

Departamento Administrativo
Nacional de Estadística



Producción Estadística
PES

Dirección de Metodología y Producción Estadística / DIMPE

METODOLOGÍA GENERAL GRAN ENCUESTA INTEGRADA DE HOGARES GEIH

Jun/2023



	METODOLOGÍA GENERAL GRAN ENCUESTA INTEGRADA DE HOGARES GEIH	CÓDIGO: DSO-GEIH-MET-001 VERSIÓN: 11 FECHA: 16/Jun/2023
PROCESO: Producción Estadística		OPERACIÓN ESTADÍSTICA: GEIH - GRAN ENCUESTA INTEGRADA DE HOGARES

TABLA DE CONTENIDO

INTRODUCCIÓN

1. ANTECEDENTES

2. DISEÑO DE LA OPERACIÓN ESTADÍSTICA

2.1. DISEÑO TEMÁTICO

2.1.1. Necesidades de Información

2.1.2. Formulación de objetivos

2.1.3. Alcance

2.1.4. Marco de referencia

2.1.5. Definición de variables y construcción de indicadores estadísticos

2.1.6. Plan de resultados

2.1.7. Estándares estadísticos utilizados

2.1.8. Diseño del cuestionario

2.1.9. Normas, especificaciones o reglas de edición e imputación de datos

2.2. DISEÑO ESTADÍSTICO

2.2.1. Universo de estudio

2.2.2. Población objetivo

2.2.3. Cobertura geográfica

2.2.4. Desagregación geográfica

2.2.5. Desagregación temática

2.2.6. Fuentes de datos

2.2.7. Unidades estadísticas

2.2.8. Periodo de referencia

2.2.9. Periodo de recolección/acopio

2.2.10. Marco estadístico (censal o muestral)

2.2.11. Diseño muestral

2.2.12. Ajustes de cobertura (o ajuste de cobertura por no respuesta)

2.2.13. Especificaciones de ponderadores

2.3. DISEÑO DE LA RECOLECCIÓN/ ACOPIO

2.3.1. Métodos y estrategias de recolección o acopio de datos

2.3.2. Estructura organizacional del operativo y conformación del equipo

2.3.3. Esquema de entrenamiento de personal

2.3.4. Invitación pública de selección de personal

2.3.5. Proceso de sensibilización y acuerdos de intercambio

2.3.6. Elaboración de manuales

2.3.7. Diseño de las estrategias de comunicación y plan de contingencias

2.3.8. Diseño de la estrategia de seguimiento y control

2.3.9. Diseño de sistemas de captura

2.3.10. Transmisión de datos

2.4. DISEÑO DE PROCESAMIENTO

2.4.1. Consolidación de archivos de datos

2.4.2. Codificación

2.4.3. Diccionario de datos

2.4.4. Revisión y validación

2.4.5. Diseño de instrumentos de edición (validación y consistencia) e imputación de datos

2.4.6. Diseño para la generación de cuadros de resultados

2.5. DISEÑO DEL ANÁLISIS

2.5.1. Métodos de análisis

2.5.2. Anonimización de microdatos

2.5.3. Verificación de la anonimización de microdatos

2.5.4. Comités de expertos

2.6. DISEÑO DE LA DIFUSIÓN Y COMUNICACIÓN

2.6.1. Diseño de sistemas de salida

2.6.2. Diseño de productos de comunicación y difusión

2.6.3. Entrega de productos
2.6.4. Estrategia de servicio
2.7. DISEÑO DE LA EVALUACIÓN DE LAS FASES DEL PROCESO
2.8. DISEÑO DE LOS SISTEMAS DE PRODUCCIÓN Y FLUJOS DE TRABAJO
3. DOCUMENTACIÓN RELACIONADA
GLOSARIO
BIBLIOGRAFÍA
ANEXOS

INTRODUCCIÓN

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.

1. ANTECEDENTES

La recolección de información sobre mercado laboral por medio de encuestas a hogares en Colombia inició en 1967 con la aplicación de encuestas de fuerza de trabajo y de ingresos y gastos por parte del Centro de Estudios sobre Desarrollo Económico (CEDE) de la Universidad de los Andes. Posteriormente, el DANE inició la aplicación ininterrumpida de estas encuestas de manera oficial desde la década de los 70. La experiencia institucional del DANE en la aplicación de encuestas a hogares se enuncia a continuación:

Encuesta Nacional de Hogares

Entre 1970 y 1975, tuvo lugar el primer periodo de desarrollo de las encuestas de hogares por parte del DANE: se realizaron diez encuestas con coberturas geográficas diferentes, en las cuales se indagaron temas como condiciones de la vivienda, ingresos y gastos y fecundidad, con preguntas adicionales a las del formulario básico sobre las características demográficas y de fuerza de trabajo de la población. Los criterios de periodicidad, cobertura y diseño muestral se unificaron a partir de marzo de 1976. En 1978, se realizó la primera medición en áreas rurales, que se replicaría en 1988. Entre 1976 y 1984, se llevó a cabo la Encuesta Nacional de Hogares (ENH), la cual inició con aplicación trimestral para cuatro ciudades y semestral para tres más. paulatinamente, se incorporaron nuevos dominios geográficos de estudio, hasta llegar en el año 2000 a siete ciudades y áreas metropolitanas definidas para la encuesta, así: Medellín - Valle de Aburrá, conformada por los municipios de Barbosa, Bello, Caldas, Copacabana, Envigado, Girardota, Itagüí, La Estrella y Sabaneta; Cali - Yumbo; Barranquilla - Soledad; Bucaramanga - Floridablanca, Girón y Piedecuesta; Manizales - Villamaría; Pereira - Dosquebradas y La Virginia; Cúcuta - Villa del Rosario, Puerto Santander, Los Patios y El Zulia, con recolección trimestral y a seis más con otras periodicidades, incluyendo las áreas rurales.

En 1988, se realizó la primera Encuesta Nacional de Hogares rural en 148 municipios distribuidos entre las regiones del país, con el objeto de estudiar a la población rural de los centros poblados y rural disperso en los siguientes temas: fuerza de trabajo, características demográficas, indicadores de pobreza y tenencia de parcelas por parte de los hogares. La encuesta supuso un importante cambio conceptual de la ruralidad, pues pasó de entenderla solo como población dispersa, a contemplar la franja de población ubicada entre la población dispersa y las 150 cabeceras municipales más grandes del país.

A partir de marzo de 1990, se presentaron resultados representativos para las siete ciudades con el área metropolitana definida por la encuesta. En diciembre de 1991, se inició la medición nacional. Esta medición continuó todos los años en los meses de septiembre a partir de 1992. En 1996, se amplió la muestra, lo que permitió presentar resultados a nivel de 24 departamentos.

Encuesta Continua de Hogares

En 1998, se analizó la adopción de un sistema de recolección continuo y la actualización de la metodología de la ENH, a partir de las resoluciones aprobadas por la Oficina de la Organización Internacional del Trabajo en 1983, el nuevo Sistema de Cuentas Nacionales de 1993 (SCN-93), experiencias y prácticas internacionales y, especialmente, las particularidades del mercado laboral colombiano.

Para el desarrollo del proyecto de la Encuesta Continua de Hogares (ECH), el DANE contó con la asistencia técnica de la OIT- Ginebra a través del experto internacional Sr. Ralf Hussmans y de un Comité Técnico de Empleo. Se examinaron los alcances y limitaciones del sistema colombiano y las ventajas adicionales de un modelo alternativo de recolección de datos básicos, de manera que se facilitara captar el comportamiento integral de los fenómenos durante todas las semanas del año, para que así fuera más real, objetivo y representativo de las variables estudiadas y de forma particular, las laborales. Además, se buscaba acortar los periodos de estimación y publicación de resultados, mejorar la calidad de las mediciones de campo y reducir los costos de aplicación de la encuesta.

En enero de 1999, se inició en Bogotá la aplicación de la Encuesta Continua de Hogares (ECH) como prueba piloto que se extendió a partir del segundo semestre a Medellín y Cali, y desde septiembre a Barranquilla, Bucaramanga, Manizales y Pasto. Una vez realizados los análisis correspondientes a la prueba piloto, se aplicó la ECH en forma definitiva a partir de enero de 2000 en las siguientes ciudades con las áreas metropolitanas¹ definidas para la encuesta: Bogotá, Cali - Yumbo, Medellín - Valle de Aburrá conformada por los municipios de Barbosa, Bello, Caldas, Copacabana, Envigado, Girardota, Itagüí, La Estrella y Sabaneta, Barranquilla - Soledad, Bucaramanga - Floridablanca, Girón y Piedecuesta, Manizales - Villamaría, Pasto, Pereira - Dosquebradas y La Virginia, Cúcuta - Villa del Rosario, Los Patios, Puerto Santander y El Zulia, Villavicencio, Ibagué, Montería y Cartagena. Esta operación estadística incorporó los desarrollos metodológicos y conceptuales vigentes de la Conferencia Internacional de Estadísticos del Trabajo (CIET) y las dinámicas del mercado laboral colombiano.

Durante el año 2000, se realizaron en paralelo la Encuesta Nacional de Hogares (encuesta tradicional de aplicación trimestral) y la Encuesta Continua de Hogares (operación estadística de aplicación permanente) para proporcionar información que permitiera analizar las diferencias en los resultados de las dos encuestas e identificar los criterios básicos para el empalme de las series. Lo anterior, permitió afinar los procesos y los procedimientos operativos, armonizar las encuestas urbanas y rurales y definir el sistema de procesamiento y producción de resultados, para así determinar la cantidad de indicadores y la periodicidad con la cual debían publicarse.

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

Gran Encuesta Integrada de Hogares

En 2005, se propuso la integración de las tres encuestas a hogares más importantes del DANE: la Encuesta Continua de Hogares (ECH), la Encuesta Nacional de Ingresos y Gastos (ENIG) y la Encuesta de Calidad de Vida (ECV), con el propósito de ampliar el alcance temático de la operación estadística y reducir el costo de aplicación. Como resultado de esta propuesta, tan solo se pudieron aplicar algunos módulos de la ECV y la ENIG, por lo cual, la integración de estas encuestas no fue efectiva en su totalidad. Este ejercicio permitió el enriquecimiento temático de cada una, sin afectar la calidad en la recolección de los datos. Las mejoras desarrolladas en este proceso contemplaron la implementación de nuevos desarrollos logísticos e innovaciones como la introducción del informante (entrevistado/encuestado) directo, que permitió realizar una mayor indagación sobre variables

claves como los ingresos y las condiciones de trabajo de las personas y la captura de información mediante dispositivos móviles (DMC), lo que permitió una mayor oportunidad en la producción de resultados.

En 2006, se inició el diseño de la Gran Encuesta Integrada de Hogares (GEIH) a partir de la revisión de experiencias internacionales y con el propósito de cumplir el alcance temático de las encuestas a hogares. Para el desarrollo de este proyecto se contó con el apoyo del *Programa regional para el mejoramiento de las encuestas y la medición de las condiciones de vida en América Latina y el Caribe* (MECOVI) del Banco Interamericano de Desarrollo (BID), el Banco Mundial y la Comisión Económica para América Latina y el Caribe (CEPAL) de las Naciones Unidas. El apoyo financiero del Banco Mundial se centró en los diseños temático, muestral y logístico, así como en el entrenamiento y en las pruebas de campo para avanzar en la estrategia de integración de las encuestas.

A partir del 10 julio de 2006, se amplió la cobertura de la ECH a 11 ciudades capitales: Tunja, Florencia, Popayán, Valledupar, Quibdó, Neiva, Riohacha, Santa Marta, Armenia, Sincelejo y San Andrés, que contaban con las mismas condiciones de representatividad de las trece ciudades con las áreas metropolitanas definidas para la encuesta.

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

Entre abril y diciembre de 2008, se aplicó la encuesta en la muestra diseñada de forma paralela, a partir del marco muestral del Censo 2005. En enero de 2009, se inició la aplicación de la GEIH con el nuevo marco de muestreo, el cual utilizó las proyecciones de población que surgieron del Censo General de 2005 para las estimaciones de los dominios.

La GEIH es el resultado de un marco estructural del Sistema de Encuestas Sociales en un lapso de diez años (período intercensal), donde en el primer año se integró la Encuesta Continua de Hogares (ECH) con la Encuesta Nacional de Ingresos y Gastos (ENIG) y, posteriormente, se integró la ECH con la Encuesta de Calidad de Vida (ECV). En las etapas siguientes, se implementaron módulos periódicos para profundizar en el estudio de fenómenos específicos o para monitorear fenómenos que se midieron en profundidad en la GEIH el primer año.

Mediante los módulos incorporados a la encuesta, en 2006 se recopilaron datos sobre Otros ingresos, Turismo y Negocios y actividades agrícolas, pecuarias, forestales y mineras. A partir del segundo trimestre de 2007, se incluyeron los módulos de Televisión y Tecnologías de información y comunicación, y en el cuarto trimestre de este año se aplicó el módulo de Trabajo infantil². En 2008, ingresó de forma permanente el módulo de Fecundidad; posteriormente, en las diferentes etapas de aplicación de la encuesta, se incluyeron los módulos de Empleo rural, Consumo de arroz y Precepción sobre los servicios y los mecanismos de atención al ciudadano. En el tercer trimestre de 2009, se incorporó el módulo de Impunidad y acceso a la justicia. En 2012, comenzó la aplicación continua del módulo de Migración, y en 2013 el de Micronegocios. Este mismo año se incorporó Formación para el trabajo, con recolección periódica en el segundo trimestre de cada año y Tipo de identificación en el cuarto trimestre, éste último con aplicación permanente desde el último trimestre de 2017. En el segundo semestre de 2015, se aplicó el módulo de Limitaciones en el funcionamiento humano.

Acorde con los Principios Fundamentales de las Estadísticas Oficiales de la ONU (Resolución 68/261 de 2014) y en línea con el noveno principio referente a la utilización de normas internacionales, el DANE realizó la actualización de parámetros con la revisión estadística correspondiente, sustentada a través de la Encuesta de Microestablecimientos, donde se encontró evidencia de que las empresas que ocupan entre seis y nueve empleados generan empleo formal.

De esta forma, en diciembre de 2009, se comenzó a usar la Resolución 3 sobre estadísticas de ocupación en los sectores informales adoptada de la 15a CIET de 1993 de la Organización Internacional del Trabajo (OIT). A partir de esta, se cambió la escala de tamaño de empresa para definir si un ocupado es formal o informal, pasando de 10 a 5 trabajadores. De esta forma, las personas que trabajaban en empresas de 5 o menos trabajadores eran consideradas informales. Con la adaptación realizada, la ocupación informal se media por las características de las unidades de producción (empresas), más que por las características de las personas o de los puestos de trabajo.

A partir del cuarto trimestre de 2012, se amplió la cobertura de la GEIH a las ocho ciudades capitales de los departamentos de la Amazonía y la Orinoquía, con recolección de información trimestral en Arauca, Leticia, Inírida, Mitú, Mocóa, San José del Guaviare, Vichada y Yopal. En 2013, la aplicación de la encuesta se hizo en un período de nueve meses, comprendido entre los meses de abril y diciembre. En 2014, se modificó la periodicidad de recolección para estas ocho ciudades, aplicando la encuesta de forma continua de enero a diciembre. De esta manera, la encuesta alcanzó cobertura en las 32 principales ciudades capitales de departamento, cabeceras y centros poblados y rural disperso de más de 443 municipios del país, en los que se encuestaban 58.200 hogares cada trimestre, aproximadamente. Con esta ampliación de la muestra, se logró cubrir la totalidad de las ciudades capitales de los departamentos del país.

A partir de abril de 2015, se inició la recolección de información de la GEIH en Buenaventura. En enero de 2018, se empezó el levantamiento de información en Barrancabermeja y Soacha, en tanto que para Rionegro (Antioquia) y Tumaco inició en 2020. Todos los anteriores dominios permiten obtener resultados con representatividad anual.

En 2019, la GEIH modificó de manera transitoria el operativo de recolección debido a la coyuntura nacional ocasionada por el COVID-19. En tal sentido, se ajustó el formulario inicial que se aplicaba de manera presencial, reduciendo su tamaño, y realizando la captura de información de manera telefónica con el fin de garantizar la continuidad, consistencia y calidad de la información recolectada.

Rediseño Gran Encuesta Integrada de Hogares (Marco 2018)

A finales de 2018 se propuso el rediseño de la GEIH en términos conceptuales, operativos y muestrales. Este proceso contempló la inclusión de innovaciones y cambios, tales como la reformulación del objetivo general y los objetivos específicos; mejorar la medición y el análisis de los indicadores de mercado laboral, de los ingresos y la pobreza monetaria; la actualización del marco y el diseño de la muestra, de acuerdo con los nuevos objetivos y los cambios ocasionados por la actualización censal.

En 2019, se contó con la consultoría de tres expertos nacionales en pobreza y desarrollo, y diseños muestrales, así como con la asesoría de expertos internacionales en estadísticas de mercado laboral de la Organización Internacional del Trabajo (OIT). De esta manera, se llevó a cabo la revisión conceptual y metodológica de la encuesta y se dio inicio al proceso de rediseño de la operación estadística, con base en las recomendaciones de la OIT emitidas a través de las resoluciones 17ª, 19ª y 20ª de la Conferencia Internacional de Estadísticos del Trabajo (CIET).

El objetivo de la actualización de la GEIH fue adaptar la encuesta a los nuevos requerimientos del país en cuanto al fortalecimiento de la medición del mercado laboral y la pobreza monetaria; así como identificar grupos poblacionales minoritarios como son la población campesina, las personas lesbianas, gays, bisexuales, transgénero e intersexuales (LGBTI) y las personas en condición de discapacidad (PCD). Otro aspecto importante que se tuvo en cuenta para el rediseño fue la actualización del marco muestral con base en los resultados del Censo Nacional de Población y Vivienda 2018 (CNPV).

En 2020, se adelantaron los procesos de diseño y construcción dentro de los diferentes componentes que tiene la GEIH, como son el temático, el estadístico, el de recolección, y el de difusión y comunicación. De esta forma, se diseñaron nuevas preguntas, se elaboró el nuevo formulario, se generaron los manuales, se definieron los indicadores y el plan de pruebas requerido. Asimismo, se realizó la estimación del tamaño de la muestra, la estratificación y selección de municipios y el diseño muestral tanto de la prueba piloto como del diseño experimental que dieron como resultado la encuesta actual. Frente a los procesos de recolección y difusión y comunicación de la encuesta, se definieron lineamientos operativos para la prueba piloto y diseños experimentales, con pruebas al sistema de monitoreo y control. De igual manera, se diseñó la estrategia de comunicación general.

