Tratamiento de datos faltantes en el ingreso y su efecto en las estimaciones de pobreza en Colombia

Presentado por
Camilo Andrés Avila Carreño

Fundación Universitaria Los Libertadores
Facultad de Ingeniería y Ciencias Básicas
Especialización en Estadística Aplicada
Bogotá D.C, Colombia
2018

Tratamiento de datos faltantes en el ingreso y su efecto en las estimaciones de pobreza en Colombia

Presentado por
Camilo Andrés Avila Carreño

en cumplimiento parcial de los requerimientos para optar al título de

Especialista en Estadística Aplicada

Dirigida por
Heivar Yesid Rodriguez Pinzon
Profesor

Fundación Universitaria Los Libertadores
Facultad de Ingeniería y Ciencias Básicas
Especialización en Estadística Aplicada
Bogotá D.C, Colombia
2018

Notas de aceptación

Firma del presidente del jurado

Firma del jurado

Firma del jurado

Las directivas de la Fundación Universitaria Los Libertadores, los jurados calificadores y el cuerpo docente no son responsables por los criterios e ideas expuestas en el presente documento. Estos corresponden únicamente a los autores y a los resultados de su trabajo.

Índice general

Tratamiento de datos faltantes en el ingreso y su efecto en las estimaciones de pobreza en Colombia

Resumen

Los datos faltantes dentro de las encuestas a hogares son uno de los problemas más comunes en el análisis de información sobre todo en la estimación de la incidencia de la pobreza monetaria, es por esto que el presente trabajo evaluar el efecto del tratamiento datos faltantes en la variable ingreso de la Gran Encuesta Integrada de Hogares y el posterior cálculo de pobreza para el caso colombiano. Dando como resultado que de los métodos de imputación seleccionados no se puede determinar si alguno de estos afecta o no en mayor o menor media la incidencia de la pobreza.

Palabras clave: Pobreza Monetaria, Imputación ingresos, ingresos

Capítulo 1

Introducción

La medición de la pobreza monetaria trae con sigo una necesidad de información muy importante para el país, es así como el Departamento Administrativo Nacional de Estadísticas - DANE ha realizado desde el año 2012 la medición de pobreza y pobreza extrema monetaria para el país. Para cumplir con su cometido el DANE realizó las estimaciones de las líneas de pobreza con la información de la Encuesta Nacional de Ingresos y Gastos - ENIG 2006-2007 y los ingresos son medidos por medio de la Gran Encuesta Integrada de Hogares-GEIH encuesta que se realiza de forma continua en el país desde 2006 (MESEP, 2012).

Las encuestas a hogares en su quehacer diario reciben rechazos por parte de las personas encuestadas ya sea totales o parciales en alguna información, específicamente en la GEIH los ingresos, que es un parte fundamental en la medición de la pobreza, presenta unos niveles de faltantes que han venido aumentando en el tiempo. Esta información faltante está muy relacionada con la complejidad de la variable de ingresos dado que es información muy sensible para la persona encuestada y por ende prefieren omitir esta información dificultando así la estimación de los ingresos y por ende la posterior estimación de la pobreza.

Dado esto en el 2012 el DANE decidió adoptar un método de imputación para estos datos faltantes conocido como Hot-Deck, gracias a esto la pobreza monetaria se ha estimado año a año sin mayores sobresaltos, pero ante la cantidad en aumento de valores vacíos puede que ya no se tengan las mismas consideraciones para realizar este

método. Adicionalmente se debe tener en cuenta que un aumento mayor de estos vacíos puede perjudicar aún más estas imputaciones por ende es necesario evaluar como funcionan tanto el método que posee el DANE como otros usados para la imputación de ingresos a la luz de una mayor cantidad de faltantes.

Es por esto que el presente documento busca evaluar diferentes técnicas para la imputación de los datos faltantes o atípicos dentro de la GEIH, a diferentes niveles de no respuesta en los datos de ingreso y como estas imputaciones pueden afectar la estimación de personas pobres. Para cumplir este objetivo el documento tiene los siguientes acapites: I. Marco Teórico, II. Metodología, III. Resultados y analisis y IV. Conclusiones

Capítulo 2

Planteamiento del Problema

La pobreza ha sido uno de los fenómenos sociales y económicos que ha sido ampliamente estudiado por diversas ciencias humanas desde el inicio de las mismas, es así como clásicos como Aristóteles afirman que uno de los males sociales es la pobreza por diversas razones, pero específicamente porque la pobreza genera revueltas y crímenes, como citó Aguirre (2012). Así mismo Platón planteo que «tanto la riqueza como la pobreza tienen consecuencias inmorales. La riqueza conlleva el afán de lucro; la pobreza, por su parte, hace al hombre menos hombre» Maceri (2009).

Por la misma naturaleza de las ciencias económicas estas no se pueden alejar del estudio de la pobreza, es así como Adam Smith, uno de los padres de la economía, considera que «Las personas más pobres, entonces, son aquellas que apenas pueden proporcionarse las necesidades de subsistencia, aun cuando disfruten mucho los pocos bienes materiales que pueden adquirir.»(Beltrán, 2000). De esta manera diferentes autores de muy diversas corrientes de pensamiento económico han escrito sobre la pobreza, ya sea en formas de medición o en formas de erradicación.

La pobreza posee múltiples conceptos desde los se pueden abordar desde diferentes dimensiones y bajo diversas aproximaciones socio-económicas. Como lo afirma Davis y Sanchez-Martinez (2014) existen visiones relativistas de la pobreza, los que tienen una visión más holística y en las visiones más contemporáneas como un tema de capacidades de las personas como lo expone Amartya Sen. Para los estudios y mediciones mismos de la pobreza es necesario usar un concepto que se pueda operativizar, es así que para el

presente documento la pobreza es, como los define el banco mundial, la falta de capacidad de las personas para la obtención de unos tipos específicos de bienes de consumo (Haughton y Khandker, 2009).

Estas capacidades de adquisición pueden ser medidas gracias a las líneas de pobreza, estas tienen dos papeles fundamentales en la medición de la pobreza. El primero es determinar cuál es el estándar mínimo de vida en la cual ya no se considera a una persona como pobre (cuenta con la capacidad de obtener ciertos bienes), el segundo rol es el de hacer comparaciones de manera interpersonal dado que para su estimación tienen en cuenta variables socio-demográficas que nos dan el estándar mínimo de vida de las personas según la región donde viven, el tamaño de su núcleo familiar, entre otros (Ravallion, 1998).

Una vez establecidas las líneas se debe buscar el mecanismo que pueda medir el ingreso o gasto de las personas para determinar cuáles están por arriba o debajo de este umbral establecido, en otras palabras, se busca generar una variable que muestre la capacidad de las personas para obtener ese mínimo estándar de vida. En la gran mayoría de países la variable elegida es el ingreso dado que este representa la capacidad actual de adquisición de los hogares.

En muchos casos la gran cantidad de población, o así mismo la falta de recursos para la consecución de variables socioeconómicas de la población, como lo es el ingreso, hace necesario la implementación de encuestas a hogares o encuestas sociales. «Las encuestas sociales juegan un rol importante en el entendimiento de situaciones sociales y económicas y en investigaciones académicas. Información acerca de los ingresos, consumo y ahorro, los cuales son factores determinantes del bienestar económico, son recogidos por medio de encuestas de ingresos o gastos» (Sano, Tada, y Yamamoto, 2015, pp.506).

Como expone of Economic y Affairs (2008), en las encuestas a hogares se selecciona una parte de la población y se le aplica una encuesta para obtener la información deseada, con esta información y por medio de inferencia estadística se pueden generar parámetros estimados de la población. Pero así como las encuestas poseen grandes

ventajas para la recolección de dato, también pueden presentar variables que por su naturaleza no son respondidas por las personas encuestas, mientras que existen otras que si poseen unas tasas de respuesta mucho más alta, una de las razones principales es la sensibilidad de las preguntas que se deben realizar, un ejemplo son las variables asociadas a los ingresos, que poseen gran sensibilidad y pueden tener entre un 5 % y 15 % de falta de información(Heeringa, West, y Berglund, 2010).

La variable ingresos posee grandes dificultades para su obtención que otras variables en las encuestas a hogares, Smith (1991) expone dos de los problemas centrales para la obtención de la información de ingresos en los hogares. La primera razón es la oposición o resistencia de entregar información personal que es considerada de gran importancia, sobre todo si se tiene en cuenta que este es uno de los factores que definen la pertenencia o no a una clase social. La segunda razón es que la información no es conocida por todos los miembros del hogar, dando como resultado que no todas las personas de un hogar encuestado puedan dar la información de ingresos para todas las personas.

Cabe resaltar que no todos los valores faltantes afectan de igual manera la inferencia de los parámetros poblacionales, esto también depende de cómo esta falta de información se aparece dentro de la población. Si los valores faltantes son totalmente aleatorios quiere decir que todas las personas poseen la misma probabilidad de no responder la variable, con esto se puede determinar que estos valores faltantes no generan ningún sesgo en la variable a estudiar, pero si algunas personas con ciertas características sociodemográficas poseen mayor probabilidad de no responder o dar información para la variable objeto de estudio , esta sesgará los resultados y por ende las inferencia sobre los parámetros poblacionales estimados

Para el tratamiento de los vacíos de la información existen diversos métodos, cada uno de ellos tienen sus ventajas y desventajas, así mismo pueden ser simples o muy complejos de realizar. La imputación de datos se refiere básicamente al remplazo de la información faltante por medio de diversas técnicas estadísticas, el objetivo principal de cualquier técnica de imputación es la de generar una base de datos que pueda ser analizada por cualquier método estadístico para datos completos (Briggs, Clark,

Wolstenholme, y Clarke, 2002).

Las diferencias que arrojan la imputación o no de los ingresos traen consigo cambios en la estimación de pobreza, por ejemplo para México se ha encontrado que «cuando se construye el índice de la tendencia laboral de la pobreza, el CONEVAL (2012) encuentra que la pobreza ha aumentado un 25 % en el periodo 2005-2012. Sin embargo, cuando se imputan los ingresos inválidos se encuentra que la pobreza laboral solo ha aumentado un 12 %.» (Campos-Vázquez, 2013, pp.52). Al observar estos resultados el mismo autor sugiere que «las instituciones encargadas de medir pobreza y asignar recursos con base en ella, como por ejemplo CONEVAL y SEDESOL, deben incluir en sus cálculos un método de corrección para los trabajadores que deciden no declarar ingresos»(Campos-Vázquez, 2013, pp.52).

Cabe resaltar que el cálculo de la incidencia de la pobreza es uno de los datos más importantes para un país, que a pesar de que no poseen un único concepto de pobreza y por ende no una única medición todos definen políticas públicas para luchar contra la pobreza, sobre todo desde que en la cumbre del milenio de la ONU se definiese que uno de los compromisos era la reducción de la pobreza a 2015 (Muñoz, 2015). Así mismo posee un impacto muy grande dentro de la población dado que reconoce o muestra que tantos avances a logrado su sociedad en terminaos de mejora de calidad de vida.

En Colombia la Gran Encuesta Integrada de Hogares - GEIH es la encuesta encargada de la captación de la información de desempleo, ocupación y adicionalmente de la captación periódica de los ingresos. Históricamente ha poseído un porcentaje de no respuesta de ingresos y valores extremos, por tal motivo, la actual metodología de medición de la pobreza contempla la corrección por omisión e imputación de falsos ceros y valores extremos MESEP (2012), esta imputación la realiza mediante el método *hot deck*. Este tipo de imputaciones son necesarias dado que se puede incurrir en el error de sobrestimar la pobreza (personas que no reportan ingresos son identificados como pobres) o subestimarla (Personas que tienen datos atípicos son considerados no pobres).

Aunque los ingresos provenientes de la GEIH ya posee un método de imputación definido por parte del Departamento Administrativo Nacional de Estadísticas - DANE,

esto no implica que no existan otros métodos de imputación que dados otros escenarios generen una mejor estimación con una varianza menor, mejorando así la estimación del parámetro poblacional. El escenario principal en los métodos de imputación deben ser evaluados es en la cada vez mayor cantidad de valores a imputar a medida que pasa el tiempo, como lo muestra la siguiente tabla.

Cuadro 1
Datos faltantes en la GEIH 2013-2017

Año	**Cantidad de personas con algún valor en ingresos**	**Cantidad de personas con datos faltantes**	**Porcentaje de datos faltantes**
2013	432.571	37.237	8,6
2014	437.272	49.752	11,4
2015	444.960	50.523	11,4
2016	439.611	53.037	12,1
2017	437.547	62.301	14,2

Elaboración propia

En el cuadro 1 se puede observar que a través de los años la cantidad de personas que poseen ingresos (ya sea observado o como un valor faltante) han aumentado en el tiempo pasando de 432.571 en 2013 a 437.547 en 2017, pero la cantidad de personas con ingresos faltantes si ha venido en aumentando en mayor medida pasando de 37.237 a 62.301 en 2017. En términos relativos esto quiere decir que la cantidad de datos imputados pasó de un 8,6 % en 2013 a 14,2 % en 2017, mostrando así la necesidad de generar diversos escenarios en los cuales exista un aumento de valores imputados, esto con el fin de reconocer cuál método de imputación es el más eficiente en la actualizada también en momentos de mayor nivel de imputación.

