



Universidad de Buenos Aires  
Facultad de Ciencias Sociales  
Carrera de Sociología

## Documento de Cátedra N°57

### ANÁLISIS DE DATOS SECUNDARIOS CUANTITATIVOS

**Alejandra Otamendi y María Pía Otero**  
**(2009)**

**Cátedra de Metodología y Técnicas de la Investigación Social**  
**Profesora Titular: Ruth Sautu**

Este Documento de Cátedra forma parte de una serie que tiene como propósito contribuir a la formación de alumnos de la Carrera de Sociología de la Universidad de Buenos Aires en cuestiones vinculadas al diseño y realización de investigaciones científicas en el campo de las ciencias sociales. Su contenido complementa los textos de metodología de lectura obligatoria y optativa incluidos en cada uno de los tres cursos que se dictan en la Carrera.

Estos documentos son material de uso interno y no pueden ser incorporados a ediciones impresas ni reproducidos comercialmente.

La Cátedra solicita a los usuarios de estos Documentos que citen a sus autores indicando las referencias completas, es decir: autores, fecha, título, número y tipo de documento (traducción, resumen, elaboración propia, etc.). En los casos en que el usuario utilice sólo parte del documento, haciendo referencia a algunos de los autores/obras originales allí incluidos, consignar que fue tomado de nuestro Documento de Cátedra. Por ejemplo:

E. O. Wrigth (1985), *Classes*, London: Verso, citado en Documento de Cátedra II.1., Plotno, G., Lederman, F. & Krause, M. (2007) "Escalas Ocupacionales".

## Introducción

El presente documento tiene un objetivo sustantivo y otro pedagógico. El primero consiste en argumentar acerca de la relevancia central que adquiere en la actualidad el análisis de datos secundarios, en particular cuantitativos, como estrategia poderosa de aprovechamiento de datos producidos por otros a fin de responder a nuevos objetivos de investigación; labor que plantea una serie de ventajas y que, a su vez, exige estar advertidos de posibles obstáculos. El segundo objetivo se refiere a la producción de una herramienta de estudio y de discusión acerca de la potencialidad del análisis de datos secundarios, de los problemas que plantea este tipo de estudios y de cómo proceder en estos casos.

Nos interesa plantear en este documento el análisis de datos secundarios, sus potencialidades y limitaciones, como una estrategia de investigación para producir conocimiento socialmente relevante que renueva su valor en el actual contexto de enorme proliferación de información. Frente a un crecimiento exponencial en la generación de bases de datos no sólo por académicos y por organismos de gobierno, sino también por empresas, ONGs y consultoras, cabe reflexionar cómo se posicionan los sociólogos ante estos 'competidores' en la producción de conocimiento. El desafío se acrecienta porque para estas fuentes alternativas la investigación que ellos realizan generalmente responde bien a sus necesidades: es productiva y 'efectiva' en sus propios términos.

En un artículo recientemente publicado en una revista científica anglosajona de sociología, dos investigadores académicos plantean esta cuestión en términos de una crisis en ciernes para la sociología empírica, aun poco reconocida y comprendida, producida por la enorme infraestructura que conforma lo que Thrift (2005, citado en Savage y Burrows, 2007) denomina *knowing capitalism*, donde proliferan los circuitos de información alimentados por numerosos tipos de tecnologías de la información. Esto ha traído consigo un cambio en el significado de la investigación empírica, que pone en cuestión el reclamo jurisdiccional que sobre ella y a partir de sus repertorios metodológicos, pueden hacer los sociólogos. Este llamado plantea una reflexión acerca de cuál es la mejor manera para que los sociólogos se relacionen con la proliferación de datos sociales recolectados por otros (Savage y Burrows, 2007). Una posible respuesta

es incorporar las tecnologías más recientes con el fin de aprovechar la posibilidad que éstas nos dan de contar con datos diferentes y más descriptivos (por ejemplo, imágenes) a ser desplegados en forma novedosa y potente. Junto con los métodos, también resulta crucial volver a plantear con Sautu (2003) la importancia esencial de la teoría como andamiaje, guía y sostén de todo el proceso de investigación empírica y, sobre todo, primordial en el análisis e interpretación de los datos.

La utilización de los datos disponibles para el estudio de lo social da la oportunidad de combinar distintas fuentes y así ganar una mayor amplitud (*scope*) y profundidad (*insight*) sobre los problemas que constituyen el objeto de la investigación. Si los datos disponibles son bases de datos, esto nos dará un mayor margen para manipular los datos que si disponemos únicamente de cuadros publicados por los autores del estudio original. Esta distinción atravesará los temas y ejemplos desarrollados en el presente artículo.

El análisis de datos secundarios puede estar referido a estudios con un abordaje cuantitativo o cualitativo y cada uno de ellos da lugar a consideraciones particulares; en el presente documento centramos la exposición y los ejemplos para el caso de la utilización de datos secundarios cuantitativos<sup>1</sup>. En su libro *Análisis Secundarios de Encuestas por Muestreo* de 1972, Herbert Hyman fue un precursor en la sistematización de principios y procedimientos de análisis secundario en la literatura sociológica y en la descripción detallada de su aplicación en diferentes tipos de estudios. Reconociendo la mencionada publicación como particularmente exhaustiva respecto de las particularidades que asume el análisis de datos secundarios cuantitativos, en el presente documento se describen resumidamente los principales beneficios y posibles obstáculos teniendo como telón de fondo y pensando en ejemplos de investigaciones en Argentina.

### **Delimitación del análisis de datos secundarios como estrategia metodológica**

En el contexto de los estudios cuantitativos los datos primarios son producidos en su totalidad por un mismo investigador o equipo, con un control sobre todas las etapas; por ejemplo, los estudios mediante encuestas en los cuales el mismo equipo de

---

<sup>1</sup> En breve un grupo de docentes de la cátedra elaborará un documento sobre el análisis de datos secundarios cualitativos que consiste en el análisis documental.

investigación decide el diseño muestral, la selección y operacionalización de los conceptos, la elaboración del instrumento de medición, la organización del trabajo de campo, la construcción de la matriz de datos y el procesamiento inicial. En contraste, el análisis de datos secundarios cuantitativos se refiere a la investigación con datos contruidos por otros -gobiernos, empresas de opinión pública, universidades, entre otros- con propósitos diferentes a los planteados en el estudio original (Hyman, 1972; González Rodríguez, 1996).

Los censos relevados por los gobiernos nacionales, tanto con propósitos administrativos como de informar al diseño de políticas públicas, constituyen el ejemplo más a mano de datos secundarios cuantitativos largamente utilizados por los sociólogos para investigar temas diversos tales como: distribución y redistribución del ingreso, patrones de migración, características del mercado de trabajo, estructuras ocupacionales, estructura de los hogares, cambios en la composición familiar, características de la población urbana y rural, entre otros (Frankfort-Nachmías y Nachmías, 1992: 291). En general, los datos producidos por censos, como por ejemplo los del último Censo Nacional de Población y Vivienda de 2001, se ponen a disposición de los usuarios en un formato ya tabulado por el organismo que lo realiza, en este caso el Instituto Nacional de Estadísticas y Censos. Dichos tabulados pueden estar o no publicados y en este último caso habrá que solicitarlos. En contraste, estudios por encuesta de gran tamaño como los que realiza el Centro de Estudios de Opinión Pública de la Universidad de Buenos Aires (CEDOP-UBA) brindan la posibilidad de contar con bases de datos enteras que permiten un mayor aprovechamiento de los datos en un análisis secundario.

