

NOCIÓN Y ESTRUCTURA DEL DATO

Por Guillermo Pablo Abritta

La palabra dato tiene su origen etimológico en el término latino "Datum" que significa "lo dado". Sin embargo, en sentido estricto, en el ámbito de la investigación científica, como señala Gil Flores, J. (1) "La mayoría de los autores asumen que el investigador desempeña un papel activo respecto de los datos: el dato es el resultado de un proceso de elaboración, es decir, el dato hay que construirlo."

Siguiendo al mencionado autor, aunque no textualmente, la noción de Dato, se puede definir como aquella información extraída de la realidad que tiene que ser registrada en algún soporte físico o simbólico, que implica una elaboración conceptual y además que se pueda expresar a través de alguna forma de lenguaje. Los principales componentes enumerados por Gil Flores, J., se incluyen en esta definición, los cuales giran alrededor de la noción de dato, ellos son los que a continuación se exponen:

- 1) Una elaboración conceptual.
- 2) Un contenido informativo.
- 3) Un registro en algún soporte físico y
- 4) La expresión de los mismos en alguna forma de lenguaje, ya sea numérica o no numérica.

Tales componentes operan durante todo el proceso de la investigación, esto es, desde la elección del tema, la elaboración del diseño hasta el informe final, pero se plasman durante la etapa de recolección de datos a través de la administración de las técnicas de investigación, ya sean cuantitativas o cualitativas. En efecto, una entrevista producirá datos de naturaleza verbal, un test de inteligencia datos de naturaleza numérica. También está la alternativa a través de la cual el investigador encuentra los datos que han sido producidos por "otros", ya sea por los sujetos investigados o por otros investigadores. A pesar de ello, el dato obtenido es el resultado de una interacción entre el investigador, con sus supuestos básicos subyacentes, su enfoque del problema, los objetivos del estudio y su adhesión a algún paradigma teórico y/o metodológico.

Galtung, J. define el término "dato" de la siguiente manera (2): "Se obtienen datos sociológicos cuando un sociólogo registra hechos acerca de algún sector de la realidad social o recibe hechos registrados para él."

También sostiene que todo dato tiene una estructura compuesta por tres elementos: unidades de análisis, variables y valores. Cualquier dato consistirá en:

1) Una unidad de análisis que

2) En una variable asumirá

3) Un determinado valor.

Así lo que constituye un dato, para este autor, son estos tres elementos considerados en forma conjunta a través de las relaciones que mantienen entre sí.

U de Analisis Ahora bien, es necesario definir estos tres términos que en forma conjunta constituyen un dato. Las Unidades de Análisis son los elementos menores y no divisibles que componen el universo de estudio de una investigación. Sobre dichos elementos se estudia el comportamiento de las variables. Las unidades de análisis se establecen de manera previa a la etapa de recolección de datos, por ende, su definición forma parte del marco teórico. Las mismas se pueden clasificar según Mayntz, R.(3) en:

1) "Individuos como seres sociales.

2) Determinados productos de la acción humana, tanto de tipo material como inmaterial (por ejemplo, ideas, representaciones valorativas, normas),

3) Colectivos sociales o grupos: a saber desde conglomerados pequeños y efímeros hasta grandes colectividades organizadas con inclusión de las sociedades globales".

Generalmente, cuando las unidades son colectivos sociales, se distingue entre unidades de análisis y unidades de observación, estas últimas son las que van a aportar la información que se va a registrar y luego se va a comunicar a través de un lenguaje numérico o no numérico. Por ejemplo: si en un estudio las unidades de análisis son las escuelas, las unidades de observación pueden ser las autoridades de la escuela, los docentes y los alumnos.

Variable En lo que atañe al término Variable, el mismo tiene un origen matemático y luego por extensión en las ciencias sociales se convirtió en sinónimo de aspecto o dimensión, Korn F. (4) lo define de la siguiente manera: " En otros términos, el significado completo de la palabra "variable", tal como es usada en ciencias sociales, contiene no sólo la connotación de "aspecto" o "dimensión" de un fenómeno, sino también la propiedad de estos aspectos o dimensiones de asumir diferentes valores."

También se puede definir el término variable, como un concepto acerca de algún aspecto y/o magnitud de un elemento o unidad de análisis capaz de asumir diferentes cualidades y/o valores.

Mientras que un Valor o categoría es una de las diferentes posiciones o alternativas que presenta la variable y adopta alguna unidad de análisis y se puede expresar cualitativamente a través de una clasificación por ausencia y presencia, por jerarquía u orden o sino cuantitativamente, es decir, a través de magnitudes.

Cabe señalar, que Samaja, J. (5), contrariamente a lo que dice Galtung, sostiene que la

estructura general del dato científico tiene cuatro componentes, a saber:

1) Unidad de análisis

2) Variables

3) Valores

4) Indicadores.

Estos cuatro elementos del dato científico se relacionan con cuatro funciones que cumplen los enunciados descriptivos:

1) La Unidad de análisis corresponde al componente "argumento" (X);

2) La Variable, a la función misma (F);

3) El Valor coincide, incluso en el nombre, con el valor de la función (Y), y

4) El Indicador con las operaciones de que está construida y que permiten calcular el valor de la función.

El autor lo ejemplifica a través de un diálogo sacado de una escena entre Sherlock Holmes y el Dr. Watson: "Por lo que veo, ha estado usted en Afganistán". Según Samaja "Independientemente de las características coloquiales que tiene la afirmación anterior, en ella hay dato." Y más adelante afirma: "La unidad de análisis (UA) es, en este caso concreto el Dr. Watson. La función de descripción (que en la jerga metodológica se denomina "variable" (V) es "lugar de residencia". El valor (R) es Afganistán. Y por último, el indicador (I) (Por lo que veo ...) es "observación de indicios físicos".

Si bien la propuesta elaborada por el autor es interesante, cabe señalar que en la jerga metodológica un Indicador, en principio, es una variable de nivel empírico que se deduce y representa a la variable teórica. Según Díez Nicolás, J. (6) "El indicador es un signo (propiedad, atributo, variable) mediante el cual nos aproximamos al conocimiento de ciertas características de un objeto que no se pueden medir directamente (de aquí que se hable de inferencia)." Por lo expuesto y como se podrá apreciar más adelante, los indicadores se ubican, en una matriz de datos, en las columnas como el resto de las variables. Por ello, se opta en este trabajo por la concepción de Galtung acerca del dato, o sea, por la estructura tripartita del dato. Además, esta postura es la que goza de mayor consenso entre los autores que abordan el tema, como por ejemplo, para citar algunos de ellos: Baranger, D. (7), Errandonea, A. (8), Yalour, M.(9), Ghiglione, R. y Matalón, B. (10).

