

DOCUMENTO METODOLÓGICO

Metodología para la recolección de información primaria

Convenio interadministrativo UPME No. CV-008-202, IPSE No.147-2021, suscrito entre la Unidad de planeación minero-energética- UPME; el Instituto de planificación y promoción de soluciones energéticas para las zonas no interconectadas - IPSE y la Universidad del Cauca.



Universidad
del Cauca

Vigilada Mineducación



PERS Cauca

Plan de energización Rural Sostenible

Documento Metodológico

Metodología para la recolección de información primaria

Sistematización, Redacción y Análisis

Maximiliano Bueno López

Director Académico del Proyecto

Andrés Mauricio Gómez Sánchez

Líder caracterización socioeconómica

Luis Felipe Jiménez Chaves

Profesional en Estadística

Cristian Eduardo García Bermúdez

Analista de datos

Juliana Isabel Sarmiento Castillo

Analista de datos

Plan de energización Rural Sostenible del Departamento del Cauca

PERS-CAUCA

Universidad del Cauca

UPME

IPSE

Popayán, Cauca

Colombia

Febrero de 2022

1. Introducción

El presente documento tiene como propósito básico explicar las metodologías y técnicas de muestreo utilizadas para la determinación de la muestra representativa para las unidades residenciales y no residenciales presentes en el Departamento del Cauca. Esto se realiza a partir del diseño de instrumentos estadísticos apropiados para el levantamiento de información primaria.

Este documento será la guía para el trabajo de campo que se desarrollará durante la ejecución del Plan de Energización Rural y Sostenible (PERS) Cauca y busca tener información en torno al aprovechamiento de la energía por fuente, tenencia y usos para diversos sectores, en especial, el residencial y otros como el institucional, comercial e industrial.

El documento consta de cuatro partes: i) los objetivos que se sustentan en la relevancia del PERS Cauca en su contexto geopolítico; ii) la definición de las viviendas residenciales para el marco muestral; iii) el diseño y estimación de la muestra en distintos tipos de unidades; iv) la distribución de las encuestas según subregión o tipo de unidad.

2. Objetivos

En seguimiento a los objetivos generales del PERS para el Departamento del Cauca, se establecen los lineamientos para la definición de una muestra representativa a nivel subregional que permita cumplir con los siguientes objetivos:

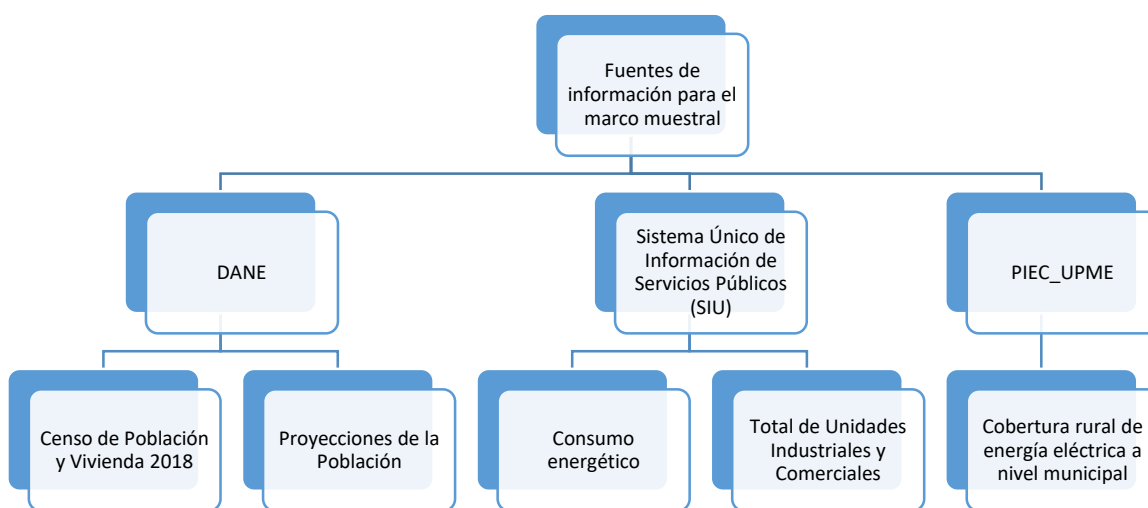
- Caracterizar el consumo básico de energía por uso y fuente en el sector residencial de zonas rurales representativas por medio de información primaria.
- Caracterizar y analizar las variables socioeconómicas de las familias que residen en viviendas de zonas rurales y cabeceras, en especial, para municipios con niveles de ruralidad superiores al 40% a partir de información obtenida a partir de encuestas primarias e información secundaria.
- Analizar la oferta de recursos energéticos en las subregiones del departamento del Cauca.
- Determinar el consumo básico de energía eléctrica para la subsistencia del sector residencial.

Para fijar e implementar un número de encuestas adecuado en cada municipio y/o subregión de acuerdo con sus características de ruralidad y de cobertura energética, se emplearon métodos de muestreo que son ajustados de acuerdo con la experiencia de otros PERS departamentales (Nariño, Putumayo y Cundinamarca).

3. Marco muestral

Para la definición del número adecuado de encuestas que permitan inferir sobre las características de uso y consumo de energía eléctrica en las viviendas rurales y/o cabeceras municipales del departamento del Cauca, así como unidades no residenciales, se emplearon múltiples fuentes de información que permitieran tener una idea del comportamiento de los parámetros poblacionales de algunos indicadores claves sobre el tema. El siguiente esquema resume las principales fuentes de información y las metodologías empleadas para la estimación y asignación de la muestra por distintas unidades de desagregación (viviendas, establecimientos comerciales, industriales e institucionales).

Esquema 1. Fuentes de información para el marco muestral del PERS Cauca



Fuente: Elaboración PERS_CAUCA

La información de las distintas fuentes de información se analizará de manera detallada en este documento, sin embargo, de manera a priori es importante precisar que de acuerdo con fuentes de información secundarias (DANE y SUI) y debido a la heterogeneidad socioeconómica del departamento muy probablemente se tendrán comportamientos territoriales no homogéneos en las unidades de referencia con respecto a los consumos y acceso a energía eléctrica.

La definición del marco muestral para distintas unidades residenciales y no residenciales se realizó en dos procesos diferentes, debido a que el Censo de Población y Vivienda 2018, no contempla unidades con fines comerciales y

económicos¹. Para el análisis de unidades comerciales, industriales e institucionales se consideró la información del Sistema Único de Información (SUI), instrumento que permite aproximarse a la distribución, número de unidades suscriptas, niveles de consumo y otras características del sector de servicios públicos domiciliarios del país. Para las viviendas residenciales el insumo clave fue la información del Censo de Población y Vivienda 2018, misma que fue apoyada por las Proyecciones de la Población instrumentos sociodemográficos y socioeconómicos publicados por el Departamento Administrativo Nacional de Estadística (DANE).

De acuerdo con los objetivos del PERS y las limitaciones de cada fuente de información, para definir el marco y diseño muestral se dio prioridad a las unidades residenciales, es decir, en este caso la técnica de muestreo contempló estratos y conglomerados que permitieran inferir para el total del departamento y también para distintas demarcaciones territoriales. Para la inferencia de las unidades no residenciales se utilizaron técnicas de muestreo que permitieran describir el comportamiento de los establecimientos comerciales, institucionales e industriales a nivel departamental. Los análisis que se realizan para estos sectores por subregiones se presentan como una guía para el levantamiento de información en el trabajo de campo.

Además, una de las características de las fuentes empleadas para los marcos muestrales es que permiten obtener información a nivel municipal, sin embargo, si se quisiera obtener indicadores representativos sobre el uso y consumo de energía eléctrica a nivel de las 42 demarcaciones territoriales que integran el departamento se tendría que optar por un tamaño y una asignación de encuestas suficientes para todos los municipios. Este proceso implicaría incrementar los costos de operación para el trabajo de campo, así como el tiempo de levantamiento y tabulación de la información.

Por lo tanto, la estrategia metodológica para contemplar la heterogeneidad del departamento en cuanto a la cobertura y consumo de energía eléctrica fue considerar la información a nivel municipal para estimar las medidas de heterogeneidad con respecto al tema de interés, que permitieron estimar los tamaños muestrales; y para la distribución de las encuestas y definición de los factores de expansión se contemplaron las subregiones que agrupan los municipios por cuestiones geográficas y por lineamientos definidos por los órganos de planeación departamental como la Gobernación (Tabla 1).

