

**IMPUTACIÓN DE INGRESOS DE HOGARES: LA EXPERIENCIA
DE LA ENCUESTA NACIONAL DE GASTOS DE LOS
HOGARES DE LA ARGENTINA**

**SAÚL KEIFMAN, GUILLERMO MANZANO,
CORINA RODRÍGUEZ Y ANA SILVIA VILER**

INDEC - Argentina

ÍNDICE

	<u>Página</u>
I. Introducción	425
II. Metodología	426
III. Resultados	428
Bibliografía	430

I. Introducción¹

Entre febrero de 1996 y marzo de 1997, el Instituto Nacional de Estadísticas y Censos llevó a cabo la Encuesta Nacional Urbana² de Gastos de los Hogares (ENGH). El Estado argentino condujo desde 1924 diversas encuestas sobre gastos e ingresos de los hogares. El antecedente más próximo de la ENGH es la Encuesta de Ingresos y Gastos realizada en 1985-1986 en el Area Metropolitana de Buenos Aires. La ENGH es la primera encuesta urbana de gastos realizada en el país que tuvo alcance nacional. Por su diseño muestral, la ENGH es representativa de las localidades de 5.000 y más habitantes y recoge la información brindada por poco más de 27.000 hogares.

La falta de respuesta a las preguntas de ingresos, es un problema común en las encuestas a hogares. El Cuadro 1 muestra las tasas de no respuesta a ingresos en la ENGH, por región y total del país. Frente a esta dificultad se pueden adoptar dos actitudes alternativas. La primera consiste en no hacer nada y presentar estimaciones de ingresos sólo para los hogares que hayan brindado información completa. La segunda consiste en imputar los ingresos de los hogares con no respuesta a ingresos y presentar los resultados de ingresos para todos los hogares.

Cuadro 1
Encuesta Nacional Urbana de Gastos de los Hogares
Tasas de no respuesta a ingresos por Región
(porcentaje de hogares no respondentes)

Región	Tasa de No Respuesta
Metropolitana del Gran Buenos Aires	27,0
Pampeana	17,1
Noroeste	8,3
Noreste	9,6
Cuyo	8,6
Patagonia	10,4
Total del País	14,7

Qué actitud conviene seguir dependerá de cuál sea el tipo de no respuesta predominante.³ La literatura reconoce la existencia de tres situaciones: a) no respuesta completamente aleatoria, b) no respuesta aleatoria y c) no respuesta no ignorable.

En el primer caso la probabilidad de no respuesta no estaría relacionada con ninguna característica del hogar, por lo tanto los hogares no respondentes de ingresos tendrían atributos similares a los de los hogares respondentes, y en consecuencia no sería necesario imputar.

En el segundo caso, la probabilidad de no respuesta estaría relacionada con características del hogar observadas en la encuesta, por ejemplo, nivel de gastos, nivel de educación de los miembros, etc.

¹ Los autores de este documento son integrantes del equipo de metodología y análisis de la ENGH. Dirección de Estudios de Ingresos y Gastos de los Hogares. INDEC. Argentina.

² Con una muestra de 114 ciudades que representa a 28 millones de habitantes, la ENGH cubre al 96% de la población urbana del país y el 84 por ciento de la población total.

³ Ver Little y Rubin (1987)

En consecuencia, los hogares no respondientes serían significativamente diferentes a los respondientes. Por lo tanto, los resultados basados en los datos brindados por estos últimos no serían representativos de la población y por ende sería necesario imputar.

En el caso de no respuesta no ignorable, la probabilidad de no respuesta estaría relacionada con el ingreso, aún cuando se controlara por otras características informadas por el hogar. También en este caso sería necesario imputar para obtener estimaciones de ingresos representativas de la población.

El Cuadro 2 presenta algunos indicadores que caracterizan a los grupos de hogares que respondieron y no respondieron ingresos en la ENGH. Del mismo se infiere que las diferencias entre estos dos grupos no son despreciables. En particular, se advierte que los valores de las variables correlacionadas positivamente (negativamente) con el nivel de ingreso de los hogares son más altas (bajas) para el grupo de no respondientes de ingresos. A título de ejemplo, el gasto medio de los hogares no respondientes es 30 por ciento mayor que el de los respondientes. Por lo tanto, la no imputación de los ingresos sesgaría significativamente hacia la baja los ingresos estimados por la encuesta, lo que motivó la decisión de explorar métodos de imputación de ingresos en la ENGH.

