

RAESTA 1 - AÑO 1 (2014)

EDITORIAL

Con este número 1 estamos incorporando, en el amplio espectro de las actividades que la UNTREF ofrece a la comunidad, nuestra Revista Argentina de Estadística Aplicada que proponemos e imaginamos como un espacio de debate científico y profesional acerca de la utilización, diseño y aplicación de técnicas estadísticas para la generación y el análisis de datos.

En este primer número de presentación y a modo de ratificación de su compromiso académico y profesional de nuestra Maestría en Generación y Análisis de Información Estadística, se presentan una serie de artículos producidos por sus graduados. Se trata de artículos que intentan resumir y/o ampliar el trabajo realizado en el transcurso de las investigaciones que los respectivos autores llevaron adelante en el marco del Plan de Estudio vigente.

Presentación de los artículos:

Florencia Sourrouille en su artículo “La pérdida de valor de las credenciales educativas en el mercado de trabajo argentino 1995-2001. Una respuesta desde los métodos estadísticos” aborda un problema clásico de las ciencias sociales: la pérdida de valor de los títulos educativos. Según la autora, tal proceso de desvalorización se centraría específicamente en los niveles educativos medios y altos (secundaria completa y superiores). Para poner a prueba su hipótesis construye pools de datos, donde evalúa la viabilidad de su utilización y realiza una serie de regresiones lineales para el cálculo de premios salariales.

El problema del abordaje de los docentes como colectivo laboral es tratado en “Los docentes en la Encuesta Permanente de Hogares. Notas metodológicas para su identificación y estudio” de Leandro Bottinelli. El artículo aborda específicamente el uso de fuentes secundarias para el estudio de este universo, donde analiza las fortalezas y debilidades de los Censos Nacionales Docentes. Entre sus debilidades se encuentra su periodicidad decenal y su dificultad para la comparación con otros grupos laborales. Por lo cual plantea el análisis de la Encuesta Permanente de Hogares (EPH) como potencial fuente para el análisis de los ocupados en la rama de enseñanza. Finaliza con una comparación de la EPH con los Censos Docentes y un análisis de las diversas variables ocupacionales que releva la encuesta para el estudio de los docentes.

El artículo de Luisa Iñigo “Especificación del modelo, datos faltantes y casos atípicos: la

“trastienda” de una investigación basada en el análisis de la varianza” se propone especificar y discutir una serie de aspectos metodológicos que dieron lugar a la tesis de la autora. Dado que en buena parte de la literatura sobre la relación entre el mercado de trabajo y la distribución personal del ingreso no suelen especificarse una serie de dimensiones metodológicas y técnicas que resultan relevantes, la autora se propone detallar de forma clara y minuciosa las decisiones tomadas en su propia investigación. Su objetivo es allanar el camino a quienes deban abordar por primera vez la información de ingresos personales relevada por la EPH mediante técnicas de descomposición de la varianza. Los aspectos que analiza el artículo serán, entonces, las decisiones tomadas en torno a la especificación del modelo, el tratamiento de los casos con información faltante, y la detección y tratamiento de los casos atípicos.

Finalmente, Ana María Capuano en “¿Qué son los Indicadores? Perspectivas y usos diferentes” presenta las diferentes perspectivas y usos de los “indicadores”. Desde las Ciencias Sociales, los indicadores han sido utilizados para la medición empírica de conceptos teóricos (operacionalización). Sin embargo, existen según la autora otros enfoques en la utilización de los indicadores: son también utilizados desde una perspectiva más pragmática, principalmente por los organismos de estadística y organismos internacionales. En el artículo se describen ambas perspectivas y se problematizan las diferencias en torno a los objetivos y a los métodos que utilizan.

Cabe remarcar que, en los números sucesivos, estos contenidos se presentarán como una sección específica dando lugar a otras secciones de Artículos, Documentos de Trabajo, Recensiones de Tesis, Reseñas bibliográficas, entre otras.

Esperamos contar con vuestro valioso aporte en este camino que comenzamos a recorrer.

ARTÍCULOS

La pérdida de valor de las credenciales educativas en el mercado de trabajo Argentino 1995-2001. Una respuesta desde los métodos estadísticos.¹

*Florencia Sourrouille*²

pool de datos

educación

La relación entre el nivel educativo de la población y su inserción en el mercado de trabajo es una temática que recorre extensamente la investigación social. Esta se aborda desde distintas perspectivas. En este artículo se propone indagar acerca del posible desajuste entre el nivel educativo de la fuerza de trabajo y el correspondiente a los requisitos de calificación de los puestos, lo cual podría estar manifestando un proceso de devaluación de las credenciales educativas.

La coexistencia durante la década de 1990 en el mercado de trabajo Argentino de un escenario macroeconómico propicio para que se pueda desarrollar un proceso de pérdida de valor de las credenciales educativas³ así como de un aumento de la desigualdad salarial hace que esa década, y en particular el período 1995 - 2001, tenga un interés particular para el análisis.

En este momento aparece, por un lado, un escenario favorable para que pueda existir un proceso de pérdida de valor de las credenciales educativas (conocido como devaluación educativa) dada la política de apertura de la economía con desempleo creciente e inestabilidad laboral (Altimir y Beccaria, 1999) junto a una oferta creciente de fuerza de trabajo con alto nivel educativo. Este hecho en particular estaría dando cuenta de una incapacidad del mercado de trabajo para absorber a la fuerza de trabajo más educada. Pero además se observa en ese mismo momento que las brechas de ingresos se amplían, es decir, existe otro proceso que es el de diferenciación de la fuerza de trabajo a partir de los salarios. Esto último pone en evidencia que no es toda la

fuerza de trabajo la que ve devaluada su inserción laboral (si fuera así se esperaría una disminución de las brechas salariales) sino que una parte ve devaluada su inserción, y otra la mejora, pudiéndose observar una segmentación del mercado de trabajo.

Así, este aumento de la desigualdad salarial plantea debates acerca de las implicancias que tiene en relación a los trabajadores más educados. Dado que a lo largo de la década de 1990 aumento del nivel educativo de la población en general y de la fuerza de trabajo en particular, pero que se registra un estancamiento de los puestos de trabajo profesionales, se discute si el mayor nivel educativo registrado en la fuerza de trabajo es requerido para el puesto o no. En el caso de no ser requerido, el trabajador se vería forzado a aceptar puestos por debajo de su nivel para evitar caer en el desempleo.

Este artículo se enmarca en la tesis de Maestría “La devaluación educativa en el mercado de trabajo Argentino 1995-2001. Una aproximación metodológica a partir de la técnica de pool de datos” elaborado por la autora y dirigida por Fernando Groisman en el marco de la Maestría en Generación y Análisis de Información Estadística (UNTREF-INDEC). A diferencia del trabajo de tesis donde se aborda la problemática planteada en el título de manera extensa, en el artículo que se presenta se trabaja solamente un aspecto específico de la problemática que es la discusión en torno a si el nivel educativo de la fuerza de trabajo es requerido o no para los puestos de trabajo en los cuáles se desempeñan desarrollando una metodología para poder dar cuenta de este fenómeno.

En primer lugar, se enuncian brevemente las distintas perspectivas con las cuáles se aborda la temática. Luego se presenta la hipótesis de trabajo y los aspectos metodológicos. En tercer lugar se desarrolla la metodología que se utiliza para el análisis. Dado que, como se observará en el desarrollo del problema, el universo de trabajo definido es pequeño y la fuente de datos que se utiliza es muestral, provocando esto que las estimaciones estén sujetas a errores muestrales grandes, se utiliza la técnica de pool de datos para resolver este problema. En este apartado se propone una metodología para poder utilizar esta técnica. En cuarto lugar se presentan los resultados y por último las conclusiones.

2. Perspectivas teóricas

Las perspectivas más sobresalientes que abordan esta temática, se pueden agrupar de la siguiente manera:

De un lado, las que dicen que el aumento del nivel educativo de la fuerza de trabajo se debe a que el mercado laboral requiere de una fuerza de trabajo más educada (González Rozada 1996, Gasparini 2003). Estos autores afirman que lo que se observa como un aumento del nivel educativo de la fuerza de trabajo se debe a que existen nuevos requerimientos del mercado laboral que provienen de cambios tecnológicos. Otros autores también, siguiendo una línea similar, sostienen que la complejización de las actividades laborales demanda de una escolarización más extensa

(Argumedo 1996, Gallart 1997). Lo que se sostiene es que existe un desajuste entre la educación formal y los cambios relacionados con la organización del trabajo. En este desajuste existe una responsabilidad del sistema educativo en no estar estrechamente articulado con el mercado de trabajo.

Una lectura diferente de la anterior es la que argumenta que el mercado de trabajo no tiene la capacidad de absorber a la fuerza de trabajo más educada, lo cual implica que exista una subutilización de la fuerza de trabajo por calificación, desaprovechándose las potencialidades de los trabajadores y dándose una pérdida de valor de las credenciales educativas.

Aquí lo que se discute es si el mercado de trabajo requiere de una fuerza de trabajo más educada dada la complejización de las actividades por cambios tecnológicos o si por el contrario hay una falta de capacidad del mercado laboral para absorber a la fuerza de trabajo más educada (Maurizio 2001, Salas 2005, Filmus 2001, Pérez 2005, Waisgrais 2005). Esto implicaría considerar que existe una subutilización de la fuerza de trabajo por calificación.

Por último Groisman (2003) propone que existe un proceso por el cual se dan dos fenómenos tiempo: la pérdida de valor de las credenciales educativas a la par que la segmentación del mercado de trabajo. Así, entre la fuerza de trabajo de mayor educación se verifica una diferenciación entre quienes logran poder acceder a puestos que están acordes a su calificación y quiénes no. En esta perspectiva se inscribe este estudio.

3. Hipótesis de trabajo y aspectos metodológicos

La hipótesis que guió la tesis de maestría fue que se habrían cubierto una cantidad relevante de puestos de trabajo con personas con calificaciones superiores a las necesarias que no impidió que progresara un proceso de heterogeneización salarial entre trabajadores del mismo nivel educativo.

La pregunta que se intenta responder en particular en este artículo es cuánto de la diferenciación salarial se debe a las distintas formaciones educativas de las personas y cuánto a las oportunidades del mercado de trabajo. ¿Cuáles son las variables que explican la diferenciación salarial? ¿Las características de formación de las personas medidas a partir de su nivel educativo, o las del puesto de trabajo? El nivel educativo de las personas ¿es requerido para el puesto de trabajo que ocupan?

Para esto se definen dos períodos de tiempo (1995 – 1998 y 1999 – 2011) en función de la evolución del PIB, sea ascendente o descendente. La fuente de datos que se utiliza es la Encuesta Permanente de Hogares (EPH) dado que se encarga de relevar la situación ocupacional y sociodemográfica en forma periódica en la Argentina.

El universo de trabajo que se utilizó en la tesis y sobre el que se reflexiona en este artículo está constituido por los asalariados plenos del sector privado de entre 25 y 55 años en aglomerados de

500.000 y más habitantes con nivel educativo secundario completo o mayor. Quedan exceptuados los trabajadores del sector estatal, la administración pública, educación, servicios sociales, de salud, comunales, servicio doméstico.

¿A qué se debe ese recorte? A que la forma de evidenciar si el nivel educativo de la fuerza de trabajo es requerido para el puesto o no se hizo a partir del cálculo de premios salariales. Este hecho vuelve relevante poder medir las diferencias en los ingresos que tengan que ver con el fenómeno que se está estudiando. Centrar el estudio en trabajadores en edad plenamente activa permite evitar introducir el sesgo de quiénes están recién empezando a construir su trayectoria laboral. Captar únicamente al sector privado se vuelve relevante dado que en este se fijan los salarios en forma distinta al sector público. Ceñir el análisis a los aglomerados de 500.000 habitantes y más permite evitar dificultades relacionadas con que en aglomerados pequeños pueden existir dinámicas distintas en relación a la capacidad de obtención de un puesto de trabajo acorde al del nivel educativo que se alcanzó.

Todas estas decisiones, consideradas necesarias para el estudio, provocaron que quede definido un universo pequeño de población y por lo tanto que las estimaciones resulten más imprecisas. Con el objetivo de poder utilizar los recursos disponibles pero a su vez realizarlo bajo el control de métodos estadísticos se utiliza la técnica de pool de datos. Esta consiste básicamente en unir datos originados en distintos relevamientos para sumar casos. El supuesto que avala realizar esta operación es que las variables no han tenido cambios significativos en el tiempo. En este trabajo se desarrolla una metodología para poder definir si las variables han tenido o no cambios significativos en el tiempo y poder decidir de esta manera en que casos se puede trabajar con pooles y en cuáles no, teniendo una medida de cuál es el efecto en los resultados de trabajar con pooles de datos.

En este artículo se presenta la metodología para la encuesta permanente de hogares puntual. En la tesis de maestría citada se presenta también la metodología para la encuesta continua. Igualmente la metodología que aquí se presenta puede ser utilizada para el trabajo con otras encuestas puntuales más allá de la EPH.

4. Desarrollo de la metodología de pooles de datos

Dado que utilizamos como fuente a una encuesta por muestreo (en nuestro caso la EPH), sus estimaciones están afectadas por el error muestral. Según la teoría del muestreo los estimadores tienden a seguir distribuciones normales y a partir del desvío standard se pueden calcular intervalos de confianza, los cuales representan la probabilidad de contener el valor verdadero que se quiere estimar. En el caso de nuestro estudio necesitamos construir intervalos de confianza para poder afirmar si dos estimaciones son significativamente distintas. Por otra parte, las recomendaciones en los manuales metodológicos de la EPH indican que las estimaciones cuyos coeficientes de variación mayores a 10% deben ser tratados con cautela.

Uno de los factores que intervienen en la dispersión y que por lo tanto están implicados en el cálculo de los estimadores y de los intervalos de confianza es la cantidad de casos de la muestra. A mayor dispersión, mayor amplitud del intervalo y será más difícil poder afirmar que dos estimaciones son significativamente distintas. Por lo tanto una estrategia para mejorar la precisión de las estimaciones es ampliar la cantidad de casos con la que trabajamos.

Para esto es que se propone trabajar con pooles de datos (Kish 1972, 1975). Esta técnica implica construir una base de datos (a partir de una onda original) que agregue registros de ondas subsiguientes con el objetivo de agrandar el tamaño de la muestra. Se basa en el supuesto que las variables utilizadas no tienen cambios significativos en su distribución en momentos cortos de tiempo. A continuación se desarrolla una metodología que permita conocer desde el punto de vista estadístico, si el supuesto de que los datos no tienen cambios significativos en el tiempo se da en la realidad o no, y de ser así, evaluar si es posible controlarlo.

En primer lugar, hay algunos criterios a cumplir y ajustes a realizar. Se evalúa la estructura de las muestras con el objetivo de determinar si se registran diferencias en las ondas con las que se quiere construir los pooles de datos (Mayo 1996 1998 – Mayo 1999 2001) en cuanto a su cobertura geográfica, el tamaño de la muestra y posibles problemas surgidos en el relevamiento.

En relación al primer punto se observa que los aglomerados de 500.000 y más habitantes (los definidos por nuestro universo) incorporados en la EPH en Mayo 96-98-99-01 se mantienen en todas las ondas.

Tampoco existen diferencias en el tamaño de la muestra. Si bien en el aglomerado de Gran La Plata se reduce la muestra el 50%, esto ocurre recién en 1999, por lo tanto para los pooles Mayo 96 98 y 99 01 el aglomerado tiene el mismo tamaño al interior de cada pool.

En Octubre de 1996 se descartó la onda de Gran Córdoba por mala calidad de los datos. En Agosto de 1998 no se relevó el aglomerado de Gran Mendoza. En esta misma onda surgieron dificultades en 5 áreas de los partidos del GBA que hicieron perder información. Al trabajar, esto llevó a utilizar las ondas de Mayo y no de Octubre.

En cuanto a la estructura de los ponderadores lo que se observa en las ponderaciones de los aglomerados con 500.000 habitantes y más para la población de entre 25 y 55 años es que en la muestra del aglomerado Gran Buenos Aires (GBA) cada caso representa a al menos 3 veces más de personas que en el resto de los aglomerados. Motivo por el cual los ponderadores deben ajustarse para que respeten esta distribución.

Cuadro N°1

Valores mínimo, máximo, promedio y desviación típica del ponderador según aglomerado

Aglomerado	Mínimo	Máximo	Media	Desviación tip.
GBA	828	1362	982	73
Gran La Plata	164	183	170	7
Gran Rosario	266	344	281	20
Gran Mendoza	220	220	220	0
Gran Córdoba	294	359	308	13
Gran Tucumán Tafi Viejo	238	266	243	8
Mar del Plata y Batán	235	3.182*	273	158

Fuente: elaboración propia a partir de la EPH.

Nota(*)1 caso. (5 registros / 736 están por encima del valor 700).

Propuesta de ajuste de los ponderadores

Una vez aumentado el tamaño muestral y obtenidas en consecuencia estimaciones más precisas se deben ajustar los ponderadores. Hay dos tipos de ajustes necesarios:

Por un lado, uno que mantenga el tamaño de muestra obtenido a partir de haber generado el pool de datos. Esto es necesario para mejorar la precisión de las estimaciones. Por otro lado, como estimaremos premios salariales a partir del uso de regresiones lineales múltiples, se necesita que ese ponderador no expanda los casos muestrales al total poblacional (dado que el tamaño de la muestra incide en los test de hipótesis que se realizan para evaluar si las afirmaciones son significativas desde el punto de vista estadístico o no). El último requisito es que respete el valor relativo de cada caso (los casos aportados por GBA representan a mayor población de los aportados por otros aglomerados). Estas necesidades quedan expresadas en el siguiente ponderador:

$$W_{media} = \frac{pondera}{\bar{x} pondera}$$

Al dividir al valor original del ponderador por su propia media, la suma de wmedia es igual al nuevo tamaño muestral.

En el siguiente ejemplo se observa cómo, suponiendo valores distintos de la variable original pondera, al dividir por su media se cumplen los requisitos enunciados: la suma de todos los casos divididos por su media vuelven a reconstruir el valor total obtenido por haber pooleado las dos muestras ($1/16+2/16+3/16+20/16+30/16+40/16= 96/16=6$), se mantienen los pesos relativos (2,50 es el doble que 1,25 así como 40 es el doble que 20), y no se expande al total poblacional.

Cuadro N°2

Ponderación corregida

N° caso	Valor pondera	pondera / \bar{x}
1	1	0,06
2	2	0,13
3	3	0,19
4	20	1,25
5	30	1,88
6	40	2,5
Total	\bar{x}	Σ
6	16	6

Fuente: elaboración propia.

Por otro lado, al momento de estimar un N, necesitamos realizarlo con un ponderador que expanda los datos al total poblacional. Con más casos se está estimando a la misma población, siendo esta una estimación más precisa. Por lo tanto hay que ajustar al ponderador en función del aumento de los casos muestrales. Este queda dado de la siguiente manera:

$$W_{pool} = pondera \cdot \left(\frac{n1}{npool} \right)$$

Donde

n1= cantidad de casos muestrales del primer relevamiento

npool= la cantidad de casos del pool resultante.

Como se observa, se achica el ponderador de forma proporcional a la cantidad de casos de cada onda. Si ambas ondas aportan los mismos casos será 0,5.

5. Pruebas de hipótesis acerca de la distribución de las variables.

Con estas se busca probar desde el punto de vista estadístico la afirmación de si las muestras fueron obtenidas de la misma población o de poblaciones con similares características. Así, si esta afirmación no puede ser rechazada, es un argumento fuerte para realizar el pool de datos. Esto porque implica que la distribución de las variables a utilizar es lo suficientemente similar para que se pueda afirmar que el hecho de haber sido relevados los datos en dos momentos distintos del tiempo no está incidiendo en los resultados.

Entonces para seleccionar la prueba de hipótesis que más se ajuste a nuestra necesidad partimos de los siguientes enunciados: tenemos 2 o más muestras, en este caso independientes, las cuáles queremos saber si provienen de poblaciones con la misma distribución.

