

MÉTODOS ESTADÍSTICOS
APLICADOS A LA
INVESTIGACIÓN
EN CIENCIAS SOCIALES
ANÁLISIS DE ASOCIACIÓN

Fernando Cortés
Rosa María Rubalcava



EL COLEGIO DE MÉXICO

CENTRO DE ESTUDIOS SOCIOLOGICOS

2. Asociación de variables no métricas en tablas de dos por dos

A continuación desarrollaremos unos cuantos conceptos estadísticos para luego centrar la exposición del material en el análisis de sus vinculaciones con lo teórico.

2.1. Distribuciones bidimensionales de frecuencias: variables dicotómicas

Supongamos que tenemos dos variables (cada una indicador de un concepto) simbolizadas por X y Y y que a la primera la dicotomizamos en B y B' y a la segunda en A y A' . El cruce de las dos variables dicotómicas genera una tabla de distribución de frecuencias como la 3:

TABLA 3

Distribución conjunta de frecuencias de dos variables dicotómicas

$Y \setminus X$	B	B'	Total
A	$N(AAB)$	$N(AAB')$	$N(A)$
A'	$N(A'AB)$	$N(A'AB')$	$N(A')$
Total	$N(B)$	$N(B')$	N

En ella debemos destacar: *i*) las distribuciones marginales de la variable X y de la variable Y , que informan sobre el número de casos en las categorías B y B' de X y sobre la cantidad de observaciones en las categorías A y A' de la variable Y . *ii*) La distribución conjunta que muestra el número de observaciones que comparten simultáneamente

¹En B' se agrupan todas las observaciones que no se incluyeron en B . Esto debido a que en una dicotomía las dos clases (en este caso B y B') son mutuamente excluyentes y exhaustivas.

las mismas características de X y Y. Así, $N(A \cap B)$ simboliza los casos que pertenecen a las clases B de X y A de Y; $N(A' \cap B)$ representa la cantidad de observaciones que son simultáneamente B en X y A' en Y. Del mismo modo se interpretan las otras frecuencias que completan la distribución bidimensional: $N(A \cap B')$ y $N(A' \cap B')$. iii) Que N representa al total de observaciones.

De lo anterior se deriva que los componentes de una tabla bidimensional de frecuencias se pueden desglosar en tres distribuciones. Las dos marginales:

X	Frecuencias	Y	Frecuencias
B	N(B)	A	N(A)
B'	$\frac{N(B')}{N}$	A'	$\frac{N(A')}{N}$

y la conjunta:

X y Y	Frecuencias
B y A	$N(A \cap B)$
B' y A	$N(A \cap B')$
B y A'	$N(A' \cap B)$
B' y A'	$N(A' \cap B')$

Para ilustrar los conceptos expuestos consideremos el cuadro 1, tomado de un trabajo de Joaquín Duque y Ernesto Pastrana,² que se categorizó para adecuarlo a los propósitos de nuestra exposición. En él se muestra una distribución bidimensional de frecuencias resultante de una encuesta aplicada a una muestra de la población económicamente activa del Gran Santiago en 1970.

CUADRO 1

Distribución de la población económicamente activa por rama de actividad y contexto ecológico

Rama de la actividad productiva	Contexto ecológico		Total
	Poblaciones	Campamentos	
Industria manufacturera	902	673	1 575
Otras	1 894	961	2 855
Total	2 796	1 634	4 430

²Joaquín Duque y Ernesto Pastrana, "La movilización reivindicativa urbana de los sectores populares en Chile", en *Revista Latinoamericana de Ciencias Sociales*, núm. 4, 1972, p. 286.

La variable contexto ecológico es indicador de la "movilización reivindicativa por la vivienda". En efecto, en Chile los trabajadores definidos como "no empleados"³ tenían derecho a afiliarse al Servicio de Seguro Social y por su intermedio hacerse de una vivienda en un conjunto habitacional denominado población. Esto quiere decir que para llegar a ser vecino de una población era necesario cumplir con los requisitos impuestos por la vía institucional. Por el contrario, los campamentos se constituían por tomas de terrenos que implicaban una modalidad de presión organizada y políticamente orientada que entrañaba de hecho un rechazo a las normas de apropiación privada del espacio urbano y una acción que desbordaba las normas estatales de acceso a la vivienda. Los campamentos se constituían por canales extrainstitucionales.

La movilización reivindicativa por la vivienda reconoce sólo dos caminos, que pueden suponerse mutuamente excluyentes: el institucional y el extrainstitucional. El indicador (o variable estadística) que trata de dar cuenta de este concepto es el "contexto ecológico", que se ha dividido en poblaciones y campamentos. Las primeras están en cotomizado directa con la "vía institucional", y los segundos con la "no institucional".

Nótese que en relación con el concepto "movilización reivindicativa por la vivienda", la variable "contexto ecológico" permite clasificar sin ambigüedad a la población económicamente activa (PEA) y que es exhaustiva en cuanto a la población que obtuvo éxito en su lucha por el techo. En esta investigación sólo interesaba aquella parte de la PEA que se podía englobar con el calificativo de "no empleado"⁴ y que se encontraba asentada en una "vivienda" propia obtenida socialmente (es decir, no por herencia ni individualmente) ya sea por vías institucionales o extrainstitucionales.

La variable "contexto ecológico" no intenta clasificar los lugares de residencia de los "no empleados" (en cuyo caso no cumpliría con el requisito de ser exhaustiva porque hay otros contextos de residencia) sino que se usa como indicador de movilización exitosa por techo; en este sentido cumple con la condición de tener categorías mutuamente excluyentes (se supone que los no empleados socialmente asentados no pueden vivir a la vez en un campamento y en una población) y exhaustivas por cuanto el camino institucional conduce a una población y la vía extrainstitucional a un campamento.

La variable "rama de la actividad productiva" es un indicador de

³Dadas las características de la población estadística en la cual se basó el estudio, la categoría "no empleado" engloba a todas aquellas ocupaciones que habitualmente se denominan manuales no calificadas.

⁴Se trata de aquella parte de la población de menores ingresos que no cotizaban, según la ley, a instituciones de seguridad específicas para empleados.

marginalidad. Los autores habían expuesto en un trabajo previo⁵ un concepto de marginalidad basado en el tipo de actividad económica que demanda el modelo de desarrollo.⁶ El argumento que permite establecer este vínculo sostiene que las distintas etapas por las que atraviesa el desarrollo de un país permiten distinguir, en cada caso, entre actividades económicas centrales y no centrales. Así por ejemplo, en la época en que el modelo de desarrollo era hacia afuera las ramas en que se basaba la dinámica económica fueron el transporte, el comercio y las actividades relacionadas con la producción del o de los bienes de exportación. En la etapa del desarrollo hacia adentro, la actividad económica que signaba el ritmo de crecimiento sería fundamentalmente la industrial.

Sobre la base de estas consideraciones, los autores plantean que las ocupaciones centrales para el desarrollo económico orientado hacia adentro serían las de la industria manufacturera. En las actividades de construcción y servicios (que conforman básicamente el grupo "otras") se llevarían a cabo labores superfluas para el modelo de desarrollo, en especial si se toma en cuenta que la PEA que se estudia es la de "no empleados" con bajos niveles de ingreso. En la categoría "otras ramas" se aglutinan vendedores ambulantes y de servicios menores, obreros de la construcción, etcétera.

Las ramas de la actividad productiva según las cuales se clasifica a la PEA indican el carácter de marginal o no marginal. Como marginales se incluye a aquellas personas con inserción laboral en la rama "otras" y como no marginales las que trabajan en la industria.

La rama de la actividad productiva que se usa como indicador (o variable estadística) de marginalidad es dicotómica en tanto cualquier miembro de la PEA labora ya sea en la industria manufacturera o en una unidad de producción clasificada en su categoría complementaria y exhaustiva por cuanto no hay ninguna persona ligada a la actividad productiva que no se pueda clasificar en una u otra categoría.

Las distribuciones marginales del cuadro de doble entrada son:

CUADRO 2

Distribuciones marginales de la PEA por contexto ecológico y rama de actividad productiva

Contexto ecológico (X)	Frecuencia	%	Rama de actividad (Y)	Frecuencia	%
Populaciones	2 796	63.1	Industria manufacturera	1 575	35.6
Campamentos	1 634	36.9	Otras	2 855	64.4
Total	4 430*	100.0	Total	4 430	100.0

5J. Duque y E. Pastrana, *Elementos teóricos para la interpretación de los procesos organizativo-políticos poblacionales*, ELAS-FLACSO, Santiago, diciembre de 1971.

6Concepto que difiere del planteado por Roger Vekemans, que más bien aludía a una marginalidad de carácter cultural.

Las distribuciones marginales muestran que los "no empleados" que han tenido éxito en su lucha por la vivienda habitan preferentemente en poblaciones (63.1%) y laboran en su gran mayoría fuera de la industria manufacturera (64.4 por ciento).

Ahora bien, el cuadro 1, que se origina en el cruce del contexto ecológico con la rama de la actividad productiva, también puede interpretarse desde el punto de vista de la teoría de la marginalidad en boga por esos años. Al respecto Jorge Giusti plantea⁷ que:

Resumiendo, dividiremos la teoría del Desarrollo Social para América Latina (DESAL) en cinco aspectos principales, que son los siguientes:

Ecológico: Desde una perspectiva ecológica o territorial, el DESAL distingue tres tipos de vivienda, tomando en cuenta las condiciones físicas de las casas y los lugares en que se hallaban emplazadas. Estos tipos son: a) los "círculos de miseria", constituidos por los grupos de viviendas precarias que forman vecindarios donde las viviendas aparecen de manera espontánea; b) las viviendas viejas y deterioradas dentro de la ciudad y c) el vecindario planificado de origen estatal o privado.

Sociopsicológico: sobre la base de la mencionada diferenciación ecológica, DESAL afirmaba que cada tipo de vivienda genera un grupo humano distinto. Centralizando su atención sobre los pobladores que viven en círculos de miseria (viviendas precarias), el DESAL desarrolló una serie de rasgos que caracterizarían a los pobladores. Dichos rasgos son:

El vocablo poblador que se utiliza para describir a las masas urbanas marginales ilustra su reducida capacidad para actuar. Ellos simplemente "pueblan" el lugar, sólo "son", y nada más.

Marginalidad significa falta de participación en los beneficios y recursos sociales y en la red de decisiones sociales.

Marginalidad es falta de integración interna. La falta de participación surge de otro rasgo constituyente de marginalidad: la falta de integración interna de los grupos "marginales".

El hombre "marginal" se caracteriza por su incapacidad de superar su situación actual por sí mismo. La marginalidad es tan radical que responde a la definición etimológica: constituye un problema que afecta las mismas raíces del ser del hombre: corroe la médula del potencial del hombre para el automejoramiento voluntario y racional.

Sociocultural: entre aquellos beneficios y recursos sociales de los que carecen los pobladores, los vinculados con su trasfondo social y cultural, se encuentran: bajos niveles de vida, de salud y de vivienda y bajos niveles educacionales y culturales.

Económico: económicamente, puede definirse al grupo humano marginal, manifestaba el DESAL, como "subproletarios", porque comparten una carencia de dos factores: ingreso y empleo. Su contribución a la productividad es mínima y resulta en salarios con los que apenas subsisten. La falta

⁷En *Organización y participación popular en Chile: el mito del hombre marginal*, Ediciones FLACSO, Buenos Aires, 1973.

de empleo estable es tal que frecuentemente no es otra cosa que desempleo disimulado.

Político: como vimos previamente, el otro aspecto de la falta de participación de la gente "marginal" se relaciona con la sociedad como una red de decisiones sociales. Es decir, que los pobladores no toman parte de las responsabilidades o tareas que deben ser emprendidas para la solución de problemas sociales en general, y para la solución de sus propios problemas en particular. Los grupos marginales, consecuentemente, no cuentan con los tipos de organización que los vincularían con la sociedad y que además, les brindarían "representación" frente a las instituciones de la sociedad participante.⁸

De las líneas anteriores se puede desprender, por una parte, que el contexto ecológico es la característica básica que determina la condición de marginalidad y, por otra, que deberíamos esperar una estrecha relación entre el lugar de residencia y la participación en la actividad económica. En buenas cuentas, según la versión de la teoría del DESAL, que nos entrega Jorge Giusti, los campamentos estaban formados por subproletarios. En cuanto a las características de los habitantes de las poblaciones no es muy explícito, pero se podría suponer que en ellas existía una cierta heterogeneidad.

De acuerdo con esta versión de la teoría de la marginalidad del DESAL se pueden especificar dos hipótesis sobre la relación entre las variables en estudio:

Si B: poblaciones , B': campamentos ,
A: industria manufacturera , A': otras industrias,

se podría sostener la proposición empírica P₁:

$$P_1: \begin{array}{c} B \text{---} \curvearrowright A \\ B' \text{---} \curvearrowright A' \end{array}$$

Donde B \curvearrowright A significa "si B entonces A". En consecuencia, esta primera proposición establece que si una persona vive en una población entonces es no marginal, y, de acuerdo con la teoría, debería trabajar en la industria manufacturera; si habita en un campamento entonces es marginal. Es decir, sólo los marginales de la sociedad usan medios extrainstitucionales para resolver su problema habitacional en tanto que los no marginales recurren exclusivamente a las vías legales.

La otra hipótesis (P₂) que se puede sostener según esta versión de

⁸Jorge Giusti, *op. cit.*, pp. 60-62.

la teoría de la marginalidad cultural se puede expresar en los siguientes términos:

$$P_2: \begin{array}{c} B \text{---} \curvearrowright (A \text{ o } A') \\ B' \text{---} \curvearrowright A' \end{array}$$

Esta proposición plantea que de acuerdo con la teoría deberíamos esperar que los habitantes de las poblaciones desempeñaran su actividad económica en cualquier rama productiva; esto quiere decir que un poblamiento regular estará compuesto tanto por marginales como por no marginales y que los campamentos debieran estar constituidos sólo por aquéllos.

La hipótesis establece que tanto los marginales como los no marginales usan el sistema institucional para resolver el problema de la vivienda. Pero que sólo los primeros recurren al medio extralegal.

Las dos formalizaciones lógicas son consistentes con el planteamiento de carácter general sustentado por Giusti. Comparan como proposición central que sólo los marginales recurren a caminos no institucionales para resolver el problema de la vivienda.

Esta versión de la teoría prácticamente asimila o subsume la marginalidad en el contexto ecológico. Son marginales aquellos que viven en campamentos y no lo son los que habitan en poblaciones. De aquí que sea más probable que la primera de las proposiciones lógicas refleje con más fidelidad el pensamiento del autor.

De la teoría pareciera derivarse que la variable explicativa (o independiente) sería el contexto ecológico y que la variable explicada (o dependiente) sería la rama de la actividad productiva en que se desempeña el trabajador. El hecho de que una persona viva en una población o en un campamento bastaría para establecer su carácter de marginal o no marginal y de acuerdo con su condición se esperaría que su inserción laboral fuese en aquellas ramas en que predomina el subproletariado o en la industria manufacturera.

El cruce de las variables contexto ecológico y rama de la actividad productiva es un punto de encuentro entre las dos teorías de la marginalidad. Aún cuando la perspectiva del Desal destaca lo cultural y la histórico-estructural privilegia lo económico, existe un vínculo entre ellas que justifica su coincidencia desde el punto de vista de lo empírico. En efecto, según el planteamiento de Giusti, la marginalidad se determina por el lugar de residencia (contexto ecológico) en tanto que de acuerdo con Duque y Pastrana por la inserción laboral. El primero de los conceptos de marginalidad se objetiva a través del contexto ecológico observable y el segundo por la inserción laboral.

El planteamiento de la "marginalidad económica" supone que el

grado de politización de la fuerza de trabajo y las dificultades para satisfacer la demanda de vivienda por los canales institucionales, crearon las condiciones para que los trabajadores buscaran como segundo camino el extralegal, aunque ello significara desconocer uno de los derechos básicos en que se apoya el capitalismo: la propiedad privada. Según las concepciones de la marginalidad económica los marginales y los no marginales participan por igual en la creación de un campamento; sólo los aglutinaba la imposibilidad de resolver el problema habitacional dentro del marco legal y una posición política relativamente compartida.

Esta postura sostiene que la marginalidad *no* determinaba el enrollamiento en la movilización reivindicativa por la vivienda. Al nivel de los indicadores esto significa que no hay relación entre la rama de la actividad (variable explicativa o independiente) y el contexto ecológico (variable explicada o dependiente).

De los desarrollos conceptuales presentados se deriva que la teoría de la marginalidad cultural proporciona elementos para que empíricamente se presente una estrecha relación entre el contexto ecológico y la rama de la actividad productiva, en que el primero sería la variable independiente y la segunda la dependiente. La teoría de la marginalidad económica no sólo se diferencia porque plantea la inexistencia de tal asociación, sino que invierte su sentido. Ahora la variable explicativa (independiente) sería la inserción laboral y la explicada el contexto ecológico.⁹

Una vez que se han especificado las hipótesis que derivan de ambas teorías, es decir, una vez conocidas las proposiciones teóricas y sus correspondientes proposiciones empíricas, a saber:

i) según la teoría de la marginalidad cultural¹⁰ debe haber una relación entre los dos indicadores, de modo que se pueda predecir la inserción laboral a partir del tipo de asentamiento urbano.

$$P_1: \quad B \text{ --- } A \\ \quad \quad B' \text{ --- } A'$$

ii) y según la marginalidad económica no debiera haber relación entre las variables, por cuanto el proceso de marginalización no determina la movilización.

$$P_3: \quad A \text{ --- } (B \text{ o } B') \\ \quad \quad A' \text{ --- } (B \text{ o } B')$$

⁹ Ambas teorías hacen descansar la explicación de los fenómenos que estudian en la marginalidad, pero este nombre remite a dos conceptos distintos que se objetivan de manera diferente, lo que hace cambiar el sentido del vínculo explicativo.

¹⁰ Por el momento nos limitaremos a trabajar sólo con la proposición P_1 ; más adelante incluiremos P_2 .

¿Cómo podemos validarlas? ¿Con cuál de las dos teorías concuerdan los datos? ¿Cuál de las conceptualizaciones entra en contradicción con la evidencia empírica?

Hasta este punto hemos examinado la relación entre las proposiciones teóricas y las proposiciones empíricas en el plano lógico. Interesa ahora (para estar en condiciones de responder a las preguntas formuladas) estudiar los vínculos que permiten unir las proposiciones empíricas con la distribución de los datos. Para este propósito será necesario dedicarse a exponer algunos principios básicos que permiten el análisis estadístico de una tabla de frecuencias de doble entrada en el caso particular en que las dos variables son dicotómicas.

2.2. Noción de independencia estadística

Tomemos la proposición empírica P_1 postulada por la teoría de la marginalidad cultural y examinemos cuál es la distribución esperada de los datos. De acuerdo con P_1 los datos del cuadro 1 se deberían ubicar *preferentemente* en las casillas que forman la diagonal principal de la tabla (la que va de izquierda a derecha). Decimos que *preferentemente*, a pesar de que la teoría sostiene que *exactamente* porque por razones de azar (o si se quiere por otros factores no considerados en el análisis, cuyo peso explicativo es insignificante) estamos dispuestos a aceptar *algunas* observaciones en la diagonal secundaria sin que ello invalide la proposición empírica y en consecuencia la teoría.

Los términos *preferentemente* y *algunas* tienen la suficiente vaguedad como para admitir que con una misma distribución bidimensional de datos un investigador decida rechazar la proposición empírica y por ende la teórica, mientras que otro no. ¿Cómo establecer un criterio que permita decidir hasta qué punto las desviaciones se pueden tolerar, de modo que no se rechacen las proposiciones?

La respuesta a esta pregunta debe buscarse por dos vías no excluyentes: una, establecer una medición de la fuerza con que se asocian las variables; otra, descartar que la asociación observada se debe al azar. Este último camino cae dentro del ámbito de la inferencia estadística y por tanto no se tratará en este libro. En adelante nos preocuparemos por la manera como procede la estadística para medir el grado de relación entre dos variables dicotómicas.

La idea de asociación se define por oposición a la de independencia. Las dos teorías de marginalidad se contradicen porque una plantea que debe existir una relación entre las variables en tanto que la otra sostiene que deben ser independientes.

En este apartado examinaremos el concepto de independencia estadística y diferentes maneras de objetivarlo, así como la forma de medir la fuerza o el grado de relación de dos variables.

