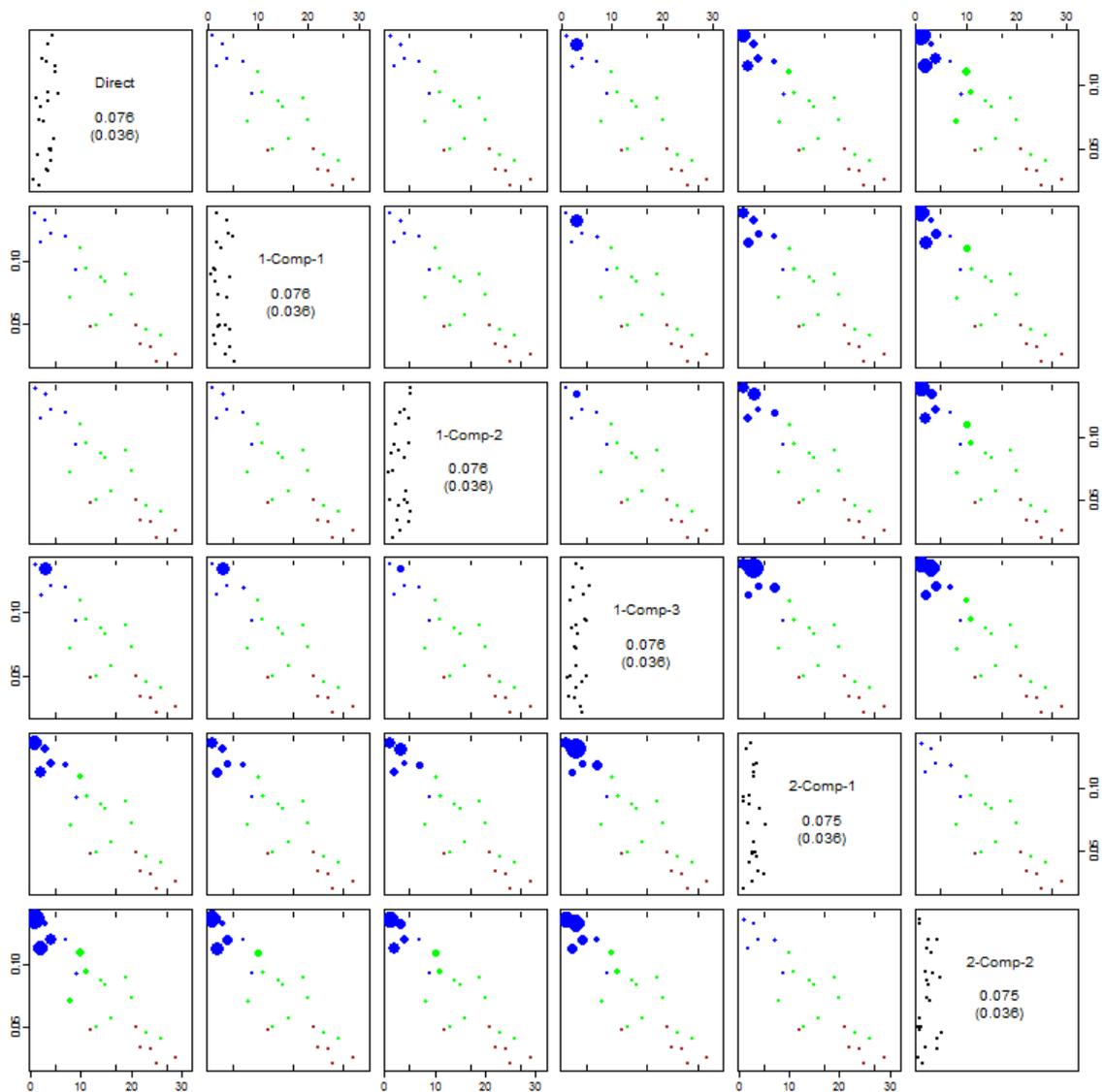
El rango de la suma del valor absoluto de los errores está entre 0,003 y 0,019 para todas las comparaciones, en general es un valor pequeño. Si se compara el conjunto de estimadores propuesto con la estimación directa, la estimación  $1 - Comp - 1$  (estimación compuesta que no usa información auxiliar ni la matriz de similaridad espacial) es la peor y la estimación  $2 - Comp - 1$  (estimación compuesta que usa información auxiliar y no usa la matriz de similaridad espacial) es la mejor.

