



ALAP 2020

IX Congreso de la Asociación Latinoamericana de Población



9 a 11 diciembre

EL ROL DE LOS ESTUDIOS DE POBLACIÓN TRAS LA PANDEMIA DE COVID-19 Y
EL DESAFÍO DE LA IGUALDAD EN AMÉRICA LATINA Y EL CARIBE

Glenn Harry Amaya Cruz, DANE, ghamayac@dane.gov.co

Lelio Alejandro Arias Vizcaino, DANE, laariasv@dane.gov.co

Andryu Enrique Mendoza Beltran,

Yenny Andrea Marín Salazar, DANE, yamarins@dane.gov.co

Mariana Francisca Ospina Bohórquez, DANE, mfospinab@dane.gov.co

Lina María Sánchez Cespedes, DANE, lmsanchezc@dane.gov.co

1. INTRODUCCIÓN

En el año 2018 el Departamento Nacional de Estadística – DANE adelantó el Censo Nacional de Población y Vivienda CNPV, censando 44.164.417 millones de personas distribuidas en 14.243.223 hogares y 13.480.729 viviendas. Este tipo de operaciones estadísticas, siempre tienen como objetivo cumplir con el total de la cobertura, sin embargo, dado el nivel de complejidad que un censo demanda, esta condición no siempre se cumple y se dejan de registrar personas de manera aleatoria en el territorio nacional, lo que se denomina omisión censal. Así pues, el presente documento tiene como objetivo informar los aspectos metodológicos y resultados obtenidos en las diferentes estimaciones de la cobertura en el nivel municipal, así como las limitaciones presentadas y las estrategias empleadas para su tratamiento.

A nivel nacional, la omisión censal fue estimada 7,2% por medio del método de la conciliación y 8,5% por el método directo *Dual System*, el cual se tomó como valor oficial. La desagregación subnacional, a municipios, de este valor se hizo a partir del análisis espacial de los factores generadores de omisión. Estos factores o tipos de omisión son tres: la ausencia de todas las personas en las viviendas a la hora de la encuesta, las limitaciones de acceso geográfico a algunas zonas y la omisión de personas dentro del hogar de manera aleatoria. En cada factor se empleó la información georreferenciada del marco geoestadístico y los resultados censales como fuentes para estimar la no cobertura, apoyado en covariados geográficos y herramientas de análisis espacial. Así, se aplicaron modelos econométricos y de autocorrelación espacial que permitieron aproximar la población.

Para la primera tipología, correspondiente a viviendas ocupadas con todas las personas ausentes, se usaron los modelos de promedios locales con granularidad geográfica que permitieron estimar la población que residía en estas viviendas. Para la segunda tipología, omisión en zonas geográficas no visitadas e incompletas, se utilizaron herramientas geoespaciales que identificaron la concentración y clúster de unidades no visitadas, como también covariables espaciales que determinaron la densidad de población. Para la tipología denominada subnumeración, u omisión de personas dentro del hogar, se usaron modelos econométricos basados en la Encuesta de Calidad de Vida- ECV2018. Estos modelos

estimaron el número de personas omitidas dentro del hogar. Finalmente, se sumaron las tres tipologías por municipio y clase para hallar el total de la omisión.

Adicionalmente, la estimación de la omisión en rutas¹ se fundamentó en la transferencia de conocimiento realizada por el equipo de GRID 3 de la Universidad de Southampton de Inglaterra, quienes a través del apoyo del Fondo de Población de Naciones Unidas UNFPA, trabajaron conjuntamente con los funcionarios DANE en modelos estadísticos y probabilísticos del comportamiento de covariables en el espacio. Estos modelos integraron la información georreferenciada de las viviendas censada y unidades esperadas en el marco censal, apoyados en otras fuentes disponibles, como imágenes de satélite de intensidad de luces nocturnas, densidad de infraestructura social (escuelas y centros de salud), distancia a centros poblados, y el índice de Necesidades Básicas Insatisfechas NBI, todas ellas variables antropocéntricas que permitirán correlacionar el volumen de población, a niveles geográficos desagregados principalmente en áreas de rutas no visitadas.

El documento está dividido en siete secciones incluyendo la presente introducción. En la segunda sección se presentan los tres tipos de omisión. En la tercera se explica la estimación de la omisión en las áreas no visitadas e incompletas, tanto en barrido como en rutas. La sección 4 presenta el procedimiento para estimar la omisión en viviendas ocupadas con personas ausentes (VOCPA). Por su parte, la sección 5 explica paso a paso la estimación de la subenumeración dentro de los hogares. La sexta sección presenta la estimación total de la omisión para los municipios. Por último, se concluye en la sección 7.

2. Estimación de cada tipología de omisión censal para el nivel municipal

La desagregación de la omisión censal a nivel municipal (cabecera y resto), se fundamentó en la conceptualización de los factores que propiciaron dejar de censar personas dentro del territorio nacional. El resumen de los factores de omisión se presenta en el siguiente esquema²:

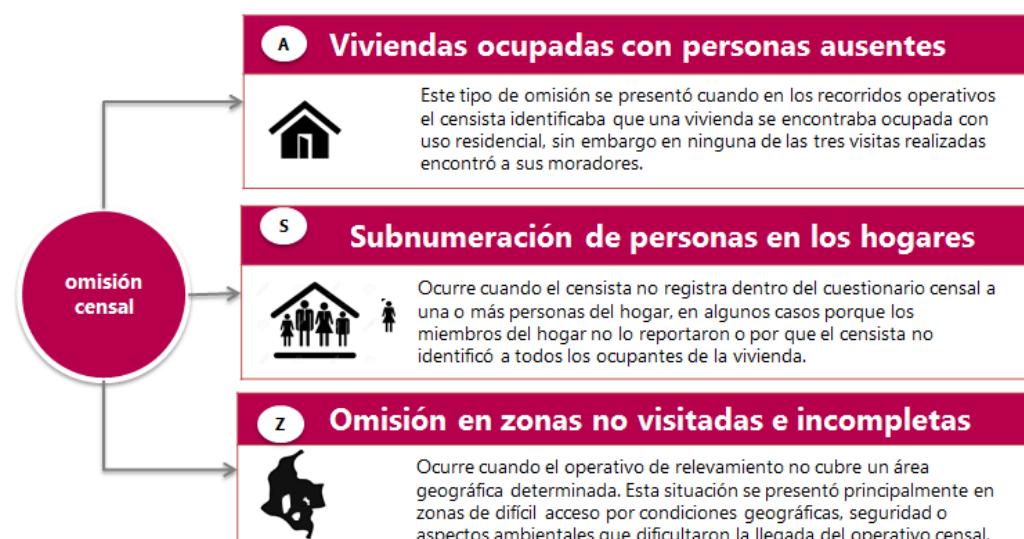


Figura 1. Esquema resumen de los factores de omisión CNPV 2018

¹ Método de recolección operativa para las áreas de mayor complejidad de acceso, el cual consiste en recoger la información sobre un recorrido estructurado a lo largo de un río y sus afluentes; o vías con sus caminos y senderos, generando un área de influencia en la cual se encuentran localizada la población.

² Basado en los errores de cobertura señalados en el documento: Los datos geográficos: alcances, limitaciones y métodos de evaluación. (CELADE, 2019, pág. 35)

Fuente: DANE - Dirección de censos y demografía CNPV 2018

Bajo este marco, el total de personas no censadas a nivel municipal (Y_m) se obtiene como la agregación de estos tres factores, con la siguiente forma:

$$Y_m = Z_m + A_m + S_m$$

Considerando la anterior expresión, tenemos que la omisión censal a nivel municipal corresponde a:

$$E_m = R_m + Y_m, \quad \% \text{ Omisión municipal} = \frac{E_m - R_m}{E_m} * 100$$

Donde,

E_m = Población Estimada para el municipio

R_m = Población Censada para el municipio

Estos tres factores que incidieron en la omisión de personas del CNPV 2018, fueron abordados metodológicamente de manera independiente, con el aprovechamiento intensivo de la información del marco censal, y la georreferenciación de las unidades observadas, aplicando modelos econométricos y de correlación espacial que permiten con valores conocidos en el espacio de un fenómeno o variable determinar la tendencia de esta misma en los lugares donde no se conoce como se muestra en el siguiente esquema:

