

**Departamento Administrativo  
Nacional de Estadística**



**Diseño  
DSO**

**Dirección de Metodología y Producción Estadística / DIMPE**

**METODOLOGÍA GENERAL GRAN ENCUESTA INTEGRADA DE HOGARES GEIH**

**Feb/2020**

**TABLA DE CONTENIDO****PRESENTACIÓN****INTRODUCCIÓN****1. ANTECEDENTES****2. DISEÑO DE LA OPERACIÓN ESTADÍSTICA****2.1. DISEÑO TEMÁTICO/METODOLÓGICO**

## 2.1.1. Necesidades de Información

## 2.1.2. Objetivos

## 2.1.3. Alcance

## 2.1.4. Marco de referencia

## 2.1.5. Diseño de indicadores

## 2.1.6. Plan de resultados

## 2.1.6.1. Diseño de cuadros de salida o de resultados

## 2.1.7. Diseño del formulario o cuestionario

## 2.1.8. Normas, especificaciones o reglas de validación, consistencia e imputación

## 2.1.9. Nomenclaturas y clasificaciones utilizadas

**2.2. DISEÑO ESTADÍSTICO**

## 2.2.1. Componentes básicos del diseño estadístico

## 2.2.2. Unidades estadísticas

## 2.2.3. Periodo de referencia y recolección

## 2.2.4. Diseño muestral (aplica en investigaciones por muestreo)

## 2.2.5. Ajustes Cobertura

**2.3. DISEÑO DE LA EJECUCIÓN**

## 2.3.1. Sistema de capacitación

## 2.3.2. Actividades preparatorias

## 2.3.3. Diseño de instrumentos

## 2.3.4. Recolección de la información

**2.4. DISEÑO DE SISTEMAS****2.5. DISEÑO MÉTODOS Y MECANISMOS PARA EL CONTROL DE CALIDAD****2.6. DISEÑO DE PRUEBAS PILOTO****2.7. DISEÑO DEL ANÁLISIS DE RESULTADOS**

## 2.7.1. Análisis estadístico

## 2.7.2. Análisis del contexto

## 2.7.3. Comité de Personas expertas

**2.8. DISEÑO DE LA DIFUSIÓN**

## 2.8.1. Administración del repositorio de datos

## 2.8.2. Productos e instrumentos de difusión

**2.9. DISEÑO DE LA EVALUACIÓN****2.10. DOCUMENTACIÓN RELACIONADA****3. GLOSARIO****4. BIBLIOGRAFÍA****5. ANEXOS**

## PRESENTACIÓN

El Departamento Administrativo Nacional de Estadística (DANE), en función de su papel como coordinador del Sistema Estadístico Nacional (SEN), trabaja para su fortalecimiento y consolidación en el marco del proyecto de Planificación y Armonización Estadística.

La producción, análisis y difusión de estadísticas; generación, adaptación y difusión de estándares; la consolidación y armonización de la información estadística; la articulación de instrumentos, actores, iniciativas y productos; para mejorar la calidad de la información estadística estratégica, su disponibilidad, oportunidad y accesibilidad, hacen parte del proceso.

Consciente de la necesidad y obligación de brindar a los usuarios los mejores productos, el DANE desarrolló una guía estándar para la presentación de metodologías que contribuye a visibilizar y a entender el proceso estadístico. En los documentos metodológicos se presentan de manera estándar, completa y de fácil lectura, las principales características técnicas de los procesos y subprocesos de cada investigación, lo que permite su análisis, control, replicabilidad y evaluación.

Esta serie de documentos favorecen a la transparencia, confianza y credibilidad en la calidad técnica de la institución para un mejor entendimiento, comprensión y aprovechamiento de la información estadística, producida en el contexto de los principios de coherencia, comparabilidad, integralidad y calidad de las estadísticas.

## INTRODUCCIÓN

Desde finales de la década de 1960 el DANE ha desarrollado e implementado de manera transversal encuestas a hogares, que consideran las temáticas: fuerza de trabajo, calidad de vida e ingresos y gastos. A partir del 10 de julio de 2006 se amplió la cobertura de la Encuesta Continua de Hogares (ECH) de 13 ciudades con sus áreas metropolitanas a 24 ciudades, al resto de cabeceras y al resto rural. El 7 de agosto de 2006 empieza la recolección de la Gran Encuesta Integrada de Hogares (GEIH), en su módulo central de mercado laboral e ingresos de los hogares.

Este documento inicialmente presenta la evolución histórica de la medición del mercado laboral en Colombia y las principales características técnicas de la GEIH. Posteriormente se identifica el diseño de la encuesta y su marco de referencia, así como el diseño temático de la encuesta respaldada por recomendaciones de organismos internacionales que regulan y estandarizan los conceptos con el fin de asegurar la comparabilidad entre los países en materia de mercado laboral. Finalmente se describe el diseño estadístico que permite tener claridad sobre el marco muestral y los procedimientos de definición de la muestra y el alcance de la misma en términos de estimación y análisis de resultados.

También se expone cada proceso de la producción estadística desde las actividades preparatorias al operativo de campo, pasando por la recolección, la transmisión y procesamiento de datos y los métodos de control de calidad, para finalizar con los procesos de análisis y difusión de los datos.

A este documento lo acompañan como anexos el formulario de la encuesta, el listado de cuadros disponibles al público por medio del banco de datos, el listado de clasificaciones internacionales adoptadas por la GEIH y el glosario de términos de mercado laboral, para mayor entendimiento del lector.

## 1. ANTECEDENTES

El DANE, ha desarrollado e implementado encuestas de hogares: fuerza de trabajo, de ingresos y gastos desde finales de la década del sesenta del siglo pasado. Durante la década de 1970, se dio inicio a la Encuesta Nacional de Hogares (ENH) que fue concebida como un sistema de encuestas de múltiples propósitos que tenía como fin producir estadísticas básicas relacionadas con las situaciones: demográfica, social y económica de la población colombiana.

Entre 1970 y 1975, se llevaron a cabo nueve etapas con diferente periodicidad, cobertura y diseño muestral, estos criterios se unificaron a partir de marzo de 1976. Posteriormente en 1978, se realizó la primera medición áreas rurales y que se replicó en 1988. Entre 1976 y 1984, en forma trimestral se investigaron las ciudades de Bogotá, Cali, Medellín y Barranquilla, con una cobertura variable en cada trimestre y en forma semestral se investigaron las ciudades de Bucaramanga, Manizales y Pasto.

En la década de 1980 se incluyeron los municipios circundantes a las siete principales ciudades, conformando la muestra trimestral las ciudades de Bogotá D.C. y Pasto; de igual manera las áreas metropolitanas<sup>1</sup> de Medellín - Valle de Aburrá (Caldas, La Estrella, Sabaneta, Itagüí, Envigado, Bello Copacabana y Barbosa), Cali -Yumbo; Barranquilla - Soledad; Bucaramanga - Girón Piedecuesta, Floridablanca; Manizales - Villamaría, semestralmente las áreas metropolitanas de Pereira - Dos Quebradas y la Virginia y Cúcuta, Villa del Rosario, Los Patios y el Zulia. Anualmente se investigaron las ciudades de Ibagué, Montería, Cartagena y Villavicencio.

A partir de marzo de 1990 se presentaron resultados representativos por área metropolitana. En diciembre de 1991 se inició la medición nacional y aunque se incorporó el concepto de ruralidad se mantuvo el concepto de distribución cabecera - resto por regiones. Esta medición continuó todos los años en los meses de septiembre a partir de 1992. En 1996, se amplió la muestra lo que permitió presentar resultados a nivel de 24 departamentos y finalmente se elimina el concepto de ruralidad.

En desarrollo del plan de trabajo, en enero de 1999 se inició en Bogotá la aplicación de la Encuesta Continua de Hogares (ECH) como prueba piloto y que se extendió a partir del segundo semestre a Medellín y Cali, y desde septiembre a Barranquilla, Bucaramanga, Manizales y Pasto. Una vez realizados los análisis correspondientes a la prueba piloto, se aplicó la ECH en forma definitiva a partir de enero de 2000 en las siguientes ciudades con sus áreas metropolitanas: Bogotá, Cali, Medellín, Barranquilla, Bucaramanga, Manizales, Pasto, Pereira, Cúcuta, Villavicencio, Ibagué, Montería y Cartagena. Todas comenzaron a formar parte del nuevo Sistema de Encuestas a Hogares que comprende la Encuesta Continua de Cobertura Nacional; este tenía diferentes niveles de desagregación geográfica en la medida que se acumulaba tamaño de muestra (cabeceras, centros poblados y rural disperso, grandes regiones y total por departamento).

Durante el 2000 se realizaron en paralelo la Encuesta Tradicional de aplicación trimestral y la Encuesta Continua, para proporcionar información que permitiera analizar las diferencias en los resultados de los dos métodos e identificar los criterios básicos para el empalme de las series. Lo anterior para: afinar los procesos y los procedimientos operativos; armonizar las encuestas urbanas y rurales, y definir el sistema de procesamiento y producción de resultados y de esta manera determinar la cantidad de indicadores y la periodicidad con la cual debían publicarse.

En 2001 se amplió la muestra maestra de 165 a más de 240 municipios, con 30.000 hogares en 13 ciudades y sus áreas metropolitanas y 7.500 en el resto de las cabeceras, centros poblados y rural disperso. En el 2003 se amplió la operación por regiones y total por departamentos. En 2004 se incrementó la muestra maestra, con un total de 44.400 hogares, con 30.000 en 13 ciudades y sus áreas metropolitanas y 14.400 en el resto de las cabeceras, centros poblados y rural disperso.

En el 2005 se propuso la integración de las tres más importantes encuestas a hogares del DANE: la Encuesta Continua de Hogares (ECH), la Encuesta Nacional de Ingresos y Gastos (ENIG) y la Encuesta de Calidad de Vida (ECV); esto con el propósito de ampliar el alcance temático de la investigación y reducir el costo de la aplicación. Como resultado de esta propuesta tan solo se pudieron aplicar algunos módulos de las ECV y ENIG. La integración de estas encuestas no fue efectiva en su totalidad.

Además, en ese proceso se revisó la documentación disponible sobre esta estrategia<sup>2</sup> y se contó con el apoyo financiero del programa del Banco Mundial para el Mejoramiento de las Encuestas y la Medición de las Condiciones de Vida en América Latina y el Caribe (MECOVI). Este apoyo financiero del Banco Mundial se centró en: el diseño temático, el diseño muestral, el diseño logístico, así como la estrategia de entrenamiento, y las pruebas de campo para avanzar en la estrategia de integración de encuestas<sup>3</sup>. En el 2006, considerando la revisión de experiencias internacionales y con el propósito de cumplir el alcance temático de las encuestas a hogares, se inició el diseño de la Gran Encuesta Integrada de hogares (GEIH).

La GEIH es el resultado de un marco estructural del Sistema de Encuestas Sociales en un lapso de 10 años (periodo intercensal) donde en el primer año se integró la ECH con la ENIG y posteriormente se integró la ECH con la ECV. En las etapas subsiguientes se realizaron módulos periódicos para profundizar en el estudio de fenómenos específicos o para monitorear fenómenos que se midieron en profundidad en la GEIH el primer año.

Consecuentemente se revisaron las variables a mejorar en cada operación, comparando el resultado de los períodos anteriores: temáticas y variables comunes para las tres encuestas. Se concluyó que el eje central de la GEIH debía ser el de Mercado Laboral ECH por su carácter de continua y por ser la de mayor tamaño y cobertura. Las otras se incluyeron como módulos específicos y que se hicieron a una submuestra de la GEIH.

El reto asumido por el DANE fue lograr esa integración temática permitiendo el enriquecimiento de cada encuesta en particular sin que el tamaño del formulario afectara la calidad en la recolección. Esto supuso la implementación de nuevos desarrollos logísticos considerando innovaciones como: la introducción del informante directo, que permitió realizar una mayor indagación sobre variables claves como son los ingresos y las condiciones de trabajo de las personas y la captura de información mediante dispositivos móviles, lo que permitió una mayor oportunidad en la producción de resultados.

A partir del 10 de julio de 2006 se amplió la cobertura de la ECH a las siguientes capitales: Tunja, Florencia, Popayán, Valledupar, Quibdó, Neiva, Riohacha, Santa Marta, Armenia, Sincello y San Andrés; todas cuentan con las mismas condiciones de representatividad de las trece principales ciudades con sus áreas metropolitanas.

La recolección de la GEIH empezó el 7 de agosto de 2006 en su módulo central de mercado laboral e ingresos y, a partir del 11 de septiembre, con su módulo de gastos de los hogares.

Desde el 2007, GEIH, ha estado sometida a la revisión y la evaluación de expertos nacionales e internacionales para establecer los impactos en la serie de mercado laboral e ingresos, teniendo en cuenta las mejoras incorporadas. Entre abril y diciembre de 2008 se aplicó en forma paralela la encuesta en la muestra diseñada teniendo como marco muestral el Censo 2005. En enero de 2009 se inició la aplicación de la GEIH con el nuevo marco de muestreo y utilizando las proyecciones de población que surgen del Censo de Población y Vivienda 2005 para las estimaciones de los dominios.

Actualmente la GEIH se ha especializado en la medición de la estructura del mercado laboral y los ingresos de los hogares. Tiene una muestra total anual de 232.000 hogares aproximadamente, lo que hace que sea la de mayor cobertura a nivel nacional. Por su parte, las mediciones de Calidad de Vida e Ingresos y Gastos de los hogares, se realizan a través de encuestas independientes (ECV y ENIG).

En cuanto a la medición de la informalidad para Colombia, la Encuesta Nacional de Hogares (ENH), aplicada de 1976 hasta el 2000, fue la primera encuesta utilizada para medir la informalidad en el país. El primer módulo de informalidad fue implementado en la ENH en junio de 1986 para las 7 principales ciudades, incluidas sus áreas metropolitanas (Bogotá, Cali, Medellín, Barranquilla, Manizales, Pasto y Bucaramanga). A partir de 1986 y hasta el 2000, el módulo se aplicó en junio con periodicidad bianual. Durante todo el período, la encuesta tuvo la misma estructura con excepción de la realizada en junio de 1990, cuando se introdujeron cambios en las opciones de respuesta a la pregunta que generaba el filtro de informalidad.

<sup>1</sup> Para la GBH se entiende por área metropolitana de la ciudad aquellos municipios que entran en la selección de la muestra.

<sup>2</sup> Elaborada por CEPAL (Juan Carlos Feres y Fernando Medina).

<sup>3</sup> Informes Proyecto MECOVI-DANE-BANCO MUNDIAL-2006.

Entre 2001 y 2006, con la sustitución de la ENH por ECH, se inició la aplicación del módulo de informalidad en los segundos trimestres de cada año, para el tercer trimestre de 2006 se empieza a aplicar la GEIH. La ECH aplicada a las 13 principales ciudades con sus áreas metropolitanas realizó la medición del sector informal mediante la denominada Encuesta 1 2 3. Esta fue una encuesta mixta realizada en tres fases, que se describen a continuación:

**Fase 1.** En esta se aplicó una encuesta de hogares que identificó a los miembros del hogar, patronos o trabajadores por cuenta propia, de una unidad económica informal visible, no visible o móvil. Con esta información se construyó un marco muestral de unidades económicas de los hogares.

