

# Colombia - Gran Encuesta Integrada de Hogares - GEIH - 2026

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

report\_generated\_on: May 12, 2026

visit\_data\_catalog\_at: <https://microdatos.dane.gov.co/index.php>

# Información general

## Identificación

NÚMERO DE ID  
DANE-DIMPE-GEIH-2026

## Información general

### RESUMEN

El Departamento Administrativo Nacional de Estadística (DANE) ha desarrollado e implementado encuestas de hogares, desde finales de la década de los años 1960, que consideran las temáticas de fuerza de trabajo, calidad de vida, e ingresos y gastos. Con el propósito de generar información que permitiera caracterizar el mercado laboral del país, inicialmente se implementó la Encuesta Nacional de Hogares (ENH), la cual luego de un proceso de mejora se convirtió en la Encuesta Continua de Hogares (ECH) y posteriormente en la Gran Encuesta Integrada de Hogares (GEIH), que a partir de 2006 inició la recolección de su módulo central de mercado laboral e ingresos de los hogares.

A partir de 2019, se inició el proceso de rediseño de la GEIH, tomando como base las recomendaciones de la Organización Internacional del Trabajo (OIT), con el propósito de adaptar la encuesta a los nuevos requerimientos que demanda el país en cuanto al fortalecimiento metodológico de la medición del mercado laboral y la pobreza monetaria; así como de avanzar en la visibilización estadística de sectores prioritarios en la política pública del país.

Entre enero de 2020 y diciembre de 2021 se implementó el Sistema General de Pruebas del proyecto de rediseño de la Gran Encuesta Integrada de Hogares, para garantizar el tránsito adecuado entre la GEIH tradicional (GEIH marco 2005) y su versión actualizada con el marco muestral del Censo Nacional de Población y Vivienda de 2018 (GEIH marco 2018). Luego de culminado el proceso de rediseño definido para la operación estadística, en el mes de enero de 2022, se inició la recolección de información de la GEIH con el total de mejoras incorporadas.

De esta forma, la GEIH proporciona información estadística sobre el tamaño y estructura de la fuerza de trabajo (empleo, desempleo y población fuera de la fuerza de trabajo), los ingresos laborales y no laborales de los hogares, la pobreza monetaria y la pobreza monetaria extrema de la población residente en el país. Las temáticas por las cuales se indagan permiten caracterizar a la población según sexo, edad, parentesco con el jefe del hogar, nivel educativo, afiliación al sistema de seguridad social en salud, grupos poblacionales como etnias, campesinos, LGBT o con algún tipo de discapacidad, otras formas de trabajo como producción de bienes y servicios para autoconsumo, trabajo en formación y voluntariado, entre otras.

Esta operación estadística se ha especializado en la medición de la estructura del mercado laboral y los ingresos de los hogares. Actualmente, la GEIH cuenta con una muestra anual aproximada de 315.000 hogares a nivel nacional, lo que hace que sea la encuesta de mayor cobertura a nivel nacional. De modo que, la GEIH permite obtener indicadores confiables y series continuas para analizar la fuerza de trabajo del país y los principales indicadores del mercado laboral, considerados como información fundamental para la toma de decisiones de política pública.

Los principales indicadores del mercado laboral son: Tasa Global de Participación (TGP), Tasa de Ocupación (TO), Tasa de Desempleo (TD), Tasa de Subocupación (TS), características del empleo (temporalidad, subocupación, rama de actividad, ocupación u oficio, posición ocupacional, ingreso, afiliación a la seguridad social, etc.), características del desempleo (tiempo de búsqueda de empleo, rama de actividad, ocupación u oficio que desempeñó en su empleo anterior), características de la población fuera de la fuerza de trabajo. Por otra parte, a través de esta operación estadística es posible conocer los ingresos de los hogares, tanto en dinero como en especie, que sirven de insumo para las mediciones sobre pobreza monetaria.

Este documento se divide en tres secciones: la primera (1) Antecedentes, en la cual se presenta la evolución histórica de la medición del mercado laboral en Colombia y las principales características técnicas de la encuesta rediseñada (GEIH marco 2018).

La segunda (2) contempla el diseño de la operación estadística y su marco de referencia. Esta sección se divide en 8 subsecciones donde se aborda el diseño temático de la encuesta, respaldado por las 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. De igual forma, 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 esta en términos de estimación y análisis de resultados.

Adicionalmente, se presenta cada proceso de la producción estadística: desde las actividades preparatorias al operativo de campo, la recolección, transmisión y procesamiento de datos, y los métodos de control de calidad, para finalizar con los procesos de análisis y difusión de los datos y el glosario de términos de mercado laboral, como complemento para la comprensión del documento.

Finalmente, la tercera (3) sección indica la documentación relacionada, como las especificaciones de validación y consistencia del aplicativo de recolección y el formulario.

#### TIPO DE DATO

Encuesta por muestreo (ssd)

#### UNIDAD DE ANÁLISIS

UNIDAD DE OBSERVACIÓN: corresponden a las viviendas, los hogares y las personas.

UNIDAD DE ANÁLISIS: también están representadas por las viviendas, los hogares y las personas.

UNIDAD DE MUESTREO: existen varias unidades de muestreo dependiendo de la etapa del diseño muestral. Las unidades primarias se relacionan con la primera etapa, las unidades secundarias con la segunda etapa y así sucesivamente.

Para la encuesta, en particular, se definen 2 etapas con las siguientes unidades de muestreo:

UNIDADES PRIMARIAS DE MUESTREO (UPM): Son todos los municipios del país.

UNIDADES SECUNDARIAS DE MUESTREO (USM): Son conglomerados de 10 viviendas contiguas en promedio, también llamados segmentos o medidas de tamaño (MT), ubicados tanto en la cabecera como en el resto de cada municipio, con límites fácilmente identificables en los que se encuestan todas las viviendas, hogares y personas.

## Ámbito

---

#### NOTAS

##### NECESIDADES DE INFORMACIÓN

Para la identificación de las necesidades de información de la encuesta, se tuvieron en cuenta los requerimientos y solicitudes tanto internacionales como nacionales sobre mercado laboral.

##### REQUERIMIENTOS INTERNACIONALES

Actualización de estándares internacionales: con relación a los requerimientos de carácter internacional, es importante aclarar que la operación estadística toma como referente conceptual las recomendaciones de la Organización Internacional del Trabajo (OIT), toda vez que es la entidad especializada de las Naciones Unidas (UN) que se ocupa de los asuntos relativos al trabajo y las relaciones laborales. Particularmente se tienen en cuenta los lineamientos dados por las Conferencias Internacionales de Estadísticos del Trabajo (CIET), con las cuales se busca estandarizar la medición estadística del mercado laboral en todos los países.

En ese sentido y como parte de las mejoras conceptuales incluidas en la GEIH, se tuvieron en cuenta dentro del rediseño de la operación estadística las recomendaciones de la 17ª (2003), 19ª (2013) y 20ª (2018) CIET.

##### REQUERIMIENTOS NACIONALES

Necesidades de política pública: Una de las motivaciones para el rediseño de la GEIH fueron las necesidades a mediano plazo de la política pública en Colombia en línea con lo planteado en el Plan Nacional de Desarrollo (PND), Ley 1995 de 2019 "Por la cual se expide el plan nacional de desarrollo 2018-2022 pacto por Colombia, pacto por la equidad".

Bajo el contexto anterior, el DANE trabajó con el Departamento Nacional de Planeación (DNP) en la recolección de necesidades de política pública con base en la GEIH. Teniendo en cuenta la magnitud de las necesidades de información de la política pública nacional, referentes a la actualización de las estadísticas de trabajo y la medición de ingresos conforme a la normatividad vigente y las buenas prácticas internacionales, fue necesario delimitar el alcance de los ajustes que podían incluirse en la encuesta sin perjudicar la recolección de información que sirve de base para la generación de los principales indicadores de mercado laboral. Para esto se establecieron cuatro criterios:

- Identificar cambios relevantes al formulario de recolección que permitieran mejorar la estandarización estadística de cifras de mercado laboral, en línea con las recomendaciones de las 17ª, 19ª y 20ª CIET de la OIT, con el objetivo de hacer comparables los resultados con los de otros países.
- Los ajustes en las preguntas del formulario no debían comprometer la información recolectada que permite medir los principales indicadores de mercado laboral.
- Contemplar las necesidades de información de los usuarios externos (ministerios, Banco de la República, universidades, entre otros), así como lo dispuesto en el PND 2018-2022, sobre superar la invisibilidad estadística de grupos minoritarios.
- No afectar el alcance ni la naturaleza de la operación estadística.

#### GRUPOS POBLACIONALES PRIORIZADOS

Otro aspecto que se tuvo en cuenta para el rediseño de la GEIH en materia de fortalecimiento de la política pública, estuvo relacionado con la superación de la invisibilidad estadística de grupos poblacionales priorizados como la población campesina, las personas en condición de discapacidad, la población LGBTI y la población con autorreconocimiento étnico.

En la identificación de esta necesidad se tomó como referencia lo expuesto en el artículo 13 5 de la Constitución de 1991 y lo dispuesto en el PND 2018-2022.

#### Sondeo de necesidades de información a usuarios internos y externos de la GEIH

Este apartado se enmarca en la fase de identificación de necesidades de los sectores económicos y sociales que hacen uso de la GEIH. La metodología utilizada por el equipo de rediseño consistió en elaborar formatos con preguntas dirigidas a los usuarios sobre los usos de capítulos, preguntas e indicadores por parte de la academia, los gremios y los particulares.

En desarrollo de las sesiones de trabajo del equipo responsable del rediseño, se construyó un plan de acción de tres meses para identificar las necesidades internas y externas a través del diseño del cuestionario con preguntas específicas de los capítulos que hacían parte de la GEIH. Este cuestionario fue enviado a entidades públicas, sector privado y organizaciones de interés, como ministerios, gremios y centros de pensamiento e investigación. De la misma manera, se diseñó una versión para ser diligenciada internamente por las direcciones territoriales y las direcciones técnicas del DANE; como complemento de la información recolectada mediante los cuestionarios dispuestos para la identificación de necesidades, se llevaron a cabo 13 mesas de trabajo con expertos en mercado laboral y pobreza, así como en temas de educación, diversidad de género y discapacidad.

El objetivo del cuestionario externo fue recopilar información que permitiera conocer la relevancia, pertinencia y uso que le daban los usuarios a la GEIH y a sus diferentes módulos, así como identificar nuevos requerimientos y expectativas en el marco del rediseño de la operación estadística. El sondeo realizado, pretendió recoger las respuestas que reflejaran la visión general de todos los usuarios de la entidad.