La afirmación que se pretende testear (que las dos muestras fueron extraídas de poblaciones con la misma distribución) queda expresada de la siguiente manera:

H_0 : La distribución de las variables es =.

H_a : La distribución de las variables es distinta.

Nuestro objetivo es no rechazar la H_0 demostrando que las distribuciones no son significativamente diferentes. Esto implica una $\text{Sig} \rightarrow 0,05$.

Las pruebas no paramétricas que podrían utilizarse para esto son principalmente el Test U de Mann Whitney, la Z de KolmogorovSmirnov y la prueba de las rachas de Wald-Wolfowitz. Estas pruebas, a diferencia de la t para muestras independientes, no tienen algunos supuestos como por ejemplo el de distribución normal, varianzas iguales, o la necesidad de una distribución intervalar. Se pueden utilizar por lo tanto con variables categóricas.

Cuadro N°3

Pruebas no paramétricas seleccionadas según características

	Test U de Mann Whitmey	Z de KolmogorovSmirnov	Prueba de las rachas de Wald-Wolfowitz
Nivel de medición	Ordinal	Ordinal	Ordinal
Recomendaciones	Muestras grandes	Muestras chicas	Muestras grandes
Potencia*	95%	96%	No es buena
Forma de la prueba	Se ordenan los puntajes en una serie y se suma cuántas veces es precedido un valor de un grupo (el más grande) por el mismo en el otro.	La frecuencia acumulada de la misma variable en las dos distribuciones debería ser igual.	Ordena a las rachas de los dos grupos en una sola serie. Si son de la misma población deberían estar entremezclados. Encuentra diferencias de distintos tipos en las poblaciones.

Fuente: elaboración propia.

*Probabilidad de no rechazar la H_0 cuando es falsa.

La potencia de una prueba es la probabilidad de rechazar la H_0 cuando es falsa en la población. En nuestro caso entonces, es importante que la prueba sea potente. Dado que estamos trabajando con una muestra grande (EPH) y que la potencia de la prueba en nuestro caso es importante, seleccionamos el test U de Mann Whitney. Dado que esta prueba de hipótesis requiere un nivel de medición de las variables como mínimo ordinal, se construyeron variables estratificadas que articulan información de sexo, edad y nivel educativo alcanzado. Estas son:

- 1- Varones por edad continua (Edadvar)
- 2- Mujeres por edad continua (Edadmuj)
- 3- Varones por edad en tres intervalos (Vedad)
- 4- Mujeres por edad en tres intervalos (Medad)

- 5- Personas de entre 25 y 34 años por nivel en 6 tramos (E1niv6)
- 6- Personas de entre 35 y 44 años por nivel en 6 tramos (E2niv6)
- 7- Personas de entre 45 y 55 años por nivel en 6 tramos (E3niv6)
- 8- Personas de entre 25 y 34 años por nivel en 7 tramos (E1niv7)
- 9- Personas de entre 35 y 44 años por nivel en 7 tramos (E2niv7)
- 10- Personas de entre 45 y 55 años por nivel en 7 tramos (E3niv7)

Las categorías de nivel educativo alcanzado en 6 y 7 tramos son las siguientes:

Cuadro N°4

Nivel educativo alcanzado en tramos

6 tramos	7 tramos
Hasta primario incompleto	Hasta primario incompleto
Primario completo y medio incompleto	Primario completo y medio incompleto
Medio completo	Medio completo
SNU o Universitario incompleto	SNU incompleto
	SNU completo
SNU completo	Universitario incompleto
Universitario completo	Universitario completo

Fuente: elaboración propia.

Los resultados que arrojan son:

Cuadro N°5

Pruebas de Mann Whitney y niveles de significancia de la población de 25 a 55 años para el conjunto de los aglomerados de 500.000 y más habitantes, por pool según variables seleccionadas

Pruebas Whitney y sig25 a 55 años aglo 500	Pool Mayo 1996- Mayo 1998		Sig. asintót. (bilateral)	Pool Mayo 1999- Mayo 2001		Sig. asintót. (bilateral)
	U de Mann-Whitney	Z		U de Mann-Whitney	Z	
EDADVAR	16.158.517,50	-0,08	0,94	15.448.972,00	-1,83	0,07
EDADMUJ	19.061.370,50	-0,98	0,33	18.600.156,50	-1,32	0,19
VEDAD	16.157.951,00	-0,08	0,93	15.456.176,00	-1,9	0,06
MEDAD	19.236.111,00	-0,11	0,91	18.681.155,00	-0,97	0,33
E1NIV6	9.129.755,50	-0,74	0,46	9.242.838,50	-0,36	0,72
E2NIV6	7.265.323,50	-1,17	0,24	6.340.240,00	-1,94	0,05
E3NIV6	6.920.667,50	-1,29	0,2	7.235.324,00	-1,18	0,24
E1NIV7	9.111.917,50	-0,91	0,37	9.216.396,50	-0,6	0,55
E2NIV7	7.263.677,50	-1,18	0,24	6.341.475,00	-1,92	0,05
E3NIV7	6.912.387,50	-1,38	0,17	7.242.714,00	-1,1	0,27

Fuente: elaboración propia a partir de la EPH. Datos ponderados por wmedia.

Aquí se observa que el nivel de significancia de las pruebas de hipótesis acerca de la distribución

de las variables son iguales o mayores a 0,05, lo cual implica que no se puede rechazar la hipótesis de que las poblaciones de las cuáles fueron extraídas las muestras son iguales. Así, se pueden formar los pooles de datos ya que no se puede decir que las poblaciones sean distintas.

6. Test del efecto de la variable relevamiento de origen.

Analizado ya si la distribución de las variables demográficas de las dos muestras que queremos unir fue extraída de las mismas poblaciones, o de dos poblaciones similares nos centramos ahora en el problema de investigación que trabaja. El objetivo de estos controles es poder evaluar si el relevamiento de origen de los datos tiene o no una incidencia significativa ya directamente en la variable que queremos explicar. Esto implica que este test está directamente vinculado al problema específico que trabajamos. La forma de poder resolver esto es a partir del cálculo de regresiones lineales múltiples donde se incluya en el modelo explicativo a la variable “relevamiento de origen”. Se la incluye como dummy, con valores 0 en el relevamiento año t y 1 en el relevamiento t+1. Así se puede conocer el efecto de haber construido el pool, directamente en la variable cuyo comportamiento se pretende explicar. Si el aporte no es significativo, se puede trabajar perfectamente con el pool. En el caso de tener una incidencia significativa se pueden tomar dos alternativas: a) no utilizar el pool porque el efecto es significativo, b) utilizarlo, manteniendo siempre la variable que lo identifica en los modelos de regresión de manera de tener controlado cual es la incidencia de dicho aporte.

Una vez conformados los pooles de datos se vuelven a calcular los coeficientes de variación para cada nivel educativo utilizando los pooles de datos, y se los compara con los obtenidos sin realizar los pooles, se puede observar la ganancia obtenida en precisión.

Cuadro N°6

Coeficientes de variación del universo de trabajo por nivel educativo alcanzado según relevamiento

	Medio completo	SNU incompleto	SNU completo	Universitario incompleto	Universitario completo	SNU UnivInc	SNU UnivCom
Mayo 1996 (CV)	4,5	18,2	13,6	6,9	8,5	-	-
Mayo 1998 (CV)	4,9	14,2	12,3	5,7	7,8	-	-
Pool M9698 (CV)	3,3	11,0	9,0	4,3	5,7	4,0	4,8
Mayo 1999 (CV)	4,8	16,7	16,0	6,4	7,9	-	-
Mayo 2001 (CV)	4,7	16,0	12,6	6,5	7,2	-	-
Pool M9901 (CV)	3,3	11,6	9,9	4,5	5,3	4,2	4,7

Fuente: elaboración propia a partir de la EPH.

7. Desarrollo del problema de investigación a partir de los pooles de datos.

Habiendo realizado los pooles de datos podemos abordar ahora el problema de investigación.

Recordemos que la pregunta que se intenta responder es cuánto de la diferenciación salarial se debe a las distintas formaciones educativas de las personas y cuánto a las oportunidades del mercado de trabajo. Más específicamente: ¿Cuáles son las variables que explican la diferenciación salarial? ¿Las características de formación de las personas medidas a partir de su nivel educativo, o las del puesto de trabajo? El nivel educativo de las personas ¿es requerido para el puesto de trabajo que ocupan?

Utilizamos los pooles de datos elaborados para disminuir los errores de muestreo, incluyendo la variable que identifica el relevamiento del cual provienen los datos del pool (onda y año) en el modelo de regresión para, como se ha explicado, poder conocer su efecto. Seleccionamos el modelo que queda conformado por las variables sexo, edad, edad al cuadrado, calificación de la tarea, nivel educativo alcanzado, cantidad de horas de trabajo, si el trabajador se encuentra registrado o no, el tamaño del establecimiento y si está inserto laboralmente en el Gran Buenos Aires o en el resto de los aglomerados de 500.000 habitantes y más. En ambos períodos este modelo explica más del 50% de la varianza de los salarios y la introducción de todas las variables es significativa.

A partir del análisis de las contribuciones marginales, se observa que en general, las variables demográficas como el sexo y la edad (e inclusive la edad al cuadrado, incluida en el modelo para dar cuenta de un posible crecimiento geométrico de los salarios en función de la edad más que de un comportamiento lineal) tienen una escasa contribución marginal al modelo. Las variables con mayor contribución son las que están relacionadas con la calificación de la tarea, el nivel educativo alcanzado, y el hecho de pertenecer al aglomerado Gran Buenos Aires.

Se observa que excluyendo del análisis al nivel educativo y manteniendo al resto de las variables en el modelo, el aporte marginal varía entre un 3 y un 4% en el período 1996-1998 y 1999-2001 respectivamente. Este cambio es más profundo cuando lo que se excluye únicamente es la calificación de la tarea, y se incorporan en el modelo el resto de las variables (entre un 11 y un 10%). Algo similar ocurre con otra característica de la oferta de trabajo (el tamaño del establecimiento) que tiene un aporte mayor al del nivel educativo (4 y 6% respectivamente). El hecho de que sean las variables que caracterizan al puesto de trabajo las que tienen mayor contribución podría alentar la hipótesis que las oportunidades están siendo marcadas por el mercado de trabajo y es sobre estas oportunidades que se integran las personas con su nivel educativo.

Por último la contribución marginal de la variable que identifica al relevamiento de origen de los datos es menor al 1%.

Cuadro N°7

Coeficiente de correlación múltiple r^2 en regresiones lineales por MC sobre el logaritmo de los

salarios horarios para el universo de trabajo por pool de datos. Modelo seleccionado y contribuciones marginales.

	Pool M96 M98	Pool M99 M01
Modelo seleccionado		
Sexo, edad, edad ² , calificación, nivel, horas, registrados, tamaño, GBA o resto aglomerados 500.000 hab, onda y año del relevamiento.	0,53	0,55
Contribución combinada de 9 variables		
Edad, edad ² , calificación, nivel, horas, registrados, tamaño, GBA o resto aglomerados 500.000 hab, onda y año del relevamiento.	0,52	0,53
<i>Aporte marginal de la variable excluida sexo</i>	0,01	0,02
Sexo, edad, calificación, nivel, horas, registrados, tamaño, GBA o resto aglomerados 500.000 hab, onda y año del relevamiento.	0,52	0,54
<i>Aporte marginal de la variable excluida edad</i>	0,01	0,01
Sexo, edad, calificación, nivel, horas, registrados, tamaño, GBA o resto aglomerados 500.000 hab, onda y año del relevamiento.	0,52	0,54
<i>Aporte marginal de la variable excluida edad²</i>	0,01	0,00
Sexo, edad, edad ² , nivel, horas, registrados, tamaño, GBA o resto aglomerados 500.000 hab, onda y año del relevamiento.	0,42	0,45
<i>Aporte marginal de la variable excluida calificación</i>	0,11	0,10
Sexo, edad, edad ² , calificación, horas, registrados, tamaño, GBA o resto aglomerados 500.000 hab, onda y año del relevamiento.	0,50	0,51
<i>Aporte marginal de la variable excluida nivel</i>	0,03	0,04
Sexo, edad, edad ² , calificación, nivel, registrados, tamaño, GBA o resto aglomerados 500.000 hab, onda y año del relevamiento.	0,50	0,51
<i>Aporte marginal de la variable excluida horas</i>	0,02	0,04
Sexo, edad, edad ² , calificación, nivel, horas, tamaño, GBA o resto aglomerados 500.000 hab, onda y año del relevamiento.	0,52	0,54
<i>Aporte marginal de la variable excluida registrados</i>	0,01	0,01
Sexo, edad, edad ² , calificación, nivel, horas, registrados, GBA o resto aglomerados 500.000 hab, onda y año del relevamiento.	0,49	0,49
<i>Aporte marginal de la variable excluida tamaño</i>	0,04	0,06
Sexo, edad, edad ² , calificación, nivel, horas, registrados, tamaño, onda y año del relevamiento.	0,50	0,51
<i>Aporte marginal de la variable excluida GBA o resto aglomerados 500.000 hab</i>	0,03	0,03
Sexo, edad, edad ² , calificación, nivel, horas, registrados, tamaño, GBA o resto aglomerados 500.000 hab.	0,52	0,54
<i>Aporte marginal de la variable excluida onda y año del relevamiento</i>	0,007	0,006

Fuente: elaboración propia a partir de la EPH. Datos ponderados por Wmedia.

Analizamos a continuación los premios salariales de las variables nivel educativo, la calificación de la tarea y el tamaño del establecimiento dado que son las que tienen las mayores contribuciones al modelo. Mediante el cálculo de coeficientes de regresión, se puede estimar cuanto es en promedio el salario que se agrega a una determinada categoría de trabajadores por sobre el resto. En el caso que estudiamos el indicador es cuánto distintas categorías de nivel educativo aportan al salario. Si trabajadores con características homogéneas en cuanto al sexo, la edad y su inserción en el mercado de trabajo el hecho de variar su nivel educativo no tiene ninguna incidencia en sus salarios, este nivel educativo del trabajador no se considera necesario para el puesto. En cambio, si el nivel educativo es remunerado, se lo considera requerido para el puesto.

Observamos que el premio a la calificación profesional es mayor al del nivel educativo a lo largo de todo el período. El primero, se encuentra por encima del 80%, mientras que el premio al nivel

universitario completo, en cambio, nunca llega al 30%. Es la calificación del propio puesto de trabajo la que tiene un efecto mayor en relación a la formación de las personas. Por otro lado se observan algunas diferencias en el comportamiento de las variables:

En el caso de la educación formal, se observa que es el premio a tener el nivel universitario completo el que es significativamente distinto al premio de los trabajadores que tuvieron estudios terciarios no universitarios (más allá de que los hayan finalizado o no) o estudios universitarios incompletos. Sin embargo ninguno de estos últimos podemos decir que tengan premios salariales significativamente distintos a los de quienes han concluido los estudios medios. Al calcular a 2 desvíos estándares los límites inferior y superior de las estimaciones los intervalos se superponen o la estimación no es significativa. Solamente el haber concluidos los estudios universitarios establece una diferenciación salarial.

En cambio, las distintas categorías de calificación si introducen cambios. Se puede observar un aumento progresivo de los salarios en forma de gradiente, a medida que aumenta la calificación del puesto de trabajo. Como se observa en la información del cuadro 4 el premio a la calificación operativa, los salarios son en promedio un 26 y 22% mayores, respecto a la calificación técnica, un 60% y en relación a la calificación profesional, el premio es de un 80%.

Cuadro N°8

Premios salariales horarios, error en la estimación, límites inferior y superior y nivel de significancia para el universo de trabajo por pool de datos. Modelo 3 variables.

	Pool Mayo 1996 Mayo 1998					Pool Mayo 1999 Mayo 2001				
	B ⁵	ET ⁶	LI ⁷	LS ⁸	Sign.	B	ET	LI	LS	Sign.
(Constante)	,842	,031	0,78	0,90	,000	,892	,028	0,84	0,95	,000
Dummy calificación operativa	,260	,031	0,20	0,32	,000	,225	,029	0,17	0,28	,000
Dummy calificación técnica	,605	,032	0,54	0,67	,000	,560	,031	0,50	0,62	,000
Dummy calificación profesional	,841	,038	0,77	0,92	,000	,803	,036	0,73	0,87	,000
Dummy de nivel superior o univinc	,092	,021	0,05	0,13	,000	,061	,022	0,02	0,10	,005
Dummy de nivel superior completo	-,027	,034	-0,09	0,04	,434	-,020	,033	-0,08	0,04	,542
Dummy de nivel univ completo	,259	,028	0,21	0,31	,000	,273	,026	0,22	0,32	,000
Dummy 6 a 25 trabajadores	,068	,024	0,02	0,12	,004	,062	,024	0,01	0,11	,010
Dummy 26 a 100 trabajadores	,167	,025	0,12	0,22	,000	,117	,025	0,07	0,17	,000
Dummy más de 100 trabajadores	,187	,022	0,14	0,23	,000	,253	,022	0,21	0,30	,000
Dummy onda y año	,030	,017	0,00	0,06	,074	-,026	,017	-0,06	0,01	,121

Fuente: elaboración propia a partir de la EPH. Datos ponderados por Wmedia.

⁵ Coeficiente beta (representa el premio salarial): Da cuenta del incremento de la variable dependiente respecto al cambio en la variable independiente. Indica cuanto aumenta en promedio el salario ante un cambio en cada una de las variables explicativas incluidas en el modelo. En el cuadro 7 las variables independientes son tipo dummy: asumen el valor de 1 cuando se cumple la condición enunciada en la misma y 0 en caso contrario. Por lo tanto, los coeficientes deben leerse como diferencias respecto de las categorías base de las variables: en el caso de la calificación el valor 0 corresponde a la categoría no calificado, en el caso del nivel educativo al nivel medio

completo, y en el caso del tamaño del establecimiento los establecimientos de hasta 5 empleados.

⁶ Error de la estimación.

⁷ Límite inferior calculado a 2 desvíos estándares.

⁸ Límite superior calculado a 2 desvíos estándares.

Como ya se ha dicho, la estimación surgida del análisis de regresión es una estimación de parámetros poblacionales dado una muestra de n observaciones. Lo que interesa demostrar es que el premio salarial dado a una categoría de trabajadores es significativamente distinto al de otra categoría. Para esto debemos poder probar que las estimaciones son distintas para un determinado nivel de confianza, siendo entonces distintas no sólo las estimaciones puntuales sino que no se superpongan las estimaciones contenidas entre el límite inferior y superior del intervalo de confianza.

La ganancia que se obtiene utilizando el pool de datos queda demostrada en el cuadro que se presenta a continuación: las estimaciones de los premios salariales obtenidas con y sin pooles de datos son distintas. En primer lugar, difieren en cuanto a los errores de la estimación. Estos se achican cuando se trabaja con pooles respecto a lo que ocurre cuando se trabaja con una sola onda de la encuesta. Además, podemos apreciar que si hubiéramos trabajado sin pooles de datos no podríamos afirmar desde el punto de vista estadístico que los premios a la calificación técnica y profesional fueran significativamente distintos, dado que al comparar las estimaciones con intervalos del 95% de confianza, los límites entre ellas se superponen en ambas ondas: la de 1999 y la de 2001. Sin embargo, al utilizar el pool de datos los intervalos dejan de superponerse y es posible afirmar que son significativamente distintos. Como por otra parte hemos incorporado al modelo la variable que identifica al origen de los datos (la onda y año del relevamiento, podemos observar que el hecho de haber unido observaciones de distintos momentos del tiempo no tiene un efecto significativo (0,12) lo cual avala el uso del pool de datos.

Cuadro N°9

Premios salariales, error en la estimación, límites inferior y superior y nivel de significancia de la calificación de la tarea y el tamaño del establecimiento para el universo de trabajo por año y pool según contribución de variables seleccionadas.