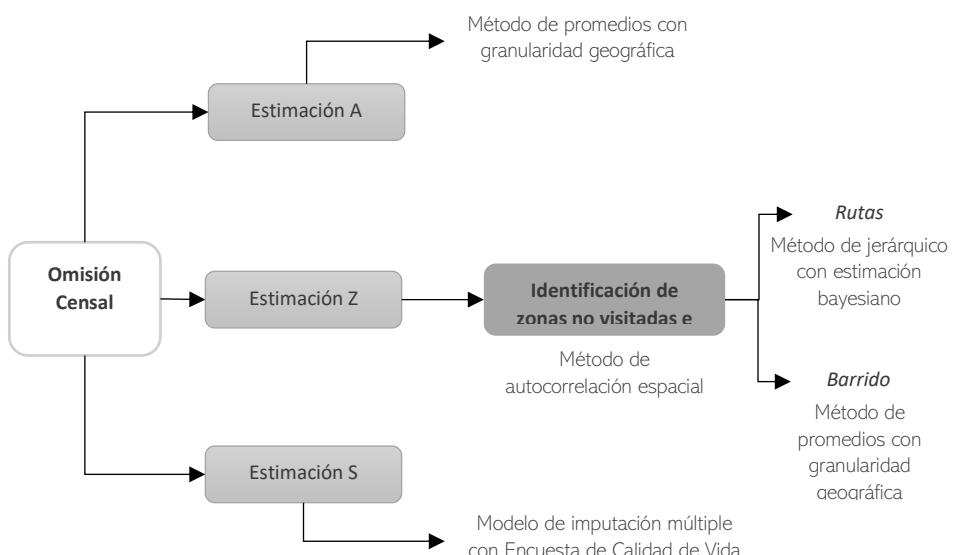


Figura 2. Diagrama general de aplicación metodológica para estimación de la omisión.

Fuente: DANE - Elaborado a partir de la implementación metodología de la estimación de la omisión censal.

A continuación, se describe el componente metodológico y los resultados de las estimaciones de la omisión en cada uno de los tres componentes.

3. Omisión de áreas geográficas no visitadas e incompletas (Z)

La estimación de la omisión en las áreas no visitadas e incompletas de barriado requirió una selección previa de las áreas susceptibles de imputación de población, es decir zonas no visitadas en donde se esperaban unidades según la información del marco censal y no censó ninguna persona, o zonas en donde con un porcentaje de observación superior al 0% e inferior a los 90%. El sistema de monitoreo

y control geográfico permitió identificar las zonas no visitadas e incompletas a partir de un indicador por cada unidad de cobertura³ con lo cual fue posible clasificar las siguientes situaciones:

Tabla 1. Clasificación de categorías de cobertura geográfica.

unidades esperadas según el marco censal	Unidades Censadas	Clasificación de categorías de cobertura
A	a	Cobertura censal 100%
A	0	Zonas no visitadas
A	b	Zonas censadas incompletas (si $b/a < 90\%$)
0	a	Zonas de marcos incompleto
B	a	Zonas de marcos incompletos
0	0	Zonas no susceptibles de imputación

Considerando $a > b$

Fuente: DANE - Dirección de censos y demografía CNPV 2018

Sin embargo, una vez realizados análisis exploratorio de la cobertura geográfica, se determinó la necesidad de establecer adecuadamente las zonas en donde se tenía $a > b$, siendo b las unidades residenciales. Es decir, se requirió un proceso de determinación de áreas incompletas sujetas de estimación con el fin de no sobreestimar zonas con otro tipo de usos.

Para ello se empleó una metodología de correlación espacial, a partir de lo cual se clasificaron las áreas que posteriormente fueron sujetas de estimación de población, abordadas por dos técnicas diferenciadas, según el método de recolección, como se desarrolla a continuación:

3.1 Análisis espacial para determinación de áreas incompletas

Este apartado presenta una metodología de selección de las áreas incompletas, basado en las relaciones espaciales de las unidades censales y el reconocimiento de su distribución particular por departamento y municipio.

A partir de la primera Ley de la Geografía de Tobler que señala que "Geográficamente todo está relacionado con todo, pero los objetos cercanos están más relacionados entre sí que los lejanos" (Tobler, 1970). Se establece uno de los principios de dependencia espacial el cual señala que "los sitios cercanos tienen valores más similares que los distantes... La ausencia de correlación espacial se refleja en el hecho de que la distancia entre los sitios no tiene influencia en la relación de sus valores" (Giraldo, 2011). La anterior premisa es el fundamento del desarrollo conceptual de la propuesta para la selección de áreas incompletas sujetas de estimación de población. "La esencia de la autocorrelación es analizar la variabilidad de un fenómeno a través del espacio geográfico... para comprender cómo se distribuye el fenómeno en el espacio analizado, y en qué grado los elementos locales pueden verse afectados por sus vecinos" (Siabato, 2019). Según (Siabato, 2019) Se definen tres tipos de autocorrelación espacial:

- *Positiva, cuando el fenómeno analizado tiende a agruparse en conglomerados o clústeres. Esto significa que, para un sitio determinado, una propiedad tiene valores muy similares en sus vecindades (Siabato et al, 2018).*
- *Negativa, cuando la variable analizada tiende a estar dispersa, es decir, los valores de la variable en las vecindades de una cierta localización son diferentes (Siabato et al, 2018).*
- *Aleatoria, lo que significa ausencia de autocorrelación: el valor de un atributo o propiedad en un lugar determinado no influye en los valores de sus vecinos (Siabato et al, 2018).*

³ Mínima unidad de planeamiento en el operativo censal, para cabeceras y centros poblados corresponde a manzanas y para zonas rurales a predios

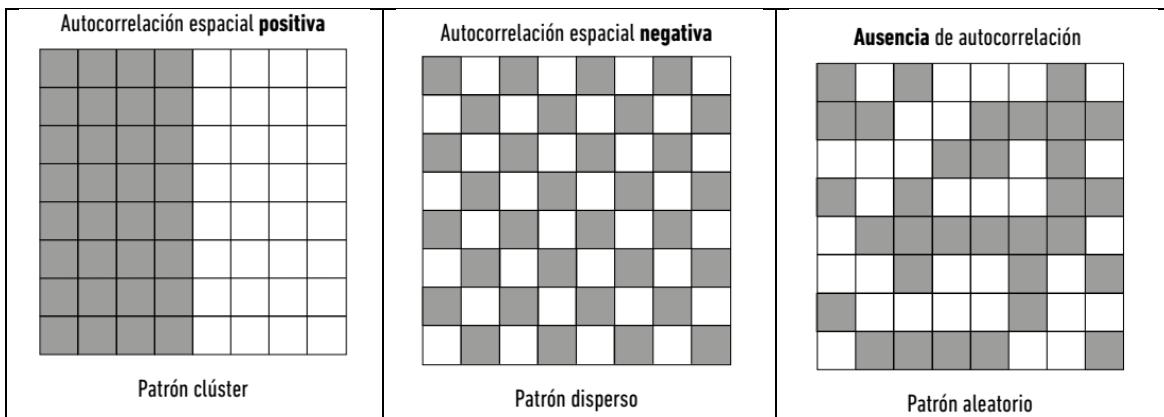


Figura 3. Tipos de autocorrelación espacial

Fuente: Siabato et al, 2019

Bajo el anterior marco conceptual y con el fin de determinar las unidades censales incompletas para el ejercicio de localización de la omisión, se completará el siguiente procedimiento:

- *Definición de variables con significancia espacial en las unidades censales (UC).*
- *Cálculo de la matriz de pesos espaciales.*
- *Cálculo de los índices de Moran y mapa de clusters, outliers y entidades no significativas.*
- *Operaciones espaciales posteriores.*

3.1.1 Definición de variables con significancia espacial

Mediante las herramientas de estadística espacial, se desea encontrar las áreas de cobertura incompleta (porcentaje de edificaciones obtenidas mayor a 0% y menor a 90%) en las que se cumplan los siguientes criterios:

- *El porcentaje de cobertura debe ser tan bajo como sea posible.*
- *Las unidades con porcentaje de cobertura bajo deben ser vecinas de otras unidades de cobertura con valores bajos.*
- *Las unidades con porcentaje de cobertura bajo deben ser vecinas de otras unidades en donde el porcentaje de otros usos (comercial, industrial, etc.) sea bajo. O, en otras palabras, unidades donde haya un predominio de viviendas.*

Lo anterior hace necesario el uso de las siguientes variables, todas provenientes de la operación censal:

- *PORCENT_U: porcentaje de edificaciones censadas con respecto a las edificaciones esperadas.*
- *P_NR_OBT: porcentaje de otros usos con respecto a las edificaciones obtenidas.*

Se escogió las edificaciones obtenidas, debido a que en estas unidades se tiene un grado mayor de certeza en cuanto al conteo de las edificaciones.

