



**El futuro
es de todos**

DNP
Departamento
Nacional de Planeación



El futuro
es de todos

DNP
Departamento
Nacional de Planeación

Un modelo SAE para pronosticar el IPM de los municipios de Colombia vía registros administrativos

Unidad de Científicos de Datos
Dirección de Desarrollo Digital

Septiembre, 2019



- 1. Descripción del proyecto**
- 2. Definición del IPM**
- 3. Introducción a los modelos de estimación en áreas pequeñas**
- 4. Resultados obtenidos**
- 5. Conclusiones**

1. Descripción del proyecto

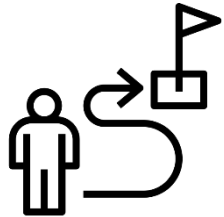


Cálculo del Índice de Pobreza Multidimensional (IPM) a nivel municipal



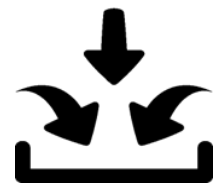
Objetivo

Identificar un modelo estadístico que permita estimar el índice de pobreza multidimensional o alguna otra medida de pobreza para todos los municipios de Colombia



Metodología

1. Estimación en áreas pequeñas
2. Modelos Lineales Generalizados Mixtos



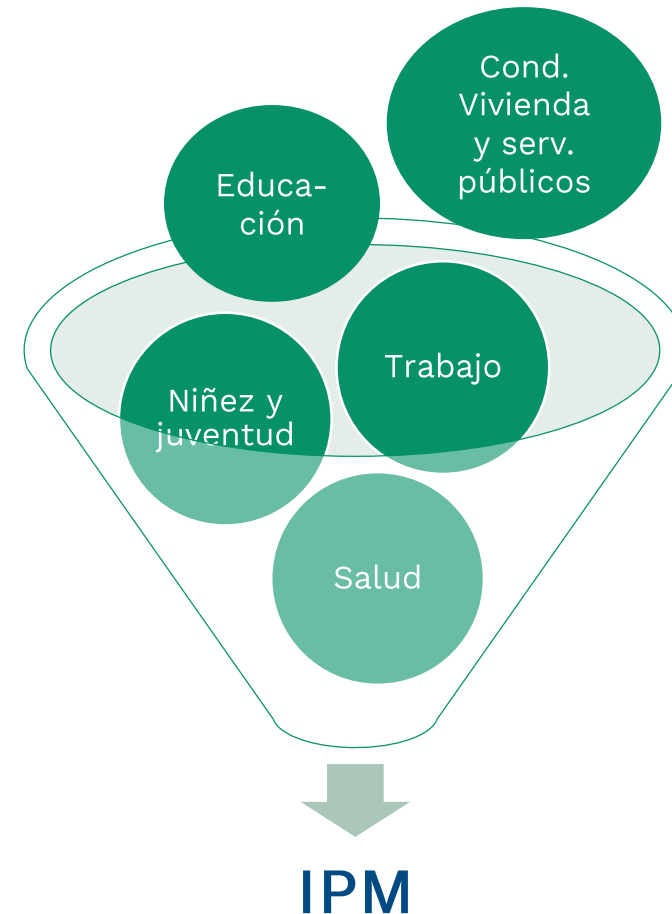
Insumos

1. SISBEN 4
2. ECV (Encuesta de Calidad de Vida)
3. GEIH (Gran Encuesta Integrada de Hogares)

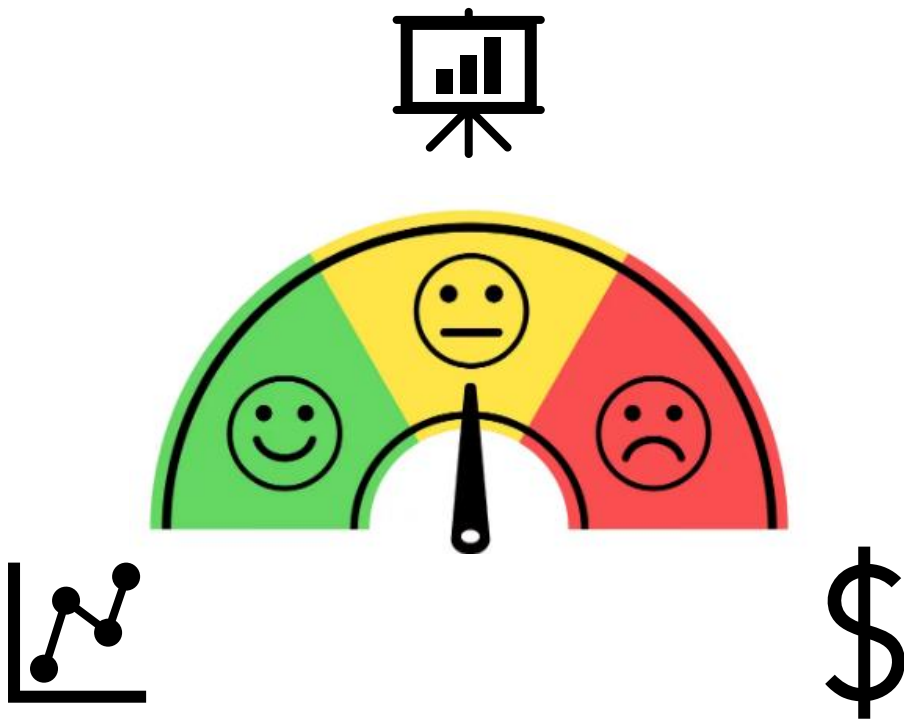
Cálculo del Índice de Pobreza Multidimensional (IPM) a nivel municipal

Introducción

El **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.



Cálculo del Índice de Pobreza Multidimensional (IPM) a nivel municipal