El análisis de datos secundarios comparte con otros tipos de estudios cuantitativos, el predominio de una secuencia lógica deductiva en la confección del proyecto de investigación. Sin embargo, este proceso se vuelve más laxo cuando los datos han sido producidos en el contexto de otro estudio. Esto ocurre porque buena parte del marco teórico y de la propuesta de medición se debe ajustar o debe incorporar las definiciones implícitas en los datos originales. Usualmente la tarea consiste en que el investigador busca y selecciona entre los datos existentes, aquellos que le permitan abordar el tema de interés; partiendo de un problema de investigación más o menos planteado y acotado. Luego tendrá que ver si puede acceder a esa información y por último deberá estudiar en profundidad las características de los datos encontrados con el fin de establecer sus

objetivos específicos en función del 'margen de acción' que le imponen las definiciones operacionales de los conceptos tal como han sido recogidos originalmente.

### **Ventajas del análisis de datos secundarios o ¿Cuándo es la opción más adecuada?**

Los autores de libros de metodología que se ocupan particularmente de las características del análisis de datos secundarios (Frankfort-Nachmias y Nachmias 1992; González Rodríguez 1996; Hyman, 1972) desarrollan las ventajas que esta estrategia proporciona y explican cuáles son las razones para promover su utilización; las cuales se pueden agrupar en cuatro tipos principales: razones conceptuales-sustantivas, razones metodológicas, razones económicas y razones sociales.

Una de las razones **conceptuales-sustantivas** para el uso de datos secundarios es que se trata de la estrategia más adecuada posible para el conocimiento de ciertos temas-problemas de investigación. Junto con esto, esta estrategia abre la posibilidad de alcanzar una mejor comprensión del contexto histórico y de describir y explicar cambios mediante el análisis de datos de diferentes momentos. De este modo, los datos secundarios nos permiten no sólo estudiar y comprender el pasado sino también comprender el presente y hacer estimaciones sobre el futuro a través del estudio del cambio social y cultural. Un ejemplo de ello son los estudios del cambio en la estructura social Argentina a lo largo del tiempo que realiza el CEDOP-UBA. El análisis de datos secundarios hace posible, además, abordar estudios con propósitos comparativos entre países o áreas geográficas dentro de un mismo país, lo cual permite generalizaciones de mayor alcance. Por ejemplo, el Latinobarómetro que realiza una misma encuesta anual sobre el apoyo a la democracia en 18 países de América Latina ([www.latinobarometro.org](http://www.latinobarometro.org)). Por último, cuando el investigador está familiarizado con las fuentes de información disponibles en su área (datos secundarios), éstas pueden contribuir a la generación de nuevas ideas de investigación.

Desde el punto de vista **metodológico**, las razones para la utilización de datos secundarios se asocian a su vez con cuatro ventajas principales (que se asocian más con los análisis a partir de bases de datos que con los de datos ya tabulados): primero, brindan la oportunidad para su replicación y por ende cuantos más estudios encuentren

resultados similares más credibilidad ganan los propios hallazgos. Asimismo, la posibilidad de reanalizar los datos con perspectivas diferentes y con técnicas analíticas distintas, permite producir nuevas interpretaciones que enriquecen la disciplina en cuestión. Segundo, la disponibilidad de mediciones periódicas, como la Encuesta Permanente de Hogares (EPH-INDEC) o las estadísticas universitarias de la Secretaría de Políticas Universitarias del Ministerio de Educación de la Nación (SPU-ME), permiten plantear un diseño longitudinal. Asimismo, se pueden tomar valores de mediciones anteriores como base para la comparación para un estudio sobre datos más recientes. En este sentido se destaca la posibilidad de construir una variedad de diseños de investigación: de tendencias, de series temporales, de cohorte, comparativos y contextuales. Tercero, para la construcción de variables e indicadores a partir de datos secundarios se realiza un análisis más detallado y exhaustivo de las variables disponibles, lo cual permite generar nuevos *insights* de los datos. Cuarto, con datos secundarios se puede aumentar el tamaño de la muestra y su representatividad de manera tal que aumenta el alcance de las generalizaciones o incluso trabajar con todo el universo, como por ejemplo con datos censales u otras estadísticas oficiales. También los datos secundarios de tipo censal contribuyen con la construcción de marcos muestrales y posteriormente con el análisis de semejanzas y diferencias entre la muestra obtenida en el estudio primario y la información censal de modo que la primera pueda ser ajustada y/o validada.

Las razones **económicas** se vinculan con el ahorro de recursos, entre ellos, de tiempo, dinero y personal involucrado dependiendo de las particularidades del estudio. Para estimar el nivel de ahorro se deberán tener en cuenta factores tales como la longitud del cuestionario, el tiempo que lleva la entrevista, el número de preguntas abiertas y el tiempo de codificación correspondiente, la ubicación geográfica y las condiciones de acceso de la población bajo estudio, el tamaño y tipo de diseño muestral y de la prueba piloto (Hyman, 1972:6). Cuando se tiene la posibilidad de acceder a bases de datos, se puede, en poco tiempo, tener información abundante, de un modo barato y de calidad. Las bases de datos de organismos oficiales pueden servir de apoyo a una extensa gama de investigaciones. Con una mirada más amplia, la utilización del análisis de datos secundarios implica una maximización de los recursos públicos que, en muchos casos son la fuente del financiamiento de los censos y de las encuestas de gran envergadura. De esta manera, también se ahorra al evitar la duplicación de investigaciones costosas

que en definitiva buscan recoger información similar a la relevada por los organismos públicos y que está disponible para los investigadores que la soliciten.

Con relación a las razones **sociales**, Hyman (1972) señala que los estudios mediante encuestas por muestreo exigen un número grande de casos y por consiguiente de recursos económicos, y los financiamientos existentes están reservados para una ‘oligarquía’ de investigadores más experimentados que además trabajan en las instituciones más reconocidas. Por ende, la utilización de datos secundarios resulta la mejor opción y oportunidad para los investigadores menos experimentados, que trabajan en instituciones menos reconocidas, o que carecen de presupuesto para estudiar temas mediante encuestas. Otra razón social por la cual el análisis de datos secundarios puede resultar la mejor estrategia a seguir se presenta cuando se está en un momento y lugar conflictivos, ya que evita el agravamiento de tensiones que puede traer consigo el trabajo de campo en algunas zonas. En efecto, ocurre en ciertas ocasiones, que las encuestas son consideradas con hostilidad e incluso interpretadas como un ‘espionaje’ cuyos propósitos serían de manipulación de datos para fines que exceden a los del conocimiento científico. Por ejemplo, durante el año 1994, la realización del Censo Nacional de estudiantes universitarios realizado en Argentina despertó una gran polémica entre algunos estudiantes que, en un contexto político más amplio de privatización de empresas y servicios públicos, veían al censo como una amenaza que podía convertirse en una herramienta para justificar el arancelamiento-privatización de las universidades estatales. Esto trajo aparejado, además de situaciones de violencia entre estudiantes, información intencionalmente errónea y una alta tasa de rechazos de respuesta. En este caso se deberían haber considerado estas circunstancias y replanteado la estrategia del estudio que permitiera su realización de todas maneras, puesto que se trata de una fuente de datos secundarios valiosa para la utilización *a posteriori* por parte de investigadores y *policy makers*. En el caso contrario, cuando la atmósfera es menos tensa, menos sensibilizada, la gente suele estar dispuesta a cooperar con las encuestas y los censos, ya que los mismos están considerablemente aceptados en la sociedad como prácticas legítimas y necesarias para el desarrollo; no obstante, resulta decisivo evitar la imposición y abusar de la buena disposición de la gente.