A primera vista, los datos del cuadro 1 parecerían aportar alguna evidencia en favor de la teoría de la marginalidad cultural ya que en las poblaciones hay más trabajadores industriales (902) que en los campamentos (673). Decimos que alguna evidencia porque 673 trabajadores industriales que habitan en campamentos parece un desvío demasiado grande. Sin embargo, ese mayor número sólo es aparente por cuanto los totales marginales son diferentes. De los 2 796 habitantes de poblaciones 902 trabajan en la industria, mientras que de las 1 634 personas que viven en campamentos 673 laboran en la manufactura. La comparación de los valores absolutos carece de sentido por cuanto se ubican en totales de distinto tamaño. Para eliminar el efecto del tamaño diferencial, las frecuencias absolutas se expresan en proporciones (que multiplicadas por 100 constituyen porcentajes) y posteriormente se procede a examinar la diferencia entre ellas.

Al calcular los porcentajes¹¹ con respecto al total de cada columna del cuadro 1, obtenemos:

CUADRO 3

Distribución porcentual de la PEA por rama de actividad económica dentro de cada contexto ecológico

Rama de actividad	Contextos ecológicos		
	Poblaciones	Campamentos	Total
Industria manufacturera	32.3	41.2	35.6
Otras	67.7	58.8	64.4
Total	100.0	100.0	100.0

Al controlar el tamaño de las poblaciones y campamentos aparece una imagen totalmente distinta de la que obtuvimos con las frecuencias absolutas. El cuadro 3 nos dice que la situación es más bien la opuesta a la que plantea la proposición empírica de la marginalidad cultural. En efecto, el porcentaje de trabajadores de la industria manufacturera es mayor en los campamentos (41.2%) que en las poblaciones (32.3%). Este resultado permite rechazar la hipótesis sostenida por la teoría de la marginalidad cultural por cuanto las observaciones no sólo no se agrupan en la diagonal principal, sino que además se comportan de manera opuesta a la hipotetizada. ¿También rechaza la proposición planteada por la teoría de la marginalidad económica?

¹¹De aquí en adelante no nos preocuparemos por distinguir con demasiada precisión entre proporción y porcentaje.

Sabemos que plantear que la movilización reivindicativa urbana no se relaciona con la marginalidad (proposición teórica) origina la proposición empírica:

$$P_3: A \text{---} \text{---} \text{---} (B \text{ o } B')$$

$$A' \text{---} \text{---} \text{---} (B \text{ o } B')$$

El examen del vínculo entre la proposición empírica y la distribución de los datos requiere que se controle el tamaño cuando las frecuencias marginales son distintas, lo que equivale a decir que se deben calcular porcentajes.

De acuerdo con la proposición empírica, los trabajadores de la industria manufacturera (no marginales) y los de las otras ramas (marginales) se deben distribuir por igual entre poblaciones y campamentos. A diferencia de la proposición P_1 de la teoría de la marginalidad cultural que predicaba que las observaciones debían concentrarse en la diagonal principal del cuadro, la de la marginalidad económica P_3 nos dice que deben distribuirse de la misma manera dentro de cada rama en ambos contextos ecológicos. La proporción de marginales en las poblaciones debiera ser similar a la de no marginales.

En consecuencia, la marginalidad y la movilización reivindicativa urbana serán independientes si comprobamos que no existe relación en el plano de los indicadores (rama de la actividad productiva y contexto ecológico). Para saber si hay o no concomitancia debemos comparar el porcentaje de marginales que viven en poblaciones con el correspondiente a no marginales que habitan en el mismo tipo de asentamiento.¹² Las variables serán independientes si ambos porcentajes son estadísticamente iguales.

CUADRO 4

Distribución porcentual de la PEA por contexto ecológico dentro de cada rama de actividad económica

Rama de actividad	Contextos ecológicos		
	Poblaciones	Campamentos	Total
Industria manufacturera	57.3	42.7	100.0
Otras	66.3	33.7	100.0
Total	63.1	36.9	100.0

¹²En una tabla de dos por dos basta con comparar esos porcentajes para juzgar acerca de la relación entre las variables, por cuanto los restantes son el complemento a cien. Es decir, si conocemos el porcentaje de trabajadores que viven en poblaciones, la fracción complementaria habita en campamentos.

Pareciera que la hipótesis de independencia que sostiene la teoría de la marginalidad económica muestra un cierto grado de adecuación con los datos. Los porcentajes de trabajadores que han logrado su asentamiento urbano a través de alguna vía institucional es similar entre los no marginales (57.3%) y los marginales (66.3%). Estos resultados llevan a Duque y Pastrana a sostener que:

Analizando las características ocupacionales de la fuerza de trabajo de las poblaciones y campamentos, puede descubrirse el carácter ideológico de las interpretaciones que categorizan estos contextos, bien sea como "marginales" o bien como formados por el "lumpen proletariado". Por el contrario, los antecedentes confirman la heterogeneidad de las situaciones de clase de los sectores populares urbanos, donde a la presencia del proletariado se suma la concurrencia del subproletariado, integrado por los trabajadores de la construcción, de los servicios personales de baja calificación, apreciables proporciones de trabajadores independientes de las más variadas actividades (artesanales, comerciales, paraproductivas industriales).¹³

Si bien podemos sostener que hay adecuación entre la proposición empírica y la distribución de los datos aún no estamos en condiciones de generar medidas estadísticas que permitan saber: *i*) si hay o no relación entre las dos variables dicotómicas y *ii*) la fuerza o grado de la relación. Para lograr este propósito es necesario establecer el concepto de independencia estadística. Antes de entrar a su desarrollo, debemos dejar constancia de un procedimiento que se encuentra implícito en los cálculos que ya hemos realizado.

Los porcentajes del cuadro 3 se obtuvieron de los totales de las columnas y los del cuadro 4 de los totales por línea. En ambos casos el cálculo se hizo tomando en cuenta la argumentación teórica: los porcentajes del cuadro 3 nos indican la distribución relativa de los trabajadores que se desempeñan en la industria manufacturera o en otras industrias en las poblaciones y en los campamentos. En el cuadro 4 señalan el peso relativo de los habitantes de cada contexto ecológico dentro de la PEA industrial y de las otras ramas. Sabemos que según los desarrollos de la teoría de la marginalidad cultural la variable independiente es el contexto ecológico y sus frecuencias marginales son los totales por columna; para la teoría de la marginalidad económica la variable independiente es la rama de la actividad productiva y sus frecuencias marginales se encuentran en las líneas. En consecuencia, siguiendo el hilo del argumento teórico, que se expresa a través de la proposición empírica, concluimos que los porcentajes los calculamos basándonos en el número de observaciones por categoría de la variable independiente. Es una convención ampliamente aceptada que los porcentajes se calcu-

¹³Duque y Pastrana, "La movilización...", *op. cit.*, p. 287.

len sobre las frecuencias marginales, de la variable independiente. Su justificación deriva de su coherencia lógica con la proposición empírica. Decimos que es una convención porque la comparación de porcentajes respecto a los totales de las correspondientes líneas o columnas. Obsérvese que las discrepancias porcentuales de los cuadros 3 y 4 son las mismas.

Considérese la siguiente tabla genérica:

TABLA 4

Distribución porcentual conjunta de dos variables dicotómicas

Y \ X	B	B'	Total
A	\rightarrow $\frac{P(A/B)}{P(A'/B)}$	$\frac{P(A/B')}{P(A'/B')}$	$\frac{P(A)}{P(A')}$
A'	$\frac{P(A/B)}{100}$	$\frac{P(A'/B)}{100}$	$\frac{P(A)}{100}$
Total	$\frac{P(A/B)}{100}$	$\frac{P(A'/B)}{100}$	$\frac{P(A)}{100}$

Se supone que X es la variable independiente y Y la dependiente. De acuerdo con la convención establecida, $P(A/B)$ y $P(A'/B)$ simbolizan los porcentajes de observaciones que tienen la característica A y A', respectivamente, en el total de la columna B. Del mismo modo $P(A/B')$ y $P(A'/B')$ indican los porcentajes que poseen las características A y A' en el total de B'. $P(A)$ y $P(A')$ son los porcentajes de observaciones que en el total tienen el atributo A y A' respectivamente (compárense estas definiciones con las dadas a propósito de los símbolos incluidos en la tabla 3).

Las variables X y Y serán estadísticamente independientes si se cumple que:

$$(1) \quad P(A/B) = P(A/B')$$

Las variables X y Y son estadísticamente independientes si el porcentaje de observaciones que poseen el atributo A (PEA ocupada en la industria manufacturera) es el mismo en B (poblaciones) que en no B (campamentos). Como se puede apreciar todos los desarrollos de esta sección han tenido como base la noción de independencia estadística. Es fácil demostrar que si en una tabla de dos por dos se cumple la

igualdad de porcentajes que supone la independencia estadística entonces también se cumple que:¹⁴

$$(2) \quad P(A/B) = P(A'/B') = P(A)$$

Es decir, ambos porcentajes son iguales al porcentaje marginal de observaciones que tienen la característica A. Del mismo modo:

$$(3) \quad P(A'/B) = P(A/B') = P(A')$$

Las triples igualdades (2) y (3) permiten derivar que:¹⁵

¹⁴Tomemos la expresión:

$$P(A) = P(A/B)P(B) + P(A'/B')P(B') = \frac{N(A \cap B)}{N(B)} \cdot \frac{N(B)}{N} + \frac{N(A' \cap B')}{N(B')} \cdot \frac{N(B')}{N} = \frac{N(A)}{N}$$

y supongamos que $P(A/B) = P(A'/B')$,

es decir que:

$$\frac{N(A \cap B)}{N(B)} = \frac{N(A' \cap B')}{N(B')}$$

entonces,

$$\frac{N(A \cap B)}{N(B)} \left[\frac{N(B)}{N} + \frac{N(B')}{N} \right] = \frac{N(A)}{N}$$

Por tanto,

$$\frac{N(A \cap B)}{N(B)} = \frac{N(A)}{N} \quad \text{y} \quad P(A/B) = P(A).$$

¹⁵Así por ejemplo si tomamos la expresión (2):

$$P(A/B) = P(A)$$

y la expresamos en términos de frecuencias, tenemos que:

$$\frac{N(A \cap B)/N}{N(B)/N} = \frac{N(A)}{N};$$

luego,

$$N(A \cap B) = \frac{N(A)N(B)}{N}$$

$$(4) \quad \left. \begin{array}{l} 4.1. \quad N(A \cap B) = \frac{N(A)N(B)}{N} \\ 4.2. \quad N(A' \cap B) = \frac{N(A')N(B)}{N} \\ 4.3. \quad N(A \cap B') = \frac{N(A)N(B')}{N} \\ 4.4. \quad N(A' \cap B') = \frac{N(A')N(B')}{N} \end{array} \right\}$$

La igualdad (1) y las ecuaciones (4) muestran dos maneras diferentes de expresar el principio de independencia estadística. La primera nos dice que las variables X y Y son estadísticamente independientes si la proporción de observaciones que son A es la misma en B y B'. Este criterio lo aplicamos en los cuadros 3 y 4 para estudiar el vínculo entre el contexto ecológico y la rama de actividad productiva. Las segundas (las ecuaciones 4) nos dicen que si las dos variables son estadísticamente independientes entonces cada frecuencia de la distribución conjunta debería coincidir con el producto de sus correspondientes marginales, dividido entre el total de observaciones.

Como las igualdades (4) se obtuvieron bajo el supuesto de independencia estadística, se concluye que deben interpretarse como las frecuencias esperadas en caso de que no hubiese relación entre X y Y. Su aplicación sistemática genera una distribución bidimensional de frecuencias esperadas cuya construcción descansa en el supuesto de ausencia de relación entre las variables.

Tenemos las distribuciones conjuntas observada y esperada de X y Y, que se traducen en dos tipos de frecuencias conceptualmente diferentes (aunque a veces coincidan numéricamente) para cada casilla, lo que hace necesario introducir una notación que permita distinguirlos. De aquí en adelante usaremos N como un operador¹⁶ que, aplicado sobre un atributo, da como resultado frecuencias observadas. Así, por ejemplo, $N(A \cap B)$ simboliza al número de casos que tienen simultáneamente las características A y B. Para representar las frecuencias esperadas usaremos E como un operador, de manera que al actuar sobre un atributo específico nos da el número de observaciones que deberíamos esperar bajo el supuesto de independencia estadística. Como las ecuaciones (4) sirven para calcular las frecuencias esperadas es necesario reescribirlas de acuerdo con esta nueva simbología.

¹⁶También representa al total de observaciones.

$$\left. \begin{aligned}
 5.1. \quad E(A\Delta B) &= \frac{N(A)N(B)}{N} \\
 5.2. \quad E(A'\Delta B) &= \frac{N(A')N(B)}{N} \\
 5.3. \quad E(A\Delta B') &= \frac{N(A)N(B')}{N} \\
 5.4. \quad E(A'\Delta B') &= \frac{N(A')N(B')}{N}
 \end{aligned} \right\} (5)$$

En la tabla 5 se muestra la distribución de frecuencias esperadas conjuntas de las dos variables dicotómicas.

TABLA 5

Distribución esperada conjunta de dos variables dicotómicas

$Y \setminus X$	B	B'	Total
A	$E(A\Delta B)$	$E(A\Delta B')$	E(A)
A'	$E(A'\Delta B)$	$E(A'\Delta B')$	E(A')
Total	E(B)	E(B')	N

No es difícil demostrar que:

$$N(B) = E(B) ; N(B') = E(B') ; N(A) = E(A) \text{ y } N(A') = E(A')$$

y, por tanto, el total de observaciones es igual en ambas distribuciones.

Al aplicar las ecuaciones (5) a los datos del cuadro 2 tenemos:

$$E(A\Delta B) = \frac{1\ 575 \times 2\ 796}{4\ 430} = 994 ;$$

$$E(A\Delta B') = \frac{1\ 575 \times 1\ 634}{4\ 430} = 581$$

$$E(A'\Delta B) = \frac{2\ 855 \times 2\ 796}{4\ 430} = 1\ 802 ;$$

$$E(A'\Delta B') = \frac{2\ 855 \times 1\ 634}{4\ 430} = 1\ 053$$

Con estas frecuencias podemos construir el cuadro 5.

CUADRO 5

Distribución esperada* de la PEA por rama de actividad productiva y contexto ecológico

Rama de actividad productiva	Contexto ecológico		Total
	Poblaciones	Campamentos	
Industria manufacturera	994	581	1 575
Otras	1 802	1 053	2 855
Total	2 796	1 634	4 430

*De acuerdo con el supuesto de independencia estadística.

Con la comparación del cuadro de frecuencias esperadas con el de frecuencias observadas (cuadro 2) queremos destacar que:

i) las distribuciones marginales son iguales y por lo tanto el número total de observaciones es el mismo.

ii) ambas distribuciones conjuntas son, en términos generales, bastante parecidas. Como la distribución observada se desvía levemente de la esperada, se puede afirmar que hay una ligazón débil entre ambas variables o, en otros términos, que los datos parecen sustentar la hipótesis de la marginalidad económica, no así la de la cultural. Esta conclusión es similar a la que habíamos obtenido con la comparación de porcentajes, y no podía ser de otro modo por cuanto los dos procedimientos son equivalentes desde el punto de vista matemático.

Es usual que cuando el análisis se efectúa por la vía de las frecuencias esperadas no se presenten dos cuadros sino sólo uno, poniéndolas entre paréntesis, inmediatamente abajo de su correspondiente frecuencia observada.

CUADRO 6

Distribuciones observada y esperada de la PEA por rama de actividad productiva y contexto ecológico

Rama de actividad productiva	Contexto ecológico		Total
	Poblaciones	Campamentos	
Industria manufacturera	902 (994)	673 (581)	1 575
Otras	1 894 (1 802)	961 (1 053)	2 855
Total	2 796	1 634	4 430

En las marginales y en el total no se anotan las frecuencias esperadas porque son iguales a las observadas. Esta tabla presenta de manera sintética las distribuciones de los cuadros 1 y 5 y permite apreciar a simple vista los desvíos de la distribución conjunta observada respecto de la independencia estadística.

Para saber si dos variables dicotómicas son o no estadísticamente independientes se puede recurrir a dos procedimientos. El primero consiste en comparar los porcentajes y en caso de que su diferencia sea pequeña o nula se dice que las variables son estadísticamente independientes. El segundo consiste en examinar las discrepancias entre las frecuencias observadas y esperadas; en caso de que coincidan o difieran poco, se dice que no hay asociación entre X y Y. Como se ve las dos vías de análisis son equivalentes desde los puntos de vista matemático y estadístico lo que necesariamente conduce a los mismos resultados sustantivos.

Hemos descrito dos maneras de analizar una distribución conjunta de dos variables dicotómicas que nos permiten apreciar si están o no relacionadas, pero no nos dan pie para construir un juicio acerca del grado o la fuerza del vínculo que las une.

2.3. Otras formas de medir asociación

El concepto de independencia estadística formalizado en la ecuación (1) contiene la idea básica para definir asociación y abordar el problema de su medición. La fuerza de la relación entre dos variables se define por su lejanía respecto a la independencia estadística. A mayor distancia mayor asociación. En lo que respecta a los porcentajes, esto quiere decir que la asociación entre X y Y será mayor mientras más grande sea la diferencia entre $P(A/B)$ y $P(A/B)$.¹⁷

De acuerdo con esta idea se podría definir como indicador del grado o de la fuerza de la relación entre las variables la diferencia:

$$(6) \quad D = P(A/B) - P(A/B)$$

En el caso de independencia estadística $D = 0$, por cuanto se cumple que $P(A/B) = P(A/B)$. Cuando $P(A/B) > P(A/B)$ el valor de D será positivo; entonces se dice que la relación entre X y Y es directa o positiva. Si $P(A/B) < P(A/B)$ entonces D será menor que cero y la asociación entre las variables se califica como inversa o negativa.

¹⁷Esta idea difiere del significado corriente que se adscribe a la palabra asociación. En su uso ordinario se dice que A y B están asociados si ambos atributos aparecen juntos con alguna frecuencia. En esta acepción A y B están asociados sólo si aparecen conjuntamente en un número mayor de casos que el esperado si ellos fuesen independientes.

Se tiene el máximo de relación positiva directa (la fuerza de la asociación es máxima) cuando:

$$P(A/B) = 100 \quad \text{y} \quad P(A/B') = 0$$

lo que implica que:

$$D = 100$$

El caso de máxima relación inversa se produce cuando

$$P(A/B) = 0 \quad \text{y} \quad P(A/B') = 100$$

y:

$$D = -100$$

Tenemos, entonces, que la discrepancia (D) puede asumir valores entre 100 y -100 con los siguientes significados:

$$(7) \quad D = \begin{cases} 100: \text{máxima relación positiva} \\ 0: \text{independencia estadística} \\ -100: \text{máxima relación negativa} \end{cases}$$

Por medio de la ecuación (6) hemos construido una escala que permite calibrar el grado o la fuerza de la relación a partir de una simple diferencia entre porcentajes. Así, por ejemplo, al calcular D con los datos del cuadro 3 se obtiene: $D = 32.3 - 41.2 = -8.9$, valor que denota una baja relación inversa, lo que apoya la hipótesis de la marginalidad económica. Este valor, mucho más cercano a 0 que a 100, permite afirmar que el contexto ecológico determina débilmente, o no determina,¹⁸ la marginalidad.

Otra vía para objetivar el concepto de independencia estadística nos la proporcionan las ecuaciones (5). En efecto, la fuerza con que relacionan las variables se puede medir a través de la discrepancia entre las frecuencias observadas y las esperadas, de acuerdo con el supuesto de independencia estadística. Este camino consiste en establecer una función que varíe con el grado de concordancia entre la distribución efectiva de las observaciones y la que debería haberse observado (esperada) si X y Y no estuviesen relacionadas. Así tendríamos:

¹⁸Esta imprecisión tal vez podría resolverse a través de una prueba de hipótesis, tema estadístico que queda fuera de los límites de este libro.

$$\begin{cases}
 8.1. D_{11} = N(A \cap B) - E(A \cap B) = N(A \cap B) - \frac{N(A)N(B)}{N} \\
 8.2. D_{12} = N(A \cap B') - E(A \cap B') = N(A \cap B') - \frac{N(A)N(B')}{N} \\
 8.3. D_{21} = N(A' \cap B) - E(A' \cap B) = N(A' \cap B) - \frac{N(A')N(B)}{N} \\
 8.4. D_{22} = N(A' \cap B') - E(A' \cap B') = N(A' \cap B') - \frac{N(A')N(B')}{N}
 \end{cases}
 \quad (8)$$

Donde D simboliza las diferencias entre frecuencias observadas y esperadas, y los subíndices permiten identificar cada casilla de la tabla. D_{11} denota la diferencia entre las frecuencias de la casilla correspondiente a las primeras línea y columna; D_{12} a la discrepancia en la casilla de la primera línea y la segunda columna, y así sucesivamente.