**Fase 2.** Se aplicó una encuesta a las unidades económicas informales, indagando por las características generales de la unidad económica tales como actividad económica, edad, formas de financiamiento, número y características del personal ocupado, costos, producción, ventas, ingresos, valor de los activos fijos, etc.

**Fase 3.** Mediante la aplicación de una encuesta de ingresos y gastos. Se indagó sobre la proporción de los gastos de consumo final de los hogares que se realiza en establecimientos informales y las razones por las cuales los hogares adquieren productos de consumo final en establecimientos informales.

Para la GEIH la batería de preguntas del módulo de informalidad se mantuvo, aunque se decidió que la aplicación del mismo debería ser de forma continua con una cobertura para total nacional, que incluye además las ciudades con sus áreas metropolitanas, cabeceras y centros poblados. La medición de informalidad se empieza a publicar en el 2007, en un informe especial de mercado laboral según el empleo informal. Esta publicación es mensual pero el acumulado de la muestra que se presenta es trimestral, posteriormente en el 2008, se presentaron los resultados de la GEIH, incluyendo la medición de informalidad, con proyecciones de población basadas en el Censo 2005.

A partir de diciembre de 2009, se comenzó a usar la resolución de 1993 de la OIT que cambió el límite de escala de personal ocupado en las empresas, pasando de 10 a 5 trabajadores. Acorde con el principio 9 de la ONU sobre la utilización de normas internacionales, el DANE realizó la actualización de parámetros con la revisión estadística correspondiente, sustentada a través de la encuesta de microestablecimientos, donde se encontró evidencia de que las empresas que ocupan entre 6 y 9 empleados generan empleo formal.

En el cuarto trimestre del 2012, la GEIH se empezó a aplicar en las ocho ciudades capitales de los nuevos departamentos: Arauca, Yopal, Mocoa, Leticia, Inírida, San José del Guaviare, Mitú, Vichada; alcanzó cobertura en las 32 principales ciudades capitales de departamento, en las cabeceras y centros poblados y rural disperso de más de 443 municipios del país, llegando aproximadamente a 58.200 hogares cada trimestre. Con esta ampliación de la muestra se logró cubrir el 100% de las ciudades capitales de todos los departamentos.

A partir abril de 2015, se inició la recolección de información de la GEIH en el municipio de Buenaventura, con representatividad en la entrega de resultados de forma anual. Mientras que en enero de 2018 se inició recolección de información para Barrancabermeja y para Soacha y para enero de 2019 se empezó a recolectar información para Rionegro (Antioquia) y para Tumaco. Todas las anteriores permiten obtener resultados con representatividad anual.

## 2. DISEÑO DE LA OPERACIÓN ESTADÍSTICA

### 2.1. DISEÑO TEMÁTICO/METODOLÓGICO

#### 2.1.1. Necesidades de Información

A partir de la medición de la estructura de la fuerza de trabajo se derivan análisis de la situación del empleo, el desempleo, la inactividad y de las características sociodemográficas de la población colombiana. Esta información es primordial para la toma de decisiones del Gobierno Nacional y para la investigación del contexto del mercado laboral del país.

#### 2.1.2. Objetivos

##### a. Objetivo general

Proporcionar información básica sobre el tamaño y estructura de la fuerza de trabajo (empleo, desempleo e inactividad) de la población del país, así como de las características sociodemográficas de la población colombiana.

##### b. Objetivos específicos

- Clasificar la población de cada uno de los dominios de estudio, según los conceptos y definiciones de la fuerza de trabajo establecidos por la Conferencia Internacional de Estadísticos del Trabajo (CIET) de la Oficina de la Organización Internacional de Trabajo (OIT) de 1983.
- Calcular los principales indicadores del mercado laboral: Tasa Global de participación (TGP), Tasa de Ocupación (TO), Tasa de desempleo (TD), Tasa de subempleo (TS) y su variación en el tiempo.
- Medir características generales de la población, vivienda, acceso a servicios públicos, sistema de protección social.
- Obtener información sobre variables sociodemográficas de la población, como: sexo, edad, estado civil, educación, etc.
- Medir las características del empleo: temporalidad, subempleo, rama de actividad, ocupación u oficio, posición ocupacional, ingreso, afiliación a la seguridad social, etc.
- Medir las características del desempleo: tiempo de búsqueda de empleo, rama de actividad, ocupación u oficio que desempeño en su empleo anterior.
- Medir las características de la inactividad y las razones por las que la población se ha retirado o no participa en el mercado laboral.
- Proporcionar información básica sobre el tamaño y la estructura de la población ocupada en empresas de hasta cinco trabajadores, así como las características sociodemográficas de esta población.
- Conocer los ingresos de los hogares tanto en dinero como en especie, que sirvan de insumo para las mediciones sobre pobreza.
- Proporcionar información sobre calidad del empleo.

#### 2.1.3. Alcance

La Gran Encuesta Integrada de Hogares - GEIH proporciona información básica sobre el tamaño y estructura de la fuerza de trabajo (empleo, desempleo e inactividad) de la población del país, así como las características sociodemográficas de la población colombiana. Las temáticas por las cuales se indagaron a través de la GEIH permiten caracterizar a la población según el sexo, la edad, el parentesco, el nivel educativo, la afiliación al sistema de seguridad social en salud entre otros. Actualmente la información se recolecta en 13 ciudades y sus áreas metropolitanas, 11 ciudades intermedias, 8 capitales de nuevos departamentos; cabeceras y centros poblados y rural disperso en más de 443 municipios del país.

A partir abril de 2015, se inició la recolección de información de la GEIH en el municipio de Buenaventura. Mientras que en enero de 2018 se inició recolección de información para Barrancabermeja y para Soacha y para enero de 2019 se empezó a recolectar información para Rionegro (Antioquia) y Tumaco. Todas las anteriores permiten obtener resultados con representatividad anual.

#### 2.1.4. Marco de referencia

##### a. Marco teórico

Lo considerado en cuanto los indicadores de mercado laboral parte de las recomendaciones presentadas por la Organización Internacional del Trabajo (OIT). En materia de estadísticas laborales, Colombia y América Latina dispone de encuestas a hogares que se levantan con regularidad desde hace más de cuatro décadas; durante este tiempo se han obtenido desarrollos teóricos y metodológicos basados en el enfoque de armonización de los principales conceptos y de las metodologías de medición del empleo.

La Conferencia Internacional de Estadísticos del Trabajo (CIET)<sup>4</sup> ha adoptado diferentes resoluciones y recomendaciones en materia de medición de las estadísticas laborales. En 1983, durante la decimotercera CIET, se estableció y normalizó el marco teórico y conceptual en el que se deben enmarcar las diferentes variables de la fuerza de trabajo: concepto y definición de cada una de las variables; períodos de referencia; actividad principal; reglas de prioridad de las variables, etc. Igualmente, se introdujo la diferenciación de la población actualmente activa de la habitualmente activa y que en el caso colombiano las mediciones se refieren a la primera.

Los indicadores de mercado laboral y específicamente la tasa de desempleo (TD) han sido utilizados como indicadores críticos para analizar el estado del comportamiento de la economía; entendiendo la tasa de desempleo como un indicador del grado de utilización de los recursos humanos de la economía. Los demás indicadores de mercado laboral se deben analizar de manera simultánea. Bajo este esquema, se busca suministrar esta información para poder entender las señales que ofrece el mercado laboral sobre las condiciones de la economía, de ahí la función del DANE de proveer esta información básica y relevante sobre el estado y las tendencias de los principales indicadores del mercado laboral.

El DANE mantenía la operación según los esquemas de la CIET, pero las resoluciones adoptadas después de 1983 se reflejan completamente y con precisión óptima en los instrumentos de recolección de información vigentes hasta 1999. Por esta razón, se inició un proceso de revisión y actualización de los marcos metodológicos, muestrales, operativos y de procesamiento de información.

<sup>4</sup> Esta conferencia de expertos de todo el mundo se reúne ordinariamente cada cinco años en Ginebra (Suiza), y produce resoluciones y recomendaciones sobre la medición de las estadísticas laborales. La última reunión se llevó a cabo en septiembre de 1998 y, entre otras resoluciones, adoptó la referente a las estadísticas de subempleo y las condiciones de empleo inadecuado. Principales CIET: XIII CIET de Conceptos de Población Económicamente Activa, XV CIET de Ingresos e Informalidad, XVII CIET de Medición del Subempleo.

Dicho proceso culminó con la adopción del Sistema de Recolección Continuo y la actualización de la metodología de la encuesta considerando: las resoluciones aprobadas por la Oficina de la Organización Internacional del Trabajo desde 1983; el nuevo Sistema de Cuentas Nacionales de 1993 (SCN-93); las experiencias y las prácticas internacionales, y las particularidades del mercado laboral colombiano. El sistema con las modificaciones señaladas empezó a regir a partir de enero del 2000 y se le denominó ECH.

La GEIH inició su diseño en diciembre de 2005 con un proyecto del grupo temático de la ECH sobre la aplicación de una encuesta integrada. Su objetivo era obtener la mayor cantidad de información posible sobre un mismo hogar, de tal forma que pueda ser utilizada para la realización de investigaciones desde diferentes perspectivas, facilitando un análisis más completo de las características económicas y sociales de la población colombiana.

Lo más importante del ejercicio al conformar el formulario de la GEIH, estuvo en la preservación de las series principales de la ECH para no perder continuidad y capacidad de comparación; por otra parte, las nuevas preguntas se diseñaron teniendo en cuenta la necesidad de trabajar la recolección de la información con informante directo en los casos que aplique.

Para la revisión del cuestionario se tuvieron en cuenta los siguientes parámetros:

- Asesoría de expertos externos nacionales e internacionales y del Comité Técnico de Empleo del DANE.
- Experiencias y prácticas internacionales.
- Particularidades del mercado laboral colombiano.

La OIT ha colaborado con Colombia a través de acciones de capacitación, estudios metodológicos y programas de asistencia técnica. Un ejemplo respecto a las últimas labores estadísticas de la OIT está relacionado con el tema del trabajo decente que ofrece un marco conceptual para la elaboración e integración de indicadores del mercado de trabajo. En efecto, el concepto de trabajo decente de la OIT está conformado por cuatro pilares: 1) normas, principios y derechos fundamentales en el trabajo; 2) empleo; 3) protección social y 4) diálogo social. Temas que sirven de clasificadores para los indicadores de trabajo decente. En América Latina, la OIT ha establecido un conjunto de 18 indicadores de trabajo decente que permiten medir el progreso logrado en la reducción del déficit de trabajo decente registrado en los países de la región.

## b. Marco conceptual <sup>5</sup>

**Población económicamente activa (PEA):** son las personas en edad de trabajar, que trabajan o están buscando empleo, también se denomina fuerza laboral. Esta población se divide en:

**Ocupado (O)** que corresponde a la persona que durante el período de referencia: trabajó por lo menos una hora remunerada en la semana de referencia; no trabajó la semana de referencia pero tenía un trabajo, y que sea trabajador familiar sin remuneración y trabajó en la semana de referencia por lo menos 1 hora.

**Desocupado (D)** que corresponde a la persona que en la semana de referencia se encontraba en una de las siguientes situaciones:

**Desempleo abierto (DA):**

- Sin empleo en la semana de referencia.
- Hicieron diligencias en el último mes.
- Disponibilidad.

**Desempleo oculto (DO):**

- Sin empleo en la semana de referencia.
- No hicieron diligencias en el último mes, pero sí en los últimos 12 meses y tienen una razón válida de desaliento.
- Disponibilidad.

**Población Económicamente Inactiva (PEI):** comprende a todas las personas en edad de trabajar (10 años y más), que en la semana de referencia no participaron en la producción de bienes y servicios porque no necesitan, no pueden o no están interesadas en tener actividad remunerada. A este grupo pertenecen las personas que son exclusivamente:

- Estudiantes
- Amas de casa
- Pensionados
- Jubilados
- Rentistas
- Incapacitado permanente para trabajar
- Personas que no les llama la atención o creen que no vale la pena trabajar.
- Otros incluidos dentro de la Población en Edad de Trabajar (PET).

**Población en edad de trabajar (PET):** Constituida por todas las personas de 12 años en adelante para las zonas urbanas y de 10 años en adelante en las zonas rurales. Se divide en población económicamente activa y económicamente inactiva.

**Población total (PT):** Constituye la población civil no institucional residente en hogares particulares. Esta población se estima con base en los censos de población, las estadísticas vitales y de migración.

**Subempleo subjetivo:** se refiere al simple deseo manifestado por el trabajador de mejorar sus ingresos, el número de horas trabajadas o tener una labor más propia de sus personales competencias.

**Subempleo objetivo:** comprende a quienes tienen el deseo, pero además han hecho una gestión para materializar su aspiración y están en disposición de efectuar el cambio.

**Subempleo por insuficiencia de horas:** ocupados que desean trabajar más horas ya sea en su empleo principal o secundario y tienen una jornada inferior a 48 horas semanales.

## Condiciones de empleo inadecuado

**Por competencias:** puede incluir todas las personas que trabajan y que durante el período de referencia, desean o buscan cambiar su situación de empleo actual para utilizar mejor sus competencias profesionales y están disponibles para ello.

**Por ingresos:** puede incluir todas las personas ocupadas que, durante el periodo de referencia, deseaban o buscaban cambiar su situación actual de empleo, con objeto de mejorar sus ingresos limitados.

## c. Marco legal

Resolución 1341 de 2010 mediante la cual se adopta el manual de elaboración de indicadores del Mercado Laboral provenientes de encuestas a hogares y sus disposiciones técnicas para la transmisión de datos estadísticos.

## d. Referentes internacionales

Los conceptos, las definiciones, el formulario básico y el sistema de recolección se enmarcan en las resoluciones y recomendaciones que produce la OIT y la CIET. El DANE también emprendió la revisión del marco conceptual y metodológico de la encuesta rural, sobre la frecuencia de su aplicación y de la utilización de conceptos y definiciones que se usan para las encuestas urbanas. Asimismo, se intentó encontrar un mecanismo técnico que permitiera incorporar a la operación dos elementos que son fundamentales: la estacionalidad de la actividad económica en las áreas rurales y la movilidad de la mano de obra. Se revisaron las experiencias al respecto de otros países de América Latina, pero no se hallaron pistas alentadoras, razón por la que se solicitó asistencia técnica nacional -expertos en el tema- e internacional (OIT).

Con este contexto, en mayo de 1998 vino en misión a Colombia Ralf Hussmanns, experto internacional de la OIT en temas relacionados con las estadísticas laborales. Atendiendo sus recomendaciones y las de los expertos nacionales, se hizo un rediseño parcial del cuestionario de las encuestas de cobertura nacional -urbana y rural- y se incluyeron los módulos de historia laboral y movilidad laboral en las encuestas de septiembre de 1998, y de marzo y septiembre de 1999. También se planeó la aplicación como prueba piloto de una Encuesta Continua de Hogares en Bogotá.

