

EL PORCENTAJE DE DESVIACIÓN MÁXIMA (PEM) UNA TÉCNICA PARA LEER DATOS CUANTITATIVOS

Abelino Martínez Röcha

RESUMEN

El autor explica una técnica concreta para evaluar la intensidad de la correlación existente en las casillas de una tabla de contingencia que proviene de la intervención de dos variables. Al mismo tiempo que subraya el carácter novedoso del procedimiento expone un ejemplo concreto de aplicación de la técnica comparándola con el uso tradicional de los porcentajes dentro de una tabla de este tipo. Sugiere finalmente su uso sistemático en el examen de las correlaciones recíprocas entre todas las variables que aparecen en una encuesta.

INTRODUCCIÓN

El objetivo de este trabajo es dar a conocer un índice para *medir la intensidad de la asociación en las casillas* (cruces entre filas y columnas), de una tabla de contingencia. Este tipo de lectura es muy corriente en el caso de cualquier lector de cuadros de contingencia que trata de interpretar el significado de los valores que se encuentran en las casillas, es decir, los datos “locales” dentro de una tabla. El índice que aquí presentamos quiere contribuir a enriquecer las diferentes técnicas existentes para una lectura de

ABSTRACT

The author explains a definite technique to evaluate the degree of correlation in the cells of a contingency table derived from the interposition of two variables. At the same time he underlines the novelty of this process by giving an specific example and explanation of its application and by comparing it with the traditional use of percentages in a similar type table. Finally, the author suggests the systematic use of this technique in the interpretation of the reciprocal correlation between the different variables appearing in an inquiry.

esta clase de datos. Este índice se denomina PEM, o Porcentaje de Desviación Máxima, y ha sido propuesto por Philippe Cibois. Es la traducción del “Pourcentage de l'écart maximum” (PEM). (Cibois, 1993, 43 y siguientes). A la vez, nos ocupamos de mostrar el uso de este instrumento de medición en el procesamiento de datos provenientes de encuestas u otras observaciones similares¹.

¹ El estudiante Enrique Membreño Martínez y la Colega Ana Delia Ramírez Calderón leyeron el borrador de este trabajo, hicieron críticas y sugerencias muy valiosas. Sólo el autor asume la responsabilidad pero no sin agradecerles de manera especial.

La exposición se organiza de tal modo que el lector es invitado permanentemente a discutir lo que se observa en una tabla de números, con cuestionamientos constantes a lo que parece evidente, lo que finalmente imprime un tono polémico a la exposición. De este modo se quiere poner en evidencia la originalidad del procedimiento propuesto y al mismo tiempo justifica las indulgencias del lector especializado en estos temas y que ya tiene resuelta la elección de procedimientos.

1 LOS DATOS DE UNA ENCUESTA

A finales de mayo y principios de junio de 1996, la EMPRESA DATANALISIS llevó a cabo una encuesta telefónica acerca del consumo de licores. Para los efectos de esta exposición, carece de relevancia lo concerniente a la calidad muestral del sondeo. Vamos a suponer que los resultados obtenidos son perfectamente generalizables a una gran parte de la población, al menos la que tiene teléfono, si bien con márgenes conocidos de confianza y error. Nos concentramos únicamente en el momento del procesamiento de los datos, cuando se busca extraer constataciones válidas que vinculan más de una variable al mismo tiempo. La exposición subsiguiente se limita al examen de un cuadro bivariado que resulta del procesamiento de la encuesta referida. La pregunta que analizaremos se formuló en los términos siguientes "*En el caso suyo, ¿cuál de las razones enumeradas es la más importante para tomar?*" Las respuestas se hallan cruzadas con la variable "sexo", en la tabla 1. Observemos lo que ocurre.

TABLA 1
RAZONES PARA TOMAR BEBIDAS ALCOHÓLICAS
SEGÚN SEXO

	Frecuencias observadas (absolutas y relativas)					
	Hombres		Mujeres		total	
	N	%	N	%	N	%
Conversar con amigos	187	64,5	96	45,9	283	56,7
Diversión	29	10,0	45	21,5	74	14,8
Descansar y tranquilizarse	50	17,2	17	8,1	67	13,4
Aperitivo antes de comer	18	6,2	34	16,3	52	10,4
Ns/Nr	6	2,1	11	5,3	17	3,4
Mientras ve T.V.	0	0,0	6	2,9	6	1,2
Total	290	100	209	100	499	100

Un observador podría interpretar que en el caso de los hombres, éstos "toman más para conversar con los amigos (64,5%) que para descansar y tranquilizarse (17,2)", (Ortíz, 1996, 126), estableciendo así un orden de prioridades como existente en la realidad, ya que esa es la "razón principal" por la cual los hombres toman. (Ortíz, 1996, 124). Para el caso de las mujeres el orden de prioridades se expresaría así: "Toman más para conversar con sus amigos (45,9%) que por diversión y entretenimiento (21,5%)" (Tabla 1) (Ortíz, 1996, 126). Aunque, como veremos en este trabajo, esta observación es inexacta en ambos sexos, centremos la atención únicamente en el caso de los hombres. El observador ve datos obvios que cualquiera puede ver en la Tabla 1, pero se condena a una lectura de lo aparente. Hace constataciones que conciernen a los cruces entre filas y columnas desconociendo los límites del instrumento que utiliza (los porcentajes ordinarios). Por ello desemboca en observaciones erróneas.

