

# El análisis de Paul Lazarsfeld en las clases de Alfredo Errandonea

Verónica Filardo \*

*El artículo refiere al modelo Lazarsfeld, una técnica que Alfredo Errandonea enseñó ampliamente en las clases de análisis multivariado, en su Cátedra de Metodología III de la carrera de Sociología. Un valioso material de apoyo docente junto a la profunda reflexión metodológica e interpretativa que caracterizó a la docencia del homenajeado.*

Agradezco la invitación a participar en una revista que se publica en homenaje a Alfredo Errandonea, para escribir en referencia a una técnica que él enseñara en sus cursos: el modelo de Lazarsfeld. Dada la consigna, este artículo se propone dos objetivos: El primero es difundir en un formato más formal que sus tradicionales y excepcionales “guías de clase” que aún se utilizan, una tipología construida por Alfredo Errandonea, y que enseñara en sus clases de análisis multivariado en su cátedra de Metodología III de la carrera de Sociología. Básicamente pretende entonces que su circulación deje de ser exclusivamente las fotocopias de sus guías escritas en la Olivetti, y con las gráficas dibujadas de su puño y letra. Estas guías más allá de constituir un apoyo docente, son el producto de su reflexión metodológica y de su preocupación constante de compartir éstas con los estudiantes.

*“Durante varios años de docencia de Metodología y Técnicas de Investigación en la Argentina (carreras de C. Política y de Sociología de la Universidad del Salvador -1976/1984’ y de sociología de la Universidad de Buenos Aires - desde 1985-, principalmente, en la organización y dictado de las clases, fui trabajando un desarrollo interpretativo de este instrumental de análisis multivariado. Al punto de que circulan versiones de apuntes de clase con fines docentes en una Guía de Clase (Errandonea 1982 última versión 1987). (Errandonea, RCS)*

El segundo es rendirle un homenaje, a partir de este modesto propósito a su larga y fértil labor docente, que desarrolló en varias universidades de América Latina, básicamente - aunque no sólo - en el área de la metodología de la investigación social. Errandonea fue docente de muchas generaciones de estudiantes de ciencias sociales, que tuvimos la suerte de tenerlo de profesor. Algunos de nosotros pudi-

mos incluso empezar nuestra carrera docente con él, como ayudantes o asistentes en sus cursos. No sólo aprendimos entonces los contenidos de esas materias. Alfredo dedicaba mucho tiempo a reuniones de coordinación de los equipos de trabajo y discutía con nosotros criterios de corrección de las evaluaciones, el armado de los contenidos de los cursos, distribuía las clases, sostenía largas charlas acerca de los procesos de aprendizaje de los estudiantes, etc., cuestiones todas que respondían a una práctica que tenía incorporada: la formación docente de los que nos incorporábamos recién en la tarea. El impulsó y lideró la conformación del área de metodología en el Instituto de Ciencias Sociales (hoy Departamento de Sociología), con el objetivo de mejorar la formación y optimizar los procesos de aprendizaje de los estudiantes integralmente, intentando favorecer la coordinación de las distintas cátedras involucradas. Es obvio que su trayectoria en la política universitaria, excede en mucho estas iniciativas referidas estrictamente a la docencia en metodología (una de los tantos aspectos que manejaba simultáneamente) pero a la que no haremos referencia ya que serán abordados en otros artículos aquí presentados.

## Paul Lazarsfeld

Paul Lazarsfeld nació en Viena en 1901, y allí se licenció en matemáticas (1925). En su juventud militó en el Partido Socialista, sus padres eran socialistas activos y su casa era un punto de encuentro cultural y político en una Viena constituida en centro intelectual y científico de relevancia europea. Ese contexto estimula a que Lazarsfeld “aun estudiando matemática pura (...) se interese más tarde por los problemas de la sociología y la psicología”. (Picó, 1998:54)

\* Profesora Adjunta del Departamento de Sociología, Master en Sociología, Facultad de Ciencias Sociales, Universidad de la República — veronica@fcsoc.edu.uy

En esta etapa, algunos de sus referentes son Alfred Adler, (psicólogo) con quien desarrolla un análisis estadístico de test mentales., y el empiriocritismo de Mach en física, reforzado por la tradición positivista austríaca representada por el Círculo de Viena que funcionan como su filosofía de referencia. Trabajando con los Bühler, incursiona en psicología social aplicada a la juventud de clase media, perfeccionando los métodos de investigación. En 1927, habiendo antes sido profesor de liceo en el 1927 entra a la universidad como profesor asistente de psicología social y psicología aplicada y estadística. En 1928 es responsable de un Instituto propio de psicología aplicada. Allí escribe el volumen de estadística para psicólogos, donde compiló no sólo las herramientas estadísticas y metodológicas sino que presenta algunas de sus primeras investigaciones en el área. Es a partir de este instituto, que Horkeimer y la escuela de Frankfurt le encargan la parte empírica en Austria de su estudio sobre "Autoridad y Familia" en colaboración con Erich Fromm. (Picó, 1998)

No obstante el punto de inflexión en la vida intelectual y política de Lazarsfeld se ubica en el momento en que emigra a los Estados Unidos, con el apoyo de una beca de la Fundación Rockefeller (1932), en un momento difícil por la emergencia del nazismo, en particular para el mundo intelectual y científico.

Allí se conecta con Robert Lynd, profesor de sociología de la Universidad de Columbia y con su ayuda obtiene un puesto en la Universidad de Newark. En 1936 la Universidad crea un centro de investigación y lo nombra director. Horkheimer lo contrata en su Instituto. Es así que desarrolla una serie de investigaciones en un contexto en que la sociología americana tiene la impronta de las universidades de Columbia y Chicago. Por otro lado el New Deal que articulaba investigaciones científicas y policy makers, funciona como un contexto favorable para el desarrollo de las actividades enmarcada en la sociología empírica, de investigaciones aplicadas sobre todo en ciencias sociales. En 1937 Lazarsfeld es nombrado director del Office of Radio Research, de la Fundación Rockefeller que en 1941 se traslada a la Universidad de Columbia, en intensa colaboración con Merton. En 1944 el Bureau of Applied Social Research se integra la Universidad y el empirismo tiene un ascenso considerable, incluso en el campo universitario, que generó una estrecha conexión con el ámbito privado Lazarsfeld aporta en los años sesenta la mitad del presupuesto de la Universidad de Columbia que proviene de contratos de investigación (Picó, 1998).