¹ En Colombia próximamente se publicarán resultados sobre el primer Censo Económico, mismos que a la fecha de elaboración de este documento no están disponibles para los usuarios

Tabla 1. Subregiones del Departamento del Cauca

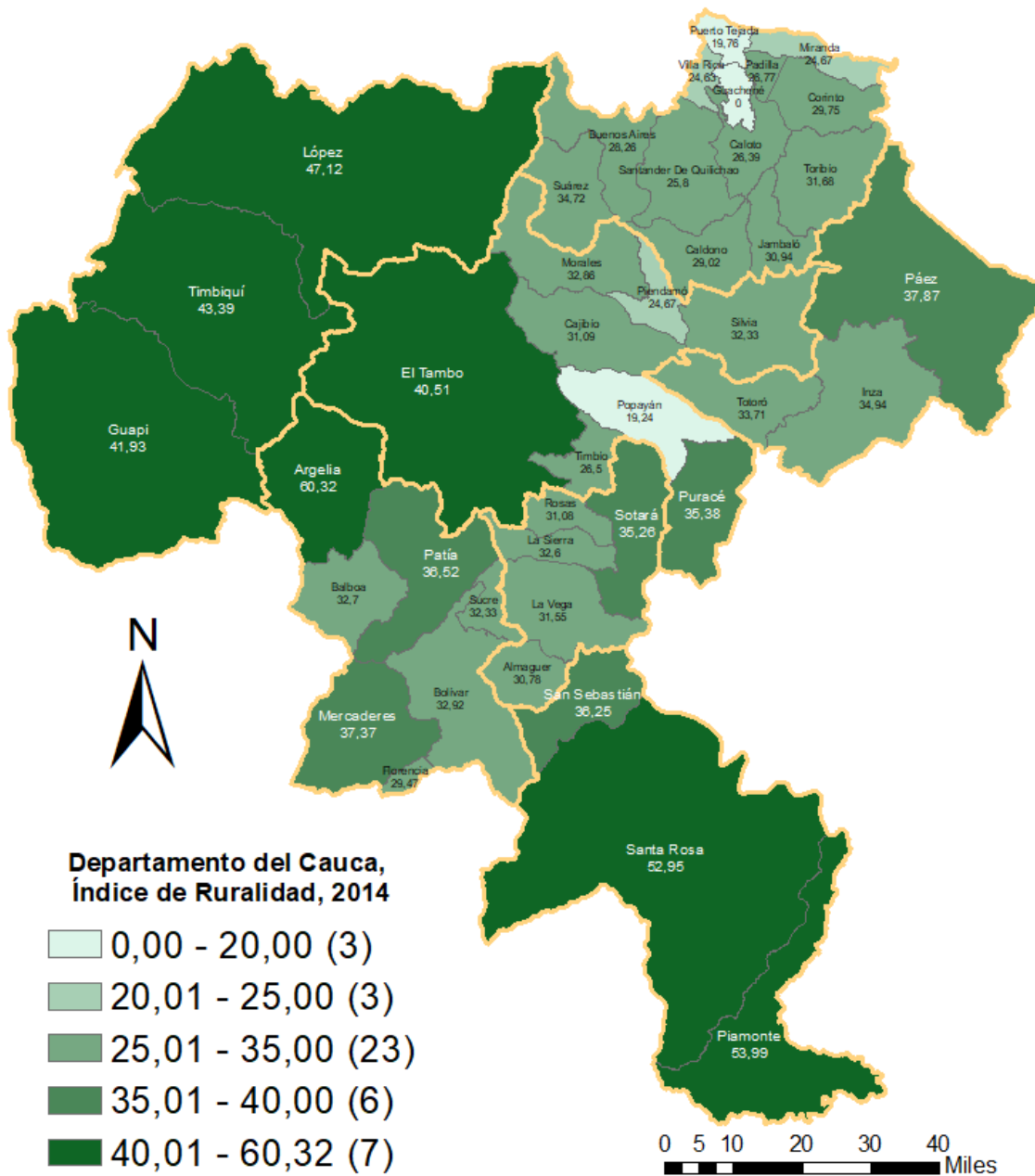
Subregión y número de municipios	Municipios
Bota Caucana (3)	Piamonte, Santa Rosa, San Sebastián
Centro (8)	Cajibío, Morales, Popayán, Piendamó, Puracé, Timbio, El Tambo, Silvia
Macizo (5)	Almaguer, La Sierra, Rosas, La Vega, Sotará
Norte (13)	Buenos Aires, Caldono, Caloto, Corinto, Guachené, Jambaló, Miranda, Padilla, Puerto Tejada, Santander de Quilichao, Suárez, Toribio, Villa Rica
Oriente (3)	Inzá, Páez, Totoró
Occidente (3)	Guapi, Timbiquí, López de Micay
Sur (7)	Argelia, Balboa, Bolívar, Florencia, Mercaderes, Patía, Sucre

Fuente: Elaboración propia con base en información cartográfica del Instituto Geográfico Agustín Codazzi.

En relación con las unidades no residenciales, es importante anotar que las cifras del SUI tienen datos no reportados o ausencia de información para algunos municipios, en especial, para la zona occidente. La información de las unidades suscritas fueron el sustento para definir los tamaños muestrales, en 2021 se reportaron 15.815 establecimientos comerciales, 7.129 unidades institucionales (centros de salud, escuelas, dependencias públicas, alumbrado público, etc.) y 2.141 industrias.

En el caso de las unidades residenciales, para el marco muestral se contempló la dinámica de la ruralidad en los municipios y/o subregiones del departamento del Cauca. La evidencia empírica a partir de fuentes de información como el Censo Nacional Agropecuario 2016, mostró que el Departamento del Cauca es una de los territorios con mayor ruralidad del país. A partir de este indicador se observó que en algunos municipios se concentra una cantidad significativa de viviendas con características rurales e incluso en sus cabeceras municipales, por lo tanto, de acuerdo con los antecedentes de PERS para otros departamentos, se decidió que en las demarcaciones con un índice de ruralidad superior al 40%, las viviendas ubicadas en cabeceras municipales se contemplan para la estimación de las encuestas.

Mapa 1. Departamento del Cauca. Índice de Ruralidad. 2014



Fuente: Informe PERS Cauca con base en PNUD Informe Nacional de Desarrollo Humano – 2011
 Nota: No se tiene información para el municipio de Guachené (Zona Norte)

Por lo tanto, para tener mayor precisión en las estimaciones derivadas de la muestra y alcanzar a inferir sobre zonas o subregiones con distintos escenarios de acceso a energía eléctrica en el departamento, el universo de estudio se centró en las viviendas rurales con distintas demandas de consumo energético distribuidas en los

municipios y/o subregiones del departamento. Es relevante mencionar que esta delimitación dio prioridad a municipios con baja cobertura del servicio, pero también permitió que las subregiones más pobladas del departamento tuvieran una participación notoria como el caso de la zona centro y norte. La tabla 2 presenta el universo de viviendas por municipio y subregión.

Tabla 2. Universo de viviendas consideradas para el diseño muestral según municipio y/o subregión