2.1 Objetivos

A continuación se detallara el objetivo general de la investigación así como los objetivos específicos.

2.1.1 Objetivo General Evaluar el efecto del tratamiento datos faltantes en la variable ingreso de la Gran Encuesta Integrada de Hogares y el posterior cálculo de pobreza para el caso colombiano.

2.1.2 Objetivos Específicos

- Definir el mecanismo de relación entre las diversas variables existentes y los datos faltantes de ingreso
- Evaluar las diferencias que posee la imputación de la variable ingresos a diversos niveles de datos faltantes.
- Identificar el método apropiado en la imputación de datos faltantes en el que minimice el cálculo de la varianza y con menor sesgo en la estimación de la razón en las variables pobreza.

Capítulo 3

Marco Teórico

En esta sección abordaremos las definiciones y conceptos teóricos de la naturaleza de los datos faltantes, las técnicas de imputación y la medición de la pobreza.

3.1 Datos faltantes

Un dato faltante es la ausencia de información, o información errónea que fue eliminada dentro de una base de datos. En las encuestas a hogares, como lo expone Median y Galván (2007), la fatiga de los informantes, el desconocer la información solicitada, el rechazo del informante a responder información sensible o el rechazo por parte del hogar o algún miembro a responder la encuesta son algunas de las principales causas para la generación de datos faltantes.

Para entender mejor la razón por la cual los datos faltantes pueden afectar las estimaciones de los parámetros Rubin (1976) presentó la siguiente definición:

> Asuma que posee un vector de variables aleatorias $U = (U_1, ..., U_n)$ definiendo F_θ como su función de probabilidad, el objetivo es hacer inferencia acerca del vector de parámetros de su densidad, es decir θ. Usualmente las variables aleatorias U pueden ser arreglado en una matriz por unidades y variables. Ahora asuma que $M = (M_1, ..., M_n)$ es un vector de variables aleatorias que se asociará al indicador de datos faltantes, donde M_i toma los valores de uno o cero. La probabilidad que M tome los valores $m = (m_1, ..., m_n)$ dado que U toma los valores $u = (u_1, ..., u_n)$ es $g_\phi(m|u)$, donde ϕ es el vector de parámetros molestos de una distribución.

La distribución condicional g_ϕ corresponden al proceso que causan los datos faltantes: si $m_1 = 1$ el valor de la variable aleatoria U_i será observado, mientras que si $m_i = 0$ el valor de la variable aleatoria U_i no será observado. De manera más precisa, defínase el vector extendido de variables aleatorias $V = (V_1, ...V_2)$ con un rango extendido para incluir el valor especial $*$ para los datos faltantes: $v_i = u_i(m_i = 1)$ y $v_i = *(m_i = 0)$. Los valores de la variable aleatoria V son observados, pero no los de la variable aleatoria U, aunque el deseo es realizar inferencia sobre la distribución U. (Rubin, 1976, pp.583)

Para conocer el tratamiento y como pueden afectar esta información carente es necesario clasificarlos según el patrón y mecanismo de relación entre los datos existentes y faltantes. Los patrones de valores perdidos se refiere a la configuración que existe entre la información faltante y la obtenida (Enders, 2010).

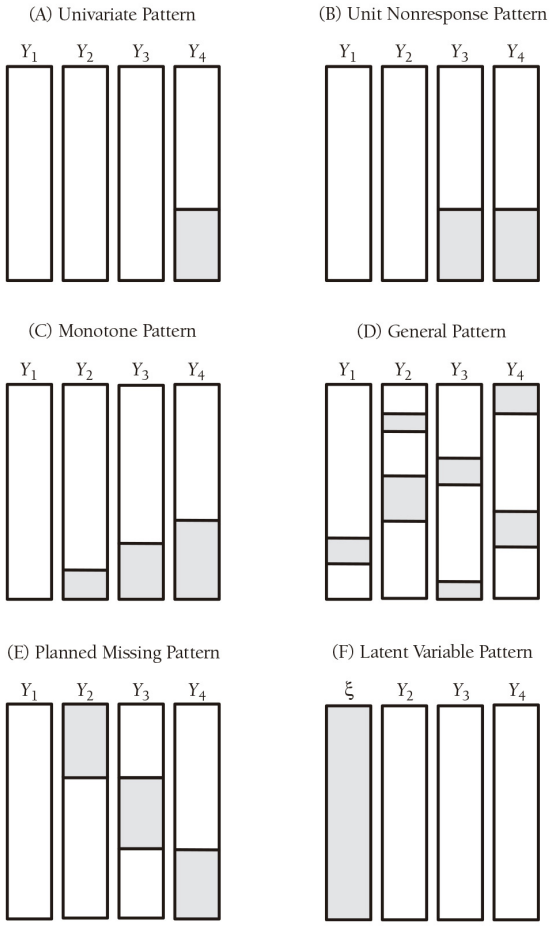

Figura 1. Patrones de datos faltantes, fuente (Enders, 2010, pp.4)

En la *Figura 1* se pueden apreciar los seis tipos de patrones faltantes definidos por

Enders (2010, pp.4). A continuación se realizará una descripción de cada uno de estos patrones:

- Patrón univariado: hace referencia a la falta de información en una única variable.
- Patrón de unida de no respuesta: es la no respuesta sistemática de al menos dos variables dentro de los datos, siempre asociada a los mismos encuestados.
- Patrón monótono: es la no respuesta que aumenta a través de las variables, esta es común en encuesta longitudinales cuando se van acumulando a través del tiempo los individuos que dejan de responder la encuesta.
- Patrón general: Es el más común de los patrones, se presenta cuando en la matiz de datos se encuentran varias variables con diferentes niveles de no respuesta.
- Patrón planificado de valores faltantes: es un diseño que de manera intencionada genera una falta de información controlada al administrar diferentes tipos de cuestionarios a la población seleccionada.
- Patrón de variable latente: Se refiere a la falta de información en un variable de análisis, como puede ser un modelos de ecuación estructural.

En cuanto a los mecanismos Zhu (2014), citando a Little y Rubin, nos presenta tres tipos de datos faltantes basados en las causas, el primero se refiere a los perdidos completamente aleatorios (MCAR por sus siglas en inglés), este ocurre cuando los datos faltantes no están relacionados con ninguna otra variable ya sea observada, o no observada. El segundo tipo son los datos perdidos aleatorios (MAR por sus siglas en inglés), se presentan cuando los datos faltantes son condicionales a alguna información observada pero son independientes de la información no observada. Por último, se encuentran los datos perdidos no aleatorios (MNAR) estos son los datos faltantes que están condicionados a la información observada y no observada. Si los datos perdidos son MCAR esto quiere decir que el mecanismo puede ser ignorado, es decir que no se necesita modelar la información faltante como parte del proceso de estimación. Si se

descarta que los datos perdidos son no MCAR los datos deben tener algún tipo de modelamiento o imputación para realizar el proceso de estimación (Soley-Bori, 2013).

Para poder determinar si los valores faltantes son completamente aleatorios (MCAR) el test más común es el propuesto por de Little. Como los expone Cheng (2013) el test χ^2 de Little posee la siguiente forma

$$d_0^2 = \sum_{j=1}^{J} n_j(\bar{y}_{oj} - \mu_{oj})^\top \Sigma_{oj}^{-1}(\bar{y}_{oj} - \mu_{oj}) \tag{1}$$

Suponga que la información y_i es una normal multivarida con el vector de medias μ y matriz de covarianzas Σ, donde algunos componentes de y_i son faltantes. Se asume que existen un total de J patrones de valores faltantes en y_i, para cada j teniendo a o_j y m_j como indices de los componentes observados y faltantes respectivamente y $p_j = |o_j|$ es el número de observaciones en el patrón j. Para los valores observado en el patrón $j - esimo$, μ_o y Σ_o serán el vector de medias y la matriz de covarianzas, mientras que $\bar{y}_o$ será la media simple de los valores observados(Cheng, 2013). Siendo este el test que se usará en el presente documento, en la metodología se hará explicita la forma en la cual se usará.

3.2 Tratamiento de los datos faltantes

Para el tratamiento de los datos faltantes uno de los primeros métodos es trabajar con la información completa, esto quiere decir que se debe descartar del estudio a todo individuo que presente información faltante, la ventaja de esté método es que es de fácil aplicación, pero es ineficiente dado que puede excluir mucha información necesaria para el estudio. Un segundo método es el trabajar con la información disponible, en esta no se descartan los datos, se trabajan con las medias de cada una de las variables según el grado de completitud de las mismas, al igual que método anterior es de fácil aplicación, pero su mayor desventaja es que se trabaja con diferentes muestras para cada variable generando problemas de comparabilidad y por ende en la estructura de covarianzas de la base de datos (Briggs y cols., 2002).

También Existe una gran diversidad de métodos de imputación, algunos de estos son:

- **Hot-Deck:** Este procedimiento se basa en la clasificación de la información en conjuntos disjuntos buscando que la distribución dentro de cada uno de los conjuntos se a lo más homogénea posible. dentro de cada uno de los conjuntos disjuntos existirán dos grupos, los donantes que poseen información completa y los receptores que no poseen la información incompleta, cada valor faltante es completado con información de los donantes dentro de cada uno de los conjuntos creados. Las variables de clasificación para la generación de los conjuntos deben estar muy ligadas a los valores faltantes y registrados para sí lograr que este método funcione correctamente.(Avila, 2002)

  Este es el método que actualmente es usado por el DANE para la imputación de los ingresos, a continuación se presenta el diagrama de proceso para la implementanción de este método.

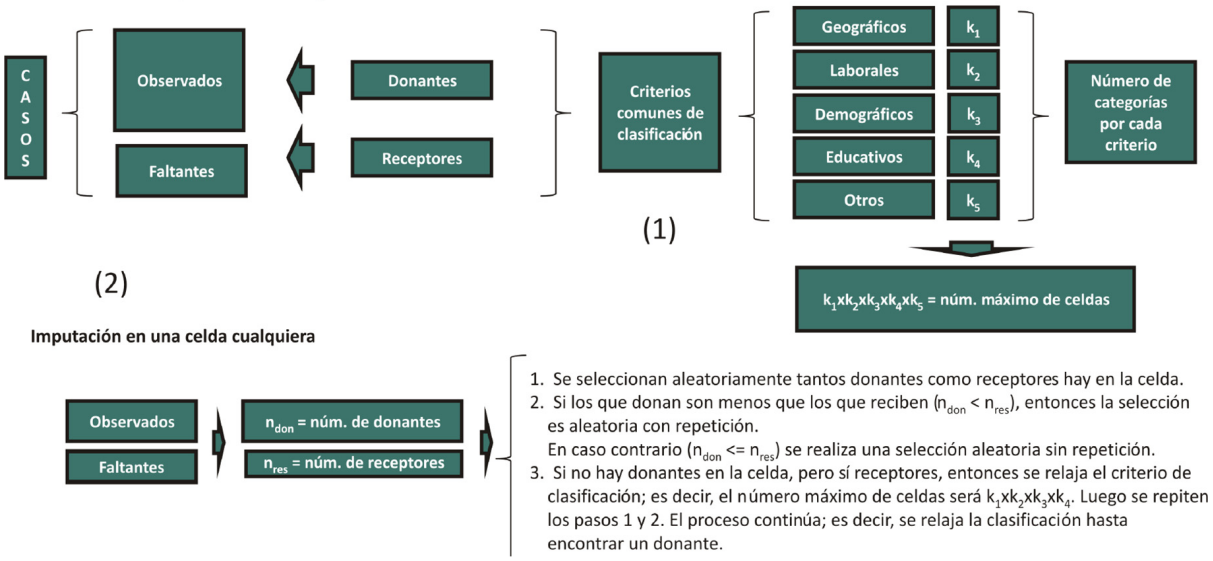

Figura 2. Proceso de imputación por Hot-deck , fuente (MESEP, 2012, pp.23)

Las algunas ventajas de este procedimiento son expuestas por Avila (2002), siendo la reducción del sesgo de no respuesta, presentación de datos limpios y consistentes y la preservación de las propiedades distribución dentro de cada uno de los conjuntos clasificados.

- **Regresión lineal:** Este procedimiento consiste en eliminar toda la información incompleta para así generar una regresión lineal y predecir los valores $\hat{y}$ que serán

los valores usados para substituir los valores faltantes, es decir las $\hat{y}$ son una media condicionada de las covariables X Median y Galván (2007). La regresión lineal esta dada por la forma:

$$Y = X\beta + e \tag{2}$$

Donde:

Y: Vector columna de tamaño $nx1$ que contiene las observaciones sobre la variable dependiente Y.

