

Denis Baranger

**CONSTRUCCIÓN
Y ANÁLISIS
DE DATOS**

**INTRODUCCIÓN AL USO DE TÉCNICAS
CUANTITATIVAS EN LA INVESTIGACIÓN SOCIAL**

Posadas, 2009

Denis Baranger

Construcción y análisis de datos : introducción al uso de técnicas cuantitativas en la investigación social /
Denis Baranger ; con colaboración de Fernanda Niño. - 3a ed. - Misiones : el autor, 2009.
100 p. ; 21x15 cm.

ISBN 978-987-05-6307-5

1. Metodología de la Investigación. I. Niño, Fernanda, colab.
CDD 001.42

Fecha de catalogación: 20/04/2009

CAPÍTULO 2

LA MATRIZ DE DATOS

Ante cualquier objeto de investigación, ya sea éste de significación teórica o de importancia meramente práctica, las decisiones metodológicas propiamente dichas tienen necesariamente lugar dentro de un cierto marco conceptual. Dentro de ese marco se debe determinar: a) el grado en que dicho objeto es susceptible de ser producido en tanto objeto científico dentro de la estructura de la matriz de datos; y b) todas las operaciones que hagan a la producción del objeto dentro de este esquema, y muy particularmente las que tengan que ver con la definición de las unidades de análisis y de las propiedades que les sean aplicables.

En este capítulo, mostraremos primero el modo en que se articulan los conceptos de ‘unidad de análisis’, ‘variable’ y ‘valor’ bajo la forma de la matriz de datos, y ejemplificaremos como se puede expresar una hipótesis en ese lenguaje. Nos encontramos ahora en condiciones de aplicar estos conceptos para desmenuzar el significado empírico de cualquier proposición.¹ Todo el problema consiste precisamente en expresar en el lenguaje de los datos el contenido de un enunciado, por lo que la tarea es propiamente la de lograr una *traducción* adecuada de ese contenido.

Comenzaremos con varios ejemplos de ejercitación de esta operación de traducción para generar en forma práctica la idea de la matriz de datos. Luego nos limitaremos a glosar las partes más pertinentes del texto de Galtung, e introduciremos algunos elementos acerca de modalidades alternativas de diseño de una matriz. Finalmente, la distinción de Lazarsfeld y Menzel entre propiedades individuales y colectivas nos permitirá terminar de dibujar un cuadro somero de las distintas posibilidades abiertas en el proceso de construcción de los datos.

1. LA FORMA DE LA MATRIZ DE DATOS

Si afirmamos: «En 1980, había en la Provincia de Misiones 50.553 hogares con necesidades básicas insatisfechas»,² podemos distinguir en este enunciado:

- 1.- Una unidad de análisis ⇔ la “Provincia de **Misiones**”;
- 2.- Una variable ⇔ el “*Número de hogares con necesidades básicas insatisfechas*”;
- 3.- Un valor ⇔ “50.553”.

Estrictamente, limitándonos a esta sola unidad de análisis, puede parecer poco pertinente hablar de una ‘variable’, puesto que ésta no variaría. No obstante, el número de hogares con necesidades básicas insatisfechas pudo haber sido otro. De hecho la ‘variable’ es tal en la medida en que hace posible la comparación entre varias unidades de análisis; así, por ejemplo, para ese mismo año de 1980 la unidad de análisis Capital Federal presentaba en esta variable un valor de “67.692”.³

Este enunciado es de los más simples, desde luego, como que se refiere a una sola característica

¹ Es esencial comprender que *todo* enunciado fáctico es, por definición, susceptible de ser reducido a estos tres elementos. No queremos decir con esto que siempre haya que proceder así (existen, por ejemplo, enunciados no-fácticos con sentido); ni tampoco que ésta sea la única ni la mejor manera de obrar en todos los contextos (a un psicoanalista este esquema no le aportaría nada para interpretar el discurso de su paciente). Sí, en cambio, sostenemos que no existe investigación empírica cuantitativa que no descansa en algún tipo de matriz de datos.

² Datos de Argentina, 1984.

³ Otra posibilidad de comparación es -como se verá luego- considerar la misma unidad de análisis en distintos puntos del tiempo.

de un objeto único. Ocurre que muchas veces estamos interesados en proposiciones que se refieren a toda una clase de objetos, y en las que se establecen relaciones entre varias de sus propiedades. Por ejemplo: « En las elecciones para Gobernador del 6 de septiembre de 1987 en Posadas, las mujeres votaron por los candidatos radicales en mayor proporción que los varones». Deberíamos distinguir aquí:

1. Varias unidades de análisis

⇒ los “electores”⁴ del 6 de septiembre de 1987 en Posadas;

2.- Dos variables

⇒ 2.1. *Sexo* ;

⇒ 2.2. *Dirección del voto*;

3.- Los valores que conforman estas variables

⇒ 3.1. “1”(Varón)/“2”(Mujer);⁵

⇒ 3.2. “1”(Radical)/“2”(No radical).⁶

La matriz es una forma de hacer inmediatamente visible la estructura tripartita de estos datos. Así, suponiendo que se haya trabajado con una muestra de 10 electores, tendríamos:

Tabla 2.1 Ejemplo de matriz de 10 x 2

Unidades de análisis	Variables	
	1. Sexo	2. Voto
1	1	2
2	2	1
3	1	2
4	1	2
5	2	1
6	2	2
7	2	1
8	1	1
9	1	2
10	2	2

Cada fila de la matriz corresponde a una unidad de análisis (identificada por un código de 01 a 10), cada una de las dos columnas a una variable, y en las celdas figuran los valores.⁷ Entre los varones, obtendríamos una proporción de 1/5 votos por el Radicalismo; entre las mujeres habría

⁴ Operacionalizar una proposición supone contar con una definición precisa de la unidad de análisis. En este caso, se podría definir como “electores” a todas las personas inscriptas para esa fecha en el padrón electoral de Posadas, por ejemplo.

⁵ Obsérvese que hemos adjudicado en forma arbitraria los códigos numéricos “1” y “2” para simbolizar respectivamente los valores “masculino” y “femenino”.

⁶ Es evidente que en cuanto a esta segunda variable ‘*dirección del voto*’ se podrían distinguir otros tantos valores como partidos se presentaron a dichas elecciones, pero dos valores son suficientes para traducir adecuadamente el significado de la proposición que nos ocupa. Este ejemplo es ilustrativo de cómo cualquier clasificación politómica (de más de dos categorías o valores) puede reducirse a una **dicotomía** (una clasificación de dos valores).

⁷ En álgebra matricial, se define más genéricamente como ‘matriz’ a “cualquier conjunto rectangular de números” (Namboodiri, 1984: 8).

3/5 votos radicales. Como $3/5 > 1/5$, deberíamos aceptar como verdadera nuestra hipótesis.⁸

Esta proposición, aun siendo verdadera, sería todavía muy puntual. Si aspiráramos a lograr un conocimiento más general del comportamiento electoral de la población argentina, podríamos tener interés en formular una hipótesis como «En 1987 las mujeres votaron por el Radicalismo en mayor medida que los varones». En este caso las variables y los valores serían los mismos, pero sería conveniente que nuestras unidades de análisis se multiplicaran hasta abarcar electores de todos los distritos del país. En este caso, lo que era una propiedad constante de nuestras unidades de análisis “estar inscripto en el padrón electoral de Misiones” pasaría a funcionar como una variable adicional que podríamos denominar ‘Provincia’.⁹ Con una muestra de 500 electores tendríamos:

Tabla 2.2: Ejemplo de matriz de 500 x 3

Unidades de análisis	Variables		
	1. Sexo	2. Voto	3. Provincia
1	1	2	17
2	2	1	21
3	1	2	4
4	1	2	7
5	2	1	8
6	2	2	14
7	2	1	13
8	1	1	1
9	1	2	1
10	2	2	2
...
499	1	2	24
500	2	1	19

Así, por ejemplo, la unidad de análisis “006” sería una “mujer” que votó “No-radical” en “Misiones”. De un modo más general, se puede describir a cualquier matriz de datos como respondiendo a la siguiente estructura:

Tabla 2.3: Forma general de la matriz de datos

	$V1$	$V2$	$V3$...	Vj	...	Vp
$UA1$	R11	R12	R13	...	R1j	...	R1p
$UA2$	R21	R22	R23	...	R2j	...	R2p
$UA3$	R31	R32	R33	...	R3j	...	R3p
...
UAi	Ri1	Ri2	Ri3	...	Rij	...	Rip
...
UAn	Rn1	Rn2	Rn3	...	Rnj	...	Rnp

Fuente : adaptado de Galtung (1966: I, 3).

⁸ Obsérvese que si la relación hubiera sido $3/5 > 2/5$, también se verificaría la hipótesis; en efecto, ésta se encuentra formulada con escasa precisión. Por otra parte, a los efectos pedagógicos prescindimos por el momento de toda consideración sobre la **significación estadística** del resultado (que sería nula, con una muestra de $n=10$).

⁹ Los valores de esta variable podrían corresponder a las 23 Provincias más Capital Federal, y se codificarían de ‘01’ a ‘24’.

En cualquier investigación se considerará un número finito 'n' de unidades de análisis (UUA) y un número finito 'p' de variables (V). UA_n será entonces la última (o la "enésima") unidad de análisis incluida en una investigación.¹⁰ Por su parte cada una de las variables se compondrá de un número 'r' de valores. En el esquema presentado R_{ij} es el valor que presenta la UA_i en la variable j. Esta es muy precisamente la forma en que aparecen los datos en el monitor de una computadora; y es por tanto la forma en que deberán estar dispuestos nuestros datos para estar en condiciones de ser procesados electrónicamente.¹¹

Nos encontramos ahora en condiciones de recapitular los tres principios fundamentales (Galtung; 1966: I,4-6) que debe satisfacer una matriz de datos en su construcción.

1. Principio de *comparabilidad*: la proposición ' UA_i, V_j da el valor correspondiente en R_k ' debe ser verdadera o falsa para cada i, j y k . En otras palabras, a la idea básica de que todas las unidades han de ser medidas en las mismas variables, se agrega la condición previa de que cada una de las combinaciones de una variable determinada con una unidad de análisis debe *tener sentido*: debe ser verdadero o falso que una combinación ($UA_i V_j$) presenta un valor determinado R_{jk} . Por medio de este principio se hacen comparables las variables, las unidades y los valores. La comparabilidad se obtiene cuando las tres series se ajustan las unas a las otras de tal manera que se satisfaga la condición enunciada en este primer principio. Así, si UA_i es una nación, V_j la variable 'estado civil' y R_{jk} la lista de los elementos "casados, solteros, viudos y divorciados", cualquier combinación no será ni verdadera ni falsa, sino que simplemente no tendrá sentido

2. Principio de *clasificación*: para cada variable V_j la serie de las categorías de respuestas R_{jk} debe producir una clasificación de todos los pares ($UA_i V_j$) ($i=1, \dots, n$). Para cada variable la serie de sus valores debe formar una clasificación; para cada par $UA_i V_j$ deberá haber un R_{jk} (exhaustividad), y sólo uno (exclusión mutua) en que el par pueda ser ubicado. Aplicado a un cuestionario de encuesta, este principio significa que cada interpelado deberá marcar una, y sólo una, respuesta o categoría.¹²

3. Principio de *integridad*: para cada par ($UA_i V_j$) debe hallarse empíricamente un valor R_k . En lo que hace al trabajo empírico de llenado la matriz, el *desideratum* es no dejar ninguna celda vacía. En la práctica, se debe intentar que la cantidad de celdas sin información se mantenga lo más baja posible.¹³

Desde el punto de vista de la matriz de datos, es factible considerar que las investigaciones pueden diferenciarse según el número de unidades de análisis y de variables que toman en cuenta. Se puede pensar así en investigaciones más o menos "intensivas" -según la cantidad de variables que consideren- y más o menos "extensivas" -de acuerdo al número de unidades de análisis que sean observadas. Combinando ambas características, generamos una tipología de las investigaciones posibles.

Según Galtung, «Lo ideal es la combinación [muchas, muchas] -tantas unidades y tantas dimensiones como sea posible. Sin embargo, nosotros suponemos que la palabra 'muchas' se utiliza de tal manera que esto es imposible, por falta de recursos tales como tiempo, energía, personal y dinero» (1966: I, 9). Del mismo modo la investigación del tipo [una, una] no tendría

¹⁰ Hemos modificado la simbología de Galtung. Así, hablamos de n (y no m) unidades de análisis, lo que presenta la ventaja de ser congruente con el uso más habitual que se le da al símbolo 'n' (o 'N') para designar al número de casos (o de UUA) seleccionados para una muestra o de elementos existentes en el universo. En aras a una mayor simplicidad de exposición, renunciamos también a diferenciar 'variable' de 'estímulo' (y 'objeto'), 'unidad de análisis' de 'sujeto', 'valor' de 'respuesta'.

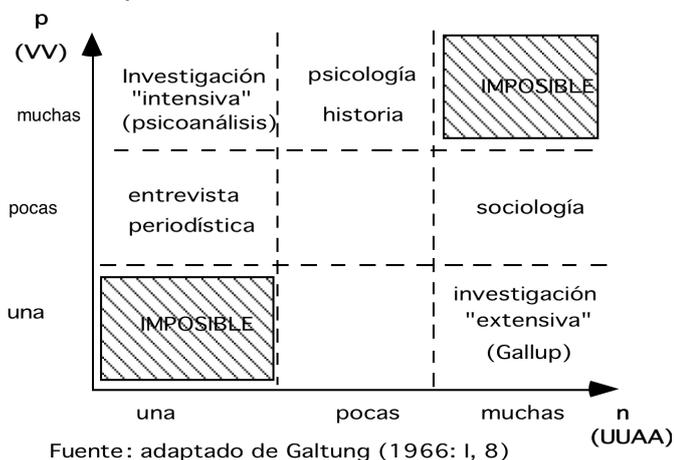
¹¹ Es también la forma más práctica de disponer los datos para su tabulación manual.

¹² En el caso en que al encuestado se le permite marcar varias respuestas, se le está presentando bajo una misma pregunta varias variables, una para cada categoría de respuesta, de tal manera que se encuentra ante dos valores posibles para cada variable: aceptación o rechazo.