3.1.2 Matriz de pesos espaciales

La *matriz de pesos espaciales* describe el nivel de asociación espacial o *vecindad* de las entidades consideradas. Existen varios criterios de vecindad, entre los que se destacan:

- *Contigüidad física de las entidades.*
- *Distancia: la dependencia espacial se debilita como función de la distancia.*
- *Ponderación de límites comunes.*

La matriz de pesos espaciales es fundamental para los cálculos de los índices locales de Moran, por lo que su adecuada determinación es crucial para la obtención de resultados fiables. Dado que los polígonos que representan las unidades censales en general no son contiguos (están separados por vías, ríos, etc.), el criterio adecuado para el cálculo de la matriz de pesos espaciales es la distancia entre sus centroides respectivos. Se escogió la opción de Adaptive Kernel por tener una mayor variedad de opciones para definir la matriz de pesos (función kernel uniforme y uso del número máximo de vecinos cercanos para definir el ancho de banda).

3.1.3 Cálculo de los índices de Moran locales univariados y bivariados

El Índice Global de Moran univariado es una Medida de la Autocorrelación espacial de una variable, basada en las ubicaciones y los valores de las entidades, de manera simultánea. Propuesto por P. Moran en 1948, su valor permite determinar si el patrón de la variable es aglomerado, disperso o aleatorio (ESRI, 2019).

Está dado por la expresión:

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

$$W = \sum_i \sum_j w_{ij}$$

En donde,

x_i, x_j : valor de la variable en la ubicación i, j

\bar{x} : valor promedio de la variable

w_{ij} : peso espacial entre i, j, dado por la matriz de pesos

N: número total de observaciones

En tanto el Índice Global de Moran Bivariado permite determinar la variación de un fenómeno a través del espacio geográfico, considerando la afectación de una segunda variable. Es importante aclarar que el índice bivariado no tiene en cuenta la correlación inherente entre las dos variables, sino entre el valor de una de ellas en un lugar y el valor de la otra en las vecindades (Siabato et al, 2018). La expresión del índice bivariado es la siguiente:

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}$$

$$W = \sum_i \sum_j w_{ij}$$

En donde,

x_i, x_j, y_i, y_j : valor de la variable x, y, en la ubicación i, j

\bar{x}, \bar{y} : valor promedio de la variable

w_{ij} : peso espacial entre i, j, dado por la matriz de pesos

N: número total de observaciones

Los índices de Moran bivariados, relacionando las variables porcentaje de cobertura y porcentaje de otros usos obtenidos, se obtuvieron con el fin de encontrar los clusters en donde se encuentren valores bajos de porcentaje de cobertura rodeados de valores bajos de porcentajes de otros usos y viceversa como se muestra en el siguiente ejemplo:

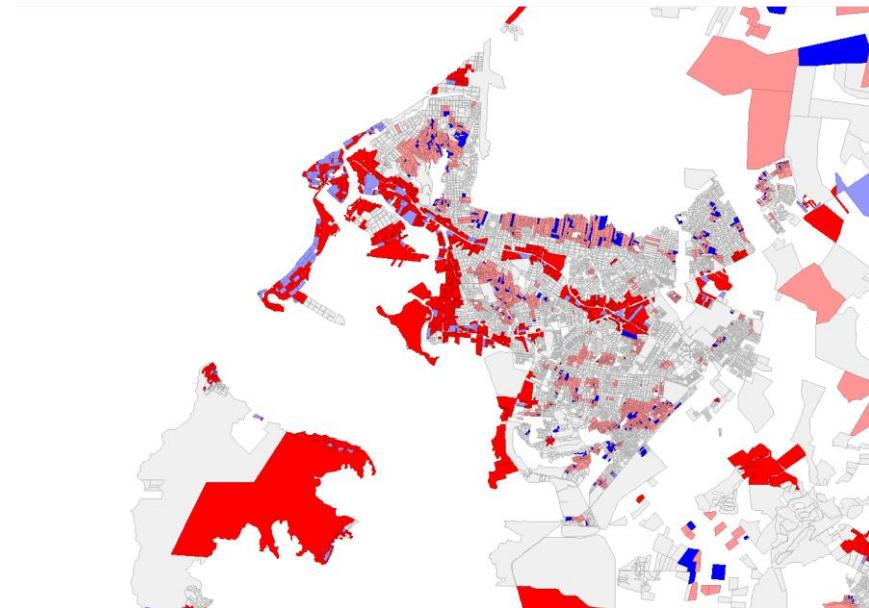


Figura 4. Mapa de clusters y outliers de Cartagena, relacionando P_U_NORM/P_NR_OBT

Fuente: DANE - Dirección de Geoestadística

En el caso del anterior ejemplo, correspondiente a la ciudad de Cartagena, los mapas muestran clusters High-High (Alto-Alto) de color rojo intenso en toda la zona hotelera y portuaria de la ciudad, lo que significa que en estas zonas predominan altos porcentajes de cobertura y altos porcentajes de otros usos en las vecindades.

Por otra parte, se encuentran aglomeraciones Low-Low (Bajo-Bajo) en zonas residenciales, lo que indica la existencia de unidades censales con bajos porcentajes de cobertura con bajos porcentajes de otros usos en las vecindades, o dicho de otra forma, de altos porcentajes de viviendas. Estas áreas son las que interesan para el análisis.

También hay valores atípicos Low-High y High-Low, pero estos no son importantes, puesto que representan bajos porcentajes de cobertura rodeados de otros usos, o altos porcentajes de cobertura rodeados de viviendas.

Finalmente, se encuentran las unidades censales en color gris, que indican una ausencia de correlación espacial y por tanto una aleatoriedad que no es de interés para el ejercicio.

3.2 Omisión en zonas no visitadas e incompletas de barrido

3.2.1 Planeamiento del metodológico

La estimación de población en las zonas no vistadas e incompletas por el método de barrido, parte de la premisa de que el tamaño de personas por hogar se encuentra influenciado por los niveles socioeconómicos, servicios sanitarios y características de las viviendas del entorno inmediato de estas

unidades, para lo cual se plantea aplicar un estimador de promedio local con granularidad geográfica que mantienen las características de tamaño de personas por vivienda de acuerdo con su contexto del municipio y la clase.

Si bien, el tamaño del hogar muestra un comportamiento diferenciado en el territorio, también se debe considerar que proporción de edificaciones son viviendas en dicho territorio, ya que no todas las edificaciones son usadas como residencia. Además, en el momento censal no todas las residencias se encontraban ocupadas. Estas consideraciones, tienen la finalidad de no sesgar la estimación del tamaño de hogar, para obtener el número de personas más probable dentro de las edificaciones de dichos territorios.

En consecuencia, se plantea un modelo que identifica en primer lugar las proporciones de viviendas por clase de uso residencial (PER), el numero promedio de estas viviendas que pueden estar ocupadas (PVO) y el promedio de personas ponderado por el número de hogares dentro de la vivienda (PPPH), obteniendo la siguiente expresión:

$$PO_X = E_e * PER * PVO * PPPH$$

Donde,

PO_X Población omitida en unidades de cobertura X

EE_e Edificaciones esperadas en las unidad de cobertura X

PER proporciones de viviendas de uso residencial por clase

PVO proporción de viviendas ocupadas por clase

PPPH promedio de personas ponderado por el número de hogares dentro de la vivienda por clase

Los pasos para obtener el tamaño de los hogares en zonas en la que se implementó el método de recolección de barrido son los siguientes:

- a) De las 19'900.593 edificaciones que se encuentran en el CNPV2018, se obtuvo la proporción de edificaciones de uso residencial (PER) y la proporción de viviendas ocupadas (PVO), por municipio y clase.
- b) Se estimó el promedio de personas ponderado por el número de hogares dentro de la vivienda (PPPH), desagregado por municipio y clase, considerando solo edificaciones de uso residencial y mixto que se encuentre ocupadas.
- c) Se selección por criterio bivariado-bivariado⁴ de las unidades en las edificaciones incompletas y no visitadas (NENV).
- d) Para cabeceras municipales y centros poblados de los municipios de barrido se estimó el producto de PER, PVO, PPPH y NENV desagregado por municipio y clase.
- e) Para rural disperso se estimó el producto de PER, PPPH y NENV desagregado por municipio y clase.
- f) Estimación de la población omitida PO por cada una de las unidades de cobertura seleccionadas.

El resultado del municipio considera la agregación de personas omitidas por municipio y por clase, en las zonas en la que se implementó el método de recolección de barrido.

3.2.2 Resultados

⁴ Ver sección 3.1

Se obtuvo que en todo el territorio nacional en las zonas no visitadas e incompletas para el método de recolección de barrido se omitieron **1.010.321** personas. En la figura 5 se muestra la distribución de las personas imputadas por departamento y clase, resaltando a los departamentos del Valle del Cauca, Bogotá, Atlántico y Bolívar como los mayores receptores de población en Rural disperso, el departamento de Nariño como el mayor receptor de población en centros poblados y los departamentos de Antioquia, Cauca, Córdoba, Cundinamarca y Tolima como los mayores receptores de población en cabeceras.