X: Es una matriz de tamaño nxp, $p \leq n$, donde la primera columna es de unos y tiene un rango igual a $k \leq p$. β: Vector columna de tamaño $px1$, de parámetros desconocidos $\beta_0, \beta_1, ..., \beta_{p-1}$. ϵ: Vector aleatorio o perturbado de tamaño $nx1$. (Jiménes, 2014)

- **Regresión lineal aleatoria:**Dado que las regresiones lineales siempre ajustan los valores al promedio, como se puede observar en la figura 3 , en ciertas ocasiones es necesario añadir un error a la predicción y así generar un patrón aleatorio. Para este patrón aleatorio se tiene en cuenta una distribución normal de los $\hat{y}$.

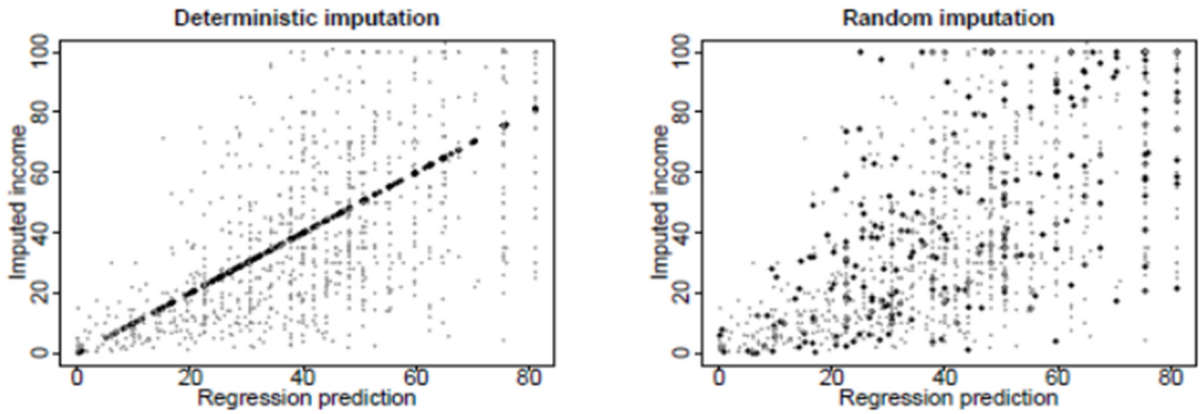

Figura 3. Regresión determinista vs. Aleatoria , fuente (Gelman, 2006, pp.537)

- **Media condicionada:** En esta técnica se basa en la construcción de grupos

relativamente homogéneos y se calcula el valor medio de los datos existentes dentro de estos grupos, siendo este valor medio el que substituye el valor faltante (Restrepo y Marín, 2019).

- **Estimación por máxima verosimilitud:** Asumiendo una distribución de datos faltantes MAR se estiman los parámetros del modelos con los datos completos con la función de máxima verosimilitud, estimando así los valores faltantes y evaluando de nuevo los parámetros con toda la información hasta logar la convergencia entre los parámetros. (Median y Galván, 2007)

- **Imputación múltiple:** Esta es la imputación del valor faltante M veces las donde los valores estimados siguen la distribución predictiva de los valores faltantes, cada uno de los valores imputados es analizado de manera separada y la variabilidad de la estadística está dado por los diversos resultados arrojados con los diferentes valores imputados (Briggs y cols., 2002).

Existen gran cantidad de referencias sobre la imputación de ingresos en diversas encuestas a nivel mundial, por ejemplo podemos encontrar trabajos como los presentados por Shenker y cols. (2006) cuyo objetivo principal era el de generar una imputación múltiple para la Encuesta Nacional de Salud (NHIS por sus siglas en inglés) en Estados Unidos, también se encuentra trabajos que buscan comparar diversos métodos de imputación e encontrar la opción más optima y viable para solucionar los datos faltantes (Starick y Watson, 2007), toda esta bibliografia lo que muestra es que según el contexto de los datos y la naturaleza misma de ellos los métodos de imputación pueden servir mejor en algunos casos que en otros, es por esto que es necesaria la evaluación de diferentes métodos para concluir cual es el más acertado. Así mismo esta bibliografía define el hot deck como uno de los métodos más usados dentro de las imputaciones de ingresos, siendo la regresión aleatoria y la media condicionada métodos mucho menos comunes, por tal motivo se eligieron estos para la evaluación en el presente documento.

3.3 Estimación de la pobreza

Para la medición de la pobreza en términos monetarios se requieren fundamentalmente dos variables, las líneas de pobreza y el ingreso, para la construcción de las primer variable el DANE siguió los lineamientos de Ravallion con ciertas variantes propias para Colombia y también se siguieron algunos aspectos metodológicos emanados de CEPAL (MESEP, 2012). Los pasos seguidos para esta construcción fueron:

I Construcción del gasto corriente per cápita a nivel de la unidad de gasto.

II Construcción del Deflactor Espacial de Precios (DEP).

III Ordenamiento de los hogares por percentil de gasto corriente per cápita deflactado.

IV Aplicación del método iterativo para la selección de la población de referencia.

V Construcción de la canasta básica de alimentos.

VI Ajuste normativo de cantidades de la canasta básica de alimentos para alcanzar los requerimientos calóricos.

VII Valoración de la canasta de alimentos ajustada: LI.

VIII Paso de la LI a la LP a partir del coeficiente de Orshansky.

Para al conformación de los ingresos la población objetivo son las personas que pertenecen a la población en edad de trabajar, es decir las personas que son de 12 años y más en las cabeceras municipales o 10 años y más en los centros poblados y rural disperso. Para la medición de la pobreza se debe conformar el ingreso corriente disponible de los hogares, según el DANE (2012) en la GEIH este se compone de:

- **Ingreso Monetario Primera Actividad IMPA**: Para asalariados es el 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)

Adicionalmente a los ingresos anteriormente descritos también se agregan los ingresos imputados por tenencia de vivienda propia, ya sea que sea propia totalmente pagada o propia la están pagando. sí el ingreso son menores que las líneas la persona será reconocida como pobre (MESEP, 2012).

La incidencia de la pobreza es:

$$ip = \frac{\sum personas\, pobres}{\sum población\, total} \tag{3}$$

Capítulo 4

Metodología

4.1 Descripción de datos

Los datos para el presente estudio son obtenidos de la Gran Encuesta Integrada de Hogares, cuyo objetivo principal es proporcionar información básica sobre el tamaño y estructura de la fuerza de trabajo(empleo, desempleo e inactividad) de la población del país, así como de las características sociodemográficas de la población colombiana, siendo uno de los objetivos específicos el conocer los ingresos de los hogares tanto en dinero como en especie, que sirvan de insumo para las mediciones sobre pobreza.

Esta encuesta posee una muestra probabilística, estratificada, de conglomerados y multietápica, a continuación se hará referencia a cada una de estas características teniendo en cuenta las definiciones del DANE (2016):

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

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

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

- **Multietápico:**En una primera etapa, la UPM, utilizando la técnica de selección controlada dentro de cada estrato. Para la segunda etapa se seleccionó en la cabecera y centros poblados la manzana, y en el rural disperso la sección o sea la USM.En la tercera etapa se seleccionó el segmento o UTM.

Los hogares encuestados se encuentran en las cabeceras y centros poblados de cerca de 437 municipios encuestados, siendo la muestra teórica de aproximadamente 248.000 hogares encuestados a nivel anual, concentrados en 22.548 segmentos. En la cuadro2 podemos encontrar los tamaños de muestra reales de la GEIH para los años 2013 a 2017.

Cuadro 2
Muestra total en la GEIH 2013-2017

Año	**Cantidad de hogares encuestados**	**Cantidad de personas encuestadas**
2013	228.944	797.877
2014	228.932	788.101
2015	232.219	787.044
2016	231.178	778.238
2017	230.909	767.867

Elaboración propia

La obtención de toda la información se logro por medio del Archivo Nacional de Datos (ANDA) y cabe resaltar que cada una de las bases de datos se manejarán de manera individual, por lo tanto cada uno de los resultados es independiente al resultado obtenido en el año anterior. La tabla específica con la cual se realizó todo el estudio es la de personas, en tabla 1 de los anexos se encuentra el diccionario de datos de esta tabla con el fin de identificar la información que esta contiene.

4.2 Test Little MCAR

Para la realización del Test de Little se usarán las bases de datos para los años 2013 a 2017, dado que las bases de datos contienen datos faltantes, que no hacen parte del

ingreso, para la aplicación de este test se usaran únicamente las variables que están completas para todas las observaciones, a continuación se identifican cuáles son estas variables:

- **Clase:** Hace referencia a la ubicación del hogar encuestado las opciones son cabeceras municipales o centros poblados y rural disperso.
- **Mes:** Mes de captación de la información, dado que la muestra está equidistribuida en cada uno de los años, cada mes posee un número similar de muestra.
- **Estrato1:** Hace referencia al estrato socioeconómico en el cual se encuentra el hogar, para las 13 principales ciudades es el estrato asociado al servicio de electricidad, mientras que para las demás cabeceras municipales y centros poblados y rural disperso es el sextil que pertenece el hogar según su Índice de Condiciones de Vida.
- **P6020:** Sexo de la persona encuestada.
- **P6040:** Edad de la persona encuestada.
- **P6210:** Nivel educativo más alto alcanzado por la persona encuestada.
- **P6210s1:** Grado en ese nivel educativo.
- **Ingtobs:** Ingreso total de las personas observado.

Para la prueba las hipótesis establecidas son:
Hipótesis nula

$$H_0 : y_{oi}|r \sim N(\mu_{oj}, \Sigma_{oj}) \tag{4}$$

Es decir que si un indicador condicional de faltantes r_i los y_{oi} se distribuyen normal con μ_{oj} y Σ_{oj} en todos los patrones se tiene la misma relación entre las variables, o mejor decirlo no existe ninguna relación por ende son MCAR.

Hipótesis alternativa

$$H_0 : y_{oi}|r \sim N(\nu_{oj}, \Sigma_{oj}) \tag{5}$$

Donde ν_{oj} es un vector de medias para cada patrón j y cada media es diferente, mostrando una relación entre los faltantes y los datos observados diferentes para cada patrón, es decir no es MCAR. Para la evaluación del test se usará el *p-valor* con un $\alpha = 0{,}05$.

4.3 Simulación de datos faltantes

Dado que el presente trabajo busca determinar si un aumento en la cantidad de valores faltantes marca diferencias entre los métodos de imputación, es por esto que es necesario generar nuevos valores faltantes dentro de las muestras. Para la selección de personas a las cuales les serán eliminados sus valores de ingreso se usará una selección de muestra sin reemplazamiento para probabilidades desiguales.

En esta técnica, como lo expone Berger y Tillé (2009), se debe considerar la selección de probabilidades p_k como

$$p_k = \frac{x_k}{\sum_{\varphi \in U}}, k \in U \tag{6}$$

El método de selección es por medio de un número aleatorio uniforme u entre 0 y 1 y se selección las unidades k de tal manera que $u_{k-1} \leq u \leq u_k$ donde

$$u_k = \sum_{\varphi=1}^{k} p_\varphi, con \quad 0 = 0 \tag{7}$$

Este proceso se realiza de manera independiente m veces.

4.4 Técnicas de imputación

A continuación, se presentarán la metodología para el uso de cada uno de los métodos de imputación evaluados en el presente documentos.

4.4.1 Hot-Deck Como se observó en la sección de la metodología el Hot-Deck necesita unos criterios comunes de clasificación para reconocer un grupo de donantes más cercanos, los cuales serán los donantes de la información. Las variables que se tuvieron en cuenta para realizar la clasificación se basan en las escogidas por el DANE en su metodología para la imputación del IMPA pero con una consideración adicional para los inactivos y desocupados en la posición ocupacional, las variables usadas fueron las :

Cuadro 3
Variables usadas como clasificación para el Hot-Deck

Variable	**Clasificación**
Dominio	1. Trece ciudades principales
	2. Resto de cabeceras
	3. Rural
Estrato	Estratos del 1 al 6
Edad	1. Menores de 18
	2. 18 a 24
	3. 24 a 44
	4. 45 y más
Educación	1. Ningún nivel
	2. Preescolar y primaria
	3. Secundaria y superior
Posición Ocupacional	1. Obreros y empleados
	2. Empleados domésticos
	3. Trabajadores cuenta propia
	4. Patronos
	5. Ocupados sin remuneración
	6. Desocupados
	7. Inactivos
Sexo	1. Masculino
	2. Femenino
Jefe de hogar	1. Jefe
	0. No es jefe

Elaboración propia

Cabe resaltar que inicialmente se tienen en cuenta todas las variables clasificadoras para la búsqueda de los donantes más cercanos, pero si no existen donantes se quitan de manera iterativa las variables iniciando con jefatura y terminando con dominio hasta encontrar al menos un donante.