CLASIFICACIÓN DE LOS DATOS:

Sierra Bravo, R. (11), establece ciertos criterios de clasificación de las investigaciones científicas, a saber: según su fin, alcance temporal, amplitud, nivel de profundidad, carácter, marco, naturaleza y fuentes. Precisamente, las fuentes de datos son clasificadas, según el último criterio mencionado como: primarias, secundarias y mixtas.

datos primarios

1) Fuentes Primarias: son aquellos datos que son relevados por el investigador para realizar el estudio. La principal ventaja es que los mismos responden al problema, delimitación, objetivos y perspectiva teórica elaborada por el investigador. La desventaja principal que tienen los estudios con datos primarios es que los costos suelen ser muy elevados, por ello, en el diseño de los mismos hay que contemplar cuidadosamente los aspectos económicos y administrativos.

datos sec

2) Fuentes Secundarias: son aquellos que no han sido relevados por el investigador pero que son usados por él para realizar la investigación. La principal ventaja que tienen es que los costos son reducidos o nulos. En cambio, entre los inconvenientes que pueden presentar, se pueden destacar los siguientes: es muy probable que dichos datos hayan sido recogidos en función de otra problemática, otros objetivos y perspectiva teórica. Además hay que analizar cuál fue el universo de estudio, las unidades de análisis, la delimitación temporal y geográfica, la definición de las variables, el sistema de categorías, etc. Lo que antecede implica efectuar una evaluación de la calidad de los datos para así determinar si se ajustan o no al enfoque que un investigador pretende dar al tema en cuestión.

Ambos tipos de fuentes pueden incluir a su vez, Datos Numéricos que se rigen por técnicas estadísticas, también denominados "Datos Duros" y Datos No Numéricos o cualitativos, como por ejemplo, notas de campo, fotos, transcripción de entrevistas, documentos oficiales y personales, diarios, películas, etc. También se los denomina "Datos Blandos", esto es, difícilmente manejables mediante procedimientos estadísticos, aunque a través de posteriores elaboraciones, como por ejemplo, con una adecuada codificación y análisis de contenido, puedan expresarse mediante números. De acuerdo con lo expuesto precedentemente, se puede elaborar una Tipología de los Datos, en la misma se cruza la clasificación de los datos en primarios y secundarios, con la clasificación de los datos en numéricos y no numéricos. El resultado de la combinación brinda un tipo de dato para cada celda, como se puede observar en la siguiente tabla:

TIPOLOGÍA DE LOS DATOS		
	PRIMARIOS	SECUNDARIOS
NUMÉRICOS	PRIMARIOS NUMÉRICOS	SECUNDARIOS NUMÉRICOS
NO NUMÉRICOS	PRIMARIOS NO NUMÉRICOS	SECUNDARIOS NO NUMÉRICOS

3) Fuentes mixtas: este tipo de fuentes implica la combinación en un mismo estudio de datos primarios y datos secundarios, tanto cuantitativos como cualitativos.

CONFECCIÓN DE UNA MATRIZ DE DATOS:

Ahora bien, para confeccionar una Matriz de Datos, el investigador previamente debe, efectuar la recolección de datos mediante la elaboración, pretest y administración de alguna técnica de investigación que se ajuste al problema, los objetivos del estudio y al carácter del estudio. Por ejemplo, si el problema y los objetivos se ajustan a un estudio de carácter cuantitativo, lo más conveniente sería utilizar como técnica la encuesta o el cuestionario y elaborar una matriz de datos, en cambio si se pretende utilizar el análisis de contenido como técnica para efectuar un estudio cuantitativo, en lugar de elaborar una matriz, lo más adecuado será construir una planilla o cédula que permita codificar y tabular la información. Si el estudio es de carácter cualitativo y se escoge como técnica la observación no estructurada, lo más oportuno sería llevar un registro a través de notas de campo. En suma, puede haber muchas variantes, lo que se pretende destacar es que no siempre es necesario construir una matriz de datos semejante a la que se va a presentar más adelante.

Además de realizar la supervisión del trabajo de campo hay que efectuar una revisión de los registros obtenidos, mediante las diferentes técnicas. Si se trata de un cuestionario o una encuesta, hay que revisar los formularios o cédulas, para controlar el porcentaje de respuestas, la calidad de las mismas, la codificación de las preguntas, el seguimiento de las preguntas filtros, etc. En suma, tiene que haber una correcta edición de los registros. Si se trabaja con técnicas cualitativas como por ejemplo, una entrevista no estructurada o la observación no estructurada, se procede de la misma forma.

Una vez realizado el trabajo de campo y la edición de los datos obtenidos, es necesario efectuar el procesamiento de los datos, es decir, que los mismos se preparan para ser analizados, para ello se apela a dos técnicas de elaboración de los datos: la codificación y la tabulación. Lo que precede es válido, en lo que atañe a la codificación, tanto para una perspectiva metodológica cuantitativa como cualitativa.

En el caso de la tabulación, cabe señalar que no necesariamente está reñida con los estudios cualitativos, el hecho de realizarla o no depende de la decisión adoptada por el investigador.

Comboni, S. y Juárez, J., (12) afirman desde una perspectiva cuantitativa que: "La codificación es un procedimiento técnico mediante el cual, los datos obtenidos se clasifican en categorías y se traducen en símbolos, ya sean cifras o letras; es decir, se asigna a cada opción de respuestas un número o una letra que permita tabularla rápidamente."

Es importante señalar que la elaboración de un sistema de categorías y la codificación se pueden efectuar en forma simultánea, pero desde un punto de vista lógico, la codificación depende del sistema de categorías o valores que adopte la variable o alternativas que presente la pregunta.

Taylor, S.J. y Bogdan, R., por su parte, sostienen desde una perspectiva cualitativa lo siguiente: (13) "... la codificación es un modo sistemático de desarrollar y refinar las interpretaciones de los datos. El proceso de codificación incluye la reunión y análisis de

todos los datos que se refieren a temas, ideas, conceptos, interpretaciones y proposiciones. Durante esta etapa del análisis, lo que inicialmente fueron ideas e intuiciones vagas se refinan, expanden, descartan o desarrollan por completo." A continuación enumeran y describen cinco (5) fases para codificar los datos cualitativos:

- 1) Desarrolle categorías de codificación. Empiece redactando una lista de todos los temas, conceptos e interpretaciones, tipologías y proposiciones identificados o producidos durante el análisis inicial.
- 2) Codifique todos los datos. Codifique todas las notas de campo, las transcripciones, los documentos y otros materiales, escribiendo en el margen el número asignado o la letra correspondiente a cada categoría.
- 3) Separe los datos pertenecientes a las diversas categorías de codificación. ...El investigador reúne los datos codificados pertenecientes a cada categoría. ...se recortan las notas de campo, las transcripciones y otros materiales y se colocan los datos de cada categoría en carpetas de archivo... .
- 4) Vea que datos han sobrado. ... Algunos de esos datos probablemente se ajusten a las categorías de codificación existentes. También se pueden plantear nuevas categorías... .
- 5) Refine su análisis. La codificación y separación de los datos permite comparar diferentes fragmentos relacionados con cada tema, concepto, proposición, etcétera, y en consecuencia refinar y ajustar las ideas."

