


Departamento Administrativo Nacional de Estadística



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

**Censo de Edificaciones
Metodología de Imputación**

Julio 2007

	CENSO DE EDIFICACIONES METODOLOGIA DE IMPUTACIÓN		CÓDIGO: ME-CEED-MET-01 VERSIÓN: 02 Página 1 Fecha: 09-07-07
ELABORÓ: Metodologo	REVISÓ: Coordinador Metodología Estadística	APROBÓ: Director de Metodología y Producción Estadística	

CONTENIDO

	Pág.
INTRODUCCION	5
1. OBJETIVOS	7
1.1 OBJETIVO GENERAL	7
1.2 OBJETIVOS ESPECÍFICOS	7
2. LINEAMIENTOS PARA EL DESARROLLO DEL TRABAJO	8
3. PLANTEAMIENTO DEL PROBLEMA	10
3.1 RECOPIACIÓN DE INFORMACIÓN DISPONIBLE	10
3.2 MARCO TEÓRICO	11
3.2.1 Crítica	11
3.2.1.1 Reglas de crítica	12
3.2.1.2 Crítica estadística	12
3.2.1.3 Crítica automática	13
3.2.2 Imputación	15
3.2.2.1 Imputación determinística	15
3.2.2.2 Imputación por substitución	15
3.2.2.3 Imputación por estimadores	16
3.2.2.4 Imputación simple	16
3.2.2.5 En frío o de expertos	17
3.2.2.6 Paquete caliente	17
3.2.2.6.1 Imputación secuencial	18
3.2.2.6.2 Selección aleatoria	18
3.2.2.6.3 El vecino más próximo	18
3.2.2.7 Imputación múltiple	18
4. ANTECEDENTES DEL CENSO	20
5. ASPECTOS GENERALES DE CRÍTICA EN EL CENSO DE EDIFICACIONES	22
5.1 PRIMER NIVEL DE CRÍTICA	22
5.2 SEGUNDO Y TECER NIVEL DE CRÍTICA	24
5.3 CUARTO NIVEL DE CRÍTICA	25
5.3 METODOLOGIA DE LA EXPERTA	26

5.3.1 Primera fase validación	26
5.3.2 Segunda fase imputación	29
6. REGLAS Y ESPECIFICACIONES	31
6.1 ESPECIFICACIONES DE VALIDACIÓN	31
6.2 ESPECIFICACIONES DE CONSISTENCIA	31
6.3 REGLAS DE IMPUTACIÓN	32
6.3.1 Formato uno	33
6.3.2 Formato dos	34
7. CONCLUSIONES	51
8. BIBLIOGRAFÍA	54

LISTA DE ANEXOS

	Pág.
Anexo A. Diseño de registro tabla Procesos. Excel	33
Anexo B. Algoritmo de imputación	39
Anexo C. Programa de imputación en SAS	44

LISTA DE CUADROS

	Pág.
Cuadro 1. Distribución de registros y porcentaje, según archivo prueba en la Investigación Censo de Edificaciones	53
Cuadro 2. Cantidad de datos a imputar, para la variable precio de venta del Metro ² , en la investigación Censo de Edificaciones, por destino según Área Metropolitana	53

INTRODUCCION

En un esfuerzo permanente por presentar al país cifras oportunas, confiables y de buena calidad, y en cumplimiento de su misión institucional, el DANE a través de la investigación Censo de Edificaciones determina el estado actual de la actividad edificadora, aplicando una metodología fundamentada en el análisis integrado de la información recopilada de fuentes directas.

Actualmente esta investigación requiere precisar técnicas y procedimientos que permitan la integridad total de los datos con el fin de asegurar la Calidad Estadística de la información y satisfacer las necesidades de los clientes o usuarios que día a día exigen que este producto intangible cumpla con todos los criterios de garantía en cuanto a utilidad, oportunidad y confiabilidad de la información.

La investigación en su interés permanente por el mejoramiento de los procesos, la optimización de la calidad para satisfacer y cumplir con los requerimientos exigidos, ha venido utilizando en el último filtro de calidad el conocimiento de expertos o especialistas en el tema, quienes generalmente realizan sus actividades en forma manual, detectando en el análisis del microdato situaciones atípicas o fuera de rango que según su conocimiento deben ser ajustados, todo ello además de hacerse de manera *artesanal y subjetiva* ya que cada caso es tratado de manera particular ha generado el no contar con procesos sistemáticos de las reglas que se aplican para dichos ajustes.

Por lo anterior, es indispensable conocer el procedimiento completo que los expertos temáticos realizan y con esa información diseñar y homogenizar en forma técnica y sistemática los criterios, reglas y especificaciones para el ajuste de la información según la naturaleza, volumen, cantidad e importancia del dato.

El presente documento consta de ocho partes: en la primera se definen los objetivos del estudio, en la segunda se proponen los lineamientos para el desarrollo del trabajo, en la tercera se hace el planteamiento del problema, en la cuarta se enuncia la metodología del Censo además se incluye el marco conceptual, en la quinta se describe la metodología aplicada por la experta en el último filtro de calidad, en la sexta se elabora con la temática las reglas y especificaciones para el ajuste de la variable precio de venta por metro², en la séptima se establecen las conclusiones y en la octava se presenta la bibliografía.



CENSO DE EDIFICACIONES METODOLOGIA DE IMPUTACIÓN

CÓDIGO: ME-CEED-MET-01

VERSIÓN: 02

Página 6

Fecha: 09-07-07

Doy un agradecimiento especial a la doctora Angela María Giraldo, por la información, sugerencias y recomendaciones, que me ayudaron a llevar a feliz término este trabajo, lo mismo que al Doctor Hernando Chacón, por su confianza, apoyo y comentarios acertados.

1. OBJETIVOS

1.1 OBJETIVO GENERAL

Conocer el proceso completo de ajuste final a la información del Censo de Edificaciones realizado por los expertos temáticos, para diseñar, definir y documentar las reglas y especificaciones de consistencia, validación e imputación que brinden de manera técnica y sistemática mayor celeridad en el proceso y que permita la comprensión del mismo en cuanto a contenido, aspectos metodológico e informático.

1.2 OBJETIVOS ESPECÍFICOS

- Conocer y documentar las estrategias de los expertos temáticos para identificar o detectar datos atípicos o inconsistentes según variable o campo.
- Conocer y documentar los procedimientos y técnicas utilizadas por los expertos temáticos para la solución definitiva de los datos atípicos según variable.
- Elaborar una metodología que permita el cálculo y detección oportuno de los datos atípicos, utilizando herramientas técnicas combinadas con el manejo de medios computacionales modernos.
- Elaborar las especificaciones de validación y consistencia de las variables de relevancia para la investigación.
- Diseñar el algoritmo imputacional.
- Construir de acuerdo a los ítems anteriores, el proceso informático de imputación para algunas variables de la investigación.

2. LINEAMIENTOS PARA EL DESARROLLO DEL TRABAJO

- 2.1 Se inicia el trabajo con el contenido de los instrumentos de recolección del Censo de Edificaciones (formulario de obras nuevas y planilla de obras paralizadas nuevas, listado de seguimiento y control de obras en proceso, listado de seguimiento y control de obras paralizadas, listado de seguimiento y control de obras culminadas).
- 2.2 En una primera fase de crítica, la cual se debe llevar a cabo con el temático para hacer una revisión textual de los procedimientos metodológicos que utiliza la investigación, en cuanto a crítica manual y automática, de esta manera establecer la coherencia lógica y los criterios de aceptación de las variables de relevancia para el estudio.
- 2.3 En una segunda fase de crítica, la temática describirá los criterios que emplea para detectar e imputar manualmente la información. En esta fase se ira documentando la metodología de aquellas características que se hacen subjetivamente; donde se aplican procedimientos con gran cantidad de tareas mecánicas y repetitivas, que los hacen más susceptibles de errores cuando son desarrollados manualmente.
- 2.4 En esta misma fase se realizará la asignación de rangos, utilizando ante todo el criterio del temático y de acuerdo con las condiciones (relaciones lógicas) que se establezcan.
- 2.5 En una tercera fase con el experto, se analizará en la base de datos que contiene el histórico de la investigación, las diferentes posibilidades de relación lógica entre variables, valores aceptables por variable, este análisis se hará en relación principalmente a los ítems relevantes para la investigación y aquellas que sirven de insumo al modelo de construcción para el cálculo de los indicadores.
- 2.6 A partir de esas charlas se establece el diseño para las especificaciones de validación, consistencia e imputación con el fin de elaborarlas en los formatos estándar de las Técnicas de Diseño y Desarrollo de Encuestas TDDE, éstas reglas serán las especificaciones para el programa de captura inteligente.

- 2.7 Elaborar las especificaciones de consistencia, de acuerdo con las especificaciones de validación y la estructura del formulario (relaciones lógicas entre variables). Con base en éstas, sistemas generará el programa de consistencia del conjunto de datos o base de datos resultante del proceso de grabación.
- 2.8 Elaborar las especificaciones de imputación, de acuerdo con los criterios de validación y consistencia definidos con el temático experto en la investigación. Sistemas implementará el programa de imputación, con el fin de obtener un conjunto de datos depurado
- 2.9 Para la implementación del programa de imputación, es necesario considerar una base de datos obtenida del registro histórico, en donde los datos presenten valores atípicos para las variables cuantitativas de interés en la investigación. Se corre el programa con el archivo de prueba y se establecen los resultados esperados, para realizar los comparativos correspondientes.
- 2.10 Se analizan los resultados esperados, con los obtenidos en la base histórica, con el fin de establecer y constatar que el programa de imputación asigna los valores “óptimos” en los campos previstos.
- 2.11 Las especificaciones deben ser claras, sencillas y directas, para una fácil comunicación entre los actores involucrados.
- 2.12 Se propone elaborar un documento (primera fase) de indicadores de calidad de los datos, donde se incluya: cantidad de datos ajustados mediante el programa de imputación, desagregados por área metropolitana y destino; porcentaje de imputación total.
- 2.13 Los indicadores de calidad que se generen deben ser documentados junto con los resultados obtenidos, para su posterior análisis y conocimiento de los usuarios de la información

3. PLANTEAMIENTO DEL PROBLEMA

Uno de los principales problemas en todo tipo de encuestas con fines estadísticos es la presencia de datos faltantes o datos atípicos para una o más de las unidades observadas, lo cual dificulta la realización de cualquier análisis. Como caso particular están las investigaciones desarrolladas por el Departamento Administrativo Nacional de Estadísticas DANE.