4.4.2 Regresión lineal aleatoria Para realizar las estimaciones por medio de la regresión lineal las variables utilizadas son e su gran mayoría las utilizadas por el DANE para la identificación de valores atípicos por medio de regresiones cuantílicas, teniendo en cuenta esto el modelo que se propone para la obtención de los $\hat{y}$ que serán la base de la imputación es:

$$\begin{aligned} Y_i = \; & \beta_0 + \beta_1 edad_i + \beta_2 edad2_i + \beta_3 horas_i + \beta_4 a\_edu_i + \beta_5 resto_i \\ & + \beta_6 capitales_i + \beta_7 O\_capitales_i + \beta_8 sexo_i + \beta_9 obrero_i + \beta_{10} emp\_domes_i \\ & + \beta_{11} cuenta\_pro_i + \beta_{12} patrono_i + \beta_{13} jefe_i + \beta_{14} numero\_per_i \\ & + \beta_{15} numero\_asala_i + \beta_{16} numero\_indep_i + \beta_{17} numero\_desocu_i \\ & + \beta_{18} numero\_infan_i + \beta_{19} numero\_adole_i + \beta_{20} num\_sin\_edu_i \\ & + \beta_{21} num\_edu\_sup_i + \beta_{22} a\_edu\_hog_i + \beta_{23} salud_i + \beta_{24} cesantes_i \end{aligned} \tag{8}$$

Donde:

$edad$ = edad de la $i-$ésisma persona.

$edad2$ = edad al cuadrado de la $i-$ésisma persona.

$horas$ = horas trabajadas por la $i-$ésisma persona.

a_edu = años de educación la $i-$ésisma persona.

$resto$ = variable dummy que identifica la zona rural como el lugar de residencia de la $i-$ésisma persona.

$capitales$ = variable dummy que identifica las trece principales ciudades como el lugar de residencia de la $i-$ésisma persona.

$O\_capitales$ = variable dummy que identifica las capitales diferentes a las trece principales ciudades como el lugar de residencia de la $i-$ésisma persona.

$sexo$ = variable dummy que identifica el sexo $i-$ésisma persona con 1 para hombre 0 para mujer.

$obrero$ = variable dummy que identifica los ocupados con posición ocupacional de la persona $i-$ésisma como obrero.

emp_domes = variable dummy que identifica los ocupados con posición ocupacional de la persona $i-$ésisma como empleado domestico.

$cuenta\_pro$ = variable dummy que identifica los ocupados con posición ocupacional de la persona $i-$ésisma como cuenta propia.

$patrono$ = variable dummy que identifica los ocupados con posición ocupacional de la persona $i-$ésima como patrono.

$jefe$ = variable dummy que identifica si la $i-$ésima persona es jefe del hogar.

$numero\_per$ = cantidad de personas que viene en el hogar de la $i-$ésima persona.

$numero\_asala$ = cantidad de personas asalariadas que viven en el hogar de la $i-$ésima persona.

$numero\_indep$ = cantidad de personas independientes que viven en el hogar de la $i-$ésima persona.

$numero\_desocu$ = cantidad de personas desocupadas que viven en el hogar de la $i-$ésima persona.

$numero\_infan$ = cantidad de personas menores de 5 años que viven en el hogar de la $i-$ésima persona.

$numero\_adole$ = cantidad de personas entre 14 y 17 años que viven en el hogar de la $i-$ésima persona.

num_sin_edu = cantidad de personas mayores de 25 años que no poseen educación y viven en el hogar de la $i-$ésima persona.

num_edu_sup = cantidad de personas con educación superior y viven en el hogar de la $i-$ésima persona.

a_edu_hog = años de educación total del hogar donde vive la $i-$ésima persona.

$salud$ = cantidad de personas con afiliadas a el sistema de salud y viven en el hogar de la $i-$ésima persona.

$cesantes$ = variable dummy que identifica la $i-$ésima persona como cesante.

Todos los resultados para cada uno de los años de los betas de esta regresión se encuentran en los anexos

4.4.3 Media Condicionada Las variables que condicionaron las medias y así mismo la imputación fueron clase, estrato y posición ocupacional, la razón principal es que se necesitan al menos dos valores dentro de cada uno de los grupos para generar una media y si se agregan más variables algunos grupos quedan vacíos, es decir sin información dada la cantidad de faltantes.

4.5 Estimación del error estándar de la pobreza

Para la estimación de los errores estándar se usará el método de Bootstrap para la estimación de errores estándar propuesto por Efron y Tibshirani (1986). Este consiste en asumir que se tiene una información observada $y = (x_1, X_2, ..., x_n)$ que proviene una muestra aleatoria de $X_1, X_2, ..., X_n$ con una distribución de probabilidad F desconocida, sobre estas observaciones se tiene el estadístico de interés $\hat{\theta}(y)$ al cual le queremos asignar el error estándar.

Ahora asumamos que $\sigma(F)$ indica el error estándar de $\hat{\sigma}$, como una función de la distribución desconocida F

$$\sigma(F) = [VAR_F \{\hat{\sigma}(y)\}]^{1/2} \tag{9}$$

El boostrap puede estimar $\hat{F}$ por ende $\hat{\sigma}$ está definido como

$$\hat{\sigma} = \sigma(\hat{F}) \tag{10}$$

Para la estimación de F es necesario implementar un algoritmo de Monte Carlo que trabaje de la siguiente manera:

- Usar un generador de números aleatorios, generando una gran cantidad de muestras bootstrap $y^*(1), y^*(2), ..., y^*(B)$.
- Para cada una de las muestras bootstrap evaluar el estadístico de interés $\hat{\theta}(b) = \hat{\theta}(y^*(b))$.
- Calcular el error el error estándar de las muestras de los $\hat{\theta}^*(b)$.

Capítulo 5

Análisis y Resultados

En este capítulo se presentan todos los resultados obtenidos de los diferentes procedimientos propuestos en la metodología, por tal motivo se presentarán las mismas secciones con los resultados respectivos.

5.1 Test Little MCAR

En la tabla 4 se pueden observar cuales fueron los resultados del Test Little para cada uno de los años, dado que los únicos valores faltantes evaluados en las bases fueron los presentados en la variable ingreso los patrones evaluados fueron 2, con y sin valores faltantes.

Cuadro 4
Resultados del Test Little MCAR

	2013	**2014**	**2015**	**2016**	**2017**
Chi cuadrado	1.385	1.837	1.826	1.732	1.716
Grados de libertad	8	8	8	8	8
P.valor	0	0	0	0	0
Patrones de faltantes	2	2	2	2	2

Elaboración propia

Los valores del χ^2 para todos los años superan los valores de 1.000 por ende los $valores - p$ son muy pequeños y dentro de la tabla son representados con un 0, esto quiere decir que con un nivel de confianza del 95 % se rechaza la hipótesis nula, por lo tanto, no existe evidencia estadística para afirmar que los datos son completamente aleatorios (MCAR).

Este resultado muestra que en alguna medida los faltantes están relacionados con otras variables, es decir, para este caso particular muestra que los ingresos faltantes no son aleatorios dentro de la población, algunas personas en particular son los que deciden no darnos los ingresos personales. Esto también permite aseverar que los métodos de imputación escogidos en el presente trabajo son acertados dada la naturaleza de los datos ya que usa variables auxiliares para la generación de los valores a imputar.

5.2 Simulación de datos faltantes

Dado que se tienen diversos niveles de datos perdidos para cada uno de los años cada uno de los años posee un valor diferente de datos adicionales que fueron eliminados, por ejemplo, para el años 2013 se tenía que para llegar a una faltante de información del 15 % se necesitó eliminar la información de 27.649 personas adicionales, mientras que para el 2017 la cantidad adicional de personas a las cuales se les eliminó el ingreso fue de 3.331. En la tabla 5 se encuentra el total de faltantes adicionales que fueron seleccionados para cada uno de los años para llegar a un nivel de 15 %, 20 % y 25 % de información faltante.

Cuadro 5
Cantidad de datos seleccionados para la simulación de faltantes

	Cantidad de personas	**Faltantes actuales**	**15 %**	**20 %**	**25 %**
2013	432.571	37.237	27.649	49.277	70.906
2014	437.272	49.752	15.839	37.702	59.566
2015	444.960	50.523	16.221	38.469	60.717
2016	439.611	53.037	12.905	34.885	56.866
2017	437.547	62.301	3.331	25.208	47.086

Elaboración propia

Cabe resaltar que para la selección de estos datos faltantes adicionales se usó un muestreo sin reemplazamiento para probabilidades desiguales, todo esto teniendo en cuenta los valores actuales de faltantes en ciudades y estratos siendo esta la información usada para generar las probabilidades estimadas para cada uno de los grupos.

5.3 Resultados de las imputaciones

El resultado inicial de todas las imputaciones fue una variable de ingresos por personas, para reconocer cuales pudieron ser los cambios que produjeron las imputaciones, según el método y la cantidad de faltantes, se generaron gráficas que mostraran la distribución kernel de la variable ingresos. En la figura 4 se puede apreciar esta distribución, para cada uno de los métodos de imputación en el año 2017, para el logaritmo natural de los ingresos, se optó por generar la distribución con el logaritmo natural dado que los ingresos al ser tan dispersos gráficamente no era apreciable algún cambio.

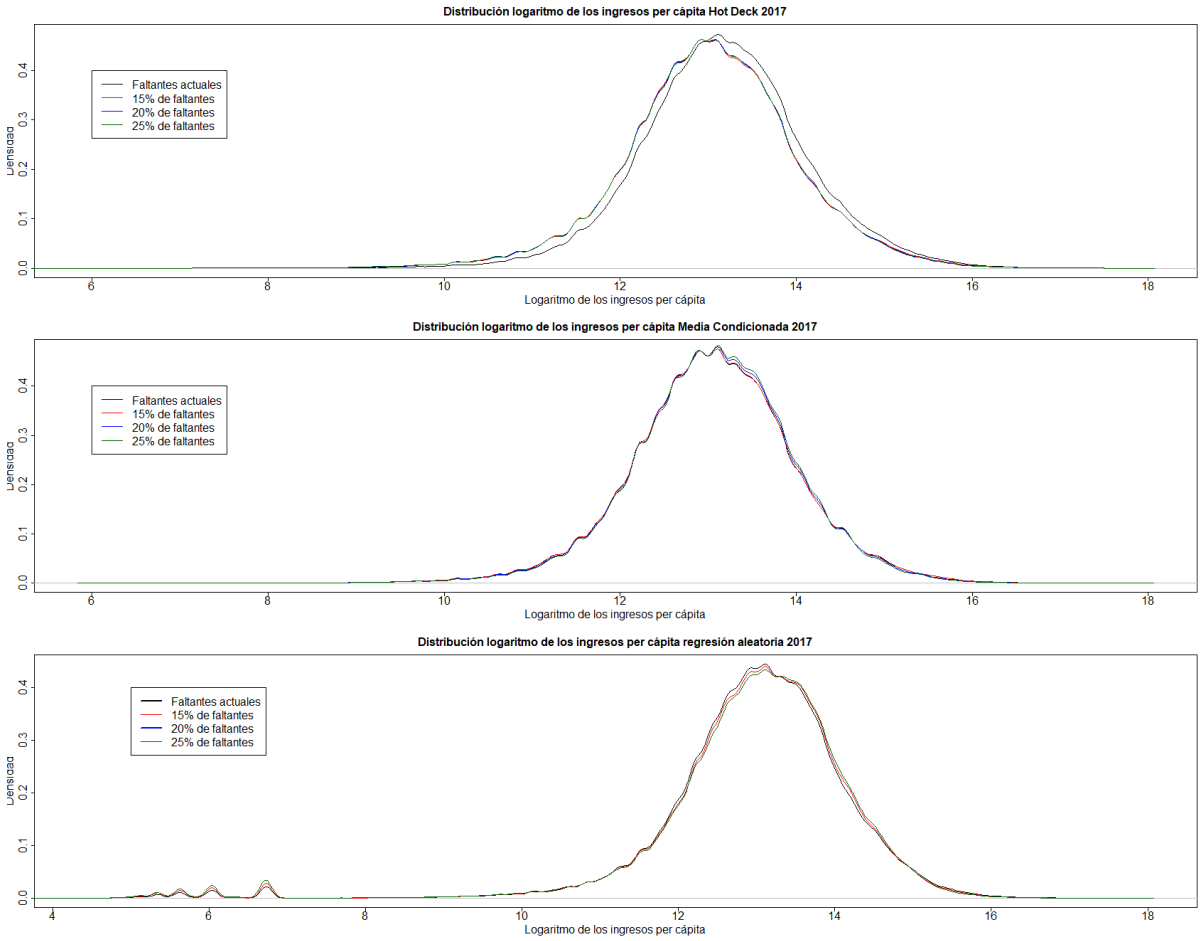

Figura 4. Distribución del logaritmo natural de los ingresos per cápita para 2017

Lo que se puede apreciar es que todos las imputaciones mantienen una distribución muy similar de ingresos, tanto entre faltantes como entre métodos de imputación, aunque el resultado de la regresión lineal aleatoria muestra una agrupación de datos en

la parte baja de la distribución que no se presenta en ninguna otras forma de imputación. Así mismo se observa que la imputación con los faltantes actuales en el Hot-Deck está más a la derecha de las demás imputaciones trayendo consigo una menor estimación de la pobreza.

