

ALGORITMO PARA LA CONFORMACIÓN DEL INGRESO PER CÁPITA PARA LA MEDICIÓN DE POBREZA A PARTIR DE LA ENCUESTA CONTINUA DE HOGARES - ECH (2002-2005) Y DE LA GRAN ENCUESTA INTEGRADA DE HOGARES - GEIH (2008-2010)¹

El insumo central para la medición de la pobreza en el país es el ingreso per-cápita disponible de las unidades de gasto. Este agregado es construido a partir de una batería de preguntas estructurada en los formularios de las encuestas periódicas realizadas a los hogares.

La Encuesta Continua de Hogares (ECH) y su versión mejorada, la Gran Encuesta Integrada de Hogares (GEIH) son instrumentos que han tenido como objeto central la medición de indicadores de mercado laboral, y adicionalmente permiten obtener información para el cálculo de indicadores de distribución del ingreso, necesidades básicas insatisfechas, índice de condiciones de vida, entre otros.

Para que la medida del ingreso per cápita de las unidades de gasto represente una estructura plausible del flujo de los ingresos dentro de los hogares, se debe tener en cuenta en su definición un conjunto de ítems lo más detallado posible que ofrezca una aproximación consistente con el ingreso disponible que éstas unidades de análisis usan para afrontar sus gastos. En este documento se describen los procedimientos empleados en la conformación de este ingreso, a partir del agregado anual de las encuestas de hogares, a saber:

1. Definición de las fuentes de ingreso.
2. Definición de valores faltantes.
3. Detección de valores extremos.
4. Detección de ‘falsos ceros’.
5. Imputación de valores faltantes, extremos y ‘falsos ceros’ (Método Hot-Deck).
6. Empalme de ingresos ECH-GEIH y GEIH-NUMA.
7. Imputación de ingreso por vivienda para propietarios y usufructuarios.

1. DEFINICIÓN DE LAS FUENTES DE INGRESO

Para realizar este procedimiento es importante tener en cuenta la siguiente clasificación laboral de la población:

Población en Edad de Trabajar² (PET)=Ocupados + Desocupados + Inactivos
Población Económicamente Activa (PEA) = Ocupados + Desocupados

Ocupados=Asalariados + Independientes + TFSR (Trabajadores familiares sin remuneración)

¹ Documento elaborado por Diana Carolina Nova, Rocio Bermúdez, Claudia Gordillo y Diana Esperanza Sánchez, miembros del equipo técnico del DANE que participó en la Misión para el Empalme de las Series de Empleo Pobreza y Desigualdad - MESEP, con base en la metodología definida por esta Misión.

² Límite poblacional definido por un país para medir los indicadores de fuerza de trabajo. En Colombia corresponde a la población de 12 o más años en la parte urbana y de 10 y más en la parte rural.

Asalariados= Obreros o empleados particulares + Obreros o empleados del gobierno + empleados domésticos (en rural incluye asalariados-jornaleros)

Independientes = Trabajadores cuenta propia + patronos + otros independientes

El archivo PERSONAS contiene información sobre variables demográficas, de educación, laborales y de ingresos para toda la población encuestada en la ECH o GEIH, según el año. A partir de esta información se pueden definir los ingresos de la PET, clasificados según su fuente de la siguiente forma:

- **Ingreso Monetario Primera Actividad – IMPA:** En la ECH este ítem incluía en una sola pregunta el ingreso monetario mensual de los asalariados correspondiente a sueldos y salarios, subsidios, horas extras, bonificaciones y viáticos. En la GEIH, el formulario utilizado abrió este rubro en los ítems implícitos en la definición anterior, es decir: ingreso salarial mensual, horas extras, subsidios (de alimentación, transporte, familiar o educativo), primas (técnica, de antigüedad, clima, orden público, otras), bonificaciones mensuales, primas anuales (navidad, vacaciones, etc). Para los independientes corresponde a la ganancia neta u honorarios de la actividad principal. En el caso rural se toma la ganancia de los últimos 12 meses y se lleva a valores mensuales, para el urbano la encuesta indaga por la ganancia del último mes.
- **Ingreso Segunda Actividad – ISA:** Aplica para todos los ocupados (asalariados, independientes y trabajadores familiares sin remuneración) que tienen otro trabajo o negocio además de su ocupación principal e incluye: ingreso en dinero y/o en especie.
- **Ingreso en Especie – IE:** Sólo aplica para asalariados e incluye: alimentos, vivienda, transporte, otros como bonos Sodexo y/o electrodomésticos.
- **Ingreso Monetario Desocupados e Inactivos – IMDI:** Incluye el ingreso por trabajo de desocupados e inactivos, recibido en el período de referencia por trabajos realizados antes de ese período.
- **Ingresos de otras fuentes - IOF:**
 - IOF1: Intereses y dividendos por inversiones
 - IOF2: Pensiones o jubilaciones por vejez, invalidez o sustitución pensional
 - IOF3: Ayudas (de hogares dentro y fuera del país, y de instituciones) y pensión alimenticia por paternidad, divorcio o separación
 - IOF6: Arriendos (efectivos)

