

El dato científico y la matriz de datos

Walter Lauphan

Introducción

Toda investigación científica organiza los hechos bajo estudio de una manera comprensible a través de un esquema que sirve de puente entre la teoría y la empiria, al que denominamos “*dato*”.

Tal como es entendido en metodología de la investigación, el *dato científico* es el valor que toma una variable en una unidad de análisis. Es decir, todo *dato* tiene - como mínimo - tres componentes - unidad de análisis, variable y valor – y por esta razón se dice que su estructura es “tripartita”. Luego veremos que algunos autores sostienen que se deben incorporar elementos adicionales a la estructura del dato.

La *matriz de datos* es una forma de ordenar los datos de manera que sea visible su estructura (tripartita ó cuatripartita) y es de suma importancia en toda investigación porque es la manera ordenada y estructurada de interpelar la realidad con la teoría para hacerla inteligible (entendible).

En sí misma es una estructura funcional a muchas etapas de la investigación, sean estas etapas exploratorias - donde recién estamos comenzando a conocer el objeto de estudio - o etapas finales - cuando emprendemos el análisis de los datos obtenidos en la investigación - en cada etapa una de ellas se utilizará la matriz de datos con mayor o menor desarrollo, aprovechando sus capacidades para resumir un esquema complejo de consideración del contexto y los componentes que incidan en nuestro objeto de estudio.

Este escrito apunta a colaborar en la comprensión de la matriz de datos explicando brevemente sus componentes y cómo se articulan entre sí.

Componentes básicos

El primero en conceptualizar el *dato científico* y la *matriz de datos* fue el sociólogo y matemático noruego Johan Galtung (1966), quien sostuvo que todo dato científico tiene una estructura compuesta por tres elementos: *Unidad de Análisis*, *Variable* y *Valor*.

Con posterioridad, el filósofo y sociólogo argentino Juan Samaja (1996) criticó y reformuló esta definición sosteniendo que en realidad el dato científico tiene cuatro componentes esenciales que lo estructuran en lugar de tres, a saber:

- 1) *Unidad de Análisis*,
- 2) *Variable (con sus dimensiones y procedimientos)*
- 3) *Valor, e*
- 4) *Indicador*

Cuadro Nº1 - Componentes inherentes a todo dato científico (*)

UdeA (Unidad de Análisis)	V (Variable)	R (Valor)
	D (Dimensión)	I (Indicador)
	P (Procedimiento)	

(*) Cuadro extractado de Samaja (1996: 159)

A efectos didácticos desarrollaremos a continuación cada uno de los componentes del dato científico basándonos en la conceptualización tripartita de Galtung y luego incorporaremos los aportes de Samaja.

Unidad de Análisis

De acuerdo con Rut Vieytes (2004), *“la unidad de análisis es el elemento mínimo de estudio observable en relación con un conjunto de otros elementos que son de su mismo tipo”*. Es sobre estos elementos que vamos a estudiar el comportamiento de las *variables*.

Cuando determinamos la UdeA de una investigación nos estamos preguntando ¿qué o quienes serán observados (o medidos) en ella?

Las mismas pueden ser individuos, organizaciones (o cualquier colectivo social), territorios, cultivos, regiones o países, especies vegetales, etc. Son ejemplos de estas unidades de análisis: una determinada especie vegetal (ejemplo el maíz), los cultivos hortícolas de fruto bajo (pimiento, tomate, berenjena), los productores agropecuarios (individuos), las organizaciones corporativas agropecuarias (SRA, FAA, CRA), las pulverizaciones en soja, etc.

También pueden ser unidades de análisis los productos de la acción humana, materiales e inmateriales (por ejemplo ideas, normas, valores, representaciones, etc.), los procesos sociales o económicos (ejemplos: el proceso de innovación tecnológica en Entre Ríos a partir de la incorporación de la siembra directa, los procesos inflacionarios e hiper-inflacionarios, etc.), los sistemas productivos, los sistemas de labranza, etc.

La UdeA debe ser concebida - valga la redundancia - como una unidad, debe ser individualizable, pero también debe ser parte de un todo general o universo (compuesto por otras unidades de su misma categoría).

Por ejemplo, si quisiéramos conocer cual es el nivel educativo de los productores agropecuarios de la provincia de Entre Ríos. Nuestras *unidades de análisis* serán cada uno de los productores, los cuales en conjunto forman el *universo* “productores agropecuarios...”

