

Departamento Administrativo Nacional de Estadística



Dirección de Metodología y Producción Es-
tadística

METODOLOGÍA DE IMPUTACIÓN MTS

JUNIO DE 2010

	METODOLOGÍA DE IMPUTACIÓN MTS	CÓDIGO: <i>DM-MTS-MET--01</i> VERSIÓN :01 Página: 2 Fecha: 23-06-2010
Elaboró: Equipo Diseño Muestrales	Revisó: Coordinador de Diseños Muestrales	Aprobó: Director DIMPE

CONTENIDO

I- MARCO CONCEPTUAL..... 3

II- METODOLOGÍA GENERAL DE IMPUTACIÓN DE DATOS FALTANTES..... 4

1. El Modelo..... 5

2. Utilización del Modelo..... 6

III- EJEMPLO..... 6

BIBLIOGRAFIA..... 9

	METODOLOGÍA DE IMPUTACIÓN MTS	CÓDIGO: <i>DM-MTS-MET--01</i> VERSIÓN :01 Página: 3 Fecha: 23-06-2010
Elaboró: Equipo Diseño Muestrales	Revisó: Coordinador de Diseños Muestrales	Aprobó: Director DIMPE

I- MARCO CONCEPTUAL.

Una gran cantidad de información, acerca de las características económicas tanto de individuos como de empresas o países, es recopilada con fines de análisis, para planear y tomar decisiones.

Al registro sistemático de mediciones u observaciones numéricas, efectuado a intervalos fijos de tiempo, se conoce como serie de tiempo y como la serie de tiempo se compone de datos numéricos, es común usar la estadística para describirla y analizarla, sea de forma descriptiva o de manera inferencial, cuyo procedimiento de esta última es la utilización de muestras, que representen a la población de estudio, para producir conclusiones válidas para toda la población.

Uno de los problemas que se presentan en el análisis estadístico inferencial es la falta de algunos registros en la serie, lo que conlleva a aumentar el error en la varianza de las estimaciones de los parámetros poblacionales; para hacer menos grave el error, se presentan dos métodos de estimación con datos faltantes que son la reponderación y la imputación.

La imputación es un método muy usado, en el cual se debe hacer el esfuerzo por imputar no más del 5% de los datos, si el porcentaje de datos imputados es muy alto se crea un error sistemático o sesgo en la varianza del estimador puntual. Pero aún si un método de imputación no produce un apreciable error, no se debe ignorar el efecto que la imputación tiene en la precisión de la varianza del estimador puntual.

La imputación es útil porque hace más viable el análisis de un conjunto de datos, asegurando consistencia en los resultados estadísticos obtenidos y reduciendo el sesgo de no respuesta.

Varias técnicas estadísticas requieren de conjuntos de datos rectangulares o en forma de matriz y que con presencia de datos faltantes, los registros se pueden restringir a un conjunto de datos completos. Esta restricción sacrifica información parcial en aquellas encuestas que no han sido diligenciadas totalmente y que se pueden utilizar o aprovechar si se hace imputación.

	METODOLOGÍA DE IMPUTACIÓN MTS	CÓDIGO: <i>DM-MTS-MET--01</i> VERSIÓN :01 Página: 4 Fecha: 23-06-2010
Elaboró: Equipo Diseño Muestrales	Revisó: Coordinador de Diseños Muestrales	Aprobó: Director DIMPE

En la literatura estadística hay una variedad de métodos que se han propuesto para imputar datos, estos métodos son clasificados según si se genera una sola imputación para cada valor faltante (imputación simple) o se generan, bajo simulaciones, m imputaciones para cada valor faltante el cual genera m conjuntos de datos completos (imputación múltiple).

Algunos métodos de imputación usan un modelo explícito como el de una regresión ajustada, una razón o la imputación por la media. En otros métodos el modelo es implícito como el de la imputación en paquete caliente (*hot deck*) y la imputación por donadores vecinos.

II- METODOLOGÍA GENERAL DE IMPUTACIÓN DE DATOS FALTANTES

En esta metodología se utiliza la información de la muestra trimestral de servicios, de tal manera que los datos imputados se aproximen a los valores reales. La metodología supone que los datos de la muestra poseen autocorrelación temporal y homogeneidad en las diferentes etapas de agregación; esto significa que la imputación debe estar de acuerdo al comportamiento de la serie histórica y de los niveles que contienen al dato faltante.

Para la imputación de registros en estado de deuda, se utiliza la razón de crecimiento de los datos, en la serie, o variación de los datos, definida como:

$$\text{Variación} = \frac{X_t}{X_{t-1}}$$

Donde

X_t = dato en el período t

X_{t-1} = dato en el período anterior $t-1$

Bajo estas consideraciones, se estimará primero la variación que tendrá el dato faltante con respecto al dato del período anterior, teniendo en cuenta el comportamiento histórico de la serie de variaciones

	METODOLOGÍA DE IMPUTACIÓN MTS	CÓDIGO: <i>DM-MTS-MET--01</i> VERSIÓN :01 Página: 5 Fecha: 23-06-2010
Elaboró: Equipo Diseño Muestrales	Revisó: Coordinador de Diseños Muestrales	Aprobó: Director DIMPE

en cada empresa y el comportamiento histórico de las variaciones, a partir de esta estimación se genera el dato faltante.