El ingreso total por persona se define como:

- a) Para Asalariados : $IMPA + IE + ISA + IOF$
- b) Para Independientes : $IMPA + ISA + IOF$
- c) Para Trabajadores Familiares Sin Remuneración: $ISA + IOF$
- d) Para Desocupados e Inactivos: $IMDI + IOF$

Para poder calcular el ingreso total, primero se debe completar la información faltante y realizar la depuración de la información mediante los procedimientos descritos a continuación. Estos se realizan para cada una de las ocho fuentes, de manera separada para ocupados y desocupados e inactivos.

2. DEFINICIÓN DE VALORES FALTANTES

En el contexto de las encuestas de hogares, un valor faltante se identifica a través de un código que indica que la información a registrar no la posee el informante o no la suministra.

Para establecer los registros con valores faltantes en las variables de ingreso se crean variables indicadoras para cada fuente, de acuerdo con los siguientes criterios:

- En todas las fuentes, los casos que reporten los códigos 98, 99, 999, 9999, 99999, 999999, 9999999, 99999999, 999999999, en el campo correspondiente al valor del ingreso.
- Si previo a la pregunta sobre el valor de un ingreso se indaga si la persona recibió o no ese tipo de ingreso, se toman como faltantes los que respondan la opción no sabe/no responde (código 9), o los que respondan afirmativamente (código 1) y no reporten información en el campo correspondiente al valor, o reporten alguno de los códigos incluidos en el punto anterior.
- Las fuentes que se componen de la suma de varios valores se consideran faltantes si alguno de sus componentes es faltante, a excepción de IMPA para asalariados, en el que se da prioridad al salario; si se reportó información sobre el salario se toma el valor como observado aún si tiene faltantes en alguno de los demás componentes (horas extras, primas, etc.)

A esos casos definidos como valores faltantes les será aplicado posteriormente el procedimiento de imputación.

3. DETECCIÓN DE VALORES EXTREMOS

Para detectar valores atípicos de todas las fuentes de ingreso y para todos los años de la serie se aplica la técnica de optimización no paramétrica denominada regresión cuantílica. Presenta algunas variaciones en las variables involucradas dependiendo de la fuente y del periodo.

Cuadro 1. Variables para el modelo de Regresión Cuantílica

VARIABLES	FUENTE	IMPA	ISA	IE	IMDI, IOF para Desocupados e Inactivos	IOF para Ocupados
Edad		x	x	x	x	x
Edad al cuadrado		x	x	x	x	x
Horas trabajadas en la actividad principal		x		x		x

Horas trabajadas en la segunda actividad		x			
Años de educación aprobados	x	x	x	x	x
Dummy Bogotá	x	x	x	x	x
Dummy Resto	x	x	x	x	x
Dummy tres grandes ciudades (Medellín , Cali y Barranquilla)	x	x	x	x	x
Dummy otras capitales	x	x	x	x	x
Sexo	x	x	x	x	x
Dummy obreros	x	x	x		x
Dummy empleados domésticos	x	x	x		x
Dummy cuenta propia	x	x			x
Dummy patronos	x	x			x
Número de meses que lleva trabajando	x	x	x		x
Dummy jefe de hogar	x	x	x		x
Número de personas en el hogar	x	x	x		x
Número de asalariados en el hogar	x	x	x		x
Número de independientes en el hogar	x	x	x		x
Número de desocupados en el hogar	x	x	x		x
Número de menores de 5 años en el hogar	x	x	x		x
Número de adolescentes en el hogar (14-17 años)	x	x	x		x
Número de ancianos en el hogar (65 y más años)	x	x	x		x
Número de personas del hogar mayores de 25 años sin educación	x	x	x		x
Número de personas del hogar con educación superior	x	x	x		x
Años de educación promedio del hogar	x	x	x		x
Número de afiliados a salud en el hogar	x		x		
Dummy cesantes				x	
Dummy estudiantes				x	
Dummy otra actividad semana anterior				x	
Dummy oficios del hogar				x	