En abril de 1999, Hussmanns cumplió una nueva misión de asesoría para evaluar los avances logrados, de la que se concluyó la viabilidad de reemplazar las encuestas trimestrales o transversales de hogares por encuestas continuas. En forma paralela, el DANE conformó el Comité Técnico de Empleo, que se reunió por primera vez el 20 de mayo del mismo año.

En relación con la Comunidad Andina (CAN), el Programa Estadístico Comunitario buscó la armonización estadística en temas como población, mercado de trabajo, cuentas nacionales, moneda y banca, balanza de pagos, entre otros. El citado programa pretendía fortalecer los sistemas nacionales de estadística y la coordinación subregional, con miras a disponer de información sobre la base de definiciones, conceptos y metodologías comunes o similares.

<sup>5</sup> Las siguientes definiciones para la clasificación de la población están fundamentadas en las recomendaciones de la OIT, específicamente del libro Encuestas de población económicamente activa, empleo y subempleo: Manual de la OIT sobre conceptos y métodos, (Hussmanns, Mehran y Verma, 1990).

El proyecto de Cooperación entre la Unión Europea y la Comunidad Andina, (ANDESTAD), lanzado en junio 2008 entre la Unión Europea y la Comunidad Andina, buscó mejorar la calidad de las estadísticas que sirviera para la preparación, la gestión y la evaluación de las políticas públicas a nivel nacional, y, en particular, de las políticas que tienen en común los países que forman la Comunidad Andina. Este proyecto propiciaba la armonización de las metodologías estadísticas para hacer comparable la información entre los países, y también una amplia difusión de la misma.

Por su parte, el Sistema de Integración Económica Centroamericana (SIECA) ha puesto en marcha un Sistema Regional de Información Laboral, con el fin de proporcionar las estadísticas laborales de los países de la sub-región por indicador en un solo portal -desagregados por edad y sexo-. A su vez, la OIT ha elaborado un estudio sobre las diferencias conceptuales y de sistemas de clasificación relativo a los indicadores del mercado de trabajo en los países centroamericanos.

#### e. Referentes nacionales

Ley 79 de 1993 por la cual se regula la realización de los censos de población y vivienda en todo el territorio nacional, en su artículo quinto estipula que las personas naturales o jurídicas, de cualquier orden o naturaleza, domiciliadas o residentes en el territorio nacional, están obligadas a suministrar al Departamento Administrativo Nacional de Estadística DANE, los datos solicitados en el desarrollo de Censos y Encuestas.

### 2.1.5. Diseño de indicadores

#### a. Principales indicadores

##### Porcentaje de PET

$$\%PET=(PET/PT)*100$$

Este indicador muestra la relación porcentual entre el número de personas que componen la población en edad de trabajar, frente a la población total.

##### Tasa Global de Participación o (TPG)

$$TGP= PEA/PET*100$$

Es la relación porcentual entre la población económicamente activa y la población en edad de trabajar. Este indicador refleja la presión de la población en edad de trabajar sobre el mercado laboral.

##### Tasa Bruta de Participación o (TBP)

$$TBP=PEA/PT*100$$

Este indicador muestra la relación porcentual entre el número de personas que componen el mercado laboral, frente al número de personas que integran la población total.

##### Tasa de Desempleo o (TD)

$$TD=DS/PEA*100$$

Es la relación porcentual entre el número de personas que están buscando trabajo (DS), y el número de personas que integran la fuerza laboral (PEA).

Está compuesta por:

##### Tasa de Desempleo Abierto (TDA)

$$TDA=DSA/PEA*100$$

##### Desempleo abierto (DSA)

Personas que en la semana de referencia se encontraban en las siguientes situaciones:

- Sin empleo en la semana de referencia.
- Hicieron diligencias en el último mes.
- Disponibilidad.

##### Tasa de Desempleo oculto (TDO)

$$TDO=DSO/PEA*100$$

##### Desempleo oculto (DSO)

Personas que en la semana de referencia se encontraban en las siguientes situaciones:

- Sin empleo en la semana de referencia.
- No hicieron diligencias en el último mes, pero sí en los últimos 12 meses y tienen una razón válida de desaliento.
- Disponibilidad.

##### Tasa de Ocupación o (TO)

$$TO=OC/PET*100$$

Es la relación porcentual entre la población ocupada (OC) y el número de personas que integran la población en edad de trabajar (PET).

##### Tasa de Subempleo o (TS)

$$TS=S/PEA*100$$

Es la relación porcentual entre el número de personas subempleadas (S), y el de personas que integran la fuerza laboral (PEA).

Está compuesta por:

##### Tasa de Subempleo por Insuficiencia de Horas (TSIH)

$$TSIH=SIH/PEA*100$$

Incluye, a las personas ocupadas que desean trabajar más horas ya sea en su empleo principal o secundario y tienen una jornada inferior a 48 horas semanales

Tasa de subempleo por condiciones de empleo inadecuado

• Tasa de Subempleo Competencias o (TSC)

$$TSC= SC/PEA*100$$

Incluye a todas las personas que trabajan y que durante el período de referencia, desean o buscan cambiar su situación de empleo actual para utilizar mejor sus competencias profesionales y están disponibles para ello.

• Tasa de Subempleo por Ingresos o (TSI)

$$TSI= SI/PEA*100$$

Incluye, todas las personas ocupadas que, durante el período de referencia, deseaban o buscaban cambiar su situación actual de empleo, con objeto de mejorar sus ingresos limitados.

## 2.1.6. Plan de resultados

Los resultados de la GEIH se publican mensualmente a través de boletín técnico, comunicado de prensa, anexos estadísticos en formato Excel, presentaciones de resultados generales y presentaciones de las 23 ciudades y áreas metropolitanas. Se publican a su vez los microdatos anonimizados<sup>6</sup> que contienen las bases de datos de la operación. Esta información se encuentra disponible con dos meses de rezago respecto al mes de la publicación de la información.

### 2.1.6.1. Diseño de cuadros de salida o de resultados

Para conocer los distintos cuadros de salida y resultados de la Gran Encuesta Integrada de Hogares, ver anexo A.

### 2.1.7. Diseño del formulario o cuestionario

El formulario de la Gran Encuesta Integrada de Hogares está compuesto por los siguientes capítulos:

- DATOS DE LA VIVIENDA: El objetivo de este capítulo es conocer las características físicas de la vivienda que permitan determinar la calidad de la misma y determinar el acceso a servicios públicos domiciliarios.
- DATOS DEL HOGAR: El objetivo de este capítulo es determinar el grado de hacinamiento que tienen los hogares, establecer las condiciones y la calidad del suministro de los servicios con los que cuentan y determinar las condiciones de salubridad teniendo en cuenta el abastecimiento de agua, el tipo de servicio sanitario y el manejo de las basuras. Así mismo, permite conocer los bienes en uso con los que disponen.
- REGISTRO DE PERSONAS: El objetivo de este capítulo es registrar a todas las personas que forman parte del hogar.
- CARACTERÍSTICAS GENERALES: El objetivo de este capítulo es caracterizar a la población de acuerdo a su edad, sexo, estado civil y parentesco con el jefe del hogar, con el objetivo de relacionar esta información con el resto de la operación.
- SEGURIDAD SOCIAL EN SALUD: El objetivo de este capítulo es identificar si las personas tienen acceso al sistema de salud, indagando por la afiliación y el tipo de régimen de seguridad social.
- EDUCACIÓN: El objetivo de este capítulo es identificar el alfabetismo, la asistencia escolar en la modalidad de educación formal, el nivel educativo alcanzado y el grado de estudios de las personas encuestadas.
- FUERZA DE TRABAJO: El objetivo de este capítulo es clasificar la fuerza de trabajo entre ocupados, desocupados o inactivos.
- OCUPADOS: El objetivo de este capítulo es conocer las características de la población ocupada y de sus empleos.
- DESOCUPADOS: El objetivo de este capítulo es conocer las características de la población desocupada y su búsqueda de empleo.
- INACTIVOS: El objetivo de este capítulo es conocer las características de la población inactiva.
- OTRAS ACTIVIDADES Y AYUDAS EN LA SEMANA PASADA: El objetivo de este capítulo es captar actividades realizadas por miembros del hogar y por las cuales no obtienen remuneración. Adicionalmente permite captar las actividades de trabajo comunitario voluntario.
- OTROS INGRESOS: El objetivo de este capítulo es determinar cuáles personas recibieron otros ingresos distintos a los laborales.

LETRA	NOMBRE DEL CAPITULO	NÚMERO DE PREGUNTAS
A	Identificación	19
B	Vivienda	5
C	Datos del Hogar	16
D	Registro de Personas	4
E	Características generales	7
F	Seguridad Social en Salud	7
G	Educación	5
H	Fuerza de trabajo	14
I	Ocupados	74
J	Desocupados	14
K	Inactivos	10
L	Otras actividades	1
M	Ingresos no laborales	4
N	Módulo de Micronegocios	6
O	Módulo de Fecundidad	3
P	Módulo de Migración	7
Q	Módulo Formación para el trabajo*	16
R	Módulo de Trabajo Infantil**	22
S	Módulo de limitaciones en el funcionamiento humano***	13

\* A partir del año 2013 se aplica en el segundo trimestre de cada año.

\*\* A partir del año 2012 se aplica en el cuarto trimestre de cada año.

\*\*\* En el año 2015 se aplicó en el segundo semestre del año.

Ver anexo B.

<sup>6</sup> Para más información dirigirse al numeral 6.1 Anonimización de las bases de datos.

### 2.1.8. Normas, especificaciones o reglas de validación, consistencia e imputación

Dentro del proceso de aseguramiento de la calidad de la GEIH, se cuenta con especificaciones y validaciones de consistencia que permiten coherencia en la información recolectada y en los flujos o pasos de preguntas en el formulario.

*Validación de rangos de acuerdo con la estructura de las preguntas*

Cuando la pregunta tiene predefinidos las opciones de respuesta, es necesario controlar las opciones a mostrar dadas las restricciones que existen sobre unicidad o valor máximo posible de ocurrencia.

*Validación de los universos*

A este proceso lo definen tres aspectos: el primero, cuando la pregunta define un flujo o salto dependiendo de la opción; el segundo cuando los datos de la vivienda se toman una sola vez, así exista más de un hogar; el tercero, cuando se define en el universo según la edad de la persona residente en el hogar<sup>7</sup>.

Para algunas variables relacionadas con la aplicación de conceptos básicos de la GEIH que son determinantes en la calidad de la estructura de la información, el DMC permite que el entrevistador confirme la respuesta que da el entrevistado. Estas variables son: número de hogares en la vivienda; nombres y apellidos; sexo; edad; parentesco, y todas las variables relacionadas con valores de los diferentes formularios.

En el diseño del programa se garantiza que en la misma pantalla se muestren el capítulo que está relacionado con las variables como las diferentes preguntas que debido a su alta correlación debe controlarse la consistencia<sup>8</sup>.

### 2.1.9. Nomenclaturas y clasificaciones utilizadas

En esta operación estadística se emplean clasificaciones estadísticas internacionales elaboradas por la CIET de la OIT. A continuación se presentan las clasificaciones empleadas<sup>9</sup>

- Clasificación Internacional Uniforme de Ocupaciones (CIUO - 1968).
- Clasificación Internacional Normalizada de la Educación Adaptada para Colombia (CINE – 2011 A.C.).
- Clasificación Internacional de la Situación en el Empleo (CISE - 1993).
- Clasificación Industrial Internacional Uniforme de Todas las Actividades Económicas (CIIU - Rev. 3)<sup>10</sup>.
- Clasificación Industrial Internacional Uniforme de Todas las Actividades Económicas (CIIU - Rev. 4)<sup>11</sup>.
- Clasificación Nacional de Ocupaciones (CNO - 70)<sup>12</sup>.

## 2.2. DISEÑO ESTADÍSTICO

### 2.2.1. Componentes básicos del diseño estadístico

#### Universo

El universo para la Gran Encuesta Integrada de Hogares está conformado por la población civil no institucional residente en hogares particulares. Esta población se estima con base en los censos de población, las estadísticas vitales y de migración.

#### Población objetivo

Corresponde a la población civil no institucional residente en todo el territorio nacional. Se excluyen de los nuevos departamentos las cabeceras municipales (no capitales de departamento) centros poblados y rural disperso, en los cuales reside aproximadamente el 1% de la población del país, y se encuentra dispersa en casi la mitad de la superficie del país.

#### Marco estadístico o marco muestral

Está constituido por el inventario cartográfico y el listado de viviendas obtenidas de la Muestra Maestra de Hogares, con actualizaciones permanentes y nuevos recuentos de edificaciones y viviendas a través de la misma.

#### Fuente de datos

Encuesta por muestreo probabilístico, multietápico, estratificado, de conglomerados desiguales y autoponderado.

#### Cobertura y desagregación geográfica

Para poder comprender la cobertura y desagregación geográfica de la GEIH primero se deben definir los conceptos de Cabeceras, centros poblados y rural disperso.

Cabecera: es el área geográfica que está definida por un perímetro urbano, cuyos límites se establecen por acuerdos del Concejo Municipal. Corresponde al lugar en donde se ubica la sede administrativa de un municipio.

Centro poblado: es un concepto creado por el DANE para fines estadísticos, útil para la identificación de núcleos de población. Se define como una concentración de mínimo veinte (20) viviendas contiguas, vecinas o adosadas entre sí, ubicada en el área rural de un municipio o de un Corregimiento Departamental. Dicha concentración presenta características urbanas tales como la delimitación de vías vehiculares y peatonales. Este concepto considera:

- **Caserío:** sitio que presenta un conglomerado de viviendas, ubicado comúnmente al lado de una vía principal y que no tiene autoridad civil. El límite censal está definido por las mismas viviendas que constituyen el conglomerado.
- **Inspección de Policía:** es una instancia judicial en un área que puede o no ser amanzanada y que ejerce jurisdicción sobre un determinado territorio municipal, urbano o rural y que depende del departamento o del municipio. Es utilizada en la mayoría de los casos con fines electorales. Su máxima autoridad es un Inspector de Policía.
- **Corregimiento municipal:** es una división del área rural del municipio, la cual incluye un núcleo de población, considerada en los Planes de Ordenamiento Territorial, P.O.T. El artículo 117 de la ley 136 de 1.994 faculta al concejo municipal para que mediante acuerdos establezca esta división, con el propósito de mejorar la prestación de los servicios y asegurar la participación de la ciudadanía en los asuntos públicos de carácter local.

Rural disperso: se caracteriza por la disposición dispersa de viviendas y explotaciones agropecuarias existentes en ella. No cuenta con un trazado o nomenclatura de calles, carreteras, avenidas, y demás. Tampoco dispone, por lo general, de servicios públicos y otro tipo de facilidades propias de las áreas urbanas.