¹³ «Como regla general de tipo práctico, puede decirse que un 10% es el máximo absoluto de celdas vacías admisibles en cualquier columna o cualquier fila de M y que un 5% es un máximo más aconsejable» (Galtung; 1966: I, 6).

sentido. El punteado de las líneas que delimitan las celdas intenta representar la dificultad en establecer un límite rígido entre las distintas cantidades de casos y de variables. El diagrama nos invita a pensar cómo, más allá de su ángulo inferior izquierdo, existe un vasto espacio de posibilidades para realizar investigaciones de naturaleza sumamente diversa y que en cuanto tales demandarán distintas habilidades. Las diferentes celdas representan en definitiva distintas **estrategias** de investigación.

Figura 2.1: Tipos de investigación, según el número de variable y de unidades de análisis involucradas



Si la matriz se reduce a una sola unidad -o a una única variable- se dice que "ha degenerado". En efecto, en ambos casos se termina renunciando a la posibilidad de la **comparación**, procedimiento intelectual que se encuentra en la base misma de la posibilidad del conocimiento. Como explica muy bien Galtung, existen varias razones por las que resulta conveniente evitar estas situaciones. Así, cabría preguntar "¿Por qué esta unidad y no otra?". Suponiendo en efecto que se pretendiera estudiar un sistema social, ¿Qué justificaría elegir un informante antes que otro?. No podemos sostener la creencia en el informante "puro"; la mejor prueba de ello es que si optáramos por otra unidad de análisis podríamos llegar a conclusiones muy distintas sobre nuestro objeto.¹⁴ La posibilidad misma de constatar variaciones y diferencias, y de evitar caer en estereotipos, requiere el poder comparar entre sí varias UA. Por la misma razón tampoco conviene trabajar con una sola variable: no existe la pregunta "pura".¹⁵ Disponer de varias variables permite comparar respuestas entre sí, y de este modo poder situar en un contexto la respuesta presentada en una variable, así como detectar patrones de respuesta.¹⁶

Finalmente, si el objetivo del conocimiento científico es establecer relaciones entre variables, su instrumento privilegiado habrá de ser la **correlación** -esto es, comparar valores en varias variables para un conjunto de UA. Y como se entenderá, no es posible el uso de la correlación sino a partir de un cierto número de unidades y contando por lo menos con dos variables.

La forma de la matriz de datos permite pensar con la mayor claridad la articulación entre las tres series de elementos (UA_i , V_j y R_{jk}) que concurren en la constitución del dato. Sin embargo, no siempre esta articulación se percibe nítidamente de modo inmediato. Un caso interesante se

¹⁴ «En Química o Física a menudo parece haber sido resuelto el problema de encontrar el caso puro. Si un químico desea comprobar una proposición acerca del sulfuro, puede tomar una cantidad cualquiera de sulfuro químicamente puro (siempre que su forma cristalina sea irrelevante para el experimento) y tratarlo como si fuera un puro y verdadero representante del sulfuro, S.», dice Galtung (1966: I, 10). Pero éste no es el caso en las ciencias sociales.

¹⁵ Constatación que se encuentra a la base misma de la necesidad de elaborar **índices** -esto es, variables complejas- para representar dimensiones complejas de una clase de fenómenos (cf. *infra*, capítulo 5).

¹⁶ El mismo Galtung señala excepciones a esta necesidad de contar con varias variables: «Cuando se busca la simple información acerca del contexto, o cuando se investiga una dimensión de actitud que ha ocurrido muy a menudo en el debate diario, quien investiga a través de una encuesta parece justificado al limitarse a una sola pregunta. Las encuestas de Gallup caen en una de estas categorías o en ambas, lo que las hace metodológicamente justificables desde este punto de vista» (1966: I, 10).

plantea cuando consideramos la misma unidad de análisis en distintos puntos del tiempo; “Misiones” en 1991 presentará muy probablemente un valor distinto en la variable “N° de hogares con NBI” que el anotado para 1980. En rigor, cabría inquirir si “Misiones” en 1991 es la **misma** unidad de análisis que en 1980 o si se trata de dos unidades distintas; es obvio que en el fondo la respuesta a una pregunta de este tipo es materia de convención.

Interesa destacar en términos prácticos cómo se puede zanjar técnicamente esta cuestión: ¿Qué hacer cuando no se trabaja con varias UUAA sino con una sola, la que ha sido medida en el mismo conjunto de variables en distintos puntos del tiempo? Supongamos que contamos con un conjunto de datos para una UA, la Universidad Nacional de Misiones, medida en dos variables ‘N° de Nuevos Inscriptos’ (NI) y ‘N° de Desertores’ en cinco años sucesivos. Podríamos, por supuesto, presentar los datos bajo la forma de una matriz que hubiera degenerado en un único renglón, como en este ejemplo:

UA	NI82	NI83	NI84	NI85	NI86	De82	De83	De84	De85	De86
UNaM	858	983	1349	2358	1996	387	561	947	1530	1242

Nada nos prohíbe presentar nuestros datos bajo la forma de este vector-fila. Procediendo de este modo, estamos midiendo nuestra única UA -la UNaM- en diez variables distintas: en efecto, NI₈₂ es una variable, NI₈₃ otra, etc. Claro que es ésta una manera impráctica de presentar este conjunto de datos; no sólo esta forma es conceptualmente oscura sino que, en especial, es impropia para ser trabajada mediante cualquier programa estadístico. Obsérvese que al disponer de una sola UA, no es posible el cálculo de ninguna correlación.

Existe una alternativa mucho más eficiente; y es tratar a nuestra UA como si fueran otras tantas UUAA como medidas en diferentes puntos en el tiempo disponemos. Es decir, se asume que una UA es idéntica a sí misma, sólo en un determinado punto del continuo temporal. Debemos entonces presuponer que $UA_t \neq UA_{t+1}$. De hecho, así es cómo se resuelve técnicamente la cuestión: UNaM₈₂ y UNaM₈₃ funcionarán como dos unidades de análisis distintas. En la matriz que generemos, cada renglón corresponderá entonces a un año diferente:

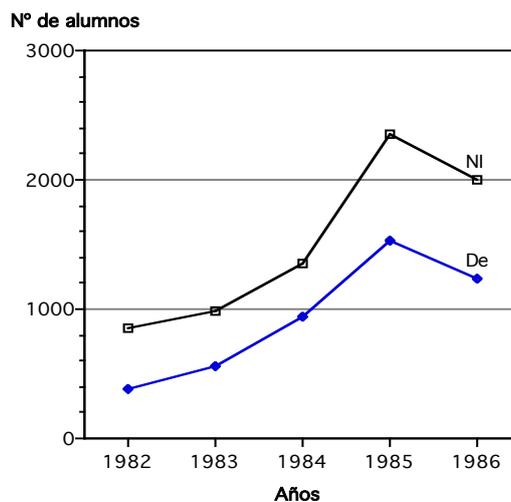
Años	Nuevos inscriptos	Desertores
1982	858	387
1983	983	561
1984	1349	947
1985	2358	1530
1986	1996	1242

Los mismos datos se han reestructurado asumiendo convencionalmente que corresponden a cinco UUAA medidas en dos variables. Bajo esta nueva presentación, salta a la vista la relación existente entre las dos variables: el crecimiento del número de desertores muestra una clara asociación con el incremento en el número de nuevos inscriptos.

Ahora bien, suponiendo que deseemos graficar estas dos series, la siguiente figura nos permite reconsiderar la cuestión desde un tercer punto de vista. Conceptualmente, se puede pensar ahora que estamos en presencia de una sola unidad de análisis -la UNaM- medida en tres variables: el año -representado en el eje horizontal-, su número de nuevos inscriptos y su número de desertores -ambas medidas en el eje vertical.¹⁷

¹⁷ Desarrollando una idea de Cattell de 1952, Rogers y Kincaid proponen la idea del “cubo de datos” (*Data-Cube*) una matriz tridimensional construida en base a (1) Unidades de Análisis (individuos o relaciones), (2) Variables, y (3) Tiempo (1981: 79-81).

Figura 2.2: Evolución de los nuevos inscriptos y desertores en la UNaM, 1982-86.



Fuente : Datos de la Secretaría G. Académica de la UNaM.

Ejemplo que debería terminar de convencernos acerca de este punto: no existe ninguna posibilidad de aplicar mecánicamente la estructura de la matriz a un conjunto de observaciones, sino que se requiere de todo un trabajo conceptual previo que es el que realmente produce -o “da”- los datos. En la sección siguiente, partiendo de la idea de ‘niveles de inclusión’, ahondaremos algo más en esta cuestión.

2. NIVELES DE INCLUSIÓN

Las variables son susceptibles de ser clasificadas de acuerdo a varios criterios, entre otros:

- el nivel de medición en que se inscriben: nominal, ordinal, intervalar;
- el papel que cumplen en el análisis: desde este punto de vista se podrá distinguir entre variables independientes, dependientes e intervinientes (cf. *infra*: capítulo 4);
- su grado de complejidad: se pueden reconocer así indicadores -variables simples- o índices.-variables complejas o compuestas- (cf. *infra*: capítulo 5)

Otro criterio para distinguir entre tipos de variables tiene que ver con el tipo de unidad de análisis al que se aplican; de acuerdo a éste, habrá que distinguir entre variables de individuo y variables de colectivo. Esta distinción es de fundamental importancia toda vez que en un proceso de investigación se encuentren involucrados diversos niveles de inclusión, y nos ocupará en esta sección.

2.1 Objeto de investigación y unidades de análisis

Retomemos nuestro concepto de ‘unidad de análisis’ (UA). De acuerdo a lo establecido en la primera sección, una UA es un sistema definido a partir de determinadas propiedades constantes. Lo primero que cabe aclarar es la diferencia entre el objeto de la investigación y la(s) UA con la(s) que se trabajará. Es importante destacar que no todo *objeto* científico constituye de por sí una UA. En primer lugar, existen objetos *teóricos* (por ejemplo, el modo de producción capitalista de Marx o el inconsciente de Freud) que son construcciones conceptuales que, aunque estén dirigidas en definitiva a “pensar” la realidad, en una investigación empírica no van a funcionar nunca como UA (ni tampoco como variables). En rigor estas construcciones conceptuales -explicitadas o no- se inscriben en el nivel de las condiciones previas a la construcción de cualquier dato científico. En términos de Althusser serían las que definen una *problemática* (1968), mientras que para Khun (1971) se ubicarían en el nivel del *paradigma*. Generalizando, se podría hablar para el caso de

construcciones *metaempíricas*.

Por otro lado, existen también objetos científicos -aunque ya no teóricos (en el sentido fuerte)- que tienen un correlato real empíricamente discernible.¹⁸ Para poner un ejemplo sencillo, podemos estar interesados en estudiar una comunidad local¹⁹ para conocerla y así planificar mejor las acciones de un programa de desarrollo tendiente a producir tales o cuales efectos. En este caso, el *objeto* de nuestro análisis estaría dado por ciertos procesos y características referidos a la comunidad como un todo la que, sin duda, en un determinado nivel funcionaría como una UA; sin embargo, más allá de alguna información referida a la ubicación geográfica de esta UA -la comunidad local-, a la situación de sus orígenes en el tiempo, a los hitos referentes a su “historia”, etc., nos encontraríamos muy pronto con que poco podríamos decir acerca de la misma si no dispusiéramos de datos acerca de los distintos integrantes -hogares, personas- que la componen: dentro de este estudio, se requeriría que cada uno de sus miembros funcionara a su vez como una *unidad de análisis*.²⁰ Describir a nuestro objeto ‘comunidad’ supondría entonces referirse a sus integrantes, a las acciones que realizan, y a las relaciones que mantienen entre sí.²¹

A título de ejemplo, podemos retomar la proposición que analizábamos anteriormente: «En 1980, había en la Provincia de Misiones 50.553 hogares con necesidades básicas insatisfechas»; considerábamos entonces a “Misiones” como la UA, al “Número de hogares con necesidades básicas insatisfechas” como una variable y a “50.553” como valor. ¿Pero cómo fue generado este dato, como se lo construyó? La respuesta es que, considerado en sí mismo, este dato supuso la definición de un universo “la Provincia de Misiones” que está compuesto por hogares, y desde este punto de vista dicho enunciado supuso considerar:

1. Varias UUAA ⇔ los **hogares** de la Provincia de Misiones;
2. Una variable ⇔ la “*Existencia de necesidades básicas insatisfechas*”;
3. Un valor ⇔ “Sí”.²²

Se puede pensar por lo tanto en la existencia de UUAA en distintos niveles de inclusión.

En la Figura 2.3, cada círculo pequeño simboliza una persona (en este caso, la unidad de menor nivel)²³, cada círculo mediano una unidad doméstica²⁴ (la unidad de segundo nivel), y el óvalo

¹⁸ Para disipar cualquier clase de equívoco, aclaremos que de atenderse a la distinción entre “objeto real” y “objeto de conocimiento” tal como la plantea Marx en el tercer apartado de la Introducción de 1857 (1971: I, 20 y ss.) los dos tipos de ‘objetos’ que distinguimos aquí formarían parte del “objeto de conocimiento”.

¹⁹ A los efectos que nos interesan, bastará con adoptar como definición de ‘comunidad local’ la de un grupo humano, de tamaño limitado, inserto dentro de una sociedad más amplia, y con una determinada base territorial.

²⁰ En realidad, hemos estado englobando en forma indiferenciada bajo esta expresión de ‘UA’ aspectos disímiles y que en tanto tales merecen ser distinguidos. Susana Torrado propone las siguientes convenciones terminológicas: «a) Unidades de análisis (UA): son las unidades “teóricamente” significativas desde la perspectiva de cada práctica o propósito analítico, por lo que su definición depende de las respectivas definiciones conceptuales.(...) b) Unidades de observación (UO): son las unidades acerca de las cuales se recogen datos a través del encuestamiento directo o indirecto». La distinción tiene su importancia porque, por lo general, la UA es operacionalizada a través de la definición de una UO. Torrado distingue también el conjunto de las Unidades de Cuenta (UC) y el de las Unidades estadísticas(UE); habitualmente, cada uno de estos conjuntos coincidirá con, o al menos incluirá, el conjunto de las UO (1983: 13-14).

²¹ El mismo Galtung incurre en algún equívoco acerca de la distinción entre ‘objeto’ y ‘UA’ cuando afirma: «A un cierto nivel en el análisis siempre se está estudiando solo una *unidad*, como cuando se hace una encuesta acerca del varón norteamericano y su conducta sexual» (1966: I, 11). De acuerdo a nuestro criterio correspondería para el caso hablar de *objeto*.