Obsérvese que tomar para tranquilizarse o para conversar con los amigos constituyen dos opciones diversas desde el punto de vista sociológico. Se trata de hechos sociales de naturalezas diferentes. Tomar para conversar con los amigos indica un efecto de convivencia y de comunicación, en cambio tomar para tranquilizarse supone más bien un efecto farmacológico, como quien ingiere una pastilla para calmar la ansiedad. Las decisiones que se puedan adoptar según algunas de estas posibilidades son distintas. El "perfil del consumidor" de licores que se puede describir, no es igual en un caso o en otro. Las estrategias políticas, sociales, clínicas o comerciales que se adopten serían diferentes en un caso y en otro. Por lo tanto consideremos que el problema no es banal y necesitamos averiguar cuál es la realidad.

En verdad aquí va a demostrarse que no es cierto que el sol gire alrededor de la tierra. Es decir, si se permite la ironía, que es al revés. Los hombres no toman para conversar con los amigos en primer lugar sino que lo hacen mayoritariamente para descansar y tranquilizarse, a pesar de la evidencia que parecen entregar los porcentajes tradicionales.

2. LO QUE SE OBSERVA A PRIMERA VISTA

En primer lugar, examinemos la tabla 2. Es la misma tabla 1, pero en ella sólo se anotan los totales de las filas y los totales de las columnas. Lo que interesa estudiar es la clase de fenómeno que tiene lugar dentro de cada una de las doce casillas. Es decir, la *dirección* y la *intensidad* del vínculo entre las dos variables, tal como se presentan en la intersección entre una fila y una columna. Para esto hay que prestar atención especial a los totales de filas y de columnas. Es como ver al mismo tiempo dos cuadros univariados. En uno de ellos está la variable sexo: en la columna se observa que hay 290 (58%) hombres y 209 (41%) mujeres en total. Todos los casos que fueron entrevistados suman 499.

En el otro margen, el de fila, se encuentran los totales de respuesta a la pregunta. Casi el 57% de los informantes toman para conversar con los amigos. Ninguna de las opciones restantes alcanza el 15%.

TABLA 2
SOLO TOTALES FILAS Y COLUMNAS

DOCE CELDAS VACÍAS				
	Hombres	Mujeres	Total	% Tot:
Conversar con amigos	A	B	283	56,71
Diversión	C	D	74	14,83
Descansar y tranquilizarse	E	F	67	13,43
Aperitivo antes de comer	G	H	52	10,42
Ns/Nr	I	J	17	3,41
Mientras ven T.V.	K	L	6	1,20
TOTAL	290	209	499	100
% Tot:	58,12	41,88	100	

Los totales, tal como existen, ejercen una influencia poderosa en los resultados tanto si se miden en columnas porcentuales como en filas por ciento. Una elemental malicia nos haría sospechar que, al calcular los porcentajes convencionales dentro de las casillas ahora vacías, necesariamente los hombres tienen que tomar más para conversar que para descansar. Así en la casilla A, es de esperar un valor numérico y porcentual muy elevado ya que allí se reúnen las posibilidades que tienen mayor peso en todas las respuestas. En

cambio en las restantes casillas la concentración necesariamente tendrá que ser menor. Esto sugiere ya una previsión: la de que es posible que se produzca una alta concentración en la casilla A, capaz por lo menos de deformar nuestra percepción e impedirnos un análisis correcto de los datos locales, es decir, los que se registran en cada una de las casillas. Se verá más adelante cómo este efecto ilusorio se debe a la influencia que ejercen los totales de columna y los totales de fila tal como existen. Iniciemos el proceso de ruptura con esta ilusión.

3. CONSTRUIR UNA HIPÓTESIS

Para fracturar las apariencias es necesario tomar como punto de partida en toda la observación, esos mismos totales, respetándolos en todo momento, ya que constituyen el marco de referencia obligado para el examen.

La pregunta que nos hacemos es: ¿Cuáles números deberían existir en cada una de las casillas del Cuadro 2, en el caso en que hubiera independencia entre las dos variables consideradas?

Es necesario imaginar una realidad a partir de estos totales de filas y de columnas. Esta realidad imaginaria debe ser útil para evaluar el significado de los datos reales. Imaginaremos pues una realidad en la cual no exista ningún grado de asociación entre la condición masculina o la femenina y la elección de alguna de las razones para ingerir licor. La hipótesis equivale a postular una ausencia de vínculos entre los dos fenómenos.

El modo para hacer este cálculo en cada casilla consiste en multiplicar el total de fila por el total de columna correspondientes a cada casilla y luego dividirlo entre el total general. Es un procedimiento muy conocido en estadística. Con esto podemos construir un cuadro puramente teórico en el cual la distribución de las opciones y los sexos es "perfecta". Es un cuadro con números hipotéticos dentro de las casillas, pero con totales reales en filas y columnas. Son los datos "esperados" en nuestra hipótesis, lo que tendría forzosamente que existir como valor en cada casilla si no existiese ninguna vinculación o asociación *local* entre las dos variables. El procedimiento se ilustra como sigue:

	Modalidad A	Total
Modalidad B	Fe	TotB
	TotA	TOTAL

La fórmula (Vessereau, 1988, 95), muy conocida como parte del cálculo del X^2 , es :

$$Fe = \frac{\text{TotA} \cdot \text{TotB}}{\text{TOTAL}} \quad \text{Donde Fe = Frecuencia esperada o teórica}$$

Para saber cuántos hombres toman, hipotéticamente, para conversar con los amigos, debemos multiplicar 290, que es el total de la columna de hombres por 283, que es el total de los que toman para conversar con los amigos, y el resultado dividirlo entre 499 que es el GRAN TOTAL, es decir todas las personas que respondieron a la encuesta².