Al tomar en cuenta ambos criterios, la suma del valor absoluto del error y el número de provincias en que mejora la estimación, el estimador  $2 - Comp - 1$  sigue siendo la mejor opción. Pese a que existen otros estimadores con valores inferiores en el primer criterio, el número de provincias en que se mejora la estimación es considerablemente mayor (23 a 16,

23 a 17 y 23 a 15). En resumen, la tabla 3 sugiere usar la estimación  $2 - Comp - 1$  para el cálculo de las proporciones de indígenas a nivel provincial.

La suma de los valores absolutos de los errores de cada provincia se compara en la figura 4.

Figura 4: Suma de errores absolutos por provincias del Ecuador



Cada panel de la figura 4 fuera de la diagonal tiene la misma escala. Cada conjunto de estimadores están representados en el eje vertical y en el eje horizontal están las provincias según el orden ascendente de sus tamaños poblacionales. El diámetro de los puntos en cada sub-gráfico muestra la magnitud de la diferencia entre la estimación del panel horizontal y vertical; sus colores indican la Región Natural a la que pertenece (Sierra: verde, Costa: café, Amazonía: azul). En la diagonal principal están las medias y las desviaciones estándar de la suma del valor absoluto de los errores así como los valores de los estimadores del eje horizontal distribuidos verticalmente en forma aleatoria.

Comparando con la estimación directa, a medida que se agrega información en los estimadores  $K - Comp - H$  se puede observar que las diferencias de las sumas de los valores absolutos crecen. Al agregar información de la matriz de similaridad espacial ( $H$ ), las diferencias en la Amazonía son mucho más notorias mientras que para las demás regiones es constante. Si se aumenta información auxiliar a la estimación, las diferencias de las regiones de la Amazonía y de la Sierra crecen más que en el caso anterior y la Costa permanece constante. Esto se debe a la proporción indígena por regiones, Sierra 11%, Costa 1% y Amazonía 33%.

Se encuentra un patrón consecuente entre el aumento de información y la mejora de la estimación dada por el conjunto de estimadores  $K - Comp - H$ . Para los fines de este trabajo, se procede a usar la estimación  $1 - Comp - 2$  para el cálculo de proporciones de indígenas a nivel provincial.

#### ***4.2.2 Estimator para los cantones de Pichincha.***

La tabla 4 resume las comparaciones de los estimadores en el caso de los cantones de la provincia de Pichincha.