	Mayo de 1999					Mayo de 2001				
	B	ET	LI	LS	Sig.	B	ET	LI	LS	Sig.
(Constante)	0,7	0,05	0,60	0,80	0,00	0,78	0,05	0,68	0,88	0,00
Dummy calificación operativa	0,3	0,05	0,20	0,40	0,00	0,22	0,05	0,12	0,32	0,00
Dummy calificación técnica	0,71	0,06	0,59	0,83	0,00	0,57	0,06	0,45	0,69	0,00
Dummy calificación profesional	0,96	0,08	0,80	1,12	0,00	0,8	0,07	0,66	0,94	0,00
Dummy de nivel superior o univinc	0,09	0,04	0,01	0,17	0,02	0	0,04	-0,08	0,08	0,99
Dummy de nivel superior completo	0,19	0,08	0,03	0,35	0,02	0,18	0,07	0,04	0,32	0,01
Dummy de nivel univ completo	0,28	0,05	0,18	0,38	0,00	0,33	0,05	0,23	0,43	0,00
Dummy6 a 25 trabajadores	0,13	0,05	0,03	0,23	0,01	0,11	0,05	0,01	0,21	0,01
Dummy26 a 100 trabajadores	0,23	0,05	0,13	0,33	0,00	0,25	0,05	0,15	0,35	0,00
Dummy más de 100 trabajadores	0,43	0,05	0,33	0,53	0,00	0,41	0,05	0,31	0,51	0,00
Dummy onda y año del relevamiento	-	-	-	-	-	-	-	-	-	-

Fuente: elaboración propia a partir de la EPH. Datos ponderados por Wmedia.

Cuadro N°10 (Continuación)

	Pool Mayo 1999 2001				
	B	ET	LI	LS	Sign.
(Constante)	0,89	0,03	0,84	0,95	0,00
Dummy calificación operativa	0,22	0,03	0,17	0,28	0,00
Dummy calificación técnica	0,56	0,03	0,50	0,62	0,00
Dummy calificación profesional	0,80	0,04	0,73	0,87	0,00
Dummy de nivel superior o univinc	0,06	0,02	0,02	0,10	0,00
Dummy de nivel superior completo	-0,02	0,03	-0,08	0,04	0,54
Dummy de nivel univ completo	0,27	0,03	0,22	0,32	0,00
Dummy6 a 25 trabajadores	0,06	0,02	0,01	0,11	0,01
Dummy26 a 100 trabajadores	0,12	0,02	0,07	0,17	0,00
Dummy más de 100 trabajadores	0,25	0,02	0,21	0,30	0,00
Dummy onda y año del relevamiento	-0,03	0,02	-0,06	0,01	0,12

Fuente: elaboración propia a partir de la EPH. Datos ponderados por Wmedia.

Por último, si tomamos por un lado a quiénes solo han concluido sus estudios medios y por el otro a quiénes han concluido los estudios universitarios podemos observar que aumentando el nivel educativo las características del puesto de trabajo discriminan más la hora de explicar los salarios. Esta evidencia también permite alentar la hipótesis que, existiendo una importante segmentación de las oportunidades laborales debido a las características de la oferta laboral y no de la demanda, en los sectores de mayor educación deben encontrarse diferencias más importantes.

Cuadro N°11 Coeficiente de correlación múltiple R cuadrado en regresiones lineales por mínimos

cuadrados sobre el logaritmo de los salarios horarios para el universo de trabajo por pool según niveles educativos seleccionados

	Pool M96-98	Pool M99-01
MC	0,25	0,26
Tamaño del establecimiento	0,11	0,09
Excluyendo la variable calificación de la tarea	0,14	0,17
Calificación de la tarea	0,18	0,20
Excluyendo la variable tamaño del establecimiento	0,07	0,06
UC	0,33	0,32
Tamaño del establecimiento	0,19	0,19
Excluyendo la variable calificación de la tarea	0,14	0,13
Calificación de la tarea	0,19	0,18
Excluyendo la variable tamaño del establecimiento	0,14	0,15

Fuente: elaboración propia a partir de la EPH. Datos ponderados por Wmedia.

8. Conclusiones

La evidencia encontrada en este breve estudio permite hipotetizar que, ante un escenario de contracción laboral y ampliación del acceso a la educación se puede observar un proceso de dos aristas. Por un lado, desde el punto de vista laboral, la exacerbación de la competencia de trabajadores, lo cual hace bajar los salarios generando un proceso de pérdida de valor de las credenciales educativas. Por el otro, el aumento de la diferenciación entre trabajadores del mismo nivel educativo sugiere la existencia de segmentación laboral.

En este estudio se evalúa cuanto de la diferenciación salarial de los trabajadores de alto nivel educativo se relaciona con las características de la formación de las personas, medidas a partir de su nivel educativo, y cuánto las características de los puestos de trabajo (medidos centralmente a partir de la calificación del puesto de trabajo y el tamaño del establecimiento), con el objetivo de contestar a la pregunta si el nivel educativo de la fuerza de trabajo es requerido o no para el puesto en el cual se desempeñan.

La calificación de la tarea caracteriza a los puestos de trabajo disponibles, siendo este un atributo de la oferta de trabajo (indicador de cuál es la estructura de oportunidades fijada por el mercado de trabajo). Por otro lado, el nivel educativo alcanzado está caracterizando a las personas que demandan trabajo. La brecha en términos de premios salariales entre la calificación de la tarea y el nivel educativo alcanzado da cuenta en nuestra opinión de la diferencia entre las capacidades de las personas y los requisitos del puesto de trabajo. Da cuenta por lo tanto de la subutilización de la fuerza de trabajo y da indicios de un proceso de devaluación de las credenciales educativas debido a que estaría poniendo de manifiesto la existencia de grupos de trabajadores cuyo nivel educativo no parece ser relevante para el puesto de trabajo.

Para poder contestar a esta pregunta existió en un primer momento un inconveniente: al quedar

definido un universo pequeño de la población los errores de muestreo eran grandes lo cual no permitía hacer afirmaciones significativas desde el punto de vista estadístico acerca de la pregunta de investigación. De un lado entonces, existía la preocupación acerca de un problema de la realidad Argentina, del otro, la necesidad de utilizar fuentes secundarias y periódicas para analizar la dinámica del mercado de trabajo, pero afectadas, al tratarse de un subuniverso pequeño de la población, por el error muestral. La necesidad del uso de fuentes secundarias, o bien, la imposibilidad de generar relevamientos específicos, ponen al investigador en el problema de cómo predicar sobre problemas con datos que no fueron diseñados para tal fin.

El desarrollo de la metodología de los pooles de datos permite contestar a una pregunta aprovechando la información disponible y realizándolo bajo el control de métodos estadísticos.

Anexo. Variables incluidas en el modelo de regresión.

Trabajamos con el logaritmo del salario horario de la ocupación principal de los asalariados plenos (entre 35 y 60 horas semanales). Para construir la variable se tomaron las siguientes decisiones:

- Se eliminaron los días trabajados mayores a 31.

- Se eliminaron los ingresos

- Se eliminaron los casos extremos del ingreso mensual correspondientes al universo de trabajo: ocupados asalariados plenos con nivel educativo secundario completo o más, de entre 25 y 55 años, de las ramas 2 a 15 en del sector privado de aglomerados con 500.000 hab y más.

- Se consideraron casos extremos al 1 % más bajo y más alto de la distribución.

- Se construyó el salario horario.

- Se deflactó por el IPC.

- Se construyó la variable logaritmo natural del salario horario.

Se trabaja con las siguientes variables predictoras:

Sexo
Varón: dsexo
Mujer: constante
Edad continua.
Edad al cuadrado.
Calificación de la tarea
Calificación profesional
Calificación técnica
Calificación operativa
No calificados: constante
Nivel educativo
Secundario completo: constante.
SNU incompleto – Universitario incompleto
SNU completo
Universitario completo
Tamaño del establecimiento
1 a 5: constante.
6 a 25
26 a 100
Más de 100
Registrados o no registrados
Registrados
No registrados: constante
Cantidad de horas de trabajo continua.
Región:
Noroeste: constante.
GBA
Cuyo
Pampeana
Relevamiento de origen
Relevamiento t: constante
Relevamiento t+1

Bibliografía

- ARGUMEDO, A. (1996): “El imperio del conocimiento. Impacto de la revolución científico-técnica” en Encrucijadas N°4, UBA, Buenos Aires.
- BECCARIA, L, ALTIMIR, O. (1999): “El mercado de trabajo bajo el nuevo régimen económico en Argentina”, Serie reformas económicas N° 28, CEPAL.
- FILMUS, D. (2001): “Cada vez más necesaria, cada vez más insuficiente. Escuela media y mercado de trabajo en épocas de globalización”, Santillana, Buenos Aires.
- GALLART, M. A. (1997): “Los cambios en la relación escuela-mundo laboral”, en Revista Iberoamericana de Educación Número 15 Micropolítica en la escuela.
- GASPARINI, L. (2003) “Argentine’s Distributional Failure.The role of integration and PublicPolicies”, CEDLAS, Universidad Nacional de La Plata.
- GROISMAN, F. (2003): “Devaluación educativa y segmentación en el mercado de trabajo del área metropolitana de Buenos Aires entre 1974 y 2000”, en Estudios del Trabajo, ASET, Primer Semestre 2003.
- KISH, L. (1995): “Survey Sampling”, New York, Wiley.
- KISH, L. (1972): “Muestreo de encuestas”, Trillas, México.
- MAURIZIO, R. (2001), “Demanda de trabajo, sobreeducación y distribución del ingreso”, Ponencia

presentada al Quinto Congreso Nacional de Estudios del Trabajo, ASET, Buenos Aires.

PÉREZ, P. (2005): "Sobreeducación en el mercado de trabajo Argentino en un período de desempleo masivo (1995-2003)", ponencia presentada en 7mo Congreso Nacional de Estudios del Trabajo, ASET.

SALAS, M. (2005): "El fenómeno de sobrecalificación: calificaciones y competencias", ponencia presentada en 7mo Congreso Nacional de Estudios del Trabajo, ASET.

WAISGRAIS, S. (2005): "Determinantes de la sobreeducación de los jóvenes en el mercado laboral Argentino", ponencia presentada en 7mo Congreso Nacional de Estudios del Trabajo, ASET.

Documentos Metodológicos:

Encuesta permanente de hogares. Base usuaria ampliada de EPH (BUA). Diferentes años.

ARTÍCULOS

Los docentes en la Encuesta Permanente de Hogares. Notas metodológicas para su identificación y estudio.

Leandro Bottinelli

encuesta permanente de hogares

trabajo docente

En las últimas décadas se han multiplicado las investigaciones que toman a los/las docentes como objeto de estudio. Desde el campo de la educación, han predominado análisis tanto de las prácticas pedagógicas de maestros y profesores como de sus trayectos de formación; la economía y las disciplinas de la administración pública han enfocado su interés en el sistema de remuneraciones y en otros aspectos que consideran al docente como recurso clave del sistema; desde la psicología y las ciencias de la salud los estudios sobre el stress y el burnout han constituido un campo muy fructífero de indagación; y desde la sociología (en particular desde la sociología del trabajo), se han desarrollado abordajes respecto de la inserción laboral, la agremiación, los procesos y condiciones de trabajo docente así como de ciertas características sociales que hacen a los trabajadores de la educación (identidad profesional, status de la profesión, pertenencia de clase, proletarización y profesionalización). Las instituciones que han cobijado los estudios son diversas: organizaciones sindicales, universidades, organismos de gobierno y organismos internacionales, lo que da cuenta de la importancia relativa que ha adquirido el tema en el campo de la investigación.

El uso de fuentes secundarias de datos para el estudio de los docentes se ha difundido en los últimos años merced a la mayor disponibilidad de información y de los medios de procesamiento. Entre las fuentes secundarias disponibles en Argentina se encuentran los Censos Nacionales de Docentes realizados en 1994 y en 2004 implementados desde el Ministerio de Educación de la Nación. También, desde el sector educativo, se realizan anualmente relevamientos a nivel de cada escuela que permiten conocer las funciones y cargos docentes de manera agregada para cada establecimiento. Para el análisis de las remuneraciones, puede considerarse el Informe indicativo

de Salarios Docentes que construye el Ministerio de Educación de la Nación a partir del relevamiento de los salarios devengados en todas las jurisdicciones del país. Ya en el campo de la información socio-demográfica, los Censos Nacionales de Población permiten caracterizar a la población ocupada en actividades de enseñanza y compararla con otros grupos ocupacionales. Por último, la Encuesta Permanente de Hogares es una fuente útil para caracterizar la ocupación docente, compararla con las ocupaciones de otras ramas y conocer la evolución de mediano plazo de ciertas variables laborales en los principales territorios urbanos del país.

Diversos trabajos han utilizado la encuesta de hogares de Argentina en los últimos años para estudiar el trabajo docente (Vegas, Pritchett y Experton, 1999; Herrero, De Santis y Gertel, 2003; Donaire, 2006; Tenti Fanfani, 2007; Dirí y Pascual, 2011). Este artículo propone un análisis de las particularidades de esta fuente en su nueva versión puesta en marcha en 2003 para el análisis laboral de los docentes así como algunas consideraciones sobre sus fortalezas y limitaciones.

2. La Encuesta Permanente de Hogares

La Encuesta Permanente de Hogares de Argentina (EPH) es un programa de investigación del mercado de trabajo urbano puesto en marcha por el Instituto Nacional de Estadística y Censos (INDEC) en el año 1974. Comenzó a aplicarse a una muestra de hogares de la Ciudad de Buenos Aires y los Partidos que la circundan para, progresivamente, extenderse a otros grandes aglomerados del país y a todas las ciudades que son capitales de provincia. En la actualidad se implementa de manera trimestral en 31 aglomerados urbanos y desde 2010 (solo un trimestre por año), en el resto de las áreas urbanas de cada provincia (consideradas de conjunto), con el proyecto de extender la cobertura también hacia zonas rurales, abarcando a la totalidad de la población del país. Actualmente la muestra comprende a unas 55 mil personas distribuidas en los 31 aglomerados.

La Encuesta clasifica a la población económicamente activa de modo análogo a cómo lo hacen los censos de población. A partir del tipo de actividad económica que declara la población, se clasifica a las personas según la ocupación a la que destinan mayor cantidad de horas semanales. Las bases de datos que se publican para los usuarios (microdatos) permiten producir información desagregada según variables económicas, demográficas o del hogar, pero con las limitaciones propias del carácter muestral de la EPH y el error asociado a este tipo de captura de información. El nomenclador utilizado para la clasificación permite discriminar a los docentes según diferentes niveles de enseñanza: Inicial y Primario; Secundario; Superior o Universitaria; Niveles combinados; Adultos y otros.

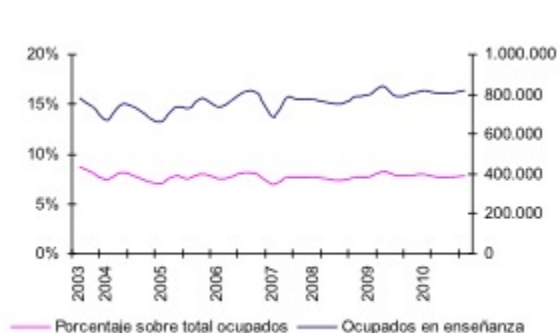
3. Identificación de los docentes en la fuente

El campo fundamental para la identificación de los docentes en la base de datos de la Encuesta es el que codifica la actividad económica en que se desempeña la población. La codificación de las respuestas espontáneas de los entrevistados se realiza en 17 grandes grupos de actividad

económica o ramas a partir del Clasificador de Actividades Económicas para Encuestas Demográficas del MERCOSUR (CAES-MERCOSUR). En la categoría “Enseñanza” se incluyen a los ocupados que se desempeñan en esa actividad más allá de que trabajen en el sistema formal de enseñanza o que lo hagan en instituciones no oficiales y/o de la educación no formal, o de que trabajen de manera particular.

Cuadro 1: Clasificación de las actividades en grandes grupos.

En la medición correspondiente al cuarto trimestre de 2010 los ocupados de enseñanza se estimaban en 820.760 personas y representaban el 7,9% de todos los ocupados en los 31 aglomerados urbanos. Las mediciones trimestrales desde 2003, con excepción de algunas correspondientes al año 2007, muestran una relativa estabilidad en la estimación de la proporción de ocupados en enseñanza, con un valor mínimo de 6,9% (2007 1er. trim.) y uno máximo de 8,8% (3er. trim. 2003). Por otra parte, se observa que en los trimestres posteriores al año 2007, la serie registra una mayor estabilidad.



Fuente: Elaboración propia en base a Microdatos EPH-INDEC, 2003 a 2010.

Gráfico 1: Ocupados en actividades de enseñanza. Absolutos y porcentajes sobre total de ocupados Total de aglomerados. Años 2003 a 2010

Los ocupados que se desempeñan en enseñanza pueden ser estudiados en cuanto a su composición interna mediante la desagregación en grupos según las diferentes variables que releva la Encuesta. Al respecto, las desagregaciones que ofrecen mayor interés para el estudio de los docentes son las siguientes.

Carácter ocupacional: los dos primeros dígitos del campo código de ocupación determinan el “carácter ocupacional” que permite clasificar a las ocupaciones según el producto generado por el proceso de trabajo específico, más allá de la rama o sector de actividad económica al cual pertenezca el establecimiento. Una de las categorías definidas por el clasificador para el carácter ocupacional es “ocupados de la educación” que, en clave del sector educativo, se correspondería en gran medida con los docentes que desempeñan funciones frente a alumnos.

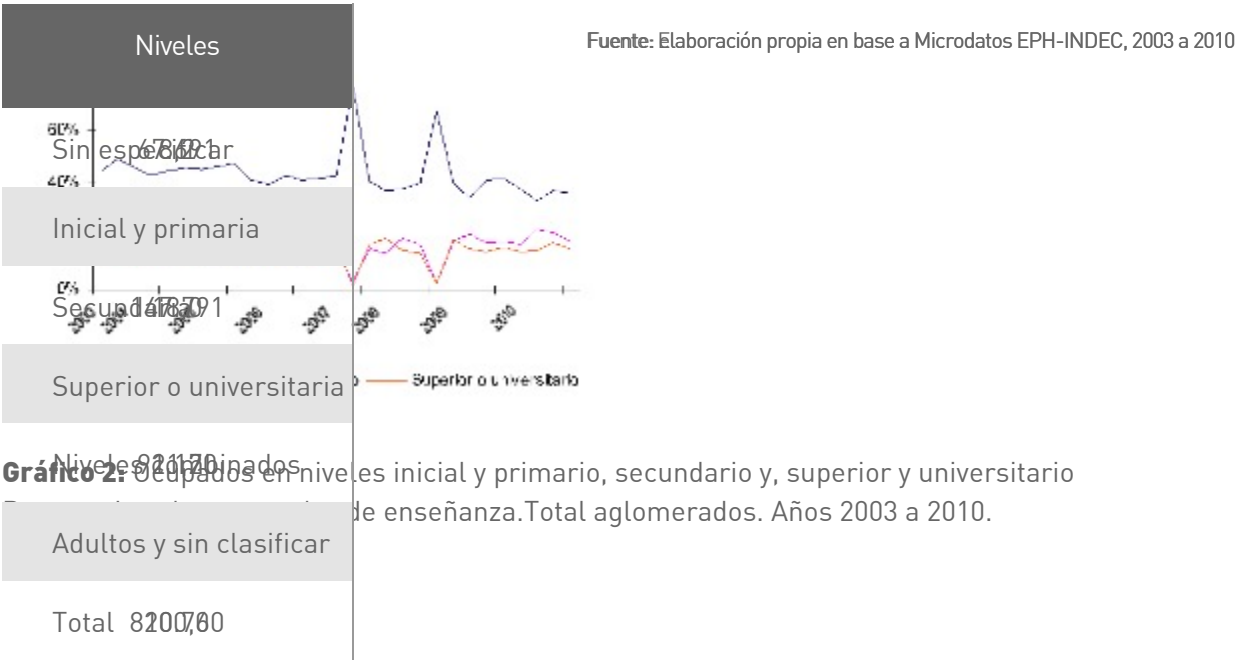
Categoría ocupacional. Es posible distinguir a los ocupados que se desempeñan en

establecimientos bajo la categoría de “obrero o empleado” de aquellos que lo hacen por cuenta propia o que son patrones (propietarios de establecimientos privados). De este modo, se recorta un perfil ocupacional relativamente homogéneo en sus condiciones de trabajo y en sus ingresos que habilita comparaciones pertinentes con asalariados de otras ramas de actividad.

