

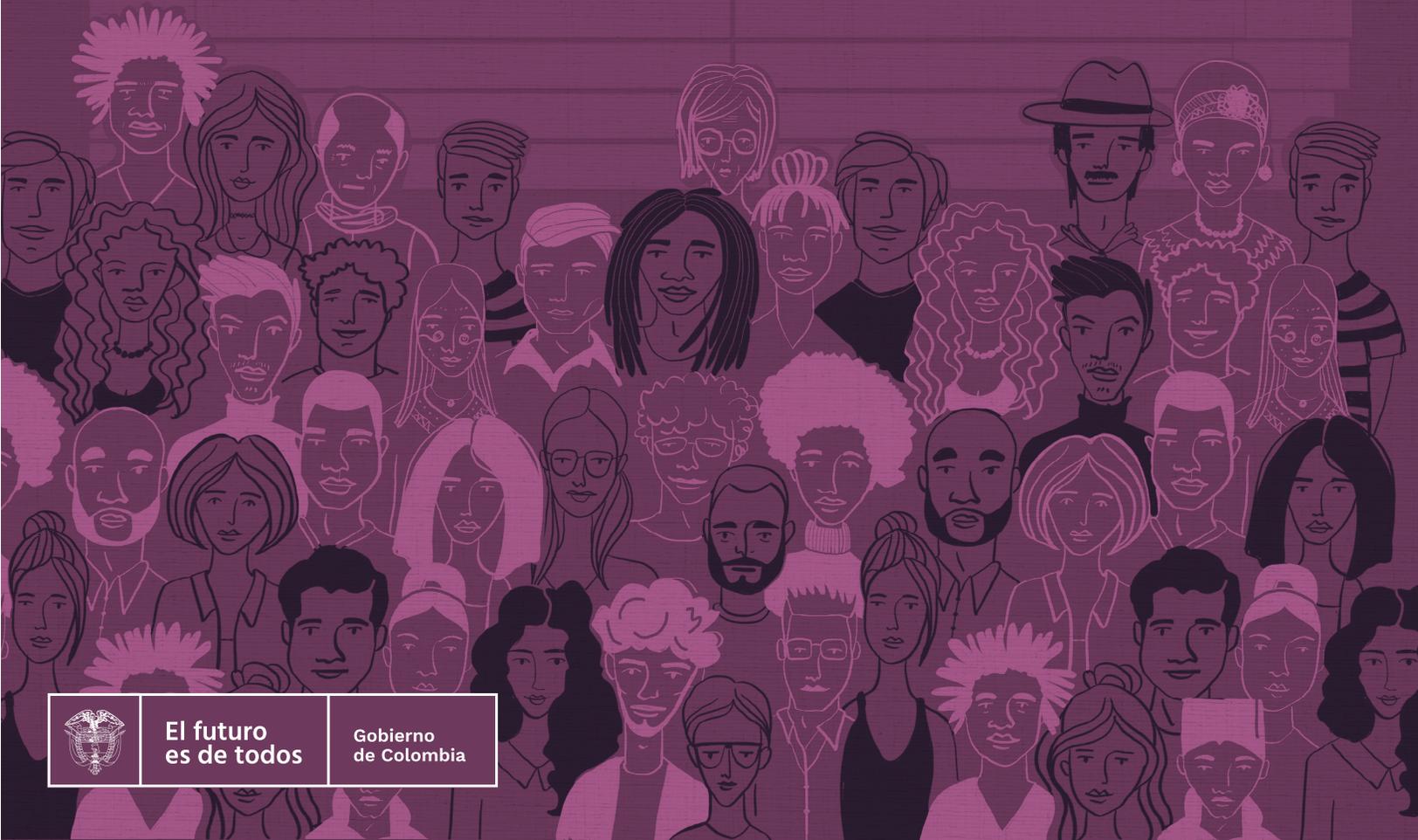
Metodologías

Demográficas Aplicadas

Número 2

Estimación de la omisión censal a nivel nacional y subnacional a partir de métodos demográficos, econométricos y geoestadísticos

Aplicación para el Censo Nacional de Población y de Vivienda - CNPV 2018



DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADÍSTICA DANE

JUAN DANIEL OVIEDO ARANGO
Director

RICARDO VALENCIA RAMÍREZ
Subdirector

MARÍA FERNANDA DE LA OSSA ARCHILA
Secretaria General

DIRECCIÓN DE DIFUSIÓN Y CULTURA ESTADÍSTICA

ANGELA VEGA LANDAETA
Dirección de Censos y Demografía

MAURICIO ORTIZ GONZÁLEZ
Dirección de Difusión, Mercadeo y
Cultura Estadística

SANDRA LILIANA MORENO MAYORGA
Dirección de Geoestadística

HORACIO CORAL DÍAZ
Dirección de Metodología y Producción
Estadística

JULIETH ALEJANDRA SOLANO VILLA
Dirección de Regulación, Planeación,
Estandarización y Normalización

JUAN PABLO CARDOSO TORRES
Dirección de Síntesis y Cuentas Nacionales

EQUIPO DE DIRECCIÓN DE CENSOS Y DEMOGRAFÍA

Yenny Andrea Marín Salazar

Lelio Alejandro Arias Vizcaino

Glenn Harry Amaya Cruz

Lina María Sánchez Céspedes

Andryu Enrique Mendoza Beltrán

Mariana Francisca Ospina Bohórquez

DIRECCIÓN DE GEOESTADÍSTICA

Gustavo Adolfo García Vélez

José Vicente Pérez Rojas

Raúl Emilio Ospina Villalobos

DIRECCIÓN DE DIFUSIÓN Y CULTURA ESTADÍSTICA EDICIÓN DE CONTENIDO Y DISEÑO

Brandon Steve Rojas Guerra

Gladys Adriana Quintero Hernández

Gabriela Wiesner Montaña

María Jimena Muñoz Saldarriaga

© **DANE, 2022**

Prohibida la reproducción total o parcial sin permiso
o autorización del Departamento Administrativo
Nacional de Estadística, Colombia.

Contenido

1.	Introducción	11
2.	Evaluación de cobertura a nivel nacional	14
2.1.	Estimación nacional por método directo	15
2.2.	Estimación nacional por método indirecto	32
3.	Evaluación cobertura a nivel municipal	49
3.1.	Causas y factores de omisión en el CNPV 2018	51
3.2.	Omisión de áreas geográficas no visitadas e incompletas (Z)	57
3.3.	Omisión en viviendas ocupadas con personas ausentes (A)	78
3.4.	Omisión por subenumeración de personas en los hogares (S)	85
3.5.	Omisión Total	94
4.	Conclusiones	102
5.	Bibliografía	104

Lista de figuras

Figura 1.	Flujo de procesos realizados en las estimaciones por método Dual System	20
Figura 2.	Esquema de mejoras en el proceso de emparejamiento	27
Figura 3.	Estructura poblacional censada vs estimada	29
Figura 4.	Población de colombianos en otros países por año de llegada en el periodo 1985-2005	38
Figura 5.	Corrección de la población de 0-1 años en un censo de población	39
Figura 6.	Evaluación de registros de nacimientos a partir de fuentes colateral	40
Figura 7.	Diagrama de Lexis: Hombres edad de 1-9 en el periodo 2009-2018	43
Figura 8.	Evaluación de población 2018. Escenario 1	45
Figura 9.	Evaluación de población 2018. Escenario 2	46
Figura 10.	Evaluación de población 2018. Escenario 3	47
Figura 11.	Evaluación de población 2018. Escenario 4	48
Figura 12.	Diagrama general de aplicación metodológica para estimación	51

Figura 13.	Seguimiento de cobertura censal- extracto urbano	_____	52
Figura 14.	Distribución departamental de las Viviendas Ocupadas con Personas Ausentes - VOCPA	_____	54
Figura 15.	Distribución I de las Viviendas Ocupadas con Personas Ausentes - VOCPA en los departamentos	_____	55
Figura 16.	Esquema resumen de los factores de omisión Censo Nacional de Población y Vivienda - CNPV 2018	_____	56
Figura 17.	Tipos de autocorrelación espacial	_____	59
Figura 18.	Covariados espaciales que permitieron la estimación de población en zonas de rutas	_____	70
Figura 19.	Covariados espaciales	_____	71
Figura 20.	Marco Censal operativo Censo Nacional de Población y Vivienda - CNPV 2018	_____	73
Figura 21.	Tamaño promedio de los hogares por departamento obtenido de la población efectivamente censada	_____	74
Figura 22.	Distribución de la estimación de personas omitidas en el método de barrido en Z por departamento y clase	_____	76
Figura 23.	Distribución de la estimación de la población omitida en zonas no visitadas e incompletas	_____	77

Figura 24.	Distribución departamental personas estimas en VOCPA _____	80
Figura 25.	Distribución porcentual departamental personas estimas en VOCPA y el impacto departamental _____	81
Figura 26.	Distribución territorial de personas estimadas en VOCPA y el impacto departamental _____	82
Figura 27.	Dispersión Omisión estimada con SISBEN vs personas imputadas en VOCPA, valores en miles _____	83
Figura 28.	Dispersión Omisión estimada con BDUA vs personas imputadas en VOCPA, valores en miles _____	84
Figura 29.	Esquema del procedimiento para obtener estimaciones de la subenumeración _____	87
Figura 30.	Distribución tamaño de hogar CNPV 2018 y Encuesta Nacional de Calidad de Vida 2018 _____	89
Figura 31.	Valores P de la estadística de Wald para el ajuste de los 17 modelos usando todas las covariables _____	91

Lista de tablas

Tabla 1.	Ficha estadística de personas Censo Nacional de Población y Vivienda - CNPV 2018 faltantes e inconsistentes _____	21
Tabla 2.	Ficha estadística de personas ECV 2018 faltantes e inconsistentes _____	22
Tabla 3.	Resultados de emparejamiento inicial _____	23
Tabla 4.	Estimadores aplicados en emparejamiento inicial _____	23
Tabla 5.	Intervalos de confianza estimación en emparejamiento inicial _____	24
Tabla 6.	Frecuencia absoluta y relativa de la diferencia en la edad declarada y la edad imputada _____	25
Tabla 7.	Resultados estimación de la omisión incluyendo registros sin número de identificación _____	26
Tabla 8.	Resultados omisión con base espejos definitivas _____	27
Tabla 9.	Estimación de la omisión en grandes grupos de edad usando la metodología dual desagregada _____	28
Tabla 10.	Estimación de la omisión usando la metodología dual desagregada _____	30
Tabla 11.	Estimación de la omisión usando la metodología dual desagregada _____	31
Tabla 12.	Estimaciones históricas y estimaciones propias de las Tasas Globales de Fecundidad - TGF _____	34

Tabla 13.	Estimaciones históricas y propias de la esperanza de vida al nacer e° por sexo _____	36
Tabla 14.	Estimaciones históricas y propias de los saldos netos migratorios SNM _____	37
Tabla 15.	Proporción de defunciones en grupos de edad menores de 1 año _____	42
Tabla 16.	Proporción de defunciones en grupos de edad menores de 1 año _____	48
Tabla 17.	Porcentaje de zonas incompletas y no visitadas según tipo de residencia _____	53
Tabla 18.	Clasificación de categorías de cobertura geográfica _____	57
Tabla 19.	Ejemplo de intersección de áreas en el municipio de Cartagena _____	63
Tabla 20.	Departamentos y municipios donde se utilizó la estrategia de rutas _____	66
Tabla 21.	Total población estimada por departamento en las zonas de rutas de mayor complejidad operativa _____	72
Tabla 22.	Estimación VOCPA desagregación por clase _____	79
Tabla 23.	Variables de interés del CNPV 2018 con su respectiva codificación dado el caso _____	90
Tabla 24.	Municipios con cierre posterior al 29 de septiembre de 2018 _____	95

Lista de ilustraciones

Ilustración 1.	Mapa de clusters y outliers de Cartagena, relacionando P_U_NORM/P_NR_OBT 50	61
Ilustración 2.	Mapa de clusters y outliers de Cartagena, relacionando P_NR_OBT/P_U_NORM	62
Ilustración 3.	Mapa del resultado de la intersección Bivariado – Bivariado, Cartagena y Bolívar	63
Ilustración 4.	Mapa del resultado de la intersección Bivariado – Bivariado, Cali (Valle del Cauca)	64
Ilustración 5.	Distribución territorial de Subenumeración municipal, presentados en porcentaje	93
Ilustración 6.	Distribución de los factores de omisión para la ciudad de Cali	98
Ilustración 7.	Ajuste de cobertura por área a nivel municipal	99
Ilustración 8.	Ajuste de cobertura a nivel departamental	100

Abstract

This document presents the methodology and results for the estimation of the coverage error of the National Population and Dwellings Census of 2018 (CNPV 2018 in Spanish) of Colombia, at a national and subnational levels. To calculate the census omission for the entire country the Dual System method was used by comparing the Quality of Life Survey (ECV in Spanish) with the census results, additionally a census reconciliation process was carried out considering the components of the demographic dynamics between the period 1985-2018. For the disaggregation into subnational areas, a bottom-up method was applied considering three factors that generate omission that are: the absence of all the people in the dwellings at the time of the survey; geographical access limitations to some areas; and the random omission of people within the household. For the subnational estimation the potential of the georeferencing of the census units was used by applying econometric, Bayesian

and geostatistical methods that allowed to obtain the non-coverage of the minimum census units ('Blocks' in capitals and 'Populated centers' or 'Section' in the rural areas). The results showed that the omission of the CNPV 2018 at the national level corresponds to 8.5% using the Dual System method, and the census reconciliation presents a range of omission between 7.2% and 8.1%, being a more appropriate method to know structure adjustments by age and sex. At the subnational level an aggregate omission of 8.1% was obtained as a result of combining the three factors described, which was raised to the accepted national estimation of 8.5%, based on specific adjustments in municipalities with operational difficulties. The potential of the applied methodology corresponds to the possibility of making aggregations at the required levels, showing that populated centers and scattered rural areas have a total omission of 15.0% compared to 6.4% in urban areas.

1.

Introducción



Introducción

En 2018¹ 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.

Es clave en el tratamiento y análisis que se deriven de esta información la evaluación de la no cobertura censal, ya que “su conocimiento, es una tarea fundamental para orientar su tratamiento futuro” (Ribotta, 2005, pág. 1). Bajo esta premisa, el DANE realizó la estimación de la Omisión Censal a nivel nacional, departamental y municipal, entendiendo el valor de la información estadística generada a partir del CNPV 2018 y su uso por parte de organismos públicos y privados, en la toma de decisiones sobre política pública, bienestar social y económico.

Así pues, el presente documento tiene como objetivo que los usuarios finales de la información censal conozcan los aspectos metodológicos y los resultados obtenidos en las diferentes estimaciones de cobertura, las limitaciones presentadas y

las estrategias empleadas para su tratamiento, como parte de la política de integralidad de la información suministrada. En este documento, se desarrollan dos grandes capítulos que presentan la omisión censal a nivel nacional y subnacional. Para el nivel nacional, se expone la aplicación tanto de la evaluación directa² como indirecta³. La primera (directa), a partir del método Dual System correspondiente a la comparación de dos mediciones independientes sobre un mismo universo, en este caso, el Censo Nacional de Población y Vivienda y la Encuesta de Calidad de Vida - ECV 2018.

La segunda (indirecta), desarrollada a partir de la conciliación censal, considera los componentes de la dinámica demográfica. No obstante, esta evaluación y cálculo de la Omisión Censal por métodos exclusivamente demográficos y de encuesta post-censal, presentan limitaciones para conocer este valor a nivel municipal y más aún en su desagregación por clase (cabecera, centros poblados y rural disperso). En este sentido, es preciso retomar los principios de Chackiel y Macio (1978) citados por CELADE 2014.

“No existe un método único capaz de proveer el mejor ajuste ni hay una técnica que reúna las cualidades de una receta universal. Por el

¹ La operación censal se realizó en una fase virtual a través del censo electrónico del 8 de febrero a 8 de marzo y una fase en terreno del 18 de abril a 31 de octubre, con recuperación de cobertura en el departamento del Valle y los municipios de Tibú, Tumaco, La Macarena y Achí entre los meses de noviembre y diciembre del 2018.

² Consisten, principalmente, en volver a empadronar, al menos una parte de la población, haciéndolo de forma tal que pueda obtenerse una medida del número de personas que fueron omitidas en el recuento censal o incluido erróneamente en el mismo (Naciones Unidas, 1953). Citado en: BID-CELADE/CEPAL, 2006.

³ Se refiere a aquella que se realiza recurriendo a medios analíticos. Ello implica la comparación de las cifras censales entre sí y con los datos obtenidos de otras fuentes o distribuciones teóricas. (BID-CELADE/CEPAL, 2006).

contrario, el trabajo de corrección y evaluación se rige por los intentos sucesivos y las pruebas alternativas que conducen al ajuste más plausible” (CELADE-UNFPA, 2014, pág. 14).

Por tal motivo, en el segundo capítulo, se desarrolla, para el nivel subnacional, la desagregación por municipio y el área obtenida, a partir del análisis espacial de los factores generadores de omisión asociados a 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 y los resultados censales como fuentes para estimar la no cobertura, apoyado en covariados geográficos y herramientas de análisis espacial, aplicando modelos econométricos y de autocorrelación espacial, que permitiera, con valores conocidos en el espacio de un fenómeno o variable, determinar la tendencia de esta misma variable en los lugares donde no se conoce.

Para la primera tipología, correspondiente a viviendas ocupadas con todas las personas ausentes, se presenta 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

describen las herramientas geoespaciales que identifican la concentración y clúster de unidades no visitadas, a partir de covariables espaciales que determinan la densidad de población. Para la tercera tipología, denominada subnumeración, se presenta los modelos econométricos variables de la ECV 2018 que estima el número de personas omitidas dentro del hogar. Finalmente se presenta la suma las tres tipologías por municipio y clase, con el cual se estimó el total de la omisión por municipio.

Lo anterior, 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 el equipo del 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 son variables antropocéntricas que permitirán correlacionar el volumen de población a niveles geográficos desagregados, principalmente en áreas de rutas⁴ no visitadas.

⁴ 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.

Evaluación de cobertura a nivel nacional



La cobertura censal se define como el grado de completitud de las unidades observadas. Por causa de algunos factores exógenos que impactan en la operación estadística, el recuento total de estas unidades no siempre es posible. Para el CNPV 2018 estos factores estuvieron relacionados con:

- *Dificultades administrativas relacionadas con el personal y transporte*
- *No respuesta por parte de algunos actores sociales*
- *Territorios extensos y de difícil acceso*
- *Orden público*
- *Factores climáticos*

Una vez entendido que la operación estadística censal está sujeta a un margen de error,

proporcionar indicadores que evidencien la calidad del CNPV se convierte en una etapa fundamental dentro de la operación estadística. Para ello se emplearon a nivel nacional dos métodos que por su naturaleza arrojan estimaciones independientes. El primero de ellos, se refiere al uso de una encuesta de control de cobertura y su emparejamiento con el CNPV (método directo). El segundo, corresponde a un análisis de los componentes de la dinámica demográfica y su evolución a partir de otros censos (método indirecto).

Estas dos estimaciones realizadas no compiten entre sí; por el contrario, se trata de presentar los escenarios evaluados, sus alcances y limitaciones, con el fin de que los usuarios conozcan los procedimientos empleados para estimar los errores de cobertura informados, de tal forma que la toma de decisiones se base en los criterios de transparencia e integralidad en la fuente.

Estimación nacional por método directo

1.1 | Método Dual System

Los métodos directos de mayor uso para estimar los tamaños poblacionales son los basados en captura y recaptura. Estas técnicas fueron desarrolladas en estudios de fauna para estimar abundancia⁵. Sin embargo, su uso en poblaciones humanas fue introducido por Wolter (1986) usando la técnica presentada por ChandraSekar, (1949). Esta metodología

básicamente consiste en el aprovechamiento de las fuentes de información proporcionadas por los sistemas estadísticos. Dichas fuentes de información pueden provenir de encuestas realizadas de forma periódica en los hogares, censos de población, registros administrativos entre muchos otros, como expuso el CELADE-UNFPA en 2014:

⁵Para algunas aplicaciones ver (Kekäläinen, 2008) y (Olsson, 2006)

“El método dual System se basa en la comparación de dos mediciones independientes sobre un mismo universo. El concepto básico subyacente a este sistema es que, dadas las siguientes categorías, es posible estimar la población total:

	Presente en el Censo	Ausente en el Censo	
Presente en la encuesta	A	B	TE
Ausente en la encuesta	C	X	
	TC		POB

A = Personas contadas en ambos.

B = Persona contadas solo en la encuesta.

C = Personas contadas solo en el censo.

TC = Total contado por el censo.

TE = Total contado por la encuesta.

X = Desconocido por ambos.

La población total puede estimarse como:

$$POB = (TC * TE) / A$$

O, lo que es lo mismo $X = (C * B) / A$, donde la población será $POB = A + B + C + X$

La cobertura censal respecto a la población estimada ($100 * TC / POB$) será entonces = $100 * (A + C) / (A + B + C + X)$ ”

La selección de este método se realizó en virtud de los avances de interoperabilidad de la información censal, con los registros administrativos y encuestas, a través de la variable de número de documento

de identidad, que por primera vez fue recopilada en Colombia durante una operación censal, con lo cual se logró la aplicación de emparejamientos determinísticos entre las bases de datos.

1.2 | Fuentes de información

Para la evaluación directa de la cobertura censal es indispensable contar con una fuente de información diferente al censo, denominada encuesta post-censal o de post-enumeración, que debe cumplir con los siguientes requerimientos (CELADE-UNFPA, 2014, pág. 56):

- “Asegurar el financiamiento adecuado para realizarla correctamente, y que no se deban tomar decisiones que violen las normas establecidas para que se cumplan con éxito los objetivos fijados. En particular, debe ser un presupuesto

independiente, que no compita con los recursos asignados al Censo.”

- *“Considerar la independencia total de los responsables de dirigirla y del personal que trabajará en el terreno respecto del censo. La idea es que sea independiente tanto de las personas que trabajaron en el censo como desde el punto de vista estadístico.”*
- *“No debe existir un desfase temporal importante con la operación censal, para evitar cambios en los hogares que modifiquen la situación que se presentaba al momento del censo, y también porque comienza a influir la memoria sobre lo declarado previamente.”*
- *“El tamaño y el diseño muestral probabilístico deben tener una correcta representatividad, no solo del total nacional, sino también de aquellas áreas*

y sectores de la población que serán objeto de evaluación de su cobertura.”

- *“El cuestionario debe ser compatible con los conceptos y definiciones utilizados en el censo.”*
- *“El personal de la encuesta, en todas las instancias, tiene que ser altamente capacitado, pues deberá resolver por ejemplo si corresponde empadronar a determinadas personas en casos complejos.”*
- *“Se debe utilizar como marco muestral la misma cartografía empleada en el censo, la que también debería tener una evaluación.”*

A partir de los anteriores criterios, se determinó que la Encuesta de Calidad de Vida - ECV cumple con los requerimientos de la encuesta post-censal, con un alcance espacial y temporal que permite obtener resultados para total nacional.

1.3 | Procedimientos

Una vez que se seleccionaron las fuentes que contienen la información de interés, se realizó un cruce entre la base de datos del CNPV y la fuente post-censal seleccionada (ECV), con el objetivo de encontrar cuantos individuos son comunes en

las dos operaciones, los resultados se resumen en tablas de contingencia 2×2 , tal que la información de las personas encontradas en la encuesta ECV se encuentra en las columnas y las del CNPV en las filas, de la siguiente forma:

CNPV	ECV		
	SI	NO	TOTAL
SI	n_{11}	n_{10}	$n_{1.}$
NO	n_{01}	n_{00}	$n_{0.}$
TOTAL	$n_{.1}$	$n_{.0}$	N

En donde se cuenta con los siguientes valores observados:

- n_{11} representa la totalidad de personas que se encuentran en la encuesta ECV y en el CNPV
- n_{10} la totalidad de personas que se encuentran en la encuesta ECV y no en CNPV
- n_{01} representa la totalidad de personas que no se encuentra en la encuesta ECV y si en el CNPV

Y por último, los valores no observados que están sujetos a estimación:

- n_{00} que representa las personas que no se encuentran ni en la encuesta ECV ni en el CNPV.
- N que representa el total de personas.

a. Estimadores empleados

Como propuesta metodológica, se consideraron los estimadores de Chandrasekar–Deming (ChandraSekar, 1949), Chapman (Brittain, 2009) y Nour (Nour, 1982), basados en las siguientes ecuaciones:

- **Chandrasekar–Deming:**

$$N = \frac{n_{1.} \cdot n_{.1}}{n_{11}}$$

- **Chapman:**

$$N_c = \frac{(n_{1.} + 1)(n_{.1} + 1)}{(n_{11} + 1)}$$

- **Nour:**

$$N_N = n_{11} + n_{10} + n_{01} + \frac{2n_{11} n_{10} n_{01}}{n_{11}^2 + n_{10} n_{01}}$$

Es importante mencionar los supuestos que se deben asumir para los dos primeros estimadores:

- Las fuentes de datos deben ser independientes.
- La población de cada fuente debe tener los mismos límites geográficos y definición de residencia.
- Debe existir una migración mínima (inmigración e emigración) de la población si la otra fuente de datos es una encuesta de hogares retrospectiva o un censo de población.

Dichos supuestos son plausibles ya que, primero, los objetivos del CNPV y la encuesta ECV son diferentes; segundo, no existe un traslape temporal ya que el censo se realizó de enero a octubre de 2018 y la ECV de noviembre a diciembre de 2018; tercero, las metodologías de selección de individuos son diferentes; cuarto, existe una independencia operativa entre la Dirección de Censo y Demografía (DCD) con la Dirección de Metodología y Producción Estadística (DIMPE).

Dentro de las ventajas con las que cuenta esta metodología, se encuentra la confiabilidad de las estimaciones, esto debido al emparejamiento directo de las fuentes y una reducción en supuestos que se deben asumir en comparación con otros métodos. Además, hay una reducción de tiempos y costos, estimaciones a bajos niveles de desagregación, además de una fácil interpretación de los resultados obtenidos.

Los estimadores de Chapman y Nour se obtienen con el fin de observar el comportamiento y comparar los resultados con Chandrasekar, dado que en (Sadinle, 2008) se encuentra que, para ciertos niveles de prevalencia, el estimador es asintóticamente insesgado*, en caso de alcanzar dichas estimaciones en los parámetros, la mejor opción en términos de parsimonia, corresponde al primer estimador propuesto.

b. Transformación Fonética

Con el fin de mitigar el sesgo de escritura presente en los nombres y apellidos, debido a que son susceptibles a presentar errores de escritura, se hizo relevante la construcción de una función que convierta una cadena de caracteres a su fonema y lograr identificar el nombre por el sonido que produce al ser pronunciado.

El objetivo básico fue codificar de la misma forma los nombres con la misma pronunciación. Para tal fin, se desarrolló una función que recibe una cadena de caracteres como parámetro, realiza la conversión de cada carácter y retorna su representación fonética. Algunos ejemplos son:

ENTRADA	SALIDA
YOHANA	YN
JOANA	YN
JENNY	YNI
YENY	YNI
GONZALES	GNZLZ
GONZALEZ	GNZLZ

c. Estimaciones propuestas

El cruce de las bases de datos se realizó en tres combinaciones diferentes, derivadas de cambios y mejoramientos en las fuentes de información, los cuales corresponde a los siguientes criterios:

- **Emparejamiento inicial:** Se desarrolló un procedimiento de búsqueda de coincidentes con las variables de nombre, fecha de nacimiento, documento de identidad y tipo de documento presentes en el CNPV y ECV. Cuando alguna de estas variables se encontró vacía, se mejoró el registro usando el Archivo nacional de identificación (ANI) o el Registro civil de nacimiento (RCN). Para este emparejamiento no se incluyeron los registros sin documento de identidad de ambas bases.

El procedimiento de cruce de las bases de datos se realizó de manera determinística, es decir, un cruce directo de datos en la base espejo de la ECV con la base espejo del CNPV, de la siguiente forma: primero se cruzan las bases usando las variables número y tipo de identificación de la persona, una vez se obtiene el cruce de personas por estas dos variables, se excluyen de las bases espejo y los remanentes se vuelven a cruzar por las variables de edad, nombres y apellidos en sus formas fonéticas.

- **Emparejamiento incluyendo registros sin identificación:** La segunda estimación se realizó a partir de las bases mejoradas aplicando el mismo procedimiento anteriormente descrito, sin embargo, en esta estimación se incluyeron todos los registros sin documento de identidad.

- **Emparejamiento con la base espejo definitiva:** Para este emparejamiento se mejoró la base espejo CNPV con nuevos algoritmos que redujeron el número de casos vacíos en la variable identificación.

Estas tres estimaciones corresponden a la evolución de la información disponible,

considerando el criterio de oportunidad para la producción de indicador de cobertura censal. En este sentido, el emparejamiento inicial se realizó durante el primer semestre de 2019 con las bases mejoradas. Este primer escenario fue evaluado por el comité de expertos, quienes en la misión de evaluación del CNPV realizada entre 10 y 14 de junio de 2019, propusieron la

inclusión de todos los registros sin identificación, esta implementación se desarrolló en la ECV y la base censal, para la presentación de resultados censales el 4 de julio de 2019. A partir de esta estimación, se implementaron algoritmos para la obtención de la base espejo definitiva con reducción de número de casos de vacíos en la variable número de identificación.

Figura 1. | Flujo de procesos realizados en las estimaciones por método Dual System.



Fuente: DANE, Dirección de Censos y Demografía

1.4 | Resultados

a. Emparejamiento Inicial

Para el emparejamiento inicial se mejoraron las bases de datos del CNPV y la ECV a partir del empalme con la variable de identificación con las bases de Archivo

Nacional de identificación - ANI o el Registro civil de nacimiento - RCN, como se mencionó anteriormente. El procedimiento empleado, tiene como resultado las bases de datos denominadas espejo que cuentan con las siguientes estadísticas de mejora:

Tabla 1. | Ficha estadística de personas CNPV faltantes e inconsistentes

Base Cruda				
NOMBRE VARIABLE	REGISTROS VACIOS	REGISTROS CON DATOS	REGISTROS INCONSISTENTES	REGISTROS CONSISTENTES
PA_TIPO_DOC	0,3%	99,7%	0,0%	99,7%
PA1_NRO_DOC	9,6%	90,4%	0,3%	90,1%
PB_1ER_APELLIDO	0,0%	100,0%	0,1%	100,0%
PB_2DO_APELLIDO	14,4%	87,6%	0,1%	87,5%
PA_1ER_NOMBRE	0,0%	100,0%	0,0%	100,0%
PA_2DO_NOMBRE	31,4%	68,6%	0,1%	68,5%
PA1_FECHA_NAC	12,4%	87,6%	0,8%	86,8%
P_SEXO	0,1%	99,9%	0,0%	99,9%

Base Espejo				
NOMBRE VARIABLE	REGISTROS VACIOS	REGISTROS CON DATOS	REGISTROS INCONSISTENTES	REGISTROS CONSISTENTES
PA_TIPO_DOC	0,2%	99,8%	0,0%	99,8%
PA1_NRO_DOC	7,8%	92,2%	0,2%	92,0%
PB_1ER_APELLIDO	0,0%	100,0%	0,1%	100,0%
PB_2DO_APELLIDO	6,2%	93,8%	0,0%	93,8%
PA_1ER_NOMBRE	0,0%	100,0%	0,1%	100,0%
PA_2DO_NOMBRE	27,2%	72,8%	0,1%	72,7%
PA1_FECHA_NAC	5,2%	94,8%	0,0%	94,8%
P_SEXO	0,1%	99,9%	0,0%	99,9%

Nota: la base cuenta con un total de registros de 44.513.219 y un total de registros sin duplicados 44.154.946, corte 29 marzo de 2019

Fuente: DANE, Dirección de Censos y Demografía. CNPV 2018

En la base CNPV se muestra una mejoría en el documento de 1.8% que representa una población de aproximadamente 800.000 personas que ingresan al análisis.