Se debe tener en cuenta que, si bien uno de los requisitos de toda UdeA es que pueda ser diferenciable de otras unidades, esta no debe ser necesariamente un sujeto. Existen también las denominadas unidades de análisis colectivas compuestas por miembros o elementos de menor nivel.

Veamos un ejemplo: en un hipotético estudio sobre el nivel de vida de los hogares rurales del distrito Isletas (Dpto. Diamante), nuestras UdeA serán cada uno de los hogares residentes en ese distrito, los cuales estarán compuesto por el conjunto de personas que los integran.

En este caso, el hogar es una unidad de análisis colectiva que se puede descomponer en miembros de menos nivel que son las personas. Por otra parte, esos hogares forman parte a su vez del universo hogares rurales residentes en el distrito Isletas.

Otro ejemplo de unidad de análisis colectiva son las organizaciones agropecuarias a las cuales hicimos referencia anteriormente.

Ahora supongamos que emprendemos una investigación para conocer cual es la capacidad de penetración que tienen las pulverizaciones en los cultivos de soja. Cada uno de los *episodios de pulverización* (por ejemplo: la pulverización efectuada con fungicida en la soja de 2da del Sr. Schneider) será una de nuestras *unidades de análisis* dentro del *conjunto de pulverizaciones efectuadas* en un determinado período.

Variable

La unidad de análisis es también concebida como un colectivo en un sentido diferente del expuesto anteriormente, esto es, como portadora de un conjunto de características o cualidades particulares.

La noción de *variable* se refiere a los rasgos relevantes que permiten conocer el perfil de la unidad de análisis a investigar, son atributos o características de las unidades de análisis que, parece obvio decirlo, pueden *variar* asumiendo distintas - *cualidades o valores* - de unidad a unidad.

Veamos algunos ejemplos de variables correspondientes a distintas UdeA:

Cuadro N°2 – Ejemplos de variables

Unidad de Análisis	Variables
Suelos	Presencia de microorganismos Porosidad Tipo de suelo (<i>Vertisoles, molisoles, haplustoles, etc.</i>) Nivel de humedad
Cultivos hortícolas de fruto bajo	Rendimiento Calidad Grado de exposición a riesgos climáticos Modalidad de cultivo (<i>bajo cubierta, al aire libre</i>)
Productores agropecuarios	Perfil de productor (<i>Familiar/empresarial</i>) Nivel-socioeconómico (<i>Alto/medio/bajo</i>) Lugar de residencia (<i>En la explotación/fuera de ella</i>) Edad
Explotaciones Agropecuarias (EAP)	Nivel de capitalización Tamaño de la explotación (<i>Según la superficie explotada en ha: Pequeñas, medianas y grandes</i>) Régimen de tenencia de tierra (<i>propiedad, arrendamiento, ocupación de hecho, etc.</i>)
Pulverizaciones en soja	Tipo de boquilla utilizada (<i>cilíndrica/cónica</i>) Método de fumigación (<i>aérea/terrestre</i>) Velocidad de avance (<i>km/h</i>) Caudal de pulverización Capacidad de penetración en el cultivo

Cuando un atributo no varía de una unidad a otra, sino que siempre tiene la misma características, deja de ser una variable para transformarse en una constante y - por lo tanto - en un dato que contextualiza a nuestras UdeA.

Tomemos el ejemplo de los *suelos* del cuadro N°2. Si nuestra investigación se circunscribiera solamente a lotes con suelos vertisoles, el *tipo de suelo* habría dejado de ser una variable para transformarse en una constante de nuestras unidades de análisis: todos los lotes fueron sembrados en vertisoles.

Valor

De acuerdo con Francis Korn (1984), el *valor* - o *categoría* - es una de las diferentes posiciones o alternativas que presenta la variable y que adopta alguna unidad de análisis. Se puede expresar *cualitativamente* - a través de una clasificación por ausencia y presencia, por jerarquía u orden - ó *cuantitativamente* a través de magnitudes (números).

Una cuestión que se desprende de la definición anterior es que toda variable tiene una *escala de valores* (Variables cuantitativas) o *sistema de categorías* (Variables cualitativas) dentro de los cuales las UdeA pueden adoptar un valor específico.