"A partir del fin de los años cuarenta el Bureau of Applied Social Research y la Universidad de Columbia fueron los centros más prestigiosos de estados Unidos y a él llegaron toda una serie de generaciones de sociólogos. Mientras la sociología de Parsons

conoció un declive a partir de fin de los años cincuenta, la sociología de Lazarsfeld, el Bureau of Applied Social Research y la universidad de Columbia impusieron su hegemonía mucho más allá de 1960." (Picó, 1998)

La lógica de investigación aplicada y empírica que promueve Lazarsfeld se extiende a nivel internacional, y se expande el "avance técnico" americano.

### El análisis de Lazarsfeld

Lazarsfeld desarrolla un modelo de análisis multivariado para variables cualitativas, entre otros, y es éste el que se trabajará aquí. La función principal de este modelo es estudiar como inciden tercera o terceras variables en una relación bivariada.

Errandonea, incorpora en sus cursos este modelo de análisis y escribe las guías de clase a que hacíamos referencia antes, para apoyar su dictado. En consecuencia se convierte en un difusor de esta técnica en América Latina, aplica este modelo en las investigaciones que realiza, a pocos años de distancia de su desarrollo. En consecuencia y por su intermedio, las generaciones de estudiantes que él forma, tienen acceso a esta técnica de análisis de datos. Hoy a pesar de la existencia de nuevos desarrollos técnicos y metodológicos más sofisticados para el análisis multivariado de datos, esta técnica continúa en los contenidos de los cursos de la licenciatura de sociología, en gran medida por la claridad y simplicidad de la lógica de control de variables que presenta, lo cual permite captar con sencillez la noción del control expost de variables, resultando de gran utilidad desde el punto de vista pedagógico. El propio Errandonea justifica la utilización de esta técnica:

*"Más allá de las modas me sigue pareciendo altamente reivindicable para la mayor parte de los análisis que hacen los científicos sociales cuando se manejan con un número limitado de variables el clásico procedimiento multivariado de control por parciales, tan utilizado desde los tiempos de Durkheim, que Lazarsfeld, formalizara a partir de 1946 (Lazarsfeld, en Lazarsfeld, y Rosemberg, 1955, Lazarsfeld, and Lerner, 1959). Porque se trata de un procedimiento que permite discriminar los papeles lógicos de cada variable, hasta donde ello es posible hacerlo estadísticamente, cualquiera sea la naturaleza de las variables y de las relaciones. Y porque la artesanidad que la elaboración requiere e implica, lejos de constituir un inconveniente, le otorga la ventaja de consistir en una lectura mucho más analítica de los datos, que entrega observaciones desagregadas, con mayor elocuencia descriptiva de las relaciones involucradas. Y éstas, a partir de ciertas asunciones teóricas, posibilitan las inferencias explicativas que procura el investigador". (Errandonea, fecha:43)*

Alfredo Errandonea no es simplemente un "divulgador, o difusor" del modelo de análisis que desarrolla Lazarsfeld, sino que avanza en el análisis que dicho autor propone, elaborando una tipología de casos empírica y lógicamente posibles, - en los que Lazarsfeld no profundiza -, los que llama "tipos de elaboración mixta".

Este artículo pretende justamente exponer este "plus" que Errandonea adiciona al modelo desarrollado por Lazarsfeld, extremadamente útil para el trabajo empírico. Por lo tanto, y básicamente sólo organiza lo que el propio Errandonea expone en sus guías de clase, y presenta un ejemplo, tal y como se presenta en los cursos de Metodología III.

### 1. El análisis multivariado para variables cualitativas propuesto por Lazarsfeld

Este tipo de análisis de datos corresponde a la estadística no paramétrica, es decir aquella que trabaja con variables no intervalas. Introduce la posibilidad de trabajar con más de dos variables simultáneamente, y **se propone estudiar como afecta una relación bivariada otra u otras variables a dicha relación**

En sus guías de clase Errandonea anota:

"En un trabajo originariamente presentado en 1946 al Congreso de Cleveland de la Sociedad Americana de Sociología, y que luego se convirtió en texto obligado para todo el mundo para la técnica de análisis multivariado, Lazarsfeld se propuso desarrollar una metodología para el análisis de más de dos variables no intervalas que hiciera las veces de la correlación parcial del análisis paramétrico. Tenía el valioso an-

tecedente del tratamiento clásico trivariado que Durkheim realiza en "el suicidio", en el cual notoriamente se inspiró. La propuesta de Lazarsfeld se ha revelado -a 4 décadas de su formulación- como un procedimiento sumamente eficaz para el tratamiento de variables como las que manejamos en las ciencias sociales, que entrega bastante información más que la propia correlación parcial como para una elaboración artesanal de la interpretación" (Errandonea, s/f)

Mediante este modelo, lo que se busca es determinar la influencia de una tercer variable en una relación entre dos variables (la relación original), y en este sentido, lo que debe hacerse es "neutralizar sus efectos". En definitiva se obtendrá la relación bivariada con una medida más depurada, si se consiguen aislar los efectos de la variable de control de la relación original.

#### La incorporación de una tercera variable:

"Puede imaginarse la incorporación de una tercera variable a un cuadro que contiene dos, según lo muestra la Fig. 7: en el cuadro I aparece un cuadro bivariado (variables "X" e "Y") con sus celdas, a las que denominamos "a", "b", "c" y "d". Cada una de ellas puede ser desagregada (desdoblada) en función de una tercer variable ("t"), con lo que "a", se convierte en "a'" ( $II_1$ ) y en "a''" ( $II_2$ ), según tenga "t<sub>1</sub>" o "t<sub>2</sub>"; y así sucesivamente con el resto de las celdas originales ("b", "c", y "d"). La suma matricial de  $II_1$  y  $II_2$  recompone el cuadro original bivariado (III). En los cuadros II se ve la relación originaria bivariada al interior de los valores de "t", es decir "controlada" por "t". Esta es la idea en la que se basa el modelo Lazarsfeld de Análisis Multivariado.