Para la detección de valores extremos se deben realizar los siguientes pasos para cada fuente de ingreso:

- A partir de la base de PERSONAS generar un archivo con los perceptores de la fuente de ingreso y la información de las variables que se incluyen en las regresiones.
- Marcar los registros con valores entre 0 y 1 en la fuente ingreso como extremos y excluirlos de la base.
- Excluir los registros con valores faltantes o cero en la fuente de ingreso.
- Calcular el logaritmo de la fuente de ingreso.
- Realizar el siguiente tratamiento a las variables, años aprobados de educación, horas trabajadas y número de meses que lleva trabajando:
 - Para los registros con información faltante para años aprobados de educación, se asigna el promedio por grupos. Para los ocupados los grupos son: Trabajadores familiares sin remuneración, obreros (asalariados con

código de oficio mayor o igual a 62), empleados (asalariados con código de oficio menor a 62), empleados domésticos, patronos, y demás independientes (trabajadores por cuenta propia y otros). Para desocupados e inactivos los grupos definidos son: cesantes, aspirantes, estudiantes, oficios del hogar, incapacitado permanente, y otros.

- Si las horas trabajadas en la actividad principal están en 999, se imputa el promedio por los mismos grupos definidos en el punto anterior para los ocupados. Si las horas trabajadas están en 998 se asigna 121 horas.
 - Si las horas trabajadas en la segunda actividad están en 999 se imputa el promedio por los mismos grupos definidos en el punto anterior. Si las horas trabajadas están en 998 se le asigna 80 horas.
 - Si el número de meses que lleva trabajando tiene valores 998 o 999, o si la relación entre la edad y el valor anualizado de esta variable es menor a 1.2 (como criterio de validación), se imputa el promedio por los grupos definidos para los ocupados.
- Calcular por hogar, el número de afiliados a salud, el promedio de años de educación aprobados, el número de personas con educación superior, el número de personas mayores de 25 años sin educación, el número de asalariados, independientes, menores de cinco años, adolescentes (entre 14 y 17 años) y ancianos (de 65 y más años de edad). Pegar estas variables a los registros de personas.
 - Teniendo ya el archivo conformado con los perceptores de la fuente que se esté tratando, se aplica la técnica de optimización no paramétrica denominada regresión cuantílica, para los cuantiles 10, 25, 50, 75, 85 y 95 del logaritmo del ingreso.
 - Se toman como extremos los valores que sean clasificados como tal, mínimo en 5 de las 6 regresiones. En cada regresión se clasifican como extremas las observaciones cuyo residuo caiga por fuera del intervalo comprendido entre -3 y 3 desviaciones estándar.
 - Los casos con valores extremos en cada fuente (incluyendo los valores entre 0 y 1 que por defecto se toman como tal) se recodifican como omitidos y se remiten al proceso de imputación.

4. DETECCIÓN DE 'FALSOS CEROS'

Para el caso de IMPA, que involucra salarios o ganancias, se realiza un procedimiento adicional que consiste en determinar cuáles valores de los declarados por los encuestados como “cero”, no lo son bajo un criterio de clasificación estadístico. Para esto se emplea el análisis discriminante no-paramétrico, de la siguiente forma:

- Se parte de un archivo con las personas ocupadas, y se excluyen los registros con valores faltantes para IMPA, o que hayan sido detectados como extremos para esta fuente de ingresos en el procedimiento anterior.
- Se clasifica la población en dos grupos de análisis, los que declaran ingreso (IMPA) cero y los que declaran ingreso (IMPA) diferente de cero. Por cada uno de los 6

estratos se calculan las proporciones de casos para estos dos eventos (información a priori).

- Luego, usando el procedimiento de análisis discriminante no-paramétrico para cada uno de los estratos, se estiman las probabilidades a posteriori de que un caso, dada cierta información (información general personal, laboral, del hogar, geográfica), tenga ingreso (IMPA) mayor que cero. Las variables explicativas incluidas en este modelo son: edad, edad al cuadrado, horas trabajadas, años de educación aprobados, variables dummy para: jefatura de hogar, Bogotá, resto (rural), tres grandes áreas metropolitanas (Medellín, Cali y Barranquilla), otras capitales, obreros, empleados domésticos, cuenta propia, y patronos; sexo, número de meses que lleva trabajando, variables referentes al hogar: número de personas, número de asalariados, número de independientes, número de desocupados, número de menores de 5 años, número de adolescentes, número de ancianos, número de personas sin educación, número de personas con educación superior, promedio de años de educación y número de afiliados a salud.
- Los casos que reportaron ingreso cero y que el procedimiento clasifica como ‘no ceros’ serán los clasificados como ‘falsos ceros’.