Durante el diseño y el desarrollo de cada una de las encuestas, se hacen muchos esfuerzos con el objeto de llegar a la estimación final con datos completos, correctos y consistentes; sin embargo, entre otras razones como la deficiente cultura estadística en nuestro país, no todas las preguntas obtienen respuesta, no todas las respuestas son correctas y se presentan casos de inconsistencias o datos atípicos.

Generalmente las investigaciones presentan reglas y especificaciones de crítica definidas, las cuales confrontan y validan la información desde la recolección hasta la grabación, actualmente de forma particular la investigación Censo de Edificaciones, necesita del desarrollo teórico e implementación estándar y sistematizada de los procesos de crítica e imputación utilizados en el último filtro de calidad por los expertos del tema.

La investigación cuenta en la actualidad con documentación básica como: "Las Especificaciones de Sistemas, "La Memoria Institucional" y "El Manual Básico de Recolección, Crítica y Codificación"; por otra parte, se tiene documentación elaborada por los pasantes estadísticos, sobre métodos de imputación aplicados a otras investigaciones.

El material bibliográfico relacionado anteriormente, sirve como soporte para el marco teórico de este trabajo.

Para el desarrollo de los objetivos propuestos, se contemplan los siguientes pasos:

3.1 RECOPIACIÓN DE INFORMACIÓN DISPONIBLE

Esta actividad incluye la lectura de manuales, documentos existentes en la investigación, bibliografía sobre crítica y métodos de imputación, así como las Técnicas de Diseño y Desarrollo de Encuestas, TDDE, Operativo de Oficina.

3.2 MARCO TEÓRICO

Los procesos de crítica e imputación surgen de la necesidad de garantizar calidad en la información, tienen por objeto determinar el grado de integridad de los datos en referencia a tres aspectos:

➤ *Compleitud*

Chequeo del cuestionario para observar que todas las variables obtuvieron respuesta, sin embargo, como última opción, que por lo menos las variables de requerimiento mínimo estén diligenciadas.

➤ *Corrección*

Chequeo de que la información plasmada en el cuestionario sea verdadera. La información es la correcta.

➤ *Consistencia interna*

Chequeo que se efectúa para constatar que la información del cuestionario es lógica, es decir, al comparar dos o más variables del cuestionario diligenciado, sus respuestas son coherentes.

3.2.1 Crítica

Es el proceso de identificar valores erróneos, faltantes o atípicos en una investigación a través de reglas establecidas. El proceso de crítica puede ser manual o automático, dependiendo del volumen de crítica en términos de datos a criticar y de la cantidad de reglas a verificar.

3.2.1.1 Reglas de crítica

La definición de las reglas se basa en el conocimiento del experto o de los especialistas en el tema de estudio, en la estructura del cuestionario y en otros datos relacionados con el tema de la encuesta. Las reglas de crítica se refieren generalmente a: los valores que en forma individual pueden asumir las variables, las relaciones que deben existir entre unas respuestas y otras y la estructura global que se espera del conjunto de datos.

Un principio fundamental en el proceso de crítica es que durante las diferentes etapas de crítica, una regla de crítica no puede contradecir reglas de crítica de otras etapas.

El diseño de una estrategia de crítica conduce a identificar registros que no se ajustan a las reglas establecidas, para que se siga el proceso de corrección correspondiente.

3.2.1.2 Crítica estadística

Es el proceso encaminado a determinar el grado de integridad de la información en referencia a completitud, corrección y consistencia interna a través de la aplicación de reglas al conjunto de datos.

Si un registro no cumple con alguna de las reglas establecidas debe ser señalado para ser corregido o cambiado. En esta crítica se estudia, analiza y valida el conjunto completo de datos.

Para estudiar el grado de integridad de los datos se plantean dos tipos de crítica:

➤ Crítica Interna

Proceso que determina las inconsistencias y/o faltantes de cada formulario y la corrección de registros individuales hasta asegurar integridad, esta crítica se aplica en la recolección de la información y en oficina hasta la captura.

Durante la recolección el encuestador hace el primer filtro de calidad de su trabajo, cuando revisa y corrige a medida que realiza la entrevista, en este caso el entrevistador se basa en reglas preestablecidas, en su sentido común y su juicio acerca de lo que son respuestas o combinaciones de respuestas aceptables.

Una vez termina la entrevista, en forma inmediata el encuestador tiene el compromiso de volver a revisar las respuestas registradas (el cuestionario diligenciado), con el fin de hacer las correcciones sin mayor contratiempo ni costo adicional.

Un segundo filtro de calidad lo realiza el supervisor, quien además de controlar la cobertura revisa en forma general el cuestionario diligenciado, con el fin de realizar las correcciones pertinentes en terreno, prestando mayor atención a las preguntas básicas del estudio (requerimiento mínimo) y verificando la consistencia de las respuestas (control de calidad de las respuestas).

Un tercer filtro de calidad se realiza en oficina, cuando el supervisor entrega su carga diaria de trabajo, allí pasa primero por la etapa de alistamiento previo al operativo de oficina. El alistamiento consiste entre otras cosas en la codificación (si la hay), la identificación con etiquetas, la construcción de paquetes o agrupamiento, el registro de formularios completos e incompletos.

Un segundo paso corresponde al proceso adicional de crítica, donde se verifican flujos lógicos, identificadores, códigos, criterios mínimos para considerar un formulario completo, etc. Y se procede a corregir en oficina o en terreno.

Un cuarto filtro de calidad se produce en la etapa de captura, la crítica más compleja y generalmente se produce durante la etapa de procesamiento de datos (crítica automática), es decir, cuando la información ha pasado de medio físico (cuestionario) a medio magnético, generalmente se aplican programas de computador diseñados para verificar las reglas de crítica (captura inteligente), los cuales verifican rangos establecidos para variables cuantitativas del cuestionario, sintaxis y códigos para variables alfanuméricas.

➤ Macro crítica

Proceso que consiste en cotejar los datos de la encuesta con otros conjuntos de datos estadísticos sobre temáticas comunes, datos macroeconómicos, estudios internacionales, registros administrativos o encuestas anteriores a fin de asegurar su coherencia.

3.2.1.3 Crítica automática

Es un proceso donde se aplica y verifica el cumplimiento de las reglas de crítica que no fueron verificadas anteriormente, así como el uso de reglas muy complejas.

Las razones para realizar la crítica automática son: corregir los datos faltantes, detectar los datos erróneos, encontrar inconsistencias y señalar los atípicos.

➤ Verificación

Es el proceso que generalmente acompaña a la grabación de los datos en un medio magnético y que consiste en garantizar que lo grabado es un fiel reflejo de lo contenido en los cuestionarios.

Se realiza generalmente a través de una segunda grabación, en que cada uno de los datos capturados va siendo comparado con la anterior captura. Puede realizarse a todo el cuestionario, o solo a las variables más importantes para la investigación.

➤ Validación

Es el proceso por el cual en forma manual o automática se controla una variable y solo una en cada momento, para determinar si los datos cumplen reglas de crítica preestablecidas de aceptabilidad o los posibles valores o rangos de respuesta válida para cada campo numérico; verifica la sintaxis numérica o alfanumérica de una determinada respuesta.

La verificación de las reglas de crítica, la puede realizar el digitador cuando al momento de grabar determina si la variable cumple o no la regla establecida, el programa, cuando se digita la información permite indexar los valores aceptables o respuestas válidas, en caso contrario el sistema emite señales de alerta como filtro de calidad, etc.

A la crítica de validación le corresponde siempre una acción en caso de encontrar un dato inválido, los cuales pueden ser:

La corrección inmediata, la imputación determinística, la señalización con banderas o códigos especiales para imputación probabilística o simplemente deja los espacios en blanco.

➤ Consistencia

También conocida con el nombre “Tablas de decisión lógica”, es el proceso que determina incoherencias a partir de la información de todo el formulario y se refiere a la preservación de las relaciones entre variables, las cuales pueden ser:

De seguimiento al flujo establecido en el formulario; de relación lógica entre dos o más respuestas de un mismo formulario.

Las especificaciones de consistencia, utilizan las condicionales "Si", "entonces" con la acción respectiva, dentro del proceso de captura.

La operación de crítica de consistencia es la identificación de datos que no cumplen las reglas y los señala para ser reemplazados por datos de registros similares, termina en retirar datos de la base, realizar una imputación determinística o marcar registros a fin de realizar la imputación probabilística en el siguiente paso.

3.2.2 Imputación

Es el proceso de corregir o asignar un valor "plausible" que cumpla con las reglas de crítica a fin de reemplazar datos inconsistentes, faltantes, erróneos y atípicos.

Si la opción ha sido recurrir a algún método de imputación, se debe tener cuidado en escoger el adecuado para ser aplicado según tipo de investigación y clase de variable.

3.2.2.1 Imputación determinística

Se utiliza cuando un dato señalado para imputar, recibe un único valor posible, que se le "puede" asignar de tal forma que cumpla la(s) regla(s) de crítica planteadas.

Este tipo de imputación se puede realizar únicamente con información del mismo registro, por lo tanto al escribir las especificaciones de crítica, de una vez se asume la acción de reemplazar el dato por el valor "correcto".

3.2.2.2 Imputación por substitución

Se utiliza especialmente para encuestas continuas, cuando se tiene una base de datos auxiliar o algún tipo de archivo administrativo, el cual contiene información referente a las mismas variables del mismo individuo en un tiempo anterior.

Es indispensable tener cuidado con la definición de la variable a imputar, ya que para diferentes estudios las definiciones de las variables pueden no ser las mismas.

3.2.2.3 Imputación por estimadores

Consiste en el tratamiento de los datos, de la misma encuesta o de una encuesta auxiliar, cuyas variables son las mismas o altamente correlacionadas con la encuesta a tratar, como por ejemplo, registros históricos de la misma encuesta o estudios que traten el mismo tema.

Este tratamiento puede ser realizado de muchas maneras, cuando se trabaja con los datos de la misma encuesta, se deben utilizar únicamente los datos correctos, esto implica procesamiento de la información, antes de realizar la imputación.

Se procede a realizar la estimación de un parámetro, como el valor esperado; (actual o histórico) de la variable, la moda o la razón. La imputación también se puede desarrollar a partir de la tendencia, en este caso, se hace necesario contar con el registro histórico, que permita a partir de todos los datos disponibles de la encuesta, establecer el comportamiento del dato actual, en comparación con los meses anteriores, para imputar la variable con dato faltante, erróneo o atípico.

Cuando se emplea la información auxiliar, se recurre a métodos como la estimación por regresión.