Tabla 4: Comparaciones de conjuntos de estimadores para cantones

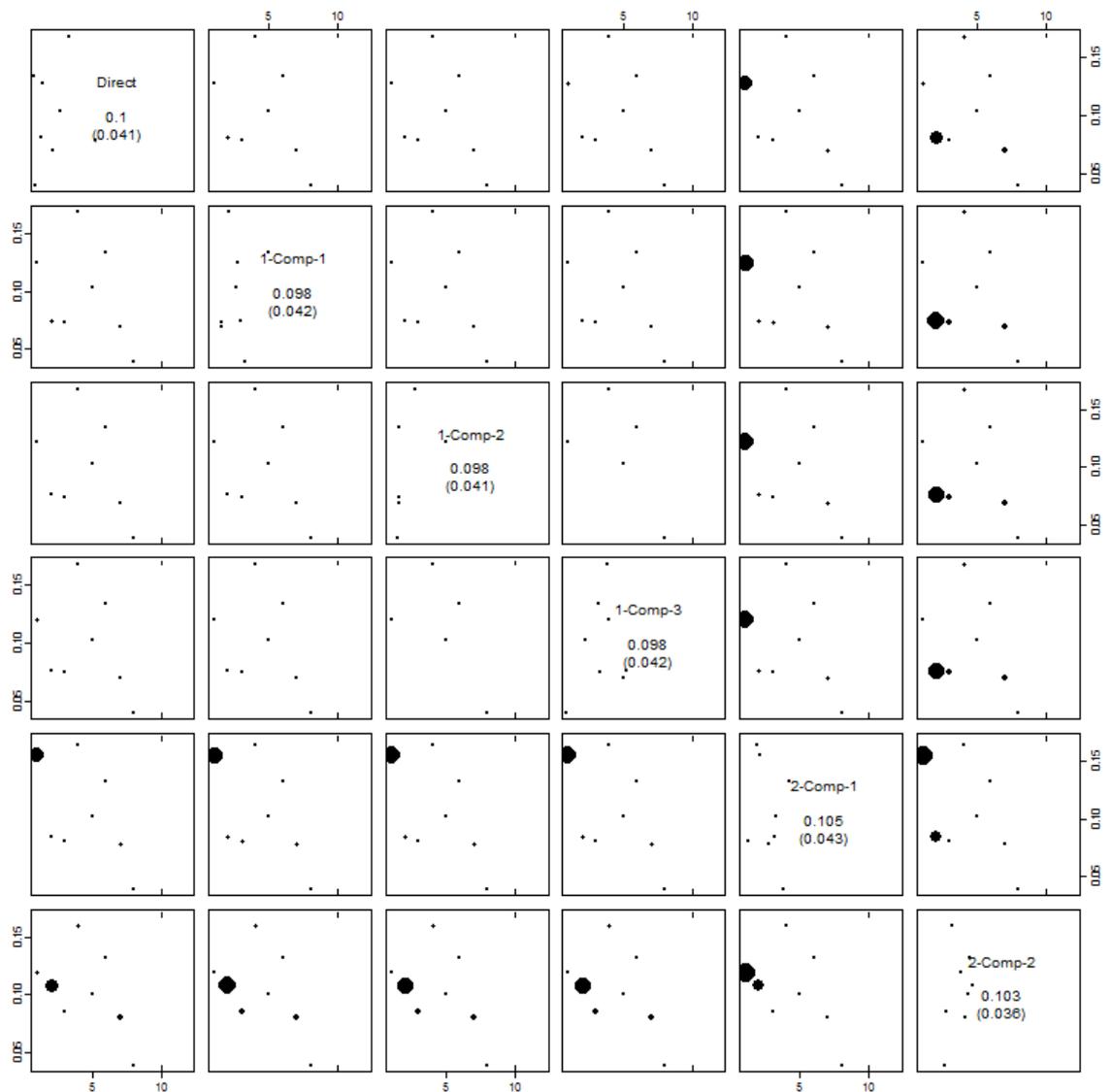
	<i>Directo</i>	<i>1-Comp-1</i>	<i>1-Comp-2</i>	<i>1-Comp-3</i>	<i>2-Comp-1</i>	<i>2-Comp-2</i>
<i>Directo</i>		0.002 (0.003)	0.002 (0.002)	0.002 (0.003)	-0.004 (0.01)	-0.003 (0.012)
<i>1-Comp-1</i>	0.016 [6]		0 (0.001)	0 (0.002)	-0.006 (0.011)	-0.004 (0.014)
<i>1-Comp-2</i>	0.018 [7]	0.008 [5]		0 (0.001)	-0.007 (0.012)	-0.005 (0.013)
<i>1-Comp-3</i>	0.02 [7]	0.009 [5]	0.004 [3]		-0.007 (0.012)	-0.005 (0.013)
<i>2-Comp-1</i>	0.048 [4]	0.063 [3]	0.062 [3]	0.065 [3]		0.002 (0.017)
<i>2-Comp-2</i>	0.065 [5]	0.074 [5]	0.067 [5]	0.067 [5]	0.072 [5]	

\*Las mismas convenciones y definiciones de la tabla 3 aplican en la tabla 4.