## **Inconvenientes en el análisis de datos secundarios**

Si bien como se pudo describir hasta aquí, el análisis de datos secundarios presenta numerosas ventajas, el investigador que será usuario de los mismos debe estar atento a la posibilidad de encontrarse con inconvenientes y limitaciones de diversa índole. A continuación se presentan las principales dificultades sistematizadas a partir tanto de las propias experiencias de investigación con datos secundarios como de la literatura consultada acerca de su uso en el campo de las ciencias sociales, principalmente la sociología. Los inconvenientes pueden ser agrupados en: problemas técnicos, problemas en la calidad y en la adecuación de los datos, y posibles consecuencias negativas en la práctica de investigación por el uso de datos secundarios.

### ***Inconvenientes técnicos***

En primer lugar, los datos secundarios plantean diferentes inconvenientes técnicos, entre ellos, en la disponibilidad de los mismos: puede suceder que los datos que uno precisa no existan o no sean utilizables por otras razones (González Rodríguez, 1996). Por ejemplo, algunos países y agencias de gobierno restringen el acceso a sus datos en respuesta a la legislación de protección del anonimato de los datos. En otros casos, sólo se puede contar con una parte de los datos necesarios, ya sea de algunos años o regiones, de forma fragmentada e incompleta. Asimismo, es posible que existan los datos, pero a un nivel de agregación distinto del que necesitamos para nuestro estudio. Por razones de confidencialidad y anonimato en muchos casos es difícil obtener datos a nivel desagregado (Boslaugh, 2007; González Rodríguez, 1996). Este es el caso de los estudios que buscan medir la calidad de la educación a partir de las evaluaciones de aprendizajes donde los resultados se presentan únicamente a nivel de las provincias y no desagregados a nivel de las instituciones educativas ni de los alumnos.

Otro problema técnico vinculado con el anterior es el tiempo que nos puede llevar dar con los datos que necesitamos. En efecto, si bien se supone que obtener datos secundarios lleva menos tiempo que generarlos, esto no siempre sucede, debido, entre otras cosas, a cuestiones burocráticas, tales como pedidos formales de acceso y de utilización de bases de datos (González Rodríguez, 1996).



Una vez localizados los datos, podemos encontrarnos que no están organizados o clasificados de forma coherente, dificultando su uso. Asimismo, pueden presentarse en formatos que ya no son compatibles con los programas de computación actuales o incluso estar directamente en papel (Fielding y Fielding, 2000). En relación con esto, los instrumentos o los métodos de recolección de datos pueden haberse modificado con el tiempo con lo cual los datos se vuelven obsoletos (Saint-Germain, 2002). En este sentido, se puede citar la escala de autoritarismo de Adorno (1950) utilizada por Gino Germani (1973) en estudios realizados en Argentina en la década del '60 donde los indicadores que daban cuenta de este fenómeno eran los apropiados para ese contexto histórico, pero que en la actualidad debieron ser reemplazados para reflejar el grado de autoritarismo con las particularidades actuales. Una última dificultad técnica la constituye el costo de compra o de préstamo de datos: por ejemplo, cada año de relevamiento del Latinobarómetro cuesta actualmente cien euros.

### ***Problemas en la calidad de los datos***

Una vez obtenidos los datos, cabe preguntarse por la calidad de los mismos en términos de confiabilidad, validez y precisión. En ese sentido, es necesario evaluar tanto la confiabilidad de los datos como de la fuente, ya que unas fuentes son más creíbles que otras (González Rodríguez, 1996). Asimismo, como el investigador no participó del proceso de planificación y ejecución de la recolección de datos, no puede saber cómo fue realizado, ni si los datos tienen problemas como la baja tasa de respuesta o si existieron dificultades de comprensión de algunas preguntas específicas del cuestionario (Boslaugh, 2007). De esta manera, puede que los datos originales contengan errores no identificables por el usuario de datos secundarios. Como lo señala Boslaugh (2007), como analistas de datos secundarios desconocemos esos pequeños inconvenientes ocultos presentes en toda recolección de datos que si bien no los invalida, deberían ser tenidos en cuenta a la hora del análisis.

Asimismo, los datos pueden haber sido modificados por el investigador (por ej. ponderados o recategorizados) sin que esto se señale o sin poder conocer el criterio seguido para tal fin. En efecto, en numerosas oportunidades encontramos en distintas

bases de datos variables que son recodificaciones de otras. Lamentablemente como desconocemos los criterios utilizados para su reagrupamiento, la reutilización de esos datos se vuelve inconveniente. Por ejemplo, todas las decisiones que se toman para clasificar la inserción ocupacional de los encuestados no siempre están explicitadas, y por lo general, esta categorización surge quizás de la combinación de respuestas abiertas que describen las tareas que realiza el encuestado con otros indicadores ocupacionales que ignoramos. Por lo tanto, cuando producimos nuestros propios datos tenemos que dejar todas las modificaciones a los mismos por escrito, ya que otro investigador puede llegar a utilizar nuestros datos en el futuro.

### ***Falta de adecuación de los datos***

Dado que los datos no fueron recogidos para responder a la pregunta específica del investigador, estos pueden ser inadecuados en varios sentidos: en cuanto a la región geográfica como al período temporal o población específica de interés. Como lo señala Boslaugh, “uno puede trabajar con los datos que existen, no con los datos que a uno le hubieran gustado recolectar” (2007: 4). Estos problemas son muy frecuentes, como por ejemplo, cuando uno cuenta con datos de una encuesta que se realizó sólo un año, lo cual no le permite comparar ni estudiar la evolución del fenómeno. En algunos casos debido a la reformulación de los indicadores, tampoco se puede construir series históricas, por lo que se deben seleccionar los años en los cuales se preguntó de igual manera o con un sentido muy similar.

Otro problema de adecuación que es incluso más importante que los anteriores es el referido a la operacionalización de los conceptos. Este problema puede surgir desde la perspectiva teórica seleccionada, ya que la misma puede ser de un nivel de abstracción inferior al que uno pretende, lo cual va a incidir en el nivel de complejidad con el que se aborda el fenómeno en estudio y su consecuente operacionalización. En efecto, dado la coherencia que se debe tener entre los objetivos de investigación, el marco teórico y la medición del fenómeno, debemos tener en cuenta las limitaciones que implica que los datos secundarios responden a otros objetivos y teorías y que no fueron medidos como a nosotros nos hubiera gustado (Brigido, 2007). Así, las variables pueden haber sido definidas o categorizadas de diferente manera a la que uno hubiera elegido: por ejemplo,

en una base de datos la edad puede haber sido recogida en categorías y no como una variable continua (Boslaugh, 2007).

