

CONVENIO INTERINSTITUCIONAL UNIDAD DE PLANEACIÓN MINERO ENERGÉTICA – UPME - CV-003/15, INSTITUTO DE PLANIFICACION Y PROMOCION DE SOLUCIONES ENERGETICAS PARA LAS ZONAS NO INTERCONECTADAS – IPSE -005/2015 Y LA UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS – UDFJC PARA LA REALIZACIÓN DEL PLAN DE ENERGIZACIÓN RURAL SOSTENIBLE PARA EL DEPARTAMENTO DEL CUNDINAMARCA – PERS CUNDINAMARCA

Plan de Energización rural sostenible del departamento de Cundinamarca PERS

Metodología para la selección de la muestra y
los municipios

V 2.0

BOGOTÁ D.C, 17 DE OCTUBRE DE 2015

Universidad Distrital Francisco José de Caldas. Carrera 8 No. 40-62. Sede Edificio Suárez Copete 3er Piso.
Unidad de Extensión Facultad de Ingeniería.
Teléfono: 3239300 Ext: 1711
E mail: perscundinamarca@udistrital.edu.co

CONVENIO INTERINSTITUCIONAL UNIDAD DE PLANEACIÓN MINERO ENERGÉTICA – UPME - CV-003/15, INSTITUTO DE PLANIFICACION Y PROMOCION DE SOLUCIONES ENERGETICAS PARA LAS ZONAS NO INTERCONECTADAS – IPSE -005/2015 Y LA UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS – UDFJC PARA LA REALIZACIÓN DEL PLAN DE ENERGIZACIÓN RURAL SOSTENIBLE PARA EL DEPARTAMENTO DEL CUNDINAMARCA – PERS CUNDINAMARCA

Metodología para la selección de la muestra y los municipios

Preparado por:

Claudia María Cardona Londoño

Revisado por:

Francisco Santamaría Piedrahita

BOGOTÁ D.C, 17 DE OCTUBRE DE 2015

Índice General

1. Introducción	5
2. Descripción del Muestreo	6
2.1 Conceptos básicos de muestreo	6
2.2 Tamaño de la Muestra	7
3. Etapa inicial	8
3.1 Software	8
3.2 Objetivos	9
3.3 Información Inicial	9
3.3.1 Ubicación del área objeto de estudio	9
3.3.2 Índice de Ruralidad (IR)	12
3.3.3 Índice de Electrificación (IER)	12
3.3.4 Índice del Complemento de Electrificación Rural (INER)	12
3.3.5 Población objetivo	16
4. Etapa Descriptiva	20
4.1 Técnicas de muestreo	20
4.2 Error Estadístico y niveles de confianza	21
5. Selección de la Muestra	23
6. Depuración de la Muestra	28
Anexo 1: Ejecución de EPIDAT	34

Índice de Figuras

<i>Figura 1. Mapa de la Ubicación Espacial de municipios en el Departamento de Cundinamarca</i>	<i>9</i>
<i>Figura 2. Distribución del número de encuestas por provincia</i>	<i>27</i>

Índice de Tablas

<i>Tabla 1. Índices IR, IER, INER para los 116 municipios de Cundinamarca</i>	<i>12</i>
<i>Tabla 2. Información detallada de Viviendas por Provincia y Municipios para 2015. Fuente: SUJ ..</i>	<i>16</i>
<i>Tabla 3. Resumen por Provincias – Viviendas y Participación</i>	<i>19</i>
<i>Tabla 4. Valores de Z según el valor de significación</i>	<i>22</i>
<i>Tabla 5. Nivel de confianza, prevalencia y precisión para la selección del tamaño muestral</i>	<i>22</i>
<i>Tabla 6. Valores máximos, medios y mínimos de la muestra</i>	<i>22</i>
<i>Tabla 7. Número Total de Encuestas y Distribución por Municipio</i>	<i>25</i>
<i>Tabla 8. Distribución de encuestas por provincia y municipio</i>	<i>29</i>
<i>Tabla 9. Distribución de encuestas por provincia</i>	<i>32</i>

1. Introducción

Para el establecimiento del PERS-Cundinamarca se tienen en cuenta aspectos como ruralidad, el número de viviendas y el consumo de energía eléctrica en cada conglomerado. El muestreo seleccionado toma a las provincias de Cundinamarca como estratos y a los municipios como conglomerados. La ruralidad se tiene en cuenta para seleccionar de manera obligatoria a los municipios que tengan un alto índice de ruralidad.

La descripción de la metodología de trabajo de campo y recolección de información por medio de encuestas en las diferentes provincias, son la base fundamental para proponer un protocolo de captura de información en campo garantizando la independencia y aleatoriedad de los resultados y poder procesar los datos con la mayor precisión.

Cuando se realiza una investigación a una población para extraer cierto tipo de información se tiene que garantizar la representación fiel a todo el universo estudiado, garantizando características de la población muestreada como representatividad aleatoriedad e independencia. La encuesta es el instrumento de recolección de información que contiene el conjunto de actividades orientadas a recolectar y resumir información acerca de las características de interés de algunas o todas las unidades de una población, usando conceptos, métodos y procedimientos bien definidos y correctamente establecidos.

Los principios de aleatoriedad representatividad e independencia junto con un nivel de confianza permiten evitar sesgos para de esta forma tener información fiel al universo, permitiendo que los análisis generados a partir de esta información tengan robustez estadística y científica.

Siendo la estadística una disciplina transversal a todas las ciencias, el desarrollo de la estadística por medio de computadoras ha facilitado el manejo de la misma, de esta manera se presentan a hoy en día una gran diversidad de software estadísticos, de esta manera se adopta el software EPIDAT que tiene muchas ventajas con respecto a los demás software estadísticos.

2. Descripción del Muestreo

2.1 Conceptos básicos de muestreo

El muestreo es un procedimiento estadístico utilizado en estudios de investigación científica, investigación económica e investigación de mercados que tiene por objeto garantizar mediciones de comportamiento de una población a través de una fracción de la misma. Además de conocer y observar el comportamiento de una población, el muestreo contribuye a resolver dos dimensiones clave en el éxito de una investigación: el tiempo y los recursos financieros. Con los mismos resultados es preferible realizar la recolección o encuesta a una muestra y no realizar un censo, toda vez que un censo implica observar a todos los elementos de la población convirtiendo el proceso en una actividad que demanda más tiempo y por supuesto exige un mayor presupuesto.

Para que una muestra sea representativa, y por lo tanto útil, debe reflejar las semejanzas y diferencias encontradas en la población, ejemplificar las características y tendencias de la misma. Cuando se dice que una muestra es representativa indica que la muestra reúne aproximadamente todas las características de la población que son importantes para la investigación. Cuando se dice que una muestra es adecuada, se hace referencia a que contiene el número suficiente de unidades de estudio de tal manera que permita aplicar pruebas estadísticas que den validez a la inferencia de los resultados a la población.

Se ha demostrado que una muestra representativa ofrece una muy buena aproximación de los resultados de la observación de experimentos, eventos o medición de fenómenos o características de una población y para que esto ocurra, el muestreo estadístico exige varias condiciones que se deben cumplir de manera rigurosa para que los resultados de la encuesta se constituyan también en el comportamiento de la población. En todo proceso de muestreo, los elementos de la muestra deben escogerse adecuadamente, de tal manera que los resultados que se observen en ella, puedan inferirse, expandirse o generalizarse al resto de la población. Para ello debe tenerse en cuenta:

- a) Tamaño de la muestra, con una característica de muestra adecuada. El tamaño de la muestra depende de la homogeneidad de la población.
- b) Condiciones de selección de la misma. Se debe garantizar que todos los miembros de la población tengan la misma probabilidad de ser seleccionados como parte de la muestra; esta es la característica de una muestra representativa.

Estas dos características darán valor estadístico a los resultados y permitirán la inferencia a la población. Es decir, los resultados permitirán referirse a toda la población y no a solo la fracción de la muestra.

2.2 Tamaño de la Muestra

A la hora de determinar el tamaño que debe alcanzar una muestra hay que tomar en cuenta varios factores: el tipo de muestreo, el parámetro por estimar, el error muestral admisible, la varianza poblacional y el nivel de confianza. Para facilitar la explicación a continuación se presenta el concepto y definición de estos factores:

Parámetros: Son las medidas o datos que se obtienen sobre la población.

Estadístico: Los datos o medidas que se obtienen sobre una muestra y por lo tanto una estimación de los parámetros.

Error muestral o error estadístico: Es una medida de la variabilidad de las estimaciones de muestras repetidas en torno al valor de la población y nos proporciona una noción clara de hasta dónde y con qué probabilidad una estimación basada en una muestra se aleja del valor que se hubiera obtenido por medio de un censo completo. Siempre se comete un error, pero la naturaleza de la investigación nos indicará hasta qué medida podemos cometerlo (los resultados se someten a error muestral e intervalos de confianza que varían muestra a muestra). El error varía según se calcule al principio o al final. Un estadístico será más preciso en cuanto y tanto su error es más pequeño. Podríamos decir que es la desviación de la distribución muestral de un estadístico y su confiabilidad.

Nivel de Confianza: Probabilidad de que la estimación efectuada se ajuste a la realidad. Cualquier información que queremos recoger está distribuida según una ley de probabilidad (Gauss o Student), así llamamos nivel de confianza a la probabilidad de que el intervalo construido en torno a un estadístico capte el verdadero valor del parámetro.

