

Dirección de Desarrollo Digital

Unidad de Científicos
de Datos



**El futuro
es de todos**

DNP
Departamento
Nacional de Planeación

ESTIMACIÓN DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL MEDIANTE MODELOS DE ESTIMACIÓN EN ÁREAS PEQUEÑAS

Entidad

- Dirección de Descentralización y Desarrollo Regional - DNP

Sector

Planeación

Lenguaje

R.

Fuente de datos

1. Terridata
2. TIPOLOGIAS DDT - Departamentales
3. TIPOLOGIAS DDT - Municipales
4. SISBEN IV
5. ECV

Presentación

El Índice de Pobreza Multidimensional (IPM), desarrollado por el Oxford Poverty & Human Development Initiative (OPHI), es un indicador que refleja el grado de privación de las personas en un conjunto de dimensiones (Condiciones educativas del hogar, Condiciones de la niñez y la juventud, trabajo, salud y Servicios públicos). La medida permite determinar la naturaleza de la privación (de acuerdo con las dimensiones seleccionadas) y su intensidad. El IPM es la combinación del porcentaje de personas consideradas pobres y de la proporción de dimensiones en las cuales los hogares son, en promedio, pobres. En el presente trabajo se utilizó un modelo de estimación en áreas pequeñas (SAE, por sus siglas en inglés) con el fin de pronosticar el valor del IPM a nivel municipal empleando la información presente en la base de datos de SISBEN IV y los datos de diferentes fuentes administrativas (Encuesta de Calidad de Vida - DANE). El modelo de pronóstico fue el insumo de un tablero de control online que presentará por medio de mapas los resultados obtenidos en el estudio.

The Multidimensional Poverty Index (MPI), developed by the Oxford Poverty & Human Development Initiative (OPHI), is an indicator that reflects the degree of deprivation of people in a set of dimensions (Educational conditions of the home, Conditions of childhood and youth, work, health and public services). The measure makes it possible to determine the nature of deprivation (according to the dimensions selected) and its intensity. The MPI is the combination of the percentage of people considered poor and the proportion of dimensions in which households are, on average, poor. In this work, a small area estimation model (SAE) was used to forecast the MPI value at the municipal level using information from the SISBEN IV database and data from different administrative sources (Quality of Life Survey - DANE). The forecast model was the input of an online control panel that will present the results obtained in the study by means of maps.

Objetivo general

Identificar un modelo de estimación en áreas pequeñas que a partir de diferentes registros administrativos genere pronósticos para el IPM por año.

Objetivos específicos

1. Estimar el valor del IPM para los municipios con información completa en el SISBEN IV.
2. Evaluar el grado de relación entre las variables auxiliares a nivel municipal con la información recolectada en el SISBEN IV.

3. Emplear técnicas de selección de variables que permita hacer una depuración de la información recolectada en el estudio.
4. Desarrollar un algoritmo que permita hacer remuestreo (Bootstrap) de la información disponible y así generar estimadores consistentes de los parámetros de interés en el estudio.

Metodología

La propuesta de IPM desarrollada por el Departamento Nacional de Planeación (DNO, DDS, SPSCV, 2011) para Colombia está conformada por 5 dimensiones y 15 variables:

- **Dimensión 1:** Condiciones educativas del hogar
Variables:
 - Logro educativo
 - Analfabetismo

- **Dimensión 2:** Condiciones de la niñez y la juventud
Variables:
 - Asistencia escolar
 - Rezago escolar
 - Acceso a servicios para el cuidado de la primera infancia
 - Trabajo infantil

- **Dimensión 3:** Trabajo
Variables:
 - Desempleo de larga duración
 - Empleo formal

- **Dimensión 4:** Salud
Variables:
 - Aseguramiento en salud
 - Acceso a servicio de salud dada una necesidad

- **Dimensión 5:** Servicios públicos
Variables:
 - Acceso a fuente de agua mejorada
 - Eliminación de excretas
 - Material de pisos
 - Material de paredes exteriores
 - Hacinamiento crítico

De acuerdo con esta medida, se considera que un hogar está en condición de pobreza si cuenta con privaciones en al menos 5 de las variables seleccionadas (33% del total de privaciones). En la Tabla 1 se presenta la definición de privación que fue considerada para realizar la estimación directa del IPM empleando los registros administrativos disponibles (Censo, 2005 y ECV).