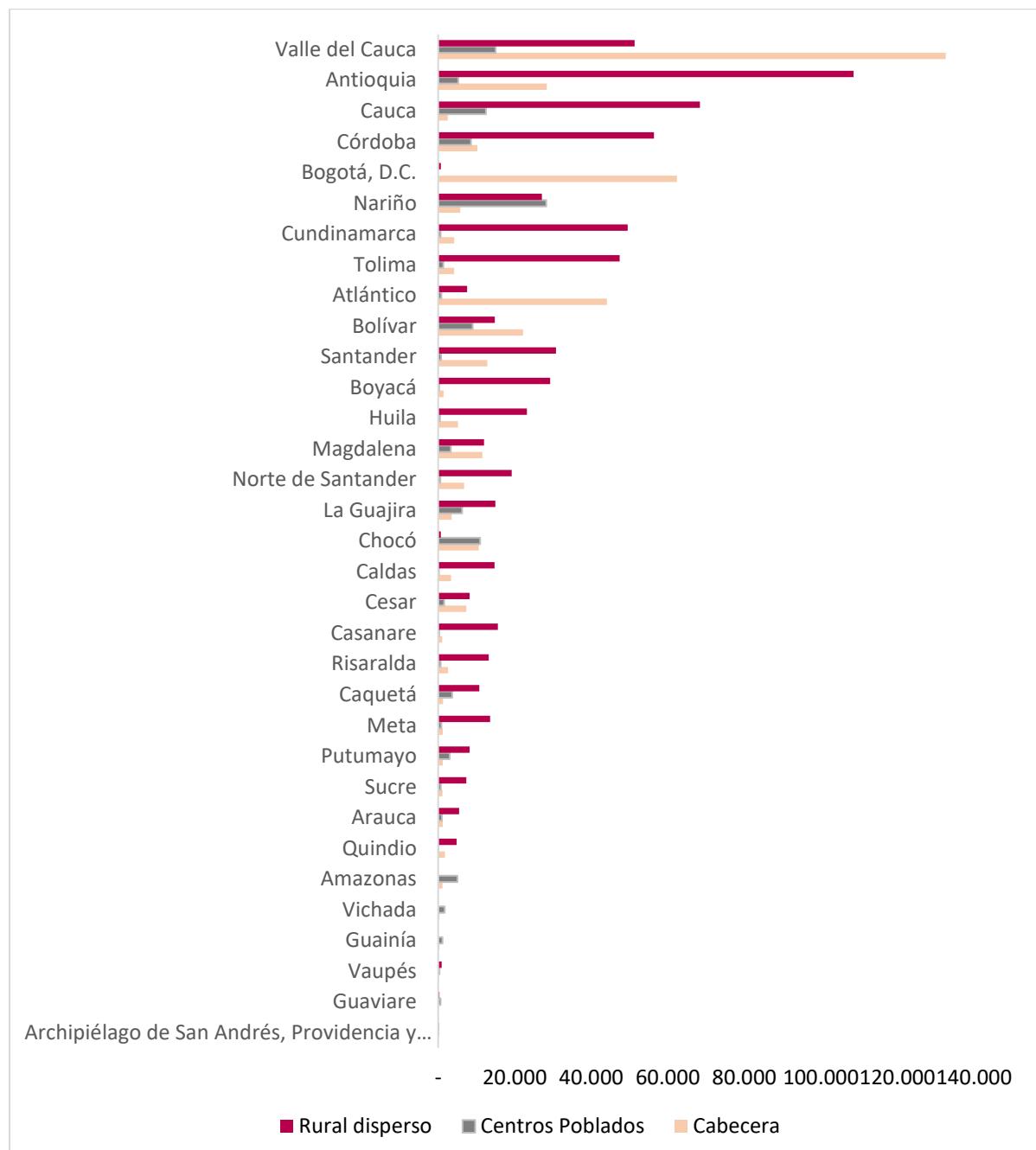


Figura 5. Distribución de la estimación de personas omitidas en el método de barrido en Z por departamento y clase
Fuente: DANE - Dirección de censos y demografía CNPV 2018

3.3 Omisión en zonas no visitadas e incompletas de rutas

3.3.1 Contexto de áreas de rutas

En Colombia existen áreas, principalmente en las regiones de la Amazonía, Orinoquía y Pacífica, que se caracterizan por su difícil accesibilidad, baja densidad poblacional, gran extensión territorial, y, además, algunas de estas presentan problemas de seguridad. La suma de estas condiciones se tradujo en mayores retos tanto en la planeación como en el operativo del CNPV 2018. Debido a esto, durante la planeación del CNPV 2018 se realizaron talleres de cartografía social, con el fin de mejorar la información disponible por el DANE de estas áreas.

A partir de los talleres de cartografía social, e información adicional de otras fuentes (3º Censo Nacional Agropecuario, Instrumentos de Ordenamiento Territorial, Planes de Desarrollo municipal y departamental, entre otras), se planificó la estrategia de recolección por rutas. Este método de recolección consistió en grupos de trabajo que recorrían la zona rural de algunos municipios y de las áreas no municipalizadas (corregimientos departamentales). Por lo general, los recorridos de las rutas fueron a lo largo de un río y sus afluentes, caminos de herradura o trochas, abarcando un área de influencia que contenía cada una de las comunidades y asentamientos existentes.

En los talleres de cartografía social se recogió información sobre la ubicación y características básicas de las comunidades o asentamientos de población. Entre la información que se recolectó, los representantes de las comunidades reportaron el número aproximado de viviendas como el de personas.

El Modelo Lineal Bayesiano Jerárquico se escogió debido a que históricamente estas áreas se han destacado por su déficit de información. Estas áreas no cuentan con información demográfica confiable, sus registros administrativos están incompletos, y no hay claridad de límites administrativos entre municipios para sus habitantes. No obstante, la tecnología a través de imágenes de satélite y sensores remotos nos brinda información, directa o indirectamente, de la presencia de personas en el territorio. Así, al obtener información sobre construcciones, vegetación, intensidad de luz, etc., se puede deducir la presencia o ausencia de personas en áreas remotas.

3.3.2 Planteamiento metodológico

Los modelos jerárquicos involucran varios parámetros de tal forma que las creencias que se tienen de unos parámetros dependen de manera significativa de los valores conocidos de otros parámetros. Por tanto, para estimar un modelo jerárquico se deben cumplir dos condiciones: 1) conocimiento inicial de la dependencia entre parámetros y 2) noción de los posibles valores iniciales de estos. La siguiente ecuación ilustra la dependencia entre parámetros para estimar la población en las áreas que se recorrieron usando rutas. Esta población se estimó con la siguiente distribución Poisson:

$$P_{dmi} \sim \text{Poisson}(\lambda_{dmi})$$

$$\lambda_{dmi} = (B_{dmi} + 0.1)O_{dmi}D_{dmi}$$

Donde P_{dmi} es el número de personas en la unidad de cobertura i que se encuentra en el municipio m y el departamento d . λ_{dmi} es el valor esperado de personas en esta unidad, el cual es la multiplicación de: B_{dmi} , que son las unidades esperadas; O_{dmi} , parámetro que especifica el porcentaje de viviendas ocupadas, y D_{dmi} , cuyo valor es igual al número de personas por vivienda ocupada. Por su parte, B_{dmi} , O_{dmi} y D_{dmi} se modelan de la siguiente forma:

B_{dmi} se modela en función a la distancia a centros poblados, la intensidad de luces nocturnas y E_{dmi} , que es una variable que estima el número de edificaciones sin importar su uso. Se decidió estimar B_{dmi} debido a que se evidenció que los representantes de los indígenas sobreestimaban el número de viviendas. O_{dmi} , por su parte, se modeló teniendo en cuenta las luces nocturnas y la elevación. Por último, D_{dmi} se estimó a partir de la densidad de instituciones educativas y el Índice de Necesidades Básicas Insatisfechas.