Incidencia de la pobreza Una vez obtenido el ingreso per cápita se comparan con las líneas de pobreza, dando como resultado la incidencia de la pobreza, en la figura 5 y 6 se pueden observar los resultados de esta incidencia de manera gráfica, en los cuadros 10 del anexo se pueden observar la cantidad de personas pobres y su incidencia para cada uno de los años.

Lo primero que se puede observar, es que, a pesar que en todos los niveles de faltantes existen diferencias según el método de imputación todas las incidencias a través del tiempo poseen la misma tendencia. Específicamente en los datos faltantes actuales es dónde existe la mayor diferencia, el Hot-Deck presenta una menor incidencia de la pobreza, como resultados de una distribución corrida a la derecha como fue presentada en la sección anterior.

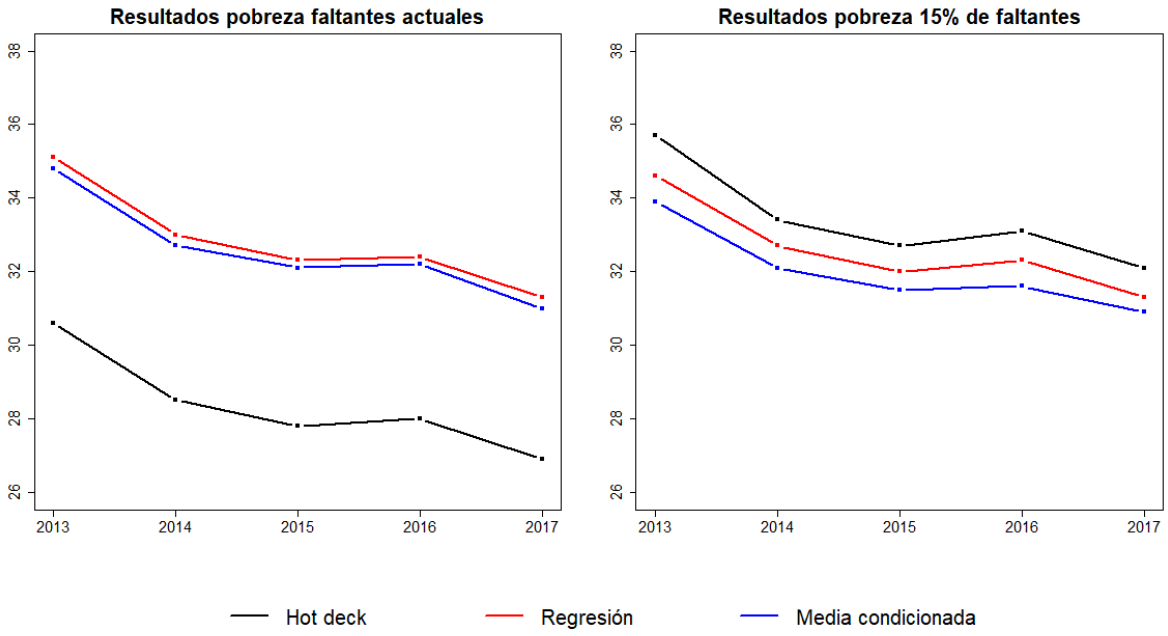

Figura 5. Incidencia de la pobreza para los años 2013-2017 con los faltantes actuales y un 15 % de faltantes

Adicionalmente en todos los faltantes simulados la media condicionada es el método

que presenta la menor incidencia de la pobreza, mientas que el Hot-Deck la mayor, esto puede estar dado a la perdida de información y por ende de donantes para la realización de este último método de imputación lo cual tiene como resultado que cada vez sean donantes menos cercanos los que sean los que den la información. Mientras que la media condicionada posee una menor cantidad de variables clasificatorias y a pesar de la perdida de información las regresiones mantienen unos betas parecidos en todos los escenarios.

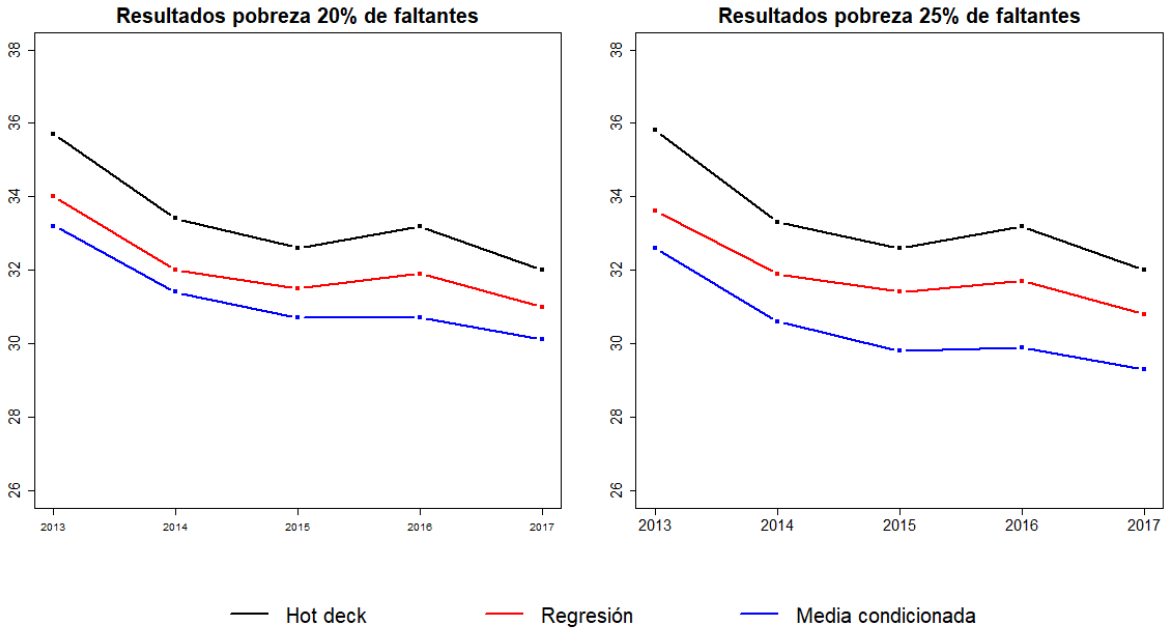

Figura 6. Incidencia de la pobreza para los años 2013-2017 con 20 % y 25 % de faltantes

Gracias a la estimación del error estándar de la estimación de la incidencia de la pobreza se pudieron generar los limites inferiores y superiores de la estimación y sus coeficientes de variación como lo se puede observar en la tabla 6 para los resultados de 2017.

Cuadro 6
Incidencia de la pobreza, intervalos de confianza y Cve para 2017

	Faltantes	**Incidencia de la pobreza**	**Limite inferior 95 %**	**Limite Superior 95 %**	**Cve %**
Hot-Deck	**Actual**	26,9	26,5	27,4	0,9
	15 %	32,1	31,6	32,5	0,7
	20 %	32,0	31,5	32,4	0,7
	25 %	32,0	31,6	32,5	0,7
Media Condicionada	**Actual**	31,0	30,5	31,4	0,8
	15 %	30,9	30,4	31,4	0,8
	20 %	30,1	29,6	30,5	0,8
	25 %	29,3	28,9	29,8	0,8
Regresión aleatoria	**Actual**	31,3	30,9	31,8	0,7
	15 %	31,3	30,8	31,7	0,7
	20 %	31,0	30,5	31,5	0,8
	25 %	30,8	30,4	31,3	0,8

Elaboración propia

Una situación particular, que se puede medir gracias a la estimación de los errores estándar del parámetro estimado, es que todas las incidencias de la pobreza , exceptuando la del Hot-Deck con los faltantes actuales y la media condicionada con un faltante del 25 %, no son estadísticamente diferentes con un nivel del 95 % de confianza.

Capítulo 6

Conclusiones y Recomendaciones

- Teniendo en cuenta los resultados del Test Little se puede afirmar que los faltantes en la variable ingresos no se presentan de una manera completamente aleatoria, es decir su mecanismo no es MCAR y por lo tanto pueden ser datos perdidos no aleatorios (MNAR) o datos perdidos aleatorios (MAR). Esto quiere decir que las imputaciones que se deben realizar a esta variable deben depender de variables adicionales.

- Los datos imputados no han cambiado las tendencias de la pobreza y dado que en la gran mayoría de los casos las incidencias calculadas no poseen una diferencia estadística, no se puede definir con certeza si existen diferencias en el cálculo de la pobreza con los tres métodos elegidos.

- No se puede determinar cuál es el método más apropiado para la imputación dado lo similar de los resultados tanto en incidencia como en la varianza de la misma, por tal motivo los efectos en la estimación de la pobreza son indeterminados y es necesario ahondar en otros métodos de imputación para reconocer si alguno genera grandes diferencias en los resultados.

- Es necesario caracterizar los hogares pobres y no pobres para todos los métodos de imputación y cantidad de faltantes, como extensión al presente trabajo. esto con el fin de determinar si a pesar de no ser muy diferentes las estimaciones los perfiles de la pobreza cambian generando una mejor identificación de los hogares

pobres, en pocas palabras conocer si los hogares que son seleccionados como pobres son los que efectivamente son pobres.

Referencias

Aguirre, V. H. M. (2012, 6). El poblema de la pobreza en la utopía aristotílica. En *Actas del vi coloquio internacional competencia y cooperación de la antigua grecia en la actualidad.* Buenos Aires, Argentica.

Avila, C. (2002). Método de imputación hot deck. (Monografia (Título Profesional de Licenciado en Estadística), Universidad Nacional Mayor de San Marcos, Lima, Perú)

Beltrán, E. P. (2000). La pobreza en smith y ricardo. , *2*, 111-130.

Berger, Y., y Tillé, Y. (2009). Sampling with unequal probabilities. , *29*.

Briggs, A., Clark, T., Wolstenholme, J., y Clarke, P. (2002). *Missing.... presumed at random: cost-analysis of incomplete data* (Vol. 12(5)).

Campos-Vázquez, R. (2013). Efectos de los ingresos no reportados en el nivel y tendencia de la pobreza laboral en méxico. , *32*, 23-54.

Cheng, L. (2013). Little's test of missing completely at random. , *13*, 795-809.

DANE (Ed.). (2012). Algoritmo para la conformación del ingreso per cápita de pobreza a partir de la encuesta continua de hogares - ech (2002-2005) y de la gran encuesta integrada de hogares - geih (2008-2010) [Manual de software informático].

DANE (Ed.). (2016). Ficha metodológica gran encuesta integrada de hogares - geih [Manual de software informático].

Davis, P., y Sanchez-Martinez, M. (2014). A review of the economic theories of poverty. , *Discussion Paper No. 435*.

Efron, B., y Tibshirani, R. (1986, 02). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci.*, *1*(1), 54–75. Descargado de `https://doi.org/10.1214/ss/1177013815` doi: 10.1214/ss/1177013815

Enders, C. (2010). *Applied data missing analysis.* New York, Estados Unidos: The Guilford Press.

Gelman, A. (2006). *Missing-data imputation.* New York, Estados unidos: Columbia University.

Haughton, J., y Khandker, S. R. (2009). *Handbook on poverty and inequality.* Washington D. C., Estados Unidos: The World Bank. Descargado de http://documents.worldbank.org/curated/en/488081468157174849/Handbook-on-poverty-and-inequality

Heeringa, S., West, B., y Berglund, P. (2010). *Applied survey data analysis.* Boca Ratón, Estados Unidos: Taylor and Francis Group.

Jiménes, J. A. (2014). *Álgebra matricial con aplicaciones estadísticas.* Bogotá, Colombia: Universidad Nacional.

Maceri, S. (2009). El concepto de la riqueza en platón en tanto impedimento para el estado justo. , *5*, 165-184.

Median, F., y Galván, M. (2007). *Imputación de datos teoría y práctica.* Santiago de Chile, Chile: CEPAL.

MESEP (Ed.). (2012). Pobreza monetaria en colombia:nueva metodología y cifras 2002-2010 [Manual de software informático]. Descargado de https://www.dane.gov.co/files/noticias/Pobrezanuevametodologia.pdf

Muñoz, J. (2015). La pobreza y las politicas públicas: del referencial global al sectorial. , *88*, 99-119.

of Economic, D., y Affairs, S. (2008). *Designing household survey samples: Practical guidelines* (U. Nations, Ed.). New York, Estados Unidos.

Ravallion, M. (1998). Poverty lines in theory and practice. , *133*. Descargado de http://documents.worldbank.org/curated/en/916871468766156239/Poverty-lines-in-theory-and-practice

Restrepo, M., y Marín, J. (2019). Imputación de ingresos en la gran encuesta integrada de hogares (geih) de 2010. , *70*, 219-243.