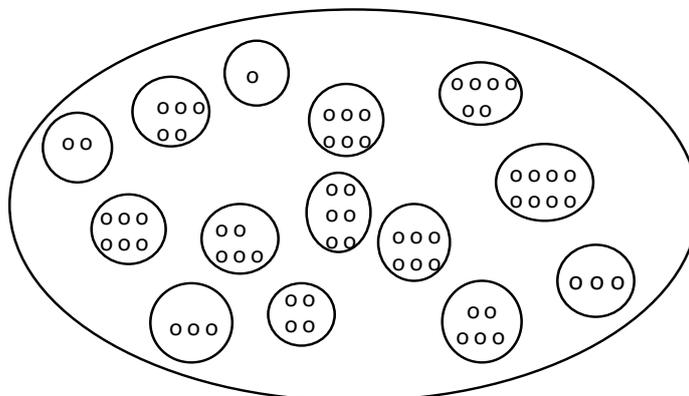
²² En el estudio sobre *La pobreza en Argentina* encontramos que los 50.553 hogares con NBI representan 39,2% de todos los hogares de Misiones. Para construir este dato, se consideraron los 128.962 hogares de la Provincia como UUAA., de los cuales 50.553 presentaban el valor “Sí” en la variable “Existencia de NBI” mientras que los restantes 78.409 presentaban el valor “No”.

²³ Como bien señalaban Lazarsfeld y Menzel, y el mismo Galtung, no es indispensable que la unidad de menor nivel coincida con personas individuales; por ejemplo, un individuo puede verse como un conjunto de roles, cada uno de los cuales constituiría la unidad de menor nivel. Una empresa puede ser considerada como un conjunto de normas, más o menos formalizadas, que se cumplen en menor o mayor medida, etc.

²⁴ Se puede definir la “unidad doméstica” como un «grupo de personas que interactúan en forma cotidiana, regular y permanente’ a fin de asegurar mancomunadamente el logro de uno o varios de los siguientes objetivos: su reproducción biológica, la preservación de su vida, el cumplimiento de todas aquellas prácticas, económicas y no económicas, indispensables para la optimización de sus condiciones materiales y no materiales de existencia»; lo que implica que

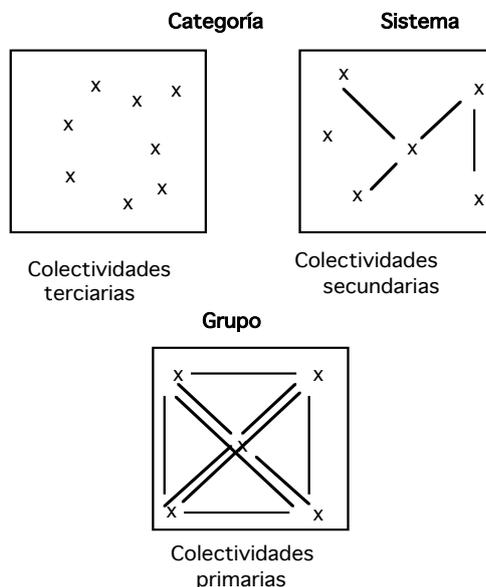
mayor representa una comunidad local (unidad de tercer nivel) integrada para el caso por catorce familias:

Figura 2.3: Unidades de análisis en tres niveles



Ahora bien, podemos creer que hay niveles de inclusión más interesantes que otros: aquellos niveles en los que se producen determinadas pautas o formas de organización, y que dan lugar a sistemas compuestos por UUA de menor nivel que mantienen relaciones entre sí. En otras palabras, niveles que se corresponden con conjuntos estructurados de sistemas. En el límite estos sistemas pueden constituirse en verdaderos “actores” sociales colectivos, con una lógica propia. Pero no es imprescindible limitarse en el análisis a este único tipo de sistemas. El mismo Galtung reconoce que, partiendo de los individuos como unidades, pueden construirse distintos tipos de *superunidades*.

Figura 2.4: Una tipología de colectividades



Fuente: Galtung, 1966: I, 37.

Así, las ‘colectividades terciarias’ corresponden a lo que se denomina más frecuentemente ‘categoría’ (o incluso ‘agregado estadístico’), es decir, un conjunto de unidades sin estructura, conjunto que puede definirse a partir de la posesión de uno o varios atributos en común pero entre

exista *unidad de residencia*, que algunos miembros estén vinculados por relaciones de *parentesco*, que esta UD opere como una *unidad de consumo* o de *funcionamiento doméstico* (1983: 15-17).

cuyas unidades no existen relaciones. Por ejemplo, todos los votantes en la ciudad de Posadas en las elecciones del 6 de septiembre de 1987; o bien, los productores de trigo de la Argentina, etc. Un segundo tipo de unidad colectiva sería lo que Galtung denomina ‘sistema’, en este caso cada una de las unidades de primer nivel se halla en relación con al menos otra unidad: podría tratarse de los socios de un club, de los empleados de un ministerio, o de los personas pertenecientes a una Facultad de Ciencias Sociales; existe ya alguna estructura del colectivo, aún cuando sea bastante laxa. Por último, en el ‘grupo’ todas las unidades se encuentran relacionadas las unas a las otras, lo que resulta en una estructura fuertemente conectada; podrá ser el caso de un grupo familiar, de los obreros en un taller, de los miembros de un pequeño departamento en una universidad, etc.

Ciertamente es difícil pensar que cualquier tipo de colectividad ‘terciaria’ (en el sentido de Galtung) llegue a funcionar como un “actor”. No obstante ello, niveles de inclusión que definen este tipo de superunidades son absolutamente legítimos y se utilizan frecuentemente en la práctica de la investigación aplicada: éste es el caso cuando se comparan entre sí unidades geográficamente determinadas como lo son los países, regiones, provincias, departamentos, etc.²⁵ Por lo tanto, y a nuestros efectos, podemos establecer que cualquier nivel de inclusión puede determinar un conjunto de UUAA susceptible de constituirse en un objeto pertinente para la investigación.

2.2 La tipología de Lazarsfeld y Menzel

Comenzaremos ahora a abordar la tipología de propiedades individuales y colectivas que se presenta en el clásico trabajo de Lazarsfeld y Menzel. Primero, aclararemos la distinción entre colectivos y miembros, lo que nos permitirá continuar ejemplificando cómo puede traducirse en datos el significado de diversas proposiciones; luego presentaremos los diferentes tipos de propiedades incluidos en la tipología de esos autores, y finalizaremos comentando algunos problemas que se suscitan en la aplicación de la tipología.

Colectivos y miembros

Entendiendo que cualquier investigación empírica va a tender a producir la corroboración de determinadas proposiciones que versan sobre objetos reales, el punto central estará en el carácter de ‘colectivos’ o de ‘miembros’ que se les pueda adjudicar a estos objetos definidos como UUAA. Las UUAA son ‘miembros’ si se las puede considerar como incluidas dentro de unidades de mayor orden denominadas ‘colectivos’. Por lo contrario, las UUAA son ‘colectivos’ cuando es posible descomponerlas en unidades de menor nivel llamadas ‘miembros’. Los de ‘colectivo’ y ‘miembro’ son por lo tanto conceptos que se definen estructuralmente: todo miembro lo es de un colectivo, y todo colectivo lo es a su vez de sus miembros. Así en nuestro ejemplo de la Figura 2.3, cada persona sería un miembro de su colectivo: la unidad doméstica; pero a su vez cada unidad doméstica podría considerarse como miembro con relación a la unidad de mayor nivel que sería la “comunidad”.²⁶

Ahora bien, para una proposición dada las UUAA podrán ser o bien miembros o bien colectivos, o bien ambas cosas, según cuales sean los niveles de inclusión implicados. Hay que diferenciar entonces tres posibilidades:

<i>Niveles de inclusión implicados</i>	<i>Carácter de las UUAA</i>
a) Un sólo nivel	Colectivos o miembros
b) Dos niveles	Colectivos o miembros
c) Más de dos niveles	Colectivos y/o miembros

²⁵ De hecho, algunas de estas unidades, como los países y las provincias, en la medida en que no son meramente geográficas, pueden perfectamente llegar a funcionar como actores sociales. Por lo demás, cualquier proyecto social debe desarrollarse en un espacio determinado; ello lleva con frecuencia a la necesidad de fijar los límites geográficos del alcance del proyecto. Las superunidades o sub-universos así delimitados pueden carecer por completo de significado sociológico y, sin embargo, aún así no dejaran de ser relevantes tanto para la implementación del proyecto como para la evaluación de sus resultados.

²⁶ Por lo demás, nada impediría considerar varias comunidades como miembros del colectivo ‘departamento’, ni varios departamentos como miembros del colectivo ‘provincia’, etc.

En el caso a), cuando sólo está implicado un nivel de inclusión, la distinción es escasamente relevante. Por ejemplo, si se afirma:

« A mayor cantidad de años de educación formal, mayor ingreso mensual».

1. Hay un conjunto de UUAA
 - ⇒ presumiblemente **personas**;
2. Hay dos variables:
 - ⇒ 2.1. “ *Nº de años de educación formal*”;
 - ⇒ 2.2. “ *Ingreso mensual*”;
3. Con sus valores:
 - ⇒ 3.1. “0”, “1”, “2”, “3”, etc.
 - ⇒ 3.2. “Hasta 100”, 101-200, 201-300, etc.²⁷

Si nuestras UUAA son personas consideradas en sí mismas, sin que se requiera hacer referencia a ningún nivel de inclusión mayor o menor, estaremos relacionando dos variables que hacen referencia a propiedades individuales.²⁸

Podemos ejemplificar el caso b) con la proposición:

« En 1980, las viviendas rurales de Misiones se encuentran electrificadas en menor medida que las urbanas»

1. Hay un conjunto de UUAA:
 - ⇒ “**viviendas**”, que se definen por una serie de propiedades constantes: “estar en Misiones”, “en 1980”;
2. Hay dos variables:
 - ⇒ 2.1. “*Condición de electrificación*”;
 - ⇒ 2.2. “ *Tipo de zona*”;
3. Con sus valores:
 - ⇒ 3.1. “Electrificada”/ “No electrificada”;
 - ⇒ 3.2. “Rural”/”Urbana”.

Aquí las unidades se definen por una parte a partir de una propiedad que les pertenece propiamente como es la “condición de electrificación”; pero la segunda propiedad que se les aplica -ser rurales o urbanas- tiene que ver con una característica del colectivo al que pertenecen.

En el caso c) las unidades pueden ubicarse, por así decirlo, “abajo”, “arriba” o “en el medio”. Si decimos:

« En las localidades rurales más aisladas, las unidades domésticas presentan un menor ingreso monetario per cápita ».

1. Hay un conjunto de UUAA:
 - ⇒ “ **unidades domésticas**”;
2. Hay dos variables:
 - ⇒ 2.1. *Condición de aislamiento de la localidad*;
 - ⇒ 2.2. *Ingreso monetario anual per cápita* ;²⁹
3. Con sus valores:
 - ⇒ 3.1. “Aislada”/ “No aislada”;
 - ⇒ 3.2. “Hasta 200”, “201-400”, “401-800”, “801-1600”, etc.

²⁷ Los valores de la variable “Ingreso mensual” aparecen agrupados por intervalos, como si los datos fueran a ser presentados en una tabla de contingencia; es claro que nada obliga a agrupar los valores de la variable: éstos podrían ser “93”, “168”, “346”, etc. (así tendrían que presentarse los valores si, por ejemplo, se quisiera realizar un análisis de variancia para diferentes niveles de educación formal).

²⁸ También podrían tomarse departamentos, o provincias, o países, como UUAA, si se consideraran, por ejemplo, las variables “*Porcentaje de personas con 12 o más años de educación formal* “ y “ *Porcentaje de personas con ingresos que superan los 500 pesos por mes* “. Ciertamente, para construir el dato habría que hacer referencia a unidades de menor nivel (las personas), pero ambas variables remitirían a propiedades colectivas.

²⁹ Variable que se podría definir, por ejemplo, como la suma en pesos de todos los ingresos monetarios de una unidad doméstica durante un año, dividida por su número de miembros.

Previamente, una breve digresión acerca del significado de esta hipótesis. A primera vista podría pensarse que la proposición hace referencia a una sola variable, interpretando que su alcance se encuentra limitado al universo de las unidades domésticas pertenecientes a las localidades rurales más aisladas. Sin embargo, la lectura cuidadosa del enunciado no debe dejar dudas sobre el hecho de que la pertenencia a una localidad rural más aislada es una característica que se debe tomar aquí como una propiedad variable (y no constante). En efecto, en la segunda parte del texto de la hipótesis se hace referencia a un **menor** ingreso per cápita, con lo que se introduce la idea de que dicho ingreso “menor” lo será con respecto a otras unidades que no comparten la propiedad de pertenecer a localidades rurales aisladas. Implícitamente, entonces, el enunciado de la hipótesis está significando:

« En las localidades rurales más aisladas, las unidades domésticas presentan un menor ingreso monetario *per capita* [que en las localidades rurales menos aisladas]». ³⁰

Retomando ahora el hilo de nuestro argumento, la hipótesis ejemplifica la posibilidad de que se encuentren implicados tres niveles, siendo que en este enunciado las UUAA se ubican en el nivel intermedio. En efecto, al hablar de ‘ingreso *per capita*’ se está haciendo referencia necesariamente a los integrantes de la unidad doméstica que funciona como un colectivo con respecto a sus miembros.³¹ Pero a su vez las UUAA -esas mismas unidades domésticas- se diferencian de acuerdo a una característica del colectivo de mayor nivel -la “localidad”, que puede estar en una situación de menor o mayor aislamiento- y con relación al cual funcionan como miembros.

Otro ejemplo de proposición que implica tres niveles, pero con una ubicación diferente de las UUAA podría ser:

« Los miembros de las unidades domésticas más recientemente asentadas en los barrios marginales de Posadas, perciben remuneraciones horarias menores por su trabajo, [que los miembros de las unidades domésticas con mayor antigüedad en esos barrios] » .

1. Hay un conjunto de UUAA:

⇨ **personas que residen en barrios marginales y trabajan;**

2. Hay dos variables:

⇨ 2.1. *Antigüedad del asentamiento de la unidad doméstica* ;

⇨ 2.2. *Remuneración horaria -promedio percibida por el trabajo* ;

3. Con sus valores:

⇨ 3.1. “Hasta 6 meses”, “7-12 meses”, “13-24 meses”, etc.;

⇨ 3.2. “Menos de 1 Peso”, “1-1,49”,

“1,50-1,99”, etc.