En la imputación por estimación, se recomienda establecer de nuevo, procedimientos de crítica después de realizada la imputación, porque pueden persistir situaciones de inconsistencia.

3.2.2.4 Imputación simple

Con el objeto de imputar un valor faltante, inconsistente o atípico, se reemplaza en la variable el valor que presenta dato faltante o inconsistente, por el valor de la media aritmética, solo es aplicable para variables cuantitativas.

Este método se puede hacer de dos formas: La primera consiste en calcular la media aritmética actual o histórica de la variable a imputar, con los valores observados que estén en el intervalo de aceptación para esta variable y reemplazar faltantes, inconsistentes o atípicos con el promedio obtenido. La segunda consiste en particionar la población de acuerdo a una variable de estratificación, calcular la media aritmética para cada estrato y reemplazar los valores a imputar en cada estrato por este valor.

Para variables cualitativas o cuantitativas discretas, se reemplaza el dato faltante o atípico por la moda.

Este método es de fácil aplicación, ya que se usa estadística descriptiva. El dato faltante es reemplazado por un valor esperado lo que implica un alto grado de estabilidad.

Su desventaja radica en que subestima la varianza, lo que conduce a alteraciones en la matriz de correlaciones.

3.2.2.5 En frío o de expertos

Consiste en usar valores o reglas predeterminadas cuando se quiere imputar el valor de una variable, el dato se obtiene de una encuesta anterior, de un experto en la temática o de un dato histórico, y se trata de encontrar un valor adecuado según el comportamiento de la variable.

Con este tipo de imputación no se tienen en cuenta las distribuciones y las relaciones entre variables, la magnitud del sesgo es incierta.

Se puede aplicar cuando los datos a imputar son pocos y otros procedimientos de imputación no arrojan resultados satisfactorios.

3.2.2.6 Paquete caliente

Consiste en sustituir el dato a imputar, por uno perteneciente a otro registro de la misma encuesta. Para este fin se divide la base de datos en dos grupos, los receptores y los donantes. En el grupo de los receptores se incluyen los registros con al menos un dato faltante, inconsistente o atípico y el grupo de los donantes, contiene los registros con todos sus datos completos y "limpios".

La idea es, que para cada registro en el grupo de receptores se encuentre un donante muy similar en el grupo de donantes para que el dato faltante, inconsistente o atípico, sea reemplazado por el mismo valor que tiene el donante para esa variable.

Esta similitud se establece mediante la determinación de las variables más correlacionadas con la variable a imputar, se toman los datos del receptor para esas variables y se busca un donante con esas mismas características. Para los casos en que no se encuentre un donante idóneo para algún receptor, se debe especificar algún otro método de imputación que será incluido dentro del paquete caliente.

Sin embargo, existen determinados métodos de imputación, que aplican únicamente según el tipo de variable, para datos de tipo categórico se utiliza un método

secuencial o de selección aleatoria y para datos de tipo cualitativo, el método más utilizado es el del vecino más próximo.

3.2.2.6.1 Imputación secuencial

Se crea una matriz de imputación en la cual se diseñan todas las posibles combinaciones de las variables de similitud. A medida que se van leyendo los datos de la encuesta, la matriz de imputación se actualiza y cuando encuentra un dato a ser imputado toma el correspondiente de esta matriz de imputación. El proceso continúa hasta que cada dato faltante, inconsistente o atípico es imputado.

El receptor tiene el mismo valor que el donante para todos los campos de similitud e imputación.

3.2.2.6.2 Selección aleatoria

Consiste en crear un subgrupo de donantes, los cuales tienen los mismos valores en las variables de similitud, y seleccionar entre ellos uno al azar como donante.

3.2.2.6.3 El vecino más próximo

Se fundamenta en el método de selección aleatoria, ya que el principio básico es construir un subgrupo de donantes. La diferencia radica en que no se busca el dato que tiene exactamente el mismo valor, sino el que más se le aproxima y que además pertenece a una clase similar en otra variable.

3.2.2.7 Imputación múltiple

Consiste en reemplazar cada dato faltante con dos o más valores representativos de la distribución de posibles valores, creando dos o más conjuntos de datos completos, y así, con cada uno de ellos es posible realizar análisis.

Droesbeke y Lavallée (1996), proponen la siguiente clasificación de la no - respuesta basada en la definición del conjunto de respuesta dada por Särndal et. al. (1992):

No respuesta total y parcial: la no respuesta total, ocurre cuando no se ha podido medir ninguna de las variables de interés. La no respuesta parcial se tiene cuando sólo se ha podido medir un subgrupo de variables de interés.

No respuesta ignorable y no ignorable: la no respuesta es ignorable cuando la probabilidad de responder a una pregunta dada no depende del valor de la variable medida, mientras que la no respuesta no ignorable si depende de este valor.

Si éste método de imputación es aplicado bajo un modelo de no respuesta aleatoria e ignorable, la variabilidad debida a los datos faltantes puede ser estimada a partir de la combinación de inferencias hechas con cada conjunto de datos completo generado bajo este esquema; garantizando un alto grado de validez en las estimaciones

Rubin y Schenker (1986) presentan la siguiente justificación teórica para la imputación múltiple:

Supóngase que el objetivo es obtener inferencias respecto a una cantidad poblacional con base en una muestra en presencia de datos faltantes, y que la distribución posterior de los datos faltantes (por ejemplo la distribución condicional dados los datos observados) es calculada. Entonces la distribución posterior de la cantidad poblacional es obtenida utilizando la distribución posterior de los datos obtenidos (por ejemplo la distribución condicional dados los datos observados y los datos faltantes) sobre la distribución posterior de los datos faltantes.

El desarrollo e implementación de la imputación estadística, estará en una relación inversamente proporcional, frente a las reglas de diseño del cuestionario y crítica manual, es decir, a mayor control previo y mejor diseño de los instrumentos de recolección (estructura lógica y secuencial, en la redacción y formulación de preguntas) menor el impacto de aquella.

El Censo de Edificaciones posee dentro de su estructura logística reglas de crítica manual y sistematizada, procedimientos que garantizan en gran medida, que la información recolectada en campo y grabada en oficina, cumple con los estándares de calidad esperados.

Sin embargo, para algunas variables ya sea el dato estimado o real, se presentan valores fuera del rango establecido, que deben ser ajustados según el criterio del temático experto; la metodología y los razonamientos aplicados no tienen un criterio homogéneo debido fundamentalmente a la relación de las variables como: tipo de edificación, destino, estrato socioeconómico y a la diversidad de precios encontrados.

4. ANTECEDENTES DEL CENSO

Tradicionalmente, el comportamiento y evolución del sector de la construcción se venía calculando a través de la actividad edificadora reportada por los indicadores de licencias de construcción, del índice de costos de la construcción de vivienda, y del consumo aparente de cemento. Sin embargo, el empleo y análisis a corto plazo de estos indicadores presentaban limitaciones al no proporcionar sino una parte del acervo estadístico para estos estudios.

El sector de la construcción se caracteriza por el dinamismo y la variedad de interrelaciones que mantiene con los demás sectores de la economía. La importancia se hace evidente al considerarlo como un gran generador de empleo y del producto interno bruto (PIB), además de participar en la creación de una considerable formación bruta de capital fijo.

Teniendo en cuenta lo anterior, a partir de diciembre de 1996, el Departamento Administrativo Nacional de Estadística (DANE), con el fin de proporcionar herramientas e indicadores que complementen y optimicen los análisis de la actividad edificadora, diseñó una metodología que permitiera medir los grandes agregados macroeconómicos a partir de información recopilada de fuentes directas, donde se incluye: obras en proceso de construcción, obras encontradas paralizadas y las que han culminado su proceso constructivo.

A partir de la realización del tercer censo (julio de 1997), se implementó y adecuó la técnica estadística del “Panel longitudinal”, lo que permite el seguimiento detallado a través del tiempo de cada una de las obras en proceso objeto de estudio, eliminando procesos deductivos y de estimación, para obtener los principales indicadores del sector edificador mediante la observación directa y el seguimiento puntual a cada elemento de la población objeto.

De otro lado, el Censo de Edificaciones desde octubre de 1999 amplió su cobertura geográfica y su universo de estudio a las ciudades de Armenia y Pereira, con el propósito fundamental de hacer un seguimiento a los procesos de reconstrucción, dados en dichas ciudades como consecuencia del sismo que afectó al eje cafetero el 25 de enero de 1999.



CENSO DE EDIFICACIONES METODOLOGIA DE IMPUTACIÓN

CÓDIGO: ME-CEED-MET-01

VERSIÓN: 02

Página 21

Fecha: 09-07-07

En julio de 2001 se amplió el alcance de la investigación en los estratos 1 y 2. A partir de esta fecha en todas las áreas metropolitanas que conforman la cobertura geográfica de la investigación; se censa todos los estratos socioeconómicos.

A partir de abril 2002, en las ciudades de Armenia y Pereira no se hará seguimiento a los procesos de reconstrucción, solo se continuara recolectando información para el Censo de Edificaciones de obras en construcción.

Además, se realizará un seguimiento individual a las obras culminadas hasta que se venda la última unidad que conforma el proyecto.

5. ASPECTOS GENERALES DE CRÍTICA EN EL CENSO DE EDIFICACIONES

En el desarrollo del Censo, se agrupan diferentes esfuerzos para garantizar que la información suministrada por la investigación tenga un alto grado de confiabilidad y veracidad. No obstante, es inevitable que en algunos casos existan datos incorrectos, inconsistentes o atípicos, que obligan a tomar valores sustitutos para reemplazar el original que presenta ese inconveniente.

En la necesidad de cumplir con estos requisitos, la parte de consolidación de la información (temática), determina un estricto filtro de calidad como último paso de crítica, el cual garantiza que la información final posea un alto nivel de veracidad y sirva de insumo para el cálculo de otros indicadores.

El Censo de Edificaciones realiza varios niveles de crítica que se pueden sintetizar en:

- Primer nivel de crítica, la realiza el encuestador y el supervisor (crítica de campo)
- Segundo nivel de crítica, recepción de información, controles de cobertura y calidad de la información responsabilidad del coordinador operativo o coordinador de zona y analista de la información (crítica de oficina)
- Tercer nivel de crítica la realiza el coordinador del operativo
- Cuarto nivel de crítica realizada en la captura y el administrador de la información

5.1 PRIMER NIVEL DE CRÍTICA

El recolector y el supervisor realizan la crítica de campo, que consiste en verificar cobertura geográfica y validar la información de los instrumentos de recolección, controlando que las preguntas cruciales dentro del cuestionario hayan sido contestadas correctamente (criterio de mínimo completamiento).