Al interior del DANE se desarrollaron dos tipos de sondeos. El primero se realizó con las direcciones territoriales que apoyan el trabajo del DANE en cada una de las sedes y subsedes responsables del proceso de producción estadística; y el segundo, se realizó con las direcciones técnicas responsables de los diferentes procesos que intervienen en la producción estadística en el DANE Central, quienes respondieron las preguntas sobre los usos y nuevas propuestas para la GEIH en el marco del rediseño.

El objetivo del sondeo realizado a las direcciones territoriales (incluidas sedes y subsedes) fue obtener información que permitiera conocer la relevancia, pertinencia, problemáticas, solicitudes, nuevos requerimientos y expectativas en el marco del rediseño de la GEIH. El cuestionario estuvo dirigido a los equipos de trabajo que intervenían en la fase operativa de la encuesta, con el propósito de recoger la mayor cantidad de observaciones del personal involucrado en la cadena de producción y de tener una mirada integral de las necesidades de la operación estadística a nivel operativo y logístico.

El sondeo realizado con la Dirección de Síntesis y Cuentas Nacionales, la Dirección de Metodología y Producción Estadística y la Dirección de Censos y Demografía permitió identificar el uso que se daba a las preguntas del formulario de la GEIH y obtener sugerencias para la inclusión de nuevas preguntas.

---

5 Artículo 13: "Todas las personas nacen libres e iguales ante la ley, recibirán la misma protección y trato de las autoridades

y gozarán de los mismos derechos, libertades y oportunidades sin ninguna discriminación por razones de sexo, raza, origen nacional o familiar, lengua, religión, opinión política o filosófica. El Estado promoverá las condiciones para que la igualdad sea real y efectiva y adoptará medidas en favor de grupos discriminados o marginados. El Estado protegerá especialmente a aquellas personas que por su condición económica, física o mental, se encuentren en circunstancia de debilidad manifiesta y sancionará los abusos o maltratos que contra ellas se cometan" Constitución Política de Colombia (1991).

## TOPICS

Topic	Vocabulary	URI
Ingresos, propiedad y ahorro/inversión	CESSDA	<a href="http://www.nesstar.org/rdf/common">http://www.nesstar.org/rdf/common</a>
Empleo	CESSDA	<a href="http://www.nesstar.org/rdf/common">http://www.nesstar.org/rdf/common</a>
Desempleo	CESSDA	<a href="http://www.nesstar.org/rdf/common">http://www.nesstar.org/rdf/common</a>
Situación económica e indicadores	CESSDA	<a href="http://www.nesstar.org/rdf/common">http://www.nesstar.org/rdf/common</a>
Economía rural	CESSDA	<a href="http://www.nesstar.org/rdf/common">http://www.nesstar.org/rdf/common</a>
Formación continua [6.4]	CESSDA	<a href="http://www.nesstar.org/rdf/common">http://www.nesstar.org/rdf/common</a>
Vivienda [10.1]	CESSDA	<a href="http://www.nesstar.org/rdf/common">http://www.nesstar.org/rdf/common</a>
Fertilidad [14.2]	CESSDA	<a href="http://www.nesstar.org/rdf/common">http://www.nesstar.org/rdf/common</a>
Disponibilidad y utilización de servicios sociales específicos [15.3]	CESSDA	<a href="http://www.nesstar.org/rdf/common">http://www.nesstar.org/rdf/common</a>

## KEYWORDS

Asalariado, Edificación, Hogar, Jefe o Jefa de hogar, Muestra, Patrón o empleador(a), Población objetivo, Trabajador por cuenta propia

## Cobertura

### COBERTURA GEOGRÁFICA

La Gran Encuesta Integrada de Hogares tiene una cobertura nacional. Se excluyen de la cobertura de los departamentos de la Amazonia y la Orinoquia, las cabeceras municipales que no son capitales de departamento, así como los centros poblados y rural disperso. También se excluye la población de Providencia y el centro poblado y rural disperso de San Andrés.

### GEOGRAPHIC UNIT

La desagregación geográfica de la encuesta está enmarcada sobre varios tipos de coberturas geográficas, que se definen de la siguiente manera:

**Cabecera:** delimitación geográfica definida por el DANE para fines estadísticos, alusiva al área geográfica delimitada por el perímetro censal. En su interior, se localiza la sede administrativa del municipio, es decir, la alcaldía.

**Centro poblado:** concepto construido por el DANE con fines estadísticos, para la identificación y localización geográfica de núcleos o asentamientos de población. Se define como una concentración mínima de veinte viviendas contiguas, vecinas o adosadas entre sí, ubicada en el área resto municipal o en un área no municipalizada (corregimiento departamental). Contempla los núcleos de población de los corregimientos municipales, inspecciones de policía y caseríos. Dicha concentración presenta características como la delimitación de vías vehiculares y peatonales.

**Caserío:** definición construida por el DANE para fines estadísticos, que corresponde a un centro poblado ubicado generalmente 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:** definición construida por el DANE, para fines estadísticos, la define como conglomerado de viviendas ubicadas en el área resto municipal que tiene por autoridad principal al inspector de policía.

**Corregimiento municipal:** división del área rural del municipio, la cual incluye un núcleo de población, considerada en los Planes de Ordenamiento Territorial (POT). El artículo 117 de la Ley 136 de 1994 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: delimitación geográfica definida por el DANE para fines estadísticos, comprendida entre el perímetro censal de las cabeceras municipales y de los centros poblados, y el límite municipal. Se caracteriza por la disposición dispersa de viviendas y de explotaciones agropecuarias existentes en ella. Corresponde al territorio que no forma parte ni de la cabecera municipal (clase 1) ni de los centros poblados (clase 2). Se caracteriza por objetos y elementos relacionados con la agricultura, predios de descanso o recreo, usos mineros o extractivos.

Resto: Corresponde al área geográfica cubierta por el centro poblado y rural disperso.

La GEIH tiene cobertura nacional y se han establecido diferentes niveles de desagregación temporal y geográfica.

A continuación, se presentan las cabeceras de las ciudades, las siete ciudades con las áreas metropolitanas establecidas en el desarrollo de la operación estadística y los departamentos .

- Cabeceras de ciudades: Bogotá, Medellín, Cali, Barranquilla, Bucaramanga, Manizales, Pasto, Pereira, Ibagué, Cúcuta, Villavicencio, Montería, Cartagena, Tunja, Florencia, Popayán, Valledupar, Quibdó, Neiva, Riohacha, Santa Marta, Armenia y Sincelejo.

- Las siete ciudades con los municipios que conforman el área metropolitana: Medellín - Valle de Aburrá comprendido por Caldas, La Estrella, Sabaneta, Itagüí, Envigado, Bello, Copacabana, Girardota y

Barbosa; Cali - Yumbo; Bucaramanga - Girón, Piedecuesta y Floridablanca; Manizales - Villa María; Pereira - Dosquebradas y La Virginia; Cúcuta - Villa del Rosario, Los Patios, Puerto Santander y El Zulia; y

Barranquilla - Soledad.

- Departamentos: Bogotá, Cundinamarca, Antioquia, Valle del Cauca, Atlántico, Santander, Caldas, Nariño, Risaralda, Tolima, Norte de Santander, Meta, Córdoba, Bolívar, Boyacá, Caquetá, Cauca, Cesar, Chocó, Huila, La Guajira, Magdalena, Quindío, Sucre, Vaupés, Arauca, Guainía, Vichada, Amazonas, Putumayo, Casanare, Guaviare.

- Cabeceras de las capitales de los departamentos de la Amazonia y Orinoquia: Mitú, Arauca, Inírida, Puerto Carreño, Leticia, Mocoa, Yopal y San José del Guaviare.

- Ciudades intermedias: Soacha, Buenaventura, Tumaco, Barrancabermeja y Rionegro.

#### UNIVERSO

El universo para la Gran Encuesta Integrada de Hogares está conformado por la población civil no institucional residente en todo el territorio nacional. Se excluyen, del universo de estudio; las cabeceras municipales de los departamentos de la Amazonia y la Orinoquia que no son capitales; para los centros poblados y rural disperso, se excluyen los departamentos de la Amazonia y la Orinoquia. También, se excluye la población de Providencia y el centro poblado y rural disperso de San Andrés.

## Productores y Patrocinadores

#### INVESTIGADOR(ES) PRIMARIO(S)

Nombre	Affiliation
Dirección de Metodología y Producción Estadística - DIMPE	

#### OTROS PRODUCTORES

Nombre	Affiliation	Role
Dirección de Metodología y Producción Estadística - Gran Encuesta Integrada de Hogares	Departamento Administrativo Nacional de Estadística - DANE	Equipo Técnico

#### FINANCIAMIENTO

Nombre	Abbreviation	Role
Departamento Administrativo Nacional de Estadística	DANE	Ejecutor

## Producción de metadato

### METADATO PRODUCIDO POR

Nombre	Abbreviation	Affiliation	Role
Ivan Rolando Castillo Prieto	ircastillo@dane.gov.co	Dirección de Regulación, Planeación, Estandarización y Normalización - DIRPEN	Coordinador Regulación
Arles Daniel Mayor Pachon	admayorp@dane.gov.co	Dirección de Metodología y Producción Estadística - DIMPE	Temático GEIH
Carlos Augusto Muñoz Alfonso	camunoza@dane.gov.co	Dirección de Regulación, Planeación, Estandarización y Normalización - DIRPEN	Actualizador DIRPEN
Katherine Andrea Rojas Acero	karojasa@dane.gov.co	Dirección de Difusión y Cultura Estadística (DICE)	Verificador DICE

### FECHA DE PRODUCCIÓN DEL METADATO

2026-03-12

### ID DEL DOCUMENTO IDD

COL-DANE-GEIH-2026

# Muestreo

## Procedimiento de muestreo

---

### MARCO ESTADÍSTICO (CENSAL O MUESTRAL)

La muestra de la operación estadística se selecciona en dos etapas de muestreo. Cada etapa tiene un marco muestral, la primera es un marco de municipios y la segunda un marco de medidas de tamaño o segmentos. El marco de segmentos tiene su origen en el marco geo-estadístico nacional (ver anexo 2.2.13.1), el cual es un producto del censo nacional de vivienda del año 2018.

### MARCO MUESTRAL DE MUNICIPIOS

El marco de municipios es un marco que enlista cada municipio del país (vigencia censo 2018). Tiene como información auxiliar las poblaciones de cabecera y centro poblado - rural disperso, indicadores del índice de pobreza multidimensional, porcentaje de población de centro poblado y rural disperso.

### MARCO MUESTRAL DE MEDIDAS DE TAMAÑO O SEGMENTOS

Es un marco construido a partir de la información estadística de densidad de viviendas en las manzanas y secciones rurales del marco geo-estadístico Nacional. Sirve para identificar y ubicar geográficamente las USM, por lo tanto, enlista cada conglomerado de 10 viviendas del país asociándole la manzana o sección rural a la que pertenece.