Las categorías pueden ser codificadas, lo que significa que se las va a identificar mediante un código numérico, antes, durante o después de efectuar el trabajo de campo. Lo que precede depende de la forma de las preguntas con las cuáles el investigador trabaje. Si las preguntas son cerradas, es decir, cuando la persona interrogada debe seleccionar entre una serie de categorías preestablecidas por el investigador, es conveniente que las categorías estén codificadas "a priori". Según Ghiglione, R. y Matalón, B. :(14) "La consigna puede adoptar formas muy diferentes; las más usuales son:

- * Indicar la respuesta más adecuada.
- * Indicar múltiples respuestas adecuadas, el número de las cuales es libre.
- * Indicar muchas respuestas cuyo número es limitado.
- * Ordenar todas las respuestas, de la más a la menos adecuada.
- * Ordenar las n (número fijo) respuestas más adecuadas."

En las preguntas siguientes se puede observar la codificación previa:

1. Sexo 2. Estado Civil
1. Masculino. 1. Soltero.
2. Femenino. 2. Casado/Unido.
3. Divorciado/Separado.
4. Viudo.
99. No sabe/No contesta.

Con referencia a las alternativas "No sabe", "No contesta", Cea D'Ancona, M. A., sostiene: (15) "...se aconseja su no inclusión expresa en la pregunta, salvo que se estime de interés en su formulación. La experiencia muestra que éstas constituyen opciones de respuesta muy recurridas, cuando el encuestado no quiere pensar o manifestar una respuesta concreta." Y más adelante agrega: "En caso que se opte por esta recomendación, se instruye al personal encargado de la grabación de los cuestionarios (ya completados) a que introduzca, automáticamente, el código correspondiente al "no contesta", a aquellas preguntas sin respuesta. Este código suele ser el número 9 o el 0, si sólo se precisa de un dígito para la grabación de las respuestas. Si precisan dos dígitos, puede optarse entre el 99 o el 00."

Por otra parte, si se estima conveniente separar las alternativas "No sabe" y "No contesta", generalmente, se le asigna a la primera el número 98 y a la segunda el número 99.

En la codificación de campo, se acostumbra a formular la pregunta a la persona interrogada, pero sin leerle las alternativas de respuesta. Entonces la respuesta es libre pero incluye alguna clasificación.

En cambio, si las preguntas son abiertas, esto es, cuando la persona interrogada responde con sus propias palabras, ya que se expresa con absoluta libertad y espontaneidad, lo que generalmente realiza el investigador es leer las respuestas y a través de la técnica de análisis de contenido, establece un sistema de clasificación de las respuestas. Dicho sistema consiste en agrupar las respuestas en categorías que obedezcan a un principio único y luego procede a codificarlas mediante números u otros símbolos. Así, de este modo el investigador o codificador procede a cerrar las preguntas abiertas.

Matriz de datos

Una vez realizada la codificación de los datos, se puede proceder a la confección de la matriz de datos. Según Galtung, J. (16): "La matriz de datos es un modo de ordenar los datos de manera que sea particularmente visible la forma tripartita."

En efecto, cada fila de la matriz corresponde a una unidad de análisis, cada columna a una variable y en cada celda, figura el valor que cada unidad asume para cada variable. De esta manera, con la articulación de estos tres elementos se configura una Matriz de Datos. A continuación, se exhibe un modelo de matriz de datos:

MODELO DE MATRIZ DE DATOS

UNIDAD DE ANÁLISIS	VARIABLES						
	V 1	V 2	V 3	V 4	V 5	V 6	VN
UA1	1	20	1	2	4	3	
UA2	1	28	1	1	5	1	
UA3	2	46	2	1	6	6	
UA4	1	34	3	4	7	1	
UA5	2	29	2	2	9	1	
UA6	2	19	1	3	4	3	
UA7	2	54	2	2	4	5	
UA8	1	67	4	5	2	4	
UA9	1	32	2	1	9	1	
UA10	2	23	3	4	5	6	
UA11	1	19	1	3	2	2	
UA12	2	18	1	3	5	3	
UA13	1	36	3	1	8	1	
UA14	2	63	4	1	3	5	
UAN							

REFERENCIA DE LA MATRIZ DE DATOS	
UA= UNIDAD DE ANÁLISIS V= VARIABLES	
<div style="border: 1px solid black; padding: 2px; margin-bottom: 10px;">V1</div> SEXO: 1. Masculino 2. Femenino	<div style="border: 1px solid black; padding: 2px; margin-bottom: 10px;">V2</div> EDAD (Pregunta Abierta)
<div style="border: 1px solid black; padding: 2px; margin-bottom: 10px;">V3</div> ESTADO CIVIL: 1. Soltero 2. Casado/Unido 3. Separado/Divorciado 4. Viudo 99. Ns./Nc.	<div style="border: 1px solid black; padding: 2px; margin-bottom: 10px;">V4</div> ¿PODRÍA DECIRME QUE LUGAR OCUPA EN SU FAMILIA? 1. Cabeza de familia 2. Esposa/o cabeza de familia 3. Hijo/a 4. Hermano/a de 5. Padre/Madre 6. Otro especificar
<div style="border: 1px solid black; padding: 2px; margin-bottom: 10px;">V5</div> NIVEL DE ESTUDIOS: 1. Sin Estudios 2. Primario Incompleto 3. Primario Completo 4. Secundario Incompleto 5. Secundario Completo 6. Terciario Incompleto 7. Universitario Incompleto 8. Terciario Completo 9. Universitario Completo o + 99. No sabe/No contesta	<div style="border: 1px solid black; padding: 2px; margin-bottom: 10px;">V6</div> SITUACIÓN OCUPACIONAL: 1. Trabaja 2. Desocupado 3. Estudia 4. Jubilado o Pensionado 5. Rentista 6. Ama de Casa

Como se puede observar, la configuración de una Matriz de Datos es muy similar a una Planilla Electrónica de Cálculo, es decir, una cuadrícula compuesta por filas y columnas, donde cada celda es la intersección de una fila y una columna.

La forma misma de la Matriz y la definición, conducen a los 3 (tres) Principios de la recolección de datos establecidos por Galtung, J.:

1) Principio de Comparabilidad: aquí se afirma que, cada combinación de Unidad de Análisis/Variable debe tener sentido, esto es, debe ser verdadero o falso que dicha combinación brinde un determinado Valor. Por medio de este principio se hacen comparables las variables, las unidades y los valores o respuestas. Por ejemplo, si las unidades son personas y una variable es la tasa de analfabetismo, entonces cualquier combinación no será ni verdadera ni falsa, simplemente no tiene sentido.

Por lo tanto, las variables han de ser de hecho, en la matriz, variables de las unidades investigadas. Por otro lado, las unidades sólo pueden ser comparadas con sentido respecto de valores de concretización sobre la variable correspondiente. Afirmaciones como "la unidad n°1 es de sexo masculino, mientras que la unidad n°2 tiene un alto grado de participación política no tienen de sentido."