Los dos caminos para obtener un indicador del grado de asociación entre variables dicotómicas (el análisis de porcentajes y la comparación de frecuencias observadas y esperadas) se definen por contraste con la independencia estadística. Mientras más grande es la discrepancia mayor es la fuerza de la relación y viceversa. Definir la asociación por oposición a la independencia produce resultados claramente interpretables sólo en caso de que no haya relación entre X y Y. Esta aseveración es doble por cuanto sostiene, por una parte, que si se lleva a cabo un análisis de porcentajes entonces se tendrá que $D = 0$; en cambio, si se sigue la vía de las frecuencias esperadas se tendrá:

$$D_{11} = D_{12} = D_{21} = D_{22} = 0$$

Por otra parte, afirma que la lejanía respecto de la independencia no es suficiente para establecer el grado de aproximación con la asociación perfecta. Como no se ha definido conceptualmente lo que se entenderá por asociación perfecta es, lógicamente, imposible interpretar el valor de D de acuerdo con esta perspectiva. A primera vista pareciera haber una respuesta inmediata, por cuanto sabemos que el rango de variación de D va de -100 a 100 para los casos de relación inversa y directa máximas, respectivamente.

Los valores límite de D establecen una posibilidad estadística que no necesariamente se encuentra ligada con cualquier proposición empírica.¹⁹ Esta puede ser de tal naturaleza que si los datos se comportan de acuerdo con la teoría, entonces D no debe alcanzar ninguno de los

¹⁹Recuérdese que la proposición empírica define el tipo de asociación. Cuando la tabla observada coincide con ella, la asociación es por definición la máxima posible.

valores extremos. Puede darse el caso de que no exista una correspondencia entre las bondades estadísticas de la medida y los requerimientos que surgen desde el campo de la teoría. La inadecuación entre el abanico de posibilidades que entrega el conocimiento técnico (usualmente como respuesta a problemas planteados dentro de su propio ámbito) y las necesidades que derivan del proceso de investigación social, delimita un campo problemático de carácter general que limitaremos al caso particular del análisis de asociación. En adelante nos preocuparemos por definir el grado de asociación entre las variables en función de las conexiones lógicas que unen las proposiciones empíricas con la distribución de los datos. Con este propósito retomemos la discusión sobre la marginalidad.

En el caso de la teoría de la marginalidad cultural habíamos concluido que no estaba formulada con la precisión suficiente para permitirnos distinguir entre:

$$\begin{aligned}
 P_1: & B \rightsquigarrow A \\
 & B' \rightsquigarrow A' \\
 P_2: & B \rightsquigarrow A \\
 & B' \rightsquigarrow (A \text{ o } A')
 \end{aligned}$$

y

La primera proposición (P_1) sostiene que si sabemos que una observación tiene el atributo B en la variable X, entonces deberá presentar el A en la variable Y, y si es B' en X deberá ser A' en Y. Si las variables se relacionaron de acuerdo con P_1 , entonces la distribución bidimensional de frecuencias deberá ser tal que todas las observaciones que comiencen el atributo B deberán ser A y que todas aquellas que presenten la característica A también deberán ser B. Por tanto, todas las observaciones deberán ubicarse en la diagonal principal, como se observa en la distribución hipotética del cuadro 7.

CUADRO 7

Distribución bidimensional de frecuencias (hipotética) para el caso en que P_1 es válida

$Y \setminus X$	B	B'	Total
A	50 (100%)	0	50
A'	0	150 (100%)	150
Total	50	150	200

Al aplicar (6) a los datos de este cuadro, tenemos que:

$$D = 100 - 0 = 100$$

Si la proposición empírica hubiese sido la inversa y los casos se hubieran distribuido de acuerdo con ella, D sería igual a:

$$D = 0 - 100 = -100$$

Estos resultados nos llevan a la conclusión de que si seguimos el camino de las diferencias porcentuales para examinar la consistencia entre la proposición empírica (del tipo P_1) y la distribución de las observaciones, entonces D podrá asumir los valores situados dentro del rango de variación limitado por ± 100 . Ya sabemos que cuando D asume el valor nulo las variables son estadísticamente independientes.

En el caso de que la segunda proposición (P_2) sea válida se derivan las siguientes consecuencias en la distribución de las observaciones: i si un caso es B en X , entonces deberá presentar A en Y y it si es B' en X entonces puede ser A o A' en Y , sin que esto contradiga la teoría que originó la proposición empírica. Toda observación que tenga el atributo B en X , deberá tener también el atributo A en Y , pero no todo A será necesariamente B . Para que esta proposición sea corroborada por los datos no se requiere que las observaciones se desplieguen sólo por la diagonal principal, sino que también aparezcan en la casilla de la primera línea y en la segunda columna.

CUADRO 8

Distribución de frecuencias (hipotética) para el caso en que P_2 es válida

$Y \setminus X$	B	B'	Total
A	50 (100%)	30 (60%)	80
A'	0	20 (40%)	20
Total	50	50	100

A pesar de que la distribución de frecuencias de las variables X y Y se ajusta a la proposición P_2 , la diferencia porcentual dista bastante de alcanzar el valor máximo:

$$D = 100 - 60 = 40$$

Este ejemplo muestra que aun cuando la proposición empírica prediga adecuadamente la distribución de las observaciones, el valor de D se puede encontrar muy lejos del valor máximo posible. Es fácil ver que, en general, cuando la proposición teórica da lugar a una proposición empírica del tipo de P_2 , el valor de D que corresponde a un ajuste perfecto con la distribución de los casos no necesariamente alcanzará el valor máximo. En general, a cada distribución bidimensional de frecuencias corresponderá un D máximo posible diferente, lo que implica que para emitir un juicio sobre el grado de relación entre dos variables será necesario confrontarlo con el valor específico de D . De este modo se podrá calibrar el valor particular que asumió la discrepancia.

La indeterminación de los límites de D , bajo las condiciones impuestas por P_2 , es una consecuencia de su definición en términos de lejanía o proximidad respecto de la independencia estadística. El sentido que tenga un valor de D dependerá de que sea posible definir conceptualmente un valor límite a la asociación, de manera que se pueda juzgar en cada caso concreto el grado de lejanía (cercanía) respecto de la independencia y simultáneamente de cercanía (lejanía) a la asociación, tal como se definió. Ahora bien, en este trabajo entenderemos por asociación el grado de concordancia entre una proposición empírica y su correspondiente distribución observada de frecuencias de X y Y .

Como conclusión parcial podríamos decir que cuando el análisis se efectúa recurriendo a la comparación de las diferencias de porcentajes (es decir cuando se calcula D) es posible formarse una idea del grado de asociación entre las variables dicotómicas X y Y si la proposición empírica es del tipo P_1 . En este caso D puede asumir todos los valores entre ± 100 con sus correspondientes interpretaciones sistematizadas en (7). Pero si la hipótesis es del tipo P_2 , la asignación de contenido a un D específico no será tan simple por cuanto sus límites variarán para cada distribución de frecuencias observada. Se podrá tener una idea de la lejanía respecto de la independencia pero no de la cercanía respecto de la asociación, a menos que se proceda a calcular en cada caso el valor máximo (mínimo si la relación es inversa) de D .

El otro camino para medir el grado de asociación entre las variables X y Y (comparar las frecuencias observadas con las esperadas) surge, al igual que el análisis de porcentajes, por contraste con la independencia estadística: el valor observado de cada casilla de la tabla se compara con el que se obtendría (esperado) si las variables fuesen independientes. Es decir, esta estrategia analítica tampoco se basa en una idea clara de la asociación que permita originar límites de un intervalo cuyos extremos correspondan a una idea bien definida de asociación entre las variables. En la próxima sección nos dedicaremos a examinar la relación

entre las proposiciones empíricas y la distribución de frecuencias conjunta; así que no abundaremos en mayores detalles.

Nos limitaremos por el momento a destacar algunas propiedades estadísticas de las discrepancias entre las frecuencias observadas y esperadas.

En primer lugar, hay que llamar la atención sobre el hecho de que en el cuadro 6 las diferencias entre las frecuencias observadas y esperadas son iguales en valor absoluto (92) en las cuatro casillas. Esto no es casual; corresponde a una propiedad fácil de entender en las tablas de dos por dos. Como las distribuciones marginales por líneas y columnas son las mismas para ambas frecuencias, entonces el valor de una discrepancia en una casilla necesariamente originará otra de la misma magnitud, pero de signo contrario. Por tanto, para saber cuánto se desvía la distribución de frecuencias observada de la distribución generada bajo el supuesto de independencia, bastará con calcular la discrepancia de una sola casilla; lo habitual es que se elija la de la primera línea y la primera columna (casilla 1,1).

También hay que notar que el valor que asume la diferencia entre la frecuencia observada y la esperada en la casilla (1,1), es diferente en los cuadros 7 y 8 a pesar de que en ambos casos la distribución de frecuencias es consistente con la proposición empírica. Para las observaciones del cuadro 7 tenemos:

$$D_{11} = 50 - \frac{50 \times 50}{200} = 37.5$$

y para las del cuadro 8:

$$D_{11} = 50 - \frac{50 \times 80}{100} = 10$$

Si D midiese la asociación como el grado de correspondencia entre la distribución que esperaríamos en virtud de los planteamientos teóricos y las frecuencias efectivamente observadas entonces ambos valores de D_{11} aunque numéricamente diferentes deberían corresponder a valores máximos. Esto nos lleva a concluir que de nuevo estamos en presencia de una escala cuyos valores extremos no están definidos, de modo que no podemos calibrar un valor particular de D_{11} obtenido en una tabla cualquiera. Sabemos que $D_{11} = 0$ significa que las variables son estadísticamente independientes y que en la medida en que nos alejemos del valor nulo aumentará el grado de asociación, sea cual sea el significado que se dé a esta palabra.

En adelante el propósito central de nuestros desarrollos será dar sentido al concepto de asociación.

2.3.1. El determinante como medida de asociación

El determinante de una tabla de dos por dos presenta un interés limitado desde el punto de vista de su aplicación. Su importancia se sitúa en el dominio teórico de la estadística, ya que es el puente para vincular los coeficientes de asociación de uso habitual con la medida de la fuerza de la relación, que descansa en las diferencias entre las frecuencias observadas y esperadas. En las secciones siguientes se estudiará la relación funcional entre el determinante y los coeficientes Q de Yule y Φ . En este apartado nos dedicaremos a estudiarlo como medida de asociación y a establecer su vínculo con D_{11} .

Es relativamente simple demostrar²⁰ que:

$$(9) \quad ND_{11} = N(A\Delta B)N(A'\Delta B') - N(A'\Delta B)N(A\Delta B')$$

Si llamamos Δ al miembro de la derecha:

$$(10) \quad \Delta = N(A\Delta B)N(A'\Delta B') - N(A'\Delta B)N(A\Delta B')$$

Donde Δ simboliza al determinante de la tabla de dos por dos y la ecuación (9) muestra que la discrepancia entre las frecuencias observada y esperada de la casilla (1,1) multiplicada por el total de observaciones (N) es igual al valor de Δ .

El primer componente del lado derecho de la ecuación (10) es el producto del número de casos en las casillas (1,1) y (2,2) y se puede interpretar como la densidad, peso o importancia de la diagonal principal. Tam-

²⁰De acuerdo con la ecuación (8):

$$D_{11} = N(A\Delta B) - E(A\Delta B) = N(A\Delta B) - \frac{N(A)N(B)}{N}$$

$$ND_{11} = N[N(A\Delta B)] - N(A)N(B)$$

Al escribir el total N y las frecuencias marginales en términos de las frecuencias conjuntas:

$$ND_{11} = [N(A\Delta B) + N(A'\Delta B) + N(A\Delta B') + N(A'\Delta B')] [N(A\Delta B)] - [N(A\Delta B) + N(A'\Delta B)] [N(A\Delta B) + N(A'\Delta B)]$$

Realizando los productos:

$$ND_{11} = [N(A\Delta B)]^2 + N(A'\Delta B)N(A\Delta B) + N(A\Delta B)N(A'\Delta B) + N(A'\Delta B)N(A\Delta B) - N(A\Delta B)^2 - N(A'\Delta B)N(A\Delta B) - N(A\Delta B)N(A'\Delta B) - N(A'\Delta B)N(A\Delta B)$$

y al cancelar términos llegamos a:

$$ND_{11} = N(A'\Delta B)N(A\Delta B) - N(A'\Delta B)N(A\Delta B') = \Delta$$

bién puede pensarse como la resultante del grado de asociación directa entre las variables ya que mientras mayor sea la fuerza de esta relación, mayor será la cantidad de observaciones ubicadas en la diagonal principal. Al contrario, mientras mayor sea el peso de las casillas (1,1) y (2,2) mayor será la fuerza de la relación positiva.

El segundo producto que compone la igualdad (10) admite una conceptualización similar. Muestra la densidad, peso o importancia de la diagonal secundaria y por lo tanto se puede interpretar como indicador del grado de relación inversa entre X y Y. El número de observaciones en la diagonal secundaria es una resultante de la fuerza de la asociación negativa.

De acuerdo con esta manera de conceptualizar los componentes de (10), Δ es un índice de asociación que compara (por diferencia) los pesos de las diagonales de la tabla o, en otros términos, que mide el grado de relación en una distribución bidimensional de frecuencias, como la resultante de la asociación positiva menos la negativa.

Ya sabemos que si las variables X y Y son estadísticamente independientes, entonces $D_{11} = 0$. Al sustituir este valor en las igualdades (9) y (10), se concluye que Δ también será nulo.

Examinemos los valores límite que puede alcanzar el determinante. Si $N(A'AB) = 0$, o $N(AAB') = 0$; o $N(A'AB) = N(AAB') = 0$ entonces tenemos que:

$$\Delta_{\max} = N(AAB)N(A'AB')$$

y si:

$$N(AAB) = 0; \text{ o } N(A'AB') = 0; \text{ o } N(AAB) = N(A'AB') = 0$$

entonces se tiene:

$$\Delta_{\min} = -N(A'AB)N(AAB')$$

Por tanto, el rango de variación para el determinante será:

$$\Delta_{\min} \leq \Delta \leq \Delta_{\max}$$

Un valor alto de Δ significará que la relación directa entre las variables es fuerte; un valor pequeño, estrecha asociación inversa. Para calcular un valor de Δ correspondiente a una tabla específica se deberá calcular cada vez su máximo y su mínimo.

Es importante notar que este coeficiente es sensible a transformacio-

nes proporcionales. Si se amplifican²¹ todas las frecuencias de la tabla por una constante K mayor que la unidad, entonces:

$$\Delta_2 = K^2 N(AAB)N(A'AB') - K^2 N(A'AB)N(AAB')$$

$$\Delta_2 = K^2 [N(AAB)N(A'AB') - N(A'AB)N(AAB')]$$

$$\Delta_2 = K^2 \Delta_1$$

donde Δ_1 y Δ_2 simbolizan, respectivamente, el determinante antes y después de aplicar K. La amplificación de todas las frecuencias de la tabla por la misma constante no modifica el grado de relación entre las variables; sin embargo, el valor de este coeficiente aumenta con el cuadrado de la constante. Por supuesto que Δ_{\max} y Δ_{\min} también varían con el cuadro de la constante, por lo que la posición absoluta del Δ calculado, se altera, no así su ubicación relativa. Por tanto, al aplicar el determinante a dos distribuciones de frecuencias, la comparación de sus valores sólo tendrá sentido si previamente se los localiza dentro de sus correspondientes intervalos de variación.

Para evitar la sensibilidad de Δ a las transformaciones proporcionales se puede recurrir a normalizar las distribuciones pasando todas las frecuencias a porcentajes respecto al total. Esto equivale a definir un nuevo determinante (Δ') relacionado con el anterior (Δ) a través de:

$$(11) \quad \Delta' = \frac{\Delta}{N^2} = \frac{N(AAB)}{N} \frac{N(A'AB')}{N} - \frac{N(A'AB)}{N} \frac{N(AAB')}{N}$$

cuyos límites son:

$$(12) \quad \Delta'_{\max} = \frac{N(AAB)}{N} \cdot \frac{N(A'AB')}{N}$$

y

$$(13) \quad \Delta'_{\min} = -\frac{N(A'AB)}{N} \cdot \frac{N(AAB')}{N}$$

Al aplicar una constante de proporcionalidad K a la distribución conjunta, tanto los numeradores como los denominadores de (11), (12) y (13) se verán afectados por K^2 y por consiguiente no se modificará el valor de Δ' .

²¹El argumento también es válido si se aplica una reducción.

A pesar de que hemos controlado el efecto de las transformaciones proporcionales, los límites del determinante siguen siendo variables de una distribución a otra, lo que implica que la comparación directa de los Δ' carece de sentido. En las secciones siguientes presentaremos dos coeficientes de asociación que corresponden a diferentes maneras de resolver el problema de los límites del determinante.

Antes de concluir este apartado queremos recordar que tanto Δ como Δ' por ser funciones de D_{11} nos proveen de una definición de asociación sólo como desvío con respecto a la independencia estadística, lo que no nos permite disponer de una interpretación razonable para juzgar el grado de relación entre X y Y .

2.3.2. Coeficiente Φ (Φ)

Este coeficiente cierra el recorrido del determinante al intervalo ± 1 , normalizándolo por la raíz cuadrada del producto de las frecuencias marginales:

$$(14) \quad \Phi = \frac{\Delta}{\sqrt{N(A)N(A')N(B)N(B')}} = \frac{N(A\Delta B)N(A'\Delta B') - N(A\Delta B')N(A'\Delta B)}{\sqrt{N(A)N(A')N(B)N(B')}}.$$

Cuando X y Y son estadísticamente independientes $\Delta = 0$ y $\Phi = 0$. Si la distribución de frecuencias es tal que todas las observaciones se ubican en la diagonal principal tenemos:

TABLA 6

Distribución bidimensional de frecuencias que origina un Φ máximo

$Y \setminus X$	B	B'	Total
A	$N(A\Delta B)$	0	$N(A)$
A'	0	$N(A'\Delta B')$	$N(A')$
Total	$N(B)$	$N(B')$	N

de manera que se cumple que: $N(A\Delta B) = N(A) = N(B)$

$$N(A'\Delta B') = N(A') = N(B') \text{ y } N(A\Delta B') = N(A'\Delta B) = 0$$

Si se reemplazan estas igualdades en la ecuación (14) tenemos:

$$\Phi = \frac{N(A)N(A') - 0}{\sqrt{N(A)N(A')N(B)N(B')}} = 1$$

De donde se concluye que cuando las frecuencias de la diagonal secundaria son nulas, el coeficiente Φ alcanza su valor máximo. Supongamos ahora que las frecuencias de la diagonal principal son nulas:

TABLA 7

Distribución bidimensional de frecuencias que origina un Φ mínimo

$Y \setminus X$	B	B'	Total
A	0	$N(A\Delta B')$	$N(A)$
A'	$N(A'\Delta B)$	0	$N(A')$
Total	$N(B)$	$N(B')$	N

Las frecuencias de esta tabla satisfacen las siguientes relaciones:

$$N(A'\Delta B) = N(B) = N(A'); \quad N(A\Delta B') = N(A) = N(B')$$

y

$$N(A\Delta B) = N(A'\Delta B') = 0$$

Al sustituir estos valores en (14) se tiene:

$$\Phi = \frac{0 - N(A)N(A')}{\sqrt{N(A)N(A')N(B)N(B')}} = -1$$

En consecuencia Φ asume su valor más pequeño cuando los casos se ubican sobre la diagonal secundaria y en la principal sólo hay valores nulos.

Tenemos entonces que:

$$-1 \leq \Phi \leq 1$$

donde el valor 1 indica que entre X y Y hay máxima asociación directa, el -1 que hay máxima relación inversa y cuando es igual a 0 que no hay asociación.²²

¿Pero qué significa que hay máxima relación directa entre X y Y ? Lo único que sabemos es que se trata del valor más alejado de la independencia estadística, pero no se sabe muy bien a qué se aproxima por

²²Para evitar complejidades matemáticas en la determinación de los límites del coeficiente Φ se optó por el camino de la comprobación en lugar de la demostración rigurosa. Se seguirá el mismo procedimiento cuando se exponga el coeficiente Q .

cuanto no se ha definido conceptualmente el sentido de los límites del coeficiente.

Con base en la distribución de frecuencias de la tabla 6, se puede decir que el coeficiente Φ mide, en una escala que va de 0 a 1, el grado de aproximación entre la proposición empírica:

$$P_1: \begin{array}{c} B \text{---} \curvearrowright \text{---} A \\ B' \text{---} \curvearrowright \text{---} A' \end{array}$$

y la distribución efectiva de los datos. Valores de Φ positivos y cercanos a cero indican no sólo independencia estadística sino también lejanía respecto a P_1 . Los valores cercanos a 1 indican alejamiento de la independencia, pero también un alto grado de adecuación entre P_1 y la distribución bidimensional de los datos.

