


Departamento Administrativo Nacional de Estadística



Dirección de Metodología y Producción
Estadística - DIMPE

Especificaciones de imputación
Muestra Mensual de Hoteles - MMH

Junio 2010

	ESPECIFICACIONES DE IMPUTACIÓN MMH	CÓDIGO: DM-MMH-EIM-01 VERSIÓN: 01 PAGINA: 2 FECHA: 18-06-2010
Elaboró: Equipo de Diseños Muestrales	Revisó: Coordinador de Diseños Muestrales	Aprobó: Director DIMPE

1. INTRODUCCIÓN

Este documento expone los pasos que se deben seguir en el proceso de imputación de variables de la Muestra Mensual de Hoteles. Estas especificaciones son especialmente dirigidas a las personas encargadas de elaborar programas y generar cifras, de tal manera que se conserven los criterios aquí establecidos.

Una de las funciones del DANE es calcular índices, variaciones y porcentajes con la información recolectada en la muestra trimestral de servicios. Pero si una o varias empresas no reportan a tiempo la información se hace necesario generar la información faltante por algún método estadístico. A esta acción se le conoce como imputación de datos.

2. OBJETIVO

Dar a conocer los pasos que se deben seguir en el proceso de imputación de información de las variables de la Muestra Mensual de Hoteles.


3. JUSTIFICACIÓN

El proceso de imputación evita que los variaciones sean calculados con pérdida de información y el sesgo de no respuesta se incremente.

3. CRITERIOS

Las fuentes se pueden imputar cuando cumplan los siguientes criterios:

- La fuente debe contar con un histórico de información mínimo de 18 meses.


	ESPECIFICACIONES DE IMPUTACIÓN MMH	CÓDIGO: DM-MMH-EIM-01 VERSIÓN: 01 PAGINA: 3 FECHA: 18-06-2010
Elaboró: Equipo de Diseños Muestrales	Revisó: Coordinador de Diseños Muestrales	Aprobó: Director DIMPE

- Las empresas que tengan novedad 4 o 5 en el mes de estudio y poseen información en los meses anteriores pueden ser imputadas por el modelo, de lo contrario la imputación la realiza el equipo de temática.
- La cantidad de fuentes a imputar no debe superar el 5% del total.
- Se verifica la completitud de la base de datos que consiste en confrontar todos los registros de las variables como novedad, ingresos, personal y idnorem.
- Las empresas que a través del tiempo tienen siempre novedad deuda o sin localizar no se imputan y no entran a cuadros.

3. ESPECIFICACIONES DE IMPUTACIÓN

El equipo de logística genera una base en formato SAS para cada mes, la cual contiene toda la información de la muestra desde el mes julio de 2004 incluyendo variables de identificación y ubicación de la empresa como son: número de orden, actividad según código CIU, tipo de inclusión (probabilística o forzosa), novedad, año y mes. El equipo de metodología estadística realiza un procedimiento de verificación y consistencia a los datos enviados por logística antes de realizar las imputaciones. Los pasos a seguir para imputación son los siguientes:

1. Se realiza seguimiento a las variaciones atípicas de las variables en el mes del año en estudio con respecto al mismo mes del año anterior, y si existe justificación por parte de logística se mantienen dado que es el comportamiento real del sector. Si no existe explicación de la variación, el grupo de logística verifica la información recolectada. Se revisan que las empresas con novedad 3 y 1 no tengan información de ingresos, personal y salarios.
2. Se verifica que las empresas que deben tener información en las variables de estudio son las que tienen novedades 99 (informa normalmente), 97 (liquidada con período parcial de operaciones) y 98 (imputada). Las otras empresas tienen cero en las demás variables.

	ESPECIFICACIONES DE IMPUTACIÓN MMH	CÓDIGO: DM-MMH-EIM-01 VERSIÓN: 01 PAGINA: 4 FECHA: 18-06-2010
Elaboró: Equipo de Diseños Muestrales	Revisó: Coordinador de Diseños Muestrales	Aprobó: Director DIMPE

3. Las empresas que tienen novedad 4 o 5 en el mes de estudio y poseen información en los meses anteriores se seleccionan y pasan a ser imputadas.
4. Con las empresas que tiene información histórica se calculan las variaciones mensuales y anuales por actividad y por empresa. Esta operación se realiza para cada una de las variables de ingresos y de personal.
5. Teniendo las variaciones mensuales y anuales calculadas se aplica el modelo como lo indica el documento DM-MMH-DIM-01.doc para calcular las variaciones de los datos imputados como un promedio ponderado de las variaciones de las variables en las empresas con novedad deuda y sin localizar. El modelo utiliza dos variables independientes con sus respectivos parámetros. Las variables independientes son:
 - a. Variación mensual del año en estudio de la empresa.
 - b. Variación anual por empresa.

$$Var_{input_emp} = \alpha_1 V_{t_emp} + \alpha_2 V_{anual_emp}$$

6. Se imputa la variación del mes en estudio con el modelo, esta imputación se tiene por empresa y para cada una de las variables.
7. Se calculan las participaciones con la información del mes anterior y se construye una base con las variaciones imputadas, participaciones y la información de las variables a imputar de las empresas que están en deuda. Con esta base se imputan las variables de ingresos y personal.
8. A las empresas imputadas se les cambia la novedad a novedad 98.
9. Con las empresas imputadas se mira el histórico de cada variable cuatro meses anteriores al mes imputado y se analiza el comportamiento de la variable durante el tiempo.
10. El archivo que contiene la imputación de las variables personal e ingresos para el mes en estudio se exporta a un archivo tipo .XLS y se envía al equipo de logística el cual es utilizado para completar los registros que tienen información faltante y de esta forma completar la base de la Muestra Mensual de Hoteles.