En este caso las UUAA pertenecen al menor nivel: se trata de individuos miembros de unidades domésticas que a su vez son miembros del colectivo más amplio “barrio”.³²

Propiedades individuales y colectivas

La tipología (Lazarsfeld y Menzel, 1979: 79) está basada en las operaciones de medición involucradas en la construcción de cada tipo de propiedad, e incluye un conjunto de propiedades para caracterizar a las UUAA que funcionen como miembros, y otro conjunto para las UUAA colectivas.

³⁰Utilizamos los corchetes para representar la parte implícita del enunciado. El caso podría ser distinto si el texto rezara: «En las localidades rurales más aisladas, las unidades domésticas presentan un **bajo** ingreso monetario per cápita» (pero ello siempre y cuando se contara con una definición **absoluta** de lo que se debe entender por un “bajo” ingreso per cápita).

³¹ El “ingreso *per capita* anual” puede ser definido como la suma de todos los ingresos monetarios de los miembros de una UD, dividida por el número de miembros de esa UD.

³² Nada impediría considerar que se trata aquí de una proposición con tres variables, haciendo que la hipótesis alcanzara a los pobladores de todos los barrios de Posadas, y no solamente los marginales. La “*Remuneración horaria-promedio*” continuaría siendo la variable dependiente, mientras que como variable independiente funcionarían tanto la “*Antigüedad del asentamiento de la UD*” como el “*Tipo de barrio*” (“Marginal”/“No marginal”).

Figura 2.5: Tipología de Lazarsfeld y Menzel

		Analítica
		Estructural
		Global
PROPIEDADES	COLECTIVAS	
	INDIVIDUALES	Absoluta
		Relacional
		Comparativa
		Contextual

En primer lugar tenemos las propiedades *analíticas*, que “son propiedades de colectivos que se obtienen a través de operaciones matemáticas con alguna propiedad de cada miembro” (1979:85). Así, el ingreso per cápita de una unidad doméstica es una propiedad analítica de este colectivo, obtenida a partir de la suma de todos los ingresos de sus miembros individuales. Los promedios son propiedades analíticas, al igual que todas las propiedades construidas como porcentajes a partir de propiedades de miembros : el porcentaje de analfabetos en la provincia de Catamarca, el porcentaje de migrantes recientes en el Partido de La Matanza, el porcentaje de habitantes de ascendencia polaca en el Departamento de Apóstoles, etc. También las correlaciones pueden usarse para caracterizar colectivos: así, en un estudio de sociología electoral sería posible caracterizar a las diferentes provincias argentinas según el grado de correlación existente entre clase obrera y voto peronista, por ejemplo.³³

Se distinguen también las propiedades *estructurales*, que “son propiedades de colectivos que se obtienen a través de algunas operaciones con los datos sobre las relaciones de cada miembro con todos o alguno de los otros miembros” (:86). Si se está tratando con clases (colectivos) de una escuela compuestas por alumnos (miembros), se puede a partir de la información que cada alumno ha dado sobre cuál es su “compañero preferido”, construir un sociograma, una representación gráfica de estas relaciones de preferencia dentro de cada grupo. Dentro de este contexto se suele hablar de “estrella” para referirse a los individuos que reciben muchas elecciones positivas (representadas por líneas que convergen hacia él). Las distintas clases de la escuela podrán entonces diferenciarse según el número de estrellas que exista en cada una, y ésta será una propiedad estructural de cada clase. Hans Zeisel presenta el siguiente ejemplo: a todos los integrantes de un grupo se les pide que expresen su actitud hacia los demás miembros, otorgándole a cada uno un puntaje variable en una escala de 5 puntos que va desde +1 (máxima aceptación) a -1 (máximo rechazo), con un punto medio neutro (0). Zeisel plantea que para medir una propiedad estructural del grupo, su nivel general de cohesión, puede utilizarse como índice el promedio aritmético de todas las puntuaciones otorgadas (Zeisel, 1980: 105-108).

Por último, tenemos a las propiedades colectivas *globales*, las “que no se basan en información sobre las propiedades de los miembros individuales” (:87). Se trata entonces de una categoría residual de propiedades colectivas: son globales todas las propiedades que no son ni analíticas ni estructurales. Así, se ha podido distinguir entre sociedades con escritura y sociedades sin escritura. Debe insistirse en que esta propiedad global de una sociedad (“el tener escritura”) es por completo independiente del porcentaje de sus miembros que saben escribir;³⁴ así, la China clásica era una

³³ Se trataría de una propiedad analítica de cada provincia obtenida a partir de la correlación entre dos propiedades analíticas (los porcentajes de votos peronistas y de obreros en cada departamento).

³⁴ De lo contrario, se trataría de una propiedad analítica.

sociedad con escritura, aunque el dominio de ésta fuera patrimonio de un grupo reducido de mandarines. También el número de habitantes de una ciudad (o de un departamento, provincia o país) es una propiedad global.³⁵ Otro ejemplo: el hecho de que la enseñanza de la religión sea obligatoria en las escuelas es una propiedad global de la Provincia de Salta, que la diferencia de Misiones. También podría registrarse para cada uno de los países de América Latina el porcentaje de su presupuesto que es destinado a la educación, una propiedad global que no siempre se ajusta al 25% recomendado por la UNESCO. Del mismo modo el número de diarios que se editan en una provincia es una propiedad global de dicho colectivo.

Las propiedades individuales más simples son las *absolutas*, “características de los miembros que se obtienen sin hacer uso de información ni sobre el colectivo ni sobre las relaciones del miembro que se describe con otros miembros” (:90). La mayoría de las propiedades que se utilizan habitualmente para describir individuos pertenecen a esta categoría: edad, sexo, estado civil, ingreso mensual, condición de alfabetismo, ocupación, religión, etc.

En cuanto a las propiedades *relacionales*, éstas “se computan a partir de la información sobre las relaciones sustantivas entre el miembro descrito y otros miembros” (:91). Así, la popularidad medida sociométricamente es una propiedad relacional de un miembro de un grupo. En el mismo ejemplo citado de Zeisel, cada miembro puede ser caracterizado tanto por la suma de los puntajes que ha obtenido de los otros miembros (“popularidad”) como por la suma de los puntajes que ha otorgado a los otros miembros (como una medida de su “sociabilidad activa”). También el poder, en su acepción weberiana clásica de “probabilidad de encontrar obediencia”, es una propiedad eminentemente relacional. Por su parte Bourdieu, refiriéndose a las clases sociales, distinguía entre la “situación” de clase -determinada a partir de un conjunto de características económicas, sociales y culturales, consideradas intrínsecamente- y la “posición de clase” -en que se encuentra una clase o sector por su ubicación dentro de una estructura. Lo que da lugar a diferenciar entre la clase determinada como una propiedad absoluta y la clase entendida como propiedad relacional: así, «la clase superior de una ciudad pequeña presenta la mayoría de las características de las clases medias de una ciudad grande» (Bourdieu, 1982: 5); si se tomara la clase como propiedad absoluta, no existiría clase superior en una ciudad pequeña.³⁶

Por su parte, las propiedades individuales *comparativas* “caracterizan a un miembro comparando su valor en alguna propiedad (absoluta o relacional) y la distribución de esta propiedad en el colectivo total del cual es miembro” (:92). Así, por ejemplo, el hecho de que un alumno tenga un CI³⁷ de 110 es una propiedad absoluta suya, pero el que este mismo alumno sea *el más inteligente* de la clase (colectivo) es una propiedad comparativa, significando que ningún otro alumno de su clase lo supera en cociente intelectual; en cambio, con un CI de 110 un alumno de otra clase puede no ser el más inteligente si en ésta existen otros alumnos con mayor CI. Del mismo modo, el ser el hijo menor en una familia (colectivo) no implica ningún valor particular -o absoluto- en la variable “edad”, sino tener una edad menor que la de los hermanos. Otro ejemplo: considerados absolutamente, los habitantes de villas-miseria aparecen como uniformemente pobres; empero en cualquier estudio de esta población es posible plantearse una estratificación de estas familias en base a indicadores económicos dividiéndolas en nivel “alto”, “medio” y “bajo”; es claro que el nivel “alto” lo será meramente en términos comparativos.

Finalmente, están las propiedades *contextuales*, las que “describen a un miembro por la propiedad de su colectivo” (:92). Si se afirma: “los pobladores de las áreas urbanas marginales

³⁵ Podría pensarse que se trata de una propiedad analítica, obtenida contando la “existencia” de todos sus habitantes, pero como lo sugieren Lazarsfeld y Menzel (1969: 96), la existencia no es una propiedad (o en todo caso no es una propiedad *variable*); antes bien, es la condición de toda propiedad. La “existencia” se define por las propiedades constantes que definen la pertenencia de un elemento a un universo de investigación.

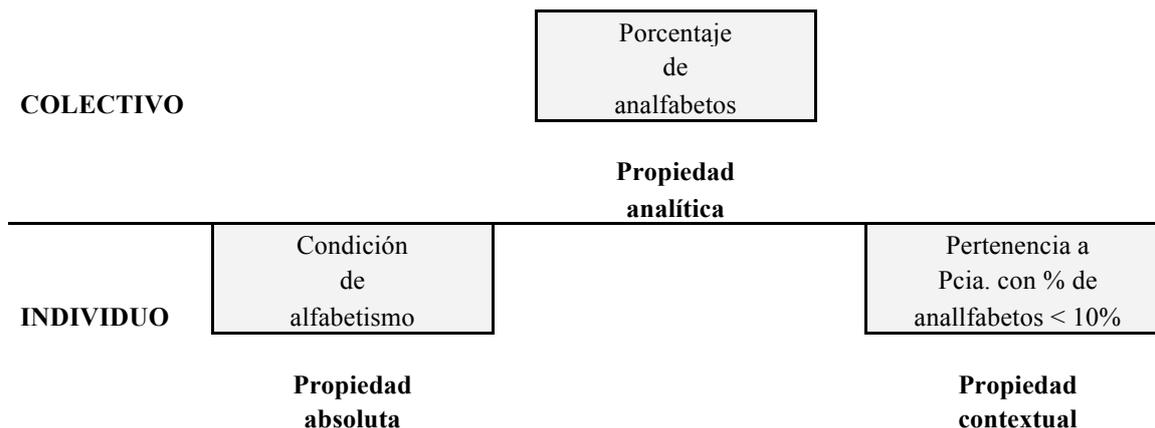
³⁶ El género es también una propiedad relacional por excelencia a la que Bourdieu trata de una manera en todo análoga a la clase. Así, hay una constancia transhistórica de «la estructura de la relación de dominación entre los hombres y las mujeres que se mantiene más allá de las diferencias *substanciales* de condición ligadas a los momentos de la historia y a las posiciones en el espacio social» (1998: 110).

³⁷ CI: Cociente Intelectual, medido a través de algún test de inteligencia.

presentan una mayor inestabilidad laboral”³⁸ , el “ser poblador de un área urbana marginal” es una propiedad contextual de los individuos. Lo dicho supone que cualquier propiedad colectiva (ya sea analítica, estructural o global) es susceptible de dar lugar a una propiedad contextual al aplicarse a cualquier miembro del colectivo.

Puede partirse, por ejemplo, de una característica absoluta de los miembros (como la condición de alfabetismo) para generar una propiedad analítica (tasa de analfabetismo) que se aplique a cada uno de los miembros como propiedad contextual, sin que el significado de ésta última se confunda con la propiedad absoluta original.

Figura 2.6: Generación de una propiedad contextual-analítica



Es claro que saber leer y escribir y pertenecer a una Provincia con más de un 10% de analfabetismo son propiedades perfectamente distinguibles. Obviamente, las propiedades contextuales sólo pueden funcionar como variables cuando se analizan UUAA pertenecientes a varios colectivos: si todas las UUAA fueran miembros del mismo colectivo, se trataría de una propiedad constante en la que todas presentarían el mismo valor. De modo análogo, tampoco las propiedades comparativas tendrán sentido si pertenecen todas al mismo colectivo: de ser el caso la propiedad comparativa queda reducida a la propiedad absoluta o relacional en la que se origina.

Algunos problemas en la aplicación de la tipología

El valor de esta tipología es esencialmente heurístico: rigurosamente considerado, el conjunto de los tipos de propiedades no constituye un sistema de categorías exhaustivas y mutuamente excluyentes.³⁹ En consecuencia, es posible que se susciten dudas al aplicarla en la operacionalización de ciertos enunciados. Por ejemplo, un mismo concepto puede llegar a ser operacionalizado mediante indicadores que remiten a propiedades de diferente tipo. Si, por caso, intentamos operacionalizar el concepto de ‘industrialización’, para poder medir el “grado de industrialización” de un país (o de una región, o de una provincia), podríamos optar por varias alternativas. Así, si tomamos como indicador el “porcentaje de la PEA ocupado en la industria”, tendremos una propiedad analítica. Mientras que si elegimos el “porcentaje del PBI generado por la industria”, se tratará de una propiedad global.⁴⁰

También hay proposiciones que se pueden operacionalizar de modos alternativos: este es el caso cuando los miembros de un conjunto de colectivos son tomados como UUAA en una hipótesis que incluye una propiedad contextual. Si se dice:

³⁸ Para restablecer el significado implícito de la hipótesis se podría agregar al enunciado: “[que los pobladores de las áreas urbanas no marginales]”.

³⁹ Aún cuando es un ejercicio de los más útiles intentar aplicarla como si lo fuera.

⁴⁰ Por más que se trate de un porcentaje, el PBI industrial no se obtiene a partir de información sobre los miembros-habitantes del país.

«Los ingresos de los pobladores de barrios marginales de Resistencia son más bajos [que los ingresos de los pobladores de barrios no marginales de Resistencia]».

Parece evidente que hay aquí:

1. Un conjunto de UUAA:

⇒ individuos “**pobladores de Resistencia**”;

2. Y dos variables:

⇒ 2.1. “*Ingreso mensual*” (que remite a una propiedad ABSOLUTA);

⇒ 2.2. “*Tipo de barrio*” (que alude a una propiedad CONTEXTUAL).

Sin embargo, también podría decirse:

« En los barrios marginales de Resistencia el promedio de ingresos es menor [que en los barrios no marginales]».

Y, en este caso, habría que reconocer la existencia de:

1. Un conjunto de UUAA colectivas:

⇒ “**barrios de Resistencia**”;

2. Y dos variables:

⇒ 2.1. “*Condición de marginalidad*” (una propiedad GLOBAL);

⇒ 2.2. “*Promedio de ingresos*” (una propiedad ANALÍTICA).