TABLA 3

TABLA DE INDEPENDENCIA

FRECUENCIAS TEÓRICAS

	Hombres	Mujeres	Total
	n	n	
Conversar con amigos	164,47	118,53	283
Diversión	43,01	30,98	74
Descansar y tranquilizarse	38,94	28,06	67
Apetito antes de comer	30,22	21,78	52
Ns/Nr	9,88	7,12	17
Mientras ven T.V.	3,49	2,51	6
TOTAL	290	209	499

Así, $290 \cdot 283 / 499 = 164$. Decimos entonces que "se esperaría" tener 164 hombres en la primera casilla. Para el caso de las mujeres que toman para conversar con los amigos la misma operación sería: $209 \cdot 283 / 499 = 119$ (También redondeado). Y así sucesivamente se procede a elaborar esta tabla "perfecta" (Tabla 3).

4. CUÁNTO OBSERVO Y CUÁNTO DEBERÍA OBSERVAR

Entramos en el segundo paso del proceso de evaluación de los datos. Ahora tomamos la tabla imaginaria y la usamos para medir el significado de la tabla real de datos observados. Es la actitud mental de quien dice, por ejemplo: "Yo debería tener 100 colones y

sólo tengo 20. Entonces: 20 menos 100 = -80. Por lo tanto tengo 80 colones menos de los que debería tener".

La operación consiste en hacer una resta: los colones que tengo menos los que debería tener. En este caso el individuo del ejemplo tiene un déficit: le faltan 80 colones. Pero si él tuviera 300 colones y debería tener los mismos 100, razona así: "300 que tengo menos 100 que debería tener = 200". En este caso el sujeto estaría muy contento porque tiene más dinero del que debería. Lo teórico, "imaginario", o "esperado", sirve de criterio para medir y comprender la realidad observable.

Apliquemos el mismo principio a la tabla de datos, siempre respetando los totales de fila y los totales de columna. En este caso hacemos también una resta para cada una de las casillas. Sustraemos a la Frecuencia observada (F_o), la frecuencia esperada (F_e). El resultado es la Desviación Observada (Do).

La formula es³: $F_o - F_e = Do$

Sustituyendo, para la casilla de los hombres que toman para conversar:

$$187 - 164 = 23.$$

La desviación (23 individuos), que es cantidad positiva, indica que entre ser hombre y tomar para conversar con los amigos existe una atracción, una correlación positiva, de 23: se deshecha la hipótesis de independencia.

Para el caso de los hombres que toman por diversión, se hace una operación equivalente: $29 - 43 = -14$.

En este caso se dice que hay un déficit. Entre ser hombre y tomar por diversión hay una relación de rechazo, pero no de independencia, pues la desviación es negativa. Si se procede de este modo con todas las casillas obtenemos la Tabla 4.

3 Observado menos teórico, es un componente muy conocido del X^2 , el cual tiene la siguiente fórmula: $X^2 = \sum \frac{(f_o - f_e)^2}{f_e}$

Donde: f_o = Frecuencia observada y f_e = frecuencia esperada o teórica.

2 En general se procura redondear las cifras, para facilitar la exposición.

TABLA 4
DESVIACIÓN OBSERVADAS

Frecuencias observadas-frecuencias técnicas	Hombres	Mujeres	Total
	Conversar con amigos	23	-23
Diversión	-14	14	
Descansar y tranquilizarse	11	-11	
Aperitivo antes de comer	-12	12	
Ns/Nr	- 4	4	
Mientras ven T.V	- 3	3	

La Tabla 4 es fundamental en toda nuestra exposición. Se trata de cifras que indican intensidad y dirección de una asociación en cada casilla, es decir, en la intersección entre filas y columnas. Este es el asunto que nos interesa y del que se ocupan todas las personas que por alguna razón deben leer tablas de contingencia.

En general se admite que hay tres posibilidades: a. Que el número sea positivo en una casilla: Significa que hay una asociación o atracción entre las dos alternativas que se concentran en dicha casilla. b. Que el número sea negativo: Significa que hay una relación de rechazo entre los dos eventos que se reúnen en esa casilla. (Ejemplo: ser hombre y tomar por diversión); y c. Que el número sea cero: en este caso la hipótesis de independencia se confirma, y se dice que no existe vínculo alguno entre los dos eventos unidos en la casilla.

Las desviaciones observadas (Do) están en el centro de la mayoría de los análisis de datos que se hacen en ciencias sociales cuando se comparan dos variables. Constituyen un ingrediente central, como punto de partida, en los programas de procesamiento de datos. (Paquete estadístico: TRIDEUX, versión 2.2 de Cibois)⁴.

4 El TRIDEUX consiste en un paquete estadístico, compuesto de una serie de programas informatizados, para el tratamiento de datos cuantitativos. En muchos de estos programas se utiliza el PEM, que es objeto de exposición en este trabajo. No disponemos de una traducción confiable del término TRIDEUX. En el Diccionario *Petit Robert* no existe la palabra. En español puede sugerir tridimensionalidad. Tri=ordenamiento; Deux=Dos. El paquete en su conjunto propone el análisis factorial de correspondencias múltiples y una serie de análisis post-factoriales.