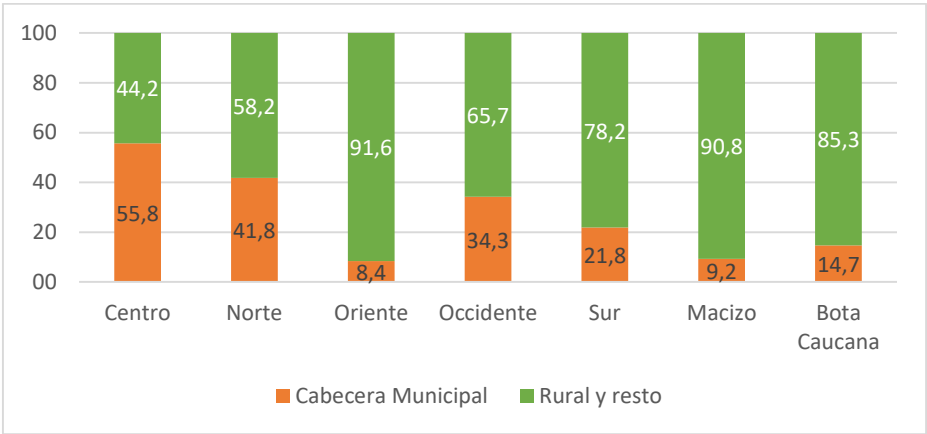
Subregion	ID_municipio	Municipio	Universo de viviendas rurales y/o cabeceras	Universo y porcentaje de viviendas por subregión	Universo de viviendas totales
Centro	19001	Popayán	19,818	85,641 (29.98%)	111,936
	19130	Cajibío	11,643		12,451
	19256	El Tambo	16,809		16,809
	19473	Morales	10,390		11,582
	19548	Piendamó	8,016		12,295
	19585	Puracé	3,677		4,280
	19743	Silvia	7,960		10,015
	19807	Timbio	7,328		11,879
Norte	19110	Buenos Aires	9,209	81,139 (28.41%)	9,644
	19137	Caldono	10,205		10,852
	19142	Caloto	6,459		8,571
	19212	Corinto	4,369		9,135
	19300	Guachene	4,267		6,244
	19364	Jambaló	3,670		4,155
	19455	Miranda	4,254		10,595
	19513	Padilla	1,721		3,353
	19573	Puerto Teja	2,144		13,984
	19698	Santander de Quilichao	16,790		36,504
	19780	Suárez	7,633		9,750
	19821	Toribio	8,462		9,231
	19845	Villa Rica	1,956		7,402
Oriente	19355	Inza	7,560	21,562 (7.55%)	8,041
	19517	Páez	8,555		9,665
	19824	Totoro	5,447		5,824
Occidente	19318	Guapi	7,067	19,227 (6.73%)	7,067
	19418	López de Mesa	5,054		5,054
	19809	Timbiquí	7,106		7,106
Sur	19050	Argelia	8,864	44,748 (15.67%)	8,864
	19075	Balboa	6,283		8,322
	19100	Bolívar	13,232		15,028
	19290	Florencia	1,558		1,966
	19450	Mercaderes	5,191		7,241
	19532	Patía	7,448		11,836
	19785	Sucre	2,171		2,772
Macizo	19022	Almaguer	6,330	24,403 (8.54%)	6,959
	19392	La Sierra	2,821		3,569
	19397	La Vega	7,298		7,824
	19622	Rosas	3,491		3,876
	19760	Sotará	4,463		4,663
Bota caucana	19533	Piamonte	2,957	8,923 (3.12%)	2,957
	19693	San Sebastián	3,956		4,247
	19701	Santa Rosa	2,009		2,009
Total				285,643	465,558

Fuente: PERS Cauca con base en DANE, Censo de Población y Vivienda 2018 y Proyecciones de la Población

Nota: Los municipios resaltados indican que el índice de ruralidad supera el 40%.

En síntesis, se tomará como universo de estudio 285.643 viviendas que representan 61.4% del total de viviendas proyectadas para el departamento del Cauca (465.558), la mayoría localizadas en los municipios de la zona centro (29.98%) y zona norte (28.41%). Sin embargo, en las otras cinco subregiones el porcentaje de viviendas rurales y resto supera la participación relativa de viviendas ubicadas en cabeceras municipales (Gráfica 1). Una nota aclaratoria, es que al momento de estimar los factores de expansión a nivel subregional podrían existir variaciones leves en el universo estudio, esto se debe a que en el proceso inferencial se asumen errores de cobertura, sin embargo, dichas diferencias en términos porcentuales son cercanas al uno por ciento.

Gráfica 1. Distribución porcentual de las proyecciones de viviendas según subregión y contexto, 2020



Fuente: PERS Cauca con base en Censo de Población y Vivienda 2018 y Proyecciones de la Población de los municipios de Colombia.

4. Diseño muestral

Para la estimación de la varianza se emplearon diversos indicadores de acceso a cobertura eléctrica o número de unidades suscriptoras del servicio. En este apartado, los resultados se centran en los indicadores seleccionados para el tamaño de muestra óptimo, sin embargo, en los siguientes segmentos se resumen las estimaciones con base en distintas fuentes de información.

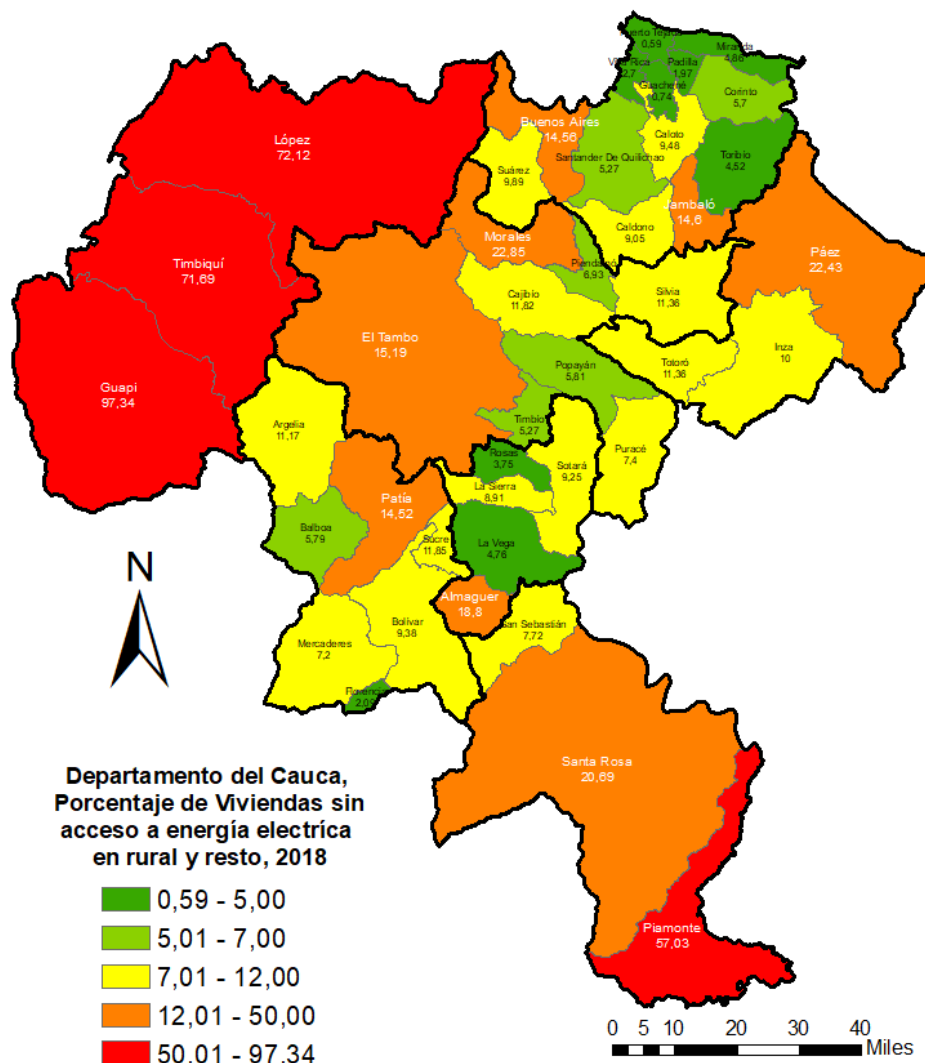
4.1. Indicadores para estimar la varianza en las unidades residenciales

Para aproximarnos a los parámetros poblacionales de las viviendas rurales se consideraron distintas fuentes de información que proporcionan la variabilidad del objeto de estudio (cobertura de energía eléctrica). Los indicadores empleados para estimar el tamaño de muestra óptimo fueron el porcentaje de viviendas sin acceso a energía eléctrica a nivel municipal a partir del Censo Nacional de Población y Vivienda 2018 y de la información del Plan Indicativo de Expansión de Cobertura de Energía Eléctrica (PIEC) de 2018 de la Unidad de Planeación Minero-Energética.

En el departamento del Cauca, el acceso a energía eléctrica difiere según el contexto de residencia. El Censo Nacional de Población y vivienda 2018 (CNPV2018) estimó que 98.97% de las viviendas localizadas en cabeceras municipales y 86.35% de las localizadas en centros poblados y rurales dispersos tenían energía eléctrica en sus unidades residenciales. Sin embargo, el UPME proyectó para 2020 que la cobertura rural de energía eléctrica departamental fue de 83.6 %. Si bien en el agregado departamental las estimaciones del DANE y UPME se asemejan, es importante, anotar que en algunos municipios persisten diferencias importantes.