Cuadro 2
Encuesta Nacional Urbana de Gastos de los Hogares
Características de los hogares respondientes y no respondientes a ingresos
Total Nacional

Características	Respondientes	No Respondientes
Gasto medio de consumo (en pesos)	817	1.063
Porcentaje de jefes con nivel universitario completo	9,24	14,43
Porcentaje de jefes con cobertura médica privada	13,20	21,20
Porcentaje de jefes inactivos	28,28	22,17
Porcentaje de jefes ocupados patrones o empleadores	5,27	7,72
Porcentaje de jefes ocupados con calificación profesional	8,31	11,91
Porcentaje de jefes mujeres	26,01	22,15
Porcentaje de hogares en estratos más altos (1, 2 y 3)*	47,28	55,5
Cantidad promedio de perceptores de ingreso en el hogar	1,68	2,03

* Para el diseño de la muestra se establecieron 6 estratos.

La información del Cuadro 2 es consistente tanto con el caso de no respuesta aleatoria como con el de no respuesta no ignorable. En caso de no respuesta aleatoria es válido utilizar la información sobre las características de los hogares de la encuesta para imputar el ingreso de los no respondientes. Esto se podría hacer mediante dos procedimientos alternativos: modelos de regresión y *hot-deck*.

II. Metodología

La ENGH cuenta con información para imputar ingresos por el método de regresión por dos procedimientos diferentes: a) los gastos y otras características del hogar, b) las características socio-ocupacionales de los perceptores.

Por el primer procedimiento podrían estimarse los parámetros de funciones de gastos o consumo de los hogares en relación a los ingresos y las características demográficas del hogar (número de

miembros, edad del jefe, etc.) para los hogares que respondieron ingresos, y utilizar luego los parámetros así estimados y la información demográfica del hogar para imputar los ingresos totales de los hogares con ingresos no válidos. Este procedimiento se basa en la existencia de una relación estable, avalada por la teoría económica, entre gastos de consumo, ingresos y otras características de los hogares.⁴

El otro procedimiento de regresión consistiría en estimar ecuaciones de determinación de ingresos para trabajadores asalariados, cuenta propia y patrones (por separado)⁵, en función de las características educacionales, ocupacionales y demográficas de los individuos que respondieron ingresos para imputar luego los ingresos de los perceptores que no lo hicieron en base a la información sociodemográfica de los mismos. La agregación de los ingresos individuales así imputados generaría el ingreso del hogar.

En caso de no respuesta a los ingresos por trabajo, sería posible emplear ambos procedimientos, lo que plantea el problema de cómo combinarlos. Al respecto parecería conveniente comparar las imputaciones resultantes de cada método con los ingresos reales de los hogares que respondieron para tratar de determinar la manera óptima de combinar ambos.

La ventaja de los procedimientos de regresión es que gracias a los paquetes estadísticos disponibles no hay gran dificultad en probar distintas especificaciones e incluir un gran número de variables explicativas, particularmente, ‘dummies’. La desventaja del método de regresión es que los errores de medición de las variables explicativas pueden sesgar significativamente los parámetros de las ecuaciones a estimar y, por tanto, las imputaciones. Este tema se retoma más adelante.

El método *hot-deck* (el más cercano) también se aplica en el nivel de los perceptores y se basa en la búsqueda de un individuo respondente con características similares a las del perceptor cuyo ingreso hay que imputar, asignándole el ingreso del primero (respondente) al segundo (no respondente). Su ventaja radica en que no sufre del problema de sesgo por errores de medición. El problema de este método, sin embargo, es que no resulta tan sencillo de implementar como el de regresión, explicado anteriormente, debido a la falta de software específico y por lo engorroso que resulta determinar los grupos de individuos.

Cabe advertir que si la no respuesta fuera “no ignorable”, tanto los métodos de regresión como el *hot-deck* producirán imputaciones sesgadas. En la práctica es casi imposible determinar si nos encontramos en el caso de no respuesta aleatoria o no ignorable, si no se cuenta con información adicional proveniente de otras fuentes tales como reentrevistas a no respondentes, registros administrativos, etc. Los resultados que se informan en el próximo punto, suponen que la no respuesta en el caso de la ENGH es aleatoria.