El diagrama general del proceso se presenta a continuación:

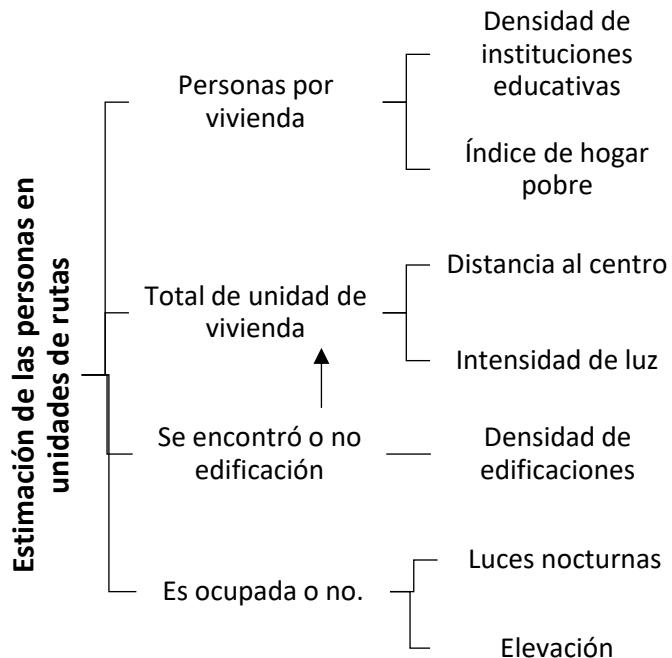


Figura 6. Covariados espaciales que permitieron la estimación de población en zonas de rutas
Fuente: DANE - Dirección de Censos y Demografía

3.3.3 Construcción de covariados espaciales

Las variables dependientes generadas para el modelo jerárquico bayesiano se procesaron de la siguiente:

- **Distancia al centro:** *Este covariado se procesó a partir de la capa de vías nacionales escala 1:100.000, y los polígonos de cabecera, a través de la función de distancia euclíadiana.*
- **Densidad construcciones:** *se calcula a partir de la capa CONSTRUCCION_P de la cartografía base del IGAC, con la herramienta Densidad Kernel, que nos da una superficie continua de densidad, con un tamaño de celda de 100 m. Posteriormente, para cada unidad de cobertura rural (predio) se calculó el promedio de esta densidad kernel, por medio de la herramienta Estadísticas Zonales, que permite obtener medidas de tendencia central a partir de un raster como el promedio, la mediana, el mínimo o el máximo.*
- **Densidad colegios:** *Este covariado se procesó a partir del directorio de instituciones educativas, posterior georreferenciación en capa tipo punto, con la cual se generó una interpolación IDW a partir del número de estudiantes en cada institución.*
- **Elevación:** *Este covariado de Modelo Digital de Terreno de DEM SRTM Misión Topográfica Shuttle Radar NASA*
- **Índice de hogar pobre:** *Este covariado se obtuvo con el cálculo de las Necesidades Básicas Insatisfechas NBI por hogar, procesados a partir de la información de CNPV 2018, y es igual a la*

proporción de hogares con NBI.

- *Imagen ráster de luces nocturnas: Este covariado se obtuvo de la Imagen NOAA/VIIRS Promedio de radiancia del 01/01- 450 metros. al 31/12/2018- Google Engiene.*

A partir de la herramienta de estadísticas zonales se vincularon los valores promedios de cada celda de la imagen ráster a las unidades de cobertura. Una vez se obtuvo esta matriz se procedió a normalizar las variables y a generar la estimación del modelo, procesado en el software estadístico R.

1.1.4 Resultados

En total se estimó una población no censada de **224.966** habitantes en las zonas de rutas que no fueron visitadas o quedaron incompletas previamente seleccionadas con los métodos de correlación espacial. El Departamento de Nariño presentó una mayor estimación de población omitida por este método, en correspondencia a las dificultades operativas de orden social presentadas principalmente en los municipios del pacífico de este departamento colombiano, seguido del departamento de Chocó en donde se presentaron condiciones de dificultades de acceso geográfico.

Tabla 2. Total, población estimada por departamento en las zonas de rutas de mayor complejidad operativa

CODIGO	DEPARTAMENTO	OMISIÓN Z EN RUTAS
5	Antioquia	15
18	Caquetá	6.010
19	Cauca	1.467
20	Cesar	3.090
27	Chocó	37.547
44	La Guajira	9.865
47	Magdalena	1.105
50	Meta	812
52	Nariño	107.051
76	Valle del Cauca	4.219
85	Casanare	834
86	Putumayo	19.944
91	Amazonas	2.072
94	Guainía	1.111
95	Guaviare	3.025
97	Vaupés	209
99	Vichada	26.591
0	Total, Nacional	224.966

Fuente: DANE- Estimación por modelo JAGS

4. Omisión en viviendas ocupadas con personas ausentes (A)

La no realización de la entrevista en las viviendas ocupadas con personas ausentes VOCPA no significa la ausencia de personas dentro de las mismas, por lo tanto, es natural asignar un número de habitantes mediante un proceso de estimación de población que las habitan. La propuesta metodológica de este proceso se encuentra descrita a continuación:

4.1 Planeamiento metodológico

Al igual que para las áreas no visitadas e incompletas del barrido censal, en VOCPA se utilizó el método de promedio local con granularidad geográfica. Sin embargo, tomar un criterio geográfico desagregado por cabecera, centros poblados y rural disperso, podría no reflejar las dinámicas de los hogares y sus integrantes, ya que factores como estratos socioeconómicos, acceso a servicios sanitarios, entre otros, están relacionados con la distribución de personas dentro del territorio, por ello, se incluyen dentro de los criterios de granularidad geográficos el área operativa.⁵ (unidad a la que se logra desagregar los promedios, haciendo uso intensivo de la información censal georreferenciada).

Esta unidad fue seleccionada dado que en las áreas rurales en donde la presencia de viviendas es más dispersa, el área operativa consigue agrupar más viviendas, garantizando una estimación más robusta, consecuentemente, los criterios de desagregación geográficos incluirán el área operativa en centros poblados y rural disperso, y quedan definidos como:

- *Desagregación geográfica: Municipio, clase, unidad de cobertura en urbano y área operativa en centros poblados y rural disperso*
- *Desagregación por tipo: Tipo de vivienda*

En consecuencia, se plantea un modelo que identifica en primer lugar las VOCPA por unidad de cobertura (UC) y se le asigna el promedio de hogar del área operativa al cual pertenece de UC, representado por la siguiente expresión:

$$PO_{AX} = \Sigma VOCPA_x * PPPH_{BX}$$

Donde,

PO_{AX} Población omitida en VOCPA en la unidad de cobertura por X

$\Sigma VOCPA_x$ Sumatoria de las viviendas ocupadas con personas ausentes en la unidad de cobertura X

$PPPH_{BX}$ promedio de personas ponderado por el número de hogares dentro de las viviendas del área operativa B al que pertenece la unidad de cobertura x

4.2 Resultados

Se toman la 13'480.729 viviendas ocupadas con personas presentes, se desagrega por municipio, clase, unidad de cobertura o área operativa (según corresponda) y tipo de vivienda. Luego se obtienen el promedio de personas presentes en la vivienda ponderado por el número de hogares dentro de ella. Este promedio se imputa en las 439.017 viviendas VOCPA, teniendo en consideración la misma desagregación con la que se computó la estimación.

Como resultado se obtiene un total de **1'520.541** personas en viviendas VOCPA, la tabla 3 muestra la distribución de esta población por clase, resaltando el territorio urbano como el mayor receptor de población. La distribución departamental de la población obtenida se muestra en la gráfica 28, resaltando que los departamentos con más VOCPA (Antioquia, Valle del Cauca, Bogotá y Atlántico), reciben las mayores cantidades de personas imputadas.

5 definida como el área de supervisión que está conformada por un número promedio de edificaciones y/o viviendas. Para la definición de este parámetro se considera la edificación y/o vivienda como la variable del diseño operativo. Este concepto es el pilar de la organización y diseño operativo y se constituye en la unidad de planeación que da lugar a la estructura de la operación estadística, al cálculo de los productos cartográficos y demás requerimientos del operativo"

Tabla 3. Estimación VOCPA desagregación por clase

CLASE	Estimación
Urbano	1.172.377
Centros Poblados	68.244
Rural disperso	279.920
Total	1.520.541

Fuente: DANE - Dirección de censos y demografía CNPV 2018

En cuanto a la distribución territorial de la omisión A, la concentración de población ajustada en este componente muestra una predominancia en la principales ciudades y capitales de departamento, en consonancia con la hipótesis de que estos territorios en donde existe mayor ausencia de los moradores de las viviendas a la hora de realizar la operación censal, dada las dinámicas económicas y sociales de las ciudades.

5. Omisión por subenumeración de personas en los hogares (S)

5.1 Planteamiento metodológico

La subenumeración se puede presentar en términos de volumen, es decir, cuántas personas se omitieron dentro de los hogares o en términos de porcentaje respecto a la población efectivamente censada. Para esta estimación, se realizaron ambos análisis con el fin de identificar las zonas geográficas con mayor y menor ajuste absoluto y relativo.

De esta forma, con base en la información obtenida de los hogares particulares que fueron efectivamente censados en el operativo del Censo Nacional de Población y Vivienda 2018, y la información de la Encuesta Nacional de Calidad de Vida (ECV) 2018 realizada por el DANE se estimó el tamaño promedio del hogar en los 1122 municipios. En este caso, la ECV se toma como la encuesta postcensales, la cual se realiza de manera independiente al censo y es representativa a nivel departamental.