El cuadro-1 presenta un ejemplo, realizado para una empresa y que pertenece a la actividad 7491 (Actividades inmobiliarias, empresariales y de alquiler), de las variaciones calculadas para la variable de ingresos.

Cuadro 1. Variaciones de la variable ingresos

Periodo	Trimestres	Novedad	Variación	Ingresos
2008	4	99		3.658.916
2009	1	99	0,813	2.975.104
2009	2	99	1,092	3.248.940
2009	3	99	0,923	2.997.157
2009	4	99	1,023	3.065.047
2010	1	98	0,870	2.665.557

La variación del dato que se va a imputar se obtiene en términos de la variación histórica promedio en la empresa y en la actividad.

1. El Modelo.

El modelo para imputar la variación es:

$$Var_{imput_emp} = \alpha_1 V_{t_emp} + \alpha_2 V_{t_act} + \alpha_3 V_{anual_emp} + \alpha_4 V_{anual_act}$$

Donde,

V_{t_emp} = Variación trimestral del año en estudio de la empresa.

V_{t_act} = Variación trimestral del año en estudio de la actividad

V_{anual_emp} = Variación anual por empresa.

V_{anual_act} = Variación anual por actividad.

	METODOLOGÍA DE IMPUTACIÓN MTS	CÓDIGO: <i>DM-MTS-MET--01</i> VERSIÓN :01 Página: 6 Fecha: 23-06-2010
Elaboró: Equipo Diseño Muestrales	Revisó: Coordinador de Diseños Muestrales	Aprobó: Director DIMPE

β_i para $i = 1,2,3,4$ son coeficientes de ponderación.

Cuya suma debe ser igual a uno para que haya convergencia en la imputación.

El modelo describe la imputación de la variación del dato faltante, como un promedio ponderado de las variaciones de los datos en la empresa y en la actividad, donde los α_i son los coeficientes de ponderación de las variaciones. El propósito es estimar los parámetros desconocidos β_i , utilizando un método iterativo con el modelo de mínimos cuadrados.

2. Utilización del Modelo.

Se tendrá en cuenta los supuestos del modelo, los cuales son el de homogeneidad de los datos dentro de la actividad y la autocorrelación temporal entre las variaciones de los datos de la serie histórica dentro de cada empresa y dentro de cada actividad, supuestos que se utilizan en la estructura del modelo.

III- EJEMPLO.

Para explicar la metodología, se procederá a presentar un ejemplo de imputación de la variación y de la variable ingreso en el año 2010 primer trimestre para la empresa antes mencionada.

Utilizando la base de datos en la cual se encuentran los datos históricos, se calcularon las variaciones que ha tenido la variable por empresa y por actividad desde el cuarto trimestre del año 2008, luego se calculó el promedio de las variaciones dentro de la empresa y dentro de la actividad.

Empleando diferentes combinaciones de parámetros, con la restricción exigida que la suma sea igual a uno se procedió a imputar las variaciones y el total de la variable mencionada anteriormente.

	METODOLOGÍA DE IMPUTACIÓN MTS	CÓDIGO: DM-MTS-MET--01 VERSIÓN :01 Página: 7 Fecha: 23-06-2010
	Elaboró: Equipo Diseño Muestrales	Revisó: Coordinador de Diseños Muestrales

En el cuadro-2 se relacionan los datos reales del total y de las variaciones de la variable dentro de la empresa y de la actividad, datos que se quieren imputar para observar la bondad del ajuste del modelo.

Cuadro 2. Datos de los ingresos y de las variaciones reales para la variable total de ingresos

Año	Trimestre	Ingresos	Variación
2009	4	3.065.047	1,023

Las imputaciones obtenidas para la variación y para el total de la variable salario integral se presentan en los cuadros 3 y 4 respectivamente y se muestra en la gráfica-1 el dato real y la imputación más grande y la más pequeña del salario integral.

Cuadro 3. Imputaciones de la variación del ingreso para distintas combinaciones de parámetros

	Combinación de parámetros									
	$\alpha_1=0.25$	$\alpha_1=0.2$	$\alpha_1=0.1$	$\alpha_1=0.3$	$\alpha_1=0.4$	$\alpha_1=0.1$	$\alpha_1=0.1$	$\alpha_1=0.1$	$\alpha_1=0.2$	$\alpha_1=0.2$
	$\alpha_2=0.25$	$\alpha_2=0.3$	$\alpha_2=0.4$	$\alpha_2=0.2$	$\alpha_2=0.1$	$\alpha_2=0.5$	$\alpha_2=0.3$	$\alpha_2=0.6$	$\alpha_2=0.3$	$\alpha_2=0.3$
	$\alpha_3=0.25$	$\alpha_3=0.3$	$\alpha_3=0.4$	$\alpha_3=0.2$	$\alpha_3=0.1$	$\alpha_3=0.3$	$\alpha_3=0.5$	$\alpha_3=0.2$	$\alpha_3=0.3$	$\alpha_3=0.3$
	$\alpha_4=0.25$	$\alpha_4=0.2$	$\alpha_4=0.1$	$\alpha_4=0.3$	$\alpha_4=0.4$	$\alpha_4=0.1$	$\alpha_4=0.1$	$\alpha_4=0.1$	$\alpha_4=0.2$	$\alpha_4=0.2$
Variaciones imputadas	1,07	1,08	1,11	1,06	1,04	1,12	1,10	1,13	1,09	1,09