El rango de la suma del valor absoluto de los errores está entre 0,004 y 0,074 para todas las comparaciones, las variaciones cantones con 4 veces mayores a las provinciales. La tabla 4 sugiere que sea la estimación  $1 - Comp - 1$  es la mejor en términos de la suma de la diferencia de los valores absolutos. Asimismo, serían las estimaciones  $1 - Comp - 2$  y  $1 - Comp - 3$  las candidatas a ser seleccionadas al considerar el criterio del número de cantones que presentan mejora.

Complementariamente a las sugerencias que muestra la tabla 4, se presenta una medición individual de los cantones en la figura 5.

Figura 5: Suma de errores absolutos por cantones de la provincia de Pichincha.



La figura 5 no sugiere un patrón de disminución del error a medida que aumenta la población. Cuatro de los 8 cantones de la provincia de Pichincha tienen valores de la proporción de indígenas muy cercanos a cero. Esto corrobora la falta de evidencia de un

patrón poblacional. Pedro Moncayo muestra un comportamiento aberrante, es el punto atípico en todos los casos. Esto se debe a que entre el 2001 y el 2010 la proporción de la población que se considera indígena para este cantón es la que más creció (7,8%).

Es claro que los estimadores compuestos mejoran los resultados en todos los casos. Sin embargo no es posible determinar uno en particular para la estimación a partir de la evidencia mostrada en la tabla y figura 4. Para el caso de los cantones de Pichincha, se procederá a realizar el cálculo de las proporciones con el supuesto de que a mayor información, mejor resultado ( $2 - Comp - 2$ ).

#### 4.2.3 Estimador para los cantones de Guayas: Adulto mayor.

La tabla 5 resume las comparaciones de los estimadores en el caso de los cantones de la provincia de Guayas para la estimación de la proporción de adultos mayores.