Dimensión	Variable	Un hogar se considera privado cuando...
1. Condiciones educativas del hogar	Logro educativo	La educación promedio de las personas de 15 años y más es menor a 9 años
	Analfabetismo	Al menos una persona mayor de 15 años no sabe leer y escribir.
2. Condiciones de la niñez y juventud	Asistencia escolar	Al menos uno de los niños entre 6 y 16 años en el hogar (si los hay) no asiste al colegio
	Rezago escolar	Alguno de los niños entre 7 y 17 años en el hogar (si los hay) tiene rezago escolar. Es decir: <ul style="list-style-type: none"> • Tiene 7 años y no tiene al menos 1 año de educación • Tiene 8 años y no tiene al menos 2 años de educación • Tiene 9 años y no tiene al menos 3 años de educación • Tiene 17 años y no tiene al menos 11 años de educación
	Acceso a servicios para el cuidado de la primera infancia	Al menos uno de los niños entre 0 y 5 años del hogar (si los hay) no tiene acceso simultáneo a salud, nutrición adecuada y educación inicial. <ul style="list-style-type: none"> • Se considera como privado un niño menor de 5 años si no está asegurado a salud; o pasa la mayor parte del tiempo con su padre o madre en el trabajo, en casa solo, o en casa únicamente con menores de edad; o si no recibe almuerzo en la institución donde permanece la mayor parte del tiempo. • Se considera que un niño de 5 años está privado si no asiste a una institución educativa, o no se encuentra asegurado a alguna entidad de seguridad social o no recibe almuerzo o algún refrigerio en el plantel educativo al que asiste.
	Trabajo infantil	Hay niños entre 12 y 17 años que están en condición de trabajo infantil
3. Trabajo	Desempleo de larga duración	Alguna persona económicamente activa del hogar se encuentra en desempleo de larga duración (más de 12 meses)
	Empleo informal	Alguno de sus ocupados no se encuentra afiliado a pensiones
4. Salud	Aseguramiento en salud	Al menos uno de los miembros del hogar, mayores de cinco años, no está asegurado al Sistema de Seguridad Social en Salud.
	Acceso a servicio de salud dada una necesidad	Al menos un miembro del hogar en los últimos 30 días tuvo una enfermedad, accidente, problema odontológico o algún otro problema de salud que no haya implicado hospitalización y que para tratar este problema de salud NO acudieron a un médico general, especialista, odontólogo, terapeuta o institución de salud.
5. Servicios públicos domiciliarios	Acceso a fuente de agua mejorada	<ul style="list-style-type: none"> • Si es urbano, no tiene servicio público de acueducto en la vivienda.

y condiciones de la Vivienda		<ul style="list-style-type: none"> Si es rural, obtiene el agua para preparar los alimentos de pozo sin bomba, agua lluvia, río, manantial, carro tanque, aguatero u otra fuente.
	Eliminación de excretas	<ul style="list-style-type: none"> Si es urbano, no tiene servicio público de alcantarillado. Si es rural, tiene inodoro sin conexión, bajamar o no tiene servicio sanitario.
	Material de pisos	Tienen pisos en tierra
	Material paredes exteriores	<ul style="list-style-type: none"> Si es urbano, el material de las paredes exteriores es madera burda, tabla, tablón, guadua, otro vegetal, zinc, tela, cartón, deshechos y sin paredes. Si es rural, material de las paredes exteriores es guadua, otro vegetal, Zinc, tela, cartón, deshechos y sin paredes.
	Hacinamiento crítico	Personas que duermen por cuarto (excluyendo cocina, baño y garaje e incluyendo sala y comedor): <ul style="list-style-type: none"> Urbano: mayor o igual a 3 personas. Rural: Más de 3 personas por cuarto.