Un examen similar, pero esta vez aplicado a la tabla 7, nos permite afirmar que los valores de Φ entre 0 y -1 indican el grado de adecuación entre los datos y la proposición empírica:

$$P_1: \begin{array}{c} B \text{---} \curvearrowright \text{---} A' \\ B' \text{---} \curvearrowright \text{---} A \end{array}$$

De manera análoga al caso de la relación directa, los valores de Φ se ubican no sólo respecto al punto de independencia estadística ($\Phi = 0$) sino también en función del máximo grado de adecuación entre P_1 y la distribución bidimensional de frecuencias ($\Phi = -1$).

En conclusión tenemos que:

$$\Phi = \begin{cases} 1 & \text{Si las observaciones se distribuyen según:} \\ & P_1: \begin{array}{c} B \text{---} \curvearrowright \text{---} A \\ B' \text{---} \curvearrowright \text{---} A' \end{array} \\ 0 & \text{Si } X \text{ y } Y \text{ son estadísticamente independientes} \\ -1 & \text{Si los casos se distribuyen de acuerdo con:} \\ & P_1: \begin{array}{c} B \text{---} \curvearrowright \text{---} A' \\ B' \text{---} \curvearrowright \text{---} A \end{array} \end{cases}$$

2.3.3. Educación y estatus socioeconómico: una aplicación de Φ

Para ilustrar el uso del coeficiente Φ tomaremos un ejemplo desarrollado por Raymond Boudon, quien dice:

Consideremos el siguiente problema. En la perspectiva generalmente aceptada de que las sociedades industriales son ampliamente meritocráticas, el nivel educativo alcanzado es uno de los principales determinantes del estatus. A partir de esto, es grande la tentación de concluir que el logro educativo debería ser también el principal factor de la movilidad social. En otras palabras, si el logro educativo es realmente un poderoso determinante del estatus, la probabilidad de que un individuo tenga mayor estatus que su padre debería ser mayor, en la medida que su nivel educativo es mayor. Inversamente, la probabilidad de que un individuo caiga a un estatus más bajo que el de su padre debería ser mayor, en tanto menor sea su nivel educativo.²³

Boudon mide el logro educacional a través de la variable logro educacional del hijo con relación al del padre (X). También el estatus socioeconómico vincula al del hijo en relación con el del padre (Y). Es evidente que ambas variables son tricotómicas, pero para los propósitos del ejemplo las hemos dicotomizado en:

$$X = \begin{cases} B & \text{Mayor nivel educacional del hijo en relación con el del padre} \\ B' & \text{Igual o menor nivel educacional del hijo en relación con el del padre} \end{cases}$$

$$Y = \begin{cases} A & \text{Mayor estatus socioeconómico del hijo con relación al del padre} \\ A' & \text{Igual o menor estatus socioeconómico del hijo con relación al del padre} \end{cases}$$

Ahora bien, del planteo de Boudon se puede deducir que la proposición empírica que se deriva del párrafo citado es del tipo:

$$\begin{array}{c} B \text{---} \curvearrowright \text{---} A \\ B' \text{---} \curvearrowright \text{---} A' \end{array}$$

Es decir, en las sociedades industriales modernas si los hijos alcanzan un nivel educativo mayor que el del padre, entonces su estatus socioeconómico también será mayor. Si, por el contrario, logran un grado de estudio, menor o igual, dicho estatus debería ser igual o menor que el del padre.

El cuadro que nos servirá para examinar el grado de concordancia entre la proposición empírica y la distribución de los datos se tomó de un estudio realizado por Centers en Estados Unidos:²⁴

²³Raymond Boudon, *Education, Opportunity & Social Inequality: Changing prospects in western society*, John Wiley & Sons, Nueva York, 1974, p. 3.

²⁴Robert Centers, "Education and Occupational mobility", *American Sociological Review*, 14, 1949, pp. 143-144. Cita tomada de R. Boudon, *op. cit.*, p. 3.

CUADRO 9

Estatus socioeconómico del hijo con respecto al del padre como función de su nivel educativo en relación con el del padre*

	Nivel educacional del hijo en relación con el del padre			Total
	Y \ X	Mayor	Igual o menor	
Estatus socioeconómico del hijo en relación con el del padre	Mayor	134	30	164
	Igual o menor	157	95	252
	Total	291	125	416

*En este cuadro presentamos las dos variables dicotomizadas, pero en el original están tricotomizadas. Se optó por formar una sola categoría con el igual y el menor porque de esa manera se respeta cabalmente la estructura de los datos de la tabla de Centers.

Iniciaremos el análisis con el examen de las distribuciones marginales para luego establecer el grado de relación entre X y Y. La variable X muestra el resultado de la expansión del sistema educativo en la sociedad industrial del siglo XX: la escolaridad ha aumentado continuamente, lo que conduce a que una alta proporción (70%) de los hijos tenga un nivel de educación mayor que el de sus padres. Un fenómeno totalmente opuesto ocurre con el estatus socioeconómico; pareciera que el desarrollo de los Estados Unidos proporciona o pone a disposición de la sociedad una cantidad decreciente de ubicaciones sociales mejores para los hijos que para los padres (39.4%). Tenemos entonces que el proceso de desarrollo educativo capacita (supuestamente) para ingresar a las mejores posiciones sociales, pero éstas son cada vez más escasas. El cuadro 9 nos muestra que hay 291 hijos con mayor nivel educativo que el de sus padres, pero sólo 164 puestos sociales que se ubican por encima del que desempeñan o han desempeñado los padres.

El análisis de marginales nos alerta en el sentido de que no puede haber una correspondencia perfecta entre la proporción empírica y la distribución de los datos al no cumplirse el requisito aritmético de que $N(A) = N(B)$ y $N(A') = N(B')$.

El cálculo del coeficiente Φ da el siguiente resultado:

$$\Phi = \frac{134 \times 95 - 157 \times 30}{\sqrt{164 \times 252 \times 125 \times 291}} = \frac{12\,730 - 4\,710}{38\,772.49} = \frac{8\,020}{38\,772.49}$$

$$\Phi = 0.21$$

Valor que denota un bajo grado de relación que lleva a concluir que el logro educativo no parece ser, al nivel de la sociedad global, el principal factor de movilidad social.

Para terminar con este ejemplo sería importante destacar que la conclusión que se obtuvo no invalida la meritocracia. En efecto, aun cuando la educación sea la palanca con la que cada individuo alcanza un mayor estatus socioeconómico, la restricción que impone el volumen de estatus mayores para los hijos hace que algunos que han obtenido un mayor nivel educativo sólo tengan acceso a posiciones sociales menores o iguales que las de sus padres. Este resultado nos muestra que a pesar de que existe una relación en lo individual entre educación y estatus socioeconómico, ésta no se manifiesta al nivel de lo social.²⁵

2.3.4. Coeficiente Q de Yule

Este coeficiente de asociación aplicable a tablas de dos por dos proviene de otra opción para cerrar el recorrido del determinante al intervalo ± 1 :

$$(15) \quad Q = \frac{\Delta}{N(A\Delta B)N(A'\Delta B') + N(A\Delta B')N(A'\Delta B)} = \frac{N(A\Delta B)N(A'\Delta B') - N(A'\Delta B)N(A\Delta B)}{N(A\Delta B)N(A'\Delta B') + N(A'\Delta B)N(A\Delta B)}$$

Sabemos que si X y Y son estadísticamente independientes entonces $\Delta = 0$ y por tanto $Q = 0$.

Para que Q alcance el valor máximo basta con que una frecuencia de la diagonal secundaria sea nula:

TABLA 8

Distribuciones bidimensionales de frecuencias que originan un Q máximo

Y \ X	B	B'	Total	Y \ X	B	B'	Total
A	N(A\Delta B)	0	N(A)	A	N(A\Delta B)	N(A\Delta B')	N(A)
A'	N(A'\Delta B)	N(A'\Delta B')	N(A')	A'	0	N(A'\Delta B')	N(A')
Total	N(B)	N(B')	N	Total	N(B)	N(B')	N

(a) (b)

En la tabla 8(a) se cumple que $N(A\Delta B) = N(A)$, $N(A\Delta B') = 0$ y $N(A'\Delta B') = N(B')$. Sustituyendo en (15) obtenemos:

$$Q = \frac{N(A)N(B') - 0}{N(A)N(B') + 0} = 1$$

²⁵Boudon analiza ampliamente esta situación en su libro, *Efectos perversos y el orden social*, Premia Editora de Libros, México, 1980.

y en la tabla 8(b) se tiene que $N(A\Delta B) = N(B)$, $N(A'\Delta B) = 0$; y $N(A\Delta B') = N(A')$. Al remplazar estas igualdades en (15) se llega a:

$$Q = \frac{N(B) N(A') - 0}{N(B) N(A') + 0} = 1$$

Este índice de asociación alcanza su valor máximo si $N(A\Delta B') = 0$ o $N(A'\Delta B) = 0$. Es evidente que también será igual a uno si $N(A\Delta B) = N(A'\Delta B) = 0$.

El coeficiente Q de Yule asume su valor mínimo cuando una frecuencia de la diagonal principal es nula:

TABLA 9

Distribuciones bidimensionales de frecuencias que originan un Q mínimo

$Y \setminus X$	B	B'	Total	$Y \setminus X$	A	A'	Total
A	0	$N(A\Delta B')$	$N(A)$	A	$N(A\Delta B)$	$N(A\Delta B')$	$N(A)$
A'	$N(A'\Delta B)$	$N(A'\Delta B')$	$N(A')$	A'	$N(A'\Delta B)$	0	$N(A')$
Total	$N(B)$	$N(B')$	N	Total	$N(B)$	$N(B')$	N
				(a)			
				(b)			

En la tabla 9(a) se cumple que $N(A\Delta B') = N(A)$; $N(A\Delta B) = 0$, y que $N(A'\Delta B) = N(B)$. Al sustituir estas igualdades en (15) se llega a:

$$Q = \frac{0 - N(B) N(A)}{0 + N(B) N(A)} = -1$$

De la tabla 9(b) se desprende que $N(A'\Delta B) = N(A)$; $N(A'\Delta B') = 0$; y que $N(A\Delta B') = N(B')$. Al sustituir en la ecuación (15) se tiene:

$$Q = \frac{0 - N(A') N(B')}{0 + N(A') N(B')} = -1$$

Se obtiene el valor mínimo de Q si $N(A\Delta B) = 0$ o $N(A'\Delta B') = 0$. Si ambas frecuencias son nulas entonces también $Q = -1$.

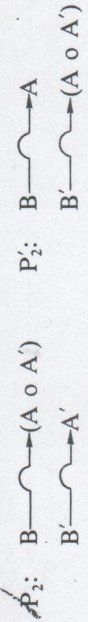
El rango de variación de la medida de asociación Q es:

$$-1 \leq Q \leq 1$$

El valor 1 se obtiene cuando hay máxima relación directa entre X y Y. El -1 cuando hay máxima relación inversa y es cero en los casos en que X y Y son estadísticamente independientes.

Como Q corresponde a una normalización particular de Δ también mide la asociación como lejanía a la independencia y deja indeterminada la definición.

Las distribuciones de frecuencias de la tabla 8 nos permiten sostener que el coeficiente Q de Yule mide, en una escala limitada por 0 y 1, el grado de aproximación entre la distribución de los datos y las proposiciones empíricas.



Los valores de Q positivos y cercanos a cero señalan simultáneamente cercanía respecto de la independencia y lejanía respecto de P_2 o P_2' . Los valores cercanos a la unidad nos dicen que hay un buen grado de ajuste entre P_2 o P_2' y la distribución de las observaciones. Hay que destacar que un valor específico del coeficiente de Yule no permite distinguir entre ambas proposiciones empíricas (que normalmente corresponderían a proposiciones teóricas diferentes), por lo que será necesario examinar la conformación de la distribución bidimensional de frecuencias para estar en condiciones de emitir un juicio acerca del grado de bondad de ajuste entre P_2 o P_2' (según sea el caso) y la información estadística.

CUADRO 10

Dos distribuciones bidimensionales (hipotéticas) diferentes que originan un mismo valor de Q

$Y \setminus X$	B	B'	Total	$Y \setminus X$	B	B'	Total
A	50	5	55	A	50	20	70
A'	20	10	30	A'	5	10	15
Total	70	15	85	Total	55	30	85
				(a)			
				(b)			

El cuadro 10 nos muestra dos distribuciones con un mismo valor para Q.

$$Q = \frac{50 \times 10 - 5 \times 20}{50 \times 10 + 5 \times 20} = \frac{400}{600} = 0.667$$

y

$$Q = \frac{50 \times 10 - 20 \times 5}{50 \times 10 + 20 \times 5} = \frac{400}{600} = 0.667$$

Sin embargo, la configuración del cuadro 10(a) tiende a mostrar un mayor grado de concordancia con P_2 que con P'_2 . La distribución del cuadro 10(b) parece reflejar un mayor nivel de adecuación con P'_2 que con P_2 .

En el caso de que las variables X y Y estén positivamente asociadas tenemos dos proposiciones empíricas cuyo grado de acuerdo con una distribución de frecuencias no se puede determinar sólo a partir del valor de Q por cuanto es incapaz de distinguir entre P_2 y P'_2 . Por tanto, para juzgar la bondad de ajuste entre una proposición empírica y una distribución de frecuencias específica hay que considerar la configuración de los casos.

Un análisis similar al que hemos realizado a partir de la tabla 8, pero esta vez realizado sobre la tabla 9 nos permite sostener que los valores de Q entre -1 y 0, miden el grado de armonía entre las observaciones y alguna de las siguientes proposiciones empíricas:

$$P'_2: B \rightsquigarrow A' \qquad P''_2: B \rightsquigarrow (A \text{ o } A')$$

$$B' \rightsquigarrow (A \text{ o } A') \qquad B' \rightsquigarrow A$$

Para distinguir entre P'_2 y P''_2 a partir de un valor negativo de Q es necesario realizar una inspección de la distribución en las cuatro casillas. En conclusión tenemos que:

1 Si las observaciones se distribuyen según:

$$P_2: B \rightsquigarrow (A \text{ o } A') \text{ o } P'_2: B \rightsquigarrow A$$

$$B' \rightsquigarrow A' \qquad B' \rightsquigarrow (A \text{ o } A')$$

Q = 0 Si X y Y son estadísticamente independientes.

1 Si los casos se distribuyen de acuerdo con:

$$P'_2: B \rightsquigarrow A' \text{ o } P''_2: B \rightsquigarrow (A \text{ o } A')$$

$$B' \rightsquigarrow (A \text{ o } A') \qquad B' \rightsquigarrow A$$

Para distinguir entre las dos proposiciones que corresponden al tramo positivo o negativo del recorrido de Q, es necesario examinar la configuración de los casos en las casillas.

2.3.5. Desarrollo social y movimiento obrero: una interpretación del caso argentino (1915-1945)

Para ilustrar el uso del coeficiente Q de Yule tomaremos como ejemplo una investigación que analiza el carácter de la lucha obrera en función del tipo de desarrollo socioeconómico:

Los sectores estratégicos o fundamentales para los diferentes modelos de acumulación son los que adquieren mayor peso político en tanto organización sindical. Hay una relación concreta entre los sectores clave del modelo de acumulación y desarrollo impuesto por las clases dominantes, la división internacional del trabajo y los niveles de organización, conciencia, práctica política y poder de dichos sectores sindicales. Ello hace que generalmente sean vanguardia de las luchas sindicales y que, por otra parte, tengan mayor poder de negociación con respecto al aparato de dominación estatal.

Analizamos ahora esta relación en el modelo de acumulación agroexportador implementado desde 1915 hasta 1930 para corroborar la hipótesis.²⁶

Un poco más adelante dice:

Todo el sistema de transporte terrestre se construía en función del proyecto exportador. El ferrocarril, que era el medio de transporte más eficaz del momento, se desarrolló construyendo todas las redes hacia las ciudades puerto, sin conexiones internas.²⁷

También plantea que:

Intentaremos demostrar cómo el sector de transportes que constituía el sector económico estratégico del modelo de acumulación del proceso político económico, es el que se convierte en vanguardia, en términos políticos. O por lo menos es el que posee mayor poder político y de movilización de sus agremiados.

Vamos a corroborar técnicamente que este sector fue el más combativo; por otra parte, que sus luchas no fueron específicamente salariales, sino que en su mayoría se entablaron por motivos de organización y de mejoras en las condiciones de trabajo.²⁸

Sostiene que a partir de 1930 se produce una reorientación del modelo de acumulación y que:

Hay una clara relación entre el proceso de industrialización y el detrimento de la ganancia en el sector transportes. Otro será el sector económicamente

²⁶ Ana Jaramillo, "Movimiento obrero y acumulación de capital: el caso argentino", en *Revista Mexicana de Ciencias Políticas y Sociales*, núm. 89, julio-septiembre, 1977, p. 181.

²⁷ *Ibidem*, p. 181.

²⁸ *Ibidem*, p. 182.

estratégico en el nuevo modelo de acumulación, y el capital británico tratará de vender el sector que hasta ahora había sido la columna vertebral de su dominación.

Este cambio que se produce en la organización y estrategia del capital tiene repercusiones en la organización y estrategia del trabajo.

El sector de vanguardia dejará de ser el sector transportes, dejará de ser el más conflictivo y la industrialización llamará a una nueva distribución ocupacional que trae aparejadas dos consecuencias fundamentales: el proceso de industrialización, con la incorporación masiva de fuerza de trabajo a la industria y, por otra parte, el proceso de urbanización como consecuencia lógica.

Los sectores de sustitución de importaciones en los cuales se basaba la industrialización, pasan a ser los más conflictivos (...) Analizando entonces el periodo 1915-1930 y el posterior 1930-1945, explicaremos los sectores estratégicos de la economía y su relación con el proceso de acumulación y el proceso político sindical. Por otra parte, los analizamos a partir de su forma específica de lucha: el conflicto, la huelga.²⁹

La autora sostiene que debe haber una relación entre "hegemonía económica" y "proceso político sindical". El análisis histórico del tipo de desarrollo que ha seguido Argentina le permite concluir que hay un punto de inflexión al inicio de la cuarta década de este siglo, cuando se pasó de un modelo orientado hacia afuera a otro volcado hacia el mercado interno; por tanto, para que la distribución de las observaciones concuerde con la proposición debe haber un cambio en los sectores específicos de donde provienen los obreros que conforman la vanguardia.

El concepto hegemonía económica se objetiva distinguiendo las ramas de la actividad en que se basa el modelo de acumulación de aquellas que son accesorias y dentro de esos grupos elige una de cada tipo.³⁰ Según este criterio define como variable explicativa (X) las ramas de la actividad productiva y supone que ella sólo puede asumir dos valores: ferrocarriles y vestido. El primero es el sector hegemónico indiscutido en el modelo agroexportador y sobre el segundo descansa el desarrollo del modelo industrializador.

Por otra parte en el texto se sostiene que el proceso político sindical se objetiva a través de la lucha: el conflicto, la huelga. La variable explicada o dependiente (Y), tipo de lucha, la clasifica según su naturaleza en salarial y organizacional. Esta dicotomización lleva implícitos los supuestos de que es posible calificar un conflicto o una huelga en sólo una de ambas categorías y que no hay otras causas que provoquen su estallido.

²⁹*Ibidem*, p. 186

³⁰El procedimiento es discutible por cuanto no trabaja con una dicotomía.

Tenemos entonces que la variable independiente es "ramas de la actividad productiva", con ferrocarriles y vestido, y que la variable explicada es "conflictos laborales", dicotomizada en salariales y organizacionales.

La proposición teórica dice que los conflictos desarrollados por las organizaciones obreras de las actividades hegemónicas obedecen fundamentalmente a motivos de organización. No hay una declaración explícita acerca del carácter de los conflictos en las actividades no hegemónicas, pero a juzgar por el tratamiento de los datos, el planteo es que en estos sectores se pueden presentar ambos tipos de conflictos sin que ello invalide la proposición.