La GEIH tiene cobertura nacional y se han establecido los siguientes niveles de desagregación temporal y geográfica:

- **Mensuales:** Para el conjunto de las 13 grandes ciudades capitales y sus áreas metropolitanas, el grupo de las 11 ciudades capitales intermedias y para el total nacional.
- **Trimestrales:** Por ciudad capital (grandes e intermedias) con su correspondiente área metropolitana y para el total del país por zona (cabecera, centros poblados y rural disperso).
- **Semestrales:** Por ciudad capital (grandes e intermedias), por grandes regiones (Atlántica, Oriental, Central, Pacífica y Bogotá) y para cabeceras y centros poblados y rural disperso y para el total nacional por zona (cabeceras y centros poblados y rural disperso).
- **Anuales:** Por ciudad capital con su área metropolitana, por grandes regiones y zona (cabeceras y centros poblados y rural disperso) y por departamentos.

<sup>7</sup> Existen preguntas que se realiza a: toda las personas; a las de 5 años o más; a las de 12 años o más.

<sup>8</sup> Los anteriores aspectos hacen parte del documento Especificaciones de normas técnicas de validación y consistencia que tienen de forma detallada cada una de las preguntas de los formularios. Para mayor información ver Anexo C.

<sup>9</sup> Ver anexo D de estas clasificaciones con sus principales niveles de desagregación.

<sup>10</sup> A partir de julio de 2001.

<sup>11</sup> A partir de abril de 2016.

<sup>12</sup> Elaborada en 1970 entre el Servicio Nacional de Aprendizaje (SENA) y el DANE.

Las 13 grandes ciudades con sus áreas metropolitanas, las 11 ciudades intermedias y las 8 ciudades capitales de los nuevos departamentos, que son en su orden:

• **13 grandes ciudades con sus áreas metropolitanas:**

- Bogotá
- Medellín - Valle de Aburrá
- Cali - Yumbo
- Barranquilla - Soledad
- Bucaramanga - Floridablanca - Girón - Piedecuesta
- Manizales - Villamaría
- Pasto
- Pereira - Dosquebradas - La Virginia
- Ibagué
- Cúcuta - Villa del Rosario - Los Patios - El Zulia
- Villavicencio
- Montería
- Cartagena

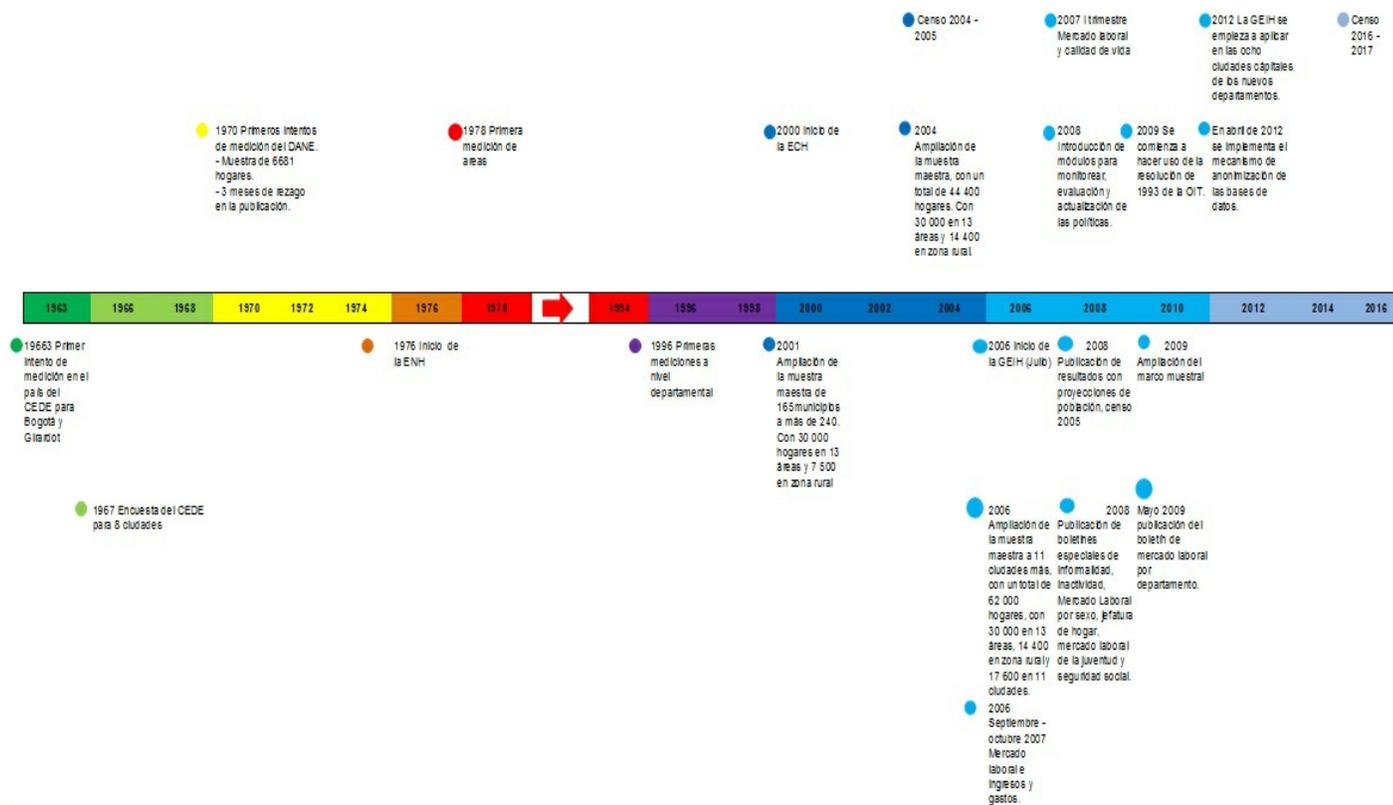
• **11 ciudades intermedias:**

- Tunja
- Florencia
- Popayán
- Valledupar
- Quibdó
- Neiva
- Riohacha
- Santa Marta
- Armenia
- Sincelejo
- San Andrés<sup>13</sup>

• **8 ciudades capitales de los nuevos departamentos nacionales:**

- Mitú
- Yopal
- Inírida
- Leticia
- Arauca
- Mocoa
- Puerto Carreño
- San José del Guaviare

### Línea y avances de la investigación a lo largo del tiempo



Fuente: DANE

<sup>13</sup> Por el tamaño de esta ciudad, a partir de 2012 se entregan sus estimaciones semestralmente.

## Desagregación temática

Los resultados de los principales indicadores de mercado laboral se publican con la siguiente desagregación temática:

- *Total nacional y total 13 ciudades y áreas metropolitanas, información mensual:* Población total, en edad de trabajar, económicamente activa, ocupados, desocupados (abiertos y ocultos), inactivos y subempleados. (en miles) y porcentaje de la población en edad de trabajar, tasa global de participación, de ocupación, de desempleo (abierto y oculto) y de subempleo.
- *Total nacional, total cabeceras, centros poblados y rural disperso, total 13 ciudades y áreas metropolitanas y 23 ciudades y áreas metropolitanas información trimestral:* Población total, en edad de trabajar, económicamente activa, ocupados, desocupados (abiertos y ocultos), inactivos y subempleados. (en miles) y porcentaje de la población en edad de trabajar, tasa global de participación, de ocupación, de desempleo (abierto y oculto) y de subempleo.
- *Jefes de hogar para el total nacional, total cabeceras, centro poblados y rural disperso y total 13 ciudades y áreas metropolitanas, información trimestral:* Población total, en edad de trabajar, económicamente activa, ocupados, desocupados (abiertos y ocultos), inactivos y subempleados. (en miles) y porcentaje de la población en edad de trabajar, tasa global de participación, de ocupación, de desempleo (abierto y oculto) y de subempleo.
- *Total nacional, total cabeceras, centro poblados y rural disperso y total 13 ciudades y áreas metropolitanas, información trimestral:* Población ocupada según ramas de actividad y posición ocupacional. Población desocupada cesante según ramas de actividad anterior. Población inactiva según tipo de inactividad.
- *San Andrés, información trimestral de 2007 a 2011 e información semestral de 2012 en adelante:* Población total, en edad de trabajar, económicamente activa, ocupados, desocupados (abiertos y ocultos), inactivos y subempleados. (en miles) y porcentaje de la población en edad de trabajar, tasa global de participación, de ocupación, de desempleo (abierto y oculto) y de subempleo.
- *Boletines técnicos y anexos estadísticos adicionales con los principales indicadores de mercado laboral, información trimestral:* Mercado laboral según sexo, inactividad, empleo informal y seguridad social, juventud, parentesco, regiones, departamentos, nuevos departamentos y educación.

### 2.2.2. Unidades estadísticas

La unidad de observación son aquellos conjuntos de elementos que componen el universo con una característica específica. En este caso la unidad básica de observación es el hogar.

La unidad de análisis corresponde a la entidad objeto de estudio, que en el caso de la GEIH corresponde a las viviendas, los hogares y las personas.

La unidad de muestreo es la Medida de Tamaño - MT, o segmento. La medida de tamaño o segmento está conformada por un promedio de diez viviendas las cuales se indagan todas; se incluyen todos los hogares de cada vivienda y cada persona de cada hogar.

### 2.2.3. Periodo de referencia y recolección

#### Periodo de referencia

Para mercado laboral se considera período de referencia la semana inmediatamente anterior a la que se realiza la encuesta.

- Semana de referencia o semana pasada. Se define como la semana calendario completa, de lunes a domingo, que precede inmediatamente a la semana en que se efectúa la encuesta.
- Con respecto a la información relacionada con los ingresos laborales se maneja como periodo de referencia el mes pasado<sup>14</sup>
- Para el caso de la clasificación de la población desocupada en las categorías de desempleo abierto u oculto y la identificación de los ingresos no laborales se manejan como periodos de referencia el último mes y los últimos doce meses<sup>15</sup>

Las preguntas básicas que el encuestador formulará se aplican a las condiciones existentes en la *semana pasada o semana de referencia*.

#### Periodo de recolección

- Semana de recolección. Se define como la semana calendario durante la cual se lleva a cabo la recolección u obtención de la información.
- El periodo de recolección es semanal para las 32 ciudades con sus áreas metropolitanas y para las cabeceras, centros poblados y rural disperso se maneja un promedio de 21 días.

### 2.2.4. Diseño muestral (aplica en investigaciones por muestreo)

#### a. Tipo de muestreo

Teniendo en cuenta los objetivos y las características del marco, se optó por una muestra probabilística, estratificada, de conglomerados y multietápica.

##### Probabilística

Cada unidad de la población objetivo tiene una probabilidad de selección conocida y superior a cero. Este tipo de muestra permite establecer anticipadamente la precisión deseada en los resultados principales, y calcular la precisión observada en todos los resultados obtenidos.

##### Estratificada

Es la clasificación de las unidades de muestreo del universo en grupos homogéneos, en función de variables independientes, altamente asociadas con los indicadores de estudio y poco correlacionadas entre sí, con el objeto de maximizar la precisión de los resultados. Este método asegura una mejor precisión de la muestra, al disminuir la varianza de las estimaciones.

Los criterios de estratificación son:

A nivel macro, el país se clasifica en dos subuniversos: el primero corresponde a las 24 ciudades con sus áreas metropolitanas y el segundo, al resto del país constituido por las demás cabeceras municipales, y sus correspondientes zonas rurales centros poblados y rural disperso

Cada capital o área metropolitana es autorrepresentada y tiene probabilidad de selección 1. Para la estratificación y selección de la muestra, el marco de cada municipio se organizó según la nomenclatura cartográfica establecida, sectores, secciones y manzanas, con la información del número de viviendas y hogares y el estrato socioeconómico.

Para el segundo subuniverso, el proceso de selección contempló la previa estratificación de las Unidades Primarias de Muestreo - UPM. Adicionalmente, la selección de las UPM a partir de los estratos elaborados se realizó con la Técnica de Selección Controlada (Goodman, and Kish L. 1950) que es un proceso extendido de estratificación, que optimiza aún más la selección.

Las UPM se estratificaron de acuerdo con los siguientes criterios:

- Geográficos, a nivel de departamento.
- Socioeconómicos, a nivel municipal, con los siguientes indicadores:
  - Tamaño poblacional del municipio o UPM.
  - Nivel de urbanización, en términos de la cantidad de población de las cabeceras municipales.
  - Estructura urbano-rural de la población municipal (% de población en cabecera).
  - Proporción de la población con necesidades básicas insatisfechas (NBI), que a su vez se clasificó en 4 intervalos: A [0-29,8%]; B [29,8%-42,7%]; C [42,7%-57,3%]; y D [57,3%-100%].

Cada municipio con 7.000 o más habitantes en su población total, se constituyó en Unidad Primaria de Muestreo (UPM). Los de menor población se agruparon con un vecino para constituirse en UPM.

Las UPM cuyo tamaño era similar al tamaño promedio de los estratos, se constituyeron en estratos de certeza; para efectos probabilísticos tuvieron probabilidad 1. Las demás UPM se agruparon en estratos de no certeza, aplicando las variables de estratificación en el orden de prioridad especificado antes.

##### De conglomerados

Corresponde a la unidad final de muestreo, que es la medida de tamaño o segmento; es el área que contiene un promedio de diez viviendas, en la que se investigan todas las viviendas, todos los hogares y todas las personas.

La concentración de la muestra en conglomerados presenta grandes beneficios en el nivel operativo, pues reduce, en forma considerable, los tiempos y los costos, y facilita la supervisión.

##### Multietápica

Unidades Primarias de Muestreo (UPM). Hay UPM de inclusión forzosa y de inclusión probabilística:

<sup>14</sup> Mes pasado: se refiere al mes inmediatamente anterior al que se está realizando la encuesta.

<sup>15</sup> Últimos doce meses: se refiere a los últimos 12 meses inmediatamente anteriores al mes en que se está realizando la encuesta.

- Son de inclusión forzosa las áreas metropolitanas y el resto de capitales departamentales que son autorrepresentadas, pues su tamaño muestral permite estimaciones trimestrales separadas para cada una de ellas, y aquellos municipios de tamaño intermedio cuya población es similar o mayor que el tamaño promedio de los estratos de su departamento.

- Son de inclusión probabilística las UPM seleccionadas con PPT de los estratos conformados por dos o más UPM (estratos de no certeza). La técnica de selección controlada utilizada permitió mediante un procedimiento probabilístico incrementar las probabilidades de selección de las combinaciones preferidas de UPM que podían ser seleccionadas en un departamento, adicionalmente a lo obtenible con un muestreo estratificado corriente.

En el caso de la Muestra Maestra la selección controlada proporcionó un mejor balance interregional en las características de las UPM seleccionadas. El resultado final de este ejercicio fue lograr la máxima reducción posible en el error estándar de estimación de los resultados.

**Unidades Secundarias de Muestreo (USM).** En las cabeceras municipales y centros poblados, las USM son grupos de manzanas contiguas del mismo municipio, que contienen un mínimo de 12 medidas de tamaño (MT) con un promedio de 10 viviendas, y en la zona rural dispersa, la USM es una sección o sector censal. El tamaño de la USM permite un control adecuado de la distribución y rotación de la muestra en cada etapa. Para la selección de las USM se aplica el siguiente procedimiento (PPT):

- Conformación, delimitación y listado de las USM por estrato socioeconómico (solo cabeceras municipales) de cada municipio seleccionado, siguiendo el orden de la nomenclatura de la cartografía censal: sector, sección y manzana.