La variable precio de venta del metro², es de requerimiento mínimo y durante la capacitación y en el manual se hace énfasis para que siempre aparezca algún valor bien sea real o estimado, cuyo procedimiento es el siguiente:

“Si el precio es obtenido directamente por la fuente, el encuestador demarca la casilla correspondiente a R (real), si por el contrario, el precio obtenido fue estimado ya sea por que la información no fue dada en el sitio de obra o porque la edificación es para uso propio ó no se vende, el encuestador diligencia la casilla correspondiente a E (estimada). Para este último caso el equipo de trabajo estima en campo teniendo en cuenta los siguientes criterios:

- Para todos los casos la estimación la realiza el supervisor en campo con un mínimo de tres obras censadas. La estimación obtenida en campo debe ser cotejada con los demás medios de estimación.
- Cuando se trate de destinos ubicados en el estrato 1 y 2, la estimación igualmente la realiza el supervisor tomando como referencia los precios de vivienda de interés social en la ciudad.
- Para los estratos 3 al 6, la estimación se realiza en campo y es analizada con los demás medios en oficina mediante: revistas guías, Finca raíz, periódicos y otros, este procedimiento esta descrito en el Manual Básico vigente de Recolección, Crítica y Codificación.

El formulario de obras nuevas en el numeral 5, presenta dos formatos correspondientes al tipo de edificación que se está construyendo, con las siguientes características:

En el formato uno, si el destino es:

- Para la venta: son aquellas obras que se están construyendo para ser vendidas. Por esta característica, poseen una fuente de información (sala de ventas). El dato obtenido para la variable precio de venta, debe ser real, porque generalmente hay un informante idóneo.
- Destinos para uso propio: son aquellas edificaciones que se están construyendo y no son comercializables. Poseen la particularidad de no tener un informante idóneo, esta clase de obra por lo general en terreno se estima la variable precio.

Para los dos casos, la base fundamental de su valoración la constituye el concepto de precios de venta o de mercado, lo que quiere decir, que dicho concepto se encuentra implícito en la estructura de tipo costo incluido el margen de utilidad.

En el formato dos, destinos no comercializables:

- Son edificaciones no comercializables, por tal motivo, poseen la característica de que se recolecta el costo directo, que incluye entre otros: materiales, mano de obra y subcontratación.
- Cuando la información para estos destinos no es obtenida en campo, tienen imputación en oficina, se realizan los procesos de crítica y filtros de calidad al igual que los destinos del formato uno. La verificación debe hacerse con medios especializados en costos como la revista Construdata que tiene un capítulo de análisis de costos de la construcción.

Cuando se incorporan o captan destinos del formato dos, automáticamente el sistema los convierte a precios de mercado (comercializables), adicionando al costo directo un coeficiente del 55%, lo que quiere decir, que el coeficiente es lo que le hace falta al valor obtenido para completar la estructura de precio de las edificaciones que son valoradas a precio de mercado.

es preciso aclarar que el coeficiente del 55% se obtuvo desde el inicio de la investigación, trabajo que realizó el Asesor Oscar Borrero. Este coeficiente corresponde al peso que tiene el costo directo en la estructura de precio de las edificaciones.

Por otra parte en la actualidad, este coeficiente esta siendo revisado y analizado teniendo en cuenta variables de asociación como son: área metropolitana, destino de la obra y estrato socioeconómico, variables hexógonas y endógenas, con el fin de ser evaluado técnicamente. Una vez conocidos los resultados metodología y temática ajustaran o mantendrán dicho porcentaje de imputación.

Finalmente, cuando el temático realiza la verificación de la variable precio de venta, en los destinos no comercializables y encuentra inconsistencias o datos atípicos, debe descomponer nuevamente como un proceso inverso para analizar de manera objetiva el costo directo de esa variable.

5.2 SEGUNDO Y TECER NIVEL DE CRÍTICA

El segundo nivel de crítica se realiza con la recepción del material diligenciado en campo, quedando debidamente clasificado: completo y pendiente de verificación.

En el tercer nivel de crítica, se verifica la cobertura geográfica y la calidad de la información. Si el cuestionario, por algún motivo aparece con dato faltante en la variable precios al llegar a crítica de oficina, este valor debe inferirse o estimarse por medio de los precios que aparecen en: revistas guías tales como: finca raíz, periódicos, internet, etc.

El valor del precio generalmente debe ser analizado teniendo en cuenta, las características propias de la obra, el área y tipo de edificación, estrato socio económico, destino, etc. y las observaciones que hace el encuestador.

Con todos estos argumentos, la persona designada para realizar dicha tarea establece un valor razonable homologado de acuerdo al mercado y al comportamiento real del sector en ese trimestre y para esa ciudad.

Otro filtro de calidad, lo realiza el coordinador operativo, revisa todos los cuestionarios que presentan valores estimados para la variable precio por metro² y hace un análisis profundo del comportamiento de esa variable, valiéndose de medios de información, anteriormente mencionados.

Para complementar el análisis, también utiliza el conocimiento del sector y su experiencia, así como la información adicional del cuestionario. Si el precio estimado es coherente según las normas establecidas metodológicamente, se avalan los datos estimados, en caso contrario nuevamente con las revistas especializadas en el tema y las características de la construcción se realiza el ajuste pertinente.

5.3 CUARTO NIVEL DE CRÍTICA

Durante la captura de la información, se realiza otro filtro de calidad, en donde se establecen rangos de aceptación en el programa de grabación para las variables cuantitativas, en este caso, en la variable precios el dato es validado a través de un valor mínimo y un máximo.

El responsable de efectuar la labor de captura, detecta los valores que se encuentran fuera de los parámetros establecidos, el sistema automáticamente emite un aviso de advertencia. Este proceso implica normas que se deben tener en cuenta y que se describen a continuación:

- Para digitar información fuera del rango establecido en el programa de captura, el administrador de la información debe ingresar con clave de usuario.

- Si el valor (precio de venta o costo directo), es real y se ha ingresado con clave de usuario y la información se encuentra por fuera del rango establecido, esta se verifica directamente (vía telefónica o en terreno), y posteriormente es avalada por el coordinador general, dejando por escrito un antecedente técnico que justifique el comportamiento de esa variable.
- Verificado el valor del precio, el responsable de la tarea (administrador de la información), solicita la clave de administrador al coordinador general, lo que determina que el valor incorporado cuente con todos los estándares de calidad.
- De otro lado, si el dato que se encuentra fuera de los parámetros establecidos en el programa, es incorporado al sistema con clave digitador, y no pasa por verificación del coordinador, el sistema automáticamente asume el valor del precio del Censo pasado y lo incorpora al registro. Este valor a primera vista no ha pasado por los filtros de calidad o crítica establecidos.
- Así mismo, el sistema arroja listados de inconsistencias en donde se determina que el dato incorporado está fuera de rango. Estos listados deben estar “*sustentados técnicamente*”, para que en el momento de realizar el último filtro en el DANE Central, no se vuelva a verificar.

5.3 METODOLOGIA DE LA EXPERTA

En el DANE Central, temática realiza el último proceso de calidad a la información del Censo. Dicho proceso consiste en un análisis económico y de verificación de variables cuantitativas como áreas y precio de venta por metro². Para la realización de este análisis se toman cuatro bases de datos (culminadas, paralizadas, proceso e histórico), sin embargo para efectos del análisis con la experta, solo se tendrán en cuenta dos bases (proceso e histórico), puesto que tienen registros trimestrales que permiten efectuar comparaciones, trimestre actual versus trimestre anterior.

5.3.1 Primera fase validación

Se hace un análisis detallado sobre la base de datos de obras en *proceso*, cuyos registros tienen pareja; utilizando la herramienta de Excel (tablas dinámicas) y aprovechando combinaciones de variables que permiten visualizar a nivel de ciudad, los registros que presentan inconsistencias.

- Se calcula el valor máximo y mínimo del precio de venta por metro²; para cada ciudad, por destino y estrato socioeconómico, identificando registros fuera de rango.

$$\text{Sea } A = \{X / X = \text{"Precio de venta por metro}^2\}$$

Donde:

x_{ij}	Valor del precio de venta por metro ² , para cada ciudad, según destino y estrato socioeconómico
M_{ij}	Valor máximo del precio de venta por metro ² , para cada ciudad, según destino y estrato socioeconómico
m_{kj}	Valor mínimo del precio de venta por metro ² , para cada ciudad, según destino y estrato socioeconómico
i	1, 2, 3,10 Destino de la obra
j	1, 2, 3,..... 6 Estrato socioeconómico

entonces,

$$\text{Valor máximo de } A = \{X / \forall x_{ij} \in A \exists x_{ij} > M_{ij}\}$$

$$\text{Valor mínimo de } A = \{X / \forall x_{ij} \in A \exists x_{ij} < m_{ij}\}$$

Para efectos de análisis, esta técnica permite observar el comportamiento de la variable, llegando hasta el nivel de micro dato, describiendo el número del formulario que posee el valor fuera de rango (la obra con todos los datos en el trimestre), si la Dirección Territorial y Subsede justificó técnicamente ese dato se admite como correcto, en caso contrario se marca para ser corregido.

- Se obtiene el precio de venta promedio ponderado por tamaño de obra, calculado con tablas dinámicas, por destino y estrato socioeconómico, total ciudad (área metropolitana).

$$\text{Sea } B = \{Y / Y = \text{"Metro}^2 \text{ construido"}\}$$

Donde:

y_{ij} Metro² construido, según destino y estrato socioeconómico

Media aritmética ponderada por tamaño de obra

$$\overline{Xi} = \frac{x_{i1}y_{i1} + x_{i2}y_{i2} + \dots + x_{i6}y_{i6}}{y_{i1} + y_{i2} + \dots + y_{i6}} = \frac{\sum_{j=1}^6 x_{ij} y_{ij}}{\sum_{j=1}^6 y_{ij}}$$

La ponderación se obtiene de relacionar el destino, el valor del área en metro² construido sin incluir garaje cubierto, se multiplica el metro² construido por el precio de venta del metro² construido según destinos; la sumatoria de estos resultados para el destino se divide por la suma total de metros² construidos del destino.

- Se comparan los promedios ponderados del trimestre, por estrato y destino (actual frente pasado) y se obtienen las variaciones trimestrales.