Como la forma en que las variables están medidas implica cierta definición de las mismas, cuando utilizamos datos secundarios, debemos asumir las mismas definiciones originales y reconstruir estas definiciones conceptuales de las variables a partir de los indicadores utilizados. Así, pueden existir referencias sub-culturales y expresiones idiomáticas que debemos utilizar aún sin comprender cabalmente el significado de las mismas. Lo anterior puede conducir a diferentes interpretaciones sobre el sentido de las preguntas del cuestionario, conduciendo a errores de interpretación (González Rodríguez, 1996). En efecto, cada vez que se traduce un cuestionario o que se utiliza el mismo cuestionario en distintas culturas como en el caso del Latinobarómetro o la Encuesta Internacional de Victimización (*The International Crime Victims Survey*: <http://www.unicri.it/wwd/analysis/icvs/data.php>) tenemos que tener en cuenta que pueden existir diferentes interpretaciones de las mismas preguntas.

Además, podemos encontrarnos con problemas de adecuación de los datos debido a la forma en que éstos se presentan, por ejemplo, en cuadros donde figuran promedios y porcentajes, que si bien “ofrecen la enorme ventaja de facilitar la comparación, no permiten discriminar, u ocultan, situaciones extremas al interior de cada jurisdicción y entre jurisdicciones, que pueden ser muy relevantes” (Brigido, 2007:4). Evidentemente, siempre es mejor contar con las bases de datos que con los tabulados, pero lamentablemente esto no siempre es posible por lo que debemos trabajar directamente con los cuadros que nos proporcionan los autores de los artículos.

Por último, muchas veces el tipo de datos disponible impone también restricciones en lo que respecta a las técnicas estadísticas que se pueden utilizar en la investigación (Brígido, 2007). En este sentido, si las variables fueron medidas a nivel nominal u ordinal, limitarán en gran medida el nivel de análisis estadístico. No obstante, en muchos casos se puede recodificar dichas variables asignándoles valores que permitan un mejor tratamiento estadístico al convertirlas en variables cuantitativas.

### ***Posibles derivaciones negativas para la práctica de investigación***

A continuación veremos algunas consecuencias indeseadas del uso de datos secundarios que conviene tener en cuenta a la hora de utilizarlos. Una de ellas es que el trabajo con datos secundarios obliga a tomar decisiones que no se hubieran adoptado en caso de disponer de información *ad hoc* para la investigación (Brígido, 2007). Como se podrá ver en los dos ejemplos que se presentan más adelante, el usuario de datos secundarios debe tomar numerosas decisiones para adaptar los datos disponibles a los propios objetivos. Cada decisión tomada debe ser mencionada para que el lector juzgue si fue una decisión acertada o no y así validar los resultados del estudio.

Otra posible consecuencia negativa del uso de datos secundarios es la propensión a caer en el llamado “empirismo burdo”. Acercarse a grandes bases de datos con la pretensión de que una masa de información va a generar por sí misma ideas, análisis e interpretaciones relevantes y de interés, constituye una ingenuidad. Los datos no hablan por sí mismos. En este sentido, la aproximación a los datos secundarios debe hacerse con una perspectiva teórica detrás, ya que esta es la única manera de encontrar la información relevante para nuestra investigación. No sirve pues la estrategia del “cuanto más mejor” o la utilizada en el análisis consistente en “cruzar todo contra todo, a ver que sale”. En una base cualquiera existen muchas variables interesantes, pero la base de datos no nos dice cuáles debemos seleccionar para nuestro estudio sino nuestra perspectiva teórica (Ander-Egg, 1992; González Rodríguez, 1996).

Otra consecuencia indeseada del uso de datos secundarios es la posible declinación en la creatividad de los investigadores. En efecto, si los investigadores se limitan a utilizar sistemáticamente los mismos datos, sin incorporar nuevos, se corre el riesgo de repetir siempre lo mismo, con lo que no tendrían cabida ni la creatividad, ni la intuición ni la innovación en la investigación en las ciencias sociales. El riesgo más importante radica en que problemas sociales emergentes tardarían mucho tiempo en incluirse en bases de datos. Es por eso que la misma actividad de generación de nuevas fuentes de datos, sobre todo por parte de los organismos públicos, debería ir introduciendo innovaciones en cuanto a la información que se genera y almacena (González Rodríguez, 1996). El Censo Nacional de Población, Hogares y Viviendas de la Argentina en el último censo

de 2001, respondió a esa necesidad incorporando preguntas sobre discapacidad, migraciones y pueblos indígenas.

Por último, otra consecuencia indeseada del uso de datos secundarios radica en la imposibilidad en muchos casos de disponer del contexto original en el que se produjeron los datos, limitando una fuente fundamental del análisis (Fielding y Fielding, 2000). Si bien parte del macro-contexto se puede reconstruir mediante la revisión de los medios periodísticos del momento, en muchos casos no contamos con las observaciones o comentarios de los investigadores primarios sobre el contexto particular en el que se recolectaron los datos, sus impresiones del lugar, del momento y de las personas presentes en el relevamiento. Lamentablemente la práctica de registrar notas de campo no es muy frecuente en los estudios cuantitativos a diferencia de los estudios cualitativos donde se tiene mayor conciencia sobre la importancia de dichas notas.

A pesar de los inconvenientes mencionados hasta aquí, son numerosas las ventajas que tiene el uso de datos secundarios. En efecto, sólo se los señala con el fin de que el investigador tome los recaudos suficientes, pero no para desalentarlo en el uso de los mismos. A continuación, veremos los mecanismos para proceder con cautela en el uso de estos datos.

### **Pautas para el trabajo de análisis de datos secundarios o ¿cómo proceder?**

Una parte importante del trabajo con datos secundarios consiste en conocer en la máxima medida posible el proceso de producción o construcción de los datos que utilizaremos. Si no disponemos de la base de datos sino que partimos de tabulados, será necesario además dilucidar la estrategia analítica utilizada, es decir el conjunto de criterios y operaciones que permitieron pasar de los datos a los resultados reportados.

La información con la cual es imprescindible contar para evaluar y decidir la utilización de los datos de interés es la siguiente:

- definiciones conceptuales que permitirán identificar la perspectiva teórica del relevamiento;
- descripción del diseño muestral: tipo, tamaño, posibles sesgos, errores muestrales, unidad de recolección, tasa de respuesta, estimadores y ponderadores

recomendados para realizar inferencias estadísticas y alcances de la representatividad muestral -nacional, provincial, municipal-;

- instrumento de recolección: cuestionarios, fichas, mapas;
- manual de códigos;
- informe acerca del trabajo de campo, manual del encuestador;
- descripción de los lugares y momentos en los cuales se desarrollaron las distintas etapas del estudio;
- descripción del contexto institucional de producción de los datos;
- informe sobre el modo de tratar a las no respuestas y a las dificultades en la comprensión de alguna pregunta u otros problemas que pudieran haberse producido durante la recolección de los datos.

Esta información puede estar publicada o sistematizada en documentos escritos y aquella que no lo estuviera deberá ser solicitada, por ejemplo, a través de consultas por correo electrónico, teléfono o visita a los investigadores o técnicos a cargo del estudio.