Las dos interpretaciones son, por supuesto, igualmente legítimas. El principal mérito de la tipología de Lazarsfeld y Menzel es arrojar luz sobre la multiplicidad de operaciones que requiere la construcción de los datos, además de aclarar los mecanismos obrantes en determinadas falacias.

En nuestra última hipótesis, consideramos una UA colectiva “el barrio”, y definimos a la condición de marginalidad como remitiendo a una propiedad global. Sin embargo, esta caracterización merece ser discutida. En efecto, la condición de “marginal” que se le pueda otorgar a un barrio puede basarse en varias características.⁴¹ Podríamos definir como ‘marginal’ a un barrio habitado por una alta proporción de “marginales”; y se trataría entonces de una propiedad analítica.⁴² Pero también se puede sostener que la marginalidad es una característica global del barrio, con lo que continuaríamos tratando con un enunciado de dos niveles: el barrio-colectivo, y el individuo-miembro. La cuestión es: ¿ En qué se basa esta definición de la “marginalidad” como propiedad global? Ateniéndonos a la etimología de esta noción, a su sentido ecológico inicial, debemos considerar ‘marginales’ los barrios localizados en la periferia de las ciudades. Vale decir barrios que se definen por mantener una determinada relación espacial con respecto a los otros barrios de la ciudad, por una ubicación específica dentro de la estructura urbana. Ahora bien, no hay modo de incluir este tipo de definición dentro de la tipología presentada, a no ser que consideremos que se trata de una propiedad relacional, para lo cual será necesario definir la pertinencia de tres niveles de inclusión en la hipótesis. Nuestras UUAA -los barrios- deberán ahora ser consideradas como miembros integrantes del colectivo ciudad, cuando al mismo tiempo podrán continuar funcionando como colectivos con relación a sus habitantes para poder definir al

⁴¹ De acuerdo a Gino Germani, «En América Latina el término ‘marginalidad’ empezó a usarse principalmente con referencia a características ecológicas urbanas, es decir a los sectores de población segregados en áreas no incorporadas al sistema de servicios urbanos en viviendas improvisadas y sobre terrenos ocupados ilegalmente. De aquí el término se extendió a las condiciones de trabajo y al nivel de vida de este sector de la población» (1973: 12-13).

⁴² La nota anterior debería bastar para eximirnos de defendernos de cualquier acusación de circularidad en esta definición.

“promedio de ingresos” como una propiedad analítica; se trata por lo tanto de un caso en que la UA se encuentra en una posición intermedia, funcionando como colectivo respecto a un nivel inferior y como miembro respecto a un nivel de inclusión superior.

En el trabajo original de Lazarsfeld y Menzel, hay dos ejemplos de proposiciones en tres niveles. Uno de ellos está tomado de *Union Democracy*, el célebre estudio de Lipset, Trown y Coleman sobre los procesos políticos en el Sindicato Internacional de Tipógrafos⁴³ y no presenta mayores dificultades en cuanto a su estructura puramente lineal. El otro ejemplo, en cambio, ofrece mayor interés:

«Los clubes de mujeres que están divididos internamente en camarillas tienen relaciones más difíciles con otros clubes de mujeres, que aquellos que no están divididos de esta manera. Aquí los elementos⁴⁴ (“clubes de mujeres”) primero son categorizados de acuerdo con una variable estructural (división interna en camarillas), y luego se afirma algo acerca de una propiedad relacional (relaciones con los otros clubes) de cada una de las categorías estructurales» (101).

Lo que sorprende a primera vista en este ejemplo es la aparente inversión entre los niveles y las propiedades que se les aplican. Si los clubes son colectivos tendrán una estructura basada en las relaciones entre sus miembros; pero para entablar relaciones con otros clubes, habrá que considerarlos como individuos (puesto que las propiedades relacionales lo son de miembros). La propiedad estructural pertenece al colectivo, pero la propiedad relacional caracteriza también a este mismo colectivo (y no a sus miembros, como podría esperarse). Implícitamente, se requiere suponer la existencia de un tercer nivel,⁴⁵ cuya definición sólo podría producirse en términos lógicos: el conjunto de todos los Clubes de mujeres. Pero, como los autores evidencian una saludable preocupación por la sustantividad de sus reflexiones metodológicas, son renuentes a definir ese tercer nivel como un colectivo.

Ha sido señalada hace tiempo una ambigüedad en el planteo de Lazarsfeld y Menzel. Así, F. Korn escribía: «No queda claro si un colectivo es cualquier conjunto lógico...o es sólo un conjunto tal que posea la propiedad de poder ser tratado como un “actor social”» (1969:15). Para obviar este problema, F. Korn proponía que sólo se consideren como colectivos “los conjuntos que posean por lo menos una propiedad “global”, es decir, que puedan ser definidos por alguna propiedad ajena a las propiedades de sus miembros” (1969:15). Esta sería sin duda una solución, pero que no resolvería la pequeña duda que se genera en el ejemplo del Club de mujeres. Es cierto que cada uno de estos clubes puede funcionar como un “actor” y, por lo tanto, **relacionarse** con otros actores -entre los cuales habrán otros clubes femeninos. Como se ha visto, Lazarsfeld y Menzel plantean como un requisito para predicar propiedades relacionales que se trate de relaciones **sustantivas**. Pero, en cambio, no parece posible que el colectivo “conjunto de todos los clubes de mujeres” posea una estructura en algún grado significativo; parece más bien tratarse de una simple “categoría”.⁴⁶ En el ejemplo se encuentra implícita la solución al problema: es la posibilidad de que existan propiedades relacionales de los miembros, sin que se produzcan correlativamente propiedades estructurales a nivel del colectivo, y por lo tanto de que éste sea definible a partir de criterios exclusivamente lógicos.

Comparando los dos ejemplos, el nuestro y el de Lazarsfeld y Menzel, se advierte que el problema que presentan es exactamente inverso; mientras los barrios forman parte de un colectivo sociológicamente existente (una ciudad) pero no mantienen entre ellos más que relaciones espaciales (y no “sustantivas”), por su parte los Clubes pueden relacionarse entre sí como “actores”

⁴³ Hay tipógrafos (miembros) que votan a favor o en contra de su predisposición, y cuyo voto es comparado con la dirección mayoritaria del voto en su taller (colectivo-miembro), y en el local sindical (colectivo) al que pertenece el taller. Se tiene entonces una propiedad absoluta del miembro, y dos propiedades contextuales correspondientes a los colectivos de menor y mayor nivel. Se determinó así que « mayor cantidad de gente votaba en forma contraria a la propia predisposición cuando ésta estaba en conflicto con la mayoría de su taller (pero no de su local) que cuando estaba en conflicto con la mayoría de su local (pero no de su taller).» (Lazarsfeld y Menzel, 1969: 101).

⁴⁴ Es decir, las UUAA, en la terminología que utilizamos aquí.

⁴⁵ No tan “implícitamente”, ya que inmediatamente antes, se nos ha dicho que se van a abordar ahora “ejemplos de proposiciones de tres niveles” (1969: 100).

⁴⁶ Al menos no se deduce del ejemplo otra posibilidad: no se menciona ni siquiera una hipotética -aunque por cierto muy plausible- Federación de Clubes de mujeres.

sin por ello dar lugar a un colectivo cuya existencia sea algo más que lógica. Hay una manera simple de llevar remedio a esta confusión entre criterios sustantivos y lógicos, y es producir una distinción estricta entre ambos. Así, desde el punto de vista de la tipología de propiedades individuales y colectivas, debería bastar con atender a criterios puramente lógicos, sin preocuparse ni por la “existencia sociológica” del colectivo ni por el carácter “sustantivo” de las relaciones entre los miembros. Una cosa es la existencia de una posibilidad lógica; otra muy distinta es que resulte pertinente hacer uso de esa posibilidad, lo que debería resolverse exclusivamente sobre una base pragmática.

CAPÍTULO 4

TÉCNICAS ELEMENTALES DE ANÁLISIS

En la secuencia típica del proceso de investigación, el análisis de datos se inscribe dentro de las últimas etapas. Respondiendo a un esquema previamente establecido, las observaciones han sido realizadas, luego codificadas y tabuladas. El resultado es una serie de cuadros estadísticos a los que habrá que “leer” en función de los objetivos de la investigación, tratando de destacar lo esencial de la información en ellos contenida.

En este capítulo expondremos algunas de las técnicas más elementales que son de uso frecuente en la investigación social. Comenzaremos concentrándonos en la tabla de contingencia, puesto que éste es el modo en que tradicionalmente se presentan muchos datos en las investigaciones sociales. Acordaremos una atención privilegiada a la tabla de 2×2 , en la que se plantean al nivel más simple los problemas lógicos del análisis. En este contexto, habré de referirme al uso de los porcentajes, al test de χ^2 y a algunos de los coeficientes de asociación más simples. Luego abordaremos brevemente otras formas de presentación de los datos como las distribuciones multivariantes conjuntas, lo que permitirá introducir algunas nociones sobre las relaciones entre variables pertenecientes a niveles más elevados de medición.

1. LA TABLA DE CONTINGENCIA Y EL USO DE LOS PORCENTAJES

Una tabla de contingencia es el resultado del **cruce** (o tabulación simultánea) de dos o más variables. Nos ocuparemos solamente de tablas bivariadas (o ‘bivariantes’), que también reciben los nombres de ‘clasificación cruzada’ o ‘tabulación cruzada’. Esta forma de presentación de los datos es muy típica de la investigación en ciencias sociales, que se caracteriza por un uso predominante de variables (o atributos) definidas en los niveles de medición nominal y ordinal.¹ La tabla de contingencia consiste en un cierto número de celdas en las que, como resultado de un proceso de tabulación, realizado en forma manual, mecánica o electrónica,² se han volcado las frecuencias (número de casos) correspondientes a cada combinación de valores de varias variables.

Forma lógica de la tabla de 2×2

Para analizar la forma lógica de este tipo de tablas, consideraremos la estructura más sencilla, la llamada tabla de “ 2×2 ”, o sea de dos valores por dos valores, que resulta del cruce de dos dicotomías, los atributos ‘X’ e ‘Y’ en los que hemos clasificado un conjunto de unidades de análisis

¹ Por cierto, la tabla de contingencia puede también utilizarse para volcar datos provenientes de mediciones realizadas en el nivel intervalar, pero a costa de una gran pérdida de información; como veremos, existen otras técnicas mucho más precisas, matemáticamente hablando, para el análisis de tales variables.

² En la era actual de difusión masiva de las computadoras personales, es francamente desaconsejable recurrir a modos manuales de tabulación o bien a antigüedades tales como las tarjetas tipo McBee, las máquinas clasificadoras de tarjetas tipo Hollerit, etc.

Tabla 4.1: Forma lógica de la tabla de 2 x 2

		Atributo X		Total
		No	Sí	
Atributo Y	Sí	-X Y	X Y	Y
	No	-X -Y	X -Y	-Y
Total		-X	X	n

‘n’ representa el *total* de unidades de análisis incluidas en la muestra, lo que se suele denominar ‘la frecuencia de *orden cero*’. Por su parte, ‘-X’, ‘X’, ‘Y’ y ‘-Y’ son las *frecuencias marginales* o de *primer orden*; así, por ejemplo, ‘-X’ representa el número total de casos que no presenta el atributo X, independientemente de que posean o no el atributo Y.

Por último, ‘-X Y’, ‘X Y’, ‘-X -Y’ y ‘X -Y’ representan las *frecuencias condicionales*, o de *segundo orden*; de este modo, ‘-XY’ significa el número absoluto de observaciones que combinan la ausencia del atributo X con la presencia de Y. Es importante notar que:

$$\begin{aligned} n &= (X) + (-X) \\ &= (Y) + (-Y) \\ &= (-X Y) + (X Y) + (-X -Y) + (X -Y) \end{aligned}$$

Aplicación del modelo a un ejemplo

Este modelo puede aplicarse para cualquier población y todo tipo de variables:

Tabla 4.2: Misiones, 1980 - Número de habitantes según tipo de asentamiento y pertenencia a hogares con NBI

Tipo de asentamiento	Hogares con NBI		Total
	No	Sí	
Urbano	194.397	96.610	291.007
Rural	122.701	166.814	289.515
Total	317.098	263.424	580.522

Fuente: elaboración propia (datos de Argentina, 1984:343).

Así, en esta tabla, las unidades de análisis son personas, y los atributos ‘X’ e ‘Y’ se traducen, respectivamente, en el hecho de pertenecer o no a un hogar con Necesidades Básicas Insatisfechas (NBI), y de residir en una zona urbana o rural (es decir, ‘no-urbana’). El título ya nos informa que se trata de la *población* de Misiones en 1980; el n corresponde por tanto a 580.522 *personas*. Se verifica efectivamente que:

$$\begin{aligned} 580.522 &= 317.098 + 263.424 \\ &= 291.007 + 289.515 \\ &= 194.397 + 96.610 + 122.701 + 166.814 \end{aligned}$$

También se cumple que: $291.007 = 194.397 + 96.610$, es decir que cada marginal es igual a la suma de las frecuencias condicionales en la hilera -o la columna - correspondiente. *El primer paso de cualquier análisis es verificar si la tabla “cierra”, vale decir, si se cumplen las relaciones aritméticas que debe satisfacer cada cifra; en caso contrario es evidente que se ha producido algún*

error en la tabulación.

¿ Qué puede afirmarse en base a esta Tabla 4.2? Puede decirse que en Misiones, en el año 1980, había 580.522 personas; proposición que, a pesar de su veracidad, tiene el inconveniente de no hacer uso de **toda** la información contenida en dicha tabla.

Un modo de comenzar el análisis es describiendo los marginales. Así, se puede decir que de estos 580.522 habitantes de Misiones, 291.007 residían en zonas urbanas en tanto que 289.515 lo hacían en áreas rurales. Esto ya es más interesante, aunque la misma conclusión podría haberse derivado de una distribución de frecuencias simple:

¿ Para qué sirven los porcentajes?

Vemos que hay más habitantes urbanos que rurales; exactamente, los primeros superan a los segundos en 1.492 personas. ¿Ahora bien, es ésta una diferencia importante? Depende del contexto dentro del cual ubiquemos ese número. Lo sería sin duda si comparáramos la cifra con los datos del Censo de Población de 1970, en el que los rurales aventajaban a los urbanos en 132.886 habitantes.