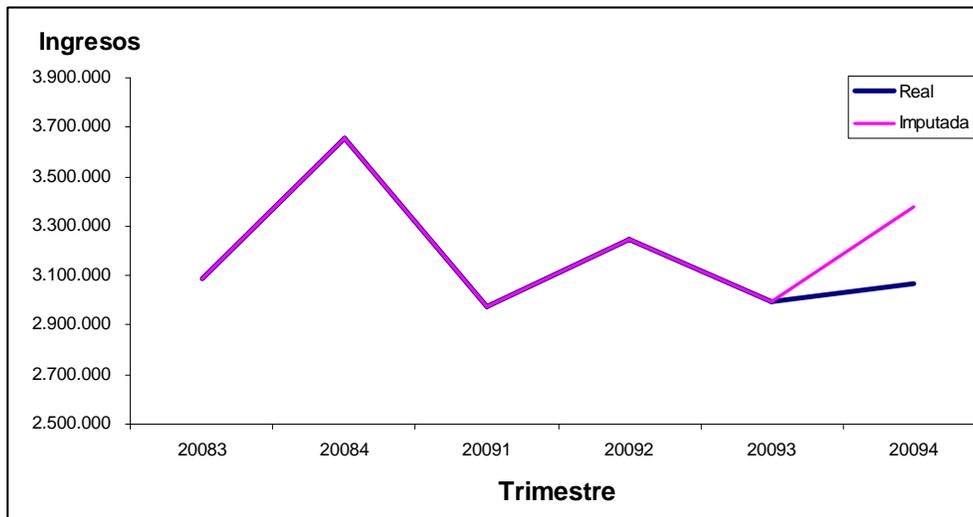
Cuadro 4. Imputaciones de ingresos para distintas combinaciones de parámetros

	Combinación de parámetros									
	$\alpha_1=0.25$	$\alpha_1=0.2$	$\alpha_1=0.1$	$\alpha_1=0.3$	$\alpha_1=0.4$	$\alpha_1=0.1$	$\alpha_1=0.1$	$\alpha_1=0.1$	$\alpha_1=0.2$	$\alpha_1=0.2$
	$\alpha_2=0.25$	$\alpha_2=0.3$	$\alpha_2=0.4$	$\alpha_2=0.2$	$\alpha_2=0.1$	$\alpha_2=0.5$	$\alpha_2=0.3$	$\alpha_2=0.6$	$\alpha_2=0.3$	$\alpha_2=0.3$

	METODOLOGÍA DE IMPUTACIÓN MTS	CÓDIGO: <i>DM-MTS-MET--01</i> VERSIÓN :01 Página: 8 Fecha: 23-06-2010
Elaboró: Equipo Diseño Muestrales	Revisó: Coordinador de Diseños Muestrales	Aprobó: Director DIMPE

	$\alpha_3=0.25$ $\alpha_4=0.25$	$\alpha_3=0.3$ $\alpha_4=0.2$	$\alpha_3=0.4$ $\alpha_4=0.1$	$\alpha_3=0.2$ $\alpha_4=0.3$	$\alpha_3=0.1$ $\alpha_4=0.4$	$\alpha_3=0.3$ $\alpha_4=0.1$	$\alpha_3=0.5$ $\alpha_4=0.1$	$\alpha_3=0.2$ $\alpha_4=0.1$	$\alpha_3=0.3$ $\alpha_4=0.2$
Totales Imputados	3.223.690	3.259.150	3.330.071	3.188.229	3.117.308	3.360.453	3.299.688	3.390.836	3.269.305

Gráfica 1. Imputaciones y dato real de los ingresos



	METODOLOGÍA DE IMPUTACIÓN MTS	CÓDIGO: <i>DM-MTS-MET--01</i> VERSIÓN :01 Página: 9 Fecha: 23-06-2010
Elaboró: Equipo Diseño Muestrales	Revisó: Coordinador de Diseños Muestrales	Aprobó: Director DIMPE

BIBLIOGRAFIA.

Roderick J.A. 1982. "Models for Nonresponse in Sample Surveys". Journal of the American Statistical Association.

Martin D. et. al. 1986. "Alternative Methods for CPS Income Imputation". Journal of the American Statistical Association.

Guerrero, V. 1991. "Análisis estadístico de series de tiempo económicas". México.

Lozano, A. 2000. "Estimación de novedades en estado de deuda". Dane.