La Tabla 4 nos dice que los hombres toman para conversar en primer lugar y en segundo lugar para tranquilizarse, pero ya la intensidad, la distancia entre estas dos opciones, no es la misma que arrojan los porcentajes simples⁵. Los porcentajes simples indican una inmensa distancia: como decía un observador de porcentajes ordinarios: los hombres "Toman para conversar con los amigos (64,5%) y para descansar y tranquilizarse (17,2%)" (Ya citado, Ortiz, 1996, 126). Ahora las cosas han cambiado. El orden de prioridades sigue siendo el mismo pero ya no es tan pronunciado, pues simplemente se duplica. Antes era una relación de 3,75 a 1 (Porque: $64,5 / 17,2 \% = 3,75$). Ahora es una relación más moderada: de 2 a 1 (Porque: $23 / 11 = 2,09$). La importancia que hemos otorgado a los totales de columna y a los totales de fila empiezan apenas a darnos sus frutos. El lente de observación empieza a ajustarse. Ya no vemos lo mismo que veíamos antes. Hemos desembocado en un conjunto de datos mucho más interesantes y mucho más fieles que el de los porcentajes simples, sean en fila o en columna. Más fieles porque evidencian de modo más exacto el tipo de fenómeno de asociación entre dos respuestas y la dirección que asume. De lo que trata esta exposición es del perfeccionamiento de este índice, instrumento fundamental en el TRIDEUX de Cibois.

5. UNA NUEVA HIPÓTESIS: LA FRECUENCIA MÁXIMA POSIBLE

A partir de ahora lo que existe es un conjunto de DESVIACIONES OBSERVADAS, (Do), es decir, reales. Más fieles a los datos originales si se les compara con los porcentajes comunes. No es aún una medición exacta ya que no resulta claro cuál es el máximo y cuál es el mínimo de la escala de medición. Conocemos la dirección de la asociación (atracción, rechazo, nulidad o independencia). Nos hemos aproximado además a una primera evaluación de la intensidad de dicha asociación (Hay cantidades variables, con signos positi-

5 El SPSS proporciona estos mismos resultados con el nombre de "residuos".

vos o negativos). Pero es claro que no podemos aún evaluar esta intensidad ya que no tenemos precisión en los límites máximos y mínimos de la escala utilizada. El PEM debe resolver esta carencia. La referencia es la Desviación Máxima, como se explicará luego. Para conocer esta Desviación Máxima primero necesitamos conocer la Frecuencia Máxima que cabe en cada casilla.

Nos preguntamos pues, en este momento, cuántas personas, como máximo, podrían existir teóricamente en cada casilla. Este esfuerzo de la imaginación tiene un límite, para que sea útil. El límite viene dado por los totales de fila y los totales de columna. Así, la respuesta es fácil (observando la Tabla 2): la mayor cantidad posible de hombres que podrían tomar para conversar con los amigos, es de 283 (Si todos los hombres y ninguna mujer hubieran respondido de este modo). La mayor cantidad de hombres que podría ingerir licor para descansar es de 67 (La misma condición).

En ambos casos ya anteriormente habíamos constatado que existía una relación de asociación o atracción, es decir la Desviación observada tenía signo positivo (Tabla 4). Para calcular el máximo imaginable en estos dos casos, bastó con adoptar el total de fila, de modo que siempre se respetó el total real de fila y de columna.

En el caso de los hombres que toman por diversión, casilla donde ya sabemos que existe una relación de rechazo, es decir, una desviación observada de signo negativo, la cantidad máxima posible imaginaria que podría existir es cero, si se mantiene el criterio de respeto a los totales de fila y de columna. Esto último se explicará más en detalle enseguida, de manera que este cero no parezca una arbitrariedad.

Resumamos y precisemos las reglas que conducen a descubrir la Frecuencia Máxima (FMAX). Existen tres posibilidades:

- a. *Que la asociación entre las dos modalidades sea positiva, según la Tabla de Desviaciones Observadas:*

En este caso hay que ubicar en la casilla correspondiente, el valor más elevado que

sea posible de acuerdo con los totales de fila y de columna existentes. Siempre debe escogerse el valor menor que existe en estos dos totales para no transgredir la estructura de los totales.

- b. *Que la asociación entre las dos modalidades sea negativa según la Tabla de Desviaciones Observadas. En general se pueden tener dos sub-variantes:*⁶

b1. Que uno de los totales (de fila o de columna) sea mayor que el complemento del total del otro margen: el valor de la casilla no puede ser cero

Ejemplo:

	Modalidad A	Otras Modalidades	Total
Modalidad B	0	?	TotB=7
Otras Modalidades	?	?	
	TotA=6	TOTAL=10	

Esta pequeña tabla tiene solamente 4 casillas: Conociendo los totales de la primera fila y de la primera columna, así como el total general, examinemos lo que ocurre en la primera casilla, intersección entre la primera fila y la primera columna. ¿Es posible el cero de la primera casilla? Si colocamos allí un cero, habiendo 6 individuos en la columna A, sólo podríamos tener 4 individuos en las otras columnas para que el TOTAL siga siendo 10. Pero esto es incompatible con el Total de la fila de la Modalidad B, que es 7. El complemento del TOTAL general (10 menos 7), es 3. El total de la Modalidad B (7) es superior a ese complemento.

El test aplicado:

Si $TOTB > (TOTAL - TOTA)$ y necesariamente $TOTA > (TOTAL - TOTB)$, entonces $FMAX =$ al número menor de las dos restas.

Sustituyendo: $7 > (10 - 6) = 4$ y necesariamente $6 > (10 - 7) = 3$. Siendo verdadera esta expresión, la Frecuencia Máxima es igual a tres, número inferior a cuatro. Nunca puede ser cero.