Tabla 2. | Ficha estadística de personas ECV 2018 faltantes e inconsistentes.

Base Cruda				
NOMBRE VARIABLE	REGISTROS VACIOS	REGISTROS CON DATOS	REGISTROS INCONSISTENTES	REGISTROS CONSISTENTES
PA_TIPO_DOC	0,0%	100,0%	0,0%	100,0%
PA1_NRO_DOC	2,0%	98,0%	7,6%	91,0%
PB_1ER_APELLIDO	3,0%	97,0%	0,1%	97,0%
PB_2DO_APELLIDO	49,0%	51,0%	0,2%	50,0%
PA_1ER_NOMBRE	0,0%	100,0%	0,0%	100,0%
PA_2DO_NOMBRE	9,8%	68,6%	0,0%	90,2%
PA1_FECHA_NAC	0,0%	87,6%	0,0%	100,0%
P_SEXO	0,0%	100,0%	0,0%	100,0%

Base Espejo				
NOMBRE VARIABLE	REGISTROS VACIOS	REGISTROS CON DATOS	REGISTROS INCONSISTENTES	REGISTROS CONSISTENTES
PA_TIPO_DOC	0,0%	100,0%	0,0%	100,0%
PA1_NRO_DOC	2,0%	98,0%	6,4%	92,1%
PB_1ER_APELLIDO	1,0%	99,0%	0,0%	99,0%
PB_2DO_APELLIDO	15,0%	85,0%	0,1%	84,9%
PA_1ER_NOMBRE	0,0%	100,0%	0,0%	100,0%
PA_2DO_NOMBRE	23,2%	77,0%	0,0%	76,8%
PA1_FECHA_NAC	0,0%	100,0%	0,0%	100,0%
P_SEXO	0,0%	100,0%	0,0%	100,0%

Nota: la base cuenta con un total de registros 282.927 corte 29 de enero de 2019
Fuente: DANE, Dirección de censos y demografía. CNPV 2018

Para la base ECV se muestra una mejoría en documentos inconsistentes de 1.2% que representa una población de aproximadamente 3400 personas que ingresan al análisis.

En el emparejamiento inicial se usaron las bases espejo, excluyendo los registros sin número

de identificación, dando como resultado un número de filas en la base espejo del CNPV de 38.165.403 y de 256.878 para a base espejo de la encuesta ECV. Estos valores corresponden a las marginales de la tabla dos por dos que se muestra a continuación:

Tabla 3. | Resultados de emparejamiento inicial.

CNPV	ECV		
	SI	NO	TOTAL
SI	207.143	38.408.260	38.615.403
NO	49.753	9.221.817	9.271.552
TOTAL	256.878	47.630.077	47.886.955

Fuente: DANE - Dirección de censos y demografía

De acuerdo a lo anterior, las bases espejo filtradas estiman la población a total nacional usando la metodología dual con el estimador de Chandrasekar en 47.886.995.

Adicionalmente, en la siguiente tabla se muestra el resultado de los otros dos estimadores con sus respectivos cálculos de omisión. Cabe

destacar, que la estimación de Chapman difiere en 85 individuos mientras que Nour, estima aproximadamente 7 millones de personas menos, lo que no refleja el comportamiento poblacional ni las condiciones sociodemográficas del país. Por tal motivo la mejor estimación posible es dada por el estimador de Chandrasekar y es el escogido para estimar la población en Colombia para el año 2018.

Tabla 4. | Estimadores aplicados en emparejamiento inicial.

	Estimación	Estimación de la cobertura con documento consistente (38.615.403)	Estimación de la omisión con documento consistente (38.615.403)	Estimación de la cobertura Total Censo (44.513.219)	Estimación de la omisión Total Censo (44.513.219)
	Total				
Chandrasekar	47.886.995	80,64%	19,36%	92,95%	7,05%
Chapman	47.886.910	80,64%	19,36%	92,95%	7,05%
Nour	39.070.323	98,84%	1,16%	113,93%	-13,93%

Fuente: DANE, Dirección de censos y demografía

Los cálculos de omisión y cobertura usan como numerador la marginal del censo, que es de 38.615.403 personas con documento válido, dando una cobertura de aproximadamente el 80.1%, sin embargo, el hecho de no contar con un documento de identificación dentro de la base de datos, no implica que el individuo no fuese censado, por lo tanto, al usar la población efectivamente censada dentro de la

base de datos el resultado de la omisión es de aproximadamente el 7%. Al aplicar la técnica de corrección de sesgo y estimación de un intervalo de confianza descrita en (Sadinle, 2009) sobre el estimador de Chandrasekar, se estima que la población al total nacional es de 47.306.362 con una cobertura estimada usando como numerador el total de personas efectivamente censada de 94.1%.

Tabla 5. | Intervalos de confianza estimación en emparejamiento inicial.

	Estimación total	Corrección del sesgo	Límite Inferior	Límite Superior	Estimación de la cobertura con documento consistente (38.615.403)	Estimación de la cobertura Total Censo (44.513.219)
Chandrasekar	47.886.995	47.306.362	47.216.780	47.395.720	81,60%	94,10%

Fuente: DANE, Dirección de censos y demografía

Producto de los anteriores resultados se tienen las siguientes observaciones para el emparejamiento inicial:

- *La muestra de la ECV tiene como probabilidad de inclusión 0 a las personas dentro de lugares especiales de alojamiento (LEA).*
- *Solo el 86.7% de los individuos en el CNPV son parte del análisis*
- *Solo el 90.7% de los individuos en la ECV son parte del análisis*
- *Excluir individuos del análisis afecta las probabilidades de inclusión y podría generar sesgos en la estimación*

- *Es una estimación preliminar, debido a no contarse en el momento de la estimación con las bases finales de ECV y CNPV*

- *No se garantiza la unicidad de individuos comunes a las dos fuentes.*

b. Emparejamiento incluyendo registros sin identificación

A partir de las recomendaciones realizadas por el comité de expertos evaluadores del CNPV, se propuso la inclusión de todos los registros sin identificación en las dos bases de datos, para su posterior emparejamiento. Ahora, dado que en el CNPV la información fue suministrada por un informante idóneo, mientras que en la ECV fue un informante directo, podría presentarse

una discrepancia en las edades registradas en las dos fuentes de información, lo que generaría un subregistro de personas presentes en las dos bases. Para mitigar esta discrepancia, se realizó un análisis univariado de la diferencia entre la edad declarada en el CNPV y la edad imputada mediante el uso del ANI y el Registro

Civil de Nacimiento -RCN como resultado se obtiene la tabla 6, en donde se evidencia que 99.76% de estas diferencias se encuentra entre -10 y 10. Por lo tanto, como criterio adicional de emparejamiento se usará la diferencia de las edades del CNPV y ECV dentro de un rango de -10 a 10 años en valor absoluto.

Tabla 6. | Frecuencia absoluta y relativa de la diferencia en la edad declarada y la edad imputada.

Diferencia edad declarada y edad imputada	Frecuencia Absoluta	Frecuencia Relativa
-10	19.142	0,04
-9	8.321	0,02
-8	11.177	0,03
-7	12.604	0,03
-6	16.714	0,04
-5	23.755	0,05
-4	36.571	0,08
-3	81.032	0,18
-2	234.847	0,53
-1	1.761.865	4,01
0	40.023.686	91,15
1	1.394.388	3,18
2	72.259	0,16
3	37.471	0,09
4	22.664	0,05
5	11.140	0,03
6	9.018	0,02
7	6.057	0,01
8	5.715	0,01
9	7.687	0,02
10	12.133	0,03
Total	43.808.246	99,76

Fuente: DANE, Dirección de censos y demografía. CNPV 2018

El resultado de la estimación, incluyendo los parámetros anteriormente descritos, se presenta a continuación:

Tabla 7. | Resultados estimación de la omisión, incluyendo registros sin número de identificación.

** Corrección del sesgo del estimador Chrandrasekar vía Bootstrap Paramétrico 10000 réplicas

CNPV	ECV			Corrección Bootstrap paramétrico		Cobertura	Omisión
	SI	NO	TOTAL				
SI	207.143	38.408.260	38.615.403				
NO	49.753	9.221.817	9.271.552		Sin corrección	91,50%	8,50%
TOTAL	256.878	47.630.077	47.886.955	47.897.237**	Corregido	91,60%	8,40%

Fuente: DANE, Dirección de censos y demografía. CNPV 2018

Al estimar el total de la población residente en viviendas particulares en Colombia para el año 2018, usando la metodología dual incluyendo todos los registros sin identificación a total nacional se obtiene un total de 47'933.579 personas y una omisión estimada del 8,5%. Este resultado tiene una mejoría de una décima porcentual una vez se aplica la corrección Bootstrap del sesgo estadístico que presenta el estimador, mostrando una población de 47'897.237 personas residentes en viviendas particulares, y una omisión estimada de 8,4%, evidenciando que la magnitud del sesgo -producto de la estimación- es despreciable.

c. Emparejamiento con la base espejo definitivas

Las bases espejo del CNPV definitivas se obtuvieron con corte 9 de septiembre de 2019. Los ejercicios previos, fueron realizados con las bases espejo

anteriormente mencionadas, algunas de las mejoras en los nuevos ejercicios consisten en identificar: duplicados dentro de la base, imputaciones de las variables de identificación de personas, identificar de registros consistentes e inconsistentes, etc. Una vez que se identifican las inconsistencias y duplicados, se procede a eliminarlos de la base para realizar el cruce determinístico por tipo y número de documento.

En los ejercicios anteriores, se realizó el conteo del número de individuos presentes en las dos bases, sin embargo, el proceso de mejoramiento de las bases, puede incluir duplicados.

Es decir, si un registro presenta inconsistencia o es vacío, el algoritmo puede mejorar el dato con un registro ya existente en la base de datos, generando duplicados no marcados, por lo tanto, no era posible identificar con certeza la población B y C de la siguiente figura:

Figura 2. | Esquema de mejoras en el proceso de emparejamiento.

		Personas presentes en la ECV 2018			NO ENTRAN
		SI	NO	TOTAL	
Personas presentes en CNPV 2018	SI	A		D	E
	NO	B			
	TOTAL				
NO ENTRAN		C			

Fuente: DANE, Dirección de censos y demografía

De este modo, una de las mejoras realizadas al proceso fue garantizar la unicidad del cruce, eliminando de las bases espejo los registros marcados como duplicados o inconsistentes, los

registros duplicados por documento de identidad y los registros de personas en Lugares Especiales de Alojamiento (LEA), obteniendo los siguientes resultados con las bases definitivas.

Tabla 8. | Resultados omisión con base espejos definitivas.

		ECV			Cobertura	Omisión
		SI	NO	TOTAL		
CNPV	SI	260.703	43574621	43.835.324		
	NO	22.309	3.728.788	3.751.097		
	TOTAL	283.012	47.303.409	47.586.421	92.1%	7,90%

Fuente: DANE, Dirección de censos y demografía

d. Desagregación de estimaciones

En esta sección se mostrará el comportamiento del estimador con diferentes desagregaciones, resaltando que, a niveles geográficos, el supuesto

de migración se debe mantener para la aplicación del método, siendo esta una debilidad de la metodología propuesta, dado el periodo extendido que presentó el CNPV. Una vez se realizaron los cruces, primero por tipo y número de documento

y segundo por nombres en forma fonética y edad, como se describió anteriormente, se desagrega el conteo por variables como edad, sexo, rural y urbano y se obtiene las siguientes combinaciones de resultados:

- Desagregación por grandes grupos de edad y sexo

En este tipo de desagregación todos los supuestos necesarios para aplicar el método son plausibles, dado que al sumar todos los valores estimados se obtiene que la población residente en viviendas particulares para el año 2018 es de 47'935.208 personas con una omisión estimada del 8.5%, reproduciendo con una diferencia de 0.003% personas el resultado del total nacional, como se observa en la siguiente tabla:

Tabla 9. | Estimación de la omisión en grandes grupos de edad, metodología dual desagregada.

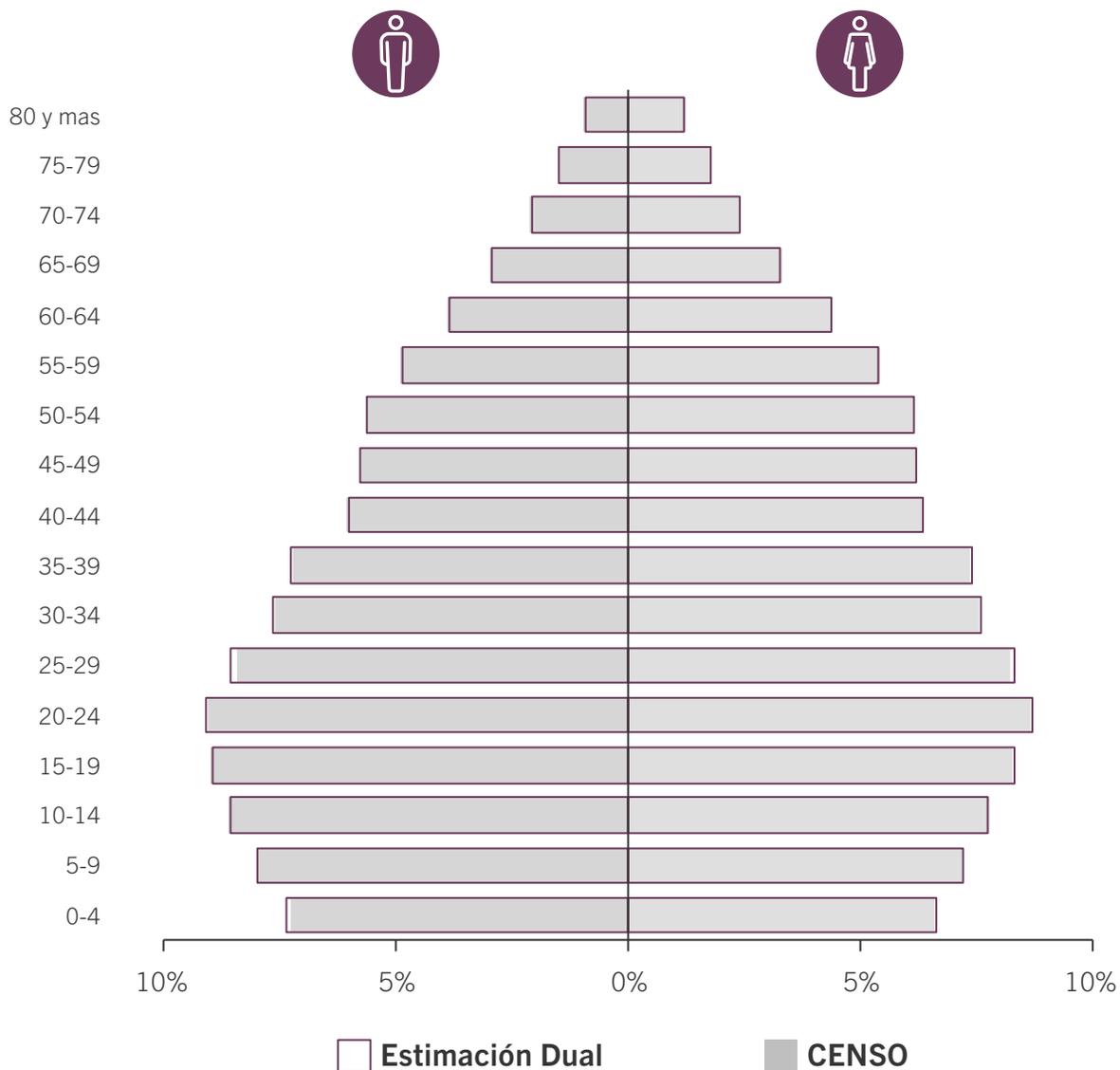
		SI censo y SI enc	" SI Enc NO Censo "	TOT CENSO	" SI Censo NO Enc "	TOT ENCUESTA	Pob Total estimada (DS)	Población omitida	% Cobertura	% Omisión
Hombres	0 a 14 años	34.459	3.292	5.089.603	5.055.144	37.751	5.575.832	486.229	91,3%	8,7%
	15 a 64 años	80.953	7.616	14.416.651	14.335.698	88.569	15.772.959	1.356.308	91,4%	8,6%
	65 años o más	11.854	897	1.820.037	1.808.183	12.751	1.957.760	137.723	93,0%	7,0%
Mujeres	0 a 14 años	32.500	3.035	4.853.649	4.821.149	35.535	5.306.905	453.256	91,5%	8,5%
	15 a 64 años	86.699	8.268	15.425.077	15.338.378	94.967	16.896.081	1.471.004	91,3%	8,7%
	65 años o más	12.458	981	2.248.605	2.236.147	13.439	2.425.670	177.065	92,7%	7,3%
Total		258.923	24.089	43.853.622	43.594.699	283.012	47.935.208	4.081.586	91,5%	8,5%

Fuente: DANE, Dirección de censos y demografía. CNPV 2018

- Desagregación por grupos quinquenales y sexo

Al igual que en la desagregación anterior, los supuestos son plausibles, mostrando una estimación de la población residente en viviendas particulares para el año 2018 de 47'949.408 personas con una omisión estimada del 8.5%, reproduciendo con una diferencia de 0.03% personas el resultado del total nacional, obteniendo la siguiente estructura de población:

Figura 3. | Estructura poblacional censada vs estimada



Fuente: DANE, Dirección de censos y demografía, CNPV 2018

La figura 3 muestra una comparación entre la pirámide poblacional de los individuos censados versus la población estimada, mostrando que la metodología de estimación conserva las estructuras, evidenciando una ligera discrepancia entre las mujeres entre los 25 a 29 años y en los

hombres de 20 a 39 años. Sin embargo, dada la homogeneidad del porcentaje de ajuste por no cobertura en los grupos quinquenales, el ajuste por edades será contrastado con otros métodos demográficos que permitan determinar la omisión por edades, como se expondrá más adelante.

Tabla 10. | Estimación de la omisión en grandes grupos de edad, metodología dual desagregada.

		SI censo y SI enc	" SI Enc NO Censo "	TOT CENSO	" SI Censo NO Enc "	TOT ENCUESTA	Pob Total estimada (DS)	Población omitida	% Cobertura	%Omisión
Hombres	0-4	10.323	1.093	1.554.157	1.543.834	11.416	1.718.711	164.554	90,4%	9,6%
	5-9	11.504	1.101	1.701.194	1.689.690	12.605	1.864.008	162.814	91,3%	8,7%
	10-14	12.632	1.098	1.834.252	1.821.620	13.730	1.993.689	159.437	92,0%	8,0%
	15-19	11.565	1.024	1.917.416	1.905.851	12.589	2.087.190	169.774	91,9%	8,1%
	20-24	9.391	955	1.929.459	1.920.068	10.346	2.125.672	196.213	90,8%	9,2%
	25-29	8.951	975	1.799.760	1.790.809	9.926	1.995.801	196.041	90,2%	9,8%
	30-34	8.413	856	1.619.445	1.611.032	9.269	1.784.219	164.774	90,8%	9,2%
	35-39	8.341	848	1.534.957	1.526.616	9.189	1.691.011	156.054	90,8%	9,2%
	40-44	7.689	684	1.291.572	1.283.883	8.373	1.406.468	114.896	91,8%	8,2%
	45-49	7.420	659	1.233.345	1.225.925	8.079	1.342.883	109.538	91,8%	8,2%
	50-54	7.380	635	1.204.754	1.197.374	8.015	1.308.415	103.661	92,1%	7,9%
	55-59	6.483	551	1.049.834	1.043.351	7.034	1.139.061	89.227	92,2%	7,8%
	60-64	5.320	429	836.109	830.789	5.749	903.532	67.423	92,5%	7,5%
	65-69	4.223	335	636.977	632.754	4.558	687.507	50.530	92,7%	7,3%
	70-74	3.068	232	455.076	452.008	3.300	489.489	34.413	93,0%	7,0%
	75-79	2.221	158	321.956	319.735	2.379	344.860	22.904	93,4%	6,6%
80-84	1.362	110	202.305	200.943	1.472	218.644	16.339	92,5%	7,5%	
85+	980	62	203.723	202.743	1.042	216.612	12.889	94,0%	6,0%	
Mujeres	0-4	9.763	998	1.480.164	1.470.401	10.761	1.631.470	151.306	90,7%	9,3%
	5-9	11.032	997	1.625.258	1.614.226	12.029	1.772.138	146.880	91,7%	8,3%
	10-14	11.705	1.040	1.748.227	1.736.522	12.745	1.903.559	155.332	91,8%	8,2%
	15-19	11.583	1.090	1.865.786	1.854.203	12.673	2.041.363	175.577	91,4%	8,6%
	20-24	10.323	1.026	1.949.509	1.939.186	11.349	2.143.270	193.761	91,0%	9,0%
	25-29	9.906	1.042	1.850.987	1.841.081	10.948	2.045.690	194.703	90,5%	9,5%
	30-34	9.331	943	1.695.061	1.685.730	10.274	1.866.366	171.305	90,8%	9,2%
	35-39	9.087	909	1.650.715	1.641.628	9.996	1.815.841	165.126	90,9%	9,1%

		SI censo y SI enc	" SI Enc NO Censo "	TOT CENSO	" SI Censo NO Enc "	TOT ENCUESTA	Pob Total estimada (DS)	Población omitida	% Cobertura	%Omisión
Mujeres	40-44	8.281	747	1.432.908	1.424.627	9.028	1.562.166	129.258	91,7%	8,3%
	45-49	8.109	715	1.397.370	1.389.261	8.824	1.520.581	123.211	91,9%	8,1%
	50-54	7.757	725	1.379.562	1.371.805	8.482	1.508.501	128.939	91,5%	8,5%
	55-59	6.824	582	1.220.572	1.213.748	7.406	1.324.671	104.099	92,1%	7,9%
	60-64	5.498	489	982.607	977.109	5.987	1.070.001	87.394	91,8%	8,2%
	65-69	4.252	341	748.401	744.149	4.593	808.421	60.020	92,6%	7,4%
	70-74	3.097	249	544.482	541.385	3.346	588.259	43.777	92,6%	7,4%
	75-79	2.313	163	402.510	400.197	2.476	430.875	28.365	93,4%	6,6%
	80-84	1.531	121	277.500	275.969	1.652	299.432	21.932	92,7%	7,3%
	85+	1.265	107	275.712	274.447	1.372	299.033	23.321	92,2%	7,8%
Total		258.923	24.089	43.853.622	43.594.699	283.012	47.949.408	4.095.786	91,5%	8,5%

Fuente: DANE, Dirección de censos y demografía CNPV 2018.
Ejercicios sin bases definitivas

- Desagregación geográfica: urbano, centros poblados-rural disperso y sexo

A diferencia de las desagregaciones etareas presentadas con anterioridad, el supuesto de migración no es plausible a estos niveles de granularidad, ya que el periodo extendido del CNPV, no garantiza que los individuos fueran censados y encuestados en la misma zona residencial. En consecuencia, la tabla 11 presenta una estimación

del total de la población colombiana residente en viviendas particulares para el año 2018 de 48'271.439 personas y una omisión del 9,2%, reproduciendo con un error de 0.7% personas el resultado del total nacional. Además de la distorsión del total nacional, temáticamente la estructura de omisión presentada va en contravía de los datos históricos y las condiciones sociodemográficas del país, ya que se presenta una cobertura mayor en centros poblados y rural disperso.

Tabla 11. | Estimación de la omisión usando la metodología dual desagregada

		SI censo y SI enc	" SI Enc NO Censo "	TOT CENSO	" SI Censo NO Enc "	TOT ENCUESTA	Pob Total estimada (DS)	Población omitida	% Cobertura	%Omisión
Hombres	Urbano	65.231	7.117	16.140.790	16.075.559	72.348	17.901.824	1.761.034	90,2%	9,8%
	Centros poblados y rural disperso	62.035	4.688	5.187.507	5.123.472	66.723	5.577.377	391.870	93,0%	7,0%
Mujeres	Urbano	74.154	7.971	17.706.306	17.632.152	82.125	19.609.601	1.903.295	93,0%	9,7%
	Centros poblados y rural disperso	57.503	4.313	4.821.036	4.763.533	61.816	5.182.637	361.601	93,0%	7,0%
Total		258.923	24.089	43.853.639	43.594.716	283.012	48.271.439	4.417.800	90,8%	9,2%

Fuente: DANE, Dirección de censos y demografía, CNPV 2018

A partir de las estimaciones realizadas por el método directo, se puede concluir que, una vez identificados los alcances y limitaciones que presenta la metodología propuesta, y la necesidad de tener estimaciones al nivel más bajo de desagregación posible, la desagregación por grandes grupos de edad presenta una estimación del total de la población acorde a los requerimientos que con esta metodología se puede alcanzar, considerando las siguientes premisas:

- *La distorsión del total nacional bajo este escenario es del 0.03%.*
- *Este escenario de estimación conserva las estructuras de la población censada, con ligeras discrepancias para mujeres entre 25 y 29 años y hombres entre 20 y 29 años.*
- *El nivel de granularidad es considerablemente bajo.*

Estimación nacional por método indirecto

2.1 | Método de conciliación censal

La evaluación de cobertura a través de la conciliación demográfica se define como:

“Conjunto de procedimientos analíticos por medio de los cuales se logra evaluar y corregir la distribución por sexo y edad de un número sucesivo de censos de población y/o datos de registros permanentes, compatibilizándolos con las estimaciones de la mortalidad de la fecundidad y de la migración neta de cada uno de los periodos intercensal.” (Rincón & Fajardo, 2007).

Este método evalúa la información censal con los datos demográficos disponibles y su interacción en la llamada ecuación compensadora, la cual presenta la siguiente forma:

$$P(t) = P(0) + N(0,t) - D(0,t) + M(0,t)$$

En dónde:

- **P(0) y P(t)** son las poblaciones en el momento 0 y t respectivamente.

- **N(0,t) y D(0,t)** son los nacimientos y defunciones entre 0 y t.
- **M(0,t)** es el saldo migratorio neto entre 0 y t

Como afirma el CELADE-UNFPA (2014): “A partir de la ecuación compensadora se propone la dinámica demográfica coherente más plausible del país, y como subproducto se obtiene la función de error censal por sexo y edades y, por lo tanto, los porcentajes de omisión para esas categorías” (CELADE-UNFPA, 2014).

Para el CNPV 2018 se realizó la conciliación censal del total nacional, se evaluó el volumen poblacional y la estructura a nivel de agregación nacional a partir de los años 1985 y 2005, estableciendo las cifras de sobrevivientes en grupos quinquenales y los nacimientos y saldos netos migratorios estimados en los quinquenios intermedios.

Para tal fin, se definieron las tendencias históricas de la evolución de los componentes y la población

ajustada de censos anteriores, con la realización de los siguientes subprocesos:

a. Evaluación de las fuentes poblacionales históricas y estimaciones propias por componentes.

b. Estimación de la población de las cohortes de 0-9 años, a partir del método de diagrama de Lexis.

c. Estimaciones de población de 10 años y más.

d. Construcción de escenarios de volumen de población 2018 a partir del juego de estimaciones propuestas e históricas.

e. Estimación de la omisión censal por el método de conciliación censal.

2.2 | Evaluación de las fuentes de información

El ejercicio de conciliación censal requiere una previa evaluación de las fuentes insumos del cálculo de los componentes de fecundidad, mortalidad, migración y la población base ajustada, así como el conocimiento de las estimaciones históricas derivadas de estos insumos para el periodo estudiado. Se revisaron los registros históricos de diferentes estimaciones disponibles, sistematizándolas para su posterior evaluación de acuerdo con cada uno de los componentes, para obtener una base de análisis de calidad de los datos a emplear.

a. Insumos para población base

Como se señaló anteriormente, los censos como toda operación estadística no está exenta de errores de cobertura. Bajo esta premisa, para el proceso de conciliación demográfica del CNPV 2018, fue necesario identificar documentalmente los reportes de cobertura establecidos para los censos históricos 1985, 1993 y 2005, como parte de la evaluación de la población base de partida, obteniendo las siguientes fuentes documentales de su evaluación de cobertura:

- **XV Censo de población y IV de vivienda 1985:** Se dispone de un estudio

de evaluación de calidad y cobertura, en donde se presentan los análisis de las estimaciones de omisión que se realizaron con métodos diferenciados. El primero referido al resultado de la muestra post censal adelantada por el Centro Nacional de Consultoría - CNC y los otros dos mediante métodos indirectos realizados por los estudios del CEDE-FEI y DANE DNP (Manrique, 1990). A partir de esta publicación se logró seleccionar el ajuste de cobertura más plausible para esta serie de datos.

- **XVII Censo de población y V de vivienda 1993:** Evaluado en cobertura a partir de método indirecto de conciliación censal (DANE, 1998). Se debe señalar que la documentación de los ajustes de cobertura es mínima y no se presentan varios escenarios que permitan identificar la mejor estimación posible.
- **Censo general de población y vivienda 2005:** De acuerdo con la documentación existente, el proceso de evaluación de este censo comprendió dos etapas. La primera correspondió a la evaluación y compensación de población por no

cobertura geográfica y por contingencia de transferencia de información censal. La segunda contempló el proceso de evaluación y ajuste mediante el método demográfico de conciliación censal. (DANE, 2008). A partir de esta documentación, se optó por revisar la estimación de la población base 2005 a partir del periodo de análisis 1985- 2005 el cual facilita el análisis en grupos quinquenales de población, procedimiento que se encuentra documentado más adelante.

b. Insumos de fecundidad

El proceso de conciliación considera los cambios del componente de fecundidad a partir de las estimaciones de las Tasas Globales de Fecundidad -TGH⁶ y las Tasas Específicas de Fecundidad⁷ -TEF⁷ históricas que se encuentran disponibles, adicional a los cálculos propios realizados⁸ por el método de Brass, para Censo 1985 y, Gompertz relacional, para los censos 2005 y 2018. Su evaluación fue fundamental para conocer las tendencias y niveles de este componente.

Tabla 12. | Estimaciones históricas y estimaciones propias de las Tasas Globales de Fecundidad- TGF

Estimación	1990	1995	2000	2005	2010
Método de hijos propios. ENDS (Vehults, 2013)	3,2	2,92	2,73	2,54	2,3

Estimación	1987-1990	1992-1995	1997-2000	2002-2005	2007-2010	2012-2015
Método historia de Maternidad. ENDS (MinSalud-Profamilia, 2015)	2,9	3	2,6	2,4	2,1	2

⁶ Número de hijos que nacerían por mujer, si la mujer tuviera sus hijos en todos sus años reproductivos de acuerdo con las tasas de fecundidad específicas por edad de la población y épocas del estudio.

⁷ El número de nacimientos que ocurren durante un determinado año o período de referencia por cada mujer en edad reproductiva clasificada en edades quinquenales.

⁸ La descripción metodológica de estos cálculos se desarrolla en el documento de proyecciones y retroproyecciones de población con base en el CNPV 2018.