La información del mapa 2 permite apreciar la heterogeneidad en cuanto al acceso a energía eléctrica en los municipios del departamento del Cauca con base en la información del DANE. En los tres municipios del departamento de la zona Occidente y Piamonte (Bota Caucana), la cobertura de acceso a energía eléctrica es limitada; en el caso de Guapi 97.34% de viviendas no tenían disponible este servicio. Ahora bien, en algunas subregiones sobresalen algunos municipios como el caso de Morales (Centro), Páez (Oriente), Santa Rosa (Bota Caucana), donde aproximadamente una de cada cinco viviendas no tenía acceso a energía eléctrica. En contraposición, Popayán y Timbío, Piendamó en la Zona Centro; Bolívar en la Sur; Santander de Quilichao y Corinto en la zona norte, destacaron por bajos porcentajes de viviendas rurales sin acceso al servicio.

Mapa 2. Porcentaje de Viviendas rurales y resto sin acceso a energía eléctrica. Departamento del Cauca. 2018



Fuente: PERS Cauca con base en Censo Nacional de Población y Vivienda 2018

Por su parte, el mapa 3 muestra el porcentaje de viviendas sin acceso a energía eléctrica en cabeceras municipales. En general, si se compara con los sectores rurales y centros poblados en todos los municipios, el acceso a energía eléctrica en sus cabeceras municipales es alto, sin embargo, sobresalen Guapi y Piamonte con porcentajes cercanos al 10%, situación que justifica la inclusión de las unidades ubicadas en las cabeceras municipales de acuerdo con el índice de ruralidad al momento de distribuir el tamaño óptimo de encuestas.

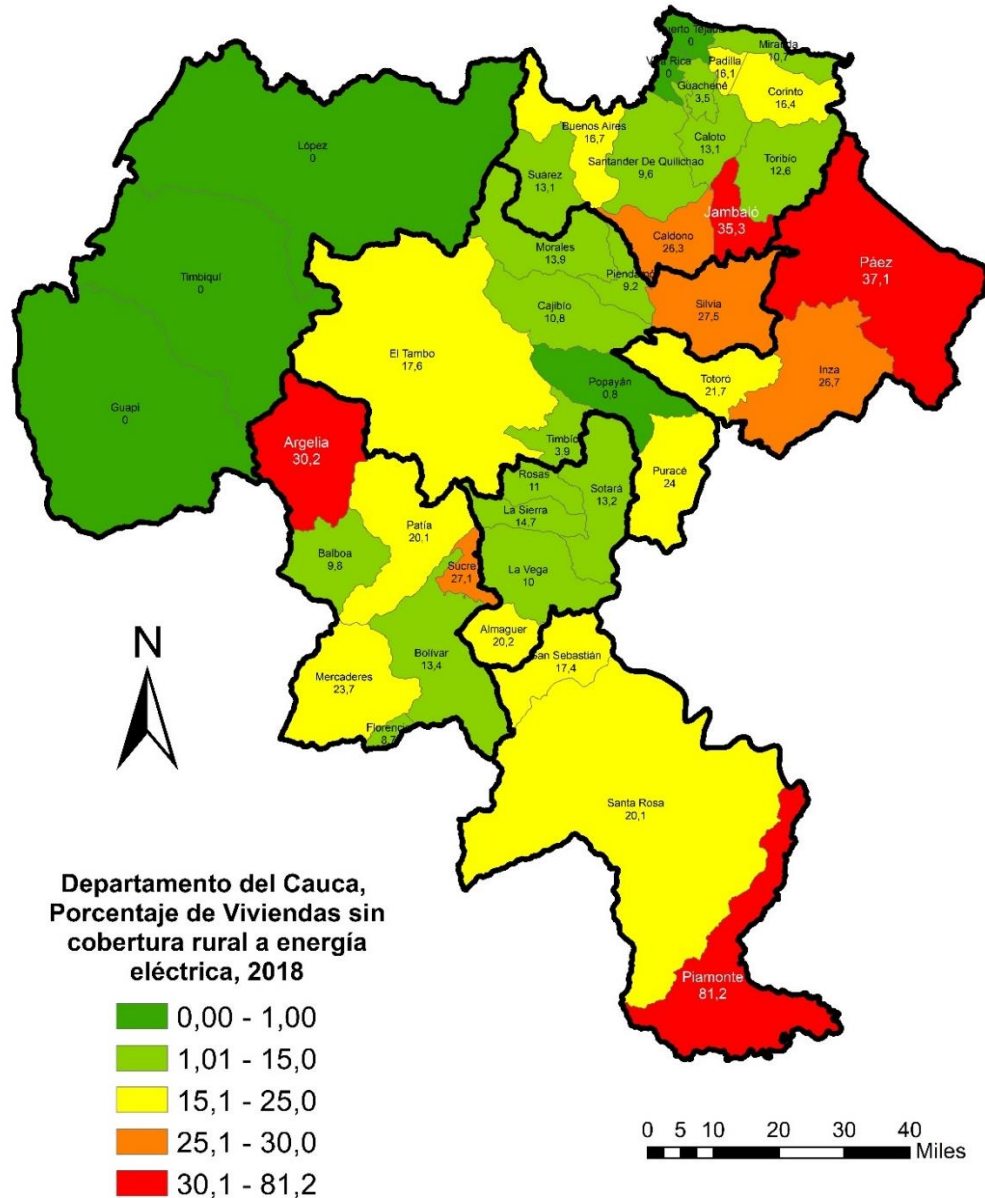
**Departamento del Cauca,
Porcentaje de Viviendas sin
acceso a energía eléctrica
en cabeceras municipales, 2018**

Color	Porcentaje de Viviendas sin acceso a energía eléctrica
Verde oscuro	0,00 - 0,64
Verde claro	0,65 - 1,66
Amarillo	1,67 - 3,53
Naranja	3,54 - 5,95
Rojo	5,96 - 10,89

0 5 10 20 30 40 Miles

Ahora bien, al analizar el indicador de porcentaje de viviendas rurales sin cobertura de energía eléctrica según la información del PIEC, se observan algunos municipios con baja cobertura como es el caso de Piamonte (81.2%), Páez (37.1%), Jambaló (35.3%) y Argelia (30.2%). En sentido contrario, resaltan con cobertura total los tres municipios de la zona occidente del departamento, estos resultados si bien no son directamente comparables con los mapas 2 y 3, muestran ordenamientos disímiles en cuanto al acceso a cobertura de energía.

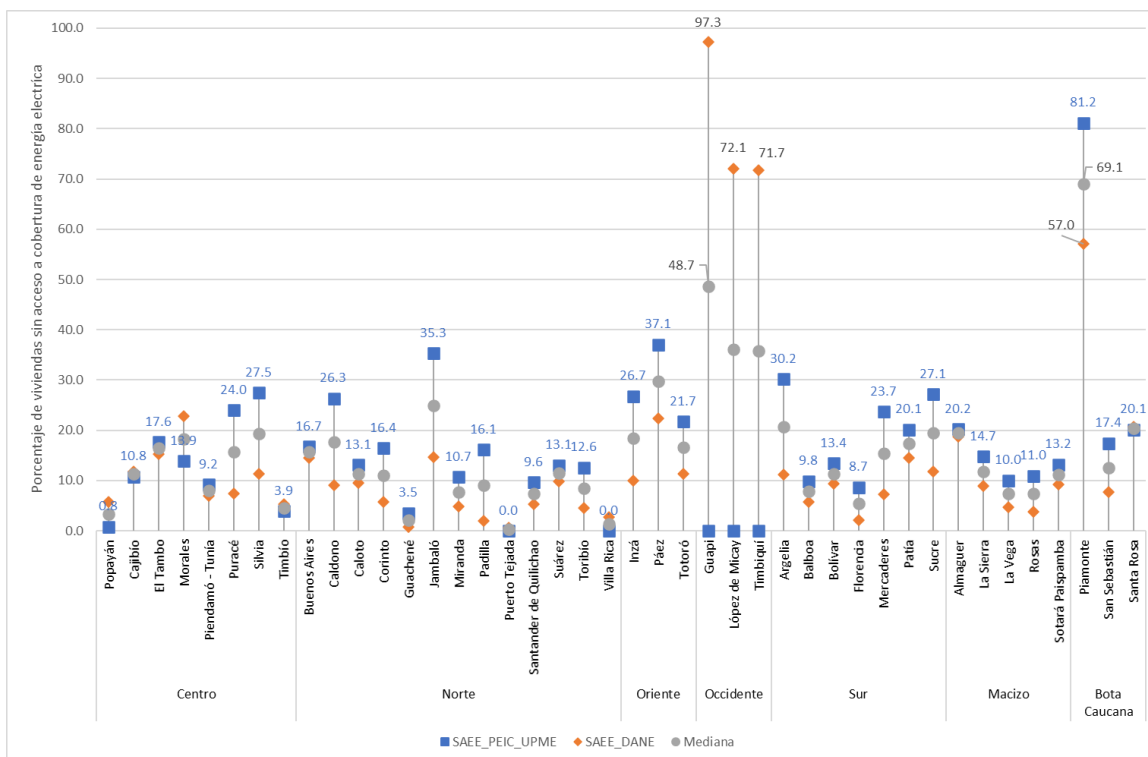
**Mapa 4. Porcentaje de viviendas rurales sin cobertura de energía eléctrica.
Departamento del Cauca. 2020**



Fuente: PERS CAUCA con base en Plan Indicativo de Expansión de Cobertura de Energía Eléctrica (PIE) de 2018 de la Unidad de Planeación Minero-Energética.