Sector del establecimiento. Permite determinar el tipo de gestión del establecimiento en que el docente desempeña su ocupación principal (estatal, privado o de otro tipo).

Código de Actividad. En su primer nivel de desagregación, los ocupados de enseñanza pueden clasificarse según los niveles educativos en los que desempeñan en su ocupación principal: a) inicial y primaria; b) secundaria; c) universitaria y superior; d) niveles combinados y; e) adultos y servicios de enseñanza no clasificados previamente. En las categorías correspondientes a los niveles de enseñanza de la educación común, se observa en los últimos años una participación ascendente tanto en secundaria como en superior y universitaria, y descendente en inicial y primaria. La serie evidencia problemas en el segundo trimestre de 2007 y en el cuarto de 2008 en los que se habría sobre-estimado los ocupados en inicial y primario a costa del resto. Por otra parte, el nomenclador fue revisado y modificado en el primer trimestre de 2012, no permitiendo desde entonces la desagregación de los ocupados de enseñanza según nivel en el que se desempeñan.

Cuadro 2: Ocupados de enseñanza según nivel educativo en que se desempeñan
Absolutos y porcentajes. Total aglomerados. 4to. trim. 2010



También pueden establecerse otras desagregaciones para el análisis de los ocupados de enseñanza a partir de la estructura del hogar en que residen, si son o no jefes de hogar, del aglomerado en que se localizan o del rango de ingresos que perciben entre otros. El límite para esta estrategia viene dado por la variabilidad (error) que tendrán los resultados en función del carácter muestral del relevamiento.

4. Comparación con los resultados del Censo Nacional de Docentes 2004

Para analizar la precisión de la identificación de docentes en la Encuesta Permanente de Hogares se realizó una comparación de los resultados del cuarto trimestre de 2004, con los que surgen del Censo Nacional de Docentes (CND) de 2004 (Ministerio de Educación, 2006).

Para realizar un pareo lo más ajustado posible en cuanto a cobertura territorial de ambas fuentes, se seleccionaron en la base de datos del CND los docentes que se desempeñan en establecimientos ubicados en las localidades que conforman los 28 aglomerados urbanos relevados por la EPH. Además se delimitó un universo de docentes de la educación básica obligatoria (inicial, primaria y secundaria) en el que la captación del CND se asume exhaustiva (dado que existirían déficits de captación en la educación no formal y que la cobertura de la fuente no incluye al nivel superior universitario).

Cuadro 3: Docentes* según CND 2004 y ocupados de enseñanza** según EPH (4to trim. 2004)
28 aglomerados urbanos. Niveles Inicial, Primario y Secundario

Considerando los perfiles docentes señalados, la EPH registra una estimación puntual de cantidad de docentes muy similar a la que surge del CND, con diferencias menores al 1%. Sin embargo, es importante señalar que la variable niveles de enseñanza de EPH incluye una categoría “niveles combinados” que clasifica a aquellos docentes que desempeñan funciones en más de un nivel, sin precisar en qué niveles. En el cuarto trimestre de 2010 esa categoría reunía a unos 73 mil docentes, el 11% de todos los ocupados de enseñanza. Si se considera que todos esos docentes desempeñan al menos un cargo en inicial, primario o secundario, el total de docentes captados por la Encuesta pasaría de los 423 mil indicados en el cuadro a 496 mil, esto es, un 17% más que los relevados por el CND. Si bien no es posible confirmar que todos los docentes de la categoría “más de un nivel” se desempeñan en la educación básica (ya que algunos pueden combinar cargos en el nivel superior y en adultos), es probable que la mayoría de ellos lo hagan, lo que amplía las diferencias de los datos de ambas fuentes.

Las diferencias señaladas podrían explicarse por otros motivos, más allá de error muestral. Uno de ellos es que la fuente demográfica localiza la información según el lugar de residencia del encuestado (hogar) mientras que la educativa, utiliza el establecimiento. Otra es que la delimitación de localidades que componen los aglomerados no es suficientemente fina para definir la localización, requiriéndose una aproximación a partir de fracciones o radios censales. Podrían también considerarse problemas asociados al subregistro del CND (en particular en el

sector privado) y/o errores en la atribución de categorías ocupacionales de la EPH.

Más allá de ello la cantidad de docentes identificados por la Encuesta Permanente de Hogares se muestra consistente con la relevada en el CND 2004. La comparación realizada con el CND permite establecer que las categorías docentes de esta fuente y los ocupados de enseñanza de la EPH resultan en gran medida asimilables, si se admiten las limitaciones en la precisión de algunas estimaciones.

5. Ocupaciones principal y secundarias

Un rasgo metodológico importante del cuestionario de la Encuesta es que recaba abundante información de la ocupación principal de los encuestados, esto es, de aquella a la que destinan mayor cantidad de horas semanales. Sin embargo, del resto de las ocupaciones denominadas secundarias, solo se consulta la cantidad de horas trabajadas y los ingresos laborales. Características como la rama de actividad, el carácter ocupacional o los beneficios sociales de las ocupaciones secundarias quedan invisibilizadas en la fuente.

Por otra parte, es importante considerar que la encuesta relaciona directamente “ocupación” con “establecimiento”. Esto significa que si una persona desempeña la misma función laboral en más de una unidad productiva, por ejemplo en dos fábricas o en dos escuelas, la fuente considera que el entrevistado tiene más de una ocupación. Por eso, los ocupados de enseñanza que trabajan en más de una escuela o universidad tienen para la fuente “más de un empleo”; el cuestionario releva en profundidad las características de su escuela “principal”, limitándose a indagar solo los ingresos y las horas trabajadas en el conjunto de las otras escuelas.

En el caso de las actividades de enseñanza, la inserción en más de un establecimiento o escuela es una característica bastante frecuente. Esos ocupados tienen para la Encuesta, más de una ocupación. Por el contrario, los docentes que desempeñan diversas funciones en una escuela (por ejemplo maestra y vicedirectora) o que tienen más de un cargo (por ejemplo, en dos asignaturas en el mismo establecimiento) son clasificados como ocupados con una sola ocupación porque lo relevante es el establecimiento.

Los ocupados en enseñanza registran, en promedio, mayor cantidad de ocupaciones que los ocupados del resto de las actividades económicas. Esta característica se acentúa entre aquellos que, además de insertarse en actividades de enseñanza, se desempeñan en ocupaciones de la educación, categoría que es asimilable a la de docentes frente a alumnos del sistema educativo. Se entiende que en este tipo de funciones laborales los docentes pueden sumar horas de clase en diferentes establecimientos algo que no es tan frecuente en cargos de dirección o de apoyo a la enseñanza que no se organizan por horas o módulos. En números redondos, mientras que dos tercios de los docentes en “ocupaciones de la educación” (frente a alumnos) tienen una sola ocupación, es decir, trabajan en una sola escuela (65,7% en el 2010), en el resto de los ocupados en enseñanza los que tienen una sola ocupación son el 84,3%. En los ocupados del resto de las ramas

esta característica de “empleo único” alcanza al 91,6%.

Cuadro 4: Ocupados de actividades de enseñanza según carácter ocupacional y del resto de las actividades según cantidad de ocupaciones.

Fuente: elaboración propia en base a Microdatos EPH INDEC, 2003 a 2010.

		Cantidad de ocupaciones			Total
		Una	Dos	Tres o más	
Actividad enseñanza	Ocupados de la educación	65,7%	26,0%	8,3%	100,0%
	Resto de los ocupados	84,3%	13,6%	2,1%	100,0%
Resto de las actividades		91,6%	7,6%	0,8%	100,0%

En definitiva, si se procesaran los datos bajo el supuesto de que todas las horas e ingresos de aquellos que desempeñan ocupaciones de enseñanza como ocupación principal corresponden a este tipo de actividad, se estaría trabajando con un error máximo de hasta el 19% (pero en promedio mucho menor) y dejando fuera del análisis al 5% de los docentes que tienen en esta actividad una ocupación complementaria.

En los niveles superior y universitario –especialmente este último- es necesario remarcar que una parte no desdeñable de los docentes desempeña actividades laborales más allá de la educación, en establecimientos correspondientes a actividades que no son las de enseñanza. Muchos de ellos incluso, tienen a esas actividades extra-educativas como principal ocupación. Como referencia téngase en cuenta que de los algo más de 150 mil docentes universitarios que se desempeñan en universidades estatales, sólo el 13,4% tiene dedicación exclusiva y un 19,5% semiexclusiva (Anuario Estadísticas Universitarias 2010). Por ello, es probable que una proporción importante de los docentes universitarios no se encuentre relevada dentro de las estadísticas socio-demográficas, en tanto su ocupación principal puede pertenecer a otra rama.

6. Calidad de la variable ingresos laborales

En algunos países como es el caso de Argentina la información sobre los ingresos monetarios de la población es de difícil acceso. Se ha señalado que para muchas personas el monto de su sueldo es un tema “tabú” sobre el que resulta difícil hablar incluso dentro de círculos íntimos como los grupos de amigos y familiares. Por eso, una de las dificultades con que se enfrenta toda encuesta socio-económica es lograr declaraciones precisas de los ingresos laborales y no laborales por parte de los entrevistados.

En el caso de la EPH las dificultades que se observan en el registro de los ingresos de la población

son siempre objeto de debates. Existe consenso respecto de que en ciertas categorías de ocupados la subdeclaración es mucho menor que en otras, como es el caso de la categoría de asalariados donde reviste la gran mayoría de los docentes. También se ha señalado que tales dificultades u omisiones se pueden suponer constantes a lo largo del tiempo con lo cual las comparaciones longitudinales no se verían afectadas (Gasparini, 2000; Altimir, 2002).

Las dificultades en la captación de ingresos en la Encuesta obedecen a la no respuesta y a la subdeclaración. En cuanto al primer caso, los niveles de no respuesta en los ingresos de la ocupación principal para todos los ocupados eran del 16,5% de los casos (sin ponderar) en el segundo trimestre de 2010, valores que resultaban superiores entre los cuentapropistas (22,0%) y más aún entre los patrones (32,5%). Entre los asalariados, categoría en la que se inscribe cerca del 90% de los docentes, los porcentajes de no respuesta eran del 14,3% en el mismo período. En el grupo de los ocupados de enseñanza, los valores son muy similares a los de los asalariados (15,0%).

La no respuesta a las variables de ingreso puede ser tratada mediante la imputación de los valores faltantes por diversos métodos estadísticos. En las bases para usuarios de la Encuesta que comenzaron a publicarse en la última década, se incluyó la imputación de los valores mediante el método hot-deck (asignando valores a partir de otros casos de la base considerados como “donantes”) además de identificar qué casos fueron los imputados y en qué tipos de ingreso. Esto le permite al usuario determinar si utiliza o no los valores imputados en las bases o si decide generar una imputación por un método diferente.

En cuanto a la subdeclaración, diversos estudios han evaluado el fenómeno en la versión anterior de la EPH (la encuesta “puntual”, vigente hasta 2003) señalando magnitudes variables que, según algunos reportes, podría alcanzar el 40% (Salvia y Donza, 2000; Roca y Pena, 2001). Sin embargo, también se ha señalado que parte de las evaluaciones realizadas en las magnitudes de subdeclaración han considerado como referencia otras fuentes de información cuya cobertura territorial es mayor a la de la EPH, sobre-estimando el problema (Roca y Pena, 2001).

Por otra parte, así como la no-respuesta varía en diferentes grupos de ocupados también la subdeclaración es variable. Para este trabajo se evaluaron los ingresos laborales de los docentes en la EPH con respecto a los que se registran en los informes indicativos de salarios docentes del Ministerio de Educación de la Nación. Se tomó como punto de partida los montos ya imputados de la base datos. A partir de esta evaluación se concluyó que los ingresos laborales de los docentes captados por la EPH evidencian valores de subdeclaración que son en promedio del 7%, pero que esos valores varían a lo largo del período 2004-2010, desde un 1% a un 15%. Se detalla a continuación el procedimiento de comparación realizado.

La evaluación de los montos provenientes de la EPH y del Informe de salarios es compleja dada la dificultad para delimitar universos docentes similares en ambas fuentes. También por la diversa cobertura territorial de cada una: principales aglomerados urbanos en un caso; total país en el

otro. Mientras en las bases de la EPH se dispone de un monto de ingreso laboral declarado por el entrevistado e imputado según no-respuesta, en el Informe de salarios el dato corresponde a una declaración realizada por un funcionario provincial especializado. Además de salario bruto y costo laboral, los informes presentan los valores del salario de bolsillo que son los que se consideran en esta evaluación.

Para la comparación, se tomó como punto de partida el monto de salario de bolsillo del cargo testigo de maestro de grado con diez años de antigüedad que registran los Informes de Salarios de Ministerio de Educación. En la EPH se delimitó un perfil de ocupado-asalariado de enseñanza lo más cercano posible al cargo testigo para lo cual se tuvieron en cuenta criterios de selección como el nivel de enseñanza inicial-primario y el carácter ocupacional que permite identificar a los trabajadores docentes que desempeñan tareas frente alumnos, es decir, tareas de “educación”. Adicionalmente se incorporó como criterio que la cantidad de horas trabajadas por esos ocupados se ubicara entre 10 y 30 horas por semana para asegurar la captación de docentes de media jornada (20 horas estimadas), frente al riesgo de incluir docentes de primaria que trabajen jornada completa. No se incluyó un criterio de antigüedad porque se trabajó bajo el supuesto de que los diez años del cargo testigo expresan un promedio de esa variable en el colectivo docente. Finalmente se consideraron los ingresos de la ocupación principal de los docentes (y no los secundarios) de modo de poder establecer la comparación contra los salarios correspondientes a “un” cargo.

Una dificultad adicional que se presentó en la comparación es la ponderación jurisdiccional de la información. Como se sabe el peso del colectivo docente de cada jurisdicción en el total nacional es muy variable. Para componer un valor nacional de salario docente es necesario ponderar los salarios de cada jurisdicción por un valor de referencia que exprese el peso jurisdiccional en el total país del cargo que se quiera comparar. Para hacerlo se puede considerar, por ejemplo, el total de docentes de nivel primario por provincia que provee el Censo Nacional de Docentes de 2004 o la cantidad de secciones de primaria (asumiendo que en este nivel se trata de una buena estimación; en el secundario la equivalencia secciones-docentes es mucho más compleja).

La estrategia mencionada no permite, sin embargo, establecer una comparación directa con los ocupados docentes que capta la EPH ya que en esta fuente se expresan otras relaciones entre los valores provinciales y los nacionales. Esto se explica por la diferente cobertura poblacional de la Encuesta en cada provincia, que resulta muy variable de acuerdo a la importancia de las aglomeraciones urbanas en cada jurisdicción. Por ejemplo, en el caso la Ciudad de Buenos Aires la población alcanzada por la EPH representa el 100% de su población; en la provincia de Buenos Aires, el 75%; en La Rioja el 54% y en Salta el 44%. El mismo fenómeno se observa en la población ocupada en la enseñanza: mientras las estimaciones captan al total de los docentes de CABA, solo registran un parte menor en jurisdicciones con alta ruralidad o con importante población que reside fuera de los grandes aglomerados.

Dado que el salario docente es homogéneo al interior de cada jurisdicción, lo señalado no es un

problema para estimar valores jurisdiccionales a partir de los datos de los aglomerados urbanos, pero sí para componer un valor nacional de ingresos laborales docentes a partir de la Encuesta. Por ejemplo, mientras el CND 2004 indica que los docentes de primaria de la provincia de Buenos Aires representan un 28% del total nacional, en la EPH los ocupados de este perfil de los cuatros aglomerados de Buenos Aires representan el 40% del total de los aglomerados. En provincias de menor dimensión poblacional y mayor ruralidad, en las que el nivel primario está muy extendido y disperso territorialmente, ocurre lo contrario: en la EPH se capta una reducida proporción de los docentes primarios de la provincia que subrepresentan su participación en total nacional. Chaco, Formosa o Misiones son los ejemplos más ilustrativos.

En definitiva, se optó por una ponderación de los montos de salarios de los Informes según las cantidades de docentes por jurisdicción que surgen de los 28 aglomerados. Esto permitió asignar a cada provincia un peso relativo en el total nacional que volviera comparables ambas fuentes.

Mediante este procedimiento se estimaron valores de subdeclaración de los ingresos de los docentes que llegan a un máximo de 15% en el período 2004-2010. La subdeclaración promedio muestra variabilidad en el curso de los siete años. Esto podría explicarse por diferentes motivos: el relativamente alto nivel de error existente en desagregados de casos de la EPH tan reducidos como el que fue necesario delimitar para establecer la comparación con otra fuente; la importante variación que tuvieron los salarios-ingresos en el curso de este período o; las propias dificultades de la fuente para captar los ingresos en general. Más allá de estas dificultades, los ingresos captados por EPH para los docentes tendrían un nivel de subdeclaración por debajo del promedio que otros estudios han señalado para el conjunto de los asalariados. Esta subcaptación estaría en un orden razonable y permitiría utilizar esta información para algunas estimaciones o para observar una tendencia de mediano plazo.

Si se admite que la variabilidad en la subcaptación se explica no por las características de la fuente ni por la baja calidad de la información en la variable ingresos de los ocupados de la rama enseñanza, sino por el importante nivel de error propio del reducido grupo docente que hubo que aislar para comparar con la fuente de salarios, se puede asumir que al considerar subpoblaciones de mayor dimensión (como es el caso de los ocupados de rama de enseñanza e incluso de algunas subdivisiones por nivel o sector), la estabilidad de la información es mucho mayor y habilita estimaciones más robustas que permiten comparaciones interanuales.

Cuadro 6: Salarios docentes*, ingresos laborales** del maestro de grado y subestimación según año

Fuente: *Informe indicativo de salarios docentes del Ministerio de Educación y
**EPH-INDEC.

	Salarios Informe	Ingresos EPH	Subestimación de la EPH
2004	530	489	-8%
2005	686	641	-7%
2006	833	812	-3%
2007	1.038	1.025	-1%
2008	1.512	1.367	-10%
2009	2.011	1.713	-15%
2010	2.250	2.043	-9%

7. Coeficientes de variación de las estimaciones

Como toda información proveniente de una muestra, los resultados de la EPH contienen un cierto error asociado al muestreo. Cada estimación puntual de totales, medias o proporciones, registra una cierta probabilidad de contener al parámetro, es decir, al valor que se quiere conocer en el universo poblacional (INDEC, 2011). La teoría del muestreo señala que las estimaciones de todas las posibles muestras de tamaño n que puedan extraerse de un universo, arrojan diversas estimaciones que se distribuyen “alrededor” del parámetro. Los diseños muestrales complejos, como el de la Encuesta, tienden a incrementar el error en comparación con un muestreo aleatorio simple; sin embargo, también es cierto que un adecuado procedimiento de estratificación, que considere la o las variables clave asociadas al fenómeno bajo estudio, juega en sentido contrario y puede compensar parcialmente el incremento del error señalado.

Cuando no se trata de estimar el parámetro por única vez sino que el objetivo es conocer su evolución a través del tiempo, la forma de considerar el error muestral se modifica parcialmente. En el caso de una muestra extraída por única vez para conocer una característica poblacional, el error de la estimación es la única referencia con que cuenta el analista sobre la precisión del dato. En el marco del nivel confianza elegido, el error de esa única estimación será el punto en que deberá anclar la solidez de sus conclusiones. En cambio, cuando se dispone de una serie temporal proveniente de diversas muestras sucesivas, la evolución más o menos oscilante de las estimaciones se constituye en un punto de apoyo adicional para considerar la robustez de los resultados alcanzados. Si el analista dispone de una serie con estimaciones de un parámetro cuyo error está en el límite de lo tolerable, pero que en su evolución señalan una tendencia de mediano

plazo, cuenta entonces con una referencia adicional como fundamento para sus conclusiones respecto de la evolución de mediano plazo del fenómeno.