Entre los años 2020 y 2021, se implementó el Sistema General de Pruebas del rediseño de la GEIH, con el propósito de garantizar una transición adecuada entre la GEIH tradicional (GEIH marco 2005) y su versión actualizada con todas las temáticas del rediseño (GEIH marco 2018). El Sistema de Pruebas implementado, contempló la realización de cinco etapas sucesivas de validación de la operación estadística que incorporaron diferentes instrumentos y escenarios de aplicación y prueba del nuevo formulario: pruebas de escritorio, pruebas de simulación, prueba piloto, diseños experimentales y paralelo.

Cada una de las etapas definidas dentro del Sistema General de Pruebas³ permitió identificar las dificultades y fortalezas que presentaba el proceso de recolección con el nuevo instrumento. Los aspectos evaluados fueron: tiempo de recolección; funcionamiento del DMC; claridad del fraseo tanto de las preguntas como de las opciones de respuesta; actitud del encuestador y del informante directo; y eficiencia de los procesos de consolidación, sincronización y transmisión de la información recolectada.

Los objetivos de la prueba piloto estuvieron orientados a verificar el tiempo de diligenciamiento del formulario rediseñado el cual contaba con 58 preguntas nuevas y 22 modificadas, implementar planes de datos para la transmisión de la información de campo, probar la efectividad de establecer controles automatizados para el seguimiento al desarrollo del operativo y estimar los impactos de la implementación del Marco censal 2018 sobre los indicadores temáticos de la encuesta. Asimismo, el objetivo del diseño experimental era evaluar si la inclusión de nuevas preguntas en los módulos del formulario relativos a mercado laboral, ingresos e identidad de género, afectaban los indicadores publicados con la GEIH marco 2005. El desarrollo de estas pruebas permitió concluir que los resultados de los indicadores de mercado laboral no se vieron afectados por las modificaciones realizadas al formulario.

Los cambios incorporados en la operación estadística consistieron en la reformulación del objetivo general el cual ya no hace referencia al tamaño y estructura de la fuerza de trabajo de la población del país, sino a información estadística sobre mercado laboral, ingresos y pobreza monetaria, y la caracterización sociodemográfica de la población residente en Colombia, asimismo, se fusionaron algunos de los objetivos específicos relacionados con la medición del empleo, en tanto que los demás, se rediseñaron en línea con el nuevo alcance de la encuesta. Adicionalmente, se incluyeron preguntas para la identificación de grupos poblacionales priorizados en la superación de la invisibilidad estadística de estas comunidades y el robustecimiento de la medición y el análisis de los indicadores de mercado laboral, de los ingresos y la pobreza monetaria; la actualización de la medición de la Población en Edad de Trabajar - PET a partir de los 15 años; y del diseño muestral con base en el marco muestral del Censo Nacional de Población y Vivienda - CNPV 2018.

En enero de 2021, con el objetivo de controlar el efecto del nuevo marco sobre los indicadores de mercado laboral y de identificar si se presentaban afectaciones en la tasa de respuesta de la encuesta, así como, en la consistencia de los datos capturados, se realizó una encuesta paralela, la cual consistió en la recolección de información con la aplicación del formulario rediseñado de la GEIH de forma simultánea a la Gran Encuesta Integrada de Hogares tradicional (GEIH marco 2005); en diciembre de 2021 se terminó la recolección en paralelo. Este ejercicio sirvió de base para el empalme, que se realizaría posteriormente a los principales indicadores de mercado laboral (tasa global de participación, tasa de ocupación, tasa de desempleo, tasa de subocupación, ocupados, desocupados, población fuera de la fuerza de trabajo (inactivos) y subocupados).

En 2022, y luego de finalizada la recolección y acopio de la información, se realizó un proceso de validación donde los resultados de la encuesta tradicional (GEIH marco 2005) y la encuesta paralela (GEIH marco 2018) con el nuevo marco fueron contrastados, identificando que sistemáticamente se presentaba una diferencia entre la población ocupada y la población fuera de la fuerza de trabajo entre ambas encuestas, sin mostrar cambios significativos en los desocupados.

Esta situación evidenció un punto de inflexión en las series de tiempo, afectando la comparabilidad de esta a partir del año 2021 respecto a los anteriores periodos, por lo que fue necesario calcular un factor de empalme para las series de ocupados y fuera de la fuerza de trabajo, desde el año 2010 hasta el 2020, con lo cual se definieron unos techos poblacionales que luego se incluyeron en el modelo de calibración tradicional de la encuesta⁴.

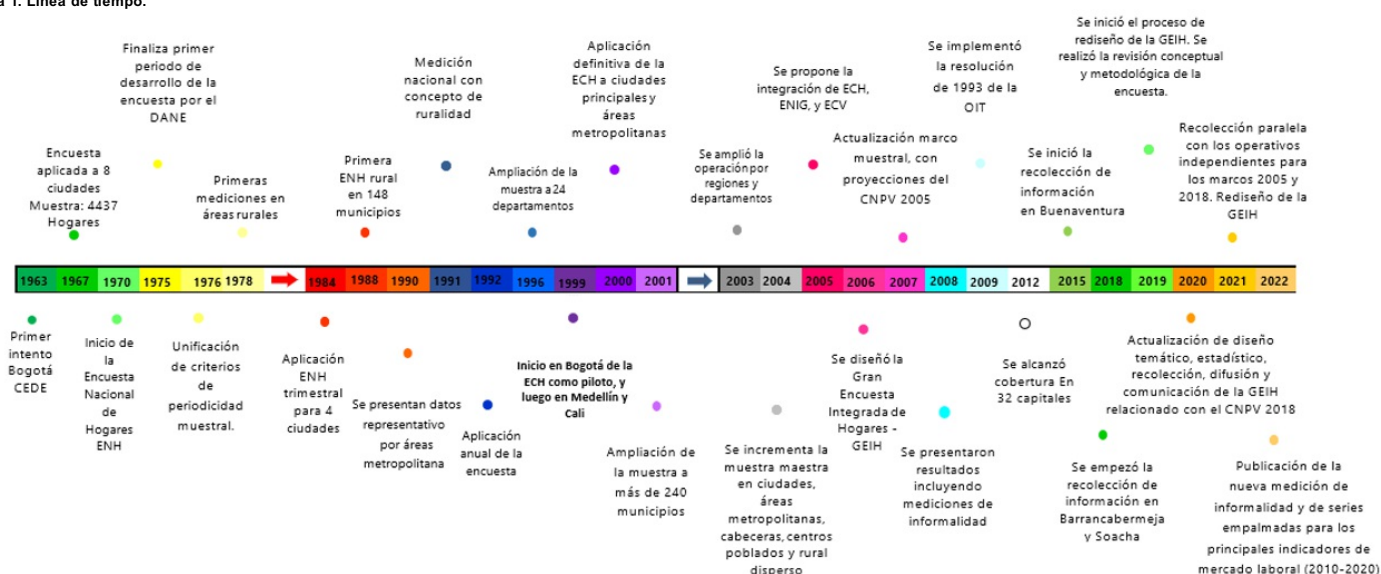
A partir de enero de 2021, los resultados de mercado laboral se producen con la información recolectada por la GEIH marco 2018. Dentro de los principales productos generados se encuentran las series de poblaciones e indicadores de mercado laboral desde 2001 hasta 2021 calculados a partir de los nuevos factores de expansión calibrados con base a las proyecciones y retroproyecciones de población del CNPV 2018; y las series de poblaciones e indicadores de mercado laboral empalmadas desde enero 2010 hasta diciembre de 2020.

Asimismo, con los datos recopilados a través de la GEIH marco 2018 se generan indicadores de mercado laboral para la población en condición de discapacidad, autorreconocimiento campesino y LGBTI y se adoptó la nueva definición para la medición de la ocupación informal, la cual se basa en la resolución de la 17ª CIET de la OIT (2003) y las recomendaciones del grupo de DELHI sobre las estadísticas del sector informal.

Actualmente, la encuesta cuenta con una muestra anual aproximada de 315.000 hogares a nivel nacional.

En la figura 1 se sintetizan los antecedentes de la encuesta, a manera de línea de tiempo:

Figura 1. Línea de tiempo.



Fuente: DANE.

¹ Con el propósito de dar continuidad a las series históricas de mercado laboral, la Encuesta Continua de Hogares y la Gran Encuesta Integrada de Hogares han mantenido la desagregación geográfica establecida en la selección de la muestra de la Encuesta Nacional de Hogares, por tanto, las áreas metropolitanas definidas para la operación estadística difieren de las legalmente constituidas actualmente.

² El DANE, en colaboración con el Instituto Colombiano de Bienestar Familiar (ICBF), el Ministerio del Trabajo y el Programa de Información Estadística y Monitoreo en Materia de Trabajo Infantil (SIMPOC) de la Oficina Internacional del Trabajo (OIT), realizaron en 2001 la primera Encuesta Nacional de Trabajo Infantil. Desde entonces, el DANE realizó el seguimiento cada dos años a los principales indicadores de trabajo infantil mediante módulos aplicados en el cuarto trimestre de la Encuesta Continua de Hogares (ECH) de 2003 y 2005, y de la GEH de 2007 y 2009. En 2011, nuevamente se aplicó la encuesta en profundidad y, a partir de 2012, se empezó a realizar el seguimiento anual mediante un módulo incluido en la GEH en el cuarto trimestre. En enero de 2023, se comienza a recolectar de manera continua durante todos los meses del año.

³ Para más información ver el cuadro anexo sobre el sistema General de Pruebas al rediseño de la GEH.

⁴ Para más información sobre el empalme de las series de mercado laboral lo invitamos a consultar el siguiente enlace: https://www.dane.gov.co/files/Investigaciones/boletines/ech/ech/Empalme_%20GEH_%202005-2018.pdf

2. DISEÑO DE LA OPERACIÓN ESTADÍSTICA

2.1. DISEÑO TEMÁTICO

2.1.1. 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 13⁵ 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).

2.1.2. Formulación de objetivos

a) Objetivo general

Proveer información estadística relacionada con mercado laboral, ingresos y pobreza monetaria, así como de las características sociodemográficas de la población residente en Colombia.

b) Objetivos específicos

1. Generar los principales indicadores de mercado laboral: ocupación, desocupación y población fuera de la fuerza de trabajo residente en Colombia.
2. Medir las características de la población residente en Colombia:
 - > Ocupada: temporalidad, subocupación, rama de actividad, ocupación u oficio, posición ocupacional, informalidad, entre otras.
 - > Desocupada: tiempo de búsqueda de empleo, rama de actividad, ocupación u oficio que desempeñó en su empleo anterior, entre otras.
 - > Fuera de la fuerza de trabajo: las razones de retiro o no participación en el mercado laboral.
3. Generar información estadística relacionada con el ingreso de los hogares, y con los indicadores de pobreza monetaria y pobreza monetaria extrema de la población residente en Colombia.
4. Medir las características sociodemográficas (sexo, estado civil, educación, salud y vivienda, entre otros) de la población residente en Colombia.

2.1.3. Alcance

Las temáticas por las cuales indaga la Gran Encuesta Integrada de Hogares 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 campesinado, personas con autorreconocimiento étnico, 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. Asimismo, la GEIH Marco 2018 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 de los hogares, la pobreza monetaria y la pobreza monetaria extrema de la población residente en el país.

Los resultados permiten la desagregación para el total nacional, total de las cabeceras municipales, total de los centros poblados y rural disperso, cada una de las 23 ciudades capitales y áreas metropolitanas, San Andrés, Buenaventura, Barrancabermeja, Soacha, Tumaco, Rionegro y las 8 ciudades capitales de los departamentos de la Amazonía y Orinoquía. La recolección se hace de forma ininterrumpida todos los meses del año.

2.1.4. Marco de referencia

I. Marco teórico

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

La Conferencia Internacional de Estadísticos del Trabajo (CIET)⁶ ha adoptado diferentes resoluciones y recomendaciones en materia de medición de las estadísticas laborales. En 1983, durante la 13a CIET, se estableció y normalizó el marco teórico y conceptual en el que se deben enmarcar las diferentes variables de la fuerza de trabajo: concepto y definición de cada una de las variables, períodos de referencia, actividad principal, reglas de prioridad de las variables, etc. Igualmente, se introdujo la diferenciación de la población actualmente activa de la habitualmente activa (en el caso colombiano, las mediciones se refieren a la primera).

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

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

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

La GEIH inició su diseño en diciembre de 2005 con un proyecto del grupo temático de la ECH sobre la aplicación de una encuesta integrada. Su objetivo fue obtener la mayor cantidad de información posible sobre un mismo hogar, de tal forma que pudiera ser utilizada para la realización de investigaciones desde diferentes perspectivas, facilitando un análisis más completo de las características económicas y sociales de la población residente en Colombia. Esto, también se tuvo en consideración para el rediseño de la GEIH.

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

Durante el proceso de rediseño de la GEIH, se incorporaron las recomendaciones de las conferencias de estadísticos del trabajo⁷ de sus ediciones 17ª, 19ª y 20ª, como se muestra en la figura 2.

Figura 2. Conferencias de estadísticos del trabajo (CIET) según etapa de implementación

Conferencia Internacional de Estadísticos del Trabajo - (CIET)	Estado Actual
CIET 13 (1982): Normas sobre la PEA, ocupados, desocupados, inactivos	
CIET 15 (1993): Resolución sobre la ocupación en el sector informal	Implementadas desde GEIH Marco 2005
CIET 16 (1998): Resolución medición del subempleo y situaciones de empleo inadecuado	
CIET 17 (2003): Directrices definición estadística del empleo informal	Implementadas en rediseño
CIET 19 (2013): Propone una definición más amplia de trabajo. Cambios en la medición de la fuerza de trabajo -> fuerza de trabajo potencial.	• Brindan un marco de medición para todas las formas de trabajo
CIET 20 (2018): Resoluciones sobre: Relaciones laborales, Migración laboral internacional y desajustes en la medición de la cualificación y las competencias de las personas ocupadas	• Proporcionan nuevos indicadores para contribuir al análisis de la dinámica del mercado laboral

Fuente: DANE.

➤ **CIET 17ª:** se adoptaron las directrices relacionadas con una definición estadística de la ocupación informal donde se contempla tanto la situación del sector informal (si la empresa donde trabaja el encuestado cuenta con registro tributario y/o lleva contabilidad), como la situación en el puesto de trabajo (si el empleado tiene seguridad social a causa de su relación laboral con la empresa que lo contrató). Para esto, dentro de la encuesta se agregaron preguntas en el capítulo de ocupados, que garantizan las variables necesarias para la aplicación de la definición de ocupación informal de la OIT y la medición de indicadores relacionados con empleo en el sector informal, empleo informal y empleo informal fuera del sector informal.

➤ **CIET 19ª:** se adoptaron las resoluciones e informes relacionados con la precisión del concepto de empleo, ampliación de las otras formas de trabajo, situación en la fuerza de trabajo, subutilización de la fuerza de trabajo y fuerza de trabajo potencial.

En cuanto al empleo, se precisó el concepto haciendo referencia únicamente al trabajo remunerado en dinero o en especie, considerando todo lo demás como trabajo fuera de la ocupación. Este ajuste implicó la inclusión de la pregunta: ¿La persona recibió remuneración en dinero o en especie a cambio de este trabajo? dentro del capítulo de fuerza de trabajo.

Respecto a la medición de otras formas de trabajo se incluyó la medición por separado de cuatro formas de trabajo específicas: (1) Empleo, como trabajo realizado a cambio de una remuneración o beneficios, (2) Trabajo en producción de bienes o servicios para el consumo final propio, (3) Trabajo voluntario no remunerado y no obligatorio y (4) Trabajo en formación. Esta profundización en la medición se ve reflejada en el ajuste del capítulo de otras formas de trabajo.

Con relación a la situación en la fuerza de trabajo, de acuerdo con el Informe II sobre Estadísticas de trabajo, ocupación y subutilización de la fuerza de trabajo, se cambió el concepto de población inactiva por población fuera de la fuerza de trabajo. Esto de acuerdo con los lineamientos del numeral 81, el cual reconoce que las personas pueden desempeñar diferentes formas de trabajo y todas las formas de trabajo implican insumos de mano de obra en actividades productivas, por lo que no sería correcto hablar de personas inactivas.

Se integraron los capítulos de desocupados e inactivos, para crear el capítulo de No ocupados, 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.

Respecto a la subutilización de la fuerza de trabajo y la situación en la fuerza de trabajo, se incluyen en los capítulos de ocupados y no ocupados, preguntas que reflejan problemas de insuficiencia de absorción de mano de obra a partir de la medición del desempleo, la subocupación por insuficiencia de horas y la fuerza de trabajo potencial. Se incluyeron preguntas que permiten identificar tres grupos de personas: (1) Buscadores no disponibles, (2) Buscadores potenciales disponibles y (3) Buscadores potenciales que desean trabajar.

En cuanto a la publicación de resultados de mercado laboral, se estableció que la Población en Edad de Trabajar iría desde los 15 años en adelante, atendiendo la recomendación de la Resolución I sobre las estadísticas del trabajo, la ocupación y la subutilización de la fuerza de trabajo, numeral 65, el cual hace referencia a los límites de edad para la PET, indicando que el límite de edad inferior debería tener en cuenta la edad mínima para acceder al mercado laboral, con las excepciones previstas en la legislación nacional. Para el caso de Colombia, la norma vigente es la Ley 515 de 1999 por medio de la cual se aprueba el "Convenio 138 sobre la Edad Mínima de Admisión de Empleo", adoptada por la 58ª Reunión de la Conferencia General de la Organización Internacional del Trabajo, Ginebra, Suiza, el veintiséis (26) de junio de mil novecientos setenta y tres (1973).

➤ **CIET 20ª:** con relación a esta conferencia la GEIH adoptó las Directrices relativas a las estadísticas de la migración laboral internacional, así como la Resolución I: relativa a las estadísticas sobre las relaciones laborales.