Tabla 5: Comparaciones de conjuntos de estimadores para cantones - Guayas

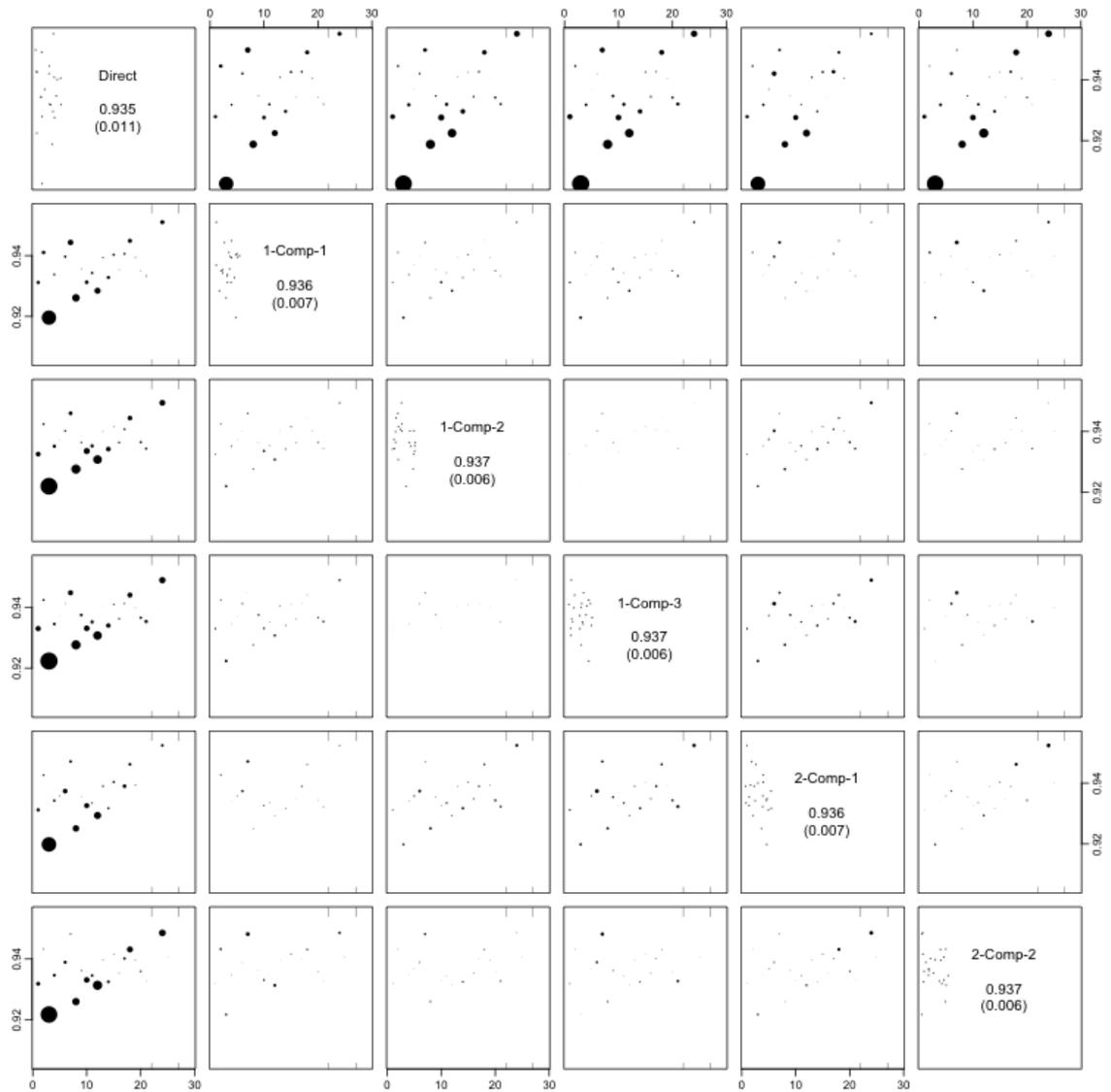
	<i>Directo</i>	<i>1-Comp-1</i>	<i>1-Comp-2</i>	<i>1-Comp-3</i>	<i>2-Comp-1</i>	<i>2-Comp-2</i>
<i>Directo</i>		-0.001 (0.004)	-0.002 (0.005)	-0.002 (0.005)	-0.001 (0.004)	-0.001 (0.005)
<i>1-Comp-1</i>	0.072 [11]		-0.001 (0.001)	-0.001 (0.001)	0 (0.001)	-0.001 (0.001)
<i>1-Comp-2</i>	0.086 [10]	0.027 [2]		0 (0.001)	0.001 (0.001)	0 (0.001)
<i>1-Comp-3</i>	0.089 [10]	0.029 [2]	0.01 [13]		0.001 (0.002)	0 (0.001)
<i>2-Comp-1</i>	0.068 [11]	0.021 [12]	0.036 [19]	0.04 [19]		-0.001 (0.001)
<i>2-Comp-2</i>	0.079 [10]	0.025 [8]	0.019 [19]	0.023 [16]	0.026 [2]	

Al comparar los valores de la suma del valor absoluto de la primera columna de la tabla 5, se aprecia que el de menor valor es la estimación  $2 - Comp - 1$  y mejora a la estimación directa en 11 cantones. Sin embargo, la estimación  $2 - Comp - 2$  supera a la anterior en dos

cantones a cambio de un ligero aumento en el error. La tabla 5 sugiere entonces el uso de la estimación 2 – *Comp* – 2 para la proporción de adultos mayores.

La figura 6 muestra el error de estimación para cada uno de los cantones de la provincia del Guayas.

Figura 6: Suma de errores absolutos por cantones de la provincia de Guayas.