2) Principio de clasificación: implica que el sistema de categorías de una variable debe cumplir con dos requisitos lógicos: debe ser exhaustivo y excluyente. Exhaustivo en el sentido que ninguna unidad quede fuera del sistema de categorías y excluyente, de manera que cada unidad asuma un valor y sólo uno en cada variable. En este último requisito puede haber excepciones, por ejemplo, en el caso de una pregunta con respuesta de elección múltiple, ya sea de número limitado o no limitado. Errandonea, A. (17), distingue un tercer requisito: "...deben obedecer al mismo criterio clasificatorio (variable)." Lo que antecede significa que el sistema de categorías de una variable debe responder a una sola idea.

3) Principio de Integridad: mientras que los dos principios anteriores aluden a la lógica de la matriz, este principio se refiere al trabajo empírico de llenar la matriz y dice simplemente esto: no dejar ninguna celda vacía. Lo que antecede significa que hay que intentar que la cantidad de celdas sin información, es decir, "No sabe"/"No contesta", se mantenga lo más baja posible. Sin embargo, en algunos casos, por ejemplo, en un estudio acerca del "Nivel de información acerca de las formas de contagio del SIDA", las alternativas mencionadas anteriormente resultan ser muy pertinentes, ya sea porque no hay conocimiento acerca del tema, no se tiene una opinión formada, no se quiere contestar o porque simplemente hay temas que no despiertan interés en la gente.

En lo que respecta, a los diferentes Tipos de Análisis que se pueden efectuar de una matriz de datos, Galtung, J. (18) describe los siguientes:

1) Análisis centrado en la variable: también denominado, por el citado autor, análisis vertical, porque las columnas se analizan en forma separada en cuanto a la información que brindan acerca de la variable correspondiente. Los valores que corresponden a las

diferentes unidades se comparan conforme al principio de comparabilidad. De esta forma se puede obtener una distribución estadística, en la cual para cada valor o categoría posible de la variable se da el número de unidades que tienen ese valor o categoría de la variable. Este tipo de análisis tiene un perfil básicamente univariable y cuantitativo.

2) Análisis centrado en la unidad de análisis: también denominado análisis horizontal. Porque en el mismo se analizan las filas separadamente en cuanto a la información que dan acerca de las unidades. Se analiza cada unidad separadamente y los valores de las diferentes variables no son comparables. Por lo tanto, este tipo de análisis nos brinda pautas o rasgos de cada unidad. Es un tipo de análisis con un perfil más cualitativo.

3) Análisis combinado: vertical y horizontal, los cuales pueden adoptar, según Galtung, diversas formas: puede ser bivariable o multivariable. Además se puede comenzar con un análisis horizontal, construyendo un índice y luego se puede efectuar un análisis vertical incorporando los valores que aporta el índice para todas las unidades, en una nueva columna.

TABULACIÓN DE LOS DATOS:

Luego de confeccionar la matriz de datos, se procede a la tabulación de los mismos. Según Rojas Soriano, R.,(19): "La tabulación es el proceso mediante el cual los datos recopilados se organizan y concentran, con base a determinadas ideas o hipótesis, en tablas o cuadros para su tratamiento estadístico."

Entonces tabular es contar las unidades que son ubicadas, ya sea en forma manual o con la utilización de una computadora, en cada categoría de una variable o unidades que son ubicadas simultáneamente en categorías determinadas de dos o más variables. Por lo tanto, la tabulación puede ser simple, esto es, univariable o cruzada, es decir, bivariable o multivariable.

Por supuesto, lo que antecede requiere un "plan de tabulación", esto es, determinar de antemano qué resultados de las variables se van a presentar y cuáles relaciones entre las mismas se van analizar, a fin de brindar respuesta al problema y los objetivos formulados.

Se presenta a continuación un ejemplo de tabulación simple o univariable elaborado a partir de la primera variable incluida en el Modelo de Matriz de Datos, o sea, la variable Sexo. Cabe aclarar que hoy en día esto se hace con la computadora, sin embargo, es bueno saber cómo se procede en forma manual. Entonces, se ubica la variable con sus categorías Masculino/Femenino codificadas 1 2 respectivamente y luego se comienza a contar. Por ejemplo, cada vez que aparece un número 1 se coloca un palote en la categoría masculino, cuando se llega a 4 se cruza el 5 para facilitar el recuento.

Sexo	
1) Masculino	IIII II
2) Femenino	IIII II

Luego se recuenta de la tabla anterior para cada categoría la frecuencia absoluta y a partir de esta se obtiene la frecuencia relativa o porcentual, como resultado de ello se obtiene la siguiente tabla con su respectivo número y título:

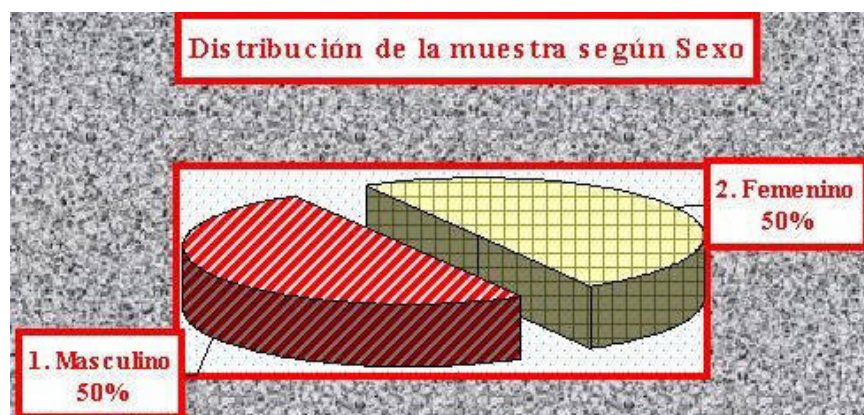
TABLA N°1:
Distribución según Sexo

Sexo	Frecuencia	Frecuencia %
1) Masculino	7	50
2) Femenino	7	50
TOTAL	14	100

En este punto es necesario definir algunos términos, como frecuencia o frecuencia absoluta (f) y frecuencia relativa o porcentual (f %).

La frecuencia absoluta es el número de veces que se observa cada valor o categoría de la variable, por ejemplo, en la tabla precedente la frecuencia de la categoría masculino es igual a 7 (siete), lo mismo ocurre con la categoría femenino.

La frecuencia relativa resulta de dividir cada una de las frecuencias absolutas de la tabla sobre el total de casos y luego se multiplica por 100 y así se obtiene el porcentaje. En este ejemplo, para la categoría masculino: $7 \div 14 = 0,50 \times 100 = 50\%$. Lo mismo ocurre con la categoría femenino. Entonces se puede afirmar que en esta pequeña muestra de catorce (14) casos, con referencia a la variable sexo no se registran diferencias de representación tanto para los hombres como las mujeres, ni en cifras absolutas ni en cifras porcentuales. También, a partir de esta distribución se puede elaborar un gráfico para visualizar la misma.