Rubin, D. (1976). Inference and missing data. , *63*, 581-592. Descargado de http://people.csail.mit.edu/jrennie/trg/papers/rubin-missing-76.pdf

Sano, S., Tada, S., y Yamamoto, M. (2015). Method of household surveys and characteristics of surveyed households: Comparison regarding household composition, annual income and educational attainment. , *11*, 505-529.

Shenker, N., Raghunathan, T., Chiu, P.-L., Maduk, D., Zhang, G., y Cohen, A. (2006). Multiple imputation of missing income data in the national health interview survey. , *101*, 924-933. doi: 10.1198/016214505000001375

Smith, T. W. (1991). *An analysis of missing income information on the general social surveys (gss methodological report no. 71).*

Soley-Bori, M. (2013). *Dealing with missing data: Key assumptions and methods for applied analysis (technical report no. 4).* Descargado de `https://www.bu.edu/sph/files/2014/05/Marina-tech-report.pdf`

Starick, R., y Watson, N. (2007). *Evaluation of alternative income imputation methods for the hilda survey* (T. U. of Melbourne, Ed.).

Zhu, X. (2014). Comparison of four methods for handing missing data in longitudinal data analysis through a simulation study. , *4*, 933-944. Descargado de `http://dx.doi.org/10.4236/ojs.2014.411088`

Apéndice A

Cuadro 1
Diccionario de datos tabla personas 1 de 4

Variable	Descripción	Variable	Descripción
Directorio	Llave de vivienda	P7070	¿cuánto recibió o ganó . . . .. el mes pasado en ese segundo trabajo o negocio?
Secuencia_p	Llave de hogar	P7090	Además de las horas que trabaja actualmente ¿..... quiere trabajar más horas?
Orden	Llave de persona	P7110	Durante las últimas 4 semanas, ¿ ..... hizo diligencias para trabajar más horas?
Clase	1. Cabecera, 2. Resto (centros poblados y área rural dispersa)	P7120	Si la semana pasada le hubiera resultado la posibilidad de trabajar más horas ¿ estaba..... disponible para hacerlo?
Dominio	Cada una de las 24 a.M., otras cabeceras y resto	P7140s1	¿Por que motivos ..... desea cambiar de trabajo o empleo: a. Para mejorar la utilización de sus capacidades o formación?
Capital	Dummy capital de departamento. 1: capital 0: no capital	P7140s2	¿Por que motivos ..... desea cambiar de trabajo o empleo: b. Desea mejorar sus ingresos?
Mes	Mes	P7150	Durante las ÚLTIMAS 4 SEMANAS, ¿..... hizo diligencias para cambiar de trabajo?
Estrato1	Estrato de energía para las 13 a.M., y sextil de icv para otras cabeceras y resto	P7160	Si le resultara un nuevo trabajo o empleo a...¿podría empezar a desempeñarlo antes de un mes?
P6020	Sexo	P7310	¿..... ha buscado trabajo por primera vez o había trabajado antes por lo menos durante dos semanas consecutivas?
P6040	¿cuántos años cumplidos tiene?	P7350	En este último trabajo era: . . . (Desocupados)
P6050	¿cuál es el parentesco de ...Con el jefe o jefa del hogar?	P7422	¿Recibió o ganó el mes pasado ingresos por concepto de trabajo?. (Desocupados)
P6090	"¿ ... Está afiliado, es cotizante o es beneficiario de alguna entidad de seguridad social en salud?	P7422s1	¿Cuánto? $_____
P6100	¿A cual de los siguientes regímenes de seguridad social en salud está afiliado:	P7472	¿recibió o ganó el mes pasado ingresos por concepto de trabajo?. (desocupados)
P6210	¿Cuál es el nivel educativo más alto alcanzado por .... y el último año o grado aprobado en este nivel?	P7472s1	¿cuánto? $_____
P6210s1	Grado escolar aprobado	P7495	El mes pasado, ¿recibió pagos por concepto de arriendos y/o pensiones?
P6240	¿En que actividad ocupó..... la mayor parte del tiempo la semana pasada?	P7500s1	¿El mes pasado, recibió pagos por: a. arriendos de casas, apartamentos, fincas, lotes, vehículos, equipos etc?
Oficio	¿qué hace. . . . . en este trabajo?	P7500s1a1	Valor mes pasado $ _____
P6426	¿cuanto tiempo lleva ... Trabajando en esta empresa, negocio, industria, oficina, firma o finca de manera continua?	P7500s2	¿El mes pasado recibió pagos por b. pensiones o jubilaciones por vejez,invalidez o sustitución pensional ?
P6430	En este trabajo es . . . . (posición ocupacional primera actividad)	P7500s2a1	Valor mes pasado $ _____
P6500	Antes de descuentos ¿cuánto ganó .... el mes pasado en este empleo?	P7500s3	¿El mes pasado recibió pagos por c. pensión alimenticia por paternidad, divorcio o separación?
P6510	¿el mes pasado recibio ingresos por concepto de horas extras?	P7500s3a1	Valor mes pasado $ _____

Cuadro 2
Diccionario de datos tabla personas 2 de 4

Variable	Descripción	Variable	Descripción
P6510s1	¿cuánto recibió por horas extras?	P7505	Durante los últimos doce meses, ¿recibió dinero de otros hogares, personas o instituciones no gubernamentales; dinero por intereses, dividendos, utilidades o por cesantias?
P6510s2	¿incluyó este valor en los ingresos del mes pasado?	P7510s1	Durante los últimos 12 meses, ¿recibió a. dinero de otros hogares o personas residentes en el país?
P6545	El mes pasado recibió a. Primas (técnica, de antigüedad, clima,orden publico, otras, etc)	P7510s1a1	Valor $ _____
P6545s1	¿cuánto recibió por primas?	P7510s2	Durante los últimos 12 meses, ¿recibió b. dinero de otros hogares o personas residentes fuera del país?
P6545s2	¿incluyó este valor en los ingresos del mes pasado ($ ____) que me declaró anteriormente?	P7510s2a1	Valor $ _____
P6580	¿el mes pasado recibió b. Bonificaciones?	P7510s3	Durante los últimos 12 meses, ¿recibió c. ayudas en dinero de instituciones del país?
P6580s1	¿cuánto recibió por bonificaciones?	P7510s3a1	Valor $ _____
P6580s2	¿incluyó este valor en los ingresos del mes pasado ($ ____) que me declaró anteriormente?	P7510s5	Durante los últimos 12 meses, ¿recibió d. dinero por intereses de prestamos o CDTs, depositos de ahorros, utilidades, ganancias o dividendos por inversiones?
P6585s1	¿el mes pasado recibió a. Auxilio o subsidio de alimentación?	P7510s5a1	Valor $ _____
P6585s1a1	¿cuánto recibió por subsidio de alimentación?	P7510s6	Durante los últimos 12 meses, ¿recibió e. dinero por concepto de cesantías y/o intereses a las cesantías?
P6585s1a2	¿incluyó este valor en los ingresos del mes pasado ($ ____) que me declaró anteriormente?	P7510s6a1	Valor $ _____
P6585s2	¿el mes pasado recibió b. Auxilio subsidio de transporte?	P7510s7	Durante los últimos 12 meses, ¿recibió f. dinero de otras fuentes diferentes a las anteriores?
P6585s2a1	¿cuánto recibió por subsidio de transporte?	P7510s7a1	Valor $ _____
P6585s2a2	¿incluyó este valor en los ingresos del mes pasado ($ ____) que me declaró anteriormente?	Pet	Población en edad de trabajar 1: sí 0: no
P6585s3	¿el mes pasado recibió c. Subsidio familiar?	Oc	Ocupado 1: sí
P6585s3a1	¿cuánto recibió por subsidio familiar?	Des	Desocupado 1: sí
P6585s3a2	¿incluyó este valor en los ingresos del mes pasado ($ ____) que me declaró anteriormente?	Ina	Inactivo 1: sí
P6585s4	¿el mes pasado recibio d. Subsidio educativo?	Impa	Ingreso monetario de la primera actividad antes de imputación
P6585s4a1	¿cuánto recibio por subsidio educativo?	Isa	Ingreso monetario de la segunda actividad antes de imputación
P6585s4a2	¿incluyó este valor en los ingresos del mes pasado ($ ____) que me declaró anteriormente?	Ie	Ingreso en especie antes de imputación
P6590	¿además del salario en dinero,¿el mes pasado recibió alimentos como parte de pago por su trabajo?	Imdi	Ingreso por trabajo de desocupados e inactivos antes de imputación

Cuadro 3
Diccionario de datos tabla personas 3 de 4

Variable	Descripción	Variable	Descripción
P6590s1	¿en cuánto estima lo que recibió? $_____	Iof1	Ingreso por intereses y dividendos antes de imputación
P6600	¿además del salario en dinero,¿el mes pasado recibió vivienda como parte de pago por su trabajo?	Iof2	Ingreso por jubilaciones y pensiones antes de imputación
P6600s1	¿en cuánto estima lo que recibió? $_____	Iof3h	Ingreso por ayudas de hogares, antes de imputación
P6610	¿normalmente... Utiliza transporte de la empresa para desplazarse a su trabajo (bus o automóvil)?	Iof3i	Ingreso por ayudas de instituciones, antes de imputación
P6610s1	¿en cuánto estima lo que recibió?$_____	Iof6	Ingreso por arriendos antes de imputación
P6620	Además del salario en dinero, ¿el mes pasado... Recibió otros ingresos en especie por su trabajo (electrodomésticos,ropa, productos diferentes aalimentos o bonos tipo sodexho)?	Cclasnr2	Estado de impa 1:faltante 0: observado
P6620s1	¿en cuánto estima lo que recibió? $_____	Cclasnr3	Estado de isa 1:faltante 0: observado
P6630s1	En los últimos 12 meses recibió ... a. Prima de servicios	Cclasnr4	Estado de ie 1:faltante 0: observado
P6630s1a1	¿cuánto recibio? $_____	Cclasnr5	Estado de imdi 1:faltante 0: observado
P6630s2	En los últimos 12 meses recibió ... B. Prima de navidad	Cclasnr6	Estado de iof1 1:faltante 0: observado
P6630s2a1	¿cuánto recibio? $_____	Cclasnr7	Estado de iof2 1:faltante 0: observado
P6630s3	En los últimos 12 meses recibió ... c. Prima de vacaciones	Cclasnr8	Estado de iof3 1:faltante 0: observado
P6630s3a1	¿cuánto recibió? $_____	Cclasnr11	Estado de iof6 1:faltante 0: observado
P6630s4	En los últimos 12 meses recibio ... D. Viáticos permanentes	Impaes	Ingreso monetario de la primera actividad imputado (sólo para faltantes, extremos o ceros inconsistentes)
P6630s4a1	¿cuánto recibió? $_____	Isaes	Ingreso monetario de la segunda actividad imputado (sólo para faltantes o extremos)
P6630s6	En los últimos 12 meses recibio ... e. Bonificaciones anuales	Iees	Ingreso en especie imputado (sólo para faltantes o extremos)
P6630s6a1	¿cuánto recibió? $_____	Imdies	Ingreso por trabajo de desocupados e inactivos imputado (sólo para faltantes o extremos)
P6750	¿cuál fue la ganancia neta o los honorarios netos de ... En esa actividad, negocio, profesión o finca, el mes pasado ?	Iof1es	Ingreso por intereses y dividendos imputado (sólo para faltantes o extremos)
P6760	¿ a cuántos meses corresponde lo que recibió?	Iof2es	Ingreso por jubilaciones y pensiones imputado (sólo para faltantes o extremos)

Cuadro 4
Diccionario de datos tabla personas 4 de 4

Variable	Descripción	Variable	Descripción
P550	¿cuál fue la ganancia neta del negocio o de la cosecha durante los últimos doce meses? (sólo para centros poblados y área rural dispersa)	Iof3hes	Ingreso por ayudas de hogares, imputado (sólo para faltantes o extremos)
P6800	¿cuántas horas a la semana trabaja normalmente.... en ese trabajo ?	Iof3ies	Ingreso por ayudas de instituciones, imputado (sólo para faltantes o extremos)
P6870	¿cuántas personas en total tiene la empresa, negocio, industria, oficina, firma, finca o sitio donde ..... Trabaja?	Iof6es	Ingreso por arriendos imputado (sólo para faltantes o extremos)
P6920	¿está... Cotizando actualmente a un fondo de pensiones?	Ingtotob	Ingreso total observado
P7040	Además de la ocupación principal, ¿.... tenía la semana pasada otro trabajo o negocio?	Ingtotes	Ingreso total imputado
P7045	¿cuántas horas trabajó ... La semana pasada en ese segundo trabajo?	Ingtot	Ingreso total
P7050	En ese segundo trabajo. . . ..es: (ocupación segunda actividad)	Fex_c	Factor de expansión anualizado

