



Unificación de información de los censos empresariales hechos por las Cámaras de Comercio de Colombia

Entregable 5: Metodología censal

Versión final

Bogotá D.C., 31 de marzo de 2019

CONTENIDO

INTRODUCCIÓN.....	4
1. DISEÑO DE LA OPERACIÓN ESTADÍSTICA.....	7
1.1. Marco conceptual.....	7
1.1.1. Objetivo	7
1.1.2. Necesidades de información.....	7
2. DISEÑO Y DOCUMENTACIÓN TEMÁTICA	9
2.1. Contenido temático.....	9
2.2. Alcance metodológico y limitaciones.....	12
2.3. Diseño estadístico	13
2.3.1. Población objetivo	15
2.3.2. Subpoblaciones de interés	16
2.3.3. Variables de interés	17
2.3.4. Indicadores de interés	18
2.3.5. Cobertura geográfica	20
2.3.6. Niveles de desagregación	20
2.3.7. Metodología muestral	21
2.3.7.1. Unidades estadísticas.....	21
2.3.7.2. Marco muestral	22
2.3.7.3. Diseño muestral	23
2.3.7.4. Tamaño de la muestra.....	24
2.3.7.5. Algoritmos de selección de muestra.....	25
2.3.7.6. Tamaño de muestra por UCU.....	25
2.3.7.7. Formatos de control y selección de muestra.....	27
2.3.8. Metodología censal.....	28
2.3.9. Esquemas de validación y consistencia	38
2.3.10. No respuesta e imputación.....	39

2.3.11.	Estructura y diseño de la base de datos	40
2.4.	Diseño del operativo de campo.....	40
2.4.1.	Aspectos relevantes	40
2.4.2.	Esquema organizacional	41
2.4.3.	Diseño de la capacitación del personal.....	44
2.4.4.	Rutas de trabajo	45
2.4.5.	Esquema de barrido de áreas geográficas.....	45
2.4.6.	Recolección de la información	47
2.5.	Prueba piloto	48
3.	PLAN DE ANÁLISIS.....	49
3.1.	Procesamiento.....	49
3.2.	Diseño de los cuadros de salida o de resultados	50
	<i>Glosario</i>	54
	<i>Bibliografía</i>	63
	Anexo I. Formulación del tamaño de la muestra.....	65
	Anexo II. Algoritmo de selección de UCU	66

INTRODUCCIÓN

El censo es un procedimiento estadístico mediante el cual se investigan las unidades de observación pertenecientes a un universo de estudio. La recolección de los datos y de la información individual de todos y cada uno de los elementos que conforman un universo de estudio, permite organizar y ampliar el conocimiento que se tenga sobre éste (UNECE, 2000). En nuestro caso, el asunto en consideración es la caracterización de los empresarios y la identificación de la informalidad en el país.

Actualmente, en Colombia es de suma importancia realizar un censo de formalización empresarial. La posibilidad de contar con una metodología actualizada, confiable y oportuna para establecer el diagnóstico del entorno empresarial e informal en Colombia permitiría definir las rutas de acción más apropiadas para fortalecer el sector y sortear la informalidad. La información obtenida a partir de este censo permitirá no sólo medir y calcular el tamaño de la informalidad empresarial en el país de manera unificada¹, sino también conocer las principales razones de la informalidad en los sectores de la industria, el comercio y los servicios. De igual manera, contribuiría a entender las consecuencias que tiene la informalidad para los empresarios, la economía del país y la región.

De acuerdo con la CEPAL (2011), determinar el tamaño de la población a censar constituye uno de los factores elementales para tener en cuenta a la hora de realizar cualquier estudio, pues sólo así se pueden determinar de manera precisa los recursos requeridos para

¹ Se estima que en el país la economía informal tiene un tamaño de alrededor el 33% del Producto Interno Bruto (PIB), mientras que la economía ilegal es de 1,9% (ANIF, 2017). En este mismo sentido, según la Gran Encuesta Integrada de Hogares (GEIH) del DANE, se estimaba que, en 2015, el 73.2% de los microestablecimientos, es decir, empresas con menos de 10 trabajadores, no estaban registradas.

el levantamiento de la información, así como el tiempo estimado para la recolección de los datos.

Según Statistics Canada (2010), las encuestas pueden llevarse a cabo a través de censos o de muestras. Las encuestas por muestreo, por un lado, recolectan la información con base en solo una fracción de la población, mientras que los censos, por el otro lado, recogen la información de toda la población de interés. En ambos casos, la información que se levanta se utiliza para calcular indicadores que permitan elaborar conclusiones sobre toda la población y algunos subgrupos que sean de interés. Statistics Canada enfatiza en que la razón principal para preferir una encuesta por muestreo sobre un censo consiste en que la primera es una forma más rápida y económica de obtener datos de calidad que satisfagan las necesidades de la investigación. Además, dado que es una operación a menor escala que un censo, también es más fácil de controlar y monitorear. Por este motivo, en las poblaciones muy numerosas se suele optar por una estrategia muestral la cual, a partir de la identificación de errores de muestreo pequeños, garantiza la precisión de los indicadores estimados. Sin embargo, en poblaciones menos numerosas lo óptimo es realizar una estrategia censal (ONU, 2010).

Este documento propone que en las poblaciones densas (numerosas) se realicen encuestas por muestreo, pues permite una toma de datos generalizada lo suficientemente representativa de la totalidad de la población. En el caso de las poblaciones menos densas, se sugiere que se emplee el esquema censal, como lo afirma la ONU (2010), puesto que aborda el 100% de los establecimientos visibles.

En este documento se propone una metodología que recoge la información por medio de un formulario estructurado, a través del cual se obtiene información estadística sobre las causas y consecuencias de la informalidad empresarial, al igual que datos sobre los insumos

para caracterizar y conocer las condiciones de los establecimientos, a saber: hace cuánto tiempo están funcionando, cuáles son sus principales actividades económicas, cuáles son las condiciones de los empleados, los ingresos, entre otras. La información obtenida permite analizar los problemas y las dificultades que enfrentan para, entre otras cosas, identificar los aspectos en los que se les debería apoyar por medio de políticas públicas. A su vez, la información recolectada servirá para cerciorarse de que las políticas públicas respondan realmente a las necesidades existentes en el país y, además, permitirá hacer un seguimiento efectivo de las variables más influyentes.

Es importante aclarar que con esta metodología se identifican, principalmente, los establecimientos visibles. En el caso de la industria oculta se podrían analizar fuentes de información de registros administrativos de empresas de servicios públicos, con el fin de establecer comportamientos inusuales de consumo en horarios considerados como laborales y que se diferencien considerablemente de los demás predios de su entorno. No obstante, la realización de este ejercicio está fuera del alcance de este documento.

Así las cosas, en este documento se exponen las propuestas metodológicas para la recolección muestral y censal. El texto tiene cuatro ejes. En primer lugar, presenta el marco general para la realización de las encuestas. En segundo, expone el diseño documental, el diseño estadístico y el diseño del trabajo de campo. En tercer lugar, presenta la propuesta de la prueba piloto. Finalmente, se explica el plan de análisis de la información obtenida. Asimismo, para mayor claridad del documento, se incluye al final un glosario con los términos utilizados.

1. DISEÑO DE LA OPERACIÓN ESTADÍSTICA

Para la realización de las encuestas se tienen en cuenta dos metodologías. En primer lugar, la *metodología censal por barrido*, óptima en poblaciones con baja densidad o que tienen problemas para contar con información cartográfica precisa. En esta metodología, un supervisor de campo identifica los establecimientos visibles por actividad económica mediante un recorrido por la zona. Así, una vez identificados, se puede proceder a la realización de la encuesta. En segundo lugar, se presenta la *metodología muestral*, que es la idónea para el caso de las poblaciones densas y que cuentan con acceso a bases cartográficas. Esta metodología, como se ha mencionado anteriormente, se fundamenta en una selección aleatoria de establecimientos.

1.1. Marco conceptual

En este apartado se presenta el marco general de la metodología censal y muestral.

1.1.1. Objetivo

Fortalecer la información sobre el universo empresarial en Colombia para mejorar el diagnóstico de la informalidad en el país y así tener un mejor enfoque a la hora de realizar las actividades de inspección, vigilancia y control (IVC) a empresas no registradas.

1.1.2. Necesidades de información

Las metodologías aquí propuestas recopilan datos a través de un formulario estructurado con el propósito de brindar la base necesaria para proponer estrategias y políticas públicas informadas que permitan cerrar las brechas existentes entre las empresas formales e

informales. El formulario utilizado en estas metodologías indaga por la siguiente información:

- Identificación del establecimiento.
- Características principales del establecimiento.
- Caracterización de la actividad económica.
- Características y condiciones del personal ocupado.
- Aspectos financieros del establecimiento.
- Expectativas.
- Aspectos relacionados con la seguridad.
- Características de la zona de influencia.

La recolección de la información mediante el formulario se realiza a través de una metodología de barrido en las áreas geográficas de interés. Como es visible, el formulario tiene un objetivo multipropósito, es decir, captura información de diferentes temáticas que robustecen el propósito de la investigación. Así, la variedad de los datos recolectados permiten el diagnóstico de estructuras formales e informales, la realización de análisis de sus actividades económicas y, además, la identificación de aquellas zonas donde se concentran.

2. DISEÑO Y DOCUMENTACIÓN TEMÁTICA

2.1. Contenido temático

El formulario planteado consta de 7 módulos y 62 preguntas. A continuación, se hace una breve descripción de cada módulo:

Módulo A. Identificación

En este módulo se obtienen las variables de la identificación de los establecimientos, es decir, su dirección, su coordenada geográfica, el sector, sección y manzana cartográfica en que se ubican.

Módulo B. Características del establecimiento

En este módulo se averiguan las variables relacionadas con el tipo de establecimiento, emplazamiento y estrato. Asimismo, se indaga por el cumplimiento de la matrícula mercantil y su renovación en el año en curso. En caso de no tenerla, se examinan las razones por que no cuenta con matrícula mercantil.

Por otro lado, en este módulo se obtienen otras variables de interés relacionadas con el tipo de actividad. Para esto, se identifica primero la zona comercial donde se encuentra ubicado y si el establecimiento es único, principal o sucursal. Además, se establece el tiempo de funcionamiento y la tenencia de este. Posteriormente, se investiga el cumplimiento de normas como el Registro Único Tributario (RUT) y el Registro de Información Tributaria (RIT).

El módulo finaliza con la variable del personal ocupado, en la que se busca tener la información para hacer algunas validaciones que permitan asegurar la calidad y consistencia de los datos que se recogen en el módulo D.

Módulo C. *Características de la actividad económica*

Este módulo indaga sobre la dedicación principal de cada establecimiento. A partir de esta información se puede clasificar su actividad económica y el sector productivo al cual pertenece: comercio, servicios o industria.

En este mismo módulo se revisa el cumplimiento de la normatividad que regula a los establecimientos. Por ello, inquiere sobre los trámites que deben ser realizados para su funcionamiento, es decir, los que se refieren al uso del suelo, seguridad contra incendios, trámites ante la secretaría de salud (bioseguridad, salubridad y sanitaria), pagos por derechos a Sayco y Acinpro, trámites para el registro nacional de turismo y cumplimiento de normas establecidas por la Secretaría de Ambiente (intensidad auditiva y horarios permitidos para el funcionamiento).

Módulo D. *Características y condiciones del personal ocupado*

En este se relacionan las condiciones de contratación laboral y el cumplimiento de los requisitos exigidos por la ley. Asimismo, se busca determinar si los socios, familiares y el personal (en tiempo completo o en tiempo parcial), cuentan con las mismas condiciones laborales. Con esta información se determinará el tipo de contratación, formas de pago, aportes a sistemas de seguridad social de salud, pensión y ARL, además de los aportes a cajas de compensación familiar.

En caso de que no exista un cumplimiento estricto sobre las condiciones del personal ocupado, se indaga por las razones que motivan a no contar con estos requerimientos. Finalmente, se determinan las jornadas de trabajo, así como a los horarios de atención al público (en caso de que tenga).