Estos casos, unidos a los detectados como valores extremos y a los faltantes, serán los que ingresen al proceso de imputación.

5. IMPUTACIÓN DE VALORES FALTANTES, EXTREMOS Y ‘FALSOS CEROS’ (MÉTODO HOT-DECK)

Este procedimiento asigna valores a los datos faltantes, o que fueron declarados extremos o “falsos ceros” (receptores), con la información de los registros con datos observados no detectados en los procesos de depuración antes mencionados (donantes). Para esto, los donantes y los receptores de cada fuente se distribuyen en celdas definidas a partir de variables que diferencian el nivel de ingreso, y se realiza una asignación aleatoria al interior de las celdas.

Los pasos necesarios para realizar este procedimiento se describen a continuación:

- i. Definir las celdas de imputación:

Con base en las Encuestas a Hogares (ECH y GEIH), para cada fuente de ingresos se definieron unos criterios de clasificación (Geográficos, Laborales, Demográficos, Educativos, Otros), con un número de categorías para cada criterio (K_1, K_2, K_3, K_4, K_5). A partir de la combinación de las categorías de cada criterio se definen las celdas. Por ejemplo, una celda está compuesta por las jefes de hogar, mujeres, empleadas, con educación superior, que tienen entre 25-45 años, de estrato cuatro, en trece áreas metropolitanas. Para cada fuente de ingreso, se crea una llave para identificar la celda a la que pertenece cada receptor.

Los criterios y categorías definidos para cada fuente se presentan a continuación:

Para el Ingreso Monetario Primera Actividad (IMPA):

- o Dominio (“13 Areas”, “Otras Cabeceras” y “Resto”) = 3 categorías

- Estrato (“1”, “2”, “3”, “4”, “5”, “6”) = 6 categorías
- Edad (“menores de 18”, “18-24”, “25-45”, “46 y más”) = 4 categorías
- Educación (“ninguno, preescolar y primaria”, “secundaria” y “superior”) = 3 categorías
- Posición Ocupacional (Ocupados) Asalariados: (“Obreros”, “empleados”, “empleados domésticos”), Independientes: (“trabajadores cuenta propia”, “patronos”) = 5 categorías
- Sexo (“1=hombre”, “2=Mujer”) = 2 categorías
- Jefe del hogar (“1=Jefe”, “0=Otra”) = 2 categorías

Para este caso, por ejemplo, la llave inicial quedaría definida como:

Llave1=dominio*1000000+estrato1*100000+edad*10000+educ*1000+posición*100+sexo*10+jefe;

En el caso del Ingreso de Segunda Actividad (ISA):

- Dominio (“13 Áreas”, “Otras Cabeceras” y “Resto”) = 3 categorías
- Estrato (“1”, “2”, “3”, “4”, “5”, “6”) = 6 categorías
- Educación (“ninguno, preescolar y primaria”, “secundaria” y “superior”) = 3 categorías
- Horas (“menor o igual a 24”, “mayor a 24”) = 2 categorías

Para Ingreso en Especie (IE):

- Dominio (“13 Áreas”, “Otras Cabeceras” y “Resto”) = 3 categorías
- Estrato (“1”, “2”, “3”, “4”, “5”, “6”) = 6 categorías
- Posición Ocupacional (Ocupados) (“Obreros”, “empleados”, “empleados domésticos”) = 3 categorías
- Sexo (“1=hombre”, “2=Mujer”) = 2 categorías

Para Ingreso Monetario Desocupados e Inactivos (IMDI):

- Dominio (“13 Áreas”, “Otras Cabeceras” y “Resto”) = 3 categorías
- Estrato (“1”, “2”, “3”, “4”, “5”, “6”) = 6 categorías
- Actividad (Desocupados e Inactivos): (“Cesante-Aspirante”, “Desocupado”, “Estudiante”, “Hogar”) = 4 categorías

En el caso de Ingreso por Otras Fuentes para Ocupados (IOF1, IOF2, IOF3, IOF6):