Estimación	1985-1990	1990-1995	1995-2000	2000-2005	2005-2010	2010-2015	2015-2020
Estimaciones elaboradas y ajustadas con base en otras estimaciones provenientes de Hijos Propios y P/F. Proyecciones a partir de curva logística (DANE, 1989)	3,13	2,92	2,75	2,61	2,5	2,41	2,34
Método P/F para el Censo 1993. Proyección logística. (DANE, 1998)	3,17	3,02	2,8	2,62	2,48	2,37	2,28
Método indirecto P/F propuesto por W. Brass (DANE, 2009)	3,34	3,14	2,86	2,6	2,45	2,35	2,29
Estimaciones propias método Gompertz linealizado. Para los censos 1985-2005-2018. Tendencias con función logística	3,12	2,82	2,58	2,42	2,27	2,09	1,97
Estimaciones WPP. CELADE. (CELADE, 2019)	3,17	3,01	2,7	2,44	2,1	1,92	1,82

Fuente: Citadas en cada estimación.

c. Insumos de Mortalidad

Los insumos de mortalidad empleados en la conciliación censal corresponden a las px (probabilidades de sobrevivencia a una edad x)⁹, contenidas en las tablas de vida. Al igual que con

los insumos de fecundidad, se revisaron las series históricas disponibles y se realizaron estimaciones propias¹⁰. El indicador de esperanza de vida¹¹ resume los cambios en las tablas de vida que se presenta a continuación:

⁹ Representa la probabilidad que tiene una persona de edad exacta x , de sobrevivir un año, es decir de llegar con vida a la edad exacta $x+1$.

¹⁰ La metodología de las estimaciones de mortalidad se presenta en el documento de proyecciones y retroproyecciones de población con base en el CNPV 2018.

¹¹ Representa el número de años que en promedio vive una persona desde la edad x en adelante hasta el final de la vida.

Tabla 13. | Estimaciones históricas y propias de la esperanza de vida al nacer e° por sexo

Esperanza de vida al nacer e° Hombres

Estimación	1985-1990	1990-1995	1995-2000	2000-2005	2005-2010	2010-2015	2015-2020
Función logística con las observaciones 1973-1985. puntos censales con el método distribución de Muertes (DANE, 1989)	65,51	66,36	67,36	68,41	69,51	70,61	71,36
Función logística con las observaciones 1985 -1993. (DANE, 2009)	64,65	65,34	67,07	69	70,67	72,07	73,08
Estimaciones propias (2019)	62,63	62,14	64,83	67,82	70,1	71,62	73,05

Esperanza de vida al nacer e° Mujeres

Estimación	1985-1990	1990-1995	1995-2000	2000-2005	2005-2010	2010-2015	2015-2020
Función logística con las observaciones 1973-1985. puntos censales con el método distribución de Muertes (DANE, 1989)	71,11	72,26	73,26	74,16	75,01	75,81	76,56
Función logística con las observaciones 1985 -1993. (DANE, 2009)	71,52	73,37	74,95	76,31	77,51	78,54	79,39
Estimaciones propias (2019)	69,34	70,95	73,08	75,67	77,58	78,61	79,58

Fuente: Citadas en cada estimación.

d. Insumos de Migración

En relación a los SalDOS Netos Migratorios - SNM¹² se dispone para los periodos intercensales 1985-2018, de la estimación de migración del Banco Mundial (2012), y las estimaciones internas resultado de la

ejecución de consultorías Recaño (2015 y 2019) y proyecciones anteriores (DANE, 2009).

Como es conocido, el componente migratorio presenta la mayor complejidad para su estimación, debido a la limitación en fuentes. Los primeros

¹² Diferencia entre los inmigrantes y los emigrantes de determinada unidad territorial.

ejercicios de conciliación se realizaron con las estimaciones disponibles provenientes de la consultoría de Recaño (2015), sin embargo, estas primeras estimaciones no contenían los efectos de la oleada migratoria de población proveniente de Venezuela que se presentó desde 2017, para

lo cual se realizaron unas segundas estimaciones en 2019, a partir del Registro Estadístico de Migración y la información de la Gran Encuesta Integrada de Hogares - GEIH¹³ por el mismo consultor que integraron esta coyuntura en el fenómeno.

Tabla 14. | Estimaciones históricas y propias de los saldos netos migratorios SNM

Saldos Netos Migratorios Hombres

Fuente	1985-1990	1990-1995	1995-2000	2000-2005	2005-2010	2010-2015	2015-2020
DANE 2009	-117.075	-119.200	-325.920	-304.741			
Banco Mundial, 2012	-121.086	-99.531	-88.044	-73.288			
Recaño 2015					-113.699	-83.804	-53.313
Joaquín Recaño, 2019					-208.450	-257.579	594.697
Estimación propia 2019	-61.956	-137.376	-132.532	-99.113			

Saldos Netos Migratorios Mujeres

Fuente	1985-1990	1990-1995	1995-2000	2000-2005	2005-2010	2010-2015	2015-2020
DANE 2009	-128.512	-149.064	-377.423	-356.410			
Banco Mundial, 2012	-132.915	-124.467	-101.956	-85.715			
Recaño 2015					-154.799	-111.951	-62.167
Recaño, 2019					-197.576	-270.992	644.099
Estimación propia 2019	-34.528	-105.595	-121.766	-113.732			

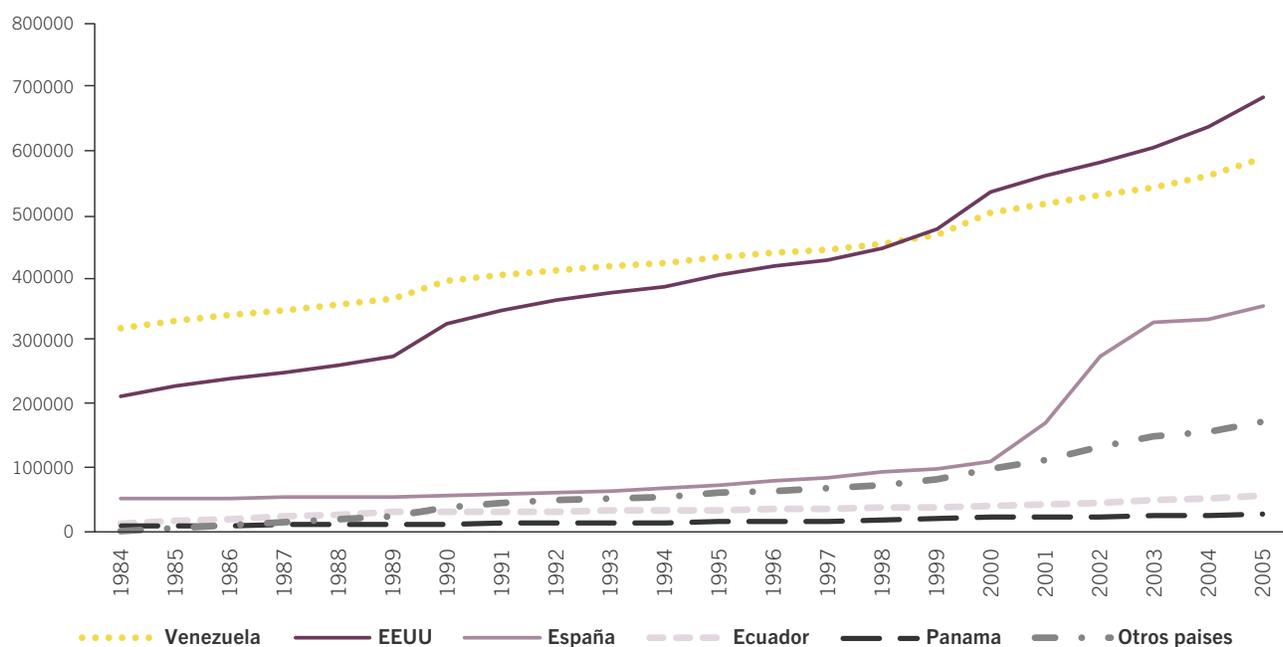
Fuente: Citadas en cada estimación.

¹³ Representa el número de años que en promedio vive una persona desde la edad x en adelante hasta el final de la vida.

Otras de las estimaciones del componente migratorio, correspondió al volumen de colombianos en el exterior periodo 1985-2005, en la cual se tomaron referencias

de estadísticas espejo, obtenidas a través de las fuentes externas que registran, censan o estiman los volúmenes de colombianos en otros países:

Figura 4. | Población de colombianos en otros países por año de llegada en el periodo 1985-2005



Fuente: DANE - Elaboración propia basada a partir de IMILA, IPUMS e INE España.

Para conocer la emigración a los países Latinoamericanos se consultaron fuentes como el sistema de Investigación de la Migración Internacional en Latinoamérica – IMILA, proyecto de la Comisión Económica para América Latina y el Caribe – CEPAL y directamente los censos de los países de interés como Ecuador, Venezuela y Panamá consultados a través de REDATAM¹⁴. Para los otros países de preferencia de migración, como Estados Unidos, se consultó el

American Community Survey ACS del Census Bureau, y para España las estadísticas disponibles en el Instituto Nacional de Estadística - INE.

El saldo Neto Migratorio fue obtenido a partir de la interpolación de la información censal 1985, 1993 y 2005 de los emigrantes y su diferencia frente al volumen de colombianos en el extranjero obtenido de las estadísticas espejo.

¹⁴ Acrónimo de recuperación de datos para áreas pequeñas por Microcomputador

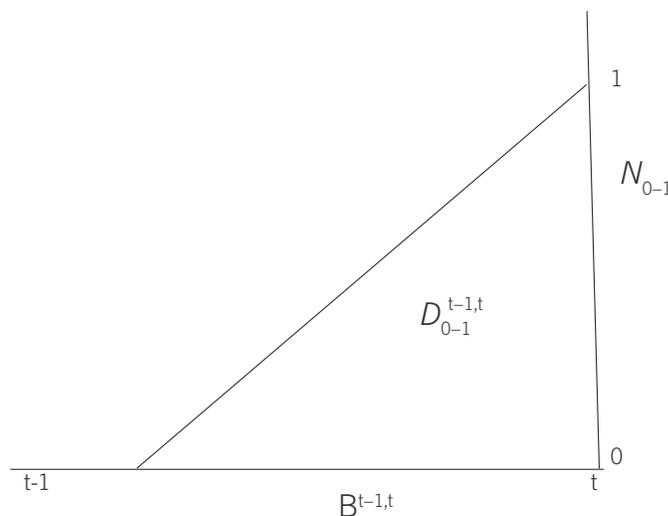
2.3 | Procedimientos

a. Estimación de la población de 0-9 años con Diagrama de Lexis.

El diagrama de Lexis es una técnica que permite calcular la evolución de una cohorte a partir de

la ecuación compensadora y su relación de la población en dos momentos. Es decir el grupo que al inicio del periodo tiene edades $x, x+1$ al final de este tendrá edades $x+1, x+2$, como se indica en la figura 5:

Figura 5. | Corrección de la población de 0-1 años en un censo de población.



Fuente: DANE, Elaboración basada en Rincón & Fajardo, Teoría y Métodos para la elaboración de estimaciones y proyecciones de población., 2007.

En donde,

- $B^{t-1,t}$ son los nacimientos del año t-1, t, corregidos.
- $\alpha D_{0-1}^{t-1,t}$ son las defunciones de menores de 1 año registradas entre el grupo de nacimientos del periodo t-1, t, ocurridas a lo largo del mismo periodo.
- N_{0-1} es la población sobreviviente de menores de 1 año, a inicio del año t.

Como se observa en el ejemplo, la población menor de 1 año en el momento t (año censal) proviene de los nacimientos ocurridos 1 año previo al mismo afectados por los fenómenos de mortalidad y migración en las edades de 0-1 años en el periodo t-1.

Bajo este marco, para la estimación de la población conciliada del CNPV 2018 menor de 10 años en edades simples a partir de diagrama de Lexis, se emplearon las estadísticas vitales de nacimientos y defunciones e información del Registro Estadístico de Migración - REMI¹⁵. A esta información se

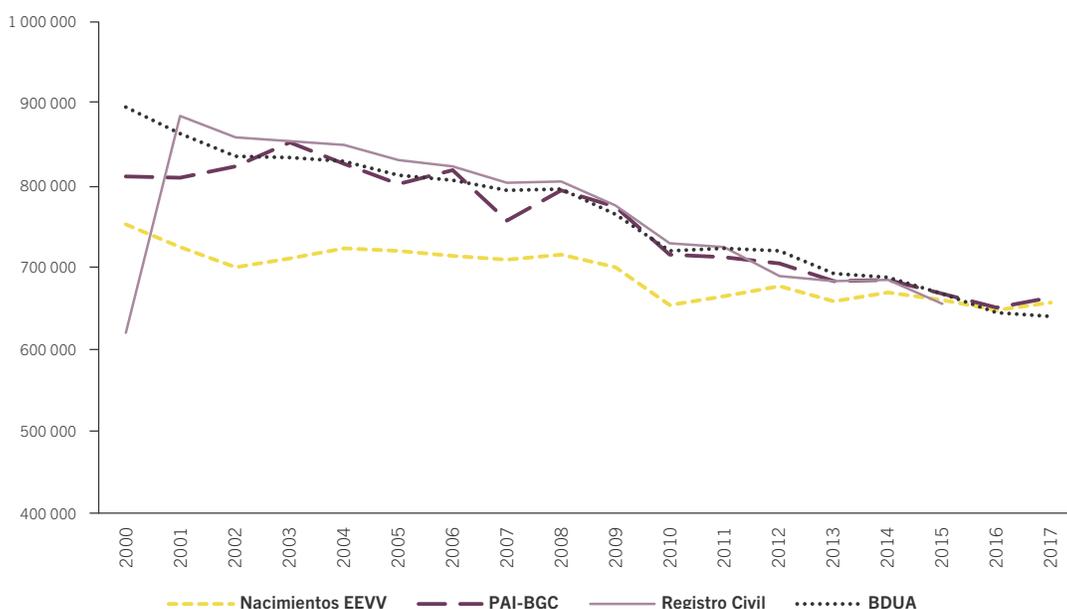
le realizó dos procesos previos, i) corrección de subregistro de las EEV de nacimientos en el periodo 2000- 2017¹⁶ y ii) estimación de los factores de separación para aplicación de las defunciones por cohortes:

1. Ajuste de nacimientos periodo 2000 -2017

Con el fin de evaluar la completitud de la serie de nacimientos de estadísticas vitales se emplearon

tres fuentes alternativas para contraste del nivel y cambio de la fecundidad en el país, la primera corresponde a una revisión de los registros de la Base de Datos única de Afiliados al Sistema de Salud - BDUA¹⁷, la segunda a una comparación de datos agregados por año de ocurrencia del evento con la información del registro civil¹⁸, y la tercera la misma comparación de valores agregados con información del Programa Ampliado de Inmunización - PAI¹⁹, como se muestra en la figura 6:

Figura 6. | Evaluación de registros de nacimientos a partir de fuentes colateral.



Fuente: DANE, elaborado a partir de EEV- DANE, BDUA, PAI (Min Salud), RCN (Registraduría Nacional del Estado Civil)

¹⁵ Aprovechamiento Estadísticos del registro de entradas y salidas de migración Colombia.

¹⁶ Se considera este periodo debido que se encuentra con registros colaterales que permiten ajustar toda las series.

¹⁷ Registro del Ministerio de Salud y Protección Social, que tiene como finalidad incorporar la información de afiliados al régimen subsidiado y contributivo de Salud.

¹⁸ Información base del sistema de identificación colombiano, que corresponde al documento público que legalmente prueba la existencia de una persona, la individualiza, le permite acceder a los beneficios que otorga el Estado y las obligaciones que tiene frente a la sociedad y la familia. (consultado en: <https://www.registraduria.gov.co/1-de-octubre-de-2012-No-68-EI.html>)

¹⁹ Programa que está a cargo en el país de la eliminación, erradicación y control de las enfermedades inmunoprevenibles. Su implementación y registro se encuentra a cargo del Ministerio de Salud y Protección Social.

A través de esta metodología, se logró evaluar la evolución de la fecundidad en el total nacional y conocer la cobertura del registro de nacimientos en el sistema de estadísticas vitales. El contraste de las fuentes colaterales evidencia una tendencia al descenso del volumen total de nacimientos y un mejoramiento de cobertura de las Estadísticas Vitales - EEV, en donde, las brechas que se presentan con otras fuentes han disminuido en las últimas décadas, logrando estimar un factor de ajuste de nacimientos para el periodo de referencia de análisis por diagrama de Lexis.

Las fuentes colaterales disponibles permiten comparar el número total de nacimientos desde el año 2000, evidenciando que los registros de las EEV presentaban una alta subcobertura en referencia a la información de Registro civil y BDUA, con una brecha del 18,3% con respecto a los registros de estas fuentes para el año de inicio de análisis, presentado mejoría en 10 años de casi 8 puntos porcentuales de la brecha con respecto a las mismas fuentes, alcanzando un valor de 10,3% en el 2010.

No obstante, la evaluación de la calidad de las EEV para el periodo reciente (últimos cinco años), presenta mayor complejidad. Si bien, la cobertura de estos hechos vitales ha mejorado con la puesta en marcha de la plataforma del Registro Único de Afiliados - RUAF, las otras fuentes también presentan registros tardíos, por tanto, su comparabilidad no genera de manera directa, requiriendo estimar las tendencias para calcular el factor de ajuste en este periodo.

Por otro lado, otros estudios referidos a la calidad de la información de las EEV, corresponden a los realizados en 2011 por DANE y UNFPA con la Evaluación de cabalidad de la EEV, (DANE-UNFPA, 2011), y los avances del método empírico de la completitud del registro de EEV (MECREV),

elaborado entre DANE, Ministerio de Salud y la Registraduría Nacional con el apoyo técnico de Vital Strategies, en el marco de la Iniciativa Datos para la Salud de Bloomberg Philanthropies (2019). Los cuales fueron contrastados con los indicadores estimados y el análisis anteriormente expuesto.

2. Estimación de los factores de separación de las defunciones de menores de 1 año.

Las defunciones del año $t+1$, que corresponde a la cohorte de los nacimientos $B^{(t-1,t)}$, se obtuvieron a partir del cálculo de los factores de separación (f_0), que distribuyan las defunciones del año calendario en las cohortes de nacimientos estudiados. Este procedimiento consistió en medir las defunciones de menores de un año basados en la relación geométrica del área que se encuentra encima de la diagonal y determinar las defunciones que ocurren en la franja superior de esta diagonal, a partir de la siguiente expresión:

$$h_i = L_i + 0.5 * A_i$$

L_i es el límite inferior, expresado en años, del intervalo de edades i .

A_i es la amplitud, también expresada en años, del intervalo i .

h_i es la proporción de muertes ocurridas en el tramo de edades i , que corresponde al año anterior.

Para ejemplificar lo expresado en la ecuación anterior, las defunciones se estiman a partir de las defunciones registradas por edad al momento de morir, separadas en los siguientes intervalos de tiempo.

Tabla 15. | Proporción de defunciones en grupos de edad menores de 1 año

Grupos de Edad	L(i)	A(i)	h(i)
Menores de 1 día	0	1	0,001
1 - 6 días	1	6	0,011
7 - 27 días	7	27	0,056
28 - 29 días	28	29	0,116
1-5 meses	30	152	0,29
6-11 meses	182	183	0,749

Fuente: DANE, Elaboración basada en CELADE, 1997

Una vez calculadas las proporciones $h(i)$ para todos los grupos de edad, se calcularon las defunciones correspondientes a las cohortes de $t-2$ y $t-1$, por sexo y año en el periodo 2009-2018. Finalmente, el factor de separación se obtuvo con la siguiente forma:

$$f_0 = \frac{\sum \delta D_i}{D_0}, \quad \delta D_i = h_i * D_i$$

Donde:

f_0 factor de separación edad 0

D_i Defunciones de los grupos de edad menores de 0 años

D_0 Defunciones edad 0

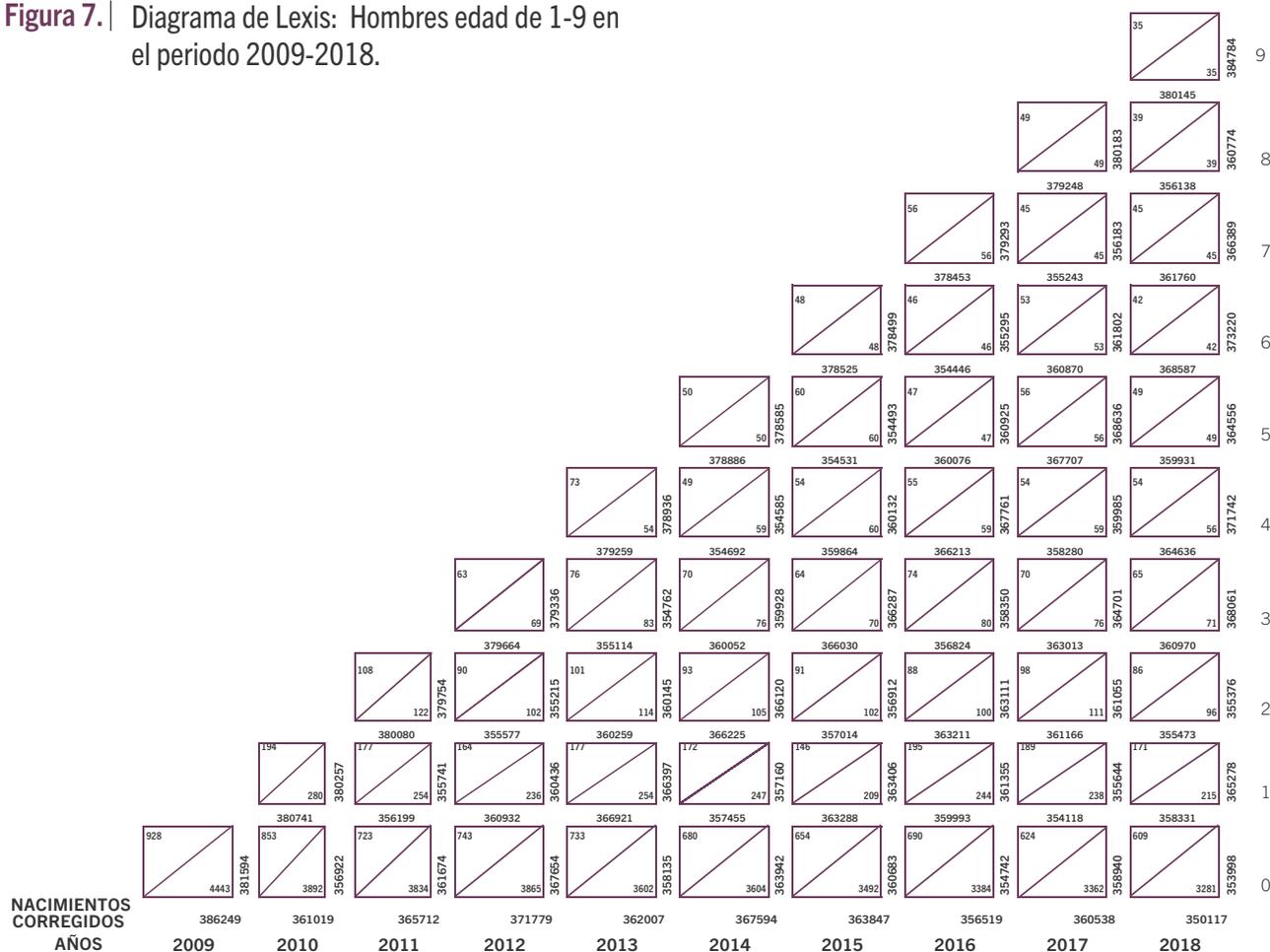
δD_i Defunciones edad 0 correspondientes a la cohorte de $t-2, t-1$

Para las edades de 1-4 se emplearon los factores de separación de Glover (Greville, Thomas., 1946) y para mayores de 5 años se asume comportamiento lineal de las defunciones con un factor de separación de igual a 0,50.

3. Estimación de la población de menor de 10 años en el diagrama de Lexis

A partir de los registros de nacimientos corregidos, las defunciones y la estimación de los saldos netos migratorios internacionales para el periodo 2009-2018 se estimó la población menor de 10 años por sexo, para el total nacional. Si bien la población de 0-4 es de mayor omisión con respecto a la población de 5-9, este diagrama permitió realizar un ajuste en edades simples que conserva la tendencia de los cambios en la fecundidad en los últimos 10 años, conservando las tendencias observadas con las estadísticas vitales, como se muestra en la figura 7:

Figura 7. | Diagrama de Lexis: Hombres edad de 1-9 en el periodo 2009-2018.



Fuente: DANE, elaborado a partir de EEV corregidas y REMI

b. Estimación de población de 10 años y más

Este proceso se basó en los referentes metodológicos de (CELADE, 2013), (Rincón & Fajardo, 2007) y consistió en la construcción de estimaciones de población por sexo y edad en grupos quinquenales, a partir de las series históricas y propias de los componentes de fecundidad, mortalidad y la migración, adicional a las poblaciones base de los diferentes años censales. Esta información fue dispuesta en un visor dinámico que permitió evaluar y obtener diferentes combinaciones con los componentes disponibles.

El visor fue construido fundamentado en la ecuación compensadora, incorporando la variable edad quinquenal y su relación en dos momentos del tiempo, evolucionando cada cohorte con los cambios de los componentes de mortalidad y migración en cada periodo, de la siguiente forma:

$$P_{x+5,x+9}^{t+5} = P_{x,x+4}^t - D_{x+5,x+9}^{t,t+5} + I_{x+5,x+9}^{t+5} - E_{x+5,x+9}^{t+5}$$

Donde,

$P_{x+5,x+9}^{t+5}$ es la población de edad de x+5 a x+9 en el año t+5, que proviene de la población x,

$x+4$ en el año t , una vez se aplican los factores de mortalidad y migración en el periodo $t+5$.

$P_{x,x+4}^t$ es la población de edad $x, x+4$ en el año t .

$D_{x+5,x+9}^{t,t+5}$ defunciones de población en el grupo de edad $x+5, x+9$ en el periodo $t+5$.

$I_{x+5,x+9}^{t+5}$ inmigrantes de población en el grupo de edad $x+5, x+9$ en el periodo $t+5$.

$E_{x+5,x+9}^{t+5}$ emigrantes de población en el grupo de edad $x+5, x+9$ en el periodo $t+5$.

Estas expresiones fueron aplicadas con las relaciones de supervivencias provenientes de las tablas de vida en quinquenios por grupos quinquenales y los SNM que resumen la diferencia entre los inmigrantes y los emigrantes, señaladas en el numeral anterior.

Los nacimientos ocurridos en cada quinquenio se estimaron a partir de las mujeres en edad fértil obtenidas con la expresión anterior, aplicando las tasas específicas de fecundidad como se muestra en la siguiente expresión:

$$N^{t,t+5} = \sum_{x=15}^{45} * P_{x,x+4}^{t,t+5} * f_{x,x+4}^{t,t+5}$$

Donde,

$N^{t,t+5}$ nacimientos en el periodo $t, t+5$

$f_{x,x+4}^{t,t+5}$ tasas específicas de fecundidad

La distribución por sexo de los nacimientos se obtuvo aplicando el índice de masculinidad al nacer de 105 hombres por cada 100 mujeres.

La combinación de estimaciones se estableció con los censos 1985 y 2005, los cuales fueron trasladados a mitad de periodo a 30 de junio de estos mismos años, el Censo de 1993 fue trasladado a 30 de junio de 1995 y la estimación resultante en el periodo 2020, fue trasladada a 30 de junio del año 2018.

Dentro de las combinaciones realizadas se evaluaron cuatro escenarios de la dinámica demográfica más plausibles de la estructura por edad y sexo del CNPV 2018, compatibilizado con los censos anteriores.

Escenario 1: Se tomó la población base de 1985 ajustado con el nivel señalado en (Manrique, 1990) previamente conciliado, con la estructura ajustada para menores de 10 con fecundidad intercensal 1973-1985. Se proyectó a 2018 con series de fecundidad, mortalidad y migración de estimaciones propias.

Escenario 2: Se tomó población base 2005²⁰. Ajuste de población menor de 10 años por diagrama de Lexis. Estimación de la migración para el periodo 2005-2018, actualizada con el efecto coyuntural de la migración de población proveniente desde Venezuela en el periodo 2005-2018. El componente de mortalidad se usaron las estimaciones propias.

Escenario 3: Se tomó población base 1985. Aplicando tasas de fecundidad ENDS, método hijos propios (Vehults, 2013), para el periodo 2015-2020 se tomó las estimadas por CELADE (2019), estimaciones propias de mortalidad y migración, proyectadas a 2018. Migración actualizada para el periodo 2005-2018.

²⁰ Este Censo fue evaluado con información del Censo 1985 corregido. Estimando una nueva población base para 2005 de 41.927.699 personas.

Escenario 4: Se tomó población base 1985. Aplicando tasas de estimaciones propias, Ajuste población menor de 10 años por diagrama de Lexis. Actualización de la migración desde 1985 con estimaciones propias actualizada con el efecto coyuntural de la migración de población proveniente desde Venezuela en el periodo 2005-2018. El componente de mortalidad se usaron las estimaciones propias.

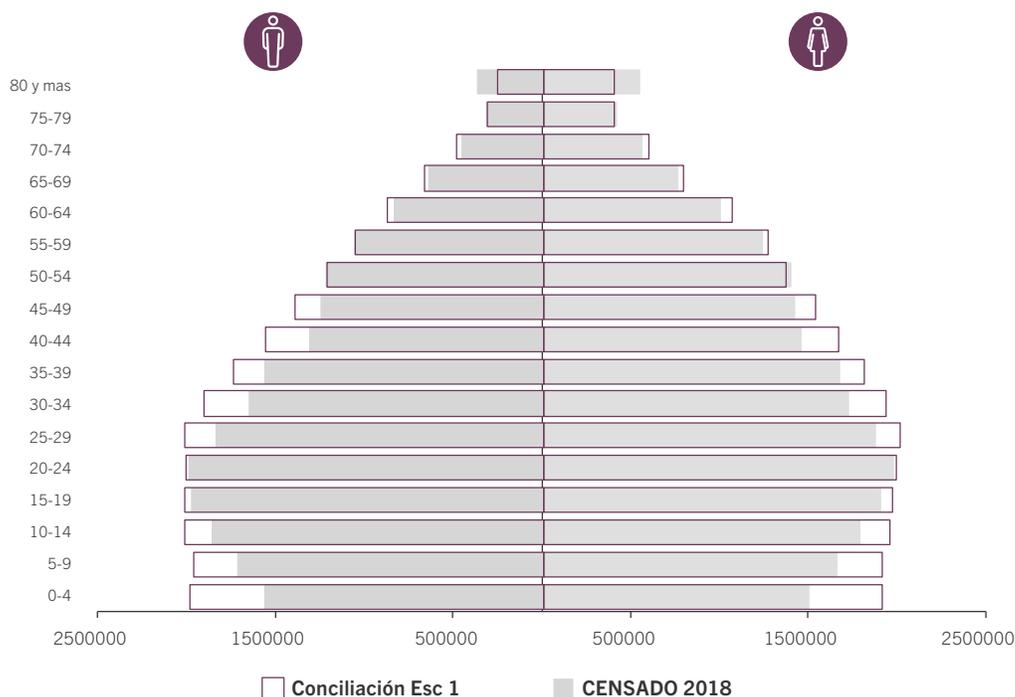
A partir de estos cuatro escenarios se compararon las estructuras por edad y sexo, manteniendo el principio de coherencia expresado Chackiel “se aplica la idea de un rompecabezas en el que todas las piezas tienen que encajar. En este caso deben encajar lo mejor posible, pues la verdad absoluta no se conoce. Lo importante es lograr la coherencia entre las poblaciones y los componentes demográficos involucrados” (2010).