Módulo E. Aspectos financieros

En este módulo, se complementa la información obtenida en el módulo C, relativa a la contabilidad y actividad económica. En particular, este trata de determinar si hay una diferenciación entre los gastos del negocio y los gastos familiares. Finalmente, busca determinar la distribución de gastos de los diferentes conceptos del negocio, así como su frecuencia y forma de pago.

Módulo F. Expectativas

En esta sección se determinan los aspectos que el informante considera apremiantes en la capacitación del personal y, además, se puntualizan las metas que tiene en el corto plazo. Con esto se busca obtener la información necesaria que luego permitiría promover actividades que motiven a los establecimientos, especialmente a aquellos que deban formalizarse.

Módulo G. Seguridad

En este módulo se indaga por los aspectos que permiten caracterizar los delitos que afectan a los establecimientos y su nivel de victimización. Con ayuda de las coordenadas geográficas se puede establecer las zonas de mayor vulnerabilidad, así como la tipificación de los delitos. Adicionalmente, este módulo busca identificar los entes a los cuales los

usuarios suelen acudir en caso de denunciar. En caso de que no se denuncie el delito, también se investigan las razones para no hacerlo.

Módulo H. Características de la zona

Este módulo pretende determinar la percepción de seguridad que se tiene de la zona en donde está ubicado el establecimiento. En caso de encontrar percepción negativa, se indaga por las razones.

2.2. Alcance metodológico y limitaciones

Aunque la *metodología censal por barrido* recorre todas las áreas geográficas es preciso señalar que esta metodología no alcanza a cubrir todas las necesidades de información, pues no identifica los datos y establecimientos ocultos. A continuación, se presentan las ventajas y desventajas de este tipo de metodologías censales frente a las muestrales, que se realizan en poblaciones más extensas.

Tabla 1. Ventajas y desventajas de un censo

Ventajas	Desventajas
• Posibilidad de desagregar la información para grupos poblacionales o áreas geográficas pequeñas. • No es necesario usar métodos estadísticos para la estimación de los indicadores de interés, aunque en ocasiones se requieren ajustes por cobertura.	• El levantamiento de información es muy costoso y toma bastante tiempo. • Debido a su complejidad operativa y administrativa, es difícil mantener el control de todas las etapas del proceso en todas las zonas que se deben recorrer.

Fuente: Adaptado de (Statistics Canada, 2010).

A continuación se presenta el diseño estadístico haciendo la distinción entre la *metodología censal por barrido* y la *metodología muestral*, mencionando los casos en que se recomienda el uso de cada una.

2.3. Diseño estadístico

El censo es una metodología que implica la cobertura del 100% de las unidades de observación. En el marco de este proyecto, dichas unidades corresponden a los establecimientos de comercio, servicios e industria. En ciudades grandes resulta difícil abarcar la totalidad de establecimientos en un tiempo razonable para lograr tener una información oportuna. Además, esto puede generar altos costos operativos, por lo cual resulta necesario acudir a una metodología diferente. Por este motivo, en los casos en que se trate de poblaciones densas, un esquema de barrido basado en una muestra aleatoria puede ser suficiente. No obstante, en los casos en que las poblaciones no tengan un área extensa o no cuenten con una cartografía, se podría llevar a cabo un barrido censal que deberá cubrir el 100% del territorio.

Debido a la dinámica del sector, donde existen empresas o negocios que funcionan a puerta cerrada, sin aviso o alguna característica que permita identificarlos desde la calle, es importante aclarar que el barrido censal se reduce a capturar la información de lo visible y no garantiza la cobertura total de los establecimientos.

En cuanto la recolección de la información, se recomienda que se haga mediante tabletas o *smartphones*, también conocidos como Dispositivos Móviles de Captura (DMC). Para ello se requieren dispositivos que permitan guardar las coordenadas geográficas y tomar fotografías, para luego asociarlas con la encuesta de cada establecimiento. También se podría

realizar la recolección en papel. Las Tabla 2 y 3 presentan las ventajas y desventajas del uso de Dispositivos Móviles de Captura (DMC) y del uso de encuestas en papel para la recolección de información.

Tabla 2. Ventajas y desventajas de la recolección de información con DMC

Ventajas	Desventajas
• Captura la ubicación del establecimiento por medio del GPS. • La información llega rápidamente a la base de datos, se trabaja sin internet y se sincroniza en el momento que haya conexión. • Se puede descargar la base sincronizada en el momento que se requiera para generar frecuencias de preguntas y así poder ir revisando su comportamiento. • La herramienta con la que se diseñan los formularios utiliza objetos y componentes visuales, preguntas filtro, saltos, avisos al entrevistador, respuesta obligada, rotaciones de códigos y de variables, validaciones, inserción de elementos multimedia y permite simular la entrevista para comprobar que su funcionamiento esté de acuerdo a los requerimientos. • Las validaciones se realizan según el flujo del formulario, el cual habilita o no la entrada a ciertas preguntas, teniendo como opciones de respuesta únicamente las establecidas. En preguntas de cifras admite rango de valores, lo cual mejora la calidad de la información recolectada.	• En caso de robo o daño de la tablet se pierden las encuestas que no fueron sincronizadas.

• Se puede grabar automáticamente el audio completo de toda la entrevista para la supervisión posterior.	
--	--

Fuente: Elaboración propia.

Tabla 3. Ventajas y desventajas de la recolección de información en papel

Ventajas	Desventajas
• Es muy poco probable que se pierda el material por robo.	• Para capturar la ubicación del establecimiento es necesario llevar un GPS y transcribir la latitud y longitud en el papel. • Para tener una base de datos se debe esperar a que lleguen las encuestas a un punto central donde se revisan y se digitan. • El control del flujo del formulario (preguntas que se deben ir haciendo) depende del entrevistador. • En caso de inconsistencias, estas se detectan en el proceso de digitación o procesamiento de datos, donde en muchos casos es necesario regresar al establecimiento.

Fuente: Elaboración propia.

2.3.1. Población objetivo

Establecimientos de comercio, servicios e industria ubicados en las cabeceras municipales del país.

2.3.2. Subpoblaciones de interés

Establecimientos de comercio

Son los establecimientos que se dedican a las actividades de compra y venta al por mayor o al por menor de mercancías (bienes que van al mercado), nuevas o usadas no sometidas a transformaciones en su naturaleza intrínseca durante las actividades propias a su comercialización. Entre estos establecimientos se encuentran supermercados, pañaleras, puestos fijos en plazas de mercado, ferias artesanales permanentes, ferreterías, marqueterías, lavados de autos, montallantas, talleres de mantenimiento y reparación de vehículos automotores, reparación de enseres domésticos, maquinaria y equipo de industria. Para mayor detalle de las actividades incluidas, se recomienda consultar la Clasificación Industrial Internacional Uniforme de todas las actividades económicas adaptada para Colombia (CIIU REV. 4 A.C).

Establecimientos de servicios

Corresponde a los establecimientos que realizan transacciones con bienes intangibles generados cada vez que son solicitados, sobre los cuales no recaen derechos de propiedad por parte del usuario y no pueden ser negociados por separado de su producción, no son transportados ni almacenados. Los servicios pueden ser comunales, personales, religiosos, sociales, bancarios, financieros, de salud, recreativos, deportivos, culturales, asesorías, consultorías, entre otros. En esta clasificación se encuentran los restaurantes, hoteles, bares, consultorios, peluquerías, salas de tatuaje, oficinas, fondos de empleados, educación no formal, juegos de azar, cabinas telefónicas, funerarias, mensajería y encomiendas, reparación de computadores y equipos de oficina. Para mayor detalle de las actividades incluidas se

recomienda consultar la Clasificación Industrial Internacional Uniforme de todas las actividades económicas adaptada para Colombia (CIIU REV. 4 A.C).

Esta investigación no tiene interés en los siguientes servicios y por lo tanto no deben tenerse en cuenta: sector de la banca (bancos, corporaciones), intermediación financiera, seguros y fondos de pensiones y cesantías, administración pública, notarias, educación pública, colegios de educación media y universidades, iglesias, juntas de acción comunal y sedes de campañas políticas.

Establecimientos de industria

Son aquellos establecimientos que se dedican a la transformación física o química de materiales o componentes en productos nuevos, ya sea que el trabajo se efectúe con máquinas o a mano, en una fábrica o a domicilio, o que los productos se vendan al por mayor o al por menor. En este caso, se exceptúa la actividad que sea propia del sector de la construcción.

Ejemplos de este tipo de actividad son: fabricación de donas, helados, pan, tamales, empanadas, buñuelos, pasteles, jugos, donde el consumo no se realiza en el mismo espacio físico identificado. Elaboración o fabricación de prendas de vestir, zapatos, artículos de cuero, carpinterías, ebanisterías, papel y cartón, impresores, sustancias y productos químicos, ornamentación, productos de hierro y acero, muebles, colchones, joyas, juguetes, artículos deportivos, entre otros.

2.3.3. Variables de interés

Las variables de interés se dividen en variables de estudio y de clasificación; las primeras buscan medir los aspectos que responden a las preguntas de investigación y a los objetivos

del estudio, mientras que las segundas permiten obtener información específica para lograr una adecuada caracterización.

- **Variables de estudio:** variables asociadas a las características del establecimiento, del personal ocupado y, en general, todas las variantes que permitan determinar la informalidad del negocio.
- **Variables de clasificación:** actividad económica, localidad/comuna, clasificación de ingresos, establecimiento formal e informal.

2.3.4. Indicadores de interés

Los indicadores de interés serán aquellos que midan la formalidad o informalidad empresarial, el tiempo en la actividad, el acceso a seguridad social, las causas de la informalidad, las ventas, entre otros. Los indicadores que se generan a partir de la información pueden ser expresados en términos de totales, proporciones o promedios. En la Tabla 4 se presentan los tipos de indicadores y algunos indicadores que pueden ser calculados a partir de las variables del formulario.

Tabla 4. Tipos de indicadores que pueden ser calculados

Tipo de indicador	Indicador	Fórmula
Totales	• Total de establecimientos con matrícula mercantil. • Total de establecimientos con RIT. • Total de establecimientos con RUT.	• Cantidad de establecimientos con matrícula mercantil. • Cantidad de establecimientos con RIT. • Cantidad de establecimientos con RUT.

Porcentaje	• Porcentaje de personas con contrato a término indefinido. • Porcentaje de establecimientos que no llevan contabilidad. • Porcentaje de personas ocupadas con aportes a EPS. • Porcentaje de establecimientos que han sido víctimas de algún delito en los últimos 12 meses.	• Cantidad de personas con contrato a término indefinido / Total de personas ocupadas. • Cantidad de establecimientos que no llevan contabilidad / Total de establecimientos. • Cantidad de personas con aportes a EPS / Total de personas ocupadas. • Cantidad de establecimientos que han sido víctimas de algún delito en los últimos 12 meses / Total de establecimientos.
Promedio	• Promedio de personas que trabajan por establecimiento. • Gasto promedio de los establecimientos.	• Cantidad de personas que laboran en los establecimientos / Total de establecimientos. • Total de gastos / Total de establecimientos.

Fuente: Elaboración propia.

Para ilustrar la forma de cálculo, se presentan algunos ejemplos de indicadores de cada tipo.