Asimismo, a la hora de buscar información secundaria, el investigador deberá realizar una evaluación de fuentes y datos. A continuación presentamos algunas consideraciones acerca del tipo de evaluación necesaria de los datos disponibles al comenzar el estudio:

- Formular nuestras preguntas de investigación primero y con ellas interrogar a las fuentes disponibles de cara a evitar el coleccionismo así como la tentación de cruzar todo contra todo a la espera de que surjan relaciones trascendentes o grandes teorías. A su vez, el investigador tendrá que ver si la información disponible cubre o no y en qué grado sus preguntas de investigación, caso contrario la combinación de diversas fuentes de datos será lo recomendable. Cuando se trata de estudios que conectan varios temas entre sí, como se verá en el ejemplo 2 del presente documento, se podrá recurrir a varias fuentes, reuniendo así la evidencia empírica necesaria.
- Juzgar la concordancia entre el marco teórico del estudio original y nuestros objetivos e hipótesis de investigación.
- Comprender lo más acabadamente posible la base de datos o los tabulados con los cuales se trabajará, por ejemplo, saber si las unidades de análisis (nivel de agregación) y el universo de estudio (recorte) al cual se refieren los datos nos permiten abordar las preguntas de investigación.
- Contar con variables relevantes y pertinentes teóricamente, clasificadas de modo tal que resulten categorías adecuadas a nuestro modelo de análisis. Por un lado,

respecto de si tienen un nivel de medición compatible con el tipo de análisis que quiero hacer, siempre constituye una ventaja si en el estudio primario las categorías están muy desagregadas porque se las podrá reagrupar del modo deseado, en tanto que resulta impracticable desagregar categorías que inicialmente fueron juntas. Por otro lado, se tendrá que juzgar si los sistemas clasificatorios utilizados para las variables de nuestro interés son apropiados en función de nuestras definiciones conceptuales.

- Comparar los objetivos del estudio que dio lugar a la recogida de datos con el nuestro respecto del recorte espacio-temporal, la amplitud de temas y el grado de profundización de los mismos, entre otros. En el campo de las actitudes y opiniones la ubicación temporal del trabajo de recogida de datos es un factor muy importante debido a que éstas son cambiantes; por lo tanto, se deberá considerar la pertinencia de su uso para responder a nuestros objetivos de investigación.
- Reconocer los propósitos con los cuales se llevó a cabo el estudio que pudieran introducir sesgos en los datos.
- Creer y confiar en el organismo que produjo los datos: considerar si se trata de instituciones confiables, con experiencia en la recolección de información ya que las instituciones de mayor trayectoria serán más confiables que otras de reciente creación.

A su vez, la validez y confiabilidad del trabajo con datos secundarios está muy vinculada con el hecho de que los datos hayan sido recogidos de manera seria, sólida y confiable; para ello, González Rodríguez (1996:266-7) presenta un conjunto exhaustivo de preguntas que vale la pena revisar para atender a estas cuestiones, pero que por falta de espacio no será reproducida aquí.

A continuación se presentan dos investigaciones basadas en el análisis de datos secundarios cuantitativos a modo de ejemplo.

## EJEMPLO 1

“Cultura política y punitividad pública en países de América del Sur” por María Alejandra Otamendi (Tesis de Maestría)

### ○ OBJETIVOS DE INVESTIGACION:

En primer lugar se buscó analizar la incidencia de factores culturales, en particular del grado de autoritarismo de la cultura política en la demanda social de mayores penas para los delincuentes (*punitividad pública*) de diez países de América del Sur en el año 2004. En segundo lugar, comparar dicha incidencia en la punitividad pública con la de factores objetivos macro-sociales, tales como las amenazas sociales y económicas. En efecto, se buscó demostrar que el autoritarismo predice el grado de punitividad en mayor medida que factores objetivos tanto sociales (el nivel de homicidios y victimización) como económicos (grado de desigualdad económica y desocupación).

### ○ FUENTE PRINCIPAL DE DATOS SECUNDARIOS UTILIZADA: ENCUESTA LATINOBAROMETRO

Para analizar la hipótesis principal planteada en el estudio, se decidió utilizar la base de datos de la encuesta del Latinobarómetro del año 2004.

#### *Características:*

Encuesta probabilística domiciliaria que se realiza anualmente desde el año 1995, relevando en la actualidad la opinión pública de 18 países de América Latina sobre temas tales como la democracia, la ciudadanía y la economía, entre otros. Para el estudio descrito arriba, se analizó la base de datos correspondiente al año 2004 hecha a 11.405 personas de diez países de América del Sur<sup>2</sup> (alredor de 1200 casos por país<sup>3</sup>) con un margen de error que va del 2,8% al 4,1 %.

#### *Ventajas de la fuente:*

1. permite realizar un estudio comparativo entre diez países porque a todos se les aplicó el mismo cuestionario en español (excepto para Brasil que fue en portugués);

---

<sup>2</sup> Argentina, Brasil, Chile, Uruguay, Bolivia, Colombia, Ecuador, Perú, Venezuela y Paraguay.

<sup>3</sup> En el caso de Paraguay fueron encuestadas 600 personas.



2. es la única encuesta que incluye preguntas sobre los temas principales que se plantea el objetivo, permitiendo estudiar la relación planteada entre cultura democrática y punitividad pública;
3. para la variable compleja “cultura democrática” el cuestionario cuenta con numerosos indicadores que combinados permite la construcción de un índice que refleja las diferentes dimensiones del fenómeno;
4. si bien tiene un costo de 100 euros por año relevado, fue posible obtenerla mediante un pedido a la Universidad, lo cual permitió contar con valiosos datos sin costo alguno y de una manera relativamente rápida;
5. el formato de la base de datos es compatible con el programa SPSS con lo cual su análisis es sencillo no siendo necesario transformar o exportar la base;
6. las consultoras de opinión que realizan el estudio en cada país cuentan con un alto grado de aprobación por la comunidad académica, lo cual cumple con los requisitos de confiabilidad y validez requeridos y
7. dada la gran cantidad de indicadores que presenta, se pueden generar numerosos estudios a partir de la misma base de datos, facilitando la emergencia de nuevos temas de investigación.

*Limitaciones de la fuente:*

1. no es seguro que los residentes de distintos países interpreten lo mismo en cada pregunta, incluso dentro de cada país (McAlister, 2006);
2. tiende a simplificar, cuantificar y generalizar puntos de vista complejos, diversos y contradictorios al formular preguntas generales y abstractas sobre el delito (Cullen et al., 2000; Hough y Roberts, 2005; Park y Hough, 2002; Roberts y Stalans, 1997). Por ejemplo, se indaga acerca del temor a ser víctima del delito directa y únicamente a partir de una pregunta (“¿Cuán frecuentemente se preocupa Ud. de que pueda llegar a ser víctima de un delito con violencia?”) sin especificar el tipo de delito ni las circunstancias en las que esto puede ocurrirle (Latinobarometro, 2007);
3. la forma de medir las actitudes tiene un alto impacto sobre las actitudes que se desean medir (Hutton, 2005). Por ejemplo, el nivel de temor al delito es mayor si se pregunta por el temor a caminar solo de noche por el barrio que si se pregunta por el temor a ser víctima del delito en abstracto;

