



Buenos Aires, 4 de diciembre de 2009

Encuesta Permanente de Hogares Programa de Actualización Metodológica

BASES DE MICRODATOS – NOVEDADES METODOLOGICAS

En las bases de microdatos de la Encuesta Permanente de Hogares (EPH) los usuarios encontrarán cambios en la presentación de los archivos, así como también algunos aspectos novedosos resultantes de la aplicación de innovaciones metodológicas desarrolladas por los equipos de trabajo de la Dirección de Metodología Estadística y la Dirección de Encuesta Permanente de Hogares.

Las definiciones de las categorías conceptuales no fueron modificadas. Tratándose en todos los casos de innovaciones en el tratamiento de la información recolectada, ellas resultan aplicables a la totalidad de la serie de datos producidos desde la instalación de la modalidad continua en la EPH. De manera tal que los usuarios disponen de la información procesada con las nuevas modalidades para la totalidad de la serie a partir del tercer trimestre de 2003, lo que permitirá empalmes y/o comparaciones por un período suficientemente prolongado.

A partir del tercer trimestre de 2009 los indicadores oficiales obtenidos de las bases de la EPH incorporarán estos cambios metodológicos. A continuación se presenta un breve resumen de las principales novedades. Para una cabal comprensión de los procedimientos realizados, se recomienda consultar los textos indicados en las notas de pie de página correspondientes a cada ítem.

1.- ARCHIVOS - Presentación

El INDEC ofrece las bases de microdatos en los formatos habituales (*Dbase IV*) e incorpora como novedad la presentación de bases en otros formatos. Los usuarios de los paquetes *SPSS* y *Stata* disponen a partir de ahora de archivos con extensión *.sav* o *.dta* respectivamente, con las variables y los valores correspondientes ya etiquetados.

En la sección *BASES DE MICRODATOS DE LA ENCUESTA PERMANENTE DE HOGARES - CONTINUA* los usuarios podrán descargar los archivos comprimidos, cuya

denominación indica tanto el período de recolección de los datos contenidos, como el formato de software en el que están grabados.

Las bases que contienen los registros de información a nivel de las personas se identifican como **Individual** (*Ind_ para la extensión .dbf*), seguido por la identificación del período y la extensión de software; por ejemplo: **Individual_t303.sav** indica que la información es obtenida principalmente del cuestionario individual, que corresponde al trimestre 3º del año 2003, y que está almacenada en el formato de software *SPSS*.

Las bases con información a nivel de los hogares y viviendas se identifican como **Hogar** (*Hog_ para la extensión .dbf*), también seguido por la identificación de período y extensión; por ejemplo **Hogar_t207.dta** indica que la información proviene de los cuestionarios de hogares y de viviendas, correspondientes al 2º trimestre de 2007, y está almacenada en el formato de *Stata*.

2. - INNOVACIONES EN EL TRATAMIENTO DE LA INFORMACIÓN

2.1.- Clasificación de actividades económicas¹ - Homogeneización de códigos

Para la clasificación de las actividades económicas o rama de actividad en las que las personas se desempeñan, la EPH en su modalidad continua utiliza el clasificador CAES-MERCOSUR.

Como consecuencia de distintos requerimientos se fueron incorporando nuevas aperturas o desagregaciones de actividades en diferentes períodos, de manera tal que la serie histórica no resulta homogénea.

En las bases de microdatos publicadas se aplican para todos los períodos los mismos criterios de clasificación, lo que implica que las desagregaciones más recientemente incorporadas fueron aplicadas a la información más antigua, y en toda la serie se dispone de la máxima apertura.

Para el tratamiento de la información se mejoró sustantivamente la codificación realizada en forma automática, duplicando el repertorio de frases clave, hasta sobrepasar

¹ Ver CLASIFICADOR DE ACTIVIDADES PARA ENCUESTAS SOCIODEMOGRÁFICAS CAES-MERCOSUR en www.indec.mecon.gov.ar

las 40.000. Las mejoras en la automatización reducen la discrecionalidad de las codificaciones manuales. En la actualidad alrededor de dos tercios de los registros obtienen su código de rama de actividad de manera automática.