Así pues, se aplicó el algoritmo de imputación múltiple donde se toman los tamaños de hogares observados del censo como datos faltantes y la ECV proporciona información completa. Teniendo en cuenta que el método de imputación múltiple proporciona m conjuntos de datos completos, es decir, se obtienen m tamaños de hogar estimados para cada uno de los hogares del censo, por lo cual, al realizar el promedio de estas m estimaciones se llega a un tamaño de hogar estimado.

El análisis anterior se realizó individualmente para las 9 regiones de la ECV, centros urbanos y zonas rurales, en el cual, para la selección de las covariables usadas en la imputación múltiple, se realiza una selección de covariables con base en la ECV exclusivamente, en la que se compara el ajuste de los modelos logísticos ordinales utilizando todas las covariables, una selección de acuerdo al criterio del experto y una selección a través del método "stepwise". Así, el modelo ordinal usado en el algoritmo de imputación es el que presenta un mejor ajuste con respecto a la selección de covariables.

Una vez se obtienen las estimaciones del tamaño del hogar para la base de datos del Censo Nacional de Población y Vivienda, se realizan los cálculos descritos en las ecuaciones descritas anteriormente, lo cual conduce a la estimación del número de personas no enumeradas al interior de los hogares.

En síntesis, el procedimiento descrito anteriormente se resume en tres grandes pasos, esto es:

1. **Homologación de las bases de datos:** En este paso se garantiza que las variables que se usarán posteriormente para el modelamiento e imputación múltiple tengan los mismos niveles, es decir, que la codificación sea igual para ambas.
2. **Modelamiento:** Con base en la información auxiliar (ECV) se seleccionan las covariables que se usan para realizar la imputación múltiple. Un punto muy importante a tener en cuenta es la naturaleza de la variable de respuesta a modelar, ya que se espera que entre mejor sea el ajuste del modelo, mejores serán las imputaciones obtenidas posteriormente.
3. **Imputación Múltiple:** Una vez seleccionadas las covariables que mejor explican la variable de respuesta, se realiza el método de imputación en el cual se obtienen m conjuntos de datos imputados que, al promediar, brindan una estimación para la variable de interés.

Una vez se obtiene el conjunto de datos imputado al final del paso (3), se produce la agrupación con respecto a la variable que se desee, en este caso, ésta se realiza a nivel municipal y se realizan los cálculos pertinentes para obtener la estimación de la subenumeración a dicho nivel. Sin embargo, al agregar a distintos niveles geográficos, se obtiene una estimación total del número de personas omitidas distinta, por lo cual, para obtener la estimación de la subenumeración a nivel departamental, primero se realiza el cálculo a nivel municipal y éste se agrega a un nivel superior (departamento).

5.1.1 Modelo de imputación

En esta sección se describe el procedimiento realizado para obtener las estimaciones de la subenumeración para cada uno de los 1122 municipios de Colombia. En primer lugar, es importante tener en cuenta que la subenumeración se calcula como

$$S_i = P_i - \hat{P}_i; \quad \text{para } i = 1, \dots, 1122.$$

donde P_i y \hat{P}_i representan la población efectivamente censada y estimada para el i -ésimo municipio. A su vez, \hat{P}_i viene dado por

$$\hat{P}_i = \delta_i \cdot H_i; \quad \text{para } i = 1, \dots, 1122.$$

donde H_i es el número de hogares efectivamente censados para el i -ésimo municipio y δ_i es la diferencia entre el tamaño de hogar promedio estimado y obtenido en el censo para el i -ésimo municipio, esto es

$$\delta_i = \bar{PPH}_i - PPH_i; \quad \text{para } i = 1, \dots, 1122.$$

con PPH_i tamaño promedio del hogar obtenido en los hogares efectivamente censados para el i -ésimo municipio y \bar{PPH}_i el tamaño promedio del hogar estimado, es decir

$$PPH_i = \frac{\sum_{j=1}^{H_i} I_j}{H_i}; \quad \text{para } i = 1, \dots, 1122.$$

$$\bar{PPH}_i = \frac{\sum_{k=1}^{\tilde{H}_i} \tilde{I}_k}{\tilde{H}_i}; \quad \text{para } i = 1, \dots, 1122.$$

donde I_j es el número de individuos en el j -ésimo hogar de la base de datos del censo, \tilde{I}_k y \tilde{H}_i es el número de individuos y hogares en el k -ésimo hogar de la base de datos del censo eliminando los hogares que no reportaron toda la información de las covariables utilizadas en los modelos ajustados y el i -ésimo municipio.

En los casos que $\delta_i \leq 0$, se toma $\delta_i = 0$ debido a que valores negativos de δ_i indican que el valor estimado del hogar está por debajo del observado, por lo cual, se supone que no se omitieron personas dentro de los hogares al momento de la enumeración y en consecuencia, no existe subenumeración

para dicho municipio. Cabe resaltar que todas las variables son categóricas a diferencia del número de dormitorios, tamaño del hogar y edad que se miden en una escala discreta continua.

Tabla 4. Variables de interés del CNPV 2018 con su respectiva codificación dado el caso.

Geográficas	Vivienda	Hogar	Jefe del hogar
-Regiones	-Tipo de vivienda (TVIV)	-Número de Dormitorios (DORM)	-Sexo
-Departamentos	-Tipo de sanitario (TSAN)	-Lugar donde se preparan alimentos	-Nivel Educativo(EDU)
-Municipios	-Material paredes (MATPA)	- Tamaño del hogar	-Trabajo(TRAB)
-Clases	-Material pisos (MATPI)		-Estado civil
	-Servicio de energía (SEE)		-Etnia
	-Servicio de acueducto (SACU)		-Edad
	-Servicio de alcantarillado (SALC)		
	-Servicio de gas natural (SGAS)		
	-Servicio de recolección de basuras (SBAS)		
	-Servicio de internet (SINT)		

FUENTE: DANE- Dirección de Censos y Demografía

5.2 Resultados

Se obtuvo que en el total del territorio nacional se subenumeraron **1.127.141** personas al interior de los hogares censados. A lo largo del país, los municipios de Bolívar, sur del Cesar, occidente del Norte de Santander, Putumayo, Nariño, gran parte de la Cordillera Occidental y San Andrés y Providencia obtuvieron porcentajes altos de subenumeración.

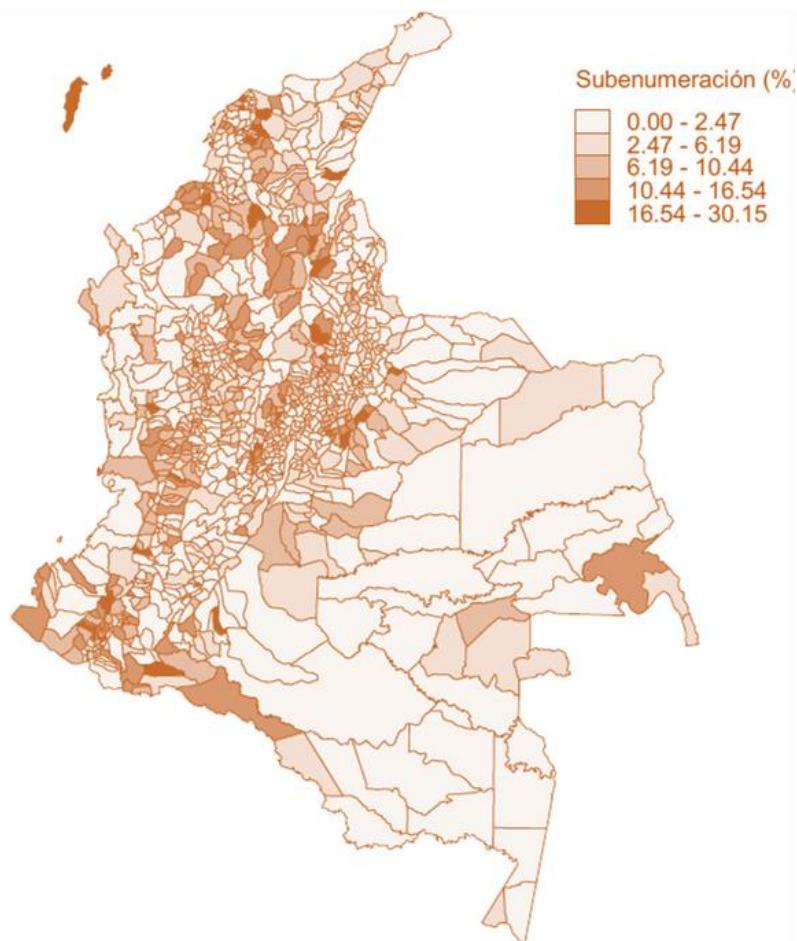


Figura 7. Distribución territorial de Subenumeración municipal, presentados como porcentaje

6. Omisión Total

6.1 Agregación de resultados.

La agregación de la omisión censal por los métodos expuestos parte de la expresión:

$$Y_N = Z_N + A_N + S_{mN}$$

Según las estimaciones realizadas

$$Z_N = 1.235.287$$

$$AN = 1.520.541$$

$$S_N = 1.127.141$$

$$Y_N = 3.882.969 \text{ Personas omitidas total Nacional}$$

$$\% \text{ Omisión total} = \frac{E_m - R_m}{E_m} * 100$$

$$\text{Omisión Nacional} = 8,1\%$$

Como se señaló en el capítulo I, la omisión aceptada e informada para el total nacional corresponde a al techo estimado con el método directo Dual System, que arroja una omisión del **8,5%** que representan 4.094.087 personas no censadas. En consonancia a continuación se presentan los procesos implementados para llevar los cálculos de omisión desagregados al techo nacional informado.