Para el caso de la migración laboral internacional, se realizaron ajustes y se incluyeron preguntas en el capítulo de migración que permitieran medir a los trabajadores migrantes internacionales y los trabajadores migrantes internacionales que regresan a su país. Esto con el fin de contribuir a las estadísticas nacionales sobre migración laboral, así como probar el marco conceptual desarrollado por la OIT.

En cuanto a las recomendaciones sobre la medición estadística de las relaciones de trabajo, hechas en la Resolución I y la necesidad de mejorar la comparabilidad internacional de las estadísticas sobre las situaciones contractuales de los trabajadores y su contribución a la medición del trabajo decente y el nivel de bienestar de los hogares y de la sociedad, se incluyeron preguntas que permiten identificar a la población catalogada como Contratistas dependientes, específicamente lo relacionado con la Clasificación Internacional de la situación en la Ocupación con base en el tipo de autoridad (CISO-18-A), en aras de establecer las relaciones de autoridad entre los trabajadores y las unidades económicas en las que trabajan o para las que realizan el trabajo.

II. Marco conceptual

En el desarrollo de la operación estadística, se utilizan los conceptos adaptados con respecto a la resolución I sobre las estadísticas del trabajo, la ocupación y la subutilización de la fuerza de trabajo de 2013 (19ª CIET), los conceptos de empleo y desempleo no presentan modificación respecto a lo vigente (13ª CIET - 1982). Asimismo, se utiliza el concepto adaptado de subocupación definido en la resolución referente a la medición del subempleo y situaciones del empleo (16ª CIET - 1998).

Población en Edad de Trabajar (PET): Población constituida por todas las personas de 15 años y más. Se divide en fuerza de trabajo y población fuera de la fuerza de trabajo.

Fuerza de Trabajo (FT): Comprende a las personas en edad de trabajar que trabajan o están buscando trabajo y están disponibles. Se divide en población ocupada y desocupada.

Población Ocupada (OC): Corresponde a las personas de 15 años y más que, durante el período de referencia: trabajaron por lo menos una hora remunerada en la semana de referencia; no trabajaron la semana de referencia, pero tenían un trabajo o negocio, y trabajaron en la semana de referencia por lo menos una hora sin remuneración.

Población desocupada (DS): Corresponde a las personas de 15 años y más que durante el período de referencia estuvieron simultáneamente: "sin empleo", es decir, que no tenían un empleo asalariado o un trabajo independiente ni se desempeñaron como trabajador(a) familiar sin remuneración; "buscando empleo", es decir que habían tomado medidas concretas para buscar un empleo asalariado o independiente en las últimas 4 semanas; y "actualmente disponibles para trabajar", es decir, disponibles para trabajar en empleo asalariado o en empleo independiente durante el período de referencia.

Población Fuera de la Fuerza de Trabajo (PFFT): Comprende a todas las personas en edad de trabajar de 15 años y más, que en la semana de referencia no participaron en la producción de bienes y servicios porque no lo necesitaban, no podían o no estaban interesadas en tener actividad remunerada. Se compone de las personas que integran la fuerza de trabajo potencial y de las personas que son exclusivamente: estudiantes, personas dedicadas a oficios del hogar, personas pensionadas, personas jubiladas, rentistas, personas incapacitadas permanentemente para trabajar, personas que no les llama la atención o creen que no vale la pena trabajar.

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

Subocupación (S): Se refiere a los trabajadores que tienen el deseo manifiesto de mejorar sus ingresos, el número de horas trabajadas o el tener una labor más acorde a sus competencias y que han hecho una gestión para materializar su aspiración y están en disposición de efectuar el cambio.

Subocupación por insuficiencia de horas (SIH): Corresponde a la población ocupada que desea trabajar más horas, ya sea en su empleo principal o secundario, y tiene una jornada inferior a 48 horas semanales.

Condiciones de empleo inadecuado:

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

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

Fuerza de Trabajo Potencial (FTP): Comprende a las personas que se encuentran fuera de la fuerza de trabajo y que tienen interés en trabajar. Está compuesta por tres grupos mutuamente excluyentes: buscadores no disponibles, buscadores potenciales disponibles y buscadores potenciales que desean trabajar.

Buscadores no disponibles: Se refiere a las personas sin empleo que están buscando empleo, pero no se encuentran disponibles para trabajar.

Buscadores potenciales disponibles: Se refiere a las personas sin empleo que no están buscando empleo, pero se encuentran disponibles para trabajar.

Buscadores potenciales que desean trabajar: Se refiere a las personas sin empleo que no están buscando empleo ni se encuentran disponibles, pero desean trabajar.

Formas de trabajo contempladas en la encuesta:

Trabajo en la ocupación: comprende el trabajo realizado para terceros a cambio de remuneración o beneficios.

Trabajo Voluntario (TV): comprende a todas las personas de 15 años y más, que en el período de referencia (últimas 4 semanas) realizaron cualquier actividad voluntaria (no remunerada y no obligatoria) con la finalidad de producir bienes o prestar servicios para terceros.

Trabajo de producción para el autoconsumo: comprende a todas las personas de 15 años y más, que en el período de referencia (últimas 4 semanas) desarrollaron una actividad no remunerada para la producción de bienes o prestación de servicios para uso final propio.

Trabajo en formación no remunerado: comprende el trabajo realizado para terceros, sin remuneración, para adquirir experiencia o competencias en el lugar de trabajo.

Población ocupada informal: Son todos los asalariados o empleados domésticos que no cuentan con cotizaciones de salud ni a pensión por concepto de su vínculo laboral con el empleador que los contrató. De igual forma, comprende a los trabajadores por cuenta propia y patrones o empleadores que hayan quedado clasificados en el sector informal, todos los trabajadores familiares sin remuneración y demás personas sobre las que no se dispone de suficiente información sobre su situación en la ocupación para ser clasificadas en las categorías anteriores. Se excluyen los obreros y empleados del Gobierno.

Sector informal: Son todas las unidades económicas de mercado o empresas que no están registradas ante las autoridades pertinentes y que no pueden ser clasificadas como cuasisociedades ya que no poseen una contabilidad completa o simplificada que les permita realizar una efectiva separación de sus gastos. Estas unidades funcionan generalmente en pequeña escala y con una organización rudimentaria en la que hay muy poca o ninguna distinción entre el trabajo y el capital como factores de producción; se excluyen las entidades del gobierno y las organizaciones sin fines de lucro.

Línea de pobreza (LP): Corresponde al costo per cápita mínimo de una canasta básica de bienes (alimentarios y no alimentarios) en un área geográfica determinada. Es el resultado de combinar dos elementos, a saber: el valor de la canasta básica de alimentos (línea de pobreza extrema) y el coeficiente de Orshansky. Resulta de multiplicar el valor de la línea a de pobreza extrema por el coeficiente de Orshansky.

Línea de pobreza extrema o indigencia monetaria (LI): valor que representa el ingreso mínimo necesario por persona para cubrir el costo de una canasta básica de alimentos, que tenga los requerimientos calóricos adecuados.

Pobreza extrema o indigencia monetaria: Ocurre cuando el ingreso mensual per cápita de un hogar se sitúa por debajo de la línea de indigencia (pobreza extrema) monetaria.

Pobreza monetaria: ocurre cuando el ingreso mensual per cápita de un hogar se sitúa por debajo de la línea de pobreza monetaria.

Las estadísticas de mercado laboral en Colombia están alineadas con las recomendaciones de la Organización Internacional del Trabajo (OIT), las cuales se discuten y difunden a través de las Conferencias Internacionales de Estadísticos del Trabajo (CIET). Los conceptos se adoptan de dichas resoluciones.

III. Marco legal o normativo

Decreto 3167 de 1968, Estatuto Orgánico del DANE, *"Por medio del cual se reorganiza el Departamento Administrativo Nacional de Estadística"*, establece en su artículo 2 las funciones que debe cumplir la entidad, dentro de las cuales se precisa que el DANE debe elaborar las investigaciones estadísticas necesarias para la formulación de planes de Desarrollo, para lo cual tendrá en cuenta las necesidades del Departamento Nacional de Planeación; planear, desarrollar y ejecutar las labores censales de todo orden que se requieran en el Gobierno Nacional, las encuestas, las estadísticas directas y derivadas, tales como las Cuentas Nacionales y las que se estimen necesarias para los establecimientos de indicadores económicos y sociales; elaborar índices de salarios, niveles de empleo y demás estadísticas laborales requeridas para la medición de las fluctuaciones del mercado de trabajo, en coordinación con el Ministerio del ramo.

Ley 79 de 1993, *"Por la cual se regula la realización de los Censos de Población y Vivienda en todo el territorio nacional"* y establece en su artículo 5 que *"Las personas naturales o jurídicas, de cualquier orden o naturaleza, domiciliadas o residentes en el territorio nacional, están obligadas a suministrar al Departamento Administrativo Nacional de Estadística DANE, los datos solicitados en el desarrollo de Censos y Encuestas"*.

Los datos suministrados al Departamento Administrativo Nacional de Estadística DANE, en el desarrollo de los censos y encuestas, no podrán darse a conocer al público ni a las entidades u organismos oficiales, ni a las autoridades públicas, sino únicamente en resúmenes numéricos, que no hagan posible deducir de ellos información alguna de carácter individual que pudiera utilizarse para fines comerciales, de tributación fiscal, de investigación judicial o cualquier otro diferente del propiamente estadístico".

El Decreto 262 de 2004, *"Por el cual se modifica la estructura del Departamento Administrativo Nacional de Estadística DANE y se dictan otras disposiciones"*, establece en su artículo 13 las funciones de la Dirección de Metodología y Producción Estadística, dentro de las cuales se encuentran la de diseñar y ejecutar las operaciones estadísticas que requiera el país para la planeación y toma de decisiones por parte del Gobierno Nacional y los entes territoriales, en materia social y económica; elaborar los estudios técnicos, en materia social y económica, que permitan definir la información estadística estratégica que deba generarse a nivel nacional, sectorial y territorial; y la de solicitar y obtener de las personas naturales o jurídicas, residenciadas o domiciliadas en Colombia, los datos que sean requeridos para dotar de información estadística al país.

Resolución 1341 de 2010, Secretaría General de la Comunidad Andina, *"Por la cual se adopta el manual de elaboración de indicadores del mercado laboral provenientes de encuestas a hogares y sus disposiciones técnicas para la transmisión de datos estadísticos"*.

IV. Referentes internacionales

Los conceptos, las definiciones, el formulario básico y el sistema de recolección se enmarcan en las resoluciones y recomendaciones que produce la OIT mediante las CIET. Dentro del proceso de rediseño de la GEIH, se revisaron las resoluciones de la CIET adoptadas después de 1998, de esta forma los conceptos, cambios e innovaciones en mercado laboral incorporados en la encuesta a partir de enero de 2022, fueron alineados con la resolución sobre las estadísticas del trabajo, el empleo y la subutilización de la fuerza de trabajo de la 19ª CIET de 2013. Asimismo, se analizaron las directrices emanadas de la 17ª 18ª y 20ª CIET en materia de empleo informal, medición del tiempo de trabajo, relaciones laborales, migración laboral internacional, desajustes en la medición de la cualificación y las competencias de las personas ocupadas.

Tabla 1. Resumen normatividad internacional de las estadísticas del trabajo

Normatividad	AÑO	Descripción	Fuente
13ª CIET, resolución I	1982	Adoptó nuevas normas relativas a las estadísticas de la Población Económicamente Activa (PEA), del empleo, desempleo y subempleo.	OIT
15ª CIET	1993	Adoptó la resolución sobre las estadísticas de ocupación en el sector informal.	OIT
16ª CIET	1998	Adoptó la resolución I, concerniente a la medición del subempleo y las situaciones de empleo inadecuado.	OIT
17ª CIET	2003	Adoptó directrices sobre una definición estadística de empleo informal.	OIT
18ª CIET	2008	Adoptó las siguientes resoluciones: I, relativa a la medición del tiempo de trabajo; II, sobre las estadísticas de trabajo infantil; III, concerniente al desarrollo de medidas de subutilización de la mano de obra; IV, sobre la labor futura en cuanto a la medición del trabajo decente; V, referente a la enmienda del párrafo 5 de la Resolución sobre estadísticas de la población económicamente activa, del empleo, del desempleo y del subempleo adoptada por la decimotercera Conferencia Internacional de Estadísticos del Trabajo (octubre de 1982).	OIT
19ª CIET	2013	Adoptó las resoluciones sobre las estadísticas del trabajo, la ocupación y la subutilización de la fuerza de trabajo y, II concerniente a la labor sobre las estadísticas de trabajo forzoso.	OIT
Resolución 2013/21 ⁸	2013	Establecen parámetros generales para que los organismos nacionales encargados de la producción estadística brinden al público información efectiva, transparente y confiable relacionada con la situación económica, demográfica, social o ambiental.	ONU
20ª CIET	2018	Adoptó las siguientes resoluciones: I, sobre las estadísticas de las relaciones laborales; II, sobre la metodología del indicador 8.8.2 de los Objetivos de Desarrollo Sostenible relativo a los derechos laborales; III, sobre la metodología del indicador 8.b.1 de los Objetivos de Desarrollo Sostenible relativo al empleo de los jóvenes; IV, relacionada con la modificación de la Resolución de la 18.ª Conferencia Internacional de Estadísticos del Trabajo (CIET) sobre las estadísticas del trabajo infantil. Igualmente, adoptó las siguientes directrices: 1. relativas a las estadísticas sobre migración laboral internacional, 2. relativas a la medición del trabajo forzoso, 3. relativas a los desajustes en la medición de la cualificación y las competencias de las personas ocupadas, 4. relativas a las estadísticas sobre las cooperativas.	OIT

Fuente: DANE, a partir de la normatividad de la OIT y la ONU.

V. Referentes nacionales

La Gran Encuesta Integrada de Hogares es la operación estadística que proporciona la información de mercado laboral, por tanto, es el referente nacional para las demás entidades del orden nacional y subnacional.

⁶ Esta conferencia de expertos de todo el mundo se reúne ordinariamente cada cinco años en Ginebra (Suiza), y produce resoluciones y recomendaciones sobre la medición de las estadísticas laborales. En la reunión llevada a cabo en septiembre de 1998, entre otras resoluciones, adoptó la referente a las estadísticas de subempleo y las condiciones de empleo inadecuado. Principales CIET: XIII CIET de Conceptos de Población Económicamente Activa, XV CIET de Ingresos e Informalidad, XVII CIET de Medición del Subempleo.

⁷ Para mayor información sobre las CIET lo invitamos a consultar el siguiente enlace de la OIT: <https://ilostat.ilo.org/es/about/standards/icls/icls-documents/>

⁸ Los diez (10) Principios Fundamentales de las Estadísticas Oficiales de las Naciones Unidas.

2.1.5. Definición de variables y construcción de indicadores estadísticos

Principales variables según capítulos del formulario de la GEH

- Vivienda:** tipo de vivienda y características físicas (material de paredes y pisos), conexión a servicios públicos, privados o comunales, conexión y uso de servicio sanitario, obtención de agua para consumo.
- Datos del hogar:** lugar y energía para preparar alimentos, eliminación de basuras, tipo de tenencia de la vivienda, productos financieros.
- Registro de personas:** identificación del residente habitual.
- Características generales:** sexo, edad, parentesco con el jefe del hogar, estado civil, identificación de población campesina, y autorreconocimiento étnico.
- Seguridad social en salud:** cobertura del Sistema General de Seguridad Social en Salud (SGSSS) por regímenes, persona que paga la afiliación y cobertura, condición física y mental (discapacidad).
- Educación:** alfabetismo, asistencia escolar, máximo nivel educativo alcanzado y último año aprobado o que esté cursando y títulos o diplomas de mayor nivel educativo obtenidos.
- Fuerza de trabajo:** clasificación de la población en edad de trabajar entre Ocupados y No ocupados (Desocupados y Población fuera de la fuerza de trabajo).
- Ocupados:**
 - Empleo principal (rama de actividad, ocupación, tipo de contrato, acceso a prestaciones, tiempo trabajado y posición ocupacional).
 - Asalariados (medios de búsqueda, remuneración mensual, horas extras, pagos en especie, subsidios, primas y bonificaciones, registro de la empresa en la que presta servicio).
 - Independientes (formas de trabajo, registro mercantil, contabilidad, ganancia u honorarios netos, decisión en: horario, producción y ventas, precio de bienes, registros del negocio, ingresos).
 - Asalariados e independientes (duración del empleo, horas normales y efectivas trabajadas, honorarios, tamaño de la empresa, sitio de trabajo, afiliación a pensiones, caja de compensación familiar y administradora de riesgos profesionales (ARP), duración entre empleo anterior y actual).

- Empleo secundario (horas trabajadas, posición ocupacional, remuneración mensual, tamaño de la empresa y sitio de trabajo).
- Empleo con insuficiencias de horas y situaciones de empleo inadecuado (por competencias e ingresos).
- Calidad del empleo.

- i) **No ocupados:** duración de búsqueda de trabajo, trayectoria laboral, ingresos y seguridad social.
- j) **Otras formas de trabajo:** actividades no remuneradas y ayudas, pasantías, autoconsumo y voluntariado.
- k) **Otros Ingresos e impuestos:** otros ingresos diferentes a los ingresos laborales y pago de impuestos.
- l) **Micronegocios:** aspectos de los trabajadores(as) por cuenta propia y patronos que trabajan en empresas hasta de 10 personas.
- m) **Fecundidad:** hijos(as) nacidos(as) vivos(as).
- n) **Migración:** nacionalidad, intención de permanencia en el país, y emigración laboral dinámicas migratorias de la población objetivo a nivel nacional.
- o) **Formación para el trabajo:** asistencia, duración, modalidad, temática del curso de formación para el trabajo y certificación de competencias.
- p) **Trabajo Infantil:** actividades y razones del trabajo infantil.
- q) **Identificación:** tipo de identificación, orientación sexual e identidad de género.

Principales indicadores

Porcentaje de la Población en Edad de Trabajar (% PET)

$$\% \text{ PET} = (\text{PET}/\text{PT}) * 100$$

Relación porcentual entre el número de personas que componen la población en edad de trabajar (PET), frente a la población total (PT).

Tasa Global de Participación (TPG)

$$\text{TGP} = \text{FT}/\text{PET} * 100$$

Relación porcentual entre la fuerza de trabajo (FT) y la población en edad de trabajar (PET). Este indicador refleja la presión de la población en edad de trabajar sobre el mercado laboral.