Todas las estimaciones diferentes de la estimación directa muestran mejoras en cuanto a precisión. Al comparar la estimación directa con la 2 – *Comp* – 2 se tiene que la estimación directa muestra un error promedio de 0,935 y la segunda de 0,937. Sin embargo la varianza de

la primera es mayor. En función de la tabla 5 y la figura 6 se puede concluir que la estimación  $2 - Comp - 2$  es una buena opción para la estimación de la proporción de adultos mayores.

### 4.3 Resultados del estimador

En esta sección se resumen los resultados obtenidos a partir de los modelos  $2 - Comp - 1$  y  $2 - Comp - 2$  para provincias del Ecuador y cantones de Pichincha y Guayas respectivamente.

La matriz de distancia entre los estimadores  $2 - Comp - 1$ , ENEMDU y CPV 2010 se presenta en la tabla 5.

Tabla 6: Matriz de distancias de los estimadores de áreas pequeñas - Provincias

	2-Comp-1	ENEMDU	CPV 2010
2-Comp-1	0.000	0.004	24.339
ENEMDU	0.004	0.000	24.343
CPV 2010	24.339	24.343	0.000

La distancia entre la estimación proveniente de la ENEMDU es mayor que la distancia que tiene  $2 - Comp - 1$  respecto de las proporciones del Censo de Población y Vivienda. Es decir que se logra el objetivo de mejorar la estimación de las proporciones de indígenas a nivel provincial.

La siguiente tabla muestra la misma información que la tabla 6, en este caso para los cantones de la Provincia de Pichincha:

Tabla 7: Matriz de distancias de los estimadores de áreas pequeñas – cantones Pichincha

	2-Comp-2	ENEMDU	CPV 2010
2-Comp-2	0.000	0.005	48.191

ENEMDU	0.005	0.000	48.194
CPV 2010	48.191	48.194	0.000

La tabla 7 muestra una distancia menor entre el CPV 2010 y el estimador 2 – *Comp* – 2 que la distancia entre la proporción obtenida a partir de la ENEMDU y el CPV 2010. Al igual que el caso anterior, mejora la estimación de la proporción de indígenas para los cantones de Pichincha.

La tabla 8 resume los resultados para la estimación de los adultos mayores en los cantones de la provincia de Guayas.

Tabla 8: Matriz de distancias de los estimadores de áreas pequeñas – cantones Guayas

	2-Comp-2	ENEMDU	CPV 2010
2-Comp-2	0	0.155	21.617
ENEMDU	0.155	0	21.705
CPV 2010	21.617	21.705	0

La estimación compuesta usando información del 2001 y la matriz de similaridad espacial es más precisa que la estimación directa. La distancia que existe entre la estimación de este trabajo y los resultados del CPV 2010 es la menor de todas en la tabla 8. Esto puede deberse a la menor heterogeneidad del indicador estimado.

## CONCLUSIONES Y RECOMENDACIONES

### Conclusiones

A través de las tablas 3 y 4 se ha puesto en evidencia que las estimaciones a través de composición superan a las estimaciones directas en todos los casos tanto para la estimación de la proporción de indígenas en las provincias del Ecuador como en los cantones de la provincia de Pichincha.

Cuando se cuenta con un número considerable de distritos, se puede observar patrones de comportamiento error-población. A medida que aumenta la población disminuye el error de estimación en todos los casos del tipo provincial.

Mayor homogeneidad en la estimación del indicador de adultos mayores refleja mayor precisión al momento de incorporar estimaciones compuestas en todos los casos.

La elección del modelo estimado depende de los criterios de superioridad en el número de distritos en los que la estimación mejora y al menor valor de del error causado.

El contar con información auxiliar suele ser uno de los grandes inconvenientes para la estimación de áreas pequeñas. En este sentido, el uso de la matriz de similaridad espacial ayuda mejorar las estimaciones directas del indicador. Por ejemplo, las estimaciones  $1 - Comp - 2$  y  $1 - Comp - 3$  mejoran la estimación en 17 y 15 provincias respectivamente.