Sea $C = \{ \overline{X} / \overline{X} = \text{"Promedio ponderado por tamaño de obra del precio de venta por metro}^{2n} \}$

Donde:

t Período actual
t-1 Período anterior

$$V = \left\{ \frac{\overline{X}_t}{(\overline{X}_{t-1} - 1)} \right\} * 100$$

La variación permite ver aumentos o disminuciones significativos, que ameritan un mejor nivel de detalle a nivel de registros. El análisis consiste en detectar variaciones fuera de los parámetros establecidos.

- Después de consolidar cada una de las zonas de cobertura geográfica, se corre la base proceso: censo actual contra censo anterior, a fin de detectar, aumentos o disminuciones superiores al 15% o -15%, respectivamente.

- Este proceso implica, seleccionar por estados de obra, aquellas que tengan dos observaciones; en nuestro caso solo aplica para las obras que continúan en proceso y cuyos códigos son "0 y 2".
- Con estas combinaciones, el sistema arroja una variación trimestral a nivel de micro dato y esto determina la evolución de los precios por obra, departamento, número de formulario, estrato, y destino.
- El análisis consiste en primera instancia en una selección de aquellos registros que tienen incrementos o disminuciones de precios iguales o superiores al 15 % y -15 % y a las cuales se les debe realizar un seguimiento de verificación y/o ajuste final.

La primera validación permite clarificar a través de la fuente, si la información es consistente. Si pese a la validación el dato es inconsistente o atípico y no posee una justificación de carácter técnico, se realiza otro tipo de análisis, incorporando la base de datos a nivel histórico, permitiendo realizar un seguimiento de la obra a través del tiempo. Este análisis se describe como la segunda fase del proceso.

5.3.2 Segunda fase imputación

Si en el análisis de la información, se detecta que el registro observado, no posee la consistencia y justificación técnica debidamente sustentada, se realiza un seguimiento a nivel de registro e históricamente, sobre el dato objeto de análisis.

Para esto, se toma la base de datos histórica, denominada precibol y se trabaja sobre tres criterios básicos de corrección: Promedios simples, Moda y Tendencia; hasta depurar el registro inconsistente. Una vez depurada la base del histórico, se realiza nuevamente un chequeo a fin de garantizar que el registro atípico haya tenido el proceso de validación e imputación pertinente.

- El promedio, que es un valor central del conjunto de valores, no admite para su aplicación, la presencia de datos extremos o variaciones significativas.
- Cuando el método de imputación es la moda de un conjunto de datos, esta corresponde a un dato con mayor frecuencia de ocurrencia (unimodal), por lo tanto, la serie de datos debe contener valores de la variable repetidos.

En algunos casos, puede existir más de una moda, es decir, para el mismo conjunto de datos existen dos o más valores de la variable que se repite con la misma frecuencia y que el valor a imputar puede ser reemplazado por cualquiera de las modas. En este caso la selección de la moda está directamente relacionada con el tipo de dato, es decir, si es real o estimado; prevaleciendo siempre el dato obtenido de fuente idónea.



CENSO DE EDIFICACIONES METODOLOGIA DE IMPUTACIÓN

CÓDIGO: ME-CEED-MET-01

VERSIÓN: 02

Página 30

Fecha: 09-07-07

- Otra manera de hacer imputación para esta investigación, es a partir de la tendencia, es decir, establecer el comportamiento del valor actual en comparación con los valores de trimestres anteriores (variaciones), relacionados con variables de similitud.

6. REGLAS Y ESPECIFICACIONES

Para detectar los valores atípicos en la variable numérica precio de venta por M², se toma como punto de partida y análisis la Base proceso, cuya estructura del archivo .EXCEL, (ver anexo 1), diseño de registro.

Se toma como herramienta de análisis esta base de datos, ya que en ella se presenta el archivo con dos períodos consecutivos; trimestres anterior y actual, que corresponde al consolidado de cada obra con las variables de interés para la investigación

6.1 ESPECIFICACIONES DE VALIDACIÓN

La investigación Censo de Edificaciones, cuenta con documentación completa de los instrumentos de recolección que lo conforman. Para las especificaciones de validación, se utilizó la estructura básica del formulario de obras nuevas, el manual básico de recolección, crítica y codificación, los rangos establecidos por las Direcciones Territoriales y Subsedes y el documento especificaciones de sistemas elaborado por Andrés del Castillo Garavito.

6.2 ESPECIFICACIONES DE CONSISTENCIA

Metodológicamente la consistencia presenta dos parámetros: de tipo lógico y de tipo estructural; desde el punto de vista lógico, se analiza a nivel de micro dato, que los valores de la variable observada sean reales y recolectados bajo condiciones normales, y no presenten fluctuaciones o saltos inesperados, que hagan inconsistente el dato. Para cumplir con este análisis de consistencia y detectar valores extremos o atípicos se establecen rangos de aceptación a través de valores máximos y mínimos avalados por la investigación.

Desde el punto de vista estructural, se analiza el comportamiento de la variable desde dos ángulos, relacionada con variables de similitud en forma directa, las cuales pueden afectar

el comportamiento de la misma en un momento determinado, y el dato actual en relación a los trimestres anteriores (análisis estacional), para determinar el comportamiento a través de variaciones.

La variable objeto de análisis, depende directamente del comportamiento de otras variables que explican el valor de la misma y que afectan directamente el precio de la obra, como la ciudad, el estrato socioeconómico cuando esta es para la venta, el destino, otras.

- Se hace el cálculo con los datos de la base proceso, para la variable precio de venta por metro² según destino, estrato socioeconómico y estado de la obra (0 y 2), calculando valores máximos y mínimos de esa variable.
- Se verifica con observaciones de la base de datos, a través de una matriz por destino, para detectar cambios en las variaciones y validar rangos o parámetros establecidos, estos son estacionales (anuales) o tendencias (trimestrales), según tipo de valor (real ó estimado) y teniendo en cuenta:

El departamento, igual número de formulario (éste identifica la obra a través del tiempo), el destino (un mismo número de formulario puede presentar dos o más destinos), el estado actual de la obra.

- Se calcula el promedio ponderado del precio por metro² por ciudad, según destino (1 a 10) y estrato socioeconómico para la variable precios, con este cálculo se obtienen las variaciones trimestrales de la variable de estudio, con el fin de analizar el comportamiento de la misma, afectado por factores de correlación.
- Sin embargo, se recomienda hacer verificación que los supuestos (tanto en rangos como en variaciones) se cumplan, para que estos tengan validez a través del tiempo.
- Por otra parte, se recomienda hacer ajustes a la metodología, en caso de cambios en la dinámica del sector, a través de la verificación sugerida para las variaciones, porque estos rangos cambian según el comportamiento del sector por ciudad.

6.3 REGLAS DE IMPUTACIÓN

La variable principal de análisis es precios, tomada de la base de datos *proceso*, donde se detectan las inconsistencias. Sin embargo, ajuste y análisis histórico, debe realizarse en la base de datos del histórico (*precibol*), en esta base se presenta el micro dato (obra),

desde que fue captada en el censo, es decir, el histórico de la obra, con todas las variables de similitud para el estudio.

En este sentido la base de datos precibol, permite consultar y verificar la información de la obra, tanto del período actual como de períodos anteriores, haciendo un seguimiento a la unidad estadística durante el transcurso del proceso de construcción y el estado en el proceso de construcción.

Los estados de la obra en construcción pueden ser:

Cuando se analiza la serie histórica, base de datos Precibol, existe una variable "estado" que se presenta codificada de la siguiente forma:

Cuando actual censo =1, obras en proceso

- Estado 0 = Obras que continúan en proceso
- Estado 1 = Obras que reinician con número de orden
- Estado 2 = Obras que ingresan por ampliación de cobertura
- Estado 3 = Obras nuevas generadas en el periodo intercensal
- Estado 4 = Obras que reinician con número de formulario

Cuando actual censo = 2, obras paralizadas

- Estado 0 = Obras que continúan inactivas
- Estado 1 = Obras inactivas nuevas del período intercensal

Cuando actual censo = 3, obras que estando en proceso culminan

Para efectos de la imputación solo se trabaja la unión de los códigos 0+2, es decir continúan en proceso ya que para efectos de cálculo de variaciones esta debe presentar mínimo dos trimestres.

Por lo tanto las reglas y especificaciones para el tratamiento de dicha variable, según formato son:

6.3.1 Formato uno

Solo si la obra continúa en proceso (estado 0 + 2), es para la venta (código=1), tipo de valor real o estimado (código=1 ó 2) y variación trimestral superior o inferior al 15% ó - 15%, no justificada técnicamente por la Dirección Territorial y Subsede, se procede así:

1. Si el valor de la variable, precio de venta por metro², repite en tres o cuatro trimestres de una vigencia censo (año de recolección de la información), para la misma obra, igual destino; el dato atípico reemplácelo por el valor que más se repite (MODA), si corresponde a un valor real.
2. Si el registro analizado para una misma obra, no posee valores que repiten en dos o más vigencias del censo, pero sus tendencias son coherentes en términos de aumento o disminución, calcule sin incluir el dato atípico, el promedio simple y reemplace el dato atípico por el valor de la media sin incluir decimales.
3. Si el registro analizado tiene varias vigencias y el valor de la variable por trimestre presenta diferencias sustanciales (valores heterogéneos), no tienen distribución definida en su comportamiento en términos de aumento o disminución y no guardan coherencia en los aumentos o disminuciones ya que se presentan oscilaciones muy fuertes, entonces:
 - Aplique incrementos o disminuciones del 10% (según rango establecido metodológicamente por ciudad), iniciando el procedimiento sobre el registro que presente el valor real más bajo, el incremento del 10%, teniendo en cuenta la vigencia del censo e iniciando el aumento en forma ascendente por vigencia y mes censo, hasta obtener incrementos acumulados sobre el valor atípico.

6.3.2 Formato dos

La imputación para las edificaciones con destino de 6 a 10, que corresponden al formato dos, es decir obras no comercializables, se realiza a la variable antes de agregar el valor porcentual del 55% al costo directo de la obra, la metodología del Censo aplica este porcentaje con el propósito de convertirlo en un valor a precio de venta o de mercado.

La metodología de imputación para la variable costo directo, se efectúa igual que al formato uno con la salvedad mencionada anteriormente.