Tasa Bruta de Participación (TBP)

$$\text{TBP} = \text{FT}/\text{PT} * 100$$

Relación porcentual entre el número de personas que componen la fuerza de trabajo (FT), frente al número de personas que integran la población total (PT).

Tasa de Desempleo (TD)

$$\text{TD} = \text{DS}/\text{FT} * 100$$

Relación porcentual entre el número de personas desocupadas (DS) y el número de personas que integran la fuerza de trabajo (FT).

Tasa de Ocupación (TO)

$$\text{TO} = \text{OC}/\text{PET} * 100$$

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

Tasa de Subocupación (TS)

$$\text{TS} = \text{PS}/\text{FT} * 100$$

Es la relación porcentual de la población ocupada que manifestó querer y poder trabajar más horas a la semana, mejorar sus ingresos y/o tener una labor más propia de sus competencias (PS) y el número de personas que integran la fuerza laboral (FT). Está compuesta por:

Tasa de Subocupación por Insuficiencia de Horas (TSIH)

$$\text{TSIH} = \text{SIH}/\text{FT} * 100$$

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

Tasa de Subocupación por Condiciones de Empleo Inadecuado:

Tasa de Subocupación por Competencias (TSC)

$$\text{TSC} = \text{SC}/\text{FT} * 100$$

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

Tasa de Subocupación por Ingresos (TSI)

$$\text{TSI} = \text{SI}/\text{PEA} * 100$$

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

Medidas de subutilización de la fuerza de trabajo⁹

La subutilización de la fuerza de trabajo se refiere a la inadecuación entre la oferta y la demanda de trabajo a consecuencia de una absorción insuficiente de la oferta de trabajo, lo que se traduce en situaciones de necesidad insatisfecha de empleo en la población. La subutilización de la fuerza de trabajo incluye las siguientes medidas:

Tasa combinada de subocupación por insuficiencia de horas y desempleo (TCSIHD):

$$\text{TCSIHD} = (\text{SIH} + \text{D}) / \text{FT} * 100$$

Es la relación porcentual de las personas ocupadas que desean trabajar más horas, ya sea en su empleo principal o secundario, y tienen una jornada inferior a 48 horas semanales (SIH) más las personas desocupadas (DS) y el número de personas que integran la fuerza de trabajo (FT).

Tasa combinada de desempleo y fuerza de trabajo potencial (TCDFTP):

$$\text{TCDFTP} = (\text{DS} + \text{FTP}) / (\text{FT} + \text{FTP}) * 100$$

Es la relación porcentual de las personas desocupadas (DS) más las personas que integran la fuerza de trabajo potencial (FTP) y el número de personas que integran la fuerza de trabajo (FT) más las personas que integran la fuerza de trabajo potencial (FTP).

Medida compuesta de subutilización de la fuerza de trabajo (MCSFT):

$$\text{MCSFT} = (\text{SIH} + \text{DS} + \text{FTP}) / (\text{FT} + \text{FTP}) * 100$$

Es la relación porcentual de las personas ocupadas que desean trabajar más horas, ya sea en su empleo principal o secundario, y tienen una jornada inferior a 48 horas semanales (SIH) más las personas desocupadas (DS) más las personas que integran la fuerza de trabajo potencial (FTP) y el número de personas que integran la fuerza de trabajo (FT) más las personas que integran la fuerza de trabajo potencial (FTP).

Línea de pobreza: resultado de multiplicar el valor de la canasta básica de alimentos (línea de pobreza extrema o indigencia monetaria) por el coeficiente de Orshansky (cociente entre el gasto total y el gasto en alimentos).

Incidencia de la pobreza monetaria

Porcentaje de personas cuyos ingresos corrientes totales son inferiores a la línea de pobreza monetaria (LP).

Incidencia de la pobreza monetaria extrema

Porcentaje de personas cuyos ingresos corrientes totales son inferiores a la línea de pobreza extrema (LI).

Para asegurar la calidad de los procesos se utilizan principalmente los indicadores de confiabilidad (cobertura de viviendas y hogares) y de calidad (tasa de respuesta), garantizando así la calidad de la recolección y los estimadores de precisión de resultados (error de muestreo relativo trimestral de 5,5% en la tasa de desempleo de las ciudades y de 4,7% anual de la pobreza monetaria en los departamentos, con un nivel de confiabilidad del 95%).

Pertinencia de las variables

Mediante la información recolectada por la GEIH, se generan diferentes variables que permiten estimar los principales indicadores del mercado laboral colombiano y caracterizar a la población según sexo, edad, estado civil, nivel educativo, acceso al sistema de seguridad social en salud, entre otros. Asimismo, permite conocer el oficio o labor que desarrollan las personas ocupadas, la posición ocupacional y la rama de actividad a la cual están aportando su fuerza laboral, tipo de contrato y calidad del empleo.

Adicionalmente, es posible identificar a la población desocupada y la población que no se encuentra ocupada para determinar la duración de búsqueda de trabajo, trayectoria laboral, ingresos y seguridad social. Luego del rediseño de la operación estadística se incorporaron nuevas variables, las cuales no afectan las series de tiempo ni los indicadores de mercado laboral, ya que se mantiene su conceptualización y forma de cálculo.

Las variables que permiten identificar la población campesina, en condición de discapacitada y LGBT incluidas en la encuesta se encuentran en línea con los objetivos, dado que permiten medir las características sociodemográficas de la población. Del mismo modo, los módulos incluidos como: Otros ingresos e impuestos, que permite identificar otros ingresos diferentes a los ingresos laborales, busca obtener información que permita realizar la medición del ingreso y de los indicadores de pobreza monetaria y pobreza monetaria extrema de la población residente en Colombia, y el módulo Otras formas de trabajo (actividades no remuneradas y ayudas, pasantías, autoconsumo y voluntariado) suministra información sobre las labores de trabajo doméstico y de cuidado no remunerado, trabajo en formación, trabajo en producción de bienes para autoconsumo y trabajo voluntario.

Pertinencia y desgaste de la fuente

El desgaste de la fuente se minimiza con la actualización y rotación de la muestra. Es importante señalar que el cuestionario se aplica a las fuentes que responden la encuesta mediante la técnica de entrevistado/encuestado directo y entrevistado/encuestado idóneo¹⁰. Por su lado, el Marco Geoestadístico Nacional tiene una vigencia anual y su actualización se hace en sus componentes cartográfico y temático, teniendo en cuenta los protocolos de actualización para los casos de: crecimiento horizontal, crecimiento vertical, incorporación de nuevos centros poblados y revisión de muestra, así como la expansión de las cabeceras municipales. De tal manera, se generan muestras trimestrales que cumplen con los criterios estadísticos y se realiza la rotación de la muestra, en donde se selecciona una manzana o sección rural y de estas se toma un segmento aleatoriamente, y para la siguiente iteración de la encuesta, se elige un segmento diferente al inicialmente seleccionado. Cuando se agota la manzana, se toma una manzana vecina para empezar de nuevo el proceso.

⁹ Oficina Internacional del trabajo OIT: Informe II Estadísticas del trabajo, el empleo y la subutilización de la fuerza de trabajo.

¹⁰ La encuesta utilizará informante directo para las personas de 18 años y más, y para los de 10 a 17 años que trabajen o estén buscando trabajo. Para los demás se acepta informante idóneo (persona del hogar mayor de 18 años, que a falta del informante directo pueda responder correctamente las preguntas).

2.1.6. Plan de resultados

Cuadros de salida

En la GEIH se producen dos tipos de cuadros, los de salida, generados por la oficina de Sistemas y los cuadros de resultados, los cuales son elaborados por el grupo interno de trabajo de temática de mercado laboral.

Los cuadros de salida generados por la oficina de Sistemas son el insumo principal para el análisis y la producción estadística de los indicadores de mercado laboral producidos por la operación estadística. Asimismo, son utilizados para el contraste de resultados de los usuarios internos de la encuesta.

Estos cuadros se diseñan de acuerdo con los requerimientos del área temática, así como con las desagregaciones requeridas (sexo, edad, ramas de actividad, posición ocupacional, entre otros) para la construcción de los indicadores. En este sentido, los cuadros de salida de la operación estadística se clasifican en cuadros de tasas, de poblaciones y de viviendas, hogares y personas.

En la producción de los cuadros de salida de tasas se calculan las tasas de Participación (Bruta y Global), Ocupación y Desempleo, según sexo y rangos de edad. En tanto que, los cuadros de salida de poblaciones presentan la PET, la fuerza de trabajo, los ocupados, los desocupados y la población fuera de la fuerza de trabajo, entre otras desagregaciones requeridas para la generación de resultados.

Todos los cuadros deben sumar los subtotales, resultado que debe ser igual al gran total. Adicionalmente, son necesarios los niveles de desagregación geográfica y temporal de los resultados.

Los cuadros de resultados producidos y publicados por la operación estadística corresponden a la proporción de población en edad de trabajar, a la tasa global de participación, de ocupación, de desempleo y de subocupación, y a la población total, en edad de trabajar, fuerza de trabajo, ocupados, desocupados, población fuera de la fuerza de trabajo, subocupados y fuerza de trabajo potencial. Asimismo, se entrega información sobre la población ocupada según ramas de actividad CIIU 4 A.C y por posición ocupacional, en tanto que, para la población fuera de la fuerza laboral se proporcionan datos según tipo de actividad.

Adicionalmente, se elaboran cuadros de resultados con los principales indicadores de mercado laboral con temáticas especiales como sexo, parentesco con el jefe del hogar, ocupación informal, para la población fuera de la fuerza laboral, campesina, LGBT, las personas en condición de discapacidad y los jóvenes.

2.1.7. Estándares estadísticos utilizados

En esta operación estadística se emplean las siguientes clasificaciones estadísticas:

- Clasificación Internacional Normalizada de la Educación Adaptada para Colombia (CINE-N 2011 A.C.).
- Clasificación Internacional de la Situación en la Ocupación Adaptada para Colombia (CISO 18 A.C.).
- Clasificación Industrial Internacional Uniforme de todas las Actividades Económicas (CIIU Rev. 4 A.C. 2020).
- Clasificación Internacional de la Situación en el Empleo (CISE 1993).
- Clasificación Internacional Uniforme de Ocupaciones (CIIU 08 A.C.).
- Codificación de la división política administrativa de Colombia (DIVIPOLA).
- Nesstar Publisher, se documenta la GEIH y se publica en el portal Archivo Nacional de Datos (ANDA).

2.1.8. 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.

L. OTROS INGRESOS E IMPUESTOS: se determina cuáles personas recibieron otros ingresos distintos a los laborales y se captan los impuestos pagados durante los últimos 12 meses.

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 INFANTIL¹²: 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.

Figura 3. Ajustes realizados al cuestionario



Fuente: DANE.

¹¹ Entre 2013 y 2022 el módulo de formación para el trabajo se aplicó en el segundo trimestre de cada año; excepto en el periodo 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.

2.1.9. Normas, especificaciones o reglas de edición e imputación de datos

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

Validación de rangos de acuerdo con la estructura de las preguntas

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

Validación de los universos

A este proceso lo definen tres aspectos: el primero, cuando la pregunta define un flujo o salto dependiendo de la opción; el segundo, cuando los datos de la vivienda se toman una sola vez, así exista más de un hogar; y el tercero, cuando se define la pregunta según la edad de la persona residente en el hogar. Para algunas variables relacionadas con la aplicación de conceptos básicos, que son determinantes en la calidad de la estructura de la información, el Dispositivo Móvil de Captura (DMC) permite que el entrevistador confirme la respuesta que da el entrevistado. Estas variables son: número de hogares en la vivienda, nombre y apellidos, sexo, edad, parentesco con el jefe del hogar, y todas las variables relacionadas con valores en los formularios. En el diseño del programa, se garantiza que en la misma pantalla se muestren el capítulo que está relacionado con las variables, como las diferentes preguntas que, por su alta correlación, se les debe verificar la consistencia.

La GEIH a través de las especificaciones de consistencia del formulario, establece los flujos de las preguntas, los rangos de valores que son objeto de validación, define los filtros de edad para las preguntas o por capítulo, asimismo, define las reglas de consistencia entre preguntas acorde con la información diligenciada en las preguntas que se relacionan.

2.2. DISEÑO ESTADÍSTICO

El diseño estadístico reúne varios aspectos entre los cuales se destacan la formulación del marco estadístico, el universo, la población, las unidades estadísticas, los períodos y el diseño muestral. A partir del año 2021, se actualizó el marco muestral con el censo del 2018; adicionalmente, se hicieron ajustes en algunos aspectos metodológicos. Esta metodología describe los procedimientos realizados a partir del año 2021.

2.2.1. Universo de estudio

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.

2.2.2. Población objetivo

La población objetivo para la Gran Encuesta Integrada de Hogares está conformado por la población civil no institucional residente en todo el territorio nacional. Se excluyen, de la población objetivo; 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.

2.2.3. 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.

2.2.4. Desagregación geográfica

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. En la siguiente tabla, se presenta la descripción de estos niveles:

Desagregación temporal ²³		Desagregación geográfica	Descripción
A n u a l	T r e m e s t r a l	Total Nacional	Total Nacional.
		Total de cabeceras de ciudades (con o sin áreas metropolitanas)	Para realizar la estimación por grupos de ciudades, debe haber al menos tres ciudades (opcional con áreas metropolitanas) por agrupación. No aplica para los departamentos de la Amazonia y la Orinoquia ni ciudades intermedias. Caso particular de publicación: total 13 ciudades con áreas metropolitanas.
		Cabeceras de las ciudades (con o sin áreas metropolitanas)	Se puede hacer la estimación para cada una de las ciudades, con o sin áreas metropolitanas.
		Grandes agrupaciones	Total nacional por zona: cabeceras, centros poblados y rural disperso.
		Regiones	Total por región: Atlántica, Oriental, Central, Pacífica y Bogotá.
		San Andrés	Estimación para la ciudad, capital de departamento.
		Departamentos	Total, por departamento.
		Cabeceras de las capitales de los departamentos de la Amazonia y la Orinoquia	Se puede hacer la estimación para cada una de las capitales de los departamentos de la Amazonia y la Orinoquia y, adicionalmente, para su agregado como conjunto.
		Ciudades intermedias	Se puede hacer la estimación para cada una de las ciudades intermedias y, adicionalmente, para su agregado como conjunto.

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 señalados en la tabla anterior:

- **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.

¹³ Nótese que cada nivel de desagregación temporal incluye las desagregaciones geográficas del anterior, además de tener desagregaciones geográficas propias. Esto es, la desagregación trimestral incluye los elementos de la mensual; la semestral incluye los de la trimestral y, por tanto, los de la mensual; y la anual incluye los de las desagregaciones mensual, trimestral y semestral.

2.2.5. Desagregación temática

El tamaño de muestra permite presentar resultados desagregados para viviendas, hogares y personas, dependiendo de los errores de muestreo asociados a cada desagregación. No se recomienda usar desagregaciones cuyos indicadores posean coeficientes de variación superiores al 15%.

2.2.6. Fuentes de datos

La fuente de información de la operación estadística está constituida por los hogares y personas seleccionados en un muestreo probabilístico. Se trata, por tanto, de una fuente primaria de información.

2.2.7. Unidades estadísticas

a) **Unidad de observación:** corresponden a las viviendas, los hogares y las personas.

b) **Unidad de análisis:** también están representadas por las viviendas, los hogares y las personas.

c) **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.

2.2.8. Período de referencia

A continuación, se presentan los períodos de referencia utilizados en la encuesta:

- **Semana pasada:** semana calendario completa, de lunes a domingo, que precede inmediatamente a la semana en que se efectúa la encuesta.
- **Mes pasado:** se define como el mes inmediatamente anterior al que se está realizando la encuesta y aplica para la información relacionada con los ingresos laborales.
- **Último mes y los últimos doce meses:** aplica para el caso de la clasificación de la población desocupada y la identificación de los ingresos no laborales, se maneja como período de referencia el último mes y los últimos doce meses, inmediatamente anteriores al mes en que se está realizando la encuesta.

2.2.9. Período de recolección/acopio

Semana de recolección o semana de la encuesta: semana calendario, de lunes a domingo, durante la cual se lleva a cabo la recolección u obtención de la información.

El período de recolección es semanal para las 32 ciudades capitales de departamento con sus áreas metropolitanas. Para las cabeceras, centros poblados y rural disperso, se maneja un promedio de 21 días.

2.2.10. 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 campo¹⁴ 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.

¹⁴ Para el DANE el 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.

2.2.11. 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).

En lo que respecta a esta encuesta se busca establecer tamaños de muestra para indicadores de proporción, con lo cual se define el tamaño de muestra (página 75, Cochran, 1977) como:

$$n = \frac{Def f \cdot T \cdot N \cdot O \cdot z^2 \cdot P \cdot Q}{N \cdot e^2 + P \cdot Q \cdot z^2}$$

Donde,

n= Número de individuos a encuestar.

N= Tamaño del universo de estudio; en otras palabras, población a representar.

Q= 1 - P.

P = Proporción objetivo de estimación.

e= Margen de error.

z= Estadístico que define el nivel de confianza deseado, basado en la distribución normal estándar.

o= Relación entre la población objetivo y el universo de estudio. Más precisamente

$$o = \text{Población del universo} / \text{Población de sub universo}$$

T= Ajuste por no respuesta esperada. Si t es la tasa (cociente) de no respuesta definida como:

$$t = (\text{Número esperado de encuestas perdidas}) / (\text{Número esperado de encuestas})$$

Entonces

$$T = (1+t)$$

Deff = Efecto de diseño.

En la construcción de número de segmentos a seleccionar se divide a n por el número de personas por hogar (PPH), por el número de hogares por vivienda (PHV) y en 10.

Es decir:

$$\text{Número de segmentos a encuestar} = n / (PPH \cdot PHV \cdot 10)$$

Para la GEIH, hay dos indicadores escogidos como relevantes a ser estimados: El primero es la Tasa de Desempleo (TD), ya que es el indicador más importante del mercado laboral; además, cuenta con el mayor error muestral relativo, debido a que es de baja prevalencia. El segundo es el Índice de Pobreza Monetaria (PM), que se caracteriza por ser un indicador sujeto a seguimiento constante para el desarrollo de política pública.