Los porcentajes, según Zeisel, H., (20): "Se utilizan para indicar con mayor claridad la dimensión relativa de dos o más números. Logran este esclarecimiento en dos formas: primera, reducen todos los números a una escala que sea fácil para multiplicar y dividir, por regla general los porcentajes son números menores de 100; segunda, transforman a uno de los números, que es la base, en la cifra 100, la cual es fácilmente divisible entre y por otros números, con lo que se facilita la determinación de su magnitud relativa."

Lo que antecede es muy importante, ya que en el ejemplo desarrollado aquí, por razones obvias, se trabajó con una muestra de catorce (14) casos donde resulta sencillo operar con cifras absolutas, pero si hay que trabajar con muestras de mayor tamaño, por ejemplo, 500 o 1000 casos, se vuelve muy engorroso operar con cifras absolutas, por eso se recomienda utilizar cifras porcentuales.

También es posible tabular dos o más variables en forma simultánea, lo cual resulta ser muy útil para poner a prueba la o las hipótesis que se han formulado como respuesta anticipada al problema de una investigación. Por supuesto que, como se señaló anteriormente esto se hace con una computadora, pero siempre es bueno saber cómo se puede hacer manualmente. A continuación se expondrá un ejemplo de tabulación cruzada, con dos de las variables incluidas en el Modelo de Matriz de Datos: V1, Sexo y V3, Estado Civil. Cabe aclarar que este ejemplo no tiene por objetivo poner a prueba ninguna hipótesis, ya que son dos variables de clasificación. El único sentido que persigue es mostrar cómo se hace una tabulación cruzada. Más pertinente hubiera sido cruzar una variable de clasificación con una de opinión o de actitud.

En principio, como resultado del cruce de estas dos variables, la tabla va a constar de ocho (8) celdas. En efecto, la variable Sexo tiene dos categorías, mientras que la variable Estado Civil, tiene cinco categorías, pero hay una de ellas, la N°5 No Sabe/No contesta, que no aparece en la matriz, entonces no se la va a considerar. Cada celda va a estar identificada con una letra, como se puede observar en el siguiente modelo:

ESTADO CML	SEXO	
	1. MASCULINO	2. FEMENINO
1. SOLTERO	a	b
2. CASADO/UNIDO	c	d
3. SEPARADO/DIVORCIADO	e	f
4. VIUDO	g	h

A continuación hay que ubicar a las unidades de análisis en las diferentes celdas, comenzando por la UA1, que para ambas variables registra el código 1, entonces deberá ubicarse en la celda "a", la segunda ídem, la UA3 asume para las dos variables el código 2, por ende deberá ubicarse en la celda "d", la UA4, asume para la variable Sexo el código 1 y para Estado Civil el código 3, por ende deberá ubicarse en la celda "e" y así sucesivamente hasta completar los catorce (14) casos. Como resultado de ello se obtendrá la siguiente tabulación:

ESTADO CML	SEXO	
	1. MASCULINO	2. FEMENINO
1. SOLTERO	///	//
2. CASADO/UNIDO	/	///
3. SEPARADO/DIVORCIADO	//	/
4. MUDO	/	/

Luego se procede a efectuar el recuento de los casos que se ubican en cada celda y se obtienen los totales, como se puede observar en la siguiente tabla:

TABLA N° 1
ESTADO CIVIL SEGÚN SEXO

ESTADO CIVIL	SEXO		TOTAL
	1. MASCULINO	2. FEMENINO	
1. SOLTERO	3	2	5
2. CASADO/UNIDO	1	3	4
3. SEPARADO/DIVORCIADO	2	1	3
4. MUDO	1	1	2
TOTAL	7	7	14

Con el resultado del recuento de los casos incluidos en la matriz, se ha elaborado una Tabla de Contingencia que contiene cifras absolutas, la misma nos va a permitir iniciar un análisis bivariante y/o multivariante de los datos.

ANÁLISIS E INTERPRETACIÓN DE LOS DATOS:

Una vez que se ha concluido con la recolección, codificación y tabulación de los datos, sigue la etapa de análisis y luego de interpretación de los datos. Según Rojas Soriano, R. (21): "Estas etapas se encuentran estrechamente ligadas, por lo cual suele confundirseles. El análisis consiste en separar los elementos básicos de la información y examinarlos con el propósito de responder a las distintas cuestiones planteadas en la investigación. La interpretación es el proceso mental mediante el cual se trata de encontrar un significado más amplio de la información empírica recabada."

En efecto, estas etapas van a permitir extraer conclusiones en función del problema, las variables, la formulación de las hipótesis y el marco teórico elaborado por el investigador. El análisis consiste en descomponer la información. El mismo requiere efectuar un plan previo, es decir, un "plan de análisis" que, de manera general, consiste en describir que tipo de tratamiento (cuantitativo y/o cualitativo) se le dará a la información. La interpretación apunta a darle sentido, esto es, efectuar eventualmente una recomposición teórica para luego redactar el informe final de la investigación.

Cuando se expuso el tema de la tabulación de los datos se desarrolló un ejemplo de análisis univariado. A continuación se expondrá un ejemplo de análisis bivariado. Previamente, es conveniente definir y describir la forma lógica de una tabla de contingencia.

Para Baranger, D. (22): "Una tabla de contingencia es el resultado del cruce (o tabulación simultánea) de dos o más variables. () ...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. 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, se han volcado las frecuencias (número de casos) correspondientes a cada combinación de valores de varias variables."

El número de celdas en una tabla de contingencia varía en función del número de categorías que tengan las variables incluidas en ella. En efecto, la combinación mínima es igual a una tabla de 2x2, o sea de dos valores por dos valores, luego pueden darse una multiplicidad de combinaciones. La forma lógica de una tabla 2x2 se puede representar de la siguiente manera:

En este tipo de tablas donde se cruzan dos variables, cada una de ellas con dos categorías:

1) "n" se denomina "frecuencia de orden cero" (0), ya que no se refiere a ninguna variable, esto es, representa a la totalidad de unidades que conforman la muestra o el universo.

2) Las frecuencias de abajo y las del costado, reciben el nombre de "frecuencias marginales" de "X" e "Y" respectivamente, o también se las denomina frecuencias de

primer orden ya que son frecuencias con respecto a una sola variable por vez.

3) Por su parte las frecuencias del interior del cuadro se las denomina "frecuencias condicionales", o también frecuencias de segundo orden, pues se refieren a la relación entre dos variables de manera simultánea. En efecto, aquí cada unidad asume un valor de X y un valor de Y, de manera combinada.

A continuación se va a aplicar este modelo a un ejemplo hipotético. Supóngase un investigador en un estudio, cuyo tema es "El grado de exposición a los medios masivos de comunicación de los habitantes de la ciudad de Buenos Aires."