- Dominio (“13 Áreas”, “Otras Cabeceras” y “Resto”) = 3 categorías
- Estrato (“1”, “2”, “3”, “4”, “5”, “6”) = 6 categorías
- Educación (“ninguno, preescolar y primaria”, “secundaria” y “superior”) = 3 categorías

- Posición Ocupacional (Ocupados) Asalariados: (“Obreros”, “empleados”, “empleados domésticos”), Independientes: (“trabajadores cuenta propia”, “patronos”) = 5 categorías
- Horas: (“menor o igual a 24”, “mayor a 24”) = 2 categorías

Para Ingreso por Otras Fuentes Desocupados e Inactivos (IOF1, IOF2, IOF3, IOF6):

- Dominio (“13 Áreas”, “Otras Cabeceras” y “Resto”) = 3 categorías
- Estrato (“1”, “2”, “3”, “4”, “5”, “6”) = 6 categorías
- Actividad (Desocupados e Inactivos): (“Cesante-Aspirante”, “Desocupado”, “Estudiante”, “Hogar”) = 4 categorías

- ii. Clasificar los perceptores de cada fuente de ingreso en donantes y receptores.

Los donantes son las personas que informaron su ingreso y que en los dos pasos anteriores no fueron declarados extremos ni “falsos ceros”. Los receptores son individuos que no declararon ingreso en la encuesta, o fueron declarados extremos o “falsos ceros”.

- iii. Por cada celda, seleccionar aleatoriamente tantos donantes como receptores hay en la celda.

Si el número de donantes es menor que el de receptores ($n_{don} < n_{rec}$) la selección es aleatoria con repetición. De otra parte, si ($n_{don} \geq n_{rec}$) se realiza una selección aleatoria sin repetición.

A los receptores se les asignan los valores de ingreso de los donantes seleccionados.

En el caso en que en alguna celda se encuentren receptores más no donantes, se omite el último criterio de clasificación³, y se repite el proceso de asignación aleatoria anteriormente descrito. Esto se realiza tantas veces como sea necesario para que todos los valores faltantes sean imputados.

Una vez se tiene el ingreso completo y depurado, se procede a conformar el *ingreso por unidad de gasto*, que corresponde a la suma de los ingresos de los miembros de cada hogar, excluyendo los que se hayan clasificado en el parentesco con el jefe de hogar como empleados del servicio doméstico y sus parientes, trabajadores o pensionistas.

³ De acuerdo al ejemplo, al relajar el criterio de clasificación para esa celda ya no se tomaran jefes, quedando los casos clasificados como mujeres, empleadas, con educación superior, entre 24-45 años, pertenecientes al estrato cuatro de las trece áreas metropolitanas. Se crearía una segunda llave para identificar las celdas, de la forma: llave2=dominio*100000+estrato1*10000+edad*1000+educ*100+posición*10+sexo;

6. EMPALME DE INGRESOS

La Encuesta a Hogares, de tipo continuo a partir del año 2001, ha tenido actualizaciones que afectan sus instrumentos de medición y el diseño y composición de la muestra sobre la que se recolecta la información. En particular, es posible distinguir dos grandes modificaciones:

1. Transición de la ECH (Ene02 – Jun06) a la GEIH (Jul06 – Dic08).
2. Cambio de marco y muestra de la GEIH (a partir de Ene09) incorporando la información del Censo 2005.

Para hacer comparables las cifras de pobreza y desigualdad a lo largo de toda la serie, se usó la información de las encuestas paralelas⁴ para determinar los factores de empalme a aplicar. Una revisión de la información provista por ambas encuestas permitió a la MESEP determinar que las cifras de la GEIH tienen mayor calidad que las que se obtenían a partir de la ECH. En este sentido se decidió aplicar los factores de empalme de adelante hacia atrás tomando como válidas las cifras de la nueva encuesta y empalmando las de la anterior para alcanzar los nuevos niveles. Por idénticas razones, se procedió de la misma forma en el empalme GEIH-NUMA (Nuevo Marco), en el cual se parte de los datos del empalme previo (ECH-GEIH) y sobre estas cifras se efectúa el empalme GEIH-NUMA.

La variable sobre la cual se efectuó el empalme fue el ingreso per cápita de la unidad de gasto.