Tabla 4.2a: Misiones 1980, Habitantes según tipo de asentamiento

Tipo de asentamiento	Número de habitantes	%
Urbano	291.007	50,1
Rural	289.515	49,9
Total	580.522	100,1

Fuente: Tabla 4.2

Pero, considerando intrínsecamente los datos de la Tabla 4.2a, la manera de apreciar la importancia de esas 1.492 personas de diferencia es poniéndolas en relación con el total de la población provincial. Es decir, considerar el **peso relativo** de cada grupo sobre el total de población. Se observa así que la diferencia entre la población urbana y la rural es muy escasa: 50,1 a 49,9%. Hemos calculado el porcentaje de población urbana mediante la siguiente operación:

$$\frac{292.007}{580522} \times 100 \quad \text{o, en general } Y/n \times 100$$

¿Cuándo redondear los porcentajes?

En realidad, el resultado de la operación aritmética anterior arroja la cifra de 50,12850503, la que nosotros hemos redondeado a un decimal anotando 50,1.³ ¿ Por qué este redondeo? El interés de los porcentajes es indicar con la mayor claridad las dimensiones relativas de dos o más números, transformando a uno de esos números, la *base*, en la cifra 100. Es indudable que:

$$291.007/580.522 = 50/100 = 50\%$$

Matemáticamente, estas son expresiones equivalentes —o casi— pero es evidente que, en un sentido psicológico, ‘50%’ es la manera más concisa, sencilla y ventajosa de denotar la relación que nos interesa. Si se conservan muchos decimales, sólo se logra tornar más engorrosa la lectura de la tabla y se pierde la ventaja de expresar las cifras en porcentajes. Por ende se puede recomendar, siempre que sea ello posible, *como regla general, prescindir totalmente de los*

³ El neologismo ‘redondear’ significa suprimir los decimales -números a la derecha de la coma-, o conservar una limitada cantidad de éstos. El redondeo se realiza observando el decimal siguiente al que se quiere conservar; en el ejemplo, el segundo decimal es un 2 -cifra comprendida entre 0 y 4- por lo que corresponde anotar ‘50,1’, mientras que, de tratarse de un número igual o superior a 5, se anotaría ‘50,2’.

decimales. ‘50,12850503’ parece más preciso que ‘50 %’; más es ésta una precisión engañosa,⁴ y que a todos los efectos prácticos o teóricos carece absolutamente de significado.⁵

Sin embargo, en la Tabla 4.2a hemos consignado ‘50,1 %’ y no ‘50 %’. ¿Por qué ya esta primera infracción a la regla que acabamos de formular? En el caso que nos ocupa, existirían al menos dos posibles justificaciones: a) trabajando con una población de 580.522 personas, cada punto del primer decimal representa 580 individuos, una cantidad relevante para muchos propósitos; y b) porque si no conserváramos el primer decimal, obtendríamos el mismo porcentaje para ambos sectores de la población, y tal vez no nos interese producir este efecto.⁶

El otro marginal de la Tabla 4.2 podría dar lugar a un análisis en un todo análogo. Se concluiría así que un 45,4 % —o un 45 %— de la población total de Misiones vivía en 1980 en hogares con NBI.

¿Cómo se lee una tabla de contingencia?

Siempre que se considera una tabla de contingencia es recomendable comenzar el análisis por las distribuciones univariadas de los marginales, para luego pasar al examen de las frecuencias condicionales, que nos permitirá aprehender el sentido peculiar de cada cruce de variables.

En un paso ulterior podríamos entonces hacer una lectura de cada una de las cifras contenidas en las celdas de la Tabla 4.2:

- ① 194.397 personas vivían en hogares sin NBI en áreas urbanas;
- ② 96.610 lo hacían en hogares con NBI en áreas urbanas;
- ③ 122.701 pertenecían a hogares sin NBI en áreas rurales;
- ④ 166.814 habitaban en áreas rurales en hogares con NBI.

Todas estas proposiciones son *verdaderas*, en el sentido de que traducen con exactitud el significado de cada cifra; pero consideradas en conjunto constituyen una lectura puramente *redundante* de la información contenida en la Tabla 4.2, y no agregan nada a lo que ésta ya está mostrando por sí misma. En general, cuando se analizan tabulaciones bi-variadas, el interés debe focalizarse en determinar si existe alguna *relación* entre las dos variables. En otros términos, partimos siempre de una *hipótesis*, más o menos explícita, acerca de la existencia o no de una relación entre las dos variables.

Modos alternativos de análisis

Hay básicamente dos modos de abordar el análisis de una tabla. De acuerdo a una distinción establecida por Zelditch (1959), se la puede analizar de manera asimétrica o simétrica. En el modo asimétrico, el interés está puesto en observar el efecto de una de las variables sobre la otra. Por lo contrario, en el análisis simétrico no se presupone que una variable funja como “**causa**” de la otra. Abordaremos sucesivamente estas dos alternativas, teniendo siempre presente que una tabla no es intrínsecamente simétrica o asimétrica, sino que esta distinción se limita al modo en que se decide encarar el análisis en función de los objetivos del investigador.

⁴ Dada la imprecisión de nuestros instrumentos de medición.

⁵ Según Galtung, « La presentación de los porcentajes con 1 o incluso con 2 decimales no tiene sentido a menos que 1) la **calidad** de la recolección sea tan buena que tenga sentido decir que el 70,1% y no el 70,2% dicen ‘sí’, etc.; 2) el **propósito** de la recolección de datos sea tal, que sea diferente para la interpretación que el 70,1% y no el 70,2% diga ‘sí’, etc. En general sugerimos que los porcentajes deben presentarse sin ningún decimal, para evitar una impresión de exactitud que es a menudo completamente espuria» (1966: II, 231). Ya Bachelard decía que « El exceso de precisión, en el reino de la cantidad, se corresponde muy exactamente con el exceso de pintoresco, en el reino de la cualidad», y veía en tales excesos las marcas de un espíritu no científico (1972: 212).

⁶ En una visión *diacrónica*, ‘50,1%’ podría tener el valor de significar la inversión de una tendencia: de un predominio histórico de la población rural, se pasa a una preponderancia de los habitantes urbanos. En cambio, desde un punto de vista *sincrónico*, podría convenir escribir ‘50%’, destacando la semejanza cuantitativa entre ambos sectores.

¿Qué es analizar una tabla asimétricamente?

El caso asimétrico: se plantea siempre que se elige considerar que una variable -la variable *independiente*-incide sobre la distribución de la otra -la variable *dependiente*. Hay tablas en las que cualquiera de las dos variables puede fungir como “causa” de la otra. Aunque también suele ocurrir que se “imponga”, por así decirlo, el análisis en una determinada dirección.

La “regla de causa y efecto” o “primera regla de Zeisel” se aplica, en palabras de su autor, «siempre que uno de los dos factores del cuadro dimensional pueda considerarse como causa de la distribución del otro factor. La regla es que *los porcentajes deben computarse en el sentido del factor causal*» (Zeisel, 1962: 37).

¿ Puede esta regla aplicarse a nuestra Tabla 4.2? Responder positivamente a esta pregunta supondrá considerar, por ejemplo, que el tipo de asentamiento de la población “determina” o “condiciona” una probabilidad diferencial de pertenecer a un hogar con NBI. Esta hipótesis es plausible, si se tiene en cuenta que, por lo general, en nuestros países subdesarrollados el nivel de vida de las poblaciones rurales es inferior al de los habitantes urbanos.

¿ Cómo se lee el título de una tabla?

El título ya es merecedor de algunas observaciones, en tanto ejemplifica un cierto código cuyas reglas debemos conocer si queremos comprender acabadamente el significado de la Tabla 4.2.1:

- ① Las dos variables se encuentran claramente identificadas; se trata, respectivamente, de ‘Pertenencia de la población a hogares con Necesidades Básicas Insatisfechas’ (que en el encabezamiento de las columnas figura como ‘Hogares con NBI’, y cuyos valores son ‘Sí’ y ‘No’) y de ‘Tipo de asentamiento’ (con los valores ‘Urbano’ y ‘Rural’).

Tabla 4.2.1: Misiones, 1980 - Pertenencia de la población a hogares con NBI según tipo de asentamiento (%)

<i>Tipo de asentamiento</i>	<i>Hogares con NBI</i>		<i>Total (= 100%)</i>
	<i>No</i>	<i>Sí</i>	
<i>Urbano</i>	66,8	33,2	291.007
<i>Rural</i>	42,4	57,6	289.515
<i>Total</i>	54,6	45,4	580.522

Fuente: Tabla 4.2

- ② Entre los nombres de las dos variables se intercala la preposición ‘según’; no hubiera sido incorrecto utilizar alguna otra preposición como ‘por’ o ‘de acuerdo’, pero cabe atender al orden en que se introducen los nombres de las variables. Tomando el ‘tipo de asentamiento’ como variable independiente, ésta es introducida a continuación de la proposición ‘según’; en efecto, la presentación de los datos en la Tabla 4.2.1 apunta a destacar esta idea: **según** sea su tipo de asentamiento tenderán las personas a diferir en cuanto al valor mantenido en la variable dependiente.

- ③ El título finaliza con la expresión ‘(%)’; ello nos indica que las cifras consignadas en las celdas son porcentajes, y no frecuencias absolutas.⁷
- ④ En el encabezamiento de la última columna aparece la expresión ‘**Total (100%)**’. Esto quisiera expresar: a) que en dicha columna las cifras **no** son porcentajes sino frecuencias absolutas; y b) que las cifras absolutas de la columna fueron tomadas como base para calcular los porcentajes de las celdas.⁸

¿Cómo se lee una cifra porcentual?

Procedamos ahora a la lectura de la Tabla 4.2.1. Habiendo tomado ‘Tipo de asentamiento’ como variable independiente, hemos en consecuencia calculado los porcentajes “en el sentido de esta variable, nuestro “factor causal”. Ello quiere decir que *las bases para el cálculo porcentual están dadas por el total de casos para cada valor de la variable independiente*.

En la celda superior izquierda de la tabla observamos ‘66,8’, y sabemos -por el título - que la cifra corresponde a un porcentaje. La lectura correcta de esta cifra tiene lugar en dos pasos, cada uno de los cuáles supone responder a una pregunta.

① Lo primero que debemos inquirir es: “¿ 66,8% **de qué?** (o ¿de quiénes?)”. La única respuesta correcta es: “del 100% constituido por los 291.007 habitantes urbanos”; es decir, buscamos primero en la tabla dónde está el 100% —en la primera hilera—, y dirigimos luego nuestra vista hacia el encabezamiento de dicha hilera leyendo: ‘Urbano’. Cumplimentado este primer paso, estaremos en condiciones de preguntarnos con éxito...

② ...“ ¿ **Qué sucede** con este 66,8%?”, y podremos responder: “viven en hogares sin NBI”. A esta segunda pregunta respondimos simplemente dirigiendo nuestra atención hacia el encabezamiento de la columna: ‘No’.

Así, el significado de la primera celda puede expresarse: «De todos los habitantes urbanos de Misiones, hay un 66,8% que pertenece a hogares sin NBI».

Igualmente correcto sería escribir:

« Un 66,8% de la población urbana vive en hogares sin NBI».

Es obvio, que existe una posibilidad cierta de optar por diferentes redacciones; pero lo fundamental es que la expresión literaria respete el significado de la cifra. Se puede pensar en dos grandes tipos de problemas que se plantean en la lectura de los porcentajes.

❶ Hablaremos de problemas “*lógicos*” cuando se produce una *falsa* lectura de la cifra porcentual. Estos errores devienen de una *confusión acerca de la base* sobre la cual está calculado el porcentaje. En cualquier tabla de doble entrada, existen potencialmente tres bases sobre las cuales es posible calcular los porcentajes, a saber,

— El total de la hilera: ‘291.007’, en este caso;

⁷ Igualmente claro sería omitir el signo de porcentaje en el título y consignarlo a continuación de cada cifra: 66,8%, 33,2%, etc.

⁸ Esta última convención, además de ser tan arbitraria como las anteriores, está lejos de ser universalmente reconocida. Otra manera habitual de proceder es hacer figurar en la última columna para todas las hileras la expresión ‘100,0’; pero aparte de ser ésta una información redundante (en el primer renglón es obvio que $33,2 + 66,8 = 100,0$), este procedimiento tiene el inconveniente de que hace desaparecer toda referencia a las frecuencias absolutas que fueron tomadas como base; en cambio, mientras éstas continúen apareciendo siempre será posible reconstruir las frecuencias absolutas correspondientes a las celdas: por ejemplo, $291.000 \times 0,332 = 96.614 \approx 96.610$ (la pequeña diferencia deviene del redondeo del porcentaje). Otra posibilidad es anotar entre paréntesis las bases de los porcentajes: ‘100,0 (291.007)’. También es posible duplicar cada cifra del cuadro consignando siempre las frecuencias absolutas y relativas -estas últimas de preferencia con algún recurso tipográfico distinto-, aunque esta práctica tiende a restarle nitidez a los datos.

- El total de la columna, '317.098'; y
- El "total total", el 'n': '580.522'.

Se comete un *error lógico* cuando un porcentaje es leído sobre una base que no fue la utilizada para calcularlo. Así, si se lee « Un 66,8% de los habitantes de Misiones son urbanos y viven en hogares sin NBI», la expresión lingüística da a entender que el porcentaje fue calculado sobre el total de la población provincial, con lo cual *el enunciado pasa a expresar una proposición falsa* (el porcentaje que correspondería a dicha expresión lingüística no sería '66,8' sino '33,5').

Igualmente erróneo sería escribir «En Misiones, un 66,8 % de las personas pertenecientes a hogares sin NBI residen en asentamientos urbanos». La construcción de esta frase supone que el 66,8% fue calculado sobre el total de personas pertenecientes a hogares sin NBI, con lo que el enunciado es también falso (para esta redacción, el porcentaje correcto sería '61,3'). Por ende, hay una sola manera de generar enunciados verdaderos; es eligiendo una construcción lingüística que dé cuenta sin ambigüedad alguna del modo en que la cifra porcentual ha sido efectivamente calculada.⁹

☉ Pero también se presentan problemas **pragmáticos**. Sucede que diferentes redacciones son susceptibles de comunicar distintos significados. Comparemos los siguientes enunciados:

a.-«**Más de** dos tercios de los habitantes urbanos viven en hogares que no presentan NBI »;

b.-« **Solamente** un 66,8% de los habitantes urbanos pertenece a hogares sin NBI ».