I			II <sub>1</sub> t <sub>1</sub>			II <sub>2</sub> t <sub>2</sub>			III		
	x <sub>1</sub>	x <sub>2</sub>	x <sub>1</sub>	x <sub>2</sub>		x <sub>1</sub>	X <sub>2</sub>		x <sub>1</sub>	x <sub>2</sub>	
y <sub>1</sub>	a	b	a'	b'	a'+b'	a''	B''	a''+b''	a'+a''	b'+b''	a'+a'' +b'+b''
y <sub>2</sub>	c	d	c'	d'	c'+d'	c''	D''	c''+d''	c'+c''	d'+d''	c'+c'' +d'+d''
	a+c	b+d	a'+c'	b'+d'	n <sub>1</sub>	a''+c''	B''+d''	n <sub>2</sub>	a'+c'+ a''+c''	b+d' b'+d''	n

Donde se deduce que I = III

### La ecuación de Lazarsfeld

La ecuación que propone Lazarsfeld es la siguiente:

$$XY = (xy, t_1) + (xy, t_2) n_2/n + (xt) (ty)$$

"Esta fórmula dice que la relación originaria entre las variables "X" e "Y" (R0) es igual a la **suma ponderada** de la relación "xy" en el valor "t<sub>1</sub>" de la variable t (Primer parcial, P1: "xy, t<sub>1</sub>"), más la relación "xy" en el valor "t<sub>2</sub>" de la variable t (Segundo parcial, P2: "xy, t<sub>2</sub>"), más el producto de la relación divariada "Xt", (subtotal) por la relación divariada "tY" (marginal)".

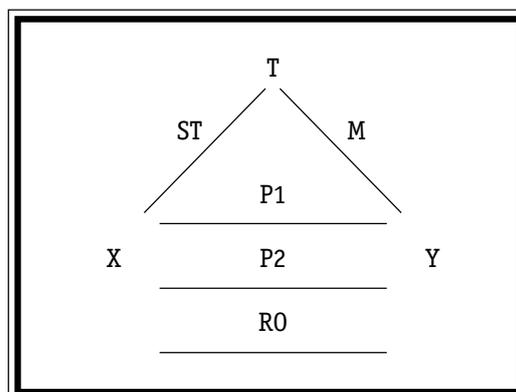
La ponderación supone que el primer parcial esté multiplicada por el coeficiente  $n_1/n$  y el segundo parcial por  $n_2/n$ . Es obvio que  $n_1 + n_2 = n$ . Puesto que  $n_1$  es el total de observaciones que pertenecen a  $t_1$  y  $n_2$  es el total de observaciones que pertenecen a  $t_2$ . La suma de ambas es el total de observaciones (n).

#### Si no se cumple la igualdad en la ecuación

"Lazarsfeld se interesa exclusivamente por la consideración analítica de cada término y no por la fórmula en general. (...) Si la igualdad que expresa la ecuación no se cumple, debe presumirse distorsión en alguno o algunos de sus términos. Ella puede obedecer al carácter espurio, por ejemplo, de algunas de las relaciones que constituyen términos de la ecuación. Si así fuera existiría alguna variable fuera del modelo que en realidad debería haber sido incluida en él, y que por alguna razón no visible con su exclusión, sólo se hace presente a través de su espuriedad. En este caso (no cumplimiento de la igualdad en la ecuación) puede afirmarse que el conjunto de las variables consideradas - y por lo tanto, el modelo puesto a prueba es inadecuado<sup>1</sup>. Pero esta prueba de adecuación es asimétrica: tan sólo descarta modelos inadecuados, pero no permite presumir adecuación en modelos no descartados. Por lo tanto su eventual utilidad práctica es demasiado relativa". (Errandonea, RCS)

"Si no hay variables externas que expliquen, la igualdad es aproximadamente cierta pero en el análisis de Lazarsfeld en realidad la utilidad de la fórmula está en que se opere con cada uno de sus elementos para interpretarlos". (Errandonea, guías de clase)

Como hemos visto, la conformación de las cinco relaciones bivariadas que es posible obtener a partir de un cuadro trivariado puede ser representado de la siguiente manera:



**Y** es el fenómeno que se quiere explicar (variable dependiente) **X** es la variable independiente (a partir de la cual se pretende explicar **Y**) y **t** es la variable de control (de la cual se quiere estudiar el papel que cumple en la relación **XY**) y que "se hipotetiza como no relacionada, interviniente, antecedente, co-determinante, verdaderamente determinante)

#### Interpretación del modelo Lazarsfeld

Como hemos visto antes, la ecuación de Lazarsfeld sostiene que la relación originaria puede descomponerse en la suma ponderada de los parciales y el producto de los marginales. Estos términos son los que requieren consideración analítica por parte del /la investigador /a, para la interpretación del modelo. Por tanto están en juego (en el caso de que la variable de control sea dicotómica<sup>2</sup>), 4 relaciones: el primer parcial (**xy, t<sub>1</sub>**), el segundo parcial (**xy, t<sub>2</sub>**), el subtotal (**Xt**) y el Marginal (**tY**).