- **Total de establecimientos con matrícula mercantil.** Se calcula a partir de la pregunta 24: “¿Este establecimiento tiene matrícula mercantil?”, todos aquellos que tengan el código 1 corresponden a los que respondieron “Sí”.
- **Porcentaje de establecimientos que obtuvieron o renovaron la matrícula mercantil en el último año.** Un porcentaje es una fracción calculada de los que cumplen con una determinada característica dentro de un subconjunto particular. En este caso serán aquellos establecimientos que en la pregunta 25: “¿La obtuvo o la renovó en el último año?” tengan el código 1, que significa que respondieron “Sí”, dividido todos los que respondieron “Sí” en la pregunta 24. La forma de calcularlo sería la siguiente:

$$\frac{\# \text{ de establecimientos que responden } \textit{¿La obtuvo o la renovó en el último año?} = \textit{Si. (P25=1)}}{\# \text{ de establecimientos que responden } \textit{¿Este establecimiento tiene matrícula mercantil?} = \textit{Si. (P24=1)}}$$

- **Total de personas ocupadas.** La pregunta 35: “Incluido usted, ¿cuántas personas trabajan en este negocio?” genera un valor para cada establecimiento encuestado, al sumarse sobre todos los establecimientos permite llegar al total de personas ocupadas.
- **Promedio de personas ocupadas por establecimiento.** La pregunta 35: “Incluido usted, ¿cuántas personas trabajan en este negocio?”, debe ser sumada en todos los establecimientos encuestados para llegar al total de personas ocupadas. Este valor se divide en el total de establecimientos encuestados y el resultado corresponde al indicador.
- **Porcentaje de personas con contrato a término indefinido:** La pregunta 47: “*Qué tipo de vinculación laboral tiene el personal ocupado por este establecimiento*” debe sumar a todos los propietarios, personal permanente, personal temporal y aprendices que tienen contrato a término indefinido en un establecimiento, para luego sumar este valor en todos los establecimientos encuestados y de esta forma llegar al total de personas que tienen esta modalidad de contratación. Finalmente, el valor obtenido se divide en el total de personas ocupadas para así obtener el indicador.

2.3.5. Cobertura geográfica

La metodología está diseñada para cubrir áreas geográficas urbanas del territorio nacional.

2.3.6. Niveles de desagregación

El nivel de desagregación geográfica es por cabecera municipal. La desagregación temática está asociada con las características del establecimiento sobre informalidad empresarial.

2.3.7. Metodología muestral

En aquellas poblaciones que por su alta densidad poblacional o grandes extensiones territoriales no sea posible llegar a todas las áreas geográficas y por ende no se pueda encuestar a todos los establecimientos, se podrá realizar una muestra aleatoria que sea representativa. Así, se podrán calcular los indicadores de interés a partir de la información recolectada en la encuesta.

2.3.7.1. Unidades estadísticas

Unidades de observación: son aquellas unidades a través de las cuales se obtiene información de variables de interés, en este caso las preguntas del formulario se refieren al establecimiento, así que las unidades de observación en este estudio corresponden a los establecimientos industriales, comerciales y de servicios.

Unidades de análisis: son aquellas unidades para las cuales se calculan los indicadores a través de la información recolectada por la encuesta. Para este estudio se considera como unidad de análisis al establecimiento.

Unidades de muestreo: corresponden a las unidades que son susceptibles de ser seleccionadas en una muestra aleatoria. En este estudio se pueden seleccionar áreas geográficas y establecimientos.

Informante idóneo: para el presente estudio se considera como informante idóneo a la persona que suministra los datos de la encuesta. El informante idóneo para responder es el dueño, gerente, socio, administrador o persona responsable del manejo del establecimiento que cuente con un conocimiento profundo que le permita responder a las preguntas del formulario.

2.3.7.2. Marco muestral

Un marco de muestreo es un listado de todos los establecimientos que se desean encuestar con su respectiva identificación y dirección. El Registro Único Empresarial (RUES) podría llegar a considerarse como el marco de muestreo, sin embargo, este listado no tiene a los establecimientos informales que son parte fundamental de esta investigación.

Al no existir un listado de todos los establecimientos de interés será necesario llegar a ellos de una forma sistemática. El listado completo puede construirse a partir de la selección de áreas geográficas y tras realizar un barrido en las zonas que sean seleccionadas. A esta fase se le suele denominar “enlistamiento”.

Para seleccionar las áreas geográficas se puede usar la base cartográfica del Departamento Administrativo Nacional de Estadística (DANE), que se encuentra a nivel de manzana y contiene la información de la cantidad de establecimientos de comercio, servicios e industria. La cartografía del DANE es de libre acceso y se puede realizar desde el geoportal² o desde redatam³. El trabajo con la cartografía del DANE requiere de un visor de datos geográficos en el computador y conocimientos mínimos en el manejo mapas.⁴

Una manzana cartográfica está identificada de manera única con un código de 14 dígitos y el visor de datos la relaciona con un lugar geográfico específico (polígono) sobre el terreno. De esta forma, el visor le permitirá generar un mapa con las indicaciones para que la persona de campo llegue al lugar a realizar el levantamiento de información. Los 14 dígitos

² <https://geoportal.dane.gov.co/?descarga-marco-geoestadistico-nacional>

³ <http://systema59.dane.gov.co/cgibin/RpWebEngine.exe/PortalAction?BASE=CG2005BASICO>

⁴ Para este propósito, puede utilizarse QGIS, el cual es un visor de datos geográficos de libre utilización.

del código corresponden a la identificación de departamento (2 dígitos) | municipio (3 dígitos) | clase (1 dígito) | sector (4 dígitos) | sección (2 dígitos) | manzana (2 dígitos). En la “clase”, el número 1 identifica las zonas urbanas; la manzana es el lugar geográfico (polígono) más pequeño dentro de la cartografía del DANE.

2.3.7.3. Diseño muestral

El objetivo de la muestra aleatoria será lograr un equilibrio entre el esfuerzo requerido para el levantamiento de la información y la precisión de los indicadores que se desean calcular. De modo que será necesario determinar la cantidad de áreas geográficas que se deben barrer y la cantidad de establecimientos que se deben encuestar.

Con este documento se hace entrega de un procedimiento que calcula la cantidad de áreas geográficas que deben ser recorridas y que además realiza la selección de la muestra. En términos generales, el procedimiento clasifica las manzanas cartográficas en tres grupos: el primero agrupa las manzanas de mayor densidad de establecimientos, el segundo aúna aquellas de densidad media y el tercer grupo reúne las manzanas de baja densidad. Para realizar la clasificación se usan los algoritmos de estratificación óptima (Lavalle & Hidroglou, 1988), (Baillargeon & Rivest, 2009). Esto es posible debido a que la cartografía DANE cuenta con información auxiliar del número de comercios, servicios e industrias en cada manzana cartográfica.

Las áreas geográficas seleccionadas se denominarán Unidades de Cobertura Urbana (UCU), que se define como un área (edificada o sin edificar) delimitada por vías de tránsito vehicular o peatonal de uso público (calles, carreras, avenidas, diagonales o senderos peatonales). Asimismo, las UCU pueden estar delimitadas por elementos naturales como caños, ríos y quebradas que sean de carácter permanente. En cada UCU será necesario

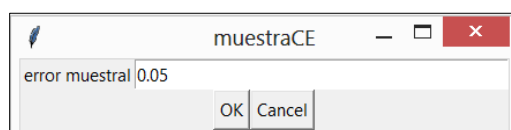
realizar un barrido censal según las indicaciones dadas en el documento de lineamientos operativos de recolección.

2.3.7.4. Tamaño de la muestra

Para establecer el tamaño de la muestra se debe determinar el margen máximo de error, este valor es fijado por el investigador y en general se recomienda que no sea superior al 5%. Esta es una medida de precisión, cuando su valor es cero significa que no existe error asociado al muestreo y solo se consigue cuando se encuesta a todos los establecimientos (censo); en la medida en que se “tolere” un margen de error más alto, el tamaño de la muestra será menor y los indicadores serán menos precisos. Por ejemplo, si a partir de la muestra aleatoria se estima que el porcentaje de establecimientos con matrícula mercantil es del 70% y el margen de error es del 5%, se debe considerar que el verdadero porcentaje, si se hubiera preguntado a todos los establecimientos, estaría en el rango de 65% a 75% (se resta y se suma el margen de error al indicador estimado).

Con este documento se entrega un video corto sobre el uso de la aplicación para la selección de la muestra, donde el usuario solo debe ingresar el error, que por defecto se fija en el 5% pero que puede ser modificado. La Figura 1 presenta el cuadro de diálogo donde se debe ingresar el valor del error de muestreo, el usuario puede elegir continuar con el valor que está por defecto y oprimir el botón OK.

Figura 1. Cuadro de diálogo para la selección de la muestra



Fuente: Elaboración propia.

El programa calcula el tamaño de la muestra y la selecciona a nivel de UCU. Las fórmulas para calcular el tamaño de la muestra a nivel de UCU se encuentran en el Anexo I.

2.3.7.5. Algoritmos de selección de muestra

Los algoritmos para la selección de la muestra se encuentran implementados en el programa. En caso de requerir mayores detalles sobre este proceso, en el Anexo II se encuentra el procedimiento o se recomienda revisar (Särndal, Swensson, & Wretman, 1992).

2.3.7.6. Tamaño de muestra por UCU

En las áreas geográficas seleccionadas se debe aplicar la encuesta en todos los establecimientos. Sin embargo, en caso de que se presenten situaciones económicas, políticas o sociales que impidan que esto se pueda realizar, se podrá seleccionar una muestra de establecimientos.

De acuerdo con el (DANE, 2007), el tamaño **mínimo** de la muestra de establecimientos a encuestar según su actividad económica en cada área geográfica seguirá las reglas establecidas en la Tabla 5. Se cuenta con tres columnas, una por cada actividad económica. En el caso de los establecimientos de comercio se deben seleccionar todas las unidades cuando la cantidad en la UCU sea menor o igual que 5. Asimismo, se deben seleccionar 5 cuando la cantidad de establecimientos en la UCU sea mayor que 5 pero menor o igual que 50, y se debe seleccionar el 10% cuando la cantidad de establecimientos de comercio en la UCU sea superior a 50. Para los establecimientos de servicios aplican las mismas reglas. En el caso de industria se deben seleccionar todas cuando la UCU tenga 12 o menos industrias y se deben elegir mínimo 12 cuando haya más de esta cantidad en la UCU.

Tabla 5. Tamaño mínimo de muestra por UCU y actividad económica

Comercio	Servicios	Industria
Si $N_c \leq 5$ hacer $n_c = N_c$	Si $N_s \leq 5$ hacer $n_s = N_s$	Si $N_I \leq 12$ hacer $n_I = N_I$
Si $5 < N_c \leq 50$ hacer $n_c = 5$	Si $5 < N_s \leq 50$ hacer $n_s = 5$	Si $N_I > 12$ hacer $n_I = 12$
Si $N_c > 50$ hacer $n_c = N_c * 10\%$	Si $N_s > 50$ hacer $n_s = N_s * 10\%$	

Fuente: Adaptado de (Departamento Administrativo Nacional de Estadística (DANE), 2007, pág. 42).

Donde:

N_c representa la cantidad de establecimientos de comercio en la manzana.

n_c es la cantidad de establecimientos de comercio que deben ser seleccionados.

N_s representa la cantidad de establecimientos de servicios en la manzana.

n_s es la cantidad de establecimientos de servicio que deben ser seleccionados.

N_I representa la cantidad de establecimientos de industria en la manzana.

n_I es la cantidad de establecimientos de industria que deben ser seleccionados.

Considérese por ejemplo una UCU donde no sea posible encuestar a todos los establecimientos y se requiera seleccionar una muestra. Suponga que la UCU tiene 63 establecimientos de comercio, 37 de servicios y 10 de industria. Según la Tabla 5, como el número de comercios es mayor que 50, se debe seleccionar al 10% de los establecimientos de comercio, que serían 6. Como la cantidad de servicios está entre 5 y 50, se deben seleccionar 5 de servicios. Y como el número de industrias en la UCU es menor que 12, entonces se deben encuestar a todas las 10 industrias.

Para complementar, considérese otro ejemplo en el que una UCU tenga 23 establecimientos de comercio, 4 de servicios y 15 de industria. Como el número de comercios está entre 5 y 50, de acuerdo con la Tabla 5, se deben seleccionar 5. Debido a que la cantidad

de establecimientos de servicios es menor que 5, entonces deben encuestarse a todos y, como el número de industrias en la UCU es mayor que 12, entonces se deben seleccionar 12.