ANEXO A. DISEÑO DE REGISTRO TABLA PROCESO.EXCEL

Número de variables 35

Nro.	Nombre de la variable	Campo	Tamaño o longitud	Valores válidos	Tipo de dato	Descripción
1	Vigencia censo	VIG_CENS	4	Año ≥ Fecha del operativo	Número	Año de diligenciamiento del censo de obras en proceso
2	Mes en que se diligencia el censo	MES_CENS	2	01= Enero 04= Abril 07= Julio 10= Octubre	Número	Código del mes en que se hace el censo
3	Departamento	COD_DPTO	2	11=Bogotá 25=Cundinamarca 05=Antioquía 76=Valle del Cauca 08=Atlántico 68=Santander 66=Risaralda	Número	Referencia geográfica donde se realiza el censo
4	Municipio	COD_MUNI	3	001=Santafe de Bogotá 754=Soacha 001=Medellín 088=Bello 266=Envigado 360=Itagüí 001=Cali 892=Yumbo 001=Barranquilla 758=Soledad 001=Bucaramanga 276=Florida Blanca 307=Girón 547=Piedecuesta	Número	Ciudad capital incluida en el censo y área metropolitana

Nro.	Nombre de la variable	Campo	Tamaño o longitud	Valores válidos	Tipo de dato	Descripción
				00=Pereira 170=Dosquebradas		
5	Localidad o comuna	LOCALIDA	2	Para Bogotá 01=Usaquén 02=Chapinero 03=Santafe 04=San Cristobal 05= Usme 06=Tunjuelito 07=Bosa 08=Kennedy 09=Fontibón 10=Engativa 11=Suba 12=Barrios Unidos 13=Teusaquillo 14=Los Mártires 15=Antonio Nariño 16=Puerto Aranda 17=La Candelaria 18=Rafael Uribe Uribe 19=Ciudad Bolívar	Carácter	El resto de municipios no poseen comunas ni localidades.
6	Clase	CLASE	1	1=Urbano ó cabecera municipal	Número	División cartográfica que permite diferenciar, cabecera municipal, centro poblado y área rural
7	Sector	SECTOR	4	$1 \leq \text{Sector} \leq 9$ secciones = a un barrio	Número	División cartográfica, generalmente equivale a un barrio
8	Sección	SECCION	2	Sección ≤ 20 manzanas contiguas, pertenecientes al mismo sector o barrio	Número	División cartográfica, que equivale a 20 manzanas contiguas del mismo sector o barrio
9	Manzana	MANZANA	2	>0	Número	Lote de terreno

Nro.	Nombre de la variable	Campo	Tamaño o longitud	Valores válidos	Tipo de dato	Descripción
					co	delimitado por vías de tránsito vehicular o peatonal
10	Número de formulario	NRO_FORM	7	Mín - Máx 10001-20000 20001-30000 30001-40000 40001-50000 50001-60000 60001-70000 70001-80000	Numérico	Rangos ciudades y áreas metropolitanas Bogotá Medellín Cali Barranquilla Bucaramanga Pereira Armenia
11	Estrato	ESTRATO	1	1=Bajo bajo 2=Bajo 3=Medio Bajo 4=Medio 5=Medio alto 6=Alto	Numérico	De acuerdo al estrato socioeconómico donde está ubicada la obra
12	Ampliación	AMPLIACION	1	1=Si 2=No	Numérico	Cuando el área construida implica una ampliación, en su diseño estructural
13	Unidad total de la obra	UNI_TOBRA	10	UNI_TOBRA ≥ número de unidades vendidas	Caracter	Variable de control para verificar con unidades vendidas
14	Destino	DESTINO	2	01=Apartamento 02=Oficinas 03=Comercios o locales 04=Casas 05=Bodegas 06=Educación 07=Hoteles 08=Hospitales 09=Administración Pública 10=Otros	Numérico	Tipo de edificación que se está construyendo
15	Tipo de	USO_PROP	1	1=Para la venta	Numérico	Identifica la obra si es o

Nro.	Nombre de la variable	Campo	Tamaño o longitud	Valores válidos	Tipo de dato	Descripción
	uso			2=Propio o no comercializable	co	no comercializable
16	Etapa A	ETAPA_A	2	$1 \leq ETAPA_A < 50$	Número	# de divisiones del proyecto
17	Subetapa	SUBETAPA	15	$SUBETAPA \geq 0$	Número	Subdivisión al interior de una etapa
18	Total etapas	T_ETAPA	2	$1 \leq T_ETAPA < 50$	Número	Total divisiones del proyecto
19	Mes inicio obra	MES_INIC	2	$1 \leq MES_INIC \leq 12$	Número	Mes de inicio de la construcción
20	Año inicio obra	VIG_INIC	4	$VIG_INIC \geq 1995$	Número	Año de inicio de la construcción
21	Capítulo del proceso constructivo	CAPITULO	1	1=Excavación y cimentación 2=Estructura y cubierta 3=Mampostería y Pañetes 4=Acabados nivel 1 5=Acabados nivel 2 6=Acabados nivel 3	Número	Procesos que se debe llevar a cabo para construir una obra, estado de avance constructivo en que se encuentra la obra para ese destino
22	Área unitaria de garaje	M2_UNIGA	5	En promedio $10 \leq M2_UNIGA \leq 16 \text{ Metros}^2$	Número	Espacio cubierto para garaje
23	Número de unidades de garajes	CANT_GAR	6	$CANT_GAR \geq 0$	Número	Total de garajes cubiertos para el destino
24	Precio unitario del garaje cubierto en miles de pesos	VR_GARAJ	12	$1000 \leq VR_GARA \leq 20000$	Número	Precio del garaje cubierto en miles de pesos
25	Tipo valor garaje	TIVR_GAR	1	R= Real E= Estimado	Caracter	Forma de obtener el precio del garaje

Nro.	Nombre de la variable	Campo	Tamaño o longitud	Valores válidos	Tipo de dato	Descripción
26	Metros ² construidos por destino	M2_CONST	6		Número	Cantidad de metros ² construidos de la obra para ese destino
27	Metros ² vendibles destino	M2_VENDI	6	1, 2 ≤ 30 Metros ² 3 ≤ 10 Metros ² 4 ≤ 40 Metros ² 5 ≤ 20 Metros ²	Número	Área vendible del destino
28	Valor del metro ² vendible sin incluir garaje cubierto	VALOR_M2	5	>0	Número	Valor del metro ² vendible sin incluir garaje cubierto. Precio de venta
29	Tipo de valor destino	TIVR_DES	1	1=Valor real 2=Estimado	Número	Sí el precio es obtenido a través de la fuente idónea o estimado
30	Unidades de destino	UND_DEST	6	>=0	Número	Unidades de construcción del destino
31	Unidades vendidas de destino	UNI_VDEST	4	>=0 y <= 500	Número	Número total de unidades vendidas hasta el momento del censo
32	Unidades por vender	UNI_XVENDE	4	= 0 ó <= 500	Número	Número de unidades que faltan por vender hasta el momento del censo
33	Grado de avance	GRAD_AVAN	3	> 0 y <= 100%	Número	Valor porcentual del capítulo constructivo
34	Actual censo	ACT_CENS	1	1=En proceso 2=Paralizada 3=Culminada	Número	Variable de control para identificar el estado de la obra en el momento del censo
35	Estado	ESTADO	1	Para actual censo=1 0=Continúa en proceso 1=Reiniciada antes con número de orden 2=Cobertura	Número	Situación en la que se encuentra la obra en el momento del censo



CENSO DE EDIFICACIONES METODOLOGIA DE IMPUTACIÓN

CÓDIGO: ME-CEED-MET-01

VERSIÓN: 02

Página 40

Fecha: 09-07-07

Nro.	Nombre de la variable	Campo	Tamaño o longitud	Valores válidos	Tipo de dato	Descripción
				3=Nuevas 4=Reiniciadas antes en proceso 9=No aplica <i>Para actual censo=3</i> 0=Continua 1=Del censo actual 9=No aplica		

ANEXO B. ALGORITMO DE IMPUTACION

Puesto que la imputación, para la variable precio de venta del metro cuadrado, en la investigación Censo de Edificaciones esta basada en estimadores, como la media, la moda y la tendencia; la base de datos debe presentar un ordenamiento lógico y cronológico, es decir:

- A cada obra en construcción, captada por el Censo de Edificaciones, se le asigna un número de formulario (único por obra), el cual identifica a la obra durante todas las captaciones de información en el mismo. Por lo tanto en la base de datos histórica, el identificador o llave para cada obra es el número de formulario.
- La base de datos *Precibol*, es decir, el histórico de la obra, se condiciona, solo se toma aquellas obras que tienen mínimo dos periodos de captación, pertenecen al destino uno y son para la venta.
- Las obras con mínimo dos periodos de captación, corresponden a edificaciones en proceso de construcción o aquellas que estando paralizadas reanudan actividad.
- La condición de que las obras tengan mínimo dos periodos de captación, es requisito indispensable para calcular variación.
- Esta base se ordena, con la llave de enlace que corresponde al número de formulario, según año, mes de recolección de la información y destino de la obra.
- Una vez ordenada la base, se calculan las variaciones trimestrales por obra, en forma condicional, de acuerdo a:

Sea:

A = Número de formulario

B = Destino

C_t = Precio de venta del metro cuadrado, en el trimestre t

V_t = Variación trimestral

t = Trimestre actual

$t-i$ = Trimestres anteriores, con $i = 1, 2, \dots, n$

Si $a_t = a_{t-1}$ y $b_t = b_{t-1}$ entonces calcule la variación, en caso contrario no haga nada y asigne punto a ese campo:

$$V_t = \left(\frac{c_t}{c_{t-1} - 1} \right) * 100$$

- Calculadas las variaciones, se marca las variaciones mayores y menores o iguales a 15 a -15, para imputar los valores respectivos de la variable C_t (precio de venta del metro cuadrado), que se consideran metodológicamente fuera de rango, es decir:

Si $V_t \geq 15$ ó $V_t \leq -15$, \Rightarrow marque C para imputar, sin tener en cuenta cero o punto.

- Se cuentan los registros marcados para imputar, con el fin de establecer la magnitud o porcentaje de datos a imputar.
- Cada registro marcado para imputar, es analizado a partir de las frecuencias de recolección en el año, esto corresponde, a cuatro meses, es decir:

1 = Enero
4 = Abril
7 = Julio
10 = Octubre

En el año 1998, la recolección del Censo se realizó en el mes octavo en todas las ciudades que cubre el Censo sin incluir Armenia.