Para construir el marco se toma el número de viviendas que hay en cada manzana o sección rural del marco geo-estadístico y luego se divide en 10. Como resultado se obtiene el número de medidas de tamaño teóricas asociadas a cada manzana y sección rural del país. Con esta información se crea un marco nuevo de muestreo donde las filas son los conglomerados. Por ejemplo, si una manzana tiene 30 viviendas entonces al dividir en 10, significa que tiene 3 medidas de tamaño y que, en consecuencia, en el marco de conglomerados, van a aparecer 3 filas para representar los 3 conglomerados.

Con este marco se selecciona directamente la muestra de segmentos en la segunda etapa de esta encuesta. Luego de la selección se identifican las manzanas o secciones asociadas a la muestra y se procede a crear un marco de viviendas segmentado. Para el caso de cabeceras y centro poblado se desarrolla un operativo de campo 14 dirigido a las manzanas asociadas a estas medidas de tamaño, donde se procede a enlistar todas las viviendas contenidas en las manzanas. Este marco posteriormente es segmentado en medidas de tamaño (conglomerados) de 10 viviendas. En lo que respecta a la parte rural dispersa, cada sección se segmenta creando polígonos que, según información de foto satelital, contengan un aproximado de 10 viviendas. En ambos casos la segmentación se enumera y posteriormente se encuesta la medida de tamaño elegida en la primera parte.

### DISEÑO MUESTRAL

La GEIH es una encuesta dirigida a hogares y personas, por lo tanto, se establece una estrategia muestral adecuada para acceder a mediciones de las unidades de análisis con procedimientos que aseguran estimadores insesgados de los parámetros bajo una precisión y confianza preestablecida de antemano. A continuación, se describirá en detalle el plan muestral.

### TIPO DE MUESTREO

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

a) Probabilística: se satisfacen los siguientes criterios, es decir:

1. Se conoce de ante mano todas las posibles muestras.
2. Cada posible muestra tiene su respectiva probabilidad de selección.
3. El procedimiento da a cada elemento del universo de estudio una probabilidad de inclusión mayor que cero.

4. Se tiene un mecanismo aleatorio dentro del cual a cada posible muestra se le asigna su probabilidad de selección.

Si el lector requiere claridad sobre la explicación, ver (Särndal, Swensson, & Wretman, Model Assisted Survey Sampling, 1992).

b) Multietápica: para lograr la muestra de viviendas, hogares y personas se seleccionaron secuencialmente las unidades de muestreo (UPM y USM) en dos etapas.

En la primera se seleccionan municipios con un muestreo sin reemplazo proporcional al tamaño. En la segunda se seleccionan conglomerados de 10 viviendas con un muestreo sistemático con un ordenamiento geográfico previo.

c) Estratificada: para la GEIH desarrollada a partir del año 2020 se definió una estratificación de UPM y otra de USM.

Respecto a la estratificación de UPM, en la primera etapa de la muestra se generan grupos de municipios dentro de cada departamento. En el primer grupo de cada departamento se establece el estrato de inclusión forzosa.

El cual va a contener un único municipio que corresponde a la capital de departamento.

El segundo grupo lo componen los municipios considerados como áreas metropolitanas de las capitales de departamento. Las otras agrupaciones de municipios se hicieron a partir de un Análisis de Componentes Principales (ACP) 15 y un posterior agrupamiento tomando en cuenta la proyección de los municipios sobre los nuevos ejes factoriales y la población. Para lograrlo, se tomaron los resultados del Censo 2018: el índice de pobreza multidimensional, sus 15 indicadores y el porcentaje de urbanización. Con el propósito de buscar un diseño muestral auto ponderado, los grupos consolidados tienen aproximadamente la misma población, aunque esta característica no se puede asegurar; por tal motivo, van a existir unos estratos más grandes que otros y algunos de inclusión forzosa de municipios con una gran población, es decir, que no corresponden al tamaño promedio de los estratos.

En la segunda etapa, la estratificación se hace a nivel de cabeceras, centros poblados y rural disperso.

Por último, se hace una estratificación implícita de los segmentos al ordenar el marco según sector, sección y manzana y luego seleccionar a través de un muestreo sistemático la muestra de USM. Este procedimiento garantiza que la muestra quede uniformemente distribuida sobre las áreas geográficas.

d) De conglomerados: los conglomerados están definidos por la unidad secundaria de muestreo (USM) y corresponden a 10 viviendas (contiguas) en promedio donde se encuestan todas las viviendas, los hogares y las personas que los conforman.

## CÁLCULO DEL TAMAÑO DE LA MUESTRA

Para establecer el tamaño de la muestra, en primera instancia, se debe identificar un indicador relevante en la investigación al cual se le asigne un tamaño de muestra suficiente para que el estimador de este indicador cumpla con ciertas características de precisión y confianza.

En los casos donde el diseño muestral es complejo, además de lo anterior, se debe tomar como referencia un muestreo aleatorio simple de elementos (MAS) estableciendo un tamaño de muestra primario, el cual deberá ser ajustado por un efecto del diseño muestral (Deff) y otros parámetros como la tasa de no respuesta esperada y las subdivisiones de la población donde, en esta última, se procura garantizar suficiente muestra para su respectiva desagregación. Se invita al lector a revisar el desarrollo teórico del tema en (Cochran, 1977) o en las metodologías publicadas por Naciones Unidas sobre encuestas de hogares (Naciones Unidas, 2008).

14 Para el DANE operativo donde se enlistan las viviendas es llamado Recuento. Este procedimiento será descrito con mayor precisión en el capítulo del diseño de la recolección/acopio.

## Desviaciones del diseño de la muestra

En el desarrollo normal de la operación estadística, se identifican varios aspectos que pueden causar sesgos: la desactualización de marco, la imposibilidad de acceder a la muestra seleccionada con el procedimiento probabilístico predefinido debido a inconvenientes en campo y al agotamiento de la fuente.

La desactualización de marco se da con el tiempo debido a que en las áreas geográficas la población puede crecer o decrecer. Para controlar la desactualización del marco se procede a tomar en cuenta estos cambios con un factor de corrección. En la medida que las manzanas crezcan o decrezcan en viviendas las medidas de tamaño también lo harán en proporción ajustándolas en su factor de expansión con un factor de corrección que toma en cuenta el cambio del número de viviendas de la manzana respecto a lo medido en el censo. La fórmula se define como el número de segmentos actuales sobre el número de segmentos que aparecían en el marco al momento de seleccionar la medida de tamaño.

Se realiza un proceso en el cual, con la información del recuento se va actualizando el marco de medidas de tamaño en lo que respecta al número de segmentos. Y luego, se ajusta la muestra según el resultado de la actualización.

Con respecto a la pérdida de muestra, lo que se solicita es hacer revisitas en la medida de lo posible. En los casos donde se presentan novedades como pérdida de segmentos por problemas de orden público o por rechazo de segmentos, se tiene protocolos para evitar la pérdida. Sin embargo, en los casos donde es inevitable la pérdida de segmentos se opta por reemplazarlos con segmentos lo más parecidos posible al segmento original.

Estos casos, afortunadamente, son muy escasos.

El agotamiento de la fuente aparece cuando se encuesta al hogar en repetidas ocasiones. Para evitar este inconveniente, se procede a rotar la medida de tamaño cada año. El procedimiento consiste en encuestar al año siguiente una medida de tamaño vecina que no haya sido encuestada por alguna operación estadística en los dos años anteriores como mínimo.

# Formularios

## Información general

### DISEÑO DEL CUESTIONARIO

En este apartado se presentan los capítulos, su objetivo y las principales temáticas que conforman el formulario de la Gran Encuesta Integrada de Hogares, haciendo énfasis en los principales ajustes incluidos al formulario luego del rediseño de la encuesta, los cuales se aplican desde 2021.

- A. **IDENTIFICACIÓN:** se registra la información referente a región, departamento, municipio, clase, sector, sección, manzana, segmento, edificación, vivienda, dirección, barrio, así como, el control de calidad de la encuesta.
- B. **DATOS DE LA VIVIENDA:** se pretende conocer las características físicas de la vivienda, que permitan determinar su calidad y el acceso a los servicios públicos domiciliarios.
- C. **DATOS DEL HOGAR:** busca 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. Asimismo, pretende conocer los productos financieros utilizados en el hogar.
- D. **REGISTRO DE PERSONAS:** se registran todas las personas que forman parte del hogar.
- E. **CARACTERÍSTICAS GENERALES:** se caracteriza la población de acuerdo con su edad, sexo, estado civil y parentesco con el jefe del hogar, se identifica la población campesina y los grupos étnicos con el objetivo de relacionar esta información con el resto de la operación.

Con el propósito de brindar información sociodemográfica y de la situación en el mercado laboral de la población campesina, se incluyeron 3 preguntas para la identificación subjetiva de esta población: 6. ¿Usted se considera campesino(a)?, 7. ¿Usted considera que alguna vez fue campesino(a)? y 8. ¿Usted considera que la comunidad en la que vive es campesina?

Debido a la situación presentada por la pandemia (COVID - 19), y con el fin de poder establecer si a las personas encuestadas, se les han presentado alguna o algunas dificultades por esta causa, se incorporó la pregunta 13. Debido a la situación que se presenta en el país con la pandemia de COVID - 19, ¿Cuáles de las siguientes dificultades se le han presentado a .? .

F. **SEGURIDAD SOCIAL EN SALUD:** se identifica si las personas tienen acceso al sistema de salud, indagando por la afiliación y el tipo de régimen de seguridad social, así como, las dificultades en el funcionamiento humano actividades de acuerdo con la escala del Grupo de Washington.

Para la identificación de personas en condición de discapacidad, se incluyó la pregunta 5. Dada su condición física y mental, en su vida diaria tiene dificultades para realizar las siguientes actividades: a. ¿Oír, aun usando audífonos medicados?, b. ¿Hablar o conversar?, c. ¿Ver, aun usando lentes o gafas?, d. ¿Mover el cuerpo, caminar o subir y bajar escaleras?, e. ¿Agarrar o mover objetos con las manos?, f. ¿Entender, aprender, recordar o tomar decisiones por sí mismo/a?, g. ¿Comer, vestirse o bañarse por sí mismo (a)? y h. ¿Relacionarse o interactuar con las demás personas?

G. **EDUCACIÓN:** se identifica el alfabetismo, la asistencia escolar en la modalidad de educación formal, el mayor nivel educativo alcanzado y el grado de estudios de las personas encuestadas.