2.3.7.7. Formatos de control y selección de muestra

Los formatos de control permiten entender las particularidades que tienen las Unidades de Cobertura Censal (UCU) en las que se va a trabajar, medir la cobertura, establecimientos desocupados e identificar los establecimientos que no hacen parte del interés de la investigación (hospitales, edificios gubernamentales, grandes cadenas, iglesias, etc.).

El formato de barrido censal permite construir el listado de establecimientos por cada manzana e identificar el sector económico al que pertenece: comercio, servicios o industria. La Figura 2 presenta una imagen del formato de barrido censal, el encabezado será diligenciado por el supervisor y podrá hacerlo desde la oficina, el resto de la información como el número, nombre comercial, dirección y si el establecimiento es visible o no desde la calle serán diligenciados por el encuestador, junto a las observaciones que puedan surgir mientras va realizando el barrido.

Figura 2. Formato de barrido censal de las unidades económicas

CENSO EMPRESARIAL FORMATO DE BARRIDO CENSAL DE LAS UNIDADES ECONÓMICAS						
		UCU			Departamento Municipio	
Sector		Sección		Manzana		Cantidad Comercio Servicios Industria Total
Nro.	Nombre comercial	Dirección (incluya local si aplica)	Actividad	Visible desde la calle		Observaciones
			Comercio	Si	No	
			Comercio	Si	No	
			Comercio	Si	No	
			Comercio	Si	No	
			Comercio	Si	No	

Fuente: Elaboración propia.

En caso de que se requiera la selección de establecimientos, se debe garantizar la rigurosidad del método de muestreo, que podrá hacerse en terreno y estará bajo la responsabilidad del supervisor de campo. Esto se hará con base en el diligenciamiento del formato de selección de unidades económicas presentado en la

Figura 3.

Figura 3. Formato de selección de unidades económica

FORMATO DE SELECCIÓN DE UNIDADES ECONÓMICAS			
Id manzana: _____	Fecha: _____		
Dirección (cruce) inicio de recorrido: _____			
	COMERCIO	SERVICIOS	INDUSTRIA
Número de:			
Tamaño de la muestra =			
Medida de Tamaño (Micros de / Muestra) =			
Tamaño del segmento =			
Hora y minutos =			
Tabla seleccionada =			
Arranque aleatorio =			

*Verifique que la cantidad en la cartografía se aproxime a lo observado, de lo contrario haga un recuento estimado de unidades económicas en la manzana

*Agregue el tamaño de muestra solicitado para cada sector en esta manzana, recuerde que el operativo se detiene cuando complete la muestra mínima

*Divida el número de unidades económicas entre el tamaño de la muestra (incluyendo la sobremuestra)

*De la medida de tamaño, tome el valor antes de la coma que separa las cifras decimales

*Registre la hora exacta (hora+minutos) de este momento

*Tome el último dígito de la hora que escribió antes, si termina en cero vaya devolviéndose hasta que encuentre un número diferente de cero

*Vaya a la "TABLA ALEATORIA-SELECCIÓN DE UNIDADES ECONÓMICAS" con el consecutivo dado en el recuadro anterior

Fuente: Elaboración propia.

El manejo y uso de estos formatos se encuentra con mayor detalle en el documento de lineamientos operativos de recolección.

2.3.8. Metodología censal

Esta metodología debe ser aplicada cuando la extensión territorial y la población no es tan amplia. Asimismo, se debe utilizar en los casos en que no se cuente con una cartografía precisa que permita realizar una selección adecuada de las manzanas.

La operación de un censo se puede hacer mediante la recolección por rutas o barrido censal completo. La metodología de recolección por rutas es usada especialmente en áreas dispersas, en general es preferida en zonas rurales por las dificultades de acceso y la baja densidad poblacional. A través de esta técnica se define un trayecto que puede tardar varios días e implica acceso por ríos, caminos de herradura o trochas. La *metodología de barrido censal* completo es usual en cabeceras urbanas y consiste en definir áreas geográficas donde se llevará a cabo un recorrido sistemático que abarque la totalidad de los establecimientos visibles.

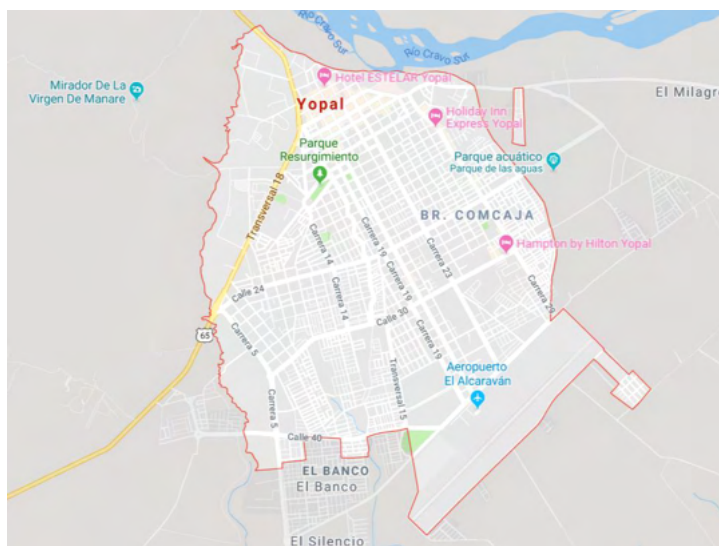
Para asegurar la cobertura, se recomienda utilizar la codificación de Divipola⁵, la cual le brinda un código e identidad única a cada departamento, municipio, corregimiento departamental y centro poblado. En cada municipio será muy útil contar con los mapas en estructura censal, en papel o en digital.⁶ En este mapa se podrá delimitar el espacio total que debe ser censado de acuerdo a zonas de coordinación operativa, áreas operativas urbanas y Unidades de Cobertura Urbana (UCU). De igual forma, se podrán apoyar en otro tipo de mapas como los generados por Google: <https://www.google.com/maps>. Sin embargo, en caso de usar esta herramienta se recomienda asignarle una codificación a cada área de trabajo con el fin de poder realizar un adecuado seguimiento de la cobertura. La

⁵ Codificación de la división político-administrativa de Colombia brindada por el DANE.

⁶ Para visualizar en dispositivos electrónicos los mapas se pueden descargar en <https://geoportal.dane.gov.co/?descarga-marco-geoestadistico-nacional> además cuenta con un manual de uso, una guía de descarga y visualización.

Figura 4 muestra un ejemplo de un mapa del municipio de Yopal, Casanare.

Figura 4. Área geográfica del municipio de Yopal



Fuente: Adaptado de Google Maps.

Una vez identificada el área a censar se debe dividir en áreas operativas teniendo en cuenta la cantidad de establecimientos económicos. El propósito es que las áreas generen aproximadamente el mismo esfuerzo de trabajo, aunque sus extensiones territoriales parezcan inequitativas. De esta manera permitirá que se realice una planeación más acertada definiendo el número estimado de días de trabajo en cada área operativa.

En caso de contar con la cartografía e información del DANE, esta podrá ser usada como apoyo. En caso contrario, se podría hacer con base en el conocimiento que el equipo técnico tiene del territorio. La Figura 5 presenta una división del municipio de Yopal en áreas operativas⁷, dependiendo del tiempo que se tenga para la recolección, se podrá disponer de

⁷ Esta división se realizó de forma arbitraria, es importante realizar esta división de una forma equitativa en cuanto a la cantidad de establecimientos que se esperan encontrar.

un equipo de trabajo en cada área, en este caso se requeriría de 6 equipos, o en su defecto conformar 3 equipos y asignar dos áreas a cada uno.

Es muy importante que el área operativa se delimite en áreas geográficas más pequeñas, preferiblemente que sean equivalentes a manzanas o secciones cartográficas. En las áreas con mayor densidad de establecimientos, se recomienda que la división sea a nivel de manzana. Google cuenta con una herramienta muy útil que permite diseñar los mapas y que puede ser usada con este propósito: <https://www.google.com/maps/d/u/0/>.

Figura 5. División del área a censar en áreas operativas



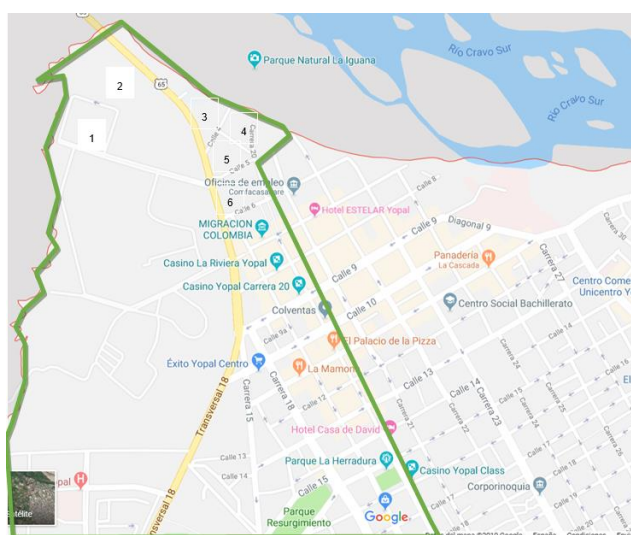
Fuente: Adaptado de Google Maps.

Para realizar un seguimiento adecuado y posteriormente un apropiado análisis de los resultados, es necesario identificar de forma clara el área geográfica donde se levanta la información y que a su vez pueda ser fácilmente verificada. Lo mejor es usar la nomenclatura DANE, pero en caso de no contar con ella se debe seguir una codificación que permita la fácil ubicación del personal en terreno sobre los mapas operativos y que desde la base de

datos pueda ser identificada con facilidad. La mínima representación geográfica estará delimitada por la Unidad de Cobertura Urbana (UCU) como se ilustra en la

Figura 6. En esta se muestra una codificación para algunas áreas, pero todas deberán codificarse: cada manzana dentro del área operativa deberá tener un consecutivo.

Figura 6. Ejemplo de codificación de áreas operativas en unidades de cobertura urbanas



Fuente: Adaptado de Google Maps.

Como se ha dicho, para facilitar la ubicación en terreno, las UCU siempre deben estar delimitadas por vías de tránsito vehicular o peatonal de uso público (calles, carreras, avenidas, diagonales o senderos peatonales) o también a partir de elementos naturales como caños, ríos, quebradas, entre otros, siempre y cuando sean de carácter permanente. En la Figura 7, la UCU codificada con el número 6 se encuentra claramente delimitada por la transversal 18 hasta la carrera 20 entre las calles 5 y 6; la correcta codificación de esta área en la base de datos debería ser 850010106, donde el código 85 corresponde al código DANE del departamento del Casanare, 001 es el código DANE del municipio de Yopal en el

departamento del Casanare, el código 01 hace referencia al área operativa y el código 06 corresponde a la UCU.

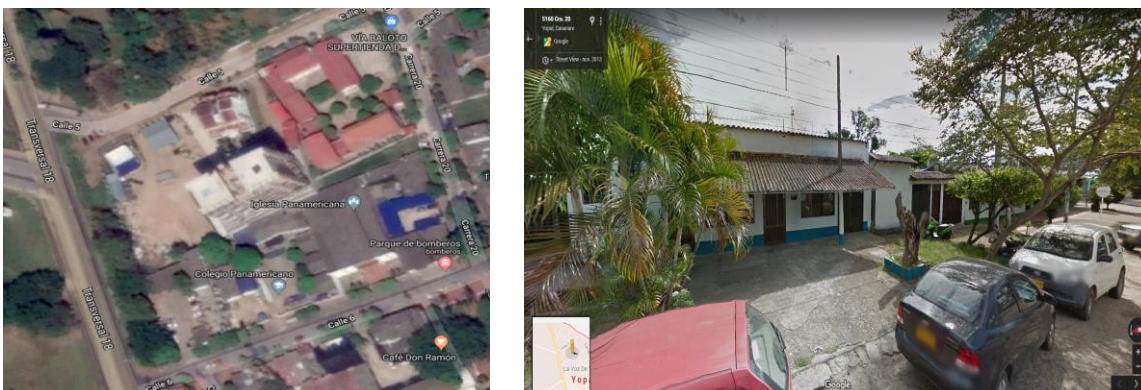
Figura 7. Delimitación de la UCU 850010106



Fuente: Adaptado de Google Maps.