Varianza poblacional: Generalmente es un valor desconocido y hay que estimarlo a partir de datos de estudios previos. Teniendo en cuenta estos factores, la fórmula para calcular el tamaño de muestra dependerá si se estima la media de la población o la proporción de una variable. Normalmente se trabaja con errores muestrales o estadísticos entre el 3% y 8% y nivel de confianza mayor o igual a 95%.

3. Etapa inicial

3.1 Software

Siendo la estadística una ciencia transversal a todas las ciencias, de acuerdo a los avances tecnológicos, el desarrollo de la estadística por medio de computadoras ha facilitado el manejo de la misma, de esta manera se presentan a hoy en día una gran diversidad de software estadístico, muchas veces especializado para campo que se trabaje a nivel estadístico, de esta manera se adopta el software Epidat 4.0 que tiene muchas ventajas con respecto a los demás software estadísticos.

Una de las más grandes ventajas es que es un software libre, de esta manera se evita el pago de licencias, con ello también se tiene una plataforma de constante actualización y alimentación con el mundo científico a través de la implementación de nuevos paquetes estadísticos que alimentan al programa, otra ventaja que tiene el Epidat 4.0, es que es un programa para la gestión y análisis de datos, utilizado en ciencias de la salud para definir con alta precisión el tamaño de una muestra, con información de importancia socio-cultural y base para el desarrollo y la gestión de calidad en la prestación de un servicio, y pudiendo comparar el componente de salud cómo un servicio muy similar a la prestación de un servicio de energización, la técnica utilizada por el programa en su caja de herramientas es propia en la etapa del diseño muestral y selección de la muestra representativa para el estudio.

El Epidat 4.0 mantiene la filosofía de estar orientado preferentemente al manejo de datos tabulados; sin embargo, en la mayoría de las opciones del programa es posible cargar los datos a partir de una tabla en formato Excel, Access u Open Office. Para algunos procedimientos, además, el programa opera con bases de datos individuales (no solo con tabulados). En esta nueva versión se ha migrado del lenguaje de programación Delphi al Java, debido a la versatilidad de este último, que permite el funcionamiento de la aplicación en distintos sistemas operativos, tales como Windows, Linux y Macintosh. Un cambio importante es que el Epidat, tiene articulación modular, de manera que, a partir de un entorno general, se podrán añadir o modificar en forma independiente distintos módulos, lo cual facilitará las actualizaciones del programa

Es que al poder tener acceso al lenguaje de programación de manera directa se pueden establecer modificaciones y atributos acorde a las necesidades y hasta caprichos de los investigadores, de esta manera se tienen mejoras gráficas, siendo así el EPIDAT, se selecciona como el programa para tratamiento estadístico en el PERS (Programa de Energización Rural Sostenible) Cundinamarca.

3.2 Objetivos

- Establecer directrices para la estandarización de datos obtenidos para el proyecto PERS-Cundinamarca
- Seleccionar la muestra representativa, explicando la selección
- Reducir tiempos en depuración de datos para realizar análisis.

3.3 Información Inicial

3.3.1 Ubicación del área objeto de estudio

La Ubicación espacial del estudio, presenta al Departamento de Cundinamarca con 15 provincias y 116 municipios que esquematizan en el mapa de la Figura 1, y a continuación se presenta la descripción de cada provincia y municipios que la conforman.



Figura 1. Mapa de la Ubicación Espacial de municipios en el Departamento de Cundinamarca

1. **Almeidas** está ubicada al nororiente del departamento de Cundinamarca. Limita al oriente con el departamento de Boyacá, al sur con la provincia del Guavio, al occidente con la provincia de Sabana Centro y al norte con la provincia de Sabana Occidente. La provincia está compuesta por los municipios de Suesca, Sesquilé, Chocontá (cabecera de provincia), Villapinzón, Machetá, Tibirita y Manta, tiene una población oficial a 2005 de 72.891 habitantes.
2. **Alto Magdalena**, al norte limita con la provincia de Magdalena Centro, al sur con el río Magdalena y el departamento del Tolima, al oriente con las provincias del Tequendama y Sumapaz y al Occidente con el río Magdalena y el departamento del Tolima. La provincia está compuesta por los Municipios de Agua de Dios, Girardot, Guataquí, Jerusalén, Nariño, Nilo, Ricaurte y Tocaima. La población oficial al 2005 para esta provincia es de 152.271 habitantes y una extensión de 4045 km².
3. La provincia de **Bajo Magdalena** está ubicada al noroccidente del departamento de Cundinamarca. Limita al oriente con las provincias de Rionegro y la Gualivá, al sur con la provincia de Magdalena Centro, al occidente con Caldas y Tolima, al norte con Boyacá. La provincia está compuesta por los municipios de Caparrapí, Guaduas y Puerto Salgar. Con una población oficial a 2005 de 60.275 habitantes.
4. **Gualivá**, limita al Norte con la provincia de Rio-Negro, al Este la provincia de Bajo Magdalena, al sur con las provincias de Magdalena Centro y Tequendama, al Oeste limita con la provincia de Sabana Occidente. Está conformada por los municipios de Albán, La Peña, La Vega, Nimaima, Nocaima, Quebradanegra, San Francisco, Sasaima, Supatá, Útica, Vergara y Villeta, ésta última declarada su capital. Su población a 2005 es de 99.108 habitantes.
5. La Provincia del **Guavio** limita al Norte con Sabana Centro y Almeidas, al Sur con la Provincia de Oriente y Medina, al Oeste con Bogotá D.C., al Este con Boyacá. Está compuesta por ocho municipios: Guatavita, Guasca, Gachetá, La Calera, Junín, Gama, Ubalá y Gachalá. Su población a 2005 es de 80.597 habitantes.
6. **Magdalena Centro** está ubicada al occidente del departamento de Cundinamarca. Limita al oriente con las provincias de Rionegro y Gualivá, al sur con la Provincia del Alto Magdalena, al occidente con el departamento del Tolima y al norte con la Provincia del Bajo Magdalena. Conformada por Beltrán, Bituima, Chaguani, Guayabal de Siquima, Pulí, San Juan de Rioseco y Viani. Su población a 2005 es de 28.121 habitantes.
7. La provincia de **Medina**, limita al Norte con la Provincia de Neira y el departamento de Boyacá, al Sur con el Departamento del Meta, al Oeste con la

- Provincia del Guavio, al este con el Departamentos del Meta y Casanare, los municipios son Paratebueno y Medina. Su población a 2005 es de 16.740 habitantes.
8. La provincia de **Oriente** limita al Norte con la provincia de Guavio, al este con el Distrito Capital, y al oeste con la provincia de Medina. Con una población de 16.509 a 2005.
 9. **Rio Negro** limita al sur, con la provincia de Gualivá, al occidente con la provincia del Bajo Magdalena, al norte con Boyacá, al oriente con la provincia de Ubaté. Está compuesta por los municipios de El Peñón, La Palma, Pacho (capital provincial), Paime, San Cayetano, Topaipí, Villagómez y Yacopí. Su población a 2005 es de 71.927 habitantes.
 10. **Sabana Centro** limita al norte con la Provincia de Ubaté, al Sur con el Distrito Capital, al Oeste con las Provincias de Sabana Occidente y la Rionegro, al Este con las Provincias del Guavio y la Almeidas. Está compuesta por once Municipios: Cajicá, Chía, Cogua, Cota, Gachancipá, Nemocón, Sopó, Tabio, Tenjo, Tocancipá y Zipaquirá. Su población a 2005 es de 347.890 habitantes.
 11. La provincia de **Sabana Occidente** limita al Norte con la provincia de Rio negro, al sur con la provincia de Soacha, al Este con las provincias de Gualivá y Tequendama, al Oeste con las provincias de Sabana Centro y Distrito Capital. Está compuesta por 8 municipios: Bojacá, El Rosal, Facatativá, Funza, Madrid, Mosquera, Subachoque y Zipacón. Su población a 2005 es de 72.891 habitantes.
 12. **Soacha** limita al Norte con Sabana Occidente, al Sur con la provincia de Sumapaz, al Este con la provincia de Tequendendama, al Oeste con el Distrito Capital. Está conformada por los municipios de Soacha y Sibaté, tiene la categoría de provincia, por sus características especiales. Su población a 2005 es de 429.461 habitantes.
 13. **Sumapaz** limita al Norte con la provincia de parte media y baja de la cuenca del río Bogotá, al Oriente con Parque Nacional Natural Sumapaz, al Sur con el Departamento del Huila, al Occidente con el Departamento del Tolima. Está compuesta por 10 municipios. Su capital es Fusagasugá, sus otros municipios son: Sylvania, Tibacuy, Pasca, Arbeláez, Pandí, San Bernardo, Venecia, Cabrera y Granada. Su población a 2005 es de 185.453 habitantes.
 14. La provincia de **Tequendama** limita al Norte con la provincia de Gualivá, al Sur con Sumapaz, al Este con las provincias de Magdalena Centro y Bajo Magdalena, al Oeste con las provincias de Soacha y Sabana Occidente. Está compuesta por los municipios de Anapoima, Anolaima, Cachipay, El Colegio, La Mesa, Quipile, Apulo, San Antonio del Tequendama, Tena y Viotá. Su población a 2005 es de 126.968 habitantes.