- Se realiza el conteo según año de recolección a partir de la cantidad de trimestres de captación en el censo, este procedimiento es necesario para establecer los criterios de imputación.
- Si el dato de la variable en estudio, fue captado para el mismo año, en los cuatro meses de recolección del Censo y el valor recolectado se repite en tres ocasiones, entonces, para variaciones fuera de rango, impute el dato por el valor que se repite, es decir:

Sea:

V_r = Valor real de la variable

V_e = Valor de la variable estimado en campo u oficina

$(C_t = C_{t-1} = C_{t-2}) \neq C_{t-3}$ y C_{t-3} fuera de rango \Rightarrow impute C_{t-3} por C_t ó C_{t-1} ó C_{t-2}

- Este procedimiento se aplica a todas las posibles combinaciones donde aparece un dato atípico, con las características antes mencionadas, ejemplo:

$(C_t = C_{t-2} = C_{t-3}) \neq C_{t-1}$ y C_{t-1} fuera de rango \Rightarrow impute C_{t-1} por C_t ó C_{t-2} ó C_{t-3}

- El procedimiento de imputación es igual, cuando los datos recolectados para la variable de estudio son estimados en campo u oficina y presentan variaciones fuera de rango.
- Si el dato de la variable en estudio, fue captado para el mismo año, en tres de los cuatro trimestres de recolección del Censo y el valor recolectado se repite en dos ocasiones, entonces, para variaciones fuera de rango, impute el dato por el valor que se repite, es decir:

$(C_t = C_{t-1}) \neq C_{t-2}$ y C_{t-2} fuera de rango \Rightarrow impute C_{t-2} por C_t ó C_{t-1}

- Si el dato de la variable en estudio, fue captado para el mismo año, en dos de los cuatro trimestres de recolección del Censo y la variación esta fuera de rango, impute el valor fuera de rango de la siguiente manera:

$C_t \neq C_{t-1} \Rightarrow$ impute C_t por C_{t-1} ó

$C_t \neq C_{t-1} \Rightarrow$ impute C_{t-1} por C_t

- Para el caso en que los dos valores de la variable son estimados en terreno u oficina y presenta variación fuera de rango impute así:

$C_t \neq C_{t-1} \Rightarrow$ impute C_t por C_{t-1}

Para el caso en que las edificaciones tienen cinco y hasta ocho trimestres de captación en el censo y presentan variaciones trimestrales fuera de rango para la variable precio de venta del metro cuadrado, proceda así:

- Si las variaciones trimestrales son diferentes de cero o punto, existe una ó más variaciones fuera de rango, entonces proceda así, sin incluir el valor fuera de rango para el cálculo de la media:

$$\text{Si } \exists C_j \text{ donde } V_j > \delta \Rightarrow \text{impute } C_j \text{ por } \bar{C}^*$$

Donde:

δ =Variación máxima permitida metodológicamente

$$\bar{C}^* = \frac{C_1 + C_2 + \dots + C_{j-1} + C_{j+1} + \dots + C_n}{n - k}$$

k = Cantidad de datos a imputar

$n \geq 5k$, con $k \geq 1$

j = 1....k

- Si la variable en estudio, fue captada en tres o más trimestres de recolección del Censo y el conjunto de valores obtenidos para la variable son diversos, no tienen distribución definida y una o más variaciones se presentan fuera de rango, impute buscando en los trimestres anteriores a la variación fuera de rango, un valor real de la variable y a partir de éste incremente en un 10%, hasta que la variación se estabilice, así:

$$\text{Si } \exists V_i \geq \delta \Rightarrow \text{impute } C_i > C_x \text{ por } \hat{C}_i$$

Donde:

$$C_x = \text{Ultimo } C_j \text{ real}$$

$$\hat{C}_i = C_x \pm 10\% C_x (i - x)$$

i = Posición del dato a imputar

x = Posición del último dato real

x = 1, 2..., n

$i > x$

- Para cada dato atípico según la cantidad de registros en la obra y sus características solo es imputado una vez, por cualquiera de los métodos expuestos.



CENSO DE EDIFICACIONES METODOLOGIA DE IMPUTACIÓN

CÓDIGO: ME-CEED-MET-01

VERSIÓN: 02

Página 45

Fecha: 09-07-07

- Una vez finalizado el procedimiento, se verifica el comportamiento de la variable calculando nuevamente la variación, con el fin de comprobar primero la imputación y segundo que no se presenten nuevos casos de variación fuera de rango.

ANEXO C. PROGRAMA DE IMPUTACION EN SAS

ESTE PROGRAMA REALIZA LA IMPUTACION PARA LA VARIABLE PRECIO DE VENTA POR M2

DATA DOS;

SET UNO;

```
        COD_DPTO1 =LAG(COD_DPTO);
        COD_DPTO2 =LAG(COD_DPTO1);
        COD_DPTO3 =LAG(COD_DPTO2);
        Nro_form1 =lag(Nro_form );
        Nro_form2 =lag(Nro_form1);
        Nro_form3 =lag(Nro_form2);
        Destino1  =lag(Destino  );
        Destino2  =lag(Destino1 );
        Destino3  =lag(Destino2 );
        Vig_c1    =lag(Vig_c    );
        Vig_c2    =lag(Vig_c1   );
        Vig_c3    =lag(Vig_c2   );
        Valor_M21 =lag(Valor_M2 );
        Valor_M22 =lag(Valor_M21);
        Valor_M23 =lag(Valor_M22);
        Tivr_des1 =lag(Tivr_des );
        Tivr_des2 =lag(Tivr_des1);
        Tivr_des3 =lag(Tivr_des2);
        Mes_cens1 =lag(Mes_cens );
        Mes_cens2 =lag(Mes_cens1 );
        Mes_cens3 =lag(Mes_cens2 );
        IF (COD_DPTO1 =COD_DPTO) AND (Nro_form1=Nro_form) and
(Destinol=Destino)
            THEN Variac=ROUND((Valor_M2/Valor_M21*100)-100);
            ELSE Variac=.;
            Z=1;
            IF Variac <=-15     THEN Varia = -15;
            IF Variac >= 15     THEN Varia = 15;
            IF Variac in (.,0) THEN Varia = Variac;
        RUN;

/*PRUEBA A LA BASE PRECIBOL*/
DATA TRES_EST (KEEP=VIG_C MES_CENS COD_DPTO NRO_FORM ACTUAL ESTRATO
DESTINO
NUM_CENSO TIVR_DES VALOR_M2 VARIAC);
SET DOS;
```



CENSO DE EDIFICACIONES METODOLOGIA DE IMPUTACIÓN

CÓDIGO: ME-CEED-MET-01

VERSIÓN: 02

Página 47

Fecha: 09-07-07

```
IF VIG_C=2006 AND MES_CENS=10; /*CAMBIAR*/
IF VARIAC IN (.,0) THEN DELETE;
IF (TIVR_DES=1 AND (VARIAC>15 OR VARIAC<-15)); /*REALES*/; RUN;
/*IF (TIVR_DES=2 AND (VARIAC=>15 OR VARIAC=<-15)); RUN; /*ESTIMADOS*/

PROC SORT DATA=TRES;
BY COD_DPTO ACTUAL DESTINO; RUN;

/*PARTICIONO LA BASE PARA EMPEZAR A VER LO QUE NECESITO DEL TRIMESTRE*/
DATA TRES1;
SET DOS;
IF VIG_C=2005 AND MES_CENS = 12;
RUN;
/*ESTADISTICA DESCRIPTIVA*/

data pru; set WILLI.Precibol_i_07_120707;
*IF NUM_CENSO=42;
if NRO_FORM='3828' AND COD_DPTO='66';
run;

run;

PROC SORT DATA=TRES1;
BY COD_DPTO NRO_FORM DESTINO; RUN;
PROC FREQ DATA=TRES1;
TABLE COD_DPTO*TIVR_DES/ NOCUM NOROW NOCOL NOPERCENT LIST;
RUN;
PROC FREQ DATA=TRES1;
TABLE COD_DPTO*TIVR_DES*USO_PROP/ NOCUM NOROW NOCOL NOPERCENT LIST;
RUN;
PROC FREQ DATA=TRES1;
TABLE COD_DPTO*TIVR_DES*USO_PROP*DESTINO/ NOCUM NOROW NOCOL NOPERCENT LIST;
RUN;
DATA TRIMESTRE (KEEP= USO_PROP COD_DPTO TIVR_DES NRO_FORM ACTUAL VIG_C MES_CENS DESTINO VARIAC VARIA);
SET TRES1;
IF VARIAC>20 OR VARIAC<-20; RUN;
DATA TRIMESTRE; SET TRIMESTRE; IF TIVR_DES=2; RUN;
DATA TRIMESTRE; SET TRIMESTRE;
IF TIVR_DES=1; IF VARIAC=. THEN DELETE; RUN;
PROC FREQ DATA=TRIMESTRE;
TABLE TIVR_DES*VARIAC/ NOCUM NOROW NOCOL NOPERCENT LIST;
RUN;
PROC FREQ DATA=TRIMESTRE;
TABLE COD_DPTO*TIVR_DES*USO_PROP/ NOCUM NOROW NOPERCENT;
```



```
RUN;
PROC FREQ DATA=TRIMESTRE;
WHERE TIVR_DES=1 ;
TABLE COD_DPTO*VARIA/ NOCUM NOROW NOPERCENT;
RUN;

/*CASOS PARA LA TEMATICA*/
DATA TRIMESTRE1 (KEEP= VIG_CENS MES_CENS COD_DPTO NRO_FORM ACTUAL
DESTINO ESTRATO TIVR_DES VARIA VARIAC );
SET TRES;
IF VARIAC= . OR VARIAC= 0 THEN DELETE;
IF TIVR_DES=1 AND (VARIAC =>30 OR VARIAC <=-30);RUN;
/*CASOS PARA RESOLVER*/
DATA Z (KEEP=COD_DPTO NRO_FORM ACTUAL VIG_CENS MES_CENS ESTRATO DESTINO
TIVR_DES CAPITULO VALOR_M2);
SET WILLI.PRECIBOL_PIB190204;
IF COD_DPTO = 63 AND (ACTUAL=1719 OR ACTUAL=2244);RUN;
/*NUEVO DISEÑO PARA IMPUTAR*/
DATA SANCIONADOS;
SET TRES1;
IF VARIA=20 OR VARIA=-20;RUN;

PROC SORT DATA=SANCIONADOS;
BY COD_DPTO NRO_FORM DESTINO ;RUN;
PROC SORT DATA=UNO;
BY COD_DPTO NRO_FORM DESTINO ;RUN;
DATA IMPUTAR;
MERGE SANCIONADOS(IN=A)UNO(IN=B);
BY COD_DPTO NRO_FORM DESTINO ;
IF A; RUN;

DATA IMPUTAR1(DROP=COD_DPTO1 COD_DPTO2 COD_DPTO3 NRO_FORM1 NRO_FORM2
NRO_FORM3 DESTINO1 DESTINO2 DESTINO3 VIG_C1 VIG_C2 VIG_C3 VALOR_M21
VALOR_M22 VALOR_M23 TIVR_DES1 TIVR_DES2 TIVR_DES3 MES_CENS1 MES_CENS2
MES_CENS3 Z);
SET IMPUTAR;
RUN;

DATA IMPUTAR;
SET IMPUTAR;
IF VARIA=. THEN VARIA=0;
IF TENDE=. THEN TENDE=0;
IF MEDIA=. THEN MEDIA=0;
IF MODA =. THEN MODA =0;
IF IMPU=. THEN IMPU=0;RUN;
DATA A(KEEP=COD_DPTO NRO_FORM DESTINO TENDE MEDIA MODA IMPU VARIA );
```