Para el adecuado trabajo de los equipos operativos se debe tener acceso a los mapas de forma impresa o digital. De igual manera, se recomienda que previo a la salida de campo se realice un reconocimiento del terreno a través de Google tal como se muestra en la Figura 8.

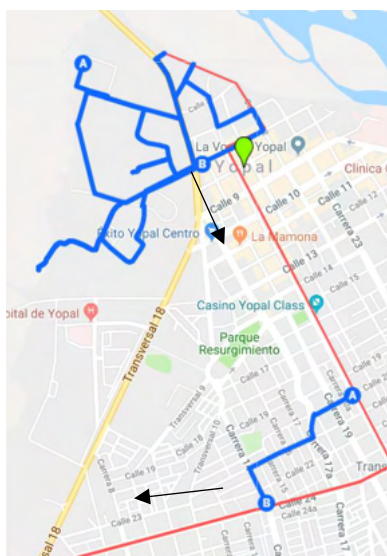
Figura 8. Reconocimiento digital de una UCU



Fuente: Google Maps.

Para la organización operativa en esquema de barrido censal, en cada área operativa los equipos deben empezar en puntos distintos hasta llegar a un punto en común. La Figura 9 muestra la ruta⁸ (parte superior) y el área (parte inferior) de trabajo para el día uno. Los equipos deberán continuar el trabajo cada día definiendo un orden en el recorrido, como se establece en las flechas, de modo que se cubra un área geográfica superior a una UCU, hasta que los dos equipos finalicen en un punto en común, con el fin de cubrir la totalidad del área operativa. Este proceso debe replicarse en las demás áreas operativas del municipio.

Figura 9. Ubicación de equipos en esquema de barrido censal



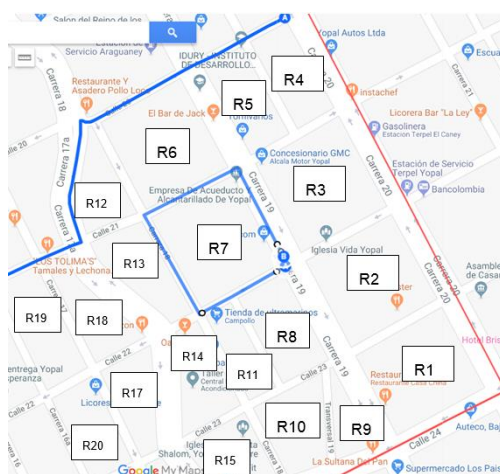
Fuente: Adaptado de Google Maps.

Para empezar el recorrido de un área se definen las UCU que se trabajarán cada día, esto dependerá principalmente de la densidad de establecimientos a encuestar. Una vez el equipo se haya ubicado en el área correspondiente, debe empezar a trabajar en la UCU desde

⁸ En este caso se define una ruta por ser un área dispersa.

la calle mayor con carrera mayor. En el caso de la Figura 9, el equipo que trabajará el polígono ubicado en la parte inferior, deberá iniciar su trabajo en la UCU que se delimita en la carrera 20 con calle 20 e ir haciendo el barrido hasta que finalice el polígono. Luego en forma de culebrilla deberá devolverse por la siguiente carrera para ir recorriendo las UCU. La Figura 10 presenta un esquema del recorrido que debe hacer el segundo equipo de trabajo, donde R1 representa la UCU que se barre en el recorrido 1 y R2 la UCU que se barre en el segundo recorrido y así sucesivamente.

Figura 10. Esquema de recorrido para barrido censal de las UCU de un área operativa



Fuente: Adaptado de Google Maps.

Una vez ubicado el equipo en una UCU, por ejemplo, en la marcada con R7 en la Figura 10, el encuestador identificará y reconocerá el terreno para lograr llegar a todos los establecimientos. Para ello el encuestador se debe ubicar en la esquina de la calle mayor con carrera mayor de cada manzana señalada en la cartografía y desplazarse siguiendo el sentido de las manecillas del reloj. Todo encuestador debe iniciar y terminar el recorrido en el mismo punto, tal como se ilustra en la Figura 11 (recorrido desde el punto A hasta el punto B).

- FU: Fuera de Universo.¹⁰

Si durante el recorrido se encuentran centros comerciales o unidades con propiedad horizontal de oficinas u otras actividades económicas, se recomienda:

- Recorrer uno a uno todos los pisos empezando por el primero.
- Iniciar el recorrido por la derecha tomando como referencia la entrada número 1 o la entrada principal.
- Averiguar si hay sótanos y si existen establecimientos en ellos.
- Entrar y salir por la misma puerta.
- Usar las escaleras para ir llevando el control de la edificación.

Los equipos deben realizar los desplazamientos diarios de forma ordenada y seguir el programa establecido para recorrer todos y cada uno de los establecimientos que componen la UCU. Debe llevarse a cabo la planeación estipulada de la recolección y, al finalizar la jornada de trabajo, el equipo deberá regresar al centro operativo para realizar los respaldos y la sincronización de la información levantada. En caso de no haber terminado de encuestar a todos los establecimientos de la UCU, al día siguiente se deberá volver para terminar con el barrido. En caso de que la planeación se empiece a ver afectada, se debe aumentar el equipo de trabajo o en su defecto aumentar los días establecidos para la operación censal.

2.3.9. Esquemas de validación y consistencia

Los esquemas de validación se encargan de crear reglas que permiten certificar la integridad y consistencia de los datos, así como controlar posibles errores de flujo del

¹⁰ Este resultado se da cuando el establecimiento no es parte de la población objeto de estudio.

formulario.¹¹ Para esto se definen algunas normas o esquemas según lo establecido por el Sistema Estadístico Nacional en lo relativo a la captura de los datos.¹²

2.3.10. No respuesta e imputación

Generalmente en una investigación se presentan dos tipos de no respuesta, una es cuando la persona en el establecimiento seleccionado se niega a responder algunas preguntas del cuestionario y la otra es cuando hay imposibilidad de levantar toda la información de la encuesta por cualquier motivo.

El primer tipo de no respuesta se puede solucionar a través de métodos de imputación que buscan asignar información plausible, usando la información de individuos similares. La mayoría de textos de análisis de datos presentan los métodos de imputación con sus ventajas y desventajas; lo más importante es controlar adecuadamente los niveles de no respuesta y tratar de evitar esta situación. No obstante, en este caso se recomienda que la no respuesta parcial (cuando solo falta la información de algunas preguntas) se maneje como un dato faltante y con la etiqueta “No responde”.

En el caso de los establecimientos que rechazan la encuesta, se recomienda reportar la cobertura que se calcula a partir de los establecimientos que respondieron a la encuesta sobre los que debían responder.¹³

¹¹ Un error en el flujo del formulario se da cuando una pregunta que debía ser contestada no se diligenció o cuando se diligenció una pregunta y no debía ser respondida.

¹² Para mayor información se recomienda ver el manual con las especificaciones de validación y consistencia para la recolección de los censos empresariales.

¹³ Esta información se encuentra en el formato de control de cobertura.

2.3.11. Estructura y diseño de la base de datos

Para el levantamiento de información se usarán Dispositivos Móviles de Captura (DMC), que permiten una sincronización automática de los datos, lo cual garantiza la confiabilidad de la información. Para salvaguardar la información ante una falla eventual en la transmisión de los datos, se recomienda que el supervisor de campo haga un respaldo diario de los archivos individuales de cada encuestador, con el fin de poder sincronizar nuevamente los datos en caso de que se requiera.

Los datos transmitidos se deben consolidar en las tablas del sistema de información donde se desarrolló el aplicativo de captura y se deben almacenar en el sistema de gestión de bases de datos con todas las variables del formulario y formatos de control de cobertura.

2.4. Diseño del operativo de campo

En esta sección se presentan algunos aspectos del diseño operativo, para un mayor detalle se recomienda remitirse al documento de lineamientos operativos de recolección.

2.4.1. Aspectos relevantes

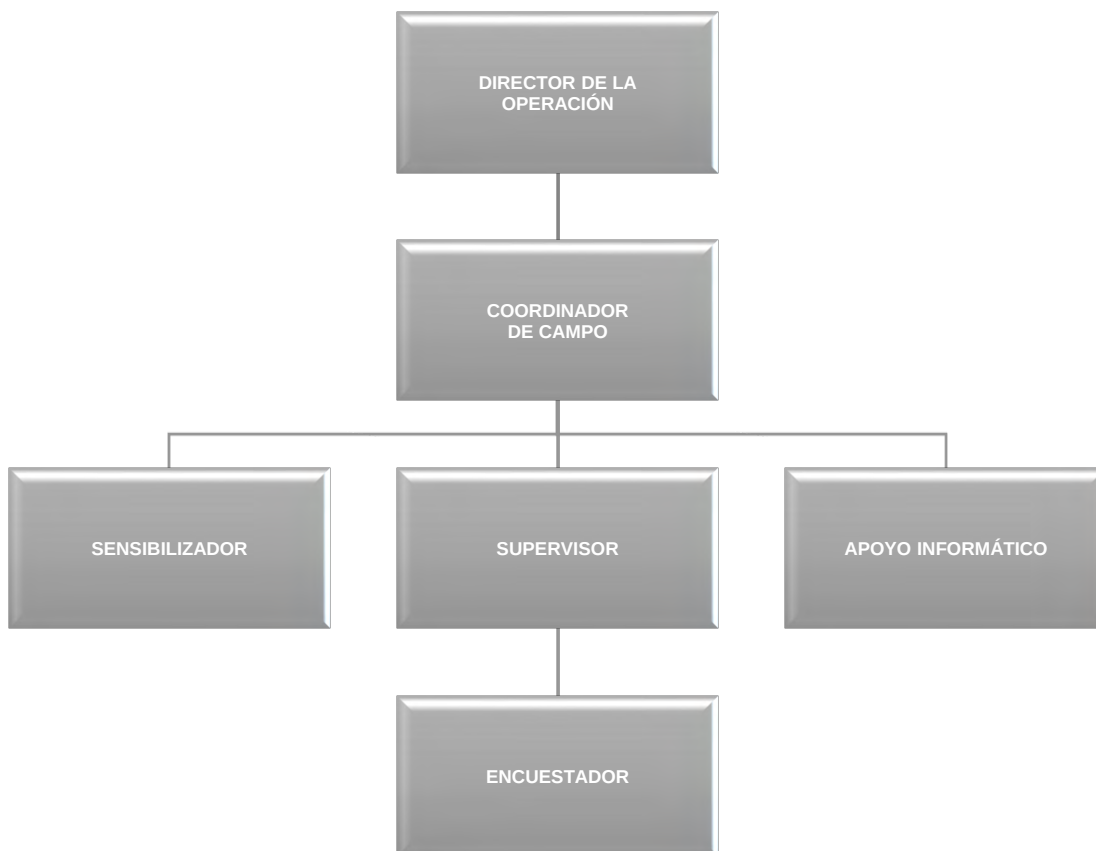
Es necesario tener clara la delimitación de las áreas geográficas. Por ejemplo, se deben excluir aquellas áreas que sean parques, estadios, aeropuertos, colegios, clínicas y establecimientos penitenciarios para evitar la sobre cobertura. Se recomienda manejar la nomenclatura DANE para identificar cada UCU. Se debe definir un cronograma para la operación y realizar la planeación asegurando que se cumplan los tiempos fijados para la misma. Es importante tener presentes las siguientes actividades:

- Plan operativo: actividades, recursos económicos y humanos, cronogramas y dispositivos para el levantamiento de la información.
- Hacer seguimiento de forma permanente durante la ejecución del operativo e ir realizando los ajustes que sean requeridos.
- Solicitar con debida anticipación la base cartográfica con la cantidad establecimientos de comercio, servicios e industria a nivel de manzana cartográfica, así como las capas necesarias para la generación de la cartografía.
- Realizar el alistamiento: probar los dispositivos de captura y elementos necesarios para la recolección.
- Anticipar situaciones que puedan poner en riesgo a las personas encargadas de la recolección, y a los equipos que se usarán.