15. **Ubaté** limita al norte, noroeste y noreste con el departamento de Boyacá, al occidente con la provincia de Rionegro, al sur con la provincia de Sabana Centro y al sureste y oriente con la provincia de Almeidas. Está compuesta por los municipios de Carmen de Carupa, Cucunubá, Fúquene, Guachetá, Lenguazaque, Simijaca, Susa, Sutatausa, Tausa y Ubaté. Su población a 2005 es de 106.325 habitantes.

3.3.2 Índice de Ruralidad (IR)

Aunque el Índice de Ruralidad (IR) es una medida continua entre 0 y 100 (entre más alto, más rural es considerado el municipio), para efectos de análisis se fija una línea de separación entre los municipios considerados rurales (IR > 40%) y los municipios no rurales (IR < 40%). Los ejercicios estadísticos basados en análisis de clúster, indican que IR = 40% es el punto que maximiza la varianza intergrupar y que minimiza la varianza intragrupos.

3.3.3 Índice de Electrificación (IER)

El índice de electrificación rural (IER) se refiere al porcentaje de viviendas rurales que poseen servicio de energía eléctrica, es decir que es una relación entre el número de abonados al servicio de energía eléctrica y el total de viviendas de una población.

3.3.4 Índice del Complemento de Electrificación Rural (INER)

Con el índice de electrificación rural de cada municipio se calculó la diferencia al cien por ciento, encontrando el porcentaje de viviendas rurales por municipio que no cuentan con el servicio de energía eléctrica. A este número lo denominamos el índice del complemento de electrificación rural (INER).

El IR y el INER son la base de la selección de la muestra. En la Tabla 1 se presentan los tres índices (IR, IER e INER) para los 116 municipios de Cundinamarca. Se resaltan los tres municipios con IR > 40%.

Tabla 1. Índices IR, IER, INER para los 116 municipios de Cundinamarca

Código DANE	Municipio	IR (%)	IER (%)	INER (%)
25001	Agua de Dios	25,47	94,20	5,80
25019	Albán	25,85	93,10	6,90
25035	Anapoima	27,18	93,30	6,70
25040	Anolaima	26,26	91,80	8,20
25053	Arbelaez	27,45	94,40	5,60
25086	Beltran	39,55	89,30	10,70
25095	Bituima	31,85	89,10	10,90



25099	Bojacá	26,88	95,70	4,30
25120	Cabrera	39,50	96,00	4,00
25123	Cachipay	23,29	91,20	8,80
25126	Cajicá	14,89	91,20	8,80
25148	Caparrapí	36,36	99,40	0,60
25151	Cáqueza	23,91	88,50	11,50
25154	Carmen De Carupa	35,11	95,40	4,60
25168	Chaguaní	35,33	96,70	3,30
25175	Chía	12,97	84,30	15,70
25178	Chipaque	28,66	99,50	0,50
25181	Choachi	30,00	95,20	4,80
25183	Chocontá	30,20	94,00	6,00
25200	Cogua	25,63	99,30	0,70
25214	Cota	19,60	99,40	0,60
25224	Cucunubá	30,55	97,30	2,70
25245	El Colegio	23,48	97,10	2,90
25258	El Peñón	33,40	93,80	6,20
25260	El Rosal	24,47	93,80	6,20
25269	Facatativá	15,90	99,20	0,80
25279	Fomeque	34,02	95,60	4,40
25281	Fosca	28,83	86,80	13,20
25286	Funza	14,37	92,30	7,70
25288	Fúquene	31,05	97,20	2,80
25290	Fusagasugá	16,77	98,60	1,40
25293	Gachala	37,85	97,50	2,50
25295	Gachancipa	22,04	98,60	1,40
25297	Gachetá	32,56	99,00	1,00
25299	Gama	33,01	95,20	4,80
25307	Girardot	16,11	91,80	8,20
25312	Granada	25,32	98,00	2,00
25317	Guachetá	30,50	98,90	1,10
25320	Guaduas	32,71	98,90	1,10
25322	Guasca	32,74	94,00	6,00
25324	Guataqui	34,63	98,70	1,30
25326	Guatavita	35,00	88,80	11,20
25328	Guayabal de Siquima	30,13	96,50	3,50
25335	Guayabetal	34,72	89,80	10,20
25339	Gutierrez	40,07	89,80	10,20
25368	Jerusalen	38,72	89,30	10,70
25372	Junín	34,77	89,20	10,80
25377	La Calera	28,43	95,90	4,10
25386	La Mesa	23,51	98,50	1,50
25394	La Palma	31,85	94,50	5,50



25398	La Peña	31,13	90,10	9,90
25402	La Vega	28,19	95,80	4,20
25407	Lenguazaque	30,66	96,40	3,60
25426	Macheta	34,47	92,60	7,40
25430	Madrid	17,49	99,90	0,10
25436	Manta	32,70	94,10	5,90
25438	Medina	41,37	67,30	32,70
25473	Mosquera	16,19	95,50	4,50
25483	Nariño	32,51	92,70	7,30
25486	Nemocón	26,87	97,20	2,80
25488	Nilo	32,27	92,80	7,20
25489	Nimaima	27,99	95,20	4,80
25491	Nocaima	28,79	86,60	13,40
25506	Ospina Pérez (Venecia)	32,57	86,90	13,10
25513	Pacho	30,34	90,50	9,50
25518	Paime	34,60	78,90	21,10
25524	Pandi	28,73	94,00	6,00
25530	Paratebueno	41,91	71,60	28,40
25535	Pasca	29,96	93,30	6,70
25572	Puerto Salgar	35,24	75,50	24,50
25580	Puli	37,51	83,70	16,30
25592	Quebradanegra	30,47	93,00	7,00
25594	Quetame	30,04	88,40	11,60
25596	Quipile	29,42	89,20	10,80
25599	Apulo	29,16	95,90	4,10
25612	Ricaurte	29,56	95,90	4,10
25645	San Antonio de Tequendama	25,61	97,30	2,70
25649	San Bernardo	30,92	92,90	7,10
25653	San Cayetano	37,54	94,80	5,20
25658	San Francisco	29,19	90,10	9,90
25662	San Juan de Rioseco	33,70	89,40	10,60
25718	Sasaima	27,89	94,60	5,40
25736	Sesquilé	29,41	98,00	2,00
25740	Sibaté	20,67	97,40	2,60
25743	Silvania	24,44	96,00	4,00
25745	Simijaca	28,03	97,70	2,30
25754	Soacha	8,49	98,90	1,10
25758	Sopo	23,10	99,10	0,90
25769	Subachoque	29,60	99,30	0,70
25772	Suesca	28,78	97,40	2,60
25777	Supata	32,69	91,40	8,60
25779	Susa	28,59	97,20	2,80
25781	Sutatausa	29,87	91,50	8,50

25785	Tabio	21,29	98,80	1,20
25793	Tausa	32,78	98,00	2,00
25797	Tena	24,37	88,40	11,60
25799	Tenjo	24,49	99,30	0,70
25805	Tibacuy	29,55	88,80	11,20
25807	Tibirita	31,41	94,80	5,20
25815	Tocaima	29,17	95,60	4,40
25817	Tocancipa	20,23	99,60	0,40
25823	Topaipí	34,49	85,30	14,70
25839	Ubalá	33,10	93,10	6,90
25841	Ubaque	28,58	94,30	5,70
25843	Ubate	21,46	96,30	3,70
25845	Une	30,77	91,30	8,70
25851	Utica	31,60	87,90	12,10
25862	Vergara	31,25	87,00	13,00
25867	Viani	29,98	90,10	9,90
25871	Villagomez	34,18	88,80	11,20
25873	Villapinzón	29,87	88,80	11,20
25875	Villeta	24,41	93,20	6,80
25878	Viota	28,72	97,10	2,90
25885	Yacopí	38,50	66,90	33,10
25898	Zipacón	27,30	92,50	7,50
25899	Zipaquirá	18,26	99,20	0,80

A partir de los valores de INER para cada municipio, se obtuvo el valor promedio para el departamento de Cundinamarca mediante la suma de los índices de los complementos de electrificación rural sobre el total de municipios (116), encontrando un promedio de 7,04%.

Teniendo en cuenta lo anterior, se propone un algoritmo de selección que involucra estos dos índices, así:

Si el índice de ruralidad es mayor al 40% “y” el complemento del índice de electrificación rural es mayor al 7,04%, el municipio es aceptado en la muestra, de lo contrario el municipio se rechaza.

Este algoritmo tiene dos condiciones: i) Todos los municipios con IR < 40% deben ser incluidos en la muestra, ii) La muestra debe contener al menos un municipio por cada provincia del departamento.

3.3.5 Población objetivo

En el Marco PERS Cundinamarca se tiene como información base para realizar la inferencia estadística, la información sobre población y número de viviendas. La Tabla 2 presenta la información obtenida del Sistema Único de Información de Servicios Públicos (SUI), con los datos de viviendas totales, de la cabecera y el resto para 2015, para cada municipio en su respectiva provincia.