Por tanto, con el objetivo de contemplar la heterogeneidad existente en este indicador a partir de ambas fuentes de información, se empleó la mediana entre ambas observaciones para la estimación de las varianzas apropiadas que permitan aplicar la estimación del tamaño muestral (Gráfica 2). Se considera importante la mediana para reducir la dispersión en el indicador de interés, en especial, para los municipios de la subregión occidente y otros particulares como Piamonte, Jambaló y Páez.

Gráfica 2. Porcentaje de viviendas rurales sin cobertura de energía eléctrica según distintas fuentes



Fuente: PERS CAUCA con base en Plan Indicativo de Expansión de Cobertura de Energía Eléctrica (PIE) de 2018 de la Unidad de Planeación Minero-Energética y Censo Nacional de Población y Vivienda 2018.

4.2. Definición del muestreo para las unidades residenciales

De acuerdo con el indicador de la mediana de la falta de acceso a cobertura de energía eléctrica, se empleó un método de muestreo que permitiera contemplar la heterogeneidad de este indicador en los 42 municipios del departamento del Cauca, es importante, reiterar que la agregación y estimaciones finales se realizarán en términos de las subregiones, si bien se utilizan los municipios como insumo para la obtención de la varianza y el planteamiento del tamaño de muestra, todos los resultados serán plasmados a nivel de las subregiones.

4.2.1. Tipo de muestreo para las unidades residenciales

Muestra probabilística, estratificada, de conglomerados, multietápica:

- Probabilística: Cada unidad de la población objetivo tiene una probabilidad conocida de selección mayor a 0.
- Estratificada: Con el fin de ganar precisión de la muestra, se propone realizar una estratificación de acuerdo con las siete subregiones:

centro, norte, oriente, pacifico, sur, macizo y piedemonte amazónica, para un total de 7 estratos muestrales.

- Conglomerados: Unidad final de muestreo compuesta por el segmento (definir cuántas viviendas por cada uno, por ejemplo, la gran encuesta de hogares utiliza 10).
- Multietápica
 - Unidad Primaria de Muestreo (UPM): para las cabeceras son grupos contiguos de acuerdo con la división política (ej: manzanas) del mismo municipio, que contiene mínimo 12 segmentos con un promedio de 10 viviendas contiguas (5 a 14 viviendas). En la zona rural, la UPM está conformada por una sección o sector censal.
 - Unidad Secundaria de Muestreo (USM): el segundo paso es seleccionar dentro de cada UPM, el número necesario de segmentos requeridos, los cuales corresponden a la USM.
 - Unidad terciaria de Muestreo (UTM): finalmente, la UTM corresponde a las viviendas que conforman cada USM seleccionada para conformar la muestra.

4.2.2. Tamaño de la muestra para las unidades residenciales

Teniendo en cuenta el muestreo planteado multietápico, el cálculo del tamaño de muestra se realizará por medio de las siguientes expresiones:

$$n = \frac{\sum_h \sum_L w_{h.L} (B_{h.L} + C_{h.L})}{V(\hat{P}) + \sum_h \sum_L w_{h.L}^2 \left(\frac{B_{h.L}}{N_{h.L}} \right)}$$

$$B_{h.L} = \left(\frac{S_{UPM(h)}^2}{\bar{X}_{0h}^2} \right)$$

$$C_{h.L} = \left(\frac{\sum_{i \in N_h} M_{hi} (1 - f_{2h}) S_{UPM(h)}^2}{M_h} \right)$$

Donde n es el tamaño de la muestra.

La varianza $V(\hat{P})$ es calculada por medio de la expresión $\left(\frac{EM}{Z} \right)^2$, donde EM representa el máximo margen de error absoluto aceptable, que de acuerdo con el

nivel de significancia será del 10%. Además, el estadístico Z dado un nivel de confianza deseado del 90%, equivale a 1.64².

$W_{h.L}$: peso de cada estrato en la etapa 1 (estrato h) y etapa 2 (estrato L) en el muestreo

$B_{h.L}$ y $C_{h.L}$: valor calculado del diseño muestral acorde a la varianza. Se estiman a priori de la información disponible, para el caso particular serán los promedios y varianzas de la población objeto de estudio.

$N_{h.L}$: representa el tamaño de las poblaciones objeto de estudio.

f_{2h} : representa el ratio para que el número de unidades (UR) residenciales al interior de cada lado de manzana en cada uno de los estratos (subregiones) sea proporcional al número de UR en el lado analizado o encuestado (auto ponderación). Para ello suponemos que dicha fracción es constante en la segunda etapa del muestreo.

$S_{UPM(h)}^2$: varianza a priori estimada desde la fuente de información

$\overline{X_{0h}}^2$: es el promedio de las estimaciones obtenidas a priori según cada estrato dentro del diseño muestral. Se obtiene de la información sobre la cual se planteó el marco muestral

Las proporciones $B_{h.L}$ y $C_{h.L}$ resaltan los pesos correspondientes a la asignación por subregión y municipio respectivamente, los pesos son calculados de la información a priori con la se cuenta. Sin embargo, es de destacar que las estimaciones fueron tomadas como un estimador de la mediana entre la fuente del PERS y el DANE. La varianza general fue calculada de dichas fuentes de información, por lo que se ajustó el tamaño de muestra a un total de 2.092 encuestas.

Para la obtención de la muestra se debe considerar la desagregación de las unidades residenciales por subregiones y dentro de estas a los municipios. Los valores de referencia que se presentan en las fórmulas anteriores permiten obtener una estimación de la muestra teniendo en cuenta el diseño de muestreo. Dentro de la estimación, se consideran las varianzas intra-region, intra-municipio y la varianza total; también se consideran los pesos de cada municipio en la subregión y de la

² En la obtención de la muestra es importante considerar que, de acuerdo con el teorema del límite central, para tamaños de n lo suficientemente grandes la prueba t y los valores z de la distribución normal son equivalentes. Para el lector interesado, consultar Latpate, *et al.* (2021).

subregión dentro del departamento, esto con la finalidad de garantizar la representatividad de cada una de las subregiones³.

Teniendo en cuenta los valores previos, fijando un nivel de confianza en el 90% y un margen de error permisible máximo del 10%, se obtiene un tamaño muestral de 2.092 encuestas para las unidades residenciales.

Es importante precisar que el ejercicio de estimación de la muestra se realizó con varios indicadores que contemplan diferentes unidades de observación y fuentes de información; sin embargo, dado el interés de investigación, el tamaño muestral se asignó con base en la variabilidad del porcentaje de viviendas sin acceso a energía eléctrica en zonas rurales y centros poblados.

Tabla 3. Estimación del tamaño de muestra según diferentes fuentes de información

Indicador y fuente de Información	n	Promedio	Varianza	Margen de error (e)	Nivel de Confianza deseado (Z)
Porcentaje de viviendas sin acceso a energía eléctrica en zonas rurales y centros poblados (DANE, Censo Nacional de Población y Vivienda 2018)	2.526	15,44	426,43	0,10	1,64
Componente de Servicios del Índice de Necesidades Básicas Insatisfechas (Censo Nacional de Población y Vivienda 2018)	2.561	7,67	107,38	0,10	1,64
Suscriptores de un servicio de energía eléctrica (Sistema Único de Información del sector de servicios públicos domiciliarios del país)	2.635	13,59	559,27	0,10	1,64
Porcentaje de viviendas sin Acceso a Energía Eléctrica (DANE, Gran Encuesta Integrada de Hogares, 2018)	2.477	7,35	56,90	0,10	1,64

³ En el anexo metodológico se profundiza en los pasos específicos para la estimación del tamaño de muestra.