2.4.2. Esquema organizacional

Los equipos de trabajo en terreno están compuestos por un supervisor, un sensibilizador y máximo 5 encuestadores en un esquema jerárquico como se presenta en la Figura 12.

Figura 12. Esquema organizacional



Fuente: Elaboración propia.

El rol del sensibilizador podrá ser asumido por el supervisor, quien se encarga de lograr que las fuentes a entrevistar respondan el censo sin desconfianza, mediante el uso de procesos comunicativos que concienticen a las fuentes donde se presentará el encuestador, señalen la entidad que representa y la importancia de la labor que realiza.

Para los roles operativos se cuenta con un coordinador de campo que es el responsable de las actividades previas a la aplicación de la encuesta y de la supervisión de los equipos de trabajo en campo. Tiene a su cargo el manejo de la encuesta en todos los aspectos operativos:

realizar la asignación de cargas de trabajo a cada equipo y verificar el cumplimiento de la muestra. Además, debe consolidar el informe de cobertura, hacer el análisis de calidad de la información recolectada y generar los informes de avance y novedades presentados durante el trabajo realizado en campo.

La persona que apoya desde el área informática se encarga de realizar el alistamiento de los Dispositivos Móviles de Captura (DMC), de descargar los formularios electrónicos necesarios para el desarrollo de las encuestas y de prestar apoyo en la solución de los errores generados en el aplicativo, con el fin de garantizar el normal funcionamiento de dispositivos.

El proceso de supervisión buscará asegurar la calidad de la información recolectada en los establecimientos establecidos en la muestra. Quien ejerza el papel de supervisor debe conocer la investigación, acompañar de forma permanente a su grupo, resolver las dudas en el momento oportuno y solucionar las inquietudes que surjan en el desarrollo del trabajo. Esta persona hará el reconocimiento del terreno y ubicará a los encuestadores en los respectivos establecimientos seleccionados. Asimismo, acompañará al equipo de encuestadores en caso de que se presenten dudas o sea necesario realizar aclaraciones; estará al tanto de las encuestas pendientes por realizar con el fin de establecer las revisitas y culminar con éxito el proceso. Finalmente, se encargará de verificar el adecuado cierre de cada encuesta en el aplicativo. Para mayor detalle sobre las funciones de cada uno se recomienda revisar los manuales operativos.

2.3.1. Verificación previa

Antes de llevar a cabo la ejecución de las encuestas se recomienda realizar una verificación que garantice que las jornadas de trabajo se desarrollarán con éxito. Algunos elementos para tener en cuenta son:

- Grupos de trabajo completos.
- Cartografía lista y cargada de las áreas geográficas a trabajar.
- Personal capacitado.
- DMC cargados con la muestra de los segmentos cartográficos y formularios.
- Kit tecnológico (cargadores, baterías, etc.).
- Kit operativo (formatos, lápiz, esferos, etc.).
- Transporte.

2.4.3. Diseño de la capacitación del personal

La información recolectada será útil y verídica si se realiza una capacitación efectiva del personal encargado, ya que son los encuestadores los encargados de generar las condiciones óptimas para que las personas que accedan a contestar la encuesta sientan la confianza para responder con la verdad.

El proceso de capacitación del personal se debe realizar de forma presencial y de acuerdo a la estructura jerárquica del equipo, es decir, se inicia con los coordinadores de campo, luego con los supervisores y finalmente con los encuestadores. El presente documento metodológico desarrolla algunos de los contenidos que se deben presentar en la capacitación. De igual forma, incluye los manuales operativos que presentan en detalle las responsabilidades de cada rol, los aspectos generales del censo, los conceptos básicos y algunas técnicas para lograr una encuesta efectiva.

Luego de realizar la capacitación se debe efectuar una evaluación que permita determinar si los candidatos comprendieron y se apropiaron de todos los conceptos y procedimientos. Del puntaje obtenido en la evaluación se seleccionarán a los aspirantes que

obtengan las mejores calificaciones. Aquellos con las puntuaciones más altas se seleccionarán como coordinadores de campo, los siguientes como supervisores y finalmente los encuestadores, siempre y cuando aprueben la evaluación.

2.4.4. Rutas de trabajo

Las rutas se deben crear de manera previa a la recolección de datos y necesitan ser definidas de acuerdo a las zonas de trabajo, preferiblemente sobre planos cartográficos. Para cada ruta se debe obtener información sobre el número aproximado de establecimientos con el fin de determinar los puntos de encuentro, tiempos de desplazamiento, posibles inconvenientes y si el grupo de trabajo debe ser grande o pequeño.

Es importante realizar una asignación eficiente de las zonas a trabajar, no solo para un día en particular sino para todo el tiempo que dure la recolección. De esta forma, se podrán definir rutas que agrupen la mayor cantidad de UCU de un modo más eficiente y a través de desplazamientos más cortos.

2.4.5. Esquema de barrido de áreas geográficas

En este tipo de esquema de recolección el equipo de trabajo recorre simultáneamente una zona o un segmento cartográfico hasta identificar todos los establecimientos. Para ello primero se debe realizar un reconocimiento del área que se va a trabajar, luego un enlistamiento de los establecimientos de la población objetivo, los cuales deben registrarse en el formato de barrido censal. En caso de que se requiera una muestra de establecimientos, el supervisor la seleccionará usando las tablas de selección.

Para el correcto reconocimiento de la manzana, primero se debe identificar el cruce de la calle mayor con carrera mayor, solo así podrá asegurarse de que el encuestador está ubicado en la manzana correcta. Una vez ubicado en la manzana, si no hay numeración, el recorrido se debe iniciar en la esquina más nororiental. Por ejemplo, si se encuentra en una manzana que está ubicada entre las calles 23 y 22 con carreras 20 y 21 se debe iniciar por la esquina ubicada en la calle 23 con carrera 21, como se muestra en la Figura 13.

Figura 13. Ejemplo del recorrido de una manzana



Fuente: Presentación esquemas de recolección DANE.

Una vez ubicado en el punto de inicio, el recorrido de la manzana se debe realizar de tal manera que, al caminar por la calle, el **hombro derecho** se encuentre siempre en dirección hacia la pared. El recorrido debe iniciar y terminar en el mismo punto. Cuando el área asignada esté compuesta por varias manzanas, se deben recorrer individualmente, realizando el ejercicio mencionado anteriormente. Es importante hacer un recorrido previo de la manzana con el fin de reconocer el espacio y confrontar la forma del terreno con la del plano.

La manzana se debe recorrer en su totalidad, independientemente de los problemas geográficos que pueda encontrar en su recorrido.

2.4.6. Recolección de la información

La principal recolección de la información está a cargo de los encuestadores, pero los supervisores son quienes definen los establecimientos que se deben encuestar. Asimismo, los supervisores tienen la tarea de ir generando los informes sobre la carga eficiente de trabajo.

En las zonas en las que hay baja concentración de establecimientos se tendrán muestras pequeñas de información, así que serán necesarios pocos encuestadores, es decir, que con un grupo de trabajo será suficiente. Este equipo debe estar liderado por un supervisor o apoyo operativo, que estará encargado de garantizar la cobertura y la calidad de la realización del cuestionario. La encuesta es presencial, la metodología de investigación se lleva a cabo a través del contacto directo con los individuos mediante una tableta o teléfono inteligente, que tendrán incorporados los esquemas para validar y garantizar la consistencia de la información de modo que se evite el riesgo de cometer errores en el diligenciamiento de los datos o que se hagan preguntas sin secuencia lógica.

Se espera que el supervisor asigne a cada encuestador una carga de trabajo diaria, en promedio que sea de 25 encuestas. También será necesario volver a visitar los establecimientos que estaban cerrados e intentar sensibilizar a los usuarios que rechacen la investigación; de este modo, se podrá evitar al máximo las fallas en el flujo del formulario y se evitará la sobre cobertura de zonas no asignadas en la planificación inicial.

El supervisor deberá realizar un control permanente del trabajo realizado, verificar el buen desempeño de las actividades de cada encuestador dentro de los límites de su manzana,

aclarar las dudas sobre la ubicación de los establecimientos económicos y resolver otras preguntas que se puedan presentar en las jornadas de trabajo. En estos términos, serán responsabilidades del supervisor:

- Verificar que la ubicación de los establecimientos económicos visitados por sus encuestadores corresponda con los establecimientos asignados inicialmente.
- Validar que los dispositivos de recolección cuenten con la información de formatos y formularios entregados y actualizados para poder realizar la correcta sincronización de la información al final de la jornada.
- Constatar que la cantidad de encuestas sincronizadas corresponda a la cantidad de encuestas teóricas que se realizaron durante el día de trabajo.
- Realizar acompañamientos de supervisión durante la aplicación de algunas encuestas, con el fin de hacer la retroalimentación al encuestador correspondiente una vez finalice la misma.

En el caso de manzanas con centros comerciales como San Andresito, edificios, bloques de apartamentos y similares, el supervisor previamente deberá haber hecho una visita a la respectiva administración informando sobre el trabajo que se va a realizar, y pedir su autorización si es necesario.

2.5. Prueba piloto

De acuerdo con la propuesta, se desarrolló la prueba piloto con tres cámaras de comercio. En esta fase fue posible validar el instrumento de recolección, formatos de control, manuales operativos y procesos logísticos. Con base en los resultados de la prueba piloto se realizaron los ajustes necesarios al formulario y a los manuales de operaciones para garantizar la comprensión de la información y la correcta ejecución de la encuesta.

3. PLAN DE ANÁLISIS

Se debe realizar un análisis en el que se describan los hallazgos sobre la información recolectada para lograr entender las tendencias. Para ello se usarán tablas de frecuencias, gráficos, medidas de tendencia y dispersión. También es necesario calcular los indicadores de interés y detectar posibles inconsistencias o valores atípicos. Se debe verificar la cobertura a nivel del área geográfica y de establecimientos para determinar la pérdida de muestra y proceder a realizar el tratamiento por la no respuesta a las entrevistas.

3.1. Procesamiento

A partir de la recolección de información se consolida una base de datos, que entra en un proceso de depuración y verificación del cumplimiento de la cobertura y se debe hacer un análisis de la no respuesta por pregunta. El procesamiento de datos debe ser de tipo cuantitativo, para lo cual se recomienda usar programas estadísticos especializados como R, STATA, SAS o SPSS.

Se deben calcular los indicadores de interés organizando la información en cuadros de salida. Algunos de los indicadores que deben ser calculados son: el total de establecimientos informales; el total de personas que trabajan en esos establecimientos; el porcentaje de personas sin afiliación a sistemas de seguridad social; el tiempo promedio de funcionamiento de los establecimientos formales e informales; ventas promedio de los establecimientos informales, entre otros.

3.2. Diseño de los cuadros de salida o de resultados

A partir de las bases de datos con la información depurada y consistente se procesan los cuadros de salida, estos se utilizan en la investigación para presentar diferentes niveles de agregación de la información, ya sea por la actividad económica, tiempo de funcionamiento, tipo de personal ocupado o por aproximación de informalidad. A continuación se relacionan los cuadros diseñados para esta investigación, estos se encuentran en un archivo anexo:

- Cuadro 1. Resumen del cumplimiento de requisitos: este cuadro presenta un el total y la proporción de establecimientos que cuenta con matrícula mercantil, RUT y RIT, además de los que llevan contabilidad y cuentan con al menos un empleado con aportes a los sistemas de seguridad social.
- Cuadro 2. Total y proporción de informantes de la encuesta según de tipo de actividad y aproximación de informalidad: presenta la cantidad y porcentaje de informantes que son propietarios, gerentes, encargados, familiares o socios.
- Cuadro 3. Tipo de identificación según actividad y aproximación de informalidad: presenta el porcentaje de establecimientos que se identifican con NIT/RUT, cédula de ciudadanía, cédula de extranjería, pasaporte o simplemente no cuentan con un número de identificación.
- Cuadro 4. Nivel de escolaridad de los propietarios o socios con la mayor participación según actividad y aproximación de informalidad: presenta el porcentaje de propietarios o socios sin nivel educativo o con nivel de primaria, bachillerato, etc.
- Cuadro 5. Total y proporción de establecimientos que son visibles y tiene acceso al público según actividad y aproximación de informalidad: este cuadro reporta información sobre los establecimientos que cuentan con acceso al público para desarrollar su actividad económica.