Para obtener mayor detalle en las opciones de respuesta de los niveles normalista en adelante (semestres) y especificación del campo de formación, se ajustaron las preguntas 4 y 5 de este capítulo, así: pregunta 4 ¿Cuál es el mayor nivel educativo alcanzado y el último grado o semestre aprobado por ....? 1. Ninguno, 2. Preescolar, 3. Básica primaria (1o - 5o), 4. Básica secundaria (6o - 9o), 5. Media académica (Bachillerato clásico), 6. Media técnica (Bachillerato técnico), 7. Normalista, 8. Técnica profesional, 9. Tecnológica, 10. Universitaria, 11. Especialización, 12. Maestría, 13. Doctorado y 99. No sabe, no informa; y pregunta 5. ¿Cuál es el título o diploma de mayor nivel educativo que ha recibido...? 1. Ninguno, 2. Media académica (Bachillerato clásico), 3. Media técnica (Bachillerato técnico), 4. Normalista, 5. Técnica profesional, 6. Tecnológica, 7. Universitaria, 8. Especialización, 9. Maestría, 10. Doctorado y 99. No sabe, no informa.

H. FUERZA DE TRABAJO: permite clasificar la Población en Edad de Trabajar (PET) entre población en la Fuerza de Trabajo (población ocupada o desocupada) o Fuera de la Fuerza de Trabajo.

Para mejorar la precisión del concepto de empleo, haciendo referencia al trabajo que se realiza a cambio de remuneración o beneficios para seguimiento más específico e identificar los ocupados ausentes por razones no tradicionales y la duración de su ausencia, se realizaron los siguientes ajustes:

Inclusión de la pregunta 2A. ¿. recibió remuneración en dinero o en especie a cambio de este trabajo? para confirmar si el encuestado se encuentra trabajando.

Inclusión de la pregunta 4A. ¿Por qué razón no trabajó la semana pasada?, con el propósito de captar las razones de ausencia temporal por las que las personas ocupadas no trabajaron en la semana de referencia, incluyendo situaciones de ausencia denominadas no tradicionales (Suspensión temporal del trabajo, Reducción de la actividad económica o cierre de la empresa, Razones climáticas o catástrofes naturales).

Por recomendación de la CIET-19, numeral 23 literal d, se requiere establecer un umbral de duración que permita determinar si se trata de ausencias de corta o larga duración. Este umbral debe tener en cuenta la duración de las licencias reconocidas en la legislación nacional, por lo cual se incorporó la pregunta 4B. ¿Cuánto tiempo estará ausente del trabajo (desde que dejó de trabajar hasta que regrese)?

En la pregunta 7. ¿Qué hizo. en las ÚLTIMAS 4 SEMANAS para conseguir un trabajo o instalar un negocio?, se modificó la alternativa de respuesta c. quedando: Hizo uso de los servicios de agencias o bolsas de empleo, Cajas de Compensación Familiar, SENA, Entes Territoriales (Servicio Público de Empleo - SPE).

I. OCUPADOS: busca conocer las características de la población ocupada y de sus empleos.

En este capítulo se incluyeron preguntas que permitieran garantizar las variables para la aplicación de la definición de empleo informal de la OIT, en cuanto a empleo en el sector informal, empleo informal y empleo informal fuera del sector informal.

En el subcapítulo de Asalariados se adicionaron las preguntas 15. La empresa, negocio o institución en la que .. trabaja ¿está registrada o tiene: a. Cámara de comercio, b. RUT, c. Personería jurídica y 15A. La empresa o negocio en la que .. trabaja tiene una oficina de contabilidad o cuenta con los servicios de un contador? a. Si, b. No, c. No sabe, no informa.

En el subcapítulo de Independientes se agregaron tres preguntas dirigidas a los que NO tienen un negocio de industria, comercio, servicios o una finca, así: 46. La empresa o negocio en la que .. realiza su trabajo ¿está registrada ante la cámara de comercio? (¿tiene registro mercantil?); 46A. La empresa o negocio en la que .. trabaja tiene una oficina de contabilidad o cuenta con los servicios de un contador, y 46B. La empresa o negocio en la que . trabaja ¿está registrada o tiene: a. RUT, b. Personería Jurídica.

Las nuevas preguntas dirigidas a los Independientes que tienen un negocio de industria, comercio, servicios o una finca, son la 47. ¿ ..... ha registrado el negocio ante la cámara de comercio? (¿tiene registro mercantil?): a. Si, b. No, la 47D. ¿A través de la contabilidad se pueden separar los gastos del negocio de los gastos del hogar? a. Si, b. No, c. No sabe, no informa; y la 47E. El negocio o actividad de.. ¿está registrado o tiene: a. RUT, b. Personería Jurídica.

Se incorporaron preguntas que permiten la identificación del grupo de contratistas dependientes en aras de establecer las relaciones de autoridad (CISO-18A) entre los trabajadores y las unidades económicas en las que trabajan o para las que realizan el trabajo, así: pregunta 28. En su trabajo, ¿quién decide sobre su horario de trabajo? ; 29. En su trabajo, ¿quién decide sobre qué producir o vender? ; y 30. En su trabajo, ¿quién decide sobre el precio de los bienes, productos o servicios prestados?

Para identificar la actividad que realizan los trabajadores independientes cuyo negocio se dedica a actividades de agricultura, ganadería, silvicultura y explotación de minas y canteras y aquellos dedicados a actividades de producción, compra y venta de mercancías o prestación de servicios, se incluyó la pregunta 37. La semana pasada, ¿qué tipo de actividad realizó en su negocio?.

Se incluyeron preguntas para identificar los gastos realizados por los trabajadores independientes. Para los que tienen negocios dedicados a actividades no agropecuarias, se agregó la pregunta 40. ¿Cuánto pagó durante el mes anterior, para el funcionamiento del negocio o actividad, por los siguientes conceptos? , en tanto que para los independientes que tienen negocios dedicados a actividades de producción agrícola, ganadera, pesquera o actividades de explotación de minas y canteras, se incorporó la pregunta 43. ¿Cuánto pagó en los últimos 12 meses, para el funcionamiento del negocio o actividad, por los siguientes conceptos?.

Para la identificación de los ingresos por autoconsumo de los trabajadores independientes que tienen negocio propio en el sector agropecuario, se incluyó la pregunta 44. ¿En el mes anterior, usted o algún miembro del hogar tomó algo de lo que cultivó, produjo o extrajo para consumo del hogar?. Asimismo, y para captar el estado de tenencia de la tierra en la que se realiza la actividad productiva de los negocios dedicados a actividades agropecuarias se agregó la pregunta 45. ¿Usted es propietario de la totalidad o parte de la tierra, terreno o granja donde realiza su actividad productiva?.

Se incorporaron preguntas para la medición del ingreso disponible tanto para los asalariados (17A.) como para los trabajadores independientes (34A) ¿Le descontaron retención en la fuente a lo que ganó el mes pasado en este empleo?.

Para la identificación del oficio y la rama de actividad del trabajo secundario se incluyeron las preguntas 70. ¿Qué hace..en este trabajo? y 71. ¿A qué actividad se dedica principalmente la empresa o negocio en la que .. realiza su trabajo? , respectivamente.

J. NO OCUPADOS: busca conocer las características de la población no ocupada y de su búsqueda de empleo.

Se integraron los antiguos capítulos de Desocupados e Inactivos en uno solo, el cual se denominó No ocupados, esto con el propósito de identificar las posibles transiciones que se dan entre las poblaciones que componen la fuerza de trabajo y la población fuera de la fuerza de trabajo. El concepto de población inactiva ahora corresponde a la población fuera de la fuerza de trabajo, tomando como referencia la terminología contenida en el numeral 81 de la 19ª CIET.

Se incluyeron tres preguntas sobre disponibilidad para trabajar dirigidas a la población fuera de la fuerza de trabajo, estas corresponden a las preguntas 4. Si le hubiera resultado algún trabajo a .. ¿estaba disponible la semana pasada para empezar a trabajar? ; 5. Si le resultara algún trabajo, ¿cuándo estaría . disponible para trabajar? : y 6. Si le hubiera resultado un trabajo la semana pasada, ¿cuántas horas a la semana estaba disponible para trabajar? . Asimismo, se agregaron preguntas relacionadas con Oficio, Actividad económica y nombre de la empresa donde la persona trabajó por última vez.

K. OTRAS FORMAS DE TRABAJO: se captan las actividades realizadas por miembros del hogar y por las cuales no obtienen remuneración. Además, permite captar las actividades de trabajo comunitario voluntario.

El capítulo de Otras actividades y ayudas en la semana pasada se modificó de manera parcial y ahora se denomina Otras formas de trabajo.

Se incorporaron subcapítulos que permiten medir las diferentes formas de trabajo, se modificaron y separaron preguntas relacionadas con trabajo en producción de servicios para el autoconsumo y en producción de bienes para el autoconsumo, adicionalmente, se incluyeron preguntas relativas a la medición del trabajo en formación, acotando su aplicación a personas de 15 a 35 años (dada la prevalencia de respuestas afirmativas en resultados de prueba piloto y diseño experimental). Por recomendación de la mesa técnica realizada con la OIT, el período de las preguntas sobre formación para el trabajo, trabajo en producción de bienes para autoconsumo y trabajo voluntario, se acotó a las últimas 4 semanas.

En el capítulo de Otros ingresos, se incluyen seis preguntas de impuestos asociadas a la medición del ingreso disponible, por lo que ahora se denomina Otros ingresos e impuestos. Las preguntas en mención corresponden a la 5. ¿Es propietario de una o varias propiedades inmuebles? ; 6.

Durante los últimos doce meses, ¿cuánto pagó por impuesto predial de su(s) propiedad(es)? ; 7.

Durante los últimos doce meses, ¿cuánto pagó por impuesto de valorización de su(s) propiedad(es)? ; 8. Durante los últimos doce meses, ¿realizó el pago de impuesto de vehículos? ; 9. Durante los últimos doce meses, ¿realizó el pago de impuestos a la renta y complementarios? y 10. Durante los últimos doce meses, ¿realizó el pago de impuestos a ganancias en juegos de azar, chances, loterías, indemnizaciones, liquidaciones, venta de propiedades, acciones, vehículos, etc.?

M. MICRONEGOCIOS: se indaga sobre los aspectos de los trabajadores(as) por cuenta propia y patronos que trabajan en empresas hasta de 10 personas.

M. FORMACIÓN PARA EL TRABAJO11 : se determinan las dinámicas de la formación para el trabajo y la participación en procesos de certificación de competencias de la población de 15 años y más.

O. FECUNDIDAD: se registran los hijos e hijas nacidos vivos.

P. MIGRACIÓN: se identifican las dinámicas migratorias de la población objetivo dentro y fuera del país y la migración laboral internacional.