Tabla 2. Información detallada de Viviendas por Provincia y Municipios para 2015. Fuente: SUI

Código DANE	Provincia	Municipio	Viviendas		
			Total	Cabecera	Resto
25183	Almeidas	Chocontá	5448	2543	2874
25426	Almeidas	Macheta	2332	390	1919
25436	Almeidas	Manta	1857	466	1383
25736	Almeidas	Sesquilé	3203	741	2460
25772	Almeidas	Suesca	4207	1375	2715
25807	Almeidas	Tibirita	1006	177	828
25873	Almeidas	Villapinzón	4813	1419	3388
25001	Alto Magdalena	Agua de Dios	3414	2705	710
25307	Alto Magdalena	Girardot	28427	27526	903
25324	Alto Magdalena	Guataquí	626	316	309
25368	Alto Magdalena	Jerusalén	843	184	659
25483	Alto Magdalena	Nariño	619	455	165
25488	Alto Magdalena	Nilo	2964	1224	1815
25612	Alto Magdalena	Ricaurte	4361	2790	1696
25815	Alto Magdalena	Tocaima	5063	3152	1919
25148	Bajo Magdalena	Caparrapí	4673	809	3868
25320	Bajo Magdalena	Guaduas	10637	4874	5698
25572	Bajo Magdalena	Puerto Salgar	5463	4081	1378
25019	Gualivá	Albán	1919	412	1505
25398	Gualivá	La Peña	2139	280	1859
25402	Gualivá	La Vega	5121	1796	3314
25489	Gualivá	Nimaima	1731	555	1153
25491	Gualivá	Nocaima	2097	624	1468
25592	Gualivá	Quebradanegra	1690	113	1575
25658	Gualivá	San Francisco	2872	875	1994
25718	Gualivá	Sasaima	3614	740	2870
25777	Gualivá	Supatá	1725	478	1244
25851	Gualivá	Útica	1463	821	644
25862	Gualivá	Vergara	2691	468	2209

25875	Gualivá	Villeta	7762	4733	2993
25293	Guavio	Gachala	1508	464	1034
25297	Guavio	Gachetá	3050	854	2183
25299	Guavio	Gama	1104	229	874
25322	Guavio	Guasca	3681	1307	2375
25326	Guavio	Guatavita	1816	601	1223
25372	Guavio	Junín	2852	301	2545
25377	Guavio	La Calera	9177	3156	5929
25839	Guavio	Ubalá	3018	262	2740
25086	Magdalena Medio	Beltrán	632	176	462
25095	Magdalena Medio	Bituima	796	143	653
25168	Magdalena Medio	Chaguaní	1163	212	951
25328	Magdalena Medio	Guayabal de Siquima	1208	272	937
25580	Magdalena Medio	Pulí	913	225	691
25662	Magdalena Medio	San Juan de Río Seco	3110	874	2233
25867	Magdalena Medio	Vianí	1065	314	749
25438	Medina	Medina	2788	1083	1706
25530	Medina	Paratebuena	1957	532	1415
25151	Oriente	Caqueza	4708	1771	2901
25178	Oriente	Chipaque	2106	621	1483
25181	Oriente	Choachí	2784	818	1951
25279	Oriente	Fomeque	4039	1288	2696
25281	Oriente	Fosca	1835	433	1397
25335	Oriente	Guayabetal	1258	424	838
25339	Oriente	Gutiérrez	1062	327	739
25594	Oriente	Quetame	1635	303	1324
25841	Oriente	Ubaque	1719	236	1482
25845	Oriente	Une	2349	949	1373
25258	Río Negro	El Peñón	1623	134	1488
25394	Río Negro	La Palma	3496	1192	2317
25513	Río Negro	Pacho	7506	3746	3695
25518	Río Negro	Paime	1433	124	1307
25653	Río Negro	San Cayetano	1463	170	1289
25823	Río Negro	Topaipí	1299	181	1107
25871	Río Negro	Villagómez	657	141	516
25885	Río Negro	Yacopí	5149	989	4124
25126	Sabana centro	Cajicá	13383	8438	4950
25175	Sabana centro	Chía	32952	25832	7122
25200	Sabana centro	Cogua	5962	1837	4124
25214	Sabana centro	Cota	5714	2942	2727

25295	Sabana centro	Gachancipá	2868	1475	1366
25486	Sabana centro	Nemocón	3353	1340	2020
25758	Sabana centro	Sopó	6380	3953	2408
25785	Sabana centro	Tabio	6736	2729	3923
25799	Sabana centro	Tenjo	4924	2321	2601
25817	Sabana centro	Tocancipá	7302	3159	4148
25899	Sabana centro	Zipaquirá	31009	27336	3679
25099	Sabana occidente	Bojacá	2441	1794	617
25260	Sabana occidente	El Rosal	3061	1930	1092
25269	Sabana occidente	Facatativá	33292	29972	3308
25286	Sabana occidente	Funza	15185	14386	815
25430	Sabana occidente	Madrid	12758	10121	2578
25473	Sabana occidente	Mosquera	18029	17137	860
25769	Sabana occidente	Subachoque	4449	1622	2829
25898	Sabana occidente	Zipacón	1767	457	1282
25740	Soacha	Sibaté	7363	5022	2343
25754	Soacha	Soacha	124438	122903	1522
25053	Sumapaz	Arbeláez	3070	1118	1929
25120	Sumapaz	Cabrera	1159	249	909
25290	Sumapaz	Fusagasugá	36743	28738	7946
25312	Sumapaz	Granada	2190	499	1691
25506	Sumapaz	Venecia	1340	349	989
25524	Sumapaz	Pandi	1545	310	1235
25535	Sumapaz	Pasca	2949	614	2334
25649	Sumapaz	San Bernardo	3024	896	2117
25743	Sumapaz	Silvania	6602	1941	4667
25805	Sumapaz	Tibacuy	1223	97	1131
25035	Tequendama	Anapoima	5241	2416	2829
25040	Tequendama	Anolaima	4506	1115	3388
25123	Tequendama	Cachipay	3004	959	2045
25245	Tequendama	El Colegio	8536	2795	5726
25386	Tequendama	La Mesa	10764	5912	4837
25596	Tequendama	Quipile	2999	242	2757
25599	Tequendama	Apulo	2891	1148	1744
25645	Tequendama	San Antonio del Tequendama	3344	209	3129
25797	Tequendama	Tena	3104	258	2846
25878	Tequendama	Viotá	4949	1397	3535
25154	Ubaté	Carmen de Carupa	2861	613	2209
25224	Ubaté	Cucunubá	1883	328	1553
25288	Ubaté	Fúquene	1539	72	1468

25317	Ubaté	Guachetá	2799	803	1987
25407	Ubaté	Lenguazaque	2498	502	1994
25745	Ubaté	Simijaca	3033	1682	1349
25779	Ubaté	Susa	2928	1122	1744
25781	Ubaté	Sutatausa	1521	476	1045
25793	Ubaté	Tausa	2534	243	2281
25843	Ubaté	Villa de San Diego de Ubaté	8972	6352	2711

El número de viviendas presentadas en esta tabla corresponde a un estimativo a partir de las proyecciones de población del DANE y el número de personas por vivienda del censo 2005.

La población objetivo del marco PERS corresponde a las viviendas de los municipios del área dispersa (RESTO en la Tabla 2) + población en cabecera de municipios con índice de ruralidad mayor a 40%. Es decir que para el caso de Cundinamarca en 113 municipios esta población corresponde a la del área dispersa (RESTO), mientras que en 3 municipios (Medina, Paratebuena y Gutiérrez) la población objetivo es el total de viviendas del municipio.

En la Tabla 3 se presenta el resumen de la población objetivo para las 15 provincias.

Tabla 3. Resumen por Provincias – Viviendas y Participación

Provincia	viviendas 2015	Participación
Almeidas	12.693	5,02%
Alto Magdalena	8.176	3,24%
Bajo Magdalena	10.944	4,33%
Gualivá	22.828	9,04%
Guavio	18.903	7,48%
Magdalena Medio	6.676	2,64%
Medina	4.744	1,88%
Oriente	16.509	6,53%
Río Negro	15.843	6,27%
Sabana centro	39.067	15,46%
Sabana occidente	13.381	5,30%
Soacha	3.864	1,53%
Sumapaz	24.947	9,87%
Tequendama	32.836	13,00%
Ubaté	21.213	8,40%
Total	252.626	100,00%

4. Etapa Descriptiva

4.1 Técnicas de muestreo

Una vez que se han definido las características de los sujetos u objetos del estudio y se ha calculado el número necesario, sólo queda determinar la manera en que serán seleccionados de la población a la que pertenecen. Se denomina muestreo al procedimiento mediante el cual se obtiene una muestra de la población.

Existen dos tipos de muestreo: el «probabilístico» y el «no probabilístico». Con el muestreo «probabilístico», todos los sujetos u objetos tienen la misma probabilidad de entrar a formar parte del estudio. La elección se hace al azar. El «no probabilístico» es aquel en el que no todos los sujetos u objetos tienen la misma probabilidad de formar parte de la muestra de estudio.

Los tipos de muestreo «probabilístico» más utilizados son: aleatorio simple, aleatorio sistemático, aleatorio estratificado y aleatorio por conglomerados.

Muestreo aleatorio simple: Para poder realizar este tipo de muestreo, todos los individuos de la población deben estar numerados en un listado. Normalmente se hace a partir de un listado de números aleatorios, con un programa estadístico, o con alguno de los programas para calcular el tamaño de la muestra que tenga la opción de generar listados de números aleatorios.