Errandonea presenta el siguiente esquema básico para la interpretación:

1 Podríamos también decir "el modelo no está bien especificado".

2 Recordemos que hay tantos parciales como categorías tenga la variable de control

ELABORACION POR MARGINALES			ELABORACION POR PARCIALES							
$xy = 0$			$xy > xy$ $xt \cdot ty \neq 0$	$xy = xt \cdot ty$		$xy = p_1 + p_2$ $p_1 < xy < p_2$ $(p_1 \neq 0 \text{ y/ó } p_2 \neq 0)$ $p_1 \neq 0 \text{ ó } p_2 \neq 0$ $p_1 = 0 \text{ ó } p_2 = 0$	$xy = p_1 + p_2$ $xy = p_1 = p_2$	$xy > p_1 + p_2$ $p_1 = p_2$ $p_1 \neq 0 \text{ y } p_2 \neq 0$	$xy > p_1 + p_2$ $p_1 = 0 \text{ y } p_2 = 0$	$xy = 0$ $p_1 \neq 0 \text{ y } p_2 \neq 0$ (uno "+x" y otro "-x")
$xt = 0$ $ty = 0$	$xt = 0$ $ty \neq 0$	$xt \neq 0$ $ty = 0$								

3 Todo este apartado es transcripción textual del anexo de la Guía de clase N° 9 de Alfredo Errandonea, s/f

### Hacia una tipología interpretativa

No obstante, más adelante él realiza un "desarrollo lógico del modelo Lazarsfeld, en base a las magnitudes de las relaciones, que se transcribe íntegra en este apartado.

*"En base a la consideración de la igualdad de la fórmula de Lazarsfeld y teniendo en cuenta las magnitudes de cada una de las relaciones (originaria, parciales, subtotal y marginal), he desarrollado una tipología de casos lógicos posibles. El precedente apartado sólo plantea algunos casos más habituales sin hacer hincapié en las magnitudes de las relaciones." (errandonea, s/f)*

$$(XY) = (xy, t_1) \frac{nt_1}{n} + (xy, t_2) \frac{nt_2}{n} \dots \dots \text{donde} \dots (XY) = (xy, t_1) = (xy, t_2)$$

### Los principios para un desarrollo tipológico<sup>3</sup>:

A. Errandonea escribe un anexo en las Guía de clase N° 9, en donde expone en detalle los fundamentos de la tipología que construye y que es nuestro objetivo, documentar aquí. A estos fines se transcribe textualmente dicho anexo.

#### 1) Elaboración por parciales

Si el modelo no es inadecuado, la suma ponderada de los parciales es igual o inferior a la medida de la relación original. No puede superarla y sólo la iguala si la variable de control no juega algún papel explicativo (especificación, de intermediación, co-determinación, etc.) que se manifieste en las relaciones marginales, en cuyo caso, si efectivamente no juega ese papel, el último término de la fórmula de Lazarsfeld se anularía al convertirse en '0'.

Si hay mera especificación pura, sin papel explicativo propio en ningún grado, el último término de la fórmula se anula (los dos o uno de los dos coeficientes de él, consisten en "cero"), o si "t" nada tiene que ver con la relación (ambas relaciones originales son "0" y los parciales son similares entre í) la suma ponderada de los parciales será igual a la relación originaria.

Pero, como se acaba de sugerir, esta igualdad puede cumplirse de dos maneras

En primer lugar, porque alguna (o algunas) de las relaciones parciales es superior a la originaria en el grado en que compensen lo menor que lo es (o lo son) la otra (u otras)...

$$(XY) = (xy, t_1) \frac{nt_1}{n} + (xy, t_2) \frac{nt_2}{n} \dots \dots \text{donde} \dots (xy, t_1) < (XY) < (xy, t_2)$$

En este caso, "t" está especificando la relación originaria. La hace más débil en alguno de sus valores, y mas fuerte en algún otro (o algunos otros); la acentúa o la atenúa. En el límite, esta especificación puede convertirse en condición, si alguno (o algunos) de su (s) parcial(es) se convierte en "0"; pero ello es difícil que ocurra sin que "t" tenga relación alguna con "Y" (también lo es si la especificación de "t" es fuerte).

En segundo lugar, la igualdad también puede cumplirse si todas las relaciones parciales son iguales o similares a la relación originaria:

Aquí, en cambio, "t" se revela como irrelevante en la relación originaria entre "X" e "Y". Sus parciales "no acusan" de ninguna manera su presencia.

Sin embargo habría una tercer manera de cumplirse esta igualdad de la suma ponderada de parciales con la relación originaria. Es cuando ella se opera algebraicamente: la relación originaria es "0" o próxima y cada parcial tiene signos contrarios, de manera tal que se compensan. En este caso la variable original constituye un cambio de contextos que implica una alteración del sentido de la relación originaria, lo que le hace a esta globalmente encubierta, no visible a primera vista al análisis bivariado (si fuera más significativa) y en todo caso es su lógica.

Por supuesto que si los marginales tienen valores, y el último término es distinto de "0" - siempre y cuando el modelo no sea inadecuado-, entonces la suma ponderada de los parciales será inferior a la relación originaria. Pero allí se reproducen las variantes de la comparación entre los parciales que hemos visto (más algunas más) y en ellas juega la lógica considerada. La diferencia estriba que para la interpretación de lo que ocurra en los parciales tiene relevancia lo que ocurra en los marginales. Es una suerte de "elaboración mixta" en la que juegan los parciales, marginales y sus magnitudes relativas.

Antes, sin embargo, corresponde considerar la lógica de la incidencia de los marginales y sus pesos cuantitativos, en lo que se ha dado en denominar "elaboración por marginales".

## 2) Elaboración por marginales

Continuando con el supuesto que el modelo no es inadecuado, si concentramos ahora nuestra atención en los marginales (o marginales y subtotales como también suelen llamarse) el producto de ellos solo puede ser igual o inferior a la relación originaria. Sólo será igual si ambos parciales son iguales a cero: como la ubicación de los parciales en la ecuación (formula general) es en suma, basta que uno sea diferente de cero para que el producto de los marginales tenga que ser inferior a la relación originaria. Entonces, lo mas habitual es que sea inferior a ella. Y si no lo es en suficiente grado como para que la relación "tY" también lo sea - no olvidar que el producto entre fracciones de 1 reducen ambos términos-. Habrá que controlar la eventualidad de que "t" no sea la verdadera determinante, girando la relación.