En el cuadro Nº 2 ya hemos mencionado ejemplos de Valores para distintas variables, veamos en detalle algunos de ellos:

Variables de índole cuantitativa

- Las *fumigaciones* (UdeA) pueden ser clasificadas según la *velocidad de avance* a la que fueron efectuadas (Variable), las cuales se pueden expresar en km/hora siendo 20, 28 y 30 distintos valores que puede asumir la misma.
- La *edad* (Variable) del *productor agropecuario* (UdeA) puede asumir distintos Valores de acuerdo a los años cumplidos: 25, 26, 27... 37... 63..., etc.

Variables de índole cualitativa

- Las *fumigaciones* (UdeA) pueden asimismo ser clasificadas de acuerdo al *tipo de boquilla utilizada* (Variable). En principio estas pueden ser *cilíndricas* ó *cónicas* (Categorías de la variable).
- En el caso de los *productores agropecuarios* (UdeA) también podríamos caracterizarlos de acuerdo a su *nivel socioeconómico* (Variable), siendo una de las clasificaciones posibles la que distingue entre *nivel alto, medio y bajo* (Categorías de la variable)

La *escala de valores* (o *sistema de categorías*) de las *variables* son previamente definidas por el investigador y para ser correctamente construidas deben cumplir ciertos requisitos:

- Las clasificaciones serán **exhaustivas**. Esto significa que debe ser posible ubicar (clasificar) a cada una de nuestras UdeA en algún valor o categoría de la variable, ninguna de ellas puede quedar al margen de la clasificación.
- Los *valores o categorías* deben ser **mutuamente excluyentes**. Toda UdeA pueda ser clasificada solamente en un valor de la variable y no en varios a la vez.
- Los sistemas clasificatorios de las variables deben ser coherentes con las unidades de análisis que pretenden clasificar.

Veamos algunos ejemplos:

En los años '90 en Argentina, el sistema de siembra directa se abrió paso frente a una forma tradicional de explotar la tierra basada en la roturación de suelos. Según datos de

AAPRESID, entre los años 1978 y 1986, la superficie promedio sembrada con este sistema era de 2.000 ha en todo el país. Durante la campaña 1990/91 la superficie tratada bajo esta modalidad fue de 270.000 ha. Ya en la campaña 2000/01, la misma alcanzó 8.670.000 ha.

Siguiendo el ejemplo anterior, si comparamos el *sistema de categorías* de la variable “tipo de labranza” en investigaciones típicas de los años 1970 y 1990, las categorías serían las siguientes:

1970	1990
Convencional (Reja y vertedera)	Convencional (Reja y vertedera)
Convencional (Disco)	Convencional (Disco)
Vertical (cincel)	Vertical (cincel)
	Mínima (Siembra directa)

Es decir, para estudiar las mismas unidades de análisis (que en este caso podrían ser “cultivos de trigo”), necesitamos introducir una nueva categoría para que “sistema de labranza” sea una variable **exhaustiva**.

Veamos ahora un ejemplo de un **sistema de categorías excluyente**: si quisiéramos conocer como se distribuye la *superficie de la tierra en la provincia de Entre Ríos* (UdeA) según los *regímenes de tenencia* (variable), podemos reconocer distintas modalidades según el Censo Nacional Agropecuario 2002:

- *Propiedad*
- *Arrendamiento*
- *Contrato accidental*
- *Aparcería*
- *Ocupación*
- *Otros*

Cualquier porción de superficie explotada puede ser clasificada en una y sólo una de las categorías de esta variable, es decir que son *mutuamente excluyentes*.

Distinto sería el caso si nuestro propósito fuera conocer como se distribuyen las *explotaciones agropecuarias* (UdeA) según el *régimen de tenencia de la tierra* (variable). Si reflexionamos un instante nos damos cuenta que una misma explotación puede combinar distintas modalidades de tenencia (propiedad y contrato accidental, por ejemplo), de hecho es muy frecuente que ello suceda. En este caso el *sistema de categorías* de la variable asumirá otra estructura:

- *EAP con toda su tierra en Propiedad*
- *EAP con toda su tierra en Arrendamiento*

- EAP con toda su tierra en Contrato accidental
- EAP con toda su tierra en Aparcería
- EAP con toda su tierra en Ocupación
- EAP con toda su tierra en Otras modalidades
- EAP que combinan propiedad con otras modalidades
- EAP con otras combinaciones sin propiedad

De los ejemplos anteriores se desprende que los sistemas clasificatorios de las variables no son rígidos, pueden reconfigurarse con el paso del tiempo (sistema de labranza) y deben ser coherentes con la Unidad de Análisis considerada (régimen de tenencia de la tierra).