Tabla 1: Definición de privación por variable del IPM.

El IPM permite obtener estimaciones de la incidencia de la pobreza multidimensional para diferentes dominios geográficos. Las estimaciones realizadas por el DNP previas al desarrollo del presente proyecto incluyen resultados para 1997, 2003 y 2008 con base en la información de la Encuesta Nacional de Calidad de Vida. Los resultados reportados por el DNP, desde el punto de vista estadístico, son estimaciones directas pues emplean los pesos de muestreo del diseño considerado por la Encuesta Nacional de Calidad de Vida para los años en cuestión y en ese sentido tendrán la precisión establecida por el grado de representatividad de dicha encuesta (a nivel región).

Bajo esta consideración, cualquier estimación que se realice a un mayor nivel de desagregación (departamental o municipal) puede tener un mal comportamiento en términos de varianza porque los tamaños de muestra pueden ser insuficientes para garantizar la precisión adecuada en la estimación de las características de interés.

El presente estudio tiene como finalidad realizar una estimación del IPM a nivel municipal mediante el uso de modelos de áreas pequeñas empleando la información agregada generada por la Encuesta Nacional de Calidad de Vida y el microdato disponible en la base del SISBEN IV. Se considera el uso explícito de estos modelos pues tienen la capacidad de generar estimaciones precisas a nivel departamental o municipal si se tiene información auxiliar de calidad que esté asociada al parámetro de interés en el estudio (IPM).

Modelos de área pequeña

En particular, los modelos de estimación en áreas pequeña involucran efectos aleatorios específicos por dominio de interés que explican la variación entre estos más allá de lo explicado por las variables auxiliares incluidas en el modelo.

El uso de modelos SAE ofrece varias ventajas:

- (i) Los diagnósticos del modelo se pueden usar para encontrar modelos adecuados que se ajusten bien a los datos. Dichos diagnósticos del modelo incluyen análisis de residuales para detectar desviaciones del modelo asumido, selección de variables auxiliares y medidas diagnósticas para detectar observaciones influyentes y/o atípicas.
- (ii) Las medidas de precisión de cada dominio pueden asociarse con cada estimación de “área pequeña”, a diferencia de las medidas globales (promedios por dominio de interés) empleados por los estimadores sintéticos.
- (iii) Los modelos lineales mixtos, así como los modelos no lineales, como los modelos de regresión logística y los modelos lineales generalizados con efectos de área aleatorios, pueden ser empleados como parte del predictor lineal. Los modelos SAE permiten, adicionalmente, manejar estructuras de datos complejas, como la dependencia espacial y series de tiempo.
- (iv) Los desarrollos metodológicos recientes para los modelos de efectos aleatorios se pueden utilizar para lograr inferencias precisas en modelos SAE.

Los modelos de estimación en áreas pequeñas se pueden clasificar en 2 tipos:

- **Modelos de nivel agregado (área)**
Son aquellos que relacionan los estimadores directos en los dominios de interés con covariables que son específicas al área.
- **Modelos a nivel de unidad**
Son aquellos modelos que relacionan los valores unitarios de una variable de estudio con las covariables específicas de la unidad.

Resulta importante señalar que los modelos a nivel de unidad asumen que los valores dentro de la muestra en los dominios de interés obedecen al modelo poblacional asumido. Para los modelos de área se asume la ausencia de muestra informativa en los dominios de interés en situaciones donde solo algunos de estos son seleccionados en la muestra. Esto implica que los valores obtenidos en las áreas muestreadas (estimadores directos) obedecen al modelo poblacional asumido.