4. toda encuesta de victimización se lleva a cabo sabiendo que los encuestados tienen escaso conocimiento de conceptos socio-legales y de la realidad del delito (Kury et al., 2002), lo cual incide en, por ejemplo, el nivel de demanda de castigo hacia los delincuentes, siendo éste mayor entre quienes poseen un menor conocimiento del funcionamiento del sistema penal (Hough y Roberts, 2005);
5. el cuestionario del Latinobarómetro que incluye preguntas que permiten estudiar la punitividad pública sólo se realizó durante el año 2004, lo cual impide llevar a cabo un estudio longitudinal del fenómeno;
6. la muestra sólo es representativa a nivel de los países, lo cual no permite evaluar las dinámicas subnacionales, tan importantes en países de gran heterogeneidad como los Sudamericanos;
7. para la medición de la punitividad pública se cuenta únicamente con un indicador, lo cual limita la validez del estudio. En efecto, en el cuestionario se indaga únicamente sobre si se considera a la “escasa dureza de las penas” como una de las principales causas de la delincuencia en el país hoy, en un listado que incluye otras posibles causas (Latinobarometro, 2004);
8. si bien se realizan varias preguntas para estudiar las causas del delito según los encuestados, hay posibles causas que no fueron tenidas en cuenta a la hora de construir el cuestionario, lo cual limita en cierta medida el análisis del fenómeno;
9. se pregunta por las causas del delito y no se pide la opinión sobre distinto tipo de castigos ni qué se debería hacer para mejorar la seguridad en los barrios, con lo cual no responde exactamente a los objetivos de investigación formulados más arriba y
10. la forma como está medida la punitividad no permite la creación de una escala actitudinal hacia el castigo. Sin embargo, al tomar los datos agregados por país, esta dificultad se ve superada al calcularse el porcentaje de respuestas afirmativas por país. De esta manera, el rango de la variable va de 0 a 100%.

#### *Decisiones tomadas:*

En primer lugar, la decisión más importante y que siempre debe hacerse con los datos secundarios es si finalmente, a pesar de todas las limitaciones que presentan, los utilizamos o no para responder a nuestro objetivo de investigación. En este caso, decidimos utilizar los datos del Latinobarómetro porque era la fuente de datos más

completa, ya que incluía los temas que queríamos estudiar, permitía la comparación entre países y la fuente era lo suficientemente confiable para ser tomada en cuenta.

Una fuente alternativa como el *International Crime Victims Survey* (ICVS) sólo contenía preguntas que indagaban las opiniones sobre el delito, pero no sobre la democracia, abarcaba algunas ciudades de América del Sur (no eran nacionales) y tampoco permitía un estudio longitudinal porque las distintas olas se habían hecho en distintas ciudades con muy pocos países repetidos. De esta manera, el Latinobarómetro era la mejor fuente disponible para nuestro estudio.

Luego, una vez que obtuvimos la base de datos, lo cual llevó dos meses de insistencia y la firma de un convenio, empezó el trabajo con las variables del estudio. Para la variable dependiente, el nivel de punitividad pública, si bien contábamos con un sólo indicador, lo cual limitaba las dimensiones de análisis del fenómeno a medir, el mismo había sido utilizado en otros estudios en el área para medir el mismo fenómeno. Así, para obtener el dato del nivel de punitividad pública a nivel de los países, se consideró al porcentaje de encuestados de cada país que sostenía que una de las causas del delito era la liviandad de las penas.

Por su parte, para medir la variable independiente principal (grado de cultura democrática) fue necesario construir un índice a partir de la suma de los valores de varios indicadores (porcentaje que en ninguna circunstancia apoyaría un gobierno militar, porcentaje que cree que un gobierno militar puede solucionar menos cosas que un gobierno democrático, entre otros) y luego evaluar la consistencia del mismo. En ambos casos, las variables creadas tenían un rango de 0 a 100 lo cual permitía ubicar a los distintos países en dichas variables de tipo cuantitativas y así facilitar el análisis estadístico.

Luego de un primer análisis bivariado, se decidió reagrupar a los países en dos sub-regiones, la Región Andina y el Cono Sur. De esta manera, los resultados del análisis multivariado, donde la sub-región pasó a ser incluida como una variable Dummy, permitió obtener un primer modelo multivariado con buen ajuste para explicar el nivel de punitividad pública en América del Sur.

○ OTRAS FUENTES DE DATOS SECUNDARIAS UTILIZADAS: UNCS, OMS, PNUD

Para testear las hipótesis alternativas, fue necesario utilizar otras fuentes de datos secundarios, entre ellas:

- United Nations Survey on Crime Trends and the Operations of Criminal Justice Systems (UNCS) que compila los datos de los registros de policía y justicia nacionales;
- la base de datos de Causas Externas de Mortalidad de la Organización Mundial para la Salud (OMS) que compila los registros de los sistemas estadísticos nacionales de salud y
- el Compendium estadístico del libro del PNUD de 2005: *Democracy in Latin America: towards a citizens' democracy. Statistical Compendium*, Buenos Aires, Programa de las Naciones Unidas para el Desarrollo, de donde se tomaron las variables económicas.

*Decisiones tomadas:*

El uso de otras fuentes para testear hipótesis alternativas, introdujo nuevos problemas, ya que no siempre se encontraban datos para los distintos países para el mismo año. Por lo tanto, se decidió incluir los datos más próximos al año 2004 que era el año de referencia de la encuesta Latinobarómetro. En otros casos, se privilegió el año en el cual se contaba con datos de la mayor cantidad de países para de alguna manera permitir la comparación entre ellos.

Asimismo, no siempre se contaba con datos para todos los países relevados, con lo cual para algunas variables fue necesario combinar dos fuentes diferentes, con todo el problema que eso conlleva. En efecto, para la tasa de homicidios, el UNCS incluía datos de todos los países, menos de Brasil. En ese caso, se decidió, luego de evaluar la consistencia de ambas fuentes, incluir el dato sobre homicidios a partir de la base de la OMS para dicho país.

Lamentablemente en algunos casos directamente no fue posible encontrar los datos correspondientes para todos los países en estudio, con lo cual el análisis de algunas variables se limitó a los países que sí contaban con datos. Todas estas decisiones y el uso de distintas fuentes, con diferentes años de relevamiento y no siempre con datos de todos los países, condujo a importantes limitaciones en el estudio, lo cual podrían llegar

a explicar la poca relevancia de los factores alternativos para comprender el fenómeno de la punitividad pública.

No obstante, considerando los importantes resultados obtenidos en la relación original planteada entre cultura política y punitividad pública y los resultados de estudios previos, además de la coherencia teórica alcanzada, permitieron concluir que la hipótesis principal había sido corroborada. En otras palabras, que el grado de autoritarismo de una cultura incide en el nivel de punitividad pública en países de América del Sur, siendo los países más autoritarios, los más punitivos, en especial, luego de dividirlos en las subregiones del Cono Sur y Región Andina. Además, se pudo señalar que la punitividad pública se encontraba mayormente asociada a este factor político-cultural que a factores macro-sociales más objetivos.

## EJEMPLO 2

“Educación universitaria y estructura social. Tendencias de la expansión universitaria y sus efectos en la inserción ocupacional y pertenencia de clase de los graduados. Un estudio con datos secundarios de Argentina y del Área Metropolitana de Buenos Aires entre 1989 y 2004” por María Pía Otero (tesis de Maestría).

### o OBJETIVOS DEL ESTUDIO:

El objetivo general consta de dos partes, la primera, describir las tendencias de la expansión de la educación universitaria en Argentina y, en particular, los grupos sociales que lograron acceder a la universidad y la transmisión intergeneracional de la educación de padres a hijos en el Área Metropolitana de Buenos Aires entre 1989 y 2003. La segunda, examinar la graduación universitaria como canal para la inserción ocupacional y el posicionamiento de clase en el AMBA en 2003/4.<sup>4</sup>

Las preguntas que orientan el estudio (objetivos específicos) son las siguientes:

- La educación universitaria ¿está restringida a ciertas categorías sociales o, a partir de la expansión, se han ido incorporando nuevos grupos antes excluidos?<sup>5</sup>
- ¿En qué medida la educación de los hijos es hereditaria de los logros educacionales de los padres?
- En un contexto de reestructuración productiva que implicó una apertura de la economía y de cambios en el mercado laboral que produjo un deterioro de las condiciones de trabajo y una disminución de la demanda de empleo ¿cuál ha sido la tendencia en la participación económica y la ocupación de los graduados universitarios?
- ¿En qué calidad de ocupaciones se insertan los graduados en 2003/4 y cuál es su nivel de ingresos? ¿Varía esta inserción entre mujeres y varones y entre distintos grupos de edad?