- Cuadro 6. Total y proporción de establecimientos que cuentan con aviso en el exterior visible al público según actividad y aproximación de informalidad.
- Cuadro 7. Total y proporción de establecimientos por tipo de emplazamiento según actividad y aproximación de informalidad: presenta la cantidad de establecimientos que funcionan en un local, oficina o fábrica; puesto fijo; dentro de una vivienda con actividad económica o en una bodega.
- Cuadro 8. Total y proporción de establecimientos por matrícula mercantil según actividad y tiempo de funcionamiento desde su creación.
- Cuadro 9. Total y proporción de establecimientos que renovaron su matrícula mercantil en el último según actividad y tiempo de funcionamiento desde su creación.
- Cuadro 10. Total y proporción de establecimientos por razones para no tener matrícula mercantil según actividad y tiempo de funcionamiento desde su creación: se calcula sobre los establecimientos que respondieron que no tienen matrícula e informa sobre los establecimientos que consideran que no es útil, es muy costoso, no la han sacado por falta de tiempo, desconocimiento o porque creen que les implica pagar impuestos.
- Cuadro 11. Total y proporción de establecimientos con Registro Único Tributario según actividad y tiempo de funcionamiento desde su creación.
- Cuadro 12. Total y proporción de establecimientos inscritos en el Registro de Información Tributaria - RIT según actividad y tiempo de funcionamiento desde su creación.
- Cuadro 13. Total y proporción de establecimientos por tipo de ubicación según actividad y aproximación de informalidad: informa si los establecimientos están ubicados en centros comerciales, plazas de mercado, edificios de oficinas, etc.
- Cuadro 14. Total y proporción de establecimientos por tiempo de funcionamiento en la dirección actual según actividad y aproximación de informalidad: presenta la

cantidad y porcentaje de establecimientos que llevan menos de un año, de uno a menos de tres años, de tres a menos de 5 años, de 5 años a menos de 10 años y más de 10 años funcionando en la misma dirección.

- Cuadro 15. Total y proporción de establecimientos por tiempo de funcionamiento desde su creación según actividad y aproximación de informalidad.
- Cuadro 16. Total y proporción de establecimientos por tipo de contabilidad según actividad y tiempo de funcionamiento desde su creación
- Cuadro 17. Total y proporción de establecimientos con ... según actividad y tiempo de funcionamiento: informa la cantidad y porcentaje de establecimientos que cuentan con certificado de seguridad contra incendios, pago de la organización Sayco y Acinpro, registro nacional de turismo, entre otros.
- Cuadro 18. Resumen de gastos y condiciones del personal ocupado según actividad y aproximación de informalidad: informa sobre el promedio de personas ocupadas por establecimiento, la cantidad y porcentaje de personas con aportes a EPS, fondo de pensiones, ARL, cajas de compensación o que no tienen ningún tipo de afiliación.
- Cuadro 19. Total y proporción de personal ocupado por tipo de contratación según actividad y aproximación de informalidad: informa sobre la cantidad y porcentaje de personas con contrato a término indefinido, a término definido, por prestación de servicios, etc.
- Cuadro 20. Total y proporción de personal ocupado por tipo de contratación según tipo de personal: informa para cada tipo de personal la cantidad y porcentaje de personas con contrato a término indefinido, a término definido, por prestación de servicios, etc.
- Cuadro 21. Total y proporción de personal por tipo de condiciones laborales según tipo de personal: informa para cada tipo de personal, la cantidad y porcentaje con aportes a EPS, fondo de pensiones, ARL, etc.

- Cuadro 22. Total y proporción de personal por tipo pago según tipo de personal y aproximación de informalidad: presenta la cantidad y porcentaje de personas que reciben un salario, pago por comisiones, por hora, etc.
- Cuadro 23. Gastos por concepto según actividad y aproximación de informalidad: presenta los gastos por concepto de arriendos, servicios, pagos de personal, etc.

Glosario

Actividad económica: combinación de operaciones y recursos (humanos y técnicos) para la producción y/o comercialización de bienes o servicios, tanto para las personas como para las empresas. En un sentido más amplio, la actividad económica es toda acción de producción y/o comercialización resultante de una interacción de medios (capital, mano de obra, procesos, etc.) que lleva a la generación y/o mercado de un bien o a la prestación de un servicio para cubrir las necesidades de una población.

Área de coordinación operativa: área geográfica de trabajo correspondiente a la agrupación de un número determinado de áreas operativas que facilitan la gestión, planeación, control y calidad de la operación censal. Los responsables de ejecutar este rol son la coordinación operativa de campo.

Área operativa: área de terreno conformada por las Unidades de Cobertura Urbana (UCU), esta zona es asignada a un equipo de trabajo compuesto por las personas supervisoras y censistas para realizar la investigación en un tiempo determinado y de acuerdo con unos rendimientos establecidos. Este lugar tiene límites fácilmente identificables en terrenos que son consistentes con la división político administrativa, que se puede consultar en el servicio de Codificación de la División Política Administrativa (Divipola) del DANE.

Cartografía: conjunto de procedimientos, mapas y croquis que se encargan de representar de forma gráfica el plano de la superficie terrestre donde se ha de realizar la investigación.

Censo: encuesta realizada sobre el conjunto completo de las unidades de observación pertenecientes a una determinada población o universo (UNECE, 2000).

Comercio: actividad de compra y venta al por mayor o al por menor de mercancías (bienes que van al mercado), nuevas o usadas no sometidas a transformaciones en su naturaleza

intrínseca durante las actividades inherentes a la comercialización. Consiste en el intercambio de mercancías, bien sea para su uso, venta o transformación.

Diseño de muestra: conjunto de estrategias metodológicas que se definen para extraer una muestra probabilística de la población bajo cierto tipo de condiciones deseables, aquí se determina el tamaño y procedimiento de selección de la muestra encaminado a reducir la varianza en sus estimaciones. También se conoce con el nombre de diseño muestral.

Elemento: unidad individual sobre la que se realizan mediciones.

Encuesta: método estadístico por el cual se investiga y analiza un fenómeno, mediante la selección de una muestra y un propósito, con el fin de inferir resultados en el total de la población.

Enlistamiento: proceso mediante el cual se identifican y registran los establecimientos de industria, comercio y servicios, que serán objeto de investigación en una manzana o grupo de manzanas.

Establecimiento: espacio independiente y separado donde se desarrolla un tipo de actividad económica, que no interfiere con otra actividad debido a que todos los establecimientos tienen elementos entre sí que los diferencian. Es un lugar dedicado de forma predominante a un tipo de actividad económica específica que funciona bajo una sola dirección y control. En caso de ser una plaza de mercado o una feria artesanal, la delimitación de los establecimientos se da por la numeración de los puestos, líneas o mesas.

Establecimiento único: unidad económica exclusiva con una única razón social correspondiente a su actividad.

Establecimiento comercial: unidad económica que bajo una sola dirección y control combina actividades y obtiene recursos para dedicarse a la venta o reventa de mercancías sin

transformación al por mayor o al por menor. Por ejemplo, tiendas, supermercados, puestos fijos en plazas de mercados, ferias artesanales, reparación o mantenimiento de vehículos automotores, motocicletas, enseres domésticos, fábricas, almacenes, etc.

Establecimiento industrial: unidad económica que combina actividades y recursos bajo una sola forma jurídica, desarrolla una actividad administradora para realizar la producción del producto más homogéneo posible de bienes manufacturados.

Establecimiento principal: unidad económica considerada como la más importante entre otras de la misma empresa. En este caso, también debe preguntarse el número de sucursales.

Establecimiento de servicio: unidad económica que bajo una sola dirección y control combina actividades y recursos para dedicarse a la prestación de servicios. Por ejemplo, restaurantes, hoteles, bares, consultorios, peluquerías y salas de belleza, oficinas, fondos de empleados y cooperativas de empleados, educación no formal, hogares de madres comunitarias, juegos de azar, celulares y accesorios, funerarias, etc.

Establecimiento sucursal: unidad económica que, situada en distinto lugar de la central principal de la cual depende, realiza las mismas actividades que está en otra ubicación. Como en el caso anterior, se debe indagar por el número de sucursales.

Estadística: función matemática que permite calcular mediante una base de datos y de variables observadas las características de un grupo de estudio.

Estimación: valor que resulta de ejecutar la estadística a un conjunto de datos particulares.

Estimador: estadística que cumple con algunas propiedades deseables y que se emplea para aproximar el resultado de un parámetro de interés.

Error muestral: grado de precisión que tiene una estadística respecto a la esperanza de su valor calculado.

Factor de expansión: valor que se le atribuye a cada individuo seleccionado en una muestra, como capacidad que tiene para representar a más individuos semejantes con su respuesta.

Imputación: metodología por la cual se aproxima hacia el valor de un resultado que no fue posible obtener en las variables o en los individuos.

Individuo: la unidad más simple que hace parte de un universo de estudio. Una persona, un lugar, un establecimiento, una unidad experimental, etc.

Industria: transformación física o química de materiales o componentes en productos nuevos, bien sea que el trabajo se efectúe con máquinas o a mano, en una fábrica o a domicilio, o que los productos se vendan al por mayor o al por menor; exceptuando la actividad que sea propia del sector de la construcción.

Informante idóneo: corresponde a la persona que suministrará la información de la encuesta, es decir, propietario, administrador o aquella persona que esté encargada del establecimiento económico y tenga conocimientos acerca del personal, valor de las ventas y, en general, de las operaciones del establecimiento.

Logística: corresponde a los procesos y actividades responsables del diseño, planeación, ejecución y control del operativo de recolección de datos.

Manzana: menor división censal cartográfica. Corresponde a un área delimitada por calles, carreras, avenidas, transversales, diagonales, o por elementos naturales como caños, quebradas y ríos, entre otros; siempre y cuando sean de carácter permanente. La manzana puede corresponder a un lote de terreno, edificado o sin edificar, pero delimitado por vías de

tránsito vehicular o peatonal. La manzana se identifica en el ámbito cartográfico con un número de dos dígitos.

Marco de muestreo: lista de elementos que permite ubicar y describir a todos los individuos que hacen parte de una población objetivo. También se conoce con el nombre de marco muestral.

Muestra: conjunto de establecimientos que hacen parte de la población de interés, que han sido seleccionados mediante la implementación de un diseño muestral.

Muestra estratificada: se obtiene dividiendo a la población de interés (marco muestral) en grupos (por ejemplo, comercios, servicios e industria), y seleccionando una muestra aleatoria dentro de cada grupo a partir de un diseño muestral particular.

No respuesta: negación que manifiesta un individuo para no permitir que se realice alguna medición de una característica sobre él o su establecimiento.

Nombre comercial: topónimo con el cual se identifica el público del establecimiento económico (aviso publicitario).

Operación estadística: conjunto de procesos y actividades que parten de la recolección sistemática de datos y conducen a la producción de resultados agregados.

Parámetro: valor de los indicadores de interés propuestos para preguntarle a todos los elementos que componen la población objetivo (total de establecimientos con matrícula mercantil, porcentaje de personas con aportes a EPS, etc.).

Personal aprendiz: se refiere a los empleados vinculados al establecimiento a través de un contrato de aprendizaje, por tiempo definido, no mayor a dos años, y a quienes se les otorga una mensualidad, como apoyo de sostenimiento.

Personal ocupado: personas que desempeñan alguna labor en el establecimiento que puede ser o no remunerada. Incluye a los propietarios, socios y familiares sin remuneración fija.

Personal ocupado permanente: se refiere al personal total contratado a término indefinido.

Personal ocupado temporal: se refiere al número de personas ocupadas temporalmente por el establecimiento.