En el presente proyecto se están empleando los modelos SAE a nivel área para realizar una estimación inicial del parámetro de interés (IPM a nivel municipal). Los mencionados modelos se definen mediante la siguiente ecuación:

$$\hat{\theta}_i = \mathbf{z}_i^T \boldsymbol{\beta} + b_i v_i + e_i$$

Donde:

- $i = 1, 2, \dots, m$ denota el subíndice asociado a cada dominio o área pequeña,
- $\hat{\theta}_i$ denota el estimador del parámetro de interés θ (IPM) en el i -ésimo dominio (municipio)
- \mathbf{z}_i^T denota el conjunto de variables auxiliares observadas para cada dominio específico. Para el presente proyecto denota la información auxiliar disponible para cada municipio de interés.
- $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ denota el vector de coeficientes de regresión ($m > p$) que representan el efecto (fijo) que tiene cada variable auxiliar sobre la estimación del IPM para cada municipio.
- b_i denota un conjunto de constantes positivas.
- Los v_i son los efectos aleatorios específicos de cada dominio o área de interés. Se supone que los v_i son independientes e idénticamente distribuidos (iid) con

$$E_m(v_i) = 0, \quad V_m(v_i) = \sigma_v (\geq 0),$$

Siendo E_m y V_m , el valor esperado y la varianza del modelo lineal considerado. En resumen, para el proyecto en cuestión se supone que cada municipio tiene un efecto aleatorio independiente que sigue una distribución de probabilidad con media 0 y varianza σ_v , es decir, $v_i \sim iid (0, \sigma_v^2)$.

El parámetro σ_v^2 es una medida de homogeneidad de los municipios (áreas) después de tener en cuenta la información disponible en las covariables \mathbf{z}_i^T .

- e_i denota el error del diseño de muestreo empleado que está asociado al i -ésimo municipio. Se supone que los e_i son independientes e idénticamente distribuidos (iid) con

$$E_p(e_i|\theta_i) = 0, \quad V_p(e_i|\theta_i) = \psi_i (\geq 0),$$

También es habitual suponer que las variaciones asociadas al diseño de muestreo, ψ_i , son conocidas.

Nota: El modelo asume que v_i y e_i son independientes.

Resultados

Empleando los datos del CENSO 2005 y los criterios de privación presentados en la Tabla 1, se realizó la estimación directa (sin uso de modelos SAE) del índice de pobreza multidimensional para todos los municipios de Colombia. En la Tabla 2 y 3 se presentan los 10 resultados obtenidos de mayor y menor valor a nivel municipal respectivamente, el listado completo no se presenta en este documento debido a la extensión de la tabla resultante; para tal fin se diseñó un tablero de control en línea que permite presentar a través de mapas la estimación del índice a nivel departamental y a nivel municipal. La URL del aplicativo, que puede ser consultado al conectarse a la red interna del DNP, es la siguiente: <https://vdatascience:3838/ipm>

Departamento	Municipio	IPM
Cauca	Argelia	1,000
Choco	Medio Atrato	1,000
Nariño	El Charco	0,990
Nariño	Magui-Payan	0,989
Choco	Jurado	0,989
Choco	Atrato	0,987
Antioquia	Vigía Del Fuerte	0,986
Choco	Rio Iro	0,984
Choco	Novita	0,984
Choco	Union Panamericana	0,980

Tabla 2: Listado de 10 municipios con mayor IPM (estimación directa).

Departamento	Municipio	IPM
--------------	-----------	-----

Antioquia	Envigado	0,143
Cundinamarca	Chía	0,165
Antioquia	Sabaneta	0,181
Cundinamarca	Madrid	0,219
Antioquia	Itagüí	0,231
Cundinamarca	Facatativá	0,240
Cundinamarca	Mosquera	0,241
Cundinamarca	Funza	0,246
Cundinamarca	Cajicá	0,250
Cundinamarca	Sopo	0,262

Tabla 3: Listado de 10 municipios con menor IPM (estimación directa).

En la Figura 1 se presenta el primer mapa de Colombia generado por el mencionado tablero de control; en este se ilustra a partir de un mapa de calor las estimaciones directas del IPM a nivel departamental.