Las preguntas iniciales son las siguientes: ¿Cuál es el grado de exposición a los medios masivos de comunicación de los habitantes de la ciudad de Buenos Aires en el año 1998?. ¿Cuáles son los factores que inciden en el grado de exposición a los medios masivos de comunicación de los habitantes de la ciudad de Buenos Aires?. Teniendo por objetivo general: Medir el grado de exposición a los medios masivos de comunicación de los habitantes de la ciudad y entre los objetivos específicos: Determinar el grado de exposición a los medios masivos según nivel socioeconómico.

Luego se procede a formular, entre otras, la siguiente hipótesis: "A mayor Nivel Socioeconómico de los habitantes de la ciudad de Buenos Aires menor grado de exposición a los medios masivos de comunicación y a la inversa." La variable Nivel Socioeconómico se la considera como independiente (X) y el grado de exposición a los medios como dependiente (Y). Ambas variables se midieron a través de la elaboración de sus respectivos índices. Finalmente, se efectúa la tabulación cruzada de los datos y se obtiene la siguiente tabla de contingencia:

TABLA N° 2

GRADO DE EXPOSICIÓN A LOS MEDIOS SEGÚN NIVEL SOCIOECONÓMICO

GRADO DE EXPOSICIÓN A LOS MEDIOS	NIVEL SOCIOECONÓMICO		TOTAL
	MEDIO BAJA	MEDIO ALTA	
MEDIO BAJA	150	150	300
MEDIO ALTA	850	150	1000
TOTAL	1000	300	1300

Baranger, D. (23) afirma lo siguiente: "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."

$$\begin{aligned}
N &= (X) + (-X) \\
&= (Y) + (-Y) \\
&= (XY) + (-XY) + (X-Y) + (-X-Y) \\
N &= 1300 \\
&= 1000 + 300 \\
&= 300 + 1000 \\
&= 150 + 150 + 850 + 150
\end{aligned}$$

Los valores de la tabla están expresados en cifras absolutas, lo cual dificulta efectuar el análisis, ya que no se pueden realizar comparaciones al no existir una base común, por lo tanto, conviene calcular los porcentajes y obtener la Diferencia Porcentual y/o aplicar algún Coeficiente de Asociación. Para ello hay que tomar una decisión relacionada con el tipo de nexo que existe entre las variables que componen la hipótesis. En efecto, la relación entre las variables puede ser asimétrica o simétrica. Una relación es asimétrica cuando $X \rightarrow Y$, es decir, la variable "X" determina a "Y", y no a la inversa. Lo que antecede significa que la relación es irreversible. Por ejemplo: "El estilo de liderazgo de los jefes determina el nivel de participación de los trabajadores". Una relación es simétrica cuando $X \leftrightarrow Y$, esto es, ninguna variable es causa de otra, por lo tanto es una relación reversible. Por ejemplo: "A mayor grado de participación, mayor grado de comunicación y a mayor grado de comunicación, mayor grado de participación."

Cabe señalar que, la línea divisoria entre asimetría y simetría no siempre es fácil de establecer. Por ejemplo: "A mayor grado de desigualdad social, mayor nivel de exclusión social y a mayor grado de exclusión social, mayor grado de desigualdad social." Aquí en principio se podría pensar que la relación es asimétrica, ya que la desigualdad actuaría como causa de la exclusión, pero también hay que considerar que un incremento en la exclusión acentuaría la desigualdad, por lo tanto, también se podría pensar que la relación sería simétrica, en el sentido que habría una especie de retroalimentación entre las variables en cuestión.

Continuando con el tema y siguiendo la orientación de Mora y Araujo, M. (24), aunque no literalmente, si la relación es asimétrica, generalmente, hay que calcular los porcentajes y las diferencias porcentuales o el coeficiente Lambda asimétrico (Goodman y Koruskal). En cambio, si la relación es simétrica hay que utilizar algún coeficiente de asociación, como por ejemplo, Q de Yule, Phi, Gamma (Goodman y Koruskal), Tau b (Kendall), Lambda simétrico, etc. Las condiciones de uso de los diferentes coeficientes depende de varios factores: el nivel de medición de las variables, el número de categorías de las variables, esto es, si la tabla o cuadro es de 2×2 , 3×3 o más y también hay que tener en cuenta si la asociación es lineal o curvilínea. También se pueden utilizar en casos de querer observar la asociación entre dos variables sin tener el objetivo de especificar si una determina a la otra en la estructura de la hipótesis.

El resultado de los mismos varía entre -1 a 1 pasando por 0 (cero). 1 (uno) es el grado máximo de asociación, 0 (cero) indica independencia, o sea, falta de asociación y -1 (uno) es el grado máximo de asociación inverso a la hipótesis. El signo indica la dirección de la asociación y depende de la ubicación de las categorías de la variable en el diseño de la tabla. Para variables de nivel de medición nominal no se lo considera, para

el resto de los niveles si y se sugiere ordenar las categorías de las variables de menor a mayor en la tabla.

Para tablas de 2 x 2 se aplican Q de Yule y Phi. El primero se utiliza para relaciones rinconales y el segundo para relaciones diagonales. Ambos se pueden aplicar a variables de nivel de medición nominal y ordinal.

Para tablas de 3 x 3 o más se aplican Gamma y Tau b. El primero para relaciones rinconales y el segundo para relaciones diagonales. Ambos se utilizan para variables de nivel de medición ordinal.

Cuando ambas variables son de nivel de medición intervalar o racional se aplica el Coeficiente de Correlación "r" de Pearson y sus derivados.

En el caso de una relación diagonal, para una tabla de 2 x 2, las celdas registrarán una fuerte concentración en la diagonal central, como se puede observar en la figura siguiente:

N	n
n	N

Donde:

N = frecuencias altas.

n = frecuencias más bajas.

0 = frecuencias nulas o bajas.

En el caso de una relación rinconal las concentraciones adoptarán una forma curvilínea, como se puede observar en la figura siguiente:

N	n
0	n

Mora y Araujo, M. (25) afirma lo siguiente: "En el caso diagonal, la probabilidad es que si "X" está presente "Y" también lo esté, y si "X" está ausente "Y" también esté ausente. En el rinconal la probabilidad es que, por ejemplo, si "X" está presente "Y" también lo

esté, pero si "X" no está entonces "Y" está o no está (una variable es condición suficiente pero no necesaria para la aparición de un valor de la otra).

Ejemplos:

Asociación diagonal: Un alto grado de organización familiar incide sobre el rendimiento escolar de los hijos.

Asociación rinconal: Con un nivel de información incompleto acerca de las formas de contagio del SIDA el nivel de prevención es malo, pero para un nivel de información completo el nivel de prevención puede ser bueno o malo.

Nota: en el Anexo I se encuentran las fórmulas de los diferentes coeficientes mencionados.

Ahora bien, retomando el ejemplo expuesto en la Tabla N° 1, se puede observar que la relación es asimétrica, por lo tanto hay que proceder a calcular los porcentajes y las diferencias porcentuales.

Zeizel, H. (26) enunció una regla para calcular los porcentajes: "La regla es que los porcentajes deben computarse en el sentido del factor causal."

Entonces hay que construir la tabla de manera tal que "X" se ubique en la parte superior de la misma e "Y" en el costado izquierdo.

