



Universidad de Buenos Aires
Facultad de Ciencias Sociales

DOCUMENTO DE CÁTEDRA 48. La matriz de datos: la construcción de variables individuales y agregadas.

Griselli, Lucía (1998); adaptado por Carolina Zuccotti (2008)

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 objetivo de este artículo es analizar una etapa del proceso de investigación cuantitativa por encuesta: la construcción de la matriz de datos. Para ilustrar el tema se tomará el estudio “Familia y rendimiento escolar: primario en un barrio de clase trabajadora” (1995.Sautu,Vujosevich,Griselli), cuyo objetivo principal fue examinar la influencia de las condiciones socio-económicas de la familia en el rendimiento escolar del alumno.

Variables del estudio, muestra y cuestionario

Los factores que afectan el rendimiento escolar de los menores pueden ser las condiciones de la familia y también la escuela. Los estudios de la CEPAL sólo toman las variables relacionadas a la familia, como ser las condiciones socio-económicas y socio culturales, pero no toman en cuenta datos sobre la escuela. Nuestro análisis se sitúa en esta línea. La variable dependiente es entonces “rendimiento escolar” y las variables independientes, aquellas relacionadas con la familia: 1)“condiciones socio-económicas”, 2)“condiciones socio-culturales” y 3)“condiciones habitacionales”¹.

A fines de explicar el proceso de construcción de la matriz de datos, nos centramos sólo en las variables independientes. Éstas se midieron a través de distintos indicadores:

1) Indicadores de las condiciones socio-económicas de la familia

- Ocupación del jefe: es un buen indicador de sus oportunidades laborales.
- Estabilidad laboral: introduce inseguridad en la fuente de ingresos
- Ingresos de la familia y tasa de dependencia laboral o económica: mide cuántos miembros del hogar son sostenidos por cada uno de los económicamente activos, es una medida del bienestar y del esfuerzo que es necesario hacer para alcanzar un ingreso. Cuánto más alta la tasa de dependencia más frágil y vulnerable es la situación económica del hogar, porque muchos miembros dependen de unos pocos que trabajan.

2) Indicadores de las características socio culturales de la familia

Estas variables definen las características del potencial estímulo cognitivo del hogar:

¹ Además se utilizaron variables básicas referidas al encuestado como sexo, edad, ocupación, etc.

- Status marital del hogar: da cuenta de la mayor o menor estabilidad de las relaciones familiares.
- Educación del jefe del hogar o adulto principal.
- Educación de la madre o mujer adulta principal.
- Porcentaje de adultos que terminaron la escuela primaria.

3) Indicadores de las condiciones habitacionales o de vivienda de la familia

Las características de la vivienda son indicadores de la capacidad de acumulación de la familia. La vivienda muestra en qué condiciones transcurre la vida del menor, si existe espacio en el cual desarrollar sus actividades escolares o si, por el contrario, el nivel de hacinamiento impide tener el propio espacio.

- Hacinamiento de personas por cuarto
- Origen del agua
- Ubicación del baño

Para llevar a cabo el estudio se realizó una encuesta a una muestra de personas que cumplieran determinados requisitos. Las variables mencionadas fueron incorporadas al cuestionario en preguntas que se formularon a cada encuestado.

La muestra fue no aleatoria de 125 familias residentes en la zona denominada Maciel - partido de Avellaneda - que comprende dos zonas definidas por su nivel de urbanización: la denominada Isla Maciel y el asentamiento poblacional denominado Villa Maciel.

La unidad de análisis fue cada uno de los menores de 6 a 19 años miembros de las familias, que pueden encontrarse o no asistiendo a un establecimiento escolar. Fueron excluidos menores de 17 a 19 años que eran compañeras/os del hijo/a del jefe, y se incluyeron como menores dependientes a los sobrinos o nietos (junto con los hijos del jefe o cónyuge). El criterio aplicado fue el de menores que se asumía habían sido criados en ese hogar.

Para la obtención de información de todo el núcleo familiar se utilizó la siguiente grilla de preguntas o cuestionario, para luego armar las variables correspondientes al estudio.

DATOS DEL GRUPO FAMILIAR

1 • Cuántas personas viven habitualmente en esta casa?

Por favor dígame nombre, parentesco, sexo y edad de cada uno.