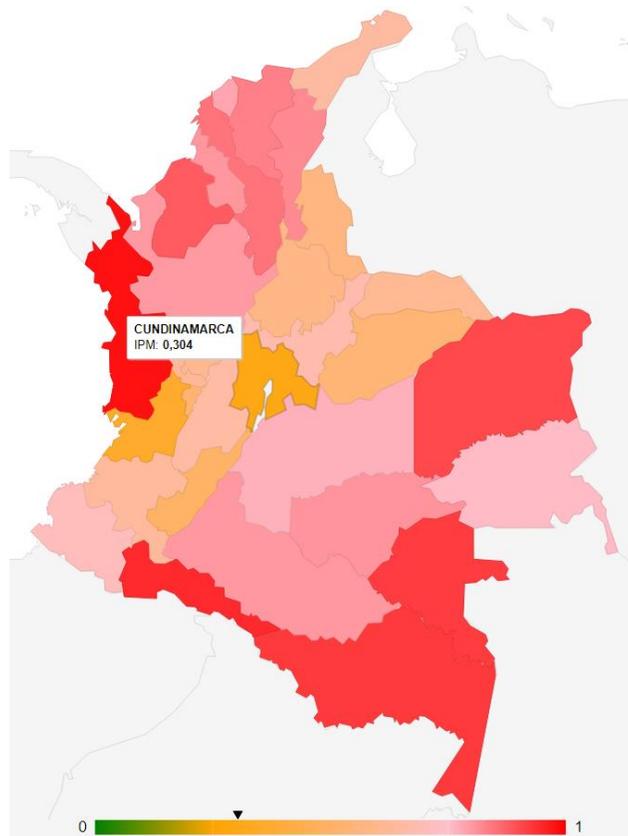


Figura 1: Mapa de la estimación directa del IPM a nivel departamental - Censo 2005.

Nota: Un mapa de calor (*heatmap*) es una ilustración que utiliza diferentes colores típicamente asociados con las temperaturas. Así, el mapa de calor clásico contiene los colores rojos (para muy caliente), naranja (para moderadamente caliente), amarillo (para caliente) y verde (para frío).

La escala de colores definida inicialmente para el proyecto en cuestión está asociada a los posibles valores del IPM el cual se estableció en el intervalo $[0,1]$, donde 0 denota que no existe ninguna privación y 1 el máximo grado de privación en el cual todos los hogares del departamento tienen todas las privaciones definidas en la Tabla 1. La paleta de colores asociada al mapa de intensidad se presenta en la Figura 2.



Figura 2: Paleta de colores para mapa de calor.

Inicialmente el mapa no presenta el valor numérico del IPM que fue calculado para cada uno de los departamentos, dicho valor es presentado utilizando un *tooltip*, el cual aparecerá una vez el usuario del tablero de control ubique el cursor en el departamento que quiera consultar.

La segunda sección del aplicativo en línea tiene un botón de consulta, el cual está diseñado para que el usuario pueda seleccionar un departamento del país (Ver selección roja en Figura 3). Una vez seleccionado el departamento, el aplicativo genera un segundo mapa de intensidad que permite observar el IPM estimado para cada municipio del departamento seleccionado por el usuario. En la Figura 3 se presenta el mapa de calor asociado a la selección de Caquetá en el tablero de control.

Resulta importante señalar que si el usuario no especifica un departamento el aplicativo genera el mapa de intensidad de Cundinamarca sin incluir Bogotá por ser Distrito Capital.