Porcentaje de viviendas sin acceso a Energía Eléctrica (SISBEN, Sistema de Identificación de Potenciales Beneficiarios de Programas Sociales, 2020)	3.263	5,88	104,91	0,10	1,64
Porcentaje de viviendas sin acceso a Energía Eléctrica según la mediana (DANE, Censo Nacional de Población y Vivienda 2018 y UPME)	2.092**+	84,50	16,5581	0,10	1,64

Fuente: Elaboración PERS-CAUCA

4.3. Distribución de encuestas por subregión para las unidades residenciales

Para realizar la asignación de la muestra, se tomará como base la estimación de la proporción de viviendas sin acceso a energía por cada subregión, de esta manera se obtendrá los pesos asociados para cada una de las subregiones del departamento del Cauca. En cuanto a las unidades de los conglomerados, se planteará una estimación de la proporción de viviendas sin acceso a energía eléctrica en las zonas rurales de los municipios.

Para asignar la muestra, se hará uso de las probabilidades de Bayes, por lo tanto, se multiplicarán las proporciones y el tamaño de muestra, en este sentido se obtendrá la muestra específica para las subregiones con baja cobertura al servicio y una participación considerable de viviendas rurales. Una vez se tienen los pesos asociados, la asignación proporcional de las estimaciones se realiza por medio de tablas en Excel⁴.

Ahora bien, de acuerdo con las estimaciones realizadas, es importante que en algunos municipios se logre obtener información para un número determinado de encuestas que permita inferir sobre el comportamiento del consumo energético al interior de cada subregión. La siguiente tabla muestra la distribución por subregión con la referencia de los municipios claves.⁵

** El redondeo de encuestas a nivel subregional equivale a 2.092 debido al margen de error alcanzado por la agregación.

+ El indicador calculado se encuentra entre 0 y 1, para su extrapolación con la población objeto de estudio se expresa en miles. (Invariabilidad en la escala)

⁴ La información utilizada para los diferentes cálculos realizados se los puede encontrar en el documento de Excel nombrado como "Soportes metodología_PERS_CAUCA" compartido en el repositorio: https://github.com/cristian42253/PERS_Cauca.

⁵ En la primera parte del anexo 1 se muestra paso a paso cómo se obtuvieron los factores de expansión y posteriormente, el código en R para determinar el tamaño de la muestra.

Tabla 4. Distribución de encuestas según Unidades residenciales por subregión

Subregión	Número de Encuestas	Factor de expansión	Unidades residenciales alcanzada con el factor de expansión
Centro	507	168	85,123
Norte	306	266	81,337
Oriente	150	144	21,649
Occidente	707	27	19,081
Sur	206	217	44,672
Macizo	107	230	24,599
Bota caucana	109	82	8,912

Fuente: Elaboración PERS-CAUCA

En la Tabla 4 se muestra la desagregación del factor de expansión a nivel subregional. Es de notar que al realizar los cálculos, la suma total de los productos comprende el 99.87% de la población total, esto se debe al margen de error atribuible a las fuentes de información. No obstante, es importante resaltar que al contemplar los intervalos de confianza, las estimaciones se asemejarán a los valores poblacionales, por lo tanto, el producto del número de encuestas y el factor de expansión no necesariamente coincidirá con la estimación de viviendas residenciales.

El cálculo de las probabilidades a nivel subregional no se puede realizar de la misma forma en las que se realizan a nivel municipal, en el *anexo 1* se agregan los resultados municipales como una referencia para la definición en el trabajo de campo de cuáles podrían ser los municipios en los que se podría centrar el levantamiento de información, sin embargo, será muy importante que se tenga información por subregiones de acuerdo con los tamaños muestrales definidos en la tabla 3.

Con la finalidad de garantizar la representatividad de las subregiones, es de suma importancia que en aquellos lugares en donde se asignó más de 100 encuestas, estas efectivamente se lleven a cabo, entendiendo que en las condiciones logísticas se podrán redistribuir de manera aleatoria aquellas asignaciones estadísticamente no significativas.

El factor de expansión corresponde a la asignación de las probabilidades de elegir cada elemento de la población según las características asociadas al objetivo del estudio. Hay que tener en cuenta que para el cálculo se utiliza el teorema de Bayes, el cual estará anidado a la población objeto de estudio y a la asignación muestral. Por tanto, en este caso siempre se realizará como la fracción de uno sobre la probabilidad de las viviendas de las zonas rurales del departamento del Cauca dado que no tienen acceso a energía eléctrica.

4.4. Definición del muestreo para las unidades no residenciales

Para las unidades no residenciales presentes en el territorio departamental, con base en el Sistema Único de Información de Servicios Públicos (SUI), se obtuvo una aproximación de la distribución de estas unidades por sector y subregión. Como se mencionó en el apartado del marco muestral, en 2021 estaban suscritas en el SUI 15.815 unidades comerciales, 2.141 industrias y 7.129 unidades institucionales de carácter oficial y público. Además, en la subregión norte y centro se concentraron la mayoría de estas unidades.

Para establecer un tamaño de muestra adecuado, es necesario tener un marco muestral de referencia para definir los estadísticos después de realizar el muestreo; información que en el momento es insuficiente dado que se requeriría de un Censo Nacional o Departamental Económico que permitiera fijar los estándares para la distribución de la muestra. De esta forma, como parte integral del PERS y siguiendo la estructura y desarrollo de experiencias de otros departamentos, se planteó en paralelo la recolección de muestra asociadas a fuentes industriales, institucionales y comerciales, en ese sentido se planeó un muestreo aleatorio estratificado con asignación proporcional. Al contar con poblaciones finitas se asumirá la siguiente estructura:

$$n = \frac{N * Z_{\alpha}^2 * p * q}{e^2 * (N - 1) + Z_{\alpha}^2 * p * q}$$

- n = tamaño de muestra buscado.
- N = Tamaño de la Población objeto de estudio
- Z = Estadístico que define el nivel de confianza deseado basado en la distribución normal estándar. Asumimos varias opciones, al 90% y 95% cuyos valores son 1,64 y 1,96 respectivamente.
- e = Error de estimación máximo tolerado (se asigna entre el 5% y 10%).
- p = probabilidad de ocurrencia de fenómenos estudiados (viviendas en resto sin servicios)
- $q = 1 - p$. Complemento de la probabilidad de ocurrencia.

Asumiendo un $p = 50\%$ (el cual nos permite estimar el máximo valor del tamaño de muestra que se puede obtener cuando no conocemos el verdadero valor de la proporción), se obtienen los resultados mostrados a continuación.⁶

⁶ Es de aclarar que se presentan estimaciones con distintos niveles de confianza y márgenes de error, por tanto con distintos tamaños de muestra para cada tipo de unidades. Esto con el fin que UPME decida de acuerdo con el presupuesto, cuál conviene. A mayor confianza, mayor precisión de estas unidades.

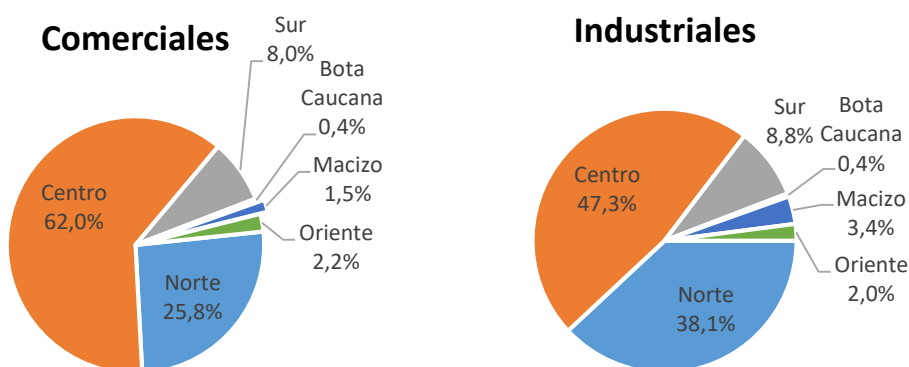
Tabla 5. Universo de unidades no residenciales consideradas para el diseño muestral

Sector	95% ---> Z=1.96		90% ---> Z=1.64		Población (N)
	<i>n</i>	<i>e</i>	<i>n</i>	<i>e</i>	
Comercial	205	0.068	145	0.068	15.815
Industrial	92	0.1	65	0.1	2.141
Oficial	95	0.1	67	0.1	7.129

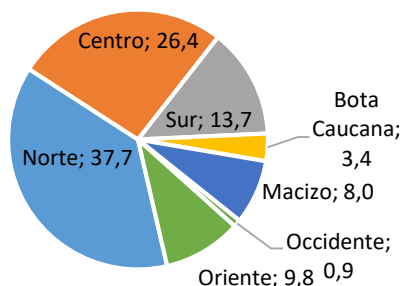
Fuente: Elaboración PERS-CAUCA con base en Sistema Único de Información de Servicios Públicos

Los resultados de la tabla 5 permiten inferir el comportamiento departamental. Sin embargo, para apoyar la planeación del trabajo de campo, en el siguiente análisis se presenta la distribución relativa de estas unidades según el sector (grafica 3). A partir de la información del SUI sobre unidades no residenciales suscritas, 63.5% fueron de naturaleza comercial, 16.1% oficiales y/o institucionales, 11.5% establecimientos de carácter público y 8.9% industrias de gran, mediana y pequeña escala.