Cada indicador tiene preestablecido una desagregación mínima temporal y geográfica diferente con el fin de garantizar estratégicamente su publicación. Para el caso de la tasa de desempleo se establece una publicación mensual de total país, trimestral por cada una de las 23 ciudades, semestral para San Andrés y anual para las capitales de los departamentos de la Amazonía y la Orinoquía, y cinco ciudades intermedias. Adicionalmente se debe asegurar una desagregación de ciudades más área metropolitana. En lo que respecta al Indicador de Pobreza Monetaria se requiere una estimación total departamento anual.

En la construcción del tamaño de muestra de las ciudades (ver tabla 2) se tienen varios criterios:

1. Se toma como tasa de referencia la Tasa de Desempleo promedio del 2018. Para calcular el Deff, la Fuerza de Trabajo, los promedios de personas por hogar y los promedios de hogares por vivienda, también se usa la información de 2018.
2. La subdivisión de población es la Fuerza de Trabajo.
3. Los tamaños de muestra son suficientes para cada cabecera de las 23 ciudades principales en el trimestre y semestral para San Andrés.
4. Se fija un error relativo (CV) de 5,5% y un intervalo de confianza de 95% para todas las ciudades.
5. La tasa de pérdida de la encuesta es de 5% ¹⁶.

Tabla 2. Distribución de la muestra de las ciudades según desagregación geográfica y temporal.

Ciudad	Desagregación temporal	Tasa de referencia	Deff	Hogares esperados
Medellín	Trimestral	10,0%	1,3	2.400
Barranquilla	Trimestral	10,0%	1,2	1.605
Bogotá D.C.	Trimestral	10,0%	1,4	2.431
Cartagena	Trimestral	10,0%	1,2	2.040
Tunja	Trimestral	12,3%	1,3	1.948
Manizales	Trimestral	11,5%	1,3	2.321
Florencia	Trimestral	15,1%	1,3	1.647
Popayán	Trimestral	11,8%	1,4	2.183
Valledupar	Trimestral	13,8%	1,3	1.502
Montería	Trimestral	12,3%	1,1	1.407
Quibdó	Trimestral	16,2%	1,9	2.299
Neiva	Trimestral	12,6%	1,4	1.900
Riohacha	Trimestral	15,3%	1,4	1.606
Santa Marta	Trimestral	10,7%	1,2	1.993
Villavicencio	Trimestral	12,9%	1,5	2.018
Pasto	Trimestral	10,0%	1,2	2.179
Cúcuta	Trimestral	16,3%	1,3	1.296
Armenia	Trimestral	16,0%	1,4	1.583
Pereira	Trimestral	10,0%	1,1	2.191
Bucaramanga	Trimestral	10,0%	1,2	1.924
Sincelejo	Trimestral	11,3%	1,3	1.660
Ibagué	Trimestral	13,7%	1,2	1.549
Cali	Trimestral	10,0%	1,3	2.299
San Andrés	Semestral	10,0%	1,5	2.426

Fuente: DANE.

Para la desagregación de ciudades más el área metropolitana, se fijó un error de muestreo relativo de 5%, con el cual se calculó un tamaño de muestra para la ciudad más área metropolitana. La diferencia entre ambos tamaños de muestra de ciudades más áreas metropolitanas (CV=5%) y ciudades (CV=5,5%) generan el tamaño de muestra asociado a las áreas metropolitanas (ver tabla 3).

Tabla 3. Distribución de la muestra de áreas metropolitanas según desagregación geográfica y temporal.

Áreas metropolitanas	Desagregación temporal	Tasa de referencia	Deff	Hogares esperados
Medellín	Trimestral	10,0%	1,3	519
Barranquilla	Trimestral	10,0%	1,2	336
Manizales	Trimestral	11,5%	1,3	485
Cúcuta	Trimestral	16,3%	1,3	276
Pereira	Trimestral	10,0%	1,1	454
Bucaramanga	Trimestral	10,0%	1,2	390
Cali	Trimestral	10,0%	1,3	499

Fuente: DANE.

En la construcción del tamaño de muestra anual por departamento, el cual está relacionado con la pobreza monetaria, se tomó en cuenta lo siguiente:

1. La información de referencia con la que se fijan los parámetros es de la encuesta de 2018.
2. Las tasas de referencia del Indicador de Pobreza Monetaria son del departamento, excluyendo las ciudades y áreas metropolitanas.
3. Las tasas de referencia se construyen asumiendo una disminución mayor, en el futuro, de los departamentos que en este momento tienen mayores niveles de pobreza monetaria, específicamente para departamentos que tienen una pobreza monetaria superior a 25%, se define su tasa de referencia igual a 25%. Para los departamentos que están entre 20% y 25%, se les reduce dos puntos porcentuales y, para los departamentos por debajo de 20%, se les deja la tasa registrada en 2018.
4. Los errores relativos (CV) se fijan en 4,7%.
5. El tamaño de la muestra es anual para cada departamento.
6. Los Deff son calculados de la experiencia obtenida del Indicador de Pobreza Monetaria de 2018.
7. La tasa de pérdida es de 5%.
8. La subdivisión de población de la encuesta es toda la población.

Como resultado de lo anterior, se obtienen los siguientes tamaños de muestra anuales para los departamentos sin ciudades y áreas metropolitanas:

Tabla 4. Distribución de la muestra de los departamentos según desagregación geográfica y temporal (excluyendo ciudades y áreas metropolitanas).

Departamento	Tasa de referencia	Deff	Hogares esperados
Antioquia	20,0%	6,4	3.821
Atlántico	25,0%	7,8	2.652
Bolívar	25,0%	6,4	2.333
Boyacá	25,0%	5,2	2.437
Caldas	25,0%	7,6	3.488
Caquetá	25,0%	6,4	2.698
Cauca	25,0%	6,3	2.684
Cesar	25,0%	6,8	2.706
Córdoba	25,0%	6,6	2.419
Cundinamarca	15,4%	5,7	3.414
Chocó	25,0%	5,7	2.316
Huila	25,0%	7,0	2.946
La Guajira	25,0%	7,7	3.087
Magdalena	25,0%	8,8	3.128
Meta	25,0%	8,3	3.789
Nariño	25,0%	5,8	2.497
Norte de Santander	25,0%	5,9	2.512
Quindío	23,4%	6,2	3.108
Risaralda	21,7%	7,9	3.024
Santander	24,2%	6,5	3.007
Sucre	25,0%	8,6	3.144
Tolima	25,0%	7,2	3.316
Valle del Cauca	23,6%	6,5	3.189

Fuente: DANE.

En la construcción del tamaño de muestra de las ciudades intermedias y los departamentos de la Amazonía y la Orinoquía se tienen varios criterios:

1. Se toma como tasa de referencia la tasa de desempleo promedio del año 2018. Para calcular el Deff, la Fuerza de Trabajo, los promedios de personas por hogar y los promedios de hogares por vivienda también se usa la información de 2018.
2. La subdivisión de población es la Fuerza de Trabajo.
3. Los tamaños de muestra son suficientes para cada cabecera de las 5 ciudades intermedias y las capitales de los departamentos de la Amazonía y la Orinoquía.
4. Se fija un error relativo de 5.5% ($Cv=5.5\%$) y un intervalo de confianza de 95%. Para todas las cinco ciudades intermedias. Para las cabeceras de las capitales de los departamentos de la Amazonía y la Orinoquía se fija un error relativo de 7%.
5. La tasa de pérdida de la encuesta es de 5%¹⁷.

En resumen, en promedio se espera encuestar mensualmente 18.500 hogares en las 32 ciudades capitales de departamento y 6.970 en el resto de cabecera, centro poblado y rural disperso, para un total de 25.470 hogares.

Selección de la muestra

En la primera etapa, la UPM o municipio se selecciona dentro de cada estrato de muestreo con un algoritmo de selección proporcional al tamaño (PPT).

Para la segunda, se ordena el marco por sector sección y manzana luego, se hace una segmentación del marco muestral, de tal manera que la USM tenga aproximadamente 10 viviendas en promedio. Luego dentro de cada estrato de la etapa se hace una selección sistemática de segmentos con único arranque. Para asegurar un tamaño de muestra fijo se usa el método de "Fractional Interval Method" el cual es su nombre en inglés.

Todos los procedimientos se hacen con PROC SURVEYSELECT del software SAS (SAS Institute Inc, 2019). Por lo cual las especificaciones de algoritmos usados se pueden revisar en la página web del programa.

A modo de ejemplo de código se tiene el siguiente código para seleccionar conglomerados de los municipios. Para este propósito el marco de conglomerados está en la base "marco" y tiene por lo menos dos variables, una identificando cada estrato lo cual es igual a municipio o "mpio" y otra identificando cada fila que corresponde al conglomerado. El tamaño de muestra está en otra base llamada "TAM_MUNICIPIOS_CAB"; en esta base hay dos columnas una que define el estrato que se llama "mpio" y otra _NSIZE_ que el programa lo entiende como el número de conglomerados a seleccionar. Luego de correr el programa, el resultado obtenido es una base que contiene la muestra de conglomerados llamada "muestra".

```
PROC SURVEYSELECT DATA =marco method = SYS OUT= Muestra seed=2
SAMPsize = TAM_MUNICIPIOS_CAB;
STRATA mpio;
RUN;
```

Mantenimiento de la muestra y manejo de novedades muestrales

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.

¹⁵ El ACP busca reducir la dimensionalidad de las variables que caracterizan los municipios y, de esta forma, facilitar la construcción de grupos homogéneos dentro y heterogéneos en un espacio multidimensional.

¹⁶ Los tamaños de muestra se presentan sin ajuste por tasa de pérdida.

¹⁷ Los tamaños de muestra se presentan sin ajuste por tasa de pérdida.

2.2.12. Ajustes de cobertura (o ajuste de cobertura por no respuesta)

El ajuste de cobertura se realiza en los casos donde se pierden viviendas, hogares y personas. El procedimiento se describió arriba cuando se definió el factor de cobertura (f_{cob}). Este factor restituye a nivel segmento las pérdidas ocasionadas por el operativo de campo. Cabe resaltar que, para evitar sesgos de selección, cuando se pierde de forma parcial o total la información de una persona dentro del hogar, el hogar se da como perdido y se toma en cuenta en el factor de cobertura.

2.2.13. Especificaciones de ponderadores

Los ponderadores usados en las encuestas de hogares que hace el DANE son construidos por etapas de la siguiente manera:

Factor de corrección de desactualización de marco (PSUB):

Es un factor a nivel de segmento cuyo objetivo es nivelar el peso que tiene el segmento al que debería tener por el aumento o disminución de viviendas en el momento actual. Se puede ver como una etapa más de la muestra, en donde el conglomerado final es partido en varios conglomerados de los cuales se selecciona sólo uno. Se expande entonces con el factor de sub-muestreo para de nuevo restituir el número de viviendas del conglomerado final.

Como ejemplo supóngase que se tiene una manzana de 20 viviendas en el censo 2018 (2 medidas de tamaño). Para el año 2020 el número de viviendas encontradas es 50 (5 medidas de tamaño). Es claro que, al segmentar estas 50 viviendas en 2, cada conglomerado final tendría un tamaño de 25 viviendas. Sin embargo, encuestar 25 viviendas es costoso, lo ideal es mantenerse en un tamaño de muestra fijo, así que se prefiere encuestar 10 viviendas y expandirlas a 25 con el peso de sub-muestreo. Es fácil verificar que el factor es aproximadamente $25/10$ (en términos de viviendas)=5/2 (en términos de medidas de tamaño)=2.5.

De esta forma, el peso de sub-muestreo se construye contrastando el número de medidas de tamaño que se definieron en la manzana a partir de la información de viviendas del censo 2018 (MTs teóricas), con las encontradas en campo en el momento de hacer la encuesta. Su fórmula es la siguiente:

$$PSUB = \frac{\text{Número de MTs actuales en la manzana}}{\text{Número de MTs teóricas en la manzana}}$$

Donde el número de MTs teóricas de la manzana es el entero más cercano de la razón entre el número de viviendas del censo 2018 y 10. Mientras el número de MTs actuales en la manzana es el entero más cercano del número de viviendas encontradas en el recuento y 10.

En el proceso de uso de este factor cabe aclarar que no se aplica de manera generalizada, sino que se restringe a unos límites con un mínimo de 0.8 y un máximo de 1.5, para evitar que el factor genere desproporciones estructurales en la medición de los indicadores.

Factor de cobertura (FCOB):

Es un factor a nivel de segmento que corrige la pérdida de viviendas, hogares y personas dentro del segmento. Su fórmula se establece de la siguiente forma:

- Controlando el tamaño del segmento: En la práctica los segmentos aparecen de diferentes tamaños, aunque se definen segmentos de 10 viviendas desde el diseño muestral. Para controlar este fenómeno se nivela la representación del segmento a 10 viviendas por segmento.

$$f_{cob1} = \frac{10}{\text{Número de viviendas encuestadas efectivas en el segmento}}$$

- Se ajusta las viviendas vacantes: La naturaleza de las áreas geográficas debe ser representada. El número de viviendas vacantes debe ser descontado del número de viviendas del segmento, para lograrlo se usa la siguiente relación.

$$f_{cob2} = \frac{\text{Número de viviendas en el segmento} - \text{viviendas vacantes}}{\text{Número de viviendas en el segmento}}$$

El factor de cobertura final es la multiplicación de los dos anteriores.

$$FCOB = f_{cob1} \cdot f_{cob2}$$

Cabe resaltar que el número de viviendas en el segmento representa al número de viviendas identificadas en el recuento y que no se excluyen ni las vacantes ni las pérdidas en el operativo. Por otro lado, el número de viviendas encuestadas efectivas en el segmento se refiere a las viviendas con encuesta completa que terminan en la base de datos.

El factor también posee una restricción de un máximo de 1.5 para evitar los sesgos generados por sobre-expansión de algunas unidades Muestrales.

Factor de segmento a municipio (FMUNI):

Es un factor asociado a todos los segmentos de un municipio en su cabecera o en su resto (siempre la operación se hace de forma independiente para ambas clases). El factor expande los individuos de la muestra al total de individuos del municipio en su cabecera o resto y tiene la siguiente formula:

$$FMUNI = \frac{Número\ de\ personas\ en\ el\ mpio\ (cabecera\ o\ resto)}{Número\ de\ personas\ en\ la\ muestra\ del\ mpio\ (cabecera\ o\ resto)}$$

Factor de municipio a estrato (FESTR):

Es un factor asociado a todos los municipios de un estrato de muestreo en la primera etapa de la muestra (Estratificación de UPMs). La tarea del factor consiste en expandir la población de los municipios seleccionados en la primera etapa a la población de todos los municipios dentro de sus respectivos estratos. Como el anterior factor, este ponderador se aplica independientemente según cabecera y resto.

$$FESTR = \frac{Número\ de\ individuos\ de\ los\ municipios\ pertenecientes\ al\ estrato\ según\ cabecera\ y\ resto}{Número\ de\ individuos\ que\ conforman\ la\ muestra\ en\ el\ estrato\ según\ cabecera\ y\ resto}$$

Factor básico (d_k):

Es el factor resultado de la multiplicación de los anteriores factores de expansión. Se pone el subíndice k para identificar la vivienda k y seguir la notación de los artículos de calibración (Dewille & Sãmdal, 1996).

$$d_k = PSUB_K \cdot FCOB_K \cdot FMUNI \cdot FESTR$$

Nótese que hasta este punto las viviendas, hogares y personas dentro de un mismo segmento comparten el mismo factor de expansión básico.

Factor calibrado (w_k)

Los factores calibrados son el resultado de un proceso que toma en cuenta un modelo lineal asistido y la información auxiliar de la población proyectada por el DANE con el censo de población y vivienda 2018. La metodología de modelos asistidos se puede revisar en (Sãmdal, Swensson, & Wretman, Model Assisted Survey Sampling, 1992) o en (Gutiérrez, 2009). El propósito de este documento es la especificación técnica del modelo y su aplicación particular. Se invita al lector a consultar la bibliografía relacionada si se requiere entender mejor el proceso.

Para proceder a calcular los pesos calibrados, se define la matriz de diseño del modelo, cuyas columnas identifican cada una de las viviendas y las filas identifican los grupos de edad y sexo. De esta manera, cada celda representa el número de personas que tiene una vivienda particular según un grupo de edad y sexo determinado. Como se tienen 4 grupos de edad y dos de sexo, entonces se tienen 8 filas. Adicionalmente, se crea otra fila con el número de hogares en la vivienda. En resumen, el número de columnas de la matriz es el número de viviendas encuestadas y el número de filas equivale a 9, que es el resultado de la combinación de las categorías de sexo y edad más el número de hogares. A continuación, se definen las categorías asociadas a cada variable de clasificación.

Sexo:

- Hombres.
- Mujeres.

Grupos de edad:

- De 0 años a menores de 15 años.
- 15 años a menores de 25.
- 25 años a menores de 55.
- 55 años y más.

A modo de ejemplo se muestra la siguiente fila transpuesta de la matriz de diseño para una vivienda con 5 personas y 2 hogares.

Vivienda	Hombres < de 15 años	Mujeres < de 15 años	Hombres entre [15 -24) años	Mujeres entre [15 -24) años	Hombres entre [25 -54) años	Mujeres entre [25 -54) años	Hombres de 55 y + años	Mujeres de 55 y + años	Número de hogares
K	1	0	2	0	0	0	2	0	2

Entonces, siendo m el número de viviendas y 9 el número de combinaciones según grupos de sexo y edad más el número de hogares, cuya combinación se identifica con el sub-índice i, la matriz que contiene la información de individuos de la muestra tiene la siguiente forma:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{91} & \cdots & x_{9m} \end{bmatrix}$$

De lo cual se tiene el vector de la vivienda k

$$X'_k = [x_{1k} \ x_{2k} \ \cdots \ x_{ik} \ \cdots \ x_{9k}]$$

Las proyecciones de la población son la información auxiliar que se usa para calibrar. Estas proyecciones entran como un vector que contiene, en sus celdas, las proyecciones de población de los diferentes grupos de edad y sexo predefinidos arriba y la proyección de hogares. La información se pone en un vector de longitud 9:

$$t'_x = [t_1 \ t_2 \ \cdots \ t_i \ \cdots \ t_9]$$

También, se construye un vector con la estimación de la muestra de los 8 grupos de edad y sexo, más la estimación de hogares:

$$\hat{t}'_x = [\hat{t}_{x1} \hat{t}_{x2} \cdots \hat{t}_{xi} \cdots \hat{t}_{x9}]$$

Donde $\hat{t}_{xi} = \sum_s d_k x_{ik}$ es una estimación del total de población de cada uno de los nueve grupos identificados con el sub-índice i , con s conjunto de viviendas de la muestra y d_k factor básico.