Registrar en la planilla siguiendo el orden: **Jefe, Cónyuge, hijos en edades decrecientes y Otros miembros.**

	1	2	3	4
Nombre				
2 • Relación con el Jefe	JEFE			
3 • Sexo: 1-Varón 2-Mujer				
4 • Edad				
5 • Nivel Educativo Alcanzado 1-Nunca asistió 2-Primario. 3-Secundaria. 4-Terciario/Universitario 5-Otros Especificar				
6 • Ultimo Grado/Año aprobado o ciclo/carrera completo/a				
7 • ¿Continúa estudiando? 1-SI 2-NO				
8 • ¿Dónde nació? 1-Cap. Fed. o G.B.A. Loc. 2-Interior país Pcia. 3-Extranjero País.
9 • ¿Cuántos años hace que vive en Capital o Gran Bs. Aires?				
10 • Est. Civil (14 años y +) 1-Soltero 4-Sep/Div. 2-Casado 5-Viudo 3-Unido				
11 • Trabajo (14 y+ años) 1-Trabaja ahora				

2-Desocupado 3-Jubilado/Pensionado 4-Ama de casa 5-Estudiante 6-Otros. Especificar _____				
12 • ¿En qué trabaja ó trabajaba? ¿Qué tareas hace o hacía en ese trabajo? _____ _____				
13 • ¿Dónde trabaja o trabajaba ? ¿Qué hace la empresa, qué produce, qué vende, qué servicio presta? _____				
14 • ¿Es/era estable el trabajo ? 1-SI 2-NO				
15 • Posición ocupacional: 1-Por su cuenta sin pers./con local 2-Por su cuenta sin pers./sin local 3-Patrón Cant. Empleados, obreros 4-Jefe (Cant. Empleados a cargo) 5-Capataz (Cant. Obreros a cargo) 6-Empleado calificado 7-Empleado sin calificar 8-Obrero calificado 9-Obrero sin calificar 10-Otro. Especificar				

La construcción de datos y la matriz de datos

Luego de finalizada la etapa del trabajo de campo (reunión del material) los cuestionarios se prepararan para la etapa del procesamiento de la información. Los datos deben organizarse y sistematizarse para ser contados y poder luego dirigir la atención al análisis e interpretación. Analizar es resumir las observaciones de forma tal que proporcionar respuestas a los interrogantes y la interpretación busca significados más amplios con otros conocimientos posibles. La matriz de datos forma parte de la etapa de organización.

En la construcción de los datos intervienen tres elementos:

- 1) las *unidades de análisis*, tales como personas, naciones, provincias, etc. Es alrededor de ella que se elabora la matriz de datos. Estas unidades constituyen el objeto de la investigación.
- 2) las *variables* que uno desea conocer de las unidades de análisis.
- 3) los *valores* (o categorías) que alcanzan las unidades de análisis en las variables estudiadas, o las respuestas o los resultados que se obtienen cuando las unidades se exponen a los estímulos.

De esta forma tenemos las tres partes que constituyen la estructura de los datos sociológicos, lo que se denomina “la estructura tripartita de los datos”:

Unidades de análisis --- Variables estudiadas --- Valores o categorías de las variables

La matriz de datos es una forma de ordenar la información de manera de organizarla en forma tal que sea particularmente visible los tres elementos que constituyen el dato: las unidades de análisis, las variables y las categorías.

Unidad de análisis	Variable 1	Variable 2	Variable 3	Variable 4	n variables
unidad 1	valor	valor	valor	valor	
unidad 2	valor	valor	valor	valor	
unidad 3	valor	valor	valor	valor	
...					
n unidades	n valores				

* valor o categoría de la variable

Ésta es la forma en que se ingresan los datos en el programa estadístico para ciencias sociales SPSS (Statistical Package for the Social Sciences), uno de los programas más completos y de mayor difusión, que sirve luego para el procesamiento de los datos cuantitativos. En la matriz se pone el código del valor de la variable que corresponde a cada unidad de análisis. Codificar, es asignar un número a cada categoría de respuesta: los datos originales son transformados en números de modo tal que luego puedan ser contados. En las preguntas cerradas, aquellas donde las posibilidades de respuestas están numeradas (codificadas) de antemano, el número de código se coloca al costado de cada categoría. El código está inserto en el propio cuestionario. En las preguntas abiertas se elabora el código después de aplicado el cuestionario. Las respuestas a las preguntas se codifican después de realizadas las encuestas, dado que no es posible prever qué es lo que contestarán los entrevistados, o no es posible preverlo en un grado tal que permita confeccionar de antemano las categorías. Esta elaboración se realiza en varias etapas: primero la lectura de una muestra de las respuestas; luego se confecciona los códigos que se aplican a una muestra de entrevistas lo que permite rectificar las categorías y establecer el código definitivo.

La idea es que debe haber un valor o categoría para cada unidad de análisis en cada variable. Cuando se cuenta los números se sabe cuál es el número de casos comprendidos en una categoría determinada. En la matriz de datos no se debe dejar ninguna celda vacía, pero esto no se alcanza tan fácilmente en la práctica, por lo tanto se sigue el principio de que la cantidad de celdas que no contengan ninguna información o ninguna respuesta debe ser mantenida lo más baja posible. Aquí debemos diferenciar entre la ausencia de datos (porque el entrevistado no quiso contestar) y una mala aplicación del cuestionario por parte del encuestador (realizó una pregunta que no correspondía).