---

<sup>4</sup> La justificación para el recorte de este universo de estudio está dada, por un lado, porque el AMBA es el aglomerado urbano con mayor concentración de graduados universitarios y, por otro lado, porque permite abordar el estudio de los logros ocupacionales bajo condiciones económicas y del mercado laboral más homogéneas que si se opta por abordar un análisis a escala nacional. El recorte temporal se debe a un interés por estudiar el momento actual, específicamente 2003/4 por tratarse de los últimos datos disponibles para abordar este objetivo.

<sup>5</sup> Vale señalar que se considera el acceso en un sentido amplio que incluye la graduación del nivel y la inserción en el mercado de trabajo.

- ¿El nivel de educación más alto del sistema educativo asegura a quienes obtienen estas credenciales las ocupaciones profesionales y directivas y gerenciales propias de las clases medias y medio-altas?
- ¿Cuál es el posicionamiento de clase de los graduados universitarios? Si se ampliaron las bases de los que acceden a la graduación universitaria, estas nuevas incorporaciones, ¿tuvieron condiciones sociales favorables para que sus logros educacionales se puedan ver reflejados en sus posiciones de clase?

○ FUENTES DE DATOS SECUNDARIOS UTILIZADAS:

1) Anuarios Estadísticos publicados por la Secretaría de Políticas Universitarias del Ministerio de Educación de la Nación, organismo que releva anualmente los datos sobre estudiantes y docentes de las universidades estatales y privadas de Argentina (<http://www.me.gov.ar/spu/Publicaciones/publicaciones.html>). Las siguientes son las fuentes específicas utilizadas:

- MINISTERIO DE EDUCACIÓN, CIENCIA Y TECNOLOGÍA - Secretaría de Políticas Universitarias, (2006); *Anuario de Estadísticas Universitarias 2006*. Vol.1; Buenos Aires, Argentina. [http://www.mcyt.gov.ar/spu/guia\\_tematica/estadisticas\\_y\\_publicaciones/estadisticas\\_y\\_publicaciones.html](http://www.mcyt.gov.ar/spu/guia_tematica/estadisticas_y_publicaciones/estadisticas_y_publicaciones.html)
- MINISTERIO DE EDUCACIÓN, CIENCIA Y TECNOLOGÍA Secretaría de Políticas Universitarias, (2004); *Anuario 1999/2003 de Estadísticas Universitarias*. Buenos Aires, Argentina.
- MINISTERIO DE EDUCACIÓN, CIENCIA Y TECNOLOGÍA - Secretaría de Políticas Universitarias, (2002); *Anuario 1999-2000 de Estadísticas Universitarias. Universidades Nacionales y Privadas*. Buenos Aires, Argentina.

Con las publicaciones estadísticas de la Secretaría de Políticas Universitarias del Ministerio de Educación de la Nación, se respondió al objetivo específico de describir la expansión del subsistema de educación universitaria argentino. Este análisis incluyó el universo total de universidades argentinas, desde mediados de la década del '90 hasta mediados de 2000, recorte temporal restringido a dicho período porque se trata de la información disponible para los totales de estudiantes, nuevos inscriptos y graduados. De este modo, mediante la información reportada por las casas de altos estudios fue posible caracterizar las tendencias en el acceso y graduación de la educación

universitaria a la luz de los principales factores sociodemográficos como el sexo y la edad.

2) Bases de datos de cuatro ondas de mayo para los años 1989, 1994, 1999 y 2003 de la Encuesta Permanente de Hogares (EPH) realizada por el Instituto Nacional de Estadísticas y Censos (INDEC). Si bien la Encuesta Permanente de Hogares (EPH) se implementa con un diseño muestral de tipo panel (en donde los mismos individuos son re-entrevistados en distintos períodos) aquí se la utilizó como cortes transversales. La EPH es un relevamiento periódico que lleva adelante el Instituto Nacional de Estadísticas y Censos con el propósito de contar con información sistemática para la producción de estadísticas de empleo y desempleo. Asimismo, se miden otros indicadores relacionados con el comportamiento del mercado laboral tales como los niveles educacionales y de ingresos de la población. Vale señalar que aunque la EPH no ha sido diseñada para indagar especialmente sobre logros educacionales de la población, contiene preguntas que resultaban relevantes para abordar los objetivos de investigación de la tesis. Asimismo, la posibilidad de contar con las bases y con documentos metodológicos que describen las definiciones que la sustentan así como también detalles sobre el diseño muestral fueron cruciales al momento de decidir utilizarla como fuente para este estudio.

Una vez presentada la descripción de la ampliación de la cobertura del sistema universitario argentino en los últimos 15 años con los datos provenientes de las instituciones universitarias, con la EPH-INDEC se abordó el objetivo de analizar las tendencias educacionales de la población del AMBA<sup>6</sup>. Estos datos también se utilizan para describir las tendencias de la participación económica de los graduados universitarios en el AMBA.

3) Base de datos de encuesta aleatoria nacional (2003 y 2004) sobre temas de estratificación social realizada por el Centro de Estudios de Opinión Pública (CEDOP-UBA)<sup>7</sup> con un diseño muestral probabilístico y de alcance nacional aunque para este estudio se recortó al Área Metropolitana de Buenos Aires.

---

<sup>6</sup> Comprende la Ciudad Autónoma de Buenos Aires y los 24 Partidos del Gran Buenos Aires, que conforman un aglomerado urbano definido como tal por el INDEC.

<sup>7</sup> El Centro de Estudios de Opinión Pública (CEDOP) está dirigido por el Prof. Raúl Jorrot y pertenece al Área de Estratificación Social del Instituto Gino Germani de la Facultad de Ciencias Sociales de la UBA.



Con la encuesta del CEDOP se analizó la transmisión intergeneracional de la educación, con atención particular sobre el nivel universitario y la calidad de las ocupaciones y el posicionamiento de los universitarios en la estructura de clase.

Es importante señalar aquí las razones por las cuales se recurrió a una fuente de datos alternativa a la EPH-INDEC. Además de proporcionar información detallada acerca de las ocupaciones de los sujetos, la encuesta del CEDOP habilitó: i) la construcción de un esquema de clases sociales, puesto que se llevó a cabo en el marco de un estudio sobre estratificación social; ii) el estudio de la influencia de los padres sobre la educación de los hijos, puesto que relevó información educacional y ocupacional de los padres de los sujetos entrevistados; y iii) la comparabilidad de los datos con la EPH ya que son dos muestras probabilísticas del mismo universo (la población adulta del AMBA).