Muestreo aleatorio sistemático: Es similar al aleatorio simple, los sujetos han de estar identificados, pero no es necesario disponer de un listado. Éstos no se eligen a partir de un listado de números aleatorios, sino que se hace sistemáticamente eligiendo a uno de cada cierto número de sujetos. Este número se denomina «fracción de muestreo» (k) y se calcula dividiendo el total de la población por la muestra necesaria.

Muestreo aleatorio estratificado: En este tipo de muestreo se divide a la población en subgrupos o estratos que tienen alguna característica común; e interesa mantener estos estratos en la muestra, para que mantenga la misma composición que la población. La selección de sujetos dentro de cada estrato se realizará aleatoriamente. La estratificación se suele hacer en función de diferentes variables o características de interés, definidas previamente por el referente teórico.

Muestreo por cuotas: Consiste en seleccionar la muestra considerando una serie de características específicas presentes en la población, por lo que la muestra habrá de tenerlas en la misma proporción. Las cuotas se establecen a partir de variables

consideradas relevantes: de la información consultada y de los intereses de la investigación

Muestreo por bola de nieve: Se utiliza cuando la población es difícil de identificar o cuando es complicado acceder a ella porque tiene ciertas características que no son muy aceptadas socialmente. Consiste en ir seleccionando los individuos a partir de un solo elemento o de un grupo reducido, que va conduciendo a otros individuos que reúnen las características de estudio; éstos, a su vez, conducen a otros y así se va obteniendo el número de individuos necesario.

Muestreo accidental: Este tipo de muestreo se denomina también «consecutivo», ya que la selección de los sujetos de estudio se hace en función de su presencia o no en un lugar y un momento determinados. Es el caso, por ejemplo, de la inclusión de las mujeres a medida que van acudiendo al hospital, o el de un encuestador que en la calle entrevista a las personas que pasan en ese momento por allí. Aunque puede parecer similar al muestreo «probabilístico», es evidente que no todas las personas tienen la misma probabilidad de estar en el momento y el lugar donde se selecciona a los sujetos.

Muestreo por conglomerados: Este tipo de muestreo también se denomina en «etapas múltiples». Se emplea cuando se desea estudiar una población grande y dispersa, y no se dispone de ningún listado para poder aplicar las técnicas anteriores. En lugar de seleccionar sujetos, se empieza por seleccionar subgrupos o «conglomerados» a los que se da el nombre de «unidades de primera etapa» o «unidades primarias».

La diferencia con los estratos del tipo de muestreo anterior es que los conglomerados ya están agrupados así de forma natural (provincias, municipios, etc.). Seleccionan, de manera aleatoria, las «unidades de segunda etapa» o «unidades secundarias» a partir de las unidades primarias. Así, sucesivamente, se van eligiendo hasta llegar a las unidades de análisis, que serán los individuos que compongan la muestra de estudio.

4.2 Error Estadístico y niveles de confianza

Un aspecto importante en la metodología estadística, es el cálculo de la cantidad de participantes que deben incluirse en un estudio. El tamaño de muestra permite a los investigadores saber cuántos individuos son necesarios estudiar, para poder estimar un parámetro determinado con el grado de confianza deseado, o el número necesario para poder detectar una determinada diferencia entre los grupos de estudio, suponiendo que existiese realmente. El cálculo del tamaño de la muestra es una función matemática que expresa la relación entre las variables, cantidad de participantes y poder estadístico.

La muestra de un estudio debe ser representativa de la población de interés. El objetivo principal de seleccionarla es hacer inferencias estadísticas acerca de la población de la que proviene. La selección debe ser probabilística. Los factores estadísticos que determinan el tamaño de la muestra son: hipótesis, error alfa, error beta, poder estadístico, variabilidad, pérdidas en el estudio y el tamaño del efecto.

En esta sección se revisan las fórmulas utilizadas para el cálculo del tamaño de la muestra para PERS Cundinamarca, y las estrategias para disminuir el número necesario de sujetos. Además, se describe el procedimiento aplicado específicamente al diseño muestral y metodológico del estudio realizado en el PERS Cundinamarca. Asumiendo que los datos se comportan de manera normal, se utiliza los valores de Z, según el valor del nivel de significación α .

Tabla 4. Valores de Z según el valor de significación

A	0,10	0,05	0,01	0,001
Z α	1,645	1,960	2,576	3,291

Los diferentes tamaños muestrales mínimos, máximos y medios, según valores de nivel de confianza (α), prevalencia (p) y precisión (d) de la Tabla 5, se observan en la Tabla 6.

Tabla 5. Nivel de confianza, prevalencia y precisión para la selección del tamaño muestral

A (Z α)	p	d	N
0,05 (1,96)	0,20	0,08	97
0,05 (1,96)	0,20	0,04	385
0,01 (2,576)	0,20	0,08	166
0,01 (2,576)	0,20	0,04	664
0,05 (1,96)	0,40	0,08	145
0,05 (1,96)	0,40	0,04	576

Tabla 6. Valores máximos, medios y mínimos de la muestra

A (Z α)	P y d	N	mínimo	medio	50% (+)	30% (+)
0,05 (1,96)	0,20; 0,04	385	401	601	802	826
0,01 (2,576)	0,20; 0,04	664	691	863	1036	1346
0,05 (1,96)	0,40; 0,04	576	600	750	900	1170

Nota: El 30% más para N del 50% más sobre el valor mínimo, sumado a valor medio de las encuestas, está dando un margen del 80% de aseguramiento muestral.

Como resultado de este análisis es que es necesario realizar un máximo de **1346** encuestas con un nivel de significación de 0,01%, o en su defecto **1170** encuestas con un nivel de significación de 0,05%, cualquiera de los dos valores está justificado estadísticamente.

5. Selección de la Muestra

5.1 Marco Muestral

Para seleccionar de manera aleatoria las diferentes unidades en cada una de las etapas previstas en el diseño muestral, fue necesario construir un marco de muestreo con información de las siguientes variables: las regiones del departamento, los municipios en cada región, centros poblados en cada municipio, zonas para cada municipio y finalmente las unidades de consumo energía básica (residencial, comercial, industrial e institucional) en cada centro poblado.

Desde el punto de vista del diseño muestral, fue posible construir el marco de muestreo hasta el nivel de municipio (consideradas las unidades primarias de muestreo –UPM–), con los que se procedió posteriormente, y con base en información de los planes (o esquemas) de ordenamiento territorial de cada municipio, a seleccionar centros poblados (veredas) y finalmente, las unidades finales de muestreo (unidades de consumo básico) a partir de la elaboración de mapas de ruta para cada UPM seleccionada. Esta metodología de selección de veredas y unidades de consumo básico se detalla más adelante en la sección de trabajo de campo.

5.2 Esquema de Muestreo

Con base en los instrumentos evaluados en el programa de PERS se realiza la selección más adecuada de muestreo equiprobabilístico por conglomerados bimetálico estratificado.

Debido a alta variabilidad relativa del consumo de energía eléctrica observada en las unidades comerciales e industriales/institucionales de las diferentes provincias del departamento, comparada con las otras unidades de consumo, se consideró finalmente, seleccionar muestras independientes en todo el departamento según la unidad de consumo de acuerdo con el siguiente esquema: Residenciales:

- Muestras probabilísticas por conglomerados bietápicas, independientes para cada provincia.
- Comerciales: Muestra probabilística estratificada (por región) por conglomerados bietápica.
- Otros no residenciales (Industriales e institucionales)
- Muestra probabilística estratificada (por provincia) por conglomerados bietápica.

Como el diseño muestral de la encuesta PERS se caracteriza por ser probabilístico, los resultados obtenidos de la encuesta se generalizan a todos los dominios de la población (provincias y departamento).

A su vez es bietápico (dos etapas), estratificado y por conglomerados, según área geográfica y tamaño de la población, donde la unidad última de selección y observación es la unidad de consumo (residencial, comercial, industrial e institucional). La selección en cada una de las etapas es estrictamente aleatoria.

En la primera etapa, la población se dividió en las quince provincias con sus correspondientes conglomerados (municipios) que constituyen las Unidades Primarias de Muestreo –UPM– y que se utilizaron como Marco Muestral con sus respectivos tamaños poblacionales (proyecciones del DANE a 2015).

Se seleccionó al azar muestras independientes en cada región con probabilidades proporcionales a los tamaños poblacionales (hogares) de los conglomerados. El número de UPMs en cada región se fijó en un 35% con un mínimo de 1 UPM y máximo de 5 UPMs. Los municipios fueron seleccionados al azar mediante la aplicación del paquete estadístico EPIDAT, a partir del marco muestral que contiene el listado de provincias y los municipios.

En la etapa 2 se seleccionan el número de encuestas para cada uno de los municipios en el sector residencial, proporcionalmente al número de viviendas, Índice de Ruralidad y el Índice del Complemento de Electrificación Rural, en cada uno de los Conglomerados. De acuerdo a esto, se tiene en cuenta una Probabilidad de no menos del 2,8% y contando con la selección proporcional se obtienen como estratos a los municipios.

En el Anexo 1 de este informe se presenta en detalle el procedimiento llevado a cabo en el Programa Epidat 4.0, teniendo en cuenta el Muestreo Aleatorio Estratificado proporcional.