Cuadro 5
Resultados para el 2017

	Dependent variable:			
	Ingresos act	Ingresos 15 %	Ingresos 20 %	Ingresos 25 %
	(1)	(2)	(3)	(4)
edad	27,357.320*** (583.479)	27,300.670*** (582.198)	26,908.830*** (573.643)	26,835.330*** (581.939)
edad2	−93.362*** (6.149)	−93.448*** (6.136)	−93.888*** (6.047)	−94.431*** (6.136)
horas	7,264.875*** (136.881)	7,235.594*** (136.465)	7,148.987*** (134.165)	7,167.573*** (135.941)
a_edu	98,641.530*** (690.106)	98,163.670*** (688.112)	95,309.890*** (677.811)	94,060.450*** (687.027)
resto1	56,851.610*** (6,557.248)	54,335.980*** (6,554.349)	41,574.170*** (6,560.447)	36,582.930*** (6,791.878)
capitales1	−203,902.600*** (9,861.666)	−182,335.500*** (9,987.225)	−39,293.690*** (11,205.370)	19,443.000 (13,933.520)
O_capitales1	−311,065.600*** (8,928.300)	−288,811.000*** (9,065.975)	−131,290.100*** (10,380.620)	−55,817.790*** (13,224.020)
sexo1	189,773.100*** (4,052.131)	189,120.200*** (4,041.318)	186,240.300*** (3,980.224)	184,779.700*** (4,036.693)
obrero1	482,761.200*** (24,483.480)	483,116.500*** (24,406.260)	490,735.900*** (23,959.150)	486,837.200*** (24,309.380)
emp_domes1	303,266.100*** (27,067.680)	303,216.800*** (26,985.710)	309,561.100*** (26,504.450)	302,825.800*** (26,884.640)
cuenta_pro1	95,172.810*** (24,195.070)	95,564.500*** (24,117.380)	107,265.000*** (23,671.440)	103,601.900*** (24,013.540)
patrono	985,277.900*** (26,721.360)	986,844.400*** (26,640.430)	996,462.300*** (26,171.440)	988,447.500*** (26,566.560)
jefe1	178,232.100*** (4,446.029)	177,140.600*** (4,434.510)	173,899.100*** (4,368.756)	173,503.800*** (4,431.323)
numero_per	35,117.290*** (2,801.407)	34,636.720*** (2,792.600)	33,823.750*** (2,744.325)	32,127.030*** (2,778.322)
numero_asala	−61,576.730*** (3,235.964)	−61,433.190*** (3,228.285)	−57,211.080*** (3,180.612)	−56,386.020*** (3,232.424)
numero_indep	−35,263.940*** (3,157.190)	−35,097.680*** (3,148.197)	−35,504.290*** (3,100.917)	−34,852.610*** (3,144.150)
numero_desocu	−92,276.990*** (4,398.202)	−90,882.000*** (4,386.890)	−86,362.010*** (4,306.781)	−83,655.270*** (4,357.654)
numero_infan	32,944.800*** (4,658.990)	32,821.880*** (4,645.322)	29,921.400*** (4,566.607)	32,020.110*** (4,622.356)
numero_adole	57,500.070*** (4,057.810)	57,553.790*** (4,047.439)	52,766.790*** (3,984.797)	51,769.110*** (4,036.651)
num_sin_edu	186,637.100*** (4,875.302)	184,472.000*** (4,860.718)	170,286.500*** (4,777.904)	164,059.700*** (4,836.708)
num_edu_sup	64,781.810*** (2,531.106)	64,699.460*** (2,524.989)	62,849.820*** (2,492.020)	61,173.380*** (2,530.760)
a_edu_hog	45,642.860*** (938.675)	45,265.000*** (936.247)	43,329.580*** (923.137)	42,827.450*** (936.809)
salud	−61,302.610*** (3,675.537)	−60,932.690*** (3,664.576)	−58,409.990*** (3,603.991)	−56,317.780*** (3,651.917)
cesantes1	1,233,367.000*** (49,137.830)	1,232,906.000*** (48,983.570)	1,251,241.000*** (48,093.840)	1,239,363.000*** (48,793.120)
Constant	−2,669,568.000*** (51,336.830)	−2,680,309.000*** (51,203.380)	−2,790,864.000*** (50,554.370)	−2,832,068.000*** (51,872.610)
Observations	375,246	371,914	350,106	328,872

Note: *p<0.1; **p<0.05; ***p<0.01

Cuadro 6
Resultados para el 2016

	Dependent variable:			
	Ingresos act	Ingresos 15 %	Ingresos 20 %	Ingresos 25 %
	(1)	(2)	(3)	(4)
edad	27,636.880*** (614.466)	27,512.160*** (612.121)	26,909.560*** (603.283)	26,801.730*** (618.629)
edad2	−95.233*** (6.500)	−96.954*** (6.478)	−93.936*** (6.387)	−92.025*** (6.553)
horas	6,784.171*** (138.953)	6,771.631*** (138.270)	6,665.514*** (135.981)	6,626.974*** (139.220)
a_edu	100,096.300*** (722.164)	98,432.130*** (718.985)	96,418.830*** (708.121)	96,122.980*** (725.600)
resto1	1,029.975 (7,605.145)	6,525.888 (7,709.450)	24,322.470*** (7,875.581)	39,833.760*** (8,465.066)
capitales1	35,484.440*** (6,637.663)	43,922.180*** (6,679.563)	64,689.860*** (6,714.398)	72,535.840*** (7,051.688)
O_capitales1	−59,012.140*** (4,884.313)	−42,367.590*** (4,902.823)	−6,539.933 (4,908.422)	12,087.210** (5,135.567)
sexo1	199,232.700*** (4,258.910)	197,490.700*** (4,239.089)	195,712.600*** (4,171.666)	195,359.800*** (4,274.456)
obrero1	444,314.500*** (24,519.470)	454,834.100*** (24,518.790)	459,043.500*** (24,202.750)	464,405.900*** (25,042.770)
emp_domes1	301,762.100*** (27,377.260)	309,353.800*** (27,342.430)	314,307.400*** (26,969.520)	317,754.000*** (27,841.240)
cuenta_pro1	84,348.500*** (24,208.410)	95,232.040*** (24,212.320)	101,583.300*** (23,898.530)	107,613.600*** (24,732.190)
patrono	1,011,099.000*** (27,087.650)	1,014,737.000*** (27,078.210)	1,007,759.000*** (26,723.220)	1,019,250.000*** (27,616.300)
jefe1	183,061.300*** (4,699.315)	181,134.000*** (4,679.418)	178,404.700*** (4,605.945)	176,599.500*** (4,719.483)
numero_per	38,894.010*** (2,915.297)	38,116.440*** (2,901.346)	37,911.840*** (2,851.357)	37,607.100*** (2,919.259)
numero_asala	−53,557.010*** (3,372.942)	−53,613.710*** (3,359.930)	−54,219.540*** (3,312.689)	−54,175.460*** (3,400.329)
numero_indep	−37,185.600*** (3,316.954)	−36,564.490*** (3,304.719)	−38,138.860*** (3,255.838)	−37,510.790*** (3,338.690)
numero_desocu	−94,415.630*** (4,560.908)	−92,164.100*** (4,530.240)	−89,468.830*** (4,447.515)	−87,491.590*** (4,539.664)
numero_infan	23,819.670*** (4,802.188)	24,527.870*** (4,779.779)	22,166.850*** (4,698.468)	22,615.920*** (4,811.310)
numero_adole	61,087.970*** (4,199.281)	60,100.290*** (4,181.149)	56,335.990*** (4,119.794)	55,479.270*** (4,225.211)
num_sin_edu	176,341.900*** (5,064.953)	170,873.500*** (5,048.652)	163,675.400*** (4,983.279)	159,295.700*** (5,119.793)
num_edu_sup	74,760.340*** (2,670.093)	75,151.860*** (2,661.224)	70,975.160*** (2,623.504)	67,250.590*** (2,692.454)
a_edu_hog	41,969.120*** (985.832)	40,821.290*** (981.781)	39,474.390*** (967.120)	39,009.100*** (991.745)
salud	−67,653.220*** (3,827.140)	−66,738.490*** (3,809.863)	−64,066.110*** (3,746.312)	−62,999.700*** (3,838.366)
cesantes1	1,237,515.000*** (49,314.310)	1,251,223.000*** (49,324.990)	1,242,169.000*** (48,686.570)	1,256,952.000*** (50,376.980)
Constant	−2,928,081.000*** (50,979.850)	−2,920,576.000*** (50,983.550)	−2,893,705.000*** (50,319.370)	−2,916,219.000*** (52,038.630)
Observations	386,574	373,670	351,784	330,324

Note: *p<0.1; **p<0.05; ***p<0.01

Cuadro 7
Resultados para el 2015

	Dependent variable:			
	Ingresos act	Ingresos 15 %	Ingresos 20 %	Ingresos 25 %
	(1)	(2)	(3)	(4)
edad	27,798.020*** (573.003)	27,658.880*** (570.670)	27,338.420*** (566.327)	27,479.400*** (581.606)
edad2	−107.191*** (6.098)	−109.253*** (6.076)	−108.451*** (6.034)	−110.043*** (6.199)
horas	6,637.776*** (123.740)	6,625.334*** (122.996)	6,597.368*** (121.760)	6,591.748*** (124.767)
a_edu	94,408.070*** (668.461)	92,499.480*** (665.554)	90,496.050*** (660.163)	90,430.140*** (677.171)
resto1	−21,353.140*** (7,024.052)	−14,405.550** (7,128.707)	−604.428 (7,291.067)	9,925.976 (7,788.990)
capitales1	23,432.020*** (6,205.126)	39,226.080*** (6,276.664)	57,310.510*** (6,377.690)	56,862.600*** (6,732.577)
O_capitales1	−49,484.440*** (4,502.010)	−29,644.120*** (4,512.295)	−4,022.076 (4,522.359)	4,432.204 (4,702.867)
sexo1	200,611.900*** (3,968.128)	197,894.200*** (3,949.258)	192,879.300*** (3,911.704)	192,948.700*** (4,011.528)
obrero1	410,198.500*** (21,342.030)	412,356.500*** (21,269.990)	421,506.000*** (21,172.550)	426,160.300*** (21,786.730)
emp_domes1	281,766.000*** (24,034.790)	280,351.700*** (23,935.190)	287,877.300*** (23,794.440)	293,481.900*** (24,465.600)
cuenta_pro1	73,680.680*** (21,029.250)	75,849.970*** (20,957.710)	85,055.210*** (20,862.360)	90,616.070*** (21,469.440)
patrono	997,337.300*** (23,740.910)	986,809.500*** (23,666.220)	988,602.200*** (23,564.520)	988,512.900*** (24,267.510)
jefe1	173,241.400*** (4,377.108)	170,299.200*** (4,356.553)	164,621.200*** (4,318.788)	163,733.700*** (4,430.380)
numero_per	38,047.720*** (2,674.742)	37,934.500*** (2,659.688)	37,141.540*** (2,632.770)	36,910.780*** (2,702.215)
numero_asala	−39,527.120*** (3,129.933)	−41,134.180*** (3,120.243)	−43,035.460*** (3,099.696)	−43,241.730*** (3,186.248)
numero_indep	−23,473.510*** (3,059.912)	−24,733.280*** (3,049.408)	−25,489.390*** (3,026.827)	−25,487.890*** (3,108.717)
numero_desocu	−77,183.000*** (4,285.840)	−75,845.200*** (4,260.496)	−73,490.060*** (4,215.177)	−71,627.640*** (4,314.818)
numero_infan	28,339.810*** (4,400.011)	27,454.790*** (4,379.829)	27,985.920*** (4,342.123)	28,470.480*** (4,458.912)
numero_adole	63,011.600*** (3,848.333)	61,135.270*** (3,829.808)	58,346.520*** (3,798.378)	59,278.970*** (3,899.824)
num_sin_edu	165,675.900*** (4,721.832)	160,174.000*** (4,704.428)	153,169.300*** (4,667.598)	151,285.800*** (4,801.521)
num_edu_sup	77,420.910*** (2,491.157)	75,978.610*** (2,483.992)	73,940.680*** (2,469.598)	71,461.280*** (2,539.918)
a_edu_hog	39,031.080*** (915.525)	37,751.580*** (912.035)	36,530.160*** (904.755)	36,161.080*** (928.418)
salud	−69,049.170*** (3,515.834)	−67,905.690*** (3,498.145)	−65,693.890*** (3,467.272)	−65,047.260*** (3,559.621)
cesantes1	1,221,306.000*** (42,928.960)	1,211,607.000*** (42,788.350)	1,220,494.000*** (42,603.440)	1,223,973.000*** (43,855.830)
Constant	−2,844,511.000*** (44,561.300)	−2,810,107.000*** (44,413.310)	−2,795,624.000*** (44,209.950)	−2,805,048.000*** (45,509.610)
Observations	394,437	378,211	356,187	334,824