6.2 Ajuste al techo de omisión Censal Nacional informada

Una vez desagregada la omisión censal a nivel municipal cabecera - resto, se determinó que el prorrato no era el procedimiento más adecuado para llevar al techo la estimación municipal, puesto que asume una distribución en forma proporcional, pero no necesariamente responde a las complejidades operativas que incidieron en la posibilidad de dejar de censar personas. El procedimiento realizado para su corrección se basó en la selección de unidades no visitadas e incompletas para el ajuste Z en estos municipios. A partir del análisis de selección espacial de unidades a incompletas⁶. Se utilizaron los dos índices de Moran con los siguientes criterios:

1. De los municipios con cierre operativo antes al 29 de septiembre de 2018, usar el criterio bivariado-bivariado para estimar el número edificaciones no visitadas (**NENV**).
2. De los municipios con cierre operativo posterior al 29 de septiembre de 2018, usar el criterio bivariado-Univariado para estimar el número edificaciones no visitadas (**NENV**).

Este ajuste arroja un nuevo valor de **Z**, que incrementó en 174.111 personas omitidas en estos municipios pasando de 1.235.287 a 1.409.398.

Por el lado de la subenumeración se realizó un ajuste específico a la ciudad de Bogotá, que como se mencionó presentó una estimación distante del tamaño de hogar esperado, por tanto, se ajustó a nivel desagregación manzana los 36.957 personas subenumeradas. En total la nueva estimación de **S**, ascendió a 1.164.148, obteniendo nuevamente la población omitida por los tres componentes de la siguiente manera.

$$Z_N = 1.409.398$$

$$AN = 1.520.541$$

$$S_N = 1.164.148$$

⁶ Ver capítulo 2, Numeral 2.1

$$Y_N = 4.094.087 \text{ Personas omitidas total Nacional que determinan una omisión censal de 8.5\%}$$

6.2.1 Ajuste de la cobertura departamental

El agregado a nivel departamental muestra un patrón de mayor omisión en los departamentos del sur del país, en donde se encuentran las omisiones de Putumayo con un 28,3%, Nariño con un 24,7%, Cauca con un 18,2%, Valle del Cauca 17,8%, solo superados por el departamento Amazónico de Vichada, en donde la omisión superó el 36,5%.

En contraste, las omisiones mas bajas, estuvieron centralizadas en Bogotá con un 3.1% y Cundinamarca con un 4.3%, seguido del departamento de Sucre que alcanzó una omisión del 4.5%.

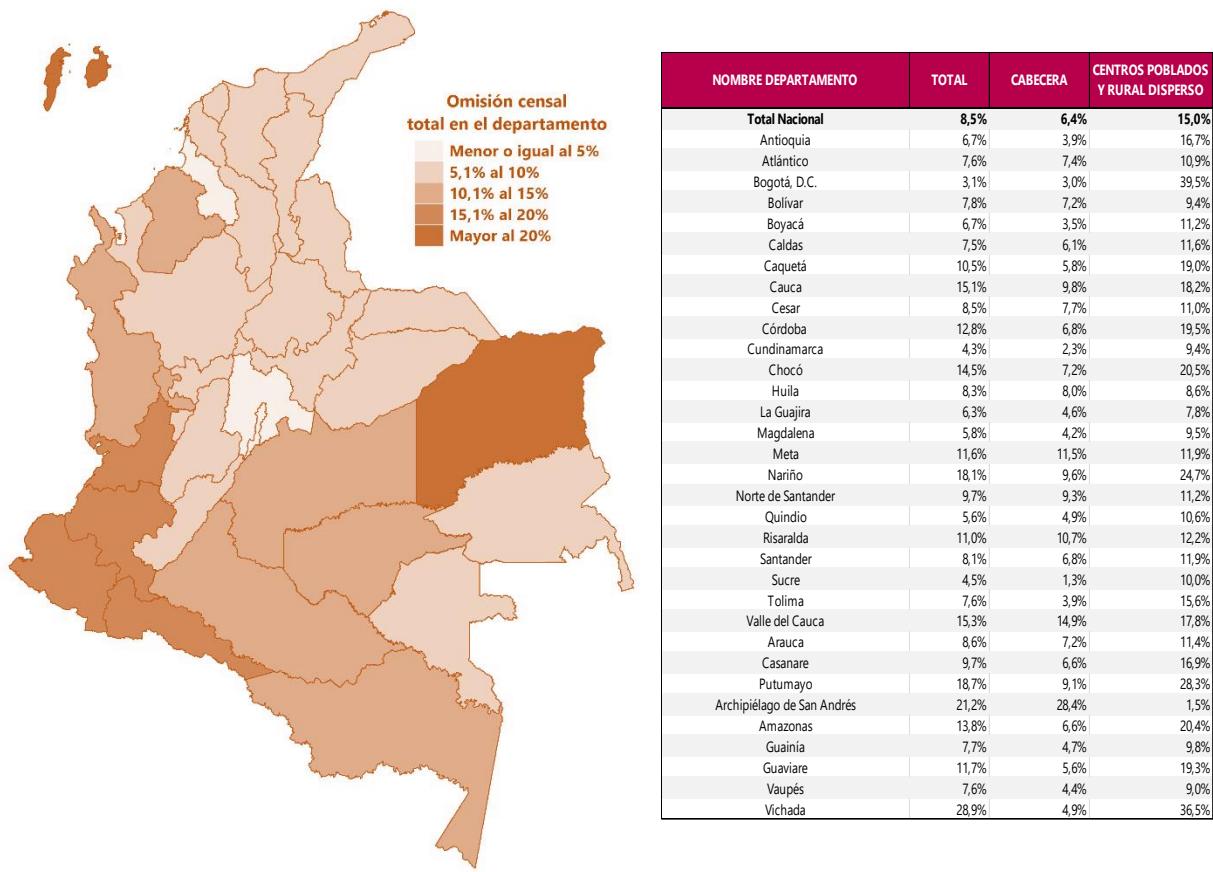


Figura 8. Ajuste de cobertura a nivel Departamental.

Fuente: DANE- Dirección de Censos y Demografía

7. CONCLUSIONES

Como principal conclusión para la estimación a nivel subnacional se evidencia el potencial de la información geográfica como insumo para desagregar la población omitida. Los procesos desarrollados permitieron ratificar la hipótesis de partida, principalmente en los factores de omisión correspondientes a viviendas con personas ausentes y zonas no visitadas e incompletas.

Para las viviendas ocupadas con personas ausentes, el estimador de promedio local con granularidad geográfica mantienen las características de tamaño de personas por vivienda de acuerdo con su entorno

geográfico, el tamaño de vivienda se encuentra influenciado por los niveles socioeconómicos, servicios sanitarios, características de las viviendas.

El análisis realizado demuestra que esto solo es posible cuando se cuenta con la información georreferenciada tanto de las viviendas ocupadas, como de aquellas ocupadas con todas las personas ausentes. Así pues, en zonas en donde el Censo se realizó en cuestionarios en papel se debe generar una georreferenciación de estas viviendas con tamaños promedios más agregados en donde se pierden las características locales de los residentes de las viviendas.

Por otro lado, en cuanto a la estimación de población en las zonas no visitadas e incompletas para rutas, indica que existen fuentes alternas espaciales que permiten inferir la presencia de asentamientos humanos principalmente todos aquellas relacionadas con las transformaciones antrópicas.

El uso de la información colateral como las imágenes de satélite y la cartografía que se encuentran disponible puede ser más explorada a partir de los sistemas de datos abiertos abriendo un camino de posibles investigaciones más avanzadas en este campo, que permitan a la demografía integrar estas herramientas como parte de las nuevas técnicas que se pueden abordar.

Para el factor de subenumeración, debido a que la metodología implementada se basa en modelos lineales generalizados, se permite el manejo de diferentes tipos de variable respuesta (continua, ordinal, nominal, conteos, etc.), lo cual genera que se puedan abordar una variedad de problemas con este enfoque.