- Cálculo del número de medidas de tamaño (MT), dividiendo por 10 (el número de viviendas) y aproximando a un valor entero.

- Cálculo de intervalo de muestreo (IM), dividiendo el total de MT existentes en una zona por el número de USM requeridas en la muestra de la zona.

- Selección de un número de arranque entre 1 y el valor del intervalo, el cual indica la primera USM de la muestra de la zona; y selección del resto de USM de la zona, sumando sucesivamente el IM al número de arranque.

**Unidades Terciarias de Muestreo (UTM).** El tercer paso es la selección dentro de cada USM de una o más UTM, dependiendo de las requeridas, del tamaño y cantidad de USM preseleccionadas en cada municipio.

Cada UTM es un segmento con un promedio de 10 viviendas contiguas (5 a 14 viviendas), es decir, una medida de tamaño (MT). En un paso intermedio, se seleccionan manzanas en las cabeceras municipales, centros poblados y segmentos de sección rural claramente delimitados, con límites naturales (quebradas, ríos, etc.) o sociales (carreteras, caminos, etc.). Los segmentos de los sectores rurales se conforman previamente con base en fotografías aéreas y en lo posible, se procura que tengan un promedio de 10 viviendas; sin embargo, por insuficiente detalle cartográfico, los segmentos pueden tener dos o más MT.

La selección se hace con probabilidad proporcional al número de MT (PPT), así una manzana o segmento rural puede contener una o más MT, y las que tienen menos de 5 viviendas se juntan con una manzana vecina del mismo estrato socioeconómico.

Las áreas urbanas no urbanizadas se anexan al área o manzana vecina muy cercana con una o más viviendas con el objeto de que tengan probabilidad de selección y para que puedan captar cualquier proceso futuro de urbanización. Si solo se requiere una MT y la manzana o porción solo tiene una MT, la muestra queda constituida por todas las viviendas y hogares existentes. Si la manzana o segmento rural tienen dos o más MT se procede al recorrido y delimitación de segmentos de una MT, en el terreno. Finalmente, se hace la asignación aleatoria del orden en que entrarán los segmentos al proceso de rotación.

**Unidades Cuartas de Muestreo (UCM).** Cada UCM es un segmento de 10 viviendas contiguas (5 a 14 viviendas), es decir una medida de tamaño MT y su selección es aleatoria.

Finalmente se hace una asignación aleatoria del orden en que entrarán los segmentos al proceso de rotación. En este proceso de selección por etapas, solo se requiere una actualización cartográfica detallada de las áreas seleccionadas.

Para el subuniverso del resto de cabeceras, centros poblados y rural disperso se seleccionó:

- En una primera etapa, la UPM, utilizando la técnica de selección controlada dentro de cada estrato.

- Para la segunda etapa se seleccionó en la cabecera y centros poblados la manzana, y en el rural disperso la sección o sea la USM.

- En la tercera etapa se seleccionó el segmento o UTM.

En este proceso de selección por etapas sólo se requiere una actualización cartográfica detallada de las áreas seleccionadas.

#### **b. Rotación de la muestra**

Dado el carácter continuo de la GEIH y con el fin de controlar la repetición de la encuesta a los mismos hogares en períodos cortos de tiempo, lo cual puede generar rechazos y la reducción en la cobertura y calidad de la encuesta, se definieron siguientes los procedimientos<sup>16</sup> que se resumen a continuación:

En las capitales de departamento se conformaron grupos de MT, mínimo 12, para asegurar que por lo menos en tres años no se repitiera un segmento.

En las UPM probabilísticas la encuesta se aplica en segmentos diferentes, aunque se repiten anualmente el mismo mes; este proceso de rotación es cuidadosamente controlado y se realiza con un trimestre de anticipación.

Se verifican las manzanas y las secciones censales donde se agotan las medidas de tamaño.

Se elabora un recuento de viviendas para conformar nuevos segmentos para el desarrollo de la encuesta.

#### **c. Definición tamaño de la muestra**

Inicialmente el tamaño de la muestra mensual correspondía aproximadamente a 23.000 hogares. En el 2000 con la implementación de la ECH la muestra maestra fue ampliada de 165 municipios a más de 240; con 30.000 hogares en 13 áreas y 7.500 en el resto de cabeceras, centros poblados y rural disperso. Durante 2004 se amplió la muestra maestra, con un total de 44.400 hogares, con 30.000 hogares en 13 áreas y 14.400 en el resto de cabeceras, centros poblados y rural disperso. Para el año 2006 con la implementación de la GEIH, se amplió la muestra a 11 ciudades más; con 17.600 hogares adicionales, para un total de 62.000 hogares. Con el marco generado por el censo 2005, la nueva muestra implementada desde 2009 está conformada por 437 municipios y anualmente se visitan aproximadamente 248.028 hogares, concentrados en 22.548 segmentos. La muestra mensual es de 20.669 hogares, 18.790 viviendas y 1.879 segmentos en las capitales de los nuevos departamentos se visitan anualmente 1.272 segmentos.

Los tamaños de muestra se calculan con una precisión deseada de la variable tasa de desempleo no superior a un error estándar relativo del 5% y una tasa de desempleo del 10%. Los cálculos se realizan con las fórmulas correspondientes al tipo de diseño muestral. Se ajustan con base en el efecto de los conglomerados en el diseño (deff) el cual es una relación para cada dominio, entre la varianza real de este diseño de conglomerados y la que se obtendría con un diseño aleatorio simple de elementos.

El tamaño de muestra está dado por la siguiente fórmula:

$$n = \frac{NPQdeff}{N(ESrelP)^2 + PQdeff}$$

Donde:

n = tamaño de muestra

N = tamaño del universo

P = probabilidad de ocurrencia del fenómeno estudiado

Q = 1 - P

ESrel = Error estándar relativo

$$deff = \frac{Var(\text{conglomerados})}{Var(MAS)}$$

Efecto de los conglomerados en el diseño

#### **2.2.5. Ajustes Cobertura**

**Factor básico de expansión (F):** aplicado a los datos muestrales, da a cada elemento de la muestra el peso o representación que le corresponde en el universo investigado. En consecuencia, mediante su aplicación, se estiman numéricamente, en forma aproximada, las características de la población objetivo.

**Peso de submuestreo (P<sub>h</sub>):** está dado por segmento y es teóricamente igual a 1 para todos los segmentos en razón de que representan una medida de tamaño. El desarrollo dinámico del marco de muestreo en algunos casos puede ser mayor o menor que 1, de acuerdo con la densidad de viviendas del segmento en el momento de hacer la encuesta, y por tal motivo modifica el factor básico de expansión en el segmento.

**Ajuste de cobertura por no respuesta (R<sub>h</sub>):** cuando las tasas de no respuesta varían en los subgrupos de la población de diferentes características, el ajuste normal es asignar a los hogares y a las personas no encuestadas el promedio de las características de los encuestados en el mismo segmento. Esto se logra corrigiendo el factor básico de expansión por un nuevo factor resultante de la razón entre el número de

<sup>16</sup> Estos se encuentran consignados en los respectivos manuales operativos.

### c. Definición tamaño de la muestra

Inicialmente el tamaño de la muestra mensual correspondía aproximadamente a 23.000 hogares. En el 2000 con la implementación de la ECH la muestra maestra fue ampliada de 165 municipios a más de 240; con 30.000 hogares en 13 áreas y 7.500 en el resto de cabeceras, centros poblados y rural disperso. Durante 2004 se amplió la muestra maestra, con un total de 44 400 hogares, con 30 000 hogares en 13 áreas y 14 400 en el resto de cabeceras, centros poblados y rural disperso. Para el año 2006 con la implementación de la GEIH, se amplió la muestra a 11 ciudades más; con 17.600 hogares adicionales, para un total de 62.000 hogares. Con el marco generado por el censo 2005, la nueva muestra implementada desde 2009 está conformada por 437 municipios y anualmente se visitan aproximadamente 248.028 hogares, concentrados en 22.548 segmentos. La muestra mensual es de 20.669 hogares, 18.790 viviendas y 1.879 segmentos en las capitales de los nuevos departamentos se visitan anualmente 1.272 segmentos.

Los tamaños de muestra se calculan con una precisión deseada de la variable tasa de desempleo no superior a un error estándar relativo del 5% y una tasa de desempleo del 10%. Los cálculos se realizan con las fórmulas correspondientes al tipo de diseño muestral. Se ajustan con base en el efecto de los conglomerados en el diseño (deff) el cual es una relación para cada dominio, entre la varianza real de este diseño de conglomerados y la que se obtendría con un diseño aleatorio simple de elementos.

El tamaño de muestra está dado por la siguiente fórmula:

$$n = \frac{NPQdeff}{N(ESrelP)^2 + PQdeff}$$

Donde:

n = tamaño de muestra

N = tamaño del universo

P = probabilidad de ocurrencia del fenómeno estudiado

Q = 1 - P

ESrel = Error estándar relativo

$$deff = \frac{Var(conglomerados.)}{Var(MAS)}$$

Efecto de los conglomerados en el diseño

### 2.2.5. Ajustes Cobertura

*Factor básico de expansión (F)*: aplicado a los datos muestrales, da a cada elemento de la muestra el peso o representación que le corresponde en el universo investigado. En consecuencia, mediante su aplicación, se estiman numéricamente, en forma aproximada, las características de la población objetivo.

*Peso de submuestreo (P<sub>h</sub>)*: está dado por segmento y es teóricamente igual a 1 para todos los segmentos en razón de que representan una medida de tamaño. El desarrollo dinámico del marco de muestreo en algunos casos puede ser mayor o menor que 1, de acuerdo con la densidad de viviendas del segmento en el momento de hacer la encuesta, y por tal motivo modifica el factor básico de expansión en el segmento.

*Ajuste de cobertura por no respuesta (R<sub>h</sub>)*: cuando las tasas de no respuesta varían en los subgrupos de la población de diferentes características, el ajuste normal es asignar a los hogares y a las personas no encuestadas el promedio de las características de los encuestados en el mismo segmento. Esto se logra corrigiendo el factor básico de expansión por un nuevo factor resultante de la razón entre el número de

hogares seleccionados en un segmento y el número de hogares encuestados en el mismo segmento.

*Factor final de expansión (W<sub>h</sub>)*: es el producto de los anteriores y se aplica a cada segmento de la ciudad o área investigada.

$$W_{h\alpha} = F * P_{h\alpha} * R_{h\alpha}$$

La estimación total para una característica Y de la población está dada por:

$$Y' = \sum_{h=1}^H \sum_{\alpha=1}^{ah} Y_{h\alpha} * W_{h\alpha}$$

$$W_{h\alpha} = F * P_{h\alpha} * R_{h\alpha}$$

Donde:

h = 1,2,... H: estratos

a = 1,2,...ah: segmentos seleccionados para la muestra en el estrato h.

ah = Número total de segmentos en la muestra del estrato h.

W<sub>hα</sub> = Factor de restitución del segmento a en el estrato h.

Y<sub>hα</sub> = Estadística Y para el segmento a del estrato h.

#### Ajuste por los valores poblacionales

Las estimaciones del universo de estudio pueden ser mejoradas, si se equiparan a nivel de las desagregaciones geográficas los totales de la población obtenidos de la muestra expandida, con los totales del censo de población proyectados a la fecha de la encuesta. Dicho mejoramiento se basa en la premisa de que la estructura por desagregaciones geográficas, proyectada a partir de un censo de población reciente, es más precisa que la estimada a partir de la muestra.

Se aplican los estimadores de regresión utilizando la información auxiliar correspondiente a las Proyecciones de Población del CENSO 2005, en este caso, se ha usado la información auxiliar por grupos de edad y sexo. Este proceso se realiza a través de los Métodos de Calibración que son procedimientos que utilizan información auxiliar relacionada con las variables de estudio, con el fin de mejorar la precisión y reducir los sesgos en las estimaciones.

Estimador del total

La expresión del estimador de la variable de estudio Y en un periodo de la encuesta es la siguiente:

$$\hat{Y} = \sum_h \frac{P_h}{p_h} \sum_{i=1}^{n_h} y_{hi} \quad (1)$$

Donde:

h = corresponde a los estratos de la ciudad o área metropolitana, la región o el total nacional.

P<sub>h</sub> = proyección de la población, por los grupos de calibración en el estrato h, referida a la mitad del mes de referencia.

p<sub>h</sub> = número de personas que habitan en las viviendas de la muestra, en el estrato h, en el momento de la entrevista.

n<sub>h</sub> = número de viviendas en los segmentos de la muestra en el estrato h.

Y<sub>hi</sub> = valor de la variable de estudio en la vivienda i-ésima, del estrato h.

La metodología de calibración mediante un modelo de regresión considera lo siguiente:  
 Sea una población  $U=\{U_1, \dots, U_n\}$  de la cual se extrae una muestra  $S=\{U_1, \dots, U_k, \dots, U_n\}$

La expresión (1) puede escribirse de la siguiente forma:

$$\hat{Y} = \sum_{k \in S} d_k y_k$$

Donde:

$Y_k$  = valor de la variable de estudio en la unidad muestral  $k$ .

$d_k$  = factor de expansión básico de la unidad  $k$  obtenido mediante la expresión

$$\frac{P_h}{p_h}$$

, siendo  $h$  el estrato al que pertenece la unidad.

$$\sum_{k \in S}$$

, sumatoria extendido a todas las unidades de la muestra  $s$ .

Para aplicar el método de calibración se establecieron los siguientes ocho grupos:

- SEXO  
 - Hombres  
 - Mujeres

Y para cada uno los siguientes grupos de EDAD:

- De 0 a menores de 12 años (para el caso cabecera, en el resto 10 años)
- 12 a menores de 25
- 25 a menores de 55
- 55 y más

Con la información del total de personas en las proyecciones de población, en cada grupo de calibración se construye el vector de totales:

$$X_U^t = \left[ \sum_U x_{1k} \quad \sum_U x_{2k} \quad \sum_U x_{3k} \quad \sum_U x_{4k} \quad \sum_U x_{5k} \quad \sum_U x_{6k} \quad \sum_U x_{7k} \quad \sum_U x_{8k} \right]$$

Donde cada variable  $x_{jk}$  da la cantidad de personas en cada grupo en la vivienda  $k$ . (Se hace a nivel de vivienda ya que es la unidad de análisis más agregada, de esta manera a la vivienda le corresponde un único factor de expansión que es el mismo de los hogares y personas que viven en ella).

A nivel de vivienda se establecen las variables de la siguiente forma:

Vivienda	Hombres < de 12 años $X_{1k}$	Mujeres < de 12 años $X_{2k}$	Hombres entre [12 - 24) años $X_{3k}$	Mujeres entre [12 - 24) años $X_{4k}$	Hombres entre [25 - 54) años $X_{5k}$	Mujeres entre [25 - 54) años $X_{6k}$	Hombres de 55 y + años $X_{7k}$	Mujeres de 55 y + años $X_{8k}$
$k$	1	0	2				2	0

Donde cada variable  $X_{jk}$  corresponde a la cantidad de personas por sexo y edad en la vivienda  $k$ .