Por lo tanto tenemos dos situaciones básicas:

- |    |   |
|----|---|
| a) | $(XY) = (Xt).(tY) \dots$ en cuyo caso: $(xy,t_1) = 0 \dots y (xy, t_2) = 0$         |
| b) | $(XY) = (Xt).(tY) \dots$ en cuyo caso: $(xy,t_1) \neq 0 \dots y/o (xy, t_2) \neq 0$ |

En el caso b)

$$(Xt).(tY) = (XY) - [(xy,t_1) \left(\frac{nt_1}{n}\right) + (xy,t_2) \left(\frac{nt_2}{n}\right)]$$

En la primera de las situaciones (a) la intervención de "t" diluye totalmente la relación originaria detectada -nivel bivariado-, Y ello puede deberse a dos razones:

En primer lugar, puede ocurrir que la relación originaria bivariada se debía que ambas variables principales son efectos de "t", que es su verdadera determinante. Al exponerse a la relación aparente bivariada "XY" a los parciales de "t", en ellos desaparece. Evidencia su naturaleza espúrea. Es el clásico caso de "espuriidad" de tradición durkheimiana, el del rí-sueño o absurdo ejemplo de la existencia de cigüeñas y las tasas de natalidad. ...

En segundo lugar, este resultado también puede deberse a que toda la relación originaria pase a través de "t". En este caso "t" es la variable intermedia-ria, o sea, -y no es un juego de palabras-, que "t" es la variable independiente de la variable dependiente "Y" y a la vez es variable dependiente de la variable independiente "X". El que toda la varianza de Y que explica X pase a través de "t" hace que cuando "XY" sea controlada por "t" la relación "XY" de los parciales de "t" desaparezca.

El que sea uno u otro caso es una cuestión teórica. Sin embargo hay algún indicio empírico que puede colaborar en a interpretación a partir de las magnitudes.

En efecto: para el segundo caso que es de INTERMEDIACION se requiere que ambos marginales (o marginales y subtotales) sean individualmente por lo menos similares a la relación originaria si es que ellos sucesivamente van a conducir la determinación que "x" tiene sobre "Y". Por la misma razón que dos cables que sucesivamente tienen que transferir una determinada corriente eléctrica cada uno de ellos debe ser capaz (tener la dimensión suficiente) para transmitir el total de la corriente que sucesivamente conducirán, so pena de fundirse cualquiera de los dos que no cumplan esa condición. En ese caso, por lo tanto, la magnitud estará entre la de "XY" para la menor de las marginales, como mínimo, y la igualdad que su producto debe arrojar con relación a "XY".

Pero esta condición - la que la menor de las marginales no sea inferior a "XY" no es suficiente para definir la **INTERMEDIACION**, aunque sí es necesaria. Se trata de una contribución asimétrica a la interpretación; aunque su incumplimiento hace descartar la INTERMEDIACION su cumplimiento no prueba que sea este el papel de "t". Máximo si se tiene en cuenta que para la otra alternativa, la de la **ESPURIEDAD**, el requerimiento de la igualdad del producto de los marginales con la relación originaria y el propio hecho de que la determinación en ambos casos debe ser lo suficientemente fuerte como para hacer que exprese relación bivariada "XY", supondrá muchas situaciones en que AMBOS marginales superen individualmente a la relación original.

En la segunda de las situaciones planteadas como básicas en este párrafo - que será mas frecuente -, el producto de los marginales es menor a la relación originaria. Vale decir que algo de ella, por lo menos se "retiene" en los parciales. Por lo tanto, ni la explicación determinante se debe íntegramente a "t", (como en el caso de espuriidad) ni toda la explicación de "X" sobre "Y" pasa por "t". (como en el caso de la intermediación). Aquí habría alguna intervención de "t", que no da cuenta mas que de una parte de la relación originaria: la de la proporción que constituye el producto de los marginales sobre "XY", o, lo que es lo mismo, la que significa la porción restante de sustraerle la suma ponderada de los parciales a la relación originaria.

En esta segunda situación, la intervención de "t" puede ser en diferentes papeles, como interviniente o antecedente, que especifica o condiciona a la relación originaria, como intermediaria parcial, en casos de doble vía de determinación.

En todos estos casos, el dirimir el papel que realmente cumple "t" se requiere de una suerte de enfoque que implique elaboración mixta: elaboración MP o PM.

### 3) Elaboración Mixta:

Como se ha visto, en muchos casos se requiere focalizar el análisis en ambos tipos de relaciones (parciales y marginales) en vez de un centramiento prevalente en alguna de ellas. Lo denominaremos "ELABORACION MIXTA".

Si se tiene en cuenta la magnitud relativa de las relaciones de los parciales y marginales, siempre suponiendo el cumplimiento de la igualdad de la fórmula general por tratarse de un modelo adecuado, tendremos básicamente dos situaciones en que los parciales serán equilibrados:

a) todas las relaciones tienen valores, entonces

$$XY > (xy, t_1) \frac{nt_1}{n} + (xy, t_2) \frac{nt_2}{n}$$

b) Alguno de los marginales o menos no tienen relación, con lo cual su producto es cero:

Entonces;

$$XY = (xy, t_1) \frac{nt_1}{n} + (xy, t_2) \frac{nt_2}{n}$$

Además si los parciales no son iguales, se reproducen estas alternativas, pero en todos los casos allí ellos reiteran la forma "especificación" o "condición" vistas en la elaboración por parciales ya vista por lo que nuestro enfoque de elaboración mixta lo circunscribimos a las dos situaciones básicas mencionadas que parten de la igualdad entre parciales.

En la situación "a" el mantenimiento de relación en los parciales en la exposición a "t" confirma la relación XY, para la presencia de los marginales y el menor valor de la suma ponderada de los parciales con respecto a la relación originaria, denotan la presencia de t en la relación. Presencia que no es especificante, y menos condicionante, porque ambos parciales se mantienen similares.