2.4 | Resultados

Escenario 1: Si bien, la población resultante se ajusta de manera consistente para las edades superiores de 5 años, la base de la estructura piramidal presenta un ensanchamiento que no representa las tendencias de la población censada. En la cúspide de la pirámide se observa ajustes a la población entre 60 y 70 años, no obstante, la población de 80 y más se estima por debajo de lo censado.

Se observan ajustes importantes en el grupo de 40-45 producto de fecundidad del Censo 1973, usados para el ajuste del Censo de 1985. Adicionalmente en este escenario los grupos de edades de 15 a 25 años presentan bajo ajuste, cambiando la tendencia censada de un ensanchamiento de la estructura en estas edades poblacionales, como se muestra la figura 8:

Figura 8. | Evaluación de población 2018. Escenario 1.

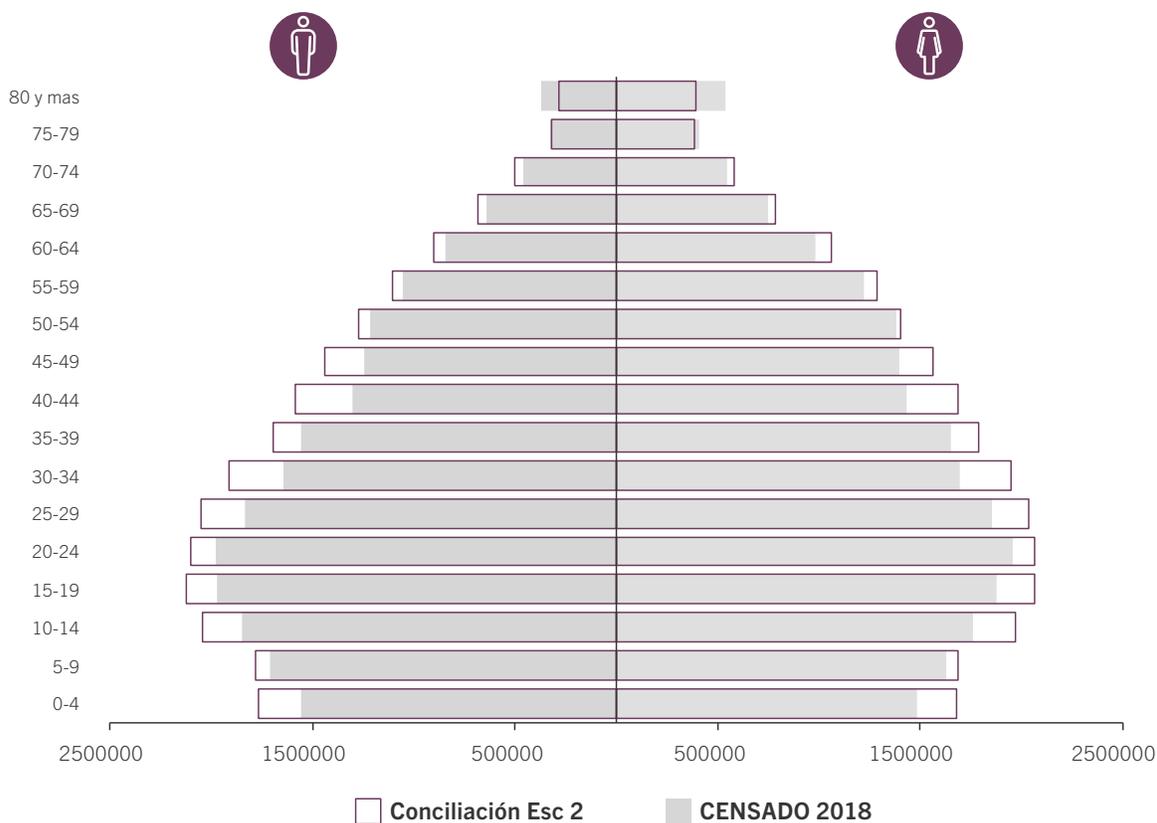


Fuente: DANE. Elaborado a partir de población base 1985 ajustada, componentes por estimaciones propias.

Escenario 2: En la figura 9, se muestra un mayor ajuste en los grupos poblacionales de 10 a 30 años, ajustados a partir de la conciliación realizada para el Censo de 2005, se observa

un base con menor predominancia de los niños de 0-10 años. No obstante, el grupo de 5-9 marca una diferencia importante con el grupo predecesor de 10-14 años.

Figura 9. | Evaluación de población 2018. Escenario 2.



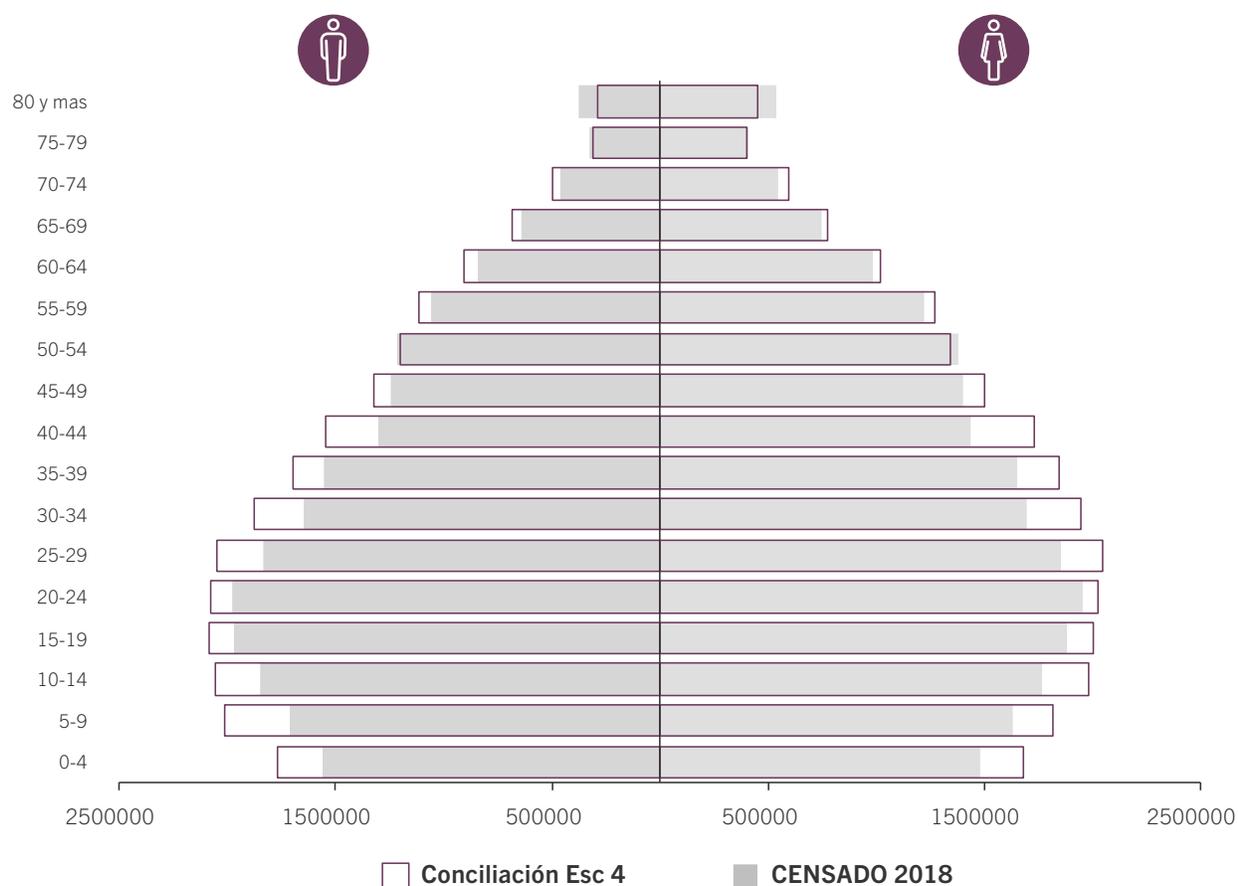
Fuente: DANE, Elaborado a partir de población base 2005 ajustada, población menor de 10 años diagrama de Lexis, Estimación de Migración Recaño 2019.

Escenario 3: Este escenario consideró fuentes externas de información para la fecundidad alternas a los Censos de Población, provenientes de la Encuesta Nacional de Demografía y Salud - ENDS estimada a través del método de hijos propios por el investigador Verhulst (2013). En este caso, la estimación se complementó con la estimación ENDS por el método historia de maternidad para el periodo 2010-2015 y por la estimación de CELADE (2019) para el periodo 2015-2020 que mantiene las tendencias observadas en la ENDS.

Esta serie de TGF, fecundidad responde adecuadamente a la estructura censal observada, corrigiendo principalmente la población de 0 a 50 años, con ajustes mínimos en las edades de adultos mayores.

En este escenario, se mantuvieron las estimaciones de mortalidad propias y la estimación de migración ajustadas con el fenómeno del éxodo de población desde Venezuela en los años 2016 a 2019.

Figura 10. | Evaluación de población 2018. Escenario 3.



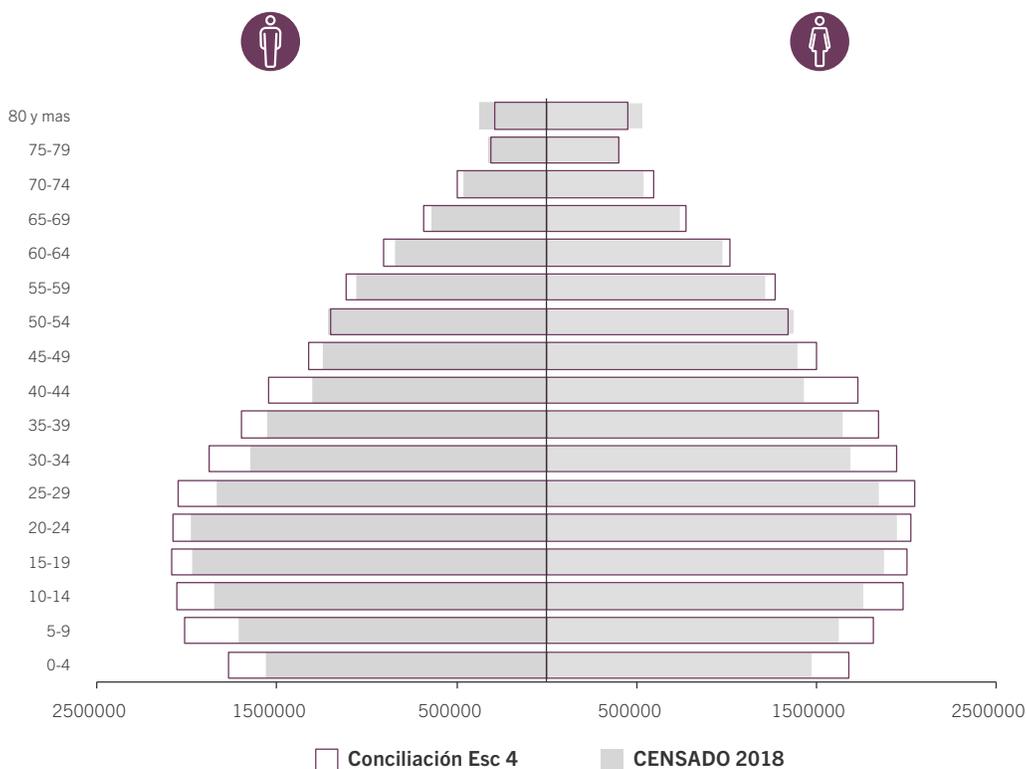
Fuente: DANE, Elaborado a partir de población base 1985 ajustada, Fecundidad ENDS Hijos propios, y CELADE 2015-2020

Escenario 4: Para el escenario 4 se contempló la estimación de migración realizada desde el año 1985 y la estimación de fecundidad propia, generando una estructura acorde a la población censada. Se empleó la estimación de menores de 5 años proveniente de diagrama de Lexis. Se observa una mejoría en los ajustes en todas las edades, no obstante, el ajuste de la población de 5-9 representa un cambio en la tendencia esperada para hombres.

Finalmente, una vez se realizaron las diferentes combinaciones se obtuvieron las siguientes

estimaciones de población que arrojaron un intervalo de omisión censal entre el 7,2% de la conciliación del escenario 1 y el 8,5% obtenido con el método directo. La información de la cobertura por conciliación censal se empleó como parámetro de referencia y comparativo de la evolución de los componentes, tomando como valor oficial de la estimación de la omisión censal el obtenido a partir del método directo Dual System correspondiente a 8.5% para el total nacional que corresponde al valor tope de la omisión calculada.

Figura 11. | Evaluación de población 2018. Escenario 4.



Fuente: DANE. Elaborado a partir de población base 1985 ajustada, fecundidad propia, población menor de 5 años ajustada por Diagrama de Lexis. Migración ajustada desde 1985

Tabla 16. | Proporción de defunciones en grupos de edad menores de 1 año.

Escenario	Población Total Estimada	Omisión Censal
Conciliación 1	47.587.826	7,20%
Conciliación 2	47.993.779	8,00%
Conciliación 3	48.041.250	8,10%
Conciliación 4	47.930.348	7,80%
Directo Dual System	48.258.494	8,50%

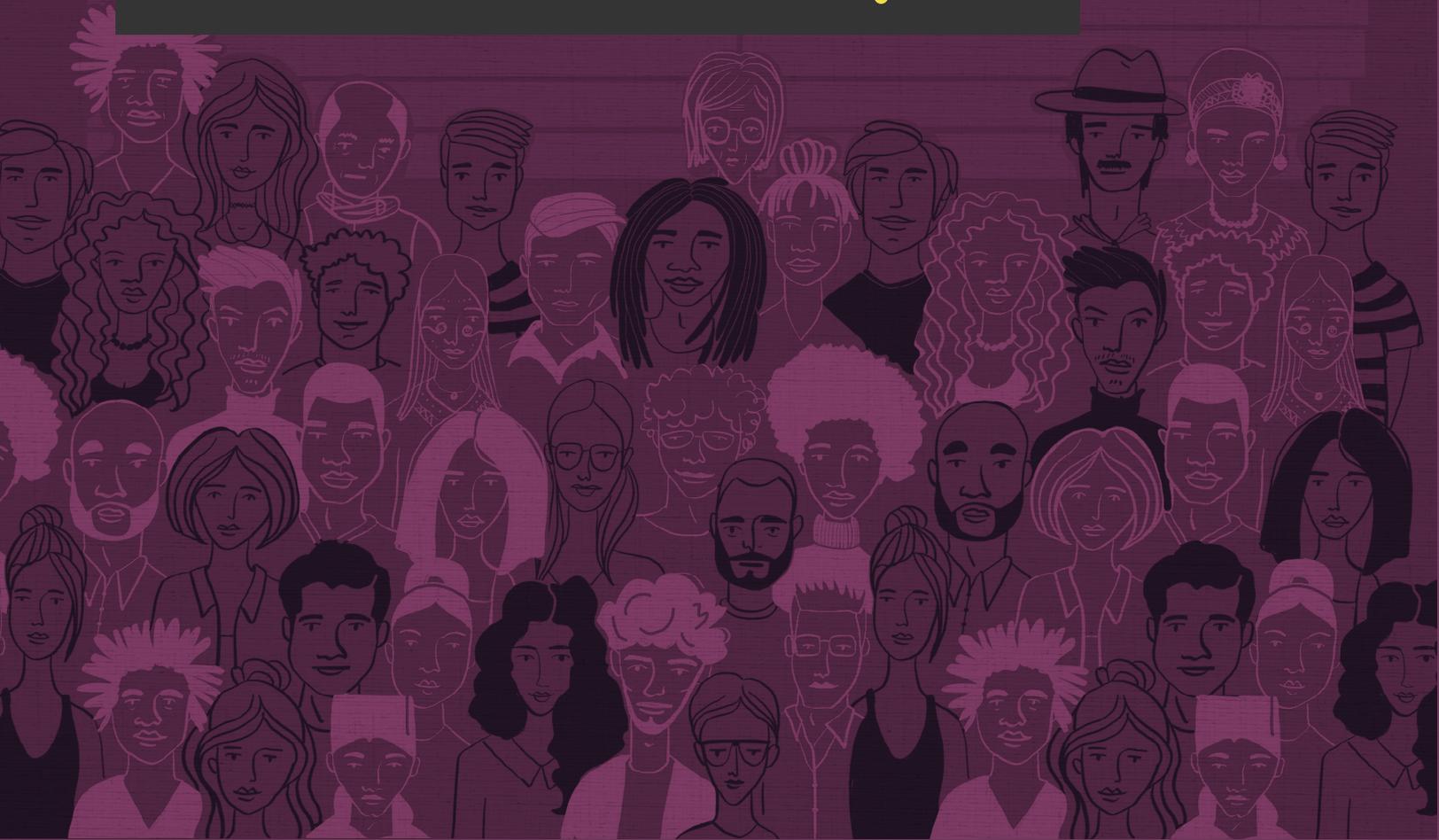
Fuente: DANE. Elaborado a partir de estimaciones de método directo e indirecto de omisión censal

Para La desagregación subnacional de la Omisión Censal se implementaron otras metodologías que permitieron niveles de desagregación geográficos, urbano – rural, departamental, municipal, que serán presentadas en el siguiente capítulo.

Las desagregaciones por sexo y edades simples se trabajaron a partir de metodologías matemáticas y estadísticas descritas en el documento metodológico de proyecciones y retroproyecciones de población.

3.

Evaluación cobertura a 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 censar personas dentro del territorio nacional, asociado a las siguientes condiciones de acceso a las unidades de observación²¹:

- *Limitaciones de acceso geográfico: Personas no censados en zonas no visitadas o incompletas (Z)*
- *Ausencia de todas las personas en las viviendas a la hora de la encuesta: Personas no censadas en las viviendas ocupadas con todas las personas ausentes. (A)*
- *Omisión de una o más personas dentro del hogar de manera aleatoria: Personas Subenumeradas al interior de los hogares (S)*

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:

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

Donde,

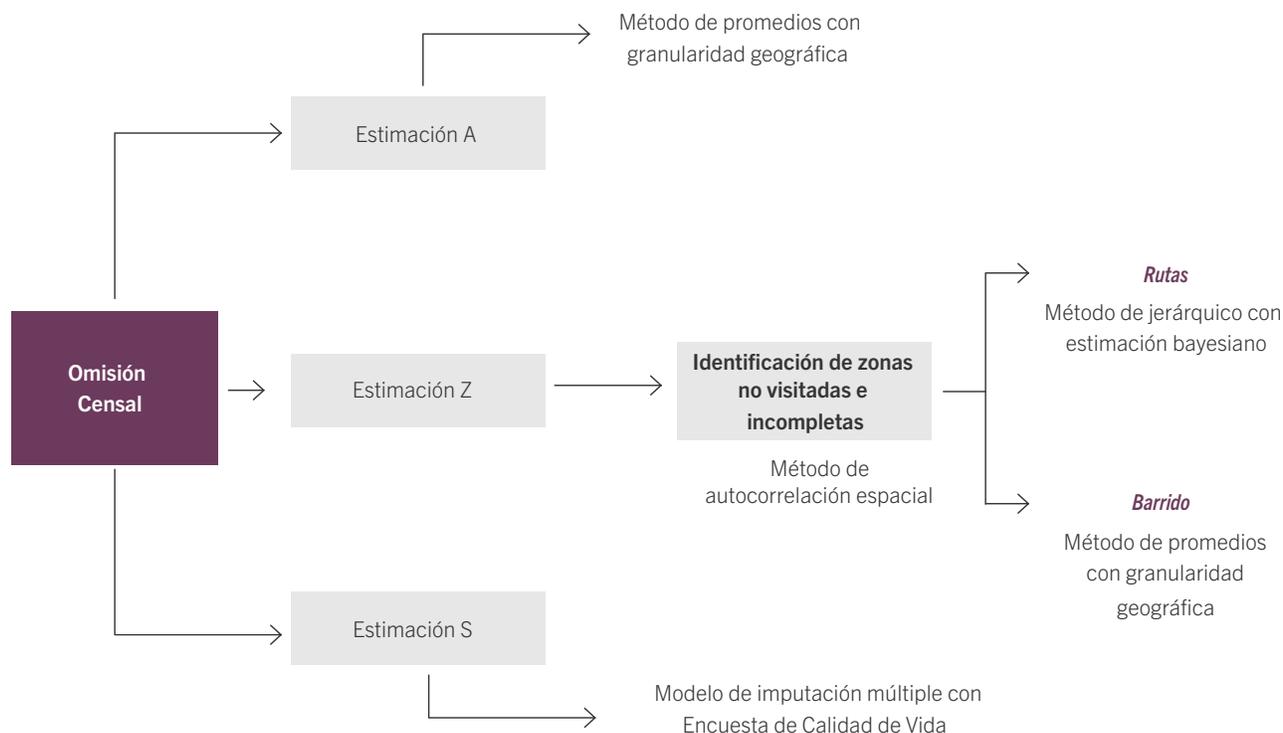
E_m = Población Estimada para el municipio

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 variable en los lugares donde no se conoce, como se muestra en el siguiente esquema:

²¹ Basado en los errores de cobertura señalados en el documento: Los datos geográficos: alcances, limitaciones y métodos de evaluación. (CELADE, 2019, pág. 35)

Figura 12. Diagrama general de aplicación metodológica para estimación.



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

Cada uno de los procesos metodológicos señalados en el anterior diagrama serán ampliados a continuación, iniciando por la descripción del contexto operativo en el cual se presentaron estos factores:

Causas y factores de omisión en el CNPV 2018

El operativo censal se encuentra diseñado para garantizar este acceso a todas las unidades de observaciones, no obstante, diferentes circunstancias limitan la consecución de este objetivo. Para el CNPV 2018 se establecieron estrategias en formatos y sistemas de seguimiento que permitieron conocer

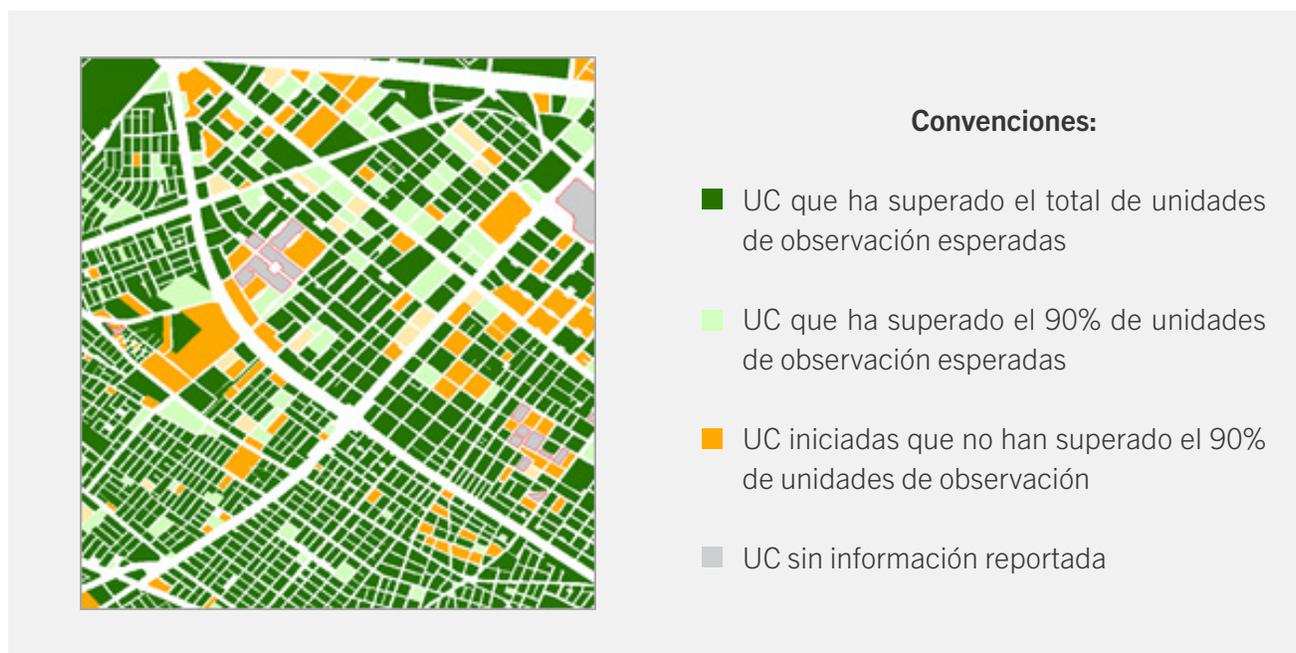
las unidades operativas que no fueron visitadas o quedaron incompletas, así como el registro de las viviendas que luego de agotar las tres visitas no fue posible encontrar a su moradores. Estos sistemas permiten caracterizar las siguientes causas y factores de omisión presentados en el CNPV.

1.1 | Omisión de áreas geográficas no visitadas e incompletas (Z):

Este tipo de omisión ocurre cuando el operativo de relevamiento no cubre un área geográfica determinada y las viviendas y personas que residen allí quedan excluidas del censo. (CELADE-UNFPA, 2014, pág. 35). Para el CNPV 2018, esta situación se presentó principalmente en zonas de difícil acceso por condiciones geográficas, seguridad o aspectos ambientales que dificultaron la llegada del operativo censal. En el total del país las unidades de cobertura no visitadas representan el 1,2% del total de unidades programadas.

Esta subcobertura se identificó en el operativo de campo a partir del seguimiento con un sistema de monitoreo geográfico del todo el país, en donde se identificaron en verde las unidades que superaron más del 90% de las unidades esperadas, naranja las que se encuentran en un rango del 1 al 90% de las unidades esperadas, y gris las unidades no visitadas, como se muestra en la siguiente figura.

Figura 13. | Seguimiento de cobertura censal- extracto urbano



Fuente: DANE, Dirección de Geoestadística CNPV 2018

Como resultado final del operativo censal, se clasificaron, a partir del sistema de monitoreo y control las zonas identificadas como no visitadas.

Estas se presentaron con mayor predominancia en la zona rural y centros poblados, en donde

alcanzaron el 6,4% y 6,1% del total de unidades esperadas. Las zonas incompletas se presentaron en menor porcentaje, predominando en los centros poblados.

Tabla 17. | Porcentaje de zonas incompletas y no visitadas según tipo de residencia.

Área	Zonas	Participación
Cabecera	Incompletas	0,50%
	No Visitadas	2,10%
Centros Poblados	Incompletas	0,90%
	No Visitadas	6,10%
Rural	Incompletas	0,60%
	No Visitadas	6,40%

Fuente: DANE, Dirección de Geoestadística CNPV 2018 – Sistema de seguimiento geográfico

1.2 | Omisión en viviendas ocupadas con todas las personas ausentes (A).

Este tipo omisión se presentó cuando en los recorridos operativos el censista identificó que una vivienda se encontraba ocupada con uso residencial, sin embargo, en ninguna de las tres visitas realizadas encontró sus moradores. En esta situación el censista, luego de agotar las tres visitas a esta vivienda, procedía a marcar en el cuestionario censal que esta vivienda se encontraba ocupada con todas personas ausentes²², por parte del informante, según sea el caso.

Para las viviendas ocupadas con todas las personas ausentes se diligenció el módulo de ubicación del cuestionario censal, lo que permite contar con la

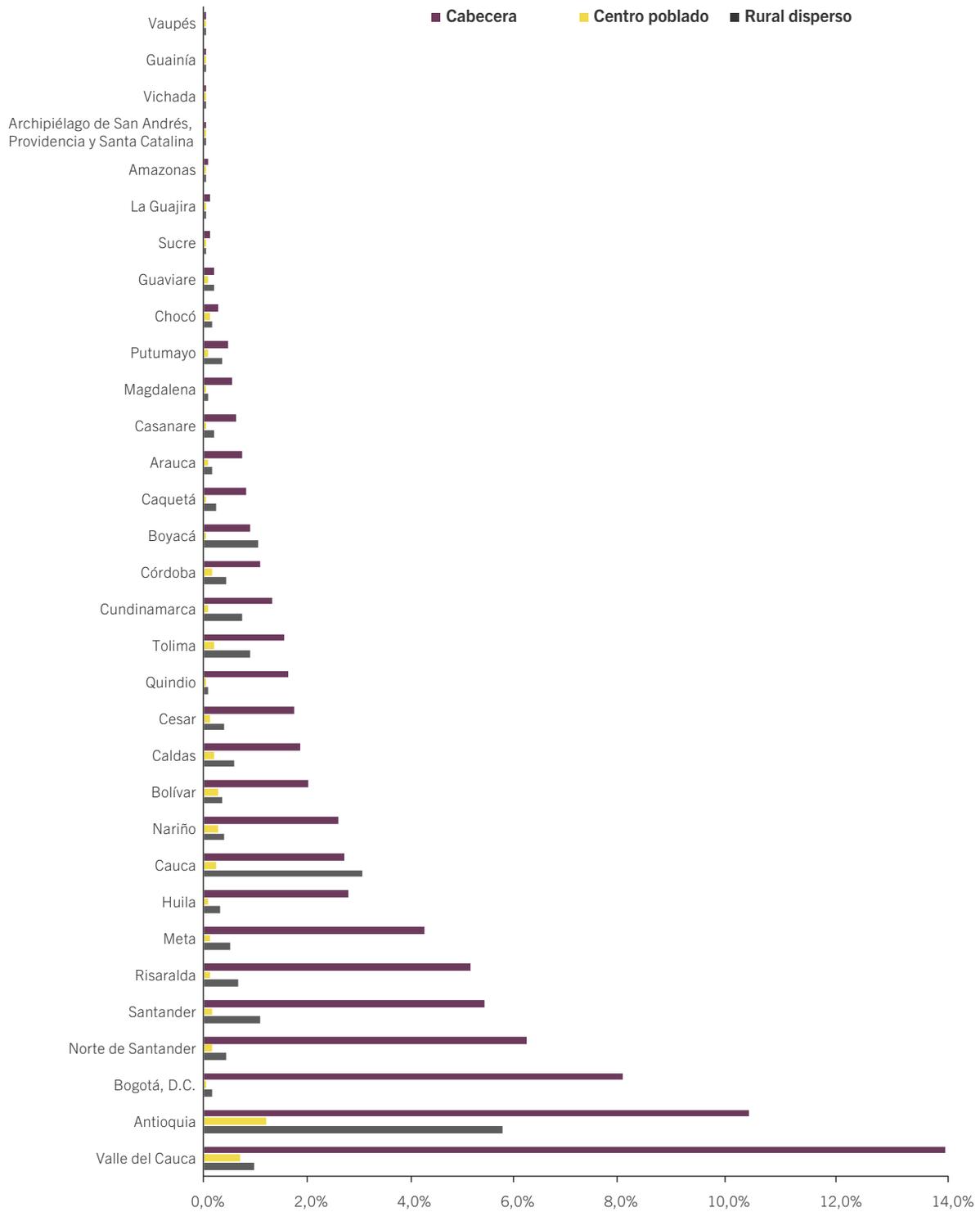
georreferenciación de estas unidades a nivel de departamento, municipio, clase, coordinación, área y unidad de cobertura²³.