Note: *p<0.1; **p<0.05; ***p<0.01

Cuadro 8
Resultados para el 2014

	Dependent variable:			
	Ingresos act	Ingresos 15 %	Ingresos 20 %	Ingresos 25 %
	(1)	(2)	(3)	(4)
edad	27,504.210*** (562.586)	27,267.050*** (565.096)	27,248.680*** (571.193)	27,020.590*** (585.138)
edad2	−111.504*** (6.004)	−111.050*** (6.034)	−111.122*** (6.102)	−108.700*** (6.255)
horas	5,862.255*** (119.681)	5,873.977*** (119.925)	5,893.048*** (120.718)	5,898.758*** (123.180)
a_edu	92,591.620*** (652.034)	91,395.200*** (654.975)	90,726.460*** (661.351)	90,613.990*** (676.608)
resto1	−18,755.370*** (6,877.347)	−12,310.190* (7,030.235)	−2,779.572 (7,304.843)	5,417.616 (7,756.637)
capitales1	14,423.700** (6,121.690)	29,057.080*** (6,252.648)	51,397.850*** (6,469.332)	66,314.690*** (6,793.360)
O_capitales1	−33,233.150*** (4,405.428)	−17,312.890*** (4,457.753)	−2,022.624 (4,554.504)	9,948.473** (4,734.608)
sexo1	193,717.500*** (3,889.228)	191,314.500*** (3,905.700)	189,719.100*** (3,941.012)	188,819.900*** (4,029.985)
obrero1	395,454.500*** (20,968.280)	399,104.500*** (21,048.940)	399,727.600*** (21,331.510)	403,946.200*** (21,958.580)
emp_domes1	259,781.300*** (23,520.670)	262,765.300*** (23,609.070)	261,451.000*** (23,898.660)	266,050.100*** (24,565.940)
cuenta_pro1	82,476.820*** (20,667.580)	87,881.990*** (20,747.720)	89,826.420*** (21,026.630)	94,620.780*** (21,650.380)
patrono	1,087,814.000*** (23,292.690)	1,089,349.000*** (23,387.730)	1,084,984.000*** (23,684.370)	1,103,730.000*** (24,352.430)
jefe1	179,218.200*** (4,301.136)	178,021.800*** (4,320.160)	176,635.600*** (4,363.423)	176,271.000*** (4,464.498)
numero_per	38,365.480*** (2,553.884)	38,002.710*** (2,560.701)	38,505.930*** (2,580.899)	39,353.220*** (2,638.298)
numero_asala	−36,677.050*** (3,065.552)	−36,099.850*** (3,081.515)	−36,235.240*** (3,114.320)	−37,372.900*** (3,191.084)
numero_indep	−29,440.920*** (2,996.720)	−29,198.810*** (3,011.254)	−29,132.060*** (3,042.818)	−29,376.430*** (3,116.163)
numero_desocu	−74,276.930*** (4,099.702)	−72,636.610*** (4,108.228)	−70,648.860*** (4,135.783)	−70,343.290*** (4,215.653)
numero_infan	18,659.810*** (4,205.633)	18,590.590*** (4,219.794)	17,627.450*** (4,259.700)	17,902.120*** (4,356.913)
numero_adole	59,893.100*** (3,735.332)	58,942.100*** (3,749.350)	59,102.990*** (3,786.625)	59,929.270*** (3,876.977)
num_sin_edu	148,672.600*** (4,559.390)	142,980.100*** (4,587.053)	138,789.400*** (4,636.896)	135,463.100*** (4,749.660)
num_edu_sup	77,696.360*** (2,433.670)	77,632.120*** (2,449.070)	77,761.590*** (2,477.599)	76,798.810*** (2,538.380)
a_edu_hog	33,629.330*** (898.733)	32,161.390*** (903.460)	30,966.800*** (912.408)	30,024.350*** (933.604)
salud	−66,618.660*** (3,371.793)	−65,376.570*** (3,383.510)	−65,555.430*** (3,415.190)	−65,759.510*** (3,494.657)
cesantes1	1,298,374.000*** (42,196.620)	1,305,411.000*** (42,363.500)	1,300,195.000*** (42,926.560)	1,321,549.000*** (44,183.040)
Constant	−2,847,895.000*** (43,790.510)	−2,837,742.000*** (43,962.060)	−2,831,514.000*** (44,525.960)	−2,851,804.000*** (45,805.600)
Observations	387,520	371,738	350,684	330,634

Note: *p<0.1; **p<0.05; ***p<0.01

Cuadro 9
Resultados para el 2013

	Dependent variable:			
	Ingresos act	Ingresos 15 %	Ingresos 20 %	Ingresos 25 %
	(1)	(2)	(3)	(4)
edad	27,436.280*** (549.916)	26,806.000*** (540.141)	26,588.170*** (553.511)	26,801.420*** (565.934)
edad2	−117.138*** (5.886)	−115.245*** (5.785)	−113.020*** (5.931)	−114.765*** (6.065)
horas	5,441.610*** (110.721)	5,386.557*** (108.342)	5,398.669*** (110.658)	5,395.627*** (112.837)
a_edu	87,330.860*** (632.639)	83,744.690*** (621.406)	83,291.570*** (636.229)	83,023.140*** (649.593)
resto1	−31,479.900*** (6,731.500)	−14,419.860** (6,827.132)	−6,031.507 (7,220.990)	976.659 (7,688.509)
capitales1	−16,982.820*** (5,986.106)	5,279.709 (6,055.413)	19,962.330*** (6,356.131)	30,386.640*** (6,681.792)
O_capitales1	−48,167.310*** (4,307.999)	−7,758.709* (4,299.765)	4,060.193 (4,470.230)	14,863.080*** (4,651.797)
sexo1	185,928.000*** (3,803.920)	180,351.500*** (3,732.861)	179,648.300*** (3,820.354)	180,644.700*** (3,900.551)
obrero1	394,589.700*** (20,685.330)	404,604.500*** (20,337.580)	410,469.400*** (20,894.470)	415,854.800*** (21,426.680)
emp_domes1	248,486.600*** (23,046.470)	253,497.500*** (22,638.400)	256,651.500*** (23,227.010)	263,498.500*** (23,792.880)
cuenta_pro1	75,398.660*** (20,409.670)	89,295.230*** (20,064.660)	95,735.740*** (20,613.040)	100,965.300*** (21,140.460)
patrono	989,765.200*** (22,699.280)	972,352.500*** (22,350.760)	984,928.600*** (22,956.450)	995,817.900*** (23,534.140)
jefe1	171,788.900*** (4,222.609)	166,204.800*** (4,144.437)	166,085.500*** (4,244.396)	163,878.100*** (4,336.187)
numero_per	35,098.030*** (2,448.282)	34,196.170*** (2,397.362)	34,001.820*** (2,450.753)	33,673.270*** (2,500.824)
numero_asala	−35,106.060*** (2,962.649)	−36,335.350*** (2,914.279)	−36,929.140*** (2,986.793)	−39,297.890*** (3,055.513)
numero_indep	−27,290.900*** (2,906.634)	−28,874.260*** (2,856.299)	−29,224.330*** (2,927.536)	−30,307.080*** (2,994.119)
numero_desocu	−87,096.630*** (3,883.446)	−81,062.550*** (3,799.253)	−80,307.060*** (3,876.752)	−80,210.870*** (3,948.435)
numero_infan	13,365.930*** (4,032.185)	12,489.770*** (3,953.980)	13,272.200*** (4,046.330)	14,281.290*** (4,131.120)
numero_adole	60,380.790*** (3,581.411)	56,768.660*** (3,514.322)	57,228.640*** (3,597.602)	57,902.230*** (3,678.601)
num_sin_edu	140,792.200*** (4,414.080)	129,876.500*** (4,336.184)	127,800.500*** (4,447.767)	126,390.000*** (4,558.980)
num_edu_sup	78,814.880*** (2,362.540)	74,924.760*** (2,328.903)	73,090.010*** (2,387.459)	70,630.790*** (2,441.828)
a_edu_hog	33,830.960*** (875.506)	32,115.290*** (860.114)	31,631.140*** (881.121)	31,720.910*** (899.933)
salud	−59,402.840*** (3,245.588)	−56,967.560*** (3,183.887)	−56,325.550*** (3,258.131)	−55,256.180*** (3,329.108)
cesantes1	1,198,105.000*** (41,561.500)	1,186,847.000*** (40,881.020)	1,204,102.000*** (42,003.440)	1,216,435.000*** (43,075.810)
Constant	−2,683,107.000*** (43,057.460)	−2,633,221.000*** (42,355.310)	−2,649,267.000*** (43,512.730)	−2,674,578.000*** (44,617.440)
Observations	395,334	367,832	347,423	328,041

Note: *p<0.1; **p<0.05; ***p<0.01

Cuadro 10

Cantidad de personas pobres e incidencia de la pobreza para todos los escenarios de faltantes 2013-2017

		2013		2014		2015		2016		2017	
		Personas	**%**	**Personas**	**%**	**Personas**	**%**	**Personas**	**%**	**Personas**	**%**
Hot Deck	**Faltantes actuales**	13.994.014	30,6	13.209.722	28,5	13.038.620	27,8	13.267.677	28,0	12.883.106	26,9
	15 % Faltantes	16.322.806	35,7	15.460.996	33,4	15.328.457	32,7	15.692.479	33,1	15.354.103	32,1
	20 % Faltantes	16.363.621	35,7	15.472.017	33,4	15.285.864	32,6	15.696.272	33,2	15.303.169	32,0
	25 % Faltantes	16.386.200	35,8	15.405.628	33,3	15.247.193	32,6	15.699.991	33,2	15.340.159	32,0
Regresión	**Faltantes actuales**	16.071.599	35,1	15.260.472	33,0	15.132.872	32,3	15.352.196	32,4	14.988.991	31,3
	15 % Faltantes	15.856.025	34,6	15.153.362	32,7	14.998.058	32,0	15.269.349	32,3	14.958.820	31,3
	20 % Faltantes	15.580.653	34,0	14.813.464	32,0	14.753.911	31,5	15.117.026	31,9	14.842.154	31,0
	25 % Faltantes	15.367.278	33,6	14.758.268	31,9	14.717.423	31,4	15.001.521	31,7	14.758.812	30,8
Media condicionada	**Faltantes actuales**	15.949.371	34,8	15.156.583	32,7	15.015.682	32,1	15.248.118	32,2	14.825.338	31,0
	15 % Faltantes	15.536.323	33,9	14.862.233	32,1	14.726.529	31,5	14.962.522	31,6	14.790.832	30,9
	20 % Faltantes	15.191.194	33,2	14.537.127	31,4	14.352.245	30,7	14.548.350	30,7	14.384.643	30,1
	25 % Faltantes	14.907.238	32,6	14.185.633	30,6	13.974.031	29,8	14.132.965	29,9	14.029.149	29,3

Elaboración propia

Cuadro 11

Cantidad de personas pobres extremos e incidendia de la pobreza extrema para todos los escenarios de faltantes 2013-2017

		2013		2014		2015		2016		2017	
		Personas	**%**	**Personas**	**%**	**Personas**	**%**	**Personas**	**%**	**Personas**	**%**
Hot Deck	**Faltantes actuales**	4.149.000	9,1	3.741.870	8,1	3.718.168	7,9	4.003.480	8,5	3.534.386	7,4
	15 % Faltantes	5.679.235	12,4	5.199.082	11,2	5.185.166	11,1	5.519.212	11,7	4.954.763	10,4
	20 % Faltantes	5.703.753	12,5	5.208.482	11,3	5.186.788	11,1	5.583.869	11,8	4.966.899	10,4
	25 % Faltantes	5.728.329	12,5	5.202.682	11,2	5.189.343	11,1	5.622.498	11,9	4.970.094	10,4
Regresión	**Faltantes actuales**	5.374.420	11,7	4.893.202	10,6	4.871.090	10,4	5.138.436	10,9	4.628.476	9,7
	15 % Faltantes	5.060.391	11,1	4.742.246	10,2	4.710.421	10,1	4.969.899	10,5	4.590.027	9,6
	20 % Faltantes	4.853.243	10,6	4.470.822	9,7	4.458.160	9,5	4.712.153	10,0	4.368.164	9,1
	25 % Faltantes	4.632.676	10,1	4.254.165	9,2	4.224.679	9,0	4.499.004	9,5	4.143.625	8,7
Media condicionada	**Faltantes actuales**	5.797.109	12,7	5.414.464	11,7	5.365.443	11,5	5.728.664	12,1	5.208.757	10,9
	15 % Faltantes	6.075.370	13,3	5.627.549	12,2	5.530.786	11,8	5.842.223	12,3	5.240.653	10,9
	20 % Faltantes	6.305.394	13,8	5.797.666	12,5	5.770.601	12,3	6.101.826	12,9	5.508.689	11,5
	25 % Faltantes	6.525.893	14,3	6.111.032	13,2	6.043.705	12,9	6.456.359	13,6	5.767.887	12,0

Elaboración propia