Se construyó para cada caso una matriz de factores de empalme⁵. Esta matriz se definió a partir de tres variables: área geográfica, educación del mayor perceptor, y edad del mayor perceptor. Los factores de empalme correspondientes se presentan a continuación:

Matriz de factores de empalme ECH-GEIH

Área geográfica	Educación de mayor perceptor	Edad del mayor perceptor	
		<25	25 +
13 áreas	Secundaria incompleta	0.993	1.029
	Secundaria completa	1.106	1.064
	superior sin título	1.098	1.105
	superior con título	1.020	1.096
	Postgrado		1.083
Resto	Secundaria incompleta	0.966	1.004
	Secundaria completa	1.086	1.052
	superior sin título	1.043	1.107
	superior con título	1.225	1.058
	Postgrado		1.160

Fuente: Cálculos MESEP con base en GEIH y ECHp 2008 – DANE.

⁴ Entre noviembre de 2007 y diciembre de 2008 el DANE aplicó la ECH con todas sus características de forma paralela a la GEIH. Entre abril y diciembre de 2008 aplicó una encuesta paralela con el nuevo marco muestral (NUMA), cuando las cifras oficiales aún correspondían a la encuesta GEIH con el marco del Censo 93. La GEIH con el nuevo marco se oficializó a partir de enero de 2009.

⁵ Para mayores detalles metodológicos remitirse a los documentos de la MESEP, *Resumen Ejecutivo Fase I y Metodología cifra pobreza 2009*.

Matriz de factores de empalme GEIH - NUMA⁶

Área geográfica	Educación de mayor perceptor	Edad del mayor perceptor	
		<25	25 +
13 áreas	Secundaria incompleta	0.971	0.956
	Secundaria completa	0.989	0.974
	superior sin título	0.995	0.948
	superior con título	1.072	0.929
	Postgrado		0.942
Resto	Secundaria incompleta	1.018	1.004
	Secundaria completa	0.997	0.991
	superior sin título	1.017	0.947
	superior con título	1.082	1.048
	Postgrado		0.995

Fuente: Cálculos MESEP con base en GEIH y paralelo NUMA abril-diciembre 2008 – DANE.

Para definir a qué celda de la matriz pertenece cada hogar, y por tanto qué factor se le debe aplicar a su ingreso, es fundamental la identificación del *mayor perceptor*. Para esto se toma el ingreso total de cada persona de la unidad de gasto, y se busca el mayor ingreso. Si hay más de una persona en la unidad de gasto con ese ingreso se toman, en su orden, los siguientes criterios de desempate:

- Parentesco: Se da prelación al más cercano al jefe de hogar, tomando el que tenga menor valor en la siguiente codificación:
 - Jefe: 1
 - Cónyuge: 2
 - Hijo: 3
 - Otros parientes: 4
 - Otros no parientes: 5
- Edad: Si persiste el empate luego de aplicar el criterio del parentesco (por ejemplo si hay dos hijos con el mismo nivel de ingreso, siendo el más alto en el hogar), se toma como *mayor perceptor* al de más edad.
- Como última opción, si persiste el empate, se elige aleatoriamente.

Adicionalmente, la metodología adoptada contempla el uso de un factor de gradualidad del tipo t/T aplicado al factor de empalme, en donde t corresponde al periodo para el cual se está calculando la cifra empalmada y T al periodo en el cual se efectúa la transición de una a otra encuesta. Este factor de gradualidad se aplica cuando puede presumirse que los efectos se introdujeron gradualmente en la encuesta, como en los casos de cambio de marco y muestra, más no en los de cambios de instrumentos. El factor toma el valor de 1 en el punto T y va disminuyendo hacia atrás hasta conseguir que la serie empalmada converja a sus valores originales en un punto del tiempo.

El empalme ECH-GEIH no tiene gradualidad y el GEIH-NUMA sí, empleando como puntos de referencia los momentos censales de ambos censos: 24 de octubre de 1993 y 20 de octubre de 2005. Entre estas dos fechas se tienen 144 meses. Con esto se obtiene:

Factor gradualidad Ene 2002 = $99/144$

Factor gradualidad Oct 2005 = $144/144$

⁶ Cuando se publicó la cifra 2009 se calcularon factores de empalme NUMA-GEIH, para quitar a la cifra de ese año el efecto del cambio de marco y que así fuera comparable con la serie anterior. La nueva serie se empalmó hacia atrás, es decir incorporando el efecto del cambio de marco a las cifras anteriores a 2009, por lo que los factores usados corresponden a los inversos de los calculados inicialmente.

Así, para empalmar la serie es necesario multiplicar el ingreso de la unidad de gasto de los años 2002-2005 por los factores de empalme ECH-GEIH y por los factores de empalme GEIH-NUMA, estos últimos ajustados por el factor de gradualidad que corresponda para cada mes⁷. Para el año 2008 sólo se requiere aplicar los factores de empalme GEIH-NUMA.