El diseño muestral de la EPH se estructura para valorar con precisión los cambios en la tasa de desempleo entre trimestres sucesivos y, por supuesto, entre períodos mayores. Para la estimación de otros parámetros la lógica de trabajo con series, es la que debe considerarse como pertinente.

Los valores de error de las estimaciones tienen relación también con la desagregación de los resultados, ya que los coeficientes de variación se incrementan cuando corresponden a estimaciones que involucran a parcelas de casos de la muestra. Si bien no existe una referencia rígida para determinar cuál es el máximo error aceptable –ya que depende del tipo de análisis que se esté desarrollando–, lo cierto es que a partir de coeficientes de variación del 20% los resultados deben ser tratados con cautela (INDEC, 2004).

En la estimación de la población ocupada en actividades de enseñanza en un semestre el coeficiente de variación es del 3,8% con un 90% de confianza (INDEC, 2011). En efecto, la estimación puntual de los ocupados de enseñanza de la rama de enseñanza para el segundo semestre de 2011 era de 823 mil personas en los 28 aglomerados urbanos. El error asociado a la estimación de este total, con el nivel confianza señalado, es de 51 mil personas ($823.000 \times 0,038 \times 1,645$), con lo que los límites inferior y superior son 772 mil y 875 mil respectivamente.

En las estimaciones provenientes de bases de datos semestrales el coeficiente de variación se reduce entre el 10 y el 20% de acuerdo a la cantidad de casos involucrados. En el grupo de los ocupados de enseñanza, que son cerca de 800 mil, la agregación semestral genera estimaciones que son un 10% más eficientes.

Cuadro 7: Coeficientes de variación de estimaciones trimestrales y semestrales

Fuente: elaboración propia en base a Microdatos EPH-INDEC, 2003 a 2010.

Cantidad de casos de la estimación	Coeficiente de estimación (CV)		Reducción del CV
	Trimestral	Semestral	
20.000	27,0	21,4	21%
40.000	19,1	15,6	18%
60.000	15,6	12,9	17%
80.000	13,5	11,3	16%
200.000	8,5	7,3	14%
500.000	5,3	4,7	11%
800.000	4,2	3,8	10%

El error asociado a la estimación de la cantidad de ocupados en la actividad de enseñanza en el total de aglomerados resulta más que aceptable según los criterios de error comúnmente aceptados. En las estimaciones que toman como referencia a subgrupos dentro de los docentes (los del nivel primario, los de escuelas estatales, etc.), los valores de error se incrementan en la medida que disminuyen la cantidad de casos involucrados. Al final de esta sección se consigan referencias de error para diferentes subpoblaciones docentes. Sin embargo, tanto la agregación de bases trimestrales en semestrales, como el trabajo con series para valorar tendencias de mediano plazo, permiten controlar el error de las estimaciones de modo de arribar a conclusiones que sean pertinentes. El uso de medias móviles, promediando las estimaciones de un período con las del inmediato anterior y posterior, es otra estrategia válida para evaluar la dirección general del cambio de un fenómeno más allá de coyunturas o variaciones debidas al muestreo.

También es necesario señalar que en las estimaciones que no refieren a totales poblaciones sino a tasas o proporciones que derivan de un cociente entre dos totales (por ejemplo un porcentaje de docentes asalariados), el error se puede estimar a partir de las tablas de error y considerando el total correspondiente al numerador de la tasa. Bajo el supuesto de que el denominador es grande (en todas las tasas sobre la población docente siempre son superiores a 250 mil casos), el error en que se incurre por seguir este procedimiento es bajo. En el caso de los promedios, se puede seguir un criterio similar y aplicar el error correspondiente al total de casos (n) involucrado en el promedio en cuestión.

Es importante señalar que el procedimiento señalado se muestra eficiente para conocer la tendencia en la evolución de fenómenos que involucran a subpoblaciones al interior de los docentes, pero que pueden no permiten obtener conclusiones de cambios entre trimestres, semestres y, en algunos casos, entre dos años sucesivos. Incluso, si se desea estimar la evolución

de la cantidad de ocupados en enseñanza entre dos trimestres sucesivos, y dicha variación es –como muy probablemente pueda ocurrir– de una magnitud inferior al 4%, la Encuesta no va a resultar una herramienta satisfactoria en tanto el error asociado a la estimación de ese total es justamente del orden del 4%. Pero si se busca evaluar el mismo fenómeno en un plazo más largo –digamos de 1 o 2 años– y se disponen de estimaciones para todos los trimestres o semestres, la tendencia que se dibuja permite sacar conclusiones consistentes no solo para ese parámetro sino para otros referidos grupos menores.

Cuadro 8: Ocupados en actividades en enseñanza según categorías de variables seleccionadas
Total aglomerados. 4to trimestre de 2010

Fuente: elaboración propia en base a Microdatos EPH, INDEC, 2003 a 2010.

	Casos (4to. trim. 2010)	Casos ponderados (4to. trim. 2010)	Coefficiente de variación trimestral (4to. trim. 2010)
Residen en la región Noroeste (NOA)	440	93.463	12,5
Son jefes de hogar	788	320.883	6,7
Alcanzaron a completar el nivel Superior (Terciario o Universitaria)	1.277	519.827	5,2
Desempeñan ocupaciones de la educación	1.356	578.509	5,0
Se desempeñan en el nivel de enseñanza secundario (en su ocupación principal)	379	147.791	10,0
Se desempeñan en establecimientos del sector privado (en su ocupación principal)	545	305.912	6,8
Son obreros o empleados	1.830	773.916	4,3
Tienen jerarquía laboral de asalariados	1.741	732.752	4,4
Tienen dos o más ocupaciones	467	224.241	8,1
Trabajan 35 o más horas semanales (en los establecimientos)	698	288.993	7,1
Están buscando trabajar más horas (subocupación demandante)	104	45.986	17,8
Trabajan en establecimientos grandes (más de 40 empleados)	859	412.249	5,9
Utilizan equipo informático en su ocupación	109	41.283	18,8
Son trabajadores no registrados (sin descuento o aporte jubilatorio)	441	161.370	9,6
Residen en hogares del primero o segundo quintil de ingresos (por persona del hogar)	261	123.069	10,9

8. Consideraciones finales

La Encuesta Permanente de Hogares es una fuente de datos útil para el estudio de los docentes

en la medida que aporta información periódica y comparable con otros grupos de ocupados. Si bien no tiene la precisión de una fuente del sector educativo como es el caso del Censo Nacional de Docentes o el Informe Indicativo de Salarios, permite ciertos tipos de análisis que pueden ser concebidos como complementarios respecto de estas fuentes más específicas. A respecto, en estos párrafos finales, se repasan las principales limitaciones y fortalezas.

De la comparación con el Censo Nacional de Docentes de 2004, se desprende que la cantidad de docentes u ocupados en actividades de enseñanza que releva la EPH resulta consistente. A partir de la comparación realizada con el CND, se puede establecer que las categorías docentes de esta fuente y los ocupados de enseñanza de la EPH resultan en gran medida asimilables, si se admiten las limitaciones en la precisión de algunas estimaciones.

Las dificultades derivadas de no poder caracterizar la naturaleza de las ocupaciones secundarias de los trabajadores es un problema importante, en particular teniendo en cuenta la alta proporción relativa de ocupados de enseñanza con otras ocupaciones. Una estrategia para abordar este problema es atribuir a la ocupación principal, las características de las secundarias (ingresos, horas trabajadas) pero con las limitaciones del error aparejado que ya se mencionaron.

Los coeficientes de variación trimestrales de las estimaciones de totales de población ocupada en actividades de enseñanza resultan más que aceptables para el total de ocupados en este sector. La desagregación de la categoría ocupados de enseñanza admite numerosas alternativas que permiten identificar en la fuente subgrupos de docentes que resulta de interés estudiar. Así pueden considerarse docentes de diferentes niveles de enseñanza, sectores de gestión, funciones educativas o pertenecientes a diversos niveles socio-económicos o tipos de hogar. Si bien algunos de los valores de error en este tipo de resultados rondan el 15%, podrá considerarse la estrategia de pool de datos para procesar resultados semestrales en lugar de trimestrales, realizando ajustes en los ponderadores.

Las dificultades que se observan en los valores de ingresos de la Encuesta son siempre objeto de debates y críticas. Existe consenso respecto de que en ciertas categorías de ocupados la subdeclaración es mucho menor que en otras, como es el caso de la categoría de asalariados donde reviste la gran mayoría de los docentes. También que se pueden suponer constantes a lo largo del tiempo los problemas de no respuesta y subdeclaración con lo cual las comparaciones longitudinales no se verían afectadas. La evaluación realizada a partir de la comparación entre los ingresos laborales registrados por la EPH y los provenientes de los Informes Indicativos de Salarios ponen en evidencia niveles de subdeclaración bajos (menores al 15%) pero variables a lo largo del período de la posconvertibilidad, por lo que su uso debe considerarse para analizar tendencias de mediano plazo (3 a 5 años).

La metodología presentada para analizar la evolución del empleo docente en base a la EPH tiene una característica que es necesario destacar: no es sofisticada en el sentido de irreproducible para usuarios con un conocimiento medio de estadística. Luego de la validación de la fuente

mediante el cruce con otras de tipo sectorial, la metodología propuesta para el análisis de las tendencias en el empleo docente, ha procurado utilizar todos los insumos que el INDEC pone a disposición del público (microdatos, documentos metodológicos y tablas de errores). Se espera que este aspecto del trabajo realizado pueda resultar en una contribución para una más eficaz utilización de esta fuente en el estudio de los docentes.

Bibliografía

- ALTIMIR, O. y otros. (2002). La distribución del ingreso en Argentina, 1974-2000, en Revista de la CEPAL, N° 78, Diciembre (Buenos Aires).
- BECCARIA, L.; GROISMAN, F. (2006). Inestabilidad, movilidad y distribución del ingreso en Argentina. Revista de la CEPAL, n° 89.
- DIRIE, C. y PASCUAL, L. (2011). La rama de enseñanza en la Provincia de Buenos Aires: Evolución reciente y condiciones laborales. Ministerio de Trabajo de la Provincia de Buenos Aires. Buenos Aires.
- DONAIRE, R. (2009). Sobre la proletarización de los trabajadores intelectuales. Un ejercicio comparativo a partir del caso de los docentes de Argentina. En NEFFEA, J.; DE LA GARZA, E.; MUÑIZ, L. (Org.). Trabajo, empleo, calificaciones profesionales, relaciones de trabajo e identidades laborales. Buenos Aires: Consejo Latinoamericano de Ciencias Sociales. p. 57-88.
- DONAIRE, R. y LLANOS POZZI, M. (2004). Población docente: reflexiones sobre los relevamientos estadísticos continuos y censales existentes. Limitaciones y propuestas de reformulación. Departamento de Estadística, Dirección de Investigación, Secretaría de Educación, Gobierno de la Ciudad de Buenos Aires. Buenos Aires.
- GASPARINI, L. y otros (2000). La distribución del ingreso en la Argentina y en la provincia de Buenos Aires. En Cuadernos de Economía, N° 49 (La Plata).
- HERRERO, V, DE SANTIS, M, y GERTEL, H. (2003). El ingreso de los docentes en la Argentina: ¿es alto o bajo? Instituto de Economía y Finanzas, Facultad de Ciencias Económicas, Universidad Nacional de Córdoba. Córdoba.
- INDEC. (2011). Errores de muestreo en la estimación de totales de personas a partir de una base trimestral. Buenos Aires: Instituto Nacional de Estadísticas y Censos.
- INDEC. (2003). La nueva Encuesta Permanente de Hogares de Argentina. Buenos Aires: Instituto Nacional de Estadísticas y Censos.
- MARSHALL, A. (1999). El empleo en el sector educativo. Ministerio de Educación de la Nación.
- MINISTERIO de Cultura y Educación de la Nación. (2011). Anuario de estadísticas universitarias 2010. Buenos Aires.
- MINISTERIO de Educación, Ciencia y Tecnología. (2006). Censo Nacional Docente 2004 - Resultados definitivos. Buenos Aires: Ministerio de Educación, Ciencia y Tecnología de la Nación.
- MORDUCHOWICZ, A. (2004). Discusiones de economía de la educación. Buenos Aires: Losada.
- ROCA, E, PENA, H. (2001). La declaración de ingresos en las encuestas de hogares. Ponencia presentada al 5to. Congreso Nacional de la Asociación de Estudios del Trabajo (ASET). Buenos Aires.
- SALVIA, A. y DONZA, E. Sin año. Problemas de medición y sesgos de estimación derivados de la no

respuesta completa a las preguntas de ingresos en la EPH (1990-1998).

TENTI FANFANI, E. (2010). Características socioeconómicas, opiniones, valoraciones y expectativas de los docentes argentinos de educación básica. Encuesta Nacional a Docentes 2010. Buenos Aires: Mimeo.

TENTI FANFANI, E. (2007). La condición docente. Buenos Aires: Siglo XXI Editores.

VEGAS, E.; PRITCHETT, L; y EXPERTON, W. (1999). Attracting and retaining qualified teachers in Argentina: Impact of the structure and level of compensation. World Bank. Washington.

ARTÍCULOS

Especificación del modelo, datos faltantes y casos atípicos: la “trastienda” de una investigación basada en el análisis de la varianza

Luisa Iñigo¹

mercado de trabajo

varianza

El propósito del presente artículo es comentar algunas decisiones metodológicas y técnicas tomadas durante el proceso de investigación de la Tesis de Maestría “Escolaridad y diferenciación de la fuerza de trabajo en el Gran Buenos Aires, 1985-2005”².

El análisis de la correlación entre el salario (o, más en general, el ingreso personal) y diversas características de los perceptores y sus trabajos, a partir de datos de la EPH, es relativamente frecuente en el campo de las ciencias sociales argentinas en los estudios referidos a la distribución personal del ingreso o a las características de los mercados de fuerza de trabajo. Para hacer dicho análisis, es necesario comenzar por tomar una serie de decisiones que van desde las más obvias, como la especificación del modelo, hasta otras más discretas, como el tratamiento que se dará a los casos con información faltante o a los casos atípicos.

Sobre este último tipo de decisiones, sólo ocasionalmente se encuentran referencias explícitas en los informes de investigación que llegan al público. En cuanto a la especificación general del modelo que representará la variabilidad salarial, normalmente la selección de variables regresoras es fundamentada a partir de un desarrollo conceptual y el foco específico de atención de cada investigación. Sin embargo, el proceso que lleva a la selección final de k variables regresoras entre el conjunto de variables que resultan candidatas, suele quedar retirado de la mirada del lector.

Este artículo apunta a visibilizar esta “trastienda” de la investigación, buscando establecer, siempre que sea posible, cuánto de los resultados obtenidos se afectan según las decisiones tomadas en la investigación. El objetivo es allanar en algo el camino a quienes deban abordar por primera vez la información de ingresos personales relevada por la EPH mediante técnicas de descomposición de la varianza.

De entre todos los aspectos en que sería posible detenerse, nos referiremos a los tres ya mencionados: la especificación del modelo, el tratamiento de los casos con información faltante y la detección y tratamiento de los casos atípicos. Las reflexiones acerca de la especificación del modelo parten de la identificación de las variables que habitualmente se proponen como “explicativas” de la variabilidad salarial en la literatura especializada; luego se detalla el conjunto de variables consideradas como “candidatas” en nuestra investigación, para, posteriormente, presentar los resultados de la aplicación del método de introducción “paso a paso” de variables al modelo, que ponen en evidencia cuáles de esas variables (y en cuántos años de la serie) presentan una correlación estadística significativa con los salarios en presencia de las restantes; por último, se compara el R^2 corregido correspondiente a los cuatro modelos finalmente aplicados.

El apartado siguiente refiere a la incidencia de la información faltante en los resultados obtenidos. Para ello, se parte de consideraciones generales acerca de la información faltante y sus medidas remediales; se continúa evaluando la incidencia y la fuente de la pérdida de información en las bases de datos utilizadas y, finalmente, se explicitan las limitaciones encontradas para aplicar medidas remediales adecuadas.

El tercer apartado trata sobre la detección y el tratamiento de los casos atípicos, partiendo de consideraciones generales acerca de estos casos en la regresión y el análisis factorial; en él se presenta una clasificación de los casos atípicos hallados en las veintiuna bases analizadas según su origen presunto y el tratamiento que se les aplicó. Por último, se evalúa la incidencia de las correcciones realizadas sobre los resultados de la regresión. Finalmente, se sintetizan las conclusiones principales.

2. La especificación del modelo

Existe una multitud de trabajos que, principal o subsidiariamente y desde perspectivas conceptuales diferentes, aplican técnicas basadas en la descomposición de la varianza (generalmente, la regresión múltiple) a la información sobre ingresos personales que surge de la EPH, transformada mediante el cómputo del logaritmo natural. Las variables incorporadas como regresoras varían de estudio en estudio, y conforman un conjunto que incluye el sexo de los perceptores, su nivel educativo, su condición de jefe de hogar, su edad, su edad al cuadrado, su estado civil, el número de hijos menores de 18 años, la categoría ocupacional, la condición de registro y la calificación de las ocupaciones, así como la rama de actividad, el sector y el tamaño de los establecimientos en que están empleados.

En nuestra investigación de Tesis, la identificación de las determinaciones principales del valor y el precio horario de la fuerza de trabajo nos llevó a seleccionar un conjunto de variables candidatas a formar parte de los modelos de regresión y de análisis factorial. Una primera instancia del análisis de la información consistió en la exploración de las correlaciones parciales entre dichas variables y el logaritmo natural del salario horario, mediante el método “paso a paso” de introducción de variables.

Las variables candidatas fueron: el sexo; la edad; el sub-aglomerado de residencia (Ciudad de Buenos Aires o partidos del conurbano); la cantidad de menores de 18 años que habitan en el hogar del asalariado; el máximo nivel de instrucción alcanzado; la calificación de la ocupación; el carácter general de la ocupación; la situación de registro; la antigüedad y la rama de actividad.

De antemano, este listado fue resultado de la exclusión de algunas variables que a primera vista parecían indicadores apropiados de ciertas determinaciones identificadas.

Para empezar, la variable correspondiente a la cantidad de menores de 18 años escolarizados que habitan en el hogar del asalariado fue excluida por dos razones. En primer lugar, por presumirse su colinealidad con la cantidad de menores de 18 años (escolarizados o no). En segundo lugar, en razón de que la inclusión en el salario mensual de un equivalente específico destinado a la educación de los hijos fue eliminada en 1996.

La exclusión de la “jerarquía ocupacional”, que sin duda presentaría correlación parcial con el salario, se debió al hecho de que sólo se encuentra disponible a partir del año 1992, año en que la codificación de la ocupación para el aglomerado GBA pasó a basarse en el Clasificador Nacional de Ocupaciones (CNO).

El “tamaño del establecimiento” en que el asalariado está empleado, por su parte, fue excluido debido a que, si bien la escala de operación del capital individual evidentemente juega un papel en la determinación del nivel del salario, la información de la EPH que habitualmente se utiliza como indicador de este aspecto (v. gr., la cantidad de personas empleadas) resulta de utilidad limitada en este sentido. Adicionalmente, se tuvo en consideración que los casos con información faltante para esta variable representaron el 11% de la muestra de asalariados de cada año del período 1985-2005, con un mínimo de 5% (en 1987) y un máximo de 22% (en 1990).