En el Censo Nacional de Población y Vivienda 2018 - CNPV 2018 fueron contadas 16'070.893 viviendas de uso residencial y mixto. De dichas viviendas, 439.017 se identificaron como ocupadas, sin embargo, ninguna persona atendió al censista, por tal motivo se encuentran categorizadas dentro de la base de datos, como Viviendas Ocupadas con las Personas Ausentes - VOCPA. Como se muestra en la distribución departamental:

²² Esta información se encuentra marcada en la pregunta 14, opción

²³ Estructura del marco censal. Ver definiciones a final de documento.

Figura 14. Distribución departamental de las Viviendas Ocupadas con Personas Ausentes - VOCPA

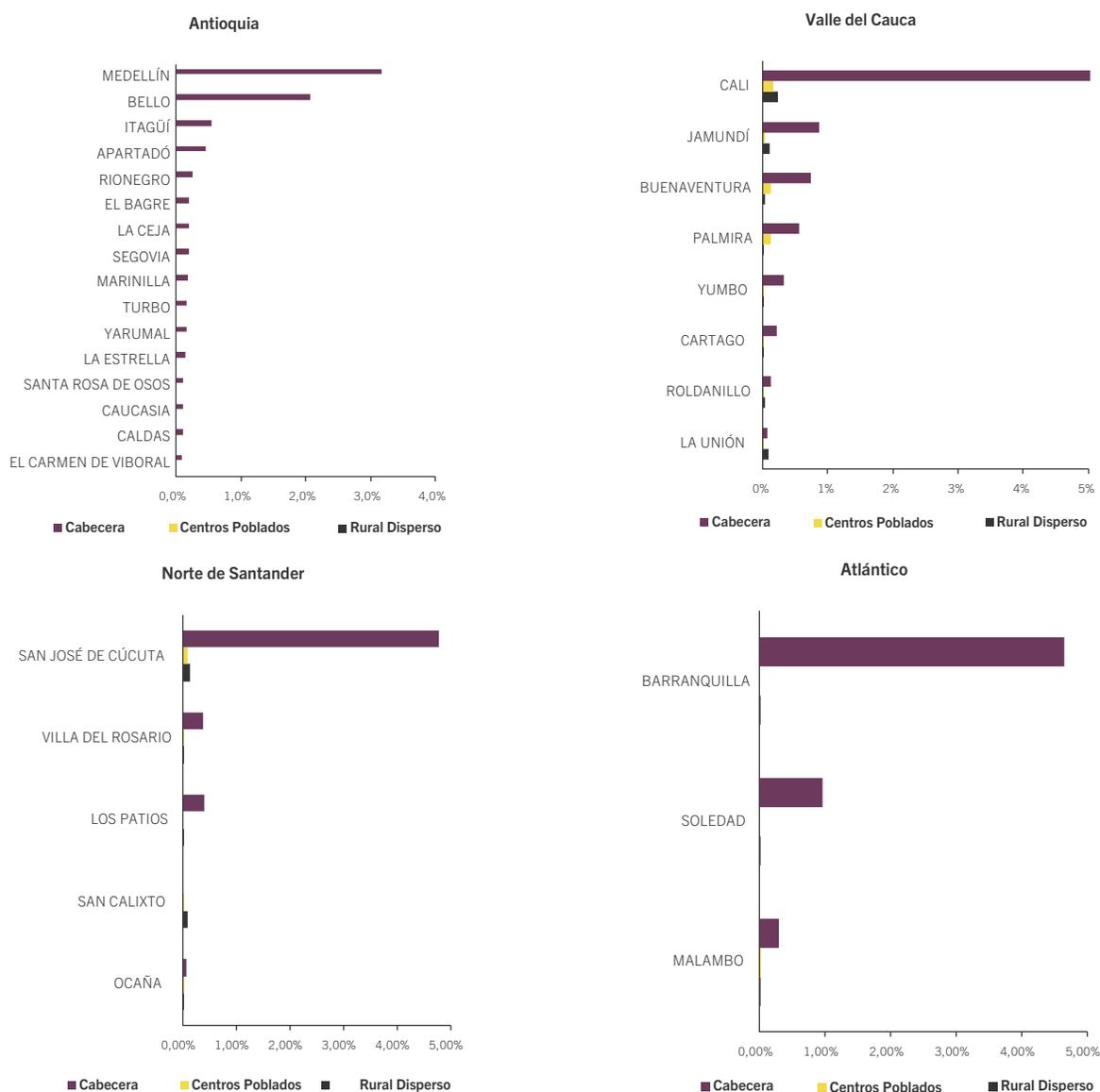


Fuente: DANE, Dirección de censos y demografía CNPV 2018

Como se observa en la figura 13, la distribución viviendas categorizadas como VOCPA, se concentran en sectores urbanos de los departamentos, sin embargo, en el departamento de Antioquia y Cauca, las VOCPA también se encuentran concentradas en rural disperso.

En cuanto al comportamiento al interior de los departamentos de mayor participación de VOCPA, en la figura 15, se observa que Medellín, Cali, Cúcuta y Barranquilla acumula en su área urbana la mayor cantidad de VOCPA, además de los municipios de Bello, Itagüí y Los Patios; los dos primeros en el área metropolitana de Medellín y Los Patios para el área metropolitana de Cúcuta.

Figura 15. Distribución I de las viviendas ocupadas con personas ausentes -VOCPA en los departamentos



Fuente: DANE, Dirección de censos y demografía CNPV 2018

1.3 | Omisión por subenumeración de personas en los hogares (S)

Este tipo de omisión se presenta cuando el censista no registra dentro del cuestionario censal a una o más personas del hogar, en algunos casos porque los miembros del hogar no lo reportaron o por que el censista no identificó a todos los ocupantes de la vivienda. Esta situación ha sido documentada en América Latina, en donde se presenta una tendencia a omitir los niños menores, lo que introduce sesgos sistemáticos que producen faltantes en población de ciertas edades (CELADE-UNFPA, 2014, pág. 36). Estos grupos Subenumerados suelen estar

representados en la omisión de población de 0-5 años, omisión sistemática de población adulta joven (Rincón, Conciliación Censal y determinación de la población base, 2008). La determinación de esta omisión se debe realizar a partir de métodos indirectos o variables colaterales que identifiquen el tamaño de hogar por municipio o la identificación de población omitida.

Como resumen de los factores de omisión se presenta el siguiente esquema:

Figura 16. | Esquema resumen de los factores de omisión CNPV 2018



Fuente: DANE, Dirección de censos y demografía CNPV 2018

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

Omisión de áreas geográficas no visitadas e incompletas (Z)

Como se presentó en la figura 12, la estimación de las áreas no visitadas e incompletas 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%.

Como se mencionó anteriormente, el sistema de monitoreo y control geográfico permitió identificar las zonas no visitadas e incompletas a partir de un indicador por cada unidad de cobertura²⁴, con lo cual fue posible clasificar las siguientes situaciones:

Tabla 18. | 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 realizado el análisis exploratorio de la cobertura geográfica, se determinó la necesidad de establecer adecuadamente las zonas en donde se tenía $a > b$

y b corresponde unidades residenciales, es decir se requirió un proceso de determinación de área incompletas sujetas de estimación con el fin de no sobreestimar zonas con otro tipo de usos.

²⁴ Mínima unidad de planeamiento en el operativo censal, para cabeceras y centros poblados corresponde a manzanas y para zonas rurales a predios

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:

2.1 | 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 fundamenta el desarrollo conceptual de la propuesta para la selección de áreas incompletas sujetas de estimación de población.

Como afirma Siabato (2019) “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”. En este contexto a partir de la aplicación de estas técnicas de correlación espacial se identificarán clústeres* de áreas incompletas censadas en

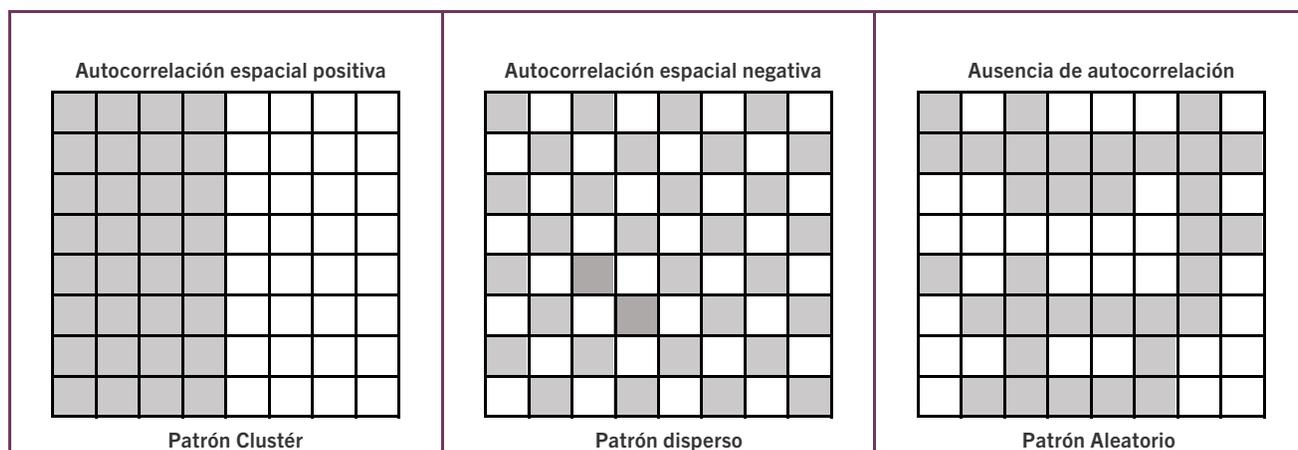
donde se aplicaron indicadores locales y globales de asociación espacial. Este tipo de análisis es recomendado para “explorar y mostrar de manera clara la distribución de las variables, los valores atípicos, la asociación espacial, agrupamientos espaciales, así como los puntos calientes/fríos y analizar la influencia en las relaciones de contigüidad espacial” (Aguilar & Santana, 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).*

* Agrupación geográfica del fenómeno

Figura 17. | 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.

a. 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:

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

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

b. 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.

Existen varios programas de propósito general y específicos para el cálculo de esta matriz, habiendo seleccionado el software GeoDa® para todos los procedimientos de estadística espacial, dado su nivel de especialización, alta eficiencia y manejo intuitivo.

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).

c. 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

$w_{i,j}$: 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:

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

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

En donde,

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

x, y : valor promedio de la variable

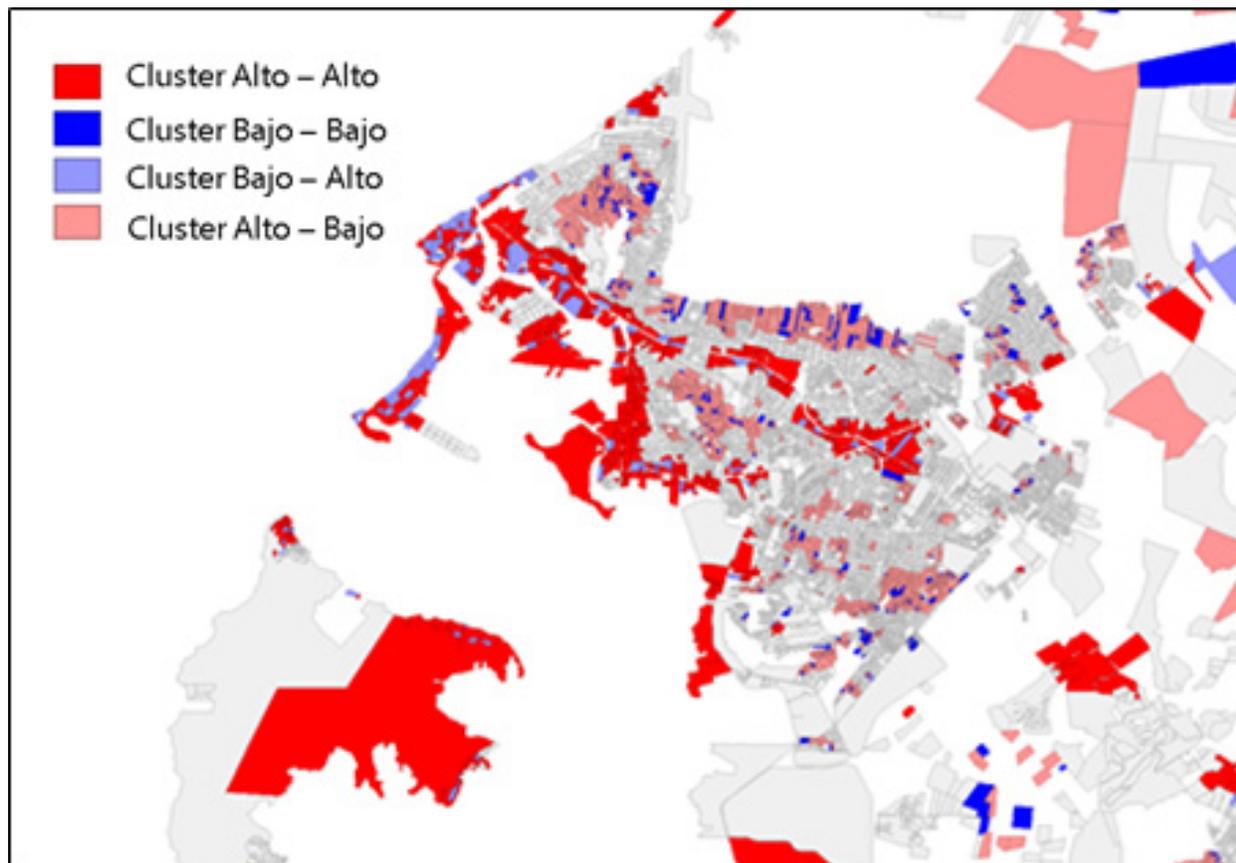
$w_{i,j}$: peso espacial entre i , j , dado por la matriz de pesos

N : número total de observaciones

Tomado de (Siabato et al, 2019):

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 cómo se muestra en el siguiente ejemplo:

Ilustración 1. | 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

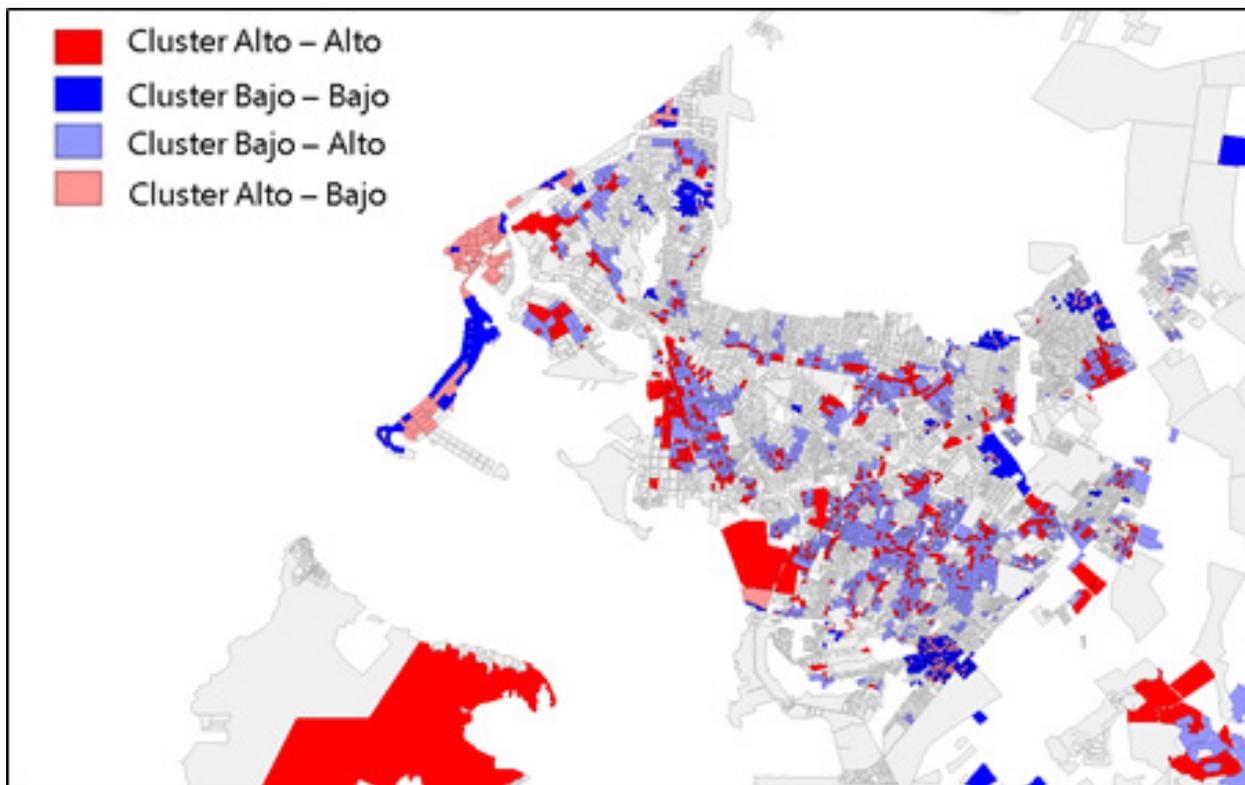
representan bajos porcentajes de cobertura rodeados de otros usos, o altos porcentajes de cobertura rodeados de viviendas.

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.

GeoDa® ofrece la opción de guardar estas categorías de clusters o outliers como valores de 0 a 4 para tomar el valor de interés (en este caso 2 correspondiente a Low-Low)

A continuación, se muestran los resultados para la relación P_NR_OBT/P_U_NORM (la inversa a la anterior: valores de porcentajes bajos de otros usos rodeados de valores bajos de cobertura).

Ilustración 2. | Mapa de clusters y outliers de Cartagena, relacionando P_NR_OBT/P_U_NORM



Fuente: DANE, Dirección de Geoestadística

Es evidente que la distribución de clusters es distinta tan solo al intercambiar las variables: se aprecia un mayor número de zonas Low-Low y de outliers Low-High hacia el sector del Cerro de la Popa y hacia el sur de la ciudad.

d. Operaciones espaciales posteriores.

Los mapas anteriores muestran que hay sectores que comparten dos características esenciales: porcentaje

bajo de cobertura (clusters fríos o low – low) y alto porcentaje de viviendas (clusters calientes o high-high). Estos son los sectores y unidades censales que se desea identificar, de manera que se garantice la imputación sobre áreas de predominancia residencial, con cobertura incompleta. Para separar plenamente las áreas que tienen estas características en común, es necesario realizar una intersección de áreas. La siguiente tabla ilustra los resultados de las intersecciones propuestas:

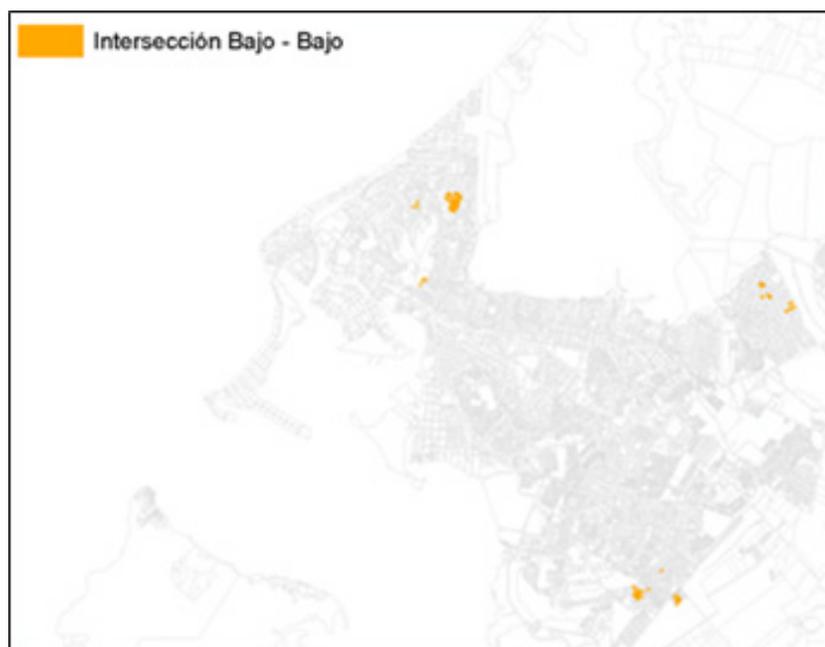
Tabla 19. | Ejemplo de intersección de áreas en el municipio de Cartagena.

CLUSTER 1	CLUSTER 2	UC EN INTERSECCIÓN ²⁵
P_U_NORM/P_NR_OBT (Bivariado): BAJO - BAJO	P_NR_OBT/P_U_NORM (Bivariado); BAJO - BAJO	535

Fuente: DANE, Dirección de Geoestadística

Los resultados se muestran en las siguientes imágenes:

Ilustración 3. | Mapa del resultado de la intersección Bivariado – Bivariado, Cartagena y Bolívar



Fuente: DANE, Dirección de Geoestadística

²⁵ Resultado para Cabecera, centros poblados y rural

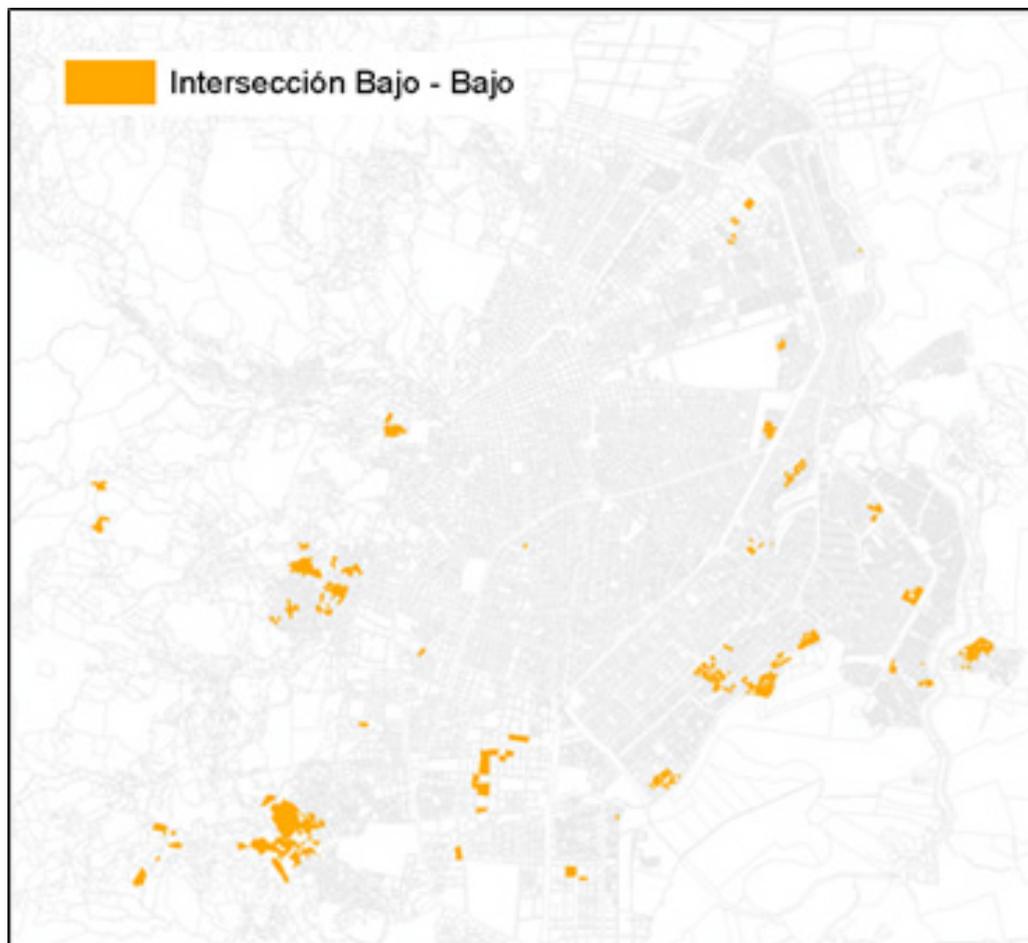
La intersección de ambas relaciones bivariadas evidencia unas zonas específicas de la ciudad de Cartagena y del departamento de Bolívar en donde simultáneamente se presentan bajos porcentajes de cobertura censal y altos porcentajes de viviendas (o bajos porcentajes de otros usos de edificaciones).

Si se realiza el mismo ejercicio con los clusters univariados P_U_NORM (porcentaje de cobertura comparado contra sí mismo en las vecindades) y P_NR_OBT (porcentaje de otros usos), se obtiene exactamente el mismo resultado.

El resultado en Cartagena muestra que la intersección se ubica en los sectores del Cerro de la Popa (Norte) y Arroz Barato (Sur), los cuales fueron mencionados en los informes de campo como aquellos donde no se realizó una cobertura completa.

Al repetir el ejercicio para una ciudad como Cali, se obtienen unidades censales ubicadas en zonas identificadas como de difícil acceso por motivos de orden público o de restricciones por el tipo de vivienda (invasiones, condominios) o donde se realizó una operación de barrido rápido sin garantizar un cubrimiento del 100% de edificaciones.

Ilustración 4. | Mapa del resultado de la intersección Bivariado – Bivariado, Cali (Valle del Cauca)



Fuente: DANE, Dirección de Geoestadística

2.2 | Omisión en zonas no visitadas e incompletas de rutas

a. 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 (3° 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). La Tabla 20 lista los municipios y departamentos donde se usó esta estrategia. 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. Ambas cifras fueron insumos del Modelo Lineal Bayesiano Jerárquico que se utilizó para estimar la población en las 394 rutas.

El Modelo Lineal Bayesiano Jerárquico se escogió para estimar la población en áreas rurales de los municipios de la Tabla 20, debido a que históricamente estas áreas se han destacado por su déficit de información. Estos municipios 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 las luces nocturnas, etc., se puede deducir la presencia o ausencia de personas en áreas remotas. A nivel estadístico las ventajas del enfoque Bayesiano sobre el frecuentista son: primero, la decisión de rechazar o no una hipótesis no depende del tamaño de la muestra. Segundo, el modelo del análisis Bayesiano exige tener en cuenta el conocimiento previo.

Tabla 20 . | Departamentos y municipios donde se utilizó la estrategia de rutas

DEPARTAMENTO	MUNICIPIO	DEPARTAMENTO	MUNICIPIO
VAUPÉS	MITÚ	GUAINÍA	INÍRIDA
	CARURU		BARRANCO MINAS ¹
	PACOA1		MAPIRIPANA ¹
	TARAIRA		SAN FELIPE ¹
	PAPUNAU ¹		PUERTO COLOMBIA ¹
	YAVARATE ¹		LA GUADALUPE ¹
PUTUMAYO	PUERTO ASIS	CAQUETÁ	CACAHUAL ¹
	PUERTO GUZMAN		CAMPO ALEGRE ¹
	PUERTO LEGUIZAMO		MORICHAL NUEVO ¹
	PUERTO CAICEDO		CARTAGENA DEL CHAIRA
GUAVIARE	SAN JOSE DEL GUAVIARE	META	SAN VICENTE DEL CAGUAN
	CALAMAR		SOLANO
	EL RETORNO		SOLITA
	MIRAFLORES		PUERTO RICO
	LETICIA		CURILLO
AMAZONAS	EL ENCANTO ¹	VICHADA	LA MACARENA
	LA CHORRERA ¹		PUERTO GAITÁN
	LA PEDRERA ¹		MAPIRIPÁN
	LA VICTORIA ¹	ANTIOQUIA	PUERTO CARREÑO
	MIRITI - PARANA ¹		LA PRIMAVERA
	PUERTO ALEGRIA ¹		CUMARIBO
	PUERTO ARICA ¹		MURINDÓ
	PUERTO NARIÑO	VIGIA DEL FUERTE	
	PUERTO SANTANDER ¹		
	TARAPACA1		

DEPARTAMENTO	MUNICIPIO
CHOCÓ	QUIBDO
	ACANDI
	ALTO BAUDO
	ATRATO
	BAGADO
	BAHIA SOLANO
	BAJO BAUDO
	BOJAYA
	CARMEN DEL DARIEN
	CERTEGUI
	CONDOTO
	EL CANTON DE SAN PABLO
	EL CARMEN DE ATRATO
	EL LITORAL DEL SAN JUAN
	ISTMINA
	JURADO
	LORO
	MEDIO ATRATO
	MEDIO BAUDO
	MEDIO SAN JUAN
	NOVITA
	NUQUI
	RIO IRO
	RIO QUITO
	RIOSUCIO
	SAN JOSE DEL PALMAR
	SIPI
	TADO
UNGUIA	
UNION PANAMERICANA	

DEPARTAMENTO	MUNICIPIO
VALLE DEL CAUCA	BUENAVENTURA
	DAGUA ²
CAUCA	BOLÍVAR ²
	LÓPEZ
	GUAPI
	TIMBIQUI
	EL TAMBO ²
	SANTA BARBARA
	EL CHARCO
NARIÑO	LA TOLA
	OLAYA HERRERA
	MOSQUERA
	FRANCISCO PIZARRO
	ROBERTO PAYAN
	MAGUI
	BARBACOAS
	TUMACO
	RICAURTE
	CUMBITARA ²
NORTE DE SANTANDER	POLICARPA ²
	EL ROSARIO ²
	EL CARMEN ²
	CONVENCIÓN ²
	EL TARRA ²
	TEORAMA ²
	TIBÚ ²
TOLEDO ²	
CHITAGÁ ²	

DEPARTAMENTO	MUNICIPIO
LA GUAJIRA	URIBIA
	ALBANIA ²
	MAICAO
	RIOHACHA ²
	MANAURE
	DIBULLA ²
	SAN JUAN DEL CESAR ²
CESAR	VALLEDUPAR ²
	PUEBLO BELLO ²
MAGDALENA	SANTA MARTA ²
	CIÉNAGA ²
	ARACATACA ²
	FUNDACIÓN ²
	EL CARMEN ²
NORTE DE SANTANDER	CONVENCIÓN ²
	EL TARRA ²
	TEORAMA ²
	TIBÚ ²
	TOLEDO ²
	CHITAGÁ ²
	BOYACÁ
	CUBARÁ ²
ARAUCA	TAME ²
	FORTUL ²
	SARAVENA ²
CASANARE	SÁCAMA ²
	TÁMARA ²
CÓRDOBA	TIERRALTA ²
	PUERTO LIBERTADOR ²

DEPARTAMENTO	MUNICIPIO
RISARALDA	PUEBLO RICO ²
	MISTRATO ²
CALDAS	RIOSUCIO ²
CASANARE	HATO COROZAL ²
	PAZ DE ARIPORO ²
	LA JAGUA DE IBIRICO ²
CESAR	BECERRIL ²
	AGUSTÍN CODAZZI ²
	LA PAZ ²
	MANAURE BALCÓN DEL CESAR ²
	DABEIBA ²
ANTIOQUIA	FRONTINO ²
	URRAO ²
	ITUANGO ²
	PUERTO CONCORDIA ²
META	PUERTO RICO ²
	PUERTO LLERAS ²
	VISTA HERMOSA ²
	LA URIBE ²
	MESETAS ²

Fuente: DANE- 2016 Notas: 1. Corregimientos departamentales o Áreas No Municipalizadas (Decreto 2274 de 1991). 2. Regiones o territorios con municipios que incluyen además de Rutas (en las rutas duraban 28 días o más en campo), áreas de Recolección Especial (ruta en la que duraban menos de 28 días en campo), o una parte de su resto municipal utiliza el método de barrido. Sin embargo, las cabeceras municipales en estos territorios tuvieron su propio operativo de barrido.

b. Antecedentes del método

Los modelos jerárquicos bayesianos son una herramienta que ha surgido como alternativa a los modelos clásicos de regresión y que han sido ampliamente utilizados por su flexibilidad para abordar diferentes problemas; además, su estructura jerárquica permite ir desde el componente más simple y sencillo hasta modelos más complejos donde cada componente se junta en cada nivel, y así, se van creando interacciones las cuales permiten, al final, resolver problemas más grandes y complejos. Una de las ventajas de los modelos bayesianos es que podemos encontrar la función de distribución a posteriori de los parámetros, la cual nos permite obtener intervalos de credibilidad y hacer inferencia.