Gráfica 3. Departamento del Cauca. Distribución de unidades comerciales e industriales según subregión de localización. 2021



Institucionales



Fuente: Elaboración PERS-CAUCA con base en Sistema Único de Información de Servicios Públicos

Al analizar la distribución de estas unidades según ubicación por subregión, es de notar que en su gran mayoría las unidades comerciales, institucionales e industriales se concentran en las zonas centro y norte. Por lo tanto, la distribución de las unidades no residenciales por subregión es una sugerencia de la distribución a priori de las encuestas de cada sector.

La Tabla 6 presenta la distribución de las encuestas por subregión.

Tabla 6. Asignación de encuestas por subregión

Subregión	Nivel de Confianza 95% ---> Z=1.96			Nivel de Confianza 90% ---> Z=1.64		
	Comercial	Industrial	Oficial	Comercial	Industrial	Oficial
Centro	127	24	45	90	17	32
Norte	53	35	36	37	25	26
Sur	16	13	8	12	9	6
Bota Caucana	1	3	0	1	2	0
Macizo	3	7	3	2	5	2
Oriente	5	9	2	3	6	1
Occidente	0	1	0	0	1	0
Total Encuestas	205	92	95	145	65	67

Fuente: Elaboración PERS-CAUCA

5. Anexo 1. Metodología para la muestra de unidades residenciales

5.1. Introducción

El presente acápite tiene como propósito profundizar en los detalles técnicos relacionados con los procesos de estimación de la muestra y los factores de expansión necesarios para alcanzar representativa muestral subregional, de acuerdo con la localización residencial rural y/o cabecera urbana en el Departamento del Cauca. En tal sentido, este apartado se presenta como una guía metodológica para el trabajo de campo que se desarrollará durante la ejecución del PERS Cauca, tomando como base los instrumentos estadísticos diseñados para el levantamiento de información primaria, como se informó en el capítulo 4⁷.

5.2. Factores de expansión

Como se señaló en el capítulo 4, para el cálculo de los factores de expansión se utilizarán las probabilidades de obtener cada uno de los elementos de la población objeto de estudio, en concordancia con los postulados teóricos derivados del Teorema Bayes. La representación general de la regla de Bayes viene dada por la siguiente expresión:

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{\int_{\theta} P(x|\theta)P(\theta)d\theta}$$

Donde x representa la información disponible relacionada con las unidades objetivo de estudio (viviendas y su acceso a energía eléctrica), tanto a tanto a nivel primario como secundario (información a priori). El parámetro θ captura la probabilidad estimada que la vivienda no cuente con el servicio de energía eléctrica; mientras que la expresión $\int_{\theta} P(x|\theta)P(\theta)d\theta$ representa la probabilidad total para el universo de estudio, es decir, la probabilidad conjunta para las viviendas de la zona rural y/o cabeceras municipales del departamento del Cauca. Finalmente, vale la pena señalar el supuesto información completa desde el enfoque bayesiano, la cual se presenta como condición necesaria para el cálculo de estas probabilidades.

5.3. Proceso de estimación de los factores de expansión para el departamento del Cauca

Para el cálculo de los factores de expansión empleados en el diseño muestral (Tabla 4), se estimaron a priori las probabilidades asociadas a la carencia de energía eléctrica en los hogares del departamento del Cauca, controlando geográficamente

⁷ En el repositorio *GIT-Hub* se almacenan las versiones de los *script* del paquete estadístico R empleados, así como los archivos necesarios para el desarrollo del proyecto, los cuales se pueden consultar en: https://github.com/cristian42253/PERS_Cauca.

a nivel subregiones. A fin de aportar una mayor claridad relacionado con el proceso de su estimación, a continuación se describe el proceso estadístico empleado.

Uno de los factores primordiales en el diseño muestral se encuentra fuertemente ligado a la representatividad que la muestra pueda tener sobre el universo poblacional. En este sentido, el primer paso consiste en garantizar la condición de aleatorización de las unidades de objeto de estudio, para lo cual se asume que la escogencia de cada subregión es equiprobable, es decir, cada una de las viviendas ubicadas al interior de las zonas rurales y/o cabeceras municipales de las subregiones del departamento del Cauca tienen idénticas probabilidad de ser elegidas.

En concordancia con los datos del tamaño muestral definido, se tiene la siguiente representación estadística: $P(Viviendas \cap No \text{ Energía})$; para la cual se requiere un IR superior al 40%, es decir, 2 de cada 5 viviendas ubicadas en zonas rurales y/o cabeceras municipales del departamento del Cauca no cuenten con suministro eléctrico. En este sentido, siguiendo el Teorema de Bayes, tenemos que:

$$P(Viviendas|No \text{ Energía}) = \frac{P(Viviendas \cap No \text{ Energía})}{P(Viviendas)} \quad (1)$$

De la ecuación anterior, es posible derivar la probabilidad de obtener una vivienda que no cuenta con fluido eléctrico, a través de la siguiente expresión:

$$P(Viviendas \cap No \text{ Energía}) = P(Viviendas) * P(Viviendas|No \text{ Energía}) \quad (2)$$

Ahora bien, considerando que la expresión $P(Viviendas)$ en la ecuación (2) representa la probabilidad de obtener una vivienda dentro de cada subregión, esta se obtiene dividiendo el número total de viviendas localizadas al interior de cada subregión entre el total de las subregiones. Por otro lado, la expresión $P(Viviendas|No \text{ Energía})$ se obtiene considerando la proporción equivalente de la submuestra a través de la asignación planteada en el diseño muestral y que satisfacen la condición de ausencia del fluido eléctrico. Al multiplicarse estas dos probabilidades, se obtiene la expresión $P(Viviendas \cap No \text{ Energía})$.

Como la representatividad estadística de la expresión anterior es solo al interior de la subregión, y se requiere para las 7 subregiones; este resultado debe multiplicarse finalmente por la probabilidad de 1/7, independientemente de la subregión analizada, ya que como se mencionó anteriormente, la escogencia de cada una de

las subregiones es equiprobable. Con esto, la representatividad ahora ya no es al interior de la subregión, sino entre subregiones.