Tanto 'a' como 'b' expresan correctamente el porcentaje, desde una perspectiva puramente lógica. Sin embargo, es evidente que ambos enunciados no tienen el mismo significado: ciertamente 'a' trasunta una visión de la situación más optimista que 'b'. Sucede que, como lo explicara hace tiempo el lingüista Roman Jakobson, no es posible denotar sin connotar: las operaciones de selección y combinación que están necesariamente en obra en la producción de todo discurso introducen en él una dimensión ideológica.¹⁰ Podemos probar de eliminar los adverbios en nuestros enunciados 'a' y 'b', con lo que obtenemos expresiones cuyo valor lingüístico es muy similar:

a1.-« Dos tercios de los habitantes urbanos viven en hogares que no presentan NBI »;

b1.-« Un 66,8% de los habitantes urbanos pertenece a hogares sin NBI ».

Aparentemente habríamos eliminado así toda valoración, permaneciendo sólo la fría cifra. Pero esto es creer que el significado de un enunciado individual sólo depende de su contenido intrínseco. Lo cierto es que este enunciado se inserta en un contexto más amplio, el discurso al que pertenece, cuyo significado global concurre a producir, pero que a la vez determina grandemente su propia significación. Lo expresado abona la idea de que estos problemas que hacen a la pragmática del discurso son inevitables. A lo sumo puede intentarse limitarlos controlando en alguna medida la adverbialización y la adjetivación.

⁹ Puesto que siempre existen tres alternativas para el cálculo de los porcentajes, es un hecho tan lamentable cuanto inevitable que para leer un porcentaje siempre existan dos posibilidades de equivocarse y sólo una de acertar...

¹⁰ Verón caracterizó a la ideología « como un **nivel de significación** de todo discurso transmitido en situaciones sociales concretas, referido al hecho inevitable de que, por su propia naturaleza, todo mensaje transmitido en la comunicación social posee una dimensión connotativa» (Verón, 1972: 309).

¿Cómo se lee un conjunto de porcentajes?

Podemos ahora leer el conjunto de las cifras de la Tabla 4.2.1, que traducimos en la siguiente serie de enunciados:

1. - Un 66,8% de los habitantes urbanos pertenece a hogares sin NBI;
2. - Un 33,2% de los habitantes urbanos pertenece a hogares con NBI;
3. - Un 42,4% de los habitantes rurales pertenece a hogares sin NBI;
4. - Un 57,6% de los habitantes rurales pertenece a hogares con NBI;
5. - Un 54,6% de todos los habitantes pertenece a hogares sin NBI; y
6. - Un 45,4% de todos los habitantes pertenece a hogares con NBI.

Todos estos enunciados son verdaderos. Empero, su mera enumeración no constituye una “buena” lectura de la Tabla 4.2.1. En efecto, este conjunto de enunciados a) es en gran medida redundante; y, sobre todo, b) no apunta a destacar lo fundamental, esto es, la relación entre las variables que postula nuestra hipótesis y que es la única razón por la que los datos han sido presentados como se lo ha hecho, calculando los porcentajes en una dirección determinada.

En cuanto a la redundancia, debe resultar claro que el contenido del enunciado 2 ya está incluido —implícitamente— en el enunciado 1: si un 66,8% de los habitantes urbanos pertenece a hogares sin NBI, y nos encontramos tratando con una variable dicotómica, ello implica que *necesariamente* hay un 33,2% de los habitantes urbanos en hogares con NBI.¹¹ Y viceversa: si es verdadero el enunciado 2, necesariamente lo será también el 1. Es evidente que la misma relación se da para los pares de enunciados 3 y 4, y 5 y 6.

Aunque ello no resulte tan obvio, también son redundantes en cierto modo los porcentajes correspondientes a la hilera del total. Así, el que 45,4% de todos los habitantes pertenezcan a hogares con NBI no es más que el resultado de un promedio ponderado entre el 33,2% de urbanos y el 57,6% de rurales que presentan esta característica. Es ésta una propiedad interesante de los porcentajes marginales: necesariamente su valor se ubicará dentro de un rango limitado por los valores porcentuales consignados en las celdas correspondientes; en este caso, el porcentual del marginal deberá ser superior a 33,2 e inferior a 57,6; el que se encuentre “más cerca” de una u otra de estas cifras dependerá sólo del peso relativo de ambos grupos (el ‘rural’ y el ‘urbano’) sobre la población total.¹² Es por esta razón que frecuentemente se omite la presentación de los porcentajes marginales.¹³

En suma, si intentamos reducir al mínimo la redundancia en la lectura de la tabla, podemos considerar que lo esencial de la información está contenido en los enunciados 2 y 4 (o, indiferentemente, en los 1 y 3). De este modo, destacaremos el sentido fundamental que queremos prestarle a los datos: en estas dos cifras -33,2% y 57,6%-¹⁴ está resumido lo que la tabla significa para nosotros. Comparando estos dos porcentajes, nuestra lectura pone en evidencia la relación estocástica entre las dos variables postulada por nuestra hipótesis:

«Mientras que en la población urbana hay un 33,2% de habitantes en hogares con NBI, entre los pobladores rurales este porcentaje asciende al 57,6%».

Se corrobora por lo tanto la existencia de una probabilidad diferencial de pertenecer a un hogar con NBI en función del tipo de asentamiento de la población.

Una alternativa interesante para presentar esta información puede ser mediante un gráfico de columnas. Se ve claramente cómo ambas poblaciones son de tamaños similares y cómo la

¹¹ 33,2 es el complemento necesario para alcanzar al 100,0%; en efecto, $100,0 - 66,8 = 33,2$.

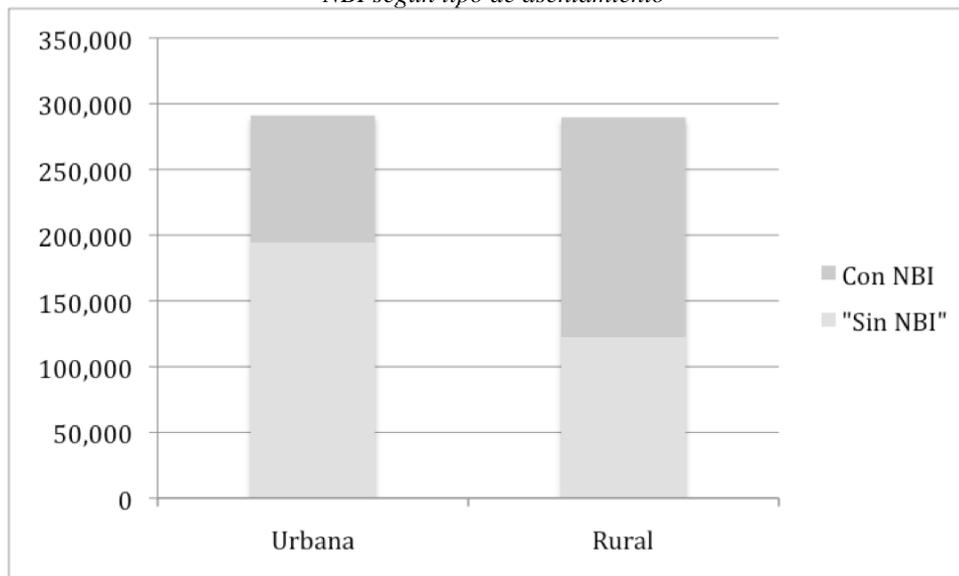
¹² En el caso particular de la tabla que nos ocupa, ambos grupos tienen aproximadamente el mismo peso, por lo que podría calcularse: $(33,2 + 57,6) / 2 = 45,4\%$.

¹³ Sin embargo, cuando se trabaja con tablas de mayores dimensiones -y no basadas en dicotomías-, el porcentaje marginal puede funcionar como un punto de referencia útil que facilita la atribución de un significado a los porcentajes en las celdas. Incluso en nuestra misma Tabla 4.1.1 podría decirse que, frente a un 45,4% del total de la población que pertenece a hogares con NBI, el 33,2% de los ‘urbanos’ es **comparativamente bajo**.

¹⁴ O, alternativamente, en el par: 66,8% y 42,4%.

proporción de personas pertenecientes a hogares con NBI es mucho mayor en el campo:¹⁵

Figura 4.1: Misiones, 1980 - Pertenencia de la población a hogares con NBI según tipo de asentamiento



Fuente: Tabla 4.2.1

Primera regla de Zeisel

Nos encontramos ahora en condiciones de completar nuestra formulación de la primera regla de Zeisel, referida al caso del análisis asimétrico:

« LOS PORCENTAJES SE CALCULAN EN EL SENTIDO DE LA VARIABLE INDEPENDIENTE, Y SE COMPARAN EN EL SENTIDO DE LA VARIABLE DEPENDIENTE.»¹⁶

En efecto, esto es todo lo que hemos hecho: hemos computado los porcentajes en el sentido horizontal (en este caso), y los hemos leído en el sentido vertical.¹⁷ Lo fundamental es calcular los porcentajes en la dirección adecuada, esto es, en el sentido de la variable a la que asignaremos el rol de 'independiente' en nuestro análisis.

¿Cuál es la variable independiente?

Ahora bien, cuál sea la variable independiente, es una cuestión que está enteramente supeditada a los objetivos de nuestro análisis. Así como en la Tabla 4.2.1 se eligió el 'Tipo de asentamiento', así puede tomarse también como independiente la variable 'Pertenencia a hogares con NBI':

¿ En qué sentido puede pensarse que el hecho de pertenecer o no a un hogar con NBI "determine" el tipo de asentamiento de las personas? ¿ Puede sostenerse una brumosa hipótesis según la cual los "pobres" preferirían residir en áreas rurales?.

¹⁵ Los gráficos tienen un gran poder de comunicación y son un recurso al que se puede apelar en la presentación de informes de investigación. Es posible que muchas personas perciban mejor una relación expresada visualmente, aunque se pierda algo de precisión con respecto a los datos numéricos.

¹⁶ Esta fórmula pertenece a Zelditch (1959). Galtung dice: «sacar siempre los porcentajes perpendicularmente a la dirección de la comparación» (1966: II, 233).

¹⁷ Algunos autores sostienen la conveniencia de presentar siempre la variable independiente en el encabezamiento del cuadro y la variable dependiente en las hileras, con lo que los porcentajes se computarían siempre en el sentido vertical. Hay un único fundamento razonable para esta práctica, y es el mantener la analogía con el tratamiento de variables cuantitativas graficadas en un diagrama de ejes cartesianos, en el que la x -variable independiente- aparece siempre en la abscisa.

Tabla 4.2.2: Misiones, 1980 - Distribución de la población por tipo de asentamiento y pertenencia a hogares con NBI

Tipo de asentamiento	Hogares con NBI		Total
	No	Sí	
Rural	61	37	50
Total	(317.098)	(263.424)	(580.522)

Fuente: Tabla 4.2

La Tabla 4.2.2 nos permite pensar la diferencia que media entre las expresiones ‘variable independiente’ y ‘causa’. Es evidente que en este caso no tiene demasiado sentido pensar en la pobreza como “causa” del tipo de asentamiento.¹⁸ Pero es perfectamente posible leer:

« En las zonas rurales de la Provincia se concentra el 63% de las personas pertenecientes a hogares con NBI, frente a sólo un 39% de las que pertenecen a hogares no carenciados ».

Se trata de una presentación de los datos de la Tabla 4.2 que tiende a destacar cómo la pobreza se concentra mayoritariamente en las áreas rurales de Misiones. No solamente la Tabla 4.2.2. es tan “verdadera” como la 4.2.1, sino que ambas son igualmente válidas. Aún cuando la Tabla 4.2.1 indujera en nosotros un mayor sentimiento de satisfacción, esta segunda interpretación no sería menos legítima por ello.¹⁹

Segunda regla de Zeisel

Existe sin embargo una limitación al sentido en que es lícito computar los porcentajes, cuando se trabaja con datos muestrales. No siempre las muestras tienen la virtud de ser autoponderadas. Por diversas razones, puede ocurrir que una muestra no sea representativa de la población en algún sentido; hemos visto en el capítulo anterior que el caso es frecuente al utilizar diseños de muestra estratificados o por cuotas.

Imaginemos que queremos investigar acerca de la conformidad de un grupo de estudiantes de las carreras de Trabajo Social y de Turismo con el sistema de promoción por examen final; a tales efectos, seleccionamos una muestra de 40 alumnos de cada carrera, a sabiendas de que los totales de alumnos eran de 160 para Trabajo Social y de 80 para Turismo.

En casos semejantes sólo cabe calcular los porcentajes en el sentido en que se lo ha hecho en el ejemplo de la Tabla 4.3, y se podrá concluir que la proporción de disconformes con el sistema de aprobación por examen final es más elevada entre los alumnos de Turismo (82%) que entre los de Trabajo Social (60%).

¹⁸ Estrictamente, para poder hablar de una relación causal entre las variables X e Y se requiere contar con evidencia de tres tipos: a) variación concomitante de X e Y; b) precedencia temporal de X con respecto a Y; y c) eliminación de otros posibles determinantes de Y (cf. Selltitz *et al.*, 1968: 100 y ss.).

¹⁹ En los estudios de mercado es usual distinguir entre dos tipos de porcentajes, según la dirección en la que han sido calculados. Así, el porcentaje ‘de penetración’, también denominado ‘cuota de mercado’ se calcula sobre el total de integrantes de una categoría –de edad, sexo, nivel educativo, etc.– e indica cuántos de este total consumen el producto (o manifiestan su intención de votar por un candidato, si se trata de *marketing* político); en cambio, el porcentaje ‘de composición’ indica sobre el total de consumidores del producto (o de votantes del candidato), qué proporción corresponde a una categoría en particular (cf. Antoine; 1993: 33 y ss.). Por analogía, mientras que la Tabla 4.2.1 estaría indicando una mayor penetración de la pobreza en áreas rurales (57,6% de los rurales son pobres), la tabla 4.2.2 mostraría el peso mayoritario de los habitantes rurales en la composición de la población con NBI (63% de los pobres son rurales).

Tabla 4.3: Conformidad con el sistema de examen final según carrera

Conformidad con el examen final	Carrera		Total
	Turismo	Trabajo Social	
Sí	7 18%	16 40%	23 29%
No	33 82%	24 60%	57 71%
Total	40 100%	40 100%	80 100%

Fuente: elaboración propia.

En general, la segunda regla de Zeisel -que no es más que una limitación a la primera - afirma:

« CUANDO UN CONJUNTO DE MARGINALES NO ES REPRESENTATIVO DE LA POBLACION, LOS PORCENTAJES DEBEN COMPUTARSE EN LA DIRECCION EN QUE LA MUESTRA NO ES REPRESENTATIVA ».²⁰

En efecto, en nuestra muestra la relación entre los alumnos de las dos carreras es de 1:1 (40 en cada una), en tanto sabemos que en la población la relación real es de 1:2 (hay el doble de alumnos en Trabajo Social). Como nuestra muestra no es representativa por carrera, los porcentajes sólo pueden calcularse en esa dirección: sobre el total de alumnos de cada carrera.