1. Inicialmente se selecciona la base de datos previamente depurada.
2. Los municipios seleccionados en la etapa anterior se toman como estratos para asegurar su selección.
3. Luego se realiza el reparto proporcional de estratos para que las 1346 encuestas sean distribuida de manera proporcional a cada municipio.
4. Mediante este proceso se obtuvo el listado de los municipios y el número de encuestas en cada uno de ellos, con alta confiabilidad y robustez estadística, como se observa en la Tabla 7.

Tabla 7. Número Total de Encuestas y Distribución por Municipio

Código DANE	Provincia	Municipio	Probabilidad	Ponderación	Selección	Numero encuestas
25183	Almeidas	Chocontá	36,1	2,8	38,7	26
25426	Almeidas	Macheta	36,2	2,8	43,9	34
25772	Almeidas	Suesca	35,9	2,8	39,8	28
25001	Alto Magdalena	Agua de Dios	35,5	2,8	20,9	11
25307	Alto Magdalena	Girardot	38,9	2,6	8,0	7
25324	Alto Magdalena	Guataquí	35,5	2,8	31,8	22
25488	Alto Magdalena	Nilo	36,2	2,8	50,1	34
25320	Bajo Magdalena	Guaduas	36,5	2,7	38,0	27
25402	Gualivá	La Vega	35,9	2,8	37,9	28
25491	Gualivá	Nocaima	36,4	2,8	48,0	36
25718	Gualivá	Sasaima	35,6	2,8	48,2	31
25851	Gualivá	Útica	36,5	2,7	51,0	24
25875	Gualivá	Villeta	36,8	2,7	32,2	21
25297	Guavio	Gachetá	36	2,8	36,9	27
25326	Guavio	Guatavita	36,6	2,7	43,5	30
25377	Guavio	La Calera	35,8	2,8	26,9	19
25839	Guavio	Ubalá	36,3	2,8	52,3	37
25095	Magdalena Medio	Bituima	36,6	2,7	51,0	37
25580	Magdalena Medio	Pulí	36,2	2,8	57,9	38
25662	Magdalena Medio	San Juan de Río Seco	36	2,8	46,2	36
25438	Medina	Medina	36	2,8	56,4	40
25530	Medina	Paratebueno	36,4	2,8	58,4	40
25151	Oriente	Caqueza	36,5	2,7	39,5	27
25181	Oriente	Choachí	35,8	2,8	42,3	29
25279	Oriente	Fomeque	35,7	2,8	45,1	35
25339	Oriente	Gutiérrez	35,9	2,8	42,5	33
25594	Oriente	Quetame	36,1	2,8	49,9	35
25394	Río Negro	La Palma	36,6	2,7	39,9	30
25513	Río Negro	Pacho	36,2	2,8	38,0	25
25823	Río Negro	Topaipí	36,1	2,8	54,1	39
25885	Río Negro	Yacopí	36,1	2,8	64,8	44
25126	Sabana centro	Cajicá	36,5	2,7	24,1	19

25175	Sabana centro	Chía	29,4	3,4	9,4	5
25799	Sabana centro	Tenjo	36,7	2,7	46,5	33
25889	Sabana centro	Zipaquirá	36,4	2,8	9,8	8
25269	Sabana occidente	Facatativá	38,1	2,6	9,3	8
25286	Sabana occidente	Funza	26,7	3,8	6,0	4
25473	Sabana occidente	Mosquera	34,8	2,9	14,4	8
25740	Soacha	Sibate	35,9	2,8	14,9	14
25290	Sumapaz	Fusagasugá	35,3	2,8	17,3	12
25524	Sumapaz	Pandi	36,7	2,7	44,1	33
25743	Sumapaz	Silvania	36,5	2,7	40,3	31
25035	Tequendama	Anapoima	36,8	2,7	35,0	25
25123	Tequendama	Cachipay	35,9	2,8	33,9	28
25386	Tequendama	La Mesa	35,8	2,8	36,5	24
25645	Tequendama	San Antonio del Tequendama	36,5	2,7	40,5	31
25154	Ubaté	Carmen de Carupa	35,9	2,8	47,7	33
25407	Ubaté	Lenguzaque	36,5	2,7	46,4	31
25779	Ubaté	Susa	36,5	2,7	35,1	31
25843	Ubaté	Ubaté	36,4	2,8	32,0	38
TOTAL ENCUESTAS						1.346

Como resultado de este proceso estadístico se estableció que los 50 municipios listados en la Tabla 7 son representativos de Cundinamarca, y que por lo tanto en ellos deberían llevarse a cabo las 1.346 encuestas distribuidas como se presenta en la misma tabla. Adicionalmente, se establece que los municipios seleccionados en cada provincia son representativos de la población total de dicha provincia y por lo tanto las encuestas realizadas en dichos municipios no representan a ningún municipio en particular sino a la provincia en general.

La Figura 2 presenta un diagrama de barras con la distribución de encuestas por provincia, en donde se evidencia que provincias con altos IR e INER, como Gualivá y Oriente, requieren de un mayor número de encuestas, mientras que provincias con alta densidad de población pero bajos índices IR e INER, como Soacha, requieren de una menor cantidad de encuestas.

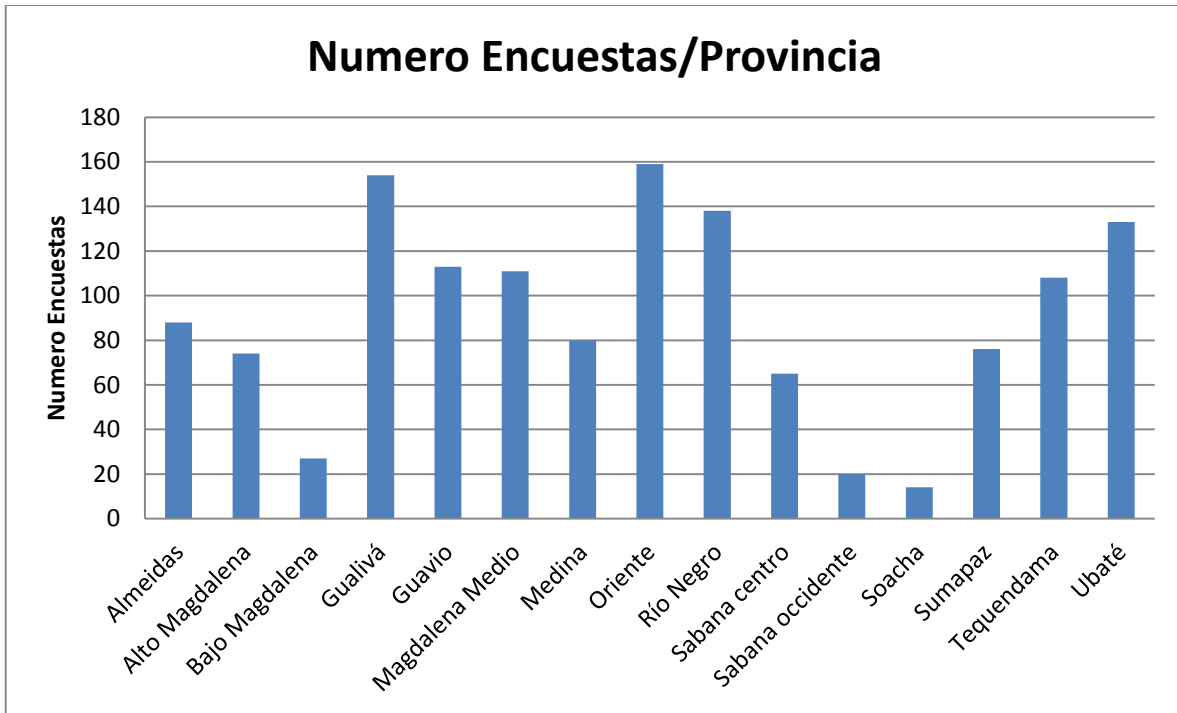


Figura 2. Distribución del número de encuestas por provincia

Además del análisis estadístico existen otros factores a tener en cuenta al momento de determinar la muestra final. Estos factores adicionales se presentan en la siguiente sección.

6. Depuración de la Muestra

La selección realizada en la sección anterior cubre las 15 provincias de Cundinamarca; sin embargo:





1. Las provincias más electrificadas del departamento (Soacha, Sabana Occidente, Bajo Magdalena y Sabana Centro), tienen un número de encuestas muy bajo comparativamente con las demás provincias.
2. Seis municipios (Girardot, Chía, Zipaquirá, Facatativá, Funza y Mosquera) presentan un número de encuestas menor a 10.
3. Ninguno de los municipios de la provincia Sabana Occidente requiere de más de 10 encuestas.
4. Se debe optimizar el uso de los recursos, por lo tanto no resulta conveniente realizar desplazamientos hasta municipios que no tengan un número significativo de encuestas.
5. La logística para realizar 1 o más encuestas es la misma.