De manera que para cada vivienda se establece el vector:

$$x_k^t = [x_{1k} \quad x_{2k} \quad x_{3k} \quad x_{4k} \quad x_{5k} \quad x_{6k} \quad x_{7k} \quad x_{8k}]$$

La idea es ajustar los pesos para que:

$$X_U = \sum_k w_k x_k^t$$

Es decir sumando sobre las variables  $x$  definidas expandidas, se obtienen los valores proyectados de población en cada grupo de edad y sexo, por estrato de diseño.

Estos nuevos pesos se obtienen de manera que sean lo más cercano posible a los pesos dados por el diseño  $d_k = W_k$

Los nuevos pesos son obtenidos mediante minimización de la distancia lineal entre  $w_k$  y  $d_k$ :

Los valores  $V_k$  o sea la corrección de pesos se obtiene de manera iterativa hasta que cumplan con la ecuación de calibración y con los límites establecidos.

El planteamiento del problema es encontrar unos valores  $W_k$  que hagan mínima la expresión:

Siendo:

$G$  = Función de distancia.

$X$  = Vector de dimensión  $(J,1)$  con los totales de las variables auxiliares.

$X_k$  = Vector de dimensión  $(J,1)$  con los valores de las variables auxiliares en la unidad muestral  $k$ .

La solución del problema depende de la función de distancia  $G$  que se utilice.

$$z = \frac{w_k}{d_k};$$

Si se considera la función de distancia lineal de argumento

$$G(z) = \frac{1}{2}(z-1)^2, \quad z \in R$$

La ecuación se resuelve mediante la utilización de los multiplicadores de Lagrange que conducen a la obtención de un conjunto de factores que verifican las condiciones de equilibrio y proporcionan las mismas estimaciones que el estimador de regresión generalizado.

De esta manera la estimación de un total para una variable de estudio se da como:

$$\hat{Y} = \sum_r w_k y_k \quad (1)$$

El cálculo del ajuste por calibración se realiza utilizando la macro Clan 97 v3.1 software creado por Statistics Sweden que corresponde a un conjunto de rutinas en el programa SAS para la estimación puntual y de errores en encuestas por muestreo.

Estimador de la varianza para un total

Para el estimador de un total, la varianza se estima de la siguiente forma:

$$\hat{V}(\hat{Y}) = \sum_h^H \hat{V}_h(\hat{Y}_h) = \sum_h^H \frac{1}{a_h - 1} \left[ a_h \sum_{\alpha}^{a_h} (\hat{e}_k d_k)^2 - \left( \sum_{\alpha}^{a_h} \hat{e}_k d_k \right)^2 \right]$$

Sea h=1,2,...,H los estratos de la muestra

A=1,2,...,a<sub>h</sub> los segmentos seleccionados en el estrato h

ah= Número total de segmentos en la muestra del estrato h

El término  $\hat{e}_k$  se calcula a nivel de vivienda, se suma la variable en estudio para obtener el total Y<sub>k</sub> a nivel de vivienda. Entonces:

$$\hat{e}_k = y_k - x_k^t B_D$$

Donde  $x_k^t$  es el vector de variable auxiliar y

$$B_D = \left( \sum_{S_D} w_k x_k x_k^t \right)^{-1} \left( \sum_{S_D} w_k x_k y_k \right)$$

Es un vector calculado en cada dominio de ajuste.

*Estimador de razón*

Las tasas, razones, proporciones y promedios, generadas a partir de este diseño muestral son de la forma de una razón, en la cual el numerador y el denominador son variables aleatorias, así: Estimador de la varianza para una razón

Para estimar la varianza del estimador de una razón se debe transformar la variable en una función lineal, para esto se utiliza el método de Linealización de Taylor, y se aplica la metodología anterior.

$$\hat{u}_k = \frac{1}{Z} (y_k - z_k \hat{r})$$

Utilizando la fórmula de varianza de un total, para una razón se cambia Y<sub>k</sub> por Con

lo cual se obtiene:

$$\hat{V}(\hat{r}) = \sum_h^H \frac{1}{a_h - 1} \left[ a_h \sum_{\alpha}^{a_h} (\hat{e}_k d_k)^2 - \left( \sum_{\alpha}^{a_h} \hat{e}_k d_k \right)^2 \right]$$

Sea h=1,2,...,H, cada uno de los estratos

A=1,2,...,a<sub>h</sub> segmentos seleccionados en el estrato h

ah= Número total de segmentos en la muestra en el estrato h.

El término  $\hat{e}_k$  se calcula a nivel de vivienda, se suma la variable en estudio para obtener el total Y<sub>k</sub> a nivel de vivienda

Donde:

$$\hat{e}_k = \hat{u}_k - x_k^t B_D$$

Donde  $x_k^t$  es el vector de variable auxiliar y es un vector calculado en cada dominio de ajuste.

$$B_D = \left( \sum_{S_D} w_k x_k x_k^t \right)^{-1} \left( \sum_{S_D} w_k x_k \hat{u}_k \right)$$

#### Coefficiente de variación

Para medir la magnitud de la variabilidad de la distribución muestral del estimador, denominado error muestral, se usan el error estándar y el coeficiente de variación.

El coeficiente de variación se define como la relación porcentual del error estándar o raíz cuadrada de la varianza del estimador y el estimador, multiplicado por 100.

$$CV = \frac{\sqrt{V(\hat{p})}}{\hat{p}} * 100$$

El valor de este coeficiente, expresado en porcentaje, permite evaluar la calidad de un procedimiento de estimación.

#### Calculo de Precisión de los resultados

Se diseñó una muestra para obtener una precisión esperada medida en términos del error de muestreo menor o igual a 5% con un nivel de confiabilidad del 95%, para los principales indicadores de fuerza laboral nacionales.

### 2.3. DISEÑO DE LA EJECUCIÓN

#### 2.3.1. Sistema de capacitación

Para el caso de la GEIH, por ser una encuesta continua y contar con personal operativo que posee experiencia en la encuesta, se realiza un reentrenamiento de repaso y consolidación de conceptos. El reentrenamiento deberá llevarse a cabo con periodicidad trimestral y sin excepción se aplicará un examen de conocimientos que respalde el conocimiento adquirido en la investigación.

El reentrenamiento es un proceso educacional a corto plazo mediante el cual el personal operativo refuerza los conceptos necesarios para realizar las actividades y ejercer las funciones de su rol.

Para llevar a cabo esta actividad se deben tener en cuenta las siguientes recomendaciones:

- Cumplir con las fechas establecidas en el cronograma.
- No descuidar el llamado a lista.
- Desarrollar actividades con un alto contenido práctico.
- Hacer uso de los manuales.
- Efectuar trabajos de campo, en donde se practiquen los temas que se hayan detectado como débiles en el grupo.
- Realizar realimentación de cada uno de los talleres, resolviendo dudas y aclarando conceptos.
- Evaluar por escrito los temas profundizados durante el reentrenamiento a través de un examen técnico.
- Calificar cada uno de los exámenes de manera clara y objetiva.
- Enviar los resultados de los exámenes al área de Logística del DANE Central.

#### 2.3.2. Actividades preparatorias

##### a. Sensibilización

El objetivo del proceso de sensibilización es, lograr una respuesta oportuna y de calidad por parte de las fuentes de información para la operación estadística al DANE. Esta busca lograr un ambiente de colaboración, acercamiento y conocimiento de la población objetivo mediante la visita a cada uno de los hogares a los que les será aplicada la encuesta, por parte de los sensibilizadores, con el fin de informarles que el hogar ha sido seleccionado para ser encuestado. En esta primera visita, se pueden establecer citas y horarios en los cuales los encuestadores del DANE pueden aplicar la encuesta. En caso de que el sensibilizador no encuentre a los miembros de los hogares en la primera visita, deberá realizar visitas durante el transcurso de la semana con el fin de sensibilizar a la cantidad de personas que encuentre en las viviendas y hogares asignados en la muestra.

La sensibilización en las ciudades capitales y áreas metropolitanas, es llevada a cabo por una persona contratada para tal fin. En el caso rural, la sensibilización es realizada por el supervisor de campo junto con su equipo de encuestadores.

##### b. Selección y entrenamiento del personal

La selección y entrenamiento del personal se realizará siguiendo los lineamientos establecidos desde el GIT Área de Logística y Producción de Información de DANE Central, en el procedimiento de selección de personal operativo para las operaciones estadísticas del DANE (PES-002-PD-001), el cual se encuentra disponible en ISOLUCION, al igual que los documentos correspondientes para la contratación de servicios profesionales y de apoyo a la gestión.

En los siguientes documentos se puede consultar la información referente a la convocatoria, selección, entrenamiento y contratación de personal:

- [PROCEDIMIENTO PARA LA SELECCIÓN DEL PERSONAL OPERATIVO DE LAS OPERACIONES ESTADÍSTICAS DEL DANE - PES-002-PD-001](#)
- [PROCEDIMIENTO CONTRATACIÓN DE SERVICIOS PROFESIONALES Y DE APOYO A LA GESTIÓN - GCO-020-PD-001](#)

#### 2.3.3. Diseño de instrumentos

Dentro del proceso de recolección de la información de la GEIH, se cuenta con unos manuales temáticos y operativos que acompañan dicho proceso. Los principales manuales son:

- *Manual de recolección y conceptos básicos*: el equipo de Temática Social elabora este manual que acompaña al formulario, el cual es guía para los supervisores y encuestadores para realizar la recolección de información en campo con el fin de garantizar la homogeneidad y calidad de los resultados.

- *Manual operativo área urbana y rural*: en este documento se consignan los diferentes aspectos que se deben tener en cuenta para el operativo de campo de la encuesta en las ciudades, áreas metropolitanas, cabeceras municipales y centros poblados y rurales dispersos.

- *Manual del usuario*: en el cual se especifican los pasos a seguir en el inicio y fin de la recolección de información en campo para la Gran Encuesta Integrada de Hogares, a través de los dispositivos móviles de captura – DMC.

- *Manual del Supervisor de campo*: en el cual se encuentra la información que el supervisor de campo de la GEIH requiere, junto con las normas establecidas e instrucciones necesarias que deberá seguir y aplicar en el cumplimiento de los objetivos propuestos para este proyecto.

- *Manual para la Sincronización y envío de información*: el cual tiene como finalidad orientar a los Apoyos Informáticos sobre la instalación y manejo del aplicativo para garantizar el envío al DANE Central de toda la información recolectada en los diferentes operativos.

- *Manual de recuento*: el cual tiene por objeto dar a las personas encargadas de realizar el recuento las pautas para el desarrollo de su trabajo, el cual consiste en el conteo y registro de las unidades o estructuras y el uso o destino de éstas, con el propósito de obtener el recuento y segmentación de unidades y viviendas.

- *Manual del Coordinador de Campo*: el manual del Coordinador de Campo ha sido elaborado para proporcionar una referencia que permita asegurar el cumplimiento de los objetos de la Gran Encuesta Integrada de Hogares en el cual se establecen las actividades que debe realizar el Coordinador en cada Sede o Subselección con el fin de lograr una óptima ejecución de las actividades del operativo de la encuesta.

### 2.3.4. Recolección de la información

#### a. Organigrama operativo



Fuente: DANE

#### b. Esquema operativo, método y procesamiento para la recolección

Para manejar eficientemente el trabajo de campo, se conforma un equipo operativo de carácter temporal que funciona sobre la base de cuatro niveles jerárquicos:

*Asistente de encuesta Gran Encuesta Integrada de Hogares.* A nivel regional es el responsable de la organización operativa de la encuesta, de él dependen los grupos de trabajo adscritos a su Dirección Territorial, con quienes debe mantener comunicación permanente.

*Coordinador de campo Gran Encuesta Integrada de Hogares.* En las Direcciones Territoriales del DANE, es el responsable de las actividades preliminares de la encuesta, de la supervisión a los grupos de trabajo en campo y de la realización de por lo menos una re-encuesta semanal a un hogar perteneciente a cada grupo de trabajo, con fin de confirmar si la información recolectada es consistente. En las subsecciones tiene a su cargo el manejo de la encuesta en todos sus aspectos operativos.

*Apoyo Logístico Gran Encuesta Integrada de Hogares.* El Apoyo Logístico colabora con las actividades de recuento y logística, en la elaboración de los informes operativos. Es el encargado de la revisión, preparación y distribución de los requerimientos para la ejecución de los procesos de recolección de la información.

*Apoyo informático.* Es el encargado de transmitir la información y realizar el cruce de segmentos faltantes en las Direcciones Territoriales del DANE.

*Recuentista:* Es la persona encargada de realizar el reconocimiento y recorrido de un área previamente asignada y de efectuar el conteo de las unidades y viviendas que la conforman con el fin de generar la correcta asignación de medidas de tamaño [1] de los segmentos a trabajar. A nivel rural el recuento es efectuado por los encuestadores de manera simultánea con la recolección.

*Sensibilizador:* Se encarga, antes de la recolección, de dar a conocer la encuesta en los hogares en donde se va a realizar, mediante el barrido de cada segmento objeto de la muestra; utilizando las piezas comunicativas, dirigidas a las personas y hogares que serán fuentes de información.

*Supervisores.* Realizan seguimiento y controlan el trabajo en los segmentos en la ciudad y en los municipios que se les asignen y, reportan su trabajo al coordinador de campo de la encuesta.

*Encuestador.* Conocido también como recolector, es la persona encargada de obtener la información requerida, digitándola en el DMC conforme a las normas y los conceptos establecidos. La labor del encuestador es exigente y requiere un conocimiento preciso de los conceptos y las normas que rigen la encuesta, además de un compromiso orientado a lograr la cobertura y los niveles de calidad propuestos.

Para la ejecución del trabajo de campo, se utiliza un folder por segmento (carga de trabajo) denominado *Carpeta de muestra*, en cuya carátula se encuentra la identificación del área, que se va a trabajar el cual contiene:

- Un plano (cuando así se requiera) del municipio o del segmento en que se encuentra ubicado para orientación del grupo de trabajo.
- El formato de recuento GEIH1.
- Listado de edificaciones y viviendas GEIH2.

La encuesta utiliza un informante directo <sup>17</sup> para las personas de 18 años y más, y para aquellas de 10 a 17 años que trabajen o estén buscando trabajo. Para los demás se acepta informante idóneo (persona del hogar mayor de 18 años, que a falta del informante directo pueda responder correctamente las preguntas). No se acepta información de empleados del servicio doméstico, pensionistas, vecinos o menores, excepto cuando el menor de edad es el jefe del hogar o cónyuge.

El sistema utilizado para la recolección es el de barrido y rutas, para el resto de cabeceras, centros poblados y rural disperso.

#### Dispositivos Móviles de Captura – DMC

La información se recolecta mediante dispositivos móviles de captura (DMC), los cuales contienen el formulario en forma digital, validaciones y flujos de consistencia.