CENSO DE EDIFICACIONES METODOLOGIA DE IMPUTACIÓN

CÓDIGO: ME-CEED-MET-01

VERSIÓN: 02

Página 49

Fecha: 09-07-07

```
SET WILLI.IMPUTAPRE_241103;
RUN;
DATA NUEVO;
SET UNO;
IMPU =000000;RUN;
DATA UNO(DROP=TENDE MODA MEDIA IMPU);SET UNO;RUN;
PROC SORT DATA=NUEVO;
BY COD_DPTO NRO_FORM DESTINO ;RUN;
PROC SORT DATA=PRECIBOLPIB191104;
BY COD_DPTO NRO_FORM DESTINO ;RUN;

DATA UNO;
SET WILLI.PIB_IMPU180804;RUN;
DATA DOS; SET WILLI.PRECIBOL_PRELIMINAR_180804;RUN;
PROC SORT DATA=UNO;
BY COD_DPTO NRO_FORM DESTINO ;RUN;
PROC SORT DATA=DOS;
BY COD_DPTO NRO_FORM DESTINO ;RUN;

DATA IMPUTADO;
MERGE NUEVO (IN=A) PRECIBOLPIB191104(IN=B);
BY COD_DPTO NRO_FORM DESTINO ;
RUN;
DATA IMPUTADO (DROP=VIG_C VARIAC VARIA);
SET IMPUTADO; IF IMPU>0 THEN VALOR_M2=IMPU;
IF MES_CENS=12 THEN MES_CENS =1;RUN;
DATA IMPUTADO (DROP=IMPU); SET IMPUTADO ;RUN;

DATA IMPUTADO1(DROP=VIG_C);
SET IMPUTADO;
IF MES_CENS=12 THEN MES_CENS=1;
IF TENDE=0 THEN TENDE=.;
IF MEDIA=0 THEN MEDIA=.;
IF MODA=0 THEN MODA=.;
IF IMPU=0 THEN IMPU=.;
IF IMPU>0 THEN VALOR_M2=IMPU;RUN;
PROC SORT DATA=IMPUTADO1;
BY COD_DPTO NRO_FORM DESTINO VIG_CENS MES_CENS;RUN;

DATA PRECIBOL (DROP=VARIA TENDE MEDIA MODA IMPU);
SET IMPUTADO1;
RUN;
```

CUADRO 1.

Distribución de registros y porcentaje según archivo prueba

Archivo prueba, destinos del 1 al 5	Registros	%
Total base de datos limpia	48.647	100
Observaciones consistentes	30.462	63
Observaciones con variación igual 0	15.422	31
Observaciones para imputar	2.763	6

FUENTE : DANE - Censo de Edificaciones. Archivo histórico de obras

CUADRO 2.

Cantidad de datos a imputar para la variable precio de venta del metro², en la investigación Censo de Edificaciones por destino, según Área Metropolitana

Área Metropolitana	Destinos					Total
	Apartamentos	Oficinas	Comercio	Casas	Bodegas	
Medellín	92	9	35	57	7	200
Barranquilla	57	5	44	59	5	170
Bogotá D. C.	582	123	253	593	174	1725
Soacha	17	1	3	24	8	53
Armenia	20	13	31	181	5	250
Pereira	18	3	11	69	8	109
Bucaramanga	30	4	19	38	5	96
Cali	23	5	29	93	10	160
Total	839	163	425	1114	222	2763

FUENTE : DANE - Censo de Edificaciones. Archivo histórico de obras

7. CONCLUSIONES

Este documento centró toda la atención a la variable precio de venta por metro² (formato uno), y costos directos (formato dos), analizando la relación con otros factores de similitud, que influyen directa e indirectamente sobre el mismo.

1. Se analizó las modificaciones que fueron realizadas manualmente, utilizando ante todo el criterio del investigador, demandando ello tiempo y esfuerzos considerables, introduciendo sesgos en la variable, ya que el ajuste se realizaba solo para los datos con mayores anomalías, especialmente los que estaban incrementando la dirección de los índices para los cuales sirve de insumo.
2. Se dejaba por fuera del análisis, otros datos que quizás necesitaban también del ajuste, pero por no presentar variabilidad considerable, no se tuvieron en cuenta.
3. Hasta la fecha y como se planteó en el documento, el ajuste realizado por la experta según variable y comportamiento de la misma, depende en gran medida de la correlación con variables de asociación, como el destino, el estrato socioeconómico, otras. La imputación se basó en estimadores, como la media aritmética, la tendencia y la moda.
4. Este método de imputación basado en estimadores, requiere procesamiento y manejo de los datos para su aplicación. Es por esta razón que los datos atípicos o inconsistentes se detectan trimestralmente en la base de datos proceso, pero el procedimiento para imputación amerita valores correctos, por ello, se hace necesario el manejo de todo el histórico de la obra.
5. Metodológicamente la imputación se aplica para: corregir o estimar valores atípicos, inconsistentes o faltantes y no para imputar datos que no presenten estas características.
6. Por otra parte, la imputación se efectúa a los datos de la variable precio de venta del metro², sin tener en cuenta si el precio fue estimado o real, esto con el fin de depurar en su totalidad la variable en estudio.
7. Una vez se ajusta el histórico del registro que presenta inconsistencia o dato atípico, este requiere en el futuro, *únicamente* controlar la variación trimestral, porque la serie

de captación de la obra en el censo queda depurada en su totalidad y no vuelve a presentar este tipo de inconvenientes, es decir este procedimiento no es iterativo.

8. A través del análisis se pudo observar, que existe un filtro de calidad sistemático, que no se lleva a cabo en su totalidad, y que corresponde a los valores reales fuera de rango, que deben ser reportados al Dane Central con justificación técnica, la sustentación de esta afirmación se basa, en que en muchas oportunidades, la experta debe realizar un detallado filtro de calidad y solicitar a la Dirección Territorial y Subsede la debida sustentación o en su defecto la corrección del registro fuera del intervalo.
9. Si metodológicamente existen rangos de aceptación para la variable, estos deben ser respetados a menos que existan argumentos técnicos válidos por parte del funcionario responsable de ese filtro de calidad, para que sea aceptado como dato correcto.
10. Por lo tanto, se deben analizar los argumentos que las Direcciones Territoriales y Subsedes envían cuando el porcentaje de datos fuera de rango es considerable, para determinar la fuente de variabilidad de los precios, ya que los intervalos de depuración de la variable deben estar acorde con la estructura y evolución del sector y la ciudad, y de ser necesario, modificarlos para que reflejen el comportamiento real de la construcción.
11. Al examinar el comportamiento de la variable costo directo, se observa la existencia de valores extremadamente altos o bajos para el rango de aceptación de la variable, y no permite captar fácilmente en el programa de consistencia los valores que pueden ser reales pero atípicos, por lo tanto no se puede proceder con la imputación hasta tanto esta no sea depurada y controlada de forma lógica y estructural esta variable.
12. Por otra parte, se debe tener en cuenta que al aplicar el método de imputación basado en estimadores, puede persistir situaciones de inconsistencias o datos atípicos, primero por manejo de la base de datos, segundo porque las reglas de crítica exigen procedimientos matemáticos, esto implica, que una vez realizada la imputación se debe volver a calcular la variación para verificar que no existen valores fuera de rango.
13. Por todo lo anterior es aconsejable, que el procedimiento descrito para imputación se pruebe, ajuste y finalmente se aplique en la base de datos del histórico. Una vez imputados los valores atípicos o inconsistentes en la variable precio, en forma histórica, desde la captación de la obra en el censo, no se requiere volver a aplicar el método, es decir *el procedimiento no es iterativo*.

14. De tal forma, que depurada la serie histórica, solo se requiere controlar los precios trimestralmente mediante variaciones y así, se esta controlando el histórico o registro de cada obra a través del tiempo.
15. Cuando se realiza el análisis y ajuste sobre el registro histórico, se garantiza que la información del censo que sirve como insumo para los indicadores como el PIB sub sector edificaciones, índice de precios de vivienda nueva IPEN y el índice de precios de edificaciones nuevas IPEN, poseen los estándares de calidad que requiere la investigación, porque se esta controlando la variabilidad.

8. BIBLIOGRAFÍA

1. Consulta del material metodológico de la investigación Censo de Edificaciones.
 - DANE. *ESPECIFICACIONES DE SISTEMAS*. Dirección Técnica de Metodología y Producción Estadística. Bogotá, D.C.
 - DANE. *FICHA METODOLÓGICA*. Dirección Técnica de Metodología y Producción Estadística. Bogotá, D.C.
 - DANE. *LA FORMACIÓN BRUTA DE CAPITAL FIJO Y EL CENSO DE EDIFICACIONES (APROXIMACIÓN EN SU MEDICIÓN)*, Dirección Técnica de Metodología y Producción Estadística. Bogotá, D.C.
 - DANE. *METODOLOGÍA DEL CENSO DE EDIFICACIONES*. Dirección Técnica de Metodología y Producción Estadística. Bogotá, D.C.
2. Revisión bibliográfica referente al tema de critica e imputación.
 - DANE. MURCIA CLAUDIA B, MALUENDAS ALEXIS V. (2001). DOCUMENTO PRELIMINAR: UNA METODOLOGIA DE IMPUTACIÓN PARA LA ENCUESTA CONTINUA DE HOGARES E INGRESOS Y GASTOS.
 - DANE. VILLAMIZAR, M. E.; BAUTISTA, L. (1997). ADAPTACION: *TECNICAS DE DISEÑO Y DESARROLLO DE ENCUESTAS TDDE, OPERATIVO DE OFICINA*, documento de apoyo docente.