En síntesis, las variables candidatas a integrar el modelo de regresión fueron las que se detallan a continuación:

Cuadro 1: Variables candidatas para el modelo de regresión

	Variable original	Variables derivadas (dicótomas)	Nombre de la variable derivada	Nivel de medición	Categoría de control
Atributos del asalariado	Sexo	Mujer	dumsex	Nominal	Varón
	Edad	--	--	Intervalar (años)	--
	Edad al cuadrado	--	--	Intervalar (años al cuadrado)	--
	Sub-aglomerado de residencia	Habita en Ciudad de Bs. As.	aglomerado	Nominal	Partidos del conurbano
	Cantidad de menores de 18 años en el hogar del asalariado	--	--	Intervalar (cant. de personas)	--
	Nivel de instrucción	Medio incompleto	dumemi	Nominal	Hasta primario completo
		Medio completo	dumemc	Nominal	
		Superior no universitario inc. ocomp.	dumesnu	Nominal	
		Universitario incompleto	dumeui	Nominal	
		Universitario completo	dumeuc	Nominal	
Atributos de la ocupación	Calificación de la ocupación	Operativa o técnica	dumcal1	Nominal	No calificada
		Profesional	dumcal2	Nominal	
	Carácter general de la ocupación	2	ocupadm	Nominal	Ocup. de producción y reparación de bienes y ocup. auxiliares de la prod.
		3	ocupcom	Nominal	
		4	ocuptran	Nominal	
		5	ocupserv	Nominal	
		6	ocupvigi	Nominal	
	Situación de registro	No registrado	ennegro	Nominal	Registrado
Atributos del establecimiento	Rama de actividad	Antigüedad en el puesto	--	Intervalar (años)	--
		2	electya	Nominal	Industria manufacturera y servicios de reparación
		3	constr	Nominal	
		4	comercio	Nominal	
		5	transp	Nominal	
		6	finan	Nominal	
		7	servsope	Nominal	
		2	electya	Nominal	

Nota: Las categorías de las variables “carácter general de la ocupación” y “rama de actividad” se corresponden con la siguiente codificación (continúa):

Carácter general de la ocupación:

1. Tareas de producción, auxiliares de producción y de reparación de bienes,
2. Tareas de gestión administrativa, jurídico-legal y contable (incluye tareas de dirección, véase nota 2),
3. Tareas de comercialización,
4. Tareas de transporte, almacenaje y telecomunicaciones,
5. Tareas de servicios,
6. Tareas de vigilancia y seguridad.

Rama de actividad:

1. Industria manufacturera y servicios de reparación.
2. Generación y distribución de electricidad, luz y agua.
3. Construcción.
4. Comercio al por mayor y menor y restaurantes y hoteles.

5. Transporte, almacenamiento y comunicaciones.
6. Intermediación financiera, servicios empresariales y de alquiler.
7. Servicios comunales, sociales y personales.
8. Sin dato y otras ramas de actividad.

El método “paso a paso” de introducción de variables regresoras al modelo conservó las variables correspondientes a la calificación de la ocupación, la antigüedad, la ocupación en tareas “de gestión administrativa, jurídico-legal y contable”¹¹ el máximo nivel educativo alcanzado por el asalariado (especialmente, los niveles iguales o superiores al medio completo) y la situación de registro del asalariado en todos los años en que el conjunto completo de las variables candidatas se encontraba presente (1986-1988 y 1991-2002)¹².

Por su parte, la “cantidad de menores de 18 años en el hogar del asalariado” fue excluida por el proceso en catorce de los quince años observados. Las variables dicótomas referidas a las ocupaciones que involucraban tareas de transporte, almacenaje y telecomunicaciones y a las ramas de actividad ligadas con el transporte, el almacenamiento y las comunicaciones resultaron excluidas en trece y doce años, respectivamente, verificándose que su exclusión no se debía a su probable colinealidad recíproca¹³. La variable dicótoma correspondiente a tareas “de servicios” fue excluida en diez años. Por su parte, las variables referidas a las ramas de “generación y distribución de electricidad, luz y agua”, “servicios sociales y personales” y “construcción” y las ocupaciones que involucraban “tareas de comercialización”, lo fueron en nueve.

En síntesis, las variables que aparecieron asociadas más fuertemente con la variabilidad del logaritmo del salario horario en presencia de las restantes fueron aquéllas ligadas con la complejidad del trabajo realizado, con el ejercicio de la representación del capital en la producción y la circulación y con la condición más o menos manifiesta de sobrante para las necesidades del capital total de la sociedad¹⁴. En cambio, el “carácter de la ocupación”, esto es, en un sentido muy general, el tipo de desgaste corporal que tiene lugar durante el trabajo, no apareció jugando un papel en la determinación de diferencias salariales, a excepción de lo que sucedió con las tareas de “vigilancia y seguridad”, normalmente peor pagas que el resto. Por su parte, en las actividades de “comercio, restaurantes y hoteles” por lo general se pagaron los peores salarios, mientras que en las de “intermediación financiera, servicios empresariales y de alquiler” los salarios fueron más altos que en el resto en casi la mitad de los años aquí considerados. Fuera de éstas, tampoco la rama de actividad pareció establecer diferencias salariales sustanciales. Por último, el sexo, la edad, la edad al cuadrado, el sub-aglomerado de residencia y el nivel medio incompleto determinaron diferencias salariales significativas en la mayor parte de los años, si bien en general moderadas.

En nuestra investigación de tesis comparamos la bondad de ajuste de modelos diversos, especificados mediante la incorporación o la eliminación de variables individuales, a partir de un modelo básico aplicable a todos los años del período que se observó en la investigación. Así, al modelo que incluía las variables dicótomas derivadas de la calificación de la ocupación, el nivel educativo, la situación de registro, el sexo, la rama de actividad y el carácter de la ocupación y la

edad, se incorporó primero el sub-aglomerado de residencia y luego la antigüedad en el puesto (quitándose entonces la edad, para prevenir una posible colinealidad). En un ejercicio posterior, se incorporó la edad al cuadrado a este último modelo.

Formalmente, los modelos quedaron especificados como se detalla a continuación:

Modelo 1:

$$\ln(wh) = \hat{\beta}_0 + \hat{\beta}_1 cal1 + \hat{\beta}_2 cal2 + \hat{\beta}_3 dumemi + \hat{\beta}_4 dumemc + \hat{\beta}_5 dumesnu + \hat{\beta}_6 dumeui + \hat{\beta}_7 dumeuc \\ + \hat{\beta}_8 ennegro + \hat{\beta}_9 mujer + \hat{\beta}_{10} electya + \hat{\beta}_{11} constr + \hat{\beta}_{12} comercio + \hat{\beta}_{13} finan + \hat{\beta}_{14} servsope \\ + \hat{\beta}_{15} ocupadm + \hat{\beta}_{16} ocupcom + \hat{\beta}_{17} ocupser + \hat{\beta}_{18} ocupvigi + \hat{\beta}_{19} edad + \hat{e}$$

Modelo 2:

$$\ln(wh) = \hat{\beta}_0 + \hat{\beta}_1 cal1 + \hat{\beta}_2 cal2 + \hat{\beta}_3 dumemi + \hat{\beta}_4 dumemc + \hat{\beta}_5 dumesnu + \hat{\beta}_6 dumeui + \hat{\beta}_7 dumeuc \\ + \hat{\beta}_8 ennegro + \hat{\beta}_9 mujer + \hat{\beta}_{10} electya + \hat{\beta}_{11} constr + \hat{\beta}_{12} comercio + \hat{\beta}_{13} finan + \hat{\beta}_{14} servsope \\ + \hat{\beta}_{15} ocupadm + \hat{\beta}_{16} ocupcom + \hat{\beta}_{17} ocupser + \hat{\beta}_{18} ocupvigi + \hat{\beta}_{19} edad + \hat{\beta}_{20} aglom + \hat{e}$$

Modelo 3:

$$\ln(wh) = \hat{\beta}_0 + \hat{\beta}_1 cal1 + \hat{\beta}_2 cal2 + \hat{\beta}_3 dumemi + \hat{\beta}_4 dumemc + \hat{\beta}_5 dumesnu + \hat{\beta}_6 dumeui + \hat{\beta}_7 dumeuc \\ + \hat{\beta}_8 ennegro + \hat{\beta}_9 mujer + \hat{\beta}_{10} electya + \hat{\beta}_{11} constr + \hat{\beta}_{12} comercio + \hat{\beta}_{13} finan + \hat{\beta}_{14} servsope \\ + \hat{\beta}_{15} ocupadm + \hat{\beta}_{16} ocupcom + \hat{\beta}_{17} ocupser + \hat{\beta}_{18} ocupvigi + \hat{\beta}_{19} aglom + \hat{\beta}_{20} antig + \hat{e}$$

Modelo 4:

$$\ln(wh) = \hat{\beta}_0 + \hat{\beta}_1 cal1 + \hat{\beta}_2 cal2 + \hat{\beta}_3 dumemi + \hat{\beta}_4 dumemc + \hat{\beta}_5 dumesnu + \hat{\beta}_6 dumeui + \hat{\beta}_7 dumeuc \\ + \hat{\beta}_8 ennegro + \hat{\beta}_9 mujer + \hat{\beta}_{10} electya + \hat{\beta}_{11} constr + \hat{\beta}_{12} comercio + \hat{\beta}_{13} finan + \hat{\beta}_{14} servsope \\ + \hat{\beta}_{15} ocupadm + \hat{\beta}_{16} ocupcom + \hat{\beta}_{17} ocupser + \hat{\beta}_{18} ocupvigi + \hat{\beta}_{19} aglom + \hat{\beta}_{20} antig + \hat{\beta}_{21} edad^2 + \hat{e}$$

La comparación del R2 corregido muestra cómo mejora el ajuste de la función de regresión a medida que se incorporan el sub-aglomerado de origen, la antigüedad y la edad al cuadrado como variables regresoras (cuadro 2).

Cuadro 2: Bondad del ajuste de modelos de regresión (R2 corregido)* GBA. 1985-2005

Año	Modelo			
	1	2	3	4
1985	0,372	-	-	-
1986	0,351	0,351	0,357	0,36
1987	0,426	0,426	0,434	0,439
1988	0,45	0,451	0,456	0,46
1989	0,356	-	-	-
1990	0,34	-	-	-
1991	0,311	0,312	0,312	0,315
1992	0,264	0,266	0,263	0,265
1993	0,284	0,285	0,289	0,292
1994	0,318	0,321	0,32	0,325
1995	0,358	0,362	0,377	0,381
1996	0,343	0,346	0,344	0,352
1997	0,381	0,383	0,371	0,386
1998	0,431	0,435	0,444	0,449
1999	0,382	0,385	0,403	0,405
2000	0,42	0,42	0,434	0,437
2001	0,421	0,423	0,431	0,434
2002	0,49	0,494	0,49	0,498
2003	0,411	0,414	-	-
2004	0,419	0,426	-	-
2005	0,441	0,443	-	-

Fuente: Elaboración propia a partir de bases usuarias EPH, octubre 1985-2002 y 4º trim. 2003-2005.

Nota (*) La información correspondiente al sub-aglomerado de residencia no se encuentra disponible en los años 1985, 1989 y 1990. De 2003 a 2005, la información correspondiente a la antigüedad en el puesto no fue publicada.

Como detallamos en el informe de tesis, la incorporación de la antigüedad en el puesto como variable regresora supone algunas diferencias de magnitud de los coeficientes de correlación parcial dignas de mención, con respecto a los modelos 1 y 2. Entre ellas, se encuentra: la reducción de la correlación parcial (positiva) entre el salario y las variables derivadas del nivel educativo, con la excepción del nivel universitario completo y la reducción del valor absoluto del coeficiente negativo de las variables “mujer” y “trabajador no registrado”.

En conclusión, nuestra investigación sugiere la pertinencia de centrar el análisis de la variabilidad del salario horario en su correlación con el máximo nivel de instrucción alcanzado, la calificación de la ocupación, el carácter general de la ocupación y/o la jerarquía ocupacional (según la extensión del período estudiado, debido a la disponibilidad de la segunda de estas variables), la antigüedad, la situación de registro, el sexo, la edad y la edad al cuadrado.

3. Los casos con información faltante

Little y Rubin señalan que “los paquetes estadísticos normalmente excluyen del análisis a aquellas unidades con código de no respuesta para alguna de las variables involucradas. Esta

estrategia, tomada como regla general, es inapropiada, ya que lo que interesa al investigador es hacer inferencias sobre la población completa, más que sobre la porción que efectivamente responde a todas las variables relevantes en el análisis¹⁴. Los análisis basados simplemente en las unidades completas son sencillos de desarrollar y pueden resultar apropiados cuando la cantidad de datos faltantes es pequeña aunque, como contrapartida, pueden derivar en sesgos en las estimaciones¹⁷.

Además del peso relativo de los casos con información faltante en la muestra, al evaluarse la pertinencia de aplicar medidas que remedien la pérdida de información, resulta relevante la consideración del *mecanismo generador* de la información faltante¹⁸. A éste se lo puede clasificar según su correlación con los valores de la variable dependiente o las independientes. Si la probabilidad de que un dato de Y se encuentre faltante es independiente tanto de los valores de Y como de los de X se dirá de los datos que se encuentran *perdidos completamente al azar* (PCA). Si, en cambio, dicha probabilidad depende de los valores de X pero no de los de Y , se dirá que están *perdidos al azar* (PA): los valores observados de Y no necesariamente serán una muestra aleatoria de los casos muestrales, pero sí de los valores que se encuentren dentro de las subclases definidas por los valores registrados de X . Si, por último la probabilidad de que un dato de Y se encuentre faltante depende de los valores de Y (y, posiblemente, también de los de X) el mecanismo generador de información faltante resulta *no ignorable*, en el sentido de que los análisis basados en los casos completos producirán necesariamente estimaciones sesgadas¹⁹.

Si los casos se encuentran perdidos completamente al azar, su ausencia no obsta para que se puedan realizar inferencias válidas a partir de los casos completos de la muestra. Cuando no puede suponerse que los datos perdidos son PCA pero sí PA, las técnicas disponibles para subsanar la existencia de casos con información faltante consisten en la alteración de la *ponderación* de las observaciones según la probabilidad de respuesta, por un lado, y la imputación de los valores perdidos, por otro. Cuando, en cambio, el mecanismo generador de información faltante es no ignorable, la estimación corregida debería basarse en la maximización de la verosimilitud para el conjunto de unidades respondientes y no respondientes, sobre la base de una modelización del mecanismo generador de información faltante, antes que en procedimientos que supongan la similitud de comportamiento entre unidades observadas y no observadas con respecto a la variable de interés. Aún entonces, el sesgo por no respuesta conserva un impacto potencialmente significativo, cuya dimensión depende de los valores que se asignen a los parámetros de la función de respuesta R (es decir, de la distribución condicionada de la probabilidad de respuesta efectiva²⁰), los cuales por lo general no pueden ser fidedignamente estimados a partir de los datos. La solución óptima al problema de no respuesta no ignorable consiste en reducir la no respuesta o en acumular información acerca de cómo los no respondientes difieren de los respondientes en las variables respuesta bajo estudio²¹. El seguimiento de unos pocos no respondientes puede ser de gran ayuda para la reducción de la sensibilidad de las inferencias²².

Ha sido frecuentemente señalado que las encuestas de hogares encuentran dificultades para la

correcta captación de los ingresos de los entrevistados²². El rechazo a responder la encuesta completa o la negativa a responder las preguntas relativas a los ingresos, generan una considerable pérdida de información. Al respecto, Beccaria y Herrero, citando a Minardi (2002), afirman que “luego de aparear hogares sin respuesta en una encuesta de Argentina con su respectiva cédula censal, la probabilidad de no respuesta a la encuesta aumenta en los hogares con características asociables a mayores ingresos familiares (cantidad de coberturas de salud, habitar viviendas grandes y ser propietario de vivienda y terreno)”²³. Felcman, Kidyba y Ruffo (2003) señalan que la negativa a responder puntualmente la pregunta sobre ingresos laborales en la Encuesta Permanente de Hogares entre mayo de 1993 y mayo de 2002 se mostró “asociada a características particulares relacionadas con el ingreso”. En definitiva, se ha mostrado que, en la información de ingresos registrada mediante las encuestas de hogares en la Argentina, tanto la no respuesta total como la parcial resulta no ignorable.

De acuerdo con Felcman, Kidyba y Ruffo, el rechazo de encuesta completa afecta a alrededor del 15% de la muestra del total de aglomerados en las ondas comprendidas entre mayo de 1993 y mayo de 2002. La no respuesta parcial habría afectado, en el promedio del período, a entre el 6% y el 7% de los asalariados de todos los aglomerados relevados por la encuesta²⁵. En el caso de las muestras en que se basó nuestra investigación²⁶, los asalariados encuestados que no declararon sus ingresos (no respuesta parcial) representaron un mínimo de 0% y un máximo de 16% de los asalariados entrevistados (ver cuadro 2, fila *a*). Adicionalmente, en especial durante la primera década considerada y en los dos últimos años de la segunda, hubo asalariados que, por diversas circunstancias, no percibieron ingresos en los treinta días previos a la realización de la encuesta. Este hecho incidió en la imposibilidad del cómputo del logaritmo natural de su salario horario y en que el software estadístico los considerara como “casos perdidos” para dicha variable. Estos casos representan entre el 0% y el 8% del total de asalariados encuestados durante el período (tabla 3, fila *b*). En total, los casos sin información relativa al logaritmo del salario horario representaron hasta un 21% de los asalariados encuestados, variando este porcentaje de acuerdo con el año (tabla 3, fila *c*).

Cuadro 3: Bondad del ajuste de modelos de regresión (R^2 corregido)* GBA. 1985-2005

		1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	
Total casos		3.515	3.679	3.815	4.073	4.349	2.721	2.938	2.974	3.068	2.923	
<i>Ingreso desconocido (a)</i>	<i>n</i>	112	0	155	222	679	422	431	358	243	190	
	%	3%	0%	4%	5%	16%	16%	15%	12%	8%	7%	
<i>Sin ingreso (b)</i>	<i>n</i>	29	130	0	15	11	14	7	4	17	1	
	%	1%	4%	0%	0%	0%	1%	0%	0%	1%	0%	
<i>Total casos excluidos</i>		<i>n</i>	141	0	155	222	679	422	431	358	243	190
<i>(c = a + b)</i>		%	4%	0%	4%	5%	16%	16%	15%	12%	8%	7%
		1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Total casos		2.980	3.000	3.226	3.414	3.368	3.359	3.012	1.585	2.560	2.778	3.102
<i>Ingreso desconocido (a)</i>	<i>n</i>	173	222	191	198	239	189	262	131	371	418	394
	%	6%	6%	5%	5%	6%	6%	10%	4%	12%	15%	13%
<i>Sin ingreso (b)</i>	<i>n</i>	0	15	10	0	0	24	0	0	0	153	245
	%	0%	0%	0%	0%	0%	1%	0%	0%	0%	6%	8%
<i>Total casos excluidos</i>		<i>n</i>	173	222	191	198	239	213	262	131	371	394
<i>(c = a + b)</i>		%	6%	6%	5%	5%	6%	6%	10%	4%	12%	13%

Fuente: Elaboración propia a partir de bases usuarias EPH, octubre 1985-2002 y 4º trim. 2003-2005.

Otra fuente de distorsión de la información referida a ingresos proveniente de la EPH (y de las encuestas de hogares en general), consiste en la subdeclaración de los mismos. La probabilidad de subdeclaración también se encuentra relacionada con el nivel de ingresos del respondiente. No contamos con una estimación de la proporción de subdeclaración que puedan presentar los ingresos de asalariados registrados por la EPH, si bien cabe esperar que ésta no resulte digna de consideración.