## 2.4. DISEÑO DE SISTEMAS

### a. Especificaciones de consistencia para el formulario

Este documento consta de reglas de validación y consistencia de las variables, representadas en cada una de las preguntas del formulario de recolección. Cuenta además, con los flujos que deben tener las preguntas para cada uno de los casos que se puedan presentar a lo largo de la realización de la encuesta. Adicionalmente cuenta con los flujos de preguntas entre cada uno de los módulos que hacen parte del formulario. Estas validaciones concuerdan con el proceso de validación interno del Dispositivo Móvil de Captura (DMC), revisado conjuntamente entre las áreas de temática y sistemas.

### b. Transmisión de datos a DANE Central

La transmisión de datos de las ciudades de la encuesta se realiza a través del FTP (siglas en inglés de File Transfer Protocol - Protocolo de Transferencia de Archivos) que es un protocolo de red para la transferencia de archivos entre sistemas conectados a una red TCP (siglas en inglés de Transmission Control Protocol), basado en la arquitectura cliente-servidor. Desde un equipo cliente se puede conectar a un servidor para descargar archivos desde él o para enviarle archivos, independientemente del sistema operativo utilizado en cada equipo. El FTP como medio, facilita el proceso de la encuesta asignándole a cada ciudad un buzón en el servidor de FTP del DANE Central, allí se coloca la información de las encuestas completas que se recolectaron diariamente.

#### Consolidación de archivos

Los datos provenientes del área geográfica en donde se aplica la operación se consolidan diariamente, y pueden ser monitoreados por el área temática y sistemas.

Para enviar la información al DANE Central, se utiliza la herramienta Swin que genera un archivo comprimido asignándole un nombre, que indique la operación, ciudad y fecha a que corresponde la información contenida; luego se copia al servidor destinado para el proceso de transmisión, utilizando el protocolo FTP en la carpeta asignada para ello y confirmar inmediatamente que la información está en dicho servidor, se envía mediante correo electrónico. De igual forma, en el DANE Central se realizan las copias de respaldo necesarias y se aplican los planes de contingencia que garanticen el normal funcionamiento del operativo.

### c. Procesamiento de datos

Verificación de la consistencia interna de los datos y ajustes. El proceso para controlar e identificar los datos faltantes, inválidos o inconsistentes se lleva a cabo mediante el análisis de la variable Incompleta que toma valor 0 cuando la encuesta se encuentra completa y se puede utilizar directamente de la base de datos, y toma valor 1 cuando la encuesta debe ser revisada por datos faltantes inválidos o inconsistentes.

<sup>17</sup> Informante directo, se refiere a la persona que da la información sobre sí mismo.

Imputación y/o ajustes de cobertura. El procesamiento es centralizado por DANE Central, e inicia desde el momento en que las Direcciones Territoriales envían la información; los datos son transmitidos por FTP (File Transfer Protocol) y posteriormente cargados en la base de datos Oracle.

En esta operación estadística no se realiza el proceso de imputación, cuando los miembros de los hogares se niegan a responder la encuesta, esta no respuesta es ajustada con los factores de expansión, lo que se explica en el numeral 3.6.1 Componentes del factor de expansión.

Ponderadores. Los resultados muestrales se llevan a un ajuste por variable exógena, que es un estimador independiente de población, y permite mejorar las estimaciones referentes a las desagregaciones geográficas; se igualan los totales estimados con base en la muestra, con los totales del censo de población proyectados a la fecha de la encuesta. Lo anterior se apoya en la premisa de que la estructura por desagregaciones geográfica, proyectada a partir de un censo de población reciente, es más exacta que la estimada a partir de la muestra.

Generación de cuadros de salida:

• Para generar los cuadros de salida, se realizan dos fases:

- Primera fase: a nivel de base de datos, se conforman vistas y vistas materializadas con los principales agregados requeridos por los cuadros de acuerdo a las especificaciones suministradas por el equipo temático de la encuesta.

- Segunda fase: a nivel del software SAS, se utilizan los objetos de base de datos, que se construyeron en el ítem anterior, y se realiza la programación para el cálculo de variables auxiliares (porcentajes, máximos y mínimos) y la generación del formato de salida. Una vez los cuadros se generan pasan al equipo de muestras para calcular los coeficientes de variación.

#### **d. Anonimización de las bases de datos**

Dando cumplimiento al punto número 6, de los principios que rigen las actividades internacionales de estadística de la Oficina de Estadística de las Naciones Unidas: "Los datos individuales recolectados sobre personas naturales o entidades legales, o sobre pequeñas agrupaciones que están sujetas a reglas de confidencialidad nacional, deben mantenerse estrictamente confidencial y ser usadas exclusivamente para fines estadísticos o para propósitos establecidos por la legislación."<sup>18</sup> En abril de 2012 la Oficina de Sistemas del DANE implementó mecanismos para evitar que se tenga acceso directo a los microdatos, mitigando el riesgo de fuga de esta información. Para lo anterior se definen los siguientes casos de acceso a microdato:

- Asignación de permisos a funcionarios del DANE que requieren acceso directo al microdato para el cumplimiento de sus funciones al interior del departamento.
- Entidades externas que requieren acceso a microdatos anonimizados.
- Utilización de microdatos anonimizados para ambientes de pruebas de desarrollo de *software*.

Se establecen los criterios utilizados para garantizar que la información entregada al público no permita la identificación de la fuente, que para el caso de la GEIH pueden ser hogares o personas miembros del hogar, a continuación se relacionan las variables:

*Identificación personal:*

Todas las variables relacionadas con identificación son eliminadas de las bases de datos, estas variables sólo son utilizadas durante la recolección de información para acceder al hogar y contactar a los encuestados.

*Las variables de identificación son:*

Municipio, Barrio, Dirección, Teléfono, Sector, Sección, Manzana, Segmento, Estrato de diseño, Edificación, Vivienda Número.

Estas variables, que son absolutamente confidenciales, no se manejan en las bases de datos de acceso público y tampoco se encuentran dentro de las bases de datos utilizadas al interior del DANE para la generación de estadística básica.

*Datos de la Vivienda y datos del hogar*

Las variables se capturan con opciones de respuesta que no permiten la identificación particular de las viviendas, sino la caracterización por tipo de vivienda. Los datos del hogar están enmarcados en la medición de las condiciones de quienes lo habitan.

Datos de las personas

Las variables relacionadas con nombres y apellidos no son utilizadas en ninguna etapa posterior a la captura de información y no se encuentran en las bases de datos de uso público. Como parte del proceso de anonimización, se realiza una modificación a la variable Fecha de nacimiento. Esta variable se entrega con formato de mes y año, eliminando el día de nacimiento.

Fuerza de trabajo, ocupados, desocupados e inactivos

Las variables objeto de anonimización son:

- Nombre de la empresa, negocio, industria, oficina, firma o finca donde trabaja, que permite, mediante el procesamiento interno, determinar si la empresa es pública o privada.
- Otras variables que se modifican para la publicación del microdato, con propósitos estadísticos de codificación de la rama de actividad y el oficio son: ¿Qué hace en este trabajo?, ¿A qué actividad se dedica principalmente la empresa o negocio en la que realiza su trabajo? y ¿cuál es la actividad de la empresa o persona que lo contrató? Para la codificación de estas variables se emplean nomenclaturas y clasificaciones internacionales que permiten el nivel de agregación necesario y que garantizan la comparabilidad internacional de los resultados.

## **2.5. DISEÑO MÉTODOS Y MECANISMOS PARA EL CONTROL DE CALIDAD**

### **a. Instrumentos de control para supervisión**

La recolección de la información está bajo la dirección del supervisor, quien fundamentalmente controla el trabajo de campo que realizan los encuestadores y, cuyos resultados dependen en gran parte de la iniciativa para la solución de problemas y dificultades. Además, del liderazgo y la capacidad organizativa del trabajo.

Las actividades de control que debe llevar a cabo un supervisor son:

- Control de desempeño de las actividades diarias del encuestador.
- Control de calidad, con el objetivo de verificar la calidad y consistencia de la información recolectada mediante la verificación de conocimientos de conceptos del encuestador o por medio de reentrevistas.
- Control de visitas dirigido a revisar que el encuestador haya concurrido directa y personalmente a los hogares de las viviendas que debe entrevistar, y el cumplimiento de las citas concertadas para lograr la encuesta completa.
- Control de cobertura, cuyo propósito es controlar que los hogares de las viviendas encuestadas correspondan exactamente a las viviendas seleccionadas en la muestra y en caso de rechazo por parte del hogar, tratar de persuadir al informante de la importancia de la operación, recordando su confidencialidad para que suministre los datos.
- Como parte de las funciones del supervisor también se encuentra la evaluación de la entrevista, para verificar el buen manejo de los DMC por parte del encuestador y seguimiento a las instrucciones de recolección dadas para recoger la información de mercado laboral.

### **b. Indicadores para el control de calidad de los procesos de la operación**

Para asegurar la calidad de los procesos se utilizan principalmente los Indicadores de Confiabilidad (cobertura de viviendas y hogares) y de Calidad (tasa de respuesta) asegurando la calidad de la recolección y los estimadores de precisión de resultados (error de muestreo menor o igual a 5% con un nivel de confiabilidad del 95%), para los principales indicadores de fuerza laboral nacionales.

Los principales indicadores que se utilizan en el análisis operativo de los resultados son:

Cobertura de viviendas: se obtiene de cruzar la información de viviendas iniciales seleccionadas, contra viviendas finales encontradas.

Cobertura de hogares: se obtiene de cruzar la información del total de hogares con encuesta completa contra el total de hogares encontrados.

Calidad de la recolección: se obtiene de identificar el número de errores cometidos durante la recolección.

Errores e inconsistencias: se obtiene de la realización de diversas pruebas que ayudan a constatar que los cálculos estén bien hechos y que los datos sean coherentes.

<sup>18</sup>[http://unstats.un.org/unsd/methods/statorg/Principles\\_stat\\_activities/principles\\_stat\\_activities.asp](http://unstats.un.org/unsd/methods/statorg/Principles_stat_activities/principles_stat_activities.asp)

Alarmas: se obtienen al hacer una búsqueda de datos que se salgan de los estándares, como: alto número de viviendas vacantes, alto número de rechazos, bajo promedio de personas, entre otros.

Informe de cobertura campo vs Sistemas: se obtiene de cruzar la información que las sedes y subsedes envían al DANE Central en los resúmenes de cobertura sobre viviendas, hogares y personas encontradas, contra la información que es transmitida al área de sistemas.

Indicador de puntualidad: busca garantizar que la información enviada por las sedes y subsedes sea oportuna.

Tasa de respuesta: se obtiene al medir en número de encuestas efectivas respecto a las esperadas. El objetivo de este indicador es identificar el grado de eficacia en el que se encuentra la operación con respecto al proceso operativo de recolección.

Para garantizar la calidad de la información es necesario tener en cuenta las siguientes normas:

- El equipo operativo debe de ser evaluado periódicamente.
- Los recolectores deben ser rotados de supervisor y de zonas de la ciudad.
- Al finalizar cada etapa de recolección, el asistente técnico y/o coordinador de campo debe enviar al DANE Central un informe de la justificación de la variación de las tasas de mercado laboral según formato establecido.
- Las cargas de trabajo deben ser evacuadas en su totalidad en la semana respectiva, y por ningún motivo dejar trabajo de recolección para la semana siguiente.
- El personal de campo debe tomar un día de descanso a la semana. Este día se seleccionará teniendo como prioridad no afectar la calidad y cobertura de la información recolectada durante la semana, en este sentido es importante que el número de encuestas pendientes sea el mínimo posible. Es importante tener presente que durante el día seleccionado para descanso el grupo de trabajo debe reunirse para tratar aspectos propios del operativo como:
  - Dar a conocer las fortalezas y debilidades observadas durante el acompañamiento en campo, tanto por parte de los supervisores como del coordinador de campo.
  - Solucionar dudas e inquietudes operativas.
- Presentar los cambios: logísticos, temáticos, cartográficos, sistemas y muestrales para que sean entendidos y asimilados por el grupo de campo.

Además de lo anterior se cuenta con el Manual de diligenciamiento de formatos, que tiene por objetivo registrar y mantener un control en la calidad y cobertura del operativo de campo de la GEIH a través de varios formatos que han sido diseñados para este fin:

- Listado de Recuento de Unidades Estadísticas
- Listado de Unidades Seccionadas
- Resumen Acumulado de Viviendas, Hogares y Personas por Segmentos
- Informe Diario del Supervisor de Campo
- Reentrevista
- Informe técnico del coordinador de campo
- Segmentos no trabajados
- Novedades cartográficas
- Control de la muestra segmentos con submuestreo
- Segmentos agotados
- Asignación y control diario de áreas de trabajo
- Indicadores de calidad
- Resumen operativo de cobertura
- Informe de contexto
- Justificación en la variaciones de las tasas de desempleo, tasas de ocupación y tasas globales de participación
- Informe justificaciones comparativo Sistemas VS Campo
- Hoja de croquis de segmento

## 2.6. DISEÑO DE PRUEBAS PILOTO

No aplica

## 2.7. DISEÑO DEL ANÁLISIS DE RESULTADOS

### 2.7.1. Análisis estadístico

Este procedimiento está orientado al análisis tanto de la consistencia interna de los datos como del comportamiento del mercado laboral y su relación con otras variables macroeconómicas que sirvan de base para la formulación, seguimiento y evaluación de políticas.

Para realizar el análisis de contexto se llevan a cabo diferentes estudios como son:

- El análisis descriptivo ayuda a observar el comportamiento de la muestra en estudio, a través de tablas, gráficos, intervalos de confianza y estadísticas de tendencia y dispersión.
- Se analiza en la muestra la estructura de los indicadores por dominios de estudio a partir de la distribución de frecuencias, y se detectan posibles inconsistencias y valores atípicos. Se verifica la cobertura a nivel de segmento, viviendas, hogares y persona, y se observa la distribución de la pérdida de muestra para realizar los respectivos ajustes de la no respuesta.
- El análisis inferencial tiene como objetivo realizar la estimación de la muestra a la población objetivo. En este proceso se aplican los factores de expansión y se revisa la inferencia a la población objetivo establecido para los correspondientes dominios de estudio.
- Se verifica que los ajustes del factor de expansión no generen sesgos en las estimaciones, y que sus errores muestrales sean aceptables de acuerdo con los parámetros establecidos por dominios de estudio.
- El grupo temático de la GEIH revisa la consistencia de la información sobre los principales Indicadores de Mercado Laboral (IML). Posteriormente elabora los productos de publicación como son el boletín técnico y las presentaciones de resultados. En este proceso se realizan análisis comparativos entre periodos de los diferentes años y se analiza la consistencia interna de la información por dominios geográficos y demás variables según conceptos de fuerza de trabajo.

### 2.7.2. Análisis del contexto

Los informes de contexto y justificación de la variación de las tasas, son elaborados mensualmente por cada una de las sedes y subsedes. En estos informes se documenta toda información que permita dar un primer acercamiento de la dinámica económica de cada ciudad y que pueda influir en el comportamiento de los principales indicadores de mercado laboral. Se puede utilizar información como noticias, prensa, observación directa de los grupos de recolección, entre otros.