En una situación como ésta, el procedimiento de FMAX equivale a: $FMAX = TOTA - (TOTAL - TOTB) = TOTB - (TOTAL - TOTA) = TOTA + TOTB - TOTAL$.

6 Aquí seguimos el mismo ejemplo propuesto por Cibois, *ob cit.*, Pág 60.

b2. Que uno de los totales (de fila o de columna) sea menor o igual al complemento del total del otro margen: el valor de la casilla es cero.

	Mod A=Hombres	Mujeres	Total
B=Diversión	0		TotB=74
Otras Modalidades	?	?	?
	Tot A= 290		TOTAL=499

Examinemos la primera casilla de este pequeño cuadro que proviene de la segunda fila de la Tabla 5.

TotB > (TOTAL-TotA) y necesariamente TotA > (TOTAL-TotB). Sustituyendo: $74 > (499-290)=209$ y necesariamente $290 > (499-74)=425$. Siendo falsa esta expresión, TotB es por fuerza menor o igual al valor del complemento del total del otro margen.

TABLE 5
FRECUENCIAS MÁXIMAS IMAGINABLES
TEÓRICAS

	Hombres n	Mujeres n	Total
Conversar con amigos	283		283
Diversión	0		74
Descansar y tranquilizarse	67		67
Aperitivo antes de comer			52
Ns/Nr			17
Mientras ven T.V.			6
TOTAL	290	209	499

Luego, en la primera casilla debemos colocar un valor de cero. Procediendo de este modo se elabora la Tabla 5 (que se limita aquí a tres casillas para efectos de exposición)⁷. Es una tabla teórica, pero respetuosa, como ya vimos, de los límites impuestos en la tabla 4: de Desviaciones Observadas.

c. *Que la asociación entre las dos modalidades sea cero, es decir que se verifique la hipótesis de nulidad absoluta, según la tabla de Desviaciones Observadas*

En este caso de situación pura de independencia el PEM es también nulo. Esto resul-

ta porque la Desviación Observada en cada casilla es el numerador de la fórmula para obtener el PEM. Además, una desviación puramente nula no se encontrará de ordinario en el procesamiento de una encuesta.

6. UNA HIPÓTESIS MAXIMALISTA: EL MÁXIMO IMAGINABLE DE DESVIACIÓN

A partir de la tabla 5 es posible construir una tabla teórica o imaginaria nueva. Es una en donde identificamos las Desviaciones Máximas que sería posible haber esperado. Es la intensidad mayor de desviación que los datos nos autorizan a confeccionar.

La fórmula para hallar la desviación máxima (D_{MAX}) en cada casilla es: $F_{MAX} - F_e = D_{MAX}$.

Desembocamos así en la tabla 6. Es una nueva resta. A la Frecuencia máxima posible (hipótesis, Tabla 5), le restamos la frecuencia teórica, (Tabla 3: F_e), que también era una hipótesis. El resultado de una resta entre dos hipótesis necesariamente será una nueva hipótesis: la Desviación Máxima esperada (D_{MAX}).

TABLE 6
DESVIACIÓN MÁXIMA POSIBLE
FRECUENCIA MÁXIMA-FRECUENCIA TEÓRICA

	Hombre Max-Teo	Mujeres Max-Teo	Total
Conversar con amigos	119		
Diversión	-43		
Descansar y tranquilizarse	28		

Para los hombres que toman licor para conversar con los amigos tendríamos: 283 (Tabla5) -164 (Tabla3) y el resultado es 119 (redondeado). Así se procede para cada una de las casillas de la tabla. Lo más que podrían

Máxima: Ésta debe respetar los totales. Por ejemplo en la primera casilla en cuestión, el lector puede intentar una Frecuencia Máxima que no sea 283, sino 290. Se verá que esta cifra no se sostiene ya que no cabe en el total de la fila correspondiente. Es la regla que corresponde a la posibilidad "a" ya explicada.

7 Los totales de Filas y de Columnas en realidad corresponden a la Tabla 1. Se consignan en la Tabla 5 para hacer ver la coherencia de la Frecuencia.

atraerse entre sí la modalidad "hombre" con la práctica de tomar para conversar con los amigos es 119. Lo más que podría rechazarse esta misma condición con la práctica de tomar por diversión es de -43. Y lo más que cabría imaginar que la modalidad "hombre"⁸ se atraería con "tomar para descansar y tranquilizarse", es 28.

Hemos construido un cuadro maximalista de atracciones y rechazos. Este cuadro viene autorizado por los datos existentes ya que han salido de ellos. Lo que nos queda es ahora volver los ojos a la realidad a fin de saber cuanto de lo posible ideal, existe en lo observable.

7. ¿QUÉ PORCENTAJE DE LA HIPÓTESIS MAXIMALISTA ES OBSERVADO EN LA REALIDAD?

Cuánta desviación observada existe, es algo ya sabido, y lo encontramos en la tabla 4, que expresa la cantidad de desviaciones realmente existentes. La tabla 6 es irreal pero consistente. La tabla 4 es real. Lo que interesa ahora es saber cuánto tenemos en porcentaje de la tabla irreal 6. Esto equivale a encontrar el PEM de Cibois. Es una simple fórmula para extraer un porcentaje:

$$PEM = Do / DMAX \cdot 100$$

Es decir, Desviación Observada dividido entre Desviación Máxima y multiplicado por cien. Lo que obtenemos es un porcentaje. Una escala de medida con extremos máximos y mínimos.