Las mejoras aplicadas al módulo de migración están acordes con las definiciones aprobadas en las directrices sobre

estadísticas de la migración laboral internacional y con las recomendaciones de la consultoría realizada con expertos de la OIT. De esta manera, se incluyeron las siguientes preguntas: 2. Nacionalidad, 3. ¿. tiene la intención de quedarse a vivir en Colombia? , 4. ¿Alguna vez ha vivido en otro(s) país(es) durante más de 6 meses? , 5. ¿Mientras estuvo viviendo en el exterior ., trabajó? , 6. ¿A qué país viajó por última vez a trabajar? , 6A. ¿Durante cuánto tiempo estuvo trabajando en ese país? , 7. ¿. en qué año viajó a ese país? , 8. ¿. en qué año salió de ese país? , y 9. ¿. en qué año volvió a vivir en Colombia? .

Q. TRABAJO INFANTIL 12: se busca identificar las actividades y razones del trabajo infantil.

R. TIPO DE IDENTIFICACIÓN: se indaga por la clase de documento de identidad, así como, por la orientación sexual y el género con el cual se reconoce la persona.

Las preguntas incluidas en el módulo de Tipo de identificación, sobre orientación sexual e identidad de género, hacen parte de las estrategias de política pública para superar la invisibilización estadística y el subregistro de que son objeto los grupos poblacionales LGBTI. Se incorporaron dos preguntas, así: 4. ¿Usted siente atracción sexual o romántica por? a. Hombres, b. Mujeres, c. Ambos sexos, d. Otro ¿Cuál?, y 4A ¿Usted se reconoce como? a. Hombre, b. Mujer, c. Persona no binaria (ni hombre ni mujer), d. Prefiere identificarse de otra forma ¿Cuál?.

Las mejoras realizadas al cuestionario de la GEIH fueron la inclusión de preguntas nuevas en los capítulos de Características generales, Seguridad social en salud, Fuerza de trabajo, Ocupados, Otras formas de trabajo, Otros ingresos e impuestos, Migración y Tipo de identificación; la reformulación de algunas preguntas de los módulos de Educación, Fuerza de trabajo, Ocupados y Otras formas de trabajo y la integración de los capítulos de Desocupados e Inactivos, el cual se denomina No Ocupados.

11 Entre 2013 y 2022 el módulo de formación para el trabajo se aplicó en el segundo trimestre de cada año; excepto en el período abril-junio 2020, debido al cambio en el operativo de recolección de la GEIH por la pandemia del COVID-19. A partir de 2023, se recolecta información durante todos los meses del año.

12 Entre 2012 y 2022 el módulo de trabajo infantil se aplicó en el cuarto trimestre de cada año. A partir de 2023, se recolecta información durante todos los meses del año.

## Recolección de datos

### Fecha de recolección de datos

Start	End	Cycle
2026	2026	N/A

### Time Periods

Start	End	Cycle
2026		Semanal, Mensual, Trimestral y Anual

### Modo de recolección de datos

Entrevista personal asistida con DMC (dispositivo móvil de captura)

### Notas de recolección de datos

#### ESTRUCTURA ORGANIZACIONAL DEL OPERATIVO Y CONFORMACIÓN DEL EQUIPO

A continuación, se describen las actividades correspondientes a cada rol:

**Responsable de Encuesta:** en las sedes, es la persona responsable de los operativos, prueba piloto, diseños experimentales y paralelo, en todos sus aspectos, de éste dependen los equipos de trabajo adscritos a su sede, con quienes debe mantener comunicación permanente.

**Coordinador de Campo:** es la persona responsable de las actividades operativas y la supervisión a los equipos de trabajo en campo. Tiene a su cargo el control de calidad de la información y el manejo de la encuesta, asegura que toda la información operativa requerida por el DANE Central llegue oportunamente y con la calidad requerida.

**Apoyo Informático:** es la persona responsable de la actualización de los aplicativos de captura y de la transmisión diaria de la información a la base de sistemas de DANE Central.

**Recuentista:** en las áreas urbanas, es la persona encargada de realizar el recorrido de un área previamente asignada por el responsable de encuesta y efectuar el conteo de las unidades que la conforman, con el fin de generar la correcta asignación de las medidas de tamaño y selección de los segmentos a encuestar. En las zonas rurales, el recuento lo hace el equipo de recolección.

**Supervisor de Campo:** es la persona que se encarga de conseguir la ubicación correcta del equipo en campo, de hacer el seguimiento y control de la orientación del proceso de recolección en las áreas seleccionadas y asignadas. Tiene a su cargo dos o tres encuestadores. Debe responder por la calidad de los datos que se recogen.

**Encuestador:** es la persona que se encarga de visitar las viviendas y los hogares seleccionados y entrevistar a las personas para obtener información de acuerdo con la metodología, los procedimientos y las instrucciones contenidas en el Manual de Recolección y Conceptos Básicos de cada operativo. Reporta su trabajo al Supervisor de Campo.

En el DANE Central, los aspectos metodológicos y muestrales están a cargo de los equipos de Temática y Diseños Muestrales, respectivamente; asimismo, los aspectos logísticos, operativos y de producción, le corresponden a la Dirección de Recolección y Acopio - DRA. El desarrollo del operativo es responsabilidad de cada Dirección Territorial y las labores de procesamiento de datos, la realiza la Oficina de Sistemas. En los manuales de Sistemas y de Usuario, se encuentran los lineamientos y las normas del operativo de recolección.

#### ESQUEMA DE ENTRENAMIENTO PERSONAL

En el marco de la visión del DANE, el modelo de aprendizaje, evaluación y selección busca garantizar la adecuada preparación del personal operativo para obtener la mejor cobertura y calidad en las operaciones estadísticas.

Este modelo es un proceso de aprendizaje b-learning, en el que se combinan la modalidad de estudio presencial con la aplicación de pruebas de conocimientos virtual, de la siguiente manera:

**Aprendizaje virtual:** inicialmente, los candidatos que participan en este proceso van a tener la oportunidad de contextualizarse sobre cada investigación a partir de materiales virtuales (multimedia), que encontrarán estructurados en un curso en la plataforma @prendanet, los cuales están relacionados con las generalidades de la encuesta, la recolección de la información y las características particulares que conforman la operación estadística. Esta modalidad de estudio se lleva a cabo por auto aprendizaje y los candidatos la pueden realizar en una computadora con conexión a internet desde el sitio que deseen.

Al finalizar este período de aprendizaje virtual, los estudiantes deben desarrollar la primera prueba de conocimientos en la plataforma, la cual pueden presentar desde cualquier lugar con conexión a internet.

**Aprendizaje presencial o sincrónico:** las personas que aprobaron la primera prueba de conocimientos se deben dirigir a cada una de las sedes del DANE o ingresar al espacio virtual indicado para realizar un proceso de entrenamiento, que incluye el abordaje de los aspectos temáticos, operativos, muestrales, cartográficos y de sistemas, así como el desarrollo de ejercicios enfocados en lograr una mayor aprehensión de los conocimientos requeridos y para resolver la mayor cantidad de inquietudes posibles.

Luego de esto, los candidatos deben presentar la segunda y última prueba de conocimientos en la misma sede donde realizaron el proceso de aprendizaje presencial.

## INVITACIÓN PÚBLICA DE SELECCIÓN DE PERSONAL

La selección y entrenamiento del personal se realiza siguiendo los lineamientos establecidos desde la Secretaría General del DANE Central. El proceso de selección inicia con la inscripción en el Banco de Hojas de Vida de la entidad ( <https://www.dane.gov.co/index.php/convocatorias-y-contratacion/informacion-laboral/convocatorias-roles-operativos-para-operaciones-estadisticas> ), donde el aspirante registra su formación académica y experiencia laboral, para poder formar parte del proceso de selección del personal operativo requerido para los diferentes proyectos que adelanta la entidad a nivel nacional.

## PROCESO DE SENSIBILIZACIÓN Y ACUERDOS DE INTERCAMBIO

Un mes antes de la recolección y previo a iniciar la sensibilización se realiza el recuento, proceso por el cual se localiza, delimita y se reconocen las áreas urbanas seleccionadas en la muestra para obtener una identificación detallada de las manzanas correspondientes a segmentos con rotación. En las áreas rurales se realiza en simultáneo con el levantamiento de la información.

En este proceso, el recuentista hace el levantamiento del croquis en las clases 1, 2 y 3 (ciudades, áreas metropolitanas, cabeceras centros poblados y rural disperso).

Adicionalmente, los recuentistas deben enviar a la Dirección de Recolección y Acopio (DRA) los siguientes reportes: novedades cartográficas, segmentos agotados y segmentos con submuestreo. Asimismo, se reportan todas las novedades cartográficas encontradas durante el recuento o la recolección, vía correo electrónico a la DRA para su correspondiente gestión, seguimiento y respuesta.

Por su parte el objetivo de la sensibilización es lograr una respuesta oportuna y de calidad por parte de las personas que conforman el hogar a encuestar. Además, busca generar un ambiente de colaboración y hacer un primer acercamiento a los hogares, de manera que el sensibilizador presente a todos los miembros del hogar los objetivos de la encuesta y la importancia que tiene para el país la información que se va a suministrar, así como, explicar que la información suministrada por cada persona es confidencial (Ley 79 de 1993) y que es utilizada solo para fines estadísticos. Por otra parte, informa que posteriormente el hogar será visitado tanto por el encuestador como por el supervisor de la operación estadística para aplicar la encuesta a cada persona del hogar.

En el operativo urbano, la persona encargada de la sensibilización, la realiza una semana antes de la aplicación de la encuesta. Mientras que, para el operativo rural, el personal encargado de aplicar la encuesta es el mismo que realiza la sensibilización y el recuento, debido a que para este operativo se organiza una ruta y el desplazamiento a los segmentos seleccionados puede durar varios días. Es importante que, en el desarrollo de la sensibilización rural, se busque persuadir a los miembros del hogar para aplicar la encuesta en ese mismo momento y, de no ser posible, establecer una cita en la cual se pueda realizar la recolección de la información.

Adicional a la sensibilización del hogar, el sensibilizador debe diligenciar el formato "Listado de unidades sensibilizadas", con el fin de registrar la información del hogar seleccionado en la muestra. Este formato se relaciona directamente con la información levantada previamente en el recuento y relaciona la cantidad de viviendas, hogares y personas del segmento seleccionado. Este formato, contiene la información base para medir la cobertura de hogares.

## Formularios

### DISEÑO DEL CUESTIONARIO

En este apartado se presentan los capítulos, su objetivo y las principales temáticas que conforman el formulario de la Gran Encuesta Integrada de Hogares, haciendo énfasis en los principales ajustes incluidos al formulario luego del rediseño de la encuesta, los cuales se aplican desde 2021.