Las *escalas de valores* de una variable pueden variar también en diferentes contextos. En la producción agropecuaria estas diferencias se pueden observar, por ejemplo, cuando consideramos la superficie cultivada por las EAP (UdeA) en distintas regiones del país. A continuación presentamos la variable “escala de extensión” empleada por el INDEC en la presentación del Censo Nacional Agropecuario 2002, para explotaciones dedicadas al *cultivo de la vid* (Mendoza) y a la *producción de cereales y oleaginosas* (Entre Ríos):

Escala de extensión (ha)

Superficie implantada con Vid	Superficie implantada con Cereales y oleaginosas
Hasta 1	Hasta 5
1,1 - 3	5,1 - 10
3,1 - 5	10,1 - 25
5,1 - 7	25,1 - 50
7,1 - 10	50,1 - 100
10,1 - 15	100,1 - 200
15,1 - 25	200,1 - 500
25,1 - 35	500,1 - 1.000
35,1 - 50	1.000,1 - 2.500
50,1 - 70	2.500,1 - 5.000
70,1 - 100	5.000,1 - 7.500
Más de 100	7.500,1 - 10.000
	10.000,1 - 20.000
	Más de 20.000

Fuente: INDEC - Censo Nacional Agropecuario 2002.

Matriz de Datos

Llegados a este punto, estamos en condiciones de efectuar una primera aproximación al concepto de matriz de datos tal cual lo conceptualizara Galtung. De acuerdo a este autor, la

matriz de datos es una forma de ordenar los datos en una disposición cuadrangular de manera que sea visible su estructura tripartita. En la misma, “*las UdeA ocupan la dirección de las hileras, las variables la de las columnas y los valores aparecerán en el cruce de hileras y columnas*” (Samaja: 1996).

A continuación exponemos la matriz de datos de un hipotético estudio sobre productores agropecuarios (UdeA) de quienes hemos obtenido información referida a tres Variables: Edad, Nivel socioeconómico y Lugar de residencia. En las distintas celdas podemos visualizar los Valores que asumen los productores en las distintas variables.

UdeA \ Variables	Edad	Nivel socioeconómico	Lugar de residencia
Productor agropecuario 1	46	Alto	En la explotación
Productor agropecuario 2	54	Medio	En la explotación
Productor agropecuario 3	37	Medio	En la explotación
Productor agropecuario 5	52	Medio	En la explotación
Productor agropecuario 6	26	Alto	Fuera de la explotación

- Si se desea conocer las características de una unidad específica se recorre todo el renglón. Por ejemplo: el Productor N°2 tiene 54 años de Edad, posee un Nivel Socioeconómico medio y tiene su Residencia fijada en la explotación.
- Si se desea conocer cómo se distribuyen las unidades en los distintos valores de una variable se recorre la columna. De este modo, mientras que 5 productores residen en la explotación, 1 de ellos reside fuera de la misma.

Una de las finalidades mas importantes de la matriz de datos es la de sistematizar la información recolectada durante una investigación científica con el objetivo de resolver el problema que se planteado por la misma. Desde este punto de vista la matriz es una estructura sumamente útil en las últimas etapas de la investigación, cuando ya hemos recolectado nuestros datos y debemos emprender la tarea de procesar y analizar los mismos.

Sin embargo, la comprensión de la estructura del dato científico es fundamental desde el inicio de una investigación, dado que es el momento donde el investigador plantea cuales serán las unidades de análisis y los conceptos relevantes en su investigación.

Dimensiones e indicadores

Como indicamos anteriormente, Samaja sostiene que la estructura del dato científico tiene cuatro elementos en lugar de tres. Para dar cuenta de ello incorpora dos nociones fundamentales: Dimensión e Indicador.