A manera de ejemplo, para la subregión centro se tiene que:

$$P(Viviendas_{centro}) = \frac{85.641}{285.643}$$

$$P(Viviendas_{centro} | No Energía_{centro}) = \frac{1}{7,1956}$$

$$P(Viviendas_{centro} \cap No Energía_{centro}) = \frac{85.641}{285.643} * \frac{1}{7,1956} \approx \frac{1}{24}$$

El último paso para lograr la representatividad deseada de la zona centro, es multiplicar por (1/7) el resultado anterior, y por tanto:

$$P(Viviendas_{centro} \cap No Energía_{centro}) = \frac{1}{7} * \frac{1}{24} = \frac{1}{168} = 0,005952$$

Finalmente, el factor de expansión (FE) es obtenido como el inverso de la probabilidad de encontrar una vivienda sin energía en la subregión:

$$FE = \frac{1}{P(Viviendas_{centro} \cap No Energía_{centro})} = \frac{1}{0,005952} = 168$$

Siguiendo el procedimiento anterior para las restantes subregiones, se presentan en la siguiente tabla, los cálculos para los demás factores de expansión:

Tabla 7. Estimación de los factores de expansión por subregión

Subregión	Viviendas por subregión (1)	Total Viviendas en 7 subregiones (2)	P(Viviendas) (3)=(1)/(2)	P(Viviendas No Energía) (4)	P(Viviendas ∩ No Energía) * (5)=(3)*(4)	P(sub región) (6)	P(Viviendas ∩ No Energía) ** (7)=(5)*(6)	Factor de Expansión (8)=1/(7)
Centro	85641	285643	0.2998	1/7.19560	0.0417	1/7	0.0060	168
Norte	81139	285643	0.2841	1/10.7942	0.0263	1/7	0.0038	266
Oriente	21562	285643	0.0755	1/1.55285	0.0486	1/7	0.0069	144
Occidente	19227	285643	0.0673	1/0.25960	0.2593	1/7	0.0370	27
Sur	44748	285643	0.1567	1/4.86396	0.0322	1/7	0.0046	217
Macizo	24403	285643	0.0854	1/2.80710	0.0304	1/7	0.0043	230
Bota	8923	285643	0.0312	1/0.36593	0.0854	1/7	0.0122	82

Fuente: Elaboración PERS-CAUCA

Nota: */ Representatividad al interior de la subregión, **/ Representatividad entre subregiones

Cabe señalar que los factores de expansión obtenidos comprenden el 99.87% de la población total, dado que al considerar la variabilidad de la fuentes de información y el error asociado al muestreo no se alcanza el 100% de la población.

5.4. Obtención Muestra

Siguiendo el muestreo multietápico planteado en el capítulo 4, el cálculo del tamaño muestral se realiza por medio de las siguientes expresiones:

$$n = \frac{\sum_h \sum_L w_{h.L} (B_{h.L} + C_{h.L})}{V(\hat{P}) + \sum_h \sum_L w_{h.L}^2 \left(\frac{B_{h.L}}{N_{h.L}} \right)} \quad (3)$$

$$B_{h.L} = \left(\frac{S_{UPM(h)}^2}{\bar{X}_{0h}^2} \right) \quad (4)$$

$$C_{h.L} = \left(\frac{\sum_{i \in N_h} M_{hi} (1 - f_{2h}) S_{UPM(h)}^2}{M_h} \right) \quad (5)$$

Para el cálculo del tamaño muestral, se debe considerar que los lados de manzanas (M_{hi}) se asumen constantes e iguales a 0,02381 debido al desconocimiento de las estructuras municipales y subregionales; lo cual trae como consecuencia que el factor M_h también sea idéntico a la sumatoria de los M_{hi} , es decir, iguales a 7. Así mismo, la expresión $(1 - f_{2h})$, se asume proporcional y equivalente a 0,5.

El peso del conglomerado dentro de la subregión ($w_{h.L}$) se plantea equidistribuido con valor de $1/7$ y, con respecto a los municipios, será la proporción del peso que estos presentan con respecto al total de unidades residenciales. La varianza asociada al error muestral, $V(\hat{P})$, es igual a 0,003718.

Finalmente B_h es equivalente a la razón entre la varianza global y el promedio global, ya que se considera la suma desagregada por estrato y conglomerado. Los resultados globales de la estimación de la medida y la varianza para la muestra se presentaron en la Tabla 3; sin embargo, el desarrollo estadístico del diseño muestral se relaciona a continuación:

$$\sum_h \sum_L w_{h.L} = 0,142989$$

$$\sum_h \sum_L \left(\frac{w_{h.L}^2}{N_{h.L}} \right) = 225,7896$$

$$\sum_h \sum_L M_{h.L} = 22,3886$$

$$\sum_h \sum_L w_{h.L} * \sum_{i \in N_h} M_{hi} = 6582,254$$

$$M_{hi} = 0,02381$$

$$V(\hat{P}) = \left(\frac{EM}{Z} \right)^2 = \left(\frac{0,1}{1,64} \right)^2 = 0,003718$$

$$B_{hL} = \left(\frac{S_{UPM(h)}^2}{\overline{X_{0h}}^2} \right) = \frac{16,5581899}{(84,0965169)^2} = 0,002341299$$

$$C_{hL} = \left(\frac{\sum_{i \in N_h} M_{hi}(1-f_{2h})S_{UPM(h)}^2}{M_h} \right) = 7785,015144$$

Para el desarrollo de cada una de las expresiones, se consideró las expansiones de las sumatorias y coneciones algebraicas que permiten resumir cada una de las estructuras, para seguir los resultados se sugiere consultar el archivo de soporte adjunto así, como la referencias bibliógraficas utilizadas.

$$\sum_h \sum_L w_{h.L} (B_{h.L} + C_{h.L}) = 1113,173994$$

$$N_h \left\{ \begin{array}{l} N_{centro} = 85641 \\ N_{norte} = 81139 \\ N_{oriente} = 21562 \\ N_{occidente} = 19227 \\ N_{sur} = 44748 \\ N_{macizo} = 24403 \\ N_{Bota\ Caucana} = 8923 \end{array} \right. , \quad W_h \left\{ \begin{array}{l} W_{centro} = \frac{1}{7} * \frac{85641}{285643} \\ W_{norte} = \frac{1}{7} * \frac{81139}{285643} \\ W_{oriente} = \frac{1}{7} * \frac{21562}{285643} \\ W_{occidente} = \frac{1}{7} * \frac{19227}{285643} \\ W_{sur} = \frac{1}{7} * \frac{44748}{285643} \\ W_{macizo} = \frac{1}{7} * \frac{24403}{285643} \\ W_{Bota\ Caucana} = \frac{1}{7} * \frac{8923}{285643} \end{array} \right.$$

$$V(\hat{P}) + \sum_h \sum_L w_{h.L}^2 \left(\frac{B_{h.L}}{N_{h.L}} \right) = 0,003718 + (0,002341299 * 225,7896) = 0,532359$$

Reemplazando los resultados anteriores se concluye,

$$n = \frac{\sum_h \sum_L w_{h.L} (B_{h.L} + C_{h.L})}{V(\hat{P}) + \sum_h \sum_L w_{h.L}^2 \left(\frac{B_{h.L}}{N_{h.L}} \right)} = \frac{1113,1739940}{0,532359} = 2091,0564$$

Vale la pena señalar que, para el cálculo del tamaño de la muestra bajo la estructura del muestreo por etapas usando conglomerados y estratos, se busca tener una estimación del promedio y la varianza total que sea representativa de la población. Asimismo, cuando se desagrega la estructura de los estadísticos $B_{h.L}$ y $C_{h.L}$ se pueden considerar como una media y varianza global. Adicionalmente, se debe tener en cuenta que al tener estructuras globales de la media y la varianza el muestreo por etapas, será equivalente asintóticamente al cálculo de una muestra simple, para lo cual se realizó una prueba de potencia para el tamaño muestral (*power.t.test*), permitiendo definir estadísticamente el tamaño de muestra uniformemente más potente⁸. Estos resultados fueron validados mediante una prueba de hipótesis a dos colas para validar diferencia estadística entre las poblaciones analizadas.

⁸ Para el lector interesado consultar Teorema 7.5 (p.91) en Latpate, *et al.* (2021).

Finalmente en la imagen 1, se presenta el código de programación utilizado en el cálculo y obtención de la muestra representativa⁹.

Imagen 1. Código de programación en R para obtención de la muestra

```
## packages necesaries
packages = c("tidyverse", "pwr",
             "Sample.Size", "sampling", "samplesize")

## Now load or install&load all
package.check <- lapply(
  packages,
  FUN = function(x) {
    if (!require(x, character.only = TRUE)) {
      install.packages(x, dependencies = TRUE)
      library(x, character.only = TRUE)
    }
  }
)

# Load packages
load_pkg <- rlang::quos(tidyverse, pwr, Sample.Size, sampling, samplesize)

invisible(lapply(lapply(load_pkg, rlang::quo_name),
                    library,
                    character.only = TRUE
                ))

# Calculate sampling
delta <- 1.45
sigma <- 14.4589
d <- delta/sigma
pwr.t.test(d=d, sig.level=.05, power = .90, type = 'two.sample')
```

Referencias

Latpate, R., Kshirsagar, J., Gupta, V. K., & Chandra, G. (2021). *Advanced Sampling Methods*. Springer Singapore.

⁹ En la obtención de la muestra es importante considerar que de acuerdo con el teorema del límite central, para tamaños de n lo suficientemente grandes la prueba t y los valores z de la distribución normal son equivalentes. Para el lector interesado, consultar Latpate, *et al.* (2021).