Si suponemos como parece obviamente necesario que Y es variable dependiente de X en estas rela-

ciones y asumimos la responsabilidad teórica de optar por la dirección de "las flechas" en las restantes, de acuerdo a la opción que hagamos, se presentan las siguientes alternativas:

- 1) que supongamos que X determina a "t" y ésta a Y, en cuyo caso tenemos una **DOBLE VÍA DE DETERMINACIÓN** a Y, directamente y a través de "t",
- 2) que supongamos que "t" determina tanto a "Y" como a "X", en cuyo caso tenemos la situación de **RESULTANTES RELACIONADAS**
- 3) Si suponemos que "t" determina a "x" pero que es determinada por "Y", estaremos presenciando un **SISTEMA CIRCULAR AUTOSUSTENTADO** (X' Y t X)

En la situación b) cuando la relación ausente en los marginales que anula su producto, es "TY" debe suponerse que la relación presente "Xt" significa determinación de X sobre "t" (de lo contrario, habiendo relación XY tY a nivel bivariado debería haber relación, en cuyo caso "t" o "Y" serían **EFFECTOS INDEPENDIENTES DE X**).

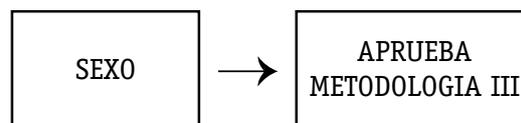
Cuando la relación ausente en los marginales que anula su producto, es "Xt", no puede suponerse que "t" sea dependiente de "Y", que entonces bivariadamente debería haber relación "Xt", por lo tanto "t" debe ser determinante de "Y", y en otro caso, estamos en presencia de **DETERMINANTES INDEPENDIENTES** entre sí ("X" y "t") del efecto "Y".

Por fin, si la nulidad del producto de los marginales se debe a que ninguna de esas dos relaciones (ni "Xt", ni "tY") existen y teniendo en cuenta la similitud de los parciales entre sí, y con la relación originaria.

### EJEMPLO:

A continuación se presenta un ejemplo<sup>4</sup>, para aplicar esta técnica, siguiendo un esquema similar al utilizado en el dictado de las clases de Metodología III, al introducir el modelo Lazarsfeld de análisis de datos.

Se quiere estudiar si el sexo de los estudiantes de ciencias sociales se asocia con que éstos aprueben o reprueben metodología III en el primer período de examen.



4 construido a partir de datos ficticios

Ambas variables son dicotómicas y nominales, por lo tanto para determinar el nivel de asociación entre ellas debe utilizarse un coeficiente de asociación. Consideremos el coeficiente PHI.<sup>5</sup>

### ESTUDIANTES QUE APRUEBAN METODOLOGÍA III EN EL PRIMER PERIODO POR SEXO

APRUEBA	SEXO		TOTAL
	mujer	hombre	
no aprueba	13	26	39
	48,1%	76,5%	63,9%
aprueba	14	8	22
	51,9%	23,5%	36,1%
TOTAL	27	34	61
	100,0%	100,0%	100,0%

El valor del coeficiente PHI es  $\phi = 0,29$ , lo que señala una relación moderadamente fuerte entre sexo y aprobación de Metodología III. De la lectura porcentual del cuadro<sup>6</sup>, se desprende que: entre los que no aprueban se encuentra el 48,1 % de las mujeres y el 76,5% de los hombres, entre los que aprueban se encuentra casi el 52% de las mujeres y menos de uno de cada cuatro hombres.

De esta tabla, por tanto se desprende que las mujeres aprueban la materia metodología III en el primer período en mayor proporción que los varones.

Una vez obtenidos esto debe "interpretarse" esta relación. ¿Esta prueba que se ha realizado, permite

sostener entonces que las mujeres son más inteligentes que los hombres y por eso aprueban la materia en mayor proporción que sus compañeros varones?

Probablemente sea necesario "aislar" los efectos que pudieran producir otras variables y que afectaran la relación entre sexo y aprobación de metodología III en el primer período, antes de sacar conclusiones aventuradas.<sup>7</sup>

Se cuenta con una encuesta realizada a los estudiantes de la materia, en la que se pregunta si trabajan, y cuantas horas semanales lo hacen. Se decide estudiar como afecta a la relación original (sexo—aprobación) el hecho de trabajar 8 horas o más, o trabajar menos de 8 horas, y aplicar un modelo Lazarsfeld para ello.

Por ejemplo, el investigador podría suponer que el tiempo que los estudiantes trabajen puede afectar el tiempo de estudio dedicado a la materia y en consecuencia la aprobación de ésta en el primer período. Por otra parte es probable que la proporción de estudiantes que trabajen 8 horas o más sea diferencial por sexo. Si este fuera el caso la variable "trabaja 8 horas o más", podría estar distorsionando la relación entre sexo y aprobación de metodología III en el primer período.

Dado lo cual debe buscarse una manera de "aislar los efectos de la variable trabajar 8 horas o más y permitir que la relación original (sexo — aprobación Metodología III) se pueda expresar sin la perturbación que el hecho de trabajar pueda provocar.

Esto puede realizarse fragmentando el universo de análisis en tantas categorías tenga la variable de control o factor de prueba, y estudiando la relación original en cada una de estos "subuniversos". En el ejemplo se prueba la relación original (sexo—aprobación) en los dos subuniversos. Estudiantes que trabajan 8 horas y más y el subuniverso estudiantes que no trabajan 8 horas y más.

5 Errandonea aconsejaba en sus cursos que cuando el investigador se enfrente a la elección del coeficiente de asociación de una relación divariada en las que ambas son variables dicotómicas, utilizar el coeficiente Phi frente al Q de Yule, justificando que el último tienen alta sensibilidad ríncional, es decir es muy sensible (incrementa rápidamente su valor) en el caso de una distribución observada en la que exista una celda (de los rincones de la tabla) que tenga una frecuencia cero o cercana a cero. Cortes en cambio ( ) discute la elección del coeficiente a utilizar en función de criterios más deductivos, es decir debe determinarse con anterioridad al cálculo de los mismos, la "distribución esperada" de las frecuencias, en función de la hipótesis que guía la investigación. En el caso que la distribución esperada sea de este tipo:


Se recomienda utilizar el coeficiente PHI

Si en cambio la distribución esperada es de este tipo:


Se recomienda utilizar Q de Yule

Para la discusión y fundamentación de este criterio se recomienda la lectura de Cortés y Ruvalcaba "Métodos estadísticos de". Colegio de México.