Para el caso del análisis demográfico existe suficiente evidencia del uso de los modelos jerárquicos bayesianos, por ejemplo: En Raftery et al. (2012,2013) Las Naciones Unidas utilizan modelación bayesiana para producir proyecciones de población probabilísticas y de esperanzas de vida. En Alkema et al. (2011) se describe un modelo de proyección bayesiana para producir Tasas de Fecundidad Total - TFR para los países considerados en el estudio. En (Sharrow et al. 2013) se presenta un enfoque bayesiano para encontrar la función de distribución de los parámetros del modelo de Heligman-Pollard (Heligman-Pollard. 1980) propuesto para medir mortalidad por alguna enfermedad.

c. 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: primero, conocimiento inicial de la dependencia entre parámetros y, segundo, 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:

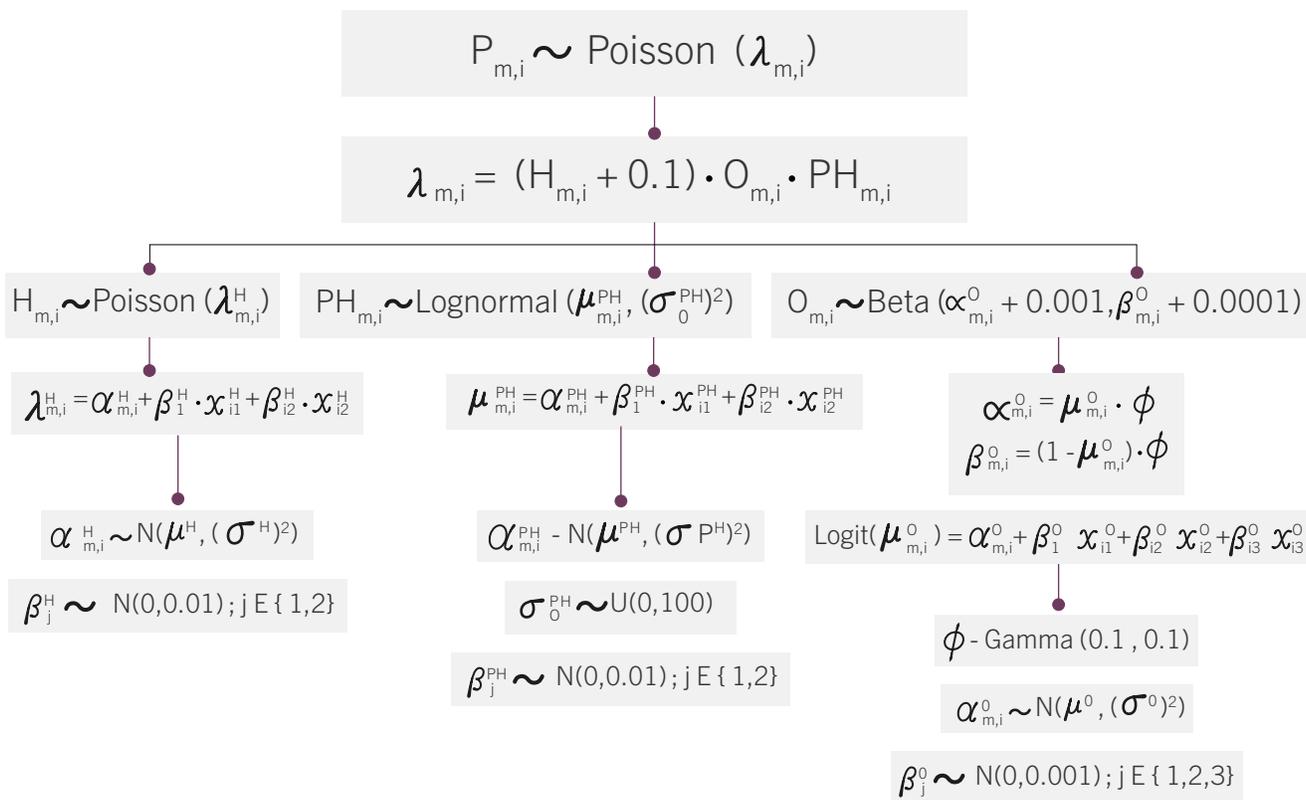
$$P_{m,i} \sim \text{Poisson} (\lambda_{m,i})$$

$$\lambda_{m,i} = (H_{m,i} + 0.1) O_{m,i} PH_{m,i}$$

Donde $P_{m,i}$ es el número de personas en la unidad de cobertura i que se encuentra en el municipio m . $\lambda_{m,i}$ es el valor esperado de personas en esta unidad, el cual es la multiplicación de: $H_{m,i}$, que es el número de viviendas; $O_{m,i}$, parámetro que especifica el porcentaje de viviendas ocupadas, y $PH_{m,i}$, cuyo valor es igual al número de personas por vivienda ocupada.

La estructura del modelo se presenta en la figura 18:

Figura 18. Estructura del modelo



P: Número de personas en la unidad de cobertura i

H: Número viviendas en i en el municipio m

PH: Personas por vivienda

O: Proporción de viviendas ocupadas

x_{i1}^H : Densidad de construcciones

x_{i2}^H : Viviendas esperadas (talleres de cartografía social)

x_{i1}^{PH} : Densidad de escuelas

x_{i2}^{PH} : Proporción de hogares con NBI

x_{i1}^0 : Elevación (m)

x_{i2}^0 : Distancia a centros poblados

x_{i3}^0 : Noches nocturnas

Fuente: DANE. Dirección de Censos y Demografía

d. Construcción de covariados espaciales

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

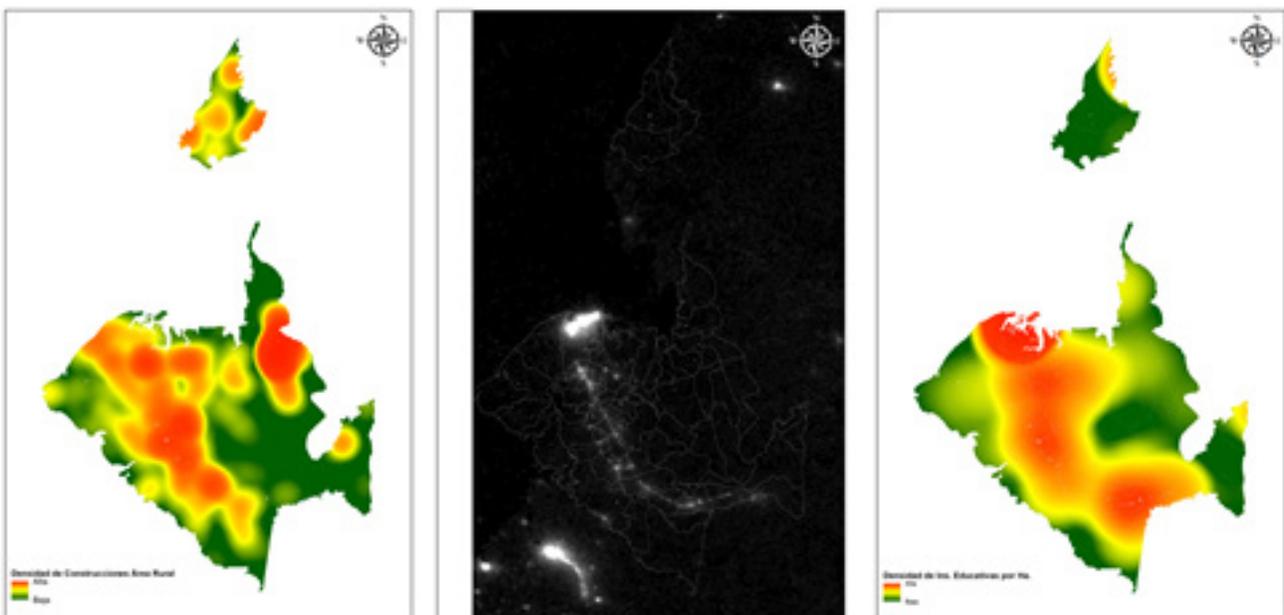
- *Distancia a centro poblados:* Este covariado se procesó a partir de la capa de vías nacionales escala 1:100.000, y los polígonos de cabecera, a través de la función de distancia euclidiana.

- *Densidad de 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 de 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.
- *Proporciones de hogares por NBI:* 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 Engine.

Figura 19. | Covariados espaciales



Fuente: DANE, Elaboración a partir de procesamiento espacial

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.

e. Resultados

En total se estimó una población no censada de **224.966** habitantes en las zonas de rutas

que no fueron vistas 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 de la zona más cercana al océano pacífico de este departamento colombiano, seguido del departamento de Chocó en donde se presentaron condiciones de dificultades de acceso geográfico.

Tabla 21 . | Total, población estimada por departamento en las zonas de rutas de mayor complejidad operativa

CÓDIGO	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

2.3 | Omisión en zonas no visitadas e incompletas de barrido

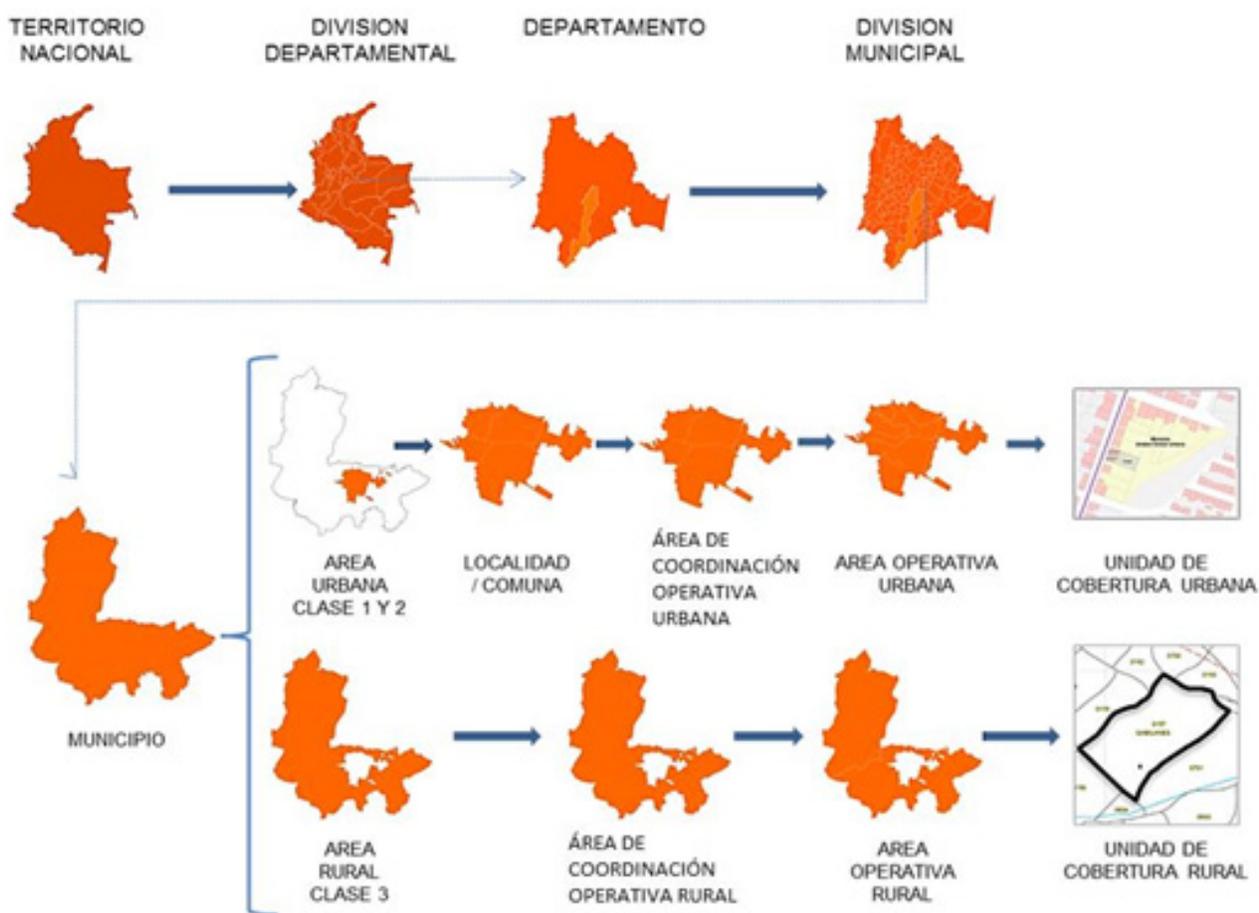
a. Contexto de áreas de barrido

En el diseño operativo del CNPV se definieron 5 métodos para la recolección de información: Electrónico, Focalizado, Ruta, Mixto y Barrido, este último consisten en recorrer las unidades de cobertura de manera sistemática y ordenada censado edificación a edificación. Este método se diseñó para el 86,6% de las unidades y se enfocó en 1.099 cabeceras municipales y 975 áreas

rurales de municipios, las restantes 23 cabeceras y 91 municipios en su área rural se diseñaron por el método de ruta, las restantes 56 áreas rurales municipales se aplicó el método mixto, para un total de 1122 municipios recorridos en el Operativo Censal.

Para el Operativo Censal se contó con un marco que permitió recorrer, una a una las unidades de observación, presentado a continuación:

Figura 20 | Marco Censal operativo CNPV



Fuente: DANE. Tomado de documento de Marco Censal- DIG.

El marco censal contiene el atributo de las Edificaciones Esperadas - EE en cada unidad de cobertura, denominador del indicador de unidades terminadas y avanzadas.

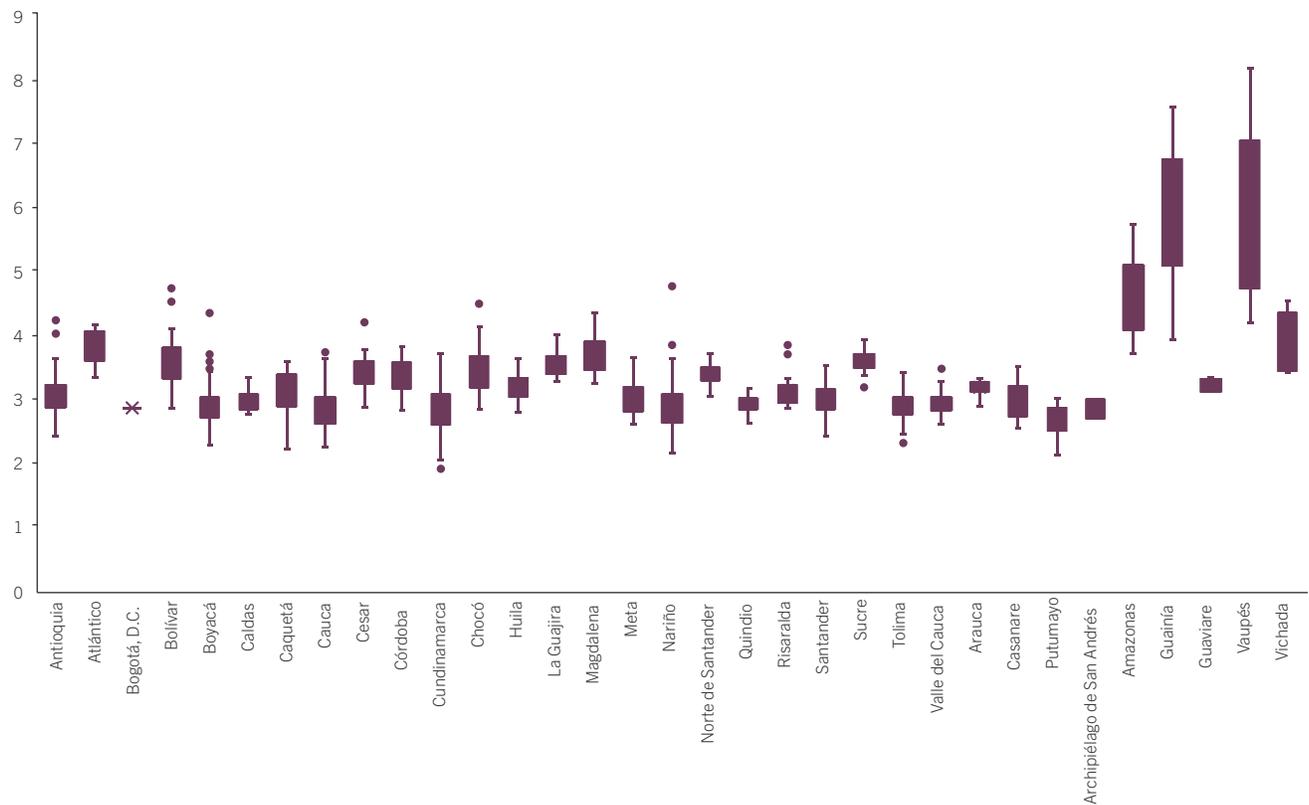
b. Planeamiento del metodológico

La estimación de población en las zonas no vistas 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 qué 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 encuentran 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.

Figura 21 | Tamaño promedio de los hogares por departamento obtenido de la población efectivamente censada



Fuente: DANE, Dirección de censos y demografía CNPV 2018

En consecuencia, se plantea un modelo que identifica: las proporciones de Viviendas por Clase de Uso Residencial - PER, el número 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_x Edificaciones esperadas en la unidad de cobertura X

PER proporción 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 las viviendas 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:

1. De las 19'900.593 edificaciones que se encuentran en el CNPV2018, se obtuvo la proporción de PER y PVO, por municipio y clase.
2. Se estimó el -PPPH, desagregado por municipio y clase, considerando solo edificaciones de uso residencial y mixto que se encuentre ocupadas.

3. Se seleccionó por criterio bivariado-bivariado²⁶ de las Unidades en las Edificaciones Incompletas y No Visitadas - NENV.

4. 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.

5. Para rural disperso se estimó el producto de PER, PPPH y NENV desagregado por municipio y clase.

6. Estimación de la población omitida PO por cada una de las unidades de cobertura seleccionadas.

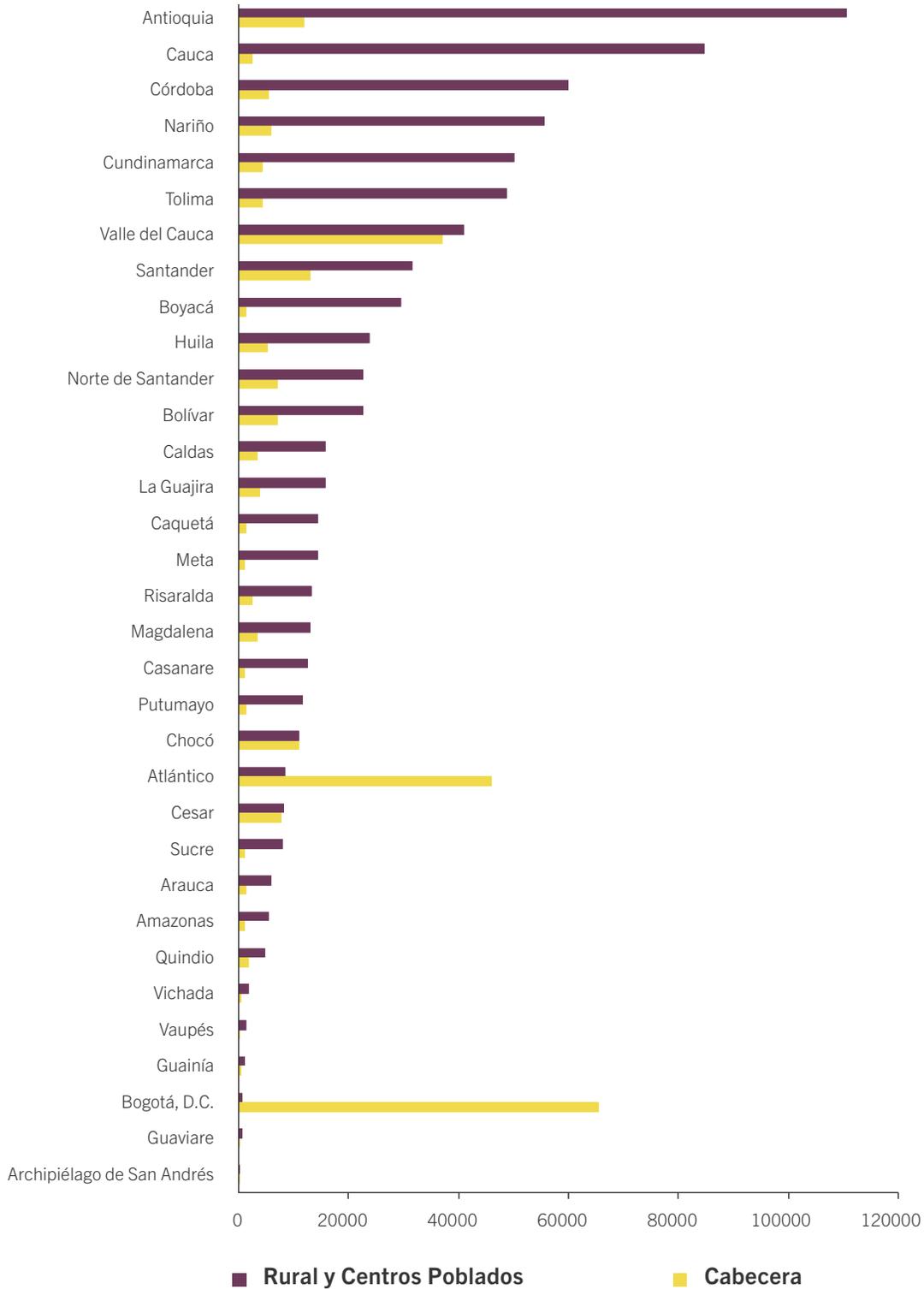
El resultado del municipio considera la agregación de personas omitidas por municipio y por clase, en las zonas en la que se implementó el método de recolección de barrido.

c. Resultados

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 26 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.

²⁶ Ver sección 3.1

Figura 22 | Distribución de la estimación de personas omitidas en el método de barrido en Z por departamento y clase

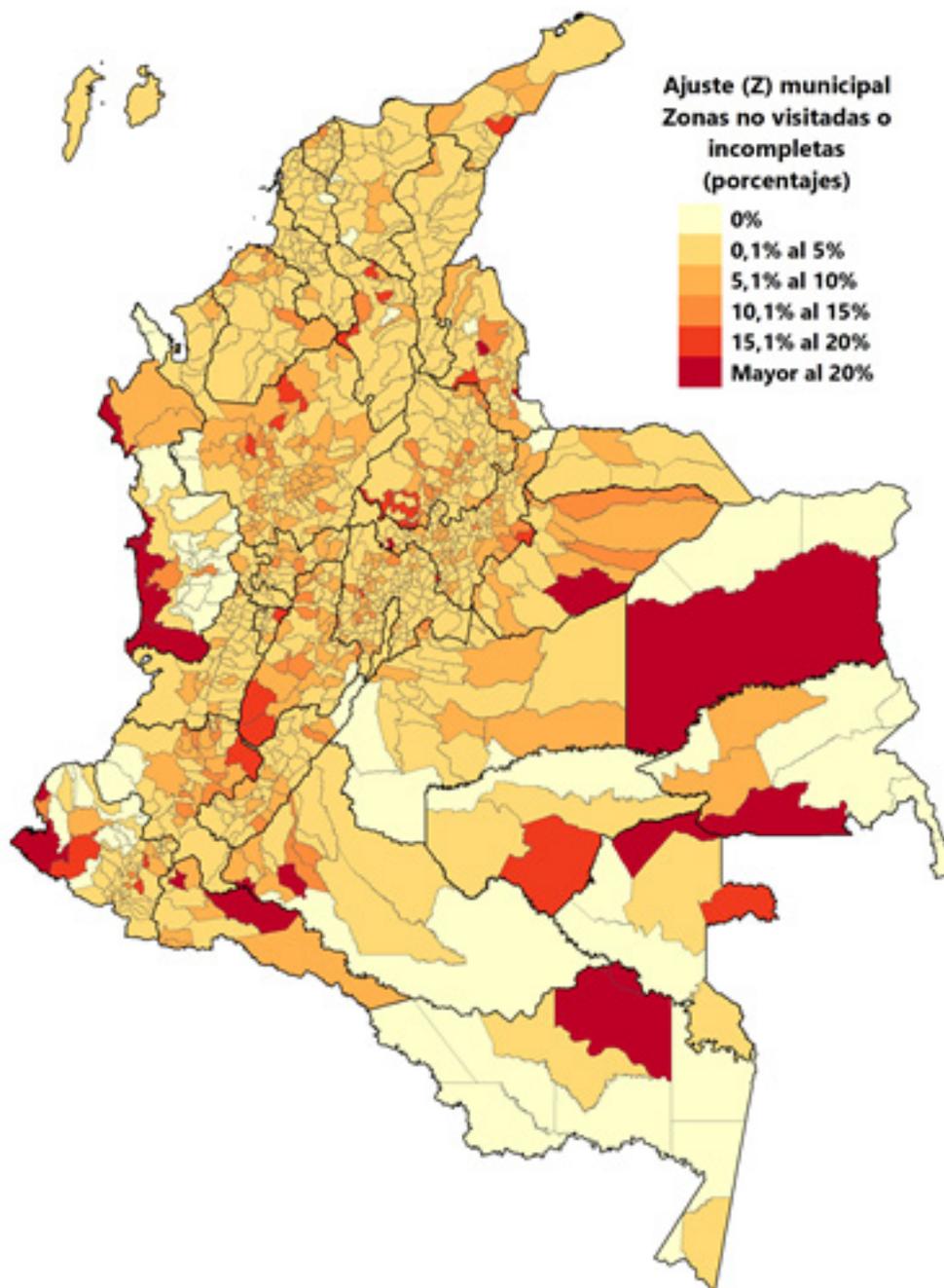


Fuente: DANE, Dirección de censos y demografía CNPV 2018

Finalmente, como consolidación de la omisión de población en zonas no visitadas e incompletas (Z) se debe agregar la omisión del método de barrido

y del método de rutas, obteniendo una población omitida **1.235.287** personas en el territorio nacional como se muestra en la figura 27.

Figura 23 | Distribución de la estimación de la población omitida en zonas no visitadas e incompletas.



Fuente: DANE, Dirección de censos y demografía CNPV 2018

La distribución territorial de la omisión en las zonas de mayor complejidad de acceso se relaciona con las periferias del país, con una marcada omisión en el pacífico colombiano, y

municipios del sur del departamento de Nariño, así como los municipios de la Orinoquia y Amazonia, y alta y media Guajira que marcaron complejidades de acceso.

Omisión en viviendas ocupadas con personas ausentes (A)

La no realización de la entrevista en 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:

3.1 | 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 operativa²⁷. (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:

²⁷ 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”

$$PO_{Ax} = \sum VOCPA_{X} * PPPH_{Bx}$$

Donde,

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

ΣVOCPAX 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

3.2 | Resultados

Se toman las 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.

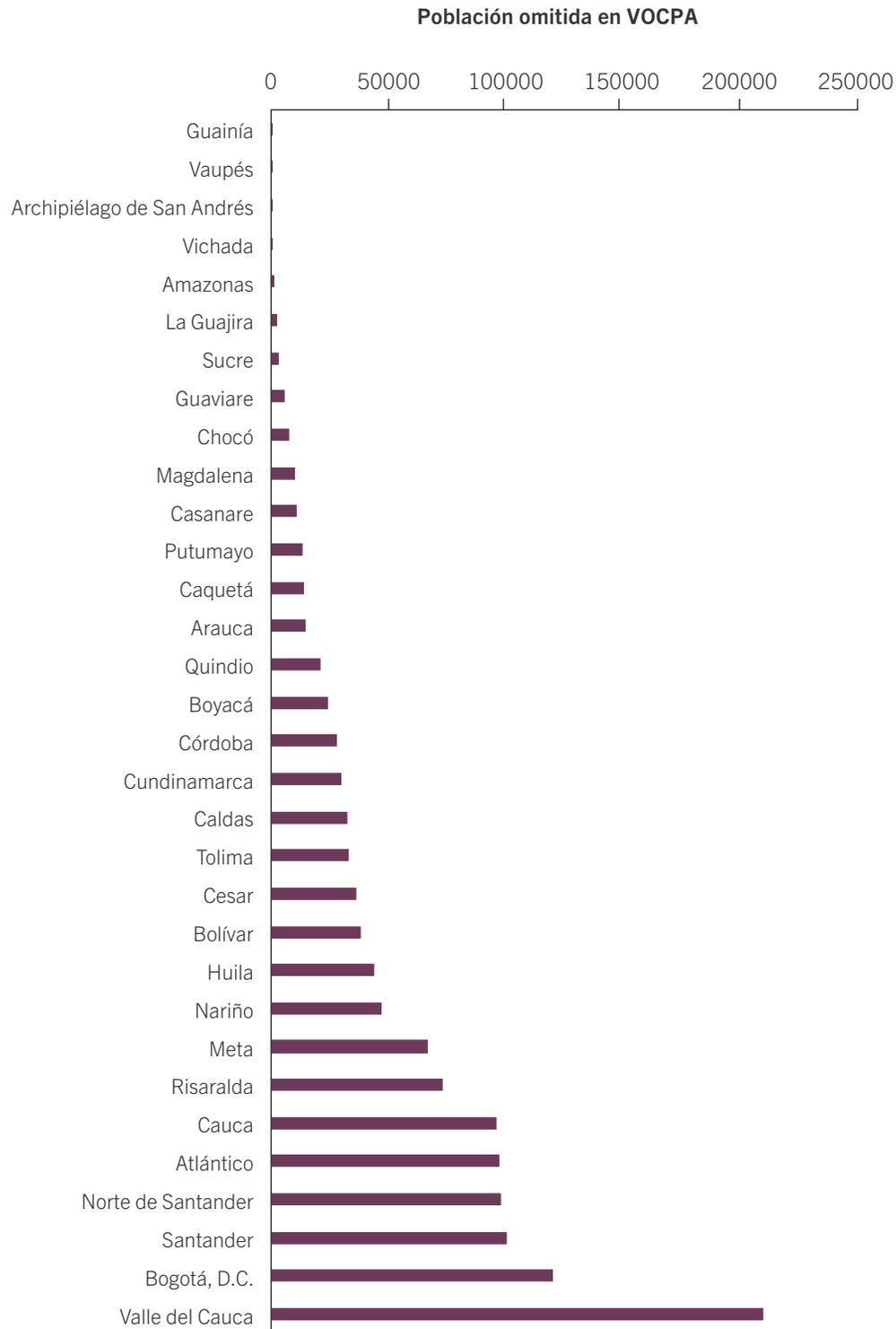
Como resultado se obtiene un total de **1'520.541** personas en viviendas VOCPA. La tabla 22 muestra la distribución de esta población por clase, resaltando el territorio urbano como el mayor receptor de población. La distribución departamental de la población obtenida se muestra en la Gráfica 28, la cual resalta que los departamentos con más VOCPA (Antioquia, Valle del Cauca, Bogotá y Atlántico), reciben las mayores cantidades de personas imputadas.