Luego se divide cada frecuencia condicional por el total de la columna y se multiplica por 100, esto es, se calculan los porcentajes en forma vertical y finalmente se comparan los porcentajes de forma horizontal.

De esta manera se calcula la Diferencia Porcentual, que sirve para observar cuanto la variable independiente influye sobre la dependiente. La Diferencia Porcentual, $d\%$, varía de 0 a 100 o de 0 a -100. Cuando la $d\%$ da 0 (cero) o próxima a 0 hay independencia entre las variables; en cambio, cuando la $d\%$ da 100 se obtiene el grado máximo de dependencia entre las variables. Entonces en principio, la $d\%$ indica si las variables son independientes o se encuentran asociadas.

En el caso de estar asociadas el resultado de la misma refleja la intensidad de la asociación y por último el signo + o -, señala la dirección de la relación y depende, al igual que en el caso de los coeficientes de asociación, de la ubicación de las categorías de las variables en el diseño de la tabla.

A continuación se presenta la Tabla N° 3 con el cálculo de los porcentajes y la diferencia porcentual:

TABLA N° 3

GRADO DE EXPOSICIÓN A LOS MEDIOS SEGÚN NIVEL SOCIOECONÓMICO

GRADO DE EXPOSICIÓN A LOS MEDIOS	NIVEL SOCIOECONÓMICO				TOTAL		Dif. %
	MEDIO BAJO	%	MEDIO ALTO	%			
MEDIO BAJA	150	15	150	50	300	15 - 50 =	-35
MEDIO ALTA	850	85	150	50	1000	85 - 50 =	35
TOTAL	1000	100	300	100	1300	Sumatoria= 0	

La siguiente tabla presenta los resultados sólo en cifras porcentuales:

TABLA N° 4

GRADO DE EXPOSICIÓN A LOS MEDIOS SEGÚN NIVEL SOCIOECONÓMICO (%)

GRADO DE EXPOSICIÓN A LOS MEDIOS	NIVEL SOCIOECONÓMICO		Dif. %	PORCENTUAR
	MEDIO BAJO	MEDIO ALTO		
MEDIO BAJA	15	50	-35	↓ (X)
MEDIO ALTA	85	50	35	
TOTAL	100	100	Sumatoria=0	
COMPARAR	→		(Y)	(n = 1300)

Cabe efectuar una aclaración con referencia a las dos tablas precedentes y es la siguiente: en una tabla o cuadro de 2 X 2, la suma de ambas diferencias porcentuales siempre tiene que ser igual a 0 (cero). Lo que antecede es una prueba de que los cálculos se han efectuado correctamente. Pero de ninguna manera esa suma igual a " 0 " (cero) debe confundirse con el resultado de la d%, que en este caso es igual a 35%. Lo mismo ocurre para el resto de las posibles combinaciones, esto es, 2 x 3, 3 x 3, etc.

Ahora cabe formular la siguiente pregunta ¿Cómo se lee y analiza una tabla de contingencia? Al respecto, existen ciertas recomendaciones, enumeradas por Mora y Araujo, M. (27) que resultan ser básicas, a continuación se expondrán algunas de ellas:

1) Toda tabla debe tener su número y el título correspondiente, indicando las unidades o cifras en que se expresan los datos. Por convención, se ubica primero a la variable dependiente y luego la variable independiente. En este caso: "Grado de exposición a los medios según Nivel Socioeconómico."

2) Leer las notas aclaratorias, explicaciones y notas de pie.

3) Identificar claramente cuáles son las unidades de análisis, las variables y sus respectivas categorías y los niveles de medición de las mismas.

4) Verificar si el cuadro o tabla debe ser tratado como asimétrico o simétrico.

a) Si es asimétrico, calcular los porcentajes y las diferencias porcentuales. Para ello hay que determinar cuál será la base de los porcentajes y efectuar las comparaciones en el sentido contrario. Si las variables tienen más de dos categorías la diferencia porcentual debe calcularse entre la primera y la última columna o por diferencia de a pares, lo más práctico es la primera alternativa. Por ejemplo:

TABLA N° 5

GRADO DE EXPOSICIÓN A LOS MEDIOS SEGÚN NIVEL SOCIOECONÓMICO

		NIVEL SOCIOECONÓMICO			TOTAL
		BAJO	MEDIO	ALTO	
GRADO DE EXPOSICIÓN A LOS MEDIOS	BAJO	20	80	120	220
	MEDIO	80	60	40	180
	ALTO	120	60	20	200
	TOTAL	220	200	180	600

(n = 1300)

TABLA N° 6

GRADO DE EXPOSICIÓN A LOS MEDIOS SEGÚN NIVEL SOCIOECONÓMICO (%)

		NIVEL SOCIOECONÓMICO			Dif. %
		BAJO	MEDIO	ALTO	
GRADO DE EXPOSICIÓN A LOS MEDIOS	BAJO	9	40	67	-58
	MEDIO	36	30	22	14
	ALTO	55	30	11	44
	TOTAL	100	100	100	Sumatoria = 0

n= 600 (220) (200) (180)

Ejemplos de las dos formas de cálculo de las diferencias porcentuales:

1) Por los dos extremos:

1ra. Fila $9 - 67 = -58$

2da. Fila $36 - 22 = 14$

3ra. Fila $55 - 11 = 44$

Sumatoria = 0

2) Por diferencia de pares:

1ra. Fila $9 - 40 / 40 - 67$

$- 31 - 27 = - 58$

2da. Fila $36 - 30 / 30 - 22$

$6 - 8 = 14$

3ra. Fila $55 - 30 / 30 - 11$

$25 - 19 = 44$

Sumatoria = 0

b) Si es simétrico, aplicar algún coeficiente de asociación, teniendo en cuenta el número de categorías de las variables, la relación que existe entre las mismas y sus niveles de medición.

5) Iniciar el análisis por las distribuciones marginales de las variables, para observar si son uniformes o no y luego continuar con las distribuciones condicionales. Por ejemplo, (Ver Tabla N° 3) en la muestra hay, por un lado, un:

77 % de nivel socioeconómico medio bajo y

33 % de nivel medio alto, y por el otro, un

23 % con exposición media baja y

67 % con exposición media alta. Luego,

a) Un 15 % de individuos de nivel socioeconómico medio bajo tienen una exposición a los medios media baja.

b) Un 85 % de nivel socioeconómico medio bajo tiene una exposición media alta.

c) Un 50 % tiene un nivel socioeconómico medio alto y una exposición medio baja.

d) Finalmente un 50 % tiene un nivel socioeconómico medio alto con una exposición media alta.