La exploración de los procedimientos de imputación en la ENGH comenzó con el método de regresión para estimar ingresos totales de los hogares en función del gasto y otras características del hogar. En la práctica, hay dos maneras de utilizar la función consumo para imputar ingresos. La estimación directa de la función consumo utiliza el gasto del hogar como variable a explicar y el ingreso y demás características de los hogares como variables explicativas. Esto requiere de la inversión de la ecuación estimada a fin de imputar los ingresos de los no respondentes.

⁴ El concepto de función de consumo tiene una larga tradición en la teoría económica, que ha sido fundamentada, entre otros, por Keynes (1936), Friedman (1957), Modigliani y Brumberg (1954). Para una aplicación empírica en América Latina, véase Musgrove (1980).

⁵ Las ecuaciones de determinación de ingresos tienen su origen último en las “diferencias compensadoras” de Adam Smith, y su estimación empírica tuvo un fuerte impulso en las últimas décadas gracias a los trabajos de Becker y Mincer. Véase Berndt (1991) para un síntesis de esta literatura.

La estimación indirecta de la función de consumo consiste en utilizar el ingreso como variable a explicar y el gasto y demás características de los hogares, como variables explicativas. Ambos métodos sufren por el problema de los errores de medición en ingresos y gastos. En efecto, cuando se grafican los promedios decílicos de los ingresos reales contra los ingresos ajustados por regresión mostró un sesgo negativo (positivo) en la estimación de la ordenada al origen y positivo (negativo) en la estimación de la pendiente para el método directo (indirecto).

Luego de una serie de pruebas efectuadas para la Región Metropolitana del Gran Buenos Aires, se decidió utilizar el método indirecto de estimación de funciones de consumo con un ajuste del sesgo mencionado en una segunda etapa basado en una regresión de los promedios cuantílicos de ingresos reales contra los de los imputados (para aquellos hogares respondentes). Este procedimiento resultó exitoso en términos de ajuste de escalas decílicas de ingresos.

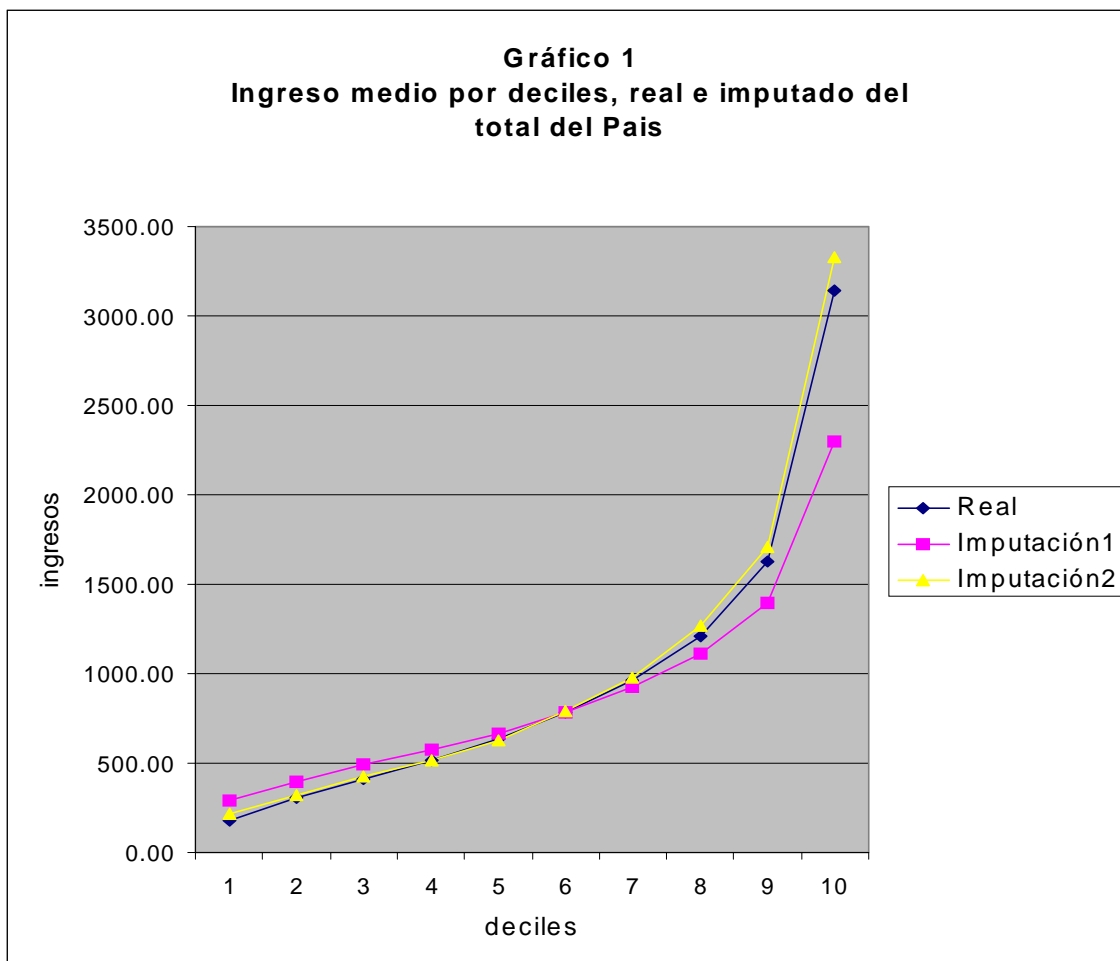
Este procedimiento se aplicó a nivel de región, subregión y provincia en algunos casos. En cada dominio geográfico se estimaron ecuaciones para ocho grupos de hogares clasificados según el régimen de tenencia de la vivienda (inquilinos y no inquilinos) y la condición de actividad y categoría ocupacional del jefe (asalariado, cuenta propia o patrón, desocupado e inactivo).