2.2.- Calibración de los factores de expansión²

La variable de ponderación es calibrada de acuerdo a la estructura de edad y sexo de la población proyectada para la población de referencia de la encuesta. A los procedimientos habituales de ajuste de la ponderación (ponderación de diseño, corrección por no respuesta y ajuste a la proyección demográfica total) se agrega una calibración final según la proyección demográfica de proporciones de la población, según sexo y según tramos de edad.

El procedimiento apunta a eliminar o reducir las fluctuaciones originadas en el muestreo, ya que trimestre a trimestre -según la muestra de viviendas que ingresa al panel- cambia la estructura demográfica de la muestra. Esto puede provocar que, por ejemplo, si en la muestra ingresan muchos niños y adultos mayores se incremente el segmento de la población en edad teóricamente inactiva y se disminuya artificialmente la tasa de actividad y de empleo, o lo inverso si entran pocos niños y adultos mayores a la muestra. Este procedimiento de calibración es de mayor importancia aún en análisis de aquellos indicadores que tienen alta variabilidad entre los sexos (educativos por ejemplo), se reducen las variaciones debidas al muestreo en el análisis de sub-poblaciones y se mejoran las estimaciones en series históricas.

Asimismo se realizó el ajuste de toda la serie a las proyecciones de población definitivas del Censo 2001, de manera que cada base está ajustada a la proyección definitiva de población total según sexo y tramos de edad correspondiente al trimestre respectivo, incluyendo las de los años 2003 y 2004.

2.3. - Corrección de la no respuesta en las preguntas de ingreso

² Ver documento "Ponderación de la muestra y tratamiento de valores faltantes en las variables de ingreso en la Encuesta Permanente de Hogares" en www.indec.mecon.gov.ar

Desde el comienzo de su modalidad continua la EPH ha tenido como objetivo el desarrollo de estrategias de imputación de datos faltantes; entretanto ha corregido la no respuesta de ingreso mediante una reponderación.

Mediante esta técnica se creaba para cada una de las variables un nuevo factor de expansión:

Pondio: Ponderador para la variable p21, Ingreso de la Ocupación principal

Pondii: Ponderador para la variable p47T, Ingreso total Individual

Pondih: Ponderador para la variable ITF, ingreso total familiar

De este modo los respondentes obtenían una ponderación igual a la original aumentada (dentro de ciertas clases de reponderación), para corregir la no respuesta, mientras que a los no respondentes, se le asignaba una ponderación igual a cero en cada una de ellas.

Este método de corrección trae aparejado varios inconvenientes:

- El usuario de las bases debe trabajar con varios ponderadores (cuatro concretamente, uno para el factor de expansión original, uno asociado al ingreso de la ocupación principal, otro asociado al ingreso total individual y un tercero asociado al ingreso total familiar)
- Si se analiza una variable de ingreso, se debe trabajar con una base parcial de datos, con lo que en la práctica obliga a trabajar con 4 bases distintas.
- Se restringen las posibilidades de análisis multivariado.
- Aparecen inconsistencias entre los resultados de las bases de personas y las de hogares, como se puede observar en las tablas que siguen, extraídas del comunicado de distribución del ingreso del 2º trimestre de 2006.