Si bien todos estos enunciados son verdaderos, no brindan una buena lectura de una tabla, ya que como señala Baranger, D. (28), por un lado resultan ser redundantes y por el otro no apuntan a destacar la relación entre las variables. En efecto, lo esencial de la información está contenida en los enunciados a) y c) o indistintamente b) y d), y se puede expresar de la siguiente forma: "Mientras un 15 % de nivel socioeconómico medio bajo tiene una exposición a los medios media baja, un 50 % de nivel socioeconómico medio alto tiene una exposición media baja." Aquí es donde surge la diferencia de 35 puntos, entre ambos estratos y por ende se verifica la hipótesis de trabajo. En efecto, a mayor nivel socioeconómico menor exposición a los medios y a la inversa. Al leer la tabla se puede observar que las mayores frecuencias, tanto en términos absolutos como relativos, se concentran en la diagonal comprendida por las celdas "b" y "c".

6) Finalmente, cuando las variables tienen muchas categorías, es conveniente eliminar o reducir la cantidad de las mismas. Esto se puede hacer:

- a) Dejando de lado los "no sabe" y "no contesta", a menos que el "no sabe" pueda ser considerado como una categoría significativa. Por ejemplo, en un ítem cognitivo.
- b) Reducir el número de categorías por la eliminación de algunas. Por ejemplo, las del medio.
- c) Combinar distintas categorías.

BUENOS AIRES, Marzo de 1999.

ANEXO I:

FÓRMULAS DE LOS COEFICIENTES DE ASOCIACIÓN:

a	b
c	d

$$\text{Phi} = \frac{(a \cdot d) - (b \cdot c)}{\sqrt{(a+b) \cdot (c+d) \cdot (a+c) \cdot (b+d)}}$$

$$\text{Q de Yule} = \frac{(a \cdot d) - (b \cdot c)}{(a \cdot d) + (b \cdot c)}$$

a	b	c	m
d	e	f	n
g	h	i	o
j	k	l	

$$\text{Tau} = \frac{P - Q}{\sqrt{(j \cdot k) + (j \cdot i) + (k \cdot l) + (m \cdot n) + (m \cdot o) + (n \cdot o)}}$$

P= es cada valor de frecuencia multiplicado por las frecuencias condicionales que quedan hacia abajo y a la derecha.

Q= es cada valor de frecuencia multiplicado por las frecuencias condicionales que quedan hacia abajo y a la izquierda.

$$\text{Gamma} = \frac{p - q}{p + q}$$

P= es cada valor de frecuencia multiplicado por las frecuencias condicionales que quedan hacia abajo y a la derecha.

Q= es cada valor de frecuencia multiplicado por las frecuencias condicionales que quedan hacia abajo y a la izquierda.

$$P = a(e+f+h+i) + b(f+i) + d(h+i) + e(i)$$

$$Q = c(d+e+g+h) + b(d+g) + f(g+h) + e(g)$$

BIBLIOGRAFÍA:

- (1) Gil Flores, Javier: Análisis de Datos Cualitativos. Aplicaciones a la Investigación Educativa, Barcelona, Edit. PPU, 1994, Cap. 1.
- (2) Galtung, Johan: Teoría y Método de la Investigación Social. Buenos Aires, Eudeba, 1966, Tomo I. Cap. 1. Pág. 1.
- (3) Mayntz, Renate; Holm, Kurt y Hubner, Peter: " Introducción a los métodos de la sociología empírica. Madrid. Alianza Editorial. 1988. Cap. 1. Pág. 16.
- (4) Korn, Francis: "Conceptos y Variables en la Investigación Social." Edit. Nueva Visión. Buenos Aires. 1984. Cap. 1.
- (5) Samaja, Juan: " Epistemología y Metodología. Elementos para una teoría de la investigación científica. Buenos Aires. Eudeba. 1993. Parte III, punto 3.5
- (6) Díez Nicolás, J.: "Sociología." Madrid. Guadiana. 1969. Pág. 334. Citado por Garmendia, José A.: "Sociología. Claves para el estudio y transformación de la estructura social. Madrid. Centro de Investigación Social. 1979. Pág. 142.
- (7) Baranger, Denis: "Construcción y Análisis de Datos". Introducción al uso de técnicas cuantitativas en la investigación social. Posadas - Misiones - Argentina. Edit. Universitaria Cátedra. 1992. Cap. 2.
- (8) Errandonea, Alfredo: Inédito. "Colección de Guías de Clase de Metodología de la investigación. Guía de Clase N°6". Facultad de Ciencias Políticas. Universidad del Salvador. 1986.
- (9) Hyman, Lazarsfeld, Zeisel, Sorokin, Coser: "La Investigación Social", Introducción, Notas y Selección de Textos por Yalour, Margot Romano. Buenos Aires. Centro Editor de América Latina. 1977. Pág. 15.
- (10) Ghiglione, Rodolfo y Matalón, Benjamín: "Las Encuestas Sociológicas. Teorías y Prácticas." México. Edit. Trillas. 1989. Cap. 7.
- (11) Sierra Bravo, Restituto: "Técnicas de Investigación Social. Madrid. Edit. Paraninfo. 1988. Cap. 2, pág. 34.
- (12) Comboni, Susana y Juárez, José M.: "Introducción a las técnicas de investigación." México. Edit. Trillas. 1990. Cap. 4. Pág. 92.
- (13) Taylor, J.S. y Bogdan, R.: "Introducción a los Métodos Cualitativos de Investigación. Buenos Aires. Edit. Paidós. 1986. Cap. 6. Pág. 167.
- (14) Ghiglione, Rodolfo y Matalón, Benjamín: op. cit. Cap. 4. Pág. 107.

- (15) Cea D'Ancona, M. Angeles: "Metodología Cuantitativa. Estrategias y técnicas de investigación social. Editorial Síntesis. Madrid. Año. Cap.7. Pág.255.
- (16) Galtung, Johan: op. cit, Cap. 1. Pág. 3.
- (17) Errandonea, Alfredo: "Manual de introducción a la Sociología." Montevideo. Fundación de cultura universitaria. Cap. 2. Pág. 29.
- (18) Galtung, Johan: op. cit. Tomo II. Cap. 1. Pág. 218.
- (19) Rojas Soriano, Raúl: "La investigación Social:Teoría y Praxis."Edit. Folios. México. 1986. Pág.177.
- (20) Zeisel, Hans: "Dígalo con números." Edit. F.C.E.. México.1974. Cap.I. Pág. 21.
- (21) Rojas Soriano, Raúl: op. cit. Cap. 12. Pág. 241.
- (22) Baranger, Denis: op. cit. Cap.IV. Pág. 61.
- (23) Baranger, Denis: op. cit. Cap.IV. Pág. 63.
- (24) Mora y Araujo, Manuel: "Recomendaciones para la lectura y análisis de cuadros". Centro de Documentación de Sociología. U.B.A. Ficha N° 514. Elementos de Metodología y Técnicas de la Investigación Social. 1965.
- (25) Mora y Araujo, Manuel: op. cit. Pág. 5.
- (26) Zeisel, Hans: "Dígalo con números." Edit. F.C.E.. México.1974. Cap. II. Pág. 37.
- (27) Mora y Araujo, Manuel: op. cit. Pág. 1.
- (28) Baranger, Denis: op. cit. Cap.IV. Pág. 71.