Códigos de variables (ejemplos):

Lugar de residencia:

- 1- Capital Federal - GBA
- 2- Interior del país.

Estabilidad laboral:

- 1- Trabajo estable
- 2- Trabajo no estable.

Rendimiento escolar:

- 1- desertor
- 2- asiste atrasado

- 3- asiste regular
- 4- completo

Nro.	Sexo	Edad	Lugar de residencia	Años de estudio	Concurrencia escolar	Rendimiento escolar ♦	Cantidad de miembros en el hogar	Cantidad de miembros en el hogar que trabajan	Tasa de dependencia laboral ♦	Estabilidad laboral	Tasa de hacinamiento o ♦	Nivel educativo de la madre
1	1	18	2	7	2	4	2	2	1	1	1	1
2	2	10	2	5	1	3	7	3	2.3	2	90	3
3	1	16	1	5	2	1	2	2	4.0	1	1	1
4	2	19	2	7	2	4	4	1	3.5	1	3	3
5	2	15	2	1	1	4	4	2	1.5	1	2	2
n...

♦ Variables construidas.

Planteo de las relaciones causales

El análisis de los datos se puede comenzar con las **tablas de distribuciones de frecuencias simples** de todas las variables incluidas en la investigación, que proporcionan datos meramente descriptivos. El papel de estos cuadros es también realizar controles en la base de datos. Es necesario comprobar la exactitud del ingreso de los datos, primero mediante la realización de estas distribuciones frecuencias para determinar el número de casos que corresponden a los distintos valores o categorías de la variable o realizando cuadros de dos variables.

Para establecer relaciones entre variables es preciso realizar **cuadros bivariados o multivariados** en concordancia con los planteamientos de la investigación que nos indican qué variables han de asociarse. La rapidez de las computadoras presenta la tentación de hacer cruces de cada variable con cada una de las demás más que planificar el análisis, este procedimiento es el menos recomendado y deseable.

Sólo mediante datos obtenidos por vía experimental es posible, en sentido estricto, inferir nexos causales. Podemos aproximarnos a un análisis causal cuando a partir de relaciones postuladas teóricamente constatamos la existencia de asociación empírica.

Cuando el análisis se centra en el **examen de la asociación** entre las variables, éstas se deben ordenar progresivamente en el tiempo, y se parte del supuesto de una relación causal entre estas. El análisis supone que la variable dependiente puede ser explicada de algún modo o en cierto grado por la variable independiente.

El investigador debe establecer qué variables considera dependientes y cuáles independientes. La dependencia y la independencia no son propiedades inherentes ni fijas de las variables, una variable dependiente en una relación puede convertirse en independiente en otra. Los objetivos de la investigación determinan su status.

BIBLIOGRAFIA

- Barton, A. (1955) Concepto de espacio de atributos en Sociología en Boudon & Lazarsfeld Métodos de las Ciencias Sociales, Volumen I, Barcelona, Editorial Laia
- Benson, O. (1974) El laboratorio en Ciencia Política, Buenos Aires, Ed. Amarroutu
- Blalock, H. (1982) Introducción a la Investigación Social, Buenos Aires, Editorial Amarroutu
- Boudon, R. (1970) Los Métodos en Sociología, Madrid, Editorial Arredondo
- Briones, G. (1985) Métodos y Técnicas de Investigación para las Ciencias Sociales, México, Editorial Trillas.
- Caplow, T. (1974) La Investigación Sociológica, Barcelona,
- Cortes, F. (1987) Métodos Estadísticos Aplicados a la Investigación en Ciencias Sociales. Análisis de Asociación, México, Editorial Colegio de México.
- García Ferrando, M. (1985) Socioestadística, Madrid , Alianza Editorial
- Lazarsfeld, P. & H. Menzel, (1961) Relaciones entre propiedades individuales y propiedades colectivas en Boudón y Lazarsfeld, Método de las Ciencias Sociales, Vol. II, Barcelona, Editorial Laia.
- Mayntz, R. (1980) Introducción a los Métodos de la Sociología Empírica, Madrid, Alianza Editorial
- Mora y Araujo, M. (1971) El Análisis de Datos en la Investigación Social, Buenos Aires. Editorial Nueva Visión.
- Noelle, E. (1970) Encuestas en la Sociedad de Masas, Madrid, Alianza Editorial
- Padua, F. (1969) Métodos y Técnicas de la Investigación en Ciencias Sociales, Mexico, Editorial Siglo XXI
- Polít, D. y Hungler, B. (1989) Investigación Científica, México, Editorial Interamericana
- Sautu, R., J. Vujosevich y L. Griselli (1995) “Familia y rendimiento escolar en el ciclo primario en un barrio de clase trabajadora” Sociedad N° 6, Buenos Aires
- Travers, R. (1971) Introducción a la Investigación Educativa, Buenos Aires, Paidós
- Zeizel, H. (1974) Dígalo con Números, México, FCE.