Siendo $\hat{T} = \sum_s d_k X_k X'_k$ una matriz 9x9, se definen los pesos g_k como sigue:

$$g_k = 1 + (t_x - \hat{t}_x)' \hat{T}^{-1} X_k$$

Con todo el desarrollo anterior, se construye el factor de calibración:

$$w_k = d_k g_k$$

Nótese que los w_k son iguales para todos los individuos de la vivienda y diferentes entre viviendas.

Nótese que también se pueden construir los w_k con la metodología de estimadores de calibración (Deville & Särndal, 1996), donde se demuestra que el estimador de calibración es asintóticamente equivalente al estimador de regresión bajo la métrica de la *ji-cuadrado*.

Para terminar la especificación técnica de los factores de expansión de la encuesta, se aclara que el modelo de calibración, expuesto acá, no se aplica al general de la población en un solo modelo, sino que se desarrolla un modelo independiente por particiones geográficas. Las particiones donde se define cada modelo se conocen como bloques. No necesariamente los bloques deben coincidir con la estratificación. Adicionalmente, se tienen dos calibraciones la primera se hace mensualmente y toma en cuenta las regiones y la segunda se hace una vez al año y toma en cuenta los departamentos.

Tabla 5. Distribución del número de bloques según área geográfica.

	Número de Bloques	
	Mensual regional	Anual departamental
Cabeceras de ciudades capitales de departamento	32	32
Cabeceras de áreas metropolitanas	7	7
Cabeceras de 5 ciudades intermedias	5	5
Resto de Cabeceras (Regiones)	4	23
Resto de Centro poblado y rural disperso (Regiones)	4	23

El centro poblado y rural disperso de Bogotá entra en el bloque de la región oriental cuando se calibra cada mes y en Cundinamarca cuando se calibran los departamentos. Las proyecciones usadas para calibrar cada mes son de fecha de mitad de mes. Para las calibraciones anuales se aplica 30 de junio. Acá las ciudades, áreas metropolitanas y ciudades intermedias se calibran con el promedio anual.

Estimadores

Los estimadores son funciones con las cuales se estiman parámetros poblacionales. Para la encuesta se presentan los estimadores de totales y razones con sus respectivos estimadores de varianza. Adicionalmente, con los anteriores se presentan los estimadores de los errores muestrales.

Estimador de Totales

El estimador del total para la variable y con la muestra s en el estrato h tiene la siguiente forma:

$$\hat{t}_{y_h} = \sum_{s_h} w_l y_l$$

Con l el sub-índice que representa al individuo.

El estimador del total poblacional de la variable y es entonces:

$$\hat{t}_y = \sum_{h=1}^H \hat{t}_{y_h}$$

En los casos en los cuales se requiera estimar grupos especiales de la población (dominios de estudio), la estrategia es crear otra variable que marque con 1 los individuos que pertenecen al grupo y cero los que no. Posteriormente, esta variable se multiplica a la variable y creando una nueva variable llamémosla y^* . En este punto se procede de la misma manera para estimar.

$$\hat{t}_{y^*} = \sum_{h=1}^H \hat{t}_{y_h^*}$$

Cuando se quiere estimar unidades diferentes a personas, como es el caso de hogares o viviendas, la estrategia es similar a la de dominio de estudio y consiste en crear una variable con un uno para la primera persona de cada una de estas unidades más grandes y cero para las demás personas dentro de la misma unidad.

Estimador de varianza del estimador del total

La encuesta no tiene un estimador de varianza exacto, ya que es un diseño muestral complejo, donde en su última etapa se hace un sistemático de único arranque. Por lo tanto, el acercamiento a la varianza se plantea a través del método de conglomerado último (Hansen, Hurwitz, & Madow, 1953) páginas 257-259.

$$\hat{v}(\hat{t}_y) = \sum_{h=1}^H \left(\left(\frac{n_h}{n_h - 1} \right) \sum_{i=1}^{n_h} \left(\hat{t}_{\hat{e}_{hi}} - \frac{\hat{t}_{\hat{e}_h}}{n_h} \right)^2 \right)$$

De donde i es el índice asociado a cada segmento de la muestra y toma valores desde 1 hasta n_h , el cuál es el número de segmentos dentro del estrato h . En algunos casos se adiciona a la formula el factor de finitud $(1 - N_h/n_h)$.

Si l es el índice que identifica cada individuo, k es el índice que identifica cada vivienda dentro del segmento i y L_i el número de individuos que pertenecen al segmento i .

Entonces:

$$\hat{t}_{\hat{e}_{hi}} = \sum_{l=1}^{l_{hi}} w_l \hat{e}_l$$

N_h el número de segmentos en el estrato h .

$$\hat{t}_{\hat{e}_h} = \sum_{i=1}^{n_h} \sum_{l=1}^{l_{hi}} w_l \hat{e}_l$$

$$\hat{e}_{l=} y_k - X'_k \hat{B}$$

$$\hat{B} = \hat{T}^{-1} \hat{t}$$

$$\hat{T} = \sum_{k=1}^K d_k X_k X'_k$$

$$\hat{t} = \sum_{k=1}^K d_k X_k y_k$$

Con y_k el agregado de individuos de la variable y en la vivienda k y K número de viviendas en el bloque donde se hace el modelo.

Estimador de Razones

El estimador de razón se construye haciendo el cociente de los estimadores de dos totales. Es así que si se quiere calcular la razón de los totales t_y y t_z se tiene la siguiente formula:

$$\hat{R} = \frac{\hat{t}_y}{\hat{t}_z}$$

Estimador de varianza del estimador de razones

El estimador de varianza de una razón es el mismo que el de estimador del total sólo que el total que se toma en cuenta es una variable producto de una transformación lineal. Más exactamente se usa la linealización de $\hat{t}_u$, donde $u_l = \frac{1}{\hat{t}_z} (y_l - z_l \hat{R})$.

Taylor para poner a $\hat{R}$ como combinación lineal de totales y se procede a calcular la varianza de $\hat{t}_u$. Con l sub-índice que identifica al individuo.

Entonces la varianza de una razón se define como:

$$\hat{v}(\hat{R}) = \sum_{h=1}^H \left(\left(\frac{n_h}{n_h - 1} \right) \sum_{i=1}^{n_h} \left(\hat{t}_{\hat{e}_{hi}} - \frac{\hat{t}_{\hat{e}_h}}{n_h} \right)^2 \right)$$

De donde i es el índice asociado a cada segmento de la muestra y toma valores desde 1 hasta n_h , el cuál es el número de segmentos dentro del estrato h .

Si l es el índice que identifica cada individuo, k es el índice que identifica cada vivienda dentro del segmento i y L_i el número de individuos que pertenecen al segmento i .

Entonces:

$$\hat{t}_{\hat{e}_{hi}} = \sum_{l=1}^{l_{hi}} w_l \hat{e}_l$$

N_h El número de segmentos en el estrato h

$$\hat{t}_{\hat{e}_h} = \sum_{i=1}^{n_h} \sum_{l=1}^{l_{hi}} w_l \hat{e}_l$$

$$\hat{e}_l = u_k - X'_k \hat{B}$$

$$\hat{B} = \hat{T}^{-1} \hat{t}$$

$$\hat{T} = \sum_{k=1}^K d_k X_k X'_k$$

$$\hat{t} = \sum_{k=1}^K d_k X_k u_k$$

Con u_k el agregado de individuos de la variable u en la vivienda k y K número de viviendas en el bloque donde se hace el modelo.

Donde u_k es el agregado de individuos de la variable u en la vivienda k y K número de viviendas en el estrato donde se hace el modelo.

Errores muestrales

Son varias las formas de medir la precisión de los indicadores. En la investigación se presentan los coeficientes de variación estimados (errores relativos) y los errores marginales de cada indicador. Los estimadores de total y razón se asumen con una distribución normal y sus medias y varianzas son las presentadas en la anterior sección.

El coeficiente de variación se define como la relación porcentual del error estándar del estimador y el valor esperado del estimador. Se puede estimar de la siguiente forma:

$$CVe(\hat{t}) = \frac{\sqrt{\hat{v}(\hat{t})}}{\hat{t}} \text{ Coeficiente de Variación estimado para totales.}$$

$$CVe(\hat{R}) = \frac{\sqrt{\hat{v}(\hat{R})}}{\hat{R}} \text{ Coeficiente de Variación estimado para razones.}$$

El valor de este coeficiente, expresado en porcentaje, permite evaluar la calidad de un procedimiento de estimación. Se consideran como excelentes las mediciones menores de 5%, buenas las que están entre 5% y 10% y aceptables las que están entre 10% y 15%.

El error marginal mide la amplitud media del intervalo de confianza. Con este valor se puede definir el límite inferior y superior del intervalo. Para una confianza del 95% la fórmula de la estimación es:

$$EMe(\hat{t}) = 1.96 \sqrt{\hat{v}(\hat{t})} \text{ Error Marginal estimado para totales.}$$

$$EMe(\hat{R}) = 1.96 \sqrt{\hat{v}(\hat{R})} \text{ Error Marginal estimado para razones.}$$

2.3. DISEÑO DE LA RECOLECCIÓN/ ACOPIO

A partir de la realización del Censo Nacional de Población y Vivienda (CNPV) de 2018, se hizo necesario contar con un proceso de adaptación de las encuestas realizadas por el DANE, el cual permitiría adaptar el diseño muestral de las encuestas a la nueva realidad establecida por el CNPV.

Adicionalmente, se requirió poner a prueba el nuevo formulario de recolección, la implementación de los planes de datos para la transmisión de la información de campo y la efectividad de establecer controles automatizados para efectuar seguimiento al desarrollo del operativo. Para esto, se realizaron pruebas de campo y experimentales que permitieron evaluar los cambios requeridos y por incorporar en el diseño muestral y temático de la encuesta.

El diseño de las fases de recolección de información comprende la definición de los roles, perfiles y cantidad de personal requerido para el operativo de campo, el calendario de recolección, el entrenamiento del personal que intervendrá en el proceso y los procedimientos para la recolección de los datos.

2.3.1. Métodos y estrategias de recolección o acopio de datos

La recolección de la información debe realizarse de acuerdo con las normas y conceptos establecidos por el DANE Central. La recolección tiene por objeto: obtener la información requerida de los hogares, con el fin de medir las principales características e indicadores del mercado laboral, ingresos y pobreza monetaria.

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

El sistema de recolección utilizado es el de "barrido", que consiste en que cada equipo de trabajo recorre simultáneamente un segmento hasta encuestar todas las viviendas seleccionadas con sus respectivos hogares y número de personas, incluyendo los que puedan aparecer como nuevos y no se encuentren en el listado de viviendas seleccionadas (en este último caso, se debe realizar el reporte de la novedad a DANE Central). Este sistema de trabajo contribuye en la eficacia de la supervisión de campo y en las diversas actividades de recolección.

La recolección de la información se hace en los formularios electrónicos mediante un Dispositivo Móvil de Captura (DMC), el cual contiene también la cartografía de la ciudad, georreferenciación, listas o tablas de validaciones y control de flujos.

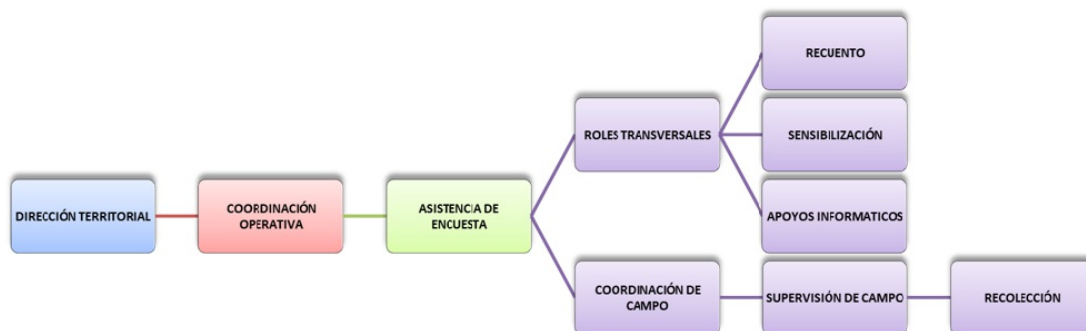
¹⁸ De manera excepcional, se permite tomar información a través del informante idóneo para los casos donde no es posible encontrar a las personas del hogar que trabajan o están buscando trabajo, por ejemplo, cuando no se encuentran disponibles en la semana de la encuesta por motivos como viajes, trabajo o enfermedad.

2.3.2. Estructura organizacional del operativo y conformación del equipo

En las direcciones territoriales y sedes del DANE, se cuenta con una organización operativa que le permite adelantar eficientemente los operativos de campo. La siguiente figura presenta la estructura operativa para la recolección de información y posterior transmisión de los datos:

Figura 4. Organigrama operativo





Fuente: DANE.

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.

2.3.3. Esquema de entrenamiento de 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 comprensión 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.

2.3.4. 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.

2.3.5. 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.

2.3.6. Elaboración de manuales

Con el fin de garantizar un óptimo desarrollo de los operativos de campo para cada encuesta se han elaborado los siguientes manuales:

Manual de diligenciamiento de formatos: presenta de forma clara el diligenciamiento de los formatos operativos que deben ser utilizados por el personal de campo, como son:

- Listado de recuento de unidades estadísticas.
- Listado de unidades seleccionadas.
- Resumen acumulado de viviendas, hogares y personas por segmentos.
- Informe diario de la supervisión de campo.
- Reentrevista.
- Informe técnico coordinación de campo.
- Segmentos no trabajados.
- Novedades cartográficas.
- Segmentos agotados.
- Asignación y control diario de áreas de trabajo.
- Resumen operativo de cobertura.
- Hoja de croquis del segmento.
- Errores y alarmas.
- Justificación de las variaciones de las tasas de mercado laboral.
- Justificación de las variaciones de las tasas de cobertura.

Manual del Coordinador de Campo: describe las actividades y consideraciones a tener en cuenta en la coordinación en campo, proporcionando una guía de las acciones a realizar en este aspecto.

Manual del Supervisor de Campo: describe las actividades y consideraciones para tener en cuenta en la supervisión en campo, proporcionando una guía de las acciones a realizar en este aspecto.

Manual de Recolección y Conceptos Básicos: detalla los procesos de recolección de la información y aclara conceptos temáticos y técnicos de la encuesta. También da pautas para realizar las entrevistas, aclara los períodos de referencia y los tipos de preguntas, y explica cada uno de los capítulos, pregunta por pregunta.

Manual de Recuento: contiene los diferentes aspectos que se deben tener en cuenta para garantizar el correcto desarrollo y seguimiento del proceso de recuento en las áreas seleccionadas en la muestra de las diferentes operaciones estadísticas.

Manual del Sistema: contiene la descripción de las características técnicas generales del Dispositivo Móvil de Captura (DMC), los diferentes procedimientos que se deben llevar a cabo para la correcta instalación y configuración del aplicativo, además de las instrucciones necesarias para el proceso de transmisión de información y para el diligenciamiento del formato de apoyo de sistemas, necesario para el control de envíos de información al DANE Central.

Manual del Usuario: contiene la descripción de los procedimientos utilizados para el diligenciamiento de información en el aplicativo de captura, la realización de copias de seguridad y el cierre de encuestas, facilitando al usuario la recolección de la información y el manejo de cada una de las preguntas de la encuesta.

Manual de Diligenciamiento de Formatos: Presenta de una manera clara y precisa el diligenciamiento de los formatos tipo registro que deben ser utilizados por el personal de campo de la Gran Encuesta Integrada de Hogares-GEIH.

2.3.7. Diseño de las estrategias de comunicación y plan de contingencias

Como estrategia de comunicación, se tendrá una sensibilización con enfoque de relacionamiento en las tres fases del proceso: preoperativo, operativo y posoperativo, dentro del cual se busca:

- Dar a conocer la importancia del DANE y de la encuesta.
- Identificar a las entidades gubernamentales, autoridades locales, agentes externos y líderes comunitarios, logrando establecer comunicación y apoyo en el relacionamiento con la comunidad donde se aplica la recolección.
- Sensibilizar a cada integrante del hogar, estableciendo comunicación directa con todos los miembros e iniciando con el jefe de hogar.
- Reportar las novedades presentadas que dificulten el acceso a las fuentes, como, por ejemplo: rechazos, restricción de acceso a conjuntos residenciales, territorios étnicos o de otros grupos de la comunidad.
- Mantener un relacionamiento adecuado con las fuentes durante la recolección de información dejando "las puertas abiertas" al DANE para próximas encuestas.

Productos definidos para la sensibilización

- Cartas de presentación: Dirigidas a las Alcaldías, autoridades locales, policía nacional, agentes externos, líderes comunitarios, Juntas de Acción Comunal, líderes de organizaciones locales, administradores, consejo de administración, porteros y/o personal de vigilancia, para contar con la colaboración necesaria para el acceso a los hogares.
- Afiches o Carteleras informativas: Dirigidas a los residentes de conjuntos residenciales o edificios con portería con el propósito de informar a los habitantes sobre la encuesta y las fechas en que el equipo operativo se encontrará en ese lugar.
- Piezas de Sensibilización: Dirigidas a los miembros de los hogares a encuestar con el propósito de informar a las personas el objetivo de la encuesta y los resultados que se pueden obtener con la información que suministran.
- Notificaciones: Dirigidas a los miembros de los hogares seleccionados en la muestra de la GEIH, con el propósito de informarles que su hogar fue seleccionado para ser encuestado.

Modificaciones por afectaciones externas o razones de fuerza mayor

En caso de una situación de fuerza mayor que impida el desarrollo habitual de la operación estadística y para garantizar la recolección de la información de la GEIH, el DANE optará por tomar medidas extraordinarias como la recolección de información de manera telefónica. Esto, con el fin de garantizar la continuidad en las series de mercado laboral y la publicación de los principales indicadores.