Métodos de estimación en áreas pequeñas:

1. La información del CENSO está desactualizada.
2. Los registros administrativos de la ECV y la GEIH son representativos para las grandes ciudades y áreas metropolitanas.
3. Registro administrativo actualizado y disponible a nivel municipal: SISBEN IV

2. Definición del IPM



Índice de pobreza multidimensional (IPM)

Dimensiones y variables que lo conforman



1. Logro educativo
2. Analfabetismo



1. Asistencia escolar
2. Rezago escolar
3. Acceso a servicios para el cuidado de la primera infancia
4. Trabajo infantil



1. Desempleo de larga duración
2. Empleo formal



1. Aseguramiento en salud
2. Acceso a servicio de salud dada una necesidad



1. Acceso a fuente de agua mejorada
2. Eliminación de excretas
3. Material de pisos
4. Material de paredes exteriores
5. Hacinamiento crítico

Índice de pobreza multidimensional (IPM)

Privaciones

Dimensión	Variable	Un hogar se considera privado cuando...
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.
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

Ejemplos de privaciones

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).

3. Introducción a los modelos SAE

Generalidades

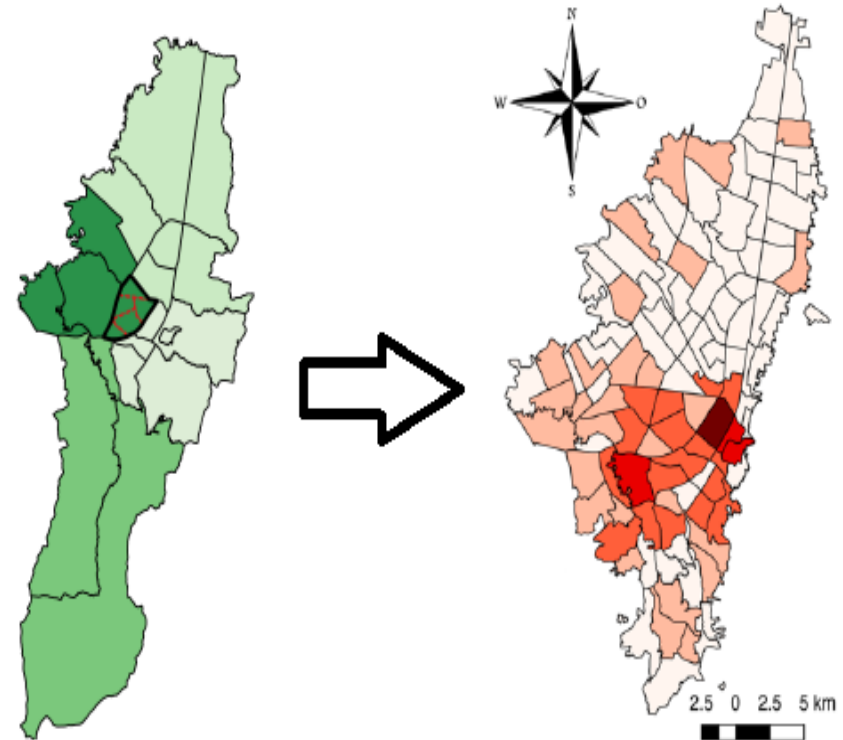
Las encuestas basadas en un diseño de muestreo hacen uso de los estimadores directos para el cálculo de parámetros como totales y promedios.