Tabla 22 . | Estimación VOCPA desagregación por clase

CLASE	Estimación	OMISIÓN Z EN RUTAS
Urbano	1.172.377	15
Centros Poblados	68.244	6.010
Rural disperso	279.920	1.467
Total	1.520.541	3.090

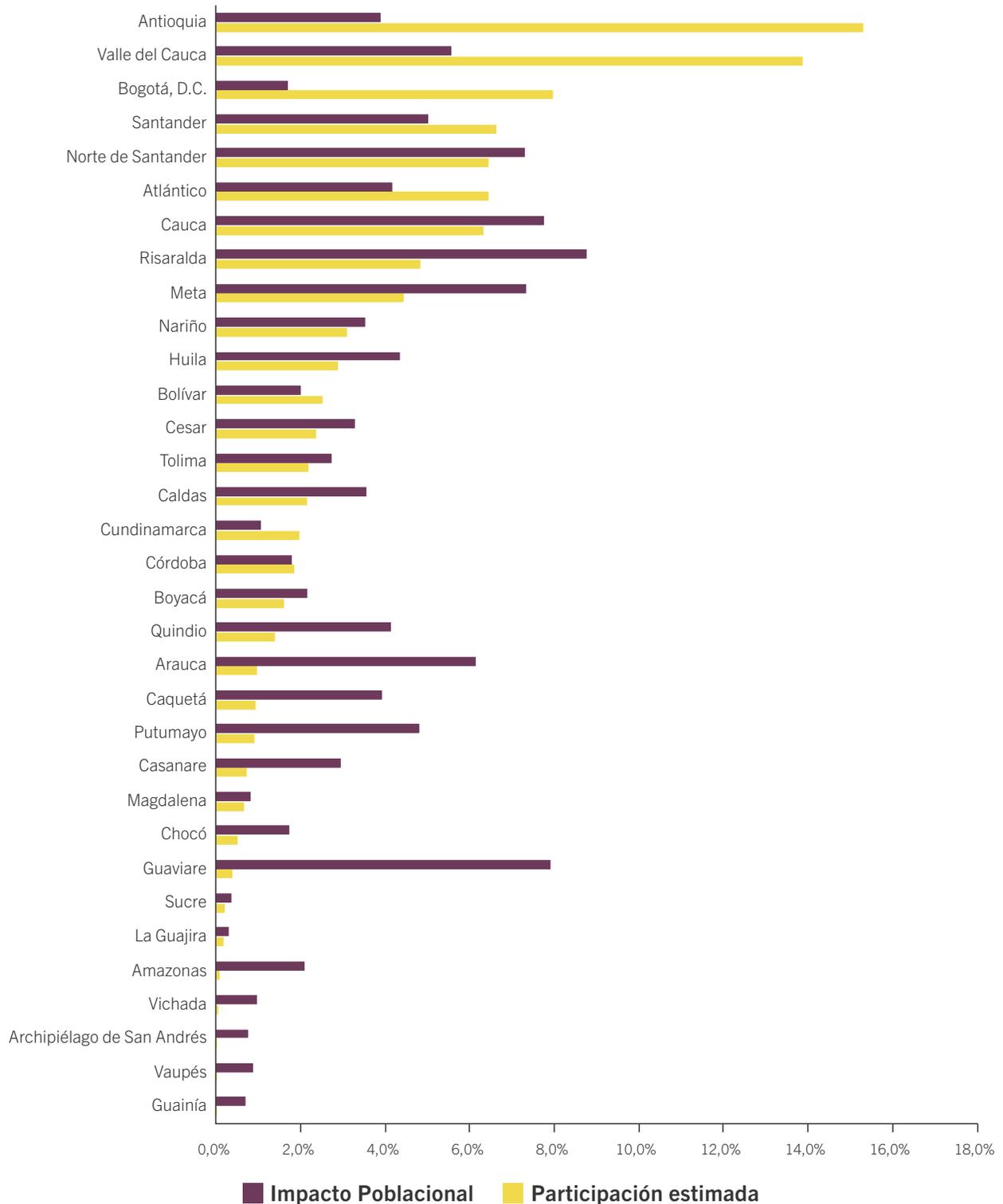
Fuente: DANE, Dirección de censos y demografía CNPV 2018

Figura 24. Distribución departamental personas estimadas en VOCPA



Fuente: DANE, Dirección de censos y demografía CNPV 2018

Figura 25. Distribución porcentual departamental personas estimadas en VOCPA y el impacto departamental

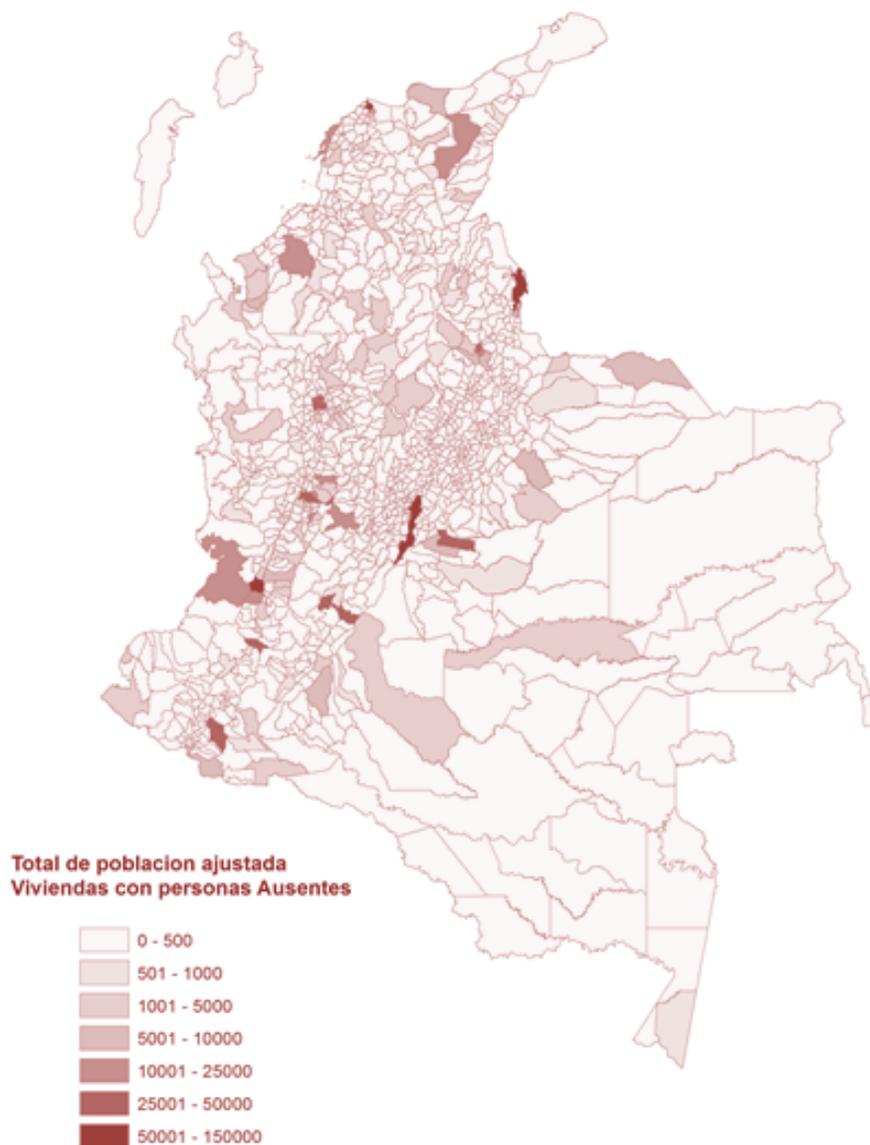


Fuente: DANE, Dirección de censos y demografía CNPV 2018

La Figura 29 muestra los departamentos con mayores cambios relativos en su población efectivamente censada. Es decir, el cambio que presentará el conteo del departamento producido de la metodología propuesta. Los departamentos con mayores cambios relativos son Guaviare, Risaralda, Cauca, Norte de Santander, Meta y Arauca. En cuanto a la distribución territorial de la omisión A,

las concentraciones de población ajustada en este componente muestra una predominancia en las principales ciudades y capitales de departamento (figura 30), en consonancia con la hipótesis de que estos territorios 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.

Figura 26. | Distribución territorial de personas estimadas en VOCPA y el impacto



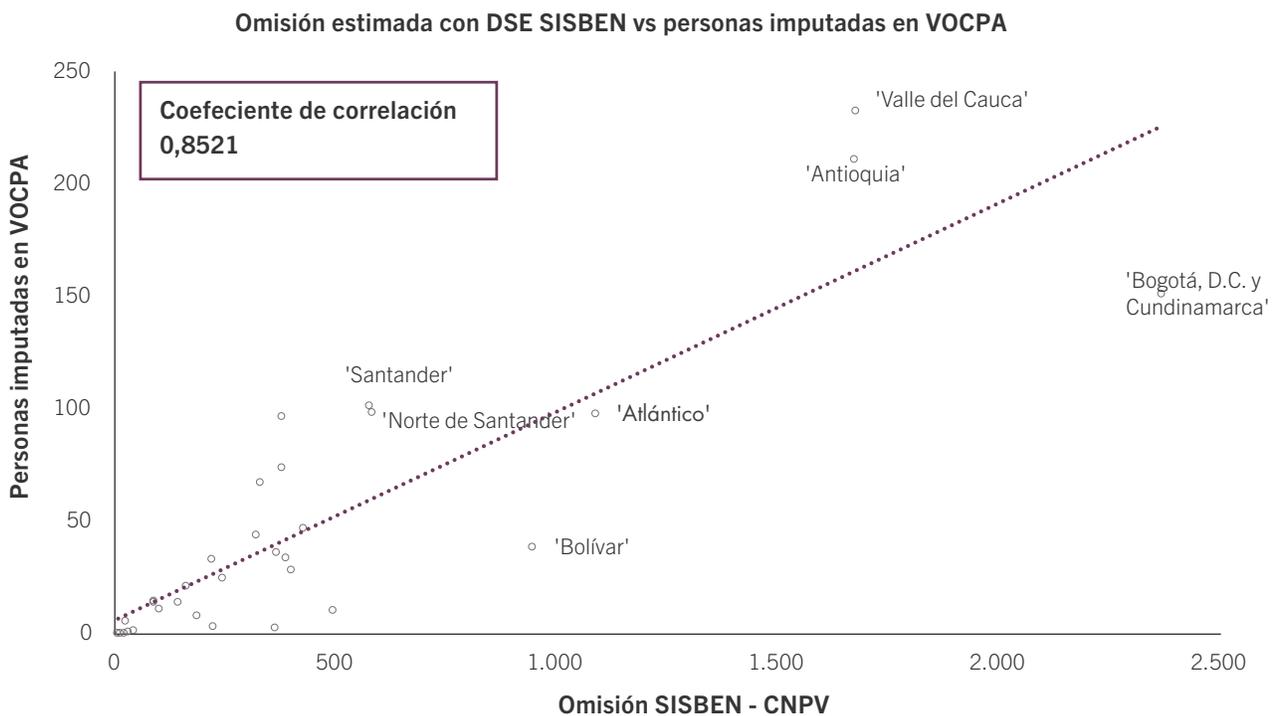
Fuente: DANE, Dirección de censos y demografía CNPV 2018

Es importante considerar, que el proceso de imputación de personas en viviendas VOCPA es un ejercicio posterior al cálculo de la omisión total del censo. Es decir, los procesos de imputación no deben superar el 8.5% de omisión en el CNPV 2018, estimada con la Encuesta de Calidad de Vida - ECV y la metodología dual de captura-recaptura - DSE. Dicha estimación de la omisión es confiable dado que la ECV cumple de forma temática todos los supuestos necesarios, para ser la encuesta de pos-enumeración y su posterior implementación (CEPAL- UNFPA, 2014) Sin embargo, aunque las estimaciones no serían confiables al usar un registro administrativo como encuesta de pos-enumeración, debido a la violación de supuestos, un registro administrativo de la calidad, como el Sistema de Selección de Beneficiarios Para Programas Sociales - SISBEN o la Base de Datos Única de Afiliados - BDU, evidenciarían la tendencia de la omisión

distribuida de forma departamental. Es por eso que se aplicó la metodología dual usando SISBEN y BDU, desagregada por departamento; considerando a Bogotá y Cundinamarca como un solo territorio.

La base del SISBEN a corte diciembre de 2018 cuenta con 37'935.355 de registros, de los cuales 28'594.397 registros fueron emparejados de forma determinística con el CNPV 2018. La base BDU está compuesta por 44'897,948 registros a corte diciembre de 2018 (con estado activo y suspendido) y 2'054,243 del régimen de excepción con corte octubre de 2018, para un total de 46'952.191. Sin embargo, del régimen de excepción 1'150,319 no cuentan ubicación, por lo tanto, para la aplicación de la metodología dual, la base BDU cuenta con 45'801.782 registros, de los cuales 35'447.435 fueron emparejados con la base CNPV 2018.

Figura 27. | Dispersión Omisión estimada con SISBEN vs personas imputadas en VOCPA, valores en miles.

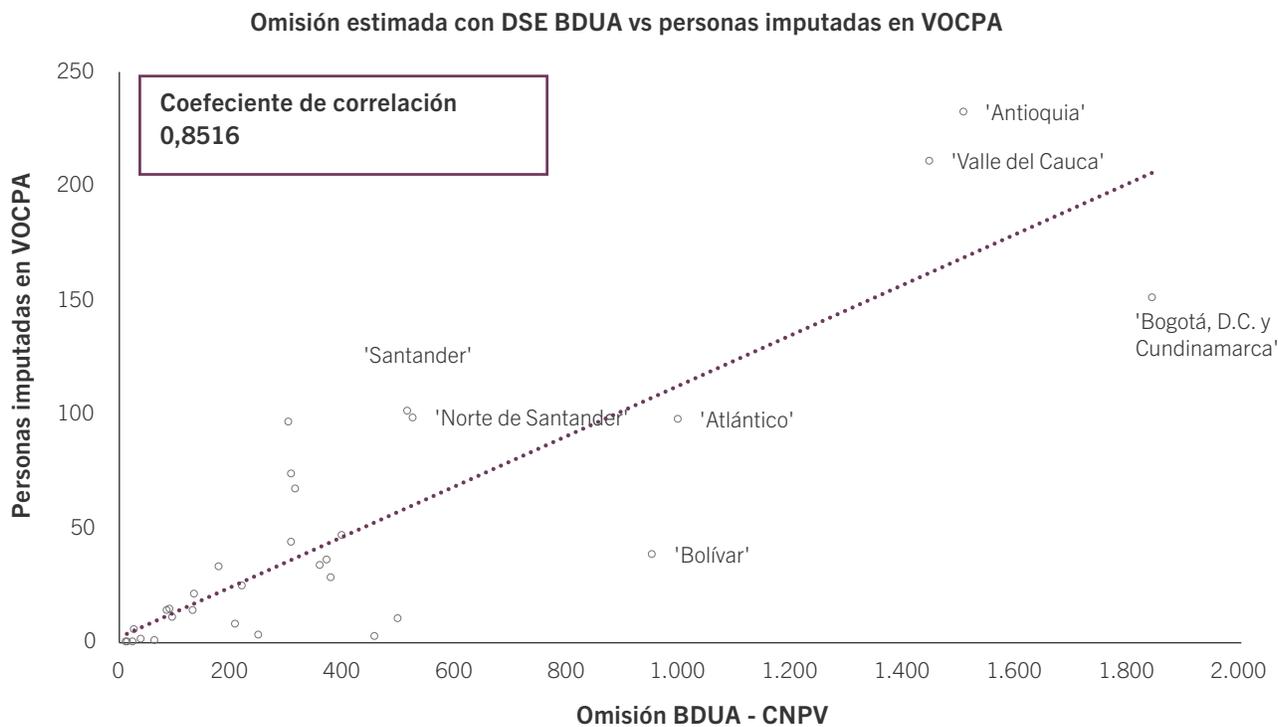


Fuente: DANE, Dirección de censos y demografía CNPV 2018

Las figuras 31 y 32, muestran el comportamiento de las omisiones estimadas por departamento con los registros SISBEN y BDUA comparadas con la imputación realizada de personas en VOCPA. Se destacan los departamentos con mayor imputación (Antioquia, Valle del Cauca, Bogotá y Atlántico), como los departamentos con mayores omisiones estimadas mediante los registros. Es decir, que en

los departamentos donde se focaliza la omisión, con la metodología propuesta, se está mitigando el fenómeno. Además, con coeficientes de correlación lineal de las omisiones estimadas y la imputación en VOCPA de 85,21% y 85,16% para SISBEN y BDUA respectivamente, se evidencia una relación directamente proporcional existente entre la omisión departamental y el proceso de imputación.

Figura 28. | Dispersión Omisión estimada con BDUA vs personas imputadas en VOCPA, valores en miles.



Fuente: DANE, Dirección de censos y demografía CNPV 2018

El proceso de imputación en viviendas VOCPA es una propuesta metodológica, que da continuidad al proceso de omisión iniciado en el CNPV, por lo tanto, no debe estimar una población superior al 8.5%, dato oficial de omisión. Con un valor de 1'520.541 personas en VOCPA se tiene un valor de personas imputadas de 3.2%, dejando un restante de 5.3% de omisión.

Se garantiza que las imputaciones realizadas reflejan las tendencias de omisión que los registros administrativos de SISBEN y BDUA muestran, al encontrar la existencia de una relación directamente proporcional entre lo estimado con DSE en registros administrativos y el valor de imputación.

El proceso de imputación en viviendas VOCPA es estable, una vez se evidencian las distribuciones de las viviendas en el territorio, es un proceso acorde con el principio de parsimonia de las estimaciones, debido a la facilidad del cálculo e

interpretación. Consecuentemente, es un proceso que depende de la calidad de los datos del CNPV. Por ende, refleja de la mejor forma posible el comportamiento de los hogares y sus integrantes dentro de las viviendas.

Omisión por subenumeración de personas en los hogares (S)

4.1 | Antecedentes del método

Con el fin de proponer una metodología apropiada para la estimación de las personas omitidas dentro del hogar es importante estudiar las características que tiene dicha población. En las investigaciones que se han realizado al respecto se encuentra que la composición del hogar y las características sociales, demográficas y económicas de los individuos están relacionadas a la probabilidad de ser omitido dentro del hogar. Por ejemplo, en Estados Unidos y el Reino Unido, se ha encontrado que las viviendas en las que vive más de un hogar, así como el tamaño del hogar y el parentesco que tienen los residentes afecta los porcentajes de omisión en un área de estudio (Martin, 2010), (National Research Council, 1972), (Baffour, 2006), (Ellis, 1995), (Brown, 1999). Así mismo, dichos autores mencionan el papel que tiene la edad y el estado civil en este problema, donde los individuos jóvenes (18 a 29 años) y los que nunca se han casado son los que más se omiten en el censo.

Uno de los métodos más utilizados para estimar la omisión para subgrupos de la población de estudio es el método dual²⁸. Sin embargo, el método dual solo permite realizar estimaciones a nivel global de

la omisión, como se señaló en la estimación de la omisión nacional. Es decir, sin distinguir entre los tres tipos de omisión mencionados anteriormente, debido a ciertos supuestos que se manejan para realizar el emparejamiento, dichas estimaciones no se pueden trabajar a nivel de departamentos o municipios.

Por otro lado, Cressie (1988) propone un modelo de regresión para estimar el nivel de omisión general libre de nivel, es decir que se pueda usar para realizar estimaciones a nivel departamental, municipal o para algún tipo de subgrupo de la población. Así mismo, Smith (2002) plantea un modelo de regresión para estimar el número promedio de personas dentro del hogar con base a variables sintomáticas como lo es el número de nacimientos, número de matriculados en último grado de secundaria e inscritos mayores de 65 años o más al seguro médico por hogar. “Al igual, Kimpel & Lowe (2007) consideran un modelo de regresión para estimar el cambio en el tamaño del hogar de 1990 a 2000 y posteriormente estimar el tamaño del hogar en el 2000.

²⁸ Método descrito en el capítulo 1

Ghosh (1994) hacen una revisión de distintas metodologías aplicadas para abordar la estimación de áreas pequeñas, la cual se puede usar para abordar el problema de omisión censal a nivel municipal, departamental o para subgrupos de la población relativamente pequeños. En particular, para la estimación del número de personas omitidas al interior de los hogares no se encuentra mucha información, sin embargo, al comparar una estimación del tamaño del hogar contra lo observado, se puede llegar a determinar el nivel

de subenumeración en algún tipo de agregación. Teniendo en cuenta que los métodos de imputación múltiple son de utilidad para obtener estimaciones de variables de interés (Rubin, 2004), se propone aplicar esta metodología por medio de información auxiliar, lo cual brinda gran parte de las ventajas que se pueden tener con un enfoque de regresión, esto, con el fin de estimar en una primera etapa un tamaño aproximado de los hogares, y posteriormente, calcular el número de personas omitidas dentro de los hogares de manera directa.

4.2 | 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 postcensal, 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 CNPV como datos faltantes y la ECV proporciona información completa. Teniendo en cuenta que el método de imputación múltiple

proporciona conjuntos de datos completos, es decir, se obtienen tamaños de hogar estimados para cada uno de los hogares del censo, por lo cual, al realizar el promedio de 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 hace una selección de covariables exclusivamente de la ECV, 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 otra 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 CNPV 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 (Figura 33), esto es:

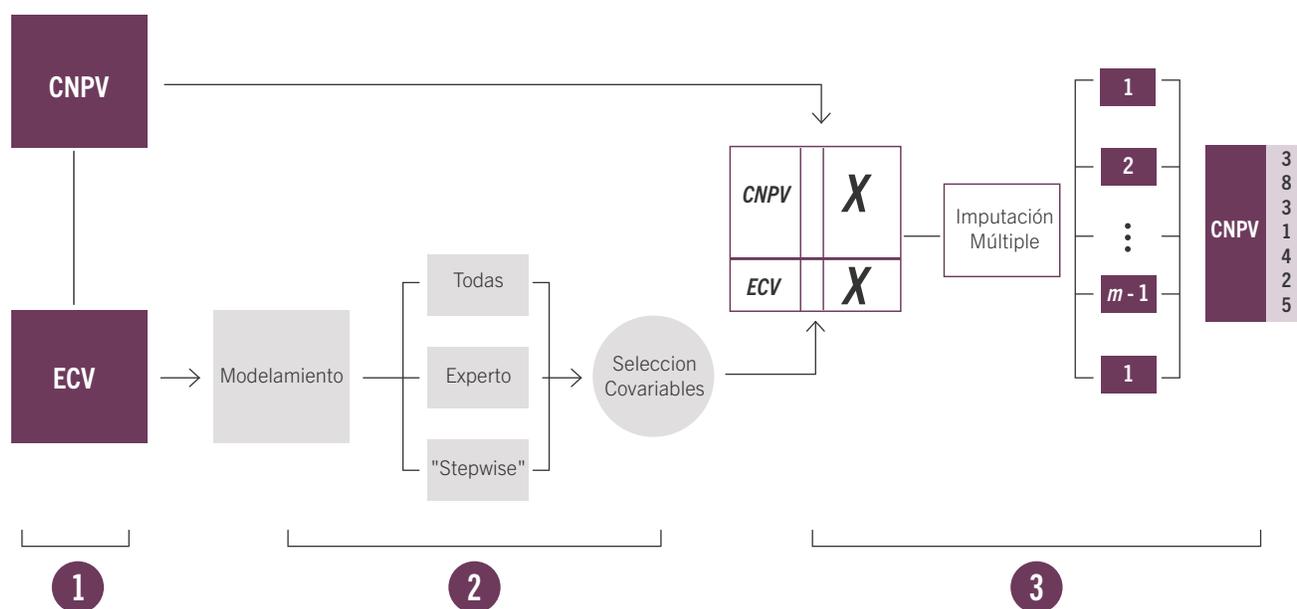
a. Homologación de las bases de datos: En este paso se garantiza que las variables que se usarán posteriormente para el modelamiento e imputación múltiple tengan los mismos niveles, es decir, que la codificación sea igual para ambas.

b. 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.

c. 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.

Figura 29. Esquema del procedimiento para obtener estimaciones de la subenumeración.



Fuente: DANE, Dirección de Censos y Demografía

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 efectúan 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 este se agrega a un nivel superior (departamento).

a. 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. Cabe resaltar que 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, P_i viene dado por

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

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

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

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

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

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

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.

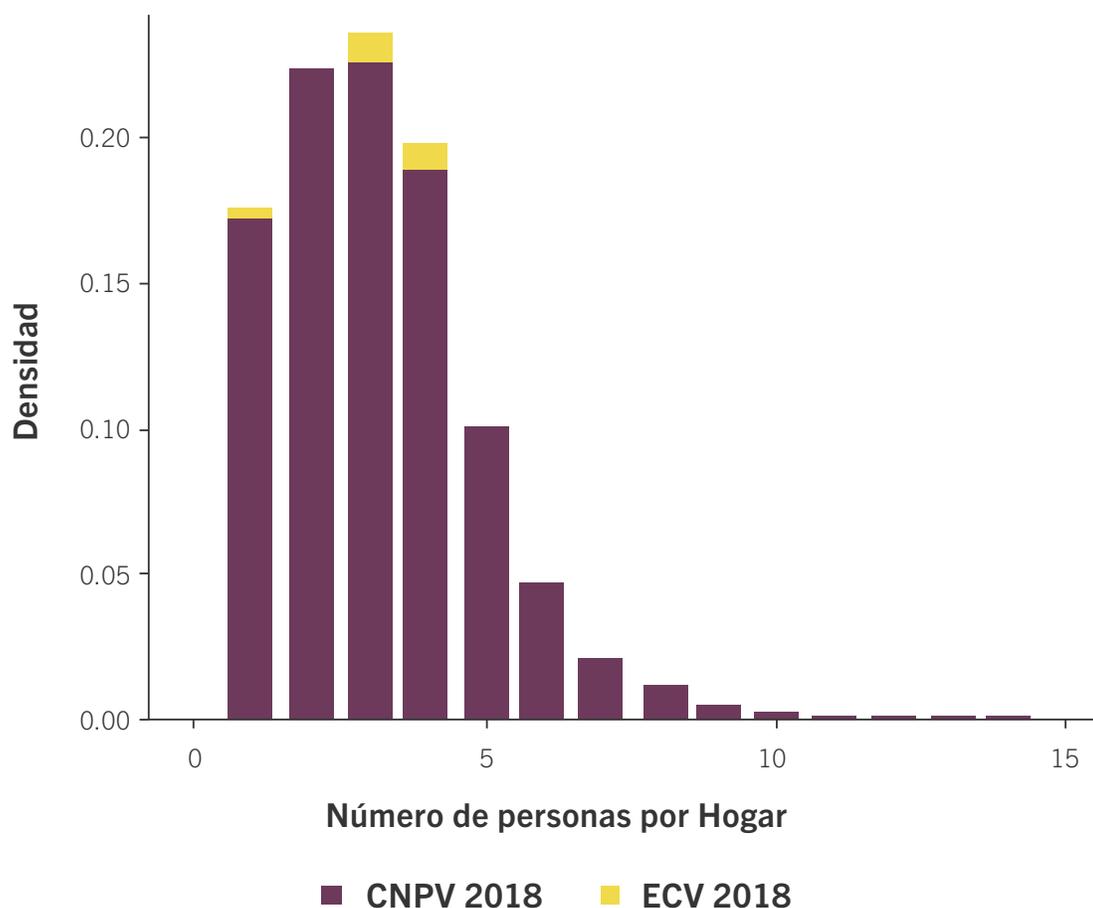
Es importante resaltar que en los casos que $\delta_i \leq 0$, se toma $\delta_i = 0$ debido a que valores negativos de δ_i indican que el valor estimado del hogar está por debajo del observado, por lo cual, se supone que no se omitieron personas dentro de los hogares al momento de la enumeración y,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

b. Covariados seleccionados

El uso de la ECV 2018 como fuente de información auxiliar está motivado por las variables que recolecta relacionadas al tamaño del hogar y dado que se considera como la encuesta postcensal, sin embargo, es importante verificar a partir de los datos la validez que tiene esta información.

Se puede apreciar que la distribución observada del número de personas por hogar a nivel nacional para el CNPV 2018 y ECV 2018 son muy similares entre sí, donde se presentan ligeras diferencias en los hogares de 3 y 4 personas, mientras que en los otros casos la distribución es casi idéntica como se muestra en la figura 34.

Figura 30. Distribución tamaño de hogar CNPV 2018 y ECV 2018



Fuente: DANE, Dirección de Censos y Demografía

Adicionalmente, una de las diferencias que se presentan en estas dos fuentes de información, con respecto al número de personas por hogar es que en el CNPV 2018 se observan hogares con máximo 39 personas, mientras que en la ECV 2018 el valor máximo es 20. Por lo cual, al momento de aplicar la metodología de imputación múltiple las estimaciones del tamaño de hogar no se van a ver influenciadas por estos casos extremos observados en el CNPV 2018. Por esta razón, se tiene que la ECV 2018

es una fuente apropiada de información auxiliar para estimar los tamaños de hogares del CNPV ya que presenta un comportamiento muy similar con respecto a esta variable.

Por otro lado, al ser una encuesta realizada anualmente, los procesos y metodologías para la recolección de información brindan una mejor calidad en las variables medidas, a comparación del CNPV 2018, que es una operación estadística de mayor complejidad.

Una vez se depuraron los datos, eliminando los hogares que no tenían información para alguna de las variables de interés, se llega a una base de

datos con información de 13 700 139 hogares particulares en que las variables se pueden categorizar en 4 grupos:

Tabla 23. | 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

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. Así mismo, en algunos casos el número de categorías que se presentan en la Tabla 23 no corresponden necesariamente con el número de categorías que se reportan en el CNPV, esto se debe a que con el fin de unir la base de datos de este CNPV 2018 y la ECV 2018, las variables deben tener las mismas categorías, por lo cual, se adaptaron las variables para que los niveles fueran lo más parecidos posibles.