Puesto que un propósito principal de nuestra investigación de tesis era la estimación de los coeficientes de regresión y los tamaños de los efectos correspondientes a las variables independientes consideradas, los métodos de corrección de la no respuesta propuestos por Felcman, Kidyba y Ruffo no habrían resultado adecuados en el marco de la misma, aun cuando los datos perdidos hubieran sido PA. Su corrección de la no respuesta parcial (es decir, la no respuesta a la pregunta de ingresos) se basa en la imputación del ingreso individual a partir de un modelo econométrico que incluye características del trabajador, el establecimiento y el puesto de trabajo. Esto es, a partir de las correlaciones parciales observadas que constituyeron el objeto mismo de nuestro estudio. Su corrección de la no respuesta total, por su parte, ajusta los ingresos de los perceptores ubicados en los cincuentiles superiores (en los que se concentran las diferencias entre la distribución de ingresos que surge de la información de EPH y la que tiene fuente en los registros del SIJP) de manera tal de compensar el efecto de la negativa a responder la encuesta completa por parte de hogares de ingresos altos. La corrección, que puede ser apropiada cuando el objetivo final es realizar estimaciones de totales y medias generales de ingresos, resultaría aventurada en nuestro caso, en que se buscaba estimar la forma de la distribución multivariada y sus efectos sobre los diferenciales salariales entre grupos de trabajadores.

Puesto que la información referida a ingresos no se encuentra perdida al azar, las estimaciones obtenidas en nuestra investigación resultaron sesgadas y, por lo mismo, su extrapolación al conjunto de la población asalariada del aglomerado GBA no puede realizarse sin tomar en cuenta esta salvedad. La ausencia de los datos debería remediarse considerando la probabilidad individual de no respuesta (tanto total como parcial). El desafío pendiente consiste en modelizar esas probabilidades diferenciales. El seguimiento de un número al menos pequeño de no respondientes, si bien supone grandes dificultades, constituiría una herramienta invaluable para realizar dicha modelización.

4. Los casos atípicos

Un caso atípico dentro de un conjunto de datos es aquella “observación (o subconjunto de observaciones) que parece inconsistente con respecto al resto”³⁴. El origen de tal comportamiento extraño puede deberse a circunstancias diversas: errores de declaración, transcripción o carga de la información, que genéricamente llamaremos “de medición”; inadecuación del modelo o la escala de medición escogidos para describir los datos; contaminación de la muestra por inclusión de individuos provenientes de una población diferente a la que se intentaba muestrear o, simplemente, inclusión en la muestra de casos cuya probabilidad de ocurrencia dada la distribución poblacional es baja, pero que aun así son expresión genuina de la variabilidad intrínseca de la población³⁵.

Como es sabido, los procedimientos de maximización matemática -como los utilizados en nuestra investigación-, son especialmente sensibles a la presencia de observaciones que se separen considerablemente del resto o “atípicas”³⁶. Si tales observaciones no resultaran ser miembros genuinos de la población principal, podrían contaminar severamente las estimaciones o pruebas de parámetros de la representación postulada³⁵.

En los modelos multivariados, “un caso atípico [...] puede distorsionar no sólo las medidas de posición o de escala, sino también aquéllas de orientación (es decir, la correlación)”³⁷. En los diseños experimentales como ANOVA, la presencia de casos atípicos puede determinar la aparición de efectos significativos aparentes o encubrir efectos significativos reales al incrementar la suma de cuadrados del error³⁷. En la estimación de regresiones, los valores atípicos pueden, dependiendo de su ubicación en el espacio de X , afectar la altura promedio de la línea de regresión o sesgar la estimación de la pendiente.³⁸

La detección de tales casos normalmente se basa en el examen de los efectos de la eliminación de una cantidad m de observaciones (generalmente una) por vez. En nuestra investigación, se realizó mediante el examen de la distancia de Cook y los valores DFFITS³⁹.

Frente a la presencia de casos atípicos, son habituales dos tipos de abordaje: 1) la utilización de una batería de técnicas de detección para su estudio detallado, con la posibilidad de excluir algunos o todos del análisis y 2) el desarrollo de procedimientos de estimación relativamente

insensibles a los casos atípicos o “técnicas de regresión robusta” [40][41].

El tratamiento que se dé a los casos atípicos deriva de la identificación de la fuente que los produce. Por esto, Cook señala que, pese a que la detección de casos atípicos y el uso de métodos robustos aparecen compitiendo entre sí, ellos no constituyen abordajes mutuamente excluyentes [42]. La utilización *a priori* de técnicas de estimación robusta puede resultar en la pérdida de información valiosa que las observaciones que se alejan de la distribución esperada pueden brindar. En particular en nuestra investigación, la utilización de estimaciones robustas a la presencia de casos atípicos conllevaba el riesgo de subestimar la dispersión efectiva del salario horario.

De acuerdo con Barnett y Lewis, cuando se está frente a un error de medición obvio, la exclusión del caso extremo o su corrección están justificadas. En este caso, el procesamiento de la observación atípica no depende de análisis estadísticos, sino que debe emplearse el “ingenio nativo” [43]; la corrección no genera dificultades particulares y la realización de una prueba de discordancia resulta innecesaria [44]. Si el problema se encuentra en la representación propuesta o en la escala de medición de las variables involucradas, son ellas, evidentemente, las que deben ser corregidas. Cuando el caso atípico surge de la inclusión en la muestra de individuos de otra población, su detección podría derivar en la exclusión de los mismos, pero también dar lugar a un modelo modificado; por ejemplo, uno de carácter mixto o que contemplara fuentes de variabilidad no consideradas previamente [45].

En nuestra investigación, sólo fueron tratados los casos atípicos detectados que resultaban atribuibles a errores de medición. Éstos se agruparon en ocho categorías no mutuamente excluyentes, según su origen presunto y los indicios principales en que se basó el diagnóstico. Existieron casos atípicos que reunían características correspondientes a más de una categoría; a éstos se los clasificó de acuerdo con la característica que se consideró más relevante para la inferencia de su origen. Los casos pertenecientes a un mismo grupo recibieron, por lo general, el mismo tipo de tratamiento, salvando pocas excepciones en las que la información disponible permitía o impedía conservar el caso en situaciones en que normalmente se lo hubiera excluido o conservado, respectivamente.

La investigación detectó ocho tipos básicos de casos atípicos producto de errores de medición. Estos son:

1) Incongruencia entre la duración declarada de la última jornada semanal y el monto del salario correspondiente a los últimos treinta días [46], con la consiguiente distorsión del promedio salarial horario [47]. Estos casos fueron excluidos del análisis. Se dio igual tratamiento a los casos en que la escasa duración de la jornada semanal dificultaba la comparabilidad del salario horario con el del resto de los asalariados de la muestra.

2) Anomalías no especificadas, señaladas por el INDEC en la base de datos mediante el registro de

ingresos iguales a cero en la variable correspondiente al ingreso total de fuente laboral de asalariados con una sola ocupación que, en cambio, mostraban datos en las variables correspondientes al ingreso mensual de la ocupación principal, la duración de la jornada semanal y el ingreso horario de la ocupación principal (variable "inghor"). Por tratarse de personas con sola ocupación, el ingreso total de fuente laboral debería haber sido coincidente con el de la ocupación principal. Este hecho, combinado con la discordancia que condujo a examinar el caso, nos llevó a excluir estos casos del análisis.

3) Declaración del ingreso correspondiente a períodos inferiores o superiores al mes, sin que esto fuera considerado en el cómputo del ingreso horario, con la consiguiente distorsión en el resultado. A raíz de la detección de estos casos, se corrigió el ingreso horario de todas las observaciones, a partir de la mensualización previa del monto de los ingresos.

4) Probable declaración de la jornada semanal correspondiente a todas las ocupaciones y el ingreso correspondiente a una sola, o viceversa, en asalariados con más de una ocupación. En primera situación, se recalculó el ingreso horario a partir del monto declarado como ingreso total como asalariado (p47_1), si éste excedía del declarado como ingreso de la ocupación principal (p21); en la segunda, se excluyó el caso del análisis cuando no se disponía de información sobre la jornada semanal de las otras ocupaciones (bases 1985-1994) o si la consideración conjunta de la cantidad de ocupaciones y la diferencia entre ingresos como asalariado e ingreso de la ocupación principal no permitía establecer si todas las ocupaciones secundarias eran asalariadas.

5) Errores en la declaración/registro de la condición de registro o del ingreso de asalariados que declararon encontrarse plenamente registrados y cuyo ingreso mensual no se correspondía con el tipo de ocupación (calificación, jerarquía) que declaraban. Estos casos fueron, por lo general, excluidos del análisis.

6) Inconsistencia de la información referida a la ocupación (distinta del ingreso). Por ejemplo, combinaciones incongruentes de dos o más de los siguientes atributos de los asalariados, sus ocupaciones y los capitales que los empleaban: nivel educativo, tamaño del establecimiento, rama de actividad, condición de registro, carácter de la ocupación, jerarquía, antigüedad en el puesto, jornada semanal. Estos casos fueron excluidos del análisis sólo si el grado de inconsistencia así lo ameritaba.

7) Errores de carga en la variable correspondiente al ingreso mensual de la ocupación principal en individuos *con una sola ocupación*, detectados por comparación con el ingreso total de fuente laboral. En estos casos, se reemplazó el valor del ingreso mensual con el del ingreso total y se recalculó el salario horario.

8) Errores diversos no correspondientes a ninguna de las categorías anteriores.

En la tabla 4 se resume la información referida, indicando el peso relativo de cada tipo de error con respecto al total de casos atípicos detectados en las veintiuna muestras sobre las que se trabajó.

Cuadro 4: Casos atípicos de regresión multivariada identificados en las bases de EPH onda octubre 1985-2002 y cuarto trimestre 2003-2005, según origen de la discordancia

Origen	Tipo de error principal	N	%	Tratamiento mayoritario
Errores de medición	1	47	17%	Exclusión
	2	18	7%	Exclusión
	3	9	3%	Corrección
	4	13	5%	Corrección
	5	11	4%	Exclusión
	6	13	5%	Conservación/Exclusión
	7	15	5%	Corrección
	8	8	3%	Varios
Subtotal		134	49%	
Otro origen		140	51%	Conservación
Total		274	100%	

Fuente: Elaboración propia a partir de EPH, ondas octubre 1985-2002 y cuarto trimestre 2003-2005.

Los casos atípicos corregidos o eliminados representaron el 0,2% sobre el total de las observaciones de las veintiuna muestras utilizadas en la investigación.


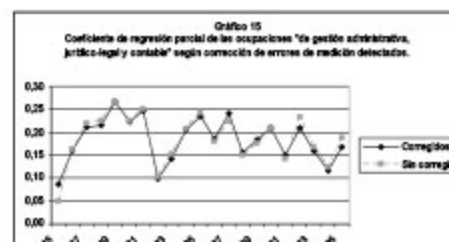
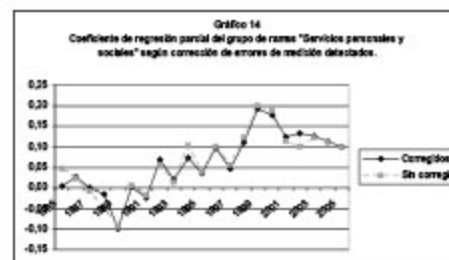
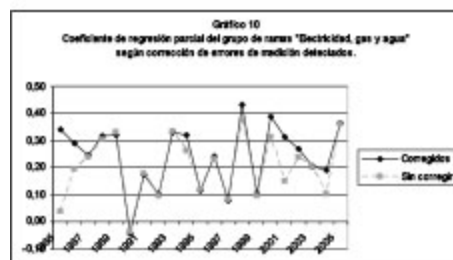
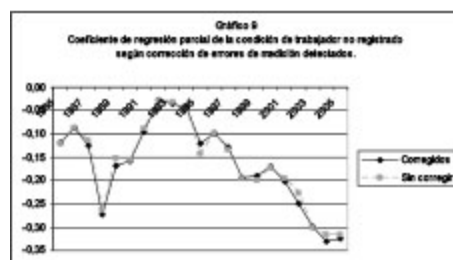
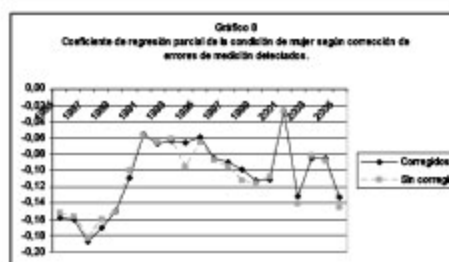
La comparación de las series de coeficientes de correlación parcial según se corrija o no los casos discordantes debidos a errores de medición muestra que, pese a sus altos valores en el estadístico de Cook y en el DFFITS, la eliminación o corrección de estos casos no altera sustancialmente los resultados de la regresión (gráficos 1 a 19) .

Lámina de gráficos 1: Coeficiente de regresión parcial para cada variable del modelo según corrección de errores detectados



ARTÍCULOS

¿Qué son los Indicadores? Perspectivas y usos diferentes

Ana María Capuano¹

método sociológico

indicadores

El presente artículo forma parte de la tesis de Maestría en Metodología de la Investigación Social denominada: 'Sistema de indicadores sociales sobre niñez y adolescencia. Propuesta conceptual y metodológica para la Argentina'.

Tanto la tesis mencionada, así como el presente artículo surgen a partir de reflexiones acerca de mi práctica profesional, en la cual diseñaba indicadores de distintas temáticas sociales, que formaban parte del Sistema de Estadísticas sociodemográficas, realizado en el Instituto Nacional de Estadística y Censos. En esa práctica, observe que la realización de indicadores tenía como objetivo generar información para el diagnóstico y seguimiento de los cambios operados en la situación social, así como ser un insumo para el monitoreo de las políticas públicas. Por otro lado, desde la perspectiva de la investigación en ciencias sociales, los indicadores refieren a la operacionalización de conceptos para medir aspectos de la realidad social, siendo un eje de debate epistemológico la idea de medir aspectos sociales como si fueran fenómenos naturales.

Desde esa escisión observada, es de donde parte este artículo, haciendo un recorrido por el concepto de indicador, desde las ciencias sociales así como de los orígenes de la construcción de indicadores 'sociales' desde una perspectiva ligada a los gobiernos y a los organismos públicos productores de estadísticas. A lo largo del artículo se observará que ambas perspectivas se desarrollaron de manera aislada, pero sin embargo, como la medición en ciencias sociales a través de indicadores empíricos necesita de la construcción de información, se concluye que se presenta necesario vincular ambas perspectivas.

2. Los indicadores como referentes empíricos de la operacionalización de conceptos. Antecedentes en la investigación social

La operacionalización de conceptos constituye uno de los principales temas en el proceso de investigación empírica en las ciencias sociales, ya que remite al problema de la medición. De acuerdo con la definición clásica del término medir significa asignar números a objetos y eventos de acuerdo a reglas (Stevens, 1951). En las ciencias sociales podría definirse como el proceso que vincula conceptos abstractos con indicadores empíricos.

Entre los principales referentes en las ciencias sociales se mencionan a Lazarsfeld y Merton, aunque hay antecedentes en la obra de Durkheim. En términos generales, los autores mencionados definen a los indicadores como instrumentos que se utilizan para el abordaje de conceptos en la investigación social.

Si bien Durkheim, uno de los principales científicos sociales afines al positivismo, no utiliza en su obra terminología explícita en relación a la operacionalización de conceptos a través de indicadores, se observan en sus trabajos aproximaciones a la temática cuando afirma que es necesario sustituir el “hecho interno” que se nos escapa por un “hecho externo” que lo simbolice y estudiar de esta manera el primero a través del segundo (Merton, 2002).

En este sentido Durkheim enuncia la siguiente regla: al iniciar la indagación de los hechos sociales, el estudioso debe dar una definición operativa de aquello que va a estudiar. Para ser objetiva, la definición debe obviamente considerar los fenómenos no como ideas, sino en términos de sus propiedades intrínsecas (Bottomore y Nisbet, 1988: 248).

En “Las Reglas del Método Sociológico” Durkheim argumenta que se necesita elaborar del principio al final, conceptos nuevos, ajustados a las necesidades de la ciencia y expresados con la ayuda de una terminología especial. Para este autor, la sociología aborda temas a los que nos referimos constantemente tales como la sociedad, la familia, el delito, entre otros y por lo tanto, el sociólogo no considera necesario ofrecer una definición previa ni rigurosa de los mismos, hecho que deriva en cierta ambigüedad y en el agrupamiento de cosas distintas bajo la misma definición.

Se observa entonces en Durkheim una primera aproximación a la necesidad de que los conceptos utilizados en el lenguaje común se transformen en conceptos científicos para las ciencias sociales.

En el siglo XX, la operacionalización de conceptos se convierte en un tema central de las ciencias sociales, principalmente dentro del desarrollo de la metodología empírica, siendo su principal referente Lazarsfeld.

Siguiendo los postulados del positivismo, Lazarsfeld pretendía desarrollar las ciencias sociales tomando como modelo el método de las ciencias naturales. Para ello, las primeras debían crear

un lenguaje de observación que consistiera en observables objetivamente definidos, generalizables y cuantificables, que permitieran conocer los atributos de los fenómenos.

Lazarsfeld propuso una estrategia empírica para poder identificar propiedades generales a través de las cuales los fenómenos fueran conocidos, tratándolas como variables -el llamado "lenguaje de las variables"- . Los términos básicos del lenguaje científico debían expresar cualidades generales de los objetos, de ahí que la investigación social adoptó términos cuasi matemáticos para hablar de los fenómenos sociales (Hughes y Sharrock, 1999).

La identificación de las variables fue un paso innovador, convirtiéndose en el eje en torno al cual surge una nueva manera de planificar y practicar la investigación social. Para la investigación social, el lenguaje de las variables ofrecía un medio de expresar relaciones entre las propiedades de los objetos (personas, instituciones, países, etc.) y como tal una manera de describir objetiva y cuantitativamente los fenómenos sociales. Todos los fenómenos de interés para la investigación social podían ser conceptualizados y medidos al menos en cierto nivel, correlacionados y manipulados de diversas maneras por las técnicas formales del análisis de variables, permitiendo formular y poner a prueba hipótesis (Hughes y Sharrock, 1999).

Lazarsfeld consideraba que las ciencias no abordan su objeto de estudio en su totalidad, sino que seleccionan determinadas propiedades del mismo e intentan establecer entre ellas relaciones recíprocas, siendo una característica de las ciencias sociales que la elección de las propiedades estratégicas se constituye en un problema esencial (Lazarsfeld, 1979).

El autor describe la operacionalización de conceptos como un proceso compuesto de cuatro etapas. En un primer momento, el concepto no es más que una entidad concebida en términos vagos, que confiere un sentido a las relaciones entre los fenómenos observados e intenta explicar las regularidades observadas. Debido a la complejidad que presentan los conceptos en ciencias sociales, el segundo paso consiste en dividirlo en dimensiones. El tercer momento es aquel en que se seleccionan los indicadores, que cubrirán las dimensiones del concepto (equivalentes empíricos de las dimensiones). Esta es la fase de la definición operacional. El indicador es un instrumento que posee una determinada relación de probabilidad con respecto al concepto, por lo tanto es necesario utilizar un número importante de indicadores para dar cuenta del mismo. Esta es una de las etapas más complejas del proceso, ya que determinar los criterios que limitarán la elección de indicadores constituye un problema. Es decir, cuando seleccionamos los indicadores ¿debemos considerarlo parte del concepto o como fenómenos independientes de este? Para encontrar una respuesta es necesario, en términos de Lazarsfeld, conocer las leyes que presiden las relaciones entre indicadores. Dentro de esta etapa se necesitan realizar pruebas de validación para poder dar cuenta de la existencia de relación entre los distintos indicadores de un concepto. La última fase consiste en reunir los indicadores obtenidos en las etapas precedentes, ponderarlos y articularlos construyendo una medida única, un índice.

Lazarsfeld fue criticado, entre otras cosas, porque se consideraba que el lenguaje de las variables

era intencionalmente ateoórico, su método resultaba una estrategia para la formación de teorías empíricas que no llegaban a ser el tipo de generalizaciones teóricas que eran el objetivo de las ambiciones de los positivistas (Hughes y Sharrock, 1999).

Posteriormente, fue Robert Merton, referente teórico del funcionalismo ², quien a diferencia de Lazarsfeld, consideró necesario que los estudios empíricos estuvieran orientados por la teoría y que ésta pudiera ser demostrable empíricamente. El interés de Merton se centraba en que hubiera una mayor vinculación entre la teoría y la investigación empírica.