Prestaciones sociales: son las remuneraciones obligatorias, especiales y pactadas, diferentes a los sueldos y salarios que el empleador debe pagar a su personal ocupado. Comprende: primas legales, extralegales, vacaciones, cesantías e intereses sobre cesantías, aportes patronales a la seguridad social (salud, pensiones, riesgos profesionales), al ICBF, al SENA y a las cajas de compensación familiar.

Población objetivo: conjunto de individuos que se pretende investigar y sobre el cual se desea obtener conclusiones.

Rechazo de entrevista: resultado final de la entrevista cuando no se puede completar una parte o la totalidad del cuestionario, luego de realizadas las visitas establecidas por la operación estadística. El motivo principal de esta situación es la negativa por parte de la persona entrevistada para atender al personal censista o de supervisión.

Recolección por barrido: consiste en realizar un recorrido sistemático de la totalidad de las unidades de observación asignadas en un área de trabajo, sincronizando periódicamente la información obtenida para generar reportes de cobertura diarios.

Recolección por rutas: la ruta es la forma como los grupos de censistas recorren la zona rural de algunos municipios caracterizados por las dificultades de acceso, baja densidad de población y grandes extensiones territoriales.

Recolección electrónica: refiere a la recolección realizada a través del diligenciamiento del cuestionario censal por la persona encuestada, mediante la interacción de una aplicación web a la que se accede por una red de internet o que se puede sincronizar en el momento de una conexión a una red.

Reglas de validación: estrategias que prueban la integridad y consistencia de los datos. Las reglas de validación se pueden utilizar para evaluar la calidad de los datos, para filtrarlos o para corregirlos.

Revisitas: son un mecanismo para mejorar la cobertura durante la recolección de los datos. Con el propósito de garantizar la cobertura censal, el equipo operativo en campo encabezado por la persona censista, realizará nuevas visitas a la unidad de observación (establecimientos) en casi de que se presenten las novedades: establecimiento cerrado en el momento de la visita, encuesta incompleta o encuesta rechazada. Asimismo, quien se encarga de la supervisión hará seguimiento a las revisitas, con el fin de cumplir y garantizar la cobertura total.

San Andresito: espacio geográfico similar a los centros comerciales, que aglomera a una cantidad relativamente alta de establecimientos que se dedican a actividades comerciales.

Sector urbano: división cartográfica censal que equivale generalmente a un barrio y que comprende entre 1 y 9 secciones. Se identifica en la cartografía por un número de cuatro dígitos encerrado en un rectángulo. Sus límites se indican con línea gruesa a trazos separados con un punto.

Sección urbana: división censal cartográfica que equivale aproximadamente a 20 manzanas contiguas pertenecientes al mismo sector o barrio. Se identifica con un número cerrado en un círculo; sus límites se indican con línea delgada o trazos. Se señala en la cartografía mediante un código de dos dígitos dentro de un círculo.

Sensibilización: tiene por objeto hacer que las fuentes a entrevistar respondan al censo sin desconfianza, consiste en la ejecución de un proceso comunicativo de concientización dirigido a las fuentes donde se presenta el encuestador, a la entidad que representa y la importancia de la labor que realiza.

Servicios: transacciones de bienes intangibles generados cada vez que son solicitados, sobre los cuales no recaen derechos de propiedad por parte del usuario y no pueden ser negociados por separado de su producción, no son transportados ni almacenados. Los servicios pueden ser comunales, personales, religiosos, sociales, bancarios, financieros, de salud, recreativos, deportivos, culturales, asesorías, consultorías, mensajería.

Sistema de monitoreo y control: modelo que permite realizar el monitoreo y la validación de la información en el permanente desarrollo de los procesos y productos que se llevan a cabo durante la ejecución a nivel de centro operativo.

Unidad de Cobertura Urbana (UCU): es la unidad mínima del marco censal a nivel urbano. Se define como un área (edificada o sin edificar) delimitada por vías de tránsito vehicular o peatonal de uso público (calles, carreras, avenidas, diagonales o senderos peatonales); también puede estar delimitada por elementos naturales como caños, ríos y quebradas que sean de carácter permanente.

Unidades de observación: elemento o conjunto de elementos sobre los que se hace la medición de las diferentes variables en una operación estadística.

Unidad económica: espacio donde se desarrolla una actividad económica independiente que no interfiere con otras actividades y tiene elementos que lo diferencian entre sí. Lugar dedicado exclusivamente a un tipo de actividad económica que funciona bajo una sola dirección. Son ejemplos de unidades económicas: almacenes, tiendas, hoteles, restaurantes, salas de belleza, bares, fábricas, etc.

Universo de estudio: población de la cual se requiere la información para estimar los parámetros sobre sus características.

Unidad muestral: unidad que se considera dentro del diseño muestral como adecuada para realizar una selección de muestra. Si el diseño es en varias etapas, también tiene varios tipos de unidades muestrales.

Variable: es una característica de la población de estudio. Toma valores determinados para cada observación que pueden diferir entre individuos.

Bibliografía

- ANIF. (2017). *Reducción del efectivo y tamaño de la economía subterránea en Colombia*. Recuperado de <http://www.anif.co/sites/default/files/investigaciones/anif-asobancaria-efectivo0517.pdf>
- Baillargeon, S., & Rivest, L.P. (2009). A general algorithm for univariate stratification. *International Stratification Review*, 77 (3), 331-344.
- CEPAL. (2011). *Guía para la elaboración de un proyecto censal*. Santiago de Chile: Publicación de las Naciones Unidas. Recuperado de https://repositorio.cepal.org/bitstream/handle/11362/5508/1/S1100203_es.pdf
- Comisión Económica para Europa de las Naciones Unidas (UNECE). (2000). Terminology on Statistical Metadata. *Conference of European Statisticians Statistical Standard and Studies*. Geneva.
- Departamento Administrativo Nacional de Estadística (DANE). (Noviembre 23 de 2007). *Microdatos DANE*. Recuperado de Microdatos DANE: http://170.238.64.38/index.php/catalog/39/related_materials
- Fundación para la Universitat Oberta de Catalunya. (2005). Bases de datos. En C. Martín Escofet, *Modulo 3 El lenguaje SQL*. Barcelona: Eureka Media.
- Fundación para la Universitat Oberta de Catalunya. (2005). Bases de datos. En R. Camps Paré, *Módulo 1 Introducción a las bases de datos*. Barcelona: Eureka Media.
- IBM. (2012). *Manual CRISP-DM de IBM SPSS Modeler*. IBM Corporation.
- Lavallée, P., & Hidirolou, M. (1988). On the stratification of skewed populations. *Survey Methodology*, 14 (1) 33-43.
- Ministerio de Comercio Industria y Turismo. (Agosto 28 de 2017). *Capítulo 45 Jurisdicción de las Cámaras de Comercio*. Recuperado de <http://www.mincit.gov.co/publicaciones/36018/info@mincit.gov.co>

- Naciones Unidas. (2010). *Principios y recomendaciones para los censos de población y habitación*. Nueva York: Publicación de las Naciones Unidas. Recuperado de https://unstats.un.org/unsd/publication/seriesm/seriesm_67rev2s.pdf
- Särndal, C., & Lundström, S. (2005). *Estimation in surveys with nonresponse*. John Wiley & Sons.
- Särndal, C., Swensson, B., & Wretman, J. (1992). *Springer series in statistics. Model assisted survey sampling*. New York, NY, US: Springer-Verlag Publishing.
- Soto, L. C., & Arnedo, C. V. (2004). *Introducción a las bases de datos Access 2003*. Publicaciones de la Universidad de Alicante: Digitalia.
- Statistics Canada. (2010). *Survey methods and practices*. Ottawa: Minister of Industry. Recuperado de https://www150.statcan.gc.ca/n1/en/pub/12-587-x/12-587-x2003001-eng.pdf?st=IR_fy8UE
- Schneider, F. (2002). “The informal economy in the EU - Accession and instability pact countries: Size, scope and challenges to the process of EU enlargement”, Centro de Investigación de Europa del Este. 15.
- Schneider, F., & Enste, D. H (2000). “Shadow economies: size, causes and consequences”, *Journal of Economic Literature*, 38(1), 77-114.
- Valliant, R., Dever, J., & Kreuter, F. (2018). *Practical tools for designing and weighting survey samples*. Pittsburgh, PA, USA: Springer.
- Vélez, F. (2006). *Estudio sobre la informalidad y las mipymes en Colombia: análisis y propuestas*. Bogotá, Proyecto ANDI-OIT. 17.
- Wooldridge, J. (2001). *Econometric analysis of cross section and panel data*. Cambridge, Massachusetts, The MIT Press.

Anexo I. Formulación del tamaño de la muestra

El tamaño de la muestra busca que las diferencias entre las estimaciones y el parámetro no supere un valor e , denominado margen de error, con un nivel de confianza de $1 - \alpha$, esto es $P(|\bar{y}_s - \bar{y}_u| < e) = 1 - \alpha$. Por el Teorema Central del Límite, la suma de variables aleatorias se ajusta a una distribución normal, por lo que

$$e = z_{1-\alpha/2} \sqrt{V_p(\bar{y}_s)}$$

Teniendo en cuenta que el efecto del diseño $p(\cdot)$, que se define como:

$$deff = \frac{V_p(\bar{y}_s)}{V_{MAS}(\bar{y}_s)} = \frac{V_p(\bar{y}_s)}{\frac{1}{n} \left(1 - \frac{n}{N}\right) s_y^2}$$

La varianza del estimador para la media puede ser expresada en función de la varianza de un muestreo aleatorio simple multiplicada por el efecto del diseño muestral, así:

$$V_p(\bar{y}_s) = \frac{1}{n} \left(1 - \frac{n}{N}\right) s_y^2 * deff$$

Al reemplazar en el error de muestreo, se tiene que $e = z_{1-\alpha/2} \sqrt{\frac{1}{n} \left(1 - \frac{n}{N}\right) s_y^2 * deff}$

Al despejar el tamaño de la muestra n , y considerando que en variables de presencia o ausencia, $s_y^2 \approx p(1 - p)$ siendo p la prevalencia del suceso, por tanto:

$$n \geq \frac{\frac{z_{\alpha}^2 p(1-p) deff}{2}}{e^2 + \frac{\frac{z_{\alpha}^2 p(1-p) deff}{2}}{N}} = \frac{N p(1-p) deff}{N\delta + p(1-p) deff}$$

Con $\delta = e^2 / \frac{z_{\alpha}^2}{2}$.

Anexo II. Algoritmo de selección de UCU

Definido el tamaño de la muestra, se debe realizar una selección aleatoria que garantice el diseño de muestreo propuesto. Para la selección de UCU en la primera etapa, se aplica un algoritmo Sunter con el fin de respetar las probabilidades de inclusión de un muestreo π_{PT} . Para ello la variable auxiliar corresponde al número de comercios y servicios y, por otra parte, a la cantidad de industrias. Los pasos del algoritmo de Sunter incian considerando la variable auxiliar (x) como el número de establecimientos que tiene la UCU, y se debe realizar el siguiente procedimiento:

1. Ordenar de forma descendente a las UCU de acuerdo con los valores que toma la variable auxiliar, x .
2. Para cada UCU, genere un número aleatorio, ζ_k con distribución de probabilidad uniforme entre 0 y 1.
3. Para $k = 1$, la primera UCU de la lista ordenada, será incluida en la muestra si y solamente si $\zeta_k < \pi_1$, donde π_1 corresponde a la probabilidad de inclusión del primer elemento.
4. Para $k \geq 2$, la k -ésima UCU de la lista ordenada, será incluida en la muestra si y solamente si

$$\zeta_k \leq \frac{n - n_{k-1}}{n - \sum_{i=1}^k \pi_i} \pi_k$$

Donde n_{k-1} representa el número de UCU que ya han sido seleccionados al final del paso k -

1. Los π_k son las probabilidades de inclusión de primer orden y se calculan utilizando la variable auxiliar x . Para mayores detalles se recomienda consultar (Särndal, Swensson, & Wretman, 1992).