Desventajas: si no se tiene suficiente información muestral de un dominio en particular los errores de muestreo pueden ser considerablemente altos.



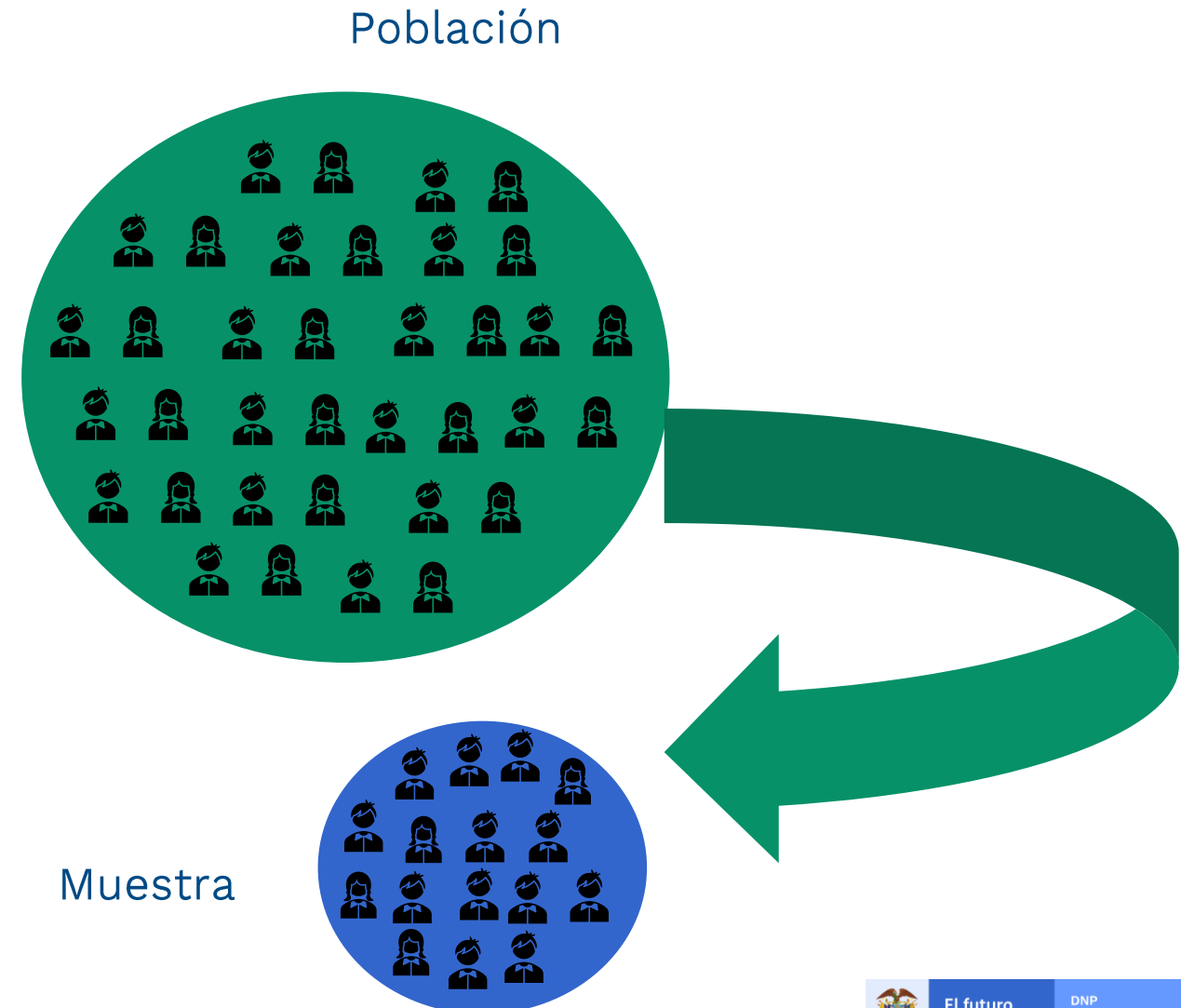
Generalidades

Muchas encuestas oficiales (como las reportadas por el DANE) fueron diseñadas para propósitos nacionales o regionales y sus diseños de muestreo no se pueden ajustar para obtener estimaciones con un mayor nivel de desagregación.