7. IMPUTACIÓN DE INGRESO POR VIVIENDA PARA PROPIETARIOS Y USUFRUCTUARIOS

Para los hogares que hayan declarado que la vivienda que ocupan es propia, sea que la estén pagando o que esté totalmente pagada, y para los que la tengan ocupada en usufructo, se imputa un valor de ingreso que corresponde a lo que tendrían que pagar por el arriendo de la misma.

En la GEIH se cuenta con la pregunta: *Si tuviera que pagar arriendo por esta vivienda ¿cuánto estima que tendría que pagar mensualmente?*, y este es el valor que se imputa⁸ teniendo en cuenta los siguientes criterios:

- Si la vivienda es propia totalmente pagada, se imputa el valor declarado.
- Si la vivienda es propia, pero la están pagando, se imputa la diferencia entre el valor declarado y la cuota de amortización reportada por el encuestado, siempre y cuando esta diferencia no sea negativa.
- Si la vivienda es ocupada en usufructo, se imputa el valor declarado, siempre y cuando ningún miembro de la familia haya reportado vivienda como ingreso en especie por su trabajo.

La Encuesta Continua de Hogares no contaba con esta pregunta, por lo que fue necesario realizar otro tipo de imputación. Para esto se empleó la metodología Hot Deck antes descrita, tomando como donantes a los arrendatarios, e imputando a los receptores el valor del arriendo declarado por los primeros.

Pero, adicionalmente, la ECH sólo incluía el módulo de vivienda para las 13 Áreas Metropolitanas en los terceros trimestres de cada año. Para los demás dominios sí se incluía en todos los trimestres. Por esta razón fue necesario implementar un procedimiento previo para estimar el estatus de tenencia de la vivienda para los periodos sin información, que se describe a continuación:

Estimación de estatus de tenencia de vivienda:

El objetivo de este procedimiento es establecer la condición de tenencia de vivienda de los hogares encuestados durante los primeros, segundos y cuartos trimestres del periodo de 2002-2005 en el dominio de las trece áreas metropolitanas.

Esta asignación se desarrolla en dos pasos que se describen a continuación:

1. Estimación de un modelo Multilogit generalizado sobre la variable tenencia de vivienda:

⁷ Para ajustar los factores de empalme GEIH-NUMA: $FE_{aj\ i} = [(FE-1) * FG_i] + 1$, donde FE es el factor de empalme sin ajuste por gradualidad y FG_i es el factor de gradualidad del mes i .

⁸ Se estableció como criterio de control que los valores no sean menores a \$10.000 ni mayores a \$3.000.000, tanto para el valor que pagaría por arriendo, como para la cuota de amortización que declaran los que están pagando la vivienda.

Mediante esta técnica se pretende estimar las probabilidades de cada uno de los hogares de pertenecer a las categorías de tenencia establecidas en el Modulo de datos de la Vivienda y del hogar en la ECH, con base en la información recolectada para los hogares en las trece áreas durante los terceros trimestres; los eventos cuya probabilidad se busca establecer son los siguientes:

La vivienda ocupada por	Propia totalmente paga	1
este hogar es:	Propia la están pagando	2
	En arriendo o subarriendo	3
	En usufructo	4

Como variables explicativas de la condición de tenencia del hogar se definió el siguiente conjunto de variables, que miden características de la vivienda, del hogar, del jefe del hogar y geográficas:

Cuadro 2.
Variables explicativas modelo Multilogit sobre tenencia de vivienda

Categoría	VARIABLES
Características de la vivienda	No. de cuartos incluido sala y comedor Dummy Estrato 1 y 2 Dummy Estrato 3, 4 y 5
Características del jefe del hogar	Edad Edad al cuadrado Años de educación Dummy Mujer Jefe de Hogar Dummy Jefes Obreros ^a Dummy Jefes Empleados Dummy Jefes Independientes Dummy Jefes Patronos Dummy Jefes Unión Libre Dummy Jefes Solteros Dummy Jefes Separados
Características del hogar	No. de personas en el hogar Dummy existencia de algún miembro del hogar que sea celador, cuidador o mayordomo ^b Dummy existencia algún miembro del hogar que sea directivo o profesional ^c Logaritmo del ingreso x unidad de gasto Dummy existencia de algún miembro del hogar que perciba ingreso en especie en vivienda
Geográficas	Dummy Medellín, Cali y B/Quilla

	Dummy Bogotá Dummy otras 13 áreas diferentes a las cuatro grandes
--	---

^a Como obrero se define al trabajador cuya posición ocupacional es empleado empresa privada, empleado público o jornalero cuyo oficio tiene una clasificación CIUO mayor o igual a 62.