O, lo que es lo mismo, para cada casilla: Tabla 4/Tabla 6·100

Es decir, de nuevo para el caso de los hombres que consumen licor para conversar con los amigos: $23/119 \cdot 100 = 19$

Hombres que toman por diversión: $-14/43 \cdot 100 = -33$

⇒ (Para llegar al -33: por convención, a fin de mantener el sentido negativo de la relación, se ha eliminado uno de los signos negativos, en este caso, el de -43, que proviene de la Tabla 6.)

Hombres que toman para tranquilizarse: $11/28.1 \cdot 100 = 39$

TABLA 7

PEM: PORCENTAJE DE DESVIACIÓN MÁXIMA (Desviación Observada/Desviación máxima)·100

	Hombres %	Mujeres %	Total
Conversar con amigos	19		
Diversión	-33		
Descansar y tranquilizarse	39		

Esta Tabla 7 contiene los Porcentajes de Desviación Máxima (PEM). Al procesar una encuesta tendríamos así un índice que se exhibe con números positivos y números negativos. Los positivos indican desviación positiva, es decir, atracción entre dos opciones de respuesta. Los negativos indican un rechazo entre dos opciones.

Sin embargo la lectura se facilita si eliminamos las cifras negativas, y convertimos todos los índices a cantidades positivas. Los resultados se estandarizan sumando cien a cada una de las casillas, con lo cual tenemos la tabla ocho.

Así: a: 100 correspondería a una relación de independencia; b: Una cifra mayor que 100, indica una relación de atracción; c: Menos que 100, hasta cero, indica una relación de rechazo. Todas las tablas de una encuesta pueden ser leídas rápidamente. La condición es abandonar el lenguaje clásico, en el que se examinan tablas de contingencia utilizando únicamente los porcentajes ordinarios. Con este nuevo índice ya no estamos leyendo cantidad de individuos en cada casilla, sino *intensidad* y *dirección* de la relación entre dos posibilidades de respuesta.

El resultado de la tabla 8 contradice al que se obtendría con porcentajes ordinarios clásicos, sean en fila o en columna.

8 Obviamente, en este trabajo no nos interesamos en discutir el problema de género que podría estar asociado a la exposición. El tema del cuadro que aquí se discute es sólo instrumental con respecto al objetivo del trabajo enunciado en el título.

TABLA 8
PEM+100

	Hombres	Mujeres	Total
Conservar con amigos	119		
Diversión	67		
Descansar y tranquilizarse	139		
Aperitivo antes y comer			

Ahora resulta que los hombres toman licor para buscar un efecto calmante de tipo farmacológico más que para una relación amistosa. El lente nos ha hecho ver otra realidad, la que se ocultaba en los datos a consecuencia del modo como se distribuyen los grupos: más hombres que mujeres, y mucho más que toman para conversar que para alguna de las otras razones.

Alguien podría pensar que este procedimiento es muy complicado y que es preferible seguir utilizando los porcentajes tradicionales para estudiar cada casilla de un cua-

dro bivariado. La tecnología contemporánea ha dejado atrás estas limitaciones. Por medios informáticos es posible extraer resultados de todas las variables de una encuesta y presentarlas ordenadas de modo que se facilite la lectura. El paquete estadístico TRIDEUX, de Cibois, ofrece esta posibilidad.

8. UNA EXPLICACIÓN SOBRE EL EFECTO ILUSORIO DE LOS PORCENTAJES CLÁSICOS

Los porcentajes tradicionales tienen una utilidad muy limitada y parcial. Mal entendidos, pueden hacernos decir falsedades. El efecto ilusorio proviene del modo como existen, en una tabla de contingencia, los totales de fila y los de columna. El modo como estos totales se distribuyen en cada una de las alternativas de respuesta, introduce un efecto de "sesgo" altamente perjudicial.

TABLA 9 1RA PARTE
RESUMEN GENERAL DEL PROCEDIMIENTO
(Para el caso de los hombres)

	Tabla 1	Tabla 3	Tabla 4	Tabla 5
	Frecuencia Observada	Independencia Teórica	Desviación Observada	Frecuencia máxima
Conversar con amigos	A	187	23	283
Diversión	C	29	-14	0
Descansar y tranquilizarse	E	50	11	67
PROCEDIMIENTO			T1-T3	

Si la dificultad proviene de los totales de fila y de columna, la solución que hemos expuesto también proviene de esos mismos totales, pero abordados desde una perspectiva muy diferente. La solución se encuentra allí donde se localiza el problema y no en otra parte. Observemos la tabla 9, (1ª y 2ª parte) en la cual se reúnen todas las tablas anteriores.

En efecto, en la fila A deberían encontrarse muchos más individuos de los que se observan, es decir, muchos más que 187, que es el valor de la Frecuencia Observada según la Tabla 1. Esta exigencia se justifica por la magnitud del total de esa fila, que es extremadamente elevada: 283 y que coincide con la Frecuencia Máxima: Tabla 5.

En cambio, en la fila E de la Tabla 1, también debería encontrarse un valor mayor que el observado. Pero, la distancia entre el observado y la frecuencia máxima es mucho menor en esta fila que en la A. En la fila A la desviación máxima es de 119, mientras que en la fila E sólo se llega a 28 (Tabla 6). El % de desviación máxima (PEM), de la fila E supera al de la fila A (Tabla 7). Por lo tanto la fila E registra una proximidad mayor a la expectativa, que la registrada en la A. Dicho de otra manera, sobre la fila A tenemos una exigencia mayor que sobre la E, debido a los totales de fila para ambos casos.