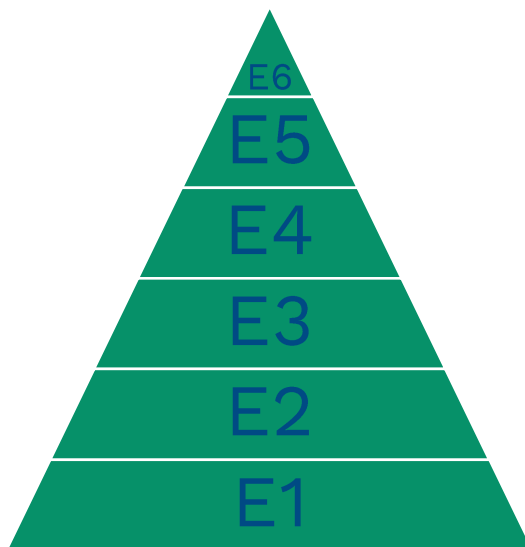
¿Qué es un área pequeña?

Según Rao y Molina (2016), un **dominio (área)** se denomina pequeña si el tamaño de muestra específico de este no es lo suficientemente grande como para garantizar estimadores directos con una precisión adecuada.

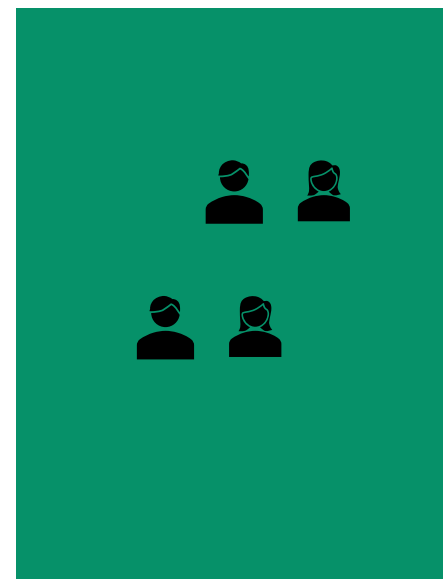


¿Qué es un área pequeña?

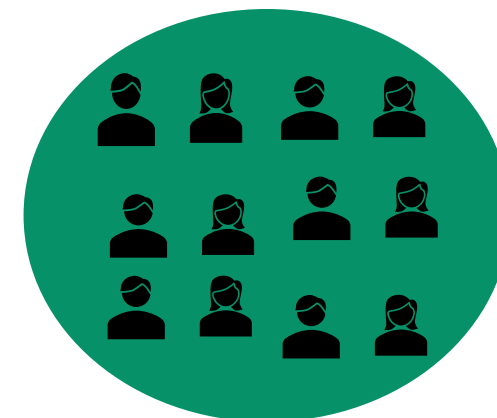
Área no siempre hace referencia a una agregación geográfica, también pueden ser subpoblaciones provenientes de una distribución social o económica.



Dominio 1



Dominio 2



El término pequeño hace referencia a la fracción entre el tamaño de muestra y la población total del dominio.

Estructura de un problema SAE

SAE es un problema de predicción o pronóstico.

El acceso de buenas fuentes de información auxiliar resulta crucial.

Información
agregada por
dominio



Estimación

Parámetros

1. Totales
2. Promedios
3. Proporciones

Microdatos



Estimación

Parámetros más
complejos

1. Razones
2. Diferencias de medias

Estimadores empleados

Estimadores directos

Estimadores basados únicamente en la información muestral del dominio específico.

Son estimadores insesgados.

$$\hat{t}_{y,\pi} = \sum_{i \in S} \frac{y_i}{\pi_i}$$

Problemas

1. Solo pueden ser empleados en aquellos dominios donde hay información muestral.
2. La varianza de estos estimadores tiende a ser grande en comparación con otros estimadores

$$\widehat{Y}_{dir} = \sum_{i \in S} \frac{\frac{y_i}{\pi_i}}{\sum_{i \in S} \frac{1}{\pi_i}}$$

Estimadores empleados

Estimadores sintéticos

Estimador basado tanto en la información muestral de un dominio específico como en un modelo de regresión (lineal) asociado a una o más covariables asociadas a la variable de interés.