III. Resultados

El Cuadro 3 resume los promedios decílicos de los ingresos reales e imputados en la primera y segunda etapa de todas las regiones y grupos de hogares de la ENGH. Se aprecia el sesgo en los ingresos ajustados en la primera etapa (sobrestimación de los deciles bajos y subestimación de los deciles altos) y la práctica desaparición del sesgo con el ajuste de la segunda etapa (salvo en el primer decil). El Gráfico 1 visualiza la bondad del ajuste de la primera y segunda etapas.

Cuadro 3
Ingreso medio por deciles, real e imputado del total del País

Deciles	Real	Imputación1	Imputación2	Imputación1/Real	Imputación2/Real
1	177.41	294.38	214.82	65.93	21.09
2	304.91	396.64	318.98	30.09	4.62
3	410.02	492.44	423.98	20.10	3.40
4	514.69	574.02	518.12	11.53	0.67
5	635.89	662.22	629.61	4.14	-0.99
6	780.09	782.29	792.39	0.28	1.58
7	959.03	922.85	980.33	-3.77	2.22
8	1209.40	1112.70	1270.33	-8.00	5.04
9	1629.70	1397.59	1707.10	-14.24	4.75
10	3141.91	2300.10	3329.17	-26.79	5.96



El Cuadro 4 resume las estimaciones de ingresos de todo el país con imputación de los hogares no respondientes de acuerdo al procedimiento antes descrito. El ingreso estimado de los hogares no respondientes es en promedio 52 por ciento superior al de los hogares respondientes y, consecuentemente, el ingreso medio estimado de todos los hogares es 10 por ciento mayor que el que se estimaría solamente en base a los respondientes.

Cuadro 4
Resultados obtenidos a nivel nacional

Clasificación de los hogares	Ingreso medio (en pesos)	Cantidad de hogares
Hogares no imputados	990	6.591.695
Hogares imputados	1.508	1.565.574
Total de hogares	1.090	8.157.269

BIBLIOGRAFÍA

Berndt, E. (1991) *The Practice of Econometrics*, MIT Press.

Friedman, M. (1957) *A Theory of the Consumption Function*, Princeton University Press.

Keynes, J. M. (1936) *The General Theory of Money, Interest and Employment*, McMillan Press.

Little, R., y D. Rubin (1987) *Statistical Analysis with Missing Data*, John Wiley and Sons.

Modigliani, F. y R. Brumberg (1954) *Utility Analysis and the Consumption Function*, en K. Kurihara (comp.) *Post Keynesian Economics*, Rutgers University Press.

Musgrove, P. (1980) *Ingreso y consumo urbano en América Latina*, OEA.

