

Uso de datos de registros administrativos para complementar la información de ingresos de la GEIH

*Gina Fernanda
Otálora Peña*

Noviembre 2024





Contenido

1. Antecedentes
2. Evaluación y uso de RR.AA en 2020
3. Gestión de los RR.AA
4. Validación de Resultados
5. Desafíos

Adaptaciones de la GEIH para garantizar la producción de indicadores laborales

Cambio de método de recolección:

- Operativo telefónico implementado entre marzo y abril de 2020 para 23 ciudades principales y áreas metropolitanas.
- Operativo presencial mantenido en centros poblados, rural disperso y otras cabeceras (dominios urbanos distintos a las 23 ciudades) y formulario completo.

Formulario reducido (marzo-julio 2020):

- De 200 a 39 preguntas para asegurar la continuidad de indicadores clave.
- Incluyó preguntas sobre:
 - Salarios y ganancias netas (representan aproximadamente el 65% del ingreso agregado de los hogares colombianos).
 - Edad, sexo, actividad económica, ocupación, ingresos principales y horas trabajadas.
 - Módulos de migración e identificación.

Diseño muestral sin alteraciones:

- Se mantuvo la misma muestra y procedimiento de selección.
- Las manzanas definidas en el diseño original permanecieron invariables.

Operativo de recuperación de información de ingresos (nov-dic 2020)

Objetivo: Completar variables de ingresos no incluidas en el formulario reducido (marzo-julio 2020) para las 23 ciudades principales.

Variables recuperadas:

- Ingresos monetarios diferentes a salarios y ganancias netas.
- Segunda actividad, ingresos en especie, ingresos de desocupados e inactivos, intereses y dividendos, pensiones y jubilaciones, pensión alimenticia y ayudas de hogares y arriendos.

Formulario de recuperación:

- Para la implementación del operativo de recuperación de información de ingresos los encuestadores usaron un aplicativo web que incluyó todas las preguntas de las fuentes de ingresos faltantes.
- Incluyó preguntas de vivienda y hogar necesarias para imputación de ingresos.

Las encuestas retrospectivas pueden enfrentar fenómenos de omisión, sub o sobre reporte en preguntas relacionadas con ingresos por posibles sesgos que se pueden derivar de la recordación humana (Ayhan & Işiksal, 2005).



Uso de RRAA como fuente de contraste e imputación



Comparación con 2019 para evaluar pertinencia.



Análisis de cobertura (posibles sesgos de cobertura falta de respuesta) y de sesgos de reporte (precisión).

Fuentes de ingresos:

- Ingreso monetario de la primera actividad (IMPA) diferentes a salarios y ganancias netas,
- Ingresos de la Segunda Actividad (ISA),
- Ingresos en Especie (IE),
- Ingreso Monetario de Desocupados e Inactivos (IMDI),
- Intereses dividendos (IOF1),
- **Pensiones y jubilaciones (IOF2)**,
- Pensión alimenticia y ayudas de hogares (IOF3H),
- **Ayudas institucionales (IOF3I)**,
- Arriendos.



Alistamiento de los RRAA

1

- Limpieza de las variables en cuanto a caracteres presentes no válidos.
- Eliminación de ceros antepuestos en los documentos de identificación
- Mejoras en las variables
- Verificación de los formatos (i.e, fechas), detección de valores por defecto.



Diseños

2

- Se realiza la respectiva estandarización del registro administrativo recibido, para adecuarlo a una estructura estándar y con la codificación de las variables de acuerdo con las tablas de referencia establecidas.



Validación de consistencia

3

- Chequeo de faltantes y consistencia de las variables de identificación (uso de tablas de referencia)
- Detección y protocolos de duplicidad por registro
- Chequeos de consistencia en cada variable, esto se determina frente a los atributos de las mismas.



Definición de la unidad estadística

4



(IOF2) Pensiones:

- Planilla Integrada de Liquidación de Aportes (PILA)
- ✓ La variable de **mesada pensional por persona** se construye sumando todas las mesadas registradas para un cotizante en un periodo, sin excluir registros por novedades.
- ✓ Las pensiones de tipo cero, según la Resolución 2388 de 2016 (actualización 2020), no se incluyen en las estadísticas.



(IOF3I) Ayudas Institucionales:

- Más Familias en Acción (FA),
 - Jóvenes en Acción (JA),
 - Colombia Mayor (CM)
 - Ingreso Solidario,
 - Compensación de IVA,
 - Programas ayudas de las alcaldías de Bogotá, Medellín y Bucaramanga.
- ✓ El monto total recibido por ayudas institucionales en el año de referencia e identificar la permanencia de los individuos dentro de los programas.
 - ✓ Debido a acumulados, retroactivos y entradas/salidas de ayudas, se usa el monto mensualizado de cobros totales observados en el periodo para cada individuo.
 - ✓ En 2020, se implementaron ayudas extraordinarias como CM, FA, JA, Ingreso Solidario, compensación de IVA, y programas locales como Bogotá Solidaria.