Cuando se dispone de los totales de las variables auxiliares para la subpoblación de estudio, se puede definir un estimador sintético de tipo regresión para el total como: $\hat{Y}_{d,GRS} = X_d^t \hat{\beta}$

Población en el dominio d

Subpoblación
1

...

Subpoblación
K

Estimadores empleados

Estimadores compuestos

Estimadores diseñados para combinar las buenas propiedades de los estimadores directos ($\widehat{Y}_{D,i}$) y los estimadores sintéticos ($\widehat{Y}_{S,i}$).

$$\widehat{Y}_{C,i} = \gamma_i \widehat{Y}_{D,i} + (1 - \gamma_i) \widehat{Y}_{S,i}$$



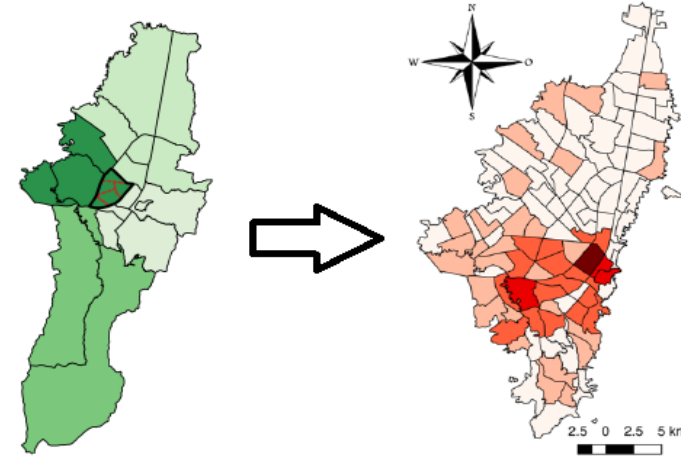
$$\hat{\gamma}_i = 1 - \frac{Var[\hat{Y}_{D,i}]}{(\hat{Y}_{M,i} - \hat{Y}_{D,i})^2}$$

Nota: $0 \leq \gamma_i \leq 1$, es un parámetro de contracción que pondera el peso relativo de los estimadores agregados en el compuesto.

Modelos SAE

Modelos a nivel de área

1. Modelos basados en estimadores directos, compuestos o sintéticos y covariables agregadas a nivel de dominio.
2. Modelos que no involucran ningún problema de confidencialidad.



$$Y_d^{FH} = \mathbf{X}'_d \beta + U_d + \epsilon_d, \quad d = 1, 2, 3, 4, \dots, D.$$

$$U_d \sim N(0, \sigma_U^2), \epsilon_d \sim N(0, \Sigma_d)$$

Modelos SAE

Modelos a nivel de unidad

1. Modelos que utilizan la información de la muestra directamente.
2. Se requiere de información auxiliar a nivel del dominio de interés.
3. Dependiendo del estudio se pueden tener problemas de confidencialidad.