TABLA 9 2DA PARTE
RESUMEN GENERAL DEL PROCEDIMIENTO
(Para el caso de los hombres)

		Tabla 6	Tabla 7	Fórmula	Tabla 8
		Desviación Máxima	Porcentaje Desviación Máxima	de Tabla 7	Porcentaje + 100 Desviación
Conversar con amigos	A	119	19	$23/119 \cdot 100$	119
Diversión	C	-43	-33	$-14/43 \cdot 100$	67
Descansar y tranquilizarse	E	28	39	$11/28 \cdot 100$	139
PROCEDIMIENTO	T5 -T3	T4/T6·100	T4/T6·100	T4/T6·100	T7 + 100

Este fenómeno es precisamente el que los porcentajes clásicos nos ocultan. Ellos se limitan a indicar la existencia de una relación, pero no son capaces de medir la intensidad de la misma. Por este motivo la lectura de los porcentajes clásicos conduce a observaciones parciales y muchas veces inexactas.

9. UN EJEMPLO DE PROCESAMIENTO INTEGRAL DE LA ENCUESTA

Este modo de procesamiento de la encuesta es integral en el sentido de que con-

siste en la presentación simultánea de una gran cantidad de cuadros de contingencia cada uno con dos variables. En la tabla 10 se pueden observar cinco cuadros en una sola tabla y constituyen sólo un fragmento de las 35 preguntas de que consta la encuesta original. Comparar varios cuadros al mismo tiempo es costumbre normal entre los científicos sociales, pero no abundan los mecanismos informatizados que les faciliten esta tarea. El programa aquí utilizado —que hace parte del TRIDEUX— construye todos los cuadros posibles en toda la encuesta en un par de minutos.

TABLA 10
CRUCE MÚLTIPLE DE VARIABLES
SEGÚN PEM (FRAGMENTO)

Pregunta de la encuesta	Modalidad	Sexo	
		Hombre	Mujer
"La última vez que tomó"	No tomó ayer	75	125
	Tomó ayer	153	47
	Tomó semana pasada	71	129
	Tomó el último mes	85	115
"Frecuencia con que Ud. toma"	Casi diario	200	0
	Varias veces por semana	156	44
	Los fines de semana	140	60
	Cada 15 días	92	108
	Cada mes	87	113
	Otra frecuencia	76	124
"Primer licor que pide en una fiesta"	No toma nunca	69	131
	Cerveza	109	91
	Ron	92	108
	Whisky	140	60
	Vodka	44	156
	Guaro	87	113
Segundo licor que pide en una fiesta"	Ginebra	37	163
	Cerveza	76	124
	Ron	71	129

Pregunta de la encuesta	Modalidad	Sexo	
		Hombre	Mujer
"Razones para tomar"	Whisky	115	85
	Vodka	106	94
	Guaro	154	46
	Ginebra	0	200
	Nada	69	131
	Conversar con los amigos	119	81
	Descansar y tranquilizarse	139	61
	Diversión	67	133
	Aperitivo	60	140
	Ver T:V:	0	200

"Hacer hablar" a la tabla 10 es extremadamente sencillo. Vamos a proponer un ejemplo de lectura. El primer dato que llama la atención es que los hombres afirmaron haber tomado algún tipo de licor el día anterior a la entrevista mientras que las mujeres dijeron no haberlo hecho. Más bien las mujeres se situaron preferiblemente dentro del grupo de personas que afirma la semana pasada (129 contra 71), o haberlo hecho el último mes (115 contra 85).

Los que toman con mayor frecuencia son masculinos, más que femeninos. Esto puede afirmarse si se examinan las respuestas a la pregunta acerca de la frecuencia con que consumen licor. Las personas que toman a diario son hombres (200 contra 0 de PEM) y no mujeres. El mismo comportamiento se produce en los que toman varias veces por semana, allí se colocan los hombres más que las mujeres (156 contra 44), y los fines de semana la relación es de 140 contra 60.

El comportamiento se invierte a medida que las opciones de respuesta implican mayores lapsos de tiempo entre una ingesta y otra. Las mujeres toman cada quince días (108 contra 92) más que los hombres aunque en este caso las diferencias son menores. No obstante esas diferencias aumentan a favor de las mujeres cuando la ingesta es cada mes (113 contra 87). Finalmente, cuando las opciones son más distantes que un mes, es decir lo que equivaldría a un consumo muy esporádico, las mujeres superan con mucha amplitud a los hombres (124 contra 76) y lo mismo ocurre cuando se elige la opción de no tomar nunca. En este caso el resultado favorece a las mujeres (131

contra 69). Esto es lo mismo que afirmar que, en el caso de los hombres, hay un rechazo a consumir en lapsos de tiempo muy pronunciados.

En los datos disponibles se observa con bastante claridad que lo primero que tienden a pedir los hombres en una fiesta es el whisky (140), mientras que las mujeres tienden a no pedirlo (60, que equivale a un negativo). Al mismo tiempo que los hombres prefieren el whisky, las mujeres demandan ginebra, vodka, o guaro, en ese orden. Sin embargo, la inclinación femenina a solicitar Guaro como primera elección en una fiesta no es muy elevada (113), a la vez que el rechazo masculino a elegir guaro es importante: 87, que indica rechazo débil.