Cuadro 4. Población ocupada según escala de ingreso de la ocupación principal. Total 28 aglomerados urbanos. Segundo Trimestre 2006

Decil	Escala de ingreso	Población		Ingreso		Ingreso medio	
		Miles	%	Miles de pesos	%	por decil	por estrato
	Pesos					Pesos	Pesos
1	7 - 160	939	10	105.292	1,2	112	
2	160 - 300	939	10	220.000	2,5	234	
3	300 - 450	938	10	346.039	4,0	369	
4	450 - 600	938	10	488.468	5,6	521	309
5	600 - 700	939	10	611.844	7,1	652	
6	700 - 850	938	10	740.185	8,5	789	
7	850 - 1.000	939	10	885.352	10,2	943	
8	1.000 - 1.279	938	10	1.044.846	12,0	1.114	874
9	1.280 - 1.800	938	10	1.428.970	16,5	1.523	
10	1.800 - 50.000	938	10	2.807.534	32,4	2.993	2.258
Ocupados con ingresos	///	9.384	95,3 ⁽¹⁾	8.678.530	100,0	925	925
Ocupados sin ingresos	///	467	4,7	///	///	///	///
Población ocupada	///	9.851	100,0	///	///	///	///

(¹) La suma del porcentaje de población ocupada por decil, corresponde al total de ocupados con ingreso

Cuadro 5. Hogares según escala de ingreso total familiar. Total 28 aglomerados urbanos. Segundo trimestre 2006

Decil	Escala de ingreso	Hogares		Población		Ingreso		Ingreso medio
		Miles	%	Miles	%	Miles de pesos	%	
	Pesos							Pesos
1	5 - 400	702	10	1.921	8,2	186.097	1,6	265
2	400 - 570	702	10	2.061	8,8	331.724	2,9	473
3	570 - 770	702	10	2.155	9,2	463.372	4,1	661
4	775 - 950	702	10	2.185	9,3	597.890	5,2	852
5	950 - 1.190	702	10	2.274	9,7	734.196	6,4	1.046
6	1.190 - 1.450	701	10	2.416	10,3	910.596	8,0	1.298
7	1.450 - 1.800	701	10	2.520	10,8	1.134.265	9,9	1.617
8	1.800 - 2.300	703	10	2.601	11,1	1.424.819	12,5	2.027
9	2.300 - 3.300	700	10	2.580	11,0	1.917.353	16,8	2.738
10	3.300 - 34.380	701	10	2.659	11,4	3.701.243	32,5	5.276
Hogares con ingresos	///	7.016	98,8 ⁽¹⁾	23.371	100,0	11.401.555	100,0	1.625
Hogares sin ingresos	///	86	1,2	181	///	///	///	///
Total	///	7.102	100,0	23.552	///	///	///	///

(¹) La suma del porcentaje de hogares por decil, corresponde al total de hogares con ingreso

El cuadro 4 expresa que la **persona** que mayor ingreso tiene en su ocupación principal percibe \$ 50.000.-

El cuadro 5 expresa que el **hogar** de mayor ingreso percibe \$ 34.380.-

Si la persona que más gana obtiene \$50.000.-, el hogar más rico no puede recibir \$34.380.- ya que los ingresos de los hogares no pueden ser inferiores a los de las personas que los componen.

Este problema se debe a que en el hogar de la persona que percibe \$50.000 hay por lo menos un **no respondente**. Por lo tanto, este hogar queda eliminado de la base que da origen a la escala de ingreso total familiar.

2.3.1.- Corrección por imputación³:

La estrategia más difundida de corrección de no respuesta parcial para la información de ingresos es la de la imputación de los valores faltantes. Esta técnica es la que aplica el INDEC en las bases de microdatos de la EPH, al igual que lo hacen las principales oficinas nacionales de estadística y los organismos internacionales.

El método de imputación de ingresos utilizado es el de hot-deck aleatorio que sintéticamente, consiste en seleccionar un 'donante' para cada valor faltante dentro de cierta sub-población definida por ciertas variables auxiliares, correlacionadas o asociadas con la variable cuya no respuesta se corrige. Ello significa que cada valor faltante es 'imputado' por un valor 'válido' que pertenece al mismo grupo de clasificación. La imputación se realiza sustituyendo el valor faltante por un valor válido en cada una de las variables.