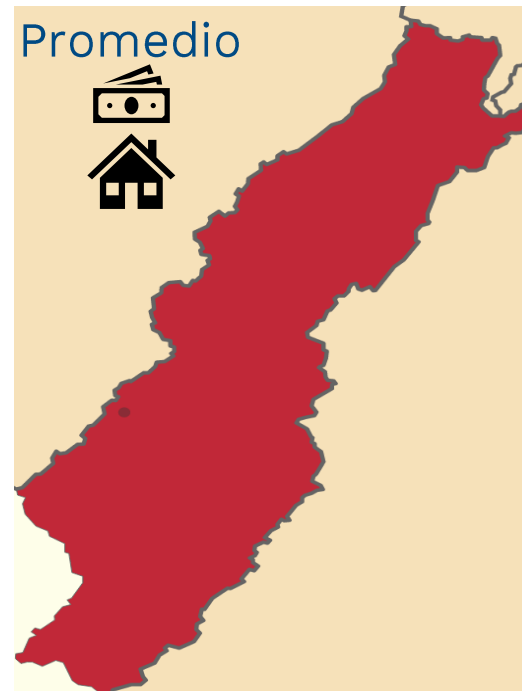
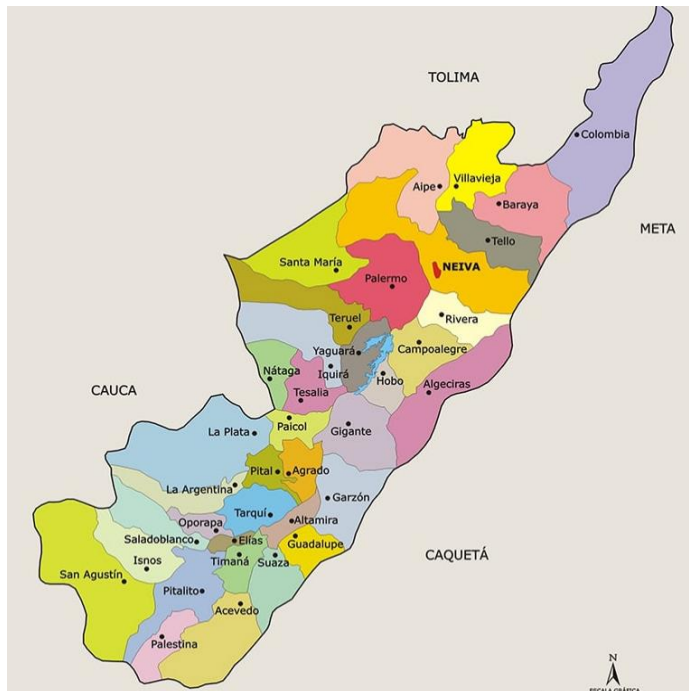
$$y_{ij} = x_{ij}^t \beta + v_i + e_{ij}$$

Supuestos

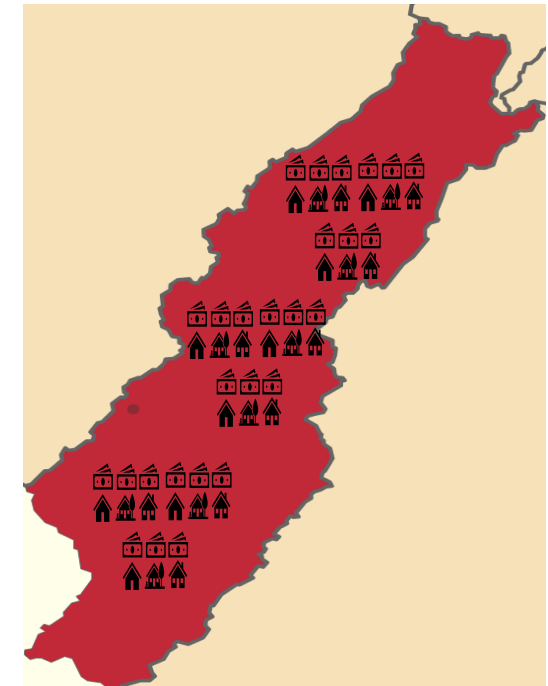
$$v_i \sim N(0, \sigma_U^2) \quad e_{ij} \sim N(0, \sigma_j^2)$$