Para Merton los conceptos constituyen las definiciones de lo que debe observarse, y entre las variables se deben buscar relaciones empíricas, en ese sentido afirma que la elección de conceptos que guían la recolección y el análisis de dato ello, tiene que idear indicadores que sean observables, bastante precisos y meticulosamente es decisiva para la investigación:

Una de las funciones del esclarecimiento conceptual consiste en hacer explícito el carácter de los datos subsumidos en un concepto (Merton, 2002: 166).

Merton destaca la importancia de la investigación empírica dentro de la investigación social. Argumenta que en general, la clarificación de conceptos se considera propia del teórico cuando en realidad es resultado de una investigación empírica y se explicita a través de la construcción de indicadores de los conceptos estudiados. Además, agrega que una tarea del análisis conceptual es establecer indicadores observables de los datos sociales que interesan a la investigación empírica, y que el problema consiste en idear indicadores de entidades inobservables o simbólicas (por ejemplo, cohesión social) que puedan sustentarse teóricamente.

En la misma línea que Durkheim, Merton argumenta que en definiciones no empíricas, es posible hablar vagamente de “moral”, “cohesión social”, sin una especificación de lo que implican esos términos, pero deben ser aclarados si el investigador pretende observar sistemáticamente tipos o grados de moral, de cohesión o de descomposición social. Para ello, tiene que idear indicadores que sean observables, bastante precisos y meticulosamente ³claros. Para Merton, el establecimiento de indicadores válidos y observables, es fundamental para el uso de conceptos en la investigación, porque:

El proceso de investigación empírica suscita cuestiones conceptuales que pueden pasar inadvertidas durante mucho tiempo en la investigación teórica (Merton, 2002: 193).

Lo expuesto hasta aquí es una síntesis de algunos de los referentes de la operacionalización de conceptos en la investigación social a través de indicadores. Sin embargo, se observa que los indicadores también son utilizados desde una perspectiva diferente a la descrita, a través de lo que se llamó movimiento de indicadores sociales, que se describe en el siguiente punto.

3. Movimiento de indicadores sociales. Antecedentes, surgimiento y consolidación

Se conoce como el “Movimiento de los Indicadores Sociales” (MIS), a la opción metodológica adoptada para el estudio de la “calidad de vida” a través de su medición empírica. Se suele ubicar el origen de este movimiento a mediados de la década del '60 en Estados Unidos.

El MIS fue un intento ambicioso para producir mediciones “precisas” y “neutrales”, evaluando el estado de la sociedad y sus cambios, a partir del uso de una variedad de datos, muchos de ellos originados en el gobierno. Se consideró importante monitorear cambios en el tiempo de un sinnúmero de cualidades de la vida, tanto a nivel de la población como un todo, como para subgrupos significativos. Esa información, combinada con otros datos, podía generar nuevo conocimiento acerca de cómo incrementar la calidad de vida por medio de políticas sociales más efectivas [Bulmer, M, 2001].

El MIS, adopta la definición de indicador propuesta por Raymond Bauer, el cual los define como “estadísticas, series de estadísticas y otras formas de prueba que nos permiten evaluar en dónde nos encontramos y hacia dónde nos dirigimos en relación con valores y objetivos” (Bauer, 1966: 1).

Si bien se encuentran antecedentes de trabajos con indicadores, en la misma línea del MIS, desde principios de siglo XX, mencionando entre algunos de ellos, el de Alfredo Nicotero y el de William Ogburn⁷, es a partir de los años '50 que comienza a producirse la demanda desde los gobiernos de contar con una metodología que evaluara aspectos económicos y pudiera medir el grado de desarrollo y evolución de las dimensiones sociales. En Estados Unidos, se genera interés por desarrollar un sistema global de información conocido con el nombre de “contabilidad social”, cuyo objetivo era medir el estado y evolución de la sociedad. Este enfoque contable, que surge a imagen de las cuentas nacionales, intentaba medir el bienestar, considerando éste concepto afín al de riqueza, renta nacional y renta per cápita.

Sin embargo, el Producto Bruto Interno (PBI)⁸ utilizado para medir el bienestar social empieza a recibir cuestionamientos ya que se consideraba que la riqueza no era sinónimo de bienestar. Entre las debilidades del PB⁹ como indicador social se destacaba en primer lugar, la incapacidad de distinguir entre actividades constructivas y destructivas, en segundo lugar, la dificultad para considerar la sustentabilidad a largo plazo y, finalmente, la imposibilidad para tomar en cuenta actividades productivas que ocurren fuera del mercado.

El cuestionamiento al PBI como indicador del desarrollo se sustentaba en que el rápido crecimiento económico traía un número de desigualdades que debían estudiarse por medio de otros indicadores que lo complementarían. A comienzos de los '60 se produjo, entonces, un crecimiento de la demanda de información social la cual contribuyó a describir aspectos no conocidos o poco estudiados de la realidad social y de los distintos grupos. La elaboración de esa información tenía como objetivo utilizarse para abordar problemas, formular metas y diseñar políticas sociales.

El objetivo que estaba en el origen de los nuevos abordajes era evaluar en qué medida tanto el crecimiento económico como la ampliación de derechos que expresaba el Estado de Bienestar excluía, en el usufructo de beneficios, a distintos grupos de población. Para poder analizar la "calidad de vida" se necesitaba tener un mayor conocimiento sobre los problemas y grupos sociales. Por todo lo dicho es que, el estudio sobre las consecuencias sociales del desarrollo económico a través de indicadores debía presentar cambios fundamentales. Este ya no podía ser abordado sólo por medio de indicadores económicos tradicionales, porque la "calidad de vida" implicaba algo más que las consideraciones económicas.

Siguiendo a Noll (2002), el crecimiento y la rápida difusión del MIS estaba relacionado, en parte, con el clima político de fines de la década del '60 y principios de los '70, y la prosperidad económica que caracterizó ese período, principalmente en los Estados Unidos.

En ese sentido, se planteó la necesidad de incluir indicadores que mostraran los niveles de vida alcanzados por la población, además de los indicadores tradicionales de factores de producción que reflejaban la distribución presupuestaria, procedimientos y procesos que se presume aumentan el bienestar (Andrews, 1989: 401).

Se ensayaron entonces la construcción de medidas estándar sobre el estado de la salud, el crimen, el bienestar, la educación y otras características sociales de la población (Bulmer, 2001: 7).

En este momento se afianza el MIS, presentando un importante impulso, no sólo desde los gobiernos y los organismos internacionales, sino también desde el ámbito científico y académico.

El rol de los organismos e instituciones internacionales fue clave en la determinación tanto de las líneas de investigación así como las pautas metodológicas, y contribuyeron de manera activa al desarrollo y consolidación del MIS.

Los distintos programas que se emprendieron en Europa en la década del '70 utilizaron la experiencia desarrollada por las instituciones mencionadas. Entre ellos se destacan los programas de la Organización para la Cooperación y el Desarrollo Económico (OCDE) y del Departamento de Asuntos Económicos y Sociales de Naciones Unidas, quienes aportaron las normativas metodológicas para la sistematización de la información estadística, así como los

procedimientos para abordar el análisis de la misma.

El desarrollo de los indicadores sociales en las Naciones Unidas fue promovido en el seno de la Comisión de Estadística del Consejo Económico y Social a través de la elaboración del Sistema de Estadísticas Sociodemográficas (SESD). El diseño del SESD constituyó un avance respecto de los trabajos previos, en el sentido que el concepto de sistema aportó una perspectiva macrosocial que se transformó en una herramienta de análisis de la sociedad concebida como un todo.

Entre las características que debía tener un sistema de indicadores sociales se mencionan:

- Estructurar la información social, mostrando qué datos es conveniente reunir sobre los seres humanos, tanto individualmente como en grupos, y sobre las instituciones con que están relacionados;
- Servir de marco de referencia a los organismos de estadística nacional garantizando un conjunto de estadísticas sociales integradas; mostrando cómo deben organizarse esos datos para obtener un sistema de información que sea útil para describir, analizar y adoptar políticas en las diferentes esferas de la vida social (Naciones Unidas, 1975: 3).

Siguiendo a Setién (1993), a partir de mediados de la década del '70, el MIS presenta tres cuestiones de distinta índole y nivel de complejidad. Se mencionan:

1. El surgimiento de actores y de problemas sociales que presentan dificultades para ser estudiados con los abordajes metodológicos tradicionales;
2. El desarrollo de programas de alcance regional y mundial que exigen parámetros de calidad para lograr homogeneidad en la elaboración de indicadores y que permitan su comparabilidad;
3. La persistencia de los problemas que tuvo el MIS desde sus inicios, entre ellos a) la falta de una teoría social que sustente el trabajo desarrollado; b) la diversidad de tipos de sistemas, modelos, etc. producida dentro del mismo movimiento, hecho que demuestra la falta de criterios unificados; c) la dificultad de encontrar una unidad común de medición; d) la primacía en los distintos enfoques de lo económico por sobre lo social.

Respecto al primer punto, se observa desde la década del '80 en el contexto internacional, cambios cualitativos, cuantitativos y temáticos en torno a los indicadores sociales basados en una serie de hechos:

a. Los foros internacionales desarrollados en la década del '90 (Cumbre Mundial a Favor de la Infancia, 1990; Cuarta Conferencia Mundial sobre la Mujer, 1995; Conferencia de las Naciones Unidas sobre el Medio Ambiente y el Desarrollo, 1992, Conferencia Internacional sobre Población y Desarrollo, 1994, entre otras) que han mostrado la creciente desigualdad social a nivel mundial y

la necesidad de contar con información específica para diversas temáticas, así como para determinados grupos poblacionales (niños y mujeres, entre otros). En estos encuentros internacionales se establecen metas que los gobiernos deben cumplir al final de un período y el cumplimiento de las mismas se evalúa a través de indicadores.

b. El surgimiento de nuevos movimientos sociales que reivindican distintos tipos de derechos. Por ejemplo, los movimientos de mujeres tuvieron un papel importante al reclamar indicadores de diferenciación socio-espacial por género. Esta reivindicación propició que los indicadores dejaran de ser globales para pasar a diferenciarse por grupos.

c. Conceptos como el de “desarrollo sustentable” o “derechos humanos” que exigen nuevos marcos conceptuales para la elaboración de indicadores.

En relación con el segundo punto, se observan distintos programas de indicadores sociales tanto en los organismos internacionales así como en los institutos de estadística. Se mencionan, en el contexto de Europa, el programa de la Comisión de la Comunidad Económica Europea, con publicaciones tales como “Europa en Cifras”, así como el Anuario Eurostat. En el contexto latinoamericano se destacan el Sistema de Estadísticas Sociodemográficas (SESD) en Argentina elaborado por el Instituto Nacional de Estadística y Censos (INDEC); Indicadores Sociodemográficos del Instituto Nacional de Estadística, Geografía e Informática de México (INEGI); Síntesis de Indicadores Sociales del Instituto Brasileño de Geografía y Estadística (IBGE), entre otros.

Desde los organismos internacionales, la Organización Panamericana de la Salud (OPS) ha desarrollado un Sistema de Datos Básicos de Salud para la región, el Fondo de las Naciones Unidas para la Infancia (UNICEF) elaboró un conjunto de indicadores sobre la situación de los niños y niñas para monitorear el progreso hacia las metas y objetivos fijados en la Cumbre Mundial de la Infancia, además de los programas de la Organización para la Cooperación y el Desarrollo Económico (OCDE) y la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), mencionando en el contexto de América Latina, el Sistema de Información de Tendencias Educativas (SITEAL).

Por último, la tercera cuestión mencionada incluye algunos de los problemas que se presentan como los más complejos de solucionar dentro del MIS.

a. Si bien no pueden desconocerse los avances logrados en el desarrollo de los indicadores sociales a partir de la experiencia previa del MIS y sus antecedentes, los mismos se enfrentan a la dificultad de la ausencia de un marco teórico y conceptual con sustento en las ciencias sociales que a través de conceptos y sus relaciones valide la propuesta de los indicadores sociales.

b. Las distintas orientaciones en los sistemas de indicadores sociales, existiendo principalmente dos tipos de enfoques, los que destacan la medición objetiva (como la Escuela Escandinava) y los

que priorizan la medición subjetiva, más característicos de Europa occidental y Estados Unidos. Los distintos tipos de enfoques demuestran la falta de homogeneidad que rige el diseño y la elaboración de los indicadores sociales en el marco del MIS.

c. El intento de encontrar una unidad común en el ámbito social, conlleva el problema de que no existe un sistema preconcebido de cómo deben funcionar las interrelaciones y los comportamientos sociales para un país en particular. Además, las personas tienen distinto tipo de aspiraciones, lo cual aumenta la complejidad de la dinámica social (Gutiérrez Espeleta, 2002).

Siguiendo a Bulmer, la principal dificultad que tuvo la construcción de una medida fue la ausencia de una unidad común de medición en relación con los fenómenos sociales, tales como la educación, la salud, el crimen, a diferencia de los indicadores económicos que tienen como unidad común de medición el dinero (Bulmer, 2001).

d. Si bien los indicadores son una herramienta de probada utilidad para el estudio de temas sociales, muchos de ellos siguen siendo abordados desde una perspectiva puramente económica; por ejemplo, la pobreza, en la mayoría de los casos es analizada desde la perspectiva de los ingresos, mostrando un papel auxiliar de lo social a lo económico. Esto explica que los indicadores sociales han estado subordinados a explicar cómo se invierten los recursos y no la satisfacción de las personas en las sociedades (Gutiérrez Espeleta, 2002).

El MIS, tuvo dificultades en la generación de un sistema que emulara al sistema de indicadores económicos, sin embargo, no debe desconocerse que la estadística social ha mostrado importantes avances en los últimos años, no sólo en Europa y Estados Unidos, sino en varios países de América Latina. Esto se observa principalmente, en la mejora alcanzada en las estadísticas administrativas y encuestas que proporcionan los datos para la construcción de los indicadores, el desarrollo de sistemas de indicadores sectoriales, así como de programas de armonización estadística internacional, entre otros. Los avances mencionados se observan de una manera más acentuada en el plano técnico metodológico que en el teórico conceptual.

4. Usos e interpretaciones diferenciales del concepto de indicador en el movimiento de indicadores y en las ciencias sociales

Si bien ambas perspectivas utilizan a los indicadores como herramientas para la medición, ambas presentan distintas concepciones sobre el método a utilizar para el conocimiento de la realidad social.

Mientras la sociología problematizó el proceso de operacionalización de conceptos teóricos a través de los indicadores para aproximarse a la medición de los fenómenos sociales, el MIS tuvo una visión pragmática sobre la elaboración y el uso de los mismos.

En tanto el enfoque metodológico de las ciencias sociales parte de la relación que el indicador

mantiene con una dimensión conceptual de alguna teoría social, la forma de aproximación a la realidad social para el MIS se basa en la aplicación del método inductivo, el cual supone que las observaciones y la medición preceden al concepto a evaluar. El método inductivo supone un aprovechamiento de las fuentes de datos disponibles para la elaboración de los indicadores, sin contemplar las dimensiones teóricas correspondientes a la temática que se investigue.

Los indicadores elaborados en el marco del MIS han sido un recurso focalizado en los procesos de toma de decisiones, teniendo como objetivo definir prioridades y metas, evaluar programas sociales y ser un insumo para la planificación de políticas.

Por otro lado, en las ciencias sociales, los indicadores son instrumentos de análisis conceptual y el método seguido es predominantemente el deductivo, en donde el concepto precede a la observación y a la medición. El método deductivo parte de un desarrollo teórico conceptual y tras el diseño de los indicadores, se plantea la posibilidad de construirlos.

Casas justamente enfatiza la diferente concepción sobre los indicadores que existe entre el MIS y la investigación social. En este sentido afirma:

Mientras que para el MIS se llega a la conceptualización a partir de la observación y la medición (método inductivo) y los indicadores son datos obtenidos a partir de la observación empírica que relacionan con un concepto; para la investigación social es condición sine qua non al uso de los indicadores alguna teoría científica que aporte, no sólo los conceptos, sino también, sus relaciones, siendo los indicadores, equivalentes empíricos de una de las dimensiones del concepto (Casas Aznar, 1989: 88).

5. Conclusiones

Es importante destacar que entre las perspectivas descritas a lo largo del artículo, existen relaciones aunque se podría argumentar que es necesaria una mayor articulación. Las encuestas realizadas por los organismos de estadística suelen ser el insumo utilizado por muchas de las investigaciones sociales en la actualidad, eso a pesar de que como advierte Marradi cada vez que se analizan datos de una investigación ajena, se está limitado a trabajar con las propiedades que le interesaron al autor de esa investigación, lo que lleva a que algunas propiedades que puedan ser relevantes para otros investigadores no se hayan registrado (Marradi, 2007). Sin embargo la dificultad que se les presenta a los investigadores sociales para que releven la información para cada una de sus investigaciones, lleva a que las encuestas realizadas por los organismos públicos sean de una enorme utilidad.

Para solucionar uno de los principales problemas mencionados del MIS que es la falta de articulación entre la teoría y el diseño de indicadores, sería deseable la vinculación entre los organismos de estadística y los institutos de investigación, para que el abordaje a los temas sociales partan de marcos teóricos conceptuales consensuados y a partir de allí se realice la

medición empírica de los fenómenos sociales.

Siguiendo a García Martínez, esto implica que el modelo teórico no debe construirse ajeno a la disponibilidad de fuentes de información, así como se debe tratar al mismo tiempo que la elección de los indicadores esté regida por el concepto que se pretende operacionalizar (García Martínez, 2000).

Bibliografía

- ANDREWS, F (1989). "The evolution of a movement", en: Journal of Public Policy, volumen 9, número 4, páginas 401-405. Cambridge University Press. London.
- BAUER, R. (1966). Social Indicators. Cambridge, The MIT Press. Massachussets/London.
- BOTTOMORE, T y NISBET, R (1988). Historia del Análisis Sociológico. Amorrortu. Buenos Aires. Argentina.
- BULMER, M. (2001). "Social Measurement: what stands in its way ?", en: Social Research: an international quarterly of the social sciences, Vol 68, no 2, pp.455-480. London.
- CASAS AZNAR, F. (1989). Técnicas de investigación social: los indicadores sociales y psicosociales. Teoría y práctica. Editorial Promociones y publicaciones universitarias. España.
- CHALMERS, A. (2000). ¿Qué es esa cosa llamada ciencia?. Editorial Siglo XXI. Buenos Aires. Argentina.
- DURKHEIM, E. (1987). Las Reglas del Método Sociológico. Editorial La Pleyade. Buenos Aires. Argentina.
- GARCÍA MARTÍNEZ, M. A (2000). "Sistemas de indicadores sociales. Una aproximación desde la estadística oficial", en: MECOVI 6to. Taller Regional. Indicadores sobre el desarrollo social. INDEC-CEPAL. Buenos Aires. Argentina.
- GUTIÉRREZ ESPELETA, E. (2002). Indicadores sociales: Una breve interpretación de su estado de desarrollo. FLACSO. Costa Rica.
- HUGHES, J., SHARROCK, W. (1999). La filosofía de la investigación social. Fondo de Cultura Económica. D.F. México.
- LAZARSFELD, P, BOUDON, R. (1979). "De los conceptos a los índices empíricos" en: Metodología de las ciencias sociales. Barcelona. España.
- MARRADI A, ARCHENTI, N, PIOVANI, J. (2007). Metodología de las Ciencias Sociales. Emece. Buenos Aires. Argentina.
- MERTON, R. (2002): Teoría y estructura social. Fondo de Cultura Económica. D.F. México.
- NOLL, H. (2002). "Social indicators and quality of life research", en: Background, achievements and current trends. In Advances in Sociological Knowledge over Half a Century, International Social Science Council, Paris.
- SETIÉN, M.L. (1993). Indicadores sociales de calidad de vida. Un sistema de medición aplicado al País Vasco. Centro de Investigaciones sociológicas. Madrid. España.
- STEVENS, S.S. (1951). Mathematics, measurement and psychophysics, en: "Handbook of Experimental Psychology" Wiley. New York.