A. IDENTIFICACIÓN: se registra la información referente a región, departamento, municipio, clase, sector, sección, manzana, segmento, edificación, vivienda, dirección, barrio, así como, el control de calidad de la encuesta.

B. DATOS DE LA VIVIENDA: se pretende conocer las características físicas de la vivienda, que permitan determinar su calidad y el acceso a los servicios públicos domiciliarios.

C. DATOS DEL HOGAR: busca 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. Asimismo, pretende conocer los productos financieros utilizados en el hogar.

D. REGISTRO DE PERSONAS: se registran todas las personas que forman parte del hogar.

E. CARACTERÍSTICAS GENERALES: se caracteriza la población de acuerdo con su edad, sexo, estado civil y parentesco con el jefe del hogar, se identifica la población campesina y los grupos étnicos con el objetivo de relacionar esta información con el resto de la operación.

Con el propósito de brindar información sociodemográfica y de la situación en el mercado laboral de la población campesina, se incluyeron 3 preguntas para la identificación subjetiva de esta población: 6. ¿Usted se considera campesino(a)?, 7. ¿Usted considera que alguna vez fue campesino(a)? y 8. ¿Usted considera que la comunidad en la que vive es campesina?

Debido a la situación presentada por la pandemia (COVID - 19), y con el fin de poder establecer si a las personas encuestadas, se les han presentado alguna o algunas dificultades por esta causa, se incorporó la pregunta 13. Debido a la situación que se presenta en el país con la pandemia de COVID - 19, ¿Cuáles de las siguientes dificultades se le han presentado a .? .

F. SEGURIDAD SOCIAL EN SALUD: se identifica si las personas tienen acceso al sistema de salud, indagando por la afiliación y el tipo de régimen de seguridad social, así como, las dificultades en el funcionamiento humano actividades de acuerdo con la escala del Grupo de Washington.

Para la identificación de personas en condición de discapacidad, se incluyó la pregunta 5. Dada su condición física y mental, en su vida diaria tiene dificultades para realizar las siguientes actividades: a. ¿Oír, aun usando audífonos medicados?, b. ¿Hablar o conversar?, c. ¿Ver, aun usando lentes o gafas?, d. ¿Mover el cuerpo, caminar o subir y bajar escaleras?, e. ¿Agarrar o mover objetos con las manos?, f. ¿Entender, aprender, recordar o tomar decisiones por sí mismo/a?, g. ¿Comer, vestirse o bañarse por sí mismo (a)? y h. ¿Relacionarse o interactuar con las demás personas?

G. EDUCACIÓN: se identifica el alfabetismo, la asistencia escolar en la modalidad de educación formal, el mayor nivel educativo alcanzado y el grado de estudios de las personas encuestadas.

Para obtener mayor detalle en las opciones de respuesta de los niveles normalista en adelante (semestres) y especificación del campo de formación, se ajustaron las preguntas 4 y 5 de este capítulo, así: pregunta 4 ¿Cuál es el mayor nivel educativo alcanzado y el último grado o semestre aprobado por ....? 1. Ninguno, 2. Preescolar, 3. Básica primaria (1o - 5o), 4. Básica secundaria (6o - 9o), 5. Media académica (Bachillerato clásico), 6. Media técnica (Bachillerato técnico), 7. Normalista, 8. Técnica profesional, 9. Tecnológica, 10. Universitaria, 11. Especialización, 12. Maestría, 13. Doctorado y 99. No sabe, no informa; y pregunta 5. ¿Cuál es el título o diploma de mayor nivel educativo que ha recibido...? 1. Ninguno, 2. Media académica (Bachillerato clásico), 3. Media técnica (Bachillerato técnico), 4. Normalista, 5. Técnica profesional, 6. Tecnológica, 7. Universitaria, 8. Especialización, 9. Maestría, 10. Doctorado y 99. No sabe, no informa.

H. FUERZA DE TRABAJO: permite clasificar la Población en Edad de Trabajar (PET) entre población en la Fuerza de Trabajo (población ocupada o desocupada) o Fuera de la Fuerza de Trabajo.

Para mejorar la precisión del concepto de empleo, haciendo referencia al trabajo que se realiza a cambio de remuneración o beneficios para seguimiento más específico e identificar los ocupados ausentes por razones no tradicionales y la duración de su ausencia, se realizaron los siguientes ajustes:

Inclusión de la pregunta 2A. ¿. recibió remuneración en dinero o en especie a cambio de este trabajo? para confirmar si el encuestado se encuentra trabajando.

Inclusión de la pregunta 4A. ¿Por qué razón no trabajó la semana pasada?, con el propósito de captar las razones de ausencia temporal por las que las personas ocupadas no trabajaron en la semana de referencia, incluyendo situaciones de ausencia denominadas no tradicionales (Suspensión temporal del trabajo, Reducción de la actividad económica o cierre de la empresa, Razones climáticas o catástrofes naturales).

Por recomendación de la CIET-19, numeral 23 literal d, se requiere establecer un umbral de duración que permita determinar si se trata de ausencias de corta o larga duración. Este umbral debe tener en cuenta la duración de las licencias reconocidas en la legislación nacional, por lo cual se incorporó la pregunta 4B. ¿Cuánto tiempo estará ausente del trabajo (desde que dejó de trabajar hasta que regrese)?

En la pregunta 7. ¿Qué hizo. en las ÚLTIMAS 4 SEMANAS para conseguir un trabajo o instalar un negocio?, se modificó la alternativa de respuesta c. quedando: Hizo uso de los servicios de agencias o bolsas de empleo, Cajas de Compensación Familiar, SENA, Entes Territoriales (Servicio Público de Empleo - SPE).

I. OCUPADOS: busca conocer las características de la población ocupada y de sus empleos.

En este capítulo se incluyeron preguntas que permitieran garantizar las variables para la aplicación de la definición de empleo informal de la OIT, en cuanto a empleo en el sector informal, empleo informal y empleo informal fuera del sector informal.

En el subcapítulo de Asalariados se adicionaron las preguntas 15. La empresa, negocio o institución en la que .. trabaja ¿está registrada o tiene: a. Cámara de comercio, b. RUT, c. Personería jurídica y 15A. La empresa o negocio en la que .. trabaja tiene una oficina de contabilidad o cuenta con los servicios de un contador? a. Si, b. No, c. No sabe, no informa.

En el subcapítulo de Independientes se agregaron tres preguntas dirigidas a los que NO tienen un negocio de industria, comercio, servicios o una finca, así: 46. La empresa o negocio en la que .. realiza su trabajo ¿está registrada ante la cámara de comercio? (¿tiene registro mercantil?); 46A. La empresa o negocio en la que .. trabaja tiene una oficina de contabilidad o cuenta con los servicios de un contador, y 46B. La empresa o negocio en la que . trabaja ¿está registrada o tiene: a. RUT, b. Personería Jurídica.

Las nuevas preguntas dirigidas a los Independientes que tienen un negocio de industria, comercio, servicios o una finca, son la 47. ¿ ..... ha registrado el negocio ante la cámara de comercio? (¿tiene registro mercantil?): a. Si, b. No, la 47D. ¿A través de la contabilidad se pueden separar los gastos del negocio de los gastos del hogar? a. Si, b. No, c. No sabe, no informa; y la 47E. El negocio o actividad de.. ¿está registrado o tiene: a. RUT, b. Personería Jurídica.

Se incorporaron preguntas que permiten la identificación del grupo de contratistas dependientes en aras de establecer las relaciones de autoridad (CISO-18A) entre los trabajadores y las unidades económicas en las que trabajan o para las que realizan el trabajo, así: pregunta 28. En su trabajo, ¿quién decide sobre su horario de trabajo? ; 29. En su trabajo, ¿quién decide sobre qué producir o vender? ; y 30. En su trabajo, ¿quién decide sobre el precio de los bienes, productos o servicios prestados?

Para identificar la actividad que realizan los trabajadores independientes cuyo negocio se dedica a actividades de agricultura, ganadería, silvicultura y explotación de minas y canteras y aquellos dedicados a actividades de producción, compra y venta de mercancías o prestación de servicios, se incluyó la pregunta 37. La semana pasada, ¿qué tipo de actividad realizó en su negocio?.

Se incluyeron preguntas para identificar los gastos realizados por los trabajadores independientes. Para los que tienen negocios dedicados a actividades no agropecuarias, se agregó la pregunta 40. ¿Cuánto pagó durante el mes anterior, para el funcionamiento del negocio o actividad, por los siguientes conceptos? , en tanto que para los independientes que tienen negocios dedicados a actividades de producción agrícola, ganadera, pesquera o actividades de explotación de minas y canteras, se incorporó la pregunta 43. ¿Cuánto pagó en los últimos 12 meses, para el funcionamiento del negocio o actividad, por los siguientes conceptos?.

Para la identificación de los ingresos por autoconsumo de los trabajadores independientes que tienen negocio propio en el sector agropecuario, se incluyó la pregunta 44. ¿En el mes anterior, usted o algún miembro del hogar tomó algo de lo que cultivó, produjo o extrajo para consumo del hogar?. Asimismo, y para captar el estado de tenencia de la tierra en la que se realiza la actividad productiva de los negocios dedicados a actividades agropecuarias se agregó la pregunta 45. ¿Usted es propietario de la totalidad o parte de la tierra, terreno o granja donde realiza su actividad productiva?.

Se incorporaron preguntas para la medición del ingreso disponible tanto para los asalariados (17A.) como para los trabajadores independientes (34A) ¿Le descontaron retención en la fuente a lo que ganó el mes pasado en este empleo?.

Para la identificación del oficio y la rama de actividad del trabajo secundario se incluyeron las preguntas 70. ¿Qué hace..en este trabajo? y 71. ¿A qué actividad se dedica principalmente la empresa o negocio en la que .. realiza su trabajo? , respectivamente.

J. NO OCUPADOS: busca conocer las características de la población no ocupada y de su búsqueda de empleo.

Se integraron los antiguos capítulos de Desocupados e Inactivos en uno solo, el cual se denominó No ocupados, esto con el propósito de identificar las posibles transiciones que se dan entre las poblaciones que componen la fuerza de trabajo y la población fuera de la fuerza de trabajo. El concepto de población inactiva ahora corresponde a la población fuera de la fuerza de trabajo, tomando como referencia la terminología contenida en el numeral 81 de la 19ª CIET.

Se incluyeron tres preguntas sobre disponibilidad para trabajar dirigidas a la población fuera de la fuerza de trabajo, estas corresponden a las preguntas 4. Si le hubiera resultado algún trabajo a .. ¿estaba disponible la semana pasada para empezar a trabajar? ; 5. Si le resultara algún trabajo, ¿cuándo estaría . disponible para trabajar? : y 6. Si le hubiera resultado un trabajo la semana pasada, ¿cuántas horas a la semana estaba disponible para trabajar? . Asimismo, se agregaron preguntas relacionadas con Oficio, Actividad económica y nombre de la empresa donde la persona trabajó por última vez.