Los informes son elaborados y enviados por las sedes y subsedes al finalizar cada etapa de recolección y son revisados y consolidados por el área de logística de DANE central antes de ser remitidos al área temática.

### 2.7.3. Comité de Personas expertas

La coordinación temática de la Gran Encuesta Integrada de Hogares, convoca mensualmente a un Comité Técnico Interno y uno Externo, con el fin de presentar los resultados obtenidos y analizar el comportamiento de los principales indicadores de mercado laboral.

## 2.8. DISEÑO DE LA DIFUSIÓN

### 2.8.1. Administración del repositorio de datos

El DANE implementó el Archivo Nacional de Datos – ANDA, que es un catálogo en el que los usuarios pueden explorar, buscar, comparar, solicitar acceso y descargar información relacionada con censos, encuestas por muestreo y uso estadístico de registros administrativos.<sup>19</sup>

Para la GEIH el ANDA contiene metadatos (manuales, formularios y metodología, entre otros) y microdatos anonimizados de acceso público<sup>20</sup>.

### 2.8.2. Productos e instrumentos de difusión

Basándose en los resultados obtenidos se elabora el boletín técnico de los principales indicadores del mercado laboral, con una periodicidad mensual. Adicionalmente se elaboran boletines de mercado laboral con temáticas especiales, los cuales se publican a manera de resumen ejecutivo en los trimestres móviles y a manera de boletín técnico en los trimestres fijos.

Publicación mensual con el siguiente contenido:

- Indicadores del mercado laboral mensual para el Total nacional y total 13 ciudades y áreas metropolitanas.
- Indicadores de mercado laboral trimestral de los jefes de hogar para el total nacional, cabeceras, centros poblados y rural disperso y total 13 ciudades y áreas metropolitanas.
- Indicadores del mercado laboral trimestre acumulado para el total nacional, total 13 ciudades y áreas metropolitanas, cabeceras, centros poblados y rural disperso y 24 ciudades y áreas metropolitanas.
- Comportamiento del mercado laboral para el trimestre acumulado para el total nacional, total 13 ciudades y áreas metropolitanas, cabeceras y centros poblados y rural disperso.

Los resultados mensuales son presentados durante una rueda de prensa; además los usuarios pueden consultar, a través de la página web del DANE, los anexos estadísticos, los boletines técnicos y las presentaciones que contienen los indicadores de mercado laboral de los diferentes dominios geográficos que cubre la operación.

## 2.9. DISEÑO DE LA EVALUACIÓN

Cumplir con los estándares de calidad es fundamental en el proceso de producción de las estadísticas, en este sentido el diseño de la evaluación en la operación de GEIH se constituye en un mecanismo importante para dar un tratamiento adecuado a esta información.

La evaluación realizada a la operación es un proceso que se realiza de manera continua, en varias etapas desde el diseño hasta el proceso de producción y difusión de la misma. Incluye entre otros:

- Auditorías del Instituto Colombiano de Normas Técnicas y Certificación (ICONTEC) y Contraloría.
- Auditorías por parte de la oficina de control interno del DANE.
- Encuesta de satisfacción al cliente: es una herramienta para conocer las necesidades de los clientes y se trabaja en conjunto con el Banco de Datos de la entidad, que es el primer canal de contacto con los usuarios.

## 2.10. DOCUMENTACIÓN RELACIONADA

En este capítulo se encuentran referenciados los documentos relacionados con el diseño temático que incluyen:

Diseño muestral: Donde se explican las premisas del plan de muestreo, se describen los elementos del diseño propiamente tal, con los detalles de los componentes de la muestra, y los métodos y procedimientos utilizados, y se presentan la definición y principios básicos de la técnica de selección controlada.

Especificaciones de consistencia y validación: empleadas como guía para evaluar la información reportada en los instrumentos de recolección.

Manuales de recolección, operativo, de usuario, coordinador, supervisor y recuento que orientan el trabajo de campo realizado en las territoriales.

<sup>19</sup> Fuente: Archivo Nacional de datos. [http://formularios.dane.gov.co/Anda\\_4\\_1/index.php/home](http://formularios.dane.gov.co/Anda_4_1/index.php/home)

<sup>20</sup> Los microdatos de la GEIH se publican con un rezago de dos meses con relación a la información que publica mensualmente la operación.

### 3. GLOSARIO

**Área metropolitana** Entidad administrativa formada por un conjunto de dos o más municipios integrados alrededor de un municipio núcleo o metrópoli. Estos están vinculados por relaciones de orden físico, económico y social, y requieren una administración coordinada para la programación, la coordinación de su desarrollo y la prestación de sus servicios públicos. (Artículo 1 ley 128 de 1994)

**Asalariados.** Trabajadores que poseen un contrato de trabajo verbal o escrito, por el cual perciben una remuneración de manera estable o regular (DANE, 2016)

**Desocupados (D).** Son las personas que en la semana de referencia se encontraban en las siguientes condiciones (OIT, 1982):

- "sin empleo", es decir, que no tenían un empleo asalariado o un trabajo independiente ni se desempeñaron como trabajador familiar sin remuneración.
- "en busca de empleo", es decir que habían tomado medidas concretas para buscar un empleo asalariado o independiente en las últimas 4 semanas y estaban disponibles para empezar a trabajar.
- Disponibilidad para empezar a trabajar.

De acuerdo con la experiencia laboral, la población desocupada se clasifica en:

- Cesantes. Las personas que habiendo trabajado antes por lo menos dos semanas consecutivas, se encuentran desocupadas.
- Aspirantes. Las personas que están buscando trabajo por primera vez

**Desempleo abierto.** Incluye a todas aquellas personas pertenecientes a la Población Económicamente Activa, que en la semana de referencia no tenían un empleo asalariado o independiente, estaban disponibles para empezar a trabajar y estuvieron en busca de empleo. En el caso colombiano, la búsqueda de empleo corresponde a las últimas cuatro semanas (OIT, 1982).

**Desempleo oculto.** Incluye a personas pertenecientes a la Población Económicamente Activa, que en la semana de referencia no tenían un empleo asalariado o independiente, estaban disponibles para empezar a trabajar y no hicieron diligencias en las últimas cuatro semanas, pero sí en los últimos doce meses y tenían una razón válida de desaliento (OIT, 1982).

**Empleo asalariado.** Empleo en instituciones públicas, privadas, corporaciones sin ánimo de lucro u hogar, en el que el trabajador percibe una remuneración en forma de salario, sueldo, comisión, honorarios, propinas, a destajo o en especie, que no depende de los ingresos de la unidad para la que trabaja, utilizando instrumentos, bienes de capital, sistemas de información e infraestructuras de propiedad de las empresas o de terceros para desarrollar sus labores bajo la supervisión del propietario o de personas contratadas por este (Elaboración propia).

**Ocupación.** Es el oficio o profesión (cuando se desempeña en ésta) de una persona, independiente del sector en que puede estar empleada, o del tipo de estudio que hubiese recibido. Generalmente se define en términos de la combinación de trabajo, tareas y funciones desempeñadas. (DANE)

**Ocupados (OC).** Personas pertenecientes a la PEA, que en la semana de referencia trabajaron por lo menos una hora remunerada o no remunerada (para el caso de los trabajadores familiares sin remuneración) en dinero o en especie o no trabajaron pero tenían un trabajo (Eurostat, CODED).

**Población Económicamente Activa (PEA).** También llamada fuerza de trabajo y comprende a las personas en edad de trabajar que en la semana de referencia realizaron (ocupados) o buscaron ejercer (desocupados) una actividad económica (OIT, 1982).

**Población Económicamente Inactiva (PEI).** Personas en edad de trabajar que en la semana de referencia no participaron en la producción de bienes y servicios porque no necesitan, no pueden o no están interesadas en tener actividad remunerada. A este grupo pertenecen estudiantes, amas de casa, pensionados, jubilados, rentistas, inválidos (incapacitados permanentemente para trabajar), personas a las que no les llama la atención o creen que no vale la pena trabajar (Eurostat, 1999).

**Población en Edad de Trabajar (PET).** Constituida por todas las personas de 12 años en adelante para las zonas urbanas y de 10 años en adelante en las zonas rurales. Se divide en población económicamente activa y económicamente inactiva. (DANE)

**Población Objetivo (PO).** Corresponde a la población civil no institucional residente en todo el territorio nacional. Se excluyen de los nuevos departamentos las cabeceras municipales (no capitales de departamento) centros poblados y rural disperso, en los cuales reside aproximadamente el 1% de la población del país, y se encuentra dispersa en casi la mitad de la superficie del país. (DANE)

**Población Total (PT).** Constituida por la población civil no institucional residente en hogares particulares. Esta población se estima con base en los censos de población, las estadísticas vitales y de migración. (DANE)

**Porcentaje de la Población en Edad de Trabajar (%PET).** Mide la relación porcentual entre el número de personas que componen la población en edad de trabajar (PET), frente a la población total (PT) y cuya fórmula es la siguiente: % PET = (PET/PT) \* 100.

**Subempleo Objetivo (SO).** Se refiere a los trabajadores que tienen el deseo manifestado de mejorar sus ingresos, el número de horas trabajadas o tener una labor más propia de sus personales competencias, pero además han hecho una gestión para materializar su aspiración, y están en disposición de efectuar el cambio. (DANE)

**Subempleo por Insuficiencia de Horas (SIH).** Se refiere a aquellas personas ocupadas que en la semana de referencia deseaban trabajar más horas en sus empleos actuales o en otros empleos, estaban disponibles para hacerlo y habían trabajado menos horas de las legalmente establecidas. Para el caso colombiano, las legales son 48 horas semanales. (OIT, 1982)

**Subempleo por Situaciones de Empleo Inadecuado (SEI):** Comprende a todas las personas que en la semana de referencia se encontraban desempeñando ocupaciones que limitaban sus capacidades y bienestar, estaban disponibles y deseaban cambiar de labor con el fin de mejorar sus ingresos o desempeñar trabajos en los que pudieran sacar mejor provecho a sus capacidades. (OIT, 1982)

**Subempleo Subjetivo (SS).** Se refiere al simple deseo manifestado por el trabajador de mejorar sus ingresos, el número de horas trabajadas o tener una labor más propia de sus personales competencias.

**Tasa de Subempleo (TS).** Relación porcentual de la población subempleada (PS) y la (PEA), y cuya fórmula es la siguiente: TS = (PS / PEA) \* 100.

**Tasa Bruta de Participación (TBP).** Mide la relación porcentual entre el número de personas que componen oferta de trabajo, frente al número de personas que integran la población total, cuya fórmula es la siguiente: TBP = (PEA / PT) \* 100.

**Tasa de Desempleo (TD).** Relación porcentual entre el número de personas que están buscando trabajo (DS) y el número de personas que integran la fuerza laboral (PEA), cuya fórmula es la siguiente: TD = (DS / PEA) \* 100.

**Tasa de Ocupación (TO).** Relación porcentual entre la población ocupada (OC) y el número de personas que integran la población en edad de trabajar (PET), cuya fórmula es la siguiente: TO = (OC/PET) \* 100.

**Tasa Global de Participación (TGP).** Relación porcentual entre la población económicamente activa (PEA) y la población en edad de trabajar (PET). Este indicador refleja la presión de la población en edad de trabajar sobre el mercado laboral y cuya fórmula es la siguiente: TGP = (PEA/PET) \* 100.

**Trabajador Familiar Sin Remuneración (TFR).** Trabajador con un empleo independiente sin remuneración, que labora en una empresa cuyo propietario es un familiar o pariente que generalmente vive en el mismo hogar (Eurostat, 1999).

**Trabajador por Cuenta Propia.** Persona natural con un trabajo, profesión u oficio, cuya remuneración depende directamente de la comercialización y venta de los bienes o servicios producidos; se caracteriza por no contratar a terceros como asalariados para lograr sus metas y objetivos (OIT, 1993).

### 4. BIBLIOGRAFÍA

DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADÍSTICA, LA MUESTRA MAESTRA DE LAS ENCUESTAS A HOGARES; 2011.

DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADÍSTICA, DANE <http://www.dane.gov.co/index.php/mercado-laboral/empleo-y-desempleo>. Antecedentes de la Investigación.

Eurostat, Conceptos y definiciones de la base de datos -CODED-.

Eurostat, 1999. Guía y programación del programa comunitario de la población y censos de hogares en 2001. Vol. 1: Guidelines, Luxembourg.

Hussmanns, Ralf, Farhad, Mehran y Vijay, Verona. Encuesta de población económicamente activa, empleo, desempleo y subempleo. Manual de la OIT sobre conceptos y métodos. Madrid, España: Centro de Publicaciones del Ministerio de Trabajo de España, 1993.

Organización Internacional del Trabajo (OIT). Base de datos sobre estadísticas del trabajo elaborada por el Departamento de Estadística de la Oficina Internacional del Trabajo (<http://laborsta.ilo.org>).

Organización Internacional del Trabajo (OIT). Resolución concerniente a la población económicamente activa, ocupados, desocupados y subempleados adoptados de la 13 conferencia Internacional de Estadísticas Laborales, Octubre 1982.

Organización Internacional del Trabajo (OIT). Resolución concerniente a la clasificación en el empleo adoptado por la Conferencia de Estadísticas Laborales 15. Enero 1993.

## 5. ANEXOS

• DSO-GEIH-MET-01 Anexos

VERSIÓN	FECHA	RAZÓN DE LA ACTUALIZACIÓN
1	15/Ene/2002	Creación del documento
2	30/Jun/2006	Actualización documental
3	01/Jul/2009	Actualización documental
5	16/Nov/2011	Actualización documental
6	16/Ago/2012	Actualización documental
7	31/Oct/2012	Actualización documental
8	22/Mar/2013	Actualización documental
9	08/Abr/2016	"Se cambia la plantilla de la metodología general de la Gran Encuesta Integrada de Hogares. Se modifica, de la metodología general de la GEIH, la prueba de campo en la selección de personal, de las actividades preparatorias. 2.3 Diseño de la ejecución; 2.3.2. Actividades preparatorias, a. Sensibilización y b. Selección y entrenamiento del personal. "
10	31/Ene/2020	Se realizó actualización en las secciones de antecedentes, alcance y actividades preparatorias.

ELABORÓ	REVISÓ	APROBÓ
<b>Nombre:</b> Diego Alejandro Martínez Jiménez <b>Cargo:</b> CONTRATISTA <b>Fecha:</b> 20/Feb/2020	<b>Nombre:</b> Geovanni Alexander Portilla Correa <b>Cargo:</b> Profesional Especializado <b>Fecha:</b> 20/Feb/2020  <b>Nombre:</b> Juan Fernando Plazas <b>Cargo:</b> Profesional Especializado <b>Fecha:</b> 20/Feb/2020	<b>Nombre:</b> Andres Mauricio Clavijo Abril <b>Cargo:</b> Director Técnico <b>Fecha:</b> 21/Feb/2020

Si este documento es impreso se considera copia no controlada