Finalmente, este documento pretende ser un instrumento detallado de los resultados obtenidas y los limitantes en las metodologías empleadas, de tal forma que se avance en la premisa de la transparencia y objetividad de los datos presentados.

BIBLIOGRAFIA

- Agresti, A. (2007). *An Introduction to Categorical Data Analysis*. Gainesville, Florida: JohnWiley & Sons, Inc., Hoboken, New Jersey.
- Aguilar, A., & Santana, G. (2019). metodo de autocorrelación espacial. In G. D. Buzai, & M. Santana, *Métodos cuantitativos en Geografía Humana* (pp. 109-129). Universidad Nacional de Lujan: Buenos Aires : Impresiones Buenos Aires Editorial, 2019.
- Baffour, B. (2006). *Modelling Census Under-Enumeration: A Logistic Regression Perspective, General*.
- Banco Mundial. (2012). *Perspectivas de la Población Mundial*. Disponibles en:
<https://datos.bancomundial.org/indicador/SM.POP.NETM?locations=CO>.
- Bay, G., & Orellana, H. (2007). *Taller de expertos en el uso de estadísticas vitales:alcances y limitaciones*. Santiago, Chile: CEPAL-UNFPA.
- BID-CELADE/CEPAL. (2006). *La Omisión Censal en América Latina* . Santiago de Chile: Publicado:
https://repositorio.cepal.org/bitstream/handle/11362/7203/1/S0600009_es.pdf.
- Brittain, S. B. (2009). Estimators in capture--recapture studies with two sources. *AStA Advances in Statistical Analysis*, 23--47.
- Brown, J. D. (1999). *A methodological strategy for a*.
- CELADE. (1997). *XX Curso Regional Intensivo de Análisis Demográfico*. Santiago de Chile: Disponible en:
https://repositorio.cepal.org/bitstream/handle/11362/34406/S9700584_es.pdf?sequence=2&isAllowed=y.
- CELADE. (2013). *III REUNION REGIONAL SOBRE EVALUACIÓN Y ESTIMACIONES DEMOGRÁFICAS CON BASE EN EVALUACIÓN Y ESTIMACIONES DEMOGRÁFICAS CON BASE EN*. Santiago de Chile: Material de curso.
- CELADE. (2019). *ESTIMACIONES Y PROYECCIONES DE POBLACIÓN A LARGO PLAZO. 1950-2100*. Santiago de Chile: Disponible en: <https://www.cepal.org/es/temas/proyecciones->

demograficas/estimaciones-proyecciones-poblacion-total-urbana-rural-economicamente-activa.

- CELADE-UNFPA. (2014). *Los datos demográficos: Alcances, limitaciones y métodos de evaluación*. Santiago de Chile: Publicación de las Naciones Unidas.
- CEPAL. (2014). *Los datos demográficos. Alcances, limitaciones y métodos de evaluación*. OCUBBRE 2014: CEPAL.
- CEPAL- UNFPA. (2014). *Los datos demográficos: Alcance, limitaciones y métodos de evaluación*. Santiago de Chile: Publicación de Naciones Unidas ISSN 1680-886x.
- CEPAL- UNFPA. (2014). *Los dtos demográficos: Alcance, limitaciones y métodos de evaluación*. Santiago de Chile: Publicación de Naciones Unidas ISSN 1680-886x.
- CEPAL-UNFPA. (2010). *Los censos de 2010 y la salud*. Santiago de Chile: Serie Seminarios y Conferencias N° 59.
- Chackiel, J. (2010). Evaluación post-empadronamiento de la cobertura de los censos de población. *Notas de Población. Año XXXVII N 91*, 43-71.
- ChandraSekar, C. a. (1949). On a method of estimating birth and death rates and the extent of registration. *J. Amer. Statist. Assoc.*, 101-115.
- Chandrasekaran, a. D. (1949). On a method of estimating birth and death rates and the extent of registration. *J. Amer. Statist. Assoc.*, 101-115.
- Chatterjee, K. &. (2015). Approximate Bayesian Solution for Estimating Population Size from Dualrecord System. *10.13140/RG.2.1.3726.7921..*
- Cressie, N. (1988). *Estimating census undercount at national and subnational levels, in 'Proceedings:...*
- DANE. (1989). *Proyecciones Nacionales de Población*. Bogotá: DANE-CELADE-DNP-CIID.
- DANE. (1998). *Colombia. Proyecciones Quinquenales de población por Sexo y Edad 1950-2050*. Bogotá.
- DANE. (2008). *Censo general 2005*. Bogotá.
- DANE. (2009). *Proyecciones nacionales y departamentales*. Bogotá.
- DANE-MinSalud-Registraduria-Bloomberg Philanthropies. (2019). *Método empírico de la completitud de Estadísticas Vitales (MECREV)*. bogotá.
- DANE-UNFPA. (2011). *Evaluación de Cabalidad de la EEVV*.
- Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1-26.
- Ellis, Y. (1995). *Examination of census omission and erroneous enumeration based on 1990 ethnographic*.
- Ghosh, M. R. (1994). *Small area estimation: an appraisal*, *Statistical science* 9(1), 55-76.
- Greville, Thomas. (1946). *United States Life Tables and Actuarial Tables 1939-1941*. United States, Departament of Commerce, Bureau of the Census: presentados en: *Curso Regional Intensivo de Demografía*, 1997CELADE.
- Joaquin Recaño. (2015 y 2019). *Estimaciones de los SNM 2005-2020*. Bogotá: Consultoria realizada para DANE.
- Juan Chackiel. (2009). *Evaluación post-empadronamiento de la cobertura en los censos de población 1*. Santiago de Chile: Notas de población N° 91 • CEPAL.
- Kekäläinen, J. N. (2008). Pike predation on hatchery-reared Atlantic salmon smolts in a northern Baltic river. *Ecology of Freshwater Fish*, 17:100-109.
- Kimpel, T. &. (2007). *Estimating household size for use in population estimates*.
- Manrique, H. (1990). *Estudios de Evaluación de la calidad y Cobertura del XV Censo Nacional de Población y IV de Vivienda*. Bogotá: DANE.
- Martin, D. (2010). Understanding the social geography of census undercount', Environment and Planning.
- Maura E. Stokes, C. S. (2003). *Categorical Data Analysis Using The SAS System*. Cary, NC, USA: SAS Institute and Wiley.
- MinSalud-Profamilia. (2015). *Resumen Ejecutivo Encuesta Nacional de Demografía Y Salud- ENDS*. Bogotá: Disponible en : <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/ENDS-libro-resumen-ejecutivo-2016.pdf>.

- Naciones Unidas. (2010). *Principios y recomendaciones para los censos de población y habitación. Revisión 2*. Nueva York: Departamento de Asuntos Económicos y Sociales, División de Estadística,.
- National Research Council. (1972). *America's Uncounted People: Report, National Academies*.
- Nour, E. S. (1982). On the Estimation of the Total Number of Vital Events with Data from Dual-record Collection Systems. *J. R. Statist. Soc. A*, 106-116.
- Olsson, A. E. (2006). Activity patterns and abundance of microchiropteran bats at a cave roost in south-west Madagascar. *African Journal of Ecology*, 44:401-403.
- Ribotta, B. (2005). *Evaluación de fuentes y estimaciones demográficas*. Córdoba- Argentina: Universidad Nacional de Córdoba- Investigación presentada para obtener el título de Magíster en Demografía.
- Rincón, M. (2008). *Conciliación Censal y determinación de la población base* .
- Rincón, M., & Fajardo, M. (2007). *Teoría y Métodos para la elaboración de estimaciones y proyecciones de población*. Bogotá.
- Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*, Vol. 81, John Wiley & Sons.
- Sadinle, M. (2008). On the Performance of Dual System Estimators of Population Size: A Simulation Study. *CERAC*, Paper, No. 13.
- Sadinle, M. (2009). Transformed Logit Confidence Intervals for Small Populations in Single Capture-Recapture Estimation. *Communications in Statistics - Simulation and Computation*, 1909-1924.
- Sardi, E. (2008). *Especificaciones del proceso de conciliación censal como Especificaciones del proceso de conciliación censal como Censo 2005* . Dipsonible en:
- Siabato, W. y.-M. (2019). *La autocorrelación espacial y el desarrollo de la geografía cuantitativa*. Bogotá: Cuadernos de geografía.
- Smith, S. K. (2002). *Regression approach to estimating the average number of*.
- Vehulst, A. (2013). Estimación de la fecundidad a partir de la Encuesta de Demografía y Salud. *Revista de Información Básica Estadística*, ISSN 2256-1552 pag, 26-41.
- Wolter, K. M. (1986). Some Coverage Error Models for Census Data. *Journal of the American Statistical Association*, 81(394):338-346.