K. OTRAS FORMAS DE TRABAJO: se captan las actividades realizadas por miembros del hogar y por las cuales no obtienen remuneración. Además, permite captar las actividades de trabajo comunitario voluntario.

El capítulo de Otras actividades y ayudas en la semana pasada se modificó de manera parcial y ahora se denomina Otras formas de trabajo.

Se incorporaron subcapítulos que permiten medir las diferentes formas de trabajo, se modificaron y separaron preguntas relacionadas con trabajo en producción de servicios para el autoconsumo y en producción de bienes para el autoconsumo, adicionalmente, se incluyeron preguntas relativas a la medición del trabajo en formación, acotando su aplicación a personas de 15 a 35 años (dada la prevalencia de respuestas afirmativas en resultados de prueba piloto y diseño experimental). Por recomendación de la mesa técnica realizada con la OIT, el período de las preguntas sobre formación para el trabajo, trabajo en producción de bienes para autoconsumo y trabajo voluntario, se acotó a las últimas 4 semanas.

En el capítulo de Otros ingresos, se incluyen seis preguntas de impuestos asociadas a la medición del ingreso disponible, por lo que ahora se denomina Otros ingresos e impuestos. Las preguntas en mención corresponden a la 5. ¿Es propietario de una o varias propiedades inmuebles? ; 6.

Durante los últimos doce meses, ¿cuánto pagó por impuesto predial de su(s) propiedad(es)? ; 7.

Durante los últimos doce meses, ¿cuánto pagó por impuesto de valorización de su(s) propiedad(es)? ; 8. Durante los últimos doce meses, ¿realizó el pago de impuesto de vehículos? ; 9. Durante los últimos doce meses, ¿realizó el pago de impuestos a la renta y complementarios? y 10. Durante los últimos doce meses, ¿realizó el pago de impuestos a ganancias en juegos de azar, chances, loterías, indemnizaciones, liquidaciones, venta de propiedades, acciones, vehículos, etc.?

M. MICRONEGOCIOS: se indaga sobre los aspectos de los trabajadores(as) por cuenta propia y patronos que trabajan en empresas hasta de 10 personas.

M. FORMACIÓN PARA EL TRABAJO11 : se determinan las dinámicas de la formación para el trabajo y la participación en procesos de certificación de competencias de la población de 15 años y más.

O. FECUNDIDAD: se registran los hijos e hijas nacidos vivos.

P. MIGRACIÓN: se identifican las dinámicas migratorias de la población objetivo dentro y fuera del país y la migración laboral internacional.

Las mejoras aplicadas al módulo de migración están acordes con las definiciones aprobadas en las directrices sobre

estadísticas de la migración laboral internacional y con las recomendaciones de la consultoría realizada con expertos de la OIT. De esta manera, se incluyeron las siguientes preguntas: 2. Nacionalidad, 3. ¿. tiene la intención de quedarse a vivir en Colombia? , 4. ¿Alguna vez ha vivido en otro(s) país(es) durante más de 6 meses? , 5. ¿Mientras estuvo viviendo en el exterior ., trabajó? , 6. ¿A qué país viajó por última vez a trabajar? , 6A. ¿Durante cuánto tiempo estuvo trabajando en ese país? , 7. ¿. en qué año viajó a ese país? , 8. ¿. en qué año salió de ese país? , y 9. ¿. en qué año volvió a vivir en Colombia? .

Q. TRABAJO INFANTIL 12: se busca identificar las actividades y razones del trabajo infantil.

R. TIPO DE IDENTIFICACIÓN: se indaga por la clase de documento de identidad, así como, por la orientación sexual y el género con el cual se reconoce la persona.

Las preguntas incluidas en el módulo de Tipo de identificación, sobre orientación sexual e identidad de género, hacen parte de las estrategias de política pública para superar la invisibilización estadística y el subregistro de que son objeto los grupos poblacionales LGBTI. Se incorporaron dos preguntas, así: 4. ¿Usted siente atracción sexual o romántica por? a. Hombres, b. Mujeres, c. Ambos sexos, d. Otro ¿Cuál?, y 4A ¿Usted se reconoce como? a. Hombre, b. Mujer, c. Persona no binaria (ni hombre ni mujer), d. Prefiere identificarse de otra forma ¿Cuál?.

Las mejoras realizadas al cuestionario de la GEIH fueron la inclusión de preguntas nuevas en los capítulos de Características generales, Seguridad social en salud, Fuerza de trabajo, Ocupados, Otras formas de trabajo, Otros ingresos e impuestos, Migración y Tipo de identificación; la reformulación de algunas preguntas de los módulos de Educación, Fuerza de trabajo, Ocupados y Otras formas de trabajo y la integración de los capítulos de Desocupados e Inactivos, el cual se denomina No Ocupados.

11 Entre 2013 y 2022 el módulo de formación para el trabajo se aplicó en el segundo trimestre de cada año; excepto en el período abril-junio 2020, debido al cambio en el operativo de recolección de la GEIH por la pandemia del COVID-19. A partir de 2023, se recolecta información durante todos los meses del año.

12 Entre 2012 y 2022 el módulo de trabajo infantil se aplicó en el cuarto trimestre de cada año. A partir de 2023, se recolecta información durante todos los meses del año.

## Data Collectors

Nombre	Abbreviation	Affiliation
Departamento Administrativo Nacional de Estadística	DANE	Gobierno Nacional

## Supervisión

### DISEÑO DE LA ESTRATEGIA DE SEGUIMIENTO Y CONTROL

Para el control de calidad de la información recopilada, se utilizan normas de validación y consistencia y controles operativos de calidad, como son:

### RESUMEN DE COBERTURA

El objetivo de este formato es llevar un control de manera resumida del número de viviendas, hogares y personas encontrados en los segmentos asignados y trabajados durante la recolección para detectar posibles inconsistencias y alarmas en la información recolectada. Es utilizado en el operativo urbano, en los centros poblados y rural disperso.

El principal usuario del formato es la persona encargada de la Coordinación de Campo en cada una de las ciudades, quien consolida la información de la supervisión a cargo para remitir semanalmente estos resúmenes a la Dirección de Recolección y Acopio en el DANE Central vía Protocolo de Transferencia de Archivos (FTP). De este reporte se obtienen los indicadores de cobertura operativa y los resultados que permiten medir el proceso de seguimiento.

### DETECCIÓN DE ALARMAS Y CORRECCIÓN DE INCONSISTENCIAS

Del resumen operativo de cobertura se obtienen las inconsistencias detectadas en la revisión semanal que se realiza sobre las variables que componen las viviendas, los hogares y las personas de cada uno de los segmentos asignados en la muestra de cada ciudad. Al realizar la revisión por parte de los grupos internos de trabajo de hogares y micronegocios de la DRA, se reportan las cifras que son motivo de justificación por parte de las sedes como: baja cobertura, segmentos con

menos de 10 viviendas encontradas, alto número de vacantes, entre otros. Las inconsistencias deben ser corregidas inmediatamente por la ciudad correspondiente, a través de la coordinación de campo y las alarmas deben ser justificadas en el siguiente reporte de cobertura enviado por cada una de las sedes.

## MONITOREO

Los monitoreos a los grupos de campo realizados por los equipos de DANE central mediante comisiones dependen de las falencias detectadas en los controles y seguimientos permanentes que se efectúan a las diferentes sedes en donde se aplica la GEIH marco 2018, como alto número de inconsistencias temáticas, baja cobertura en hogares, disminución del promedio de personas por hogar, etc. Mediante este procedimiento se observa el desempeño y se detectan falencias que pueden afectar la calidad de la información recolectada. Los monitoreos son realizados por los equipos de temática, diseños muestrales y recolección y acopio con el propósito de verificar en campo aspectos como: presentación personal, lectura adecuada de las preguntas, no inducción o deducción de respuestas, cambios adecuados de capítulo, realización de buenos sondeos, registro adecuado de los miembros del hogar y el manejo de conceptos.

Igualmente, se verifica la correcta distribución diaria de materiales de trabajo, el diligenciamiento de formatos, el manejo adecuado del transporte, la revisión de las encuestas por parte del personal de supervisión, la ubicación en el segmento y la capacidad para crear un ambiente propicio con la fuente, que permita el desarrollo óptimo de la encuesta. Todos los aspectos por mejorar por parte de los equipos de recolección, son socializados semanalmente mediante reuniones donde se presentan informes operativos preparados por cada grupo de trabajo que interviene en los procesos de la operación estadística, como la DRA, Temática, Pobreza, GEDI, Sistemas y la DIG, con los responsables de cada proceso y se hace seguimiento a la puesta en marcha de las soluciones y correctivos que sean necesarios para la mejora continua de la calidad de la información.

## INDICADORES DE COBERTURA EN VIVIENDAS , HOGARES Y PERSONAS

Se realiza seguimiento semanal a las tasas de cobertura operativa en viviendas, hogares y personas de cada una de las ciudades teniendo en cuenta el dominio (urbano y rural) de la encuesta, estos indicadores calculados por la DRA garantizan un óptimo desarrollo del operativo de campo, y permiten detectar oportunamente comportamientos inusuales en la cobertura al compararla con el período inmediatamente anterior.

Cobertura de viviendas: se obtiene de cruzar la información de viviendas iniciales contra las viviendas finales encontradas a nivel nacional, por dominio y por ciudad.

## INDICADOR DE CALIDAD DE LA REOLECCIÓN (ICR)

El objetivo de este indicador es determinar el grado de calidad y confiabilidad de la información enviada por la persona encargada de la coordinación de campo de cada sede. Se mide a través del número de inconsistencias detectadas por el supervisor en el acompañamiento al personal encuestador en campo. Esto permite hacer seguimiento continuo a las diferentes sedes que presentan bajos indicadores de calidad de la información, para determinar las posibles falencias y sugerir recomendaciones o medidas de mejoramiento en la entrevista con la fuente.

Se calcula un indicador de calidad para cada persona del grupo de recolección, el cual es obtenido con el promedio aritmético de los puntajes alcanzados por los formularios que fueron sometidos al proceso de supervisión y control de calidad, mediante un formato dispuesto especialmente para este fin. Una vez detectada la cantidad de errores en la revisión de encuestas, se realizan recomendaciones con el propósito de mejorar en el proceso de recolección de datos.