Vamos a entender por “*dimensión de la variable*”... un aspecto parcial de la misma... que es relativamente independiente de otros aspectos y que, en conjunto, constituyen su sentido total” (Samaja: 1996)

Con la noción de “*dimensión*” Samaja explicita que con frecuencia los atributos (variables) de las UdeA que deseamos estudiar presentan una complejidad tal que para establecer el valor exacto que adoptan las mismas debemos considerar una combinación de atributos o propiedades que lo componen.

Por ejemplo, si se quisiéramos evaluar la *calidad comercial* (variable) de cierto fruto como puede ser el *tomate* (UdeA), investigaremos sobre distintas dimensiones que hacen a esa propiedad general: tamaño, madurez, textura, acidez, porción comestible, presencia de daños internos y externos, etc.

Obsérvese que el estudio de cada uno de estos atributos por separado nos brindará como resultado una *faceta* de la variable, pero ninguna nos dirá por sí sola “es de buena calidad comercial”. La *ausencia de daños en el fruto* y una *adecuada madurez* - que indicarían una “buena calidad” - pueden ir acompañadas de un *tamaño pequeño* que reduce la calidad comercial del fruto.

En consecuencia, el valor de la variable *Calidad comercial* será una inferencia, un salto de las partes a un todo.

Con la noción de *Indicador* hacemos referencia a *algún tipo de procedimiento que se aplique a alguna dimensión de la variable, para establecer qué valor de ella le corresponde a una unidad de análisis determinada. Así en el caso de los tomates podríamos efectuar los siguientes procedimientos:*

- La *porción comestible* puede medirse como el porcentaje del fruto disponible para el consumo humano.
- La *madurez del fruto* podría medirse a través del color, ya sea por medio de métodos subjetivos (apreciación humana de intensidades y tonos) ó a través de medidas objetivas a partir del uso de aparatos sensitivos electrónicamente a la reflexión de la luz producida por los colores de los objetos opacos.
- Para considerar el *tamaño* deberíamos registrar el *peso* y *volumen* del tomate. Nótese que esta dimensión contiene, a su vez, dos “sub-dimensiones” que pueden ser registradas mediante el empleo de ciertos instrumentos como calibrador, balanza de precisión y probeta graduada.

Y así se procederá con cada una de las dimensiones que resulten pertinentes.

En síntesis, para comprender una variable compleja, es necesario dimensionarla, y de acuerdo a la dimensión que se evalúe, habrá que hallar el procedimiento, que aplicado a la dimensión de la variable estudiada, permitirá inferir el valor de ésta. Cada dimensión tiene un procedimiento específico. Es importante destacar que el valor del indicador no es lo mismo que el valor de la variable.

Para terminar, explicitemos con claridad las diferencias entre las nociones de “dimensión” y de “valor”, que en ocasiones pueden ser confusas

Esto quiere decir que la relación entre las dimensiones es de naturaleza “conjuntiva” (un conjunto, con muchas características distintas). Muchas dimensiones conviven en un mismo “colectivo” posible. Los valores, en cambio, deben definirse por presencia-ausencia (esto quiere decir, una relación disyuntiva). Si se percibe un valor de la variable no puede percibirse otro.

Por ejemplo, un hombre puede ser alto (medirá específicamente 1,90 m), puede ser flaco (pesará solamente 75 kg), puede ser también con gran cabellera (tendrá rulos). Las dimensiones posibles son muchas (altura, peso, tipo y cantidad de pelo...), en cambio el valor posible que le corresponde a cada dimensión es uno y solo uno de los n valores de la escala.

Bibliografía consultada

Cátedra de Metodología Aplicada al Diseño. 2007. *Ficha de cátedra: Matriz de datos*. Buenos Aires: Fac. de Arquitectura Diseño y Urbanismo - UBA. Mimeo.

Hernández Sampieri, R. et al. 1998. *Metodología de la investigación*. México DF: Mac Graw-Hill Ed.

Korn, F. 1984. *Conceptos y variables en la investigación social*. Buenos Aires: Nueva Visión.

Samaja, J. 1996. *Epistemología y metodología. Elementos para una teoría de la investigación científica*. Buenos Aires: EUDEBA.

Vieytes, Rut. 2004. *Metodología de la investigación en organizaciones mercado y sociedad: epistemología y técnicas*. - 1ª ed. – Buenos Aires: Ed. de las Ciencias.