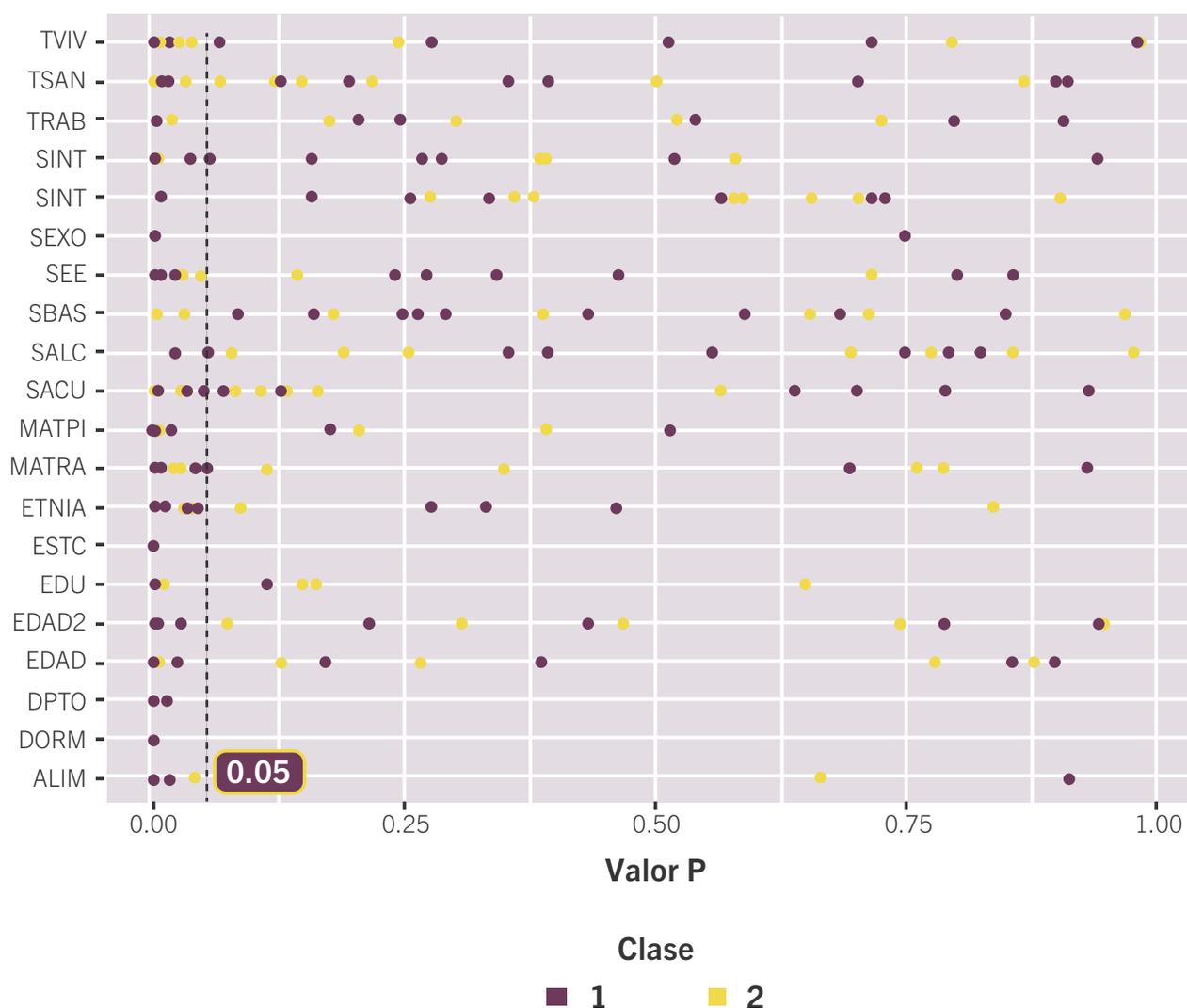
Un ejemplo de esto es la variable de clase, ya que en el CNPV 2018 esta cuenta con tres categorías: cabecera municipal, centro poblado y rural disperso. Sin embargo, en la ECV 2018 esta variable solo cuenta con dos categorías: cabecera municipal y área rural dispersa, en la que ésta última contiene centros poblados, inspecciones de policía o corregimientos.

Por último, la variable que se reporta de regiones se obtuvo a partir de la ECV 2018, ya que el CNPV 2018 no cuenta directamente con esta información. No

obstante, esta agrupación geográfica está dada por las principales regiones de Colombia (ej. Caribe, Orinoquía, Central, etc). En esta categorización de los departamentos, se separan Bogotá D.C., Antioquia y el Valle del Cauca de la región a la que deberían pertenecer, esto permite que los grupos formados sean más homogéneos con respecto a distintas características de interés.

Posteriormente, se realiza el ajuste de los modelos ordinales utilizando la ECV 2018 con el fin de seleccionar las covariables más relacionadas con el tamaño del hogar para aplicar la imputación múltiple en el CNPV 2018.

Figura 31. | Valores p de la estadística de Wald para el ajuste de los 17 modelos usando todas las covariables.



Fuente: DANE, Dirección de Censos y Demografía

Se puede apreciar que al realizar los ajustes usando todas las covariables, en la mayoría de casos las que presentan mayor significancia son el departamento, lugar donde se preparan los alimentos, número de dormitorios, estado civil y sexo del jefe del hogar (Figura 35).

Lo anterior concuerda con lo que se espera del contexto. Por ejemplo, la configuración de hogares cambia con respecto al estado civil, lo cual se ve reflejado en tamaños de hogares distintos para personas solteras, casadas, viudas, etc. Así mismo, en la zona rural dispersa (Clase 2) sobresalen las covariables de grupo, étnico del jefe del hogar, servicio de energía eléctrica y tipo de vivienda (Figura 35). Esto se puede dar debido a que, en la zona urbana, la mayoría de los hogares cuentan con servicios de energía eléctrica (99.30 %), no hay mucha diversidad con respecto al tipo de vivienda y las personas no pertenecen a algún grupo étnico

(85.20 %), lo cual resulta en covariables poco significativas en la parte urbana.

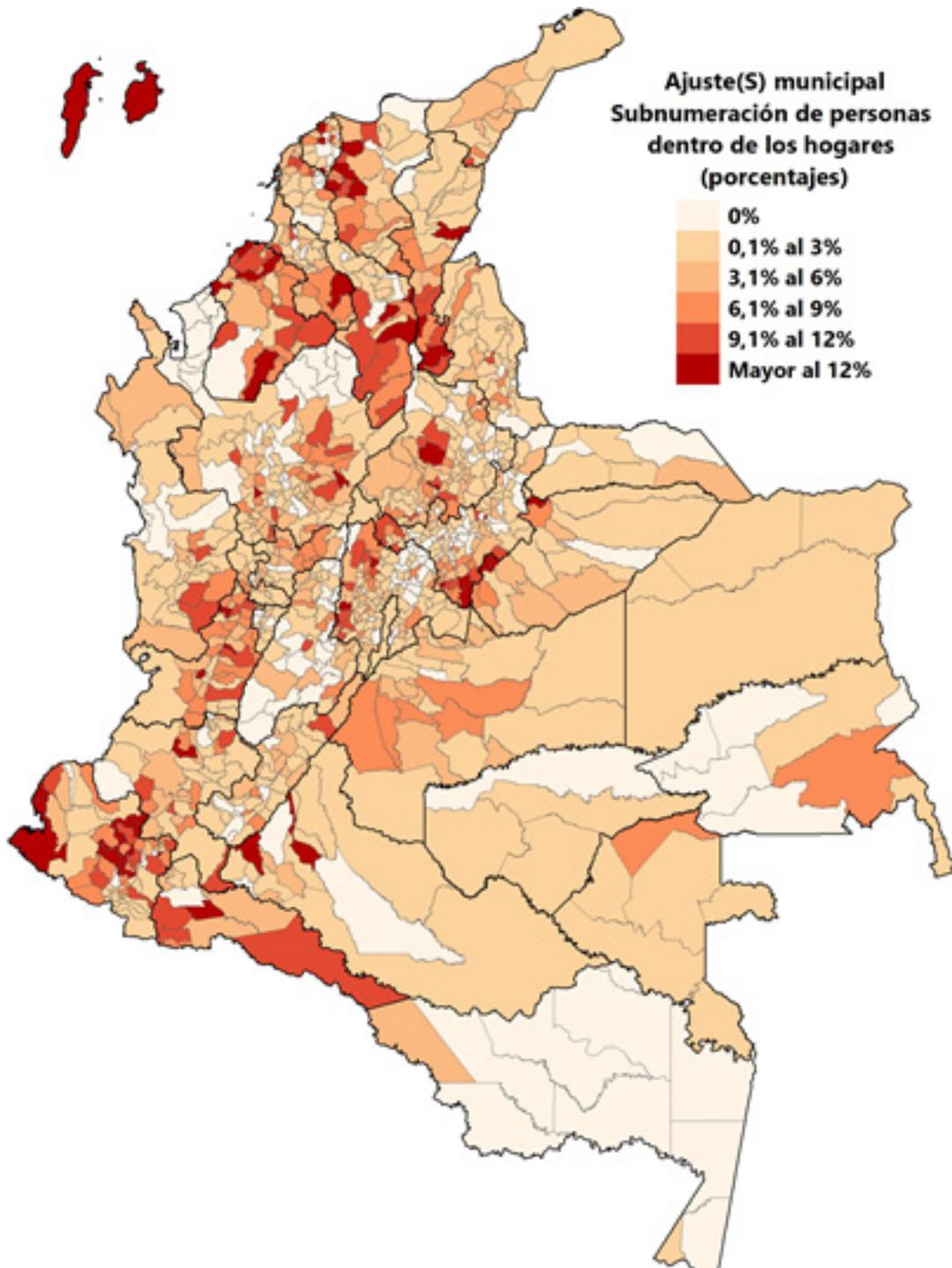
Por otro lado, en la zona urbana (Clase 1) presentan mayor significancia, a comparación de la zona rural, las covariables de nivel educativo y material de pisos y paredes de la vivienda (Figura 3). En este caso, al analizar el comportamiento de estas covariables con respecto al tamaño del hogar, ambas clases presentan un comportamiento relativamente similar, por lo cual este resultado se puede ver afectado por la relación que tienen las covariables entre sí. De igual modo, se destaca que la mayoría de covariables que evalúan si el hogar cuenta con los distintos servicios públicos son poco significativas, esto se puede dar por la presencia de multicolinealidad entre dichas variables, lo cual se confirma al analizar las matrices de correlación para las variables categóricas, ya que, en la mayoría de los modelos ajustados los servicios públicos, presentan correlaciones entre sí usualmente grandes.

4.3 | Resultados

Se obtuvo que en el total del territorio nacional se subenumeraron al 1.127.141 personas al interior de los hogares censados. Se puede apreciar en la Figura 36 que los municipios con mayor porcentaje de personas omitidas en el CNPV 2018 no presentan un comportamiento espacial claro, ya

que se encuentran distribuidos a lo largo del país, donde 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.

Ilustración 5. | Distribución territorial de Subenumeración municipal, presentados como porcentaje



Fuente: DANE, Dirección de Censos y Demografía

En la zona rural se observa un comportamiento similar al de la zona urbana a diferencia de municipios del occidente del Meta, norte del Guaviare, Vaupés y Guainía, que presentan porcentajes altos en la zona rural y bajos en la zona urbana.

Así mismo, Bogotá D.C. en la zona rural presenta uno de los ajustes más altos mientras en la zona urbana es nulo. Esto se debe a que la estimación del tamaño promedio del hogar en Bogotá D.C. en la zona urbana es de 4.44, mientras que el observado en el CNPV 2018 es de 2.87. Es decir que la metodología aplicada indica que se omitió más de una persona en todos los hogares, lo cual conduce a una estimación de más de un millón de personas.

Teniendo en cuenta que Bogotá es la capital del país, en la que se presenta una gran variedad de estructuras de hogar y que se espera que la recolección de la información en el operativo del

censo sea de buena calidad, se decide omitir este resultado en la zona urbana, alejados de las otras estimaciones de tamaño de hogar disponibles.

En el caso de San Andrés y Providencia, se identifica que en la zona urbana son uno de los municipios que presentan mayor ajuste, mientras que en la zona rural este es nulo, esto se debe a que en la ECV 2018 no existe información de éstos en la zona rural, por lo cual, la metodología de imputación múltiple no se puede aplicar.

Al analizar conjuntamente los resultados de subenumeración para la zona rural y urbana, se aprecia un comportamiento similar al descrito anteriormente para cada una de las zonas, donde los municipios que presentan mayor ajuste son los de la cordillera occidental, Putumayo, Nariño, Bolívar, Sucre, Córdoba y San Andrés y Providencia. Por el contrario, gran parte de los municipios del suroriente del país, costas del Chocó y Cauca presentan ajustes relativamente bajos. (Figura 36).

Omisión Total

5.1 | 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

$$Z_N = 1.235.287$$

$$A_N = 1.520.541$$

$$S_N = 1.127.141$$

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

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

$$\text{Omisión Nacional} = 8,1\%$$

Como se señaló en el capítulo I, la omisión aceptada e informada para el total nacional corresponde 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.

5.2 | Ajuste al techo de omisión Censal Nacional informada

Una vez desagregada la omisión censal a nivel municipal cabecera - resto, se determinó que el prorrateo 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. Por tanto, se analizó ajustar a partir de la selección de municipios con cierre operativo después del 29 de septiembre 2018 (Tabla 24), y que requieren un tratamiento diferenciado en su estimación de omisión.

Tabla 24. | Municipios con cierre posterior al 29 de septiembre de 2018.

Divipola	Departamento	Municipio
23570	Córdoba	Pueblo Nuevo
5318	Antioquia	Guarne
13006	Bolívar	Achí
76377	Valle del Cauca	La Cumbre
88001	Archipiélago de San Andrés	San Andrés
5604	Antioquia	Remedios
5674	Antioquia	San Vicente
76001	Valle del Cauca	Cali
76233	Valle del Cauca	Dagua
23574	Córdoba	Puerto Escondido
76364	Valle del Cauca	Jamundí
23068	Córdoba	Ayapel
23682	Córdoba	San José de Uré(1)
76109	Valle del Cauca	Buenaventura

5440	Antioquia	Marinilla
5686	Antioquia	Santa Rosa de Osos
23855	Córdoba	Valencia
5615	Antioquia	Rionegro
23417	Córdoba	Lorica
76520	Valle del Cauca	Palmira
23678	Córdoba	San Carlos
5036	Antioquia	Angelópolis
23660	Córdoba	Sahagún
5756	Antioquia	Sonson
76834	Valle del Cauca	Tuluá
23001	Córdoba	Montería
76892	Valle del Cauca	Yumbo
76126	Valle del Cauca	Calima
5649	Antioquia	San Carlos
47707	Magdalena	Santa Ana
76616	Valle del Cauca	Riofrío
5148	Antioquia	El Carmen de Viboral
76111	Valle del Cauca	Guadalajara de Buga
47245	Magdalena	El Banco
13001	Bolívar	Cartagena
76563	Valle del Cauca	Pradera
5376	Antioquia	La Ceja
5088	Antioquia	Bello
5266	Antioquia	Envigado
5579	Antioquia	Puerto Berrío
23807	Córdoba	Tierralta
76895	Valle del Cauca	Zarzal
23555	Córdoba	Planeta Rica
76275	Valle del Cauca	Florida

5652	Antioquia	San Francisco
5360	Antioquia	Itagui
25377	Cundinamarca	La Calera
76248	Valle del Cauca	El Cerrito
76890	Valle del Cauca	Yotoco
76250	Valle del Cauca	El Dovio
47001	Magdalena	Santa Marta
23090	Córdoba	Canalete
70717	Sucre	San Pedro
8141	Atlántico	Candelaria

Fuente: DANE, Dirección de Censos y Demografía Operativo CNPV 2018

El procedimiento realizado para su corrección se basó en la selección de unidades no vistas e incompletas para el ajuste Z en estos municipios A partir del análisis de selección espacial de unidades a incompletas²⁹. Para los municipios de la Tabla 24 se seleccionaron las unidades a partir del índice de Moran Univariado y el resto de los municipios a partir del criterio bivariado, realizando el procedimiento de ajuste en Z con la siguiente variación:

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 arrojó 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.

$$Z_N = 1.409.398$$

$$A_N = 1.520.541$$

$$S_N = 1.164.148$$

Y_N = 4.094.087 Personas omitidas total Nacional, que determinan una omisión censal de **8.5%**

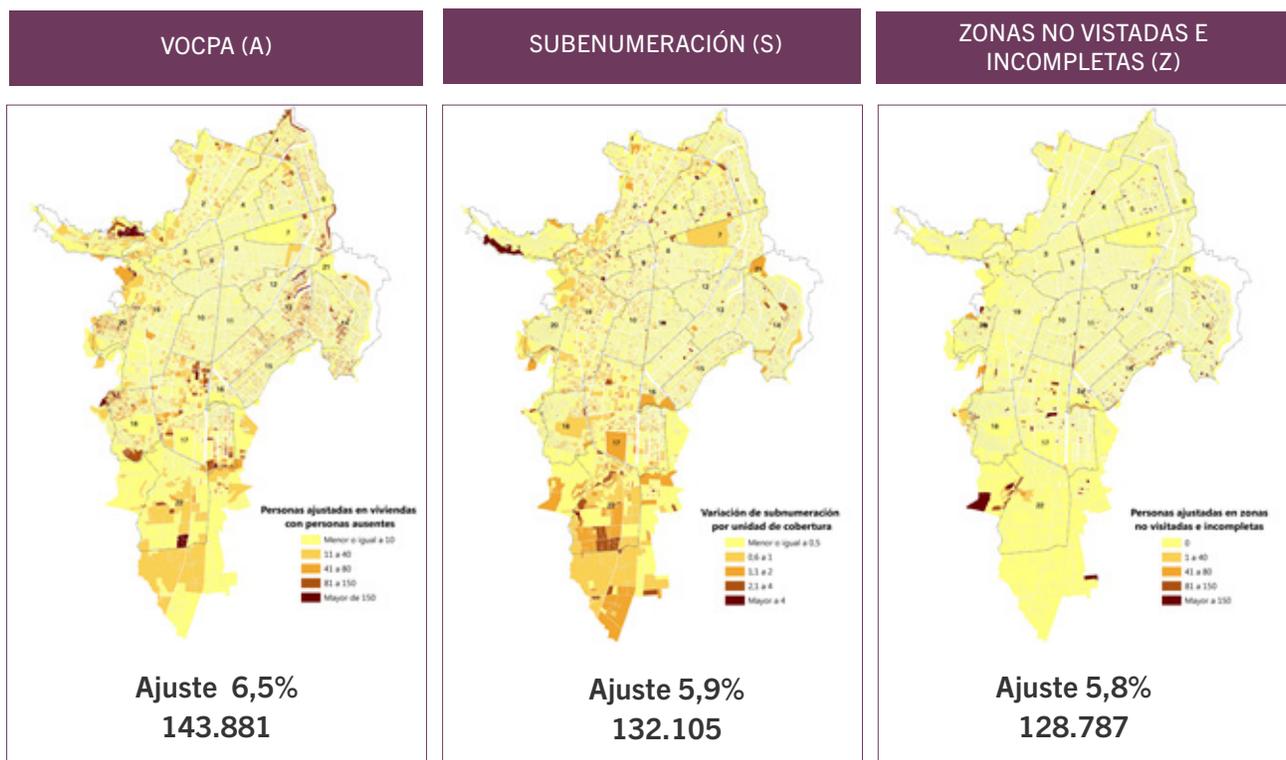
²⁹ Ver capítulo 2, Numeral 2.1

5.3 | Nivel de desagregación de la omisión censal

A partir de los métodos aplicados y del uso intensivo de la información georreferenciada, las estimaciones se generaron por unidad de

cobertura, realizando el procedimiento de agregación a este nivel, como se muestra en el siguiente ejemplo, para la ciudad de Cali:

Ilustración 6. | Distribución de los factores de omisión para la ciudad de Cali (Valle del Cauca).



Fuente: DANE, Dirección de Censos y Demografía

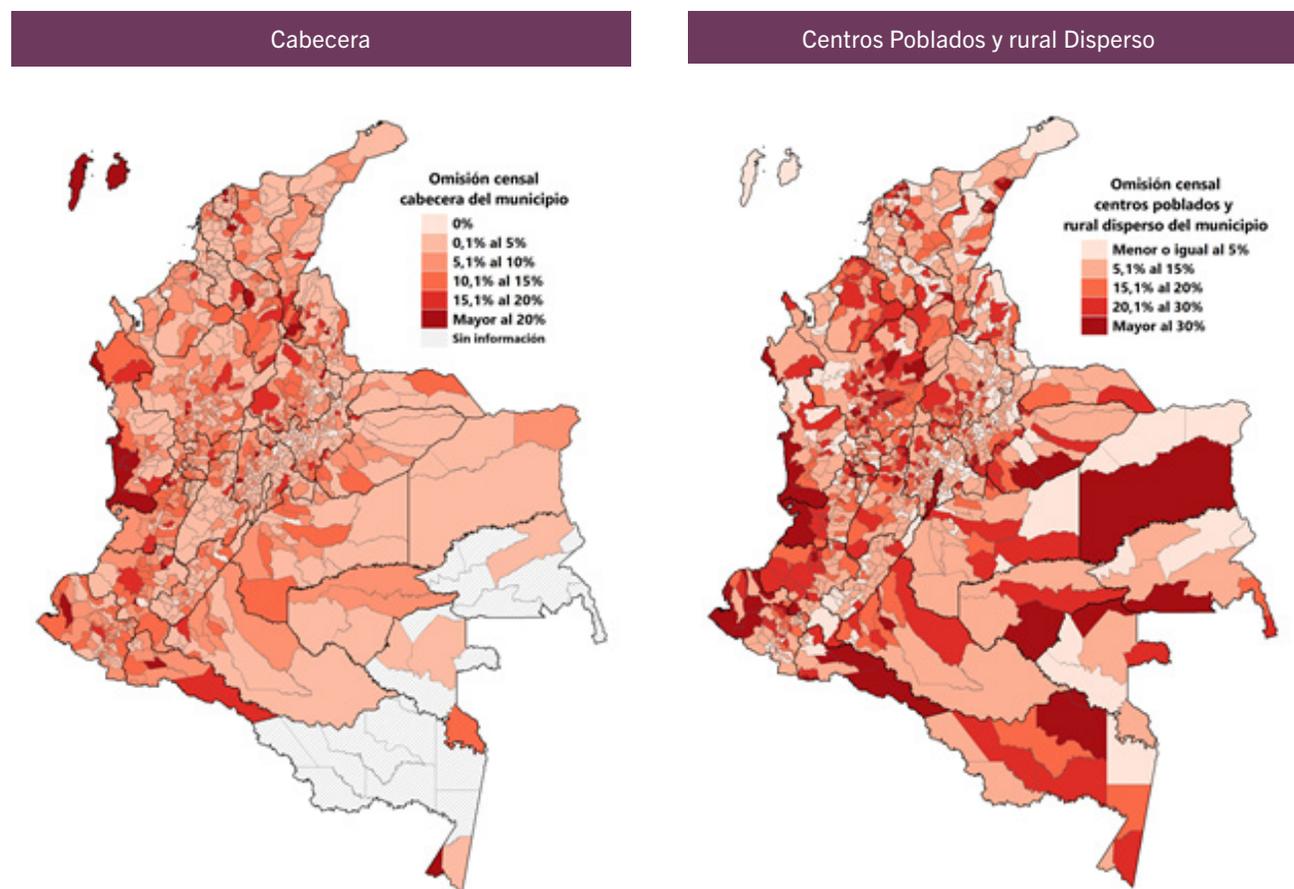
El potencial del nivel de detalle geográfico que se obtuvo con la aplicación de los métodos propuestos corresponde a la posibilidad de realizar agregaciones a los niveles requeridos, cabecera, centros poblados, rural disperso, municipal y departamental. Incluso, este es un elemento básico para las proyecciones de población en ciudades principales, dando la posibilidad de realizar agregaciones a nivel de localidad.

a. Ajuste de la cobertura municipal por área y total municipal³⁰

Los resultados de ajuste por área muestran un diferencial de ajuste de mayor omisión en las zonas de centros poblados y rural disperso con un total de omisión de 15,0%, frente a un 6,4% en las áreas urbanas.

³⁰ Los resultados del ajuste de omisión censal, pueden ser consultados en <https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/censo-nacional-de-poblacion-y-vivenda-2018>

Ilustración 7. | Ajuste de cobertura por área, a nivel municipal



Fuente: DANE, Dirección de Censos y Demografía

El mayor ajuste en las zonas rurales y centros poblados muestra las dificultades de acceso y menor capacidad de realizar visitas en estas áreas geográficas. En cuanto a las cabeceras municipales se evidenciaron las dificultades para algunas ciudades sin un patrón predeterminado.

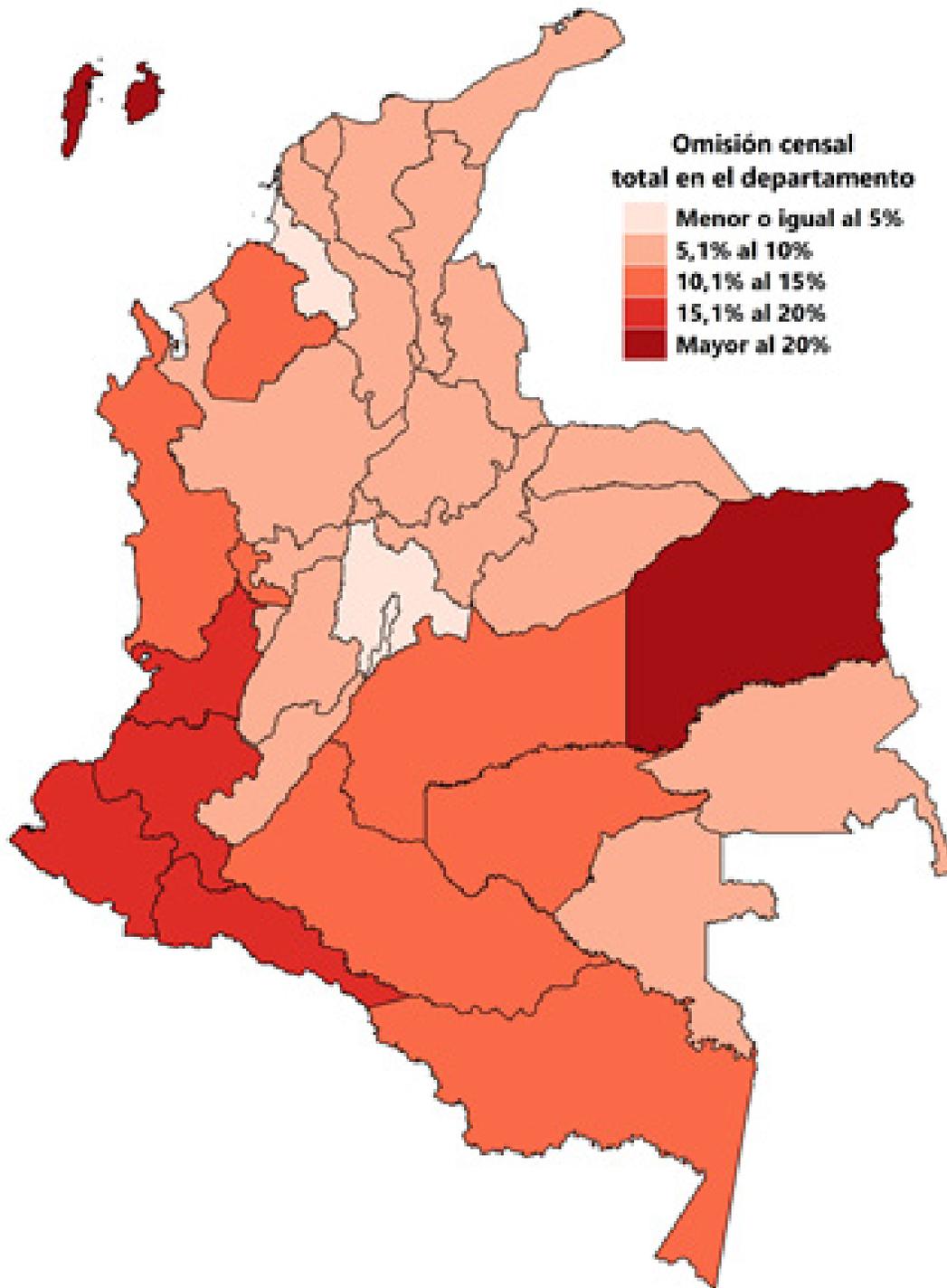
b. Ajuste de la cobertura departamental

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 más 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%.

Ilustración 8. | Ajuste de cobertura a nivel departamental.



Nombre de departamento	Total	Cabecera	Centros poblados y rural dispersos
Total Nacional	8,5%	6,4%	15,0%
Antioquia	6,7%	3,9%	16,7%
Atlántico	7,6%	7,4%	10,9%
Bogotá D.C.	3,1%	3,0%	39,5%
Bolívar	7,8%	7,2%	9,4%
Boyacá	6,7%	3,5%	11,2%
Caldas	7,5%	6,1%	11,6%
Caquetá	10,5%	5,8%	19,0%
Cauca	15,1%	9,8%	18,2%
Cesar	8,5%	7,7%	11,0%
Córdoba	12,8%	6,8%	19,5%
Cundinamarca	4,3%	2,3%	9,4%
Chocó	14,5%	7,2%	20,5%
Huila	8,3%	8,0%	8,6%
La Guajira	6,3%	4,6%	7,8%
Magdalena	5,8%	4,2%	9,5%
Meta	11,6%	11,5%	11,9%
Nariño	18,1%	9,6%	24,7%
Norte de Santander	9,7%	9,3%	11,2%
Quindío	5,6%	4,9%	10,6%
Risaralda	11,0%	10,7%	12,2%
Santander	8,1%	6,8%	11,9%
Sucre	4,5%	1,3%	10,0%
Tolima	7,6%	3,9%	15,6%
Valle del Cauca	15,3%	14,9%	17,8%
Arauca	8,6%	7,2%	11,4%
Casanare	9,7%	6,6%	16,9%
Putumayo	18,7%	9,1%	28,3%
Archipiélago de San Andrés	21,2%	28,4%	1,5%
Amazonas	13,8%	6,6%	20,4%
Guainía	7,7%	4,7%	9,8%
Guaviare	11,7%	5,6%	19,3%
Vaupés	7,6%	4,4%	9,0%
Vichada	28,9%	4,9%	36,5%

Fuente: DANE, Dirección de Censos y Demografía

4.

Conclusiones



Para la estimación nacional por método directo (Dual System) se observaron las ventajas con las que cuenta esta metodología, relacionadas a la confiabilidad de las estimaciones. Esto debido al emparejamiento directo de las fuentes y una reducción en supuestos que se deben asumir en comparación con otros métodos. Además que efectúa reducción de tiempos y costos, estimaciones a bajos niveles de desagregación, y facilita la interpretación de los resultados obtenidos.

Respecto a la conciliación censal, pese a que no fue considerada para estimar de forma oficial la omisión, permitió conocer escenarios de la estructura más aceptable para el CNPV 2018 y su concordancia con los censos anteriores. Sus resultados estuvieron dentro del margen estimado con un gran acercamiento al resultado de omisión desagregado a nivel subnacional.

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 mantiene 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 y características de las viviendas.

El análisis realizado evidencia 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 CNPV 2018 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, 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 encuentra 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 permita a la demografía integrar estas herramientas como parte de las nuevas técnicas.

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.

De igual manera, es importante notar que, a pesar de Bogotá D.C. ser un área que alberga gran cantidad de habitantes, los resultados obtenidos en la zona urbana no son consistentes con lo esperado, lo cual se puede dar por la heterogeneidad que se encuentra en esta ciudad en cuanto a la estructura de los hogares. Por esta razón, es fundamental el contexto del problema para determinar la calidad de las estimaciones y el sentido que pueden llegar a tener.

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

Bibliografía

- 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*. Sanitiago 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 demograficos: 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 demograficos: Alcance, limitaciones y métodos de evaluación. Santiago de Chile: Publicación de Naciones Unidas ISSN 1680-886x.
- CEPAL- UNFPA. (2014). Los dtos demograficos: 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 empirico 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, Department of Commerce, Bureau of the Census: presentados en: *Curso Regional Intensivo de Demografía, 1997* CELADE.
- Joaquin Recaño. (2015 y 2019). *Estimaciones de los SNM 2005-2020*. Bogotá: Consultoría 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 demograficas. 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.
- Sardinle, M. (2008). On the Performance of Dual System Estimators of Population Size: A Simulation Study. CERAC, Paper, No. 13.
- Sardinle, 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 . Diponible en: https://www.dane.gov.co/files/censos/Grupo3_Version0902.pdf?&: DANE.
- 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.
- Vehults, 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.

DANE

INFORMACIÓN PARA TODOS



@DANE_Colombia



/DANEColombia

www.dane.gov.co



/DANEColombia



@DANEColombia