El conjunto de estimadores  $K - Comp - H$  son libres de uso de un modelo. Esto puede ser una gran ventaja al dejar de lado la subjetividad que puede generarse en la especificación del modelo usado para estimadores de áreas pequeñas.

Dado que los estimadores socioeconómicos son insumos de la planificación nacional y local, de su adecuada estimación puede depender el hecho que ciertos recursos sean asignados a una determinada localidad, que una persona sea considerada o no en la participación de algún programa social, entre otros. A este respecto se concluye que, pese a que la mejora de la

estimación dada en el presente documento no luzca impactante, cualquier mejora en la estimación puede ser determinante en la toma de decisiones.

### **Recomendaciones**

Se ha utilizado como información auxiliar a los datos del CPV 2001. Es probable que esto haya afectado la precisión de la estimación compuesta que toma en cuenta este particular. Se recomienda el uso de información auxiliar más reciente de registros administrativos o encuestas especializadas. En particular, para el caso de adultos mayores se recomienda el uso de información del Registro Social que maneja el Ministerio Coordinador de Desarrollo Social.

En el presente estudio la determinación de la muestra a través de  $SSRS_d$  se realiza asumiendo una fracción de la población constante de  $1/200$ . Este valor podría ser considerado diferente para cada distrito y evaluar el rendimiento del conjunto de estimadores en este caso.

## REFERENCIAS

N. LONGFORD, *Small-area estimation with spatial similarity* , *Computational Statistics & Data Analysis*, Vol. 54, Sec. 4 (2010).

MINISTERIO COORDINADOR DE DESARROLLO SOCIAL, *Mapa de la desnutrición crónica en el Ecuador* , FESAECUADOR, 13-16, (2010).

N. LONGFORD, *Missing data and Small Area Estimators* Springer, 143-170, (2005).

J. RAO, *Small Area Estimation* , Jhon Wiley & Sons, 45-72, (2003).

INEC, InecPedia , En línea: <http://www.ecuadorencifras.com> , Consulta: 8 de enero 2014

Banco Mundial, PovMap , En línea: <http://econ.worldbank.org> , Consulta: 8 de enero 2014

## ANEXO A: CÓDIGO DE MATRIZ DE SIMILARIDAD ESPACIAL

```

"Las Provincias de Ecuador y sus vecinas"

# ProVec: CatVei

ProVec<-c(

  "AZUAY", "AZ", "EL", "LO", "ZA", "MO", "CÑ", "GU",
  "BOLIVAR", "BO", "GU", "CH", "TU", "CO", "LR",
  "CAÑAR", "CÑ", "AZ", "CH", "GU",
  "CARCHI", "CA", "ES", "IM", "SU",
  "COTOPAXI", "CO", "BO", "TU", "NA", "PI", "SD", "LR",
  "CHIMBORAZO", "CH", "CÑ", "MO", "TU", "BO",
  "EL ORO", "EL", "LO", "AZ", "GU",
  "ESMERALDAS", "ES", "MA", "SD", "PI", "IM", "CA",
  "GUAYAS", "GU", "EL", "AZ", "CÑ", "BO", "LR", "MA", "SE",
  "IMBABURA", "IM", "PI", "SU", "CA", "ES",
  "LOJA", "LO", "ZA", "AZ", "EL",
  "LOS RIOS", "LR", "BO", "CO", "SD", "MA", "GU",
  "MANABI", "MA", "SE", "GU", "LR", "SD", "ES",
  "MORONA SANTIAGO", "MO", "ZA", "PA", "TU", "CH", "AZ",
  "NAPO", "NA", "PA", "OR", "SU", "PI", "CO", "TU",
  "PASTAZA", "PA", "MO", "OR", "NA", "TU",
  "PICHINCHA", "PI", "CO", "NA", "SU", "IM", "ES", "SD",
  "TUNGURAHUA", "TU", "CH", "MO", "PA", "NA", "CO", "BO",
  "ZAMORA CHINCHIPE", "ZA", "MO", "AZ", "LO",
  "SUCUMBIOS", "SU", "OR", "NA", "PI", "IM", "CA",
  "ORELLANA", "OR", "PA", "NA", "SU",
  "SANTO DOMINGO", "SD", "LR", "CO", "PI", "ES", "MA",
  "SANTA ELENA", "SE", "GU", "MA")