Sin embargo, la recolección de información de manera telefónica (por ejemplo) tiene implicaciones en la aplicación completa de toda la encuesta. Es decir, se reducen el número de preguntas del formulario y por ende no podrá disponerse de toda la información que se recolecta habitualmente de manera presencial en los hogares. Una alternativa para la recolección de información es el aplicativo web que permite registrar la información que se recolecta de manera telefónica y carga el formulario en forma digital, validaciones y flujos de consistencia.

El DANE cuenta con un Manual de Recolección y Conceptos Básicos de la Gran Encuesta Integrada de Hogares (proceso de recolección telefónica), donde se encuentran las recomendaciones y aspectos importantes a la hora de realizar un operativo de estas características y se definen las preguntas del formulario realizadas durante estos períodos.

Este tipo de medidas adoptadas por el DANE son de carácter temporal y buscan dar continuidad y rigurosidad a la información estadística. En el momento que sea posible retomar los operativos presenciales, el DANE garantizará nuevamente la recolección y aplicación completa del formulario.

2.3.8. 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.

$$\text{Indicador de cobertura de viviendas} = \frac{\text{Viviendas finales}}{\text{Viviendas iniciales}} * 100$$

Cobertura de hogares: se obtiene de cruzar la información del total de hogares encuestados contra el total de hogares con encuesta completa a nivel nacional, por dominio y por ciudad.

$$\text{Indicador de cobertura de hogares} = \frac{\text{Hogares con encuestas completas}}{\text{Total de hogares}} * 100$$

Cobertura de personas: se obtiene de cruzar la información del total de personas encuestadas contra el total de personas con encuesta completa a nivel nacional, por dominio y por ciudad.

$$\text{Indicador de cobertura de personas} = \frac{\text{Personas con encuestas completas}}{\text{Total de personas}} * 100$$

Indicador de Calidad de la Recolección (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.

2.3.9. Diseño de sistemas de captura

La captura de datos es el proceso en el que un grupo de encuestadores obtiene la información requerida de los hogares y viviendas asignadas, registrándola en el Dispositivo Móvil de Captura (DMC), conforme a las normas y conceptos establecidos. Esto permite diligenciar y enviar los formularios para la consolidación de las bases de datos.

2.3.10. Transmisión de datos

Cada sede debe garantizar la transmisión de toda la información capturada y depurada de las encuestas realizadas durante los operativos de campo hacia DANE Central; para lo cual, al finalizar la jornada diaria de recolección, el supervisor de campo se encarga de descargar las encuestas de cada uno de los DMC y entregarlas al Apoyo Informático. Este a su vez deja en su computador una copia de dicha información y la empaqueta en un archivo .zip/.rar bajo una estructura de directorios definida, asignándole un nombre que indica la operación, la ciudad, la fecha y el código del recolector a la que corresponde la información contenida.

La transmisión de estos paquetes de datos, la realiza el Apoyo Informático a través del buzón FTP (del inglés File Transfer Protocol - Protocolo de Transferencia de Archivos), el cual es un protocolo de red para la transferencia de archivos entre sistemas conectados a una red TCP (del inglés Transmission Control Protocol - Protocolo de Control de Transmisión), basado en la arquitectura cliente-servidor: desde un equipo cliente se puede conectar a un servidor para descargar archivos desde él o para enviarle archivos, independientemente del sistema operativo utilizado en cada equipo. Cada dirección territorial tiene un buzón independiente al cual se le ha asignado un usuario y contraseña. Una vez que el Apoyo Informático ha enviado la información, notifica por correo electrónico al ingeniero de la oficina de Sistemas encargado de la operación estadística para que éste haga el descargue de los datos y correspondiente backup, como plan de contingencia que garantice el normal funcionamiento del operativo.

2.4. DISEÑO DE PROCESAMIENTO

2.4.1. 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.

2.4.2. 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 revisión 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.

2.4.3. 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.

2.4.4. 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.

2.4.5. 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.

2.4.6. 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.

2.5. DISEÑO DEL ANÁLISIS

2.5.1. 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.

2.5.2. 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 agrupaciones que están sujetas a reglas de confidencialidad nacional, deben mantenerse estrictamente confidencial y ser usadas exclusivamente para fines estadísticos o para propósitos establecidos por la legislación*"¹⁹, 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.

¹⁹ http://unstats.un.org/unsd/methods/statorg/Principles_stat_activities/principles_stat_activities.asp

2.5.3. 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.

2.5.4. 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.

2.6. DISEÑO DE LA DIFUSIÓN Y COMUNICACIÓN

La difusión y la comunicación son las estrategias que determinan la planificación, creación y gestión de contenidos institucionales y la divulgación de la información estadística producida por la entidad y se ocupa de las formas de ponerlos en circulación entre la sociedad.

2.6.1. Diseño de sistemas de salida

Los productos estadísticos para la difusión contienen los resultados de la encuesta y satisfacen las necesidades de información de los grupos de interés. Estos productos consolidan la información estadística de acuerdo con las formas de difusión establecidas en el diseño.

En la fase de difusión, se elaboran las herramientas para poner a disposición de los grupos de interés los productos estadísticos, respondiendo a las especificaciones definidas en la fase de diseño de la operación estadística. Para ello, el equipo temático responsable debe:

- Definir claramente para cada producto estadístico, a quiénes va dirigido, cómo se accede, protocolos de actualización y datos de contacto para ampliar la información del producto.
- Desagregar temática y geográficamente cada producto estadístico al máximo nivel posible.
- Proveer los metadatos de manera que describan las características del conjunto de datos por medio de fichas metodológicas, documentos metodológicos, manuales, guías, instructivos, informes y demás documentos que permitan comprender cómo se han construido, obtenido y procesado los datos, incluyendo descripciones, aclaraciones o precisiones conceptuales y técnicas sobre el origen de los datos, las variables y los indicadores que contienen, así como el lugar y la fecha o período al que se refieren.
- Especificar las nomenclaturas y las clasificaciones estadísticas que se emplearon y su respectiva versión.
- Garantizar que los datos agregados sean consistentes con las bases de datos.
- Salvaguardar la confidencialidad y la reserva estadística para proteger la identidad de las fuentes y sus datos sensibles en todos los productos y servicios.
- Usar un lenguaje claro para la presentación de los resultados, incluyendo anexos, tablas, gráficos, etc.

2.6.2. Diseño de productos de comunicación y difusión

Los resultados de mercado laboral producidos por la GEIH son presentados mensualmente en rueda de prensa, luego de finalizada los usuarios pueden realizar preguntas mediante los canales de difusión virtuales del DANE (Facebook, YouTube, entre otros) sobre la información entregada, las cuales son resueltas inmediatamente. Asimismo, pueden consultar a través de la página web de la entidad los anexos estadísticos, los anexos desestacionalizados, el boletín técnico, el comunicado de prensa y la presentación de resultados.

Las publicaciones de la GEIH se realizan periódicamente cada fin de mes, con un mes de rezago con respecto al mes de recolección de la información.

La publicación mensual está compuesta por:

- Indicadores del mercado laboral mensual para el total nacional, total 13 ciudades y áreas metropolitanas y cabeceras.
- Comportamiento del mercado laboral mensual para el total nacional y total 13 ciudades y áreas metropolitanas.
- Indicadores del mercado laboral trimestre acumulado para el total nacional, total 13 ciudades y áreas metropolitanas, cabeceras, centros poblados y rural disperso y 23 ciudades y áreas metropolitanas.
- Indicadores del mercado laboral trimestre acumulado para 7 ciudades sin área metropolitana.
- Comportamiento del mercado laboral para el trimestre acumulado para el total 13 ciudades y áreas metropolitanas y 23 ciudades.
- Indicadores del mercado laboral año móvil para las 32 ciudades.
- Tasa desempleo trimestral por sexo para el total nacional.
- Principal actividad de la población fuera de la fuerza laboral mensual para el total nacional.
- Principal actividad de la población fuera de la fuerza laboral trimestre acumulado para el total 13 ciudades y áreas metropolitanas y cada una de ellas.
- Otras formas de trabajo mensual para el total nacional.

A continuación, se describen los principales productos de comunicación y difusión:

Boletín técnico

Contiene los principales indicadores de mercado laboral y aspectos de la GEIH, los cuales se presentan con una periodicidad mensual. Adicionalmente, se elaboran boletines de mercado laboral con temáticas especiales, los cuales se publican de manera mensual, semestral o anual en la página web del DANE.

Series históricas

Contienen la información secuencial y cronológica de los principales indicadores del mercado laboral, así como las notas explicativas donde se describen la estructura de los datos, los documentos de referencia, las metodologías de tratamiento de las series y la información sobre la cobertura geográfica. Los cambios metodológicos que se realizan para el empalme de las series históricas se documentan a su vez en los anexos estadísticos y en los comunicados y boletines de prensa. Se publican con una periodicidad mensual en la página web del DANE.

Anexos

Contienen los resultados de los principales indicadores del mercado laboral, y corresponde a la presentación en PowerPoint que se muestra en la rueda de prensa. Se publica con una periodicidad mensual en la página web

del DANE al finalizar cada rueda de prensa, junto con el boletín técnico y el comunicado de prensa.

Presentación de rueda de prensa: Contiene los principales resultados de mercado laboral para total nacional y total 13 ciudades, y las tasas desestacionalizadas.

Comunicados de prensa

Sintetizan la información de los boletines técnicos y contienen la información más relevante de los indicadores del mercado laboral publicados mensualmente en la página web del DANE. Se envían periódicamente a los medios de comunicación masiva para la difusión de los resultados.

Los productos estadísticos (boletín técnico, anexos estadísticos y cuando haya lugar a presentación) son validados por la Dirección de Metodología y Producción Estadística (DIMPE) y el Comité Técnico del DANE para Mercado Laboral. El comunicado de prensa es elaborado y aprobado por la Dirección de Difusión y Cultura Estadística (DICE).

2.6.3. Entrega de productos

Una vez publicados los anexos estadísticos, anexos desestacionalizados, boletines técnicos, comunicados y presentaciones en la página web del DANE, estos quedan a disposición de los diferentes usuarios de la información y la ciudadanía en general.

A continuación, se describen algunos medios de acceso a la información estadística:

Página web de la entidad, <http://www.dane.gov.co>

El principal medio de difusión es el portal web como medio oportuno y de fácil acceso para el público en general. La información estadística de la GEIH se encuentra publicada en la sección Mercado Laboral.

Calendario anual de difusión

Se programa para los doce meses año corrido y es avalado por la dirección técnica correspondiente o a quien se delegue esta aprobación. Se publica en la página web de la entidad.

Redes sociales

Son usadas para la divulgación y circulación de información estadística, al permitir el manejo de hipervínculos y de diferentes formatos de presentación como gráficas, datos sencillos, videos, audios, infografías, tablas y contenidos en formato multimedia.

Las redes sociales son accesibles a públicos diversos e incluso posibilitan la difusión a públicos específicos y la interacción a través de una rápida retroalimentación de los mensajes enviados. En el siguiente listado se mencionan algunas de ellas y los usos que se han dado para la publicación de los resultados:

- Facebook: infografías, videos, fotografías y enlaces a noticias.
- Twitter: infografías, gráficas, hipervínculos y mensajes de texto cortos.
- YouTube: videos y ruedas de prensa.
- Instagram: infografías, gráficas e imágenes.

Respecto al uso de redes sociales, se realiza un monitoreo constante de los mensajes para conocer de primera mano la respuesta de la difusión y la interpretación que se genera en la opinión pública. Las redes sociales son usadas como un apoyo a los medios de difusión, en tanto que, las piezas gráficas y herramientas publicadas disponen de enlaces que dirigen directamente a los resultados en la página web de la entidad.

A su vez, el DANE cuenta con el Grupo Interno de Trabajo de Información y Servicio al Ciudadano, adscrito a la DICE, el cual tiene como función coordinar la implementación de la Política Nacional de Servicio al Ciudadano, de acuerdo con los lineamientos definidos por el Sistema Nacional de Servicio al Ciudadano. Frente a las inquietudes de los ciudadanos, la respuesta se da conforme a la normatividad vigente en los siguientes canales de atención:

Línea telefónica gratuita, correo electrónico y chat para la atención a los grupos de interés

Los canales de atención dispuestos por el DANE para los grupos de interés de las operaciones estadísticas que produce la entidad, con el fin de garantizar alternativas de acceso a la información estadística son: la línea telefónica, el correo electrónico y el chat en la página web para responder a sus inquietudes.

Es de advertir que los datos especialmente protegidos, tales como los datos sensibles que hayan sido suministrados, cuentan con un nivel de seguridad alto, además, se informa que las preguntas que se generen sobre este tipo de información sensible no son de respuesta obligatoria, garantizando los derechos constitucionales de privacidad, intimidad, igualdad y no discriminación.

Procesamientos especializados

El DANE cuenta con dos modalidades para el procesamiento de datos especializados:

- Información a la medida, procesada de acuerdo con las especificaciones de los usuarios.
- Sala de procesamiento, donde los grupos de interés acceden a los repositorios de almacenamiento, específicamente a las bases de microdatos confidenciales que cuentan con un nivel de anonimización diferente al de uso público en la web, para que realicen la consulta y el procesamiento de los microdatos, cumpliendo con los controles necesarios para garantizar la confidencialidad de las fuentes.²⁰

²⁰ El DANE actualizó el diseño de la interfaz del Archivo Nacional de Datos (ANDA) para que los usuarios puedan explorar, buscar, comparar, solicitar acceso y descargar información relacionada con censos, encuestas por muestreo y uso estadístico de registros administrativos, así como solicitar acceso y descargar información relacionada con las series históricas de las operaciones estadísticas a nivel de microdato anonimizado de acceso público, junto con sus correspondientes metadatos. Para la GEIH, se disponen en el ANDA los metadatos de la encuesta (manuales, formularios y metodología, entre otros), así como los microdatos anonimizados para acceso público.

2.6.4. Estrategia de servicio

El manejo de los medios de comunicación responde a la planeación anual consignada en el calendario de difusión avalado por las direcciones técnicas y se ratifica con el envío del calendario semanal. Para el caso de los productos resultantes de convenios o publicaciones especiales no continuas, su fecha de difusión se plasma en el calendario una vez sea enviada por el director técnico del área responsable de la operación estadística, previo aval de la Dirección General del DANE.

Para realizar esta labor, se generan mensajes y piezas de comunicación para la difusión de la información estadística en estrecha articulación con los técnicos productores de la operación estadística y se realizan sesiones especiales con los medios para generar comprensión sobre las diferentes operaciones que realiza la entidad y sus resultados, sobre su evolución técnica y conceptual, las mejoras metodológicas, etc.

Todos los productos de las operaciones estadísticas deben ser enviados para difusión desde las direcciones técnicas con su respectivo aval. El Comité Técnico establece que la GEIH se considera como una operación estadística de alto impacto, por lo que los resultados mensuales son presentados durante una rueda de prensa; además, los usuarios pueden consultar a través de la página web del DANE los anexos estadísticos, boletines técnicos y presentaciones que contienen los indicadores de mercado laboral de los diferentes dominios geográficos que cubre la operación.

2.7. DISEÑO DE LA EVALUACIÓN DE LAS FASES DEL PROCESO

La evaluación realizada a la GEIH es un proceso continuo desarrollado por la Dirección de Recolección y Acopio, la Oficina de Sistemas, el GIT de Gestión de Geoinformación, el GIT de Diseños Muestrales de Estadísticas Sociales, el GIT de Temática de Mercado Laboral y la Dirección de Difusión y Cultura Estadística, el cual se lleva a cabo durante las diferentes etapas de producción de la operación estadística, desde el diseño hasta los procesos de recolección, procesamiento, análisis y difusión.

Se evalúan las actividades realizadas en la operación, la documentación, los instrumentos de recolección, el funcionamiento de los equipos y los dispositivos de captura, funcionalidad de los aplicativos informáticos desarrollados y demás software empleado, el funcionamiento y la capacidad de los equipos tecnológicos utilizados, así como, la capacidad de almacenamiento, la seguridad de la información y la efectividad de los controles aplicados.

Asimismo, se valida la información que se va a difundir, se evalúan los métodos de análisis empleados y las actividades desarrolladas, finalmente, se revisa la disposición de la totalidad de la información requerida y se evalúan los productos, estrategias y medios de difusión utilizados, adicionalmente se realizan consultas sobre la satisfacción de las necesidades de información estadística de los usuarios y se plantean recomendaciones para desarrollar acciones de mejora en una siguiente iteración de la encuesta.

Criterios de seguimiento y evaluación de cada fase

Diseño de la operación estadística

- Seguimiento a las necesidades de los usuarios a través de comités internos y externos.
- Se verifica el uso de los conceptos estandarizados y de la implementación de las clasificaciones.
- Se verifica la implementación de los estándares estadísticos en los cuadros de salida y en la codificación de la base de datos.
- Se revisan continuamente las reglas de validación para aplicarlas posteriormente en la revisión de información.
- Seguimiento al cumplimiento del presupuesto asignado.

Diseño estadístico

- Se hace revisión de la cobertura geográfica.
- Se calcula el tamaño de la muestra con un error de muestreo relativo trimestral de 5,5% en la tasa de desempleo de las ciudades y de 4,7% anual de la pobreza monetaria en los departamentos.

Diseño de la recolección/acopio

- Las mencionadas en el numeral 2.3.8. Diseño de la estrategia de seguimiento y control.
- Indicadores de confiabilidad (cobertura de viviendas y hogares) y de calidad (tasa de respuesta).
- Estimadores de precisión de resultados, con un nivel de confiabilidad del 95%.

Diseño del procesamiento

- Transmisión de encuestas completas.
- Cumplimiento del cronograma para transmisión de información.
- Verificación de los cuadros de salida.
- Verificación de las variables a publicar en bases de datos anonimizadas - microdatos.

Diseño del análisis

- Revisión de las bases de datos y cuadros de salida.
- Revisión de los resultados de los indicadores generados.
- El grupo temático revisa la consistencia de la información sobre los principales Indicadores de Mercado Laboral.

Diseño de la difusión y comunicación

- Cumplimiento de cronograma de publicación de información, boletines, anexos y microdatos anonimizados.
- Se actualizan los documentos que se publican en la plataforma del archivo nacional de datos (ANDA).
- Se verifica la realización de rueda de prensa, comunicado de prensa y presentación de mercado laboral.
- Se verifica el acceso y el cargue correcto de las publicaciones realizadas en la página web: boletines, anexos y microdatos anonimizados, en cumplimiento con el calendario de publicaciones.
- Los requerimientos de los usuarios se registran en ORFEO y se realiza seguimiento para dar respuesta oportuna.