En función de la hipótesis (proposición teórica) que se maneja y definiendo:

$$\left. \begin{array}{l} X \\ B \\ B' \end{array} \right\} \begin{array}{l} \text{Ferrocarriles} \\ \text{Vestido} \end{array}$$

$$\left. \begin{array}{l} Y \\ A \\ A' \end{array} \right\} \begin{array}{l} \text{Salario} \\ \text{Organización} \end{array}$$

$$B \rightsquigarrow A'$$

$$B' \rightsquigarrow (A \text{ o } A')$$

entonces la proposición empírica válida para el periodo 1915-1930 será:

y

Según hemos visto en la sección anterior el examen del grado de adecuación entre esta proposición empírica y la distribución de las observaciones se debe realizar a través del coeficiente Q de Yule. En el estudio de referencia encontramos que:

CUADRO II

Distribución del número de huelgas, según causas (tipo de lucha) por ramas de la actividad productiva (hegemonía sectorial) en Argentina, 1919

Causas	Y \ X	Rama de actividad productiva		Total
		Ferrocarriles	Vestido	
Salario		2	53	55
Organización		28	18	46
Total		30	71	101

El análisis de las marginales permite afirmar que la distribución de conflictos por rama es altamente favorable al sector no hegemónico (a vestido corresponde un 70%) y la distribución por causas es más o menos equilibrada entre salario y organización (55 y 46 por ciento), respectivamente.

Para examinar la relación entre hegemonía económica y proceso político sindical se estudia el grado y la forma de la asociación entre sus indicadores. Lo primero se obtiene con el cálculo de un coeficiente que nos proporcione una medida de la adecuación entre la proposición empírica y la distribución de las observaciones y lo segundo con el análisis de la conformación interna del cuadro.

La proposición empírica nos dice que la asociación entre X y Y debe medirse a través de Q y su valor debe ser negativo:

$$Q = \frac{2 \times 18 - 53 \times 28}{2 \times 18 + 53 \times 28} = -0.95$$

Los datos y la proposición empírica tienen un alto grado de correspondencia. La relación es inversa y se aproxima bastante al valor límite inferior -1. La conformación de las frecuencias conjuntas también se mueve de acuerdo con la proposición empírica,³¹ los conflictos en los ferrocarriles son básicamente de índole organizacional.

Habría sido interesante examinar un cuadro con las mismas variables una vez producido el tránsito a un modelo de acumulación basado en el desarrollo del mercado interno, puesto que en tal caso la relación debería haber cambiado a directa y fuerte. Sin embargo, la carencia de información fue insoslayable, de modo que el estudio debió encamarse por otros derroteros.

Ejercicios³²

A continuación presentamos seis problemas tomados de diferentes textos. En varios tuvimos que realizar algunos cambios con el propósito de ajustarlos al caso particular de las tablas de dos por dos, lo que quizá haya desvirtuado los argumentos de los autores. Esto se puede evitar con la lectura de los trabajos originales; para ello incluimos las referencias precisas al final de esta sección.

En los primeros cinco ejercicios hay que intentar establecer la proposición teórica, siempre que sea posible,³³ y su correspondiente proposición empírica y proceder al análisis estadístico adecuado.

³¹No debe olvidarse que a un valor del coeficiente Q de Yule le pueden corresponder dos proposiciones empíricas.

³²Véase la bibliografía al final de los ejercicios.

³³En ocasiones los autores dan el mismo nombre al concepto y a la variable estadística, de modo que no resulta fácil distinguir entre los dos tipos de proposiciones.

1. Jorge N. Solomonoff, en su estudio sobre la ideología del movimiento obrero argentino nos dice que:

En el secular conflicto entre débiles y poderosos, poseedores y desposeídos, los trabajadores siempre soportaron el peso de las fuerzas represivas propias de todo sistema de dominación. Pero la visualización del conjunto de las relaciones sociales como un sistema cuyas partes juegan un papel preciso dentro del todo, implica un proceso de generalización, lo que a su vez requiere de una serie de mediaciones (p. 265).

Después de sostener que el papel ideológico de la religión decreció sistemáticamente durante el siglo XIX, plantea que las masas movilizadas, de las cuales surgía el nuevo proletariado urbano, eran más receptivas que los estratos populares para absorber puntos de vista racionales del mundo.

En seguida, señala que:

Dentro del campo de la organización obrera, las élites dirigentes disponían de mayor grado de libertad para la formulación de definiciones de la situación político-social como marco de orientación teleológica y el establecimiento de metas para la acción. En estos sistemas valorativos y metas propuestas para la acción obrera en la época, distinguimos dos líneas principales: una considera al Estado, a partir de su carácter represivo, esencialmente un elemento conservador del sistema de desigualdades económicas y políticas. De donde resulta que el establecimiento de una sociedad igualitaria requiere como condición la eliminación del aparato estatal. Es decir, la posición anarquista. La propuesta socialista espera en cambio que las clases asalariadas se apoderen del Estado, por vía de un partido que las representa, a fin de utilizarlo como instrumento de sus propias reivindicaciones. Dentro de esta conceptualización se confieren al Estado cualidades que le permiten actuar como árbitro entre los contrapuestos intereses sectoriales y su utilización por los trabajadores como agente de una distribución equitativa de la riqueza social (pp. 265-266).

A partir de los documentos en que se consignan los acuerdos y las resoluciones de los congresos obreros argentinos celebrados entre 1890 y 1915 intenta capturar:

el grado en que las definiciones teleológicas propuestas por las distintas élites (dirigentes gremiales) proveían de expresión simbólica y pautas de acción a las aspiraciones de los estratos asalariados (p. 266).

Después de comprobar que en la totalidad de los documentos anarcosindicalistas (15) y socialistas (7) examinados se conceptúa a la sociedad global dividida en clases opuestas, afirma que:

entre la visualización de la sociedad global como una estructura de grupos o de clases en conflicto y la definición del Estado, y por tanto, las expectativas respecto a las posibilidades de su utilización como medio por las distintas fuerzas sociales no existe, como hemos visto, una relación biunívoca ni independiente de los sistemas ideológicos (p. 267).

Ítems	Total	Anarcosindical	Socialista
El Estado como instrumento de dominación de la clase trabajadora	16	15	1
El Estado como mediador en los conflictos entre el capital y el trabajo	6	—	6

2. A partir de una muestra estadística de los estudiantes de Berkeley (1950), S.M. Lipset realizó un estudio que originó el artículo "La formación de opiniones en una situación de crisis". Dentro del tema central —posición ideológica y formación de opiniones—, se interesó por estudiar las vinculaciones entre la ideología de los estudiantes y sus opiniones, favorables o desfavorables, sobre que los profesores "comunistas" impartan clases. De este estudio hemos seleccionado el siguiente cuadro:

Política de empleo a profesores comunistas	Ideología de los estudiantes		Total
	Liberales	Conservadores	
En favor	130	76	206
En contra	82	102	184
Total	212	178	390

3. Ricardo Cinta, en su estudio "Burguesía nacional y desarrollo", define su posición teórico-metodológica en los siguientes términos:

El análisis integrado del desarrollo busca rescatar, de los conceptos de desarrollo y subdesarrollo, la especificidad estructural que define la relación entre centro-periferia entre los países avanzados y los de desarrollo tardío y las formas políticas de dominación, interna y externa, que posibilitan y condicionan el desenvolvimiento de las áreas atrasadas. El concepto de *dependencia*, que alude directamente a estas condiciones de existencia y funcionamiento del sistema económico y del sistema político, mostrando las vinculaciones entre ambos, tanto en lo que se refiere al plano interno de los países como al externo, se constituye así, en la *categoría analítica* central del marco interpretativo (p. 168).

Después de analizar el proceso de industrialización mexicana, con lo cual comprueba que después de la segunda guerra mundial se plantea

un nuevo tipo de vínculo con el sistema capitalista internacional y, en particular, con la economía estadounidense (p. 174) pasa a estudiar la conformación de la estructura industrial y de las élites empresariales. Plantea que:

El grado de integración y consolidación de la industria que se fue logrando en el país—a partir del proceso sustitutivo conduxo, de manera paralela, a la formación, dentro del ámbito de las clases productoras, de ciertas élites empresariales cuyas posiciones dentro del sistema industrial las coloca en una situación dominante. Los límites de dichas élites empresariales en el control del proceso económico pueden establecerse a partir del análisis de su participación dentro de la estructura productiva y de los grados de control que en ella le corresponden (p. 183).

Dentro de esta óptica presenta un cuadro que se refiere a la propiedad nacional o extranjera de las empresas del sector I (productor de bienes de capital) que se ubican entre las 200 más grandes del país.

Distribución de las mayores empresas productoras de bienes de capital (México, 1965)

	Extranjeras	Nacionales	Total
Primeras 50 más grandes	8	3	11
Desde la 51 a la 200	8	8	16
Total	16	11	27

4. En el artículo "Migración y marginalidad ocupacional en la ciudad de México", Humberto Muñoz, Orlandina de Oliveira y Claudio Stern, analizan las relaciones entre los procesos migratorios y la marginalidad ocupacional en la ciudad de México. Estudian hasta qué punto la condición migratoria es un elemento que condiciona la inserción laboral. Los datos que analizan los obtuvieron de una encuesta por muestra realizada en el área metropolitana de la ciudad de México de noviembre de 1969 a febrero de 1970. Para mayores detalles sobre la forma de elaboración de las categorías utilizadas, se puede consultar el apéndice I del artículo. A continuación incluimos el cuadro I del trabajo citado:

Distribución de la población ocupada y remunerada de 21 a 60 años de edad por condición de nativo o no nativo, y por grupos ocupacionales marginales y no marginales (porcentajes)

Grupos ocupacionales	Condición	
	Nativo	No nativo
Marginales	28.1	15.8
No marginales	71.9	84.2
Total	100.0	100.0

5. En el trabajo de Nilda Sito, "Estructura ocupacional, desarrollo y sindicalismo en los países latinoamericanos", se nos entrega la siguiente distribución bidimensional de frecuencias de 15 países de la región:

Sindicalismo	Estructura		Total
	Tradicional	Moderna	
Bajo	7	0	7
Alto	2	6	8
Total	9	6	15

La autora sostiene que las características sectoriales del patrón de desarrollo en América Latina se captan mejor por medio de la estructura ocupacional que por la participación de cada sector en el producto. El concepto estructura lo objetiva en la distribución sectorial de la población económicamente activa y sobre esta base procede a clasificar a cada país en la categoría tradicional o moderna.

Por otra parte, considera que la organización sindical es el mejor indicador de la existencia de movimientos obreros agrupados en torno a los intereses reivindicativos y se mide como el porcentaje de obreros sindicalizados sobre el total de la población económicamente activa en 1960.

6. En la obra de Emile Durkheim, *El suicidio*, leemos textualmente:

Suiza es de muy interesante estudio desde este mismo punto de vista, pues como se encuentran en ella poblaciones francesas y alemanas se puede observar separadamente la influencia del culto sobre cada una de las dos razas, y esta influencia es la misma en una y otra. Los cantones católicos producen cuatro o cinco veces menos suicidios que los de los protestantes, cualquiera que sea su nacionalidad.

Cantones franceses	Cantones alemanes		Conjunto de cantones de todas las nacionalidades
	Católicos	Protestantes	
83 suicidios por millón de habitantes	Católicos	87 suicidios	Católicos
453 suicidios por millón de habitantes	Protestantes	293 suicidios	Mixtos
			Protestantes
			86.7 suicidios
			212 suicidios
			326.3 suicidios

La acción del culto es tan poderosa que domina a las demás. Por otra parte, en numerosos casos se ha podido determinar directamente el número de suicidios por millón de habitantes de cada población con fesional; he aquí las cifras encontradas por distintos investigadores:

Suicidios en los diferentes países por un millón de sujetos de cada confesión

	Protestantes	Católicos	Judíos	Investigadores
Austria (1852-1859)	79.5	51.3	20.7	Wagner
Prusia (1849-1855)	159.9	49.6	46.4	Wagner
Prusia (1869-1872)	187	69	96	Morselli
Prusia (1896)	240	100	180	Prinzling
Baden (1852-1862)	139	117	87	Legoyt
Baden (1870-1874)	171	136.7	124	Morselli
Baden (1878-1888)	242	170	210	Prinzling
Baviera (1844-1856)	135.4	49.1	105.9	Morselli
Baviera (1844-1891)	224	94	193	Prinzling
Wurtemberg (1846-1860)	113.5	79.9	65.6	Wagner
Wurtemberg (1873-1876)	190	120	60	Durkheim
Wurtemberg (1881-1890)	170	119	142	Durkheim

Establezca las diferencias entre ambos procedimientos de validación. Además, emita un juicio acerca del grado de adecuación de la información numérica presentada para contrastar, en cada caso, la hipótesis sustentada.

REFERENCIAS BIBLIOGRÁFICAS DE LOS EJERCICIOS

- Cinta, Ricardo. "Burguesía nacional y desarrollo", en *El perfil de México en 1980*, vol. 3, Siglo XXI, México, 4a. edición, 1976.
- Durkheim, Emile. *El suicidio*, UNAM, México, 1a. ed., 1974.
- Lipset, S.M. "La formación de opiniones en una situación de crisis", en Boudon y Lanzafeld, *Metodología de las Ciencias Sociales*, II: *Análisis empírico de la causalidad*, Editorial Laia, Barcelona, 1974.
- Muñoz, Humberto, Orlandina de Oliveira y Claudio Stern. "Migración y marginalidad ocupacional en la ciudad de México", en *El Perfil de México en 1980*, vol. 3, Siglo XXI, México, 4a. ed., 1976.
- Sito, Nilda. "Estructura ocupacional, desarrollo y sindicalismo en los países latinoamericanos", en *Revista Latinoamericana de Sociología*, Centro de Investigaciones Sociales, Instituto Torcuato Di Tella, Buenos Aires, 1971.
- Solomonoff, Jorge N. *Ideologías del movimiento obrero y conflicto social*, Editorial Proyección, Buenos Aires, 1971.

3. Asociación entre tres variables dicotómicas

En sentido estricto la proposición teórica de la investigación sobre las relaciones entre desarrollo social y movimiento obrero en Argentina, citada en el capítulo anterior implica la consideración simultánea de tres conceptos: hegemonía económica, proceso político sindical y modelo de acumulación. Este último concepto se objetiva por medio de una variable dicotómica cuyo punto de corte se ubica alrededor de 1930, distinguiéndose, de este modo, entre el modelo agroexportador y el industrial.

Desde el punto de vista teórico general, el autor sostiene que la vinculación entre la hegemonía económica y el proceso político sindical es invariable ante el modelo de acumulación y que las actividades económicas que le son centrales variarían según los distintos periodos por los que atraviase.

El problema que tenemos entre manos consiste, entonces, en analizar la relación entre tres variables dicotómicas: a X y Y, ya definidas, se agrega Z, con las categorías C, modelo agroexportador y C', modelo industrial.

En el ejemplo que usamos para ilustrar el coeficiente Φ vimos que no había relación entre el nivel educativo y el estatus socioeconómico. ¿Se deberá esta carencia de asociación a la acción de un tercer factor, por ejemplo, el estatus ya alcanzado por el padre?

Estos dos casos sirven como ejemplo de un sinnúmero de situaciones en que hay que considerar simultáneamente más de dos variables. En este capítulo veremos las complejidades analíticas que surgen al tratar estas situaciones. En las primeras secciones presentaremos el análisis de covarianzas de Lazarsfeld apégándonos a su propia exposición¹ y en los últimos apartados incluiremos algunos comentarios acerca del

¹Pensamos que la exposición de Paul Lazarsfeld es de una claridad tal que resulta pedagógicamente insuperable. Por ello hemos optado por seguir sus desarrollos en "La interpretación de las relaciones estadísticas como propiedad de investigación", en R. Boudon y P. Lazarsfeld, *Metodología de las Ciencias Sociales*, vol. II, Editorial Laia, Barcelona, 1974.

uso de esta técnica cuando se incorporan consideraciones sobre las relaciones teóricas y sus correspondientes proposiciones empíricas. En consecuencia, la lógica que ha guiado la exposición experimentará un quiebre momentáneo pero se retomará más adelante. En lo inmediato, esto quiere decir que en lo que sigue no se explicitarán las proposiciones empíricas.

3.1. El análisis de covarianzas de Lazarsfeld²

Lazarsfeld empieza su exposición mostrando las relaciones entre la edad y el tipo de audición radiofónica:

CUADRO 12

Proporciones de oyentes de radio según tipos de programas radiales, por categorías de edad

Tipos de programas	Edad	
	Jóvenes	Viejos
Religiosos	17%	26%
Tribunas políticas	34%	45%
Música clásica	30%	29%
Total de casos	1 000	1 300

Es conveniente notar que el cuadro 12 constituye una síntesis de tres cuadros, uno de los cuales es el siguiente:

CUADRO 13

Distribución de los oyentes de radio según la audición de programas religiosos, por categorías de edad

Escuchan programas religiosos	Edad		Total
	Jóvenes	Viejos	
Sí	170	338	508
No	830	962	1 792
Total	1 000	1 300	2 300

²Debido a que es interesante mostrar las complicaciones que surgen en el estudio de las relaciones cuando se incluye una tercera variable, no desviaremos nuestra atención hacia el análisis de las distribuciones marginales. Nos podemos tomar esta licencia porque se trata de un trabajo de naturaleza eminentemente pedagógica; en una investigación difícilmente podríamos justificar la omisión de las marginales.

Del mismo modo se podría generar otro para edad y tribunas políticas (Y_2) y un tercero para edad y programas de música clásica (Y_3).

Del cuadro 12 se puede derivar que: *i*) habría un cierto grado de relación entre edad y audición de programas religiosos: los viejos tienden a sintonizar este tipo de programas en mayor proporción que los jóvenes (26% y 17% respectivamente). *ii*) Parece que la edad permite explicar las preferencias por los programas políticos. Los viejos muestran una marcada inclinación por este tipo de programas radiales (45%) en comparación con los jóvenes (34%). *iii*) Por último, la audición de música clásica no varía con la edad; más bien es uniforme: la sintonizan el mismo porcentaje de jóvenes (30%) y de viejos (29%), de manera que se puede concluir que ambas variables son independientes.

Sin embargo, sabemos que en las sociedades industriales ha tenido lugar un proceso de expansión en el sistema educativo lo que debiera reflejarse en que las variables edad y nivel de instrucción estarían relacionadas:

CUADRO 14

Distribución de los oyentes de programas radiales según nivel de instrucción por categorías de edad

Nivel de instrucción	Edad		Total
	Jóvenes	Viejos	
Superior	600 (60%)	400 (30.8%)	1 000
Inferior	400	900	1 300
Total	1 000	1 300	2 300

Este cuadro muestra, como esperábamos, un cierto grado de asociación entre la edad y el nivel de instrucción, que es mayor en los jóvenes que en los viejos. Las relaciones entre la edad y el tipo de programas radiales que se escuchan están afectadas por el nivel de instrucción; por tanto para medir las asociaciones que nos interesan debemos construir cuadros que permitan controlar su efecto. En este tipo de análisis el control se lleva a cabo por medio del principio de "igual presencia" de la tercera variable. Se estudia la relación entre X y Y para un mismo valor de Z; por ejemplo, se analiza la asociación entre la edad y la audición de programas religiosos (Y_1) para todos aquellos que tienen un nivel de instrucción superior (C) y por separado para las personas que poseen un nivel de instrucción inferior (C').

CUADRO 15

Distribución de los oyentes de radio según la audición de programas religiosos, por nivel de instrucción y categoría de edad

	Nivel de instrucción superior			Nivel de instrucción inferior		
	Jóvenes	Viejos	Total	Jóvenes	Viejos	Total
Escuchan programas religiosos	55 (9.2%)	*45 (11.5%)	100	115 (28.8%)	*285 (31.7%)	400
No	545	355	900	285	615	900
Total	600	400	1 000	400	900	1 300

(a)

(b)

*La suma de los viejos con nivel de instrucción superior e inferior que escuchan programas religiosos debería coincidir con el total que se presenta en el cuadro 13 (338). La diferencia de ocho casos se encuentra en el texto de Lazarsfeld y lo hemos transcrito sin modificaciones. Esta anomalía se compensa con los ocho casos de más que aparecen en los viejos que no escuchan programas religiosos. El mismo problema aparece en el cuadro siguiente.

La relación inversa que habíamos detectado entre la edad y la sintonía de programas religiosos prácticamente desaparece al efectuar el control por medio del nivel de instrucción. La subpoblación formada por las personas con un nivel de instrucción superior escucha menos emisiones radiales religiosas que la que posee un nivel inferior. Sin embargo, en los dos casos tenemos una situación próxima a la independencia estadística, lo que origina, en parte, la débil asociación que habíamos comprobado entre X y Y.

Sabemos que hay un cierto grado de relación entre la edad y la audición de tribunas políticas. Veamos qué sucede al incorporar al análisis el nivel de instrucción.

CUADRO 16

Distribución de los oyentes de radio según la audición de programas políticos, nivel de instrucción y categoría de edad*

	Nivel de instrucción superior			Nivel de instrucción inferior		
	Jóvenes	Viejos	Total	Jóvenes	Viejos	Total
Escuchan programas políticos	240 (40%)	*220 (55%)	460	100 (25%)	*360 (40%)	460
No	360	180	540	300	540	840
Total	600	400	1 000	400	900	1 300

(a)

(b)

*Véase la nota del cuadro 15.