# Procesamiento de datos

## Edición de datos

### CONSOLIDACIÓN DE ARCHIVOS DE DATOS

El área de Sistemas descarga los datos provenientes de las direcciones territoriales y los consolida en la base de datos Oracle los lunes, miércoles y viernes. Solamente en la última semana de recolección de cada etapa, la consolidación de la información se realiza de forma diaria. Este procedimiento se hace a través de la herramienta Pentaho, en la cual se ha diseñado un proceso de lectura y cargue de la información.

Posterior al cargue, se generan los primeros informes que corresponden a cobertura y se dispone la base de datos al área temática con el fin de revisar la información y detectar inconsistencias. De requerirse alguna modificación, se les solicita a las direcciones territoriales, por medio de la Dirección de Recolección y Acopio que se verifiquen y corrijan los datos inconsistentes.

Luego de realizados los ajustes, se informa al área de Sistemas para que actualice la información en la base de datos. Durante toda la etapa de recolección se realizan monitoreos por parte de todas las áreas involucradas (recolección y acopio, temática y sistemas) para garantizar la consistencia de la información.

### CODIFICACIÓN

En los capítulos de Ocupados, No ocupados y Trabajo Infantil de la encuesta, se realizan preguntas acerca del oficio que realiza la persona encuestada y la rama de actividad de la empresa donde trabaja, las cuales se capturan en campo como un texto abierto y por lo tanto requieren de codificación manual. Para realizar esta labor, se cuenta con un aplicativo a través del cual se le dispone, a las personas encargadas de asignar los códigos correspondientes, variables como la edad, el sexo, el nombre de la empresa, y el nivel educativo entre otras, con el fin de que los codificadores del DANE revisen la descripción registrada por los recolectores en las preguntas: ¿Qué hace..en este trabajo? y ¿A qué actividad se dedica principalmente la empresa o negocio en la que . realiza su trabajo?

Posteriormente, y luego de analizar dicha información se asigna el código correspondiente para cada caso. En el momento de diligenciar las preguntas que tienen como finalidad identificar la rama, se hace una descripción de la rama de actividad de la empresa con el propósito de que la codificación posterior sea más precisa.

Finalmente, el personal encargado de codificar la variable "rama" consulta la Clasificación Industrial Internacional Uniforme de todas las Actividades Económicas (CIIU Rev. 4 A.C.), para identificar la rama de actividad a la cual pertenece. Asimismo, se realiza la codificación de las respuestas relacionadas con las preguntas que indagan por la ocupación, el oficio o labor desempeñados haciendo uso de la Clasificación Internacional Uniforme de Ocupaciones (CIUO Rev. 8 A.C.). Una vez se ha identificado el código a asignar, se procede a ingresarlo en la base de datos a través del aplicativo en las variables correspondientes.

Durante el proceso de recolección y como parte de los procesos de revisión de la información recopilada, la oficina de Sistemas genera los archivos de Incodificables y de Revisión de ocupados, uno a mitad de la etapa y otro al finalizar esta.

El archivo de incodificables corresponde a registros en los cuales el texto registrado por el encuestador no es suficiente para determinar el código de la clasificación que debe ser asignado. Para la corrección de estos casos.

se hace una revisita a los hogares y se precisa la información con la persona encuestada y posteriormente se hace el reenvío de la encuesta corregida a DANE central.

El archivo de revisión de ocupados se genera para determinar si la codificación quedó bien asignada. El GIT de temática de mercado laboral realiza la verificación de los datos teniendo en cuenta el contraste de las variables codificadas con otras variables. De encontrarse inconsistencia en la información, se solicita a la oficina de Sistemas realizar el ajuste correspondiente y se le informa a la Dirección de Recolección y Acopio los cambios efectuados, para socialización con los codificadores.

### DICCIONARIO DE DATOS

Se cuenta con el diccionario de datos de la GEIH para facilitar el manejo de las bases que se publican y está dirigido tanto a usuarios internos como externos. El diccionario contiene el código y las variables correspondientes a cada una de las preguntas del formulario.

## REVISIÓN Y VALIDACIÓN

Se revisa el diligenciamiento completo de la información en las bases de datos, por lo cual se asegura que cada registro cuente con la información de la encuesta de los módulos de Identificación, Registro de personas y Tipo de Identificación. Esta información permite ubicar a los informantes o fuentes de datos para realizar verificaciones, de manera que se validan rangos de acuerdo con la estructura de las preguntas y flujo o salto, dependiendo de las opciones.

Adicionalmente, el GIT de Gestión de Datos de la oficina de Sistemas aplica el cruce de variables establecido previamente por el GIT de Temática, generando así los cuadros de salida expandidos que se envían a los GIT de Temática y Diseños muestrales para revisión de forma y consistencia de la información. En caso de encontrar inconsistencias, se informa al GIT de Gestión de Datos para que se realicen los respectivos cambios y/o reprocesamiento.

## DISEÑO DE INSTRUMENTOS DE EDICIÓN ( VALIDACIÓN Y CONSISTENCIA) E IMPUTACIÓN DE DATOS

La verificación de la consistencia interna de los datos y ajustes es parte del procesamiento de datos. 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, o toma valor 1, cuando la encuesta debe ser revisada por datos faltantes, inválidos o inconsistentes.

En esta operación estadística no se realiza el proceso de imputación. Cuando los miembros de los hogares se niegan a responder la encuesta, esta falta de respuesta es ajustada con los factores de expansión.

## DISEÑO PARA LA GENERACIÓN DE CUADROS DE RESULTADOS

Se realizan dos fases para generar los cuadros de salida:

Primera fase: a nivel de base de datos, se conforman vistas y vistas materializadas con los principales agregados requeridos por los cuadros de acuerdo con 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, 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 diseños muestrales para calcular los coeficientes de variación.

## Estimación de datos

## Estimación de error de la muestra

### MÉTODOS DE ANÁLISIS

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 sirven de base para la formulación, seguimiento y evaluación de políticas.

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

- Análisis descriptivo. Ayuda a observar el comportamiento de la muestra en estudio, por medio de tablas, gráficos, intervalos de confianza y estadísticas de tendencia y dispersión. Se analiza en la muestra la estructura de los principales indicadores de mercado laboral por dominios (total nacional, cabeceras, centros poblados y rural disperso, etc.) de estudio a partir de la distribución de frecuencias y se detectan posibles inconsistencias y valores atípicos, teniendo en cuenta la tendencia de las series y la estacionalidad de los datos. Se verifica la cobertura a nivel de segmentos, viviendas, hogares y personas, y se observa la distribución de la pérdida de muestra para realizar los respectivos ajustes por la no respuesta.

- 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 establecida 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, que para el caso de la GEIH hace referencia a que toda variable cuya proporción respecto a la Fuerza de Trabajo (FT) sea menor al 10%, tiene un error de muestreo superior al 5%, que es el nivel de calidad admisible para el DANE.

El grupo temático se encarga de dar inicio al proceso de producción de las estadísticas de mercado laboral con la recepción de los cuadros de salida: se realizan análisis comparativos de las tasas y poblaciones entre períodos de los diferentes años y se analiza la consistencia de la información por dominios geográficos y demás variables, según los conceptos de fuerza de trabajo.

Por último, el equipo temático genera boletines técnicos, anexos estadísticos y presentaciones con los principales resultados de mercado laboral. Productos mediante los cuales el DANE difunde las estadísticas de mercado laboral.

### ANONIMIZACIÓN DE MICRODATOS

Acorde con el sexto principio de la Oficina de Estadística de las Naciones Unidas, referente a la Confidencialidad de la información, que indica "Los datos individuales recolectados sobre personas naturales o entidades legales, o sobre pequeñas agremiaciones 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>19</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 al 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 y 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, Centro Poblado.

Estas variables son absolutamente confidenciales, no se manejan en las bases de datos de acceso público, ni 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 Y POBLACIÓN FUERA DE LA FUERZA DE TRABAJO

Las variables objeto de anonimización son: nombre de la empresa, negocio, industria, oficina, firma o finca donde trabaja, que permiten determinar si la empresa es pública o privada mediante el procesamiento interno.

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.

<sup>19</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)

### VERIFICACIÓN DE LA ANONIMIZACIÓN DE MICRODATOS

El DANE ofrece los niveles apropiados de Seguridad de la Información para la generación y el acceso a los microdatos anonimizados. La metodología de anonimización de la GEIH marco 2018 es definida por el grupo de temática de mercado laboral, quien genera un protocolo de anonimización para identificación de las variables que deben ser eliminadas de las bases de datos que serán objeto de publicación.

La Oficina de Sistemas es el área encargada de aplicar estas metodologías para generar archivos anonimizados, los cuales son almacenados en servidores ubicados en el centro de cómputo en un ambiente seguro. De tal manera que, se hace uso de una infraestructura de seguridad y se aplican niveles de acceso como: autenticación, autorización y registro de las actividades, con lo cual se ofrece integridad, disponibilidad y confidencialidad de la información. Posteriormente, tanto el equipo de temática como el de sistemas realizan la verificación de la implementación del protocolo en la base de datos anonimizada para su posterior publicación en página web.

El protocolo de anonimización establece los criterios utilizados para garantizar que la información entregada al público mantenga la reserva y no permita la identificación de la fuente. En este documento, se especifican las variables que debe contener cada capítulo y las variables que deben ser eliminadas de las bases de datos.

Las variables que son absolutamente confidenciales, no se manejan en las bases de datos de acceso público, ni tampoco se encuentran dentro de las bases de datos utilizadas al interior del DANE para la generación de estadísticas básicas.

Para los casos donde se requiere entregar alguna de las variables anonimizadas, la solicitud es remitida al comité de Reserva Estadística para evaluación de la viabilidad de entrega de la información requerida, teniendo en cuenta los acuerdos de confidencialidad establecidos.

### COMITÉS DE EXPERTOS

La coordinación del GIT de temática de Mercado Laboral convoca mensualmente a un comité técnico interno y a uno externo con el fin de presentar los resultados obtenidos y analizar el comportamiento de los principales indicadores de mercado laboral. Estos comités contextualizan a los usuarios de la información y son espacios que permiten generar discusiones y resolver inquietudes frente al comportamiento de los indicadores.

De acuerdo con la normatividad vigente se debe citar un comité interno y un comité externo. El comité interno se debe citar tres (3) días hábiles antes de la publicación de mercado laboral y está compuesto por los asesores de las direcciones técnicas del DANE, como la Dirección de Síntesis y Cuentas Nacionales, la Dirección de Difusión y Cultura Estadística, la Dirección de Recolección y Acopio, la Dirección de Regulación, Planeación, Estandarización y Normalización, así como los profesionales que hacen parte del Grupo Interno de Trabajo de Diseños Muestrales, Pobreza y Sistemas. Por su parte, el comité externo está compuesto por asesores de algunos organismos del Gobierno Nacional, como el Banco de la República, los ministerios y departamentos administrativos, entre otros.