¿ Qué ocurriría si calculáramos directamente los porcentajes en el sentido horizontal? Concluiríamos -erróneamente - que del total de los estudiantes que se manifiestan conformes con el sistema de examen final hay un 70% que pertenece a la carrera de Trabajo Social:

Tabla 4.3.1: Carrera según conformidad con el sistema de examen final (%)

Conformidad con el examen final	Carrera		Total
	Turismo	Trabajo Social	
Sí	30	70	100
No	58	42	100
			(n = 80)

Fuente: Tabla 4.3

Es verdad que *en la muestra* se da este 70%; pero ello ocurre debido a un factor arbitrario que es el tamaño relativo de la muestra en ambas carreras. Como en la muestra la carrera de Trabajo Social se encuentra subrepresentada con relación a su peso real en la población, y sus estudiantes son más conformistas que los de Turismo, en la población deberá ser mayor el porcentaje de conformes concentrados en aquella carrera.

Supongamos que de haber trabajado con el universo, se hubiera obtenido las mismas proporciones de conformistas en ambas carreras que las registradas en la Tabla 4.3. Los resultados serían los presentados en la Tabla 4.3.2.²¹

²⁰ La expresión de la regla pertenece a Zelditch (1959).

²¹ Para construir la Tabla 4.3.2, simplemente multiplicamos por 2 las frecuencias absolutas correspondientes a los estudiantes de Turismo, y por 4 las de Trabajo Social.

Tabla 4.3.2: Carrera según conformidad con el sistema de examen final

Conformidad con el examen final	Carrera		Total
	Turismo	Trabajo Social	
Sí	14 18%	64 82%	78 100%
No	66 41%	96 59%	162 100%
Total	80 33%	160 67%	240 100%

Fuente: elaboración propia.

Se observa que hay en realidad un 82% de los conformistas que pertenecen a Trabajo Social y que, por lo tanto, el 70% que arrojaba la Tabla 4.3.1 no podía ser tomado como una estimación válida de la proporción existente en la población. Al no ser representativa la muestra en cuanto al peso relativo de ambas carreras, el cómputo *directo* de los porcentajes sólo se puede realizar como se lo hizo en la Tabla 4.3.

Si se desea calcular los porcentajes en la otra dirección, no se lo puede hacer directamente, sino que es indispensable recurrir a algún sistema de ponderación de las frecuencias análogo al utilizado en la Tabla 4.3.2.

¿Y el modo simétrico?

Las dos reglas de Zeisel sintetizan lo esencial para el tratamiento asimétrico de tablas de contingencia. El análisis simétrico de estas tablas reviste comparativamente un interés menor. En este caso se computarán los porcentajes correspondientes a todas las frecuencias condicionales y marginales sobre la misma base del total de casos.

En el análisis asimétrico, el cálculo de los porcentajes sobre columnas -o sobre hileras - permite lograr una estandarización de las frecuencias condicionales que quedan así liberadas de los efectos de las diferencias marginales. Esto nos permitía en la Tabla 4.3 comparar un 82% de disconformes en Turismo con un 60% en Trabajo Social, aún cuando en la población hubiera el doble de Trabajadores Sociales.

En cambio, si los porcentajes se calculan todos sobre el 'n' del cuadro, no se logra ninguna estandarización, ya que las diferencias marginales continúan pesando sobre las frecuencias condicionales. En este sentido, debe resultar evidente la necesidad de que la muestra sea representativa en todos los sentidos, si se desea analizar simétricamente una tabla compuesta a partir de observaciones muestrales. Así, no cabría someter la Tabla 4.3 a un tratamiento simétrico, por la misma razón que tampoco resultaba lícito el cómputo horizontal de los porcentajes.

Pero, sobre todo, el análisis simétrico no es apto para examinar la existencia de una relación de dependencia entre las dos variables; optamos por este tipo de análisis cuando **no** interesa indagar acerca del presunto "efecto" de una variable sobre la otra. Así la Tabla 4.2 podría también ser analizada simétricamente.

Tabla 4.2.3: Misiones, 1980 - Distribución de la población por tipo de asentamiento y pertenencia a hogares con NBI

Tipo de asentamiento	Hogares con NBI		Total
	No	Sí	
Urbano	33,5	16,6	
Rural	21,1	28,8	49,9
Total	54,6	45,4	(580.522)

Fuente: Tabla 4.2.

Leeremos así que, de los 580.522 habitantes de la Provincia, hay un 33,5% que pertenece a hogares urbanos sin NBI, seguido por un 28,8% de rurales con NBI, 21,1% rurales sin NBI y 16,6 de urbanos con NBI. En esta forma de presentación de los datos, ya no se visualiza con la misma claridad el efecto de una variable sobre la otra, lo que no implica que ésta deje de ser una interpretación tan legítima como las anteriores. Simplemente, habrá variado nuestro propósito. Es posible, por ejemplo, que tengamos un interés especial en saber que un 28,8% de la población de Misiones pertenece a hogares rurales con NBI, para comparar esa cifra con el 1,8% que se registra para la misma categoría de población en la Provincia de Buenos Aires, más urbanizada y menos pobre, o con el 30,0% de la vecina Corrientes, más urbanizada y más pobre.

La Tabla 4.2 se basa en datos censales. Muchas investigaciones realizadas por muestreo pueden no perseguir el objetivo de determinar la existencia de una relación entre dos variables, sino proponerse la simple estimación de la proporción de una población que reúne determinadas características. De ser el caso, el tratamiento simétrico de los datos obtenidos por muestra permite obtener estimaciones de las proporciones de personas dentro de cada categoría de la población. Pero, si por lo contrario el objetivo es establecer una relación de dependencia entre dos variables, convendrá tratar la tabla asimétricamente.

2. EL ANÁLISIS DE LA RELACIÓN ENTRE VARIABLES

Cuando observamos mediante el tratamiento asimétrico de una tabla que una de las variables aparece determinando o afectando a la otra, podemos decir que ambas variables están *asociadas*.²² La medida de asociación más frecuentemente utilizada es, por lejos, la diferencia porcentual. Por otra parte, cuando se trata con muestras se plantea el problema adicional de determinar la significación estadística que se le puede prestar a una asociación entre variables. Abordaremos sucesivamente estos aspectos, para presentar luego algunos coeficientes de asociación.

2.1 La diferencia porcentual: una medida de la asociación

Por su simplicidad de cálculo y por la claridad de su significado, la diferencia porcentual es sin duda la medida de asociación más popular. En esencia consiste en una sistematización de la primera regla de Zeisel.

Consideremos el siguiente ejemplo, cuyos datos provienen de una muestra de 121 estudiantes,²³ partiendo de la hipótesis de que el grado de conocimiento político condiciona el grado de

²² El concepto de 'asociación' se usará para describir la existencia de una relación entre variables en una tabla de contingencia; para distribuciones multivariantes se hablará de 'correlación'.

²³ Los datos provienen del estudio "Participación política del estudiante de la FHCS-UNaM" (inédito) realizado por alumnos de Antropología Social en 1984.

participación política.²⁴

Tabla 4.4: Grado de participación política y grado de conocimiento político

Participación política	Conocimiento político		Total
	Bajo	Alto	
Alto	6	13	19
Bajo	59	43	102
Total	65	56	121

Fuente: elaboración propia.

Si lo que se quiere es comprobar el efecto del conocimiento sobre la participación, los porcentajes se deben computar en el sentido de la variable ‘conocimiento’, o sea verticalmente:

Tabla 4.4.1: Grado de participación política según grado de conocimiento político (%)

Participación política	Conocimiento político		Dif. %
	Bajo	Alto	
Alto	9	23	14
Bajo	91	77	-14
Total	100	100	(n = 121)

Fuente: elaboración propia.

Hemos simplemente aplicado la regla según la cual, los porcentajes se computan en dirección de la variable independiente y se comparan en la otra dirección. Salvo que ahora hacemos aparecer en la última columna la diferencia porcentual:

LA DIFERENCIA PORCENTUAL SE CALCULA EN LA DIRECCION EN QUE SE REALIZA LA COMPARACION

Mientras que en los alumnos de Bajo conocimiento sólo hay un 9% con alta participación, entre los de Alto conocimiento hay un 23%: es decir, hay un 14% más de alta participación política.²⁵

²⁴ Como habrá de verse a la brevedad, es igualmente plausible sostener la hipótesis de que «A mayor grado de participación política, mayor conocimiento». En términos de Zetterberg, éste es un ejemplo de relación **reversible e interdependiente** entre las variables (1968: 59 y ss.).

²⁵ Igualmente podríamos haber comparado los porcentajes de Baja participación, encontrando que en los alumnos de Alto conocimiento hay un 14% menos. Tratando con variables dicotómicas, y considerando que por definición los porcentajes deben sumar 100%, no puede sorprendernos que las diferencias porcentuales sean en ambos renglones

Ahora bien, imaginemos que quisiéramos en cambio determinar el efecto de la participación sobre el conocimiento:

Tabla 4.4.2: Grado de participación política y grado de conocimiento político

<i>Participación política</i>	<i>Conocimiento político</i>		<i>Total</i>
	<i>Bajo</i>	<i>Alto</i>	
<i>Alto</i>	32	68	100
<i>Bajo</i>	58	42	100
<i>Dif. %</i>	-26	26	(n = 121)

Fuente: elaboración propia.

Entre los altamente participativos hay un 26% más con alto conocimiento. Este 26% es también una medida de la asociación entre las variables, y tan válida como la anterior, aunque a todas luces diferente. Según sea nuestro interés, podremos optar por una u otra cifra; pero lo que muestra el ejemplo es que la diferencia porcentual no nos brinda una medida *general* de la asociación en la tabla. Sucede que los porcentajes son sensibles a los cambios en las distribuciones marginales, y que precisamente en la Tabla 4.4 estos marginales difieren en forma notable (19 y 102 para 'participación', 56 y 65 para 'conocimiento').

Cualquier uso de la diferencia porcentual como indicador resumen de la asociación en una tabla implica una gran pérdida de información. Además, este problema se magnifica al trabajar con tablas de formato mayor al 2 x 2; cuanto más elevado sea el número de valores de cada variable, se multiplicará la cantidad de diferencias porcentuales computables, y resultará aún más discutible la elección de una de las tantas diferencias posibles como medida resumen de la asociación en la tabla.

Tabla 4.5: Evaluación de la situación social según NES (%)

<i>Evaluación de la situación social</i>	<i>Nivel económico-social</i>			<i>Total</i>
	<i>Bajo</i>	<i>Medio</i>	<i>Alto</i>	
<i>Favorable</i>	35	38	47	41
<i>Neutra</i>	18	32	29	27
<i>Desfavorable</i>	45	30	24	32
<i>Total (100%)</i>	(40)	(73)	(49)	(162)

Fuente: elaboración propia.

idénticas aunque de signo contrario: necesariamente, todo aumento del porcentaje en una categoría debe implicar una disminución en la otra.

La Tabla 4.5 permite ilustrar este problema.²⁶ Puesto que el NES ha sido tomado como variable independiente, lo lógico es leer el cuadro comparando entre sí los porcentajes de cada hilera. Evidentemente, el sentido general de la tabla es que cuanto menor es el NES, más negativa resulta la evaluación de la situación: ello surge nítidamente de la comparación de los porcentajes de la última hilera. Pero es claro que no existe una única diferencia porcentual, sino nueve posibilidades distintas de cómputo de diferencias; solamente en esa última hilera, sería posible comparar 45 con 30, 30 con 24, o 45 con 24; y va de suyo que ninguna de estas diferencias es más “verdadera” que las otras.

Aún tomando en consideración estos defectos, no cabe menospreciar a la diferencia porcentual como instrumento del análisis. De hecho, en su práctica cotidiana el investigador la aplicará casi instintivamente, para tener una medida rápida de la asociación. Por lo demás, al trabajar con muestras la cuestión no radica simplemente en determinar el grado en que dos variables están asociadas, sino que se plantea un problema adicional.

¿Es esta relación estadísticamente significativa?

En la Tabla 4.4.1. la hipótesis inicial parecía corroborarse. Se observaba en efecto una diferencia positiva del 14% en cuanto a la participación de los estudiantes que contaban con una mayor grado de conocimiento político. Sin embargo, esta relación se verifica en una muestra, constituida por 121 estudiantes que eran sólo una parte de la totalidad de los estudiantes de la FHCS-UNaM en 1984. La muestra con la que trabajamos es solamente una de las tantas muestras que se hubieran podido extraer del universo de la investigación. Tal vez el azar haya sido la razón de que apareciera en la muestra este 14% más, cuando en realidad esta relación no se daba en el universo. La cuestión es: ¿Podemos considerar a esa diferencia del 14% lo suficientemente importante como para asumir que representa una diferencia existente realmente en el universo?. Cuando nos formulamos este tipo de preguntas, estamos inquiriendo si la relación es *estadísticamente significativa*.

2.2 El test de χ^2 : una medida de la significación estadística

El test de χ^2 (chi-cuadrado) es una de las respuestas posibles a esta pregunta. Dicho test es una de las pruebas de significación estadística más populares y se basa en una medida de cuánto se apartan las frecuencias condicionales observadas en la muestra de lo que serían las frecuencias esperables si no existiera ninguna relación entre las variables.

Retornemos a los datos de la Tabla 4.4 para considerar exclusivamente las frecuencias marginales:

	B	A	
A			19
B			102
	65	56	121

Examinando sólo los marginales no se puede decir nada acerca de la relación entre las variables. En cierto sentido, debemos pensar que estos marginales son lo que son. O, más exactamente, dados estos marginales, no podríamos tener **cuquiera** frecuencia dentro de la tabla.²⁷ Sin embargo, dentro de los límites establecidos por los marginales, es evidente que se puede imaginar muy diversas distribuciones de las frecuencias condicionales, y que estas distribuciones podrán ser muy diferentes en lo que hace a la relación entre las dos variables. Así, podríamos obtener:

²⁶ Los datos están tomados de un estudio (inédito) realizado por alumnos de la carrera de Antropología Social, en ocasión de las elecciones del 6 de septiembre de 1987 en Posadas.

²⁷ Así, por ejemplo, es claro que en ninguna de las dos celdas superiores podría haber una frecuencia mayor que ‘19’ (las frecuencias han de ser necesariamente números positivos).