6 La lectura porcentual es indispensable para la interpretación del modelo y para determinar cuales son los atributos de la variable dependiente y de la independiente que se dan conjuntamente. Recordemos que la porcentualización se realiza en función de las categorías de la variable independiente, y se lee en función de las categorías de la variable dependiente.

7 Recordemos la noción de "validez interna". La variable hipotetizada causal debe ser verdaderamente causa de la variable dependiente, ésta no debe ser producida por "otras" variables. De allí, justamente es que es necesario controlarlas. La noción de "control" de variables alude permanentemente a la validez interna.

Es obvio que al interior de estos dos subuniversos la relación original (sexo aprobación) no puede ser afectada por la variable "número de horas de trabajo" puesto que ésta no varía (en cada uno de los subuniversos la situación laboral es la misma: o todos trabajan 8 horas o más, o todos no trabajan 8 horas y más por día). La relación original medida en cada uno de estos subuniversos se denomina "parcial"<sup>8</sup>. Por lo tanto en cada uno de los parciales la relación

original se manifiesta sin las perturbaciones que pueda estar generando en la relación el hecho de trabajar 8 horas o más o menos de 8 horas.

Siguiendo este esquema teórico se construye la tabla trivariada, considerando número de horas de trabajo como variable de control, sexo como variable independiente y aprobación de metodología III en el primer periodo como variable dependiente.

### Estudiantes por número de horas que trabaja y sexo según aprueba o no aprueba Metodología III en el primer periodo

	Trabaja 8 horas y más			No trabaja 8 horas y más			TOTAL
	Mujeres	Hombres	Subtotal	Mujeres	Hombres	Subtotal	
<b>Aprueba Met III</b>	1 (14,3%)	2 (8,3%)	3 (9,7%)	13 (65%)	6 (60%)	19 (63,3%)	22
<b>No aprueba Met III</b>	6 (85,7%)	22 (91,7%)	28 (90,3%)	7 (35%)	4 (40%)	11 (36,7%)	39
<b>Total</b>	7 (100%)	24 (100%)	31 (100%)	20 (100%)	10 (100%)	30 (100%)	61 (100%)

De la construcción de la tabla trivariada se desprenden:

Los parciales:

**P1:** la relación original (sexo—aprueba Met III) para los que trabajan 8 horas o más. Recuadrado en

	Mujeres	Hombres	Subtotal
Aprueba Met. III	1 (14,3%)	2 (8,3%)	3 (9,7%)
No aprueba Met. III	6 (85,7%)	22 (91,7%)	28 (90,3%)
Total	7 (100%)	24 (100%)	31 (100%)

**P2:** la relación original (sexo—aprueba Met III) para los que no trabajan 8 horas y más.

	Mujeres	Hombres	Subtotal
Aprueba Met. III	13 (65%)	6 (60%)	19 (63,3%)
No aprueba Met. III	7 (35%)	4 (40%)	11 (36,7%)
Total	20 (100%)	10 (100%)	30 (100%)

8 Habrán tantos parciales como categorías tenga la variable de control. Vale la pena volver sobre algo casi elemental: el total de observaciones de cada parcial corresponde al total de observaciones de cada categoría de la variable de control. En el ejemplo: el total del P1 (n1) es 31 que corresponde al número de estudiantes que trabajan 8 horas o más. El total del P2 (n2) es 30 y corresponde al total de estudiantes que no trabajan menos de 8 horas por día. El total de estudiantes es 61 y es la suma de  $n1 + n2 = n$

**Marginal:** es la relación tY (trabaja—aprueba) que se construye con los subtotales de los parciales.

	Trabaja 8 hs. y más	No trabajan 8 hs. y más	TOTAL
Aprueba Met. III	3 (9,7%)	19 (63,3%)	22 (32,1%)
No aprueba Met. III	28 (90,3%)	11 (36,7%)	39 (63,9%)
Total	31 (100%)	30 (100%)	61(100%)

**Subtotal:** es la relación Xt (sexo—trabaja) que se construye con las filas sombreadas:

	MUJERES	HOMBRES	TOTAL
Trabajan 8 hs. y más	7 (25,9%)	24 (70,6%)	31 (50,8%)
No trabajan 8 hs. y más	20 (74,1%)	10 (29,4%)	30 (49,2%)
	27 (100%)	34 (100%)	61 (100%)

### Interpretación:

Ha sido el /la investigador/a quien le ha atribuido el "sentido" a las relaciones, es decir quien ha colocado el sentido de las flechas. La estadística no permite determinar el sentido de determinación de las variables. Es en función de la teoría que el investigador asume esta responsabilidad. Uno de los criterios más claros para determinar el sentido causal es la antecendencia temporal de las variables. En este caso es obvio que el sexo es antecedente al hecho de trabajar o no 8 hs o más por día. Por lo tanto JAMAS un investigador podría asumir una relación inversa a la planteada (el hecho de trabajar 8 hs o más no puede determinar el sexo).

El sentido de la relación entre trabajar 8 hs o más por día y la aprobación de metodología III, también se despeja por antecendencia temporal: el estudiante trabaja o no 8 hs por día antes de aprobar la materia.