Sustitución e integración GEIH - RRAA

5

- Emparejamiento de información de la GEIH con RRAA
- Pruebas de consistencia match GEIH – RRAA
- Sustitución de información



Homologación, estandarización y codificación

6



Validación de consistencia

7

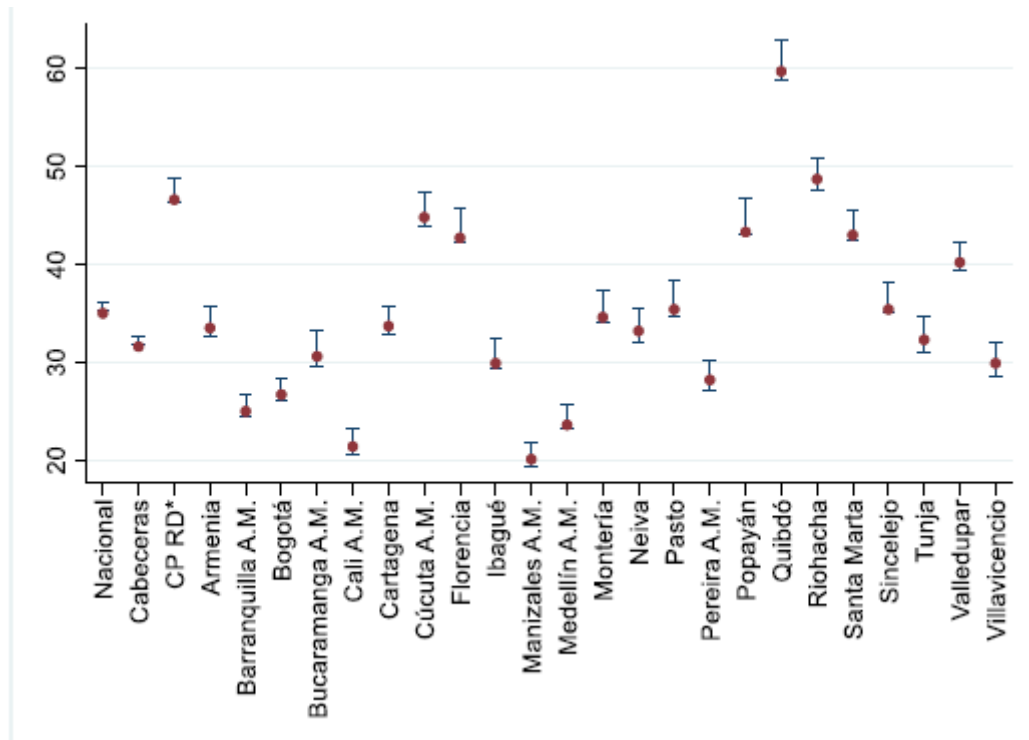
- Construcción del ICV.
- Imputación del ICV para los hogares con vacíos de información.
- Generación de sextiles.
- Imputación de estrato.
- Cálculo de faltantes y conformación de las fuentes de ingreso.
- Generación de variables para regresiones cuánticas (Son diferentes en cada fuente)
- Imputación de años de educación, tiempo trabajado y tiempo en ese trabajo. La imputación de educación se realiza para cada fuente de ingresos.
- Estimación de regresiones cuantílicas e identificación de valores externos.
- Imputación.
- Validación de consistencia output.

Validaciones de consistencias de los resultados

Sesgos en encuestas retrospectivas



Grafico 1. Incidencia de pobreza monetaria 2019 GEIH-RRAA
Intervalos de confianza publicados



- Para validar este enfoque, se estimaron las tasas de **pobreza monetaria y pobreza extrema** de 2019 integrando registros administrativos con la GEIH, y los resultados mostraron que las cifras calculadas están dentro de los **intervalos de confianza** publicados para 2019.

- Gestión de proveedores de Registros Administrativos
- Cambio en el modelo de datos de un año a otro.
- Limitaciones en la capacidad técnica y tecnológica para procesar e integrar los datos



- Unidades de análisis distintas.
- Actualizaciones en tipo de documentos en bases de años anteriores que generan cambios en los resultados y dificultan las comparaciones
- El valor de las variables puede cambiar en el tiempo.
- Actualización de los programas por cambios en la legislación y/o políticas.



Gracias!!

Gina Otálora - gfotalorap@dane.gov.co

Alejandro Rojas - jarojasv@dane.gov.co

Área de Registros Estadísticos