Teniendo en cuenta lo anterior, se establecieron las siguientes restricciones:





1. Es obligatorio que se realice el proceso de encuestas en al menos un municipio por provincia, es decir que se deben visitar las 15 provincias.
2. No se van a realizar encuestas en municipios que requieran menos de 10 encuestas.
3. Las encuestas del municipio o los municipios que no se visiten se distribuyen en otros municipios de la misma provincia. Las encuestas se transfieren al municipio más cercano.
4. En el caso de Sabana Occidente, entre los tres municipios seleccionados inicialmente (Facatativá, Funza y Mosquera), se realizó una selección aleatoria de uno de ellos (Mosquera) para la aplicación del total de las encuestas de la provincia.




Como resultado final a partir de la depuración de la muestra, se obtuvieron los resultados que se presentan en la Tabla 8, en donde para cada provincia se presentan: i) los municipios seleccionados (resaltados en amarillo) con el número de encuestas por municipio; ii) los municipios descartados en el proceso de depuración (resaltados en azul) con el número de encuestas transferidas; iii) Total de encuestas por provincia.

Tabla 8. Distribución de encuestas por provincia y municipio

No.	Ubicación Provincia	Provincia	Municipios	(Encuestas por Municipio)
1		Almedias P=72.891	Chocontá Machetá Manta Sesquilé Suesca Tibirítí Villapinzón	26 34 28 TOTAL: 88
2		Alto Magdalena P=152.271	Agua de Dios Girardot (7) Guataquí Jerusalén Nariño Nilo Ricaute Tocaima	18 22 34 TOTAL: 74
3		Bajo Magdalena P=60.275	Caparrapí Guaduas Salgar	 27 TOTAL: 27
4		Gualivá P=99.108	Albán La Peña la Vega Nimaima Nocaima Quebradanegra San Francisco Sasaima Supatá Útica Vergara Villeta	 28 36 31 24 21 TOTAL: 140

5		Guavio P=80.597	Gachalá Gachetá Gama Guasca Guatavita Junín La Calera Ubalá	27 30 19 37 TOTAL: 113
6		Magdalena Medio P=28.121	Beltrán Bituima Chaguaní Guayabal de Síquima Pulí San Juan de Rioseco Vianí	37 38 36 TOTAL: 111
7		Medina P=16.740	Paratebueno Medina	40 40 TOTAL: 80
8		Oriente P=16.509	Cáqueza Chipaque Choachí Fómeque Fosca Guayabetal Gutiérrez Quetame Ubaque Une	27 29 35 33 35 TOTAL: 159

9		Rio Negro P= 71.927	El Peñón La Palma Pacho Paimé San Cayetano Topaipí Villagómez Yacopí	30 25 39 44 TOTAL: 138
10		Sabana Centro P=347.890	Cajicá Chía (5) Cogua Cota Gachancipá Nemocón Sopó Tabio Tenjo Tocancipá Zipaquirá (8)	19 33 13 TOTAL: 65
11		Sabana Occidente P=72.891	Bojacá El Rosal Facatativa (8) Funza (4) Madrid Mosquera (8) Subachoque Zipacón	 20 TOTAL: 20
12		Soacha P= 429.46	Sibaté Soacha	14 TOTAL: 14

13		Sumapaz P= 185.453	Arbeláez Cabrera Fusagasugá Granada Pandi Pasca San Bernardo Silvania Tibacuy Venecia	12 33 31 TOTAL: 76
14		Tequendama P=126.968	Anapoima Anolaima Apulo Cachipay El Colegio La Mesa Quipile San Antonio de Tequendama Tena Viotá	25 28 24 31 TOTAL: 108
15		Ubaté P= 106.325	Cucunubá Fúquene Guachetá Lenguazaque Carmen de Carupa Simijaca Susa Sutatausa Tausa Villa de San Diego de Ubaté	31 33 31 38 TOTAL: 133

Finalmente, la Tabla 9 presenta el resumen de encuestas por provincia.

Tabla 9. Distribución de encuestas por provincia

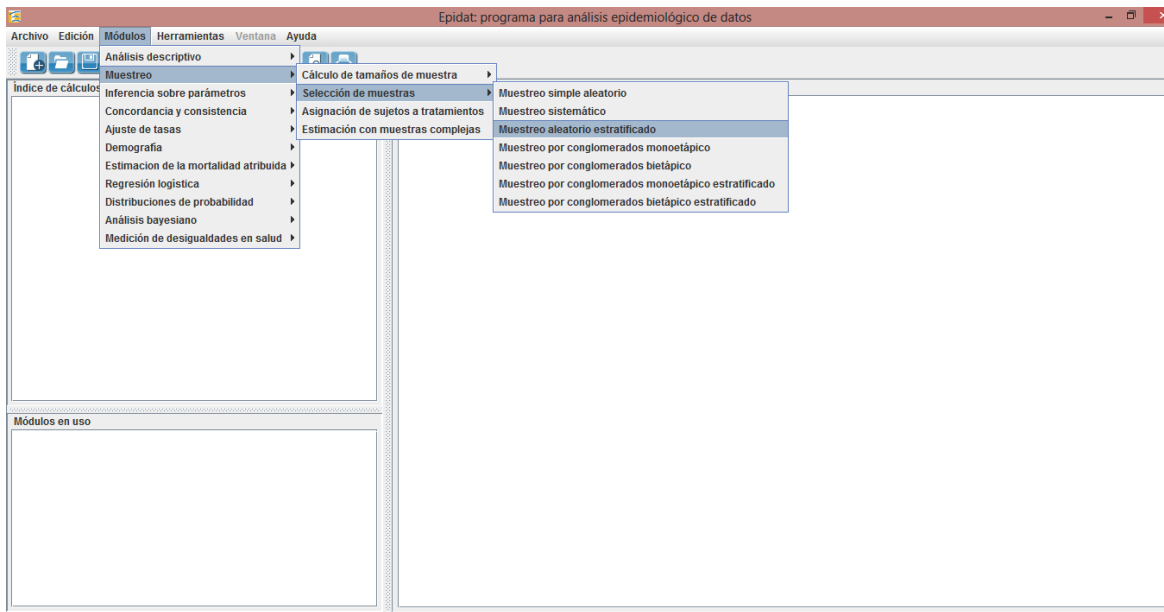
PROVINCIA	NÚMERO ENCUESTAS
Almeidas	88
Alto Magdalena	74
Bajo Magdalena	27
Gualivá	140
Guavio	113
Magdalena Medio	111

Medina	80
Oriente	159
Río Negro	138
Sabana centro	65
Sabana occidente	20
Soacha	14
Sumapaz	76
Tequendama	108
Ubaté	133
TOTAL	1346

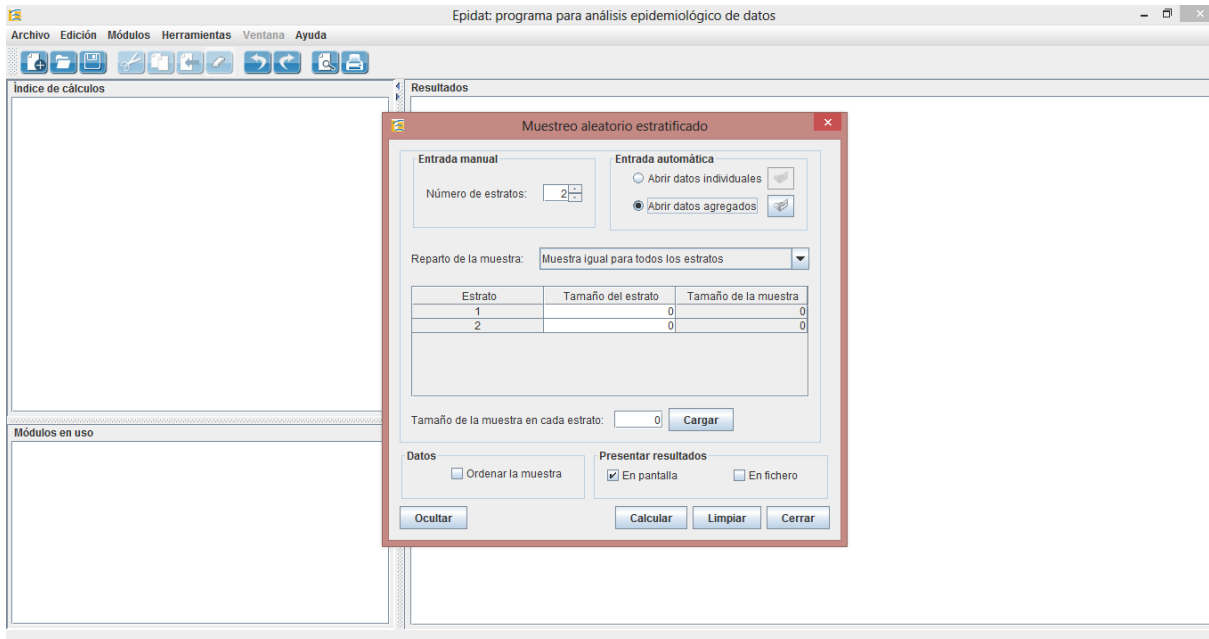
Es decir que se van a realizar 1346 encuestas en 46 municipios que abarcan las 15 provincias del departamento de Cundinamarca.