Analicemos ahora el modelo planteado a la luz de las magnitudes de los coeficientes de asociación PHI calculados para las relaciones: Partimos de una relación original sexo aprobación de 0.29 lo cual puede interpretarse como una asociación moderadamente fuerte, y que a partir de la lectura porcentual, manifiesta que las mujeres aprueban en mayor proporción que los hombres. Al controlar esta relación

### Cálculo de los coeficientes para cada una de las relaciones construidas:

PHI de la Relación Original (RO) sexo—aprueba = 0.29

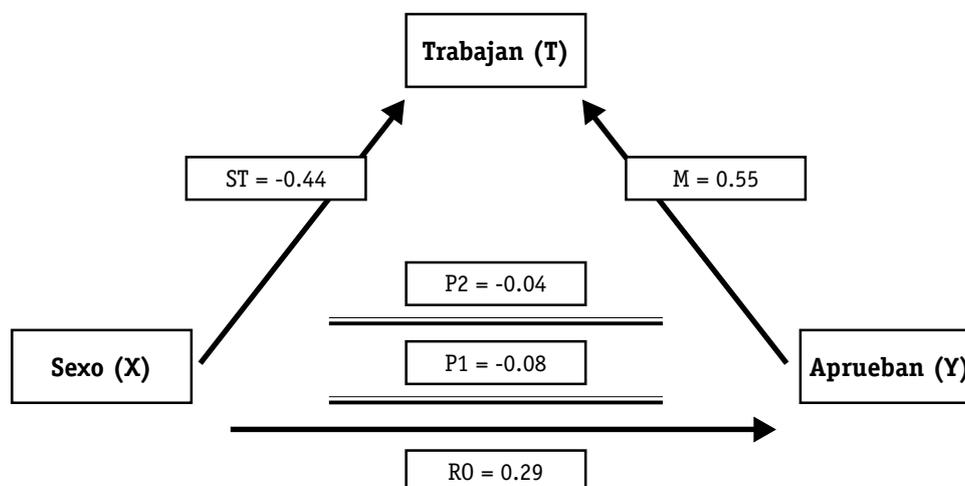
PHI del P1: (RO para los que trabajan 8 hs y más) = - 0.084

PHI del P2 (RO para los que no trabajan 8 hs y más) = -0.049

PHI del Marginal (M) (TY) Trabaja—aprueba = 0.55

PHI del Subtotal (ST) (XT) Sexo—trabaja = -0.44

### Esquema del Modelo:



por la variable número de horas que trabaja (T), se consigue "neutralizar" los efectos que esta variable T pueda estar ejerciendo sobre la relación original.

Recordemos que si lo que se pretende es neutralizar los efectos de una tercera variable en una relación original de dos variables, debe buscarse un mecanismo en la que ésta no opere. Tal es así que se mide esa relación original, para los subuniversos definidos por las categorías de la variable de control. Al interior de cada uno de los subuniversos, la variable de control permanece fija, por lo tanto no puede intervenir en la relación original.

La relación original medida en los dos parciales, bajan considerablemente considerando la R0. Ambos son cercanos a 0. Lo cual significa que al ser controlada por T, la relación original entre sexo y aprobación de Metodología III en el primer periodo desaparece.

Como habíamos visto antes, la igualdad en la ecuación de Lazarsfeld queda entonces referida al término del producto de los parciales. Veamos:

$$XY = (xy, t_1) + (xy, t_2) n_2 / n + (xt) (ty)$$

En nuestro ejemplo:

$$.293 = (-.084) 31/61 + (-0.049) 30/61 + (0.55)(-0.44)$$

$$-.293 = (-0.04) + (-0.02) + (-0.24)$$

$$-.293 \sim -.30$$

A partir de ello, podemos afirmar que la igualdad se cumple y aceptar que el modelo planteado es adecuado. Y además situarnos en un análisis del tipo "elaboración por marginales" dado que ambos parciales tienden a 0.

Dado el sentido de las flechas (que representan el sentido de determinación de las variables) estamos en un caso de intermediación. Esto se interpreta como que en realidad toda la relación que aparece (empíricamente) entre el sexo y la aprobación a Metodología III, en realidad está intermediada por el número de horas que trabajan los estudiantes. Es en realidad esta variable la "verdaderamente determinante", ya que al medir la relación original en los parciales ésta prácticamente desaparece.

De esta forma, podemos concluir que a pesar que se encuentra en la relación bivariada entre sexo y aprobación, una proporción de las mujeres superior a la de los varones que aprueban Metodología III, esto se debe a que las mujeres trabajan en una proporción menor que los varones más de ocho horas diarias. El coeficiente de asociación PHI de la relación entre sexo

y número de horas que trabajan es de .44, mucho mayor incluso al PHI de la relación original.

Incluso, más aún, dado que el marginal es superior en magnitud (el valor absoluto del coeficiente) que la relación original, podríamos estar en condiciones de "girar el cuadro". El giro hace referencia a cambiar las variables de posición, es decir ubicar la variable T (de control) como independiente y la variable independiente (X) como de control.

## Bibliografía

- Cortés, F; Rubalcava, R.M. "Métodos estadísticos aplicados a la investigación en Ciencias Sociales. Análisis de asociación. El Colegio de México 1987.
- Errandonea, Alfredo. Guías de Clase.
- Errandonea, Alfredo. "El papel lógico de las magnitudes de las relaciones en el análisis Multivariado de la asociación con parciales y marginales" Revista de Ciencias Sociales. N° 4 pgs. 43-49.
- Galtung, J. "Teoría y Métodos de la investigación social". EUDEBA Bs.As. 1978. Tomo I Cap 1 y 2.
- García Ferrando, M. "Estadística descriptiva III: Tres o más variables" En Socioestadística : Introducción a la Estadística en Sociología" Editorial Alianza Universidad Textos. España 1er ed.1985.
- Kendall, Patricia. "Introducción a las variables adicionales y la elaboración del análisis" En Hyman, H "Diseño y análisis de encuestas sociales". Amorrortu Buenos Aires 1984. (Pags 339-400).
- Lazarsfeld, Paul. "Interpretación de las relaciones estadísticas como métodos de investigación. FCU N° 61
- Lazarsfeld, P; Menzel, H. "Sobre la relación entre las propiedades individuales y colectivas. En Boudon y Lazarsfeld. Metodología de las Ciencias Sociales. Análisis empírico de la Causalidad". Editorial Laia. Barcelona, 1966 (Vol II 59-76)
- Simiand, Francois. "Sobre la explicación" (extraído de Le Salaire, l'évolution sociale et la mennaie, Paris, 1932. En Boudon y Lazarsfeld "Metodología de las Ciencias Sociales". Tomo III. Análisis empírico de la causalidad. Editorial Laia, 1966.

DESCRIPTORES: Sociología / Metodología / Análisis multivariado / Modelo Lazarsfeld