El cuadro 16 muestra que al controlar por educación se destaca la relación original. Si bien el porcentaje de personas que sintonizan tribu-

nas políticas es más alto en el nivel de instrucción superior que en el inferior, en los dos casos la proporción de viejos que escuchan programas políticos es superior en 15% a la de jóvenes.

Para los programas de música clásica tenemos una situación diferente.

CUADRO 17

Distribución de los oyentes de radio según la audición de programas de música clásica, por nivel de instrucción y categoría de edad

	Nivel de instrucción superior			Nivel de instrucción inferior		
	Jóvenes	Viejos	Total	Jóvenes	Viejos	Total
Escuchan programas de música clásica	192 (32%)	208 (52%)	400	112 (28%)	171 (19%)	283
No	408	192	600	288	729	1 017
Total	600	400	1 000	400	900	1 300

(a)

(b)

Del cuadro 17 derivamos que no hay relación entre la edad y la audición de programas de música clásica. Sin embargo, al controlar por la educación vemos que en el nivel de instrucción superior la proporción de viejos que sintonizan ese tipo de emisiones es bastante mayor que la de jóvenes, y que los porcentajes se invierten en el nivel de instrucción inferior. La carencia de asociación entre X y Y₃ se debe, en gran parte, a las relaciones de sentido opuesto que se anulan entre sí.

El ejemplo que hemos desarrollado muestra que la introducción de una tercera variable puede llevar a cualquiera de tres situaciones diferentes:

i) La proporción de viejos que sintonizan programas religiosos es superior a la de jóvenes, pero al controlar por nivel de instrucción la relación desaparece. Se comprueba una asociación entre X y Y₁, que desaparece al controlar por Z.

ii) Es mayor el porcentaje de viejos que de jóvenes que escuchan tribunas políticas. Al introducir la educación aumenta el nivel de asociación. Al examinar la relación entre X y Y₂ se concluye que las variables están ligadas, pero al introducir Z el ligamen se acentúa.

iii) La proporción de viejos y jóvenes que oyen música clásica es la misma. Sin embargo, cuando en el análisis se incluye la instrucción se tiene que en el nivel superior el porcentaje de viejos que sintonizan música clásica es mayor que el de jóvenes y que en el nivel inferior, a éstos corresponde proporción mayor. Teníamos independencia estadística entre X y Y₃, pero al introducir Z aparece una relación inversa en C (nivel de instrucción superior) y una asociación directa en C' (nivel de instrucción inferior).

En resumen cuando dos variables están relacionadas y se incluye una tercera, el vínculo puede desaparecer, permanecer o hacerse más fuerte. Si las variables son independientes, la inclusión de una tercera puede hacer que aparezca una relación.

El interés de estos tres ejemplos reside más en su forma que en su contenido. Las tres situaciones no son sino casos particulares de una estructura general que se estudiará en la sección siguiente.

3.2. La ecuación de covarianzas

El camino analítico usado en los ejemplos del apartado anterior (estudiar la relación entre dos variables a la luz de una tercera), es bastante frecuente en el proceso de investigación. Esta forma de análisis Lazarsfeld la sistematizó en una ecuación. La fórmula que propuso permite reconstituir la relación entre las variables X y Y cuando se considera Z como tercera variable, dicotomizada en C y C'. Esta última usualmente se denomina variable test o variable control:

$$(16) \quad (X \ Y) = (X \ Y; C) \oplus (X \ Y; C') \oplus (X \ Z) (Y \ Z)$$

Los términos de esta ecuación³ tienen los siguientes significados:

(X Y) simboliza la asociación entre las variables explicativa y explicada, es decir, la relación original. Por ejemplo, expresa la relación entre la edad (X) y la audición de programas religiosos (Y).

(X Y; C) denota la asociación entre X y Y en la subpoblación definida por C. En el ejemplo se trataría de la relación entre edad (X) y sintonía de programas religiosos (Y) en el nivel de instrucción superior (C).

(X Y; C') expresa la relación entre X y Y en la subpoblación C'. Representaría el grado de asociación entre la edad (X) y la audición de programas religiosos (Y) en el nivel de instrucción inferior (C').

(X Z) simboliza la relación entre las variables explicativa y control. Mostraría la ligazón entre edad (X) y educación (Z) y

(Y Z) denota el vínculo entre la variable test y la explicada. Sería la relación entre la sintonía de programas religiosos (Y) y la escolaridad (Z).

³ Quien se interese por la demostración de esta fórmula, puede consultar Paul Lazarsfeld, "El álgebra de los sistemas dicotómicos", en Boudon y Lazarsfeld, *op. cit.* En sentido estricto (X Y) simboliza el determinante de la tabla cruzada, pero la relación de Δ con los coeficientes de asociación y con las diferencias porcentuales permite generalizar su uso para cualquier medida de asociación que tenga una relación conocida con el determinante.

⊕ Representa una suma cualitativa. Para que la agregación de los componentes de la ecuación de covarianzas arroje exactamente el mismo resultado en ambos miembros de la igualdad, habría que incluir un juego de ponderaciones. Su exclusión no afecta los propósitos usuales del análisis.

Cuando en el análisis consideramos tres variables dicotómicas se generan cinco tablas cruzadas. Tres que dan cuenta de todas las relaciones posibles entre las variables tomadas de dos en dos (X con Y; X con Z; Y con Z) y dos que muestran el vínculo entre X y Y en las subpoblaciones C y C'.

El juego de tablas resultante es el siguiente:

TABLA 10

Distribuciones de frecuencias para estudiar la relación entre X y Y cuando se introduce la variable Z

Y \ X		(X Y)			(X Z)			Total
		B	B'	Total	Z \ X	B	B'	
A	N(A \ B)	N(A \ B')	N(A)	C	N(B \ C)	N(B' \ C)	N(C)	
A'	N(A' \ B)	N(A' \ B')	N(A')	C'	N(B \ C')	N(B' \ C')	N(C')	
Total	N(B)	N(B')	N	Total	N(B)	N(B')	N	

(a)

Y \ Z		(Y Z)			(X Y; C)			Total
		C	C'	Total	Y \ X	B	B'	
A	N(A \ C)	N(A \ C')	N(A)	A	N(A \ B \ C)	N(A \ B' \ C)	N(A \ C)	
A'	N(A' \ C)	N(A' \ C')	N(A')	A'	N(A' \ B \ C)	N(A' \ B' \ C)	N(A' \ C)	
Total	N(C)	N(C')	N	Total	N(B \ C)	N(B' \ C)	N(C)	

(b)

Y \ X		(X Y; C')			(X Y; C)			Total
		B	B'	Total	Z \ X	B	B'	
A	N(A \ B \ C')	N(A \ B' \ C')	N(A \ C')	A	N(A \ B \ C)	N(A \ B' \ C)	N(A \ C)	
A'	N(A' \ B \ C')	N(A' \ B' \ C')	N(A' \ C')	A'	N(A' \ B \ C)	N(A' \ B' \ C)	N(A' \ C)	
Total	N(B \ C')	N(B' \ C')	N(C')	Total	N(B \ C)	N(B' \ C)	N(C)	

(c)

Y \ X		(X Y; C')			(X Y; C)			Total
		B	B'	Total	Z \ X	B	B'	
A	N(A \ B \ C')	N(A \ B' \ C')	N(A \ C')	A	N(A \ B \ C)	N(A \ B' \ C)	N(A \ C)	
A'	N(A' \ B \ C')	N(A' \ B' \ C')	N(A' \ C')	A'	N(A' \ B \ C)	N(A' \ B' \ C)	N(A' \ C)	
Total	N(B \ C')	N(B' \ C')	N(C')	Total	N(B \ C)	N(B' \ C)	N(C)	

(d)

Y \ X		(X Y; C')			(X Y; C)			Total
		B	B'	Total	Z \ X	B	B'	
A	N(A \ B \ C')	N(A \ B' \ C')	N(A \ C')	A	N(A \ B \ C)	N(A \ B' \ C)	N(A \ C)	
A'	N(A' \ B \ C')	N(A' \ B' \ C')	N(A' \ C')	A'	N(A' \ B \ C)	N(A' \ B' \ C)	N(A' \ C)	
Total	N(B \ C')	N(B' \ C')	N(C')	Total	N(B \ C)	N(B' \ C)	N(C)	

(e)

La tabla 10(a) es producto del cruce entre las variables originales; 10(b) y 10(c) muestran las distribuciones bidimensionales de frecuencias provenientes del cruce de la tercera variable (Z) con las variables explicativa y explicada. A estas dos últimas distribuciones se les conoce como *marginales* porque sus frecuencias internas se encuentran en los márgenes de las tablas 10(d) y 10(e). Estas dos últimas muestran la distribución

conjunta de X y Y en las subpoblaciones C y C' y permiten calcular las denominadas *parciales o condicionales*.

Al cuantificar la fuerza de la relación en cada tabla se origina un componente de la ecuación (16) que hemos indicado en la parte superior de las tablas.

El uso de esta ecuación puede ayudar a entender algunos resultados paradójicos, como sería el de la relación entre el número de cigüeñas (X) y el de nacimientos (Y), que queda explicada de inmediato al introducir el tipo de comunidad (Z) dicotomizado en rural (C) y urbano (C'). La asociación desaparece en las subpoblaciones rural y urbana, es decir $(X Y; C) = (X Y; C') = 0$. Además, las variables explicativa y explicada tienen una relación no nula con la variable control. En efecto, se sabe que en las comunidades rurales hay más nacimientos y más cigüeñas que en las urbanas; por tanto, al existir un ligamen entre el control geográfico (Z) y los nacimientos (Y) y entre el primero y las cigüeñas (X), se concluye que las relaciones marginales no son nulas: $(X Z) \neq 0$ y $(Y Z) \neq 0$. La combinación de asociaciones parciales iguales a cero y marginales diferentes de cero permite explicar la relación original (X Y) por las marginales; por el vínculo de la variable control con X y Y.

La ecuación de Lazarsfeld también puede servir para encontrar explicaciones que enriquezcan la comprensión de algunas asociaciones. Por ejemplo, podríamos desarrollar el análisis de la relación entre la edad y la audición de programas religiosos usando como tercera variable el nivel de instrucción. Para cuantificar la ecuación (16) usaremos el determinante como medida de asociación. Para realizar el cálculo podremos recurrir a aplicar Δ sobre cuadros normalizados, o, lo que es equivalente, a utilizar la fórmula para Δ' de la ecuación (11) sobre cuadros no normalizados.

Al aplicar Δ' sobre la información del cuadro 13 tenemos:

$$\Delta'_{xy} = -0.022$$

y del cuadro 15 (a) y (b) se llega a:

$$\Delta'_{xyc} = -0.005 \quad \text{y} \quad \Delta'_{xyc'} = -0.006$$

El determinante para el cuadro 14 es:

$$\Delta'_{xz} = 0.072$$

Para calcular la relación marginal entre las variables control (Z) y explicada (Y) es necesario que antes se obtenga la distribución conjunta. A partir del cuadro 15 se deriva:

CUADRO 18

Distribución de los oyentes de radio, según la audición de programas religiosos por nivel de instrucción

	Y \ Z	Nivel de instrucción		Total
		Superior	Inferior	
Escuchan programas religiosos	Si	100	400	500
	No	900	900	1 800
	Total	1 000	1 300	2 300

De este cuadro se obtiene que:

$$\Delta'_{yz} = -0.051$$

Al usar el determinante para medir el grado de relación entre la edad, la sintonización de programas religiosos y la educación, la ecuación (16) toma la siguiente forma particular:

$$-0.022 = -0.005 \oplus -0.006 \oplus (0.072) (-0.051)$$

La asociación entre la edad y la audición de programas religiosos se explica, básicamente, por las relaciones marginales (nivel de instrucción con edad y nivel de instrucción con audición de programas religiosos), puesto que al controlar por nivel de instrucción las relaciones parciales tienden a desaparecer.

Desde el punto de vista analítico, de las formas particulares que asume (16) dos revisten especial interés porque ambas definen situaciones extremas contra las cuales conviene comparar cada caso concreto.

Si la tercera variable Z no está relacionada con X, entonces (16) asume la forma:⁴

$$(17) \quad (X Y) = (X Y; C) \oplus (X Y; C') \oplus (0) (Y Z)$$

Cuando en un análisis específico la ecuación de Lazarsfeld toma esta forma entonces la relación entre X y Y se debe a los vínculos que existen entre ambas variables en las subpoblaciones C y C'.

⁴La consideración de la variable Z en el análisis será pertinente si tiene una relación con Y, por lo que no se incluye explícitamente el caso $(Y Z) = 0$. Si la tercera variable no se vinculase con X ni con Y entonces sería superfluo considerarla. Desde el punto de vista estrictamente matemático en esta ecuación $(X Z) = (Y Z) = 0$ es equivalente de $(Y Z) = 0$.

Tenemos una situación enteramente diferente si:

$$(18) \quad (X Y) = 0 \oplus 0 \oplus (X Z) (Y Z)$$

en cuyo caso la asociación entre X y Y se origina en los vínculos de ambas variables con Z. Es decir, al efectuar el control por medio de la tercera variable se desvanece la relación original en cada una de las subpoblaciones. La relación se debe a las asociaciones marginales. Esta forma particular de la ecuación de covarianzas permite construir una explicación vía relaciones parciales para el ejemplo de los nacimientos y las cigüeñas y también para el caso de la audición de programas religiosos.

Tenemos, entonces, que la ecuación de covarianzas de Lazarsfeld puede asumir dos formas polares, una en que se hacen cero las marginales y la relación se debe a las parciales y otra en que se anulan las parciales y el vínculo entre X y Y se debe a las marginales. Es conveniente bautizar uno y otro polos para simplificar la comunicación. Cuando la ecuación asume la forma (17) se acostumbra decir que el análisis se hace por *elaboración de parciales* o condicionales y cuando asume la estructura de (18), que se hace por *elaboración de marginales*.

Conviene destacar que estas son las dos formas ideales que puede asumir la ecuación; por tanto, en situaciones prácticas rara vez aparecen nítidamente. Las buenas aproximaciones al caso puro de un tipo u otro de elaboración se presentan de manera bastante similar a la ecuación que calculamos para el ejemplo de la audición de programas religiosos. Pero en la mayoría de los casos prácticos se presenta una situación intermedia que resulta de una mezcla de elaboración por medio de parciales y marginales.

La ecuación de Lazarsfeld expresa en términos matemáticos una estructura de carácter lógico de validez general: el análisis de la relación entre dos variables a la luz de una tercera. Que la mayoría de las veces se aplique a estudios empíricos no invalida su uso estrictamente lógico en el campo de algunas teorizaciones (de las proposiciones lógicas). En palabras de Lazarsfeld:

Así pues, hemos reducido a su armazón lógico un procedimiento de investigación familiar. Consiste en analizar la relación entre dos variables mediante la introducción de una tercera variable. Hasta ahora lo hemos aplicado a ejemplos derivados de estudios empíricos, pero sería fácil mostrar que ciertas teorías tienen una estructura lógica idéntica (...) Durkheim sugiere que el suicidio es menos frecuente entre los católicos que entre los protestantes porque la comunidad católica está más cohesionada.⁵

⁵Paul Lazarsfeld, "La interpretación...", *op. cit.*, p. 31.

Además sostiene que la validez de su ecuación rebasa los límites impuestos por el número de variables y por el carácter dicotómico de cada una de ellas:

Para simplificar hemos restringido el análisis al caso de tres variables. La introducción de variables suplementarias no entraña ninguna modificación en la lógica del razonamiento, como se puede ver fácilmente. Por otra parte, nosotros tenemos en cuenta las variables dicotómicas. La introducción de variables que pudieran tomar un número de valores superior a dos complicaría el argumento sin cambiar la forma del mismo.⁶

Para completar su modelo, agrega una serie de consideraciones respecto a la ubicación temporal de las variables que, combinadas con las formas que puede asumir la ecuación, completan una tipología contra la cual se contrasta cada caso concreto que se analiza.

Sostiene que en cualquier investigación hay que intentar determinar la secuencia temporal de las variables. En ocasiones la identificación de la precedencia cronológica es directa o evidente, como en la relación entre la duración del matrimonio (X) y la armonía conyugal (Y), pero en otras la identificación del orden temporal podría requerir un breve análisis:

Se ha observado una relación inversa entre la importancia del ingreso y la frecuencia de pertenencia a organizaciones. Lo más verosímil es suponer una prioridad al estatus económico.⁷

En otras ocasiones no es posible determinar la secuencia temporal de las variables, en cuyo caso la validez de la tipología que presentaremos se reduciría a las dos situaciones polares que derivan de la ecuación. La postura de Lazarsfeld respecto a la importancia del tiempo es inequívoca:

De hecho siempre es teóricamente posible determinar el orden cronológico de las variables; el progreso de la investigación empírica consiste precisamente, entre otras cosas, en determinar la anterioridad entre variables. Evidentemente sería interesante discutir los procedimientos de análisis que pueden utilizarse en el caso en que el orden temporal es dudoso, pero ello nos alejaría de nuestro tema. De ahora en adelante, admitiremos que el orden temporal es conocido.⁸

⁶*Ibidem*, p. 32.

⁷*Ibidem*, p. 32.

⁸*Ibidem*, p. 33.

3.3. La tipología de Lazarsfeld

Al presentar su tipología Lazarsfeld plantea:

Desde este momento estamos en condiciones de presentar el punto esencial de nuestra argumentación: a partir de dos variables primitivas y de una variable test se pueden efectuar cuatro operaciones fundamentales. El hecho de que estas operaciones se realicen sobre datos empíricos o que correspondan a análisis teóricos no tiene ninguna importancia. Si se analiza la relación entre dos variables a la luz de una tercera, el análisis corresponderá siempre a una de estas cuatro operaciones o a una combinación de las mismas, y

Las cuatro operaciones fundamentales que constituyen la tipología surgen de la combinación de las dos formas polares que puede asumir la ecuación de covarianzas y de la ubicación temporal de la tercera variable (Z) con relación a X y Y. Si se supone que X es cronológicamente anterior a Y y se considera que Z sólo puede tomar dos posiciones, (posterior a X y anterior a Y —intermedia—, o anterior a X), entonces tenemos que las dos posiciones de la tercera variable (intermedia o anterior a X), cruzadas con las dos formas límite que puede tomar la ecuación, originan:

TABLA II

Los tipos puros de elaboración

Posición de la variable control	Forma de la ecuación	
		$(X Z) = 0$
Anterior	$(X Y; Z) \neq 0$	$(X Y; Z) = 0$
Intermedia	P.A.	M.A.
	P.I.	M.I.

La casilla identificada con P.A. (parcial anterior) corresponde al caso en que la tercera variable (Z) es anterior a X y en que la ecuación asume la forma representada por la igualdad (17). Como el producto de las marginales es igual a cero, la asociación entre X y Y se debe a la relación (X Y) en las subpoblaciones C y C'. Con una gráfica se puede llegar a una expresión sintética de las condiciones que definen P.A. (parcial anterior).

⁹Ibidem, p. 34.

GRÁFICA 1

Relaciones entre las variables en el tipo P.A.



En la gráfica 1 se supone que el tiempo es creciente en la medida en que nos movemos desde la parte superior a la inferior de la página.

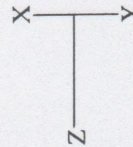
Como ejemplo del tipo P.A., tendríamos el efecto de la propaganda de una película según el grado de instrucción de los espectadores. Podríamos esperar que en la medida en que el nivel de instrucción (Z) sea menor, el efecto (Y) de la propaganda (X), sería mayor: en esta ilustración tenemos que la marginal (X Z) sería igual a cero (carencia de asociación entre instrucción y propaganda) y que la relación parcial en la subpoblación de mayor educación (X Y; C) sería menor que la observada en la subpoblación de menor educación (X Y; C'); por tanto el vínculo entre X y Y se debe a las asociaciones condicionales. En cuanto al orden temporal, tendríamos que la instrucción (Z) es anterior a la propaganda (X). Esta estructura de relación entre las variables es bastante similar a las situaciones en que tenemos una disposición (nivel de instrucción), un estímulo (propaganda) y una respuesta (efecto de la propaganda).

Es usual que una de las asociaciones parciales que dan cuenta del ligamen entre X y Y sea mayor que la otra y que a partir de este hecho surjan preguntas acerca de los factores que conducen a dicha diferencia y por esta vía se abran nuevas puertas para continuar la investigación.

El tipo P.I. (parcial intermedia) comparte con el P.A., la misma estructura de la ecuación; la diferencia se encuentra en que la tercera variable (Z) se ubica entre X y Y, como se puede apreciar en la gráfica 2.