○ VENTAJAS DEL ANÁLISIS DE DATOS SECUNDARIOS COMBINANDO FUENTES:

- permitió el abordaje de varios temas conectados entre sí a partir de la combinación de distintas fuentes, tanto de datos publicados como de bases de datos
- la utilización de bases de datos permitió la construcción de nuevas variables específicas de interés para el estudio planteado
- la disponibilidad de mediciones periódicas, como la EPH-INDEC o las estadísticas universitarias de la SPU-ME hizo posible plantear un diseño longitudinal
- aumentar el tamaño de muestra y su representatividad de manera tal que se mejoró el alcance de las generalizaciones;
- el carácter gratuito de la información nos evitó el alto costo que implica un estudio por encuesta representativo de la población de un país o de un área geográfica como el AMBA

○ INCONVENIENTES o LIMITACIONES:

La EPH cambió y pasó a ser implementada con una modalidad continua<sup>8</sup> desde el año 2004 con lo cual dejó de ser comparable con los relevamientos anteriores, o, al menos

---

<sup>8</sup> Para mayor información, ver:

[http://www.indec.mecon.ar/nuevaweb/cuadros/4/Metodologia\\_EPHContinua.pdf](http://www.indec.mecon.ar/nuevaweb/cuadros/4/Metodologia_EPHContinua.pdf)

hasta el momento los responsables de la encuesta no habían dado a conocer con documentos metodológicos el procedimiento para poder hacer comparaciones entre ambas modalidades de medición: puntual y continua. Se decidió considerar al año 2003 como momento más reciente para el análisis longitudinal con la EPH, dada la imposibilidad de comparar los resultados entre las encuestas puntuales y las continuas.

Cuando se trabaja con datos que no fueron generados especialmente para el propósito que uno tiene en su investigación, la recomendación fundamental de los investigadores formados y con experiencia es formular y responder una enorme variedad de preguntas metodológicas acerca de cómo se encuestó, qué se preguntó, por qué se preguntó, entre otras cuestiones que hacen al diseño muestral y al instrumento de recolección y su aplicación (Sosa Escudero 2005). Siguiendo estas recomendaciones se procuró leer toda la información metodológica complementaria de los datos, resultando una tarea ardua y dificultosa ya que algunos de los documentos no son publicados ni están a disposición de los usuarios en Internet. Contar con estos materiales es indispensable si se quiere garantizar la validez y confiabilidad del estudio.

Hasta aquí se describieron los datos utilizados para la mencionada tesis de Otero. A continuación se presenta una lista más amplia de fuentes de datos secundarios disponibles en Internet sobre temas de educación sólo a modo de ejemplo sin pretensión de exhaustividad.

#### ○ OTROS EJEMPLOS DE FUENTES DE DATOS SOBRE TEMAS DE EDUCACIÓN

##### **Argentina**

##### **Organismos gubernamentales**

- Instituto Nacional de Estadísticas y Censos <http://www.indec.mecon.gov.ar/>
- Ministerio de Educación Nacional - Dirección Nacional de Información y Evaluación de la Calidad Educativa (DiNIECE) <http://dineece.me.gov.ar/>
- Ministerio de Educación de la Ciudad de Buenos Aires – Dirección de Investigación y Estadística  
[http://www.buenosaires.edu.ar/areas/educacion/dirinv/?menu\\_id=11885](http://www.buenosaires.edu.ar/areas/educacion/dirinv/?menu_id=11885)

- Ministerio de Educación de la Provincia de Buenos Aires - Dirección de Información y Estadística  
<http://abc.gov.ar/lainstitucion/organismos/informacionyestadistica/default.cfm>
- Gobierno de la Provincia de Tucumán – Ministerio de Educación – Departamento de Planeamiento y Estadística <http://www.educaciontuc.gov.ar/sistema/>
- Gobierno de Catamarca – Ministerio de Educación , Ciencia y Tecnología – Dirección de Planeamiento <http://www.meccyt.catamarca.gov.ar/>
- Gobierno de La Pampa – Ministerio de Educación <http://www.lapampa.edu.ar/>
- Gobierno de Córdoba – Ministerio de Educación - Dirección de Planeamiento e Información Educativa <http://www.cba.gov.ar/vercanal.jsp?idCanal=49179>

### **Otras organizaciones**

- CTERA – Instituto de Investigaciones Pedagógicas “Marina Vilte” - Área Informaciones y Estadísticas  
[http://www.ctera.org.ar/iipmv/areas/area\\_estadisticas.htm](http://www.ctera.org.ar/iipmv/areas/area_estadisticas.htm)
- SUTEBA – Investigación Estadística <http://www.suteba.org.ar/index.php?sec=64>
- CIPPEC – [www.cippec.org.ar](http://www.cippec.org.ar)

### **Internacionales**

#### **Organismos internacionales**

- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) [http://portal.unesco.org/education/es/ev.php-URL\\_ID=48712&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/education/es/ev.php-URL_ID=48712&URL_DO=DO_TOPIC&URL_SECTION=201.html)
- UNESCO – Institute for Statistics  
[http://www.uis.unesco.org/ev\\_en.php?URL\\_ID=3753&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201](http://www.uis.unesco.org/ev_en.php?URL_ID=3753&URL_DO=DO_TOPIC&URL_SECTION=201)
- Sistema de Información de Tendencias Educativas en América Latina SITEAL  
<http://www.siteal.iipe-oei.org/>

#### **Organismos gubernamentales de otros países**

- Gobierno de Brasil - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) <http://www.inep.gov.br/>
- Gobierno de Estados Unidos – U.S. Department of Education - National Center for Education Statistics - Institute of Education Sciences <http://www.nces.ed.gov/>

## BIBLIOGRAFÍA CITADA

- Ander-Egg, E. (1992) *Técnicas de investigación social*. Buenos Aires: Humanitas.
- Boslaugh, S. (2007) *Secondary Data Sources for Public Health: A Practical Guide*, Excerpt. Cambridge, Cambridge University Press
- Brígido, A. M. (2007): *El uso de datos cuantitativos secundarios en la investigación educativa*, ponencia presentada en Jornadas PRE – ALAS “Sociología y Ciencias Sociales: conflictos y desafíos transdisciplinarios en América Latina y el Caribe”; Buenos Aires 23 y 24 de mayo.
- Cullen, F., B. Fisher & B. Applegate (2000) Public Opinion about Punishment and Corrections. *Crime and Justice*, 27, 1-79.
- Frankfort-Nachmias, C. y D. Nachmias (1992): *Research Methods in the Social Sciences*. New York: St. Martin Press. Cap. 13 “Secondary Data Analysis” pp. 291-317.
- González Rodríguez, B. (1996): “La utilización de los datos disponibles”, en M. García Ferrando y otros (Compiladores) (1996), *El análisis de la realidad social. Métodos y técnicas de investigación*, Madrid, Alianza Universidad, pp. 245-273.
- Hough, J.M. & J. Roberts (2005) *Understanding public attitudes to criminal justice*, Maidenhead, Open University Press.
- Hutton, N. (2005) Beyond populist punitiveness? *Punishment & Society*, 7, 243-58.
- Kury, H., J. Oberfell-Fuchs & M. Würger (2002) Methodological Problems in Victim Surveys: The Example of the ICVS. *International Journal of Comparative Criminology*, 2, 38-56.
- Mcalister, A. (2006) Acceptance of killing and homicide rates in nineteen nations. *European Journal of Public Health*, 16, 259-65
- Park, A. & J.M. Hough (2002) *Public attitudes towards crime and punishment*, London, National Centre for Social Research.
- Roberts, J.V. & L.J. Stalans (1997) *Public opinion, crime, and criminal justice*, Boulder, Westview.