Cuando el informante se sitúa ante la posibilidad de que el producto de su primera elección ya se hubiese agotado, las mujeres piden ginebra, ron o cerveza, en ese orden, al mismo tiempo que los hombres inclinan sus preferencias por el guaro. El guaro, como segunda elección masculina, es indiscutible (154). Es decir que el guaro es la primera opción femenina, en cambio es la segunda opción masculina. Todo esto indica que el guaro tiene presencia en ambos sexos, tiene cabida en una situación festiva, pero no ocupa el primer lugar. El lugar que tiene reservado es el de una segunda alternativa para los hombres que en la primera pidieron whisky.

Por esto mismo no debe extrañarnos esta preferencia hipotética de las mujeres que piden guaro en una fiesta. Las fiestas son ocasionales en primer lugar. En segundo lugar la pregunta permite dejar libre la espontánea imaginación del informante. Que las mujeres

pidan guaro en una fiesta no indica que ellas sean consumidoras tan frecuentes como son los hombres. Esto refuerza la idea de que el guaro tiene una presencia moderada aunque persistente entre los consumidores, tanto a nivel de sus hábitos como de sus valoraciones culturales. Esto es muy claro en el caso de las mujeres como consumidoras con un perfil de moderación. Por esto mismo, podría afirmarse como hipótesis, que los consumidores de guaro, tienen como ideal el whisky. Y de esa manera es perfectamente comprensible que en una fiesta, en donde se les ofrece gratuitamente una bebida, elijan, sin pestañear, la bebida que culturalmente les ha servido para evaluar su propia bebida habitual. Es notable observar dos elementos adicionales que confirman la tendencia que coloca a los hombres como los principales consumidores. La elección del día y la situación más propicia.

Ahora bien, los hombres toman en primer lugar para descansar y tranquilizarse y en segundo lugar para conversar con los amigos, es decir que prefieren un efecto farmacológico en primer lugar y no un efecto de convivencia social. Mientras, las mujeres lo hacen en primer lugar cuando están viendo TV, en segundo, como aperitivo, y en tercero, como diversión, en ese orden de prioridades. Aquí nuevamente las mujeres se muestran más "moderadas" que los hombres. De este modo se evidencia que prácticamente carece de sentido afirmar que las mujeres: "Toman para conversar con sus amigos (45,9%) y por diversión y entretenimiento (21,5%)" (Ortíz, 1996, 126).

10. COMENTARIO FINAL

Hemos tratado de poner en evidencia la utilidad de un procedimiento que puede ayudar a leer de manera más objetiva unos datos cuantitativos que de algún modo reflejan fenómenos sociales. Pero, dentro una encuesta particular, no hemos hecho más que mostrar un ejemplo entre muchos posibles. Así como se ha hecho con la variable "sexo", puede hacerse con cualquier otra variable de la encuesta. El resultado es lo más aproximado posible a lo que en lenguaje corriente se entiende como el "perfil" de un consumidor. Tal perfil puede constituirse combinando lec-

turas de toda la encuesta, a partir de variables como sexo, edad, ingreso, ocupación, etc.

En los programas del TRIDEUX el PEM se utiliza también para medir la intensidad de la asociación de la tabla completa. Hemos prescindido de este modo de uso para centrarnos en el estudio de la intensidad de la relación que se localiza en cada una de las casillas de una tabla. Gracias a los programas informatizados del TRIDEUX, es considerable el ahorro de tiempo en el procesamiento de los datos. El largo proceso que se ha explicado en este artículo se realiza, para todos los cruces de variables, en unos pocos minutos para encuestas de más de 1200 casos con el único límite de que las modalidades u opciones de respuesta para cada variable no superen la cantidad de 35 en cada variable o pregunta.

Las implicaciones sociológicas de las observaciones que hemos puesto aquí como ejemplo, no nos ocuparon en el presente trabajo. Nos limitamos exclusivamente a exponer el procedimiento y las ventajas ligadas a este índice.

Esta alternativa de procesamiento tiene un enorme interés para la investigación sociológica, o para los estudios de mercadeo, o cualquier otro tipo de búsqueda que pretenda ir más allá de la mirada superficial de los datos. Otro elemento de interés es la rapidez con que es posible procesar la encuesta y producir unos resultados que verbalizan las observaciones a partir de los cuadros de salida.

BIBLIOGRAFÍA

- Cibois, Philippe. *L'analyse factorielle*. Presses Universitaires de France, París, 1987.
- Cibois. *L'analyse des données en sociologie*. PUF, París, 1984.
- Cibois, Philippe. "Le PEM, Pourcentage de l'écart maximum: un índice de liaison entre modalités d'un tableau de contingence". *Bulletin de Methodologie Sociologique*, No 40, September, 1993, p.p. 43-63.
- Glass, Gene V. y Stanley, Julian C. *Métodos Estadísticos Aplicados a las Ciencias Sociales*. Prentice-Hall Hispanoamericana SA, México DF, 1988.

Hoel, Paul G. *Estadística elemental*. CECSA, México DF, 1976.

Madrigal, Jhonny y Schifter. *Primera Encuesta Nacional sobre SIDA: Informe de Resultados*. Asociación Demográfica Costarricense, San José, 1990.

Ortiz, William. "Análisis y Propuesta. Plan de Mercado, FANAL". *Mimeo*, San José, 1996.

Rojas Soriano, Raúl. *Guía para realizar Investigaciones sociales*, UNAM, México DF, 1987.

Vessereau, André. *La Statistique*, Presses Universitaires de France, París, 1988.

Abelino Martínez Rocha
Apdo 1536 - 2100
Guadalupe
E-mail: mrocha@cariari.ucr.ac.cr.