GRÁFICA 2

Relaciones entre las variables en el tipo P.I.



Para este caso tendríamos como ejemplo:

Que al estudiar la relación entre el éxito profesional y el ambiente, liberal o autoritario, de la educación se ha podido observar que los jóvenes educados en una atmósfera liberal tienen menos éxito en una situación profesional de estructura autoritaria.¹⁰

Tenemos que el tipo de educación (X, dicotomizada en liberal y autoritaria) y el éxito profesional (Y) se relacionan en virtud de que los jóvenes educados en una atmósfera liberal tienen menos éxito si el lugar de trabajo es autoritario (X Y; Z) que si es liberal (X Y; Z'). Es decir, ambas asociaciones condicionales son no nulas y diferentes. Se supone que el tipo de educación recibida no se relaciona con la atmósfera de trabajo (X Z) = 0, o, dicho de manera diferente, que los mecanismos de selección de personal no discriminan en función del tipo de ambiente educativo en que se formaron los candidatos a los diferentes puestos. Además, tenemos que la educación (X) es anterior a la variable lugar de trabajo (Z). De esta manera se concluye que el ejemplo es del tipo P.I.

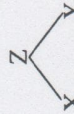
Los tipos de elaboración P.A. y P.I., tienen una relación de equivalencia. Para pasar de uno a otro basta con intercambiar Z con X. En función de esto Lazarsfeld plantea que:

Tenemos pues de hecho un tipo único de elaboración para el cual el vocablo *especificación* es sin duda el más apropiado. La consideración de la anterioridad temporal entre X y Z introduce dos subtipos de especificación.¹¹

El tipo M.A. (marginal anterior) se genera cuando Z es anterior a X y la ecuación de covarianzas asume la forma expresada por (18). El vínculo entre X y Y se origina en las relaciones marginales, es decir, en el producto de (X Z) y (Y Z).

GRÁFICA 3

Relaciones entre las variables en el tipo M.A.



Esta estructura de relación entre las variables permite entender las relaciones espurias, ilusorias o equívocas. La cantidad de cigüeñas (X) se relaciona con el número de nacimientos (Y). Pero al introducir el tipo de localidad (Z) dicotomizado en rural (C) y urbano (C') se tiene

¹⁰Ibidem, p. 35.

¹¹Ibidem, p. 35.

que $(X Y; C) = (X Y; C') = 0$; en consecuencia, la relación se debe a los vínculos que el grado de urbanización sostiene con el número de cigüeñas (X Z) y con el número de nacimientos (Y Z): en el campo hay más cigüeñas que en las ciudades y también más nacimientos. En cuanto a la ubicación temporal de las variables es claro que la urbanización es anterior a X y Y.

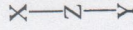
Otro ejemplo muy usado en la literatura es el de la asociación entre la importancia de los daños (Y) provocados por un incendio y el número de extintores (X). Al incluir la variable control de la gravedad del incendio (Z), desaparecen las parciales (X Y; C) y (X Y; C'). La relación se debe a las marginales (X Z) y (Y Z). Ahora bien, la gravedad del incendio es una variable anterior a los daños que provoca y a la cantidad de extintores utilizados.

Los dos casos anteriores son claros ejemplos de relaciones ilusorias, espurias o equívocas, que se pueden entender a través de la elaboración M.A. (marginal anterior).

El último tipo, el M.I. (marginal intermedia) resulta de la misma forma que la ecuación que origina M.A., pero la tercera variable se ubica temporalmente entre X y Y.

GRÁFICA 4

Relaciones entre las variables en el tipo M.I.



Considérese como ejemplo la relación entre el estado civil de las mujeres (X) y su tasa de ausentismo (Y). Si se introduce como variable control la importancia de las responsabilidades familiares (Z), entonces se tiene que:

$$(X Y; Z) = (X Y; Z') = 0$$

y la relación se debe a los vínculos que tiene la responsabilidad familiar con el ausentismo (Y Z) y con el estado civil (X Z). Además, es claro que en la secuencia temporal la variable Z es intermedia, es posterior al estado civil y anterior al ausentismo.

El lector podría analizar qué ocurre si se introduce como tercera variable la menor resistencia física, medida por la edad. ¿A qué caso corresponde esta situación? ¿Será del tipo M.I. o M.A.?

3.4. Breves comentarios sobre la aplicación de la tipología

En la investigación es bastante normal el empleo de una tercera variable para analizar la relación entre X y Y. También es frecuente que en los desarrollos sólo se tomen en cuenta la forma de la ecuación, sin considerar el tiempo. Es decir, sólo se considera, por así decirlo, los marginales por columna de la tabla II en lugar de juzgar el grado de aproximación de la situación examinada con respecto a los casos puros que presenta la tipología.

Pareciera que la incorporación del tiempo complica bastante las posibilidades analíticas. De hecho el planteamiento de Lazarsfeld incorpora el tiempo en su versión casi física. Decimos casi porque de alguna manera, aunque no explícita, reconoce complejidades que derivan de esta variable:

Ciertas variables pueden ser utilizadas de varias maneras y, en función del problema examinado, tener posiciones diferentes en la secuencia temporal. Un ejemplo típico es la edad. Las personas de 60 años se distinguen de los jóvenes de 20 no solamente por las posibilidades físicas, sino también porque nacieron y fueron educados en otra época. En la proposición *las personas de edad tienen un nivel de instrucción inferior*, lo importante es evidentemente la época en que estas personas han alcanzado la edad de la escolarización: en este caso la edad es anterior a la educación. Si por el contrario se relaciona la edad de fallecimiento y el clima, la edad evidentemente debe considerarse como posterior.¹²

Se puede usar una misma variable como indicador de conceptos que se refieren a procesos cuyas temporalidades son diferentes. En un caso la edad se utilizaría como indicador del tipo de instrucción y socialización que una persona recibió en la escuela; en el otro sería un índice del proceso de envejecimiento. Por otra parte, se pone como referencia a un proceso social que se encuentra vinculado con otros, por ejemplo la evolución de la distribución del poder entre las fracciones hegemónicas de clase, el estilo de desarrollo socioeconómico que impulsan, las normas y valores dominantes en cada período histórico, etc. La edad se ubica en relación con el desarrollo físico de los individuos: el efecto del paso de los años, que también puede incorporar (según sea el enfoque) una serie de procesos sociales, por ejemplo los relacionados con la esperanza de vida al nacer, los que determinan las condiciones de salud pública, etcétera.

Lo que importa destacar es que el tiempo es una variable bastante compleja y que cuando se trata de dar cuenta del cambio social no se

reduce a su concepción física, sino que aparece de múltiples formas en virtud del movimiento de los diferentes procesos sociales. Éste debe ser, tal vez, uno de los problemas más importantes y difíciles de resolver en las ciencias sociales.¹³ Es obvio que su tratamiento sistemático está fuera de nuestro alcance; sólo lo incorporamos con el propósito de relativizar las bondades del modelo que presentamos.

Sabemos que Lazarsfeld sostiene que el progreso de la investigación empírica consiste, entre otras cosas, en determinar la secuencia temporal de las variables. Pero, ¿qué pasa si X y Z se ubican en un mismo punto del tiempo? ¿Qué ocurre si no se puede saber cuál fue antes? ¿Es necesaria tal determinación para el avance del conocimiento? ¿Es posible producir conocimiento válido y de interés científico y político sin considerar el tiempo en la forma en que lo hace Lazarsfeld? Con el propósito de esbozar un inicio de respuesta a este tipo de preguntas retomemos el ejemplo de la relación entre hegemonía económica y carácter de la lucha sindical en Argentina. Hemos concluido que en los sectores económicos centrales para el modelo de acumulación los conflictos obedecían básicamente a razones de carácter organizativo, y en los no centrales a consideraciones reivindicativas o salariales. Su pongamos que se decide incorporar como variable control el nivel de concentración económica sectorial (Z). Se argumenta que la relación entre la hegemonía económica (X) y el carácter de la lucha (Y) no es genuina sino que se origina en los vínculos de la concentración económica sectorial con la hegemonía económica (X Z) y con el tipo de lucha (Y Z). Este argumento sostiene que si se introduce en el análisis un tercer factor, las relaciones parciales se harán nulas y que la asociación (X Y) se debe a las marginales. Es decir, a partir del esquema lógico que ha codificado Lazarsfeld se levanta una explicación alternativa que se podría contrastar.

Como la explicación alternativa a (X Y) envía a una elaboración por marginales, sería interesante saber si se trata de un caso M.A. o M.I., puesto que si se puede establecer que la concentración económica sectorial es anterior a la hegemonía, la relación sería ilusoria. Sin embargo parece bastante factible sostener que en el caso argentino sería imposible saber si Z se ubica antes de X o si cronológicamente se encuentra después de X. Es probable que sean simultáneas, en la medida en que el ferrocarril inglés nace económicamente en pocas manos y que la concentración de vestidos surge económicamente desconcentrada. Entonces, según el análisis de covarianzas no podríamos distinguir entre los tipos M.I. y M.A. ¿Qué quiere decir esto? Simplemente que no estamos en condiciones de saber si es o no ilusoria la relación entre la hegemonía

¹³Un buen tratamiento de este problema se puede ver en Fernand Braudel, *La historia y las ciencias sociales*, Alianza Editorial, Madrid, 1968, cap. 3.

¹²*Ibidem*, p. 33

económica y el tipo de conflicto, cuestión que se dirimiría ubicando Z antes o después de X.

De alguna manera, aunque no expuesto con mucha claridad en el trabajo de Lazarsfeld, el tiempo se usa para determinar la estructura causal entre las variables. Lo que sí es claro es que sus planteamientos sientan las bases para que posteriormente se exploren una serie de caminos analíticos que se englobaron con el título de análisis de causalidad.¹⁴

De este modo llegamos, a nuestro juicio, al corazón de las dificultades que presenta la incorporación del tiempo en el análisis de covarianzas. De manera implícita o explícita se trata de determinar la causalidad a partir del análisis de los datos; de este modo se llega a uno de los problemas filosóficos más candentes: ¿se puede encontrar el sentido causal a través del análisis de datos o es imprescindible hacer consideraciones de índole teórica? En otros términos, ¿sólo se puede determinar la causalidad a partir de la teoría?

No es posible dar una salida empírica al ejemplo que hemos manejado en esta sección ya que no podemos ubicar Z antes o después de X. Sin embargo, es posible esgrimir una serie de argumentos teóricos y de antecedentes históricos que ayuden a resolver el impasse entre las dos maneras de explicar la relación entre hegemonía económica y el carácter de la lucha sindical.

La dificultad para incorporar el tiempo en el análisis nos orilló a realizar una serie de consideraciones que surgen tanto de la práctica de la investigación como de la interpretación de los desarrollos posteriores a que dio lugar el planteo de Lazarsfeld.¹⁵ Encontramos que los problemas derivados del tiempo pueden provenir, en última instancia, de un intento por encontrar, a partir de los datos, el sentido causal con el que se relacionan las variables y que por tanto surge de una manera velada un serio problema filosófico: ¿cómo determinar la causalidad?

En esta sección también vimos cómo las diferentes formas que puede asumir la ecuación de covarianzas da pie a interpretaciones teóricas diferentes para explicar la ligazón entre los procesos sociales. Tal fue el caso de la inclusión del grado de concentración económica en el análisis de la relación entre hegemonía económica y lucha sindical. También sostuvimos que la simultaneidad entre X y Z, impedía establecer empíricamente si la relación es ilusoria y planteamos que un camino para

¹⁴Para examinar la relación entre el análisis de Lazarsfeld y el de causalidad, se puede ver Raymond Boudon, *L'Analyse Mathématique de la causalité*, Plon, París, 1968.

¹⁵Dentro del análisis causal se sostiene que para inferir causalidad entre dos variables es necesario que: i) presenten variación concomitante, ii) haya asimetría temporal y iii) se eliminen otros factores causales. Un buen tratamiento del tema desde el punto de vista técnico se encuentra en Herbert Asher, *Causal Modeling*, Sage Publications, California, 1976.

buscar una solución a este problema se encuentra en la reflexión teórica y en el análisis histórico.

3.5. El concepto de asociación y el análisis de Lazarsfeld

En el momento en que decidimos exponer la tipología de Lazarsfeld siguiendo lo más fielmente posible sus propios desarrollos se produjo un paréntesis en nuestra exposición. Ahora retomaremos nuestro argumento central. La ecuación de covarianzas surge del campo de la estadística y por ello tiene incorporada la definición de asociación por oposición a la independencia estadística. Sabemos que ésta es una de las limitaciones del análisis de asociación que se puede superar a través de la incorporación de proposiciones empíricas.

Hemos visto en las secciones anteriores que los coeficientes Φ y Q se pueden interpretar como una medida del grado de correspondencia entre las proposiciones empíricas particulares y la distribución de las observaciones.

Si incluimos estas consideraciones en la ecuación (16) tendríamos que en el caso de que las proposiciones empíricas fuesen del tipo:

$$P_1: \begin{array}{c} B \longrightarrow A \\ B' \longrightarrow A' \end{array} \quad \text{o} \quad \begin{array}{c} B \longrightarrow A' \\ B' \longrightarrow A \end{array}$$

se debería usar la siguiente expresión de la ecuación de covarianzas:

$$(19) \quad \Phi_{xy} = \Phi_{xyc} \oplus \Phi_{xy'c} \oplus \Phi_{xz} \oplus \Phi_{yz}$$

Si por el contrario las proposiciones empíricas fuesen del tipo:

$$P_2: \begin{array}{c} B \longrightarrow A \\ B' \longrightarrow (A \text{ o } A') \end{array} \quad \text{o} \quad \begin{array}{c} B \longrightarrow (A \text{ o } A') \\ B' \longrightarrow A' \end{array}$$

o bien

$$P'_2: \begin{array}{c} B \longrightarrow A' \\ B' \longrightarrow (A \text{ o } A') \end{array} \quad \text{o} \quad \begin{array}{c} B \longrightarrow (A \text{ o } A') \\ B' \longrightarrow A \end{array}$$

el análisis se debería basar en la siguiente forma de la ecuación:

$$(20) \quad Q_{xy} = Q_{xyc} \oplus Q_{xy'c} \oplus Q_{xz} \oplus Q_{yz}$$

Si en el ejemplo con que se pretende explicar la variación del estatus en función de la movilidad educativa, comparando a hijos con padres, incluimos como variable control el estatus socioeconómico del padre y decidimos realizar el análisis a través de coeficientes de asociación, entonces deberíamos usar la ecuación (19), pero si se tratara de la hegemonía económica y los conflictos laborales considerando el papel de la concentración económica sectorial entonces habría que usar la ecuación (20).

En la próxima sección ilustraremos el uso de estas ecuaciones en una investigación concreta.

3.6. Reforma agraria y migraciones (el caso de Chile)

En un trabajo de investigación realizado en Chile¹⁶ se estudió la migración campo-ciudad en su lugar de origen. Por medio de encuestas realizadas en una comuna se detectó la migración potencial¹⁷ en una zona geográfica específica, de manera que a cada campesino se le adjudicó el carácter de migrante (M) o de no migrante (M').

El enfoque teórico que se usó para identificar los determinantes de la migración sostiene que es necesario considerar las características estructurales de la producción campesina. Este planteamiento se erige como alternativa a la explicación de la teoría de la modernización que ponía el acento en los factores que influyen en el nivel de modernización del campesino: los campesinos modernos tienden más a migrar que los no modernos.

Una vez que se identifican tres momentos en que se producen cambios en la organización agraria¹⁸ según los diferentes periodos de dominación política, se individualizan sus correspondientes tipos de explotación agraria:

Los cambios operados en la estructura de dominación tienen una repercusión inmediata y fundamental en la organización de la estructura agraria chilena; comienzan con el inicio de un proceso profundo de movilización campesina y otro proceso importante de reforma agraria durante el gobierno de la Democracia Cristiana. Estos procesos se agudizan y se radicalizan a partir del triunfo electoral de la Unidad Popular en el año 1970. Lo interesante de estos cambios políticos para nuestro objeto de estudio es que cada uno de ellos ha dejado cristalizados diferentes tipos de organización productiva en la estructura agraria. Así tenemos como forma predominante de ex-

¹⁶Omar Argüello, "Estructura agraria, participación y migraciones internas", en *Migración y desarrollo*, 3. CLACSO, Buenos Aires, 1974.

¹⁷*Ibidem*, p. 45.

¹⁸*Ibidem*, p. 6.

plotación agrícola hasta el año 1964 los "fundos" que es el tipo de empresa privada con grandes extensiones que contrata fuerza de trabajo permanente y temporal, y que en el caso de los permanentes, paga buena parte del salario en especie, fundamentalmente con la concesión de una pequeña extensión de tierra para el usufructo del trabajador. En el periodo 1964-1970 aparece una nueva forma de organización productiva, que toma el nombre de "asentamiento" y que es un tipo de empresa cooperativa con características de autogestión, la que contrata alguna fuerza de trabajo con trabajadores temporales, pagando sus salarios en dinero efectivo. A partir del periodo que comienza en 1970 aparece otra nueva forma de organización productiva que toma el nombre de "Centro de Reforma Agraria" (CERA) que intenta ser un tipo de empresa en transición a formas socialistas de producción, con apropiación social de parte de sus excedentes.¹⁹

Una vez que el autor ha establecido la heterogeneidad de la estructura agraria chilena (en la época del estudio) continúa analizando una serie de aspectos relativos a participación, relaciones sociales de producción, etc. Para los propósitos de esta ilustración queremos destacar que:

Llevando en cuenta argumentaciones teórico-metodológicas pueden explicarse adecuadamente las variaciones encontradas en la descripción que acabamos de hacer del fenómeno migratorio, variaciones que se relacionan sistemáticamente con el tipo de organización social de la producción que ha adoptado la empresa agrícola en la cual se inserta el trabajador agrario. Se desprende de lo anterior que aquellas empresas agrícolas organizadas a partir de cambios en la estructura de dominación, según la concepción ideológica de los nuevos grupos dominantes y que se traduce en nuevas relaciones de producción y de nuevas formas de participación en otras relaciones sociales, son las empresas agrícolas que encierran dentro de sí a menores proporciones de trabajadores con proyectos de migración.²⁰

Se plantea, entonces, una relación entre el tipo de organización social de la producción que ha adoptado la empresa agrícola y las migraciones. Se nos dice, específicamente, que son las empresas del área reformada (AR) las que tienden a retener la fuerza de trabajo campesina.

Sabemos que en el momento en que se lleva a cabo el estudio coexisten tres formas productivas; las explotaciones tradicionales denominadas fundos y las que surgieron de reformas agrarias, los asentamientos y los Ceras. Al unir estos dos últimos tipos de empresas en la categoría empresas del área reformada concluimos que el factor "organización social de la producción" se objetiva a través de la variable "tipo de empresa agraria" (X), dicotomizada en fundos (empresas tradicionales) y explotaciones reformadas (AR). Tenemos:

¹⁹*Ibidem*, p. 7.

²⁰*Ibidem*, p. 54.

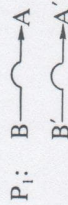
Tipo de empresa: $X \begin{cases} B: \text{Fundos} \\ B': \text{Asentamientos y Ceras (AR)} \end{cases}$

No nos detendremos en el proceso de objetivación del concepto "migración potencial",²¹ sólo señalaremos que se presenta dicotomizado en migrantes (B) y no migrantes (B')

Migración: $Y \begin{cases} A: \text{migrantes} \\ A': \text{no migrantes} \end{cases}$

A continuación debemos determinar la estructura lógica de la proposición teórica. Nos dice que en las empresas tipo AR tenderá a desaparecer la migración, pero ¿qué ocurre en las explotaciones tradicionales? Aun cuando el autor no lo sostiene explícitamente parece que se plantea que la estructura fundo promueve la salida de población. Esta afirmación se puede derivar de las fuertes corrientes migratorias de origen campesino que han afectado a la sociedad chilena.

La relación entre la organización social de la producción y la migración da origen a la siguiente proposición empírica:



El modo de organizar la producción de las explotaciones tradicionales o las relaciones sociales que las caracterizan, se muestran incapaces para retener a su fuerza de trabajo y a su población. Por el contrario, las empresas del área reformada presentan una serie de características que conducen a mantener la población agraria en su lugar de origen.

El grado de concordancia entre la proposición empírica y la distribución de los datos se mide a través de Φ aplicado a:

CUADRO 19

Distribución de los trabajadores agrarios según proyecto migratorio, por tipo de empresa agrícola

	Tipo de empresa agrícola		
	Fundos	AR	Total
Migrantes	34	8	42
No migrantes	154	378	532
Total	188	386	574

$\Phi_{xy} = 0.29$

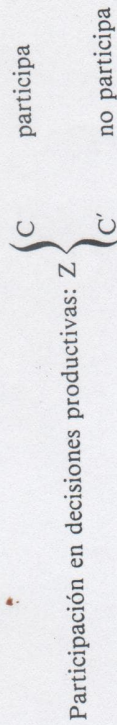
²¹El lector interesado puede consultar: *Ibidem*, sección IV: "Reforma agraria, relaciones sociales y migraciones Internas".