Diferencia entre modelos SAE



Modelos a nivel de área



Modelos a nivel de unidad

4. Resultados



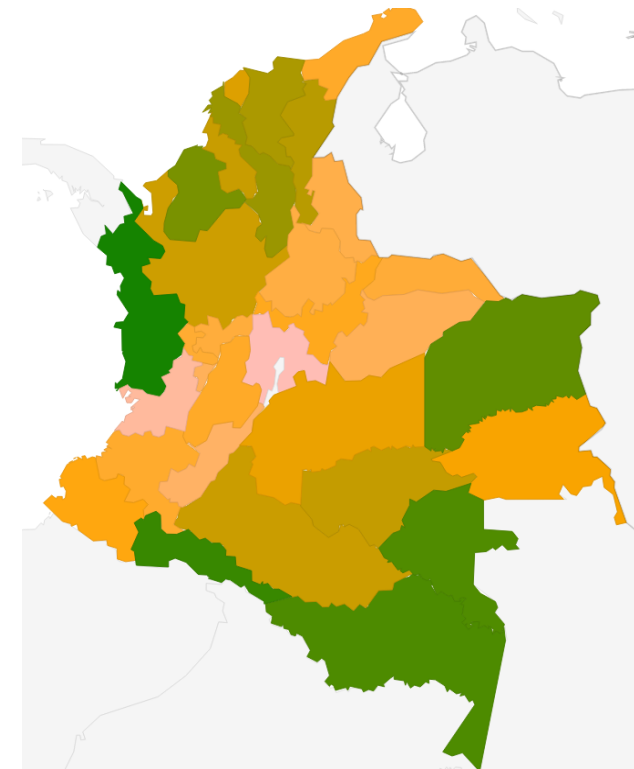
Fuentes de información

A continuación se presenta el listado de fuentes de información empleadas y su objetivo en el estudio

1. Censo poblacional 2005

Se empleó la información del censo 2005 para calcular el IPM municipal y compararlo con la estimación reportada por el DANE.

Se observaron diferencias a partir de la cuarta cifra decimal.



Fuentes de información

A continuación se presenta el listado de fuentes de información empleadas y su objetivo en el estudio

2. Terridata

Base de datos con indicadores estandarizados y comparables que dan cuenta de fenómenos sociales, económicos, ambientales y del bienestar de los hogares y personas en 1101 municipios del país.

Se empleó como variables auxiliares en los modelos SAE a nivel de área empleados.



Fuentes de información

3. ECV 2018

La ECV 2018 recopiló información para las siguientes dimensiones del bienestar de los hogares: i) Características de la vivienda, ii) Servicios del hogar, iii) Características y composición del hogar, iv) Salud, v) Atención integral de los niños y niñas menores de 5 años, vi) Educación, vii) Fuerza de trabajo, viii) Tecnologías de información y comunicación (TIC), ix) Trabajo infantil, x) Tenencia y financiación de la vivienda, xi) Condiciones de vida y tenencia de bienes.

DANE
INFORMACIÓN PARA TODOS

Fuentes de información

4. SISBEN IV

La base de datos de SISBEN IV tiene información que permite construir las dimensiones del IPM para los hogares más vulnerables de los 450 municipios priorizados en el proceso de recolección.



Modelo SAE ajustado

Covariables consideradas inicialmente:

- a) Crecimiento Poblacional
- b) Número de viviendas
- c) Número de hogares
- d) Acueducto
- e) Alcantarillado
- f) Energía
- g) Disparidades Económicas
- h) Densidad Poblacional
- i) Hurtos X 100mil habitantes
- j) Homicidios X 100mil habitantes
- k) Secuestros X 100mil habitantes
- l) Penetración Internet
- m) Inversión Ambiental
- n) Índice Desempeño Fiscal