El método adoptado ofrece como principales ventajas que :Se corrige la respuesta faltante con valores efectivamente observados, a diferencia de la imputación por regresión.

- Se preserva la distribución de la variable. Se asume que la distribución de los valores imputados es la misma que la de los valores observados dentro de cada clase de imputación.

³ Ver documento "Ponderación de la muestra y tratamiento de valores faltantes en las variables de ingreso en la Encuesta Permanente de Hogares" en www.indec.mecon.gov.ar .

Para los usuarios de las bases de microdatos comportará inmediatos beneficios:

- no será necesario alternar entre los cuatro ponderadores ni trabajar con bases parciales, lo que posibilitará nuevos análisis multivariados.
- no se presentarán inconsistencias entre estimaciones de las distintas sub-bases en relación a los ingresos de las personas y los hogares. En las nuevas bases de microdatos los usuarios disponen de un campo en el que se indica cuáles son los registros que contienen uno o más componentes de ingresos imputados y a cuáles celdas se imputó.

2.3.2.- Corrección de los agregados de ingresos para las partidas de periodicidad no mensual

El programa de actualización metodológica para la Encuesta Permanente de Hogares continúa trabajando sobre el tratamiento de la información relativa a ingresos de las personas y los hogares. Conforme a las prácticas usuales para las encuestas continuas de empleo, los ingresos relevados por la encuesta son aquellos efectivamente percibidos en el mes de referencia (mes anterior a la entrevista).

Ello trae aparejado que los resultados obtenidos se vean afectados por la inclusión de valores de ingresos no representativos del **ingreso personal disponible**. Partidas como el aguinaldo para los asalariados, los ingresos de profesionales independientes o por actividades del sector agropecuario, por ej., están sujetas a períodos de percepción distintos al mes, por lo tanto su inclusión comporta un efecto distorsivo para todas las distribuciones, de particular impacto tanto sobre las medidas de desigualdad como sobre las estimaciones de pobreza.

Un ejemplo de esto es el efecto del aguinaldo en el cálculo del Coeficiente de Gini que provoca una sobreestimación de la desigualdad en los trimestres 1º y 3º, y una subestimación de la misma en los trimestres 2º y 4º.

Los procedimientos de cálculo del **ingreso personal disponible** quedarán consolidados en los próximos meses, y sus resultados se incorporarán a las bases de microdatos semestrales.

Finalmente, asumiendo la total responsabilidad sobre los resultados finales, queremos agradecer a todos los expertos consultados que con gran generosidad

aportaron su conocimiento y experiencia para el mejoramiento de nuestra producción estadística: a nuestros colegas de la Dirección de Encuesta Permanente de Hogares, de la Dirección de Metodología Estadística, de la Dirección Nacional de Cuentas Nacionales y de la Dirección Nacional Estadísticas Sociales y de Población; a los titulares y profesionales de las Direcciones Provinciales de Estadística; a los expertos extranjeros Minezo Fujita, consultor de la Oficina de Estadística de Japón; Jacques Freyssinet, presidente del Comité Científico del Centre d' Etudes de l'Emploi de Francia; Dominique Goux, (chef de mission - animation de la recherche) de la Direction de l'Animation de la Recherche, des Etudes et des Statistiques (DARES, Ministerio de Trabajo de Francia); Giulio Barcaroli, Marco Di Zio, Gianluca Giuliani, Federica Pintaldi, Paolo Consolini y Silvano Vitaletti del ISTAT de Italia; Juan Carlos Feres, Fernando Medina, Marco Galván y Carlos Howes de la División de Estadística y Proyecciones Económicas de la CEPAL.

Ana María Edwin
Directora

Norberto Itzcovich
Director Técnico

Roberto Muiños
Director de Metodología Estadística

Claudio Comari
Director de Encuesta Permanente de Hogares

Augusto Hoszowski
Dirección de Metodología Estadística