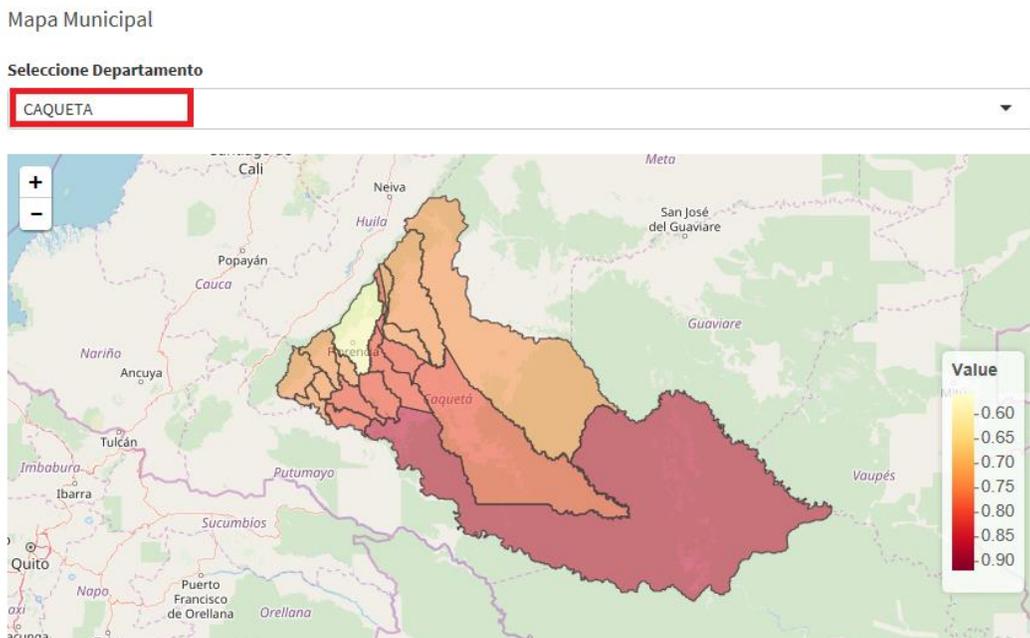


Figura 3: Mapa de la estimación directa del IPM a nivel municipal – Censo 2005.

En la siguiente etapa del proyecto se empleó la Encuesta Nacional de Calidad de Vida de 2014, 2015, 2016 y 2018 para estimar un proxy del índice de pobreza multidimensional en las ciudades principales del país (Bogotá, Medellín, Cali y Barranquilla) considerando el conjunto de variables que se presenta en la Tabla 4.

Nombre	Pregunta ECV
P6219	¿Cuál es el nivel educativo más alto alcanzado?
P6219s1	¿Cuál es el último año o grado aprobado en ese nivel?
P6213	¿En qué nivel está matriculado?
P6213s1	¿Qué grado cursa?
p6160	¿Sabe leer y escribir?
P8586	¿Actualmente estudia (asiste al preescolar, escuela, colegio o universidad)?
p6091	¿Está afiliado, es cotizante o beneficiario de alguna entidad de seguridad social en salud?
p6176	¿Dónde o con Quién permanece durante la mayor parte del tiempo entre semana?
P8578	¿Recibe almuerzo en algún establecimiento de forma gratuita o por un pago simbólico? Niños de 0 y 4 años.
P6240	¿En qué actividad ocupó la mayor parte del tiempo la semana pasada?
P6250	Además de lo anterior, ¿Realizó la semana pasada alguna actividad paga por una hora o más?
P6260	Aunque no trabajó la semana pasada, por una hora o más en forma remunerada, ¿Tenía durante esa semana algún trabajo o negocio por el que recibe ingresos?
P6270	¿Trabajó la semana pasada en un negocio familiar por una hora o más sin que le pagaran?
P6800	¿Cuántas horas a la semana trabaja normalmente en ese trabajo?
P6850	¿Cuántas horas trabajó durante la semana pasada en ese trabajo?
P6351	Si le hubiera resultado algún trabajo ¿estaba disponible la semana pasada para empezar a trabajar?
P7250	¿Durante cuánto tiempo ha estado o estuvo buscando trabajo?
P6920	¿Está cotizando actualmente a un fondo de pensiones?
p6091	¿Está afiliado, es cotizante o es beneficiario de alguna entidad de seguridad social en salud?
P8553	En los últimos 30 días, ¿tuvo alguna enfermedad, accidente, problema odontológico o algún otro problema de salud que no haya implicado hospitalización?
P6143	Para tratar ese problema de salud, ¿qué hizo principalmente?
P8520S5	¿Con cuáles de los siguientes servicios públicos, privados o comunales cuenta la vivienda? Acueducto
P8530	El agua para preparar los alimentos la obtienen principalmente de:
P8520S3	¿Con cuáles de los siguientes servicios públicos, privados o comunales cuenta la vivienda? Alcantarillado
P8526	¿Con qué tipo de servicio sanitario cuenta el hogar?
P4015	Material predominante de los pisos
P4005	Material predominante de las paredes exteriores
P5010	¿En cuántos de esos cuartos duermen las personas de este hogar?