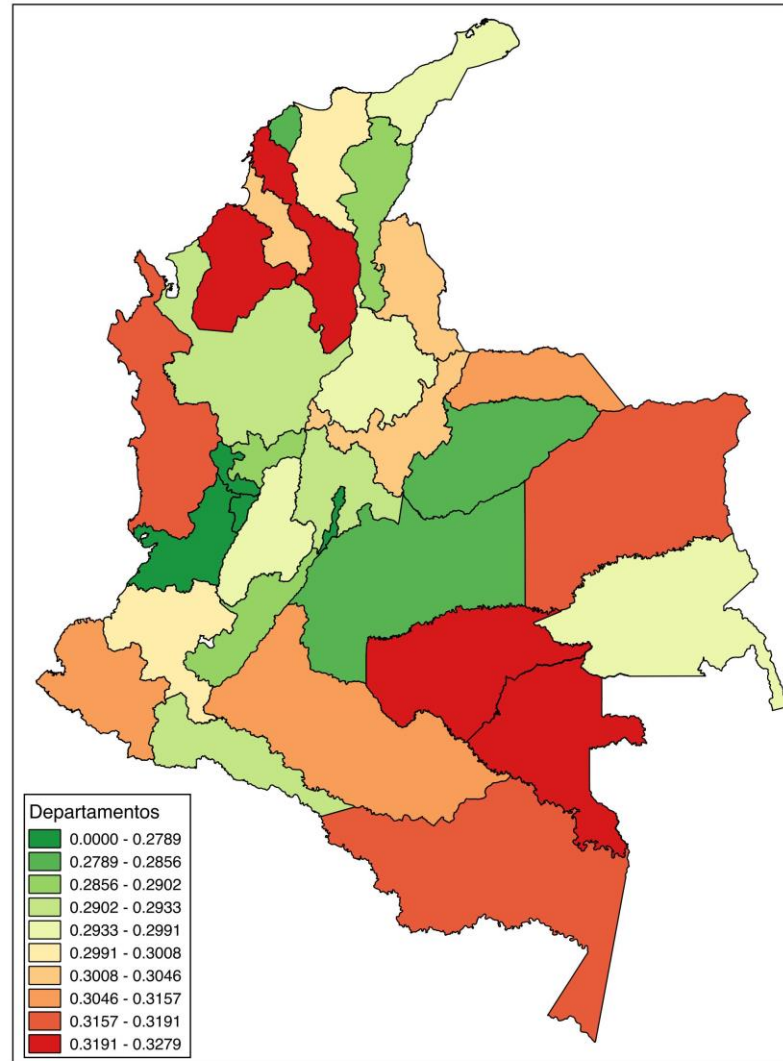
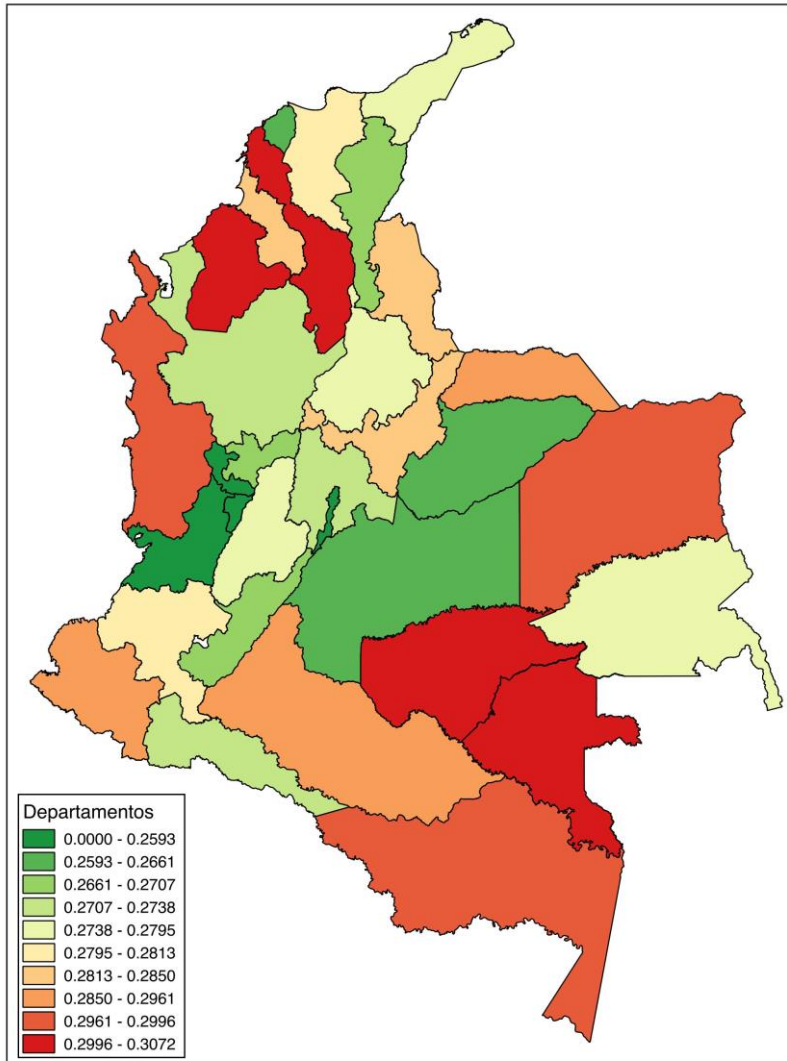
Modelo SAE 1:

$IPM \sim Población + Penetración Internet +$
 $Disparidades Económicas + Índice Desempeño$
 $Fiscal + Vivienda + Hogar + Energía +$
 $Alcantarillado + Acueducto$

Modelo SAE 2:

$IPM \sim Vivienda + Hogar + Alcantarillado +$
 $Acueducto$

Resultados agregados de los modelos SAE ajustados



4. Resultados



Conclusiones



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 contruidos 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.



**El futuro
es de todos**

DNP
Departamento
Nacional de Planeación