Adicional al seguimiento, control y evaluación realizados por los diferentes grupos que intervienen en los procesos de la operación estadística, se llevan a cabo auditorías por parte de la Oficina de Control Interno del DANE, del Instituto Colombiano de Normas Técnicas y Certificación (ICONTEC) y de la Contraloría General de la República, así como, encuestas de satisfacción.

Evaluación

Informe de evaluación consolidado, el cual registra aspectos relevantes presentados durante el período a documentar. Se presentan hallazgos y soluciones o mejoras realizadas. En caso de tener algún hallazgo que requiera un plan específico, se diseña y se le hace seguimiento continuo de avance.

2.8. DISEÑO DE LOS SISTEMAS DE PRODUCCIÓN Y FLUJOS DE TRABAJO

En el documento Descripción del Modelo Funcional de la Gran Encuesta Integrada de Hogares, se identifican y documentan los procesos y subprocesos necesarios para el funcionamiento de esta operación y la obtención de los productos. Dicho documento, describe el flujo de información desde las sedes, incluyendo las direcciones territoriales, hasta la recepción en el DANE Central. También contiene la representación gráfica de la interacción e interrelación de los procesos y subprocesos de la operación, así como de las áreas que involucra, mediante los diagramas Jerárquico Funcional, de Contexto y de Nivel.

3. DOCUMENTACIÓN RELACIONADA

Los documentos relacionados con el diseño temático incluyen:

- Especificaciones de consistencia y validación: empleadas como guía para evaluar la información reportada en los instrumentos de recolección.
- Manuales de recolección, operativo, de usuario, coordinador, supervisor y recuento, los cuales orientan el trabajo de campo realizado en las direcciones territoriales.
- Formulario de recolección: instrumento por medio del cual se recolecta la información.

GLOSARIO

Asalariado. Trabajador que posee un contrato de trabajo oral o escrito, por el cual percibe una remuneración en dinero o en especie de manera estable o regular (DANE, 2009).

Edificación. Es una construcción independiente y separada, compuesta por uno o más espacios en su interior independiente, porque tiene acceso directo desde la vía pública, caminos, senderos o a través de espacios de circulación común (corredores o pasillos, escaleras, patios) Separada, porque tiene paredes, sin importar el material utilizado para su construcción, que la delimitan y diferencian de otras Una edificación puede tener varias entradas y generalmente está cubierta por un techo (ONU, 1998).

Hogar. Es una persona o grupo de personas, parientes o no, que: ocupan la totalidad o parte de una unidad de vivienda; atienden necesidades básicas con cargo a un presupuesto común y generalmente comparten las comidas (ONU, 1998).

Jefe o Jefa de hogar. Es el residente habitual que es reconocido por los demás miembros del hogar como "jefe(a)" (ONU, 1998).

Muestra. Parte de una población, o un subconjunto de un grupo de unidades las cuales son suministradas por algún proceso u otro medio, usualmente por selección intencional de las propiedades de la población, o de un conjunto de ella, según el objeto de la investigación. (Adaptado, OCED, 2008).

Patrón o empleador(a). Son las personas que dirigen su propia empresa económica o ejercen por su cuenta una profesión u oficio, utilizando uno(a) o más trabajadores(as) remunerados(as), empleados(as) y/u obreros(as).

Población Objetivo. Conjunto de elementos de los que se desea obtener los datos y sobre los que se presentan conclusiones o resultados; está compuesta por unidades que comparten alguna característica, tienen una localización geográfica y un tiempo como periodo de referencia (DANE, 2008).

Trabajador por Cuenta Propia. Son las personas que explotan su propia empresa económica o que ejercen por su cuenta una profesión u oficio con ayuda o no de familiares, pero sin utilizar ningún(a) trabajador(a) (empleado(a) u obrero(a)) remunerado(a). Estas personas pueden trabajar solos o asociados con otras de igual condición (DANE, 2019).

BIBLIOGRAFÍA

Cochran, W. G. Sampling Techniques, 1977.

Departamento Administrativo Nacional de Estadística (DANE). Conceptos estadísticos - Sistema de consulta, 2022.

Departamento Administrativo Nacional de Estadística (DANE). Guía para la elaboración del documento metodológico de operaciones estadísticas, 2020.

Departamento Administrativo Nacional de Estadística (DANE). Manual de Recolección y Conceptos Básicos Gran Encuesta Integrada de Hogares, 2022.

Eurostat. Conceptos y definiciones de la base de datos (CODED).

Eurostat. Guía y programación del programa comunitario de la población y censos de hogares en 2001 vol. 1: Guidelines.

Freire, E. La medición del mercado laboral en Colombia. *Magazín IB de la gestión estadística* p.50-54, 2013.

Gómez, R., Suárez, A., Tinjaca, N. y Cárdenas, B. Especial Encuesta Nacional de Hogares Rural 1991 en DANE, *Boletín de Estadística* p.127-219, 1993.

Goodman, R. and Kish, L. Controlled selection a technique in probability sampling, *Journal of the American Statistical Association* vol. 45 No. 251 p.350-372, 1950.

Groves R. M. et al. *Survey methodology*, 2009.

Kish, L. *Survey Sampling*, 1965.

Kish, L. Methods for design effects, *Journal of Official Statistics* vol.11. No. 1, pp. 55-57, 1995.

Instituto Nacional de Estadística Chile. (febrero de 2018). Estadísticas de informalidad laboral: Marco conceptual y manual metodológico. Obtenido de INE Chile: <https://www.ine.cl/docs/default-source/informalidad-y-condiciones-laborales/metodologia/antecedentes-metodologicos/manual-conceptual-y-metodol%C3%B3gico-informalidad-laboral.pdf>

Marpsat, M. *Survey methods and practices*, 2010.

Organización de las Naciones Unidas (ONU). Diseño de muestras para encuestas de hogares Estudios de métodos Serie F No. 98. Directrices prácticas, 2009.

Organización Internacional del Trabajo (OIT). Acerca de la CIET. Obtenido de: <https://ilostat.ilo.org/es/resources/>

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

Organización Internacional del Trabajo (OIT). Convenio sobre estadísticas del trabajo, 1985.

Organización Internacional del Trabajo. Declaración de la OIT relativa a los principios y derechos fundamentales en el trabajo. Obtenido de: <https://www.ilo.org/declaration/thedeclaration/textdeclaration/lang-es/index.htm>

Organización Internacional del Trabajo (OIT). Resolución concerniente a la clasificación en el empleo, adoptada por la 15ª Conferencia Internacional de Estadísticos del Trabajo (CIET), 1993.

Organización Internacional del Trabajo (OIT). Resolución concerniente a la población económicamente activa, ocupados, desocupados y subempleados, adoptada de la 13ª Conferencia Internacional de Estadísticos del Trabajo (CIET), 1982.

Organización Internacional del Trabajo (OIT). Directrices relativas a una definición estadística del sector informal ocupación de la 17ª Conferencia Internacional de Estadísticos del Trabajo (CIET), 2003.

Organización Internacional del Trabajo (OIT). Resolución I: Resolución relativa a las estadísticas de trabajo, ocupación y subutilización de la fuerza de trabajo de la 19ª Conferencia Internacional de Estadísticos del Trabajo (CIET), 2013.

https://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_087622.pdf

Organización Internacional del Trabajo (OIT). Informe II - Estadísticas de trabajo, ocupación y subutilización de la fuerza de trabajo de la 19ª Conferencia Internacional de Estadísticos del Trabajo (CIET), 2013.

Organización Internacional del Trabajo (OIT). Directrices relativas a las estadísticas de la migración laboral internacional de la 20ª Conferencia Internacional de Estadísticos del Trabajo (CIET), 2018.

Organización Internacional del Trabajo (OIT). Resolución I: relativa a las estadísticas sobre las relaciones laborales de la 20ª Conferencia Internacional de Estadísticos del Trabajo (CIET), 2018.

Siabatto, C. e Hincapié, C. Encuesta Nacional de Hogares Rural. Resumen metodológico de resultados en DANE, *Boletín de Estadística* p.128-212, 1989.

Deville, J. C., & Särndal, C. E. (1996). Calibration Estimator survey sampling. *Journal of the American Statistical Association* , 376-382.

G, C. w. (1997). *Sampling Techniques*. En J. W. Sons. New York.

Gutiérrez, H. A. (2009). Estrategias de muestreo. Diseño de encuestas y estimación de parámetros. Bogotá: Univesidad Santo Tomás.

Hansen, M., Hurwitz, W., & Madow, W. (1953). *Sample Survey Methods and Theory*. New York: Wiley.

Kalton, G. (1979). Ultimate Cluster Sampling. *Journal of the Royal Statistical Society. Series A (General)*, Vol. 142.

Kish, L. (1965). *Survey Sampling*. New York: Wiley.

Naciones Unidas. (2008). Diseños de muestras para encuestas de hogares: directrices prácticas. (D. d. Departamento de asuntos económicos y sociales, Ed.) New York.

Särndal, C. E. (2007). The calibration approach in survey theory and practice. *Survey methodology*, 33(2), 99-119.

Särndal, C. E., Swensson, B., & Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.

SAS Institute Inc. (19 de diciembre de 2019). Obtenido de SAS web site: https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_surveyselect_toc.htm&docsetVersion=15.1&locale=en

Statistics Canada. (2003). *Survey Methods and Practices*. Ottawa, Canada: National library of Canada.

ANEXOS

MARCO GEOESTADÍSTICO NACIONAL

El Marco Geoestadístico Nacional es una base de datos geográfica, insumo para las diferentes fases del proceso estadístico, es así como se utiliza en el diseño, para la definición del universo de estudio, cobertura geográfica, selección de muestra, definición de rutas, elaboración de productos cartográficos y aplicativos de seguimiento; en la ejecución para el control y seguimiento del avance de la cobertura y procesos de georeferenciación; en el análisis para la validación de la distribución de los datos y en la difusión para elaboración de mapas temáticos, entre otros usos en el desarrollo de una operación estadística.

Definición del marco

El artículo 1 de la Resolución 2222 de 2018, define el Marco Geoestadístico Nacional como el sistema para referenciar la información estadística a su localización geográfica. Está constituido por áreas geoestadísticas (departamentos, municipios, cabeceras municipales, centros poblados, rural disperso, entre otras), delimitadas principalmente por accidentes naturales y culturales, identificables en terreno y su respectiva codificación.

Estructura del marco

El Marco Geoestadístico Nacional tiene un componente cartográfico y uno temático. En el componente cartográfico tenemos los niveles geográficos de: Departamento, Municipio, Clase, Centro Poblado, Sector Rural, Sección Rural, Sector Urbano, Sección Urbana y Manzana, y en el componente temático las variables asociadas a cada uno de estos niveles.

Información auxiliar del marco

Las variables asociadas al Marco Geoestadístico Nacional son: variables de ubicación que corresponden a la identificación de los niveles geográficos, la conformación del código DANE y coordenadas geográficas; y las variables temáticas asociadas del componente social (viviendas, hogares, personas, tipo de vivienda, servicios públicos, rangos de edad y nivel educativo), del componente económico (uso de las unidades, estrato) y del componente étnico (territorialidad étnica), entre otras.

Metodología de Actualización del marco

El Marco Geoestadístico Nacional, se actualizó en su componente cartográfico a partir de las novedades cartográficas registradas durante el operativo de campo del Censo Nacional de Población y Vivienda 2018, así como las reportadas por las operaciones estadísticas realizadas por el DANE y la revisión de imágenes de sensores remotos; así mismo se llevó a cabo la actualización de la variable vivienda para secciones atípicas mediante procesos de conteo de techos automático y manual, revisión de centros poblados y actualización de límites municipales. En cuanto al componente temático se realizó la asociación de 24 variables de ubicación que incluyen el Código DANE, codificación de los niveles y coordenadas geográficos, y 80 variables provenientes del CNPV, así como los conteos de techos.

Para la actualización en períodos intercensales se han implementado protocolos de actualización para los casos de: crecimiento horizontal, crecimiento vertical, incorporación de nuevos centros poblados y revisión de muestra, así como procedimientos para la actualización de la variable vivienda a partir de operaciones estadísticas y registros administrativos.

Especificaciones técnicas del marco

Para representar los niveles geográficos del MGN se emplea el sistema de referencia de coordenadas MAGNA-SIRGAS y sus características se encuentran relacionadas a continuación:

Código del sistema	
de referencia:	EPSG 4686
Datum:	GRS_1980
Prime Meridian	Greenwich
Angular Unit:	Degree

Criterios del marco

Sobre la base cartográfica se aplican reglas de consistencia topológica y se controla la trazabilidad de los cambios mediante el registro del tipo de novedad, año y caso de actualización. Los perímetros censales de las cabeceras municipales en general exceden el límite del perímetro urbano oficial, conforme a las dinámicas de expansión urbana.

La información de resguardos indígenas se encuentra registrada en la variable de territorialidad étnica. Las áreas de resguardos indígenas están principalmente contenidas en la zona rural del país sin embargo existen algunas excepciones de resguardos que cubren áreas de cabeceras municipales, en este caso y conforme a lo establecido en los planes de ordenamiento territorial se identificarán estas áreas como zona urbana.

Disponibilidad del marco

El Marco Geoestadístico Nacional se entrega al grupo de muestreo en un archivo Excel que contiene los menores niveles de desagregación (manzanas y secciones rurales) con las variables de ubicación, las de CNPV y las actualizaciones mediante conteo de techos, así como la documentación de diseño, documento de actualizaciones realizadas y diccionario de datos.

Vigencia del marco

El Marco Geoestadístico Nacional tiene una vigencia anual y su actualización se hace periódica cada año en sus componentes cartográfico y temático.

SISTEMA GENERAL DE PRUEBAS AL REDISEÑO DE LA GEIH

	PRUEBAS DE ESCRITORIO	SIMULACIONES	PRUEBA PILOTO	DISEÑOS EXPERIMENTALES	PARALELO
Validación	Aplicación de reglas de lógica en cada nueva pregunta del formulario.	Entorno de pruebas simulado de un banco de preguntas de escenarios que asemejen el trabajo de campo.	Prueba de campo donde se aplica 1 formulario de la GEIH que incluye los nuevos capítulos y preguntas en áreas del país seleccionadas. A partir de esta prueba se cuenta con personal específico para cada etapa de la encuesta: logística recolección, sistemas, procesamiento y análisis.	Prueba de campo donde se aplican 8 formularios aleatorios de la GEIH, los cuales incluyen diferentes combinaciones de las preguntas consolidadas de la encuesta con los nuevos capítulos.	Trabajo de campo donde se aplican 2 formularios de la GEIH, el formulario tradicional (GEIH marco 2005) y el nuevo formulario GEIH rediseñado (GEIH marco 2010).
Especificaciones	Prueba individual efectuada por los responsables de cada producto	Conformación de un equipo de trabajo para la realización de las simulaciones desde el área temática GEIH – DANE.	Recolección en ocho sedes: Bogotá, Barranquilla, Cúcuta, Medellín, Pereira, Pasto, Riohacha y Villavicencio las cuales incluyen muestra urbana y rural.	Recolección en doce sedes: las 8 de la prueba piloto junto con Cali, Santa Marta y Tunja, las cuales incluyen muestra urbana y rural.	Recolección en todas las sedes del país: las 24 ciudades principales, 0 nuevos territorios y 5 ciudades intermedias las cuales incluyen muestra urbana y rural.
Periodo	-	-	Recolección muestra rural Riohacha y Villavicencio: marzo 15 - 30 de 2020. Recolección demás sedes: junio 16 - julio 31 de 2020.	Septiembre 2020	Enero 2021
Finalidad	Validación de: opciones de respuesta de cada pregunta; relación de ésta con las demás preguntas de la encuesta.	Validación de: las preguntas dentro del flujo de la encuesta; duración de las preguntas; tiempo de digitación; reporte y sincronización del aplicativo.	Validación de base de datos de la prueba piloto: verificación de tasa de respuesta de las nuevas preguntas y cambio de tasa de respuesta de las preguntas de ocupación de GEIH, reporte de frecuencias.	Validación de las bases de datos de los 8 formularios experimentales: verificación de tasa de respuesta de cada uno, reporte de frecuencias y pruebas estadísticas.	Comparación de la base de datos del rediseño con la base publicable tradicional de la GEIH; comparación tasa de respuesta y frecuencias del formulario GEIH I rediseñado con el formulario tradicional de la GEIH.

VERSIÓN	FECHA	RAZÓN DE LA ACTUALIZACIÓN
1	15/Ene/2002	Creación del documento
2	30/Jun/2006	Actualización documental
3	01/Jul/2009	Actualización documental
5	16/Nov/2011	Actualización documental
6	16/Ago/2012	Actualización documental
7	31/Oct/2012	Actualización documental
8	22/Mar/2013	Actualización documental
9	08/Abr/2016	"Se cambia la plantilla de la metodología general de la Gran Encuesta Integrada de Hogares. Se modifica, de la metodología general de la GEIH, la prueba de campo en la selección de personal, de las actividades preparatorias. 2.3 Diseño de la ejecución; 2.3.2. Actividades preparatorias, a. Sensibilización y b. Selección y entrenamiento del personal. "
10	20/Feb/2020	Se realizó actualización en las secciones de antecedentes, alcance y actividades preparatorias.
11	05/Jun/2023	Se adapta la metodología de acuerdo con la estructura definida en el Modelo GSBPM y la NTC PE 1000:2020. Asimismo, se actualiza la información con las mejoras incorporadas a la GEIH luego del rediseño (GEIH marco 2018).

ELABORÓ	REVISÓ	APROBÓ
Nombre: Jeanette Consuelo Olarte Bustos Cargo: Fecha: 05/Jun/2023	Nombre: Liliana Castañeda Pardo Cargo: Coordinador GIT Temática Mercado Laboral Fecha: 15/Jun/2023 Nombre: Andrea Ramirez Pisco Cargo: Directora Técnica Fecha: 15/Jun/2023	Nombre: Leonardo Trujillo Oyola Cargo: Subdirector del Departamento Fecha: 16/Jun/2023

Si este documento es impreso se considera copia no controlada