^b Celador, jornalero o mayordomo es el trabajador cuyo oficio esta clasificado en las categorías 55, 58 o 60 de CIUO.

^c Directivo o profesional es el trabajador cuyo oficio está clasificado en las categorías 01 a 21 y 40, 50 y 51 de la CIUO.

2. Asignación de condición de tenencia a los hogares: Para realizar esta asignación se procede de la siguiente forma

- Construir una base con los registros correspondientes a primero, segundo y cuarto trimestre y calcular, usando las ecuaciones estimadas, las probabilidades de cada hogar de pertenecer a las diferentes categorías de tenencia de vivienda.
- Calcular a partir de la base para las trece áreas tercer trimestre las proporciones poblacionales⁹ observadas en cada categoría de tenencia.
- Asociar a cada registro de la base construida en el primer paso la proporción observada durante el tercer trimestre en cada una de las categorías de tenencia.
- Para cada registro de la base calcular la diferencia entre la probabilidad de pertenecer a cada categoría de tenencia generada a partir de las ecuaciones estimadas y la proporción observada en la categoría correspondiente durante el tercer trimestre.
- Ordenar los hogares por área metropolitana, estrato y diferencia entre probabilidad estimada y proporción observada en la categoría que se desea asignar.
- Seleccionar de los registros ordenados, hogares hasta que se tenga la proporción poblacional observada en la categoría por estrato y área, es decir se va sumando sobre los registros de hogares el factor de expansión ponderado por el total poblacional asociado al área y el estrato, hasta completar la proporción observada.

⁹ Al hablar de proporciones poblacionales se hace referencia a las proporciones calculadas usando factor de expansión.

- Los hogares asignados en la categoría i son extraídos de la base, y los registros restantes se ordenan de nuevo por área, estrato y diferencia en la siguiente categoría, se aplica nuevamente el criterio de asignación del paso 5, hasta que los hogares queden asignados en todas las categorías.

Una vez se tiene definido el estatus de tenencia de vivienda para todos los hogares, se procede a realizar la imputación Hot Deck, tanto del arriendo que pagarían los propietarios y usufructuarios como del valor de la cuota de amortización de quienes están pagando su vivienda. Para el primer caso los donantes son los hogares arrendatarios con información sobre valor pagado por arriendo¹⁰, y los receptores son los propietarios, sea que estén pagando la vivienda o que ya la hayan pagado, y los que la tengan ocupada en usufructo y ningún miembro de la familia haya reportado vivienda como ingreso en especie por su trabajo. A los receptores se les asigna el valor de arriendo declarado por los donantes. Para el caso de la cuota de amortización, los donantes son los hogares que están pagando la vivienda y que tienen información sobre el valor de la cuota de amortización pagada¹¹, y los receptores son los hogares propietarios que están pagando su vivienda y que no tienen información sobre el valor de la cuota.

Los criterios y las categorías empleados en estas imputaciones fueron:

- Dominio (cada una de las 13 Áreas Metropolitanas, Otras Cabeceras y Resto)= 15 categorías
- Estrato (“1”, “2”, “3”, “4”, “5”, “6”) = 6 categorías
- Número de cuartos (“1”, “2”, “3”, “4”, “5 y más”) = 5 categorías

El procedimiento de imputación es el mismo descrito en el apartado 5 de este documento.

Una vez se tiene calculado el ingreso de la unidad de gasto con imputación de ingreso por vivienda para propietarios y usufructuarios, se calcula el ingreso per cápita de la unidad de gasto, dividiéndolo por el número de personas que conforman la unidad de gasto. Este último es el ingreso que se compara con las líneas de indigencia y pobreza para establecer quiénes están por debajo, definiendo así los indigentes y los pobres.

¹⁰ Hogares arrendatarios encuestados durante el tercer trimestre en las 13 áreas metropolitanas, y de todo el año para los otros dominios, y que hayan declarado un valor de arriendo entre \$10.000 y \$3.000.000

¹¹ Hogares que manifestaron estar pagando su vivienda y que fueron encuestados durante el tercer trimestre en las 13 áreas metropolitanas, y de todo el año para los otros dominios, y que hayan declarado un valor de cuota de amortización entre \$10.000 y \$3.000.000