```

```
"Los nombres de las Provincias y sus ubicaciones en el vector"
```

```
"CatNam: ProNam"
```

```
"CatLoc: ProNom"
```

```
ProNam <- ProVec[nchar(ProVec)>2]
```

```
ProLoc <- c(match(ProNam, ProVec), length(ProVec)+1)
```

```
names(ProLoc) <- c(ProNam, c())
```

```
"Numero de Provincias"
```

```
"nCat: nProv"
```

```
nProv <- length(ProNam)
```

```
"Los acrónimos de las provincias"
```

```
"CatAcr: ProAcr"
```

```
ProAcr <- ProVec[ProLoc[seq(nProv)]+1]
```

```
"Donde termina la lista previa de vecinos"
```

```
end <- 0
```

```
"Lista de Provincias y sus vecinos"
```

```
ProVecL <- list()
```

```
lc <- 1
```

```
for (cm in ProNam)
```

```
{
```

```
  lc <- lc+1
```

```
  sta <- end + 3
```

```
  end <- ProLoc[lc]-1
```

```
  "Los vecinos de la Porvincia cm"
```

```
  ProVecL[[cm]] <- ProVec[sta:end]
```

```

}
"Formar la matriz de vecinos/similaridad"
"CatSim: ProSim"
ProSim <- matrix(0, nProv, nProv, dimnames=list(ProAcr, ProAcr))

"Marcamos los vecinos de cada Provincia"
for (lc in seq(nProv))
  ProSim[lc,ProVecL[[lc]]] <- 1
"Encontramos los vecinos una vez removidos"
ProRem <- function(cm, Mat)
{
  " cm Numero de fila de nombre (acrónimo de la Prov)"
  " Mat Matriz de similaridad"

  "Convierte a número si es necesario"
  if (!is.numeric(cm)) cm <- match(cm, dimnames(Mat)[[1]], nomatch=0)
  "Los índices"
  ind <- seq(dim(Mat)[1])
  "Los vecinos"
  vei <- ind[Mat[cm,]>0]
  "Los vecinos de los vecinos"
  vei2 <- Mat[vei,]==1
  if (length(vei)>1) vei2 <- apply(vei2,2,max)
  Mat[cm,] + (max(Mat)+vei2) * (Mat[cm,]==0)*(vei2>0)
}
"Fin de la funcion ProRem"
while (min(ProSim)==0)
  ProSim <- apply(matrix(seq(nProv),nrow=1),2,ProRem, ProSim)

```

```

"Ponemos etiquetas"

dimnames(ProSim) <- list(ProAcr, ProAcr)

"Distancia consigo misma es cero"

diag(ProSim) <- 0

```

## ANEXO B: ACRÓNIMOS DE CANTONES - GUAYAS

NOMBRE CANTÓN	ACRÓNIMO
GUAYAQUIL	GU
ALFREDO BAQUERIZO MORENO	AL
BALAO	BO
BALZAR	BA
COLIMES	CO
DAULE	DA
DURAN	DU
EMPALME	EM
EL TRIUNFO	EL
MILAGRO	MI
NARANJAL	NR
NARANJITO	NA
PALESTINA	PA
PEDRO CARBO	PE
SAMBORONDON	SB
SANTA LUCIA	SA
URBINA JADO	UR
YAGUACHI	YA
PLAYAS (GENERAL VILLAMIL	PL
SIMON BOLIVAR	SI
CORONEL MARCELINO MARIDUEÑA	CM
LOMAS DE SARGENTILLO	LO
NOBOL (VICENTE PIEDRAHITA)	NO