Tabla 4: Variables de la ENCV empleadas para estimar el IPM en las ciudades principales de Colombia.

Los resultados obtenidos se presentan en la Tabla 5.

Ciudad	2014	2015	2016	2018
Barranquilla	0.6327	0.6345	0.6773	0.7003
Bogotá	0.9060	0.9130	0.9410	0.9480
Cali	0.7009	0.7168	0.7273	0.7442
Medellín	0.7971	0.8407	0.8691	0.8717

Tabla 5: Valores estimados del IPM en las ciudades principales de Colombia.

Los resultados obtenidos en la segunda etapa del proyecto fueron empleados como información auxiliar en el proceso de estimación final que resulta del ajuste de modelos SAE.

En la tercera etapa del proyecto se empleó la Base de Datos de SISBEN IV entregada por la Subdirección de Promoción Social y Calidad de Vida del DNP para realizar la estimación directa del IPM para los 470 municipios de los que se tiene información hasta el momento. Resulta importante señalar que la información de los hogares que contiene la base de SISBEN IV no tiene factores de expansión debido a que es el resultado de un censo de los hogares en mayores condiciones de vulnerabilidad de los municipios. Lo anterior implica que cualquier estimación generada a partir de esta base no “expande” a todos los hogares del municipio pues excluye a los hogares de estratos 3, 4, 5 y 6 que puedan tener los municipios en Colombia.

Para dar una solución temporal que permita hacer uso de los modelos SAE se realizó el cálculo de factores de expansión asumiendo dos posibles diseños de muestreo:

- Muestreo aleatorio simple (MAS)
- Muestreo estratificado - Se considera como estrato de diseño el estrato socioeconómico de los municipios.

Assumiendo estos factores de expansión, se ajustó el modelo SAE de área definido en la sección Metodología considerando como variables auxiliares las generadas por el DNP en el estudio de caracterización territorial que permitió identificar características propias de cada municipio y departamento en relación con seis temáticas que precisan las condiciones territoriales en las cuales se espera adelantar intervenciones sectoriales: funcionalidad urbano-regional, dinámica económica, calidad de vida, medio ambiente, seguridad, y desempeño institucional.

La base de datos empleada se encuentra disponible en el siguiente enlace:

https://www.dnp.gov.co/programas/desarrollo-territorial/Estudios-Territoriales/Estudios-y-Ejercicios/Paginas/Tipologias.aspx?fbclid=IwAR2jFirY3EVb8VxEqNwRqkfKu9DgCF7WINTeS5tdF8kDy_sp0SLXuXoLxnM

Conclusiones y recomendaciones

1. Los modelos SAE pueden ser empleados para realizar la estimación de IPM en aquellos municipios en los cuales hay ausencia de información en la base de SISBEN IV (actualmente 650 municipios). Para tal fin se hizo necesaria la generación de factores de expansión construidos con los pronósticos de población reportada por el DANE.

2. Actualmente los pronósticos obtenidos muestran unas diferencias considerables respecto al IPM estimado empleando la información del CENSO 2005. Estas diferencias se observan principalmente en los municipios intermedios para los cuales no se tiene ninguna información en la Base de SISBEN IV ya que el modelo SAE no tuvo la oportunidad de entrenarse.
3. El modelo SAE puede ser ajustado según las covariables que la Subdirección de Ordenamiento y Desarrollo Territorial y a la Subdirección de Promoción Social y Calidad de Vida del DNP consideren deban ser analizadas dentro del estudio.

Socialización

Los resultados obtenidos de los modelos SAE fueron enviados a la Subdirección de Ordenamiento y Desarrollo Territorial y a la Subdirección de Promoción Social y Calidad de Vida del DNP para su revisión.

Adicionalmente los resultados fueron socializados en un seminario técnico realizado en el DNP al igual que un evento denominado "*Desafíos e innovaciones en la medición de la pobreza y el seguimiento del ODS1*" en el marco el Seminario Regional de la CEPAL