Anexo 1: Ejecución de EPIDAT

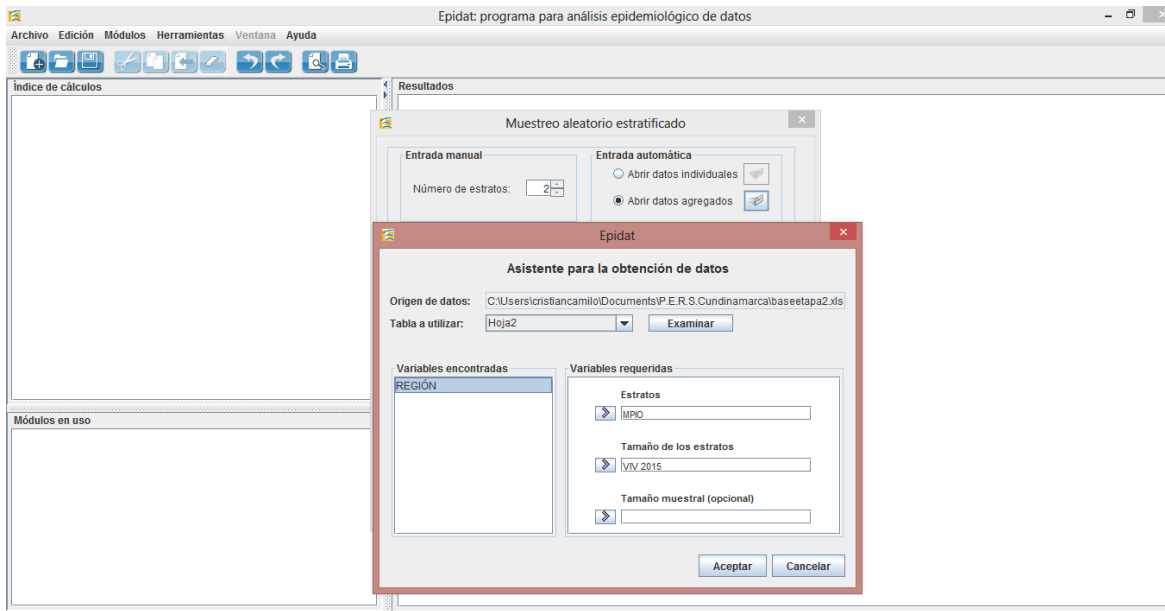
Como se describió anteriormente, la selección del número de encuestas para cada uno de los municipios es proporcional al número de viviendas, Índice de Ruralidad y al Índice de no electrificación rural. Teniendo a los municipios como estratos, se empleó el Muestreo Aleatorio Estratificado proporcional en el Programa EPIDAT 4.0.



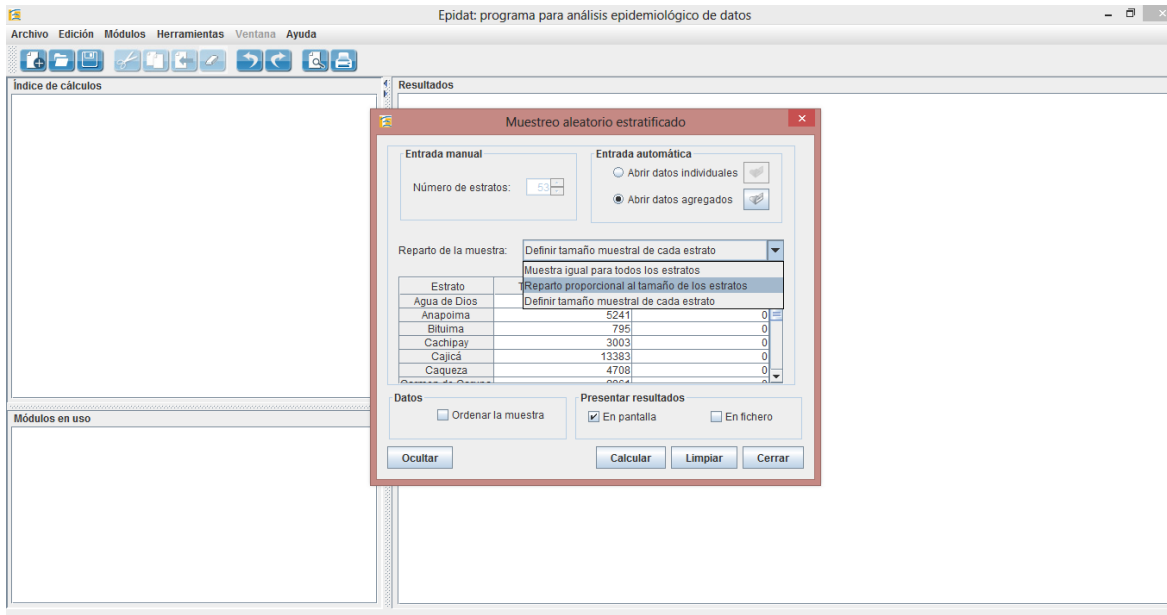
Seleccionando la opción de Datos Agrupados, se selecciona la base de datos previamente depurada:



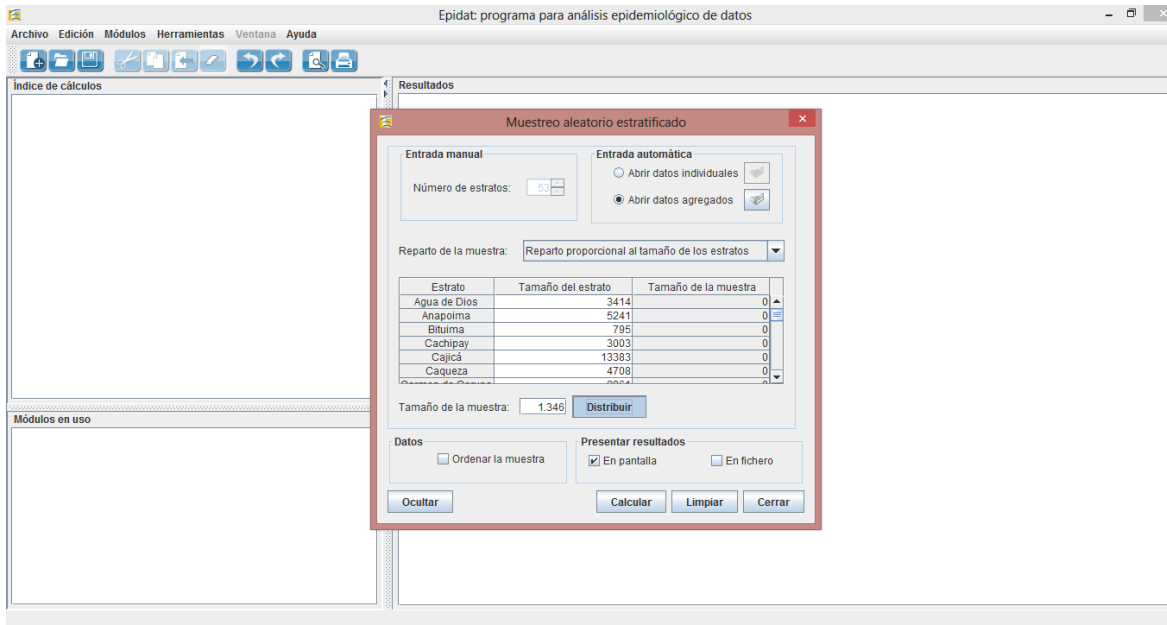
Ya que los municipios deben estar con selección al 100% de los seleccionados en la etapa anterior, se toman como estratos para asegurar su selección, mientras que la proporción viene dada por el índice de selección (sumatoria de los índices), de la siguiente manera:



Luego de esto la opción de reparto proporcional a los estratos, para que la muestra total de encuestas que es 1346 sea distribuida de manera proporcional a cada municipio, de la siguiente manera:



Con la Selección de Distribuir y con el número de encuestas en la casilla que se muestra a continuación, se realiza el reparto de acuerdo a algoritmos matemáticos que se manejan en esta sección y permiten el reparto, como se muestra a continuación:



Por último se calcula en Fichero, y se muestra el resultado en pantalla como en un archivo de salida de Excel 2007, de esta manera se tiene el listado de encuestas a realizar en cada municipio.

Epidat: programa para análisis epidemiológico de datos

Archivo Edición Módulos Herramientas Ventana Ayuda

Índice de cálculos

[1] Muestreo aleatorio estratificado
[2] Muestreo aleatorio estratificado

Módulos en uso

Muestreo aleatorio estratificado

Resultados

Archivo de trabajo: C:\Users\cristiancamilo\Documents\P.E.R.S.Cundinamarca\baseetapa2.xlsx
Tabla: BASE2
Variables:
Estratos: MP10
Tamaño de los estratos: INDICESELECCIÓN

Datos:

Reparto de la muestra: Reparto proporcional al tamaño de los estratos
Tamaño de la muestra: 1.346

Estrato	Tamaño del estrato	Tamaño de la muestra
Agua de Dios	31	11
Anapoima	68	25
Bituma	101	37
Cachipay	78	28
Cajicá	52	19
Caqueza	74	27
Carmen de Car...	92	33
Chá	17	5
Choschí	81	29
Chocotá	72	26
Facatativá	21	8
Fomeque	98	35
Funza	15	4
Fusagasugá	34	12
Gachetá	75	27
Girardot	18	7
Guaduas	74	27
Guataquí	62	22
Guatavita	82	30

Al final se tiene como resultado el listado de los municipios y el número de encuestas en cada uno de ellos, con alta confiabilidad y robustez estadística.