Si bien hay una relación entre el tipo de empresa y la migración potencial, el grado de adecuación de los datos (a juzgar por el valor de Φ) a los planteamientos desplegados es débil.

Una vez establecida la relación se continúa el argumento:

Podría sostenerse que las argumentaciones teóricas y la información empírica presentadas para la explicación sociológica del fenómeno migratorio son válidas al nivel del contexto productivo con sus diversas formas de organizarse socialmente, pero que esas afirmaciones no han sido probadas al nivel individual. En otras palabras, cabe preguntarse si no estaremos en presencia de una falacia ecológica o de nivel equivocado, que consistiría en atribuir a un nivel de análisis explicaciones que son válidas a un nivel diferente. En este caso significaría preguntarse si dentro de los contextos productivos de mayor participación, y que presentan las más bajas proporciones de trabajadores con proyectos de migración, son también los individuos con mayor participación los que menos piensan en migrar, o si ocurre el caso inverso.²²

Se trata ahora de analizar el vínculo entre la organización social de la producción y las migraciones a la luz de la participación de los campesinos. Este último concepto se objetiva a través de la participación en las decisiones productivas (Z) que se dicotomiza en participa y no participa:



La incorporación de la tercera variable tiene que ver con la construcción de una proposición teórica alternativa: la migración obedecería a factores de carácter individual y no estructural. En caso de que la explicación a través de la participación se muestre más adecuada, entonces para disminuir la migración campo-ciudad no sería necesario realizar reformas agrarias profundas que alteran las relaciones sociales de producción; bastaría con modificar los niveles de participación del campesinado.

Examinemos lo que acontece al incorporar la participación en las decisiones productivas:

²²*Ibidem*, p. 54.

CUADRO 20

Distribución de los trabajadores agrarios según proyecto migratorio, por participación en las decisiones productivas y tipo de empresa agrícola

Proyecto migratorio	Participan		No participan		Total
	Fondos	AR	Fondos	AR	
Migrantes	0	5	34	3	37
No migrantes	2	319	152	59	211
Total	2	324	186	62	248

(a)

(b)	

Antes de proceder al análisis hay que aclarar que en las conclusiones que se obtengan se deberá tomar en cuenta que en el cuadro 20 (a) hay una categoría casi inexistente, pues sólo tenemos dos personas que participan en las decisiones productivas en los fondos. Esta limitación en las observaciones, que prácticamente hace constante la variable tipo de empresa en la subpoblación que muestra participación en las decisiones productivas, cuestiona la significación estadística de los hallazgos. Esto no pasó inadvertido al autor, quien en el párrafo siguiente al último que transcribimos nos dice:

La pregunta metodológica es correcta, sólo que en el caso de nuestro estudio la asociación entre el tipo de organización productiva y la participación de los trabajadores es tan fuerte que difícilmente no sólo que se encuentre esa falacia de nivel equivocado sino que, incluso, difícilmente se la someta a prueba. De hecho ocurre que en la categoría de participación se concentran casi todos los casos de trabajadores del sector reformado con ausencia casi total de trabajadores de fondos, mientras en las categorías de no participación ocurre la situación inversa.²³

Hecha la advertencia realicemos un somero análisis de los cuadros condicionales 20 (a) y (b). En la subpoblación conformada por los que no participan en las decisiones productivas se mantiene el mismo tipo de relación que en la tabla original, aunque un poco más debilitado: el coeficiente Φ alcanza el valor 0.16 ($\Phi_{\text{yvc}} = 0.16$). Para el cuadro 20(a), tenemos que $\Phi_{\text{yvc}} = -0.01$, es decir, habría independencia estadística entre el tipo de empresa y la migración potencial; sin embargo, este resultado no es confiable por el escaso número de observaciones en los fondos en que sí hay participación en las decisiones productivas.

En el cuadro 21 mostramos la información que permite calcular las asociaciones marginales:

²³Ibidem, p. 54.

CUADRO 21

Distribución de los trabajadores agrícolas

Z \ X	Fondos		Total
	Participan	AR	
a) Según participación en las decisiones productivas, por tipo de empresa agrícola	2	324	326
No participan	186	62	248
Total	188	386	574

Y \ Z	Participación en las decisiones productivas		Total
	Participan	No participan	
b) Según proyecto migratorio, por participación en las decisiones productivas	5	37	42
Migrantes	321	211	532
No migrantes	326	248	574

Los coeficientes de asociación son: $\Phi_{\text{yz}} = -0.79$ y $\Phi_{\text{yz}} = -0.26$. En consecuencia, la ecuación de covarianzas de Lazarsfeld asume la siguiente forma particular:

$$0.29 = -0.01 \oplus 0.16 \oplus (-0.79) \oplus (-0.26)$$

La estructura de esta ecuación no corresponde a ninguno de los casos puros de la tipología de Lazarsfeld. La asociación original se conforma tanto por parciales como por marginales. A pesar de los problemas de la muestra (carencia de significación estadística de Φ_{yvc}) podemos afirmar que al introducir la participación en las decisiones productivas las relaciones parciales son más débiles que la original. Esto llevaría a privar el análisis de las marginales, donde encontramos una fuerte relación entre las variables explicativa y de control.

Sabemos que si vamos a hacer una elaboración por marginales la ubicación temporal de la variable Z es de fundamental importancia para estar en condiciones de detectar si la relación es ilusoria. La participación en las decisiones productivas es posterior a X, porque una de las consecuencias de la reforma agraria consistió en democratizar el manejo de las unidades productivas agrarias. Esta conclusión concuerda con la del autor:

Los datos muestran lo que habíamos sostenido anteriormente. Por un lado, dentro de los fondos encontramos sólo dos trabajadores que aparecen con participación en las decisiones productivas de la empresa. Por otro lado, dentro de los Asentamientos y de los Cera aquellos trabajadores que participan en esas decisiones productivas, la abrumadora mayoría, son los que muestran una proporción menor de quienes tienen proyecto de migrar. De esta manera probamos que en nuestro estudio no existe una falacia de nivel equivocado cuando afirmamos que las nuevas relaciones sociales constitutivas

de las nuevas formas de organizar la producción llevan a una menor presencia de proyectos de migración.²⁴

La relación entre el tipo de organización social de la producción y las migraciones está mediada por la participación de los trabajadores agrícolas. Esto quiere decir que si se llevan a cabo modificaciones en las unidades productivas, su efecto no se hará sentir necesariamente en las migraciones. Para retener a la población rural se deben propiciar mayores niveles de participación.

3.7. Algunas limitaciones del método para estudiar la relación entre más de dos variables dicotómicas

Una de las restricciones que se deja sentir con fuerza cuando se intenta trabajar con varias variables simultáneamente es el número de casos de que se dispone. Las casillas crecen geométricamente con la introducción conjunta de variables adicionales, tal como se ve en la siguiente tabla:

TABLA 12

Relación entre el número de variables dicotómicas cruzadas simultáneamente y el número de casillas

Variables	Casillas
2	4
3	8
4	16
5	32
6	64
.	.
.	.

Como el total de casillas crece geométricamente con razón dos, el cruce simultáneo de variables demanda una cantidad creciente de información. Si combinamos este hecho con las limitaciones presupuestarias para realizar la investigación social, entonces resulta evidente que el volumen de información constituye una severa restricción a los análisis que pretenden considerar distribuciones conjuntas definidas por varias variables.

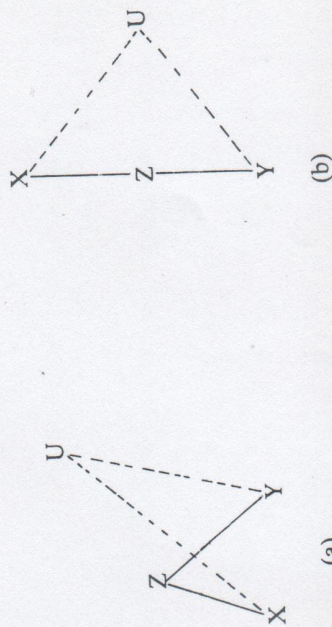
Pero, ¿por qué considerar que el número de casos es una limitación seria? ¿No bastaría con examinar la relación entre X y Y a la luz de

²⁴*Ibidem*, p. 55.

una variable control cada vez? ¿Para qué cruzar simultáneamente varias variables? Si para evitar la restricción que nos impone el número de casos sólo consideramos una variable control cada vez, puede ocurrir que las relaciones parciales no se anulen debido a la acción de una cuarta variable no incluida en el sistema de relaciones:

GRÁFICA 5

Relaciones entre cuatro variables



La gráfica 5(a) nos muestra que si la relación entre X y Y es del tipo M.A., es decir, corresponde a una asociación ilusoria y si controlamos por Z, entonces las parciales no se anularán debido a que hay una cuarta variable (U) no controlada que mantiene el vínculo entre X y Y. En la gráfica 5(b) se muestra una situación equivalente pero esta vez referida al tipo marginal intermedia. En los casos representados las parciales se anularían sólo al controlar simultáneamente por ambas variables.

El método que consiste en analizar la relación entre dos variables incluyendo una tercera cada vez, no permitiría detectar asociaciones ilusorias que se explican por dos o más variables test, ni dar cuenta de vínculos genuinos en que participan dos o más variables control. De aquí deriva la importancia de usar simultáneamente varias variables control.

Por último, quisiéramos recordar que hemos señalado que una de las limitaciones más serias para aplicar la tipología de Lazarsfeld radica en las complejidades que derivan de la ubicación de la secuencia temporal de las variables. Pero aun haciendo caso omiso de esta restricción, la utilidad de la tipología se ve menoscabada en tanto las situaciones examinadas escapan a los casos puros que la constituyen.

Ejercicios ²⁵

1. A continuación se presentan dos tablas de asociación, provenientes de la investigación de Lipset ya aludida en el ejercicio 2 del capítulo 2.

	Estudiantes que empiezan estudios		Estudiantes que terminan estudios	
	Liberales	Conservadores	Liberales	Conservadores
Los profesores comunistas deben enseñar				
Sí	74	25	118	16
No	74	38	72	32
Total	148	63	190	48

Entre las variables control que introduce hemos seleccionado el tiempo de permanencia del estudiante en la universidad. Se le pide: i) que establezca algunas proposiciones teóricas acerca del porqué se introduce esta variable test, y ii) que con la información entregada lleve a cabo un análisis de variables múltiples.

2. Con el propósito de indagar la contribución que haría la diferencia sexual a la formación de la población económicamente activa, hemos construido, a partir del censo de 1970, la siguiente tabla, referida al área metropolitana de la ciudad de México:

Población	Sexo	
	Hombres	Mujeres
Económicamente activa	1 548 575	1 456 238
No económicamente activa	2 092 079	2 320 687
Total	3 640 654	3 776 925

Un breve análisis nos conduce a comprobar la existencia de un cierto grado de relación entre las variables. Con el propósito de investigar este extraño resultado, hemos decidido incluir como variable control la procedencia de la población que lo ha originado:

Población	Migrantes		No migrantes	
	Hombres	Mujeres	Hombres	Mujeres
Económicamente activa	831 755	847 579	716 820	608 659
No económicamente activa	370 773	555 175	1 721 306	1 765 512
Total	1 202 528	1 402 754	2 438 126	2 374 171

²⁵Véase la bibliografía al final de los ejercicios.

Con el objetivo de lograr una adecuada comprensión de estos hechos empíricos, se le pide que: i) intente una construcción conceptual que permita interpretar los resultados numéricos. (Sugerencia: a través de la noción de desarrollo social desequilibrado, característico de las formaciones sociales capitalistas, introduzca las peculiaridades del desarrollo capitalista mexicano y el posible aporte de las migraciones en la formación del mercado de trabajo.) ii) Destaque las diferencias entre la idea de asociación y la de nivel y iii) ¿tiene o no sentido, en este caso concreto, aplicar la ecuación de covarianzas de Lazarsfeld?

3. En un trabajo de Américo Saldívar, titulado "Diferencias ideológicas entre obreros y empleados", encontramos la siguiente tabla:

Clase social que debería tener el poder	Empleados		Obreros	
	Estrato social		Estrato social	
	Alto	Bajo	Alto	Bajo
Alta	404	141	178	134
Trabajadora	180	101	81	121
Total	584	242	259	255

Se le pide:

- Realizar una discusión teórica sobre el posible sentido que podrían tener estas tablas. Para ello le sugerimos que intente movilizar las categorías de "conciencia real" y "conciencia atribuida".
- Abordar el análisis de la información en función de la teorización realizada, y
- Presentar ordenadamente las conclusiones a las que haya llegado. Éstas deberían mostrar la articulación entre las categorías conceptuales y la información numérica provista por los cuadros estadísticos. Vale decir, no se trata de hacer una simple descripción de resultados.

4. Silvia Sigal, en su artículo "Crisis y conciencia obrera: la industria azucarera tucumana", se dedica a examinar las consecuencias de la amenaza de desocupación sobre las orientaciones obreras.

La información que proporciona proviene de un conjunto de entrevistas con trabajadores pocos meses después de la intervención de los ingenios en 1966.

Como indicador de la amenaza de desocupación usa la situación del ingenio, ordinalmente tricotomizada en "intervenidos", "amenazados" y "seguros".

Las respuestas a las entrevistas se clasificaron en aquellas que: i) apuntaban hacia la solución de problemas inmediatos; ii) buscaban resolver la crisis con el mantenimiento de la actividad en la industria azucarera y iii) planteaban que los problemas de la provincia debían

resolverse con la diversificación productiva, es decir, con la creación de fuentes opcionales de trabajo.

El cuadro 1 del estudio permite concluir que la propensión hacia la movilidad laboral fuera del azúcar aumenta al pasar de ingenios seguros a intervenidos. Del cuadro 2 se desprende que frente al "problema del azúcar" aparece una clara tendencia hacia la diversificación de la actividad productiva.

De allí la autora concluye que frente a la amenaza de desocupación aumentan las respuestas de cambio individual (movilidad ocupacional) y de cambio global (diversificación económica). Pero entre estos dos tipos de respuesta no hay una transcripción directa puesto que no hay asociación entre aspiraciones de movilidad social y respuestas frente al problema del azúcar.²⁶

La explicación de este vínculo se encuentra en la acción ideológica de la actividad sindical, dirigida siempre a la defensa de la industria como una manera de defender el trabajo y al mismo tiempo abogar por la diversificación como una solución de largo plazo.

A continuación presentamos el cuadro 3 de la investigación de referencia; en él se muestra la distribución de las respuestas a la pregunta ¿le gustaría trabajar en algo que no fuera el azúcar? Para los propósitos de este ejercicio hemos fundido en una sola las categorías de ingenios amenazados e intervenidos y omitimos los casos en que no hubo respuesta.

Participación sindical

	Alta			Baja		
	Seguros	No seguros	Total	Seguros	No seguros	Total
Sí	27	59	86	35	43	78
No	31	27	58	27	28	55
Total	58	86	144	62	71	133

Sobre la base de esta información examine la argumentación presentada.

5. En su estudio de las determinantes sociales de la violencia en Colombia, Paul Oquist establece un vínculo entre la polarización y la continuidad de los partidos políticos colombianos y la presencia de violencia. La objetivación del concepto polarización-continuidad se realizó a través de un índice (denominado "de identificación partidaria") que permitió clasificar los municipios en aquellos cuyo porcentaje prome-

²⁶Este juicio se basa en información empírica.

dio de votos en las elecciones de 1930 y 1946: *i*) favoreció en un 80% o más a un partido; *ii*) presentó una votación en favor de uno de los partidos entre 60 y 79% y *iii*) una categoría que incluye todos los casos restantes. A la primera clase la denominó "hegemonía" en virtud de que corresponde a una situación en que un solo partido domina toda la escena política. A la segunda, "control" porque corresponde a un caso en que hay cierto equilibrio, aun cuando de todas maneras un partido es depositario de la mayoría, y a la tercera le llamó "competencia" ya que corresponde a una situación de equilibrio.

Una vez que establece el vínculo entre identificación partidaria y violencia, introduce como control el partido político que ejerce la dominación y genera los siguientes cuadros:²⁷

Identificación partidaria

	Conservador			Liberal		
	Hegemonía	Control	Total	Hegemonía	Control	Total
Ausencia de violencia	96	81	177	124	76	200
Presencia de violencia	30	23	53	69	58	127
Total	126	104	230	193	134	327

Sobre la base de esta información, ¿qué puede concluir respecto a la violencia en Colombia?

6. De un estudio sobre opiniones de niños mexicanos realizado por Rafael Segovia extrajimos el siguiente párrafo referido a la intervención del Estado en la economía:

Todo esto depende en gran medida de antecedentes familiares, pero también de la escuela, la cual discrimina desde un punto de vista social porque previamente los niños están discriminados desde un punto de vista económico. De ahí que las tendencias observadas estén aún más marcadas (las actitudes se desarrollan antes) entre los escolares de los establecimientos privados, y sean más reticentes frente a la intervención del Estado en la economía (p. 137).

i) Exprese a través del modelo de covarianzas de Lazarsfeld una relación entre tres variables (acorde con el párrafo); *ii*) indique el papel de cada variable en el modelo; *iii*) identifique a cuál de los cuatro tipos propuestos por Lazarsfeld corresponde el análisis y porqué; *iv*) suponga que las tres variables son dicotómicas y proponga para cada una

²⁷Para los propósitos de este ejercicio se omitió la categoría competencia en la identificación partidaria.

las dos categorías de manera que pueda llegar a alguna(s) proposición(es) empírica(s) derivada(s) del párrafo citado.

REFERENCIAS BIBLIOGRÁFICAS DE LOS EJERCICIOS

- Lipset, S. M. *op. cit.*
 Oquist, Paul. *Violencia, conflicto y política en Colombia*, Instituto de Estudios Colombianos, Bogotá, 1978, cap. II.
 Saldívar, Américo. "Diferencias ideológicas entre obreros y empleados", en H. Muñoz, Orlandina de Oliveira y Claudio Stern: *Migración y desigualdad social en la ciudad de México*, El Colegio de México e Instituto de Investigaciones Sociales, UNAM, 1981.
 Segovia, Rafael. *La politización del niño mexicano*, El Colegio de México, segunda edición, 1977.
 Sigal, Silvia. "Crisis y conciencia obrera: la industria azucarera tucumana", en *Revisia Latinoamericana de Sociología 70/1*. Centro de Investigaciones Sociales, Instituto Torcuato di Tella, Buenos Aires, 1970.

4. Asociación en tablas de $r \times c$.

En el capítulo anterior presentamos una extensión del análisis de asociación al considerar más de dos variables dicotómicas. En éste nos dedicaremos a exponer las complejidades que surgen al estudiar la distribución conjunta de variables pluricotómicas. Se trata, entonces, de considerar tablas con r renglones o líneas y c columnas.

TABLA 13

Distribución de frecuencias conjuntas de dos variables pluricotómicas

$Y \setminus X$	B_1	B_2	B_3	...	B_c	Total
A_1	$N(A_1AB_1)$	$N(A_1AB_2)$	$N(A_1AB_3)$...	$N(A_1AB_c)$	$N(A_1)$
A_2	$N(A_2AB_1)$	$N(A_2AB_2)$	$N(A_2AB_3)$...	$N(A_2AB_c)$	$N(A_2)$
A_3	$N(A_3AB_1)$	$N(A_3AB_2)$	$N(A_3AB_3)$...	$N(A_3AB_c)$	$N(A_3)$
.
A_r	$N(A_rAB_1)$	$N(A_rAB_2)$	$N(A_rAB_3)$...	$N(A_rAB_c)$	$N(A_r)$
Total	$N(B_1)$	$N(B_2)$	$N(B_3)$...	$N(B_c)$	N

Las variables X y Y tienen c y r categorías respectivamente. La frecuencia conjunta que corresponde a la casilla genérica formada por A_i y B_j , se simboliza por $N(A_iAB_j)$ y representa el número de casos que presentan simultáneamente ambos atributos. También tenemos dos distribuciones marginales, que nos muestran cómo se distribuyen el total de las observaciones con respecto a la variable X y a la variable Y . La frecuencia del renglón genérico A_i se simboliza por $N(A_i)$ y la que corresponde a la columna B_j por $N(B_j)$; ambas frecuencias denotan el número de casos en cada categoría.

Presentaremos dos grandes líneas de desarrollo. La primera versará sobre aquellas medidas que surgen de la definición de asociación por